STATISTICAL MODELS OF CLAIM AMOUNT DISTRIBUTIONS IN GENERAL INSURANCE

by

YOUSSEF ZIAI, M.Sc., A.I.A.

A Thesis Submitted for the Degree of Doctor of Philosophy

Actuarial Science Division
Department of Mathematics
The City University, London.

April 1979

CONTENTS

ACKNOWLEDGEMENTS

I would like to express my gratitude to Professor Bernard Benjamin for his helpful supervision, constructive criticism and kind encouragement during the preparation of this thesis.
I am additionally grateful to Mr. Stewart Coutts for suggesting some of the problems considered in this thesis and for his rigorous criticism and constructive advice throughout.
The applications of the theoretical aspects of this work are demonstrated on actual data provided by a British insurance company whose co-operation is gratefully acknowledged.
I would also like to thank the library staff of the Institute of Actuaries and of the City University whose sympathetic assistance aided my research.
Finally, I wish to record my appreciation of the work of Miss Lancaster and Miss Williams in typing the manuscript.

ABSTRACT

This work examines the following statistical distributions as possible models for the distribution of claim amounts in general insurance:

1 - The lognormal

2 - The Weibull

3 - The inverse Gaussian
(A new 3-parameter form is introduced)

4 - The Pareto

5 - The truncated lognormal
(as a model for large claim amounts)

6 - The gamma
(as a model for the distribution of the square root of claim amounts)

The properties of the above distributions are investigated and various methods of estimation of their parameters are explored. The method of multinomial maximum likelihood for estimating the parameters is favoured because data on claim amounts is generally in grouped frequency format. To find these estimates a computing technique is proposed which avoids solving a complicated set of non-linear equations. A procedure which avoids solving non-linear equations is also suggested for the least squares estimation of the 3-parameter lognormal, 3-parameter Weibull and the Pareto distribution of the second kind. In order to show how the various methods work in practice they are applied to an actual set of accidental damage claim amounts. Goodness-of-fit tests are used to judge the agreement between the model and sample values. The Chi-square and the Kolmogorov-Smirnov tests are reviewed and a new test statistic is

proposed which measures the overall agreement between the model and sample values in monetary terms. The application procedures for all these tests are described.

Inflation is likely to be the main cause of the increase in the size of claim over time. Therefore, its effects on the parameters of various models are examined. A method is suggested for predicting the future distribution of claim amounts which uses the parameters of a past model after being adjusted for inflation. This predictive method is demonstrated on the accidental damage data whenever a suitable model is found.

CHAPTER 1

INTRODUCTION

1.1 Definition

General insurance[1] is defined as all classes of insurance other than life insurance (or assurance). For a thorough comparison between life and general insurance reference may be made to Benjamin (1977). Fundamentally, general insurance has the following distinguishing features as compared with life assurance:

1 - Claim size not known in advance and often without limit.

2 - Premium changes from year to year because contracts are normally renewed annually.

3 - More than one claim can arise under the same policy.

4 - Volatility; large variances of both claim frequency and claim amount especially the latter.

1.2 Basic Problems of Insurance

In insurance we are faced with the problems of charging adequate premiums to cover a certain risk and the setting up of reserves adequate to meet the cost of future claims with some margin of profit. In life assurance, where the claim amount is known in advance or can be determined by actuarial methods, these problems have been solved by the use of the life table (which provides a model for the probability of survival) in conjunction with discounting functions. In general insurance, however, where the size of claim is most usually not known in advance, a prior estimate of the future cost of claims is essential to the calculation of reserves. This cost is a combination of the frequency of claims and their size. It

1 - Sometimes referred to as Non-life insurance.

is possible to treat each of these components of the cost separately by collecting separate statistics on the frequency and size of claims. In this work, we are concerned with the size of claims only. No claim amount table exists for the calculation of the probability of occurrence of a claim of a certain size.

1.3 Statistical Modelling

It is known that many random factors affect the size of the claim. Statistical modelling is recognized as a rational tool of analysis for problems in all areas of science and engineering where data variation cannot be ignored. Therefore, it commands considerable attention in the solution of the basic insurance problems. Statistical modelling assumes that there is a claim amount distribution underlying the risk process. This distribution, once determined, enables us to calculate the probability that if a claim occurs it will be not greater than a certain size. The shape of the claim amount distribution is important in premium determination and reserves calculations. As Beard (1974) states, "a good theoretically derived model would be of considerable help in dealing with practical estimation problems arising from the random fluctuations which arise in the relatively small samples (in terms of the large claims) which commonly are all that is available". In some classes of general insurance business reinsurance is sought due to the likelihood of occurrence of very large claims. In that case, the examination of the area (the probability) under the upper tail of the distribution is necessary for the insurance company's decision about 'retention[1]' before reinsurance. In motor insurance the policyholder may opt to pay some first part of any own damage claim ('voluntary excess') in return for a reduction in the premium. The appropriate deduction,

1 - That part of the risk which the insurance company wishes to bear without help from the reinsurer.

in this case, may be calculated by examining the lower tail of the claim amounts distribution. The essence of statistical modelling approach is, therefore, to find a statistical distribution as a model for the claim amounts experience of the particular class of insurance in which we are interested. There are two stages involved. At first we have to demonstrate from theoretical considerations of the problem that a specific statistical distribution can model the claim amounts experience. Sometimes it is not readily obvious how the model could be theoretically derived and we have to start immediately from stage two. That is, by fitting our intuitive model to samples of data from past experience and, perhaps, by using statistical goodness-of-fit tests satisfy ourselves that the specific model actually agrees with the claim amounts experience for our particular class of business. Once we have found a model and gathered sufficient knowledge about its parameters and how they behave with respect to time, we will be able to use statistical techniques to predict the distribution of the claim amounts arising during any future period. This is our main objective in this approach.

In practice a solution to our problem, even starting from stage two as mentioned above, is not easily obtainable because of many undesirable factors such as insufficiency of the data, heterogeneity of the data and data being only available in a certain form (e.g. grouped or/and truncated). The empirical distributions of claim amounts are by nature skewed to the right, i.e., there are many small claims and much fewer larger claims. This leads us to the examination of positively skewed statistical distributions as possible models. It is also important to study truncated distributions since in practice, as in reinsurance, data on claims above a certain size only may be available. Some of the models which have been more often employed are referred to below.

1.4 A Review of the Applications of Statistical Distributions to Claim Amounts Data in General Insurance

Studies have been made which involve fitting statistical distributions to claim size data. Beard (1955) has fitted the lognormal to the American fire insurance property damage claim size data. Beard (1957) gives a numerical example of the application of the lognormal and log-Pearson type I distributions to an experience of fire claims in Denmark. In Beard (1964) a lognormal distribution is fitted to a sample of settled motor insurance claims, property damage and liability claims mixed. Benckert (1962) fits lognormal distributions to claims data from fire insurance, accident insurance and motor third party insurance. Harding (1968) uses a truncated lognormal distribution as a model for the original amount of a claim falling under the excess of loss reinsurance of motor business contracts. Ferrara (1971) fits lognormal distributions to fire insurance claim size data from several different industries. Benckert and Jung (1974) fit lognormal and Pareto distributions to data on claims in fire insurance of dwelling houses reported between 1958 and 1969 by Swedish fire insurance companies. Finger (1976) uses the lognormal distribution as a model for claim amounts in liability insurance. Bickerstaff (1972) uses the lognormal as a model for the distribution by size of auto collision claims.

The lognormal seems to be the most successfully used model in general insurance. However, the above references do not deal extensively with the various methods of estimating the parameters of the lognormal distribution, and the efficiencies of these methods, nor do they examine some of the other skewed distributions.

1.5 Objectives and Outline of the Study

It was with the above remarks in mind that the present work was started. We study several skewed distributions as models for claim amounts in general insurance. In order to show the applications of these models in practice we apply their theoretical methods to a set of real data from motor insurance Accidental Damage (AD) claim amounts. A description of the data will be provided in section 1.6.

Goodness-of-fit tests are used frequently in the present work to examine the agreement between a model and sample values. Therefore, in chapter 2 we consider several of these tests. The more widely used Chi-square test is reviewed and a new test statistic is proposed to supplement it. This statistic measures the overall agreement between a model and sample values in monetary terms and, therefore, its value can be easily interpreted. The importance of this statistic is demonstrated in examining the agreement between predicted and actual distributions since, in that case, it indicates by how much we have overpredicted or underpredicted the total cost of claims.

The Kolmogorov-Smirnov test for goodness-of-fit is a well established, but less frequently used, test which is also reviewed in Chapter 2. The application procedures for all three statistics are described and, in later chapters, demonstrated on the AD data.

Because of the importance of the lognormal distribution in general insurance it is extensively studied in Chapter 3. The two and three parameter cases are considered. After defining the distribution its properties are examined and a theoretical justification for the emergence of the lognormal distribution as a model for claim amounts is provided. Tests of lognormality and various estimation methods are suggested for the two parameter case when data only in grouped form are available. These are later demonstrated on the AD data. A computer simulation is carried out to measure the efficiency of different methods of estimation. The multinomial maximum likelihood (MML) method which is most suitable for grouped data is studied and, with the wide availability of computers, a technique is suggested for finding the estimates of the parameters. This iterative procedure maximizes the loglikelihood function directly, via a search for the optimum solution, starting from a set of initial values. The effects of inflation on the parameters are then considered and predictions are made for the distribution of claim amounts, in a future period, by using different indices of wages and prices. The agreement between the actual and predicted claim amounts are tested by the goodness-of-fit tests described in Chapter 2. The 3-parameter case is then studied. A method of estimation which involves the least squares technique and a search for the location parameter is suggested which avoids solving non-linear equations. The method of MML is also modified for the 3-parameter case. These methods are then applied to the AD data. The effects of inflation on the parameters are studied and distributions of claim amounts during future periods are predicted.

In Chapter 4 the Weibull distribution which belongs to the exponential family is studied. The two and three parameter cases are examined, very much on the same lines as for the lognormal distribution. The same methods of estimation as for the 3-parameter lognormal are modified for the 3-parameter Weibull distribution.

Chapter 5 is devoted to the study of the inverse Gaussian (or inverse normal) distribution. This is a skewed distribution with a shape similar to the lognormal, the gamma and the Weibull distributions.

Chhikara and Folks (1978) state that several sets of empirical data which they have investigated seem to be equally well represented by the lognormal and the inverse Gaussian distributions. In the absence of other considerations they recommend the use of the inverse Gaussian distribution on the basis of the convenience of working with it. Therefore, it seemed important to study and test this distribution on the AD claims. Initially, the properties of the distribution are examined and the methods of moments and MML are suggested for the estimation of the parameters. These methods are applied to the AD claims data and the effects of inflation are investigated. We then introduce the 3-parameter version of this distribution by bringing in a threshold parameter. No mention of this case is made in Chhikara and Folks (1978) or in Johnson and Kotz (1970). We suggest using the MML method of estimation. This method is then applied to the data. The effects of inflation on the parameters are studied and predictions for future periods are made and comapred with the actual experience by using goodness-of-fit tests.

Chapter 6 looks at the Pareto distributions of the first and second kinds. The properties and a graphical test are studied. Various estimation methods are examined. In Chapter 7 we use the method of MML to fit this distribution to the upper tail of the AD claims data, i.e., claims greater than £600.

To study the tail of the claim amounts distribution, which is of interest in reinsurance, we deal with the truncated lognormal model in Chapter 7. The method of MML is developed and applied to the truncated samples of AD data. The effects of truncation at different points are studied.

In Chapter 8 we examine the gamma distribution. Beard (1978 - personal communication) suggests the gamma distribution as a model for the distribution of the square root of the claim amount. We would like to test this to see if by taking the square root of the claim amount we can arrive at a better fitting model. For this reason a different set of AD data was obtained where the claims are grouped into different bands according to the square root of their size. The data are better described in Chapter 8. The properties of the gamma distribution are studied and the methods of moments and MML are suggested for the estimation of the parameters. These are then applied to the data. The effects of inflation on the parameters are studied and predictions for future periods are made.

A final discussion and a summary of the findings of the study are presented in Chapter 9.

The tables of results for every chapter are presented at the end of the chapter.

All the computer programs used in this work have been written by the author in Fortran 4 language and have been run via the interactive terminals on the City University's ICL 1905E Computer. The texts of the programs are presented in the Appendix.

1.6 The Accidental Damage (AD) Data

The important part of any data analysis is to have a reliable set of data. Since we are looking for models of the distribution of claim amounts we must be sure that the data used in the analysis has been collected only from the experience of the particular class of business we are investigating. In other words, the data must be free from heterogeneity in every respect. In addition, a considerable amount of data is required and we need to look at the experience over several periods.

A medium-sized, U.K., general insurance company provided us with some data of Accidental Damage[1] (AD), excluding windscreen, claim amounts in respect of claims which occurred during certain periods of time in the past. These periods are referred to as periods of accident. The portfolio from which these claims come was comprehensively insured private cars. Data combined for all age groups, vehicle groups, districts and type of use were available. For claims up to £570 we were given the number of open and settled claims grouped in bands of £30, i.e. £1-30, £31-60, ...,£541-570. For amounts greater than £571 details of the individual claims were made available. After some investigation we grouped these in the following bands:

£571 - £600, £601 - £700,£701 - £800, etc.

until the band containing the maximum claim amount (or amounts) was reached.

Data was provided in respect of seven periods of accident. Each period covers three months of the calendar year. We were given data from the quarter starting on 1.10.1973 to the one ending on 30.6.1975. For convenience sake we refer to the period, say, from 1.10.1973 to 31.12.1973 as '73/4th quarter'. There are, therefore, seven samples of AD claim amounts each corresponding to one of the periods from 73/4th quarter to 75/2nd quarter. In each case the data had been collected at least six months after the end of the period of accident. The incurred but not reported (IBNR) claims are assumed to be so few as not to present a problem. This is because experience (of the insurance company) shows that almost all AD claims are reported and settled within six months after the end of the period of accident. Zero claims[2] were not included in the data.

1 - Damage to the policy holder's own vehicle.
2 - These are those claims in respect of which no payment is made by the insurance company either because no payment is required or because the insurance company recovers the cost from another insurer.

The relevant information for the construction of the AD claim amount samples were extracted from the computer print-outs of the insurance company's claim files. These were then stored in data subfiles on our computer.

Computer program P1 was written to print the frequency distribution of a given sample of AD data. The sample cumulative distribution function and various relevant sample statistics are also calculated and printed out. The program was run with the AD samples and the results are given in tables (1.1) to (1.7). In each table the column headed by 'NCLM > LB. AMOUNT' gives the number of those claims in the sample whose amounts are greater than the amount given by the lower boundary of each interval. From this information we can, for example, see that in 73/4th quarter there were 89 claims with amounts greater than £601. The tables show that the number of such claims has increased over time. The total number of claims in each sample is on average about 2600. For the calculation of the sample moments we have assumed that in each interval all the claims are for an amount equal to the mid-point of that interval. An inspection of the insurance company's claim file showed that this assumption was justified since the average amount of claim in each band was in fact approximately equal to the mid-point of that claim amount interval. For the calculation of the median and mode linear interpolations in the claim amount intervals were used.

From the tables we can see that the mean and the standard deviation of the claim amounts have increased over time while the coefficient of variation has remained quite stable at about 1.1. For each sample, the frequency distribution, the coefficients of skewness and kurtosis as well as the relative positions of the mode, the median and the mean all indicate the skewness and sharp peakedness of the claim amounts distribution. The sample statistics given in tables (1.1) to (1.7) will be of use in later chapters.

1.7 Tables

Table (1.1)

***** 73/4TH QUARTER *****

AMOUNT £	NO. OF CLAIMS	CUM. %	NCLM > LB. AMOUNT
1- 30	478	15.70	3045
31- 60	518	32.71	2567
61- 90	461	47.85	2049
91- 120	359	59.64	1588
121- 150	239	67.49	1229
151- 180	213	74.48	990
181- 210	148	79.34	777
211- 240	102	82.69	629
241- 270	81	85.35	527
271- 300	58	87.26	446
301- 330	66	89.43	388
331- 360	45	90.90	322
361- 390	39	92.18	277
391- 420	35	93.33	238
421- 450	34	94.45	203
451- 480	20	95.11	169
481- 510	29	96.06	149
511- 540	14	96.52	120
541- 570	8	96.78	106
571- 600	9	97.08	98
601- 700	29	98.03	89
701- 800	18	98.62	60
801- 900	20	99.28	42
901-1000	6	99.47	22
1001-1100	4	99.61	16
1101-1200	4	99.74	12
1201-1300	1	99.77	8
1301-1400	3	99.87	7
1401-1500	1	99.90	4
1501-1600	0	99.90	3
1601-1700	1	99.93	3
1701-1800	0	99.93	2
1801-1900	0	99.93	2
1901-2000	1	99.97	2
2001-2100	0	99.97	1
2101-2200	0	99.97	1
2201-2300	0	99.97	1
2301-2400	1	100.00	1

TOTAL NO. OF CLAIMS= 3045
MEAN CLAIM AMOUNT =150.36
STANDARD DEVIATION =175.48
COEFF. OF VARIATION= 1.17
MEDIAN ....= 95.97
MODE ....= 42.87
SQRTB1 =SKEWNESS = 3.50
B2 ... =KURTOSIS = 24.20
G1 ....=-0.107
G2 ....=-0.506

G1 IS COEFF. OF SKEWNESS FOR LOG OF CL. AMOUNTS
G2 IS EXCESS KURTOSIS FOR LOG OF CL. AMOUNTS.

Table (1.2)

***** 74/1ST QUARTER *****

AMOUNT £	NO. OF CLAIMS	CUM. %	NCLM > LB. AMOUNT
1- 30	381	15.61	2441
31- 60	428	33.14	2060
61- 90	351	47.52	1632
91- 120	334	61.20	1281
121- 150	211	69.85	947
151- 180	133	75.30	736
181- 210	98	79.31	603
211- 240	82	82.67	505
241- 270	54	84.88	423
271- 300	52	87.01	369
301- 330	53	89.18	317
331- 360	36	90.66	264
361- 390	29	91.85	228
391- 420	26	92.91	199
421- 450	22	93.81	173
451- 480	22	94.72	151
481- 510	17	95.41	129
511- 540	10	95.82	112
541- 570	19	96.60	102
571- 600	4	96.76	83
601- 700	26	97.83	79
701- 800	21	98.69	53
801- 900	11	99.14	32
901-1000	10	99.55	21
1001-1100	5	99.75	11
1101-1200	2	99.84	6
1201-1300	1	99.88	4
1301-1400	2	99.96	3
1401-1500	0	99.96	1
1501-1600	0	99.96	1
1601-1700	0	99.96	1
1701-1800	1	100.00	1

TOTAL NO. OF CLAIMS= 2441
MEAN CLAIM AMOUNT =149.85
STANDARD DEVIATION =172.20
COEFF. OF VARIATION= 1.15
MEDIAN ....= 95.93
MODE ....= 41.87
SQRTB1 =SKEWNESS = 2.83
B2 ... =KURTOSIS = 14.24
G1 ....=-0.076
G2 ....=-0.503

G1 IS COEFF. OF SKEWNESS FOR LOG OF CL. AMOUNTS
G2 IS EXCESS KURTOSIS FOR LOG OF CL. AMOUNTS.

Table (1.3)

***** 74/2ND QUARTER *****

AMOUNT £	NO. OF CLAIMS	CUM. %	NCLM > LB. AMOUNT
1- 30	351	14.73	2383
31- 60	380	30.68	2032
61- 90	382	46.71	1652
91- 120	295	59.09	1270
121- 150	211	67.94	975
151- 180	142	73.90	764
181- 210	114	78.68	622
211- 240	101	82.92	508
241- 270	57	85.31	407
271- 300	51	87.45	350
301- 330	39	89.09	299
331- 360	36	90.60	260
361- 390	25	91.65	224
391- 420	24	92.66	199
421- 450	27	93.79	175
451- 480	18	94.54	148
481- 510	21	95.43	130
511- 540	17	96.14	109
541- 570	12	96.64	92
571- 600	11	97.10	80
601- 700	30	98.36	69
701- 800	13	98.91	39
801- 900	11	99.37	26
901-1000	4	99.54	15
1001-1100	7	99.83	11
1101-1200	0	99.83	4
1201-1300	1	99.87	4
1301-1400	1	99.92	3
1401-1500	0	99.92	2
1501-1600	1	99.96	2
1601-1700	0	99.96	1
1701-1800	0	99.96	1
1801-1900	1	100.00	1

TOTAL NO. OF CLAIMS= 2383
MEAN CLAIM AMOUNT =151.57
STANDARD DEVIATION =167.98
COEFF. OF VARIATION= 1.11
MEDIAN ....= 98.48
MODE ....= 61.17
SQRTB1 =SKEWNESS = 2.88
B2 ... =KURTOSIS = 15.91
G1 ....=-0.144
G2 ....=-0.464

G1 IS COEFF. OF SKEWNESS FOR LOG OF CL. AMOUNTS
G2 IS EXCESS KURTOSIS FOR LOG OF CL. AMOUNTS.

Table (1.4)

***** 74/3RD QUARTER *****

AMOUNT £	NO. OF CLAIMS	CUM. %	NCLM > LB. AMOUNT
1- 30	362	12.93	2799
31- 60	427	28.19	2437
61- 90	383	41.87	2010
91- 120	356	54.59	1627
121- 150	283	64.70	1271
151- 180	194	71.63	988
181- 210	137	76.53	794
211- 240	97	79.99	657
241- 270	86	83.07	560
271- 300	71	85.60	474
301- 330	64	87.89	403
331- 360	45	89.50	339
361- 390	44	91.07	294
391- 420	25	91.96	250
421- 450	26	92.89	225
451- 480	22	93.68	199
481- 510	25	94.57	177
511- 540	14	95.07	152
541- 570	14	95.57	138
571- 600	17	96.18	124
601- 700	32	97.32	107
701- 800	34	98.54	75
801- 900	17	99.14	41
901-1000	4	99.29	24
1001-1100	9	99.61	20
1101-1200	4	99.75	11
1201-1300	0	99.75	7
1301-1400	1	99.79	7
1401-1500	3	99.89	6
1501-1600	0	99.89	3
1601-1700	1	99.93	3
1701-1800	0	99.93	2
1801-1900	0	99.93	2
1901-2000	0	99.93	2
2001-2100	0	99.93	2
2101-2200	1	99.96	2
2201-2300	0	99.96	1
2301-2400	0	99.96	1
2401-2500	0	99.96	1
2501-2600	1	100.00	1

TOTAL NO. OF CLAIMS= 2799
MEAN CLAIM AMOUNT =166.13
STANDARD DEVIATION =188.70
COEFF. OF VARIATION= 1.14
MEDIAN ....=109.67
MODE ....= 48.39
SQRTB1 =SKEWNESS = 3.42
B2 ... =KURTOSIS = 24.04
G1 ....=-0.179
G2 ....=-0.352

G1 IS COEFF. OF SKEWNESS FOR LOG OF CL. AMOUNTS
G2 IS EXCESS KURTOSIS FOR LOG OF CL. AMOUNTS.

Table (1.5)

***** 74/4TH QUARTER *****

AMOUNT £	NO. OF CLAIMS	CUM. %	NCLM > LB. AMOUNT
1- 30	394	12.86	3064
31- 60	452	27.61	2670
61- 90	426	41.51	2218
91- 120	348	52.87	1792
121- 150	272	61.75	1444
151- 180	219	68.90	1172
181- 210	154	73.92	953
211- 240	124	77.97	799
241- 270	105	81.40	675
271- 300	78	83.94	570
301- 330	75	86.39	492
331- 360	58	88.28	417
361- 390	55	90.08	359
391- 420	29	91.02	304
421- 450	43	92.43	275
451- 480	24	93.21	232
481- 510	22	93.93	208
511- 540	24	94.71	186
541- 570	19	95.33	162
571- 600	14	95.79	143
601- 700	42	97.16	129
701- 800	28	98.07	87
801- 900	20	98.73	59
901-1000	17	99.28	39
1001-1100	8	99.54	22
1101-1200	5	99.71	14
1201-1300	1	99.74	9
1301-1400	5	99.90	8
1401-1500	0	99.90	3
1501-1600	1	99.93	3
1601-1700	0	99.93	2
1701-1800	0	99.93	2
1801-1900	0	99.93	2
1901-2000	0	99.93	2
2001-2100	1	99.97	2
2101-2200	0	99.97	1
2201-2300	0	99.97	1
2301-2400	1	100.00	1

TOTAL NO. OF CLAIMS= 3064
MEAN CLAIM AMOUNT =174.19
STANDARD DEVIATION =194.03
COEFF. OF VARIATION= 1.11
MEDIAN ....=112.91
MODE ....= 51.21
SQRTB1 =SKEWNESS = 2.97
B2 ... =KURTOSIS = 17.65
G1 ....=-0.199
G2 ....=-0.428

G1 IS COEFF. OF SKEWNESS FOR LOG OF CL. AMOUNTS
G2 IS EXCESS KURTOSIS FOR LOG OF CL. AMOUNTS.

***** 75/1ST QUARTER *****

AMOUNT £	NO. OF CLAIMS	CUM. %	NCLM > LB. AMOUNT
1- 30	324	12.43	2607
31- 60	387	27.27	2283
61- 90	345	40.51	1896
91- 120	289	51.59	1551
121- 150	253	61.30	1262
151- 180	187	68.47	1009
181- 210	138	73.76	822
211- 240	114	78.14	684
241- 270	93	81.70	570
271- 300	67	84.27	477
301- 330	63	86.69	410
331- 360	44	88.38	347
361- 390	44	90.07	303
391- 420	35	91.41	259
421- 450	25	92.37	224
451- 480	26	93.36	199
481- 510	18	94.05	173
511- 540	18	94.74	155
541- 570	22	95.59	137
571- 600	17	96.24	115
601- 700	39	97.74	98
701- 800	19	98.47	59
801- 900	18	99.16	40
901-1000	12	99.62	22
1001-1100	3	99.73	10
1101-1200	1	99.77	7
1201-1300	1	99.81	6
1301-1400	0	99.81	5
1401-1500	1	99.85	5
1501-1600	0	99.85	4
1601-1700	1	99.88	4
1701-1800	0	99.88	3
1801-1900	0	99.88	3
1901-2000	0	99.88	3
2001-2100	1	99.92	3
2101-2200	1	99.96	2
2201-2300	0	99.96	1
2301-2400	0	99.96	1
2401-2500	0	99.96	1
2501-2600	0	99.96	1
2601-2700	0	99.96	1
2701-2800	0	99.96	1
2801-2900	0	99.96	1
2901-3000	0	99.96	1
3001-3100	0	99.96	1
3101-3200	0	99.96	1
3201-3300	0	99.96	1
3301-3400	0	99.96	1
3401-3500	0	99.96	1
3501-3600	1	100.00	1

TOTAL NO. OF CLAIMS= 2607
MEAN CLAIM AMOUNT =173.40
STANDARD DEVIATION =195.30
COEFF. OF VARIATION= 1.13
MEDIAN ....=116.19
MODE ....= 48.50
SQRTB1 =SKEWNESS = 4.39
B2 ... =KURTOSIS = 48.47
G1 ....=-0.237
G2 ....=-0.394

Table (1.7)

***** 75/2ND QUARTER *****

AMOUNT £	NO. OF CLAIMS	CUM. %	NCLM > LB. AMOUNT
1- 30	302	12.10	2495
31- 60	374	27.09	2193
61- 90	332	40.40	1819
91- 120	277	51.50	1487
121- 150	235	60.92	1210
151- 180	187	68.42	975
181- 210	122	73.31	788
211- 240	110	77.72	666
241- 270	80	80.92	556
271- 300	72	83.81	476
301- 330	47	85.69	404
331- 360	39	87.25	357
361- 390	40	88.86	318
391- 420	38	90.38	278
421- 450	29	91.54	240
451- 480	21	92.38	211
481- 510	30	93.59	190
511- 540	19	94.35	160
541- 570	17	95.03	141
571- 600	11	95.47	124
601- 700	36	96.91	113
701- 800	22	97.80	77
801- 900	22	98.68	55
901-1000	11	99.12	33
1001-1100	4	99.28	22
1101-1200	3	99.40	18
1201-1300	3	99.52	15
1301-1400	6	99.76	12
1401-1500	2	99.84	6
1501-1600	1	99.88	4
1601-1700	0	99.88	3
1701-1800	1	99.92	3
1801-1900	0	99.92	2
1901-2000	1	99.96	2
2001-2100	0	99.96	1
2101-2200	0	99.96	1
2201-2300	1	100.00	1

TOTAL NO. OF CLAIMS= 2495
MEAN CLAIM AMOUNT =180.08
STANDARD DEVIATION =204.79
COEFF. OF VARIATION= 1.14
MEDIAN ....=116.44
MODE ....= 49.45
SQRTB1 =SKEWNESS = 3.07
B2 ... =KURTOSIS = 17.70
G1 ....=-0.184
G2 ....=-0.376

G1 IS COEFF. OF SKEWNESS FOR LOG OF CL. AMOUNTS
G2 IS EXCESS KURTOSIS FOR LOG OF CL. AMOUNTS.

CHAPTER 2

Tests for Goodness-of-Fit

2.1 Introduction

The natural order of any statistical analysis involving the fitting of a theoretical distribution to a set of sample values is to fit the theoretical model first and then to test its agreement with the observed distribution of the sample values. This problem of testing "the goodness-of-fit" (i.e. the adherence of the model to the data) will arise at various stages of the present work when we consider different statistical distributions as models for the distribution of claim amounts. Before considering the fitting methods for various distributions it is deemed convenient to devote one chapter to the study of the theoretical bases and application procedures of some goodness-of-fit tests.

Goodness-of-fit tests are performed to examine the agreement between the theoretical distribution of a random variable and its empirical distribution represented by a set of sample values. In other words, if $x_1,x_2,\ldots,x_n$ are independent observations of a random variable X (for instance the claim amount) with an unknown distribution function F(x), then we are required to test the null hypothesis that

$$H_o : F(x) = F_o(x)$$

where $F_o(x)$ is some particular distribution function. Any test of H_o is called a test of fit. Hypotheses of fit, H_o, may be classified as simple or composite. H_o is a simple hypothesis if it specifies the values of all the parameters of $F_o(x)$. If the values of none of the parameters or only of some of them are specified by the null hypothesis then H_o is called a composite hypothesis.

The reason for formulating the null hypothesis of a test of fit in terms of the distribution function is that the parametric hypothesis testing methods do not provide the means of testing whether observations come from a particular distribution with unspecified parameters. In addition, by our intuition we expect that the distribution of sample observations would closely approximate the true distribution (Kendall and Stuart (1973)). It is in this sense that a goodness of fit test is a measure of the discrepancy between the sample and theoretical distribution functions. Savage (1953) characterizes a goodness of fit test by the following four properties:

1 - It is defined for samples from some large class of distributions.

2 - The null hypothesis is either some specified distribution or a class of distributions of which the functional form is known.

3 - For all null hypotheses, the test statistic used has the same distribution (at least asymptotically).

4 - The test is consistent[1].

In our work a test of goodness of fit will be required in two circumstances:

(i) Fitting and Testing

Let us assume that we have a sample of data representing the observed distribution of the claim amounts. We can postulate the form of the population distribution of this variable and use one of the appropriate estimation techniques, for that particular distribution, to estimate its parameters and thus specify it completely. We will

---

1- A test of hypothesis H_0 against a class of alternatives H_1 is said to be consistent if, when any member of H_1 holds, the probability of rejecting H_0 tends to 1 as sample size(s) tends to infinity.

then need to test how well the theoretical and observed distributions agree. A good agreement will be taken as evidence that the assumed family of distributions is the correct form for the distribution of claim amounts. The null hypothesis in this case is of the composite type.

(ii) Prediction and Testing

We may predict a theoretical distribution, whose form and parameters are completely specified, as the distribution of claim amounts in a particular period of accident occurrence. If a set of data representing the observed distribution for the same period already exists, we will be able to test how well the predicted and actual distributions agree. In this situation the null hypothesis of test is of a simple type. Evidence of a good fit can be used to recommend a prediction technique and support the assumption about the theoretical model.

In this chapter several goodness of fit tests will be studied. In Section 2.2 the Chi-square goodness of fit test will be dealt with in detail. To supplement the Chi-square test, and to avoid some of its shortcomings, a test statistic based on the weighted sum of the actual minus expected number of claims, in different intervals, will be proposed in section 2.4. The Kolmogorov-Smirnov test of goodness of fit, which is generally believed to be more powerful than the Chi-square test, will be studied in detail in section 2.6. Two other test statistics will be mentioned in section 2.9 but because they are not applicable to grouped data we will not study them in detail.

2.2 The Chi-square Goodness-of-Fit Test

This was the first goodness of fit test and it was introduced by Karl Pearson in 1900.

Let us assume that we have a sample of n independent observations of a random variable X, with distribution function F(x). Pearson's test involves grouping the observations into, say, k mutually exclusive categories such that n_i be the number of observations in category i and $n = \sum_{i=1}^{k} n_i$.

If the population distribution is completely specified by the null hypothesis as $F(x) = F_o(x)$ (i.e. the form and all the parameters of $F_o(x)$ are known) and we assume that H_o is true, then the probability, p_{oi}, of a random observation falling into any category i can be calculated. If p_{oi} is multiplied by the total number of observations, n, we will find the expected frequency of class i, say n_{oi}, under the hypothesis H_o. Apart from sampling variation there should be close agreement between n_i and n_{oi}, the observed and expected frequencies respectively. The Chi-square goodness of fit test provides a probability basis for computing and deciding whether the discrepancy is too large to have occurred by chance. The test statistic, proposed by Pearson, is:

$$X^2 = \sum_{i=1}^{k} \frac{(n_i - n_{oi})^2}{n_{oi}}$$

Large values of X^2 indicate an overall lack of agreement between the observed and expected distributions. The null hypothesis which resulted in n_{oi}'s should, therefore, be rejected for large values of X^2.

If when H_o is assumed true the sampling distribution of a test statistic is known and tractable, tables of its percentage points can be constructed. The sampling distribution of X^2 is very complicated when the sample size is finite, but Cramer (1946) has shown that its limiting distribution, under H_o, is approximately Chi-square, χ^2, with k-1 degrees of freedom (we assumed $F_o(x)$

as completely specified).

In some situations the null hypothesis is composite such that usually only the form of the postulated distribution is known but not any, or some, of its parameters. If these parameters have to be estimated from the sample data, then the limiting distribution of X^2 may depend on the method of estimation. With a poor method of estimation X^2 may frequently have a large value even if the theory is correct (Cochran (1952)). In a general proof of the distribution of X^2 the method of estimation must be asserted. A method that yields those values of the parameters which minimize X^2 (minimum Chi-square method) may seem the most suitable. Fisher (1924) has shown that in the limit in large samples this method becomes equivalent to the method of multinomial maximum likelihood. The following theorem due to R.A. Fisher, states the principal theoretical result for the distribution of the Chi-square test statistic when parameters of the distribution of X under the null hypothesis, $F_o(x)$, are estimated from the sample data.

Theorem

> If $F_o(x)$, whose form is known, has r unknown parameters, and if the corresponding multinomial maximum likelihood estimates are substituted for the unknown parameters, then X^2 is distributed, in the limit, as the Chi-square, χ^2, distribution with k - r - 1 degrees of freedom.
>
> (A proof of this can be found in Cramer (1946)).

Cochran (1952) states that any efficient method of estimation gives estimates which in the limit become identical with the maximum likelihood estimates. Thus, the Chi-square distribution with the appropriate reduction in the degrees of freedom is valid, as the distribution of X^2 statistic, for any efficient method of estimation. If fully efficient ordinary maximum likelihood estimators are used, then X^2 does not have an asymptotic Chi-square distribution. There will be a partial recovery of the r degrees of freedom lost by the multinomial maximum likelihood estimators and the distribution of the Chi-square statistics, X^2, will be bounded between a χ^2_{k-1} and a χ^2_{k-r-1} variable. Therefore the critical values should be adjusted upwards. As k becomes large χ^2_{k-1} and χ^2_{k-r-1} become so close together that the difference can be ignored (Kendall and Stuart (1973)).

If individual observations are available in the sample, and the null hypothesis is that they follow a continuous density function, then the investigator must first group the observations into different mutually exclusive classes. Cochran (1952) mentions that the investigator has the choice of both the number of classes and the division points between them, but that his choice will affect the sensitivity of the test. According to Kendall and Stuart (1973) the whole asymptotic theory of the Chi-square test is valid as long as the k classes into which the observations are grouped are determined without reference to the observations because there has been no provision in the theory for the class boundaries themselves being random variables. A rule suggested by Mann and Wald (1942) and by Gumbel (1943) is to choose the classes so that the expected frequencies are all equal to n/k where k, the number of classes, is assumed given. Mann and Wald (1942) have developed a technique for finding the optimum number of classes for any sample size n such that the power of the test is never less than ½. The problems of the choice of the number of classes and class boundaries do not arise when only a sample of already grouped observations is available. Such is the case with our Accidental Damage data and the above problems are therefore pursued no further here.

In the derivation of the sampling distribution of the Chi-square test statistic, a multinomial distribution is approximated by a multinormal distribution (see Cramer (1946) or Kendall and Stuart (1973)). When the number of classes is large, and the expected frequencies are small, this approximation may not be satisfactory. It has been suggested by some authors that the expected frequency in any interval should not be less than 5. For any expected frequency less than 5, the usual procedure is to pool the adjacent classes together until this condition is removed. The number of degrees of freedom should then be calculated on the basis of the number of classes actually used, after pooling together, in the calculation of X^2. Since the discrepancy between an observed and a postulated distribution is often most apparent in the tails, the sensitivity of the Chi-square test is likely to be decreased by excessive pooling at the tails (Cochran (1952)). Consequently the rule of minimum expected frequency of 5 should not be considered as inflexible. Cochran (1942) has shown that there is little disturbance at the 5% level when a single expected frequency is as low as ½. At 1% level the same is true if the degrees of freedom of the Chi-square distribution are greater than 6. He states that two expected frequencies as low as 1 may be allowed with negligible disturbance to the 5% level.

The Chi-square goodness-of-fit test is commonly used when there is no clear alternative hypothesis. This necessarily precludes the computation of power. If there exists an alternative hypothesis, then the distribution of X^2 under the alternative hypothesis is asymptotically a "non-central χ^2" with k-r-1 degrees of freedom

(where r parameters have been estimated by the multinomial maximum likelihood estimators) and a non-centrality parameter which depends on the sample size. For the method of calculation of the limiting power function of the test see Kendall and Stuart (1973).
The dependence of the power of the Chi-square test on the sample size is a weakness of this test, since with a small sample, an alternative hypothesis which has a large departure from the null hypothesis, H_0, may have a small probability of yielding a significant value of the test statistic. On the other hand, for a large sample, rather small and unimportant departures from the null hypothesis are likely to yield a significant value of the test statistic.

*There are two major shortcomings of the Chi-square test. When testing the goodness-of-fit of a continuous distribution, grouping the observations into classes necessarily implies the loss of information by such grouping. Another shortcoming is that the X^2 test statistic is based on the squares of the deviations between the actual and expected frequencies. This implies that the Chi-square test will not be sensitive to the pattern of signs of the deviations.

*The Chi-square goodness-of-fit test is applicable to situations in which the alternative hypotheses are expressed in vague and general terms. Its main advantage is that when the hypothesized distribution is not completely specified the test can still be performed, in the same way as for a simple null hypothesis, simply by replacing the unknown parameters with efficient estimators and reducing the degrees of freedom by the number of estimated parameters.

It was mentioned that the Chi-square goodness-of-fit test ignores the signs of deviations of the observed frequencies from the expected ones. It is, therefore, sometimes informative to examine the pattern of the signs of deviations. David (1947) has shown that if the null hypothesis is simple then all patterns of signs are equiprobable.

Therefore, when H_0 is simple, if the hypothetical distribution is the true population distribution, we expect the signs of deviations to have a random pattern rather than form a few clusters of deviations of the same signs. To test the randomness pattern of the signs, it is possible to use the "runs"[1] test (see, for instance, Bury (1975)). A simpler alternative method is to compare the number of changes of signs of deviations with the number of non-changes[2]. These two numbers should be approximately equal, provided that a large number of deviations exists and the pattern of signs is random (Benjamin and Haycocks(1970)).

For a composite null hypothesis, where all the parameters have to be estimated from the sample, Fraser (1950) has shown that all patterns of signs of deviations are not equiprobable. Therefore, the "runs" test for randomness or the simple comparison of the number of changes and non-changes of signs cannot be applied in the case of a composite null hypothesis.

2.3 Application Procedure for the Chi-square Test

The frequency distributions of our AD claim amounts samples, which were presented in Chapter 1, are skewed to the right and in grouped form. In this work, the Chi-square goodness-of-fit tests will be performed in situations (i) and (ii) mentioned in section 2.1. If estimates of the parameters of a hypothesized distribution are required, they will be calculated from the grouped data, as given, without any amalgamation of the intervals. Where necessary, computer programs will be written to provide result tables, giving the actual, A, and the expected, E, frequencies as well as deviations, A-E, for

1- A "run", in this case, is defined as a sequence of consecutive plus signs or minus signs.

2- In the sequence +++--+ there are two changes and three non-changes of signs.

every interval. The value of $\frac{(A-E)^2}{E}$ for any interval in which the expected frequency is greater than or equal to 5, as well as the total of such values are also printed. For other intervals, i.e. where $E < 5$, enough intervals will have to be pooled together to remove this condition. The values of $\frac{(A-E)^2}{E}$ for the pooled intervals will then have to be calculated, on an electronic desk calculator, and added to the total already given in the table. The number of degrees of freedom is calculated on the basis of the number of intervals actually used in the calculation of the final value of X^2. Considering that the number of intervals used in the calculation of X^2 is always more than 20, the value of X^2 is expected to be large. Because of the skewness of the data the major contributions to the value of X^2 are from the intervals in the lower tail of the distribution. The number of claims in the upper tail of the distribution, for instance, claims of amounts greater than £1200, is very small. Therefore the contributions to the total X^2 from, at most one or two, pooled intervals in the upper tail will be rather small, when compared with the value of X^2, and may be ignored in many instances. This means that the total of $\frac{(A-E)^2}{E}$ values are calculated by the computer, for the intervals with an expected frequency of greater than or equal to 5, may be safely taken as the value of the Chi-square statistic. However, for the calculation of the degrees of freedom, the number of pooled intervals will be taken into account. The above procedure saves us many unnecessary calculations on the desk calculator.

Let us now assume that X^2 is the value of our test statistic and v is its relevant number of degrees of freedom. We look up a table of cumulative percentage points of the Chi-square distribution, with v degrees of freedom, and note the probability, P, that a χ^2_v random variable will exceed the value of our calculated X^2. In other words we find P such that

$$P = \Pr(\chi^2_v \geq X^2) .$$

We classify the result of the test, according to different values of P, as follows:

If (i) $P \geq 0.05$

or (ii) $0.01 \leq P < 0.05$

or (iii) $0.001 \leq P < 0.01$

or (iv) $P < 0.001$

then the difference between the observed and postulated distributions, under the null hypothesis, based on the given sample is respectively:

(i) Not Significant

(ii) Almost Significant

(iii) Significant

(iv) Highly Significant

The above arbitrary classification is based on a 5% significance level. A different level of significance, say 10%, may be used if stronger confidence is required from the test.

2.4 The Total Expected Loss Statistic, T

It is not just the signs of the deviations that are considered important but their magnitude needs some attention too. It is generally expected that for a good fit the magnitude of the deviations should be small. For a random variable such as the claim amount whose distribution is skewed to the right, the frequencies in the intervals of the lower tail of the sample histogram will be much greater than those in the upper tail intervals. Therefore, the magnitude of the deviations in the lower tail intervals will be generally greater than those in the upper tail. On the other hand, the claim amounts are larger in the upper tail than in the lower. Therefore, it is not very informative merely to look at the magnitude of the deviations. The above can perhaps be explained better by providing an example. Suppose that the (Actual-Expected) frequency in the £1-£30 interval is equal to +100. In money terms this difference is on average equal to

$$100 \times £15.5 = £1550$$

where £15.5 is the average claim amount in that interval. This is, in monetary terms, equivalent to a deviation of only 1 in the £1501 - £1600 interval. We, therefore, suggest looking at the weighted deviation for each interval where the weight is the average amount of claim in the interval (the examination of our data showed that the mid-point of each interval is approximately equal to the average claim amount in that interval). We can also define the test statistic T as the sum of the weighted deviations, i.e.

$$T = \sum_{i=1}^{k} f_i(n_i - n_{oi})$$

where f_i is the mid-point of interval i,

n_i is the actual frequency in interval i,

n_{oi} is the expected frequency in interval i,

and k is the number of intervals.

This statistic is a measure of overall agreement, in monetary terms, between a hypothesized distribution and the actual sample values. For a good fit we expect the value of T to be small. Every component of T, such as

$$f_i(n_i - n_{oi})$$

indicates, in money terms, a difference (or a loss) in interval i which we should expect to find between the observed distribution of claim amounts and its hypothesized distribution under the null hypothesis. Hence we call T the total expected loss statistic. As a better indicator of the difference between the observed and the postulated distributions we can look at the ratio of T to the total actual cost of claims (i.e. $\sum_{i=1}^{k} f_i n_i$) for the sample. For a good fit this ratio, expressed as a percentage, should be small. To illustrate this point, suppose that the total actual cost of claims for a sample of data, representing a particular period of accidents, is equal to £1,000,000. If the above defined ratio is equal to 2%, say, then the loss we would be incurring by adopting the hypothesized distribution under H_0 as the true distribution of claim amounts would be equal to £20,000. It would then have to be decided, on the basis of the situation at hand, whether such a discrepancy can be allowed. When testing the goodness-of-fit of a predicted model to actual data, T indicates, in monetary terms, how far from reality our model is predicting. A set of values of T calculated from samples of actual data collected in the past and their predicted models can indicate the reliability and consistency of a prediction technique. This knowledge will be valuable when setting up reserves to meet the cost of future claims as calculated from a predicted distribution of claim amounts.

We have given some consideration to finding the sampling distribution of this statistic. We may use the method of derivation of the asymptotic distribution of the Chi-square test statistic and argue as follows: If the null hypothesis is simple, so that all the parameters of the postulated distribution are known, then the probability p_{oi} that an observation (claim amount) will fall into interval i can be calculated on the assumption that H_o is true. In the derivation of the asymptotic distribution of the Chi-square statistic it is shown that the quantities $(n_i-np_{oi})/\sqrt{np_{oi}}$ are approximately unit normal variates, and the χ^2 distribution emerges as the sum of the squares of these quantities (see Kendall and Stuart (1973)). If the null hypothesis is composite so that, say, r parameters have to be estimated from the sample, then p_{oi} will be a function of the r unknown parameters. In the proof of the asymptotic distribution of the X^2 statistic it is again shown that when the r unknown parameters are estimated as the solution to a set of r homogeneous linear equations in the n_i (for instance, estimation by maximum likelihood method), then the quantities $(n_i-np_{oi})/\sqrt{np_{oi}}$ will be unit normal variates (see Kendall and Stuart (1973)). Therefore, for the purpose of finding the sampling distribution of T, if the above condition in the case of estimation of the parameters holds, we can take it that in both cases of simple and composite hypotheses the quantities

$$(n_i-n_{oi}) \quad \text{where } n_{oi} = np_{oi}$$

are normally distributed with mean zero and variance n_{oi}, i.e. $N(0,n_{oi})$.

Therefore, $f_i(n_i-n_{oi})$ are independent $N(0,f_i^2n_{oi})$

for $i = 1,2,\ldots\ldots,k$

and

$$T = \sum_{i=1}^{k} f_i(n_i-n_{oi}) \text{ is distributed as } N(0,\sum_{i=1}^{k} f_i^2n_{oi})$$

It is thus possible to compare the standardized value of T with the critical values of the standard normal distribution. To justify the multinormal approximation to the multinomial distribution, as in the case of the X^2 statistic, we believe that the expected frequency in each interval should be greater than or equal to 5. If this is not so, then enough intervals will need to be pooled together to remove this condition.

2.5 Application Procedure for the T Statistic

In this work we shall use the statistic T in addition to the formal Chi-square goodness-of-fit test. We are interested in T as a measure of the discrepancy, in monetary terms, between the observed and the postulated distributions. We shall not, therefore, concern ourselves with its sampling distribution or making comparisons of its standardized value with tables of standard normal cumulative percentage points. Hence intervals will not be pooled in the calculation of T. When fitting a distribution to a sample or comparing a predicted distribution with an observed one, the computer programs will produce a table of results. "Expected Loss", as defined in 2.4, will be calculated for each interval and printed in the table. The total expected loss, T, will also be produced. The ratio of the total expected loss to the total actual cost of claims, calculated as mentioned earlier in 2.4, will be shown at the bottom of the table as a percentage.

2.6 * The Kolmogorov-Smirnov Goodness-of-Fit Test

Two criticisms of the Chi-square goodness-of-fit test when used for continuous distributions were the necessity for grouping the individual observations and the adoption of large intervals for small sample sizes. Both of these procedures result in loss of information. Besides, when there are k intervals, the X^2 test statistic is based on k comparisons between the observed and expected class frequencies, while there are n observations in the sample. Therefore, in such circumstances, it is preferable to have available test statistics based on individual observations. Several goodness-of-fit test statistics exist which are based on the individual sample observations and are functions of the deviations between the observed cumulative distribution of the sample (the empirical distribution function) and the cumulative distribution function under the null hypothesis. Let us first define the empirical distribution function:

For a sample of n random observations $x_1, x_2, \ldots, x_n$ we define the empirical distribution function $S_n(x)$ as

$$S_n(x) = \begin{cases} 0 & x < x_{(1)} \\ r/n & x_{(r)} \leqslant x < x_{(r+1)} \\ 1 & x_{(n)} \leqslant x \end{cases}$$

The $x_{(r)}$ are the order statistics of the sample. Hence $S_n(x)$ is simply a step function which gives the proportion of the observations less than or equal to x.

The best known statistic of the above form is the Kolmogorov-Smirnov test statistic. This goodness-of-fit test was first proposed by Kolmogorov in 1933 and then developed by Smirnov in 1939. If $F_0(x)$ is assumed to be a continuous and completely specified population distribution function under the null hypothesis and $S_n(x)$ to be the step function of the sample, then the test makes use of the statistic

$$D = \max_x |S_n(x) - F_o(x)|$$

It is expected that $S_n(x)$, for a random sample of n independent observations, is fairly close to the specified distribution function. If it is not close enough, then the distribution under the null hypothesis is not the correct population distribution.

The maximum deviation D is a random variable whose sampling distribution is known and is independent of $F_o(x)$, when the null hypothesis holds, provided that $F_o(x)$ is continuous (Massey (1951)). Therefore, D is a distribution free statistic. Its limiting distribution was derived by Kolmogorov himself. Smirnov (1948) gave a tabulation of the limiting distribution of D. Massey (1950-a) provided the method for evaluating the distribution of D for small samples. Tables for determining the significance of D in finite samples were given by Birnbaum (1952). A table of the critical values of the test statistic D at different significance levels for sample sizes n = 1 to 20, n = 25, 30, 35 and n > 35 was given by Massey (1951). For the sake of convenience, the critical values of D for large sample sizes (n > 35) are reported, from his paper, in table (2.1) at the end of this chapter.

When data is only available in grouped form, it is possible to calculate the deviations $|S_n(x_i) - F_o(x_i)|$ at each point x_i where x_i is the upper boundary of interval i. Massey (1951) states that grouping the observations into intervals tends to lower the value of D, and he asserts that for grouped data the appropriate significance levels are smaller than those given in his table. For large samples, however, grouping causes little change in the appropriate significance levels. If the number of categories is small, then important changes can be expected in the significance levels for any sample size. According to Massey (1951), the Kolmogorov-Smirnov Statistic, D, is correctly used only if the distribution $F_0(x)$ is continuous and completely specified as regards form and all its parameters. The distribution of the maximum deviation, D, is not known when certain parameters of the distribution have to be estimated from the sample values. When we estimate the parameters of the population distribution from the data, we are in effect adjusting these parameters according to the sample values, and in consequence we should be making a closer fit of the hypothesized distribution to the sample values. Hence we expect that at the same significance level the critical value of D will be smaller than when $F_0(x)$ is completely specified. Therefore, in these circumstances, if the maximum absolute deviation exceeds the critical value $D_\alpha(n)$, corresponding to a significance level α and read from an appropriate table of the critical values of D (for large sample sizes, n, see table (2.1)), then we can safely reject the null hypothesis and conclude that the population distribution is not $F_0(x)$.

The distribution of the Kolmogorov-Smirnov test statistic when the parameters of $F_0(x)$ are estimated from the sample values depends on the form of $F_0(x)$ and is very difficult to find analytically. Monte Carlo techniques can be used to calculate the approximate distribution function of this test statistic for each particular family of distributions (say, normal) $F_0(x)$ under the null hypothesis. Lilliefors (1967) gives a table, based on Monte Carlo calculations, for use with the Kolmogorov-Smirnov statistic when testing whether a set of observations is from a normal population whose mean and variance are not specified but must be estimated from the sample. He suggests using the sample mean and variance (with denominator n-1) as estimates of the mean and variance of the normal population to specify $F_0(x)$. For large sample sizes ($n > 40$) the critical values of D at various significance levels are reproduced, from his paper, in table (2.2) at the end of this chapter.

Lilliefors (1969) gives a similar table to be used when testing whether a set of observations is from an exponential population with unspecified mean. He suggests using the sample mean as the mean of the exponential population.

2.7 Comparison Between the Chi-square and the Kolmogorov-Smirnov Goodness-of-Fit Tests

Massey (1951) argues that the Kolmogorov-Smirnov test may be always more powerful than the Chi-square test. He also points out that the K-S test, at least at the 50 per cent power level, will detect smaller deviations between the observed and hypothesized distributions than will the Chi-square test. Not enough is known about the power of either test to justify the preference for using X^2 or D for testing a completely specified hypothesis (Birnbaum (1952)). However, Massey (1950-b) has established a lower bound to the power of the K-S test in large samples.

We recall that two criticisms of the Chi-square test were the grouping of the observations when individual observations were available and the adoption of large intervals for small samples. Both of these procedures result in loss of information. The Kolmogorov-Smirnov test, however, uses individual observations and hence may utilize information more completely than the Chi-square test.
For very small samples the Chi-square test is not applicable at all

because its sampling distribution is not distribution free for finite sample sizes and is not known. The K-S test, however, may be used for very small samples.

The major shortcoming of the K-S test is that when the parameters of the postulated distribution must be estimated from the sample values the test is not applicable because the sampling distribution of D is not distribution free and is not known. In such circumstances the limiting distribution of the Chi-square is easily modified by reducing the degrees of freedom.

2.8 Application Procedure For K-S Test

In our work, we can apply the Kolmogorov-Smirnov test to examine the goodness-of-fit of a predicted distribution of claim amounts to actual data (i.e. in situation classified under (ii) in Section 2.1). In such cases the null hypothesis is of the simple form and we can use critical values of D given, for large n, in table (2.1). However, in situations classified under (i) in section 2.1, when we fit a distribution with unspecified parameters to a sample of actual data, the null hypothesis will be composite and, as mentioned earlier, we cannot in general use the K-S goodness-of-fit test because tables of the critical values of D do not exist in these circumstances. There is, however, an exception in the case of the lognormal distribution.

We say a random variable X is distributed lognormally if and only if Y = log X is distributed normally (see Chapter 3). Y = log X is a one-to-one function and hence we can use a test of normality for Y as a test of lognormality for X. Therefore, for the lognormal distribution we may use the Kolmogorov-Smirnov test statistic along with table (2.2) of its critical values for large sample sizes. As we mentioned earlier, this table has been produced by Lilliefors (1967) by using Monte Carlo techniques, and it is for testing the goodness-of-fit of a normal distribution with unknown mean and variance.

We mentioned earlier that if parameters are estimated from the sample values then the critical values of the K-S statistic would be smaller than those in the standard tables (of Massey (1951), for instance). This provides us with a means of safely rejecting a postulated distribution when its parameters have been estimated from the sample. For this purpose we need only to check that the absolute maximum deviation, D, exceeds the critical value $D_\alpha(n)$, given in table (2.1) for large n, to conclude that the hypothesized distribution should be rejected at significance level α. If D does not exceed $D_\alpha(n)$ then we cannot decide whether the null hypothesis should be rejected or accepted.

Our accidental damage data are in grouped form. Therefore, we explain the method of calculating the Kolmogorov-Smirnov test statistic, D, for this type of data. Let us assume that n observations have been grouped into k intervals such that x_i and n_i are respectively the upper boundary and the observed frequency of interval i. Suppose that under some null hypothesis the expected frequency for each interval has been obtained and that for interval i it is equal to n_{oi}. The above is all the information we need to calculate the value of the K-S test statistic, D, without resorting to the calculation of the empirical and theoretical distribution functions. This is because we can easily show that

$$D = \max_{x_i} \left| S_n(x_i) - F_o(x_i) \right| = \frac{1}{n} \max_{i} \left| \sum_{j \leq i} (n_j - n_{oj}) \right|$$

where i = 1,2,....,k.

Therefore, if in a table which gives the above information, a column for (Actual-Expected) frequencies exists, we can rapidly calculate the D test statistic. For this purpose we need to add the successive values in this column, starting from the first interval, and to find the maximum absolute value of the cumulative sums which we obtain. We then divide this maximum absolute value by n, the sample size, and the result will be the value of the K-S test statistic D.

Because the distribution of claim amounts is skewed to the right we expect that the largest deviations of the observed from expected frequencies would occur in the lower tail of the distribution. The deviations usually change sign from every interval, or every few intervals, to the next and so it is expected that the maximum absolute deviation, in the cumulative sum of the (Actual-Expected) frequencies, will occur somewhere in the lower tail of the distribution. Therefore, we need only to add the values of (Actual-Expected) frequencies for a few lower intervals to be able to calculate D.

If the null hypothesis is simple, then we compare the value of D with the critical values of its distribution given in table (2.1) for large n. If H_0 is composite and the distribution under the null hypothesis is the two-parameter lognormal distribution, then we use table (2.2).

Let $P = Pr\ (D(n) \geq D)$ where $D(n)$ is some value in the table of the critical values of D. In other words, P is the probability of finding a value from the distribution of the K-S statistic which is greater than that found from our sample. According to different values of P we can classify the result of the test as follows:

If (i) $P \geq 0.05$

or (ii) $0.01 \leq P < 0.05$

or (iii) $P < 0.01$

then the difference between the observed and the postulated distributions, under the null hypothesis, based on the given sample is respectively:

(i) Not Significant

(ii) Almost Significant

(iii) Significant

The above arbitrary classification is based on a 5% significance level. A different level of significance, say 10%, may be adopted if stronger confidence is required from the test.

2.9 Other Goodness-of-Fit Tests

There are two other goodness-of-fit tests which are, like the Kolmogorov-Smirnov test, based on the deviation between the sample empirical distribution function and the hypothesized distribution function. These are:

(i) Cramer-Von Mises

(ii) Anderson-Darling

test statistics. For the calculation of these statistics we require the sample order statistics, and hence we need to know the values of the individual observations. Most data on claim amounts, including our own AD data, are only available in grouped frequency form which does not allow the exact calculation of these statistics. We therefore do not consider these statistics in the present work.

2.10 Tables

Table (2.1)

Critical values of the Kolmogorov-Smirnov Statistic $D_\alpha(n)$, for testing completely specified distributions.
sample size = n > 35

Level of significance (α)	0.20	0.15	0.10	0.05	0.01
$D_\alpha(n)$	$\frac{1.07}{\sqrt{n}}$	$\frac{1.14}{\sqrt{n}}$	$\frac{1.22}{\sqrt{n}}$	$\frac{1.36}{\sqrt{n}}$	$\frac{1.63}{\sqrt{n}}$

Table (2.2)

Critical values of the Kolmogorov-Smirnov Statistic, $D_\alpha(n)$, for testing a normal distribution with unspecified mean and variance. sample size = n > 40

Level of significance (α)	0.20	0.15	0.10	0.05	0.01
$D_\alpha(n)$	$\frac{0.736}{\sqrt{n}}$	$\frac{0.768}{\sqrt{n}}$	$\frac{0.805}{\sqrt{n}}$	$\frac{0.886}{\sqrt{n}}$	$\frac{1.031}{\sqrt{n}}$

CHAPTER 3

The Lognormal Distribution

3.1 Introduction

Over the past century, the lognormal distribution has emerged as one of the most widely applied distributions in practical statistical work. Aitchison and Brown (1957) studied this distribution thoroughly and in 1957 published a book entitled "The Lognormal Distribution" in which almost all the results previously found by other people are collated. An extensive bibliography is also included. Their review of the literature shows that this distribution has successfully fitted data from various branches of science and engineering. A more up to date bibliography is provided by Johnson and Kotz (1970).

References to the application of the lognormal distribution in the field of insurance were made in section 1.4. The papers by Benckert (1962) and Ferrara (1971) seem most relevant to the present work. In both papers, the lognormal distribution is fitted to several sets of claim size data from different branches of general insurance. The former, however, does not define or present the data and uses only maximum likelihood estimation formulae when individual observations are available. In the latter reference, grouped data are available and the three parameter lognormal distribution is fitted by a combination of the methods of quantiles and least squares. We shall comment further on this procedure when we consider the estimation problem of the three parameter case.

Aitchison and Brown (1957) deal extensively with the estimation problem where individual observations are available. Their treatment of estimation from grouped data is, however, so brief that it is contained in a few paragraphs.

Most data on claim amounts, including our accidental damage data, are available in grouped form only. A substantial part of this chapter is, therefore, devoted to the problem of estimation from grouped data. The two and three parameter lognormal distributions are initially defined and some of their properties are derived. A theoretical justification for the emergence of the lognormal model for the distribution of claim amounts is put forward. In section 3.6, several tests of lognormality for the 2-parameter case are examined and the accidental damage data is then tested. A simulation exercise is carried out to see how these tests perform when applied to actual samples of lognormal data. Estimation from grouped data, for the 2-parameter distribution, is studied in section 3.7. Several methods are considered and, in particular, a special technique is proposed for estimating the parameters by the method of multinomial maximum likelihood. A computer simulation is performed to measure the efficiency of various methods of estimation. In 3.8, the 2-parameter model is fitted to the accidental damage data and the results are analysed. The effects of inflation on the parameters of the model are next discussed and a technique for predicting the distribution of claim amounts during a future period is suggested. This is then tested on our AD data. Several indices of prices and wages are examined to find the appropriate index for changes in the accidental damage claim amounts over time.

The 3-parameter case is dealt with next. A graphical test of lognormality is suggested which provides us with an approximate method of estimating the location parameter. In section 3.12, the estimation problem for grouped data is considered. A special technique for estimating the parameters by the method of least squares is suggested. The multinomial maximum likelihood method is also modified for the 3-parameter case. These methods are then used to fit the 3-parameter model to our accidental damage data. The results are analysed in 3.13. The theoretical effects of the rate of inflation on the parameters of the model are studied in 3.14. Predictions are made using accidental damage data from the far past and the results are compared with actual data from the past. Finally, the findings of this chapter will be discussed in the conclusion section 3.15.

3.2 Definition

A random variable X is said to be lognormally distributed if and only if Y = log X is normally distributed. Let μ and σ^2 be the parameters, mean and variance respectively, of the distribution of Y.. We denote its probability density and distribution functions by $f_N(y;\ \mu,\sigma^2)$ and $N(y;\mu,\sigma^2)$ respectively, where

$$f_N(y;\ \mu,\sigma^2) = (\sigma\sqrt{2\pi})^{-1} \exp\left[-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2\right] \qquad (3.2\text{-}1)$$

and

$$N(y;\mu,\sigma^2) = \int_{-\infty}^{y} f_N(t;\ \mu,\sigma^2)dt \qquad (3.2\text{-}2)$$

for $-\infty < y < \infty$; $-\infty < \mu < \infty$; $0 < \sigma^2 < \infty$

The probability density function (p.d.f.) of X can be derived, by using the transformation Y = log X, as

$$f_{LN}(x;\ \mu,\sigma^2) = (\sigma x\sqrt{2\pi})^{-1}\exp\left[-\frac{1}{2}\left(\frac{\log x-\mu}{\sigma}\right)^2\right] \qquad (3.2\text{-}3)$$

and its distribution function, $LN(x;\mu,\sigma^2)$, will be

$$LN(x;\mu,\sigma^2) = N(\log x;\mu,\sigma^2) \qquad (3.2\text{-}4)$$

for $x > 0$; $-\infty < \mu < \infty$; $0 < \sigma^2 < \infty$

$f_{LN}(x;\mu,\sigma^2)$, as defined in (3.2-3), is the p.d.f. of the two-parameter lognormal distribution. It is evident that μ and σ^2 are not location or scale parameters for X.

If we rearrange (3.2-3) into

$$f_{LN}(x; \mu, \sigma^2) = [\sigma\rho(x/\rho)\sqrt{2\pi}]^{-1} \exp\left\{-\frac{1}{2}\left[\frac{\log(x/\rho)}{\sigma}\right]^2\right\} \quad (3.2\text{-}5)$$

where $\rho = e^{\mu}$

then it becomes obvious that $\rho = e^{\mu}$ and σ are the scale and shape parameters respectively. It will be shown later that e^{μ} is the median of the two-parameter lognormal distribution.

We can introduce a location parameter, c, into the model by replacing x by x - c in (3.2-3). The parameter c serves as a threshold below which a lognormal variable X is not realized. The p.d.f. of the three-parameter lognormal distribution is therefore

$$f_{LN}(x; c, \mu, \sigma^2) = [\sigma(x - c)\sqrt{2\pi}]^{-1} \exp\left\{-\frac{1}{2}\left[\frac{\log(x - c) - \mu}{\sigma}\right]^2\right\} \quad (3.2\text{-}6)$$

for $x > c$; $-\infty < \mu < \infty$; $0 < \sigma^2 < \infty$

and its distribution function will be

$$LN(x; c, \mu, \sigma^2) = N(\log(x - c) ; \mu, \sigma^2) \quad (3.2\text{-}7)$$

In the 3-parameter case it is $\log(X - c)$ which is distributed normally and not $\log X$.

3.3 Properties of the Two-Parameter Lognormal Distribution

The following results can easily be derived for the two-parameter lognormal distribution (see, for instance, Aitchison and Brown (1957)).

The distribution is unimodal and has a mode at

$$x_{mode} = \exp(\mu - \sigma) \quad (3.3\text{-}1)$$

The median of the distribution is at

$$x_{median} = \exp(\mu) \quad (3.3\text{-}2)$$

Moments of all orders exist and in particular the rth moment of X about zero is

$$\theta'_r = E(X^r) = \exp(r\mu + \tfrac{1}{2}r^2\sigma^2) \qquad (3.3\text{-}3)$$

Heyde (1963) has shown that this distribution cannot be uniquely determined by its moments because there exist other distributions with the same moment sequence as $\{\theta'_r\}$.

The mean of the distribution, which for simplicity we denote by α , is at

$$\alpha = E(X) = \theta'_1 = \exp(\mu + \tfrac{1}{2}\sigma^2) \qquad (3.3\text{-}4)$$

The rth central moment, θ_r , can be expressed in terms of the rth and lower moments of X about zero. In particular

$$\theta_2 = \theta'_2 - \theta'^2_1 \qquad (3.3\text{-}5)$$

$$\theta_3 = \theta'_3 - 3\theta'_1\theta'_2 + 2\theta'^3_1 \qquad (3.3\text{-}6)$$

$$\theta_4 = \theta'_4 - 4\theta'_1\theta'_3 + 6\theta'^2_1\theta'_2 - 3\theta'^4_1 \qquad (3.3\text{-}7)$$

Hence the variance of X, denoted by β^2, is

$$\mathrm{Var}(X) = \beta^2 = \theta_2 = \exp(2\mu + \sigma^2)(e^{\sigma^2} - 1)$$

which on using (3.3-4) gives

$$\beta^2 = \alpha^2\lambda^2 \quad \text{where} \quad \lambda^2 = e^{\sigma^2} - 1 \qquad (3.3\text{-}8)$$

$\beta = \alpha\lambda \doteq \alpha\sqrt{e^{\sigma^2}-1}$

From(3.3-8) it is obvious that λ is the coefficient of variation of the distribution which depends on σ^2 only.

From (3.3-6) and (3.3-7), θ_3 and θ_4 can be expressed in terms of α and λ as

$$\theta_3 = \alpha^3(\lambda^6 + 3\lambda^4)$$

$$\theta_4 = \alpha^4(\lambda^{12} + 6\lambda^{10} + 15\lambda^8 + 16\lambda^6 + 3\lambda^4)$$

Hence the coefficients of skewness and kurtosis of the distribution, $\sqrt{\beta_1}$ and β_2 respectively, are

$$\sqrt{\beta_1} = \frac{\theta_3}{\beta^3} = \lambda^3 + 3\lambda > 0 \qquad (3.3\text{-}9)$$

$$\beta_2 = \frac{\theta_4}{\beta^4} = \lambda^8 + 6\lambda^6 + 15\lambda^4 + 16\lambda^2 + 3 > 3 \qquad (3.3\text{-}10)$$

which indicate that all lognormal densities are skewed to the right and that they are more peaked than their related normal densities. For small values of λ , and hence of σ^2 , $\sqrt{\beta_1}$ and β_2 are close to 0 and 3 respectively. In such cases, the central portion of a lognormal frequency curve resembles a normal curve and may be approximated by it.

From(3.3-3) it is evident that θ_r' is a product of $e^{r\mu}$ and $e^{\frac{r^2\sigma^2}{2}}$ which Laurent (1963) defines as the two "functional characteristics" of the distribution. The mean and variance of the lognormal distribution are non-functional in the sense that two lognormal populations may have the same mean or variance although they have different parameters μ and σ^2. This motivates the use of the median, e^{μ} , and the function e^{σ^2} as measures of central tendency and dispersion respectively.

We can show that a simple relationship exists between the quantiles of the same order of the lognormal and the standard normal distributions. Let z_q and x_q be the quantiles of order q of $N(z;\ 0,1)$ and $LN(x;\ \mu,\sigma^2)$ respectively. Then by definition

$$N(z_q;\ 0,1) = q = LN(x_q;\ \mu,\sigma^2)$$

i.e. (where $Y = \log X$)

$$= P(X \le x_q) = P(\log X \le \log x_q) = P(Y \le \log x_q)$$

$$= P\left(\frac{Y-\mu}{\sigma} \le \frac{\log x_q - \mu}{\sigma}\right)$$

$$= P\left(Z \le \frac{\log x_q - \mu}{\sigma}\right)$$

(where Z is the standardized normal variable)

$$= N\left(\frac{\log x_q - \mu}{\sigma};\ 0,1\right)$$

Hence $$z_q = \frac{\log x_q - \mu}{\sigma} \tag{3.3-11}$$

or $$x_q = \exp(\mu + \sigma z_q) \tag{3.3-12}$$

In particular, the median of $N(z; 0,1)$ is at $z_{0.50} = 0$, hence from (3.3-12) the median of the distribution of X is, as mentioned earlier, at e^{μ}.

Before concluding this section, we may note that the relative positions of the mode, the median and the mean, i.e.

$$e^{\mu-\sigma}, \quad e^{\mu} \quad \text{and } e^{\mu + \frac{1}{2}\sigma^2} \text{ respectively}$$

provide additional evidence that all lognormal densities are positively skewed.

3.4 Properties of the 3-Parameter Lognormal Distribution

The transformation $x \to x - c$ to obtain the 3-parameter distribution is only a translation along the x-axis. It leaves the shape of the frequency curve unchanged and only shifts it by an amount c along the x-axis.

The location measures are, therefore, increased by c and hence

$$x_{mode} = c + \exp(\mu - \sigma) \tag{3.4-1}$$

$$x_{median} = c + \exp(\mu) \tag{3.4-2}$$

$$\alpha' = x_{mean} = c + \exp(\mu + \tfrac{1}{2}\sigma^2) \tag{3.4-3}$$

The moments about the mean and the dispersion measures remain the same as for the 2-parameter case. In particular the variance, β^2, and the coefficients of skewness and kurtosis are the same for both the two and three parameter lognormal distributions because they are all functions of the central moments. The coefficient of variation becomes

λ' where

$$\lambda' = \frac{\beta}{\alpha'}$$

i.e. $$\lambda' = \frac{\beta}{c+\alpha} = \frac{\lambda}{1+c/\alpha} \tag{3.4-4}$$

which is a function of all three parameters.

The relationship between quantiles of order q of $LN(x ; c,\mu,\sigma^2)$ and $N(z ;0, 1)$ will, from (3.3-12), become

$$x_q = c + \exp(\mu + \sigma z_q) \tag{3.4-5}$$

which is equivalent to $$z_q = \frac{1}{\sigma}\log(x_q - c) - \frac{\mu}{\sigma} \tag{3.4-6}$$

3.5 The Lognormal Distribution as a Model for Claim Amounts

Let X_o be the amount which an insurance company at the inception of a policy expects to pay in case of a claim on that policy (X_o can be considered closely related to the net premium). Suppose that there are n factors which affect the size of a claim on this policy, such as age, type of car, district, driving skill, climate, occupation, etc. Let X_j be the size of claim due to the effect of factors 1 to j only. If we assume that the effect of factor j is to modify X_{j-1} to X_j by multiplying it by a random perturbation U_j, i.e.

$$X_j = X_{j-1} U_j$$

$$= X_o U_1 U_2 \ldots U_j$$

then for a policy subject to n different factors we have

$$X_n = X_o U_1 U_2 \ldots U_n$$

where $\{U_1, U_2, \ldots , U_n\}$ is assumed to be a sequence of random variables with known joint distribution. Consequently

$$\log X_n = \log X_0 + \sum_{j=1}^{n} \log U_j$$

provided that the joint distribution of U's is such that the central limit theorem applies to the sum of their logarithms, it follows that $\log X_n$ will tend to normality for large n. Hence the amount of the claim will be (approximately) lognormally distributed . Ferrara (1971) and Finger (1976) also justify the lognormal model with arguments very similar to the above.

3.6 Tests of Lognormality :- The 2-Parameter Case

An important step in the identification of the statistical model is to test if our sample is likely to be from a population whose distribution belongs to a particular family of distributions. In other words, we should, as a first step, test if our sample is, say, from a lognormal population. At this stage we are not interested in the parameters of the model but would just like to know if, on the basis of the sample values, the assumption of lognormality is reasonable enough to allow further analysis according to this model.

We consider three tests for the lognormal distribution.

3.6.1 Graphical Test

From equation (3.3-11) , i.e.

$$z_q = \frac{1}{\sigma} \log x_q - \frac{\mu}{\sigma}$$

it is obvious that the locus of the points $(\log x_q , z_q)$ is a straight line. When these points are calculated from a sample and plotted on a rectangular co-ordinate system then, if the sample is from a two parameter lognormal population, the points should lie approximately on a straight line.

To calculate z_q, quantile of order q of $N(z; 0,1)$ distribution, we need to know q. This can be calculated from the sample empirical distribution function which we define as

$$F(x) = P(X \leq x) = \frac{\text{Total number of claims} \leq x}{\text{Total number of claims in the sample}} \qquad (3.6\text{-}1)$$

Therefore $F(x)$ is simply the proportion of claims with amounts less than or equal to x. If we express this ratio as a percentage, then we call $F(x)$ the sample cumulative percentage function. Hence, at the point x_q in the sample we take $q = F(x_q)$ and enter a table of the cumulative distribution function of the standard normal distribution in order to find the value of the variate z_q. The point $(\log x_q, z_q)$ is thus determined. The use of a special graph paper called the "logarithmic probability paper" makes the above task easier. This graph paper has one of its axes graduated logarithmically while the other is graduated according to a standard normal probability scale. Hence we only need to know and plot $(x_q, q\%)$. The logarithmic axis converts x_q to $\log x_q$ while the probability axis converts the percentage proportion q to its corresponding standard normal variate z_q.

3.6.2 The Skewness and Kurtosis Tests

As mentioned earlier, if X has a lognormal distribution then $\log X$ is normally distributed. This suggests that a test of normality for $\log X$ is equivalent to a test of lognormality for X. Therefore, we may apply the skewness and kurtosis tests of normality on the transformed values, $\log x$, and infer from its result whether the original sample values are from a lognormal population.

This test is based on g_1 and g_2, the sample coefficients of skewness and excess kurtosis respectively, for the distribution of $\log X$, i.e.

$$g_1 = \frac{m_3}{m_2^{3/2}} \qquad (3.6\text{-}2)$$

and

$$g_2 = \frac{m_4}{m_2^2} - 3 \qquad (3.6\text{-}3)$$

where m_r is the sample central moment of order r.

Exact expressions for means and variances of g_1 and g_2 are given in Cramer (1946) as :

$$E(g_1) = 0 \qquad (3.6\text{-}4)$$

$$\text{var}(g_1) = \frac{6(n-2)}{(n+1)(n+3)} \qquad (3.6\text{-}5)$$

$$E(g_2) = -\frac{6}{n+1} \qquad (3.6\text{-}6)$$

$$\text{var}(g_2) = \frac{24n(n-2)(n-3)}{(n+1)^2(n+3)(n+5)} \qquad (3.6\text{-}7)$$

where n is the sample size.

We can, therefore, see by how many standard deviations g_1 and g_2 differ from their mean values. If the difference is less than, for instance, three standard deviations, then it is not significant, while if it is greater than or equal to, say, three standard deviations, then the difference indicates significant deviation from the assumption of normality under which the mean and variance have been calculated.

Geary and Pearson (1938) have given the 5% and 1% probability points of g_1 and $(g_2 + 3)$ for various sample sizes. In large samples, however, the rough test of normality provided by comparing g_1 and g_2 with the approximate values of their standard errors, namely $\sqrt{\frac{6}{n}}$ and $\sqrt{\frac{24}{n}}$ respectively, would be sufficient.

3.6.3 Test in the (β_1, β_2) Plane

Let $\sqrt{\beta_1}$ and β_2 be the coefficients of skewness and kurtosis, respectively, for a distribution. Barton and Dennis (1952) divide the (β_1, β_2) plane into different regions of unimodal frequency curves. They show that in the (β_1, β_2) plane the lognormal distribution is represented by a straight line. This information provides us with another test of lognormality.

Suppose that we are given several samples of independent observations on a random variable (say, claim amounts incurred during different periods of accident). We can calculate $\sqrt{b_1}$ and b_2, i.e. the sample coefficients of skewness and kurtosis respectively. In large samples, $\sqrt{b_1}$ and b_2 should be close to their population values $\sqrt{\beta_1}$ and β_2. Therefore, we can plot the points (b_1, b_2), calculated for different samples, on a rectangular co-ordinate system of axes. If the underlying distributions of the populations from which our samples were derived are lognormal, we would expect the points (b_1, b_2) to lie, approximately, on a straight line. Therefore, with the availability of several samples, we can test whether the underlying model for the distribution of claim amounts is lognormal.

3.6.4 Testing the Accidental Damage Data for 2-Parameter Lognormality

We first applied the graphical test. Accidental Damage data for seven different periods of accident were presented in tables (1.1) to (1.7). For each sample, data is in grouped form and the sample cumulative percentage function, $F(x_i) = q_i\%$, has been calculated at each point x_i which is the claim amount equal to the upper boundary of interval i. Considering the range of the claim amounts, the points (x_i, q_i) were plotted on a 3-cycle logarithmic probability paper which has its log axis (for claim amount) graduated for increases of x up to one thousandfold. The plots for different periods of accident are presented in figures (3.1-a) and (3.1-b). To avoid producing a too voluminous thesis, these

Figure (3.1-a) - Points (x , F(x)) plotted on logarithmic probability paper .

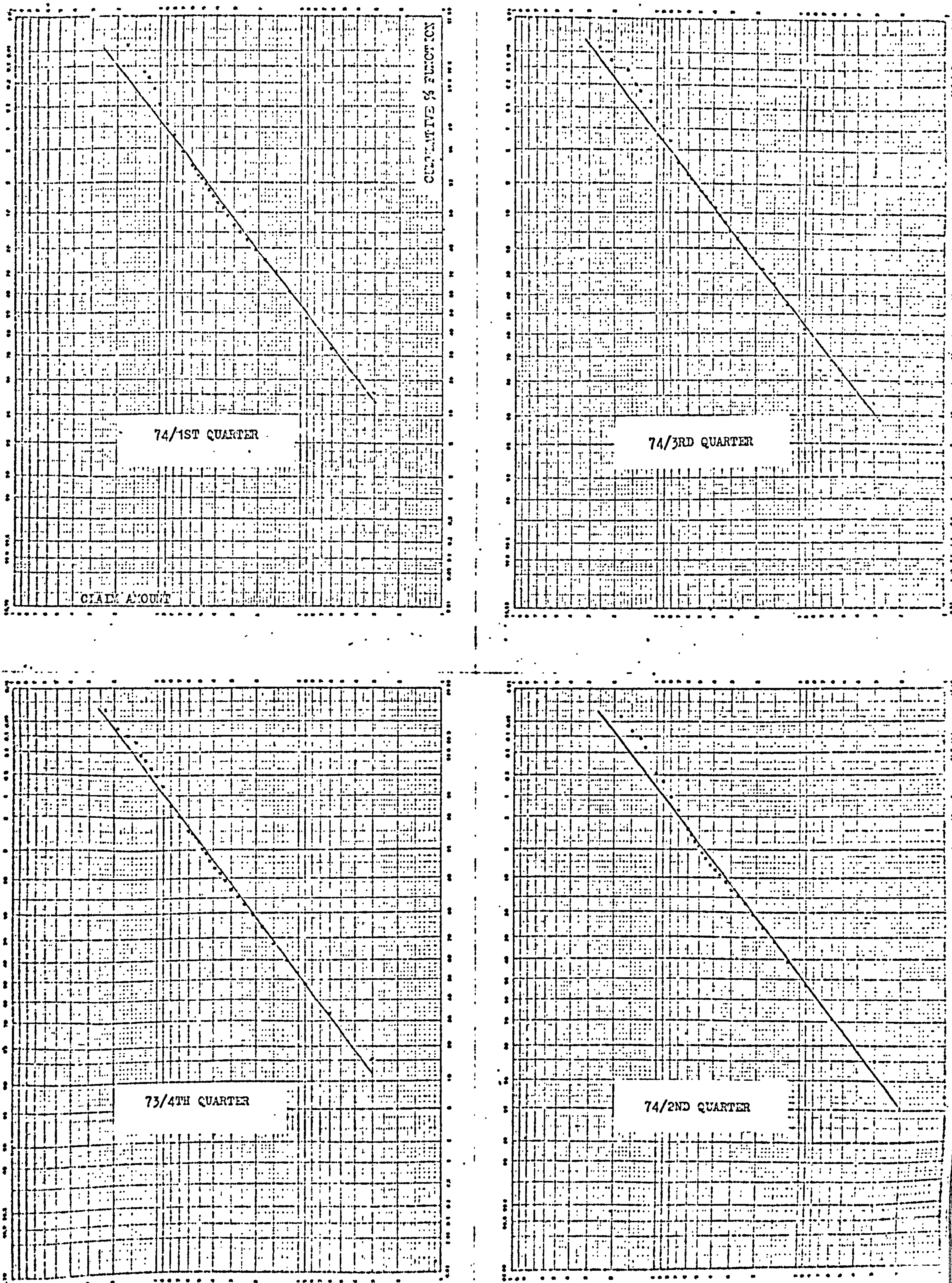
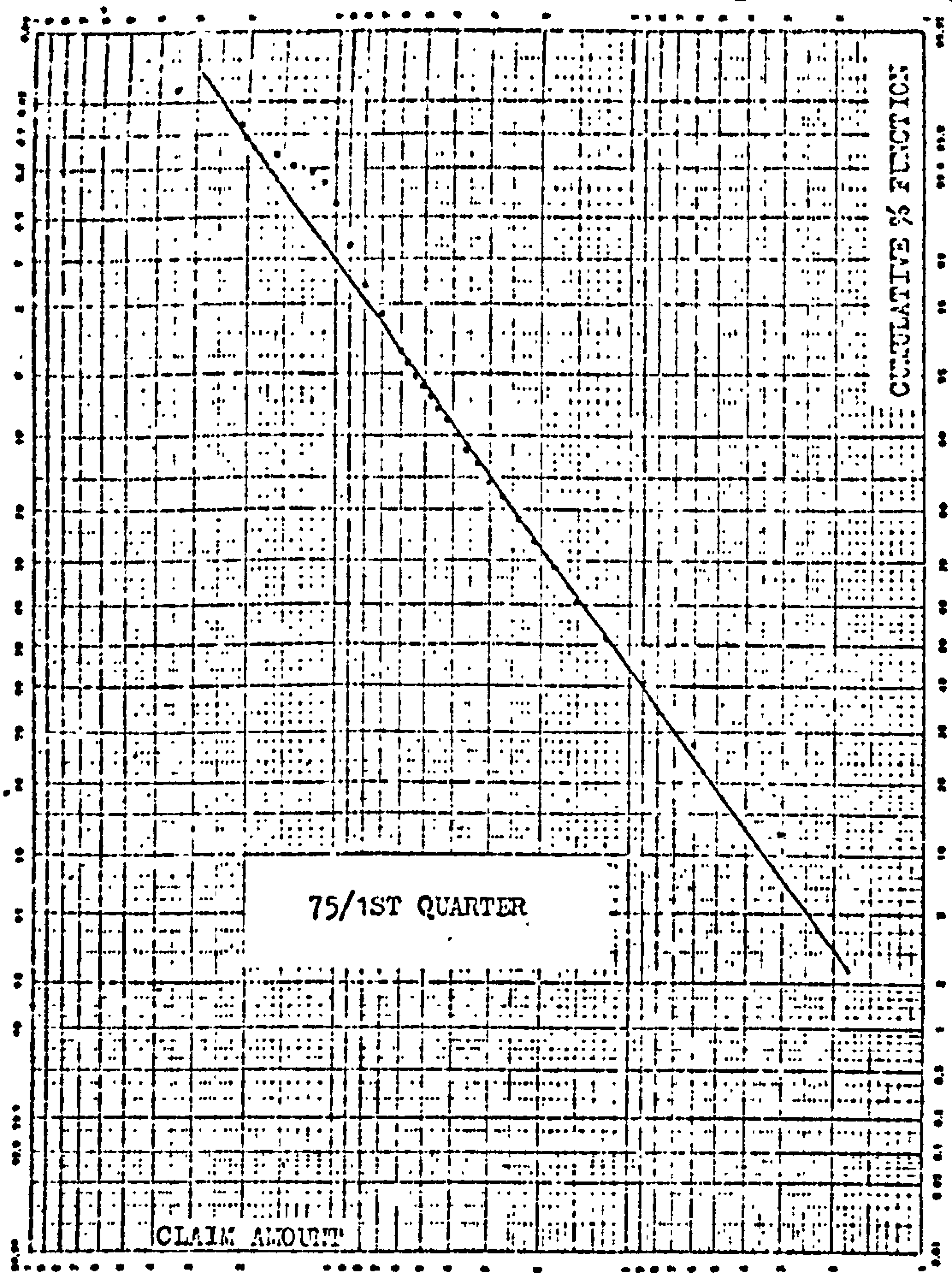
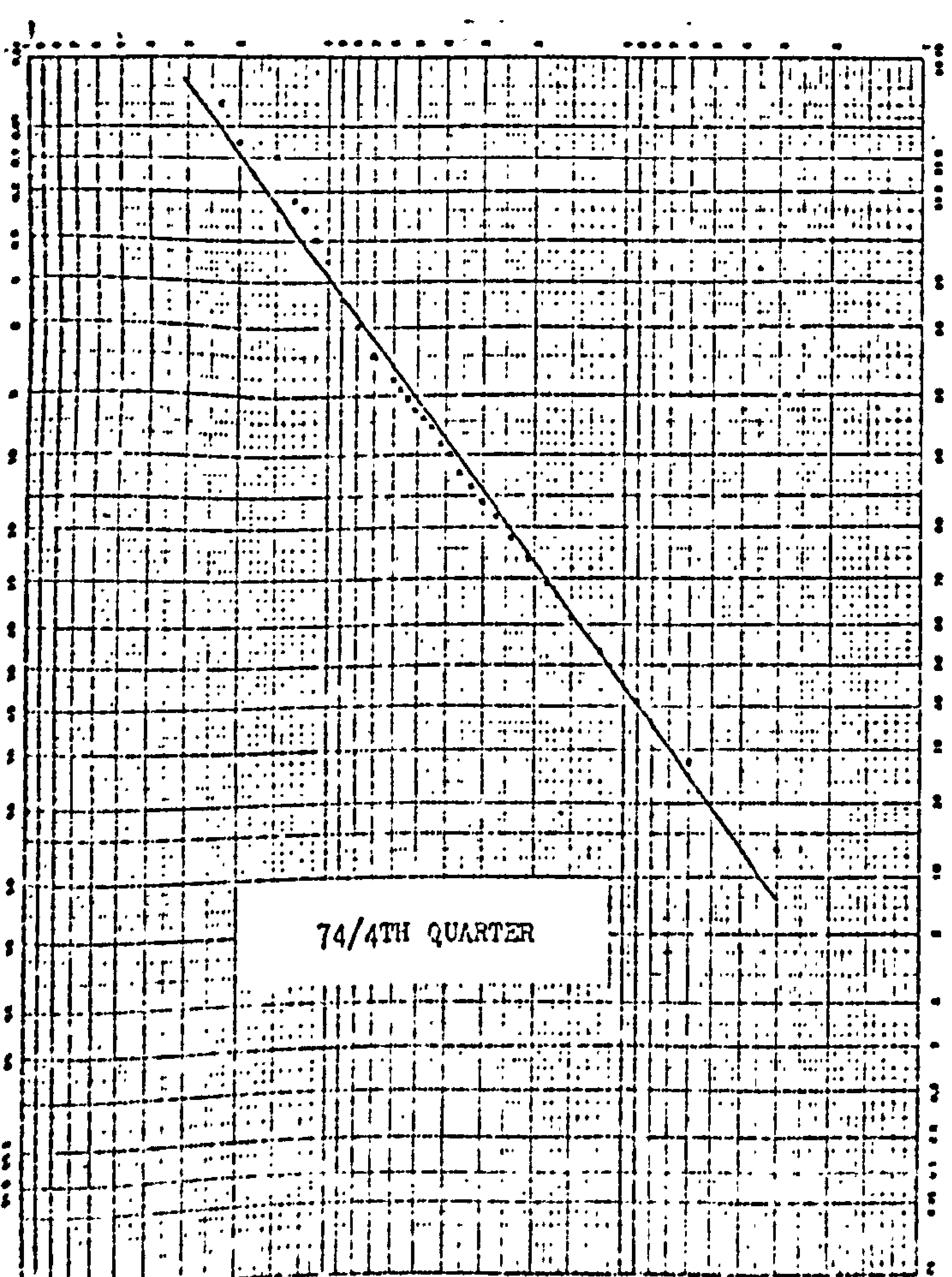
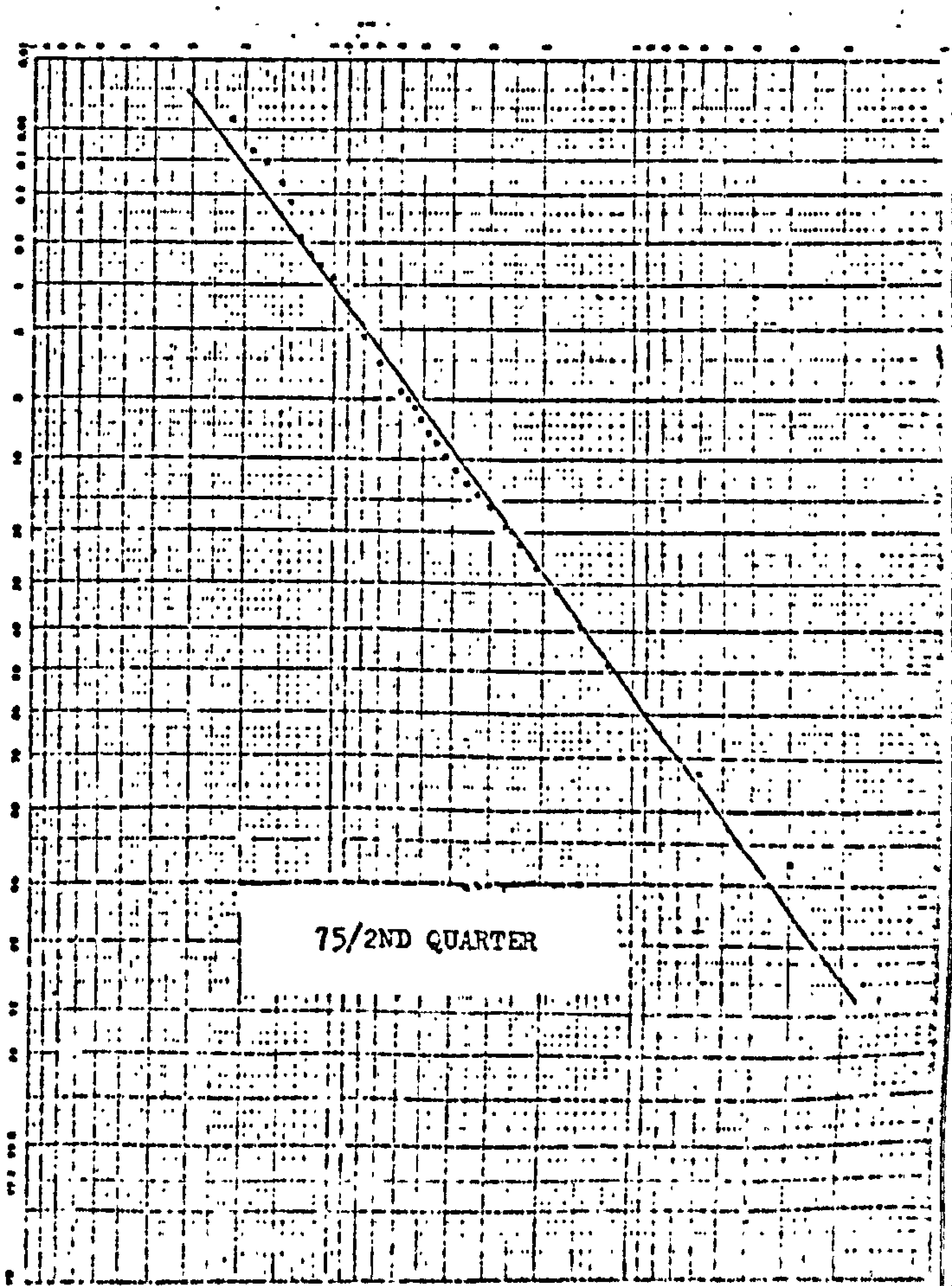


Figure (3.1-b) - Points (x , F(x)) plotted on logarithmic probability paper .







and other graphs have been reduced from their originals. In this case, they have been reduced to $\frac{1}{4}$ of their original size. However, for the purpose of comparison between plots for different samples, it is more convenient to present these graphs as in figures (3.1-a) and (3.1-b). For each sample the points appear to lie approximately on a straight line. We have fitted this line by eye. The pattern of formation of the points about the line is the same for all the samples. The deviations from the line seem to be greater at the tails of the distribution.

It was deemed necessary to see what an actual sample of two-parameter lognormal observations from a population similar to those from which our accidental damage data have been derived would look like when plotted on the logarithmic probability paper. A computer simulation exercise was, therefore, performed. We generated ten random samples each consisting of 2500 lognormal observations from the population with $\mu = 4.5$ and $\sigma^2 = 1$. It will be shown later that these values of μ and σ^2 are close to our estimates of the population parameters for the accidental damage data. The size of each sample was adopted as 2500 because this figure is close to the size of our AD samples.

For each simulated sample, the random lognormal variates were grouped according to the same grouping format as for the AD samples, i.e. up to 600 in bands of 30 and afterwards in bands of 100. The computer program P2 performed the task of simulation and then printed out a table of grouped data with its corresponding sample cumulative percentage function. The ten simulated samples, thus generated, were then each plotted on logarithmic probability papers in the same way as described for the accidental damage data. The plots are presented in figures (3.2-a), (3.2-b) and (3.2-c).

In each case the points seem to lie approximately on the straight line

Figure (5.2-a) - Points (x , F(x)) plotted on logarithmic probability paper.

(SIMULATED SAMPLES)

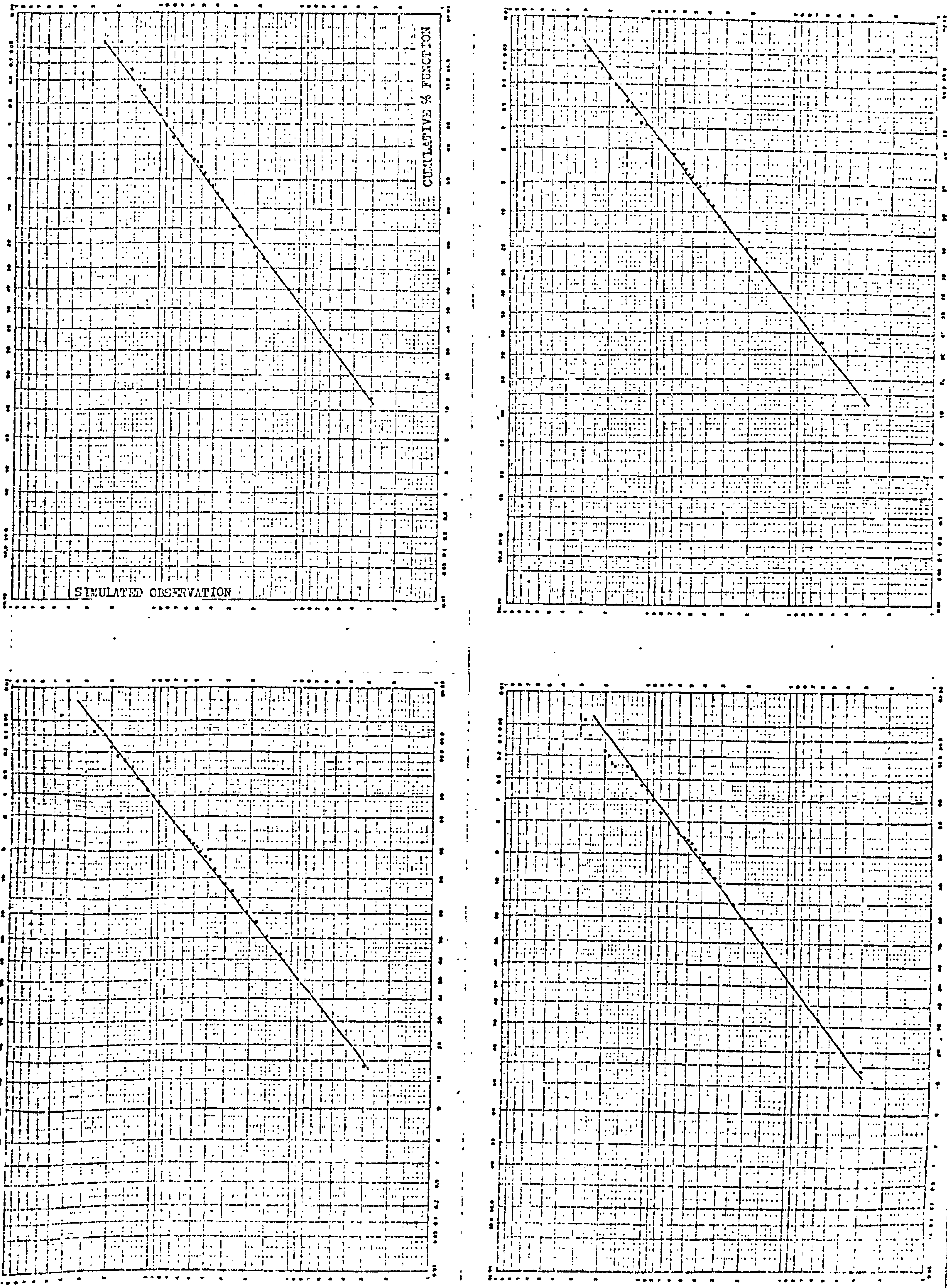


Figure (3.2-b) - Points (x ; F(x)) plotted on logarithmic probability paper .
(SIMULATED SAMPLES)

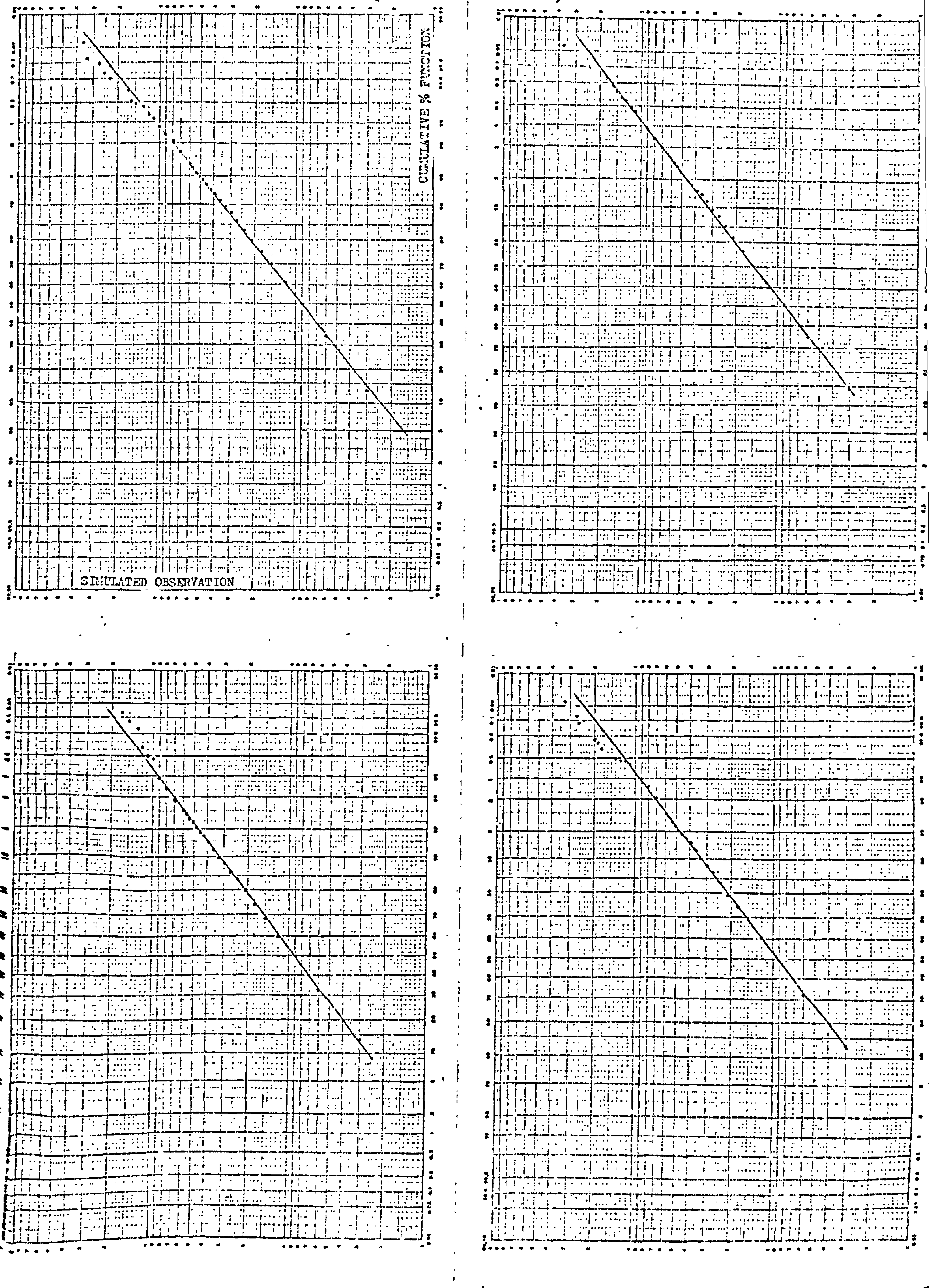
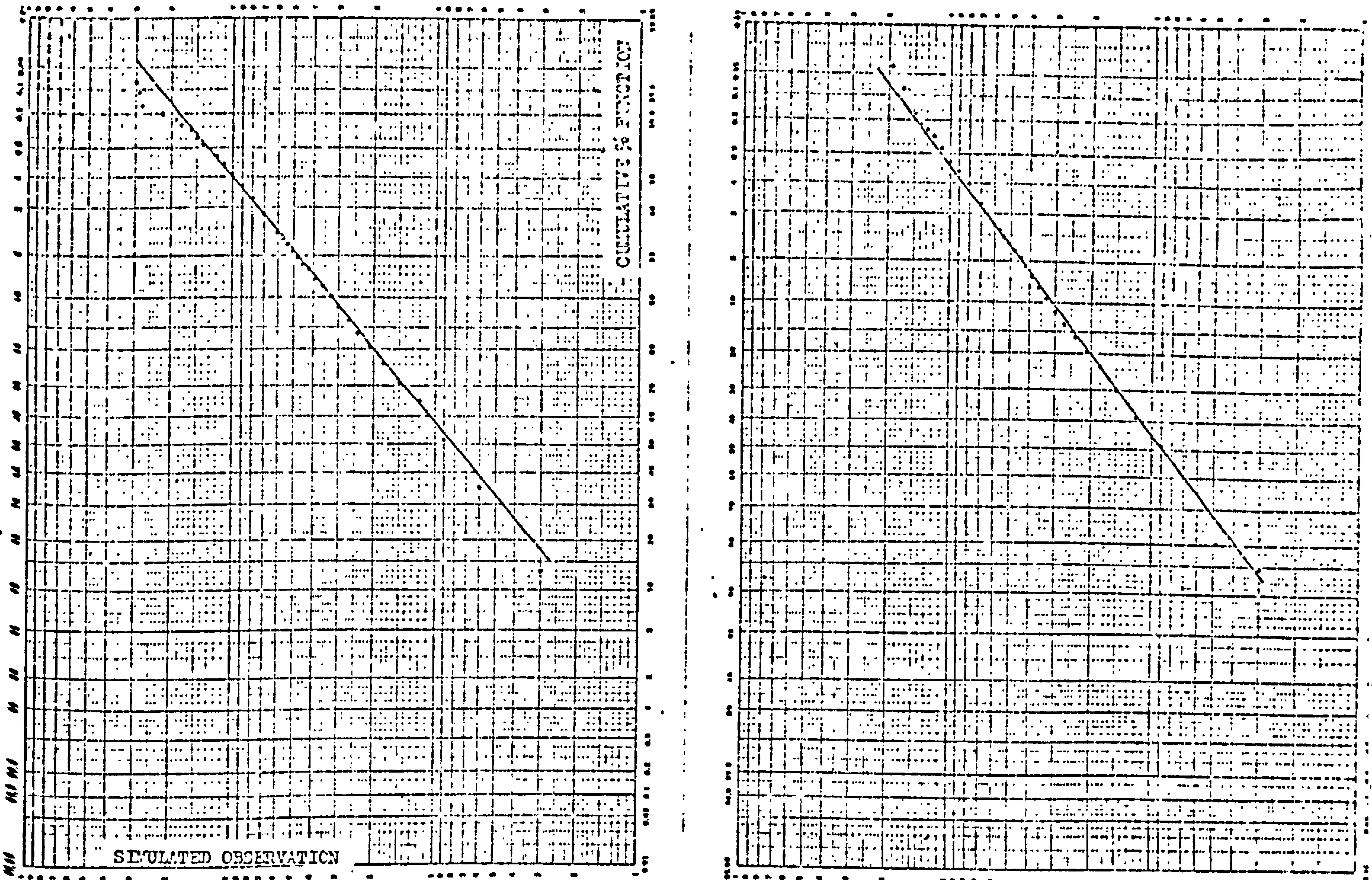


Figure (3.2-c) - Points (x , F(x)) plotted on logarithmic probability paper .
(SIMULATED SAMPLES)



which has been fitted by eye. In the lower tail of the distribution the points seem to lie almost exactly on the line , unlike the behaviour of the accidental damage samples. The pattern of the points in the upper tail of the simulated samples is similar to that of the accidental damage data, i.e. the deviations from the line are more marked. This indicates that even from a sample of 2500 lognormal observations we cannot obtain a close fit to the straight line in the upper tail of the distribution. This is because the lognormal is a very skew distribution and observations in the tail are scarce. Therefore, we need a much larger sample size before we can see a true picture of the upper tail of the distribution. Hence we can safely assume that the pattern of the points in the upper tail of the accidental damage samples is by no means unusual to the lognormal distribution and should, therefore, feel satisfied that the distribution fits the upper tail reasonably well. The simulated samples, however, tell a different story in the lower tail of the distribution. Here the points lie very much closely on the straight line and the deviations from it are much smaller than in the case of the accidental damage samples. Therefore, the pattern of the points in the lower tail of the accidental damage samples is not consistent with that of the two-parameter lognormal distribution. We believe that this inconsistency could be partly because of the insufficiencies and inaccuracies of the data in the lower tail. Most policy holders do not claim for small amounts for fear of losing their full entitlement to no claim discount which, in some cases, is worth more than the amount they may recover by a claim. Therefore, the data in the lower intervals may be incomplete. On the whole, the assumption of two-parameter lognormality as tested by this method seems reasonable enough to allow further analysis.

Two other tests of lognormality, mentioned in 3.6.2, are the skewness and kurtosis tests of normality. The values of the coefficients

g_1 and g_2 for each sample of the accidental damage data were calculated by program P1. These were presented in tables (1.1) to (1.7) and are produced in table (3.1) at the end of this chapter. The standard errors of these coefficients, for large samples, have been calculated according to the formulae of section 3.6.2. These are presented in table (3.1) as well. From this table it is apparent that in four periods, namely 74/3rd quarter to 75/2nd quarter, the value of g_1 is greater than three times its standard error. Therefore, in these four cases, the underlying assumption of normality for $\log X$, and hence of two-parameter lognormality for X (the claim amount) should be rejected. For the remaining three periods, namely 73/4th quarter to 74/2nd quarter, the lognormal assumption for the distribution of claim amounts is supported. The values of g_2 are significantly larger than their standard errors. Therefore, this test suggests that the assumption of two-parameter lognormality should be rejected. g_1 and g_2 were calculated from grouped data with the assumption made, in section 1.6, about the concentration of the claims in each band at the mid-point of that band. This could affect the values of g_1 and g_2 calculated from grouped data. A Sheppard's correction for the calculation of moments of $\log X$ is not applicable exactly because the intervals of $\log X$ are not of equal length.

It was decided to put the ten simulated lognormal samples to these tests. The values of g_1 and g_2 were calculated from the simulated grouped data in exactly the same way as for the accidental damage data. Their standard errors for a sample of size 2500 were also calculated. The results are presented in table (3.2). The values of g_1 compare favourably with its standard error and, therefore, this test performs very well. The values of g_2 are in most cases acceptable, but in the case of samples 1, 7 and 8, g_2 is greater than three times its standard error.

Therefore, the coefficients g_1 and g_2, even when calculated from grouped data, can be used to test normality for $\log X$ and, on the basis of the simulated samples, it seems that g_1 can indicate normality more accurately than g_2. The comparison of the values of g_1 and g_2 calculated from accidental damage samples with those for the simulated samples shows that the assumption of two-parameter lognormality is, on the whole, reasonable.

Another test of lognormality is the test in (β_1, β_2) plane which was mentioned in section 3.4.3. The values of $\sqrt{b_1}$ and b_2 for the AD samples were calculated by program P1 and they were shown in tables (1.1) to (1.7). Figure (3.3) shows a plot of the points (b_1, b_2). These points, although some of them are very near each other, seem to lie approximately on a straight line. This supports the assumption of lognormality as the underlying model for the distribution of claim amounts.

To see how this test performs when applied to actual lognormal samples, b_1 and b_2 were calculated for each of the ten simulated samples. The points (b_1, b_2) are plotted in figure (3.4). They appear to lie approximately on a straight line. Therefore, this test can be applied to test lognormality. However, for reliable results we need a large number of points and hence a large number of samples.

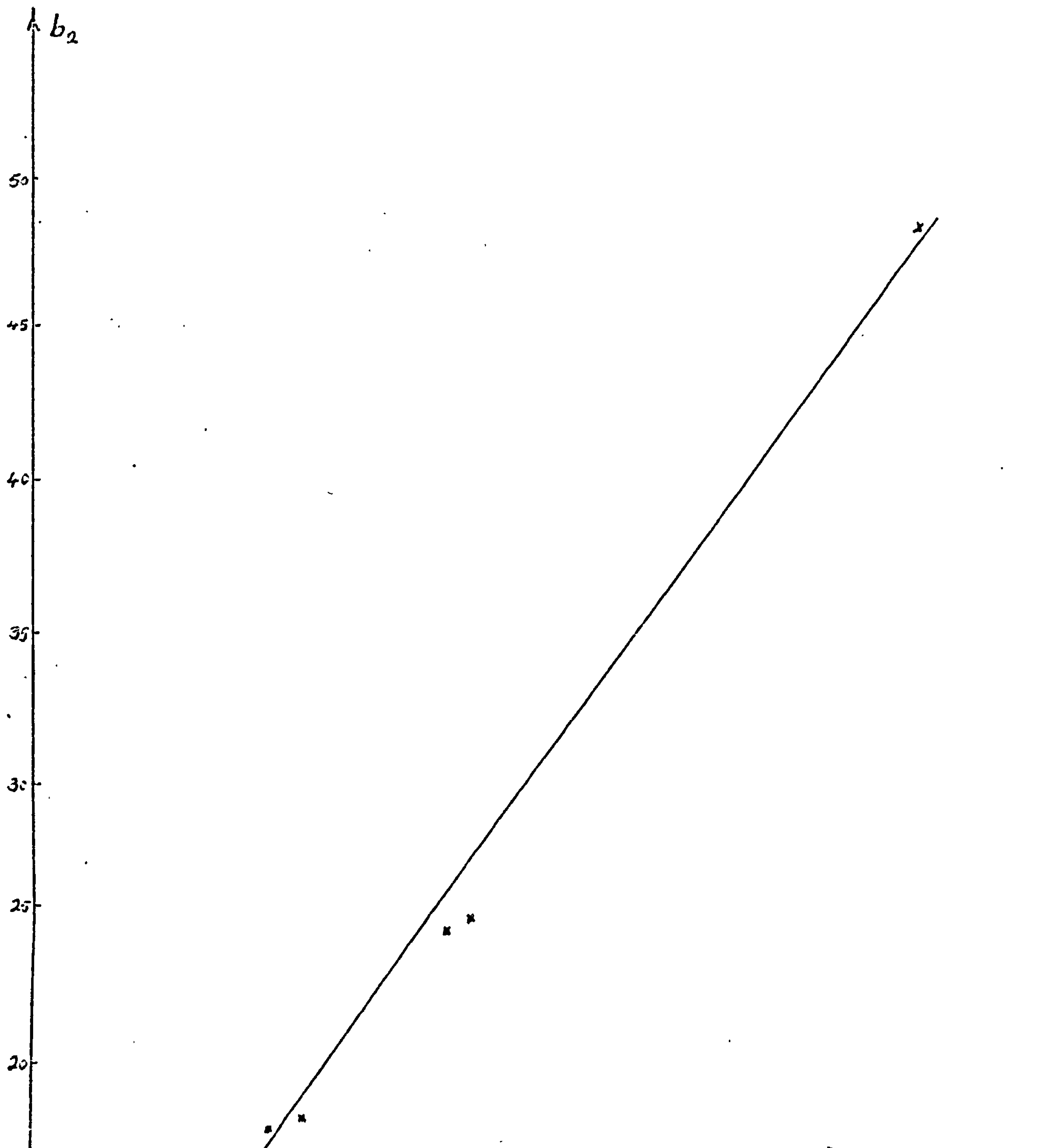


Figure (3.3) - Plot of the (b1,b2) points for the Accidental Damage data .

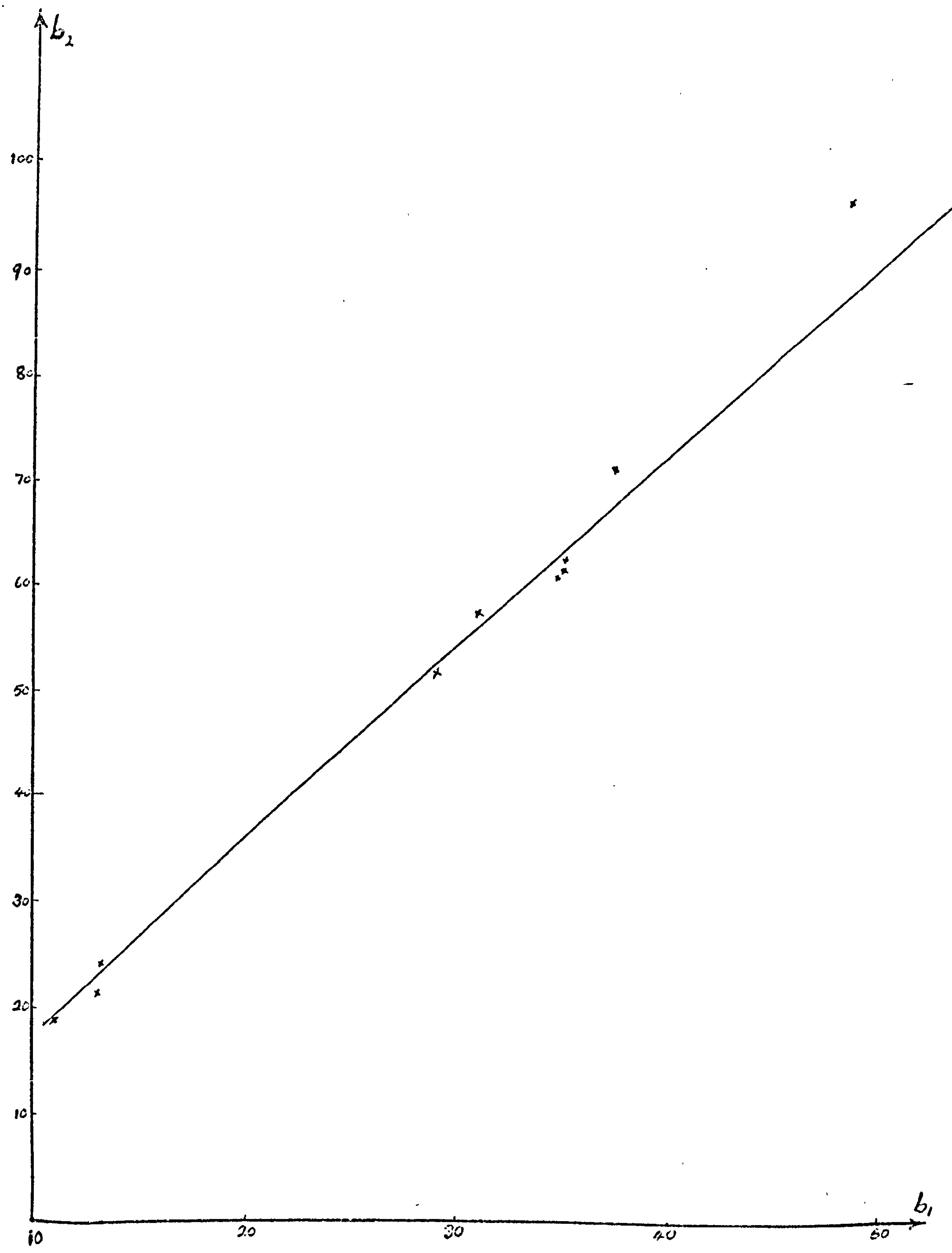


Figure (3.4) - Plot of the (b1,b2) points for the simulated samples.

3.7 Estimation of the Parameters of the 2-Parameter Lognormal Distribution

Aitchison and Brown (1957) investigated thoroughly the problem of estimating the parameters of the lognormal distribution when the value of each individual observation in the sample is available. However, their treatment of the problem in the case of grouped data is very brief and they point out that there are certain difficulties in the various methods of estimation. In this section we study the problem of estimation from grouped data. Some of the methods, described below, have been modified from methods of estimation for when individual sample values are given. Some others, such as the graphical, the least squares, and the multinomial maximum likelihood methods are directly suitable for grouped data.

Let us assume that we have a sample of grouped data where n independent random observations on a random variable X (in our case, the claim amount) have been grouped according to their size into k intervals. Further, let n_i be the number of observations (claims) in the class interval (x_{i-1}, x_i) such that

$$n = \sum_{i=1}^{k} n_i$$

It is usual, in the case of grouped data, to assume that all the values within any interval are concentrated at the mean of the portion of the distribution over that interval. If the distribution is the correct model for the population from which the sample values have been derived, then for every relatively small interval, the mean of the portion of the distribution over that interval can be considered as equal to the mean of sample values in the same interval. Information on the average amount of claim in any interval was provided by the insurance company. This showed that for every interval the average amount was very nearly equal to the value of the mid-point of that interval. We, therefore, feel justified in assuming that within any interval (x_{i-1}, x_i) the claims are concentrated at the point

$$\dot{x}_i = \frac{1}{2}(x_{i-1} + x_i) \tag{3.7-1}$$

*3.7.1 The Method of Moments

Let S_1 and S_2 be respectively the first and the second moments of the sample values about zero, i.e.

$$S_1 = \frac{1}{n}\sum_{i=1}^{k} n_i\dot{x}_i \tag{3.7-2}$$

$$S_2 = \frac{1}{n}\sum_{i=1}^{k} n_i\dot{x}_i^2 \tag{3.7-3}$$

Estimation by the method of moments consists of putting S_1 and S_2 equal to their corresponding theoretical values from the distribution. Hence, by using equation (3.3-3) with $r = 1$ and $r = 2$ respectively, we will have

$$S_1 = \exp(\mu + \tfrac{1}{2}\sigma^2) \tag{3.7-4}$$

$$S_2 = \exp(2\mu + 2\sigma^2) \tag{3.7-5}$$

Solving the above simultaneous equations for μ and σ^2 we obtain

$$\hat{\mu} = 2\log S_1 - \tfrac{1}{2}\log S_2 \tag{3.7-6}$$

and

$$\hat{\sigma}^2 = \log S_2 - 2\log S_1 \tag{3.7-7}$$

where $\hat{\mu}$ and $\hat{\sigma}^2$ are the estimates of μ and σ^2 respectively. We have given some consideration to the question of applying a Sheppard's correction to the second sample moment. However, numerical investigations showed that this correction does not have a significant impact on the values of $\hat{\mu}$ and $\hat{\sigma}^2$. It has been suggested by some authors (see for instance, Geary and Pearson (1938)) that unless the grouping interval is more than one-third of the standard deviation of the sample, it is not really necessary to apply any corrections to the moments. A glance at tables (1.1) to (1.7) shows that for our AD samples the interval width for the majority of the bands, containing some 95% of the total number of claims, is 30, while the standard deviation in each sample is at least six times that. From the above considerations a correction to the second moment is not deemed necessary.

3.7.2 The Approximate Maximum Likelihood Method

The word "approximate" in the title of this method is used because of the assumption of concentration of the claims in every interval at the mid-point of that interval. For a grouped sample, using the notation already adopted, the likelihood function L is :

$$L = (2\pi\sigma^2)^{-n/2} \prod_{i=1}^{k} \dot{x}_i^{-n_i} \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{k} n_i(\log \dot{x}_i - \mu)^2\right] \quad (3.7\text{-}8)$$

The same values of μ and σ^2 maximize both L and $\log L$ where

$$\log L = -\frac{n}{2}\log 2\pi - \frac{n}{2}\log\sigma^2 - \sum_{i=1}^{k} n_i \log \dot{x}_i - \frac{1}{2\sigma^2}\sum_{i=1}^{k} n_i(\log \dot{x}_i - \mu)^2 \quad (3.7\text{-}9)$$

The values of μ and σ^2 which maximize $\log L$ are the solutions of the following two simultaneous equations,

$$\frac{\partial \log L}{\partial \mu} = 0$$

$$\frac{\partial \log L}{\partial \sigma^2} = 0$$

Hence, it can be easily shown that $\hat{\mu}$ and $\hat{\sigma}^2$ the estimates of μ and

σ^2 respectively are

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^{k} n_i \log \dot{x}_i \tag{3.7-10}$$

and

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{k} n_i (\log \dot{x}_i)^2 - \hat{\mu}^2 \tag{3.7-11}$$

3.7.3 The Method of Quantiles

The relationship (3.3-12) between the same order quantiles of the lognormal and the standard normal distributions, namely

$$x_q = e^{\mu + \sigma z_q}$$

is the basis of this method. Let x_{q_1} and x_{q_2} be two different quantiles of orders q_1 and q_2 calculated from the sample. These can be either selected as the upper boundaries of two different intervals or be calculated by interpolation in the intervals of x. The corresponding standard normal quantiles, z_{q_1} and z_{q_2} respectively, can be calculated by reference to a table of N(0,1) distribution function. $\hat{\mu}$ and $\hat{\sigma}^2$ the estimates of μ and σ^2 respectively, are the solutions of the following simultaneous equations

$$x_{q_1} = e^{\hat{\mu} + \hat{\sigma} z_{q_1}}$$

$$x_{q_2} = e^{\hat{\mu} + \hat{\sigma} z_{q_2}} .$$

Hence

$$\hat{\mu} = \frac{z_{q_2} \log x_{q_1} - z_{q_1} \log x_{q_2}}{z_{q_2} - z_{q_1}} \tag{3.7-12}$$

and

$$\hat{\sigma}^2 = \left(\frac{\log x_{q_2} - \log x_{q_1}}{z_{q_2} - z_{q_1}} \right)^2 \tag{3.7-13}$$

It can be shown that this method is most efficient when the two quantiles are symmetrically placed (Aitchison and Brown (1957)). Let, therefore

$$q_1 = q \quad \text{and} \quad q_2 = 1 - q \quad \text{where } q < 1/2$$

then

$$z_{q_1} = z_q = - t \text{ , say, where } t > 0 \text{ since } q < 1/2$$

and

$$z_{q_2} = z_{1-q} = t$$

Hence $\hat{\mu}$ and $\hat{\sigma}^2$ will be

$$\hat{\mu} = \frac{1}{2} (\log x_{1-q} + \log x_q) \qquad (3.7\text{-}14)$$

and

$$\hat{\sigma}^2 = \frac{1}{4t^2} (\log x_{1-q} - \log x_q)^2 \qquad (3.7\text{-}15)$$

Aitchison and Brown (1957) state that quantiles of order 27% and 73% estimate μ with 81% efficiency and quantiles of order 7% and 93% estimate σ^2 with an efficiency of 65%. We will later use these two sets of quantiles to estimate μ and σ^2 from the AD samples. To obtain the values of the quantiles linear interpolation in the intervals of x will be used.

3.7.4 The Method of Median and Coefficient of Variation

This is a very simple method which involves calculating the sample median and coefficient of variation, x_{median} and V respectively, and putting them equal to their theoretical values from the distribution. To find x_{median}, interpolation in the intervals of x may be necessary. For the sample coefficient of variation, V , we need to calculate the sample mean and standard deviation. The expressions for the median and coefficient of variation for the two-parameter lognormal distribution were given by (3.3-2) and (3.3-8) respectively. Hence

$$x_{median} = e^{\mu}$$

and

$$V = (e^{\sigma^2} - 1)^{\frac{1}{2}}$$

The estimates of μ and σ^2 will therefore be

$$\hat{\mu} = \log x_{median} \qquad (3.7\text{-}16)$$

and $$\hat{\sigma}^2 = \log (1 + V^2) \qquad (3.7\text{-}17)$$

This method can be considered as a combination of the methods of moments and quantiles.

3.7.5 The Graphical Method

This method is derived as a follow-up of the graphical test of lognormality mentioned earlier in section 3.6.1. Therefore, we know that the locus of the points $(\log x_q, z_q)$ is the straight line

$$z = \frac{1}{\sigma} \log x - \frac{\mu}{\sigma}$$

Let us assume that we have followed the procedure of section 3.6.1 and plotted the points $(\log x_q, z_q)$ from a sample of grouped data. Let us further assume that these points seem to lie approximately on a straight line. We can fit this line to the points by eye. The inverse of the gradient of the line will give an estimate of σ and then from the intercept of the line we find the estimate of μ .

Aitchison and Brown (1957) suggest the following method when the logarithmic probability paper has been used to plot the points :
From a table of the standard normal distribution function we notice that

$$z_{16\%} \doteqdot z_{15\cdot 8\%} = -1$$

and

$$z_{84\%} \doteqdot z_{84\cdot 13\%} = +1$$

$$z_{50\%} = 0$$

Therefore, if we put the lognormal population quantiles of orders 16%, 50% and 84% equal to their corresponding values read from the straight line we will have

$$x_{16\%} = e^{\mu-\sigma} \tag{3.7-18}$$

$$x_{50\%} = e^{\mu} \tag{3.7-19}$$

$$x_{84\%} = e^{\mu+\sigma} \tag{3.7-20}$$

Hence from (3.5-18)

$$\hat{\mu} = \log x_{50\%} \tag{3.7-21}$$

Then from (3.5-19)

$$e^{\hat{\sigma}} = \frac{x_{50\%}}{x_{16\%}}$$

and from (3.5-20)

$$e^{\hat{\sigma}} = \frac{x_{84\%}}{x_{50\%}}$$

a better estimate of e^{σ} would be

$$e^{\hat{\sigma}} = \frac{1}{2}\left(\frac{x_{50\%}}{x_{16\%}} + \frac{x_{84\%}}{x_{50\%}}\right)$$

Hence

$$\hat{\sigma}^2 = \left\{\log\left[\frac{1}{2}\left(\frac{x_{50\%}}{x_{16\%}} + \frac{x_{84\%}}{x_{50\%}}\right)\right]\right\}^2 \tag{3.7-22}$$

For grouped data, this method has the advantage that the data is already in the form required by it.

3.7.6 The Method of Least Squares

This is a variation of the graphical method. Instead of fitting a straight line by eye to the set of points $(\log x_q, z_q)$ we use the least squares regression method to determine its equation. The theory of linear regression tells us that if the unweighted least squares line through a set of m points (U_i, V_i), where i = 1, 2, ... , m , is

$$V = aU + b \tag{3.7-23}$$

then the coefficients a and b are

$$a = \frac{\overline{UV} - \overline{U}\,\overline{V}}{\overline{U^2} - \overline{U}^2} \tag{3.7-24}$$

and

$$b = \overline{V} - a\overline{U} \tag{3.7-25}$$

where

$$\overline{U} = \frac{1}{m}\sum_{i=1}^{m} U_i \quad , \quad \overline{V} = \frac{1}{m}\sum_{i=1}^{m} V_i$$

$$\overline{U^2} = \frac{1}{m}\sum_{i=1}^{m} U_i^2 \quad \text{and} \quad \overline{UV} = \frac{1}{m}\sum_{i=1}^{m} U_iV_i$$

As we mentioned earlier, the point (log x ,z) lies on the straight line

$$z = \frac{1}{\sigma}\log x - \frac{\mu}{\sigma} \tag{3.7-26}$$

Therefore, we can let

$$U_i = \log x_{q_i} \quad \text{and} \quad V_i = z_{q_i}$$

where x_{q_i} is the upper boundary of interval i, q_i is the value of the sample cumulative percentage function at the point x_{q_i} and z_{q_i} is the standard normal variate corresponding to q_i, i.e.

$$q_i = (\sqrt{2\pi})^{-1} \int_{-\infty}^{z_{q_i}} \exp\left(-\frac{t^2}{2}\right)dt$$

The problem then is to calculate the coefficients a and b of the line

$$V = aU + b$$

If our grouped data consists of k intervals, then i takes integer values from 1 to k - 1 because for the final interval $q_k = 1$ which results in an indeterminate value of z_q. The values of z_q can be obtained from a table of the standard normal cumulative distribution function. If this estimation procedure is programmed on the computer, an approximate formula for calculating z_q from q should be used. Hastings (1968) gives the following approximate formula for calculating the standard normal variate , z , from , p , the area under the upper tail of its frequency curve :

$$z = t - \frac{2 \cdot 515517 + 0 \cdot 802853t + 0 \cdot 010328r}{1 + 1 \cdot 432788t + 0 \cdot 189269r + 0 \cdot 001308rt} \qquad (3.7\text{-}27)$$

where

$$p = (\sqrt{2\pi})^{-1} \int_z^{\infty} \exp\left(-\frac{\theta^2}{2}\right) d\theta \quad , \quad 0 < p \leq 0 \cdot 5$$

$$r = \log \frac{1}{p^2} \qquad \text{and} \qquad t = \sqrt{r}$$

To use this formula in conjunction with the estimation method, we have to take

$p = 1 - q_i$ where q_i is as defined above.

Let, therefore, V = aU + b be the equation of the least squares line where a and b have been calculated according to the formulae (3.7-24) and (3.7-25). A comparison between the equation of this line and that of its equivalence given by (3.7-26) shows that

$$a = 1/\hat{\sigma} \qquad \text{and} \qquad b = -\hat{\mu}/\hat{\sigma}$$

Hence $$\hat{\mu} = -b/a \qquad (3.7\text{-}28)$$

and $$\hat{\sigma}^2 = a^{-2} \qquad (3.7\text{-}29)$$

3.7.7 The Multinomial Maximum Likelihood Method

This is by far the most suitable method for estimation of the parameters from grouped data. Let us assume that a sample of data as defined in section 3.7 is available. If we denote by p_i the probability that the amount of a claim will fall in the interval $(x_{i-1} - x_i)$, then

$$p_i = LN(x_i;\ \mu,\sigma^2) - LN(x_{i-1};\ \mu,\sigma^2) \qquad (3.7\text{-}30)$$

$$p_i = N\left(\frac{y_i - \mu}{\sigma};\ 0,1\right) - N\left(\frac{y_{i-1} - \mu}{\sigma};\ 0,1\right) \text{ where } y = \log x \qquad (3.7\text{-}31)$$

Therefore, L , the likelihood function of the sample will be

$$L = \frac{n!}{\prod_{i=1}^{k} n_i!} \prod_{i=1}^{k} p_i^{n_i} \qquad (3.7\text{-}32)$$

and that part of the loglikelihood function which depends on the parameters μ and σ^2 will be

$$\log L = \sum_{i=1}^{k} n_i \log\left[N\left(\frac{y_i - \mu}{\sigma};\ 0,1\right) - N\left(\frac{y_{i-1} - \mu}{\sigma};\ 0,1\right)\right] \qquad (3.7\text{-}33)$$

The multinomial maximum likelihood estimates of μ and σ^2 are obtained by maximising L (or log L) simultaneously with respect to μ and σ^2. Therefore, $\hat{\mu}$ and $\hat{\sigma}^2$ are the solutions of the simultaneous equations

$$\frac{\partial \log L}{\partial \mu} = 0$$

and

$$\frac{\partial \log L}{\partial \sigma^2} = 0$$

From these equations we cannot find an analytic solution for $\hat{\mu}$ and $\hat{\sigma}^2$. Gjeddebaek (1949) has considered this problem and introduces and tabulates two functions to facilitate the interpolation involved in the solution of the above equations. His method, however, becomes very laborious even when the number of intervals is greater than 3. It is possible to use an iterative procedure to solve the system of non-linear equations in μ and σ^2. Tallis and Young (1962) apply this method in the case of the 3-parameter lognormal distribution.

We instead propose the use of an iterative method which maximizes $\log L$ itself with respect to the parameters μ and σ^2. There are now many computer routines available for maximizing a general function of severable variables. We use one such routine provided by the NAG[1] library of programs. This routine minimizes the function $-\log L$ (i.e. it maximizes $\log L$) with respect to the unknown parameters by using a revised quasi-Newton method implemented by Gill and Murray (1972). Analytical derivatives of the function need not be supplied. To start the iterative process the initial values for $\hat{\mu}$ and $\hat{\sigma}^2$ are obtained by one of the simpler estimation methods described earlier. The procedure uses estimates of the gradient and curvature of the objective function to generate a sequence of points which are intended to converge to a local maximum of $\log L$. $\hat{\mu}$ and $\hat{\sigma}^2$ thus obtained can then be used as the starting point in order to make sure that they are indeed the optimum points. If on inspecting a maximum found by the routine it is suspected that a 'better' solution exists, the remedy is to re-run the routine from another starting point closer to the expected solution. With the wide availability of the computer this is very much easier than the methods of Gjeddebaek or Tallis and Young.

In conclusion, we point out that there is another suitable method of estimation from grouped data which is called 'the modified Chi-square minimum method'. This involves minimizing a modified Chi-square type statistic with respect to the unknown parameters. Cramer (1946),

1 - Nottingham Algorithmic Group

however, shows that this method is identical with the multinomial maximum likelihood method. Therefore, we shall not consider the modified Chi-square minimum method in this work.

3.7.8 Measuring the Performance of Various Methods of Estimation

A simulation exercise was carried out to see how well different methods of estimation perform in practice. 100 samples, each of size 2500, were generated from the lognormal population with parameters $\mu = 4.5$ and $\sigma^2 = 1$. These values for the parameters will be shown to be close to the values for our accidental damage (AD) data. The size of the simulated samples, 2500 , is typical of the size of our accidental damage samples. In this way, simulated samples resembling the actual AD samples were produced. The computer program P3 was written for this task and samples of data grouped in the same format of intervals as for the accidental damage data were created.

We define the numerical measures of efficiency of a particular method of estimating μ and σ^2 as

$$\mathrm{Eff}(\mu) = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (\hat{\mu}_i - \mu)^2} \qquad (3.7\text{-}34)$$

$$\mathrm{Eff}(\sigma^2) = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (\hat{\sigma}_i^2 - \sigma^2)^2} \qquad (3.7\text{-}35)$$

where $\mu = 4.5$, $\sigma^2 = 1$ and $N = 100$, $\hat{\mu}_i$ and $\hat{\sigma}_i^2$ are the estimates of μ and σ^2 respectively, arrived at by the method in question from sample i. For each pair of estimates ($\hat{\mu}_i$, $\hat{\sigma}_i^2$) we can find, from formulae (3.3-4) and (3.3-8), $\hat{\alpha}_i$ and $\hat{\beta}_i$ which are the estimates of α and β the mean and standard deviation of the lognormal distribution respectively. $\hat{\alpha}_i$ and $\hat{\beta}_i$ are then considered as estimates obtained by the same method of estimation which was used to find $\hat{\mu}_i$ and $\hat{\sigma}_i^2$ from sample i. Because $\hat{\alpha}$ and $\hat{\beta}$ are functions of both $\hat{\mu}$ and $\hat{\sigma}^2$, it is important to know how well a particular method of estimation performs in estimating α and β. Therefore, we can similarly define measures of efficiency for α and β as

$$\mathrm{Eff}(\alpha) = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (\hat{\alpha}_i - \alpha)^2} \qquad (3.7\text{-}36)$$

$$\mathrm{Eff}(\beta) = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (\hat{\beta}_i - \beta)^2} \qquad (3.7\text{-}37)$$

where, from $\mu = 4{\cdot}5$ and $\sigma^2 = 1$ we have, $\alpha = 148{\cdot}41$ and $\beta = 194{\cdot}54$. When comparing two different methods of estimating a certain parameter, the one whose calculated measure of efficiency is smaller will be judged as more efficient and hence will be preferred.

The computer program P4 was written to estimate μ and σ^2, from each of the 100 simulated samples, by the methods of moments, approximate maximum likelihood, 27 and 73% quantiles, 7 and 93% quantiles, median and coefficient of variation and the least squares regression. The graphical method was excluded from the exercise because the task of plotting 100 sets of data points would have been very laborious. Besides, the efficiency of this method depends on how experienced the analyst is in using the graph paper and fitting the line by eye. The multinomial maximum likelihood method was not included because it is the most suitable method for estimating the parameters from grouped data and is, therefore, sure to be very efficient.

Program P4 produced estimates of μ, σ^2, α and β, by various methods, from each sample and calculated the efficiency of each method in estimating each parameter. Table (3.3) shows the results of this simulation exercise. It is apparent that amongst the methods considered the approximate maximum likelihood is the most efficient in estimating the parameters from the particular type of data we are dealing with. The method of 27 and 73% quantiles is the second best and its use should be recommended when simplicity in calculations rather than maximum accuracy is the criterion. The efficiency of the method of median and coefficient of variation for estimating μ is the same as that of the method of 27 and 73% quantiles, and for estimating σ^2 is the same as that of the method of moments. Therefore, the median can be used to obtain a good estimate of μ. The method of 7 and 93% quantiles is not very accurate for our type of data. The methods of moments and least squares regression appear to perform with nearly the same degree of accuracy.

3.8 Application of the 2-Parameter Lognormal Model to the Accidental Damage Data

The results of the tests of lognormality carried out in section 3.6.4 seemed reasonable enough to encourage us to fit the two-parameter lognormal model to the accidental damage data presented in tables (1.1) to (1.7). An extensive computer program, P5 , was written to estimate the parameters μ and σ^2 by all the methods described in section 3.7, except the methods of median and coefficient of variation and multinomial maximum likelihood. The former method was excluded because it can be considered as a combination of the methods of moments and quantiles. The latter will be dealt with separately. For each sample of data supplied as input, the program provides six relevant tables of results. For a particular period of accident, namely the 4th quarter of 1973 which was presented in table (1.1), the results of a run of program P5 are presented in tables (3.4-a) to (3.4-f). The first table, (3.4-a), presents the estimates of μ and σ^2 by different methods as well as some of their corresponding relevant distribution statistics. Table (3.4-b) gives the expected number of claims, in each interval, calculated by using various sets of estimates. A table of actual minus expected frequencies is produced as table (3.4-c). To allow the calculation of the Chi-square statistic, values of $\frac{(A - E)^2}{E}$ for each interval up to £1000 and each set of parameters are presented in table (3.4-d). It was considered that for claim amounts greater than £1000 the expected number of claims in each interval would be less than 5. The total of the $\frac{(A - E)^2}{E}$ values for each method of estimation and for intervals up to £1000 is provided, but these totals need to be adjusted for intervals above £1000. To allow comparison between the actual cost of claims and that expected under the model by different methods of estimation, table (3.4-e) is produced. Each entry in this table is the product of the number of claims in an interval, and the average amount of claim in the same interval (the average is assumed to be equal to the mid-point of the interval). Table (3.4-f) provides the expected loss and the T statistic, as defined in section 2.4 of chapter 2. Program P5 was run with data from other periods of accident as well. To avoid inclusion of a large number of tables, only the first table of each run, similar to table (3.4-a) are presented here in tables (3.5) to (3.10). Relevant statistics such as the Chi-square, the T and the ratio of T to total actual cost have been added to each of these tables.

Let us now closely examine the results produced for 73/4th quarter. A glance at table (3.4-a) shows that with the exception of the method of 7 and 93% quantiles all the methods have produced almost similar results. Estimates of μ are about 4·5 and of σ^2 about 0·9. The skewness and kurtosis of the frequency curve of the model are greater than those of the sample (see table (1.1)), hence indicating that the frequency curve of the model has a longer tail and is more peaked than the histogram of the sample values. Table (3.4-c) indicates that the disagreement between the actual and expected number of claims is more marked in a few of the lower tail intervals. These differences are smallest for the approximate maximum likelihood method. Table (3.4-d) in fact shows that the Chi-square statistic is smallest for that method although all the values of this statistic indicate a significant difference between the sample values and the model. It can be observed that the major contribution to the Chi-square statistic is from one or two cells in the lower tail of the distribution. We may recall that when we tested the data for lognormality, the points in the lower tail did not fit the straight line very well. We blamed the insufficiencies of the data in the lower tail intervals for this behaviour. The fit of the model to the data would have been very satisfactory if it were not for these one or two large contributions. In chapter 7, where we investigate the truncated lognormal distribution, the effects of discarding various intervals in the lower tail will be studied.

Table (3.4-e) indicates that despite the large Chi-square values, the total cost of claims from the sample and the model are fairly close to each other. It is also inferred that although in terms of the number of claims only about 4% of the total lies in the upper tail (greater than £600), in terms of the cost of claims about 17% of the total comes from the upper tail. Table (3.4-f) shows that the total expected loss is not, in fact, very large, except for 7 and 93% quantiles method and that the ratio of T to the total actual cost is quite small.

For the method of approximate maximum likelihood we can apply the Kolmogorov-Smirnov test of goodness-of-fit in conjunction with table (2.2). The value of this statistic is $D = 0.023$ which gives $P < 0{\cdot}01$ (see section 2.8) . Therefore this test, like the Chi-square, indicates significant differences between the model and the actual sample.

For other periods of accident results similar to those of the 73/4th quarter can be observed from tables (3.5) to (3.10).

The above analysis supports the findings of the simulation exercise of section 3.7.8. Amongst the methods considered, the approximate maximum likelihood gives the best fit of the model to the actual data in terms of the goodness-of-fit tests. The fact that for each sample, the Chi-square statistic is smallest for this method is not surprising if we recall the remarks of chapter 2 about the estimation method and its effects on the Chi-square statistic. The method of 27 and 73% quantiles also produces satisfactory results and its use is recommended when a quick solution is required.

Before commenting on the suitability of this model for our type of data we will consider the estimation of the parameters by the method of multinomial maximum likelihood.

Program P6 estimates the parameters by the technique described in section 3.7.7 and provides a table of results. For our accidental damage data, the results are presented in tables (3.11) to (3.17). For each sample, estimates of μ and σ^2, and the values of the mean and standard deviation of the claim amount are provided. The tables are self-explanatory. If, for each sample, we compare the results of this method with those of the approximate maximum likelihood, we will notice that the estimates are very close to each other but a slightly better fit has been provided by the present method. The Chi-square and the total expected loss statistics are smaller for each sample. The values of the Chi-square statistic still indicate significant differences between the model and the sample values. This can be seen to be supported by the Kolmogorov-Smirnov statistic. The actual minus expected frequencies are again large in just a few of the intervals and the pattern appears to be consistent for all the samples. It is, however, encouraging that the values of the test statistics are not drastically large but that they are just about significant.

The analysis so far shows that the method of multinomial maximum likelihood, as we expected, produces more satisfactory results when estimating the parameters from grouped data. Therefore, guided by this information, it will be the principal method which we shall use throughout the forthcoming stages of the present work.

To see how closely the frequency curve of the model approximates the histogram of the sample values, computer program P7 was written to plot both of them. The graphs for various samples are presented in figures (3.5.1) to (3.5.7).

(This section is continued after Figure (3.5.7)).

Figure (3.5.1)

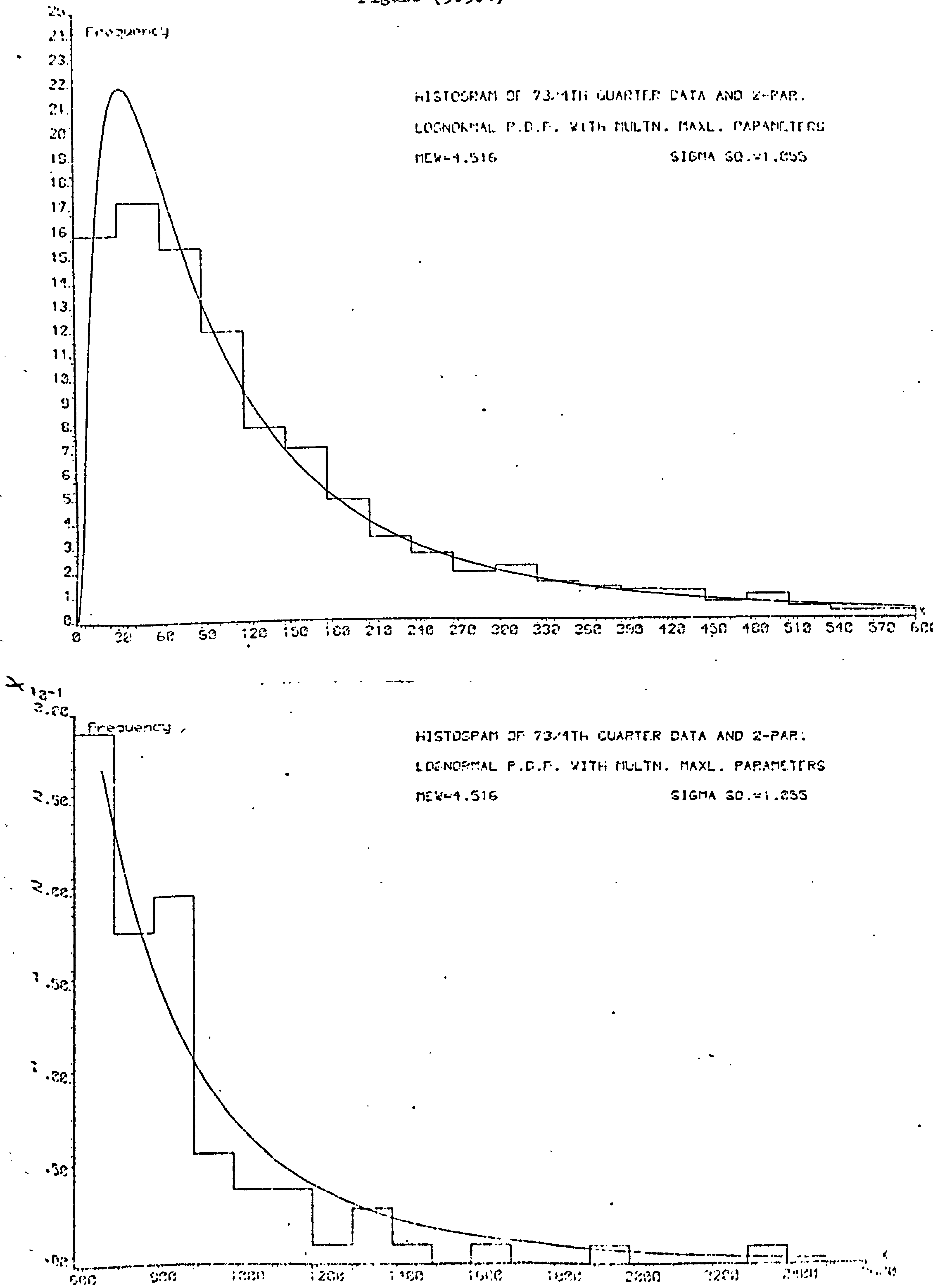
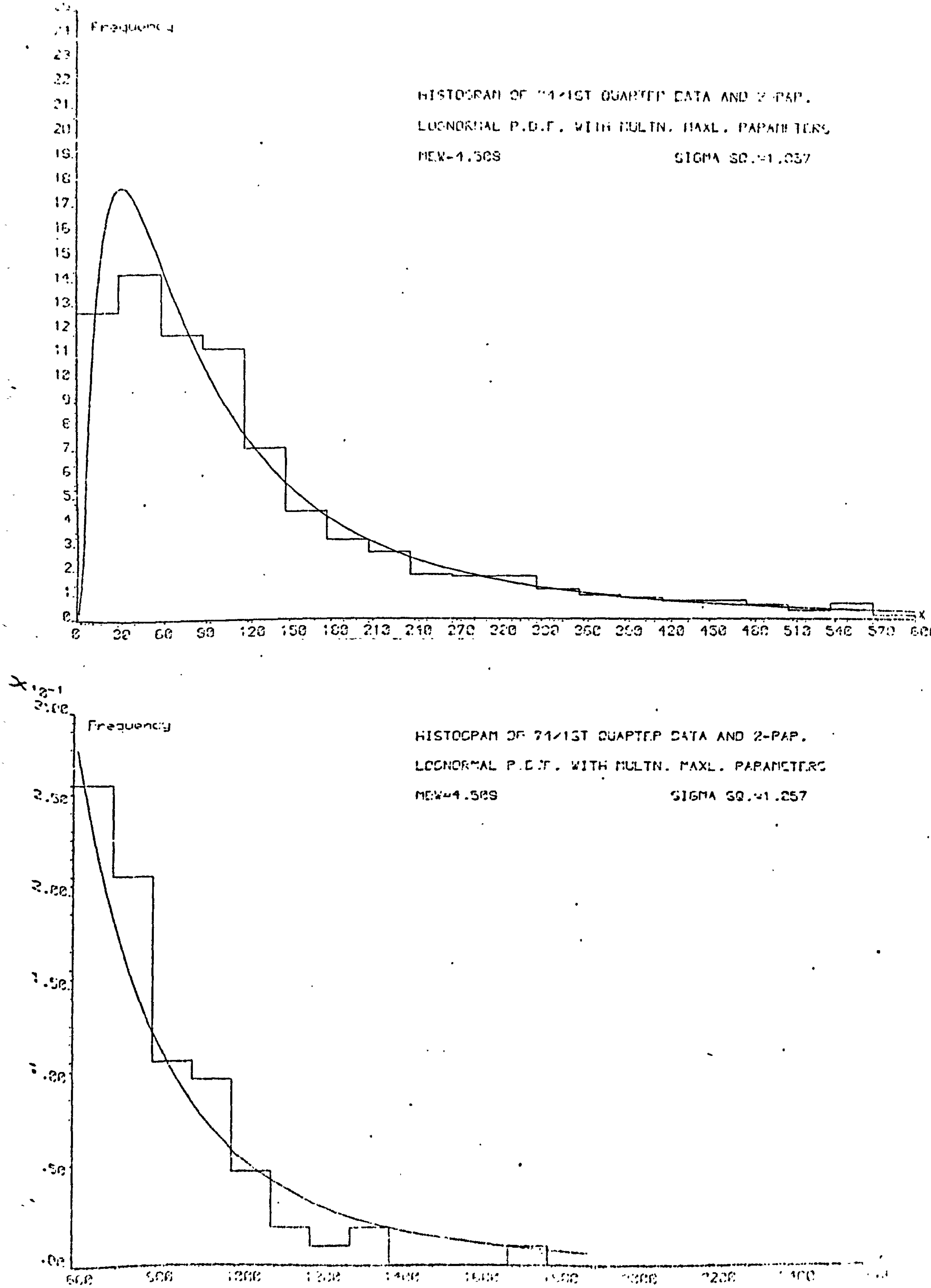
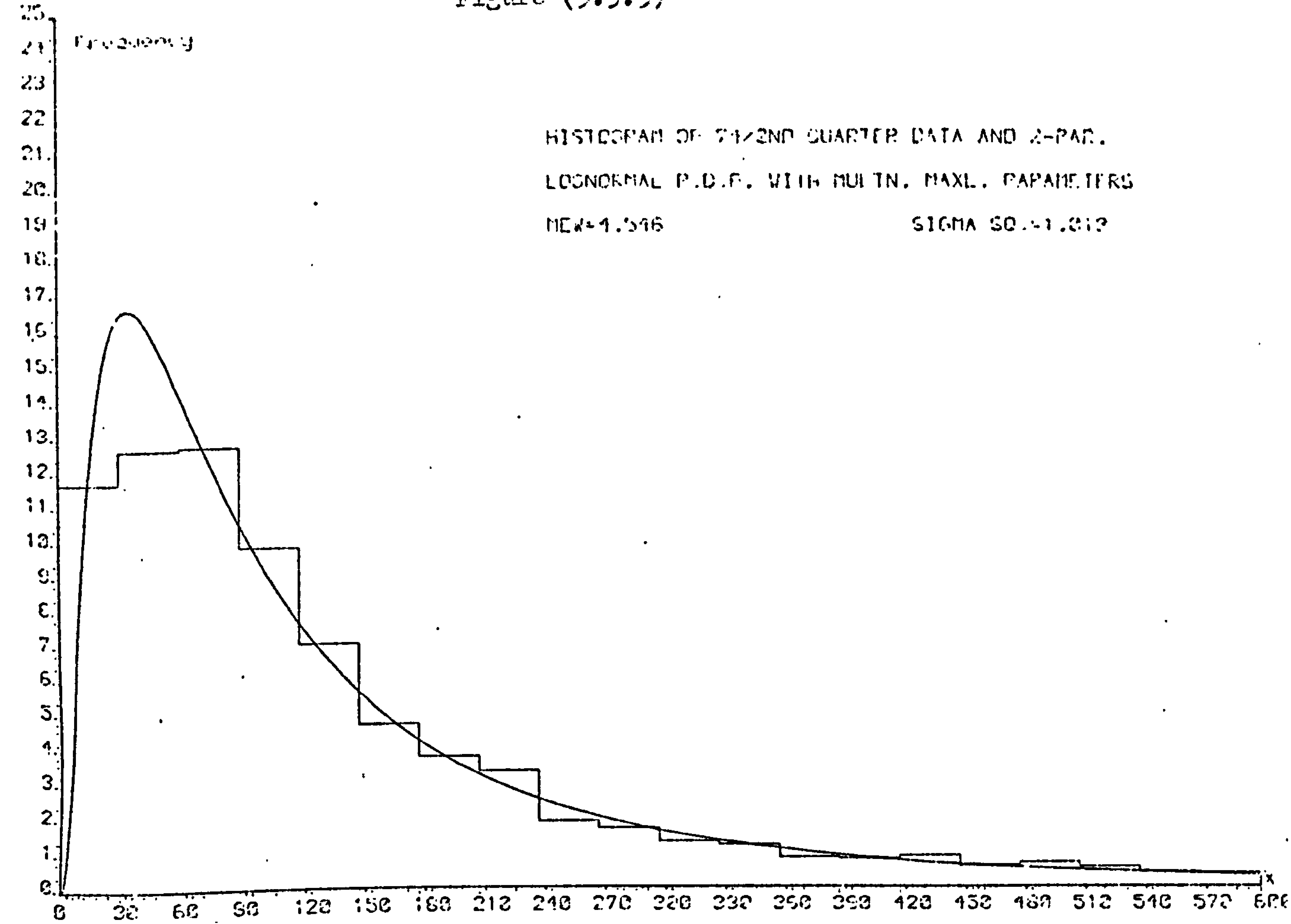

Frequency
HISTOGRAM OF 73/4TH QUARTER DATA AND 2-PAR.
LOGNORMAL P.D.F. WITH MULTN. MAXL. PARAMETERS
MEW=4.516 SIGMA SQ.=1.255
0 30 60 90 120 150 180 210 240 270 300 330 360 390 420 450 480 510 540 570 600
x 10-1
Frequency
HISTOGRAM OF 73/4TH QUARTER DATA AND 2-PAR.
LOGNORMAL P.D.F. WITH MULTN. MAXL. PARAMETERS
MEW=4.516 SIGMA SQ.=1.255
2.50
2.00
.50

Figure (3.5.2)



Frequency
HISTOGRAM OF 71/1ST QUARTER DATA AND 2-PAR.
LOGNORMAL P.D.F. WITH MULTN. MAXL. PARAMETERS
MEW=4.509
SIGMA SQ.=1.057
0 30 60 90 120 150 180 210 240 270 300 330 360 390 420 450 480 510 540 570 600
Frequency
HISTOGRAM OF 71/1ST QUARTER DATA AND 2-PAR.
LOGNORMAL P.D.F. WITH MULTN. MAXL. PARAMETERS
MEW=4.509
SIGMA SQ.=1.257
3.00
2.50
2.00
1.50
1.00
.50
.00

Figure (3.5.3)


Frequency
HISTOGRAM OF 74/2ND QUARTER DATA AND 2-PAR.
LOGNORMAL P.D.F. WITH MULTN. MAXL. PARAMETERS
MEW=4.546
SIGMA SQ.=1.013
x

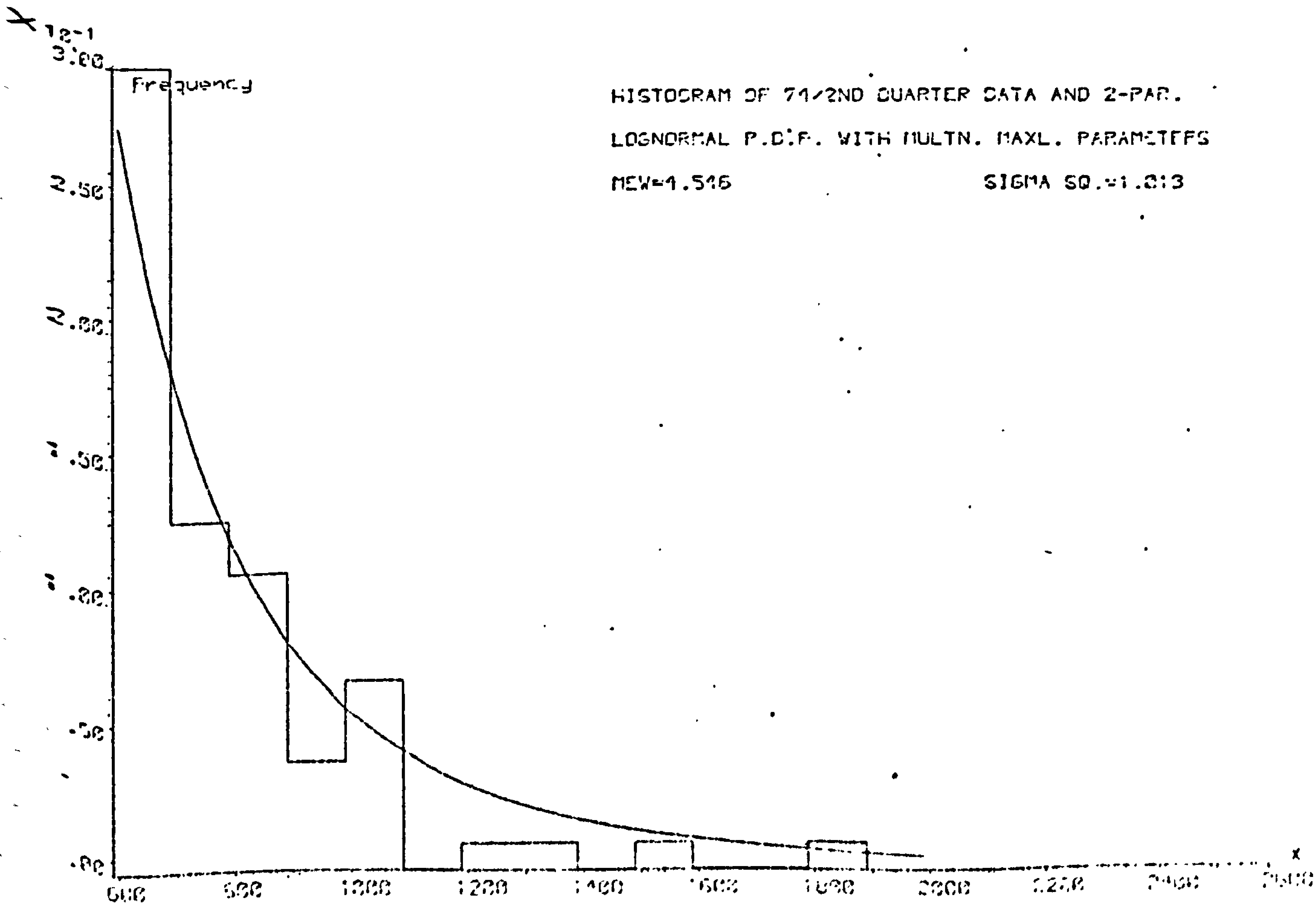
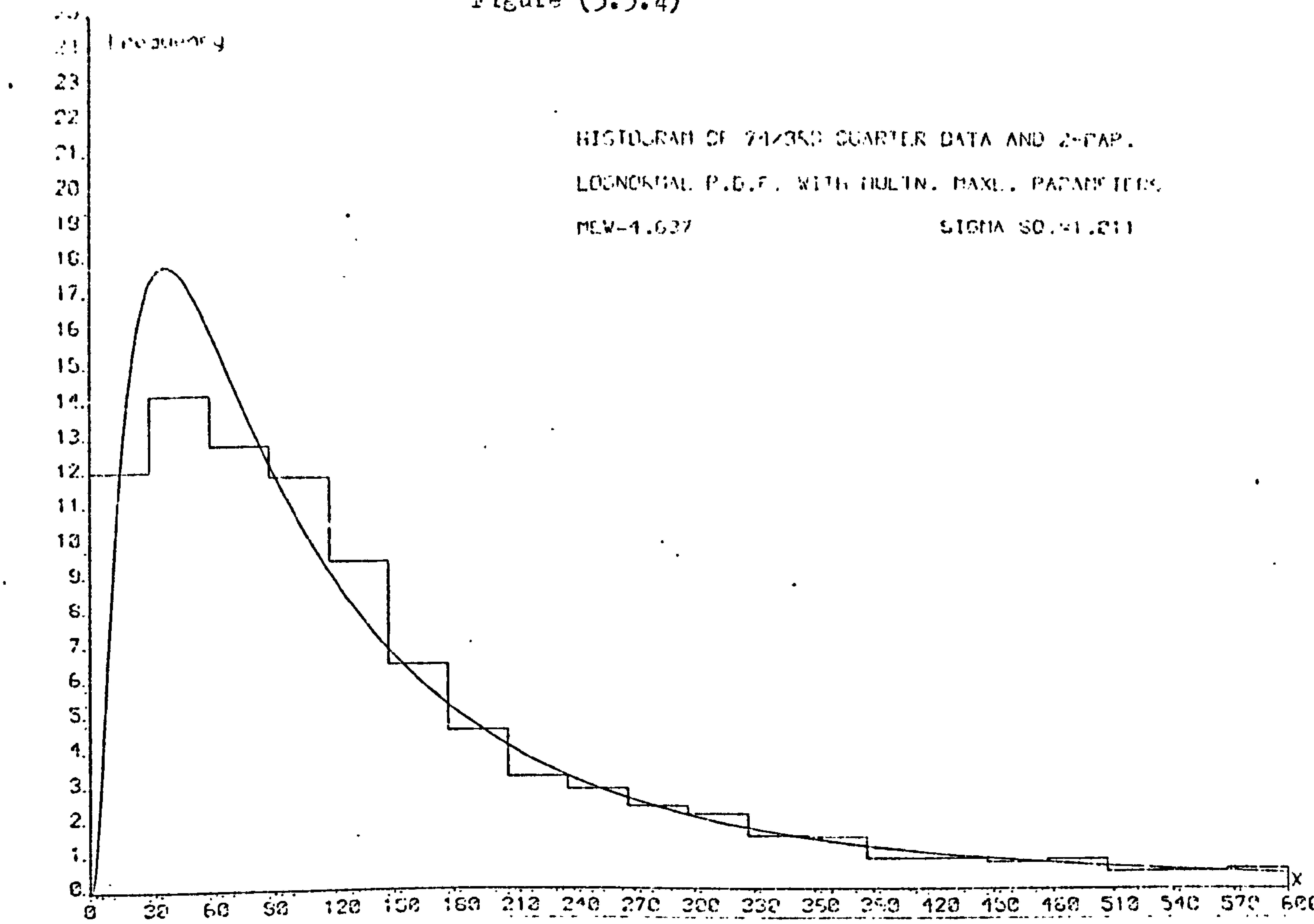

Frequency
HISTOGRAM OF 74/2ND QUARTER DATA AND 2-PAR.
LOGNORMAL P.D.F. WITH MULTN. MAXL. PARAMETERS
MEW=4.546
SIGMA SQ.=1.013
x

Figure (3.5.4)


Frequency
HISTOGRAM OF 74/35D QUARTER DATA AND 2-PAR.
LOGNORMAL P.D.F. WITH MULTN. MAXL. PARAMETERS
MEW=4.637
SIGMA SQ.=1.211
X

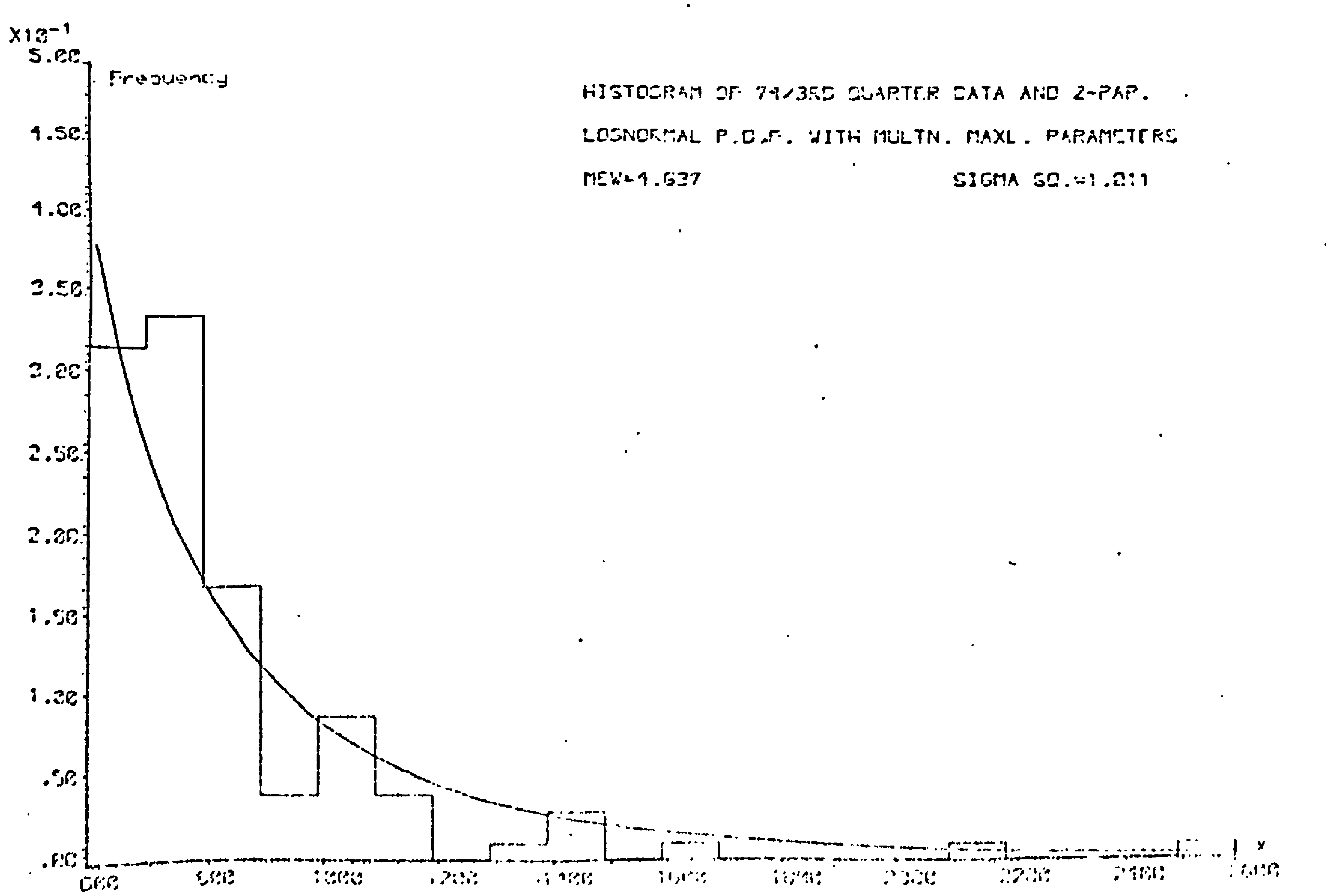
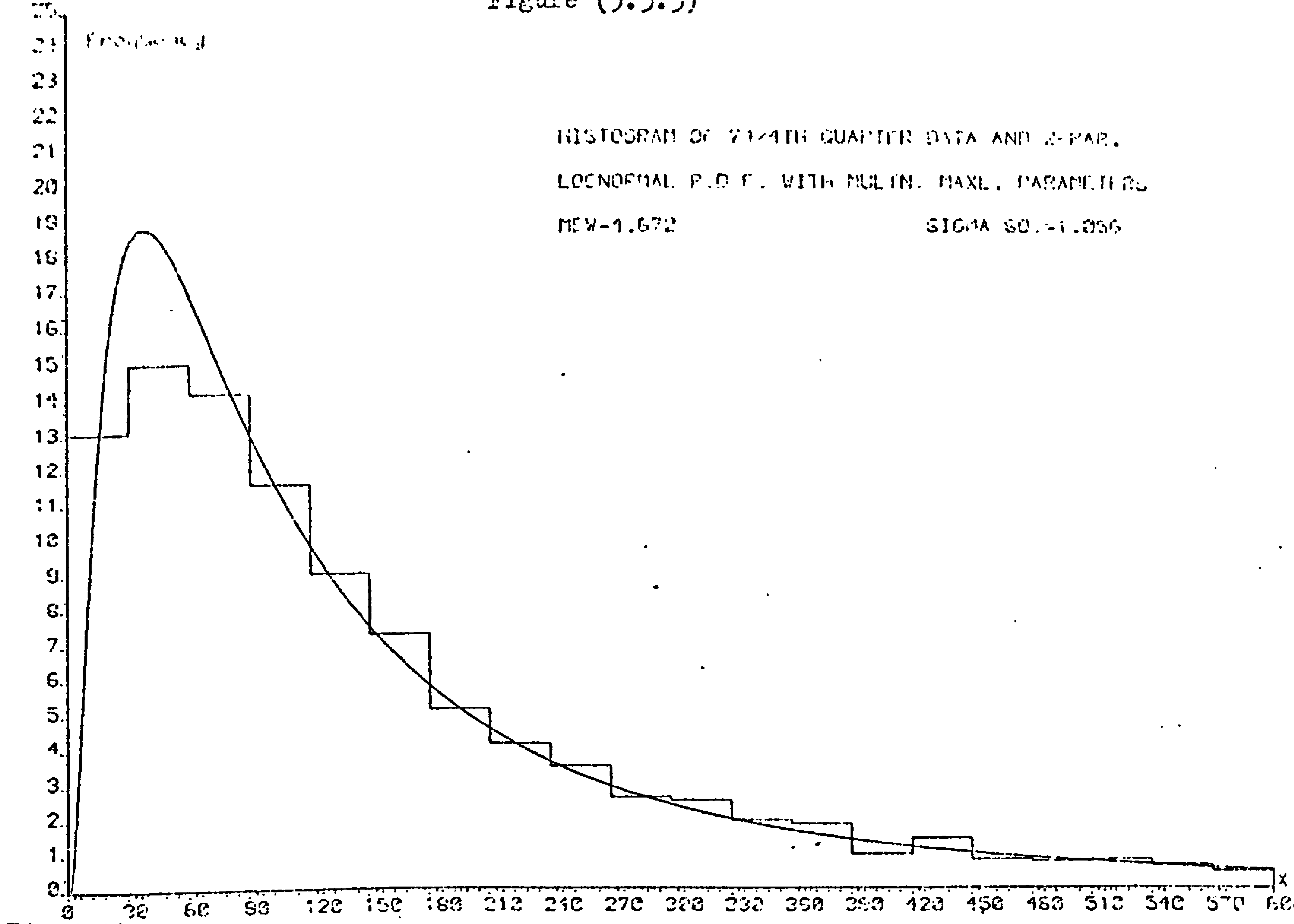
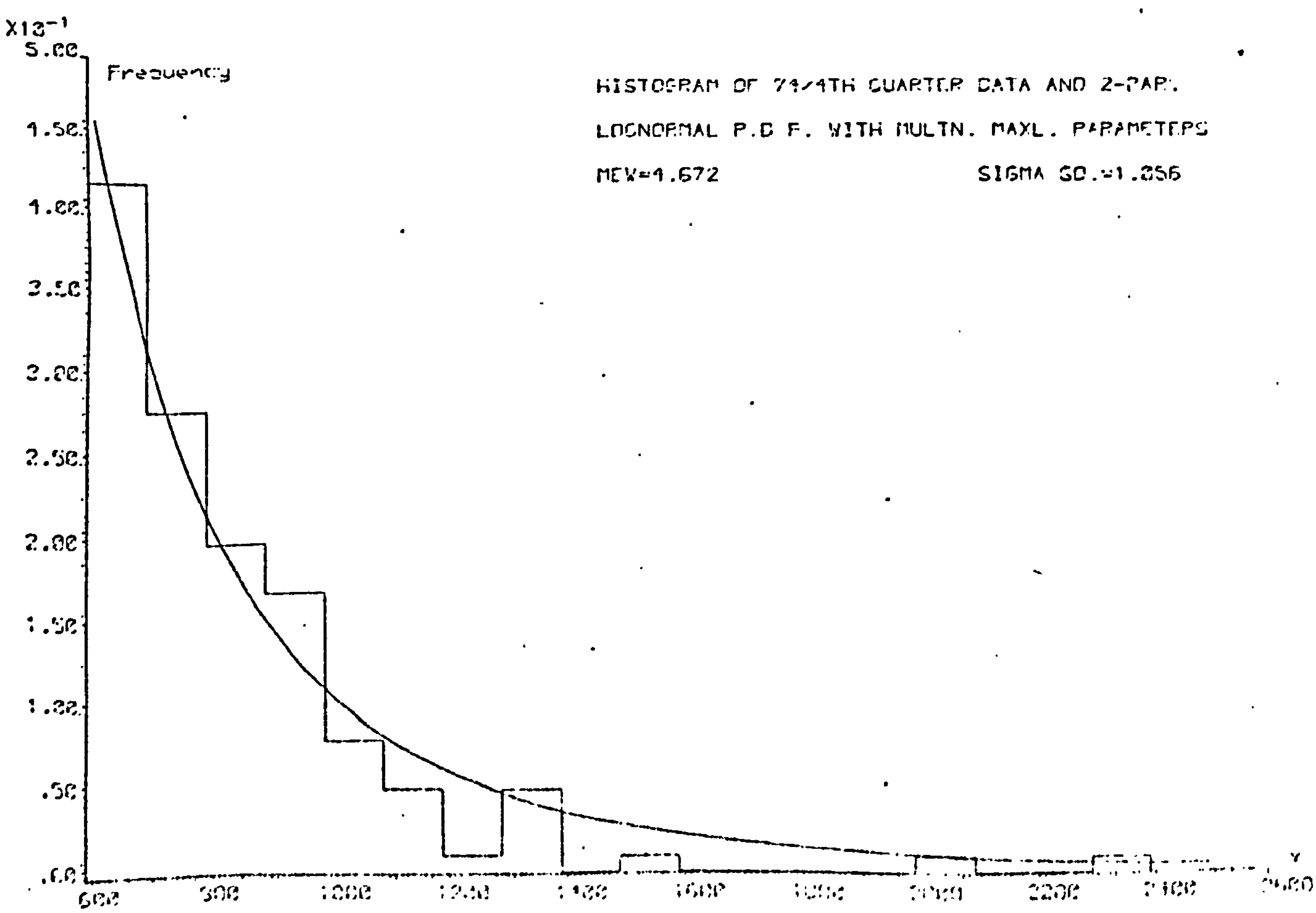

X10^{-1}
Frequency
HISTOGRAM OF 74/35D QUARTER DATA AND 2-PAR.
LOGNORMAL P.D.F. WITH MULTN. MAXL. PARAMETERS
MEW=4.637
SIGMA SQ.=1.211
5.00
4.50
4.00
3.50
3.00
2.50
2.00
1.50
1.00
.50
.00
X

Figure (3.5.5)

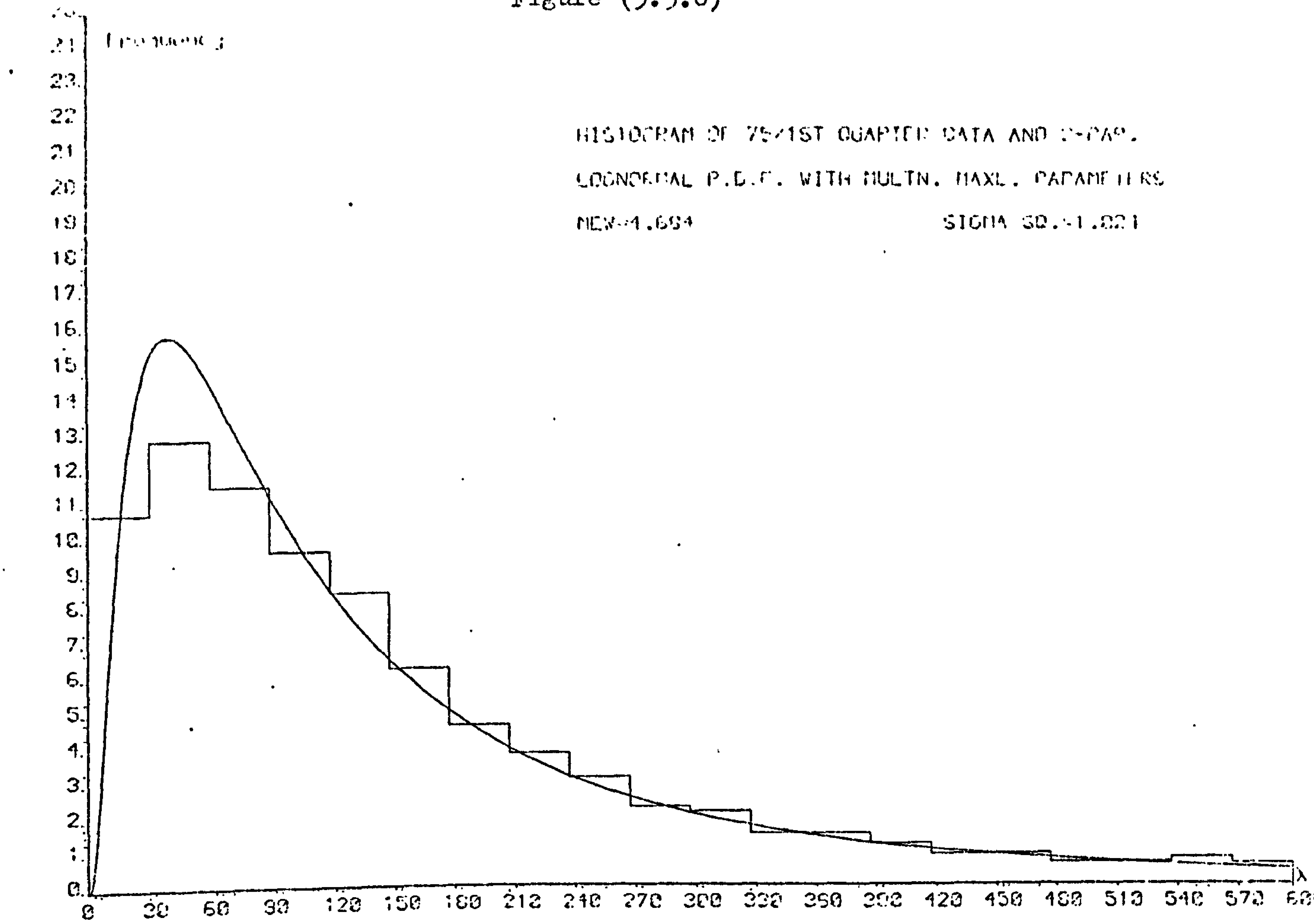


Frequency
MEW-1.672
X
0
30
60
90
120
150
180
210
240
270
300
330
390
420
450
510
540
570



X10⁻¹
5.00
Frequency
HISTOGRAM OF 71/4TH QUARTER DATA AND 2-PAR.
LOGNORMAL P.D F. WITH MULTN. MAXL. PARAMETERS
MEW=4.672
SIGMA SQ.=1.056
4.50
4.00
3.50
3.00
2.50
2.00
1.50
1.00
.50
600
800
1000
1600
2200

Figure (3.5.6)


HISTOGRAM OF 75/1ST QUARTER DATA AND 2-PAR.
LOGNORMAL P.D.F. WITH MULTN. MAXL. PARAMETERS
MEW=4.684
SIGMA SQ.=1.021
Frequency
0
30
60
90
120
150
180
210
240
270
300
330
360
390
420
450
480
510
540
570
600

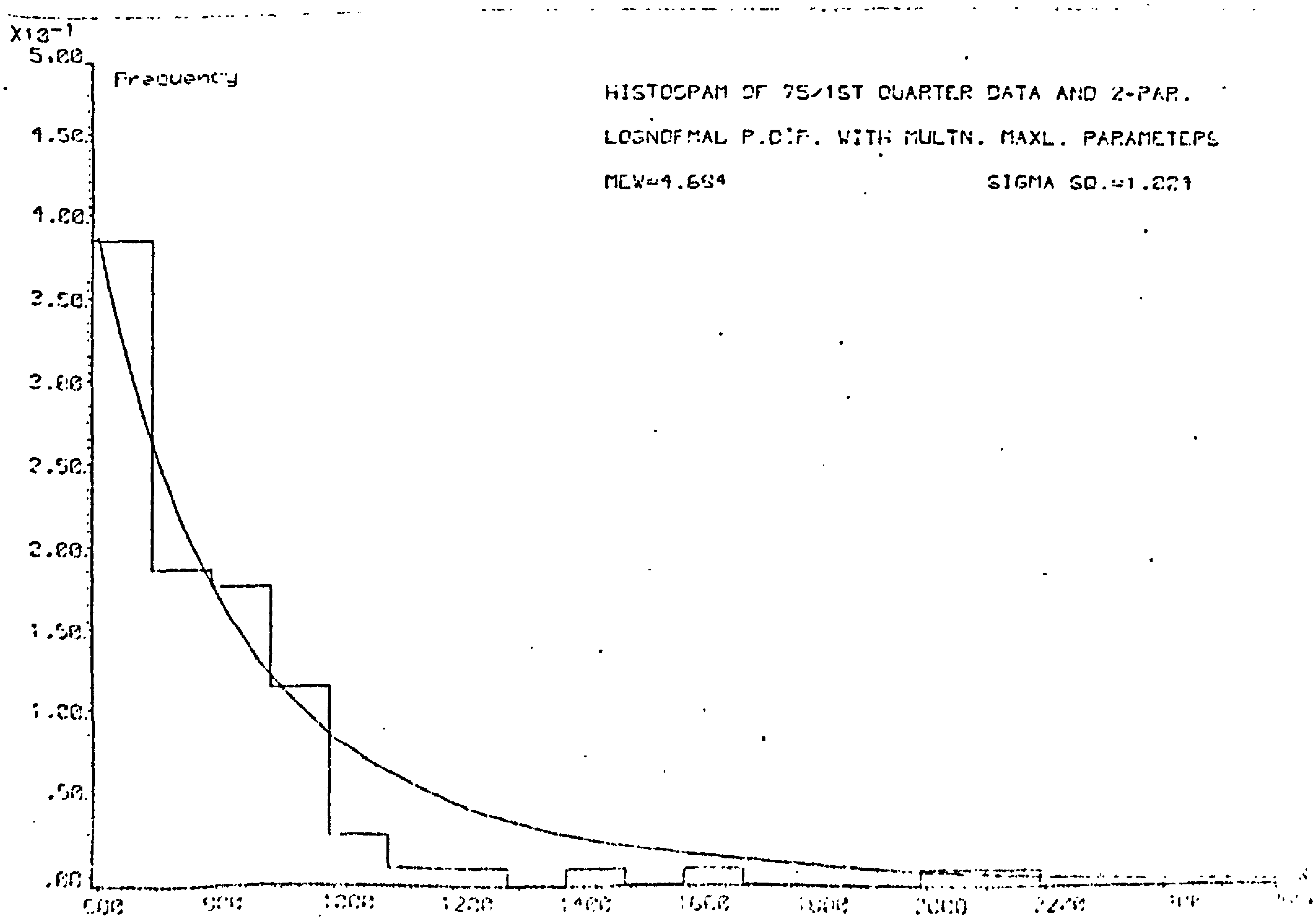
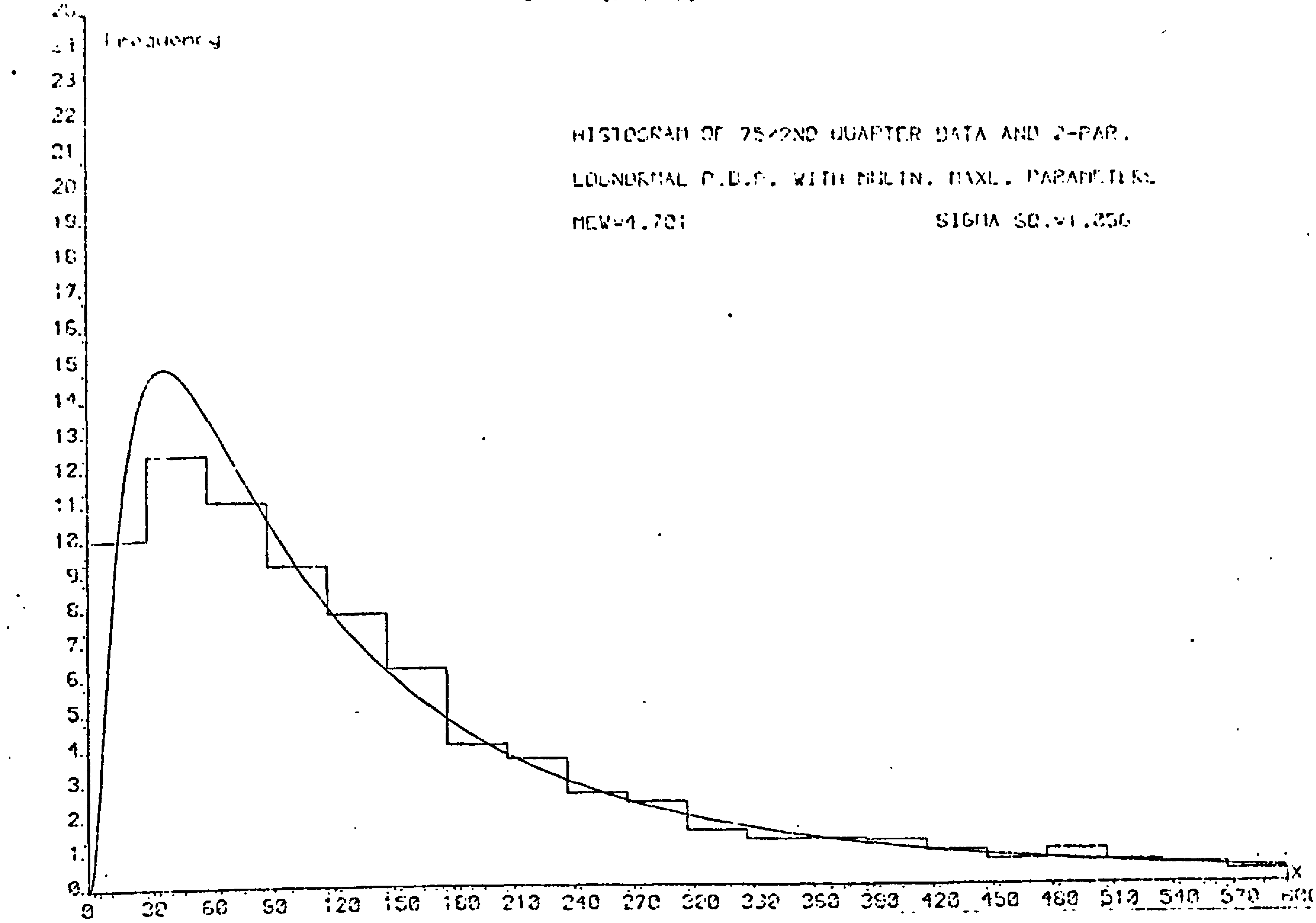
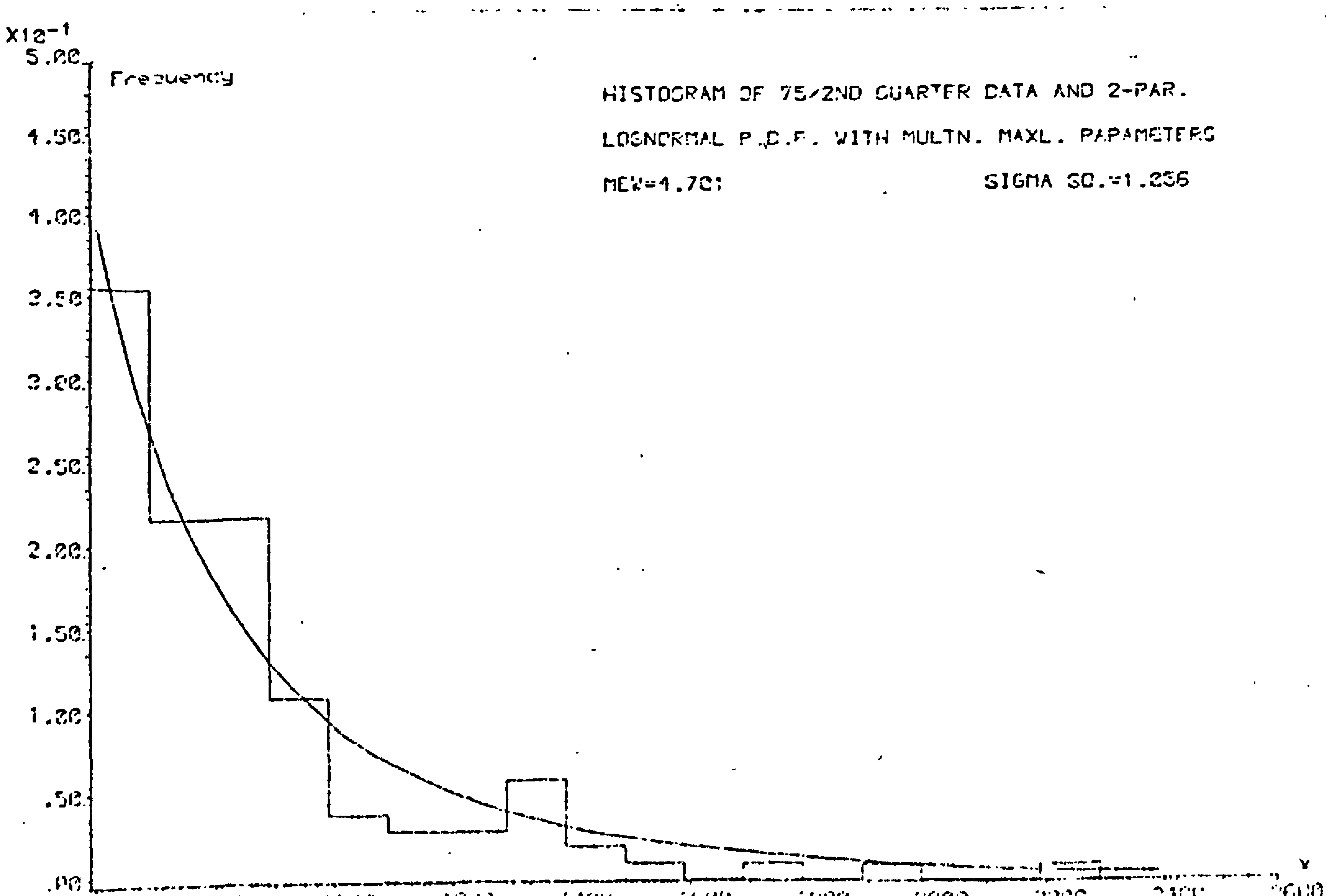

X10⁻¹
Frequency
HISTOGRAM OF 75/1ST QUARTER DATA AND 2-PAR.
LOGNORMAL P.D.F. WITH MULTN. MAXL. PARAMETERS
MEW=4.684
SIGMA SQ.=1.021
5.00
4.50
4.00
3.50
3.00
2.50
2.00
1.50
1.00
.50
.00
2000

Figure (3.5.7)


Frequency
HISTOGRAM OF 75/2ND QUARTER DATA AND 2-PAR.
LOGNORMAL P.D.F. WITH MULTN. MAXL. PARAMETERS
MEW=4.701
SIGMA SQ.=1.256
X
0 30 60 90 120 150 180 210 240 270 300 330 360 390 420 450 480 510 540 570


X10⁻¹
Frequency
HISTOGRAM OF 75/2ND QUARTER DATA AND 2-PAR.
LOGNORMAL P.D.F. WITH MULTN. MAXL. PARAMETERS
MEW=4.701
SIGMA SQ.=1.256
5.00
4.50
4.00
3.50
3.00
2.50
2.00
1.50
1.00
.50
.00
2000 2200

(section 3.8 continued)

The multinomial maximum likelihood estimates have been used to plot the frequency curve. To represent clearly the behaviour of the tail of the distribution, each curve consists of two parts. The graph at the top of the page represents the distribution up to £600 claim amounts, i.e. the lower tail. The bottom graph shows the upper tail of the distribution, i.e. claim amounts greater than £600.
As previous results indicated, the discrepancy between the frequency curve and the histogram for each sample is more marked at the lower tail of the distribution. The possible reasons for this were mentioned earlier. The curve appears to model the histogram satisfactorily at the middle part and the upper tail of the distribution. The distinct mode and the long tail of the histogram are both portrayed by the frequency curve. The graphs show that shifting the curve, slightly, along the x axis (i.e. a 3-parameter lognormal model) or ignoring the lower tail (i.e. a truncated distribution) would produce models which would fit the actual data more closely.

Judging by the results of the foregoing analysis, we are not satisfied that the two-parameter lognormal distribution is the correct model for the distribution of accidental damage claim amounts. However, the analysis has been constructive in showing how the model should be applied in practice. It has also been shown that, for our data, a modification of the model would provide better results. To complete the work on the two-parameter lognormal distribution we will study, in the next section, the effect of inflation on the parameters of the model and will suggest a means of predicting the model of the future distribution of claim amounts.

3.9 Prediction of the Claim Amount Distribution :- The 2-Parameter Lognormal Model

We are interested in the cost of future claims for both the calculation of premium rates and the setting up of reserves. We cannot, therefore, wait until all the claims have been notified and settled in order to find the cost of the claims. One way of solving this problem is by predicting the cost of the future claims. In this section we will suggest a technique by which the distribution of the claim amounts during any future period can be predicted. From this distribution, and assuming that we have a knowledge of the number of claims in the particular period , all the required information about the cost of claims can be obtained. The technique will be tested by using past data. The performance of this method will be examined by goodness-of-fit tests. In particular, the total expected loss statistic, T, will indicate, in monetary terms, how accurately the technique is performing in predicting the total cost of future claims.

3.9.1 The Effects of Inflation on the Parameters of the Model

In practice, it is usually observed that the same kind of claim costs more to settle than it previously did in the past. From tables (1.1) to (1.7) we can see that the mean claim amount has increased from one quarter to the next. It is generally believed that inflation is the main cause of the increase in the cost of claims over time. The increase may be due to other apparent factors the real roots of which lie in inflation. For instance, in the third party liability claims, the decision of the courts in awarding higher compensations stems from the rise in the cost of living and higher wages.

Let us assume that the effect of inflation is to increase a claim of size X to X(1 + i) after a period of time where i is the effective rate of inflation during that period. If we assume that the distribution of X is the two-parameter lognormal with parameters (μ,σ^2), by a transformation of variables it can be shown that X(1 + i) is distributed as the two-parameter lognormal with parameters $(\mu+\delta,\sigma^2)$ where we call $\delta = \log(1 + i)$ the force of inflation. Therefore, inflation affects only μ and leaves σ^2 unchanged. If we look at the multinomial maximum likelihood estimates of μ, given in tables (3.11) to (3.17), we notice that they have generally increased from one period to the next. The mean and the standard deviation of the lognormal distribution, α and β respectively, are functions of both μ and σ^2 (see section 3.3) and hence inflation increases both of them over time. After a period of time in which i is the effective rate of inflation, α and β will be increased to $\alpha(1 + i)$ and $\beta(1 + i)$ respectively. Tables (3.11) to (3.17) show that the mean and the standard deviation of the claim amounts have increased from one period to the next.

3.9.2 The Prediction Technique

For our accidental damage claims, investigation shows that nearly 70% of the total cost of claims are reported and settled within the actual period (quarter) of accident. Another 25% are settled by the end of the second quarter. Therefore, it is reasonable to assume that all AD claims are settled in the middle of the period of accident. Hence, if the distribution of claim amounts during a period A, for instance, is $LN(x;\ \mu,\sigma^2)$, their distribution during a future period B will be $LN(x;\ \mu+\delta,\sigma^2)$ where it is assumed that the rate of inflation from the middle of period A to the middle of period B is i and that $\delta = \log(1 + i)$. If sizeable parts of the total cost of claims are settled over various periods and, furthermore, if they take a long time after the end of the period of accident to be settled (e.g. for the third party liability claims), then the rate of settlement must be taken into consideration. A combined rate of inflation and settlement will be required for prediction purposes.

Other factors such as seasonality may be affecting the number and amount of claims during different periods of the year. For instance, in autumn and winter, due to the bad weather conditions, it is likely to have more accidents and more claims for larger amounts. The data of tables (1.1) to (1.7) show that the number of claims for the 4th quarter of the year is greater than for the other quarters. To nullify the effects of such factors we suggest that each period of accident should be used to predict the distribution of claim amounts during the same period but in a future year(s). For instance, the model for the first quarter of 1974 should be used to predict that for the first quarter of 1975 , etc.

We can use our available AD data to test this technique for three different periods of accident, namely, we can use :

73/4th quarter to predict 74/4th quarter,

74/1st quarter to predict 75/1st quarter,

and 74/2nd quarter to predict 75/2nd quarter.

The question remains as to what should be the appropriate rate of inflation. It would be advantageous if we could show that changes in the parameters of the model, over time, correspond to changes in the value of a particular index of prices or wages. We shall consider three such indices

1 - The General Index of Retail Prices.

2 - The Index of Average Earnings, Miscellaneous Services.

3 - The Index of Motor Vehicles Repair Costs.

Indices 1 and 2 are published in the "Monthly Digest of Statistics" which is a publication of the Central Statistical Office. Two of the constituents of index 2 are the wages of motor repairers and garages.

Index 3 is based on the experience of the Royal Insurance Company and is published by the Economic Advisory Group which is part of the British Insurance Association. The value of this index at the beginning of every quarter of the year is given. By taking the average of the index values at the beginning and at the end of every quarter we can find the average index for that particular quarter. For indices 1 and 2 we take the average of the monthly values as the value of the index for each quarter. The percentage change in the value of an index from one quarter to a future quarter indicates the rate of inflation to be used for prediction purposes. After we have predicted the distribution of claim amounts for one of the quarters mentioned above we can perform goodness-of-fit tests between the predicted and actual distributions. A consistently good fit for different quarters will not only show the success of the prediction technique but will also indicate the particular index which should be used when predicting the distribution of accidental damage claim amounts. Naturally, we have to use a forecast value of that index for periods in the future.

3.9.3 Prediction for the AD Data

To test the prediction technique, for the two-parameter lognormal model, we wrote computer program P8. We considered it important to see, first of all, what the predicted distribution would be like if inflation was ignored. For the three quarters mentioned in the previous section, the program was used with $i = 0$. The results are presented in tables (3.18) to (3.20). Let us look at table (3.18) for 74/4th quarter. The program has produced a comprehensive table. The details of the predicted quarter, the quarter used for prediction and its parameters as well as the rate of inflation and the index used for its calculation are provided.

The prediction parameters are computed and given as well as the actual ones. The means and the standard deviations of the predicted and actual claim amount distributions are also given. The output format for the rest of the table is similar to tables (3.10) to (3.17) and the column headings are self-explanatory. From tables (3.18) to (3.20) it is clear that when inflation is ignored parameter μ is underestimated and the predicted mean and standard deviation of the claim amount are much smaller than the actual ones. Therefore, we should expect a major disagreement between the actual and the predicted distributions. This is shown by the large values of the Chi-square and total expected loss statistics. The large positive values of the total expected loss, T, indicate that the distribution is heavily underpredicted when inflation is ignored. Tables (3.21) to (3.23) show the results when inflation, according to the Index of Motor Vehicles Repair Costs, is taken into consideration. The results are now better, but still not quite satisfactory, as the large values of the Chi-square and T statistics indicate. Here, the negative values of T show that the distribution is over-predicted. Next we used the Index of Average Earnings. The results are given in tables (3.24) to (3.26). They show an improvement over the previous index. The differences between the predicted and actual parameters are smaller than before but the Chi-square statistics still show significant differences between the predicted and actual distributions. The absolute values of T are smaller but still show a large over-prediction. For the General Index of Retail Prices, the results are presented in tables (3.27) to (3.29). This index has produced the most satisfactory results amongst the indices considered. The prediction and actual parameters are close to each other and have produced the smallest values for the Chi-square and T statistics. The Chi-square statistics still indicate significant differences between the predicted and actual distributions.

The total expected loss statistics, T , still indicate over-prediction. If we calculate the values of the Kolmogorov-Smirnov statistic, D , from each of the tables (3.27) to (3.29) and compare them with the critical values of D in table (2.1), we shall find that respectively $p \doteqdot 0.18$, $p \doteqdot 0.15$ and $p \doteqdot 0.02$ (see chapter 2, section 2.8). Therefore, on the basis of this test, the differences between the predicted and the actual distributions are not significant for the 74/4th and 75/1st quarters but are significant for the 75/2nd quarter. The large values of the Chi-square statistics are due to large contributions from one or two lower tail intervals. If it had not been for these intervals, the Chi-square values would have been satisfactory too. This again indicates that we should consider a modification of the model. The prediction technique performs satisfactorily, especially when used in conjunction with the General Index of Retail Prices.

All the indices considered resulted in over-prediction. In predicting the claims cost, and hence in reserving, it is better to over-predict and be safe rather than under-predict and end up in a loss. However, there must be a balance between over-prediction and tying up too much funds as reserves. Therefore, we would prefer the General Index of Retail Prices to the Index of Motor Vehicles Repair Costs. Besides, the former results in closer agreement between the predicted and the actual distributions. This indicates that changes, over time, in the amounts of accidental damage claims correspond to changes in the General Index of Retail Prices. The major cost of any repair, in particular for serious accidents and hence for large claims, is for the replacement of the damaged parts rather than for the labour charge. Therefore, it seems reasonable that accidental damage claim amounts should be subject to inflation according to an index of prices rather than wages. The inflation rates calculated from the Index of Motor Vehicles Repair Costs are too high. This index has been calculated from the experience of just one company and we can see that it does not agree with the experience of the company which has provided our data. One possible reason could be that the former company may be insuring more expensive or luxurious cars. In the future stages of the present work we shall, therefore, only use the General Index of Retail Prices in our predictions. Before concluding this section it is not out of place to comment on the importance of our proposed total expected loss statistic, T . Table (3.27), for example, shows that we have predicted the total claims cost with a difference (or an error) of £16,844 which is 3.16% of the total actual cost. This is easily understood in a way that, say, the Chi-square or the Kolmogorov-Smirnov statistics are not. It is thus that we recommend the use of this statistic in actuarial work.

3.10 The 3-Parameter Lognormal Model

Our accidental damage data are from a portfolio of comprehensively insured private motor cars. There is a voluntary excess, of amount c (where $c \geq 0$), on these policies. For every claim of amount $X + c$, the insured pays the amount c and the insurance company pays the amount X. Therefore, it is reasonable to assume that the logarithm of the total amount of claim, i.e. $\log(X + c)$, and not just $\log X$, is distributed normally.

In a portfolio, c may have several different values. Our data consists of payments by the insurance company only and the value of c for every claim is not known. We assume that c is the same for all policies and treat it as an unknown parameter. Hence, we are faced with a 3-parameter lognormal distribution. This was defined in section 3.2 and some of its properties were given in section 3.4.

3.11 Test of Lognormality:- The 3-Parameter Case

3.11.1 The Graphical Test

It is useful to have a procedure for testing whether a given sample of data is from a 3-parameter lognormal population. Aitchison and Brown (1957) mention a graphical method of roughly estimating the unknown location parameter c. We suggest using this method as a test of 3-parameter lognormality.

From equation (3.4-6) we know that if the random variable X is distributed as $LN(x; -c, \mu, \sigma^2)$ then the locus of the points $(\log(x + c), z)$ is the straight line

$$z = \frac{1}{\sigma}\log(x + c) - \frac{\mu}{\sigma}$$

Let us assume that we are given a sample of observations from a 3-parameter lognormal distribution. If the value of c was known, then the set of points $(\log(x + c), z)$, calculated from the sample, should lie almost on a straight line (line 1 in figure (3.6)). If c is unknown, and we

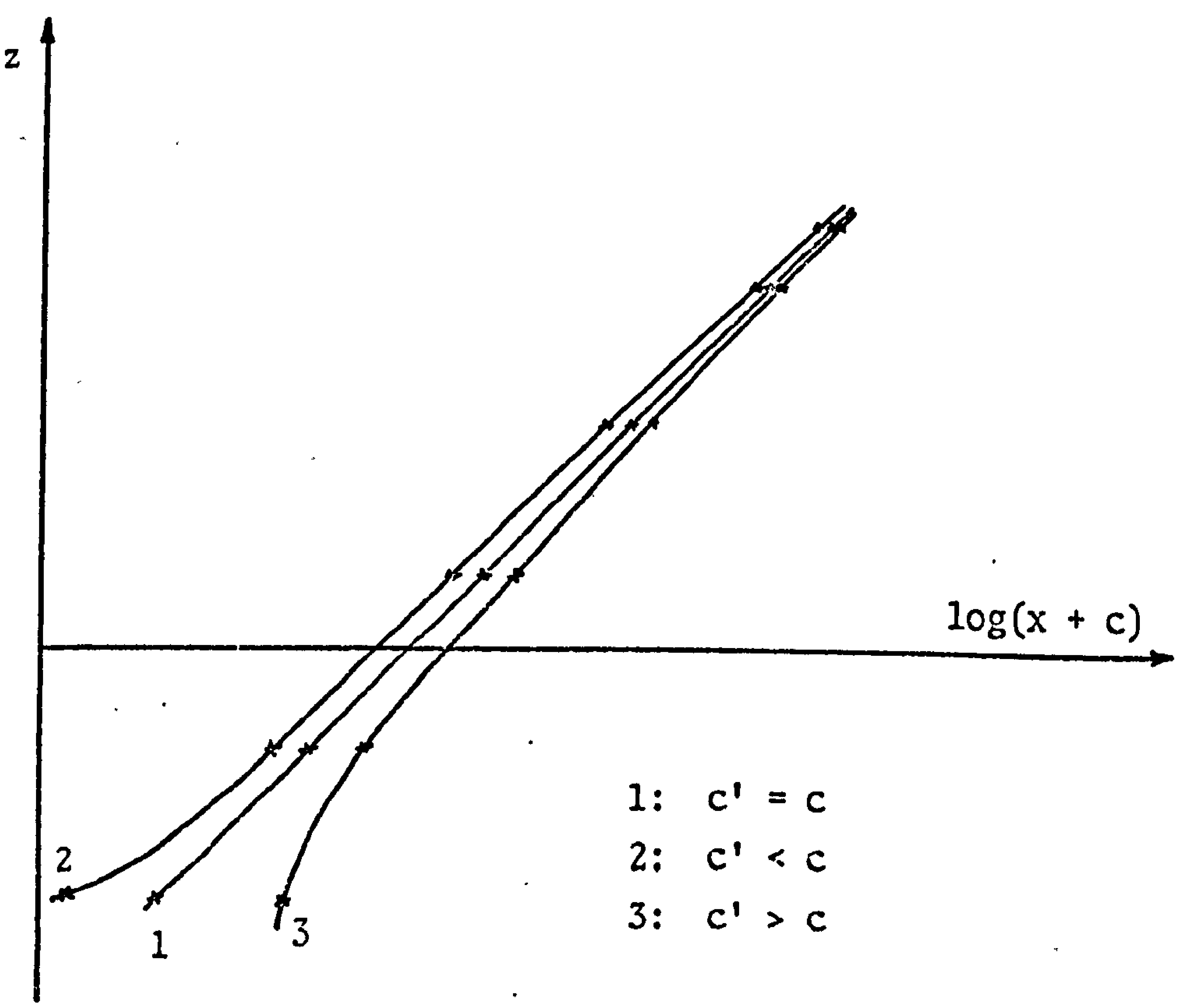


Figure (3.6) - Straight line plot for a sample from the 3-parameter lognormal distribution underestimate it as c' (i.e., use $c' < c$) then the points will show a marked curvature in the lower tail values of x (curve 2 in figure (3.6)). If we over-estimate c (i.e. use $c' > c$) then the curvature would be in the opposite direction (curve 3, figure (3.6)).

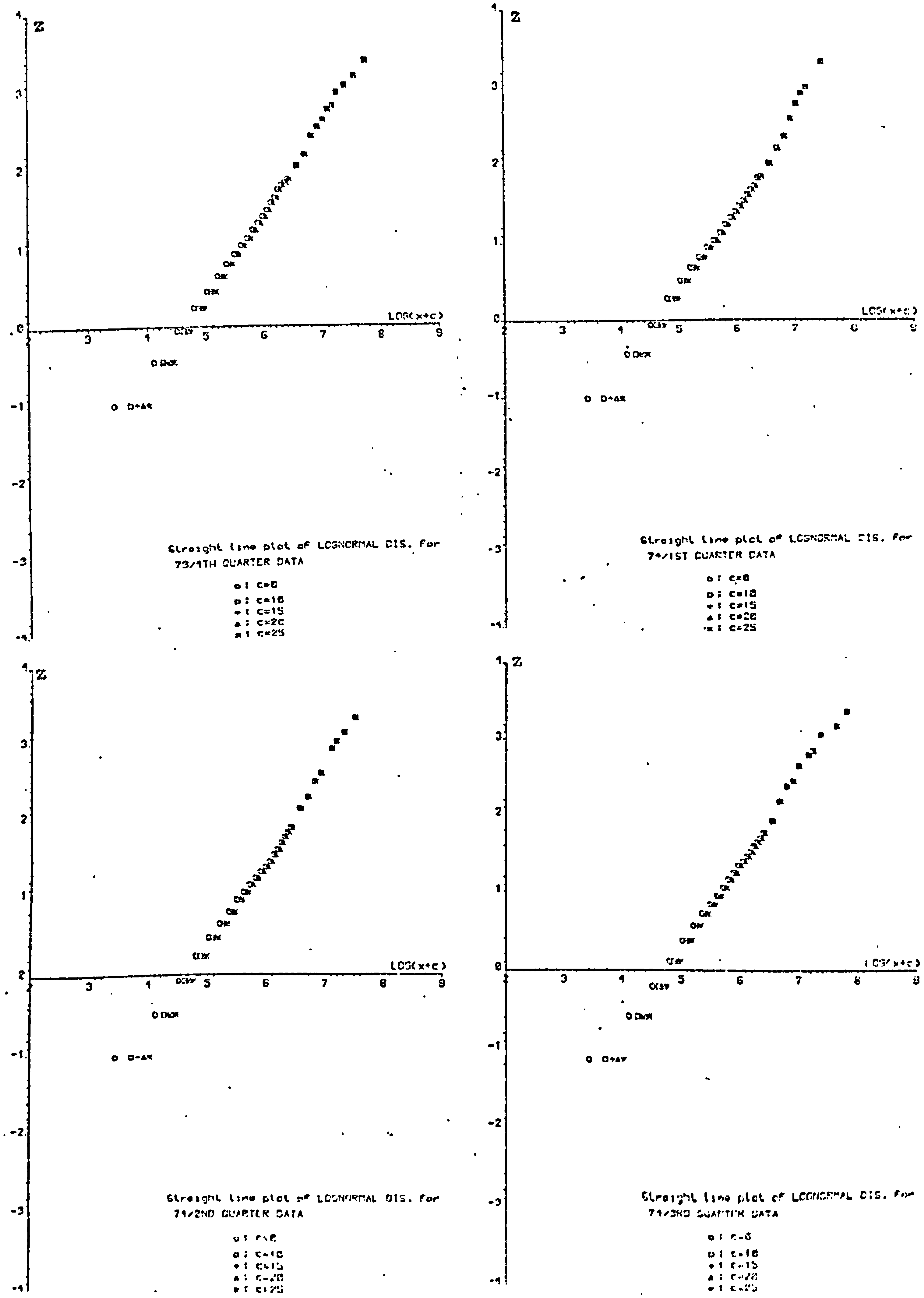
Therefore, as a test of 3-parameter lognormality for a given sample of data, we can initially plot the array of points (log x,z). If the sample is from a 3-parameter lognormal distribution, its graph will resemble curves 2 or 3 as above. According to the shape of the graph we choose a value of c in the required direction towards the straight line. We plot the array of points ($\log(x + c)$, z) and, judging by the resulting graph, accept c or modify it again in the required direction. If a value c is found which makes the points ($\log(x + c)$,z) lie approximately on a straight line, we will be satisfied that the sample is from a 3-parameter lognormal distribution. This c can be considered as an estimate of the unknown location parameter of the distribution. By using this estimate we can then, as in the 2-parameter case, find estimates of μ and σ^2.

3.11.2 Testing the AD Data for 3-Parameter Lognormality

Figures (3.1-a) and (3.1-b) showed that in the seven AD samples, the points (log x,z) deviated markedly from the straight line for the smaller values of x. We can now look at these plots and suggest that the correct model is the 3-parameter lognormal. They all show that $\hat{c} = 0$ has underestimated c and that some $\hat{c} > 0$ should be found to make the points tend towards the straight line.

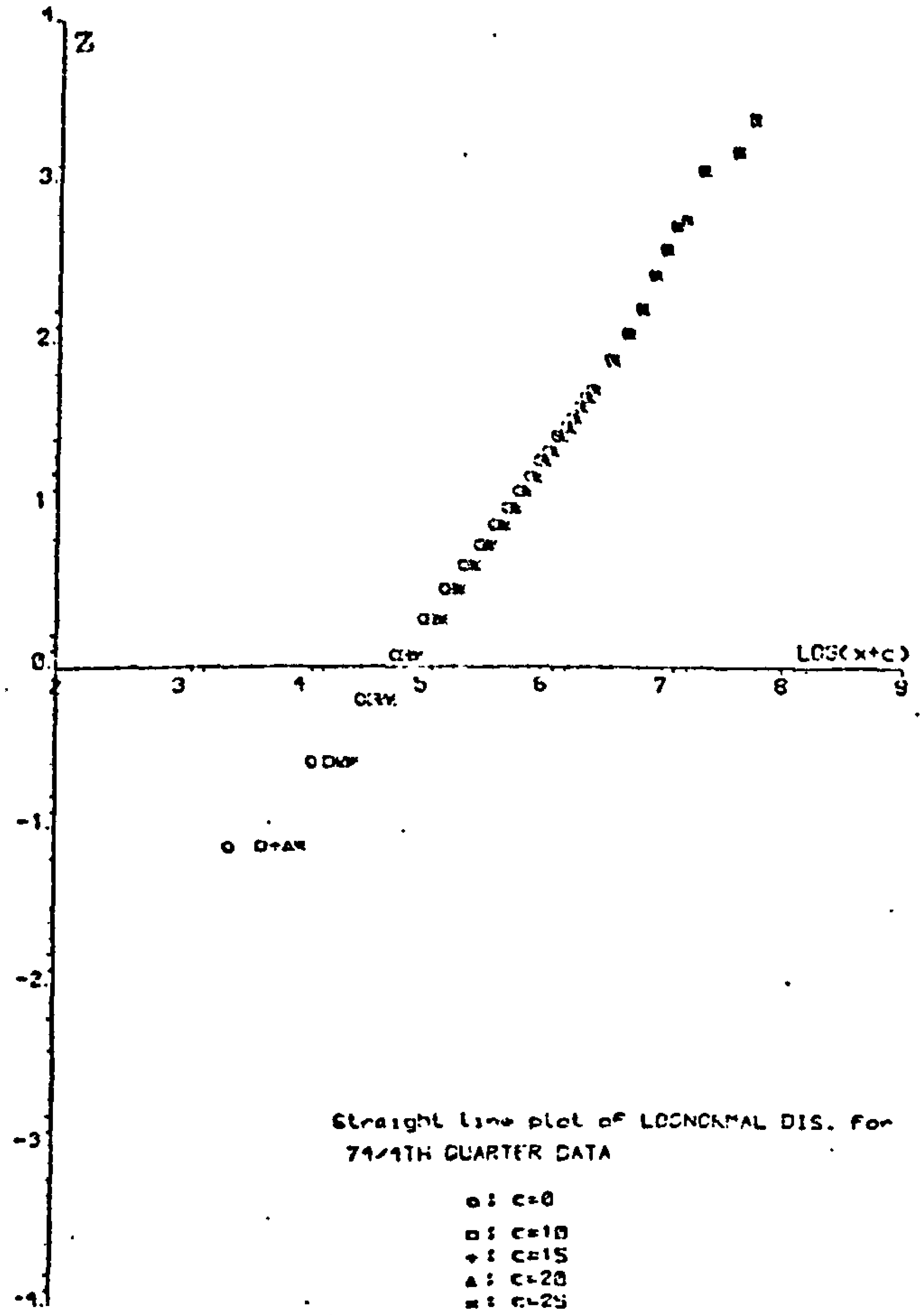
We wrote computer program P9 to plot the array of points ($\log(x + c)$, z) for values of $c = 0, 10, 15, 20, 25$. To find these points from the sample the procedure of section 3.6.1 and formula (3.7-27) were used. The plots for the seven samples of data are presented in figures (3.7-a)

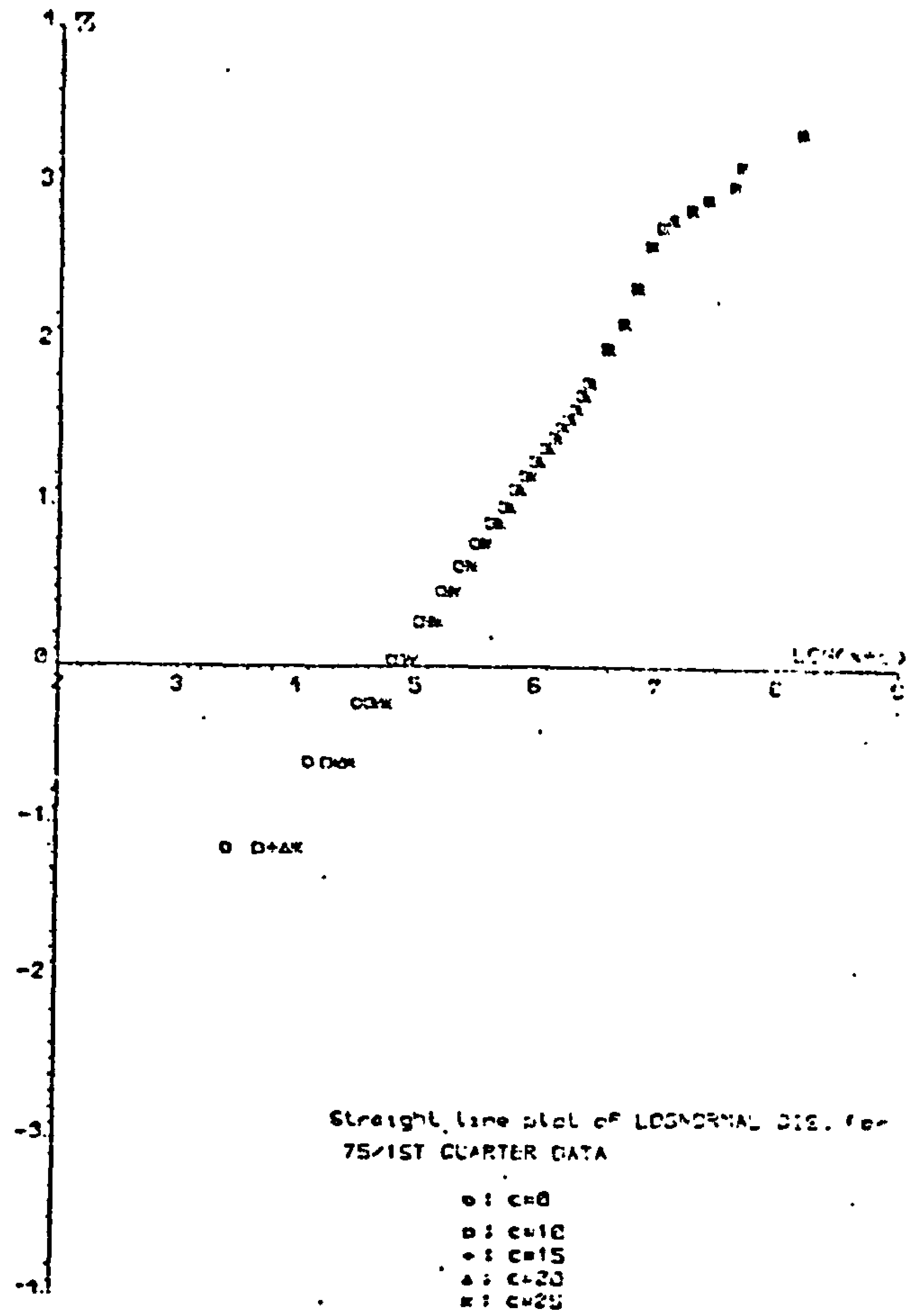
Figure (3.7-a)

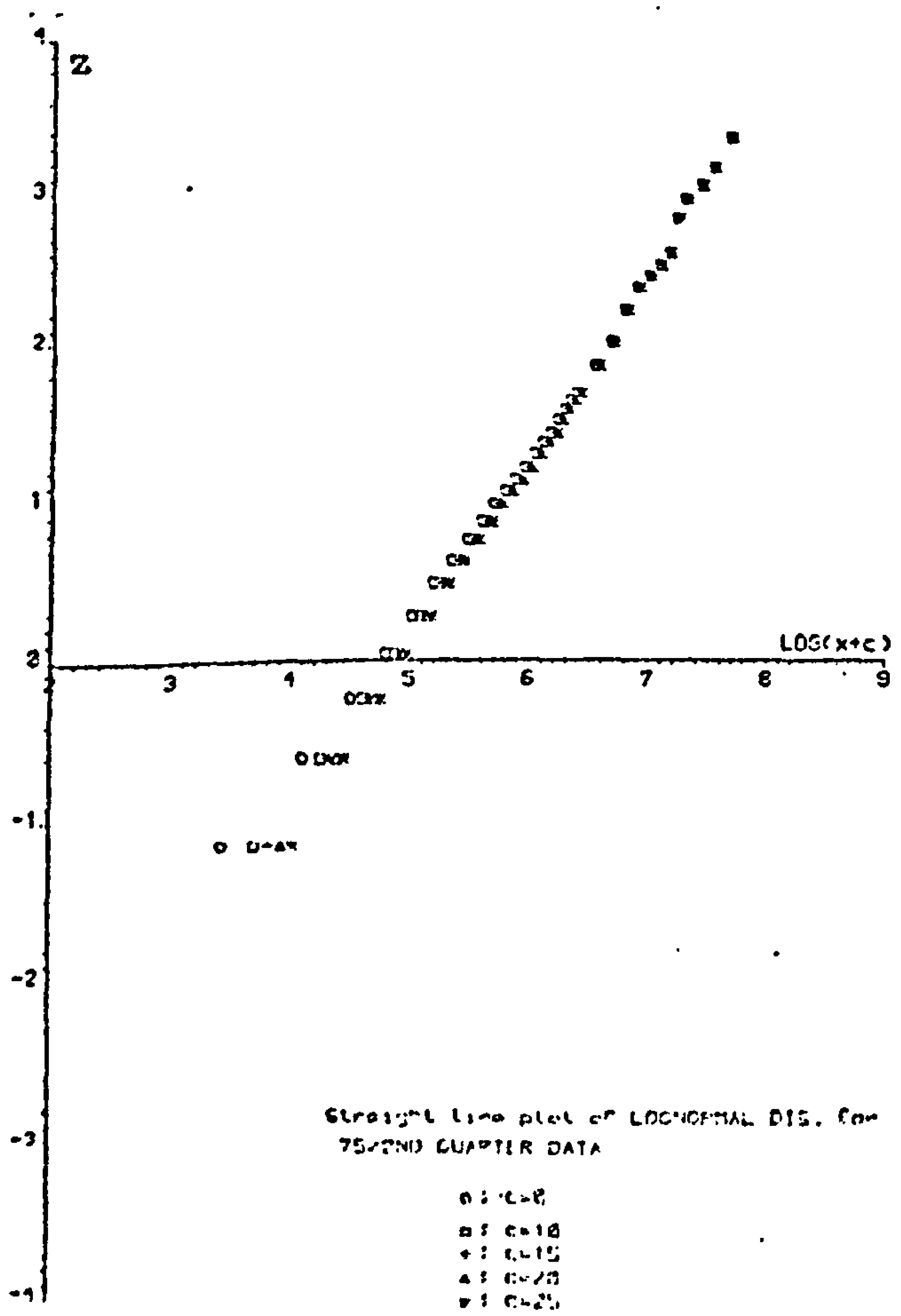


Z
LOG(x+c)
Straight line plot of LOGNORMAL DIS. for
73/4TH QUARTER DATA
o : c=0
□ : c=10
+ : c=15
▲ : c=20
* : c=25
Z
LOG(x+c)
Straight line plot of LOGNORMAL DIS. for
74/1ST QUARTER DATA
o : c=0
□ : c=10
+ : c=15
▲ : c=20
* : c=25
Z
LOG(x+c)
Straight line plot of LOGNORMAL DIS. for
74/2ND QUARTER DATA
o : c=0
□ : c=10
+ : c=15
▲ : c=20
* : c=25
Z
LOG(x+c)
Straight line plot of LOGNORMAL DIS. for
74/3RD QUARTER DATA
o : c=0
□ : c=10
+ : c=15
▲ : c=20
* : c=25

Figure (3.7-b)


Z
LOG(x+c)
Straight line plot of LOGNORMAL DIS. for 74/4TH QUARTER DATA
o: c=0
□: c=10
+: c=15
▲: c=20
*: c=25


Z
Straight line plot of LOGNORMAL DIS. for 75/1ST QUARTER DATA
o: c=0
□: c=10
+: c=15
▲: c=20
*: c=25


Z
LOG(x+c)
Straight line plot of LOGNORMAL DIS. for 75/2ND QUARTER DATA

and (3.7-b). It can be seen that the directions of the curvatures for c = 0 and 10 are opposite to that for c = 25. For c = 15 and 20 the points appear to lie more accurately on a straight line. Therefore, we are satisfied that our samples are from 3-parameter lognormal distributions whose location parameters c have values between 10 to 20.

3.12 Estimation of the Parameters of the 3-Parameter Lognormal Distribution

The existence of the unknown location parameter c, in addition to μ and σ^2, makes the estimation problem more difficult. Aitchison and Brown (1957) do not deal with this problem for grouped data and only examine the case when values of the individual observations in the sample are given. They report methods of moments and quantiles where three moments or quantiles, respectively, of the sample are put equal to their corresponding population values. The method of maximum likelihood is also considered which results in a non-linear equation in c. They also report Kemsley's (1952) method which consists of equating the mean and two quantiles of the sample to their corresponding population values. We can make assumptions, similar to the ones we made for the two-parameter case, and modify the above methods for use with grouped data (see section 3.7). However, as the analysis in the two-parameter case showed, these methods would not provide very satisfactory results. Their solutions may be of use, for example, in yielding initial values for the parameters to be used in an iterative method of estimation. Therefore, we will not consider these methods here.

Johnson and Kotz (1970) report the methods mentioned in Aitchison and Brown (1957) and they do not consider estimation from grouped data. In actuarial literature, Ferrara (1971) uses the method of quantiles to determine $\hat{c}$ first and then finds $\hat{\mu}$ and $\hat{\sigma}^2$ by least squares regression (as in the 2-parameter case).

In this section, we will examine two methods which are directly suitable for estimation from grouped data. The first one is the least squares regression method. We will suggest a procedure for finding the parameters which avoids solving non-linear equations. The second method is the multinomial maximum likelihood which produced most satisfactory results in the two-parameter case. This method will be modified and used for the 3-parameter case. As we mentioned in section 3.7.7, Tallis and Young (1962) consider this method but they suggest equating the partial derivatives of the loglikelihood function, with respect to the parameters, to zero and solving iteratively the resulting set of non-linear equations. This is more laborious than our proposed procedure.

3.12.1 The Method of Least Squares

Let us assume that a sample of grouped data as defined in section 3.7 is available. From section 3.4 we know that for the $LN(x; -c, \mu, \sigma^2)$ distribution relationship (3.4-6), i.e.,

$$z_q = \frac{1}{\sigma}\log(x_q + c) - \frac{\mu}{\sigma}$$

holds. If the random variable X (the claim amount) is assumed to be distributed as $LN(x; -c, \mu, \sigma^2)$, we can use the least squares regression technique to find the estimates of the parameters. This consists of minimizing SSD (the sum of the squares of deviations) simultaneously with respect to c, μ and σ where,

$$SSD = \sum_{i=1}^{k-1} [z_i - \frac{1}{\sigma}\log(x_i + c) + \frac{\mu}{\sigma}]^2 \qquad (3.12\text{-}1)$$

One way to solve this problem would be to equate the partial derivatives of SSD, with respect to c, μ and σ, to zero. But this will result in a system of non-linear equations in the parameters which has to be laboriously solved by an iterative technique. We, however, suggest a different method. Instead of searching for the set of parameters at which SSD attains its minimum, we propose finding the minimum value of SSD and taking its corresponding estimates of the parameters ($\hat{c}$, $\hat{\mu}$, $\hat{\sigma}$) as the required solution. Let us explain this method in more detail.

If our data is from a 3-parameter lognormal distribution, then there exists a c at which SSD attains its minimum - see figure (*) below.

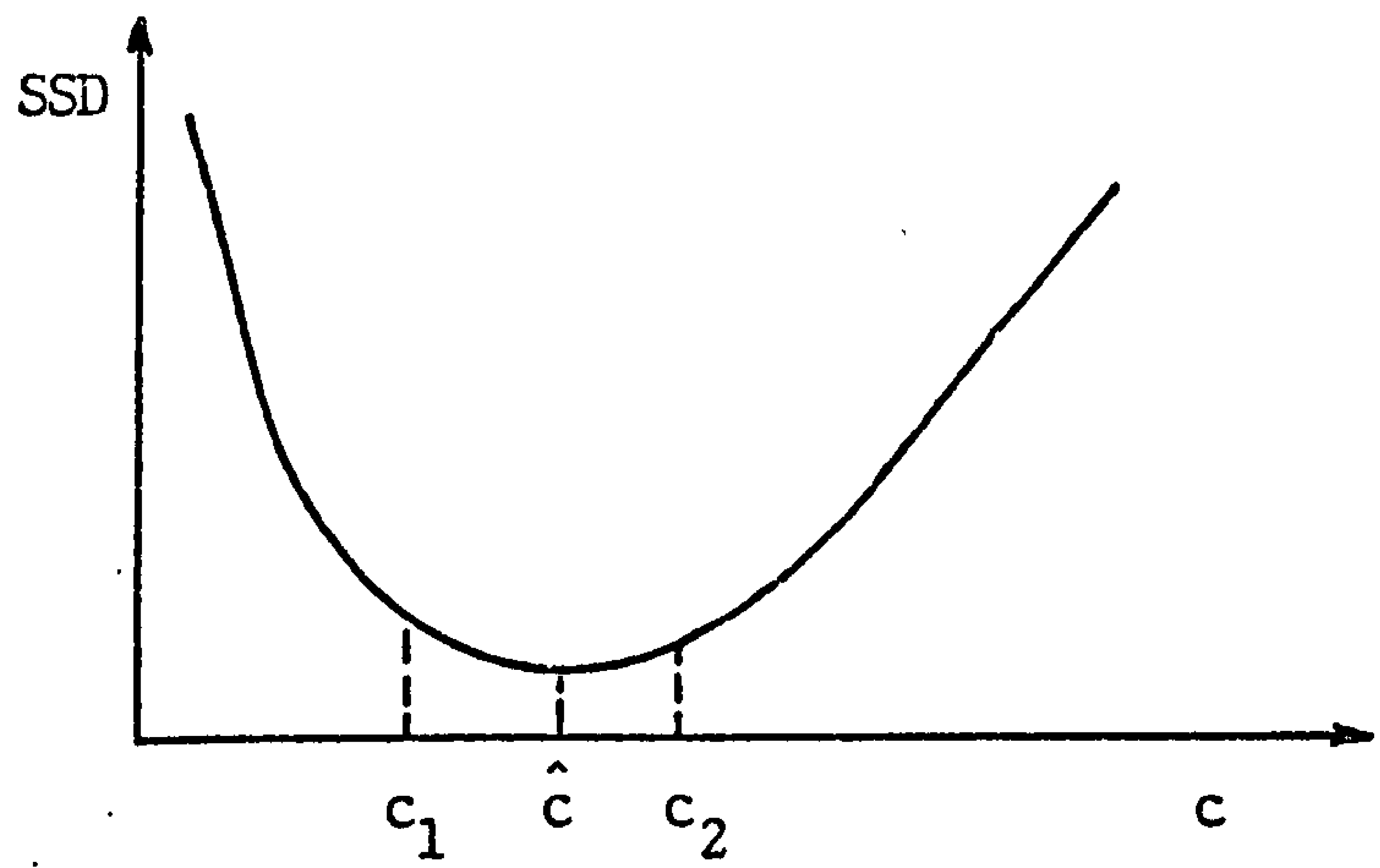


Figure (*) - The Graph of SSD with respect to c.

If we let $c = c_0$ where c_0 is some known value then the values of $\log(x_i + c_0)$, for all i, will be known and hence (3.12-1) becomes similar to the SSD for the 2-parameter lognormal distribution. Therefore, the values of μ_0 and σ_0 corresponding to c_0 can be found by the least squares regression technique described in section 3.7.6. It will then be possible to calculate the value of SSD, from (3.12-1), corresponding to (c_0, μ_0, σ_0). By choosing several different values for c and calculating their corresponding SSDs we can find a range of values of c, say (c_1, c_2), in which the SSD changes from a decreasing function to an increasing function. The minimum of SSD lies in this range. By taking d, the required degree of accuracy for $\hat{c}$, into consideration, we can choose suitable values of c in (c_1, c_2) and hence reduce the range gradually to $(\hat{c}, \hat{c} + d)$ where SSD attains its minimum value at $\hat{c}$. The set $(\hat{c}, \hat{\mu}, \hat{\sigma})$ will be the required least squares estimates of (c, μ, σ) respectively. This technique can be programmed on the computer and will then be much simpler and faster than solving a system of non-linear equations.

Any prior knowledge about the maximum value of c should be utilized in choosing the initial range of c. For example, in the case of our accidental damage data we know that c, the voluntary excess, is at most £30. Therefore, to examine the behaviour of SSD, we should initially try c = 0, 5, 10, 15, 20, 25 and 30.

3.12.2 The Multinomial Maximum Likelihood Method

This method was considered for the two-parameter lognormal distribution in section 3.7.7. In this section we will present the likelihood function for the 3-parameter case. Let us again assume that a sample of grouped data as defined in section 3.7 is available. If X is assumed to be distributed as $LN(x; -c, \mu, \sigma^2)$, by adopting the notation of section 3.7.7,

$$p_i = LN(x_i; -c, \mu, \sigma^2) - LN(x_{i-1}; -c, \mu, \sigma^2)$$

$$p_i = N\left(\frac{\log(x_i + c) - \mu}{\sigma}; 0,1\right) - N\left(\frac{\log(x_{i-1} + c) - \mu}{\sigma}; 0,1\right)$$

Therefore the sample likelihood function will be proportional to L where,

$$L = \prod_{i=1}^{k} p_i^{n_i}$$

and the loglikelihood function will be

$$\log L = \sum_{i=1}^{k} n_i \log [N(z_i; 0,1) - N(z_{i-1}; 0,1)] \qquad (3.12\text{-}2)$$

$$\text{where} \quad z_i = \frac{\log(x_i + c) - \mu}{\sigma}$$

We will adopt the technique, described in section 3.7.7, of maximizing $\log L$ with respect to (c,μ,σ^2). To start the iteration process we can use the least squares, or any other, estimates of the parameters. The set of estimates $(\hat{c},\hat{\mu},\hat{\sigma}^2)$ which maximizes $\log L$ will be the required multinomial maximum likelihood estimates.

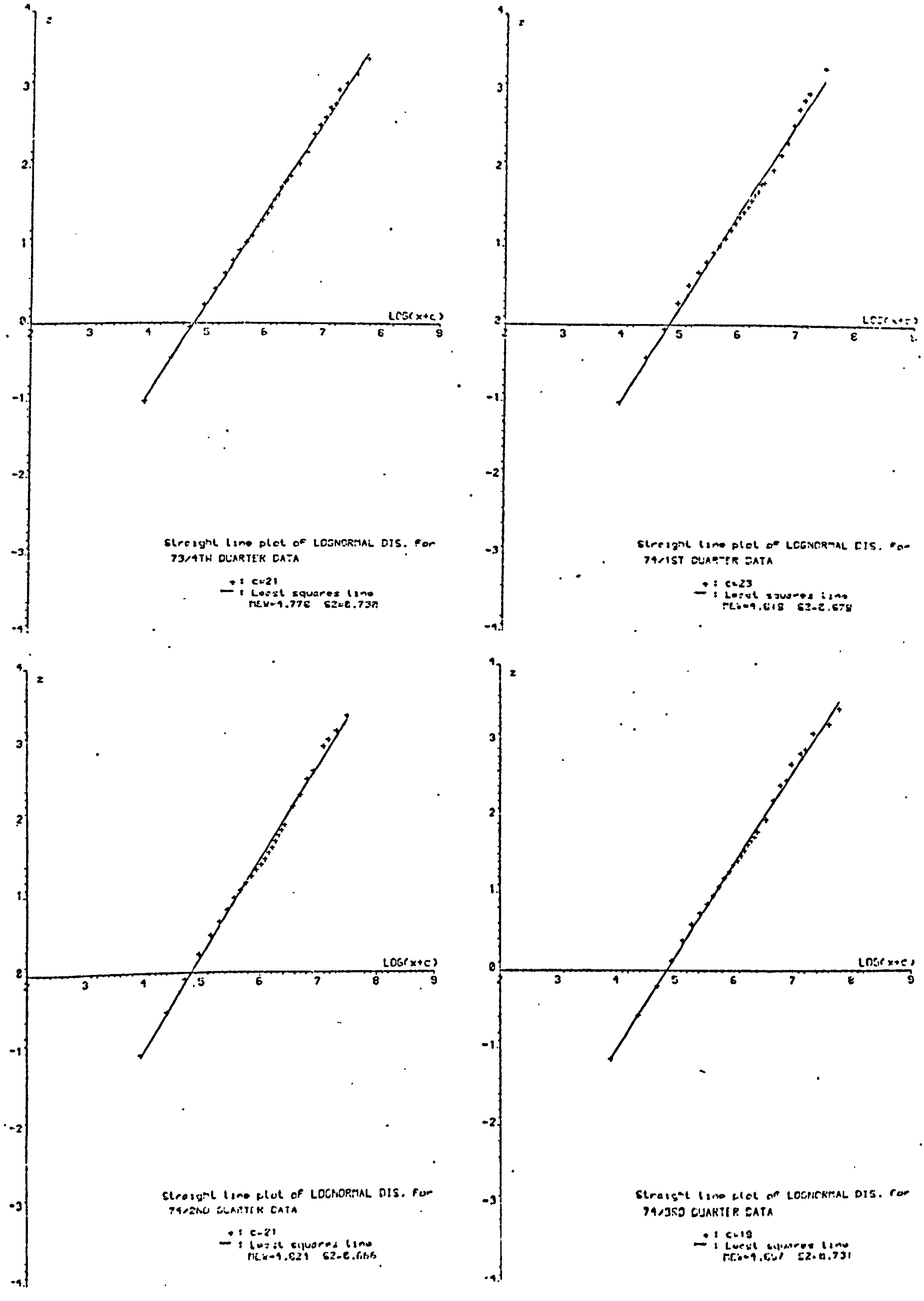
3.13 Application of the 3-Parameter Lognormal Model to the AD Data

We wrote computer program P10 to use the least squares method to estimate the parameters of a 3-parameter lognormal distribution from a sample of grouped data. The program was run on an interactive terminal which made the task of estimation even simpler and faster. It was used to estimate the parameters for our seven samples of accidental damage data which were presented in tables (1.1) to (1.7). We explain the procedure for a particular sample, say, 73/4th quarter. After the sample data has been supplied to the program, it will require a value for c along with an integer 'IPRINT' which is either 0 or 1. IPRINT = 0 indicates that only the value of SSD is required. IPRINT = 1 indicates that a comprehensive table of results is also required. The results of the computer run to find the estimate of c for 73/4th quarter data are presented in table (3.30). We started by c = 0 and increased c in steps of 5. For each c its corresponding SSD was printed. We noticed that in the range (c = 15, c = 25) SSD changed from a decreasing function to an increasing one. Therefore, as a next step we tried c = 18 whose SSD showed that the optimum c is in the range (18,25). Because an accuracy of 1 is acceptable in the estimate of c, we then tried c = 19 whose SSD indicated that the optimum c should be in the interval (20,25). We then tried c = 21 whose SSD was less than that of c = 20. Therefore, the optimum range was reduced to (21,25). As the next step we tried c = 22 whose SSD was greater than that of c = 21. Therefore, with our required accuracy, $\hat{c}$ = 21 is the least squares estimate of c. This value was then supplied to the program, with IPRINT = 1, and the estimates of μ and σ^2 along with an extensive table of results were produced (see table (3.31)). Similar tables were obtained for other samples and the results are presented in tables (3.32) to (3.37). We can see that the value of c for our data ranges from 13 to 25 which is reasonable. $\hat{\mu}$ has generally increased over time from 4.78 to 4.98 while $\hat{\sigma}^2$ ranges from 0.67 to 0.77. The means and standard deviations of the fitted models are very close to the means and standard deviations of the samples. The Chi-square statistics are smaller than in the 2-parameter case but still indicate significant differences between the fitted distributions and the actual sample values, except in the case of 74/4th, 75/1st and 75/2nd quarters where the X^2 values are not significant. The total expected loss statistics are very small and are at most 2% of the total actual cost. This indicates an overall agreement between the fitted models and the samples. It was important to see how closely the actual sample points lie on the least squares straight line. Therefore, we wrote computer program P11 to plot the sample points $(\log(x_i + \hat{c}) , z_i)$ and the least squares line

$$z = \frac{1}{\hat{\sigma}} \log(x + \hat{c}) - \frac{\hat{\mu}}{\hat{\sigma}}$$

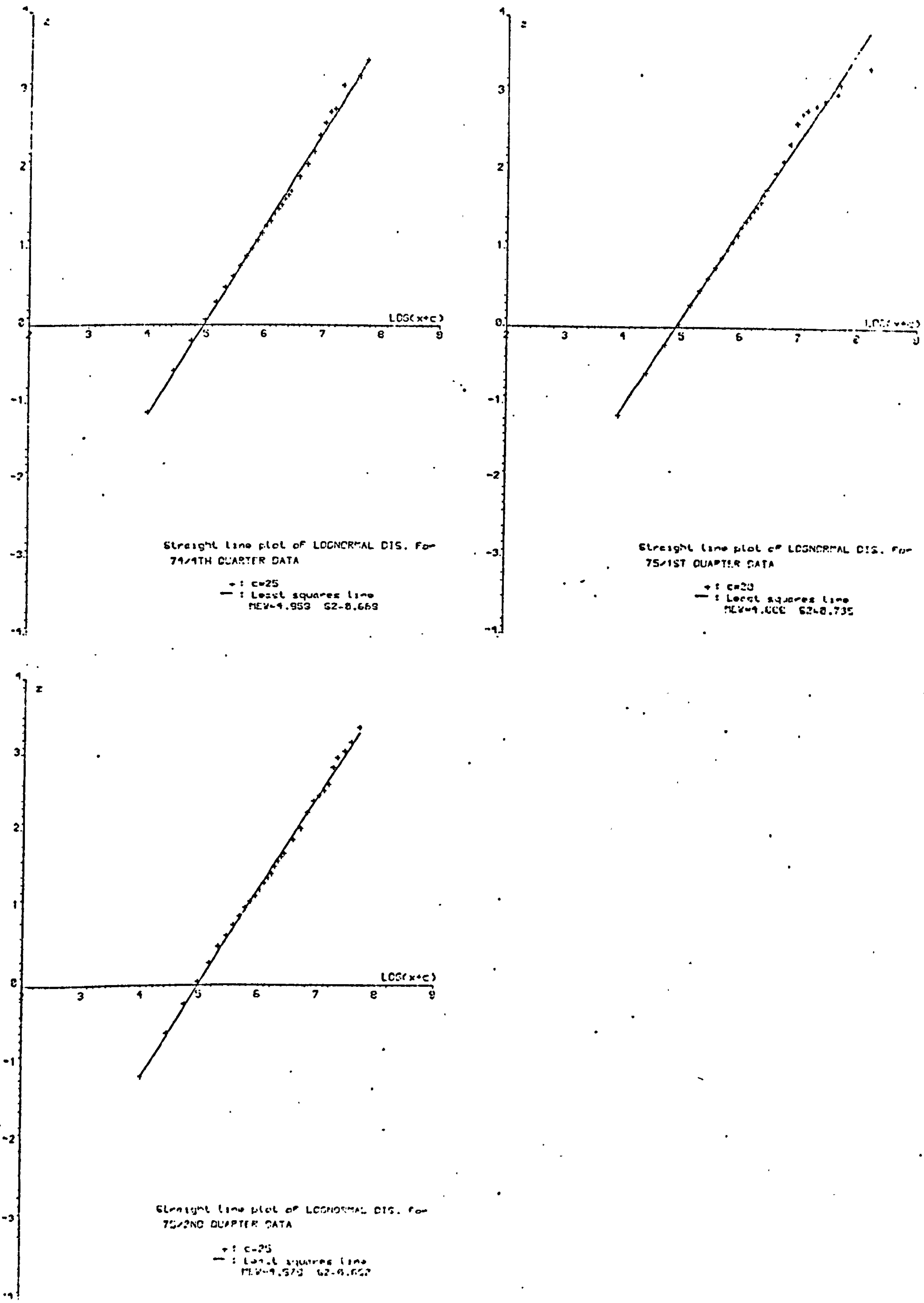
For the seven AD samples the graphs are presented in figures (3.8-a) and (3.8-b). The points, for each sample, generally lie closely on the line.

Figure (3.8-a)



z
LOG(x+c)
Straight line plot of LOGNORMAL DIS. For
73/4TH QUARTER DATA
+ : c=21
— : Least squares line
z
Straight line plot of LOGNORMAL DIS. For
74/1ST QUARTER DATA
+ : c=23
— : Least squares line
z
LOG(x+c)
Straight line plot of LOGNORMAL DIS. For
74/2ND QUARTER DATA
+ : c=21
— : Least squares line
z
LOG(x+c)
Straight line plot of LOGNORMAL DIS. For
74/3RD QUARTER DATA
+ : c=19
— : Least squares line

Figure (3.8-b)


z
LOG(x+c)
Straight line plot of LOGNORMAL DIS. For
74/4TH QUARTER DATA
+ : c=25
— : Least squares line
MEV-4.953 S2-0.669
z
Straight line plot of LOGNORMAL DIS. For
75/1ST QUARTER DATA
+ : c=20
— : Least squares line
z
LOG(x+c)
Straight line plot of LOGNORMAL DIS. For
75/2ND QUARTER DATA
— : Least squares line

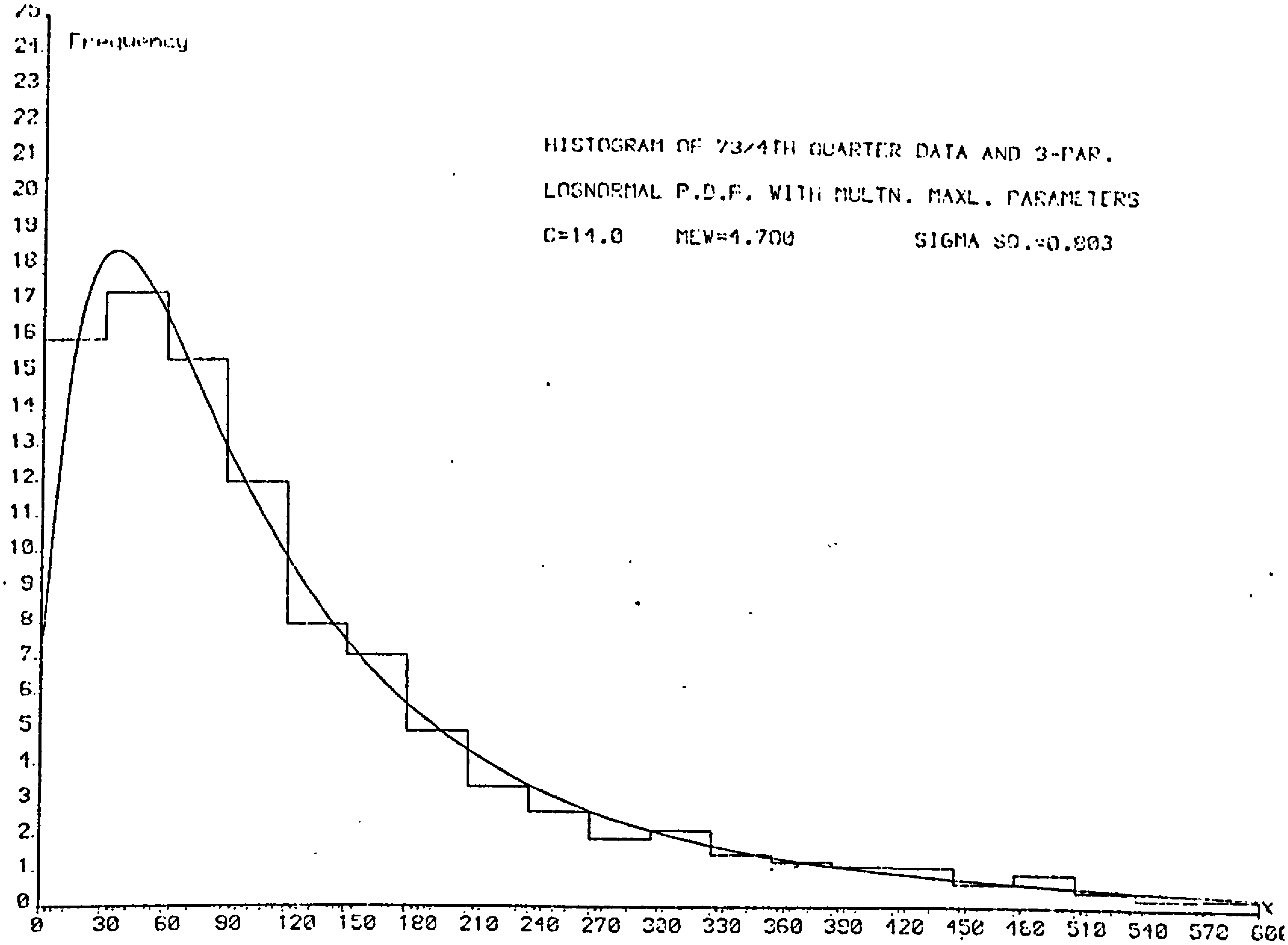
In the lower tails the fits are very good, unlike the large deviations which were observed in the 2-parameter case. It is encouraging to see that the introduction of parameter c into the model has produced more satisfactory results.

Computer program P12 was written to estimate the parameters by the method of multinomial maximum likelihood. This program was used with the samples of accidental damage data. For each sample, the least squares estimates of (c,μ,σ^2) were supplied, to the program, as the starting values for the iteration process. The results for the seven samples are presented in tables (3.38) to (3.44). In each case the estimates of the parameters, the mean and the standard deviation of the fitted distribution and an extensive table have been produced. The estimates of c range from 10 to 21 with an average of about 15 which is reasonable. The values of $\hat{\mu}$ generally show an increasing pattern and are in the 4.65 to 4.91 range. $\hat{\sigma}^2$ ranges from 0.71 to 0.86. We can observe that the Chi-square values are smaller than their corresponding values for the least squares method. They do not indicate significant differences, except for 74/1st and 74/3rd quarters, between the fitted distributions and the sample values. For 74/1st and 74/3rd quarters the differences are, respectively, highly significant and significant, but we should note that the major contribution to the X^2 values comes from just one cell in each case. The total expected loss statistics are small and are at most 1.2% of the total actual cost of claims. This indicates a very satisfactory general agreement between the theoretical models and the samples values. Values of the Kolmogorov-Smirnov statistics for all the samples were calculated, by using the (Actual-Expected) column in each table, and were compared with the critical values of this statistic as given in table (2.2). Only for 74/1st quarter this statistic indicated almost significant differences between the fitted distribution and the sample values. In all other cases the differences were satisfactorily not significant.

The least squares method is very easy to use but, as in the case of the 2-parameter lognormal, the multinomial maximum likelihood method has produced better results in terms of the goodness-of-fit test statistics. We, therefore, again recommend the use of this method in estimating the parameters of the distribution from grouped data.

In order to see the agreement between the frequency curve of the fitted distribution and the histogram of the sample values we plotted them, for each of our seven AD samples, by using a modified version of program P7 to suit the 3-parameter lognormal model. The resulting graphs are presented in figures (3.9.1) to (3.9.7). For each sample, the multinomial maximum likelihood parameters have been used to plot the frequency curve. Each figure represents two graphs. One for claim amounts $\leq$ £600 and the other for claim amounts > £600. The agreement between the frequency curve and the histogram is very satisfactory for all samples. The curve portrays the mode of the histogram. The large deviations in the lower tail between the curve and the histogram which were observed in the 2-parameter case (see figures (3.5.1) to (3.5.7)) have now disappeared. The long tail of the histogram has also been modelled very well by the distribution. We are, therefore, satisfied that the 3-parameter lognormal distribution offers a correct model for our accidental damage claim amounts. Hence, we shall next consider the problem of predicting the distribution of claim amounts during a future period.

Figure (3.9.1)


Frequency
HISTOGRAM OF 73/4TH QUARTER DATA AND 3-PAR.
LOGNORMAL P.D.F. WITH MULTN. MAXL. PARAMETERS
C=14.0 MEW=4.700 SIGMA SQ.=0.803
X

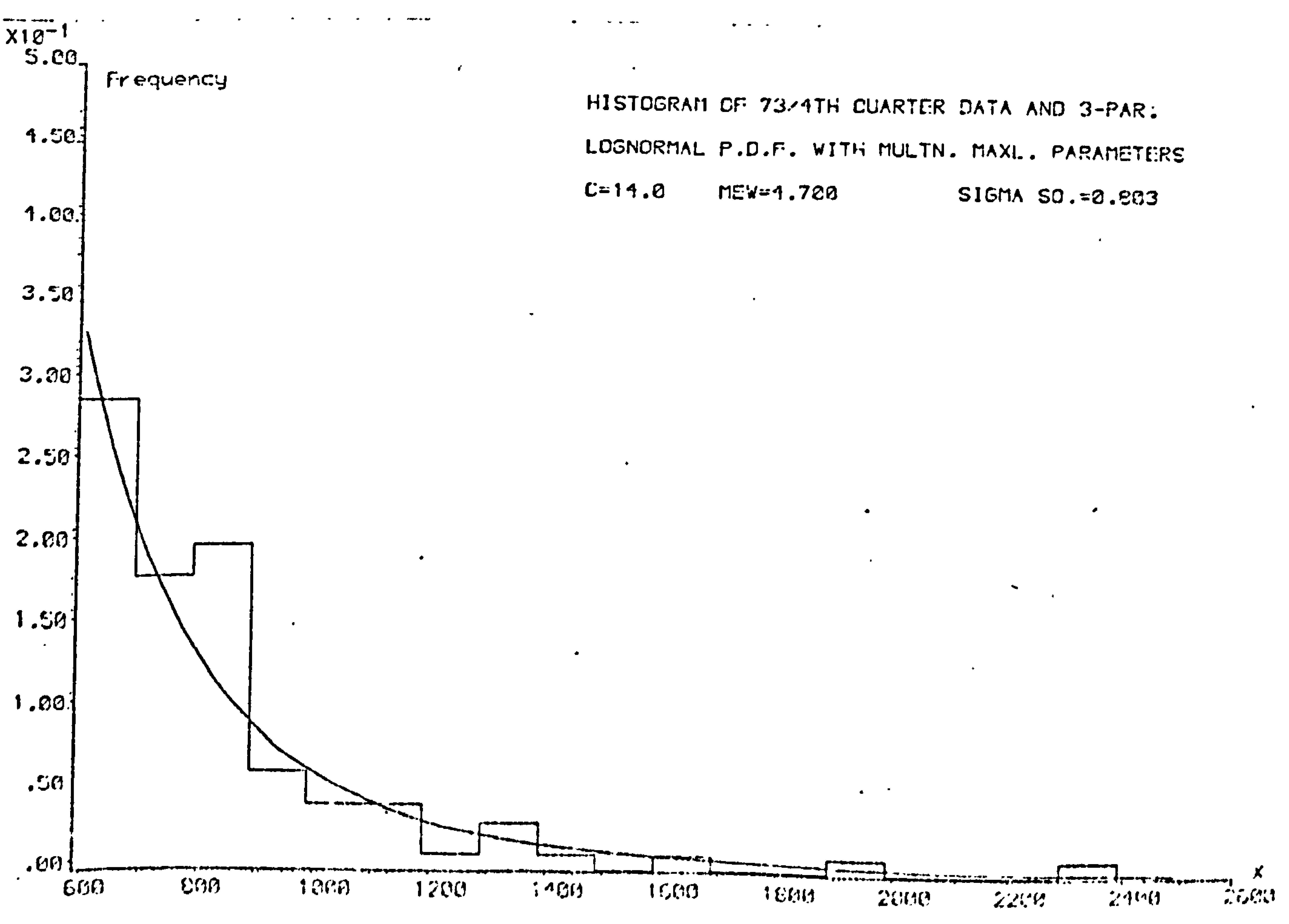
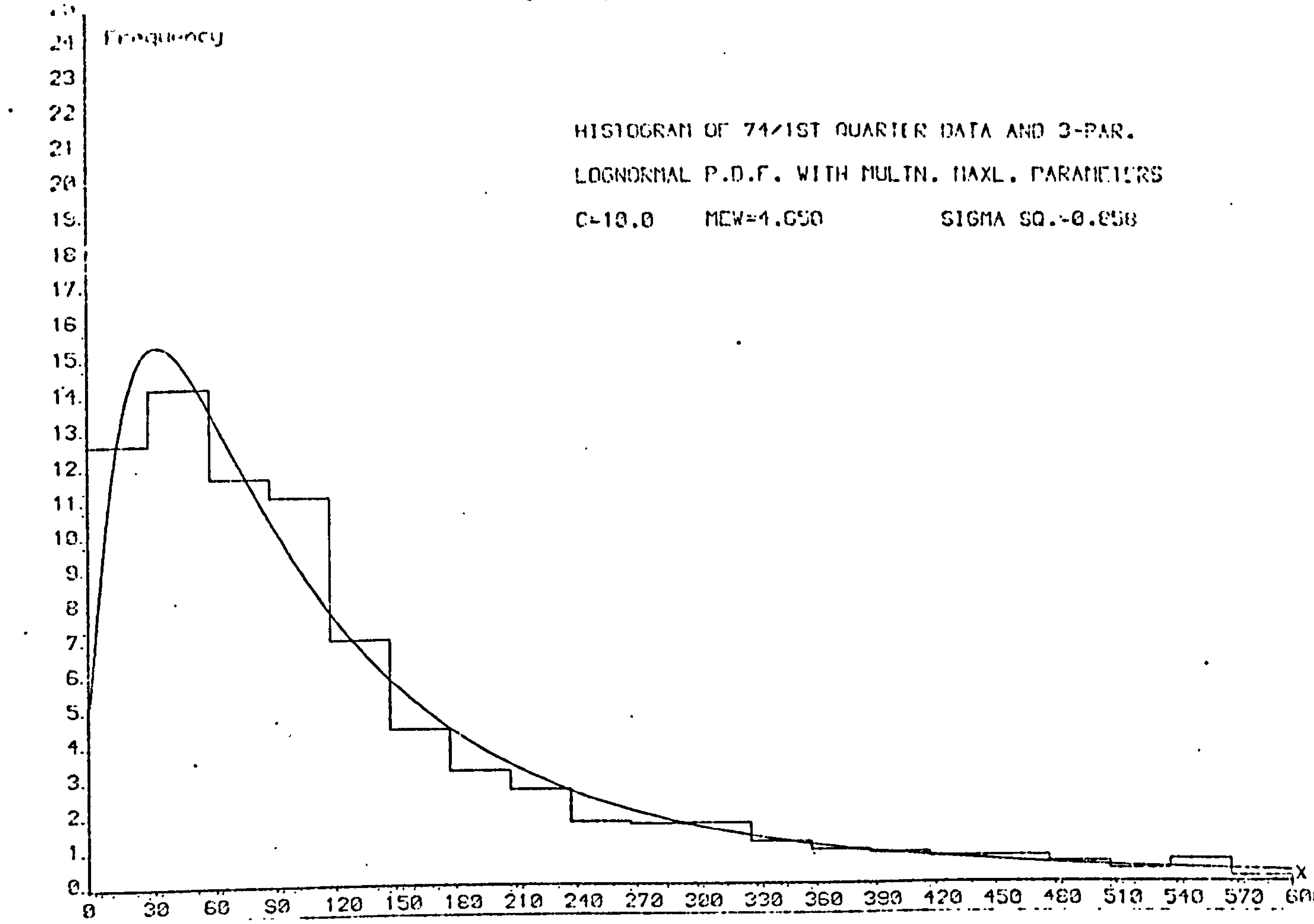

X10^-1
Frequency
HISTOGRAM OF 73/4TH QUARTER DATA AND 3-PAR.
LOGNORMAL P.D.F. WITH MULTN. MAXL. PARAMETERS
C=14.0 MEW=4.700 SIGMA SQ.=0.803
X

Figure (3.9.2)


Frequency
HISTOGRAM OF 74/1ST QUARTER DATA AND 3-PAR.
LOGNORMAL P.D.F. WITH MULTN. MAXL. PARAMETERS
C=10.0 MEW=4.650 SIGMA SQ.=0.858
X
0 30 60 90 120 150 180 210 240 270 300 330 360 390 420 450 480 510 540 570 600

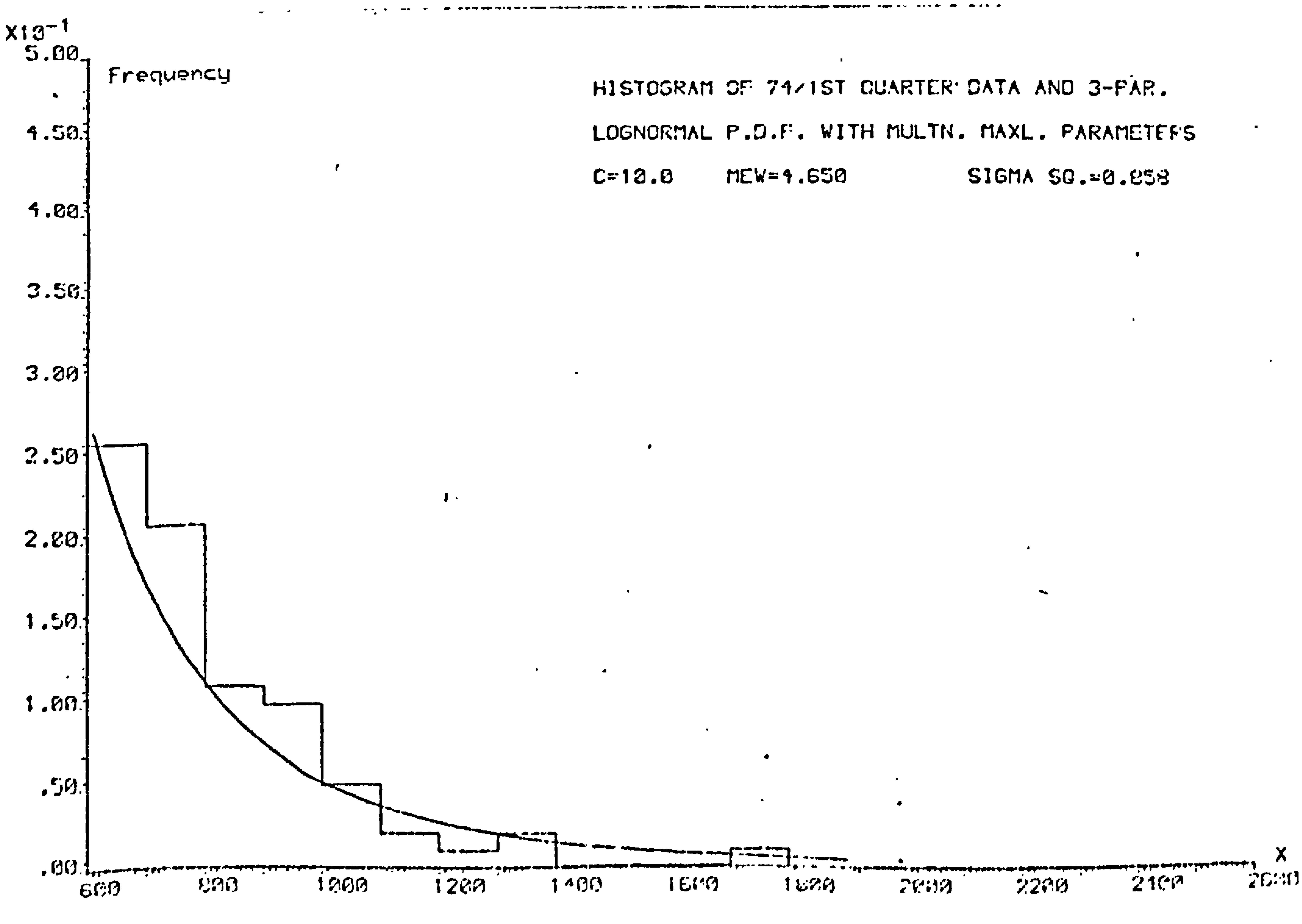
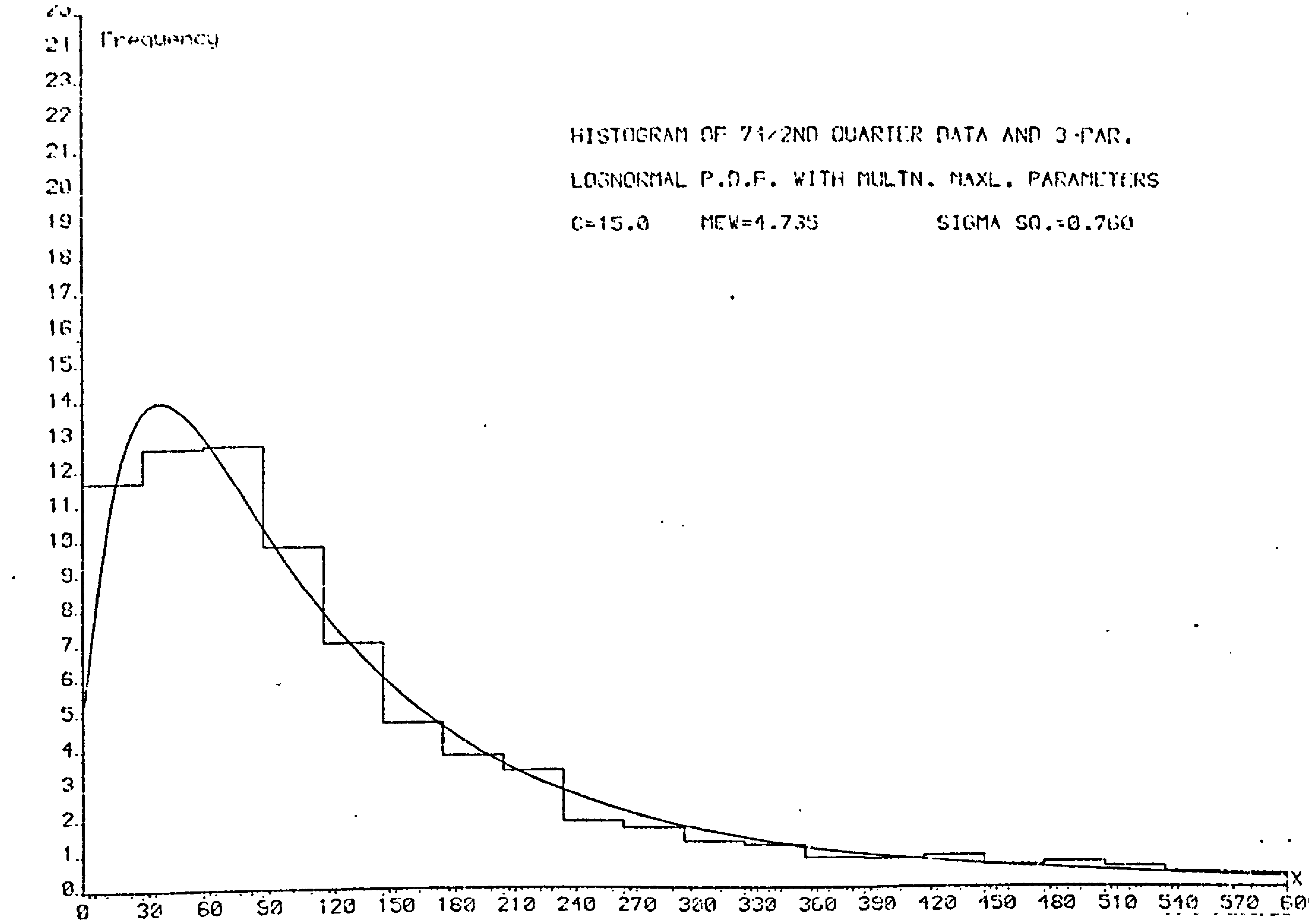

X10⁻¹
Frequency
HISTOGRAM OF 74/1ST QUARTER DATA AND 3-PAR.
LOGNORMAL P.D.F. WITH MULTN. MAXL. PARAMETERS
C=10.0 MEW=4.650 SIGMA SQ.=0.858
X
5.00 4.50 4.00 3.50 3.00 2.50 2.00 1.50 1.00 .50 .00
600 800 1000 1200 1400 1600 1800 2000 2200 2400 2600

Figure (3.9.3)


Frequency
HISTOGRAM OF 74/2ND QUARTER DATA AND 3-PAR.
LOGNORMAL P.D.F. WITH MULTN. MAXL. PARAMETERS
C=15.0 MEW=4.735 SIGMA SQ.=0.760
X

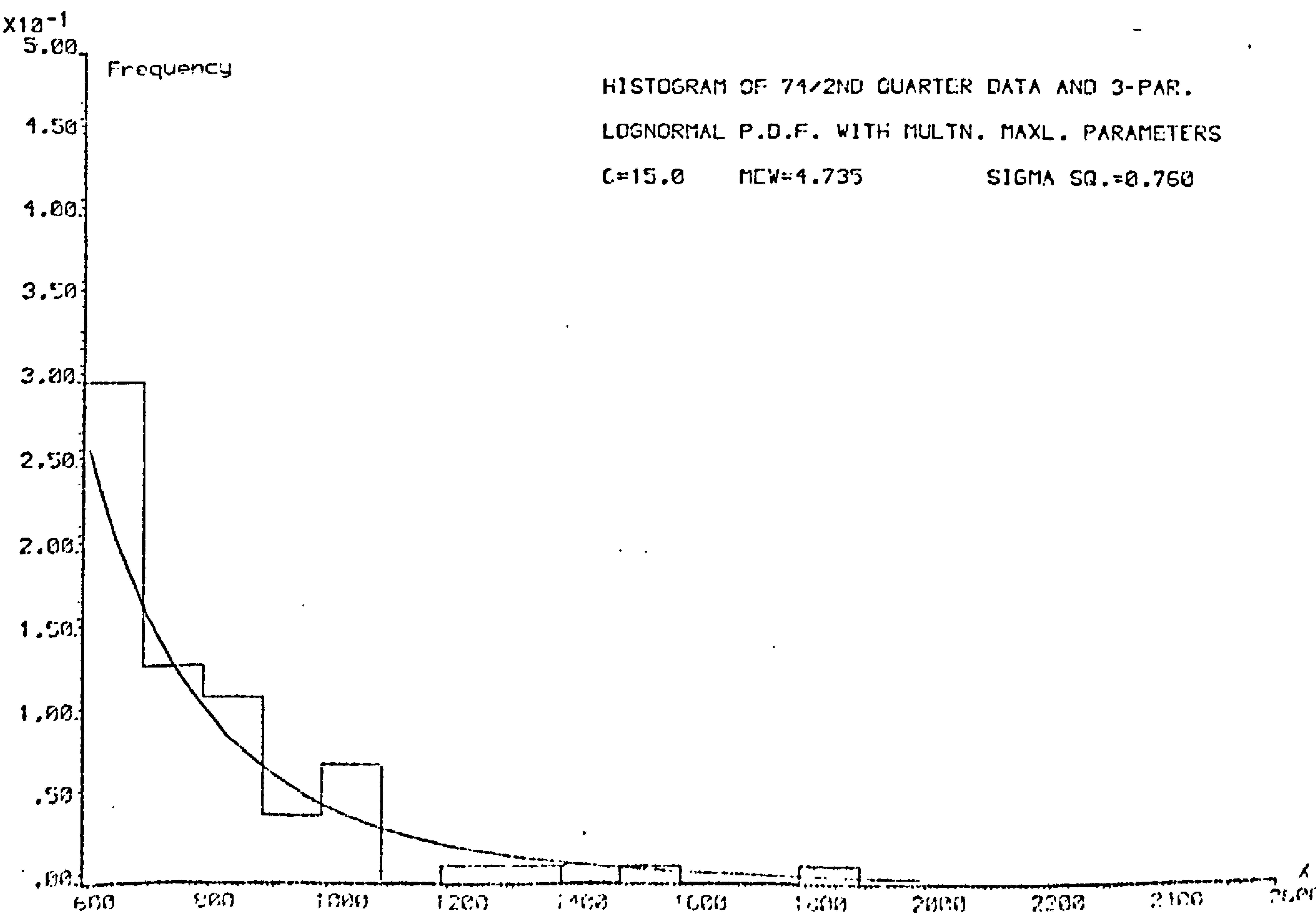
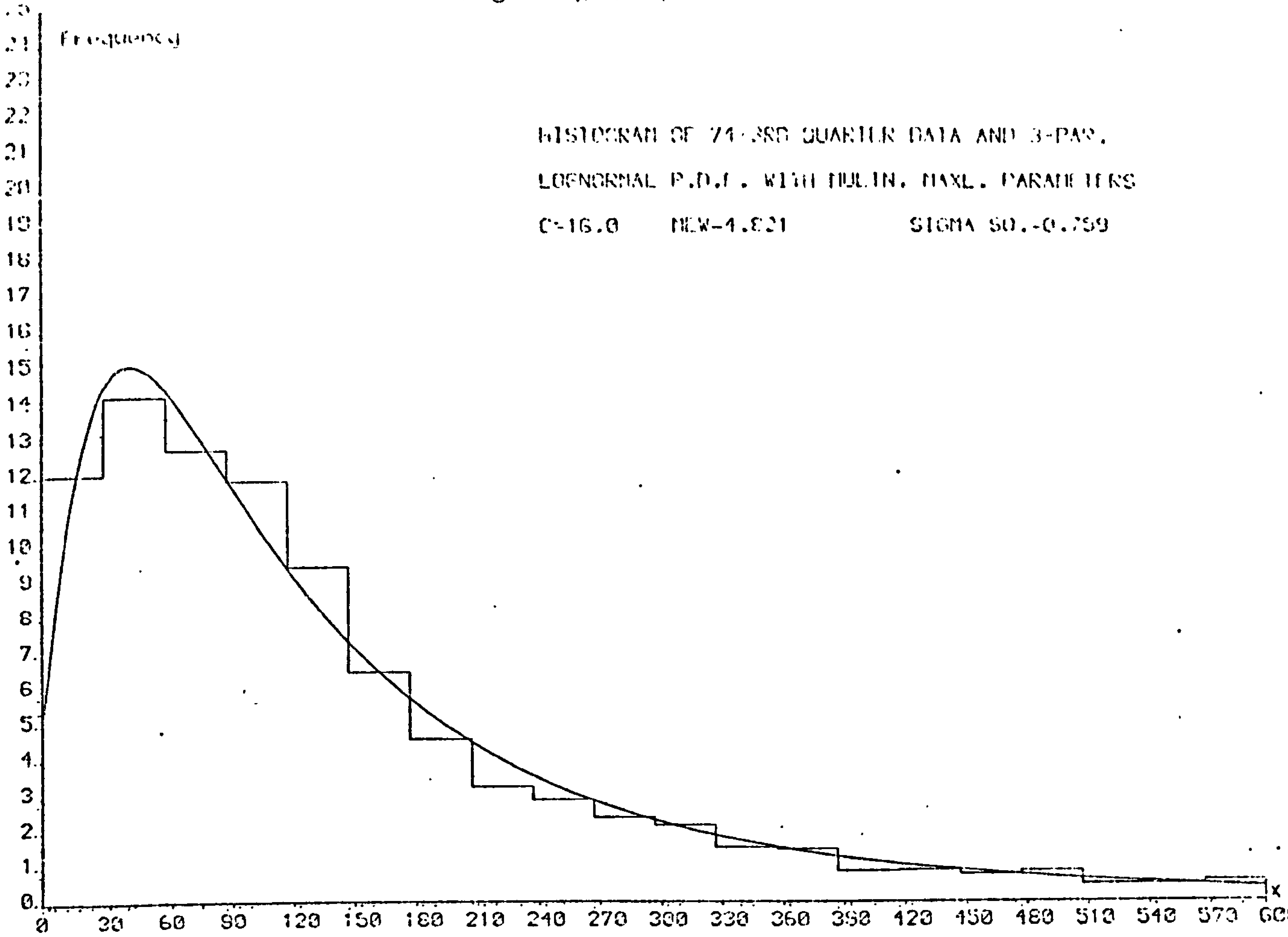
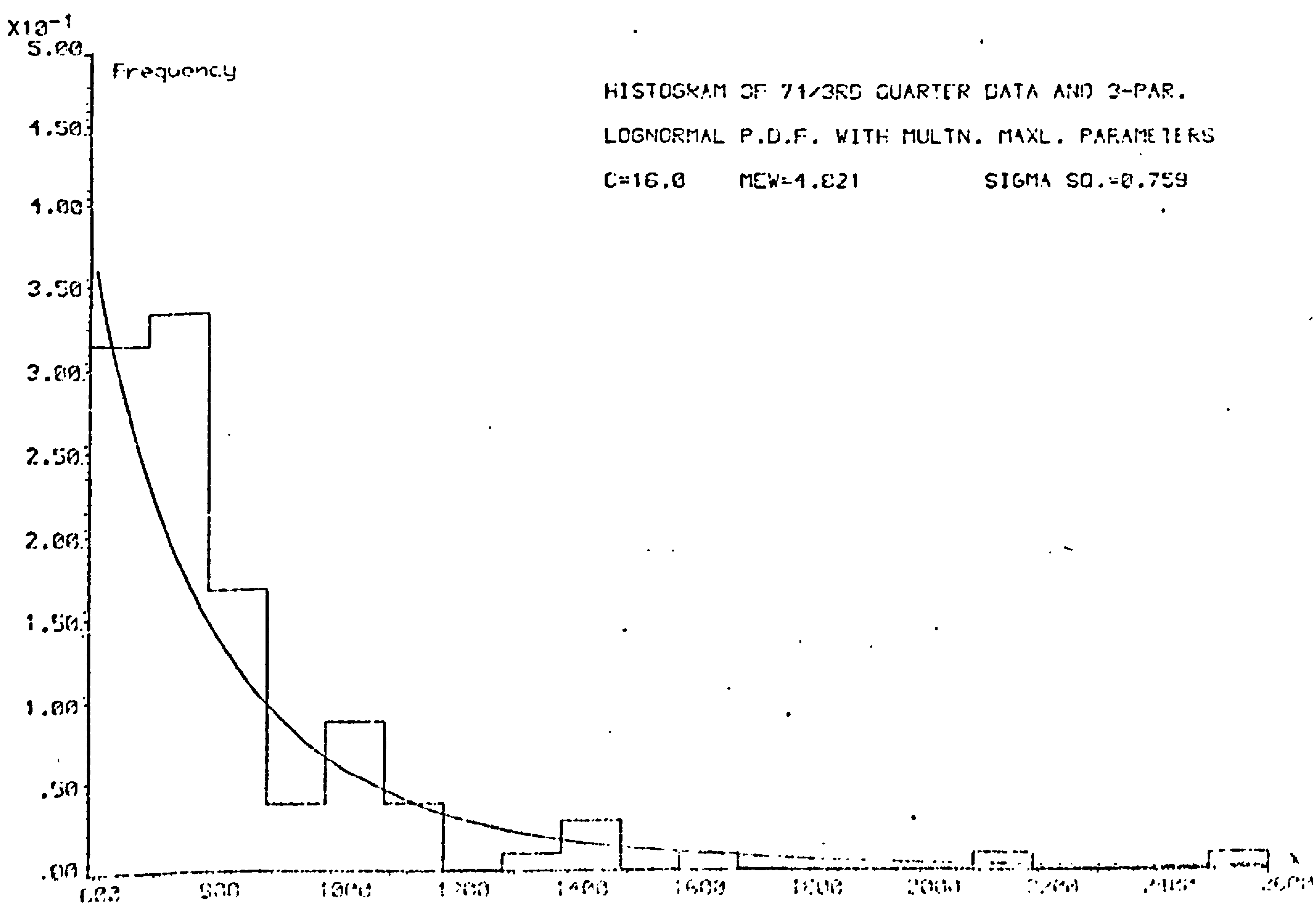

X10⁻¹
Frequency
HISTOGRAM OF 74/2ND QUARTER DATA AND 3-PAR.
LOGNORMAL P.D.F. WITH MULTN. MAXL. PARAMETERS
C=15.0 MEW=4.735 SIGMA SQ.=0.760
X

Figure (3.9.4)

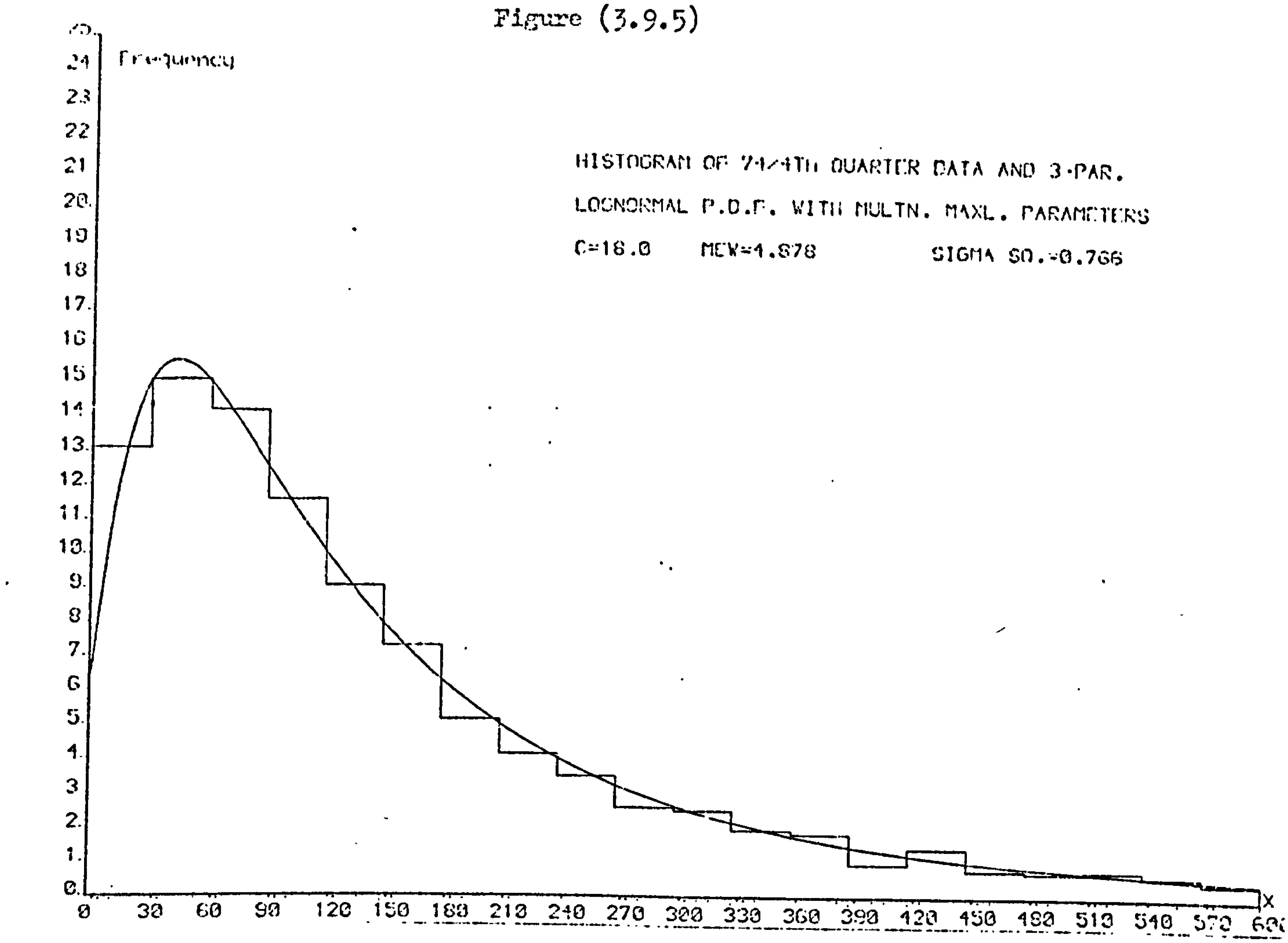


Frequency
HISTOGRAM OF 71-3RD QUARTER DATA AND 3-PAR.
LOGNORMAL P.D.F. WITH MULTN. MAXL. PARAMETERS
C=16.0 MEW=4.821 SIGMA SQ.=0.759
0 30 60 90 120 150 180 210 240 270 300 330 360 390 420 450 480 510 540 570



X10⁻¹
Frequency
HISTOGRAM OF 71/3RD QUARTER DATA AND 3-PAR.
LOGNORMAL P.D.F. WITH MULTN. MAXL. PARAMETERS
C=16.0 MEW=4.821 SIGMA SQ.=0.759
5.00 4.50 4.00 3.50 3.00 2.50 2.00 1.50 1.00 .50 .00
800 1000 1400 1600 2000

Figure (3.9.5)


Frequency
HISTOGRAM OF 74/4TH QUARTER DATA AND 3-PAR.
LOGNORMAL P.D.F. WITH MULTN. MAXL. PARAMETERS
C=16.0 MEV=4.878 SIGMA SQ.=0.766
0 30 60 90 120 150 180 210 240 270 300 330 360 390 420 450 480 510 540 570
X

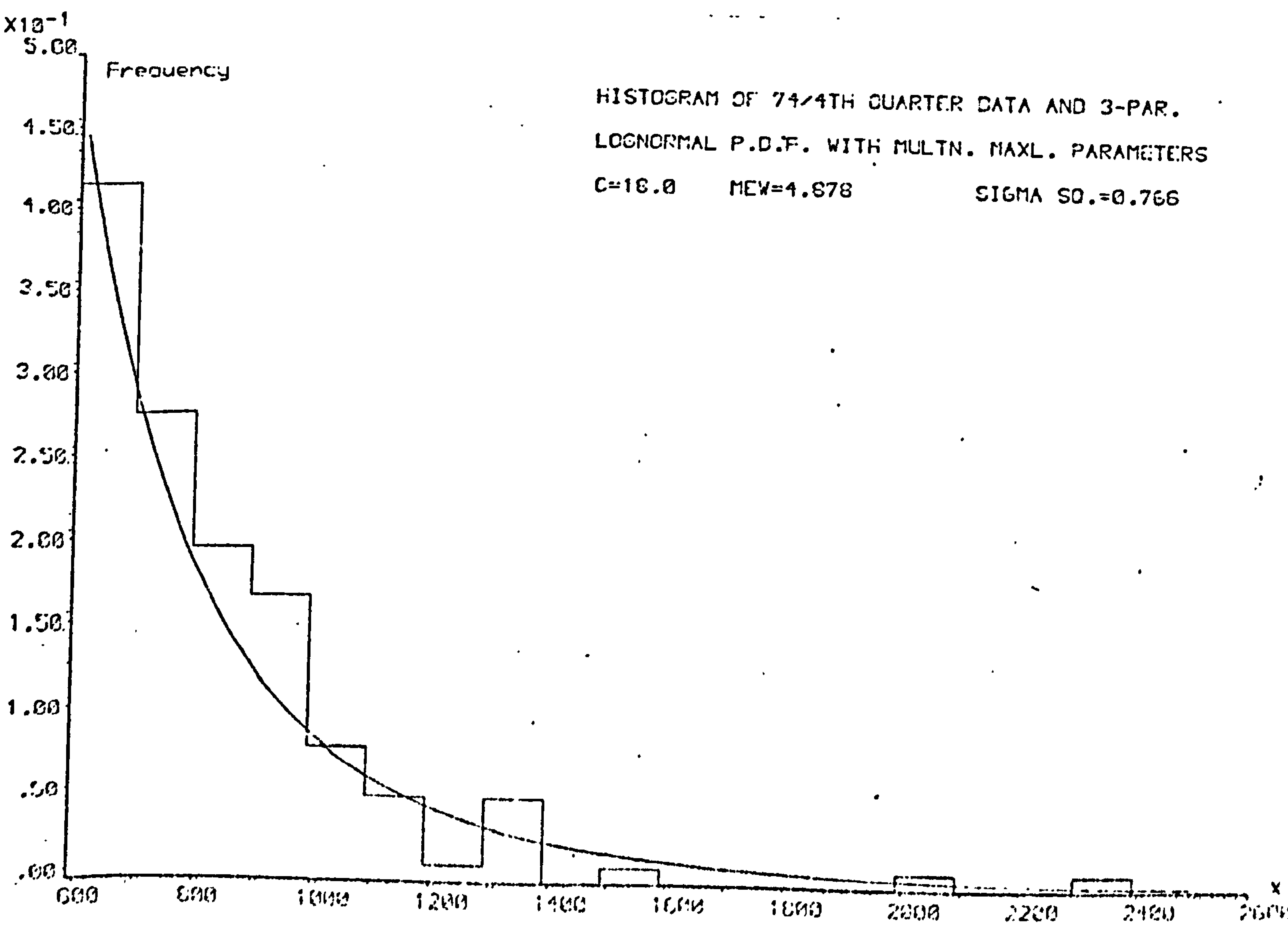
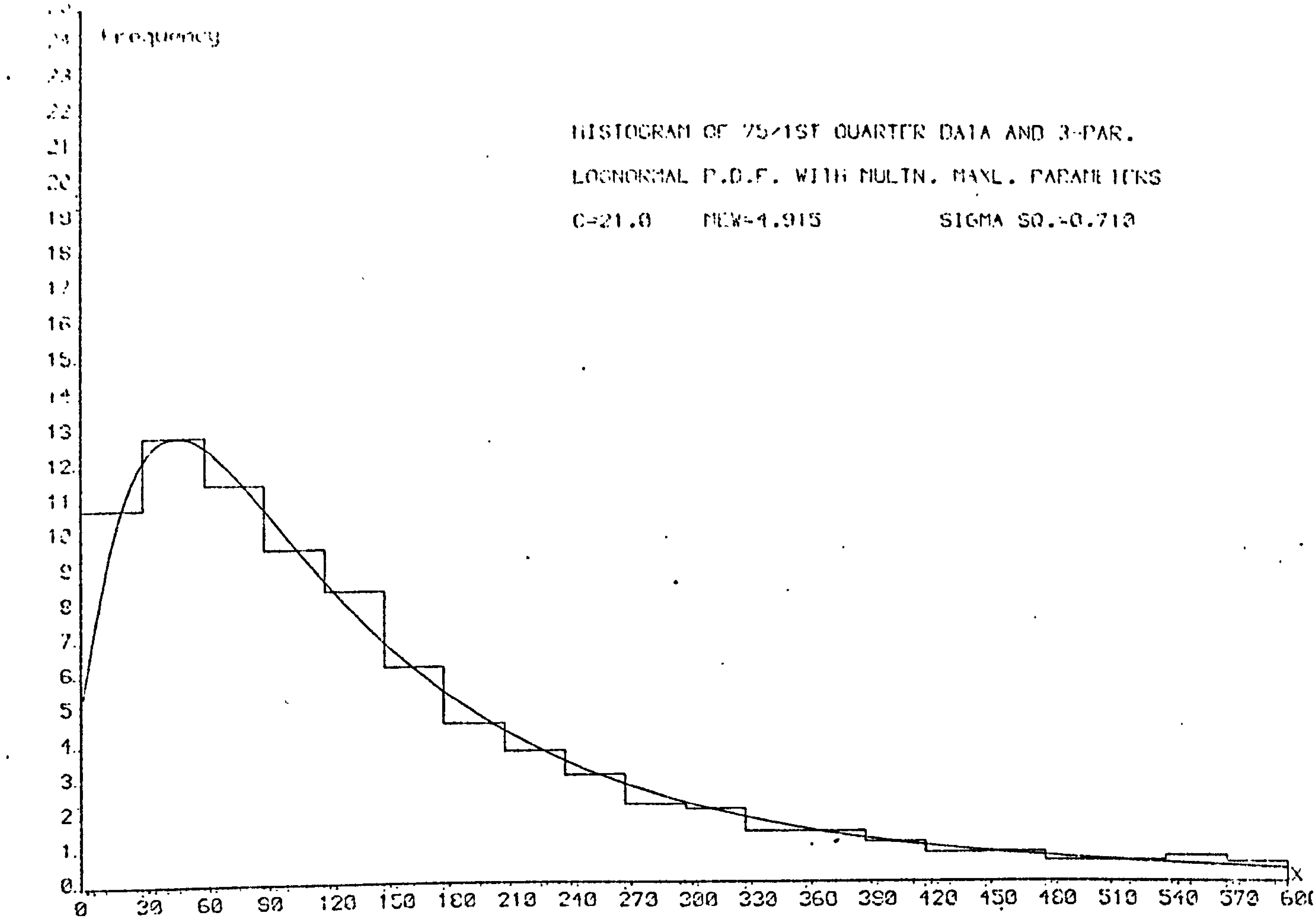

X10-1
Frequency
HISTOGRAM OF 74/4TH QUARTER DATA AND 3-PAR.
LOGNORMAL P.D.F. WITH MULTN. MAXL. PARAMETERS
C=16.0 MEV=4.878 SIGMA SQ.=0.766
5.00 4.50 4.00 3.50 3.00 2.50 2.00 1.50 1.00 .50 .00
600 800 1000 1200 1400 1600 1800 2000 2200 2400 2600
X

Figure (3.9.6)


Frequency
HISTOGRAM OF 75/1ST QUARTER DATA AND 3-PAR.
LOGNORMAL P.D.F. WITH MULTN. MAXL. PARAMETERS
C=21.0 MEW=4.915 SIGMA SQ.=0.710
0 30 60 90 120 150 180 210 240 270 300 330 360 390 420 450 480 510 540 570 600
X

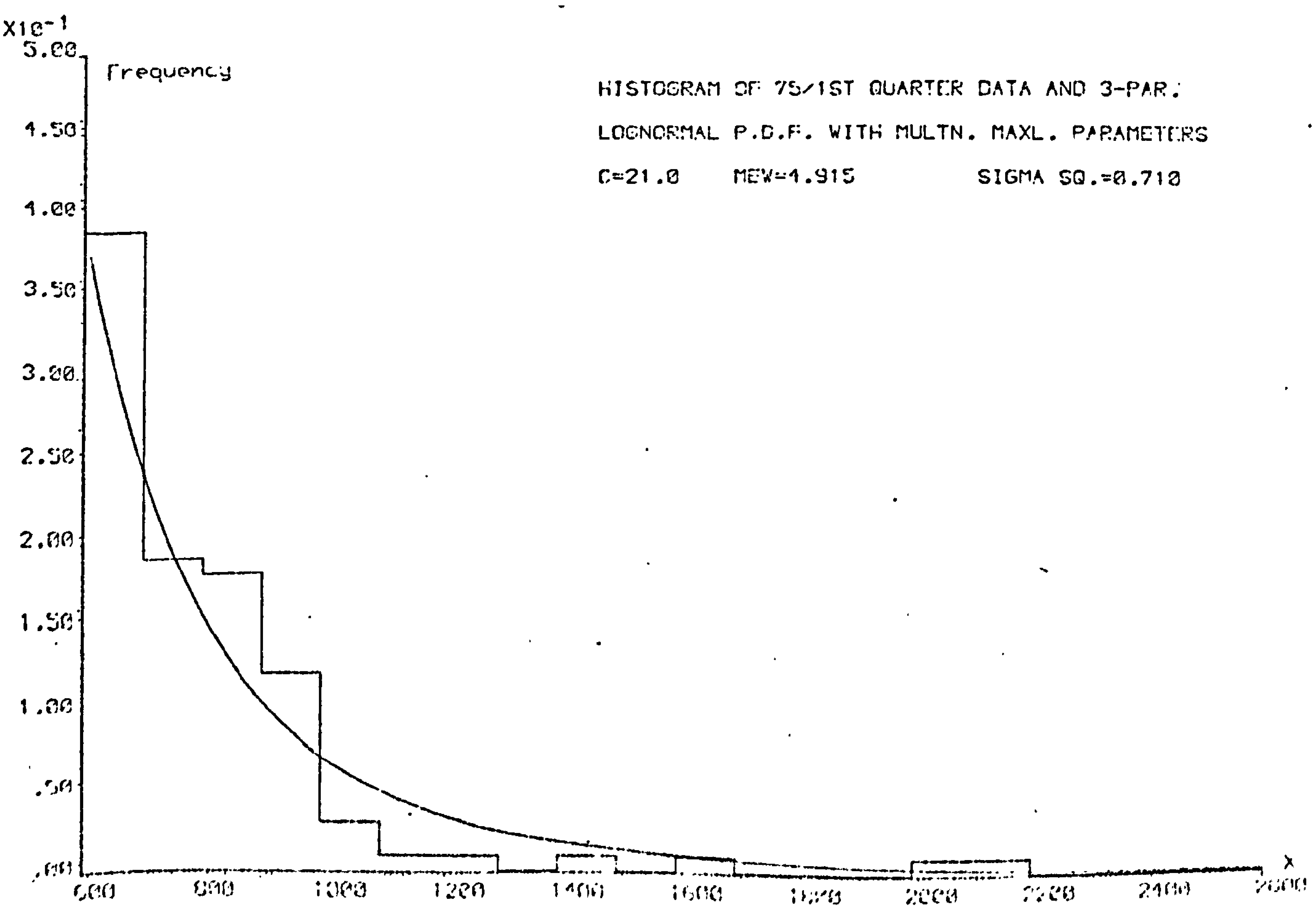
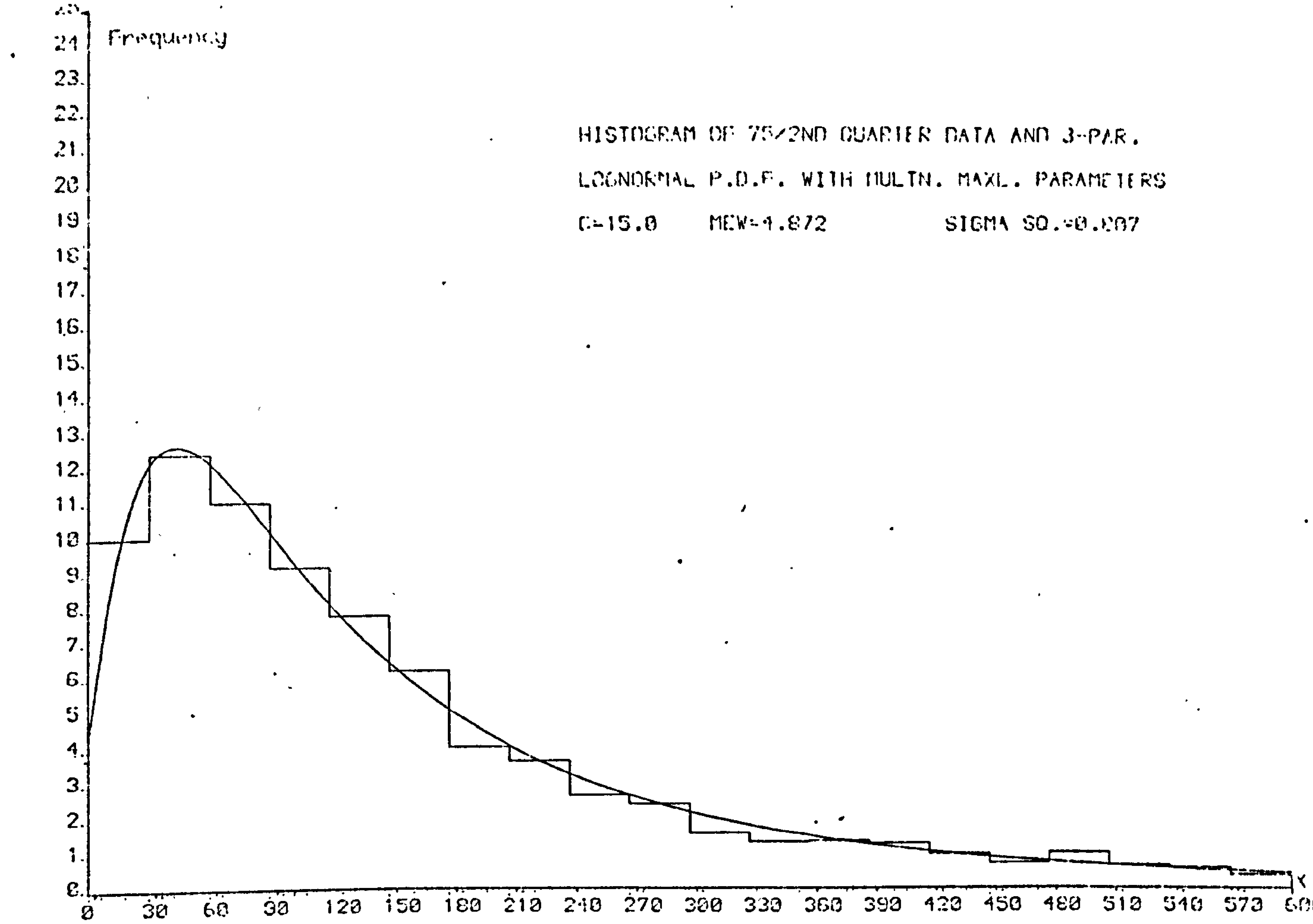
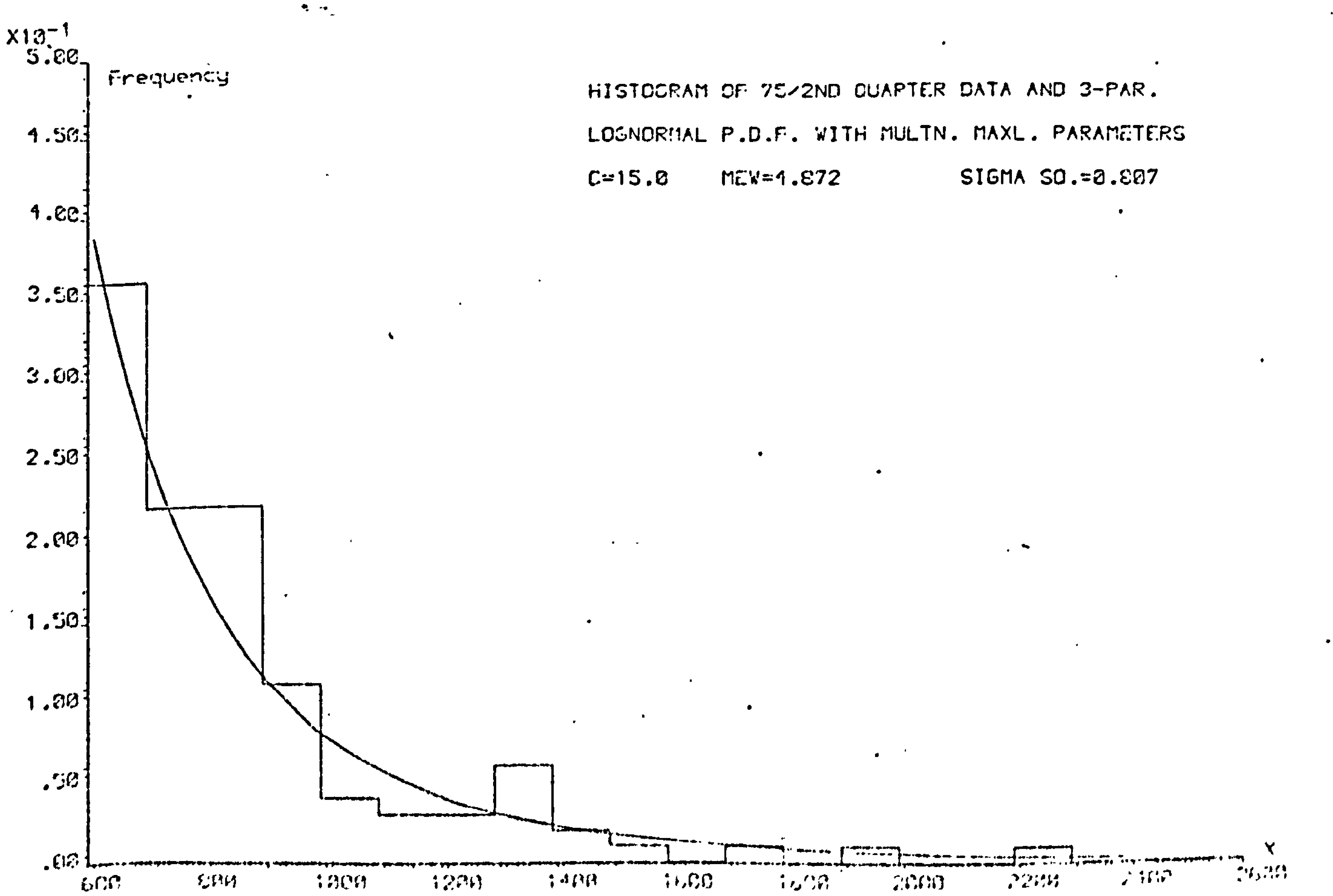

X10^-1
5.00
4.50
4.00
3.50
3.00
2.50
2.00
1.50
1.00
.50
Frequency
HISTOGRAM OF 75/1ST QUARTER DATA AND 3-PAR.
LOGNORMAL P.D.F. WITH MULTN. MAXL. PARAMETERS
C=21.0 MEW=4.915 SIGMA SQ.=0.710
X

Figure (3.9.7)


Frequency
HISTOGRAM OF 75/2ND QUARTER DATA AND 3-PAR.
LOGNORMAL P.D.F. WITH MULTN. MAXL. PARAMETERS
C=15.0 MEW=4.872 SIGMA SQ.=0.807
24. 23. 22. 21. 20. 19. 18. 17. 16. 15. 14. 13. 12. 11. 10. 9. 8. 7. 6. 5. 4. 3. 2. 1. 0.
0 30 60 90 120 150 180 210 240 270 300 330 360 390 420 450 480 510 540 570
X


X10^-1
Frequency
HISTOGRAM OF 75/2ND QUARTER DATA AND 3-PAR.
LOGNORMAL P.D.F. WITH MULTN. MAXL. PARAMETERS
C=15.0 MEW=4.872 SIGMA SQ.=0.807
5.00 4.50 4.00 3.50 3.00 2.50 2.00 1.50 1.00 .50 .00
600 800 1000 1200 1400 1600 2000 2200
X

3.14 Prediction of the Claim Amount Distribution :- The 3-Parameter Lognormal Model

The importance of predicting the future cost of claims was mentioned in 3.9 and here we will adopt techniques similar to those of that section for predicting the future distribution of claim amounts. To save us repeating some of the arguments of that section we suggest referring back to it at this stage.

3.14.1 The Effects of Inflation on the Parameters of the Model

Let us again assume that the effect of inflation on a claim of amount X, during a period of time, is to increase it to $U = X(1 + i)$ where i is the rate of inflation, according to some appropriate index, for that period. If X is assumed to be distributed as $LN(x; c,\mu,\sigma^2)$, then by a transformation of variables we can show that $U = X(1 + i)$ will be distributed as :

$$LN(u; rc, \mu + \delta , \sigma^2) \text{ where } r = 1 + i \text{ and } \delta = \log(1 + i)$$

Therefore, when the distribution of X is known, we can modify its parameters to obtain the distribution of claim amounts in a future period. The foregoing shows that inflation increases both c and μ in time but does not affect σ^2 . This supports our intuition that the amount of voluntary excess, c , should go up in time due to the inflation. For our accidental damage data, the increase, over time, in the values of μ is obvious from tables (3.38) to (3.44). The values of c for the later samples are also larger than for the earlier ones.

Based on the results for the two-parameter case, we suggest that the technique of section 3.9.2 be used for predicting the distribution of accidental damage claim amounts during a future period. The appropriate index proved to be the General Index of Retail Prices and we will use it here again.

3.14.2 Prediction for the AD Data

To test the prediction technique on the accidental damage data, we modified computer program P8 for use with the 3-parameter model. For reasons presented in section 3.9.2, we used the parameters of the distributions for 73/4th, 74/1st and 74/2nd quarters to predict the distributions of claim amounts in 74/4th, 75/1st and 75/2nd quarters respectively. The appropriate rate of inflation in each case was calculated from the General Index of Retail Prices, as described in section 3.9.2. The results, in the form of extensive tables which allow comparisons between the actual and expected (predicted) distributions, are presented in tables (3.45) to (3.47). In all three cases the Chi-square statistics are relatively small, considering the number of degrees of freedom, and do not indicate any significant differences between the predicted distributions and the actual sample values. The total expected loss statistics are small and show an over-prediction of the total cost which is more satisfactory than an under-prediction. They are a small proportion of the total actual cost in all cases. The number of changes and non-changes of signs in the 'A - E' (i.e. , the actual minus expected frequencies) column are roughly equal. The Kolmogorov-Smirnov statistics were calculated from the results tables and in each case were compared with the critical values given in table (2.1). They showed that the differences between the actual and predicted distributions were not significant for 74/4th and 75/1st quarters but were almost significant for 75/2nd quarter.

These results show a great improvement over those given in tables (3.27) to (3.29) for the 2-parameter model. The consistent satisfactory outcomes of the goodness-of-fit tests indicate the success of the prediction technique and the appropriateness of the General Index of Retail prices for inflating accidental damage claim amounts. Various assumptions about the future rate of inflation or values of the parameters can be tested by this technique and the computer program associated with it. A value for the future rate of inflation and a set of parameters can be adopted as standard values and other rates and sets of parameters can then be tested against them. The total expected loss statistic, T , in each case will indicate, in terms of the easily understood amount of money, by how much the assumptions differ from each other.

It may appear simpler just to increase the actual mean claim amount in a particular period by the effective rate of inflation between that period and a future one and to take the result as the mean claim amount for the latter period. Using the total number of claims we can then calculate the total expected cost of claims for the future period. But, it can be shown that this method will result in a much larger over-prediction of the total claims cost than by our distribution theory approach. For example, the sample mean claim amount for the 73/4th quarter is £150.36 and the effective rate of inflation from this period to 74/4th quarter is 18.2% which yields a mean claim amount of £177.7 for the latter quarter. The number of claims for that quarter is 3064 and hence the total expected cost would be £544,551 which shows an over-prediction of £10,884, as compared with £4,093 by our approach. The former discrepancy is 2.03% of the total actual cost while the distribution theory resulted in a ratio of 0.77% only. Similar results can be observed for the other two quarters we have considered. Therefore, one of the merits of our distribution theory approach is that it results in accurate but smaller reserves. This means that the insurance company will not need to tie up funds in unnecessary reserves.

3.15 Conclusions

In this chapter the two and three parameter lognormal distributions were studied as models for the distribution of claim amounts, in general insurance and in particular for motor insurance accidental damage claims. Simulation exercises showed that the graphical method is a good way of testing for lognormality. In the 3-parameter case, this test also provided a rough estimate of the location parameter c. It was observed that, even in large samples of (size 2500) observations from an actual lognormal population, the points (log x, z) deviate markedly from the straight line for values of x in the upper tail of the distribution. Therefore, in testing exercises, we should expect to find that the points in the upper tail deviate from the straight line which may fit the rest of the distribution.

Several methods of estimation for the 2-parameter case were considered and their efficiencies were calculated. It was observed that the approximate maximum likelihood method was the most efficient amongst them. It was mentioned that the most efficient method for estimation from grouped data is the multinomial maximum likelihood method. A technique for estimating the parameters by this method was proposed and shown to provide better estimates (in terms of a closer agreement between the model and the actual data). For the 2-parameter case, the approximate maximum likelihood and the multinomial maximum likelihood methods gave approximately the same results and hence the use of the former method in this case should be recommended. In the 3-parameter case our proposed technique for estimating the parameters by the least squares method was shown to be very simple when programed on the computer. However, the results produced by this method were not as satisfactory as those of the multinomial maximum likelihood method.

For the accidental damage data, it was found that the 3-parameter form of the distribution is the appropriate model. The inflation was shown to be the main cause of the increase of the claim amount over time. We proposed a technique for predicting the distribution of the claim amounts in a future period. This involved updating the parameters c and μ with respect to an appropriate rate of inflation. The General Index of Retail Prices was shown to provide the appropriate rate of inflation for accidental damage claims. We tested the prediction technique on data from the past and it proved to be very successful.

The total expected loss statistic, T , was suggested in chapter 2 as a goodness-of-fit measure. In the present chapter it was put to use and shown to provide an easily understood measure (in monetary terms) of overgraduation or undergraduation. Its use is particularly recommended in prediction exercises where it indicates the extent to which a model is overpredicting or underpredicting the actual total cost of the claims.

3.16 Tables

Coefficients of skewness (g_1) and excess kurtosis (g_2) and their standard errors for the distribution of log X, where X = claim amount

Table (3.1)
Accidental Damage Samples

Period of accident	g_1	$Se(g_1)$	g_2	$Se(g_2)$
73/4th quarter	-0.107	0.044	-0.506	0.089
74/1st "	-0.076	0.049	-0.503	0.099
74/2nd "	-0.144	0.050	-0.464	0.100
74/3rd "	-0.179	0.046	-0.352	0.092
74/4th "	-0.199	0.044	-0.428	0.088
75/1st "	-0.237	0.048	-0.394	0.096
75/2nd "	-0.184	0.049	-0.376	0.098

Table (3.2)
Simulated Samples

Sample	g_1	g_2
1	-0.049	-0.328
2	0.105	-0.143
3	0.065	-0.167
4	0.122	-0.122
5	-0.007	-0.170
6	-0.013	-0.209
7	-0.037	-0.412
8	-0.039	-0.369
9	0.030	-0.223
10	0.067	-0.272
For all samples: $Se(g_1) = 0.049$, $Se(g_2) = 0.098$		

Table (3.3)

Efficiencies of various methods of estimation for the Two-Parameter Lognormal Distribution

Method of estimation	Eff(μ)	Eff(σ^2)	Eff(α)	Eff(β)
Moments	0.044	0.095	4.145	17.710
Approx. Max. Likelihood	0.021	0.036	3.944	9.768
7 & 93% quantiles	0.131	0.188	7.267	24.493
27 & 73% quantiles	0.026	0.050	4.634	12.186
Median & Coeff. of variation	0.026	0.095	7.525	24.561
Least squares regression	0.051	0.104	4.363	18.999

Table (3.4-a)

***** 73/4TH QUARTER *****

FITTING THE TWO PARAMETER LOGNORMAL DIS. BY DIFF. METHODS

TABLE OF ESTIMATED PARAMETERS AND STATISTICS

METHOD-->	MOMNTS	MAXIMUM LIKELIHOOD	7&93 % QUANTILES	27&73 % QUANTILES	GRAPHICAL	REGRESSION
STATISTICS						
MEWHAT	4.583	4.503	4.325	4.540	4.554	4.568
SIGMA2HAT	0.860	1.089	1.321	1.018	0.893	0.876
ALPHA=MEAN	150.359	155.669	146.318	155.935	148.495	149.333
BETA=S.D.	175.482	218.539	242.476	207.387	178.399	176.791
COF OF VAR	1.167	1.404	1.657	1.330	1.201	1.184
MEDIAN	97.832	90.315	75.596	93.713	95.000	96.363
MODE	41.417	30.400	20.179	33.846	38.882	40.125
SKEWNESS	5.091	6.978	9.523	6.342	5.338	5.211
KURTOSIS	71.228	153.819	341.219	121.219	79.719	75.268

Table (3.4-b)

***** 73/4TH QUARTER *****

EXPECTED NO. OF CLAIMS BY DIFF. METHODS OF ESTIMATION

TWO PARAMETER LOGNORMAL DIS.

AMOUNT £	ACT. NO.	MOMNTS	MAX.LIK.	7&93%QS	27&73%QS	GRAPH	REG.
1- 30	478.	318.	454.	654.	405.	349.	334.
31- 60	518.	602.	613.	634.	607.	615.	609.
61- 90	461.	501.	458.	423.	469.	496.	499.
91- 120	359.	373.	329.	290.	342.	365.	369.
121- 150	239.	274.	240.	207.	250.	266.	270.
151- 180	213.	203.	179.	153.	187.	197.	200.
181- 210	148.	153.	136.	116.	142.	148.	150.
211- 240	102.	117.	106.	90.	110.	113.	115.
241- 270	81.	90.	83.	71.	86.	88.	89.
271- 300	58.	71.	67.	57.	69.	69.	70.
301- 330	66.	56.	54.	47.	56.	55.	55.
331- 360	45.	45.	44.	38.	45.	44.	45.
361- 390	39.	37.	37.	32.	37.	36.	36.
391- 420	35.	30.	31.	27.	31.	29.	30.
421- 450	34.	25.	26.	23.	26.	24.	24.
451- 480	20.	21.	22.	20.	22.	20.	20.
481- 510	29.	17.	19.	17.	19.	17.	17.
511- 540	14.	14.	16.	15.	16.	14.	14.
541- 570	8.	12.	14.	13.	14.	12.	12.
571- 600	9.	10.	12.	11.	12.	10.	10.
601- 700	29.	25.	30.	28.	30.	25.	25.
701- 800	18.	16.	20.	19.	19.	16.	16.
801- 900	20.	10.	14.	14.	13.	10.	10.
901-1000	6.	7.	10.	10.	9.	7.	7.
1001-1100	4.	5.	7.	7.	7.	5.	5.
1101-1200	4.	3.	5.	6.	5.	3.	3.
1201-1300	1.	2.	4.	4.	4.	3.	2.
1301-1400	3.	2.	3.	3.	3.	2.	2.
1401-1500	1.	1.	2.	3.	2.	1.	1.
1501-1600	0.	1.	2.	2.	2.	1.	1.
1601-1700	1.	1.	1.	2.	1.	1.	1.
1701-1800	0.	1.	1.	1.	1.	1.	1.
1801-1900	0.	0.	1.	1.	1.	1.	0.
1901-2000	1.	0.	1.	1.	1.	0.	0.
2001-2100	0.	0.	1.	1.	1.	0.	0.
2101-2200	0.	0.	1.	1.	0.	0.	0.
2201-2300	0.	0.	0.	1.	0.	0.	0.
2301-2400	1.	0.	0.	1.	0.	0.	0.
TOTAL	3045.	3043.	3043.	3043.	3044.	3043.	3042.

Table (3.4-c)

***** 73/4TH QUARTER *****

(ACTUAL - EXPECTED) CLAIM NO.S BY DIFF. METHODS

TWO-PARAMETER LOGNORMAL DIS.

AMOUNT £	MOMENTS	MAX. LIK.	7&93%QS	27&73%QS	GRAPHCL	REG.
1- 30	160.	24.	-176.	73.	129.	144.
31- 60	-64.	-95.	-116.	-89.	-97.	-91.
61- 90	-40.	3.	38.	-8.	-35.	-30.
91- 120	-14.	30.	69.	17.	-6.	-10.
121- 150	-35.	-1.	32.	-11.	-27.	-31.
151- 180	10.	34.	60.	26.	16.	13.
181- 210	-5.	12.	32.	6.	0.	-2.
211- 240	-15.	-4.	12.	-8.	-11.	-13.
241- 270	-9.	-2.	10.	-5.	-7.	-8.
271- 300	-13.	-9.	1.	-11.	-11.	-12.
301- 330	10.	12.	19.	10.	11.	11.
331- 360	0.	1.	7.	0.	1.	0.
361- 390	2.	2.	7.	2.	3.	3.
391- 420	5.	4.	8.	4.	6.	5.
421- 450	9.	8.	11.	8.	10.	10.
451- 480	-1.	-2.	0.	-2.	0.	0.
481- 510	12.	10.	12.	10.	12.	12.
511- 540	0.	-2.	-1.	-2.	0.	0.
541- 570	-4.	-6.	-5.	-6.	-4.	-4.
571- 600	-1.	-3.	-2.	-3.	-1.	-1.
601- 700	4.	-1.	1.	-1.	4.	4.
701- 800	2.	-2.	-1.	-1.	2.	2.
801- 900	10.	6.	6.	7.	10.	10.
901-1000	-1.	-4.	-4.	-3.	-1.	-1.
1001-1100	-1.	-3.	-3.	-3.	-1.	-1.
1101-1200	1.	-1.	-2.	-1.	1.	1.
1201-1300	-1.	-3.	-3.	-3.	-2.	-1.
1301-1400	1.	0.	0.	0.	1.	1.
1401-1500	0.	-1.	-2.	-1.	0.	0.
1501-1600	-1.	-2.	-2.	-2.	-1.	-1.
1601-1700	0.	0.	-1.	0.	0.	0.
1701-1800	-1.	-1.	-1.	-1.	-1.	-1.
1801-1900	0.	-1.	-1.	-1.	-1.	0.
1901-2000	1.	0.	0.	0.	1.	1.
2001-2100	0.	-1.	-1.	-1.	0.	0.
2101-2200	0.	-1.	-1.	0.	0.	0.
2201-2300	0.	0.	-1.	0.	0.	0.
2301-2400	1.	1.	0.	1.	1.	1.

Table (3.4-d)

***** 73/4TH QUARTER *****

CHI SQ. GOODNESS OF FIT TEST FOR DIFF. METHODS :

TWO-PARAMETER LOGNORMAL DIS.

AMOUNT $	MOMNTS	MAX.LIK.	7&93%QS	27&73%QS	GRAPHCL	REG.
			(A-E)**2/E			
1- 30	80.503	1.269	47.364	13.158	47.682	62.084
31- 60	11.721	14.723	21.224	13.049	15.299	13.598
61- 90	3.194	0.020	3.414	0.136	2.470	2.894
91- 120	0.525	2.736	16.417	0.845	0.099	0.271
121- 150	4.471	0.004	4.947	0.484	2.741	3.559
151- 180	0.493	6.458	23.529	3.615	1.299	0.845
181- 210	0.163	1.059	8.828	0.254	0.000	0.027
211- 240	1.923	0.151	1.600	0.582	1.071	1.470
241- 270	0.900	0.048	1.408	0.291	0.557	0.719
271- 300	2.380	1.209	0.018	1.754	1.754	2.057
301- 330	1.786	2.667	7.681	1.786	2.200	2.200
331- 360	0.000	0.023	1.289	0.000	0.023	0.000
361- 390	0.108	0.108	1.531	0.108	0.250	0.250
391- 420	0.833	0.516	2.370	0.516	1.241	0.833
421- 450	3.240	2.462	5.261	2.462	4.167	4.167
451- 480	0.048	0.182	0.000	0.182	0.000	0.000
481- 510	8.471	5.263	8.471	5.263	8.471	8.471
511- 540	0.000	0.250	0.067	0.250	0.000	0.000
541- 570	1.333	2.571	1.923	2.571	1.333	1.333
571- 600	0.100	0.750	0.364	0.750	0.100	0.100
601- 700	0.640	0.033	0.036	0.033	0.640	0.640
701- 800	0.250	0.200	0.053	0.053	0.250	0.250
801- 900	10.000	2.571	2.571	3.769	10.000	10.000
901-1000	0.143	1.600	1.600	1.000	0.143	0.143
TOTAL CHISQ	133.8	54.7	174.3	59.9	102.3	116.5
D.F.	24	26	26	25	24	24
P<	0.001	0.001	0.001	0.001	0.001	0.001

Table (3.4-e)

***** 73/4TH QUARTER *****

ACTUAL AND EXPECTED CLAIMS COST BY DIFF. METHODS

TWO PARAMETER LOGNORMAL DIS.

AMOUNT £	ACTUAL	MOMNTS	MAXLIK.	7&93%QS	27&73%QS	GRAPHCL	REG.
1- 30	7409.	4929.	7037.	10137.	6277.	5409.	5177.
31- 60	23569.	27391.	27891.	28847.	27618.	27983.	27710.
61- 90	34805.	37825.	34579.	31936.	35410.	37448.	37674.
91- 120	37874.	39351.	34709.	30595.	36081.	38507.	38929.
121- 150	32384.	37127.	32520.	28048.	33875.	36043.	36585.
151- 180	35251.	33596.	29624.	25321.	30948.	32603.	33100.
181- 210	28934.	29911.	26588.	22678.	27761.	28934.	29325.
211- 240	23001.	26383.	23903.	20295.	24805.	25481.	25932.
241- 270	20695.	22995.	21206.	18140.	21973.	22484.	22740.
271- 300	16559.	20270.	19128.	16273.	19699.	19699.	19985.
301- 330	20823.	17668.	17037.	14829.	17668.	17352.	17352.
331- 360	15547.	15547.	15202.	13129.	15547.	15202.	15547.
361- 390	14644.	13893.	13893.	12016.	13893.	13518.	13518.
391- 420	14193.	12165.	12570.	10948.	12570.	11759.	12165.
421- 450	14807.	10887.	11323.	10016.	11323.	10452.	10452.
451- 480	9310.	9775.	10241.	9310.	10241.	9310.	9310.
481- 510	14369.	8423.	9414.	8423.	9414.	8423.	8423.
511- 540	7357.	7357.	8408.	7882.	8408.	7357.	7357.
541- 570	4444.	6666.	7777.	7222.	7777.	6666.	6666.
571- 600	5269.	5855.	7026.	6440.	7026.	5855.	5855.
601- 700	18864.	16262.	19515.	18214.	19515.	16262.	16262.
701- 800	13509.	12008.	15010.	14259.	14259.	12008.	12008.
801- 900	17010.	8505.	11907.	11907.	11056.	8505.	8505.
901-1000	5703.	6653.	9505.	9505.	8554.	6653.	6653.
1001-1100	4202.	5252.	7353.	7353.	7353.	5252.	5252.
1101-1200	4602.	3451.	5752.	6903.	5752.	3451.	3451.
1201-1300	1250.	2501.	5002.	5002.	5002.	3751.	2501.
1301-1400	4051.	2701.	4051.	4051.	4051.	2701.	2701.
1401-1500	1450.	1450.	2901.	4351.	2901.	1450.	1450.
1501-1600	0.	1550.	3101.	3101.	3101.	1550.	1550.
1601-1700	1650.	1650.	1650.	3301.	1650.	1650.	1650.
1701-1800	0.	1751.	1751.	1751.	1751.	1751.	1751.
1801-1900	0.	0.	1850.	1850.	1850.	1850.	0.
1901-2000	1950.	0.	1950.	1950.	1950.	0.	0.
2001-2100	0.	0.	2050.	2050.	2050.	0.	0.
2101-2200	0.	0.	2150.	2150.	0.	0.	0.
2201-2300	0.	0.	0.	2250.	0.	0.	0.
2301-2400	2350.	0.	0.	2350.	0.	0.	0.
TOTAL CL COST	457843.	451757.	465582.	434792.	469117.	447327.	447541.

Table (3.4-f)

***** 73/4TH QUARTER *****

EXPECTED LOSS BY DIFF. METHODS: (A-E)*(CLAIM AMOUNT)

TWO-PARAMETER LOGNORMAL DIS.

AMOUNT £	MOMNTS	MAX.LIK.	7693%QS	27673%QS	GRAPHCL	REG.
1- 30	2480.	372.	-2728.	1131.	1999.	2232.
31- 60	-3822.	-4322.	-5278.	-4049.	-4413.	-4140.
61- 90	-3020.	227.	2869.	-604.	-2642.	-2869.
91- 120	-1477.	3165.	7279.	1793.	-633.	-1055.
121- 150	-4742.	-135.	4336.	-1490.	-3658.	-4200.
151- 180	1655.	5627.	9930.	4303.	2648.	2151.
181- 210	-978.	2346.	6256.	1173.	0.	-391.
211- 240	-3382.	-902.	2706.	-1804.	-2480.	-2931.
241- 270	-2299.	-511.	2555.	-1277.	-1788.	-2044.
271- 300	-3711.	-2569.	286.	-3140.	-3140.	-3426.
301- 330	3155.	3786.	5994.	3155.	3470.	3470.
331- 360	0.	346.	2418.	0.	346.	0.
361- 390	751.	751.	2628.	751.	1126.	1126.
391- 420	2027.	1622.	3244.	1622.	2433.	2027.
421- 450	3919.	3484.	4790.	3484.	4355.	4355.
451- 480	-466.	-931.	0.	-931.	0.	0.
481- 510	5946.	4955.	5946.	4955.	5946.	5946.
511- 540	0.	-1051.	-525.	-1051.	0.	0.
541- 570	-2222.	-3333.	-2777.	-3333.	-2222.	-2222.
571- 600	-585.	-1756.	-1171.	-1756.	-585.	-585.
601- 700	2602.	-651.	651.	-651.	2602.	2602.
701- 800	1501.	-1501.	-751.	-751.	1501.	1501.
801- 900	8505.	5103.	5103.	5953.	8505.	8505.
901-1000	-951.	-3802.	-3802.	-2851.	-951.	-951.
1001-1100	-1050.	-3151.	-3151.	-3151.	-1050.	-1050.
1101-1200	1150.	-1150.	-2301.	-1150.	1150.	1150.
1201-1300	-1250.	-3751.	-3751.	-3751.	-2501.	-1250.
1301-1400	1350.	0.	0.	0.	1350.	1350.
1401-1500	0.	-1450.	-2901.	-1450.	0.	0.
1501-1600	-1550.	-3101.	-3101.	-3101.	-1550.	-1550.
1601-1700	0.	0.	-1650.	0.	0.	0.
1701-1800	-1751.	-1751.	-1751.	-1751.	-1751.	-1751.
1801-1900	0.	-1850.	-1850.	-1850.	-1850.	0.
1901-2000	1950.	0.	0.	0.	1950.	1950.
2001-2100	0.	-2050.	-2050.	-2050.	0.	0.
2101-2200	0.	-2150.	-2150.	0.	0.	0.
2201-2300	0.	0.	-2250.	0.	0.	0.
2301-2400	2350.	2350.	0.	2350.	2350.	2350.
TOTAL	6086.	-7739.	23051.	-11274.	10516.	10301.
TOTAL EXP. LOSS / TOTAL ACT. COST	1.33%	-1.69%	5.03%	-2.46%	2.30%	2.25%

Table (3.5)

***** 74/1ST QUARTER *****

FITTING THE TWO PARAMETER LOGNORMAL DIS. BY DIFF. METHODS

TABLE OF ESTIMATED PARAMETERS AND STATISTICS

METHOD--> STATISTICS	MOMNTS	MAXIMUM LIKELIHOOD	7893 % QUANTILES	27673 % QUANTILES	GRAPHICAL	REGRESSION
MEWHAT	4.589	4.497	4.342	4.517	4.511	4.581
SIGMA2HAT	0.842	1.089	1.338	0.973	0.975	0.853
ALPHA=MEAN	149.853	154.599	150.076	148.995	148.175	149.534
BETA=S.D.	172.204	216.991	251.666	191.132	190.411	173.485
COF OF VAR	1.149	1.404	1.677	1.283	1.285	1.160
MEDIAN	98.371	89.707	76.866	91.603	91.000	97.628
MODE	42.391	30.204	20.164	34.625	34.322	41.615
SKEWNESS	4.965	6.976	9.746	5.959	5.977	5.042
KURTOSIS	67.145	153.672	362.560	104.021	104.779	69.626
TOTAL CHISQ.	139.3	56.4	142.2	68.9	67.9	128.2
D.F.	23	24	24	23	23	23
$p <$	0.001	0.001	0.001	0.001	0.001	0.001
T	3876	1197	20559	10161	11812	5391
$\frac{T}{\text{ACT. COST}}$ =	1%	0.3%	5.6%	2.8%	3.2%	1.5%

Table (3.6)

***** 74/2ND QUARTER *****

FITTING THE TWO PARAMETER LOGNORMAL DIS. BY DIFF. METHODS

TABLE OF ESTIMATED PARAMETERS AND STATISTICS

METHOD-->	MOMNTS	MAXIMUM LIKELIHOOD	7893 % QUANTILES	27873 % QUANTILES	GRAPHICAL	REGRESSION
STATISTICS						
MEWHAT	4.620	4.531	4.377	4.576	4.564	4.609
SIGMA2HAT	0.801	1.057	1.306	0.938	0.913	0.818
ALPHA=MEAN	151.574	157.531	152.969	155.186	151.505	151.201
BETA=S.D.	167.980	215.867	250.932	193.447	184.977	170.169
COF OF VAR	1.108	1.370	1.640	1.247	1.221	1.125
MEDIAN	101.543	92.862	79.622	97.107	96.000	100.430
MODE	45.572	32.269	21.572	38.023	38.544	44.308
SKEWNESS	4.686	6.684	9.335	5.677	5.483	4.802
KURTOSIS	58.669	138.092	324.022	92.424	84.993	62.099
TOTAL CHISQ.	136.3	52.2	142.5	64.3	87.1	119.3
D.F.	23	24	25	24	23	24
P <	0.001	0.001	0.001	0.001	0.001	0.001
T	3786	-4373	16804	-1293	4025	5616
T / TAL ACT. COST =	1%	-1.2%	4.6%	-0.4%	1%	1.5%

Table (3.7)

***** 74/3RD QUARTER *****

FITTING THE TWO PARAMETER LOGNORMAL DIS. BY DIFF. METHODS

TABLE OF ESTIMATED PARAMETERS AND STATISTICS

METHOD-->	MOMNTS	MAXIMUM LIKELIHOOD	2893 % QUANTILES	27973 % QUANTILES	GRAPHICAL	REGRESSION
STATISTICS						
MEWHAT	4.698	4.622	4.469	4.652	4.654	4.674
SIGMA2HAT	0.829	1.062	1.253	0.920	0.971	0.864
ALPHA=MEAN	166.134	172.866	163.212	166.034	170.663	165.058
BETA=S.D.	188.202	237.711	258.079	203.982	218.674	193.457
COF OF VAR	1.136	1.375	1.581	1.229	1.281	1.172
MEDIAN	109.791	101.669	87.237	104.913	105.000	107.132
MODE	47.937	35.168	24.922	41.769	39.746	45.133
SKEWNESS	4.873	6.726	8.697	5.540	5.948	5.126
KURTOSIS	64.263	140.246	269.651	87.143	103.519	72.401
TOTAL CHISQ.	147.3	69.4	171.3	84.1	76.5	111.6
D.F.	24	26	26	24	24	24
P <	0.001	0.001	0.001	0.001	0.001	0.001
T	3556	-9108	19541	8812	-4503	8161
T / ACT. COST =	0.8%	-2%	4.2%	1.9%	-1%	1.8%

Table (3.8)

***** 74/4TH QUARTER *****

FITTING THE TWO PARAMETER LOGNORMAL DIS. BY DIFF. METHODS

TABLE OF ESTIMATED PARAMETERS AND STATISTICS

METHOD-->	MOMNTS	MAXIMUM LIKELIHOOD	7693 % QUANTILES	27873 % QUANTILES	GRAPHICAL	REGRESSION
STATISTICS						
MEWHAT	4.757	4.658	4.491	4.702	4.727	4.730
SIGMA2HAT	0.807	1.103	1.278	1.021	0.836	0.833
ALPHA=MEAN	174.186	182.958	168.915	183.668	171.600	171.769
BETA=S.D.	194.033	259.553	271.767	244.844	196.111	195.838
COF OF VAR	1.114	1.419	1.609	1.333	1.143	1.140
MEDIAN	116.361	105.410	89.168	110.215	113.000	113.264
MODE	51.927	34.990	24.848	39.687	49.001	49.248
SKEWNESS	4.724	7.111	8.991	6.368	4.921	4.902
KURTOSIS	59.784	161.274	293.892	122.451	65.763	65.176
TOTAL CHISQ.	215.3	61.3	166.1	63.8	160.8	167.6
D.F.	25	27	27	27	25	25
P<	0.001	0.001	0.001	0.001	0.001	0.001
T	6102	-15853	32546	-20139	12375	13712
$\frac{T}{\text{TOTAL ACT. COST}}$ =	1%	-3%	6%	-3.8%	2.3%	2.6%

Table (3.9)

***** 75/1ST QUARTER *****

FITTING THE TWO PARAMETER LOGNORMAL DIS. BY DIFF. METHODS

TABLE OF ESTIMATED PARAMETERS AND STATISTICS

METHOD-->	MOMNTS	MAXIMUM LIKELIHOOD	7893 % QUANTILES	27923 % QUANTILES	GRAPHICAL	REGRESSION
STATISTICS						
MEWHAT	4.746	4.668	4.504	4.711	4.745	4.704
SIGMA2HAT	0.819	1.076	1.248	1.012	0.848	0.865
ALPHA=MEAN	173.402	182.457	168.680	184.385	175.688	170.119
BETA=S.D.	195.301	253.698	265.805	243.961	202.913	199.396
COF OF VAR	1.126	1.390	1.576	1.323	1.155	1.172
MEDIAN	115.128	106.531	90.382	111.176	115.000	110.414
MODE	50.750	36.317	25.949	40.418	49.273	46.513
SKEWNESS	4.908	6.860	8.640	6.296	5.006	5.127
KURTOSIS	62.271	147.335	265.099	119.562	69.442	72.413
TOTAL CHISQ.	149.6	70.9	164.7	72.2	125.4	113.7
D.F.	24	27	27	27	24	24
P<	0.001	0.001	0.001	0.001	0.001	0.001
T	3370	-14909	24317	-21324	-368	14956
T / ... ACT. COST =	0.75%	-3.3%	5.4%	-4.7%	-0.1%	3.3%

Table (3.10)

***** 75/2ND QUARTER *****

FITTING THE TWO PARAMETER LOGNORMAL DIS. BY DIFF. METHODS

TABLE OF ESTIMATED PARAMETERS AND STATISTICS

METHOD-->	MOMNTS	MAXIMUM LIKELIHOOD	7&93 % QUANTILES	27&73 % QUANTILES	GRAPHICAL	REGRESSION
STATISTICS						
MEWHAT	4.778	4.687	4.544	4.720	4.754	4.754
SIGMA2HAT	0.830	1.103	1.270	1.021	0.828	0.854
ALPHA=MEAN	180.083	188.317	177.515	186.914	175.482	177.882
BETA=S.D.	204.790	267.145	284.045	249.142	199.193	206.621
COF OF VAR	1.137	1.419	1.600	1.333	1.135	1.162
MEDIAN	118.919	108.501	94.078	112.170	116.000	116.056
MODE	51.857	36.018	26.423	40.397	50.688	49.401
SKEWNESS	4.882	7.111	8.897	6.367	4.868	5.052
KURTOSIS	64.551	161.244	285.982	122.392	64.111	69.946
TOTAL CHISQ.	145.1	50.6	121.8	52.3	140.2	116.5
D.F.	24	27	27	27	24	27
P<	0.001	0.004	0.001	0.003	0.001	0.001
T	4871	-5951	25634	-10189	15935	11176
T / TOTAL ACT. COST =	1%	-1.3%	5.7%	-2.3%	3.5%	2.5%

*** TWO-PARAMETER LOGNORMAL DIS. ***

73/4TH QUARTER DATA

ESTIMATION BY MULTINOMIAL MAX.LIKELIHOOD METHOD :-

MEW = 4.5163 SIGMA2= 1.0553

MEAN= 155.080 S.D.= 212.237

AMOUNT £	ACTUAL CL. NO.	EXPECTED CL. NO.	ACTUAL-EXPECTED	EXPECTED LOSS	(A-E)**2/E
1- 30	478.	434.	44.	682.	4.461
31- 60	518.	613.	-95.	-4322.	14.723
61- 90	461.	463.	-2.	-151.	0.009
91- 120	359.	335.	24.	2532.	1.719
121- 150	239.	245.	-6.	-813.	0.147
151- 180	213.	182.	31.	5130.	5.280
181- 210	148.	139.	9.	1759.	0.583
211- 240	102.	107.	-5.	-1127.	0.234
241- 270	81.	84.	-3.	-767.	0.107
271- 300	58.	67.	-9.	-2569.	1.209
301- 330	66.	55.	11.	3470.	2.200
331- 360	45.	45.	0.	0.	0.000
361- 390	39.	37.	2.	751.	0.108
391- 420	35.	31.	4.	1622.	0.516
421- 450	34.	26.	8.	3484.	2.462
451- 480	20.	22.	-2.	-931.	0.182
481- 510	29.	19.	10.	4955.	5.263
511- 540	14.	16.	-2.	-1051.	0.250
541- 570	8.	14.	-6.	-3333.	2.571
571- 600	9.	12.	-3.	-1756.	0.750
601- 700	29.	30.	-1.	-651.	0.033
701- 800	18.	19.	-1.	-751.	0.053
801- 900	20.	13.	7.	5953.	3.769
901-1000	6.	9.	-3.	-2851.	1.000
1001-1100	4.	7.	-3.	-3151.	1.286
1101-1200	4.	5.	-1.	-1150.	0.200
1201-1300	1.	4.	-3.	-3751.	
1301-1400	3.	3.	0.	0.	
1401-1500	1.	2.	-1.	-1450.	1.778
1501-1600	0.	2.	-2.	-3101.	
1601-1700	1.	1.	0.	0.	
1701-1800	0.	1.	-1.	-1751.	
1801-1900	0.	1.	-1.	-1850.	
1901-2000	1.	1.	0.	0.	
2001-2100	0.	1.	-1.	-2050.	
2101-2200	0.	0.	0.	0.	
2201-2300	0.	0.	0.	0.	
2301-2400	1.	0.	1.	2350.	2.286
TOTAL	3045.	3045.		-6640.	53.375

D.F. = 25
P < 0.001

TOTAL EXP. LOSS / TOTAL ACT. COST = -1.5 %

KOL.-SMIRNOV D= 0.017
P<0.05

Table (5.12)

*** TWO-PARAMETER LOGNORMAL DIS. ***

74/1ST QUARTER DATA

ESTIMATION BY MULTINOMIAL MAX.LIKELIHOOD METHOD :-

MEW = 4.5088 SIGMA2= 1.0572

MEAN= 154.070 S.D.= 211.156

AMOUNT £	ACTUAL CL.. NO.	EXPECTED CL. NO.	ACTUAL-EXPECTED	EXPECTED LOSS	(A-E)**2/E
1- 30	381.	352.	29.	450.	2.389
31- 60	428.	493.	-65.	-2957.	8.570
61- 90	351.	372.	-21.	-1585.	1.185
91- 120	334.	268.	66.	6963.	16.254
121- 150	211.	195.	16.	2168.	1.313
151- 180	133.	145.	-12.	-1986.	0.993
181- 210	98.	110.	-12.	-2346.	1.309
211- 240	82.	85.	-3.	-677.	0.106
241- 270	54.	67.	-13.	-3321.	2.522
271- 300	52.	54.	-2.	-571.	0.074
301- 330	53.	43.	10.	3155.	2.326
331- 360	36.	35.	1.	346.	0.029
361- 390	29.	29.	0.	0.	0.000
391- 420	26.	24.	2.	811.	0.167
421- 450	22.	20.	2.	871.	0.200
451- 480	22.	17.	5.	2327.	1.471
481- 510	17.	15.	2.	991.	0.267
511- 540	10.	13.	-3.	-1576.	0.692
541- 570	19.	11.	8.	4444.	5.818
571- 600	4.	9.	-5.	-2927.	2.778
601- 700	26.	24.	2.	1301.	0.167
701- 800	21.	15.	6.	4503.	2.400
801- 900	11.	11.	0.	0.	0.000
901-1000	10.	7.	3.	2851.	1.286
1001-1100	5.	5.	0.	0.	0.000
1101-1200	2.	4.	-2.	-2301.	
1201-1300	1.	3.	-2.	-2501.	2.286
1301-1400	2.	2.	0.	0.	
1401-1500	0.	2.	-2.	-2901.	
1501-1600	0.	1.	-1.	-1550.	
1601-1700	0.	1.	-1.	-1650.	
1701-1800	1.	1.	0.	0.	2.286
TOTAL	2441.	2433.		2329.	56.886

D.F. = 24

P < 0.001

TOTAL EXP. LOSS / TOTAL ACT. COST = 0.6 %

KOL - SMIRNOV D = 0.023

P < 0.01

*** TWO-PARAMETER LOGNORMAL DIS. ***

74/2ND QUARTER DATA

ESTIMATION BY MULTINOMIAL MAX.LIKELIHOOD METHOD :-
MEW = 4.5462 SIGMA2= 1.0126

MEAN= 156.407 S.D.= 207.061

AMOUNT £	ACTUAL CL. NO.	EXPECTED CL. NO.	ACTUAL-EXPECTED	EXPECTED LOSS	(A-E)**2/E
1- 30	351.	312.	39.	605.	4.875
31- 60	380.	473.	-93.	-4231.	18.285
61- 90	382.	367.	15.	1132.	0.613
91- 120	295.	268.	27.	2848.	2.720
121- 150	211.	197.	14.	1897.	0.995
151- 180	142.	147.	-5.	-828.	0.170
181- 210	114.	112.	2.	391.	0.036
211- 240	101.	87.	14.	3157.	2.253
241- 270	57.	68.	-11.	-2810.	1.779
271- 300	51.	54.	-3.	-857.	0.167
301- 330	39.	44.	-5.	-1577.	0.568
331- 360	36.	36.	0.	0.	0.000
361- 390	25.	29.	-4.	-1502.	0.552
391- 420	24.	24.	0.	0.	0.000
421- 450	27.	21.	6.	2613.	1.714
451- 480	18.	17.	1.	466.	0.059
481- 510	21.	15.	6.	2973.	2.400
511- 540	17.	13.	4.	2102.	1.231
541- 570	12.	11.	1.	555.	0.091
571- 600	11.	9.	2.	1171.	0.444
601- 700	30.	23.	7.	4553.	2.130
701- 800	13.	15.	-2.	-1501.	0.267
801- 900	11.	10.	1.	851.	0.100
901-1000	4.	7.	-3.	-2851.	1.286
1001-1100	7.	5.	2.	2101.	0.800
1101-1200	0.	4.	-4.	-4602.	
1201-1300	1.	3.	-2.	-2501.	5.143
1301-1400	1.	2.	-1.	-1350.	
1401-1500	0.	2.	-2.	-2901.	
1501-1600	1.	1.	0.	0.	
1601-1700	0.	1.	-1.	-1650.	
1701-1800	0.	1.	-1.	-1751.	
1801-1900	1.	1.	0.	0.	3.125
TOTAL	2383.	2379.		-3498.	51.803

D.F. = 24

P < 0.001

TOTAL EXP. LOSS
--------------- = -1.0 %
TOTAL ACT. COST

KOL - SMIRNOV D = 0.023

P < 0.01

*** TWO-PARAMETER LOGNORMAL DIS. ***

74/3RD QUARTER DATA

ESTIMATION BY MULTINOMIAL MAX.LIKELIHOOD METHOD :-

MEW = 4.6371 SIGMA2= 1.0112

MEAN= 171.169 S.D.= 226.355

AMOUNT £	ACTUAL CL. NO.	EXPECTED CL. NO.	ACTUAL-EXPECTED	EXPECTED LOSS	(A-E)**2/E
1- 30	362.	315.	47.	728.	7.013
31- 60	427.	518.	-91.	-4140.	15.986
61- 90	383.	421.	-38.	-2869.	3.430
91- 120	356.	317.	39.	4115.	4.798
121- 150	283.	238.	45.	6097.	8.508
151- 180	194.	181.	13.	2151.	0.934
181- 210	137.	140.	-3.	-587.	0.064
211- 240	97.	110.	-13.	-2931.	1.536
241- 270	86.	87.	-1.	-256.	0.011
271- 300	71.	70.	1.	286.	0.014
301- 330	64.	57.	7.	2208.	0.860
331- 360	45.	47.	-2.	-691.	0.085
361- 390	44.	39.	5.	1877.	0.641
391- 420	25.	33.	-8.	-3244.	1.939
421- 450	26.	27.	-1.	-436.	0.037
451- 480	22.	23.	-1.	-466.	0.043
481- 510	25.	20.	5.	2477.	1.250
511- 540	14.	17.	-3.	-1576.	0.529
541- 570	14.	15.	-1.	-555.	0.067
571- 600	17.	13.	4.	2342.	1.231
601- 700	32.	32.	0.	0.	0.000
701- 800	34.	21.	13.	9756.	8.048
801- 900	17.	15.	2.	1701.	0.267
901-1000	4.	10.	-6.	-5703.	3.600
1001-1100	9.	7.	2.	2101.	0.571
1101-1200	4.	5.	-1.	-1150.	0.200
1201-1300	0.	4.	-4.	-5002.	
1301-1400	1.	3.	-2.	-2701.	5.143
1401-1500	3.	2.	1.	1450.	
1501-1600	0.	2.	-2.	-3101.	
1601-1700	1.	2.	-1.	-1650.	0.667
1701-1800	0.	1.	-1.	-1751.	
1801-1900	0.	1.	-1.	-1850.	
1901-2000	0.	1.	-1.	-1950.	
2001-2100	0.	1.	-1.	-2050.	
2101-2200	1.	1.	0.	0.	
2201-2300	0.	0.	0.	0.	
2301-2400	0.	0.	0.	0.	
2401-2500	0.	0.	0.	0.	
2501-2600	1.	0.	1.	2550.	1.800
TOTAL	2799.	2796.		-4818.	69.274

D.F. = 26

P < 0.001

TOTAL EXP. LOSS
------------------ = -1.0 %
TOTAL ACT. COST

KOL - SMIRNOV D =0.029

P < 0.01

*** TWO-PARAMETER LOGNORMAL DIS. ***

74/4TH QUARTER DATA

ESTIMATION BY MULTINOMIAL MAX.LIKELIHOOD METHOD :-

MEW = 4.6722 SIGMA2= 1.0562

MEAN= 181.329 S.D.= 248.325

AMOUNT £	ACTUAL CL. NO.	EXPECTED CL. NO.	ACTUAL-EXPECTED	EXPECTED LOSS	(A-E)**2/E
1- 30	394.	340.	54.	837.	8.576
31- 60	452.	547.	-95.	-4322.	16.499
61- 90	426.	447.	-21.	-1585.	0.987
91- 120	348.	339.	9.	950.	0.239
121- 150	272.	257.	15.	2032.	0.875
151- 180	219.	198.	21.	3475.	2.227
181- 210	154.	154.	0.	0.	0.000
211- 240	124.	122.	2.	451.	0.033
241- 270	105.	98.	7.	1788.	0.500
271- 300	78.	79.	-1.	-286.	0.013
301- 330	75.	65.	10.	3155.	1.538
331- 360	58.	54.	4.	1382.	0.296
361- 390	55.	45.	10.	3755.	2.222
391- 420	29.	38.	-9.	-3649.	2.132
421- 450	43.	32.	11.	4790.	3.781
451- 480	24.	28.	-4.	-1862.	0.571
481- 510	22.	24.	-2.	-991.	0.167
511- 540	24.	20.	4.	2102.	0.800
541- 570	19.	18.	1.	555.	0.056
571- 600	14.	16.	-2.	-1171.	0.250
601- 700	42.	39.	3.	1951.	0.231
701- 800	28.	26.	2.	1501.	0.154
801- 900	20.	18.	2.	1701.	0.222
901-1000	17.	13.	4.	3802.	1.231
1001-1100	8.	10.	-2.	-2101.	0.400
1101-1200	5.	7.	-2.	-2301.	0.571
1201-1300	1.	5.	-4.	-5002.	3.200
1301-1400	5.	4.	1.	1350.	
1401-1500	0.	3.	-3.	-4351.	0.800
1501-1600	1.	3.	-2.	-3101.	
1601-1700	0.	2.	-2.	-3301.	
1701-1800	0.	2.	-2.	-3501.	5.143
1801-1900	0.	1.	-1.	-1850.	
1901-2000	0.	1.	-1.	-1950.	
2001-2100	1.	1.	0.	0.	
2101-2200	0.	1.	-1.	-2150.	
2201-2300	0.	1.	-1.	-2250.	
2301-2400	1.	1.	0.	0.	2.667
TOTAL	3064.	3059.		-10147.	56.382

D.F. = 27

P < 0.001

TOTAL EXP. LOSS / TOTAL ACT. COST = -1.9 %

KOL - SMIRNOV D = 0.020

P < 0.02

Table (3.16).

75/1ST QUARTER DATA

ESTIMATION BY MULTINOMIAL MAX.LIKELIHOOD METHOD :-

MEW = 4.6839 SIGMA2= 1.0241

MEAN= 180.540 S.D.= 241.173

AMOUNT £	ACTUAL CL. NO.	EXPECTED CL. NO.	ACTUAL-EXPECTED	EXPECTED LOSS	(A-E)**2/E
1- 30	324.	275.	49.	760.	8.731
31- 60	387.	463.	-76.	-3458.	12.475
61- 90	345.	384.	-39.	-2944.	3.961
91- 120	289.	293.	-4.	-422.	0.055
121- 150	253.	223.	30.	4065.	4.036
151- 180	187.	171.	16.	2648.	1.497
181- 210	138.	133.	5.	978.	0.188
211- 240	114.	105.	9.	2029.	0.771
241- 270	93.	84.	9.	2299.	0.964
271- 300	67.	68.	-1.	-286.	0.015
301- 330	63.	56.	7.	2208.	0.875
331- 360	44.	46.	-2.	-691.	0.087
361- 390	44.	39.	5.	1877.	0.641
391- 420	35.	32.	3.	1216.	0.281
421- 450	25.	28.	-3.	-1306.	0.321
451- 480	26.	23.	3.	1396.	0.391
481- 510	18.	20.	-2.	-991.	0.200
511- 540	18.	17.	1.	525.	0.059
541- 570	22.	15.	7.	3888.	3.267
571- 600	17.	13.	4.	2342.	1.231
601- 700	39.	33.	6.	3903.	1.091
701- 800	19.	22.	-3.	-2251.	0.409
801- 900	18.	15.	3.	2551.	0.600
901-1000	12.	11.	1.	951.	0.091
1001-1100	3.	8.	-5.	-5252.	3.125
1101-1200	1.	6.	-5.	-5752.	4.167
1201-1300	1.	4.	-3.	-3751.	
1301-1400	0.	3.	-3.	-4051.	5.142
1401-1500	1.	3.	-2.	-2901.	
1501-1600	0.	2.	-2.	-3101.	
1601-1700	1.	2.	-1.	-1650.	3.571
1701-1800	0.	1.	-1.	-1751.	
1801-1900	0.	1.	-1.	-1850.	
1901-2000	0.	1.	-1.	-1950.	
2001-2100	1.	1.	0.	0.	
2101-2200	1.	1.	0.	0.	
2201-2300	0.	1.	-1.	-2250.	
2301-2400	0.	0.	0.	0.	
2401-2500	0.	0.	0.	0.	
2501-2600	0.	0.	0.	0.	
2601-2700	0.	0.	0.	0.	
2701-2800	0.	0.	0.	0.	
2801-2900	0.	0.	0.	0.	
2901-3000	0.	0.	0.	0.	
3001-3100	0.	0.	0.	0.	
3101-3200	0.	0.	0.	0.	
3201-3300	0.	0.	0.	0.	
3301-3400	0.	0.	0.	0.	
3401-3500	0.	0.	0.	0.	
3501-3600	1.	0.	1.	3550.	1.500
TOTAL	2607.	2603.		-9423.	59.742

KOL-SMIRNOV D=0.027 P<0.01 D.F. = 26 P < 0.001

$$\frac{\text{TOTAL EXP. LOSS}}{\text{TOTAL ACT. COST}} = -2.1\ \%$$

Table (3.17)

*** TWO-PARAMETER LOGNORMAL DIS. ***

75/2ND QUARTER DATA

ESTIMATION BY MULTINOMIAL MAX.LIKELIHOOD METHOD :-

MEW = 4.7006 SIGMA2= 1.0562

MEAN= 186.551 S.D.= 255.479

AMOUNT £	ACTUAL CL. NO.	EXPECTED CL. NO.	ACTUAL-EXPECTED	EXPECTED LOSS	(A-E)**2/E
1- 30	302.	264.	38.	589.	5.470
31- 60	374.	435.	-61.	-2775.	8.554
61- 90	332.	360.	-28.	-2114.	2.178
91- 120	277.	276.	1.	105.	0.004
121- 150	235.	211.	24.	3252.	2.730
151- 180	187.	163.	24.	3972.	3.534
181- 210	122.	127.	-5.	-978.	0.197
211- 240	110.	101.	9.	2029.	0.802
241- 270	80.	81.	-1.	-256.	0.012
271- 300	72.	66.	6.	1713.	0.545
301- 330	47.	55.	-8.	-2524.	1.164
331- 360	39.	45.	-6.	-2073.	0.800
361- 390	40.	38.	2.	751.	0.105
391- 420	38.	32.	6.	2433.	1.125
421- 450	29.	27.	2.	871.	0.148
451- 480	21.	23.	-2.	-931.	0.174
481- 510	30.	20.	10.	4955.	5.000
511- 540	19.	17.	2.	1051.	0.235
541- 570	17.	15.	2.	1111.	0.267
571- 600	11.	13.	-2.	-1171.	0.308
601- 700	36.	34.	2.	1301.	0.118
701- 800	22.	23.	-1.	-751.	0.043
801- 900	22.	16.	6.	5103.	2.250
901-1000	11.	11.	0.	0.	0.000
1001-1100	4.	8.	-4.	-4202.	2.000
1101-1200	3.	6.	-3.	-3451.	1.500
1201-1300	3.	5.	-2.	-2501.	0.800
1301-1400	6.	4.	2.	2701.	
1401-1500	2.	3.	-1.	-1450.	1.143
1501-1600	1.	2.	-1.	-1550.	
1601-1700	0.	2.	-2.	-3301.	
1701-1800	1.	1.	0.	0.	1.800
1801-1900	0.	1.	-1.	-1850.	
1901-2000	1.	1.	0.	0.	
2001-2100	0.	1.	-1.	-2050.	
2101-2200	0.	1.	-1.	-2150.	
2201-2300	1.	1.	0.	0.	1.800
TOTAL	2495.	2489.		-4142.	44.605

D.F. = 27

P < 0.02

TOTAL EXP. LOSS
--------------- = -0.9 %
TOTAL ACT. COST

KOL - SMIRNOV D = 0.020

P < 0.02

Table (3.18)

*** TWO-PARAMETER LOGNORMAL DIS. ***

PREDICTION OF 74/4TH QUARTER CLAIMS COST
USING 73/4TH QUARTER MULTINOMIAL MAXLIK. PARAMETERS :

MEW= 4.516 SIGMA2=1.056
INFLATION RATE I= 0.0% CALCULATED FROM :
INFLATION IGNORED.

PREDICTION PARAMETERS ARE :- MEW= 4.516 SIGMA2=1.056
MEAN CLAIM AMOUNT= 155.09 S.D.=212.36
ACTUAL 74/4TH PARAMETERS :- MEW= 4.672 SIGMA2=1.056
MEAN CLAIM AMOUNT= 181.27 S.D.=248.21

AMOUNT £	ACT.NO.	EXP.NO.	A-E	EXP.LOSS	(A-E)**2/E
1- 30	394.	437.	-43.	-667.	4.231
31- 60	452.	616.	-164.	-7462.	43.662
61- 90	426.	466.	-40.	-3020.	3.433
91- 120	348.	337.	11.	1160.	0.359
121- 150	272.	246.	26.	3523.	2.748
151- 180	219.	183.	36.	5958.	7.082
181- 210	154.	139.	15.	2933.	1.619
211- 240	124.	108.	16.	3608.	2.370
241- 270	105.	85.	20.	5110.	4.706
271- 300	78.	68.	10.	2855.	1.471
301- 330	75.	55.	20.	6310.	7.273
331- 360	58.	45.	13.	4491.	3.756
361- 390	55.	37.	18.	6759.	8.757
391- 420	29.	31.	-2.	-811.	0.129
421- 450	43.	26.	17.	7403.	11.115
451- 480	24.	22.	2.	931.	0.182
481- 510	22.	19.	3.	1486.	0.474
511- 540	24.	16.	8.	4204.	4.000
541- 570	19.	14.	5.	2777.	1.786
571- 600	14.	12.	2.	1171.	0.333
601- 700	42.	30.	12.	7806.	4.800
701- 800	28.	20.	8.	6004.	3.200
801- 900	20.	13.	7.	5953.	3.769
901-1000	17.	9.	8.	7604.	7.111
1001-1100	8.	7.	1.	1050.	0.143
1101-1200	5.	5.	0.	0.	0.000
1201-1300	1.	4.	-3.	-3751.	
1301-1400	5.	3.	2.	2701.	
1401-1500	0.	2.	-2.	-2901.	
1501-1600	1.	2.	-1.	-1550.	
1601-1700	0.	1.	-1.	-1650.	
1701-1800	0.	1.	-1.	-1751.	
1801-1900	0.	1.	-1.	-1850.	
1901-2000	0.	1.	-1.	-1950.	
2001-2100	1.	1.	0.	0.	
2101-2200	0.	0.	0.	0.	
2201-2300	0.	0.	0.	0.	
2301-2400	1.	0.	1.	2350.	
TOTAL	3064	3062		66786.	128.509

CHI SQ. STAT. =132.7 , D.F.=28 P<0.001

TOTAL ACTUAL COST = 533707.
TOTAL EXPECTED COST = 466921.

TOTAL EXP. LOSS
--------------- = 12.51 %
TOTAL ACT. COST

*** TWO-PARAMETER LOGNORMAL DIS. *** Table (3.19)

PREDICTION OF 75/1ST QUARTER CLAIMS COST
USING 74/1ST QUARTER MULTINOMIAL MAXLIK. PARAMETERS :

MEW= 4.509 SIGMA2=1.057
INFLATION RATE I= 0.0% CALCULATED FROM :
INFLATION IGNORED.

PREDICTION PARAMETERS ARE :- MEW= 4.509 SIGMA2=1.057
MEAN CLAIM AMOUNT= 154.08 S.D.=211.14
ACTUAL 75/1ST PARAMETERS :- MEW= 4.684 SIGMA2=1.024
MEAN CLAIM AMOUNT= 180.55 S.D.=241.17

AMOUNT £	ACT.NO.	EXP.NO.	A-E	EXP.LOSS	(A-E)**2/E
1- 30	324.	376.	-52.	-806.	7.191
31- 60	387.	527.	-140.	-6370.	37.192
61- 90	345.	397.	-52.	-3926.	6.811
91- 120	289.	286.	3.	316.	0.031
121- 150	253.	209.	44.	5962.	9.263
151- 180	187.	155.	32.	5296.	6.606
181- 210	138.	118.	20.	3910.	3.390
211- 240	114.	91.	23.	5186.	5.813
241- 270	93.	72.	21.	5365.	6.125
271- 300	67.	57.	10.	2855.	1.754
301- 330	63.	46.	17.	5363.	6.283
331- 360	44.	38.	6.	2073.	0.947
361- 390	44.	31.	13.	4681.	5.452
391- 420	35.	26.	9.	3649.	3.115
421- 450	25.	22.	3.	1306.	0.409
451- 480	26.	18.	8.	3724.	3.556
481- 510	18.	16.	2.	991.	0.250
511- 540	18.	13.	5.	2627.	1.923
541- 570	22.	12.	10.	5555.	8.333
571- 600	17.	10.	7.	4098.	4.900
601- 700	39.	25.	14.	9107.	7.840
701- 800	19.	16.	3.	2251.	0.563
801- 900	18.	11.	7.	5953.	4.455
901-1000	12.	8.	4.	3802.	2.000
1001-1100	3.	6.	-3.	-3151.	1.500
1101-1200	1.	4.	-3.	-3451.	
1201-1300	1.	3.	-2.	-2501.	
1301-1400	0.	2.	-2.	-2701.	
1401-1500	1.	2.	-1.	-1450.	
1501-1600	0.	1.	-1.	-1550.	
1601-1700	1.	1.	0.	0.	
1701-1800	0.	1.	-1.	-1751.	
1801-1900	0.	1.	-1.	-1850.	
1901-2000	0.	1.	-1.	-1950.	
2001-2100	1.	0.	1.	2050.	
2101-2200	1.	0.	1.	2150.	
2201-2300	0.	0.	0.	0.	
2301-2400	0.	0.	0.	0.	
2401-2500	0.	0.	0.	0.	
2501-2600	0.	0.	0.	0.	
2601-2700	0.	0.	0.	0.	
2701-2800	0.	0.	0.	0.	
2801-2900	0.	0.	0.	0.	
2901-3000	0.	0.	0.	0.	
3001-3100	0.	0.	0.	0.	
3101-3200	0.	0.	0.	0.	
3201-3300	0.	0.	0.	0.	
3301-3400	0.	0.	0.	0.	
3401-3500	0.	0.	0.	0.	
3501-3600	1.	0.	1.	3550.	
TOTAL	2607	2602		60567.	135.703

Table (3.19) - continued

CHI SQ. STAT. =141.7 , D.F.= 27 P<0.001

TOTAL ACTUAL COST = 452059.
TOTAL EXPECTED COST = 391491.

$$\frac{\text{TOTAL EXP. LOSS}}{\text{TOTAL ACT. COST}} = 13.40\ \%$$

Table (3.20)

*** TWO-PARAMETER LOGNORMAL DIS. ***

PREDICTION OF 75/2ND QUARTER CLAIMS COST
USING 74/2ND QUARTER MULTINOMIAL MAXLIK. PARAMETERS :

MEW= 4.546 SIGMA2=1.013
INFLATION RATE I= 0.0% CALCULATED FROM :
INFLATION IGNORED.

PREDICTION PARAMETERS ARE :- MEW= 4.546 SIGMA2=1.013
MEAN CLAIM AMOUNT= 156.41 S.D.=207.14
ACTUAL 75/2ND PARAMETERS :- MEW= 4.701 SIGMA2=1.056
MEAN CLAIM AMOUNT= 186.61 S.D.=255.51

AMOUNT £	ACT.NO.	EXP.NO.	A-E	EXP.LOSS	(A-E)**2/E
1- 30	302.	327.	-25.	-388.	1.911
31- 60	374.	496.	-122.	-5551.	30.008
61- 90	332.	384.	-52.	-3926.	7.042
91- 120	277.	281.	-4.	-422.	0.057
121- 150	235.	206.	29.	3929.	4.083
151- 180	187.	154.	33.	5461.	7.071
181- 210	122.	117.	5.	978.	0.214
211- 240	110.	91.	19.	4284.	3.967
241- 270	80.	71.	9.	2299.	1.141
271- 300	72.	57.	15.	4282.	3.947
301- 330	47.	46.	1.	316.	0.022
331- 360	39.	37.	2.	691.	0.108
361- 390	40.	31.	9.	3379.	2.613
391- 420	38.	26.	12.	4866.	5.538
421- 450	29.	21.	8.	3484.	3.048
451- 480	21.	18.	3.	1396.	0.500
481- 510	30.	15.	15.	7432.	15.000
511- 540	19.	13.	6.	3153.	2.769
541- 570	17.	11.	6.	3333.	3.273
571- 600	11.	10.	1.	585.	0.100
601- 700	36.	24.	12.	7806.	6.000
701- 800	22.	16.	6.	4503.	2.250
801- 900	22.	11.	11.	9355.	11.000
901-1000	11.	7.	4.	3802.	2.286
1001-1100	4.	5.	-1.	-1050.	0.200
1101-1200	3.	4.	-1.	-1150.	
1201-1300	3.	3.	0.	0.	
1301-1400	6.	2.	4.	5402.	
1401-1500	2.	2.	0.	0.	
1501-1600	1.	1.	0.	0.	
1601-1700	0.	1.	-1.	-1650.	
1701-1800	1.	1.	0.	0.	
1801-1900	0.	1.	-1.	-1850.	
1901-2000	1.	1.	0.	0.	
2001-2100	0.	0.	0.	0.	
2101-2200	0.	0.	0.	0.	
2201-2300	1.	0.	1.	2250.	
TOTAL	2495	2491		67002.	114.147

CHI SQ. STAT. = 118.5 , D.F. = 27 $P<0.001$

TOTAL ACTUAL COST = 449308.
TOTAL EXPECTED COST = 382306.

TOTAL EXP. LOSS / TOTAL ACT. COST = 14.91 %

Table (5.21)

*** TWO-PARAMETER LOGNORMAL DIS. ***

PREDICTION OF 74/4TH QUARTER CLAIMS COST
USING 73/4TH QUARTER MULTINOMIAL MAXLIK. PARAMETERS :

MEW= 4.516 SIGMA2=1.056
INFLATION RATE I=29.2% CALCULATED FROM :
INDEX OF MOTOR VEHICLES REPAIRS COST.

PREDICTION PARAMETERS ARE :- MEW= 4.772 SIGMA2=1.056
MEAN CLAIM AMOUNT= 200.38 S.D.=274.36
ACTUAL 74/4TH PARAMETERS :- MEW= 4.672 SIGMA2=1.056
MEAN CLAIM AMOUNT= 181.27 S.D.=248.21

AMOUNT £	ACT.NO.	EXP.NO.	A-E	EXP.LOSS	(A-E)**2/E
1- 30	394.	287.	107.	1658.	39.892
31- 60	452.	501.	-49.	-2229.	4.792
61- 90	426.	430.	-4.	-302.	0.037
91- 120	348.	337.	11.	1160.	0.359
121- 150	272.	262.	10.	1355.	0.382
151- 180	219.	205.	14.	2317.	0.956
181- 210	154.	162.	-8.	-1564.	0.395
211- 240	124.	130.	-6.	-1353.	0.277
241- 270	105.	106.	-1.	-256.	0.009
271- 300	78.	87.	-9.	-2569.	0.931
301- 330	75.	72.	3.	947.	0.125
331- 360	58.	60.	-2.	-691.	0.067
361- 390	55.	51.	4.	1502.	0.314
391- 420	29.	43.	-14.	-5677.	4.558
421- 450	43.	37.	6.	2613.	0.973
451- 480	24.	32.	-8.	-3724.	2.000
481- 510	22.	27.	-5.	-2477.	0.926
511- 540	24.	24.	0.	0.	0.000
541- 570	19.	21.	-2.	-1111.	0.190
571- 600	14.	18.	-4.	-2342.	0.889
601- 700	42.	47.	-5.	-3252.	0.532
701- 800	28.	32.	-4.	-3002.	0.500
801- 900	20.	22.	-2.	-1701.	0.182
901-1000	17.	16.	1.	951.	0.063
1001-1100	8.	12.	-4.	-4202.	1.333
1101-1200	5.	9.	-4.	-4602.	1.778
1201-1300	1.	7.	-6.	-7503.	5.143
1301-1400	5.	5.	0.	0.	0.000
1401-1500	0.	4.	-4.	-5802.	
1501-1600	1.	3.	-2.	-3101.	
1601-1700	0.	3.	-3.	-4951.	
1701-1800	0.	2.	-2.	-3501.	
1801-1900	0.	2.	-2.	-3701.	
1901-2000	0.	1.	-1.	-1950.	
2001-2100	1.	1.	0.	0.	
2101-2200	0.	1.	-1.	-2150.	
2201-2300	0.	1.	-1.	-2250.	
2301-2400	1.	1.	0.	0.	
2401-2500	0.	1.	-1.	-2450.	
2501-2600	0.	1.	-1.	-2550.	
TOTAL	3064	3063		-68464.	67.603

CHI SQ. STAT. =83.2 , D.F.=31 p<0.001

TOTAL ACTUAL COST = 533707.
TOTAL EXPECTED COST = 602172.

$$\frac{\text{TOTAL EXP. LOSS}}{\text{TOTAL ACT. COST}} = -12.83\ \%$$

*** TWO-PARAMETER LOGNORMAL DIS. *** Table (5.22)

PREDICTION OF 75/1ST QUARTER CLAIMS COST
USING 74/1ST QUARTER MULTINOMIAL MAXLIK. PARAMETERS :

MEW= 4.509 SIGMA2=1.057
INFLATION RATE I=33.9% CALCULATED FROM :
INDEX OF MOTOR VEHICLES REPAIRS COST.

PREDICTION PARAMETERS ARE :- MEW= 4.801 SIGMA2=1.057
MEAN CLAIM AMOUNT= 206.32 S.D.=282.72
ACTUAL 75/1ST PARAMETERS :- MEW= 4.684 SIGMA2=1.024
MEAN CLAIM AMOUNT= 180.55 S.D.=241.17

AMOUNT £	ACT.NO.	EXP.NO.	A-E	EXP.LOSS	(A-E)**2/E
1- 30	324.	233.	91.	1410.	35.541
31- 60	387.	415.	-28.	-1274.	1.889
61- 90	345.	361.	-16.	-1208.	0.709
91- 120	289.	286.	3.	316.	0.031
121- 150	253.	223.	30.	4065.	4.036
151- 180	187.	176.	11.	1820.	0.688
181- 210	138.	140.	-2.	-391.	0.029
211- 240	114.	113.	1.	225.	0.009
241- 270	93.	92.	1.	256.	0.011
271- 300	67.	75.	-8.	-2284.	0.853
301- 330	63.	63.	0.	0.	0.000
331- 360	44.	53.	-9.	-3109.	1.528
361- 390	44.	44.	0.	0.	0.000
391- 420	35.	38.	-3.	-1216.	0.237
421- 450	25.	32.	-7.	-3048.	1.531
451- 480	26.	28.	-2.	-931.	0.143
481- 510	18.	24.	-6.	-2973.	1.500
511- 540	18.	21.	-3.	-1576.	0.429
541- 570	22.	18.	4.	2222.	0.889
571- 600	17.	16.	1.	585.	0.063
601- 700	39.	41.	-2.	-1301.	0.098
701- 800	19.	28.	-9.	-6754.	2.893
801- 900	18.	20.	-2.	-1701.	0.200
901-1000	12.	14.	-2.	-1901.	0.286
1001-1100	3.	11.	-8.	-8404.	5.818
1101-1200	1.	8.	-7.	-8053.	6.125
1201-1300	1.	6.	-5.	-6252.	4.167
1301-1400	0.	5.	-5.	-6752.	5.000
1401-1500	1.	4.	-3.	-4351.	
1501-1600	0.	3.	-3.	-4651.	
1601-1700	1.	2.	-1.	-1650.	
1701-1800	0.	2.	-2.	-3501.	
1801-1900	0.	2.	-2.	-3701.	
1901-2000	0.	1.	-1.	-1950.	
2001-2100	1.	1.	0.	0.	
2101-2200	1.	1.	0.	0.	
2201-2300	0.	1.	-1.	-2250.	
2301-2400	0.	1.	-1.	-2350.	
2401-2500	0.	1.	-1.	-2450.	
2501-2600	0.	0.	0.	0.	
2601-2700	0.	0.	0.	0.	
2701-2800	0.	0.	0.	0.	
2801-2900	0.	0.	0.	0.	
2901-3000	0.	0.	0.	0.	
3001-3100	0.	0.	0.	0.	
3101-3200	0.	0.	0.	0.	
3201-3300	0.	0.	0.	0.	
3301-3400	0.	0.	0.	0.	
3401-3500	0.	0.	0.	0.	
3501-3600	1.	0.	1.	3550.	
TOTAL	2607	2603		-71538.	74.701

Table (3.22) - continued

CHI SQ. STAT. =85.1 , D.F.=30 $P<0.001$

TOTAL ACTUAL COST = 452059.
TOTAL EXPECTED COST = 523597.

TOTAL EXP. LOSS
----------------=-15.82 %
TOTAL ACT. COST

Table (3.23)

*** TWO-PARAMETER LOGNORMAL DIS. ***

PREDICTION OF 75/2ND QUARTER CLAIMS COST
USING 74/2ND QUARTER MULTINOMIAL MAXLIK. PARAMETERS :

MEW= 4.546 SIGMA2=1.013
INFLATION RATE I=33.3% CALCULATED FROM :
INDEX OF MOTOR VEHICLES REPAIRS COST.

PREDICTION PARAMETERS ARE :- MEW= 4.833 SIGMA2=1.013
MEAN CLAIM AMOUNT= 208.50 S.D.=276.12
ACTUAL 75/2ND PARAMETERS :- MEW= 4.701 SIGMA2=1.056
MEAN CLAIM AMOUNT= 186.61 S.D.=255.51

AMOUNT £	ACT.NO.	EXP.NO.	A-E	EXP.LOSS	(A-E)**2/E
1- 30	302.	199.	103.	1596.	53.312
31- 60	374.	385.	-11.	-500.	0.314
61- 90	332.	345.	-13.	-982.	0.490
91- 120	277.	277.	0.	0.	0.000
121- 150	235.	219.	16.	2168.	1.169
151- 180	187.	173.	14.	2317.	1.133
181- 210	122.	138.	-16.	-3128.	1.855
211- 240	110.	111.	-1.	-225.	0.009
241- 270	80.	91.	-11.	-2810.	1.330
271- 300	72.	75.	-3.	-857.	0.120
301- 330	47.	62.	-15.	-4732.	3.629
331- 360	39.	52.	-13.	-4491.	3.250
361- 390	40.	44.	-4.	-1502.	0.364
391- 420	38.	37.	1.	406.	0.027
421- 450	29.	32.	-3.	-1306.	0.281
451- 480	21.	27.	-6.	-2793.	1.333
481- 510	30.	24.	6.	2973.	1.500
511- 540	19.	21.	-2.	-1051.	0.190
541- 570	17.	18.	-1.	-555.	0.056
571- 600	11.	16.	-5.	-2927.	1.562
601- 700	36.	40.	-4.	-2602.	0.400
701- 800	22.	27.	-5.	-3752.	0.926
801- 900	22.	19.	3.	2551.	0.474
901-1000	11.	14.	-3.	-2851.	0.643
1001-1100	4.	10.	-6.	-6303.	3.600
1101-1200	3.	8.	-5.	-5752.	3.125
1201-1300	3.	6.	-3.	-3751.	1.500
1301-1400	6.	5.	1.	1350.	0.200
1401-1500	2.	4.	-2.	-2901.	
1501-1600	1.	3.	-2.	-3101.	
1601-1700	0.	2.	-2.	-3301.	
1701-1800	1.	2.	-1.	-1751.	
1801-1900	0.	2.	-2.	-3701.	
1901-2000	1.	1.	0.	0.	
2001-2100	0.	1.	-1.	-2050.	
2101-2200	0.	1.	-1.	-2150.	
2201-2300	1.	1.	0.	0.	
2301-2400	0.	1.	-1.	-2350.	
2401-2500	0.	1.	-1.	-2450.	
TOTAL	2495	2494		-63269.	82.792

CHI SQ. STAT. =91.9 , D.F.=30 P<0.001

TOTAL ACTUAL COST = 449308.
TOTAL EXPECTED COST = 512577.

$$\frac{\text{TOTAL EXP. LOSS}}{\text{TOTAL ACT. COST}} = -14.08\ \%$$

Table (3.24)

*** TWO-PARAMETER LOGNORMAL DIS. ***

PREDICTION OF 74/4TH QUARTER CLAIMS COST
USING 73/4TH QUARTER MULTINOMIAL MAXLIK. PARAMETERS :

MEW= 4.516 SIGMA2=1.056
INFLATION RATE I=23.5% CALCULATED FROM :
INDEX OF AVE. EARNINGS, MISCELLANEOUS SERVICES.

PREDICTION PARAMETERS ARE :- MEW= 4.727 SIGMA2=1.056
MEAN CLAIM AMOUNT= 191.54 S.D.=262.26
ACTUAL 74/4TH PARAMETERS :- MEW= 4.672 SIGMA2=1.056
MEAN CLAIM AMOUNT= 181.27 S.D.=248.21

AMOUNT £	ACT.NO.	EXP.NO.	A-E	EXP.LOSS	(A-E)**2/E
1- 30	394.	310.	84.	1302.	22.761
31- 60	452.	522.	-70.	-3185.	9.387
61- 90	426.	438.	-12.	-906.	0.329
91- 120	348.	338.	10.	1055.	0.296
121- 150	272.	260.	12.	1626.	0.554
151- 180	219.	202.	17.	2813.	1.431
181- 210	154.	159.	-5.	-978.	0.157
211- 240	124.	127.	-3.	-677.	0.071
241- 270	105.	102.	3.	767.	0.088
271- 300	78.	83.	-5.	-1427.	0.301
301- 330	75.	69.	6.	1893.	0.522
331- 360	58.	57.	1.	346.	0.018
361- 390	55.	48.	7.	2628.	1.021
391- 420	29.	41.	-12.	-4866.	3.512
421- 450	43.	35.	8.	3484.	1.829
451- 480	24.	30.	-6.	-2793.	1.200
481- 510	22.	26.	-4.	-1982.	0.615
511- 540	24.	22.	2.	1051.	0.182
541- 570	19.	19.	0.	0.	0.000
571- 600	14.	17.	-3.	-1756.	0.529
601- 700	42.	43.	-1.	-651.	0.023
701- 800	28.	29.	-1.	-751.	0.034
801- 900	20.	20.	0.	0.	0.000
901-1000	17.	15.	2.	1901.	0.267
1001-1100	8.	11.	-3.	-3151.	0.818
1101-1200	5.	8.	-3.	-3451.	1.125
1201-1300	1.	6.	-5.	-6252.	4.167
1301-1400	5.	5.	0.	0.	0.000
1401-1500	0.	4.	-4.	-5802.	
1501-1600	1.	3.	-2.	-3101.	
1601-1700	0.	2.	-2.	-3301.	
1701-1800	0.	2.	-2.	-3501.	
1801-1900	0.	2.	-2.	-3701.	
1901-2000	0.	1.	-1.	-1950.	
2001-2100	1.	1.	0.	0.	
2101-2200	0.	1.	-1.	-2150.	
2201-2300	0.	1.	-1.	-2250.	
2301-2400	1.	1.	0.	0.	
2401-2500	0.	1.	-1.	-2450.	
TOTAL	3064	3061		-42168.	51.237

CHI SQ. STAT. =65.0 , D.F.=30 P<0.001

TOTAL ACTUAL COST = 533707.
TOTAL EXPECTED COST = 575876.

TOTAL EXP. LOSS
--------------- = -7.90 %
TOTAL ACT. COST

*** TWO-PARAMETER LOGNORMAL DIS. *** Table (3.25)

PREDICTION OF 75/1ST QUARTER CLAIMS COST
USING 74/1ST QUARTER MULTINOMIAL MAXLIK. PARAMETERS :

MEW= 4.509 SIGMA2=1.057
INFLATION RATE I=28.3% CALCULATED FROM :
INDEX OF AVE. EARNINGS, MISCELLANEOUS SERVICES.

PREDICTION PARAMETERS ARE :- MEW= 4.758 SIGMA2=1.057
MEAN CLAIM AMOUNT= 197.69 S.D.=270.90
ACTUAL 75/1ST PARAMETERS :- MEW= 4.684 SIGMA2=1.024
MEAN CLAIM AMOUNT= 180.55 S.D.=241.17

AMOUNT £	ACT.NO.	EXP.NO.	A-E	EXP.LOSS	(A-E)**2/E
1- 30	324.	251.	73.	1131.	21.231
31- 60	387.	432.	-45.	-2047.	4.687
61- 90	345.	368.	-23.	-1736.	1.437
91- 120	289.	287.	2.	211.	0.014
121- 150	253.	222.	31.	4200.	4.329
151- 180	187.	174.	13.	2151.	0.971
181- 210	138.	137.	1.	195.	0.007
211- 240	114.	110.	4.	902.	0.145
241- 270	93.	89.	4.	1022.	0.180
271- 300	67.	73.	-6.	-1713.	0.493
301- 330	63.	60.	3.	947.	0.150
331- 360	44.	50.	-6.	-2073.	0.720
361- 390	44.	42.	2.	751.	0.095
391- 420	35.	36.	-1.	-406.	0.028
421- 450	25.	31.	-6.	-2613.	1.161
451- 480	26.	26.	0.	0.	0.000
481- 510	18.	23.	-5.	-2477.	1.087
511- 540	18.	20.	-2.	-1051.	0.200
541- 570	22.	17.	5.	2777.	1.471
571- 600	17.	15.	2.	1171.	0.267
601- 700	39.	39.	0.	0.	0.000
701- 800	19.	26.	-7.	-5253.	1.885
801- 900	18.	18.	0.	0.	0.000
901-1000	12.	13.	-1.	-951.	0.077
1001-1100	3.	10.	-7.	-7353.	4.900
1101-1200	1.	7.	-6.	-6903.	5.143
1201-1300	1.	6.	-5.	-6252.	4.167
1301-1400	0.	4.	-4.	-5402.	
1401-1500	1.	3.	-2.	-2901.	
1501-1600	0.	3.	-3.	-4651.	
1601-1700	1.	2.	-1.	-1650.	
1701-1800	0.	2.	-2.	-3501.	
1801-1900	0.	1.	-1.	-1850.	
1901-2000	0.	1.	-1.	-1950.	
2001-2100	1.	1.	0.	0.	
2101-2200	1.	1.	0.	0.	
2201-2300	0.	1.	-1.	-2250.	
2301-2400	0.	1.	-1.	-2350.	
2401-2500	0.	1.	-1.	-2450.	
2501-2600	0.	0.	0.	0.	
2601-2700	0.	0.	0.	0.	
2701-2800	0.	0.	0.	0.	
2801-2900	0.	0.	0.	0.	
2901-3000	0.	0.	0.	0.	
3001-3100	0.	0.	0.	0.	
3101-3200	0.	0.	0.	0.	
3201-3300	0.	0.	0.	0.	
3301-3400	0.	0.	0.	0.	
3401-3500	0.	0.	0.	0.	
3501-3600	1.	0.	1.	3550.	
TOTAL	2607	2603		-50770.	54.845

Table (3.25) - continued

CHI SQ. STAT. = 68.0 , D.F. = 30 P < 0.001

TOTAL ACTUAL COST = 452059.
TOTAL EXPECTED COST = 502837.

TOTAL EXP. LOSS
----------------=-11.23 %
TOTAL ACT. COST

Table (3.26)

*** TWO-PARAMETER LOGNORMAL DIS. ***

PREDICTION OF 75/2ND QUARTER CLAIMS COST
USING 74/2ND QUARTER MULTINOMIAL MAXLIK. PARAMETERS :

MEW= 4.546 SIGMA2=1.013
INFLATION RATE I=29.5% CALCULATED FROM :
INDEX OF AVE. EARNINGS, MISCELLANEOUS SERVICES.

PREDICTION PARAMETERS ARE :- MEW= 4.805 SIGMA2=1.013
MEAN CLAIM AMOUNT= 202.55 S.D.=268.25
ACTUAL 75/2ND PARAMETERS :- MEW= 4.701 SIGMA2=1.056
MEAN CLAIM AMOUNT= 186.61 S.D.=255.51

AMOUNT £	ACT.NO.	EXP.NO.	A-E	EXP.LOSS	(A-E)**2/E
1- 30	302.	210.	92.	1426.	40.305
31- 60	374.	396.	-22.	-1001.	1.222
61- 90	332.	350.	-18.	-1359.	0.926
91- 120	277.	279.	-2.	-211.	0.014
121- 150	235.	218.	17.	2304.	1.326
151- 180	187.	172.	15.	2482.	1.308
181- 210	122.	136.	-14.	-2737.	1.441
211- 240	110.	109.	1.	225.	0.009
241- 270	80.	89.	-9.	-2299.	0.910
271- 300	72.	73.	-1.	-286.	0.014
301- 330	47.	60.	-13.	-4101.	2.817
331- 360	39.	50.	-11.	-3800.	2.420
361- 390	40.	42.	-2.	-751.	0.095
391- 420	38.	36.	2.	811.	0.111
421- 450	29.	31.	-2.	-871.	0.129
451- 480	21.	26.	-5.	-2327.	0.962
481- 510	30.	23.	7.	3468.	2.130
511- 540	19.	20.	-1.	-525.	0.050
541- 570	17.	17.	0.	0.	0.000
571- 600	11.	15.	-4.	-2342.	1.067
601- 700	36.	39.	-3.	-1951.	0.231
701- 800	22.	26.	-4.	-3002.	0.615
801- 900	22.	18.	4.	3402.	0.889
901-1000	11.	13.	-2.	-1901.	0.308
1001-1100	4.	10.	-6.	-6303.	3.600
1101-1200	3.	7.	-4.	-4602.	2.286
1201-1300	3.	5.	-2.	-2501.	0.800
1301-1400	6.	4.	2.	2701.	
1401-1500	2.	3.	-1.	-1450.	
1501-1600	1.	3.	-2.	-3101.	
1601-1700	0.	2.	-2.	-3301.	
1701-1800	1.	2.	-1.	-1751.	
1801-1900	0.	1.	-1.	-1850.	
1901-2000	1.	1.	0.	0.	
2001-2100	0.	1.	-1.	-2050.	
2101-2200	0.	1.	-1.	-2150.	
2201-2300	1.	1.	0.	0.	
2301-2400	0.	1.	-1.	-2350.	
TOTAL	2495	2490		-44057.	65.984

CHI SQ. STAT. = 72.6 , D.F.=30 P<0.001

TOTAL ACTUAL COST = 449308.
TOTAL EXPECTED COST = 493365.

TOTAL EXP. LOSS
--------------- = -9.81 %
TOTAL ACT. COST

Table (3.27)

*** TWO-PARAMETER LOGNORMAL DIS. ***

PREDICTION OF 74/4TH QUARTER CLAIMS COST
USING 73/4TH QUARTER MULTINOMIAL MAXLIK. PARAMETERS :

MEW= 4.516 SIGMA2=1.056
INFLATION RATE I=18.2% CALCULATED FROM :
GENERAL INDEX OF RETAIL PRICES.

PREDICTION PARAMETERS ARE :- MEW= 4.683 SIGMA2=1.056
MEAN CLAIM AMOUNT= 183.32 S.D.=251.00
ACTUAL 74/4TH PARAMETERS :- MEW= 4.672 SIGMA2=1.056
MEAN CLAIM AMOUNT= 181.27 S.D.=248.21

AMOUNT £	ACT.NO.	EXP.NO.	A-E	EXP.LOSS	(A-E)**2/E
1- 30	394.	334.	60.	930.	10.778
31- 60	452.	542.	-90.	-4095.	14.945
61- 90	426.	445.	-19.	-1434.	0.811
91- 120	348.	339.	9.	950.	0.239
121- 150	272.	258.	14.	1897.	0.760
151- 180	219.	199.	20.	3310.	2.010
181- 210	154.	155.	-1.	-195.	0.006
211- 240	124.	123.	1.	225.	0.008
241- 270	105.	99.	6.	1533.	0.364
271- 300	78.	80.	-2.	-571.	0.050
301- 330	75.	66.	9.	2839.	1.227
331- 360	58.	55.	3.	1036.	0.164
361- 390	55.	46.	9.	3379.	1.761
391- 420	29.	39.	-10.	-4055.	2.564
421- 450	43.	33.	10.	4355.	3.030
451- 480	24.	28.	-4.	-1862.	0.571
481- 510	22.	24.	-2.	-991.	0.167
511- 540	24.	21.	3.	1576.	0.429
541- 570	19.	18.	1.	555.	0.056
571- 600	14.	16.	-2.	-1171.	0.250
601- 700	42.	40.	2.	1301.	0.100
701- 800	28.	27.	1.	751.	0.037
801- 900	20.	19.	1.	851.	0.053
901-1000	17.	13.	4.	3802.	1.231
1001-1100	8.	10.	-2.	-2101.	0.400
1101-1200	5.	7.	-2.	-2301.	0.571
1201-1300	1.	6.	-5.	-6252.	4.167
1301-1400	5.	4.	1.	1350.	
1401-1500	0.	3.	-3.	-4351.	
1501-1600	1.	3.	-2.	-3101.	
1601-1700	0.	2.	-2.	-3301.	
1701-1800	0.	2.	-2.	-3501.	
1801-1900	0.	1.	-1.	-1850.	
1901-2000	0.	1.	-1.	-1950.	
2001-2100	1.	1.	0.	0.	
2101-2200	0.	1.	-1.	-2150.	
2201-2300	0.	1.	-1.	-2250.	
2301-2400	1.	1.	0.	0.	
TOTAL	3064	3062		-16844.	46.748

CHI SQ. STAT. = 55.5 , D.F.=30 P<0.003

TOTAL ACTUAL COST = 533707.
TOTAL EXPECTED COST = 550551.

KOL - SMIRNOV D = 0.019

TOTAL EXP. LOSS
--------------- = -3.16 %
TOTAL ACT. COST

P = 0.18

Table (3.28)

*** TWO-PARAMETER LOGNORMAL DIS. ***

PREDICTION OF 75/1ST QUARTER CLAIMS COST
USING 74/1ST QUARTER MULTINOMIAL MAXLIK. PARAMETERS :

MEW= 4.509 SIGMA2=1.057
INFLATION RATE I=20.3% CALCULATED FROM :
GENERAL INDEX OF RETAIL PRICES.

PREDICTION PARAMETERS ARE :- MEW= 4.694 SIGMA2=1.057
MEAN CLAIM AMOUNT= 185.36 S.D.=254.00
ACTUAL 75/1ST PARAMETERS :- MEW= 4.684 SIGMA2=1.024
MEAN CLAIM AMOUNT= 180.55 S.D.=241.17

AMOUNT £	ACT.NO.	EXP.NO.	A-E	EXP.LOSS	(A-E)**2/E
1- 30	324.	280.	44.	682.	6.914
31- 60	387.	457.	-70.	-3185.	10.722
61- 90	345.	377.	-32.	-2416.	2.716
91- 120	289.	288.	1.	105.	0.003
121- 150	253.	220.	33.	4471.	4.950
151- 180	187.	170.	17.	2813.	1.700
181- 210	138.	133.	5.	978.	0.188
211- 240	114.	105.	9.	2029.	0.771
241- 270	93.	85.	8.	2044.	0.753
271- 300	67.	69.	-2.	-571.	0.058
301- 330	63.	57.	6.	1893.	0.632
331- 360	44.	47.	-3.	-1036.	0.191
361- 390	44.	39.	5.	1877.	0.641
391- 420	35.	33.	2.	811.	0.121
421- 450	25.	28.	-3.	-1306.	0.321
451- 480	26.	24.	2.	931.	0.167
481- 510	18.	21.	-3.	-1486.	0.429
511- 540	18.	18.	0.	0.	0.000
541- 570	22.	16.	6.	3333.	2.250
571- 600	17.	14.	3.	1756.	0.643
601- 700	39.	35.	4.	2602.	0.457
701- 800	19.	23.	-4.	-3002.	0.696
801- 900	18.	16.	2.	1701.	0.250
901-1000	12.	12.	0.	0.	0.000
1001-1100	3.	9.	-6.	-6303.	4.000
1101-1200	1.	6.	-5.	-5752.	4.167
1201-1300	1.	5.	-4.	-5002.	3.200
1301-1400	0.	4.	-4.	-5402.	
1401-1500	1.	3.	-2.	-2901.	
1501-1600	0.	2.	-2.	-3101.	
1601-1700	1.	2.	-1.	-1650.	
1701-1800	0.	2.	-2.	-3501.	
1801-1900	0.	1.	-1.	-1850.	
1901-2000	0.	1.	-1.	-1950.	
2001-2100	1.	1.	0.	0.	
2101-2200	1.	1.	0.	0.	
2201-2300	0.	1.	-1.	-2250.	
2301-2400	0.	1.	-1.	-2350.	
2401-2500	0.	0.	0.	0.	
2501-2600	0.	0.	0.	0.	
2601-2700	0.	0.	0.	0.	
2701-2800	0.	0.	0.	0.	
2801-2900	0.	0.	0.	0.	
2901-3000	0.	0.	0.	0.	
3001-3100	0.	0.	0.	0.	
3101-3200	0.	0.	0.	0.	
3201-3300	0.	0.	0.	0.	
3301-3400	0.	0.	0.	0.	
3401-3500	0.	0.	0.	0.	
3501-3600	1.	0.	1.	3550.	

Table (3.28) - continued

TOTAL	2607	2606	-23439.	46.941

CHI SQ. STAT. =57.7 , D.F.=29 P<0.003

TOTAL ACTUAL COST = 452059.
TOTAL EXPECTED COST = 475498.

$$\frac{\text{TOTAL EXP. LOSS}}{\text{TOTAL ACT. COST}} = -5.19\ \%$$

KOL - SMIRNOV D = 0.022
P = 0.15

Table (3.29)

*** TWO-PARAMETER LOGNORMAL DIS. ***

PREDICTION OF 75/2ND QUARTER CLAIMS COST
USING 74/2ND QUARTER MULTINOMIAL MAXLIK. PARAMETERS :

MEW= 4.546 SIGMA2=1.013
INFLATION RATE I=24.3% CALCULATED FROM :
GENERAL INDEX OF RETAIL PRICES.

PREDICTION PARAMETERS ARE :- MEW= 4.764 SIGMA2=1.013
MEAN CLAIM AMOUNT= 194.42 S.D.=257.48
ACTUAL 75/2ND PARAMETERS :- MEW= 4.701 SIGMA2=1.056
MEAN CLAIM AMOUNT= 186.61 S.D.=255.51

AMOUNT £	ACT.NO.	EXP.NO.	A-E	EXP.LOSS	(A-E)**2/E
1- 30	302.	226.	76.	1178.	25.558
31- 60	374.	412.	-38.	-1729.	3.505
61- 90	332.	357.	-25.	-1887.	1.751
91- 120	277.	280.	-3.	-316.	0.032
121- 150	235.	217.	18.	2439.	1.493
151- 180	187.	170.	17.	2813.	1.700
181- 210	122.	134.	-12.	-2346.	1.075
211- 240	110.	107.	3.	677.	0.084
241- 270	80.	86.	-6.	-1533.	0.419
271- 300	72.	70.	2.	571.	0.057
301- 330	47.	58.	-11.	-3470.	2.086
331- 360	39.	48.	-9.	-3109.	1.687
361- 390	40.	41.	-1.	-376.	0.024
391- 420	38.	34.	4.	1622.	0.471
421- 450	29.	29.	0.	0.	0.000
451- 480	21.	25.	-4.	-1862.	0.640
481- 510	30.	21.	9.	4459.	3.857
511- 540	19.	19.	0.	0.	0.000
541- 570	17.	16.	1.	555.	0.063
571- 600	11.	14.	-3.	-1756.	0.643
601- 700	36.	36.	0.	0.	0.000
701- 800	22.	24.	-2.	-1501.	0.167
801- 900	22.	17.	5.	4252.	1.471
901-1000	11.	12.	-1.	-951.	0.083
1001-1100	4.	9.	-5.	-5252.	2.778
1101-1200	3.	7.	-4.	-4602.	2.286
1201-1300	3.	5.	-2.	-2501.	0.800
1301-1400	6.	4.	2.	2701.	
1401-1500	2.	3.	-1.	-1450.	
1501-1600	1.	2.	-1.	-1550.	
1601-1700	0.	2.	-2.	-3301.	
1701-1800	1.	2.	-1.	-1751.	
1801-1900	0.	1.	-1.	-1850.	
1901-2000	1.	1.	0.	0.	
2001-2100	0.	1.	-1.	-2050.	
2101-2200	0.	1.	-1.	-2150.	
2201-2300	1.	1.	0.	0.	
TOTAL	2495	2492		-26028.	52.728

CHI SQ. STAT. =57.3 , D.F.=29 P<0.003

TOTAL ACTUAL COST = 449308.
TOTAL EXPECTED COST = 475336.

TOTAL EXP. LOSS / TOTAL ACT. COST = -5.79 %

KOL - SMIRNOV D = 0.030
P = 0.02

Table (3.30)

*** 3-PARAMETER LOGNORMAL DISTRIBUTION ***

ESTIMATING PARAMETER C BY THE LEAST SQUARES REGRESSION METHOD

73/4TH QUARTER DATA

INPUT

C IPRINT

0. 0

C= 0.00 SSD=0.15923

5. 0

C= 5.00 SSD=0.10475

10. 0

C= 10.00 SSD=0.07228

15. 0

C= 15.00 SSD=0.05536

20. 0

C= 20.00 SSD=0.04983

25. 0

C= 25.00 SSD=0.05281

18. 0

C= 18.00 SSD=0.05088

19. 0

C= 19.00 SSD=0.05018

21. 0

C= 21.00 SSD=0.04982

22. 0

C= 22.00 SSD=0.05013

Table (3.31)

73/4TH QUARTER DATA

C=21.00
PARAMETERS OF THE LOGNORMAL DIS. ESTIMATED BY L.S. REGRESSION :-
MEW = 4.778 SIGMA SQ.= 0.730

MEAN= 150.219 S.D.= 177.489

SSD=0.050

AMOUNT £	ACTUAL CL. NO.	EXPECTED CL. NO.	ACTUAL-EXPECTED	EXPECTED LOSS	(A-E)**2/E
1- 30	478.	499.	-21.	-325.	0.884
31- 60	518.	504.	14.	637.	0.389
61- 90	461.	429.	32.	2416.	2.387
91- 120	359.	337.	22.	2321.	1.436
121- 150	239.	260.	-21.	-2845.	1.696
151- 180	213.	200.	13.	2151.	0.845
181- 210	148.	154.	-6.	-1173.	0.234
211- 240	102.	121.	-19.	-4284.	2.983
241- 270	81.	95.	-14.	-3577.	2.063
271- 300	58.	75.	-17.	-4853.	3.853
301- 330	66.	61.	5.	1577.	0.410
331- 360	45.	49.	-4.	-1382.	0.327
361- 390	39.	40.	-1.	-376.	0.025
391- 420	35.	33.	2.	811.	0.121
421- 450	34.	27.	7.	3048.	1.815
451- 480	20.	23.	-3.	-1396.	0.391
481- 510	29.	19.	10.	4955.	5.263
511- 540	14.	16.	-2.	-1051.	0.250
541- 570	8.	13.	-5.	-2777.	1.923
571- 600	9.	11.	-2.	-1171.	0.364
601- 700	29.	27.	2.	1301.	0.148
701- 800	18.	17.	1.	751.	0.059
801- 900	20.	11.	9.	7654.	7.364
901-1000	6.	7.	-1.	-951.	0.143
1001-1100	4.	5.	-1.	-1050.	0.200
1101-1200	4.	3.	1.	1150.	
1201-1300	1.	2.	-1.	-1250.	0.000
1301-1400	3.	2.	1.	1350.	
1401-1500	1.	1.	0.	0.	
1501-1600	0.	1.	-1.	-1550.	
1601-1700	1.	1.	0.	0.	
1701-1800	0.	1.	-1.	-1751.	
1801-1900	0.	0.	0.	0.	
1901-2000	1.	0.	1.	1950.	
2001-2100	0.	0.	0.	0.	
2101-2200	0.	0.	0.	0.	
2201-2300	0.	0.	0.	0.	
2301-2400	1.	0.	1.	2350.	0.167
TOTAL	3045.	3044.		2660.	35.940

D.F. = 23
P = 0.04

TOTAL EXP. LOSS / TOTAL ACT. COST = 0.6 %

Table (3.32)

74/1ST QUARTER DATA

C=23.00
PARAMETERS OF THE LOGNORMAL DIS. ESTIMATED BY L.S. REGRESSION :-
MEW = 4.819   SIGMA SQ.= 0.679

MEAN= 150.820   S.D.= 171.348

SSD=0.223

AMOUNT £	ACTUAL CL. NO.	EXPECTED CL. NO.	ACTUAL-EXPECTED	EXPECTED LOSS	(A-E)**2/E
1- 30	381.	377.	4.	62.	0.042
31- 60	428.	396.	32.	1456.	2.586
61- 90	351.	346.	5.	378.	0.072
91- 120	334.	276.	58.	6119.	12.188
121- 150	211.	214.	-3.	-406.	0.042
151- 180	133.	166.	-33.	-5461.	6.560
181- 210	98.	128.	-30.	-5865.	7.031
211- 240	82.	100.	-18.	-4059.	3.240
241- 270	54.	79.	-25.	-6387.	7.911
271- 300	52.	62.	-10.	-2855.	1.613
301- 330	53.	50.	3.	947.	0.180
331- 360	36.	40.	-4.	-1382.	0.400
361- 390	29.	33.	-4.	-1502.	0.485
391- 420	26.	27.	-1.	-406.	0.037
421- 450	22.	22.	0.	0.	0.000
451- 480	22.	18.	4.	1862.	0.889
481- 510	17.	15.	2.	991.	0.267
511- 540	10.	13.	-3.	-1576.	0.692
541- 570	19.	11.	8.	4444.	5.818
571- 600	4.	9.	-5.	-2927.	2.778
601- 700	26.	21.	5.	3252.	1.190
701- 800	21.	13.	8.	6004.	4.923
801- 900	11.	8.	3.	2551.	1.125
901-1000	10.	5.	5.	4752.	5.000
1001-1100	5.	4.	1.	1050.	
1101-1200	2.	2.	0.	0.	
1201-1300	1.	2.	-1.	-1250.	0.167
1301-1400	2.	1.	1.	1350.	
1401-1500	0.	1.	-1.	-1450.	
1501-1600	0.	1.	-1.	-1550.	
1601-1700	0.	0.	0.	0.	
1701-1800	1.	0.	1.	1751.	0.200
TOTAL	2441.	2440.		-109.	65.438

D.F. = 22

TOTAL EXP. LOSS
--------------- = -0.0 %
TOTAL ACT. COST

P ⟨ 0.001

Table (3.33)

74/2ST QUARTER DATA

C=21.00
PARAMETERS OF THE LOGNORMAL DIS. ESTIMATED BY L.S. REGRESSION :-

MEW = 4.824 SIGMA SQ.= 0.666

MEAN= 152.721 S.D.= 169.077

SSD=0.133

AMOUNT £	ACTUAL CL. NO.	EXPECTED CL. NO.	ACTUAL-EXPECTED	EXPECTED LOSS	(A-E)**2/E
1- 30	351.	333.	18.	279.	0.973
31- 60	380.	386.	-6.	-273.	0.093
61- 90	382.	344.	38.	2869.	4.198
91- 120	295.	276.	19.	2004.	1.308
121- 150	211.	215.	-4.	-542.	0.074
151- 180	142.	166.	-24.	-3972.	3.470
181- 210	114.	129.	-15.	-2933.	1.744
211- 240	101.	100.	1.	225.	0.010
241- 270	57.	79.	-22.	-5621.	6.127
271- 300	51.	62.	-11.	-3140.	1.952
301- 330	39.	50.	-11.	-3470.	2.420
331- 360	36.	40.	-4.	-1382.	0.400
361- 390	25.	32.	-7.	-2628.	1.531
391- 420	24.	26.	-2.	-811.	0.154
421- 450	27.	22.	5.	2177.	1.136
451- 480	18.	18.	0.	0.	0.000
481- 510	21.	15.	6.	2973.	2.400
511- 540	17.	12.	5.	2627.	2.083
541- 570	12.	10.	2.	1111.	0.400
571- 600	11.	9.	2.	1171.	0.444
601- 700	30.	21.	9.	5854.	3.857
701- 800	13.	13.	0.	0.	0.000
801- 900	11.	8.	3.	2551.	1.125
901-1000	4.	5.	-1.	-951.	0.200
1001-1100	7.	3.	4.	4202.	
1101-1200	0.	2.	-2.	-2301.	0.800
1201-1300	1.	2.	-1.	-1250.	
1301-1400	1.	1.	0.	0.	
1401-1500	0.	1.	-1.	-1450.	
1501-1600	1.	1.	0.	0.	
1601-1700	0.	0.	0.	0.	
1701-1800	0.	0.	0.	0.	
1801-1900	1.	0.	1.	1850.	0.200
TOTAL	2383.	2381.		-829.	37.100

D.F. = 22
P = 0.02

TOTAL EXP. LOSS / TOTAL ACT. COST = -0.2 %

Table (3.34)

74/3RD QUARTER DATA

C=19.00
PARAMETERS OF THE LOGNORMAL DIS. ESTIMATED BY L.S. REGRESSION :-
MEW = 4.857 SIGMA SQ.= 0.731

MEAN= 166.520 S.D.= 192.635

SSD=0.103

AMOUNT £	ACTUAL CL. NO.	EXPECTED CL. NO.	ACTUAL-EXPECTED	EXPECTED LOSS	(A-E)**2/E
1- 30	362.	369.	-7.	-108.	0.133
31- 60	427.	433.	-6.	-273.	0.083
61- 90	383.	388.	-5.	-378.	0.064
91- 120	356.	315.	41.	4325.	5.337
121- 150	283.	248.	35.	4742.	4.940
151- 180	194.	195.	-1.	-165.	0.005
181- 210	137.	153.	-16.	-3128.	1.673
211- 240	97.	121.	-24.	-5412.	4.760
241- 270	86.	97.	-11.	-2810.	1.247
271- 300	71.	78.	-7.	-1998.	0.628
301- 330	64.	63.	1.	316.	0.016
331- 360	45.	51.	-6.	-2073.	0.706
361- 390	44.	42.	2.	751.	0.095
391- 420	25.	35.	-10.	-4055.	2.857
421- 450	26.	29.	-3.	-1306.	0.310
451- 480	22.	24.	-2.	-931.	0.167
481- 510	25.	21.	4.	1982.	0.762
511- 540	14.	17.	-3.	-1576.	0.529
541- 570	14.	15.	-1.	-555.	0.067
571- 600	17.	13.	4.	2342.	1.231
601- 700	32.	31.	1.	651.	0.032
701- 800	34.	19.	15.	11257.	11.842
801- 900	17.	12.	5.	4252.	2.083
901-1000	4.	8.	-4.	-3802.	2.000
1001-1100	9.	6.	3.	3151.	1.500
1101-1200	4.	4.	0.	0.	0.000
1201-1300	0.	3.	-3.	-3751.	
1301-1400	1.	2.	-1.	-1350.	3.200
1401-1500	3.	2.	1.	1450.	
1501-1600	0.	1.	-1.	-1550.	
1601-1700	1.	1.	0.	0.	
1701-1800	0.	1.	-1.	-1751.	
1801-1900	0.	1.	-1.	-1850.	
1901-2000	0.	0.	0.	0.	
2001-2100	0.	0.	0.	0.	
2101-2200	1.	0.	1.	2150.	
2201-2300	0.	0.	0.	0.	
2301-2400	0.	0.	0.	0.	
2401-2500	0.	0.	0.	0.	
2501-2600	1.	0.	1.	2550.	0.000
TOTAL	2799.	2798.		1095.	46.268

D.F. = 24
P = 0.004

TOTAL EXP. LOSS / TOTAL ACT. COST = 0.2 %

Table (3.35)

74/4TH QUARTER DATA

C=25.00

PARAMETERS OF THE LOGNORMAL DIS. ESTIMATED BY L.S. REGRESSION :-

MEW = 4.959 SIGMA SQ.= 0.669

MEAN= 174.089 S.D.= 194.322

SSD=0.124

AMOUNT £	ACTUAL CL. NO.	EXPECTED CL. NO.	ACTUAL-EXPECTED	EXPECTED LOSS	(A-E)**2/E
1- 30	394.	382.	12.	186.	0.377
31- 60	452.	434.	18.	819.	0.747
61- 90	426.	406.	20.	1510.	0.985
91- 120	348.	342.	6.	633.	0.105
121- 150	272.	277.	-5.	-678.	0.090
151- 180	219.	221.	-2.	-331.	0.018
181- 210	154.	177.	-23.	-4496.	2.989
211- 240	124.	141.	-17.	-3833.	2.050
241- 270	105.	114.	-9.	-2299.	0.711
271- 300	78.	92.	-14.	-3997.	2.130
301- 330	75.	75.	0.	0.	0.000
331- 360	58.	61.	-3.	-1036.	0.148
361- 390	55.	50.	5.	1877.	0.500
391- 420	29.	42.	-13.	-5271.	4.024
421- 450	43.	35.	8.	3484.	1.829
451- 480	24.	29.	-5.	-2327.	0.862
481- 510	22.	25.	-3.	-1486.	0.360
511- 540	24.	21.	3.	1576.	0.429
541- 570	19.	18.	1.	555.	0.056
571- 600	14.	15.	-1.	-585.	0.067
601- 700	42.	37.	5.	3252.	0.676
701- 800	28.	23.	5.	3752.	1.087
801- 900	20.	15.	5.	4252.	1.667
901-1000	17.	10.	7.	6653.	4.900
1001-1100	8.	7.	1.	1050.	0.143
1101-1200	5.	5.	0.	0.	0.000
1201-1300	1.	3.	-2.	-2501.	
1301-1400	5.	2.	3.	4051.	0.200
1401-1500	0.	2.	-2.	-2901.	
1501-1600	1.	1.	0.	0.	
1601-1700	0.	1.	-1.	-1650.	
1701-1800	0.	1.	-1.	-1751.	
1801-1900	0.	1.	-1.	-1850.	
1901-2000	0.	0.	0.	0.	
2001-2100	1.	0.	1.	2050.	
2101-2200	0.	0.	0.	0.	
2201-2300	0.	0.	0.	0.	
2301-2400	1.	0.	1.	2350.	1.500
TOTAL	3064.	3065.		1059.	28.647

D.F. = 24

P > 0.10

TOTAL EXP. LOSS / TOTAL ACT. COST = 0.2 %

C=13.00

Table (3.36)

PARAMETERS OF THE LOGNORMAL DIS. ESTIMATED BY L.S. REGRESSION :-

MEW = 4.831 SIGMA SQ.= 0.773

MEAN= 171.413 S.D.= 199.147

SSD=0.506

AMOUNT £	ACTUAL CL. NO.	EXPECTED CL. NO.	ACTUAL-EXPECTED	EXPECTED LOSS	(A-E)**2/E
1- 30	324.	298.	26.	403.	2.268
31- 60	387.	411.	-24.	-1092.	1.401
61- 90	345.	370.	-25.	-1887.	1.689
91- 120	289.	299.	-10.	-1055.	0.334
121- 150	253.	235.	18.	2439.	1.379
151- 180	187.	184.	3.	496.	0.049
181- 210	138.	144.	-6.	-1173.	0.250
211- 240	114.	114.	0.	0.	0.000
241- 270	93.	91.	2.	511.	0.044
271- 300	67.	73.	-6.	-1713.	0.493
301- 330	63.	59.	4.	1262.	0.271
331- 360	44.	49.	-5.	-1727.	0.510
361- 390	44.	40.	4.	1502.	0.400
391- 420	35.	33.	2.	811.	0.121
421- 450	25.	28.	-3.	-1306.	0.321
451- 480	26.	23.	3.	1396.	0.391
481- 510	18.	20.	-2.	-991.	0.200
511- 540	18.	17.	1.	525.	0.059
541- 570	22.	14.	8.	4444.	4.571
571- 600	17.	12.	5.	2927.	2.083
601- 700	39.	30.	9.	5854.	2.700
701- 800	19.	19.	0.	0.	0.000
801- 900	18.	12.	6.	5103.	3.000
901-1000	12.	8.	4.	3802.	2.000
1001-1100	3.	6.	-3.	-3151.	1.500
1101-1200	1.	4.	-3.	-3451.	
1201-1300	1.	3.	-2.	-2501.	3.571
1301-1400	0.	2.	-2.	-2701.	
1401-1500	1.	2.	-1.	-1450.	
1501-1600	0.	1.	-1.	-1550.	
1601-1700	1.	1.	0.	0.	
1701-1800	0.	1.	-1.	-1751.	
1801-1900	0.	1.	-1.	-1850.	
1901-2000	0.	0.	0.	0.	
2001-2100	1.	0.	1.	2050.	
2101-2200	1.	0.	1.	2150.	
2201-2300	0.	0.	0.	0.	
2301-2400	0.	0.	0.	0.	
2401-2500	0.	0.	0.	0.	
2501-2600	0.	0.	0.	0.	
2601-2700	0.	0.	0.	0.	
2701-2800	0.	0.	0.	0.	
2801-2900	0.	0.	0.	0.	
2901-3000	0.	0.	0.	0.	
3001-3100	0.	0.	0.	0.	
3101-3200	0.	0.	0.	0.	
3201-3300	0.	0.	0.	0.	
3301-3400	0.	0.	0.	0.	
3401-3500	0.	0.	0.	0.	
3501-3600	1.	0.	1.	3550.	1.125
TOTAL	2607.	2604.		9876.	30.733

D.F. = 23

P > 0.10

TOTAL EXP. LOSS / TOTAL ACT. COST = 2.2 %

Table (3.37)

75/2ND QUARTER DATA

C=25.00
PARAMETERS OF THE LOGNORMAL DIS. ESTIMATED BY L.S. REGRESSION :-

MEW = 4.979 SIGMA SQ.= 0.692

MEAN= 180.416 S.D.= 205.254

SSD=0.086

AMOUNT £	ACTUAL CL. NO.	EXPECTED CL. NO.	ACTUAL-EXPECTED	EXPECTED LOSS	(A-E)**2/E
1- 30	302.	309.	-7.	-108.	0.159
31- 60	374.	345.	29.	1319.	2.438
61- 90	332.	323.	9.	680.	0.251
91- 120	277.	273.	4.	422.	0.059
121- 150	235.	222.	13.	1761.	0.761
151- 180	187.	179.	8.	1324.	0.358
181- 210	122.	144.	-22.	-4301.	3.361
211- 240	110.	116.	-6.	-1353.	0.310
241- 270	80.	94.	-14.	-3577.	2.085
271- 300	72.	76.	-4.	-1142.	0.211
301- 330	47.	63.	-16.	-5048.	4.063
331- 360	39.	52.	-13.	-4491.	3.250
361- 390	40.	43.	-3.	-1126.	0.209
391- 420	38.	36.	2.	811.	0.111
421- 450	29.	30.	-1.	-436.	0.033
451- 480	21.	25.	-4.	-1862.	0.640
481- 510	30.	21.	9.	4459.	3.857
511- 540	19.	18.	1.	525.	0.056
541- 570	17.	15.	2.	1111.	0.267
571- 600	11.	13.	-2.	-1171.	0.308
601- 700	36.	33.	3.	1951.	0.273
701- 800	22.	21.	1.	751.	0.048
801- 900	22.	13.	9.	7654.	6.231
901-1000	11.	9.	2.	1901.	0.444
1001-1100	4.	6.	-2.	-2101.	0.667
1101-1200	3.	4.	-1.	-1150.	
1201-1300	3.	3.	0.	0.	0.143
1301-1400	6.	2.	4.	5402.	
1401-1500	2.	2.	0.	0.	
1501-1600	1.	1.	0.	0.	
1601-1700	0.	1.	-1.	-1650.	
1701-1800	1.	1.	0.	0.	
1801-1900	0.	1.	-1.	-1850.	
1901-2000	1.	0.	1.	1950.	
2001-2100	0.	0.	0.	0.	
2101-2200	0.	0.	0.	0.	
2201-2300	1.	0.	1:	2250.	2.000
TOTAL	2495.	2494.		2905.	32.591

D.F. = 23

TOTAL EXP. LOSS / TOTAL ACT. COST = 0.6 %

P =0.09

Table (3.38)

*** 3-PARAMETER LOGNORMAL DIS. ***

73/4TH QUARTER DATA

ESTIMATION BY MULTINOMIAL MAX. LIKELIHOOD METHOD :-
C= 14.01 MEW= 4.700 SIGMA SQ.= 0.803

MEAN= 150.159 S.D.= 182.166

AMOUNT F	ACTUAL CL. NO.	EXPECTED CL. NO.	ACTUAL-EXPECTED	EXPECTED LOSS	(A-E)**2/E
1- 30	478.	477.	1.	15.	0.002
31- 60	518.	535.	-17.	-774.	0.540
61- 90	461.	443.	18.	1359.	0.731
91- 120	359.	340.	19.	2004.	1.062
121- 150	239.	257.	-18.	-2439.	1.261
151- 180	213.	196.	17.	2813.	1.474
181- 210	148.	150.	-2.	-391.	0.027
211- 240	102.	117.	-15.	-3382.	1.923
241- 270	81.	92.	-11.	-2810.	1.315
271- 300	58.	73.	-15.	-4282.	3.082
301- 330	66.	58.	8.	2524.	1.103
331- 360	45.	47.	-2.	-691.	0.085
361- 390	39.	39.	0.	0.	0.000
391- 420	35.	32.	3.	1216.	0.281
421- 450	34.	26.	8.	3484.	2.462
451- 480	20.	22.	-2.	-931.	0.182
481- 510	29.	18.	11.	5450.	6.722
511- 540	14.	16.	-2.	-1051.	0.250
541- 570	8.	13.	-5.	-2777.	1.923
571- 600	9.	11.	-2.	-1171.	0.364
601- 700	29.	27.	2.	1301.	0.148
701- 800	18.	17.	1.	751.	0.059
801- 900	20.	11.	9.	7654.	7.364
901-1000	6.	7.	-1.	-951.	0.143
1001-1100	4.	5.	-1.	-1050.	0.200
1101-1200	4.	4.	0.	0.	
1201-1300	1.	3.	-2.	-2501.	0.571
1301-1400	3.	2.	1.	1350.	
1401-1500	1.	1.	0.	0.	
1501-1600	0.	1.	-1.	-1550.	
1601-1700	1.	1.	0.	0.	
1701-1800	0.	1.	-1.	-1751.	
1801-1900	0.	0.	0.	0.	
1901-2000	1.	0.	1.	1950.	
2001-2100	0.	0.	0.	0.	
2101-2200	0.	0.	0.	0.	
2201-2300	0.	0.	0.	0.	
2301-2400	1.	0.	1.	2350.	0.167
TOTAL	3045.	3042.		5721.	33.441

D.F. = 23
P > 0.07

TOTAL EXP. LOSS / TOTAL ACT. COST = 1.2 %

KOL - SMIRNOV D = 0.007
P > 0.20

Table (3.39)

*** 3-PARAMETER LOGNORMAL DIS. ***

74/1ST QUARTER DATA

ESTIMATION BY MULTINOMIAL MAX. LIKELIHOOD METHOD :-

C= 10.40 MEW= 4.650 SIGMA SQ.= 0.858

MEAN= 150.222 S.D.= 187.204

AMOUNT £	ACTUAL CL. NO.	EXPECTED CL. NO.	ACTUAL-EXPECTED	EXPECTED LOSS	(A-E)**2/E
1- 30	381.	379.	2.	31.	0.011
31- 60	428.	444.	-16.	-728.	0.577
61- 90	351.	359.	-8.	-604.	0.178
91- 120	334.	271.	63.	6646.	14.646
121- 150	211.	203.	8.	1084.	0.315
151- 180	133.	154.	-21.	-3475.	2.864
181- 210	98.	118.	-20.	-3910.	3.390
211- 240	82.	91.	-9.	-2029.	0.890
241- 270	54.	72.	-18.	-4599.	4.500
271- 300	52.	57.	-5.	-1427.	0.439
301- 330	53.	46.	7.	2208.	1.065
331- 360	36.	37.	-1.	-346.	0.027
361- 390	29.	30.	-1.	-376.	0.033
391- 420	26.	25.	1.	406.	0.040
421- 450	22.	21.	1.	436.	0.048
451- 480	22.	17.	5.	2327.	1.471
481- 510	17.	15.	2.	991.	0.267
511- 540	10.	12.	-2.	-1051.	0.333
541- 570	19.	11.	8.	4444.	5.818
571- 600	4.	9.	-5.	-2927.	2.778
601- 700	26.	22.	4.	2602.	0.727
701- 800	21.	14.	7.	5253.	3.500
801- 900	11.	9.	2.	1701.	0.444
901-1000	10.	6.	4.	3802.	2.667
1001-1100	5.	4.	1.	1050.	
1101-1200	2.	3.	-1.	-1150.	0.000
1201-1300	1.	2.	-1.	-1250.	
1301-1400	2.	2.	0.	0.	
1401-1500	0.	1.	-1.	-1450.	
1501-1600	0.	1.	-1.	-1550.	
1601-1700	0.	1.	-1.	-1650.	
1701-1800	1.	1.	0.	0.	2.000
TOTAL	2441.	2437.		4457.	49.207

D.F. = 22

P < 0.001

TOTAL EXP. LOSS / TOTAL ACT. COST = 1.2 %

KOL - SMIRNOV D = 0.020

P < 0.03

Table (3.40)

*** 3-PARAMETER LOGNORMAL DIS. ***

74/2ND QUARTER DATA

ESTIMATION BY MULTINOMIAL MAX. LIKELIHOOD METHOD :-
C= 14.94 MEW= 4.735 SIGMA SQ.= 0.760

MEAN= 151.518 S.D.= 177.583

AMOUNT £	ACTUAL CL. NO.	EXPECTED CL. NO.	ACTUAL-EXPECTED	EXPECTED LOSS	(A-E)**2/E
1- 30	351.	348.	3.	46.	0.026
31- 60	380.	411.	-31.	-1410.	2.338
61- 90	382.	349.	33.	2491.	3.120
91- 120	295.	272.	23.	2426.	1.945
121- 150	211.	207.	4.	542.	0.077
151- 180	142.	158.	-16.	-2648.	1.620
181- 210	114.	122.	-8.	-1564.	0.525
211- 240	101.	94.	7.	1578.	0.521
241- 270	57.	74.	-17.	-4343.	3.905
271- 300	51.	59.	-8.	-2284.	1.085
301- 330	39.	47.	-8.	-2524.	1.362
331- 360	36.	38.	-2.	-691.	0.105
361- 390	25.	31.	-6.	-2253.	1.161
391- 420	24.	25.	-1.	-406.	0.040
421- 450	27.	21.	6.	2613.	1.714
451- 480	18.	17.	1.	466.	0.059
481- 510	21.	15.	6.	2973.	2.400
511- 540	17.	12.	5.	2627.	2.083
541- 570	12.	10.	2.	1111.	0.400
571- 600	11.	9.	2.	1171.	0.444
601- 700	30.	21.	9.	5854.	3.857
701- 800	13.	13.	0.	0.	0.000
801- 900	11.	8.	3.	2551.	1.125
901-1000	4.	6.	-2.	-1901.	0.667
1001-1100	7.	4.	3.	3151.	
1101-1200	0.	3.	-3.	-3451.	0.000
1201-1300	1.	2.	-1.	-1250.	
1301-1400	1.	1.	0.	0.	
1401-1500	0.	1.	-1.	-1450.	
1501-1600	1.	1.	0.	0.	
1601-1700	0.	1.	-1.	-1650.	
1701-1800	0.	0.	0.	0.	
1801-1900	1.	0.	1.	1850.	0.667
TOTAL	2383.	2380.		3626.	31.248

D.F. = 22

P = 0.07

TOTAL EXP. LOSS / TOTAL ACT. COST = 1.0 %

KOL - SMIRNOV D = 0.013

P > 0.20

Table (3.41)

*** 3-PARAMETER LOGNORMAL DIS. ***

74/3RD QUARTER DATA

ESTIMATION BY MULTINOMIAL MAX. LIKELIHOOD METHOD :-
C= 15.67 MEW= 4.821 SIGMA SQ.= 0.759

MEAN= 165.694 S.D.= 193.304

AMOUNT £	ACTUAL CL. NO.	EXPECTED CL. NO.	ACTUAL-EXPECTED	EXPECTED LOSS	(A-E)**2/E
1- 30	362.	359.	3.	46.	0.025
31- 60	427.	446.	-19.	-865.	0.809
61- 90	383.	395.	-12.	-906.	0.365
91- 120	356.	318.	38.	4009.	4.541
121- 150	283.	249.	34.	4607.	4.643
151- 180	194.	194.	0.	0.	0.000
181- 210	137.	152.	-15.	-2933.	1.480
211- 240	97.	120.	-23.	-5186.	4.408
241- 270	86.	95.	-9.	-2299.	0.853
271- 300	71.	76.	-5.	-1427.	0.329
301- 330	64.	62.	2.	631.	0.065
331- 360	45.	50.	-5.	-1727.	0.500
361- 390	44.	41.	3.	1126.	0.220
391- 420	25.	34.	-9.	-3649.	2.382
421- 450	26.	28.	-2.	-871.	0.143
451- 480	22.	24.	-2.	-931.	0.167
481- 510	25.	20.	5.	2477.	1.250
511- 540	14.	17.	-3.	-1576.	0.529
541- 570	14.	15.	-1.	-555.	0.067
571- 600	17.	12.	5.	2927.	2.083
601- 700	32.	30.	2.	1301.	0.133
701- 800	34.	19.	15.	11257.	11.842
801- 900	17.	12.	5.	4252.	2.083
901-1000	4.	8.	-4.	-3802.	2.000
1001-1100	9.	6.	3.	3151.	1.500
1101-1200	4.	4.	0.	0.	
1201-1300	0.	3.	-3.	-3751.	1.286
1301-1400	1.	2.	-1.	-1350.	
1401-1500	3.	2.	1.	1450.	
1501-1600	0.	1.	-1.	-1550.	
1601-1700	1.	1.	0.	0.	
1701-1800	0.	1.	-1.	-1751.	
1801-1900	0.	1.	-1.	-1850.	
1901-2000	0.	0.	0.	0.	
2001-2100	0.	0.	0.	0.	
2101-2200	1.	0.	1.	2150.	
2201-2300	0.	0.	0.	0.	
2301-2400	0.	0.	0.	0.	
2401-2500	0.	0.	0.	0.	
2501-2600	1.	0.	1.	2550.	0.125
TOTAL	2799.	2797.		4956.	43.828

D.F. = 23

P = 0.006

$$\frac{\text{TOTAL EXP. LOSS}}{\text{TOTAL ACT. COST}} = 1.1\ \%$$

KOL - SMIRNOV D = 0.015

P > 0.05

Table (3.42)

*** 3-PARAMETER LOGNORMAL DIS. ***

74/4TH QUARTER DATA

ESTIMATION BY MULTINOMIAL MAX. LIKELIHOOD METHOD :-
C= 18.16 MEW= 4.878 SIGMA SQ.= 0.766

MEAN= 174.458 S.D.= 206.685

AMOUNT £	ACTUAL CL. NO.	EXPECTED CL. NO.	ACTUAL-EXPECTED	EXPECTED LOSS	(A-E)**2/E
1- 30	394.	393.	1.	15.	0.003
31- 60	452.	462.	-10.	-455.	0.216
61- 90	426.	414.	12.	906.	0.348
91- 120	348.	338.	10.	1055.	0.296
121- 150	272.	269.	3.	406.	0.033
151- 180	219.	212.	7.	1158.	0.231
181- 210	154.	168.	-14.	-2737.	1.167
211- 240	124.	134.	-10.	-2255.	0.746
241- 270	105.	108.	-3.	-767.	0.083
271- 300	78.	87.	-9.	-2569.	0.931
301- 330	75.	71.	4.	1262.	0.225
331- 360	58.	59.	-1.	-346.	0.017
361- 390	55.	49.	6.	2253.	0.735
391- 420	29.	40.	-11.	-4460.	3.025
421- 450	43.	34.	9.	3919.	2.382
451- 480	24.	29.	-5.	-2327.	0.862
481- 510	22.	24.	-2.	-991.	0.167
511- 540	24.	21.	3.	1576.	0.429
541- 570	19.	18.	1.	555.	0.056
571- 600	14.	15.	-1.	-585.	0.067
601- 700	42.	37.	5.	3252.	0.676
701- 800	28.	24.	4.	3002.	0.667
801- 900	20.	16.	4.	3402.	1.000
901-1000	17.	11.	6.	5703.	3.273
1001-1100	8.	7.	1.	1050.	0.143
1101-1200	5.	5.	0.	0.	0.000
1201-1300	1.	4.	-3.	-3751.	
1301-1400	5.	3.	2.	2701.	0.143
1401-1500	0.	2.	-2.	-2901.	
1501-1600	1.	2.	-1.	-1550.	
1601-1700	0.	1.	-1.	-1650.	
1701-1800	0.	1.	-1.	-1751.	
1801-1900	0.	1.	-1.	-1850.	
1901-2000	0.	1.	-1.	-1950.	
2001-2100	1.	0.	1.	2050.	
2101-2200	0.	0.	0.	0.	
2201-2300	0.	0.	0.	0.	
2301-2400	1.	0.	1.	2350.	3.125
TOTAL	3064.	3060.		3722.	21.044

D.F. = 24
P > 0.10

TOTAL EXP. LOSS / TOTAL ACT. COST = 0.7 %

KOL - SMIRNOV D =0.007
P > 0.20

Table (3.43)

75/1ST QUARTER DATA

ESTIMATION BY MULTINOMIAL MAX. LIKELIHOOD METHOD :-

C= 21.09 MEW= 4.915 SIGMA SQ.= 0.710

MEAN= 173.207 S.D.= 197.536

AMOUNT £	ACTUAL CL. NO.	EXPECTED CL. NO.	ACTUAL-EXPECTED	EXPECTED LOSS	(A-E)**2/E
1- 30	324.	325.	-1.	-15.	0.003
31- 60	387.	383.	4.	182.	0.042
61- 90	345.	352.	-7.	-528.	0.139
91- 120	289.	292.	-3.	-316.	0.031
121- 150	253.	234.	19.	2574.	1.543
151- 180	187.	186.	1.	165.	0.005
181- 210	138.	147.	-9.	-1759.	0.551
211- 240	114.	117.	-3.	-677.	0.077
241- 270	93.	94.	-1.	-256.	0.011
271- 300	67.	76.	-9.	-2569.	1.066
301- 330	63.	62.	1.	316.	0.016
331- 360	44.	51.	-7.	-2418.	0.961
361- 390	44.	42.	2.	751.	0.095
391- 420	35.	35.	0.	0.	0.000
421- 450	25.	29.	-4.	-1742.	0.552
451- 480	26.	24.	2.	931.	0.167
481- 510	18.	21.	-3.	-1486.	0.429
511- 540	18.	17.	1.	525.	0.059
541- 570	22.	15.	7.	3888.	3.267
571- 600	17.	13.	4.	2342.	1.231
601- 700	39.	31.	8.	5204.	2.065
701- 800	19.	19.	0.	0.	0.000
801- 900	18.	13.	5.	4252.	1.923
901-1000	12.	8.	4.	3802.	2.000
1001-1100	3.	6.	-3.	-3151.	1.500
1101-1200	1.	4.	-3.	-3451.	
1201-1300	1.	3.	-2.	-2501.	
1301-1400	0.	2.	-2.	-2701.	3.571
1401-1500	1.	2.	-1.	-1450.	
1501-1600	0.	1.	-1.	-1550.	
1601-1700	1.	1.	0.	0.	
1701-1800	0.	1.	-1.	-1751.	
1801-1900	0.	1.	-1.	-1850.	
1901-2000	0.	0.	0.	0.	
2001-2100	1.	0.	1.	2050.	
2101-2200	1.	0.	1.	2150.	
2201-2300	0.	0.	0.	0.	
2301-2400	0.	0.	0.	0.	
2401-2500	0.	0.	0.	0.	
2501-2600	0.	0.	0.	0.	
2601-2700	0.	0.	0.	0.	
2701-2800	0.	0.	0.	0.	
2801-2900	0.	0.	0.	0.	
2901-3000	0.	0.	0.	0.	
3001-3100	0.	0.	0.	0.	
3101-3200	0.	0.	0.	0.	
3201-3300	0.	0.	0.	0.	
3301-3400	0.	0.	0.	0.	
3401-3500	0.	0.	0.	0.	
3501-3600	1.	0.	1.	3550.	1.125
TOTAL	2607.	2607.		2510.	22.425

KOL - SMIRNOV D = 0.007 P > 0.20

D.F. = 23 P > 0.10

TOTAL EXP. LOSS / TOTAL ACT. COST = 0.6 %

Table (3.44)

*** 3-PARAMETER LOGNORMAL DIS. ***

75/2ND QUARTER DATA

ESTIMATION BY MULTINOMIAL MAX. LIKELIHOOD METHOD :-
C= 15.16 MEW= 4.872 SIGMA SQ.= 0.807

MEAN= 180.409 S.D.= 217.852

AMOUNT £	ACTUAL CL. NO.	EXPECTED CL. NO.	ACTUAL-EXPECTED	EXPECTED LOSS	(A-E)**2/E
1- 30	302.	302.	0.	0.	0.000
31- 60	374.	376.	-2.	-91.	0.011
61- 90	332.	337.	-5.	-378.	0.074
91- 120	277.	275.	2.	211.	0.015
121- 150	235.	218.	17.	2304.	1.326
151- 180	187.	173.	14.	2317.	1.133
181- 210	122.	137.	-15.	-2933.	1.642
211- 240	110.	110.	0.	0.	0.000
241- 270	80.	89.	-9.	-2299.	0.910
271- 300	72.	72.	0.	0.	0.000
301- 330	47.	59.	-12.	-3786.	2.441
331- 360	39.	49.	-10.	-3455.	2.041
361- 390	40.	41.	-1.	-376.	0.024
391- 420	38.	34.	4.	1622.	0.471
421- 450	29.	29.	0.	0.	0.000
451- 480	21.	24.	-3.	-1396.	0.375
481- 510	30.	21.	9.	4459.	3.857
511- 540	19.	18.	1.	525.	0.056
541- 570	17.	15.	2.	1111.	0.267
571- 600	11.	13.	-2.	-1171.	0.308
601- 700	36.	33.	3.	1951.	0.273
701- 800	22.	21.	1.	751.	0.048
801- 900	22.	14.	8.	6804.	4.571
901-1000	11.	10.	1.	951.	0.100
1001-1100	4.	7.	-3.	-3151.	1.286
1101-1200	3.	5.	-2.	-2301.	0.800
1201-1300	3.	4.	-1.	-1250.	
1301-1400	6.	3.	3.	4051.	0.571
1401-1500	2.	2.	0.	0.	
1501-1600	1.	2.	-1.	-1550.	
1601-1700	0.	1.	-1.	-1650.	
1701-1800	1.	1.	0.	0.	
1801-1900	0.	1.	-1.	-1850.	
1901-2000	1.	1.	0.	0.	
2001-2100	0.	0.	0.	0.	
2101-2200	0.	0.	0.	0.	
2201-2300	1.	0.	1.	2250.	0.500
TOTAL	2495.	2497.		1669.	23.897

D.F. = 24
P > 0.10

TOTAL EXP. LOSS / TOTAL ACT. COST = 0.4 %

KOL - SMIRNOV D= 0.01
P > 0.20

*** 3-PARAMETER LOGNORMAL DIS. ***

PREDICTION OF 74/4TH QUARTER CLAIMS COST
USING 73/4TH QUARTER MULT. MAXLIK. PARAMETERS :

C= 14.0 MEW= 4.700 SIGMA2=0.803
INFLATION RATE I=18.2% CALCULATED FROM :
GENERAL INDEX OF RETAIL PRICES

PREDICTION PARAMETERS ARE :- C= 16.5 MEW= 4.867 SIGMA2=0.803
MEAN CLAIM AMOUNT= 177.62 S.D.=215.53
ACTUAL 74/4TH PARAMETERS :- C= 18.2 MEW= 4.878 SIGMA2=0.766
MEAN CLAIM AMOUNT= 174.47 S.D.=206.72

AMOUNT £	ACT.NO.	EXP.NO.	A-E	EXP.LOSS	(A-E)**2/E
1- 30	394.	394.	0.	0.	0.000
31- 60	452.	464.	-12.	-546.	0.310
61- 90	426.	412.	14.	1057.	0.476
91- 120	348.	335.	13.	1371.	0.504
121- 150	272.	266.	6.	813.	0.135
151- 180	219.	210.	9.	1489.	0.386
181- 210	154.	167.	-13.	-2541.	1.012
211- 240	124.	133.	-9.	-2029.	0.609
241- 270	105.	107.	-2.	-511.	0.037
271- 300	78.	87.	-9.	-2569.	0.931
301- 330	75.	71.	4.	1262.	0.225
331- 360	58.	59.	-1.	-346.	0.017
361- 390	55.	49.	6.	2253.	0.735
391- 420	29.	41.	-12.	-4866.	3.512
421- 450	43.	34.	9.	3919.	2.382
451- 480	24.	29.	-5.	-2327.	0.862
481- 510	22.	25.	-3.	-1486.	0.360
511- 540	24.	21.	3.	1576.	0.429
541- 570	19.	18.	1.	555.	0.056
571- 600	14.	16.	-2.	-1171.	0.250
601- 700	42.	39.	3.	1951.	0.231
701- 800	28.	25.	3.	2251.	0.360
801- 900	20.	17.	3.	2551.	0.529
901-1000	17.	12.	5.	4752.	2.083
1001-1100	8.	8.	0.	0.	0.000
1101-1200	5.	6.	-1.	-1150.	0.167
1201-1300	1.	4.	-3.	-3751.	
1301-1400	5.	3.	2.	2701.	
1401-1500	0.	2.	-2.	-2901.	
1501-1600	1.	2.	-1.	-1550.	
1601-1700	0.	1.	-1.	-1650.	
1701-1800	0.	1.	-1.	-1751.	
1801-1900	0.	1.	-1.	-1850.	
1901-2000	0.	1.	-1.	-1950.	
2001-2100	1.	1.	0.	0.	
2101-2200	0.	0.	0.	0.	
2201-2300	0.	0.	0.	0.	
2301-2400	1.	0.	1.	2350.	
TOTAL	3064	3061		-4093.	16.599

CHI SQ. STAT. = 20.67 , D.F.=28 P = 0.85

TOTAL ACTUAL COST = 533707.
TOTAL EXPECTED COST = 537801.

KOL - SMIRNOV D = 0.01
P > 0.20

TOTAL EXP. LOSS
--------------- = -0.77 %
TOTAL ACT. COST

Table (3.46)

PREDICTION OF 75/1ST QUARTER CLAIMS COST
USING 74/1ST QUARTER MULT. MAXKIK. PARAMETERS :

C= 10.4 MEW= 4.650 SIGMA2=0.858
INFLATION RATE I=20.3% CALCULATED FROM :
GENERAL INDEX OF RETAIL PRICES

PREDICTION PARAMETERS ARE :- C= 12.5 MEW= 4.835 SIGMA2=0.858
MEAN CLAIM AMOUNT= 180.71 S.D.=225.20
ACTUAL 75/1ST PARAMETERS :- C= 21.1 MEW= 4.915 SIGMA2=0.710
MEAN CLAIM AMOUNT= 173.32 S.D.=197.69

AMOUNT £	ACT.NO.	EXP.NO.	A-E	EXP.LOSS	(A-E)**2/E
1- 30	324.	321.	3.	46.	0.028
31- 60	387.	404.	-17.	-774.	0.715
61- 90	345.	355.	-10.	-755.	0.282
91- 120	289.	285.	4.	422.	0.056
121- 150	253.	225.	28.	3794.	3.484
151- 180	187.	177.	10.	1655.	0.565
181- 210	138.	140.	-2.	-391.	0.029
211- 240	114.	112.	2.	451.	0.036
241- 270	93.	90.	3.	767.	0.100
271- 300	67.	73.	-6.	-1713.	0.493
301- 330	63.	60.	3.	947.	0.150
331- 360	44.	50.	-6.	-2073.	0.720
361- 390	44.	42.	2.	751.	0.095
391- 420	35.	35.	0.	0.	0.000
421- 450	25.	29.	-4.	-1742.	0.552
451- 480	26.	25.	1.	466.	0.040
481- 510	18.	21.	-3.	-1486.	0.429
511- 540	18.	18.	0.	0.	0.000
541- 570	22.	16.	6.	3333.	2.250
571- 600	17.	14.	3.	1756.	0.643
601- 700	39.	34.	5.	3252.	0.735
701- 800	19.	22.	-3.	-2251.	0.409
801- 900	18.	15.	3.	2551.	0.600
901-1000	12.	10.	2.	1901.	0.400
1001-1100	3.	7.	-4.	-4202.	2.286
1101-1200	1.	5.	-4.	-4602.	3.200
1201-1300	1.	4.	-3.	-3751.	
1301-1400	0.	3.	-3.	-4051.	
1401-1500	1.	2.	-1.	-1450.	
1501-1600	0.	2.	-2.	-3101.	
1601-1700	1.	1.	0.	0.	
1701-1800	0.	1.	-1.	-1751.	
1801-1900	0.	1.	-1.	-1850.	
1901-2000	0.	1.	-1.	-1950.	
2001-2100	1.	1.	0.	0.	
2101-2200	1.	0.	1.	2150.	
2201-2300	0.	0.	0.	0.	
2301-2400	0.	0.	0.	0.	
2401-2500	0.	0.	0.	0.	
2501-2600	0.	0.	0.	0.	
2601-2700	0.	0.	0.	0.	
2701-2800	0.	0.	0.	0.	
2801-2900	0.	0.	0.	0.	
2901-3000	0.	0.	0.	0.	
3001-3100	0.	0.	0.	0.	
3101-3200	0.	0.	0.	0.	
3201-3300	0.	0.	0.	0.	
3301-3400	0.	0.	0.	0.	
3401-3500	0.	0.	0.	0.	
3501-3600	1.	0.	1.	3550.	
TOTAL	2607	2601		-10102.	18.297

Table (3.46) - continued

CHI SQ. STAT. = 25.76 , D.F. = 28 P> 0.50

TOTAL ACTUAL COST = 452059.
TOTAL EXPECTED COST = 462161. KOL - SMIRNOV D= 0.01

P> 0.20

TOTAL EXP. LOSS
--------------- -2.23 %
TOTAL ACT. COST

Table (3.47)

*** 3-PARAMETER LOGNORMAL DIS. ***

PREDICTION OF 75/2ND QUARTER CLAIMS COST
USING 74/2ND QUARTER MULT. MAXLIK. PARAMETERS :

C= 14.9 MEW= 4.735 SIGMA2=0.760
INFLATION RATE I=24.3% CALCULATED FROM :
GENERAL INDEX OF RETAIL PRICES

PREDICTION PARAMETERS ARE :- C= 18.5 MEW= 4.953 SIGMA2=0.760
MEAN CLAIM AMOUNT= 188.44 S.D.=220.81
ACTUAL 75/2ND PARAMETERS :- C= 15.2 MEW= 4.872 SIGMA2=0.807
MEAN CLAIM AMOUNT= 180.29 S.D.=217.79

AMOUNT £	ACT.NO.	EXP.NO.	A-E	EXP.LOSS	(A-E)**2/E
1- 30	302.	279.	23.	356.	1.896
31- 60	374.	349.	25.	1137.	1.791
61- 90	332.	325.	7.	528.	0.151
91- 120	277.	273.	4.	422.	0.059
121- 150	235.	222.	13.	1761.	0.761
151- 180	187.	178.	9.	1489.	0.455
181- 210	122.	143.	-21.	-4105.	3.084
211- 240	110.	116.	-6.	-1353.	0.310
241- 270	80.	94.	-14.	-3577.	2.085
271- 300	72.	77.	-5.	-1427.	0.325
301- 330	47.	63.	-16.	-5048.	4.063
331- 360	39.	52.	-13.	-4491.	3.250
361- 390	40.	44.	-4.	-1502.	0.364
391- 420	38.	37.	1.	406.	0.027
421- 450	29.	31.	-2.	-871.	0.129
451- 480	21.	26.	-5.	-2327.	0.962
481- 510	30.	22.	8.	3964.	2.909
511- 540	19.	19.	0.	0.	0.000
541- 570	17.	16.	1.	555.	0.063
571- 600	11.	14.	-3.	-1756.	0.643
601- 700	36.	35.	1.	651.	0.029
701- 800	22.	23.	-1.	-751.	0.043
801- 900	22.	15.	7.	5953.	3.267
901-1000	11.	10.	1.	951.	0.100
1001-1100	4.	7.	-3.	-3151.	1.286
1101-1200	3.	5.	-2.	-2301.	0.800
1201-1300	3.	4.	-1.	-1250.	
1301-1400	6.	3.	3.	4051.	
1401-1500	2.	2.	0.	0.	
1501-1600	1.	2.	-1.	-1550.	
1601-1700	0.	1.	-1.	-1650.	
1701-1800	1.	1.	0.	0.	
1801-1900	0.	1.	-1.	-1850.	
1901-2000	1.	1.	0.	0.	
2001-2100	0.	0.	0.	0.	
2101-2200	0.	0.	0.	0.	
2201-2300	1.	0.	1.	2250.	
TOTAL	2495	2490		-14487.	28.850

CHI SQ. STAT. = 29.9 , D.F.= [illegible] P>0.30

TOTAL ACTUAL COST = 44930[illegible] KOL - SMIRNOV D = 0.0[illegible]
TOTAL EXPECTED COST = 463795. P = 0.[illegible]

TOTAL EXP. LOSS / TOTAL ACT. COST = -3.22 %

CHAPTER 4

The Weibull Distribution

4.1 Introduction

The Weibull distribution is named after Waloddi Weibull, the Swedish physicist, who in 1939 derived it empirically from practical considerations of the size effect on failures in solids. He used this distribution as a model for the breaking strength of materials.

The Weibull is a flexible distribution which is related to the exponential family. For a suitable choice of parameters, it is positively skewed with a tail longer than that of the exponential distribution. Over the years, it has been used to represent the distribution of random outcomes of various phenomena in science and engineering. Weibull (1951) has fitted this distribution to data on :

1 - yield strength of Bofors steel ;
2 - size distribution of fly ash ;
3 - fibre strength of Indian cotton ;
4 - statures for adult males, born in the British Isles ;

etc.

In recent years the Weibull distribution has become one of the most widely used statistical distributions in the fields of reliability, life and fatigue testing of engineering systems and their components. A bibliography on this distribution and its applications appears in Johnson and Kotz (1970).

The (negative) exponential distribution has been used in general insurance, and in particular in fire insurance, as a model for the distribution of claim amounts (see, for instance, Ramachandran (1970)). It is important to see how the more flexible Weibull distribution would perform in this respect. In addition, the fact that the Weibull has successfully represented the distribution of positively skewed random variables, in various fields, encourages us to consider its application as the model for the distribution of claim amounts in general insurance. Therefore, in this chapter, the Weibull distribution will be initially defined and some of its properties will be mentioned. Then a graphical procedure will be suggested for testing whether a particular set of data may be regarded as a sample from a Weibull population. The accidental damage data will then be tested accordingly. In section 4.6 the least squares and multinomial maximum likelihood methods of parameter estimation from grouped data will be considered. These methods will then be used to fit the Weibull model to the samples of accidental damage data. The effects of the inflation on the parameters of the model will be studied in section 4.8. Finally, the conclusions of this chapter will be discussed in section 4.9.

4.2 Definition

A random variable X is said to have a 3-parameter Weibull distribution if its distribution function, denoted by

$W(x\ ;\ C, A, B)$ is of the form

$$P(X \leq x) = 1 - \exp\left[-\left(\frac{x - C}{A}\right)^B\right] \; ; \; x > C \qquad (4.2\text{-}1)$$

$$= 0$$

for $A > 0,\ B > 0,\ -\infty < C < \infty$

This distribution is related to the standard exponential distribution via the transformation

$$Y = \left(\frac{X - C}{A}\right)^B \qquad (4.2\text{-}2)$$

which leads to $P(Y \leq y) = 1 - e^{-y}$, $y > 0$.

Hence Y has the standard exponential distribution (i.e. with mean = 1). The probability density function (p.d.f.) of X, denoted by $f_W(x;C,A,B)$, is

$$f_W(x;\, C,A,B) = \frac{B}{A}\left(\frac{x - C}{A}\right)^{B-1} \exp\left[-\left(\frac{x - C}{A}\right)^B\right] \; ; \; x > C \qquad (4.2\text{-}3)$$

It is apparent that C is a location parameter below which the values of X are not realized. C is often called the threshold parameter. A and B are scale and shape parameters respectively. If in (4.2-1) and (4.2-3) we put C = 0 we will have the 2-parameter Weibull distribution and probability density functions respectively.

4.3 Properties of the Weibull Distribution

In this section we will briefly report some properties of the 3-parameter Weibull distribution. More details on this can be found in Johnson and Kotz (1970), Bury (1975) or Mann et al. (1974). The properties of the 2-parameter distribution will be obtained by putting C = 0 in the following.

The Weibull distribution can represent various shapes through the parameter B. In particular, when $B < 1$ the p.d.f. is monotone decreasing, with a tail longer than that of the exponential distribution, and there is no mode. For $B = 1$ this distribution becomes equivalent to an exponential distribution with C and A as location and scale parameters respectively. For $B > 1$ the p.d.f. of the Weibull distribution is unimodal, with the mode at

$$x = C + A\,[(B - 1)/B]^{1/B} \qquad (4.3\text{-}1)$$

In this case the p.d.f. is positively skewed for $B < 3.6$ (approx.), and negatively skewed for $B > 3.6$ (approx.). For $B = 3.6$ (approximately) the shape of the Weibull distribution will be similar to that of a normal distribution.

To find the moments of X about zero we use the fact that Y, as defined in (4.2-2), has the standard exponential distribution.

Let $Z = \dfrac{X - C}{A}$, therefore $Y = Z^B$ and $X = C + AZ$.

The rth moment of Z about zero is

$$E(Z^r) = E\left[(Z^B)^{r/B}\right] = E(Y^{r/B}) = \frac{r}{B}\text{ th moment of Y about zero}$$

hence from the theory of exponential distribution it follows that

$$E(Z^r) = \Gamma\left(\frac{r}{B} + 1\right) \qquad (4.3\text{-}2)$$

Therefore moments of Z and hence of X can be obtained. For $r = 1$ we have $E(Z) = \Gamma\left(\frac{1}{B} + 1\right)$.

Therefore the mean of X will be

$$E(X) = C + A\,\Gamma\left(\frac{1}{B} + 1\right) \qquad (4.3\text{-}3)$$

and by using r = 2 in (4.3-2) we can derive the variance of X as

$$\mathrm{var}(X) = A^2 \left\{\Gamma\left(\frac{2}{B} + 1\right) - \left[\Gamma\left(\frac{1}{B} + 1\right)\right]^2\right\} \qquad (4.3\text{-}4)$$

The coefficients of skewness and kurtosis can be obtained as complicated expressions involving the Gamma functions. These are derived in Bury (1975).

The median of the distribution can easily be shown to be at

$$x = C + A\,(\log 2)^{1/B} \qquad (4.3\text{-}5)$$

The quantile of order q is at

$$x = C + A\left[\log \frac{1}{1-q}\right]^{1/B} \qquad (4.3\text{-}6)$$

In many applications there is usually no explicit theoretical justification for the use of the Weibull distribution as a model. However, if a negative exponential model can be justified, or shown to be reasonable, for a random phenomenon, then its replacement by a Weibull distribution will allow greater flexibility in the model. It is with these remarks in mind, and the fact that the exponential distribution has been used as a model for claim amounts, that we study the Weibull distribution in this chapter.

4.4 The Graphical Test for the Weibull Distribution

We explain this test for the 3-parameter case. For the 2-parameter Weibull distribution the argument holds by putting C = 0.

Let us denote the cumulative distribution function $W(x; C,A,B)$ simply by $W(x)$ where

$$W(x) = 1 - \exp\left[-\left(\frac{x-C}{A}\right)^B\right]$$

Hence

$$1 - W(x) = \exp\left[-\left(\frac{x-C}{A}\right)^B\right]$$

by taking natural logarithm twice on both sides and simple manipulations we can show that

$$\log\log\frac{1}{1-W(x)} = B\log(x-C) - B\log A \qquad (4.4\text{-}1)$$

Therefore the locus of the points $\left(\log(x-C),\ \log\log\frac{1}{1-W(x)}\right)$ is a straight line. This provides us with a means of testing if a sample of data is from a Weibull population. Let us assume that we are given a sample of n independent observations on the random variable X. At a particular value x_i we define the sample empirical distribution function $F(x_i)$, which is an estimate of $W(x_i)$, as

$$F(x_i) = \text{the proportion of observations} \leq x_i \qquad (4.4\text{-}2)$$

Therefore if the value of C is known, for the 3-parameter case, and the sample is from a Weibull population we would expect the points $\left(\log(x_i - C)\ ,\ \log\log\frac{1}{1-F(x_i)}\right)$ to lie approximately on a straight line - say, line 1 in figure (4.1). The parameters A and B can be estimated from the slope and intercept of this line. The use of a special graph paper called the Weibull probability paper makes the tasks of calculations, for plotting the points, and of estimation of the parameters easier. On this graph paper, one of the axes is scaled

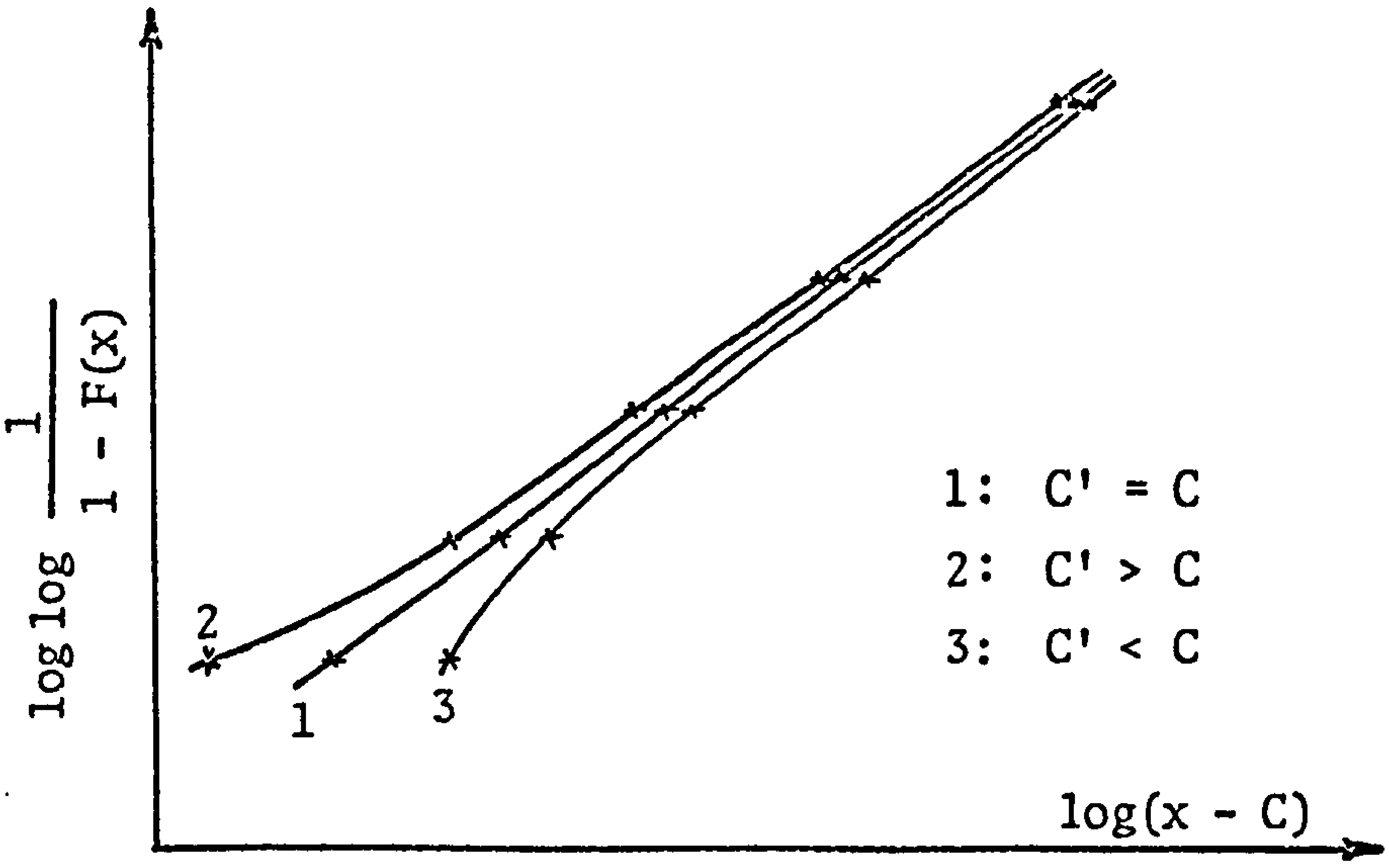


Figure (4.1) - Graph of the points $\left(\log(x - C), \log\log \frac{1}{1 - F(x)}\right)$

logarithmically while the other is graduated such that the proportions F(x) are converted to $\log\log \frac{1}{1 - F(x)}$. Therefore, for plotting, only the points (x - C, F(x)) need to be calculated. There are special scales provided on the paper which allow, after a successful fit of a straight line to the points by eye, the estimates of the scale and shape parameters to be read from the graph.

So far we have assumed that the value of C is known. If this is not so then the values of x - C cannot be calculated. In this case C has to be estimated beforehand. If we use C' for C where C' > C (i.e. if we overestimate C) then, because this will have a larger impact on log(x - C) for smaller values of x, we will find that the points lie on a convex curve - plot 2 in figure (4.1). If we underestimate C (i.e. use C' < C) then a concave curve will be obtained - plot 3 in figure (4.1). We faced the same problem in section 3.11 on testing for 3-parameter lognormality. The procedure which was applied in that section can be adopted here. Therefore we try various values of C and for each value plot the points

$\left(\log (x - C) \;,\; \log\log \frac{1}{1 - F(x)}\right)$. By observing the resulting plots we decide on the best value of C which rectifies the points to lie approximately on a straight line. If, in fact, such a C exists then we can conclude that our sample is from a 3-parameter Weibull population. If necessary, the parameters A and B can then be estimated, as before, from the slope and intercept of the line.

The initial value of C to be tried always is C = 0 which will either indicate that the sample is from a 2-parameter Weibull population or, alternatively, that by choosing C = 0 we are underestimating or overestimating C. If it indicates underestimation then the model will involve a positive location parameter and its distribution function can be expressed as

$$W(x;\, C,A,B) \;=\; 1 - \exp\left[-\left(\frac{x - C}{A}\right)^{B}\right] \qquad (4.4\text{-}3)$$

If C = 0 indicates overestimation then the model may involve a negative location parameter. Hence, for the sake of avoiding the use of negative values for parameter C, the model can be expressed as

$$W(x;\, -C,A,B) \;=\; 1 - \exp\left[-\left(\frac{x + C}{A}\right)^{B}\right] \;;\quad C > 0 \qquad (4.4\text{-}4)$$

4.5 The Weibull Graphical Test on AD Data

We wrote the computer program P13 to use a sample of grouped data, as input, and to plot the points $\left(\log(x - C),\; \log\log \frac{1}{1 - F(x)}\right)$

for various values of C. For each sample of accidental damage data, as presented in tables (1.1) to (1.7), values of C = 0, 10, 15, 20 and 25 were supplied to the program. For each C, the program then calculated and plotted the resulting array of points. The graphs for each of the seven samples of data are presented in figures (4.2-a) and (4.2-b). It can be seen that for each sample, C = 0 has produced a concave curve with a marked deviation from a straight line in the lower values of x. Therefore C = 0 has underestimated the true value of C. On the other hand, for C = 25, for each sample, a convex curve has been produced indicating an overestimation of C. For each sample C = 10 also indicates an underestimation of C. Therefore the true value of C, in each case, must lie between 10 to 25. In fact it is apparent from the graphs that C = 15 or 20 are about the best choice to make the array of points lie on a straight line. There are slight deviations from the straight line in the upper tail values for each sample. These are, as in the case of the lognormal distribution, due to lack of sufficient observations in the long tail of the sample histogram.

We feel justified that a 3-parameter Weibull model with a positive location parameter C, in the range of 10 to 25, and cumulative distribution function of the form (4.4-3) may be used to represent the distribution of our accidental damage data.

We know that some policy holders do not claim for small amounts because their no claim bonus is worth more than the amount they would recover by a claim. Considering that our proposed Weibull model involves terms of the form x - C we may interprèt parameter C as the amount below which claims are not made.

Figure (4.2-a)

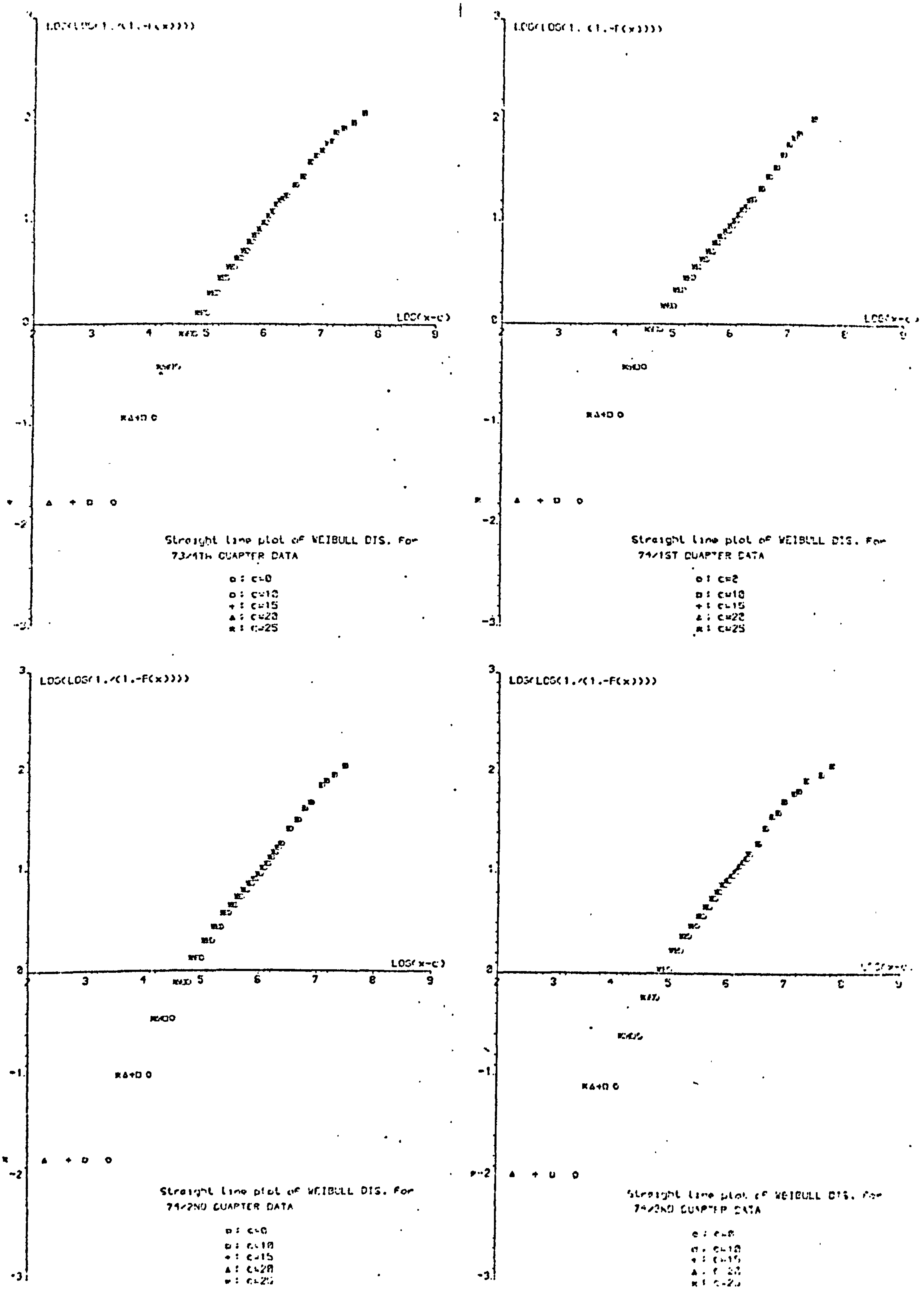
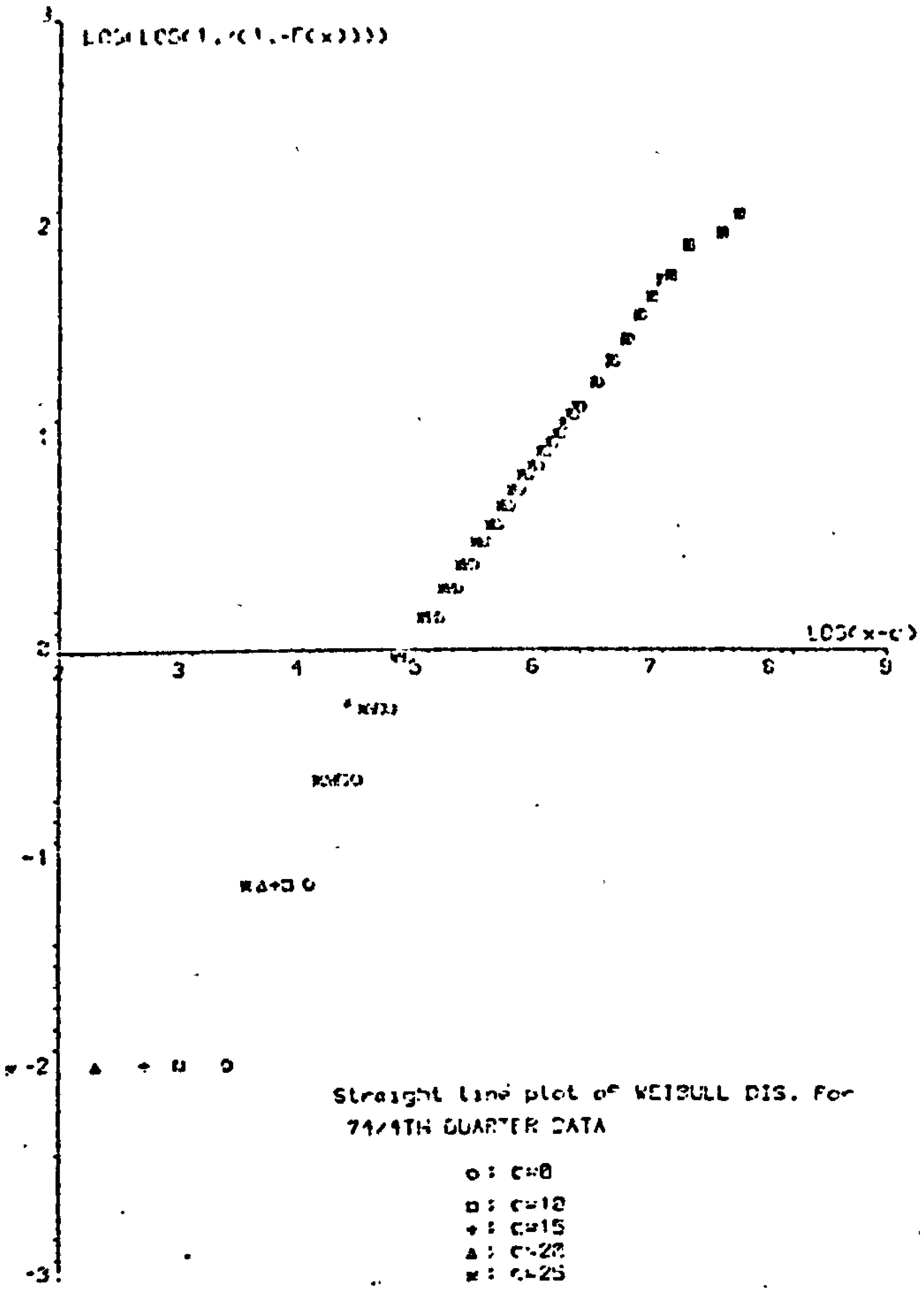
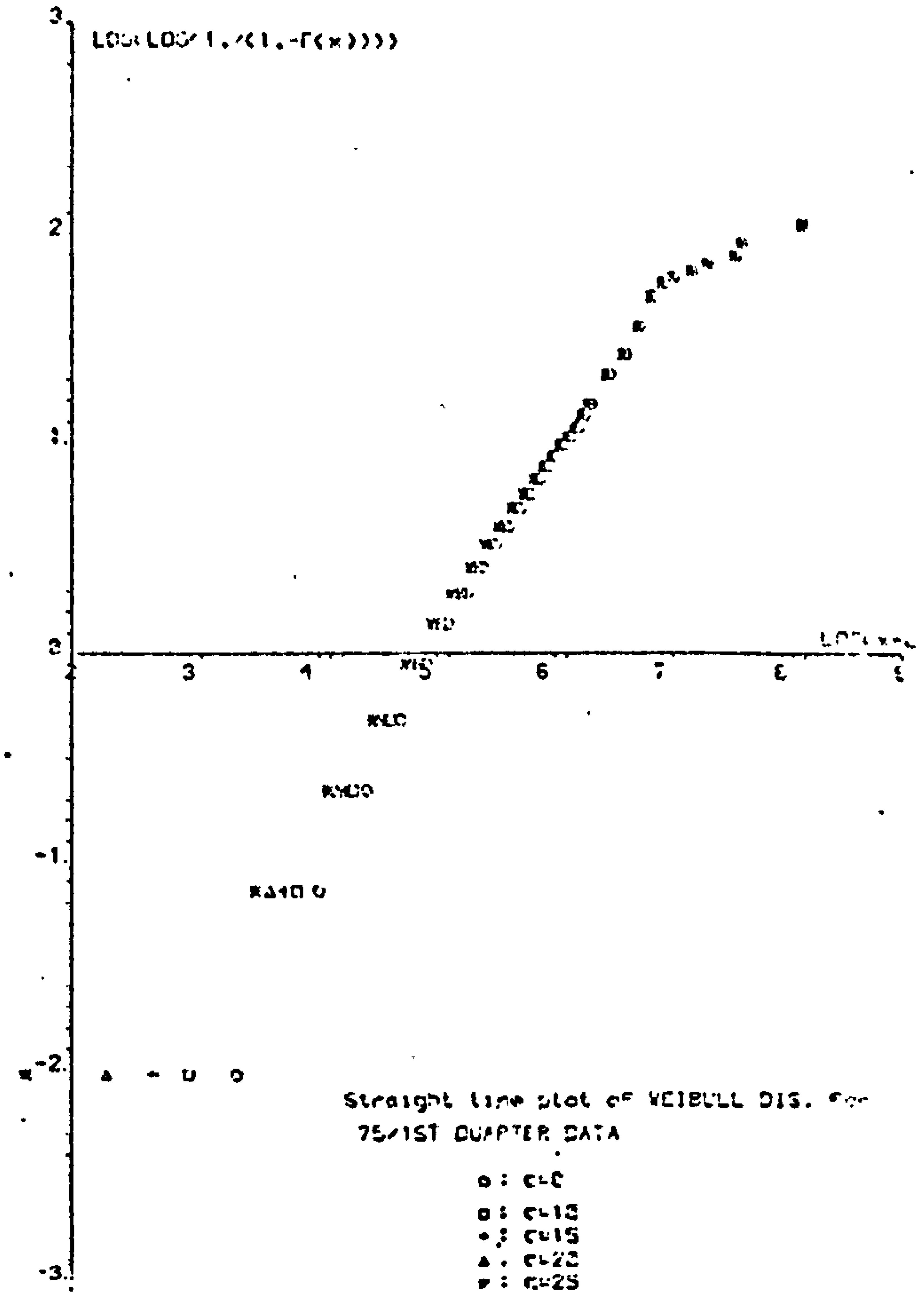
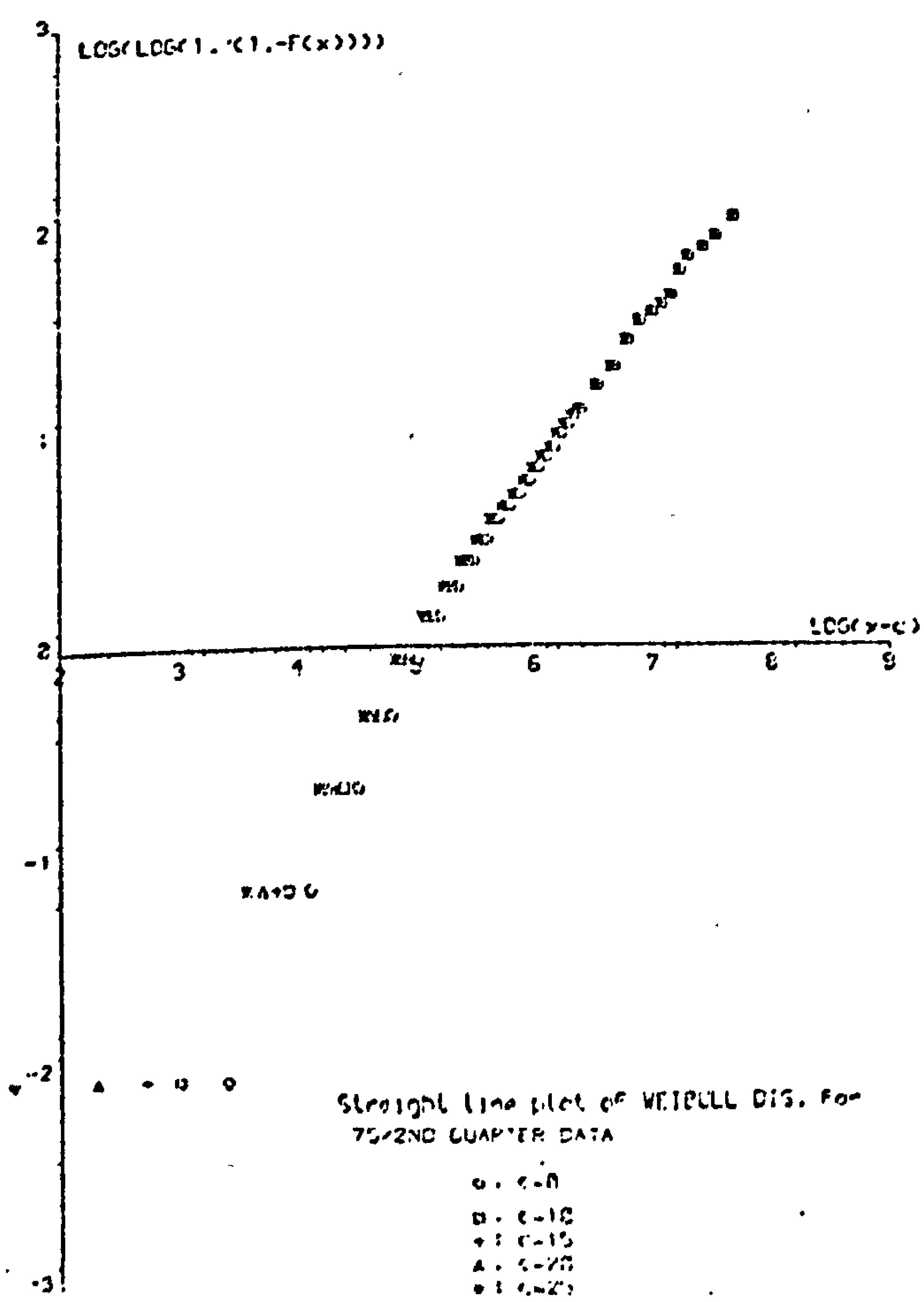

LOG(LOG(1./(1.-F(x))))
LOG(x-c)
Straight line plot of WEIBULL DIS. For 73/4TH QUARTER DATA
Straight line plot of WEIBULL DIS. For 74/1ST QUARTER DATA
Straight line plot of WEIBULL DIS. For 74/2ND QUARTER DATA
c=0
c=10
c=15
c=20
c=25

Figure (4.2-b)


LOG(LOG(1./(1.-F(x))))
LOG(x-c)
Straight line plot of WEIBULL DIS. For
74/4TH QUARTER DATA
o : c=0
□ : c=10
+ : c=15
▲ : c=20
* : c=25


LOG(LOG(1./(1.-F(x))))
Straight line plot of WEIBULL DIS. For
75/1ST QUARTER DATA
o : c=0
□ : c=10
+ : c=15
▲ : c=20
* : c=25


LOG(LOG(1./(1.-F(x))))
LOG(x-c)
Straight line plot of WEIBULL DIS. For
75/2ND QUARTER DATA

4.6 Estimation of the Parameters of the Weibull Distribution

The estimation problem for the Weibull distribution has been dealt with by various authors. However, most of the concentration appears to have been on estimation from ungrouped data and for the 2-parameter distribution. Kao (1956) considered the methods of least squares and maximum likelihood for ungrouped data when the scale and shape parameters, A and B respectively, are unknown. In addition, in Kao (1958) he dealt with estimation from grouped data, 2-parameter case, by the methods of multinomial maximum likelihood and minimum Chi-square.

The method of moments can be used which involves equating the first two (or three in the 3-parameter case) sample moments to their corresponding population values and solving the resulting equations for the unknown parameters. Dubey (1963) reports that he has found the asymptotic efficiencies for the moment estimators of the location and scale parameters when the shape parameter is known. These, he states, depend on the value of the shape parameter and are quite high in some cases.

The method of quantiles may also be used for estimation. This consists of equating two (or three in the 3-parameter case) sample quantiles to their corresponding population values. Dubey (1967) has applied this method for estimating parameters of the 2-parameter distribution. He states that the 17 and 97% quantiles give an estimate of the shape parameter which is asymptotically 66% efficient when compared with the maximum likelihood estimator. The 40 and 82% quantiles estimate the scale parameter with an asymptotic efficiency of 82%.

A relatively quick and easy method of estimating the parameters of the Weibull distribution is the graphical method which we explained as part of the Weibull graphical test in section 4.4. This method has been fully explored by Kao and we can, for example, refer to Kao (1959).

Haan and Beer (1967) deal with estimation from ungrouped data by the ordinary maximum likelihood method for the 3-parameter Weibull. In this case a system of equations involving powers of the unknown parameters need to be solved iteratively.

When data only in grouped form is available the estimation problem usually becomes more complicated. It is possible to assume that the observations in every interval are concentrated at the mid-point of that interval and hence use the estimation procedures for ungrouped data. This assumption, even when reasonably justified, results in some loss of efficiency and may introduce certain biases in the estimators. Therefore, as we showed in the 2-parameter lognormal case, it is best to use methods which are directly suitable for grouped data. Of these, the most efficient method is the multinomial maximum likelihood. This, as we mentioned earlier, was considered for the 2-parameter Weibull by Kao (1958). In the present work we will extend this method to the case when all three parameters are unknown. In addition, we will consider the method of least squares and will propose a computing technique for estimating all three unknown parameters.

Let us assume that we have a sample of grouped data where n independent random observations on a random variable X (in our case, the claim amount) have been grouped according to their size into k mutually exclusive intervals. Further, let n_i be the number of observations (claims) in the class interval (x_{i-1}, x_i) for $i = 1,2,\ldots,k$ such that

$$n = \sum_{i=1}^{k} n_i$$

Also, let us assume that X has a 3-parameter Weibull distribution with the cumulative distribution function of the form

$$W(x;\, C,A,B) = 1 - \exp\left[-\left(\frac{x - C}{A}\right)^B\right];\ x > C \qquad (4.6\text{-}1)$$

We define the sample empirical distribution function as

$$F(x) = \text{the proportion of observations} \leq x \qquad (4.6\text{-}2)$$

In the following sections we shall use the above notation. Although we will be considering the estimation problem of the 3-parameter Weibull distribution, the methods can be directly used for the 2-parameter Weibull by putting $C = 0$.

4.6.1 The Method of Least Squares

We showed in section 4.4 that if our sample is from a Weibull population then the points $\left(\log(x_i - C),\ \log\log \frac{1}{1 - F(x_i)}\right)$, for $i = 1, \ldots, k\text{-}1$ are expected to lie on the straight line

$$\log\log \frac{1}{1 - W(x)} = B \log(x - C) - B \log A \qquad (4.6\text{-}3)$$

It should be noted that we cannot use x_k because $F(x_k) = 1$ and hence $\frac{1}{1 - F(x_k)}$ will be indeterminate. To find the unknown parameters we can use the least squares regression technique which consists of minimising SSD, the sum of squares of deviations, where

$$SSD = \sum_{i=1}^{k-1} \left[\log\log \frac{1}{1 - F(x_i)} - B \log(x_i - C) + B \log A \right]^2 \quad (4.6\text{-}4)$$

simultaneously with respect to A, B and C. If we equate to zero the partial derivatives of SSD, with respect to A, B and C, we will produce a set of non-linear equations in the parameters which can be solved, laboriously, by an iterative method. In addition, for this purpose starting values for the parameters need to be obtained by a quick method of estimation, say, the graphical.

We instead suggest using the computing technique which we proposed for the 3-parameter lognormal distribution in 3.12.1.

Let $V_i = \log\log \frac{1}{1 - F(x_i)}$ and $U_i = \log(x_i - C)$.

If C was known, or was fixed, to be equal to C_o, say, then the points (U_i, V_i) would be completely determined from the sample. Hence a straight line, say

$$V = aU + b \quad (4.6\text{-}5)$$

could be fitted by the least squares regression technique with values of a and b calculated from (3.7-24) and (3.7-25) of chapter 3. The estimates of the unknown parameters A and B would then be

$$\hat{B} = a \quad (4.6\text{-}6)$$

and

$$\hat{A} = \exp\left(-\frac{b}{a}\right) \quad (4.6\text{-}7)$$

Because the value of C is unknown our computing procedure, described fully in section 3.12.1, is to find directly the minimum value of SSD simultaneously with respect to C, A and B. For this purpose we fix C, systematically, at different values and for each value estimate A and B, according to the previous paragraph, and find their corresponding value of SSD from equation (4.6-4). Hence we find that value of C, with the required degree of accuracy, and its corresponding parameters A and B which give the minimum value of SSD. These values of A, B and C are the least squares estimates of the Weibull parameters. Notice that for this procedure we do not require any starting values for the parameters. The procedure when programmed on the computer is very fast (depending on the required degree of accuracy) and easy to use on the interactive terminals.

4.6.2 Multinomial Maximum Likelihood Method

This method was previously considered for the lognormal distribution in section 3.7.7. Here we will give the loglikelihood function for the 3-parameter Weibull distribution. The computing technique to find the estimates of the parameters, by maximizing the loglikelihood is that described in section 3.7.7.

Let us assume that we have a sample of grouped data as defined in section 4.6. Let p_i be the probability that an observation (claim) occurs in the interval (x_{i-1}, x_i). Adopting the notation of 4.6, we have

$$p_i = W(x_i;\, C,A,B) - W(x_{i-1};\, C,A,B)$$

$$\text{for } i = 2,3, \ldots , k$$

and

$$p_1 = W(x_1;\, C,A,B) - W(C;\, C,A,B)$$

Using (4.2-1) we have

$$p_i = \exp\left[-\left(\frac{x_{i-1} - C}{A}\right)^B\right] - \exp\left[-\left(\frac{x_i - C}{A}\right)^B\right]$$

$$\text{for } i = 2, 3, \ldots , k .$$

and

$$p_1 = 1 - \exp\left[- \left(\frac{x_1 - C}{A} \right)^B \right]$$

The likelihood function of the sample is proportional to L where

$$L = \prod_{i=1}^{k} p_i^{n_i}$$

and the loglikelihood function, log L, is

$$\log L = \sum_{i=1}^{k} n_i \log p_i$$

$$= n_1 \left\{ 1 - \exp\left[- \left(\frac{x_1 - C}{A} \right)^B \right] \right\} +$$

$$+ \sum_{i=2}^{k} n_i \left\{ \exp\left[- \left(\frac{x_{i-1} - C}{A} \right)^B \right] - \exp\left[- \left(\frac{x_i - C}{A} \right)^B \right] \right\} \quad (4.6\text{-}8)$$

The maximum likelihood estimates of the parameters are obtained by simultaneously maximizing log L with respect to A, B and C. The computing procedure of section 3.7.7. requires starting values for the parameters. The least squares, or some other, estimates of A, B and C may be used for this purpose.

4.7 Application of the Weibull Model to the AD Data

We wrote computer program P14 for estimating the parameters of the Weibull distribution, by the least squares method, from a sample of grouped data. This program was run with samples of accidental damage data as presented in tables (1.1) to (1.7). For each sample first $C = 0$ (i.e. the 2-parameter Weibull) was tried. The program provided estimates of the parameters, A and B, as well as an extensive table of results. The Chi-square statistic, X^2, was calculated by the program

as a measure of agreement between the fitted model and actual sample values. A summary of the results is presented as part of table (4.1). It is apparent that A has increased over time while B is, approximately constant, about 0.9. The X^2 statistics are large for all the samples, hence indicating a general disagreement between the model and actual sample values. Examination of the components of the Chi-square statistics showed that very large contributions are made by a small number of the intervals in the lower tail of the distribution.
For each sample, the 3-parameter distribution was then fitted by finding the optimum value of C as described in section 4.6.1. The results are summarized in table (4.2). It can be seen that estimates of C, the amount below which claims are not made, is about £20. The values of A and C have generally increased over time. Because the mean and standard deviation of the Weibull distribution are functions of all three parameters, they showed an increasing trend over time as well. The Chi-square statistics are considerably smaller than in the 2-parameter case. However, they still indicate highly significant differences between the model and actual sample values. Again the major contributions to X^2 values are from 2 or 3 of the intervals only. The total expected loss statistics are relatively small as indicated by R, the ratio of T to total actual cost of claims. Therefore, the 3-parameter Weibull distribution provides a better fit, in terms of the X^2 statistic, to the actual data.

To see how closely the least squares line fits the sample points the computer program P15 was written to plot them. The graphs for the accidental damage data are presented in figures (4.3-a) and (4.3-b). Despite the large values of the X^2 statistic we observe that the sample points lie very closely on the least squares line. This, of

Figure (4.3-a)

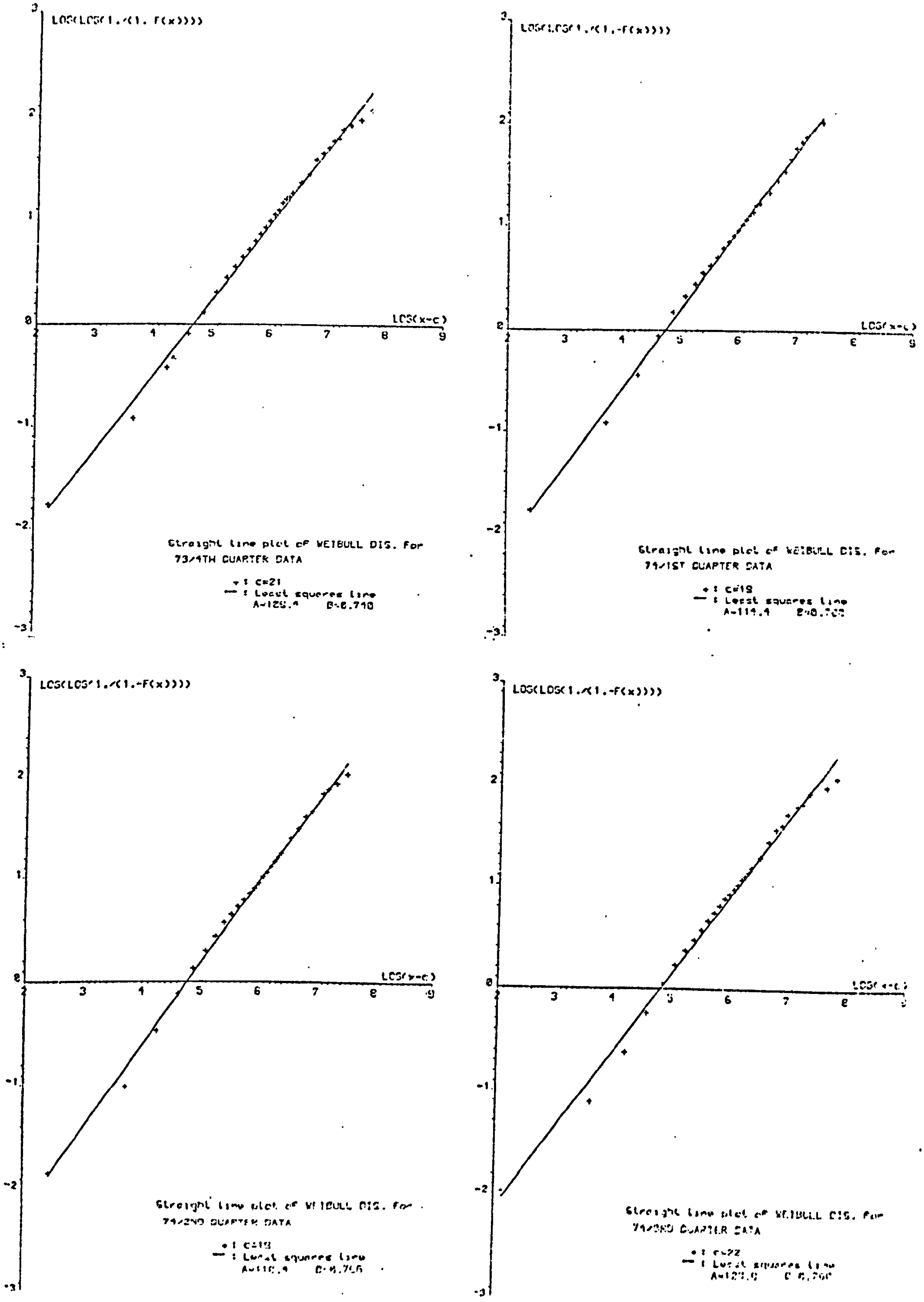
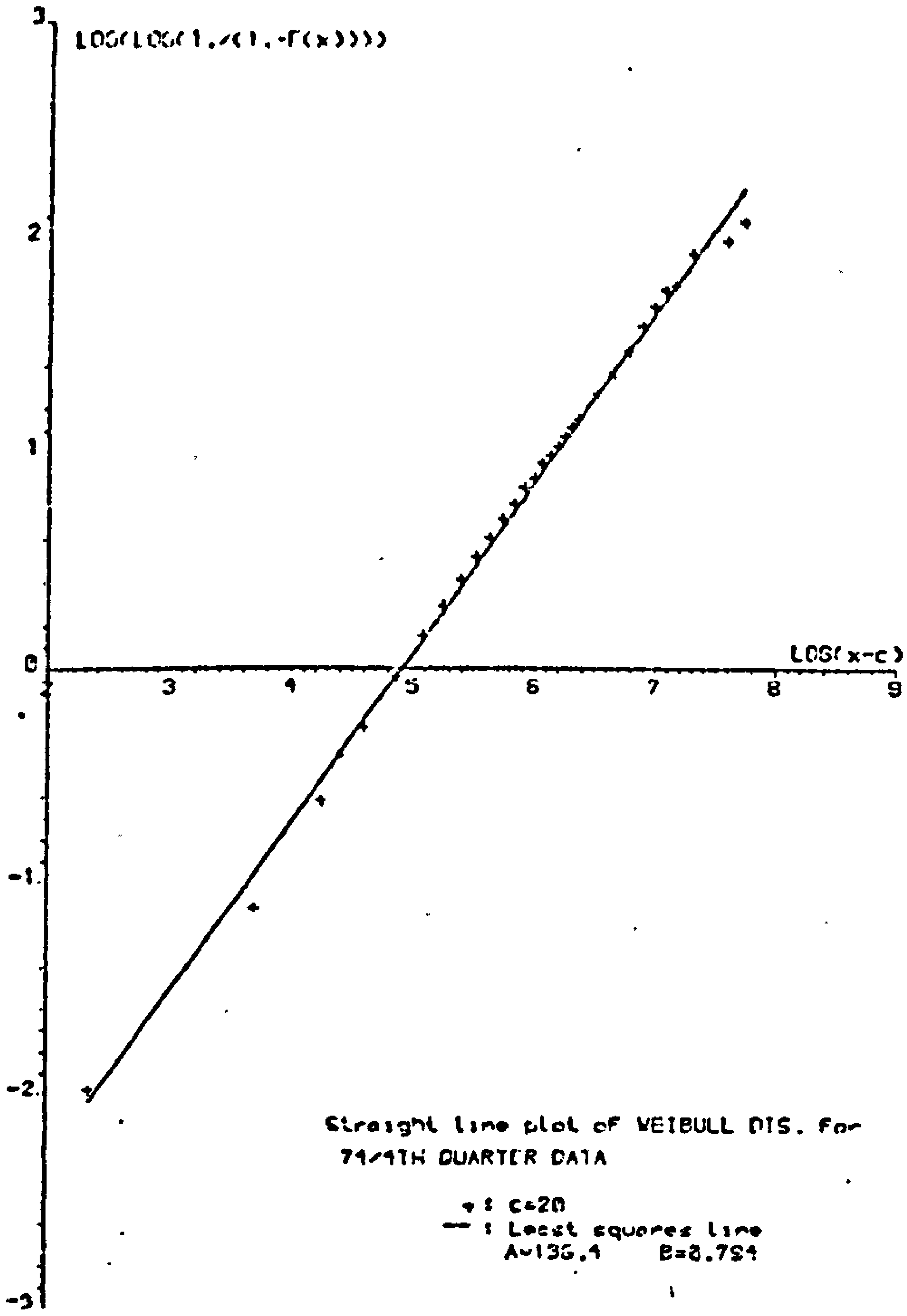
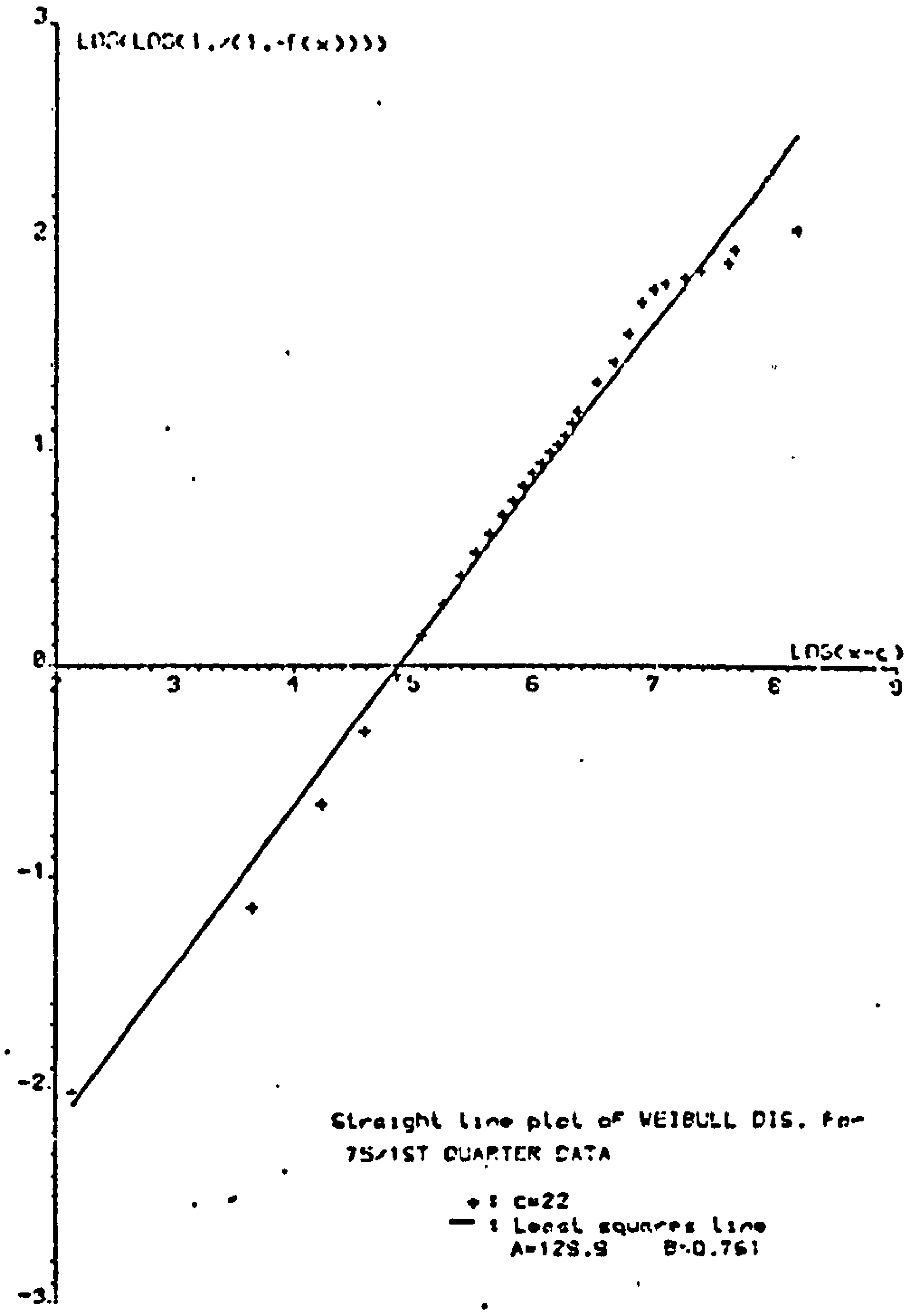
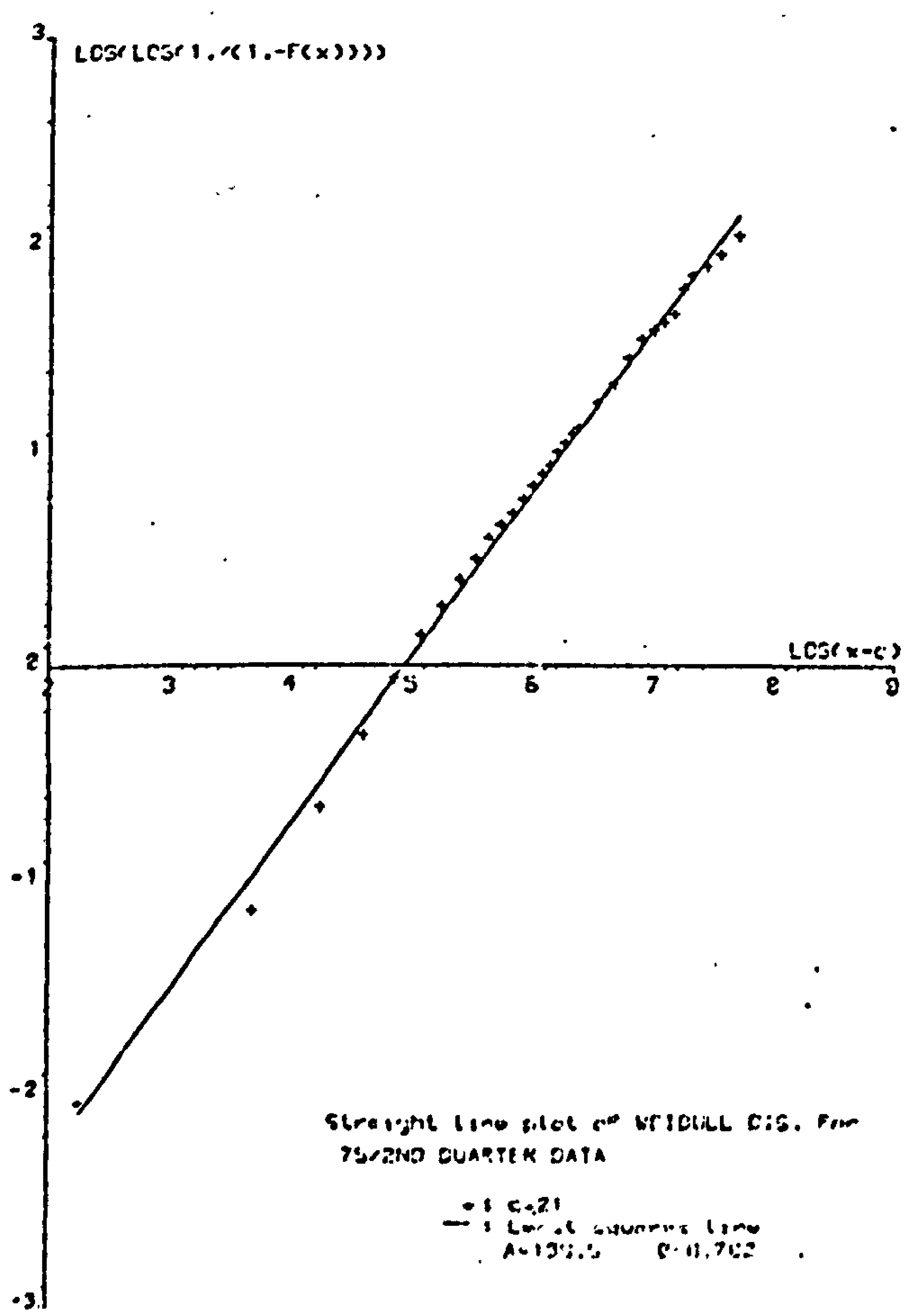

LOG(LOG(1./(1.-F(x))))
LOG(x-c)
Straight line plot of WEIBULL DIS. For
73/4TH QUARTER DATA
+ : c=21
— : Least squares line
Straight line plot of WEIBULL DIS. For
74/1ST QUARTER DATA
+ : c=19
— : Least squares line
A=114.4
Straight line plot of WEIBULL DIS. For
74/2ND QUARTER DATA
+ : c=19
— : Least squares line
Straight line plot of WEIBULL DIS. For
74/3RD QUARTER DATA
+ : c=22
— : Least squares line

Figure (4.3-b)


LOG(LOG(1./(1.-F(x))))
LOG(x-c)
Straight line plot of WEIBULL DIS. For
74/4TH QUARTER DATA
+ : c=20
— : Least squares line
A=136.4 B=0.794


LOG(LOG(1./(1.-F(x))))
LOG(x-c)
Straight line plot of WEIBULL DIS. For
75/1ST QUARTER DATA
+ : c=22
— : Least squares line
A=128.9 B=0.761


LOG(LOG(1./(1.-F(x))))
LOG(x-c)
Straight line plot of WEIBULL DIS. For
75/2ND QUARTER DATA
+ : c=21
— : Least squares line

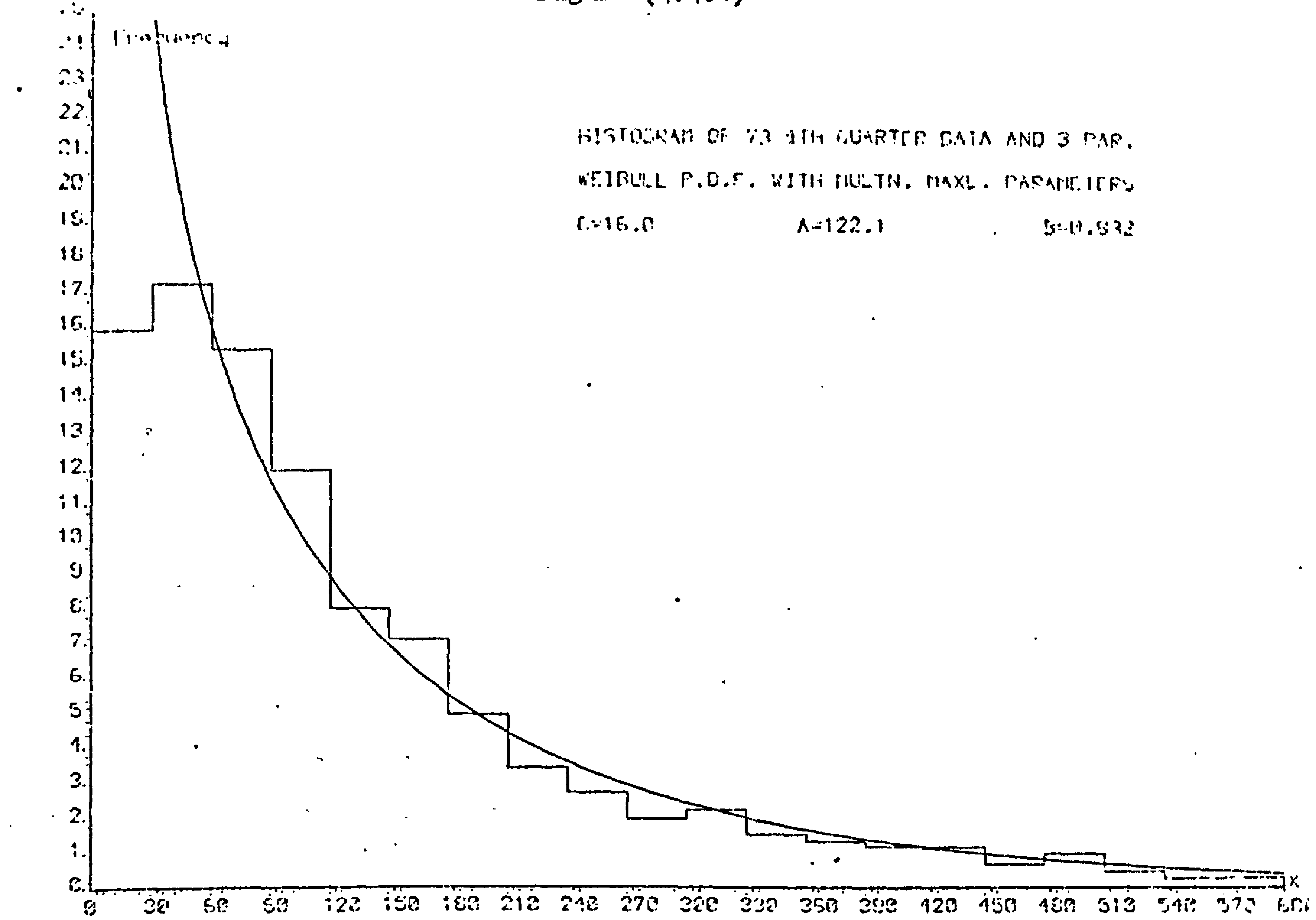
course, was expected from the graphical test which we carried out previously.

We next wrote computer program P16 to estimate the parameters of the 2-parameter Weibull distribution by multinomial maximum likelihood (MML) method in order to see how the results compare with those of the least squares method. For each of the accidental damage samples the least squares estimates of A and B were used as starting values for the iteration process. A summary of results is presented as part of table (4.1). We can see that the estimates of A are very nearly the same as their corresponding values by the least squares method but the values of $\hat{B}$ are slightly larger. The Chi-square values, however, are considerably smaller even when the slightly smaller degrees of freedom are allowed for. The required pooling of the intervals, in the upper tail of the distribution, for the calculation of the X^2 statistics were different from those of the least squares method. This resulted in the difference in degrees of freedom. The X^2 statistics, as before, show significant differences between the model and actual sample values. Based on the Chi-square goodness of fit test it is observed that the 2-parameter Weibull distribution does not provide a satisfactory model for the distribution of accidental damage claim amounts.

Computer program P17 was written to estimate the parameters of the 3-parameter Weibull distribution by MML method. For accidental damage samples, the least squares estimates of the parameters A, B and C, as given in table (4.2), were used as starting values for the iteration process. For each sample the program prints estimates of the parameters, the mean and the standard deviation of the fitted distribution of claim amounts and also an extensive table which, for each interval, provides the actual, the expected and the actual minus expected number of claims as well as the expected loss and the contribution to the Chi-square statistic (for those intervals where the expected number is greater than 5). The results for the accidental damage samples are presented in tables (4.3) to (4.9). The estimates of C are about 15 and the estimates of A are smaller than their corresponding values by the least squares method. The values of $\hat{A}$, the estimates of A, have generally increased over time. The X^2 statistics are smaller than their corresponding values by the least squares method. However, they still indicate significant differences between the model and the actual sample values. The major contribution to the Chi-square statistic is from few intervals only. Otherwise the agreement in most of the intervals is satisfactory. The total expected loss statistics are small and in every case are less than 1% of the total actual cost. The estimates of B are smaller than 1 for each of the samples. This means that the Weibull frequency curve resembles the exponential curve, but with a longer tail, and has no mode. Therefore, the model ignores the distinct mode of the histograms of the accidental damage data.

To see how closely the frequency curve of the 3-parameter Weibull model agrees with the histogram of the sample values we modified computer program P7 to plot them. The graphs are presented in figures (4.4.1) to (4.4.7). In each case the exponential shape of the curve, with no mode, is observed and the tail of the curve is seen to be shorter than the tail of the histogram. Large deviation between the curve and the histogram can be observed at several intervals. This could, of course, be expected by looking at the large values of actual

Figure (4.4.1)


Frequency
HISTOGRAM OF 73 4TH QUARTER DATA AND 3 PAR.
WEIBULL P.D.F. WITH MULTN. MAXL. PARAMETERS
C=16.0
A=122.1
B=0.932
X

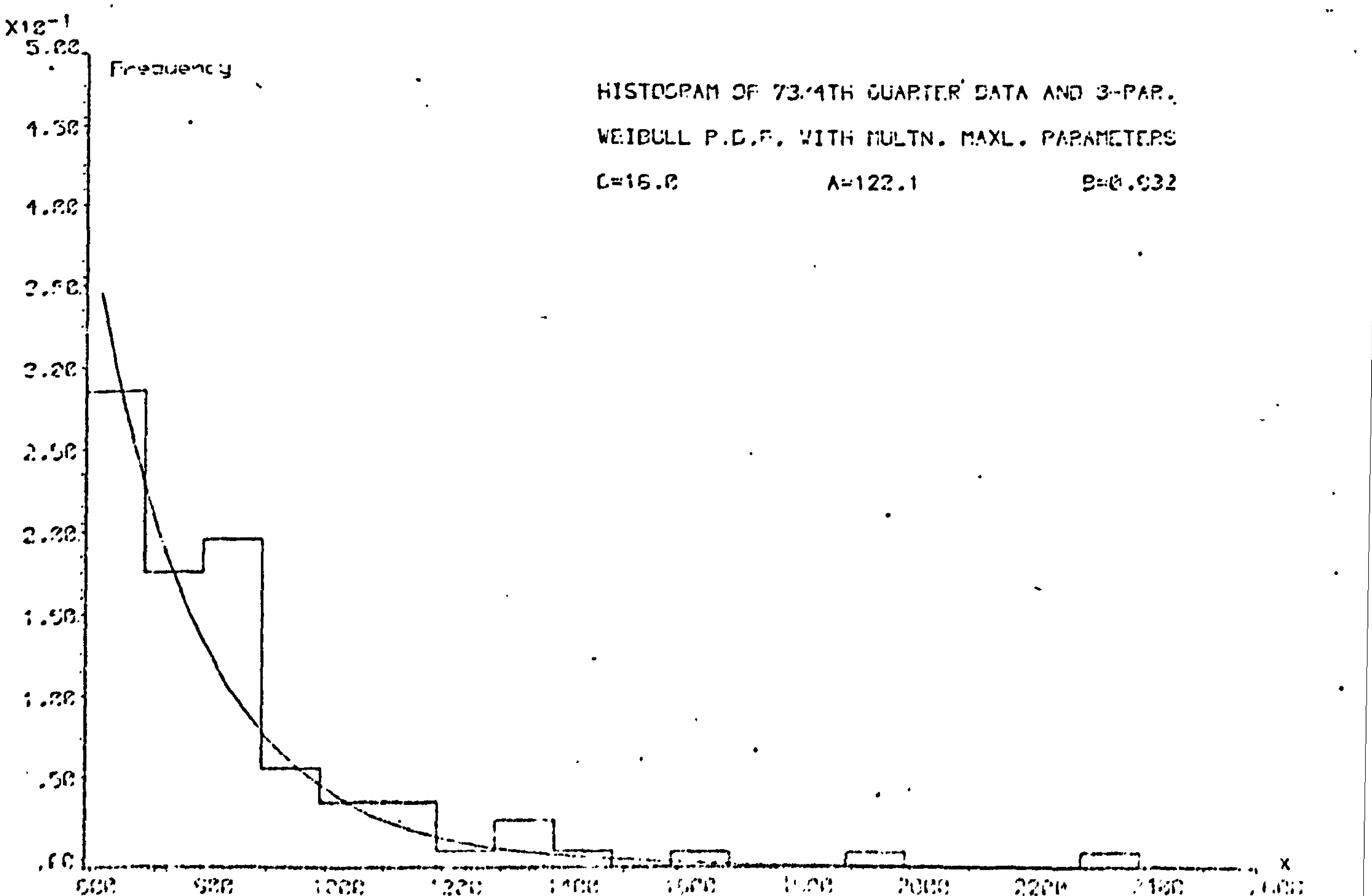
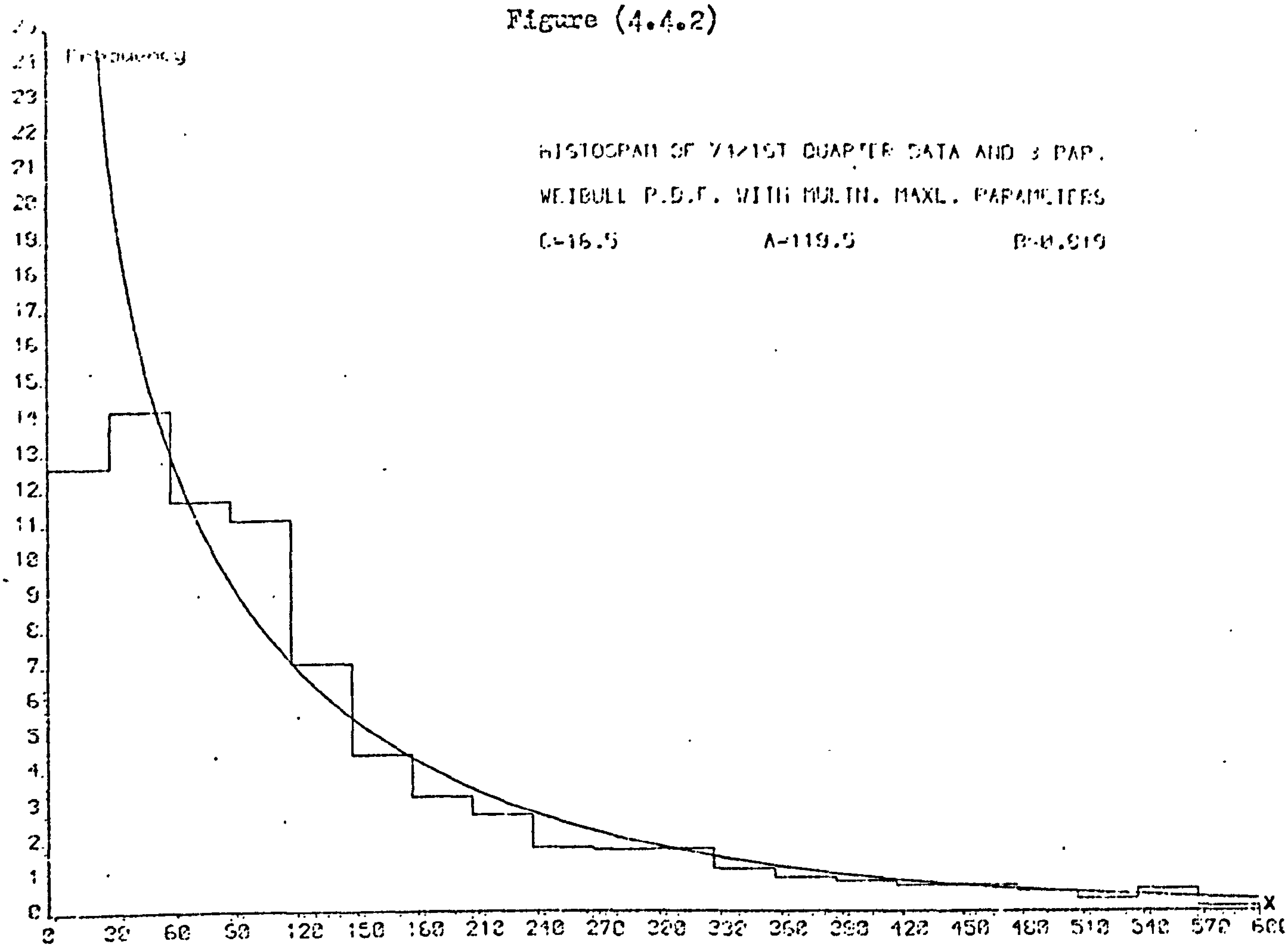

X10⁻¹
Frequency
HISTOGRAM OF 73/4TH QUARTER DATA AND 3-PAR.
WEIBULL P.D.F. WITH MULTN. MAXL. PARAMETERS
C=16.0
A=122.1
B=0.932
X

Figure (4.4.2)


Frequency
HISTOGRAM OF 71/1ST QUARTER DATA AND 3 PAR.
WEIBULL P.D.F. WITH MULTN. MAXL. PARAMETERS
C=16.5
A=119.5
B=0.919
0 30 60 90 120 150 180 210 240 270 300 330 360 390 420 450 480 510 540 570 600
X

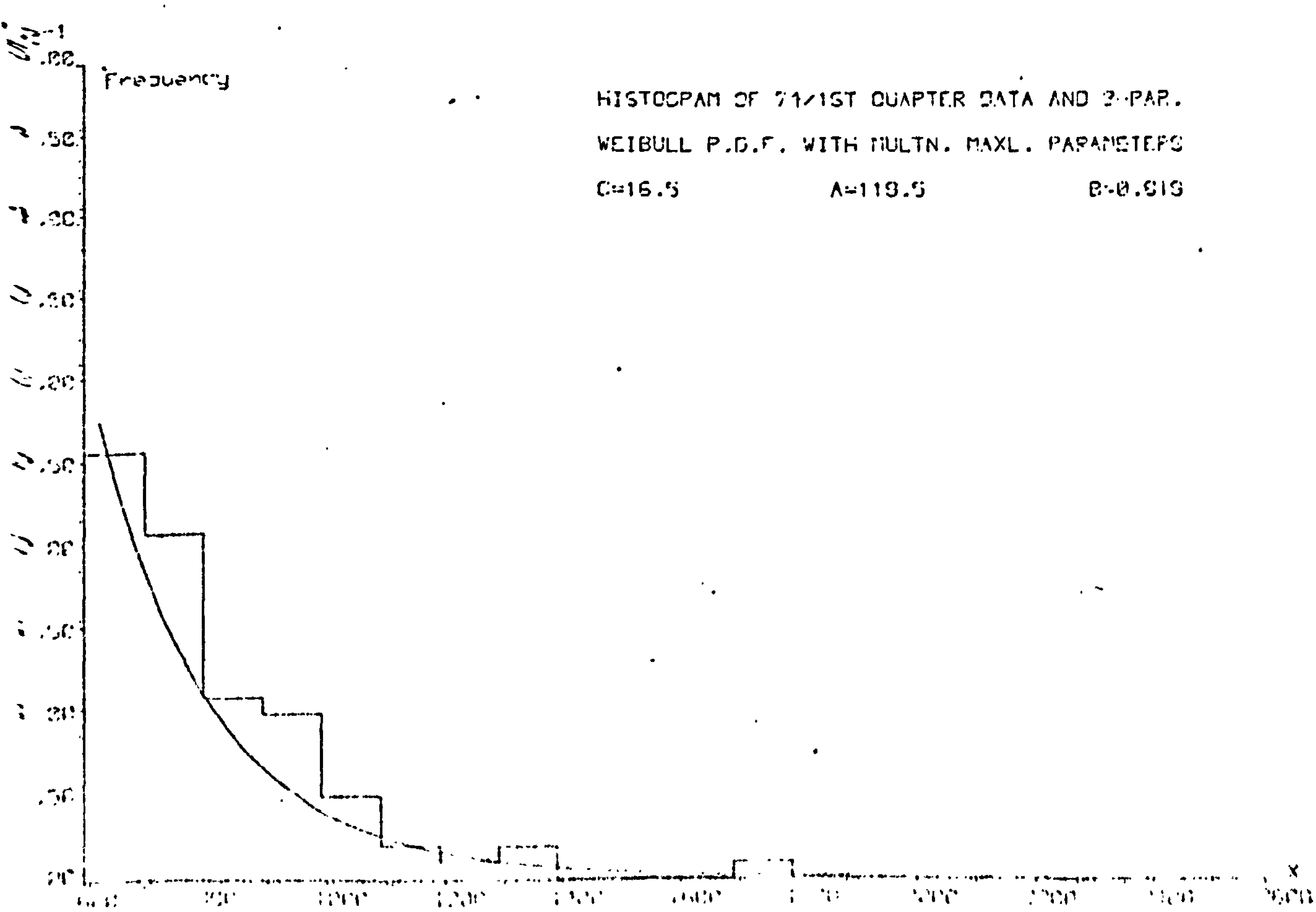
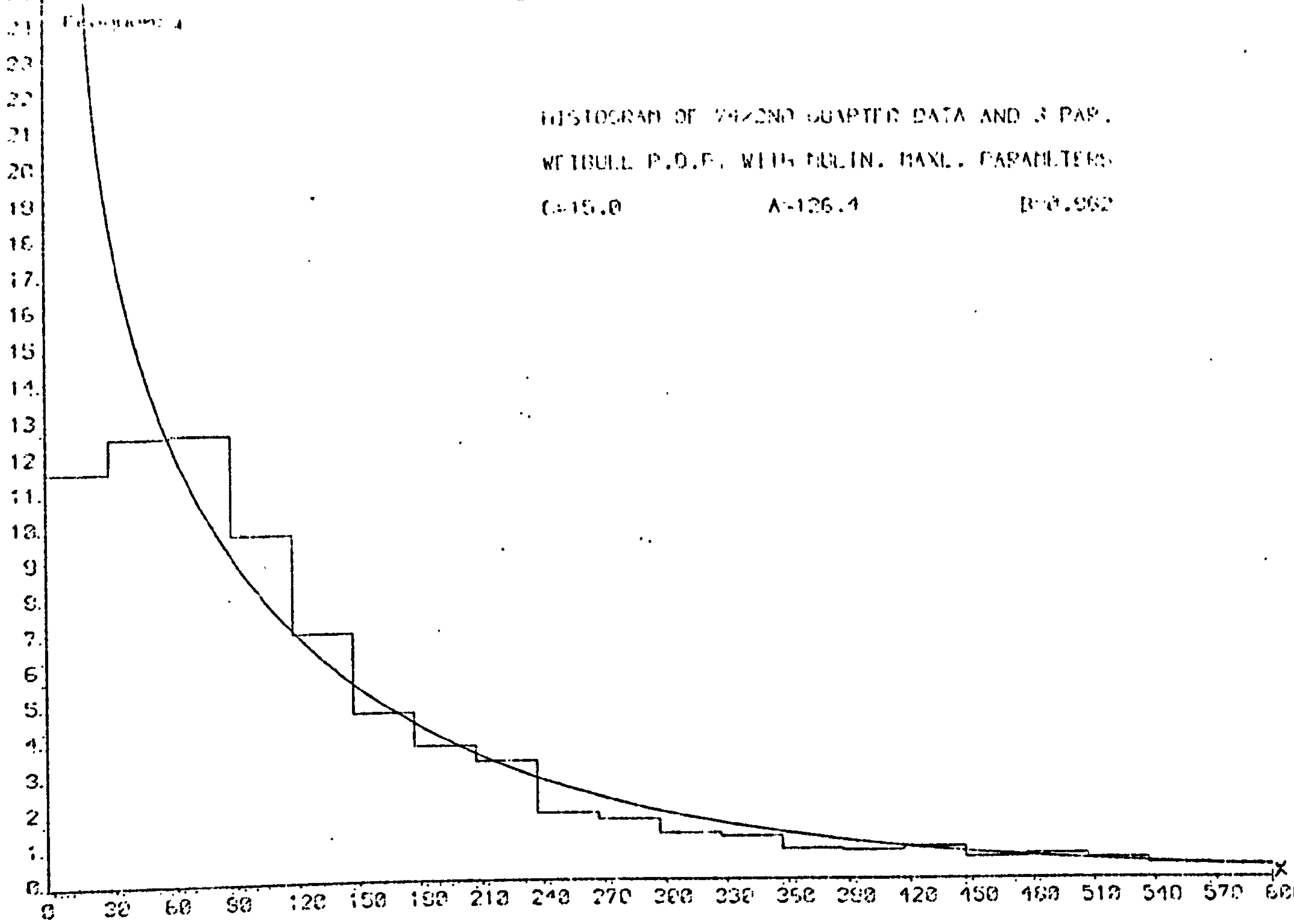

Frequency
HISTOGRAM OF 71/1ST QUARTER DATA AND 3-PAR.
WEIBULL P.D.F. WITH MULTN. MAXL. PARAMETERS
C=16.5
A=119.5
B=0.919
X

Figure (4.4.3)


Frequency
HISTOGRAM OF 74/2ND QUARTER DATA AND 3 PAR.
WEIBULL P.D.F. WITH MULTN. MAXL. PARAMETERS
C=15.0
A=126.4
B=0.862
0 30 60 90 120 150 180 210 240 270 300 330 360 390 420 450 480 510 540 570 600
X

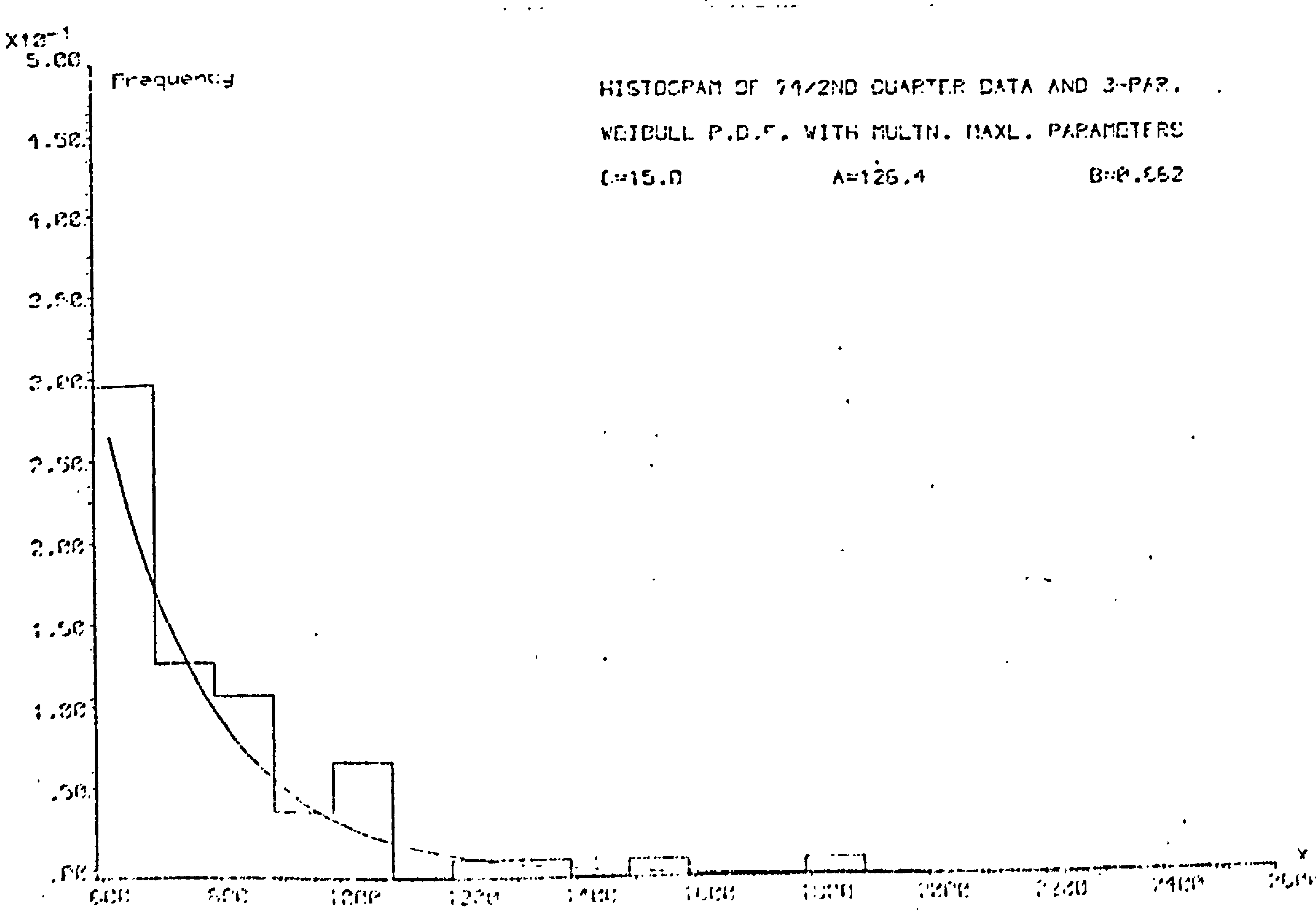
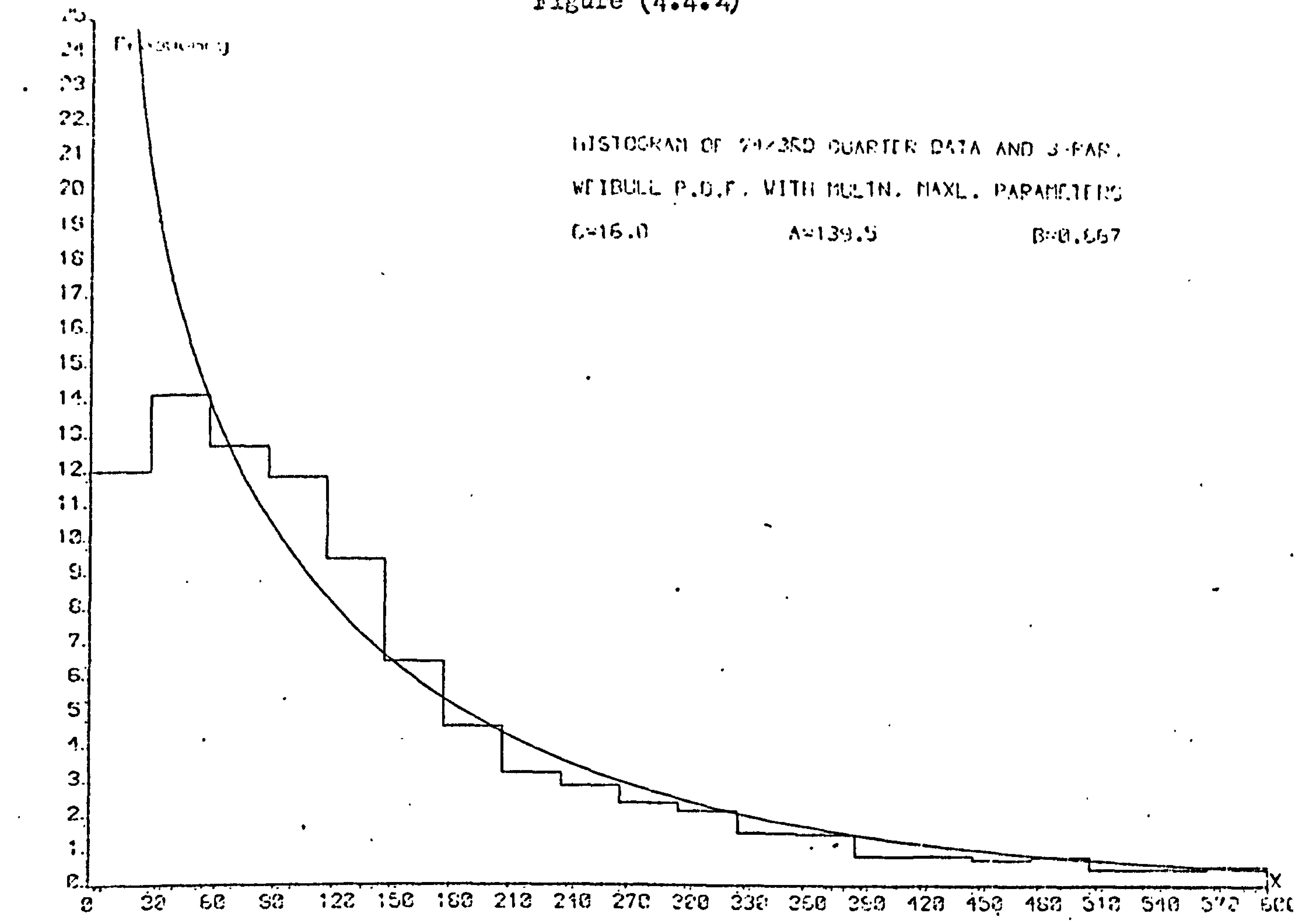

X10^-1
5.00
Frequency
4.50
4.00
3.50
3.00
2.50
2.00
1.50
1.00
.50
.00
HISTOGRAM OF 74/2ND QUARTER DATA AND 3-PAR.
WEIBULL P.D.F. WITH MULTN. MAXL. PARAMETERS
C=15.0
A=126.4
B=0.862
X

Figure (4.4.4)


Frequency
HISTOGRAM OF 74/3RD QUARTER DATA AND 3-PAR.
WEIBULL P.D.F. WITH MULTN. MAXL. PARAMETERS
C=16.0
A=139.5
B=0.667
X

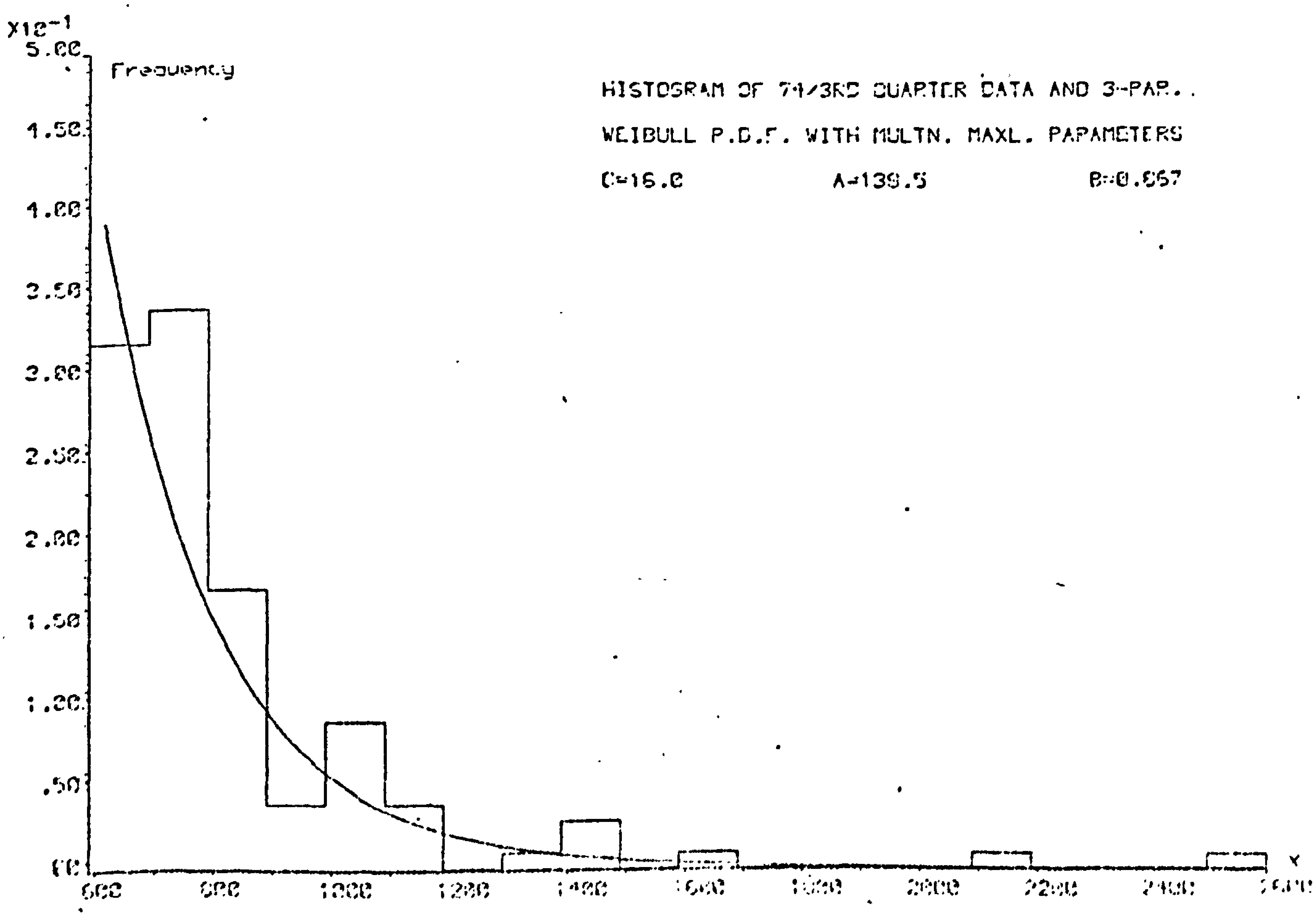
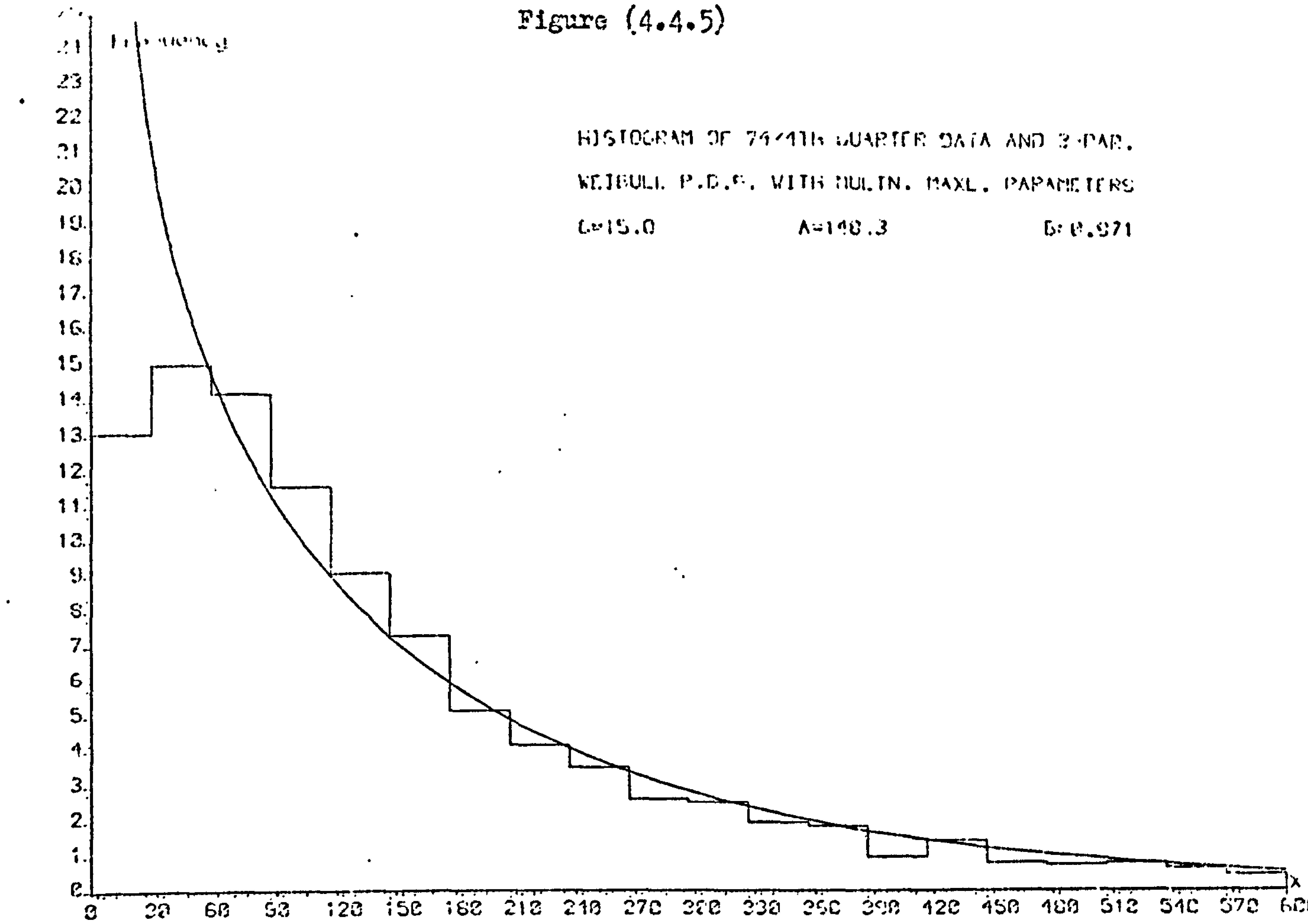

X10^-1
Frequency
HISTOGRAM OF 74/3RD QUARTER DATA AND 3-PAR.
WEIBULL P.D.F. WITH MULTN. MAXL. PARAMETERS
C=16.0
A=139.5
B=0.667
X

Figure (4.4.5)


Frequency
HISTOGRAM OF 74/4TH QUARTER DATA AND 3-PAR.
WEIBULL P.D.F. WITH MULTN. MAXL. PARAMETERS
C=15.0
A=148.3
B=0.971
X

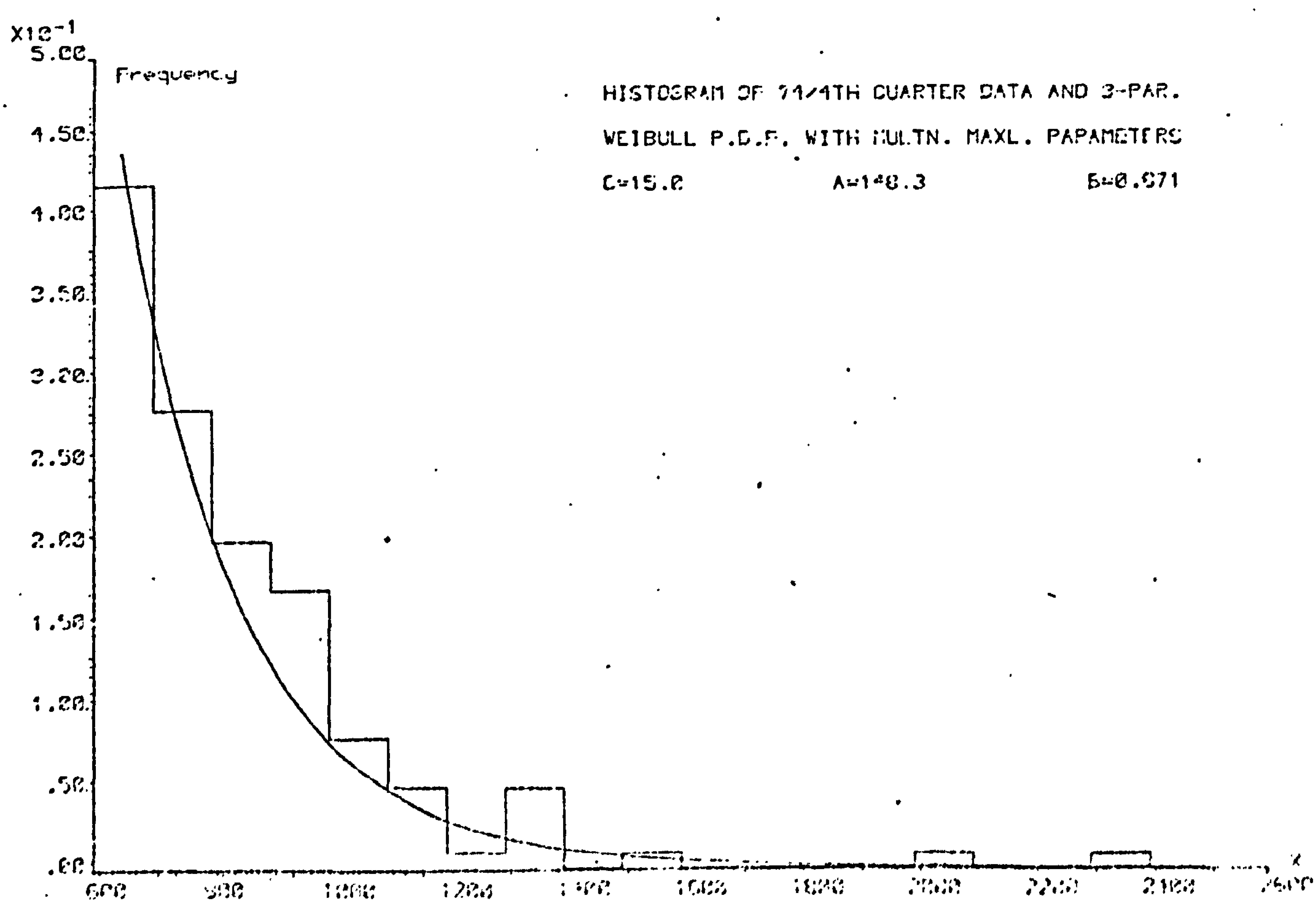
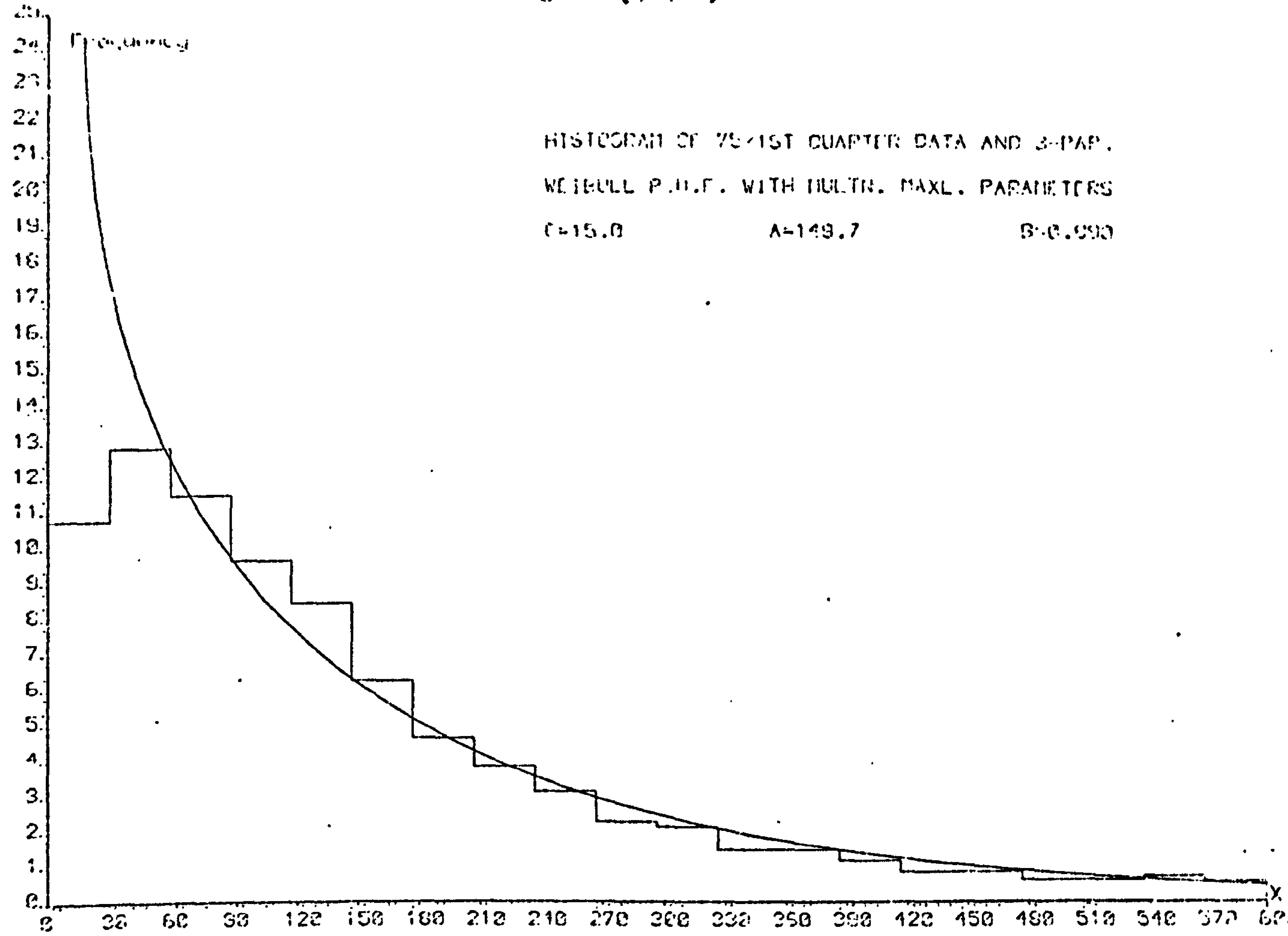

X10⁻¹
Frequency
HISTOGRAM OF 74/4TH QUARTER DATA AND 3-PAR.
WEIBULL P.D.F. WITH MULTN. MAXL. PARAMETERS
C=15.0
A=148.3
B=0.971
X

Figure (4.4.6)


Frequency
HISTOGRAM OF 75/1ST QUARTER DATA AND 3-PAR.
WEIBULL P.D.F. WITH MULTN. MAXL. PARAMETERS
C=15.0
A=143.7
X

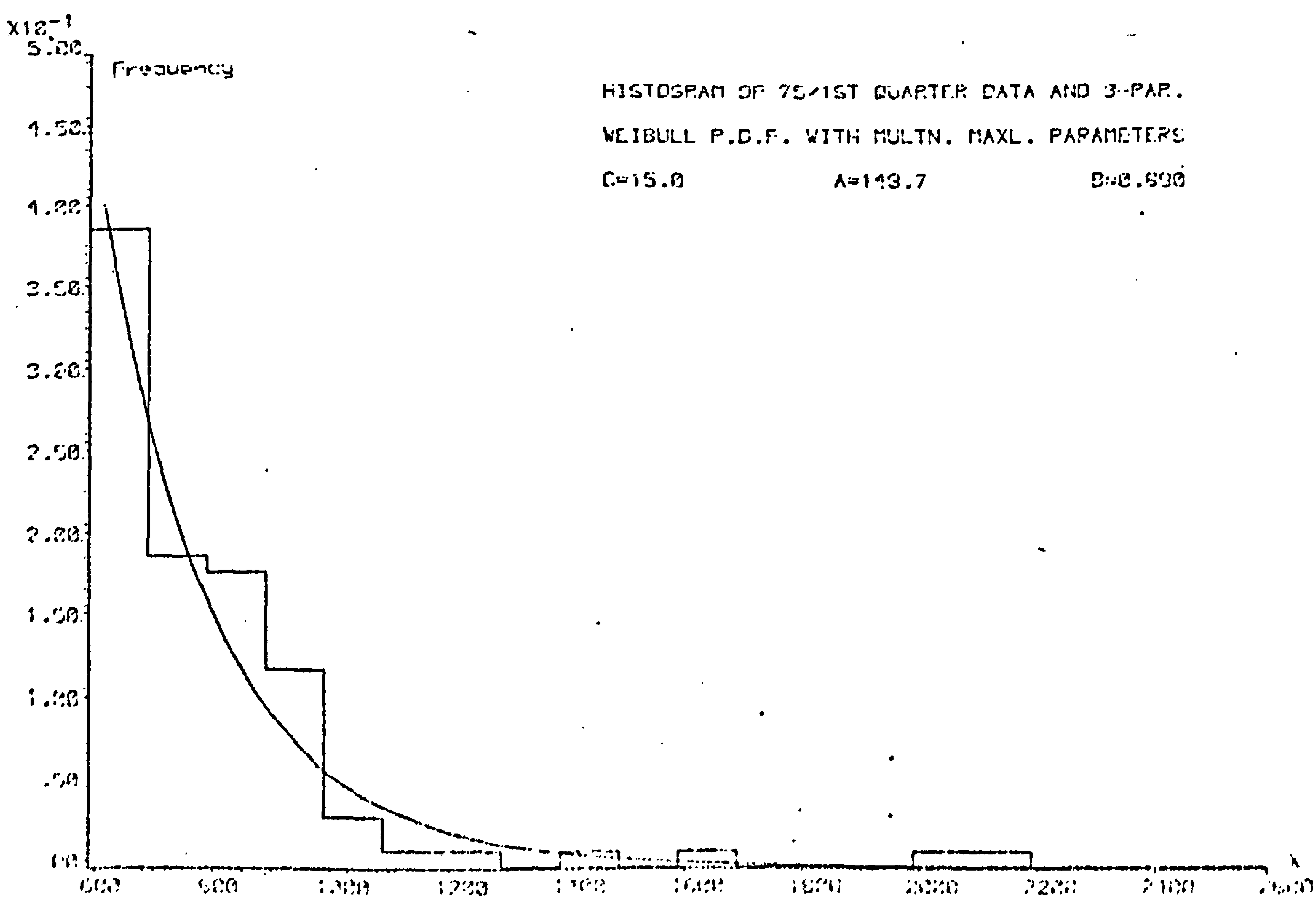
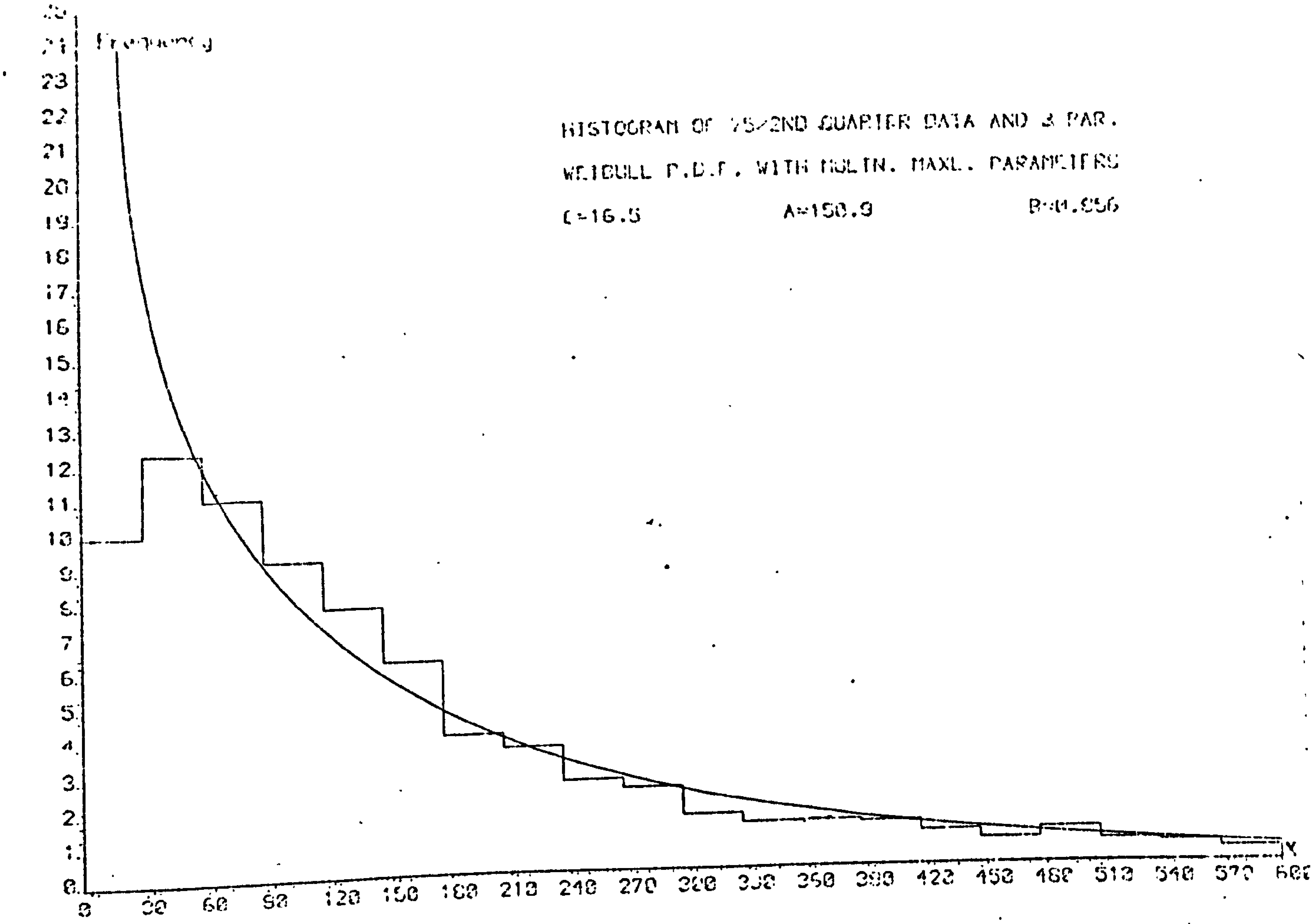
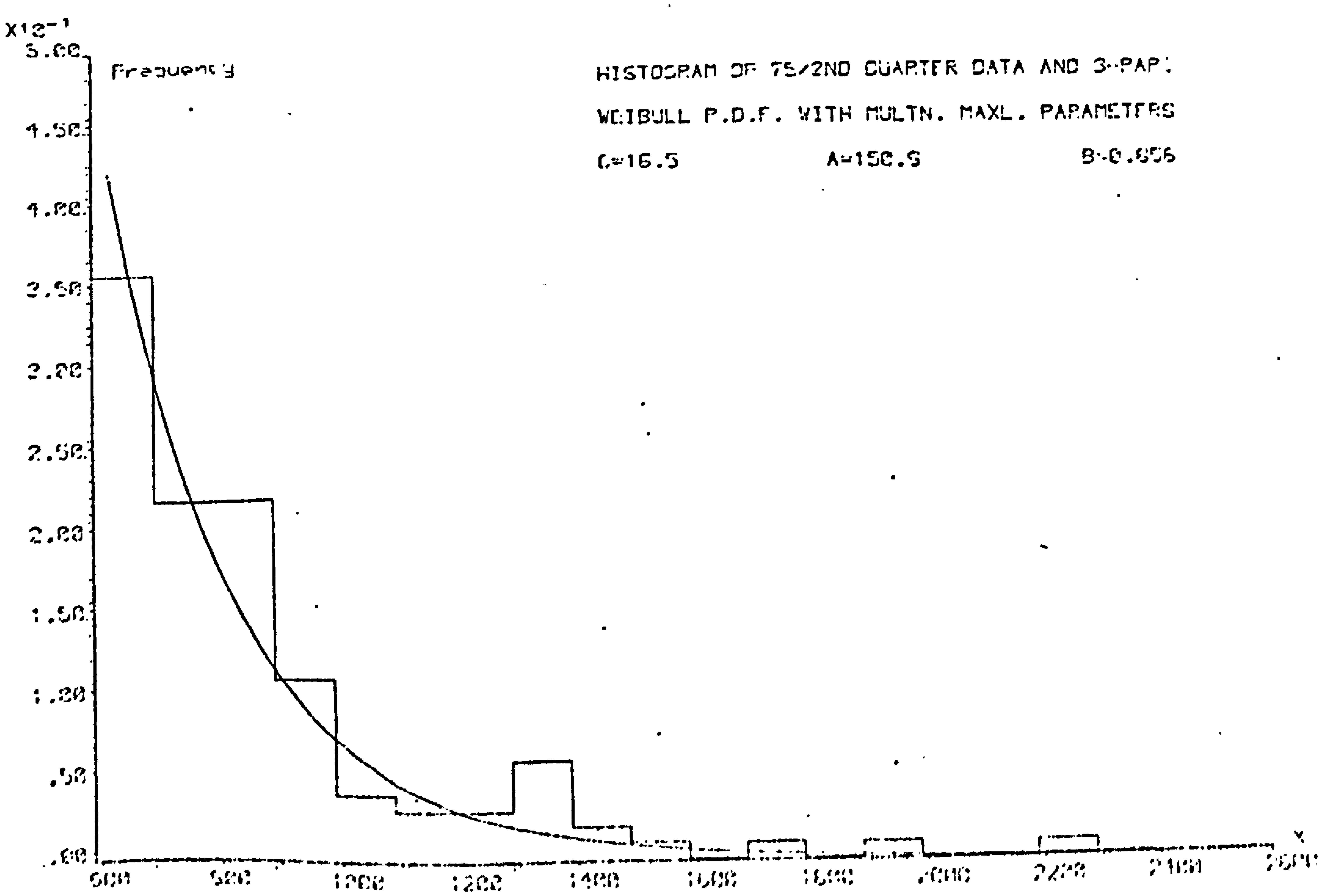

x10^-1
Frequency
HISTOGRAM OF 75/1ST QUARTER DATA AND 3-PAR.
WEIBULL P.D.F. WITH MULTN. MAXL. PARAMETERS
C=15.0
A=143.7

Figure (4.4.7)


Frequency
HISTOGRAM OF 75/2ND QUARTER DATA AND 3 PAR.
WEIBULL P.D.F. WITH MULTN. MAXL. PARAMETERS
C=16.5
A=150.9
0
30
60
90
120
150
180
210
240
270
300
360
390
420
450
480
510
540
570
600


Frequency
HISTOGRAM OF 75/2ND QUARTER DATA AND 3-PAR.
WEIBULL P.D.F. WITH MULTN. MAXL. PARAMETERS
C=16.5
A=150.9
B=0.856
5.00
4.50
4.00
3.50
3.00
2.50
2.00
1.50
1.00
.50
.00
1400
2200

minus expected frequencies in the tables. It can be noticed that the frequency curve fits the sample histogram rather well in the middle but not at the tail values. The agreement between the frequency curve of the model and the histogram of the data is not satisfactory for any of the samples.

From the above analysis we can conclude that the Weibull distribution does not provide an adequate model for the distribution of the accidental damage claim amounts.

4.8 The Effects of Inflation on the Parameters of the Weibull Model

Following the argument of chapter 3 about the effects of inflation on the parameters of a model, it is considered important to study these effects theoretically for the Weibull distribution. We will need to allow for such effects when predicting the future distribution of claim amounts for a class of general insurance business in which the Weibull model has been found to represent the distribution of claim amounts.

Let us again assume that the effect of inflation is to increase a claim of amount X to X(1 + i) over a period where i is the effective rate of inflation for that period. If X is distributed as a 2-parameter Weibull with distribution function W(x; A,B) then by a transformation of variables we can show that Z = X(1 + i) is distributed as a Weibull W(z; Ag,B) where g = 1 + i.
If X has a 3-parameter Weibull distribution, W(x; C,A,B), then again by a transformation we can show that Z = X(1 + i) is distributed as W(z; Cg, Ag, B). Therefore inflation increases the parameters C and A but leaves the shape parameter B unchanged. The mean and the standard deviation of the Weibull distribution are functions of all three parameters and hence both increase over time due to inflation. Although it has been found that the Weibull model is not completely adequate for the distribution of accidental damage claim amounts, nevertheless the increasing trends of C and A can be observed from the tables of results.

Since the Weibull is not a very satisfactory model for our data we do not use it to predict the distribution of claim amounts during future periods.

4.9 Conclusions

The 2 and 3 parameter Weibull distributions were studied in this chapter as models for the distribution of claim amounts in general insurance and in particular for the accidental damage claim amounts. The Weibull distribution has been regarded as a more flexible model than the exponential distribution and hence its use in various classes of general insurance, where the latter has been successfully applied, is recommended. An interpretation was offered for parameter C as the amount below which claims were not made. The multinomial maximum likelihood method is preferred to the least squares method for estimating the parameters of the Weibull distribution since the former method produced smaller Chi-square statistics and hence a better agreement between the model and actual sample values. The 3-parameter distribution represents our accidental damage data better than the 2-parameter distribution when judged by the Chi-square goodness-of-fit test. However, even the 3-parameter model does not satisfactorily represent the distribution of accidental damage claim amounts. This was indicated by the significant differences between the model and the actual sample values as judged by the values of X^2. The value of the shape parameter B was found to be less than 1 which indicates that the model has an exponential-like shape with no mode. Hence the model ignored the distinct mode which is observed in the histograms of our sample values. It was observed that for each sample the frequency curve of the model had a shorter tail than the histogram of the data.

The inflation was shown to affect the location and scale parameters, C and A respectively, but to leave the shape parameter, B, unchanged. The effect of inflation in increasing C over time is plausible as it is expected that with inflation the amount below which claims are not made should be increased.

We may note down in conclusion that the 3-parameter Weibull model was not found to be as adequate as the 3-parameter lognormal model in representing the distribution of our accidental damage claim amounts.

4.10 Tables

Table (4.1)

Two-parameter Weibull model fitted to accidental damage data by two methods of estimation

	Least Squares Method			Mult. Max. Likelihood Method		
Period of accident	$\hat{A}$	$\hat{B}$	X^2 (D.F.)	$\hat{A}$	$\hat{B}$	X^2 (D.F.)
73/4th quarter	143.3	0.853	224.4(24)	147.4	0.969	136.8 (21)
74/1st "	146.5	0.899	182.0(22)	146.6	0.966	148.0 (21)
74/2nd "	150.6	0.914	157.1(22)	150.7	0.998	115.8 (21)
74/3rd "	163.6	0.886	235.6(24)	165.5	1.000	160.6 (22)
74/4th "	172.5	0.911	162.8(24)	173.0	0.994	122.4 (22)
75/1st "	169.5	0.878	181.3(24)	176.5	1.040	87.5 (22)
75/2nd "	178.6	0.900	166.3(24)	178.4	0.988	118.4 (22)

Table (4.2)

Three-parameter Weibull model fitted to accidental damage data by least squares method

Period of accident	C	$\hat{A}$	$\hat{B}$	X^2 (D.F.)	R%
73/4th quarter	21	108.4	0.740	100.5 (23)	-0.2
74/1st "	19	114.4	0.780	83.3 (22)	-0.2
74/2nd "	19	118.4	0.796	75.1 (22)	-0.8
74/3rd "	22	123.9	0.760	137.0 (24)	-0.6
74/4th "	20	136.4	0.794	75.3 (24)	-0.2
75/1st "	22	129.9	0.761	102.0 (24)	-0.2
75/2nd "	21	139.5	0.782	74.2 (24)	-0.3

N.B. - X^2 is the Chi-square statistic with (D.F.) degrees of freedom. R is the ratio of total expected loss statistic, T , to the total actual cost.

Table (4.3.1)

*** 3-PARAMETER WEIBULL DIS. ***

73/4TH QUARTER DATA

ESTIMATION BY MULTINOMIAL MAX. LIKELIHOOD METHOD :-
C= 15.73 A= 122.088 B= 0.832

MEAN= 150.349 S.D.= 162.617

↑

AMOUNT £	ACTUAL CL. NO.	EXPECTED CL. NO.	ACTUAL-EXPECTED	EXPECTED LOSS	(A-E)**2/E
1- 30	478.	482.	-4.	-62.	0.033
31- 60	518.	590.	-72.	-3276.	8.786
61- 90	461.	407.	54.	4077.	7.165
91- 120	359.	304.	55.	5802.	9.951
121- 150	239.	234.	5.	678.	0.107
151- 180	213.	184.	29.	4799.	4.571
181- 210	148.	147.	1.	195.	0.007
211- 240	102.	119.	-17.	-3833.	2.429
241- 270	81.	96.	-15.	-3832.	2.344
271- 300	58.	79.	-21.	-5995.	5.582
301- 330	66.	65.	1.	316.	0.015
331- 360	45.	54.	-9.	-3109.	1.500
361- 390	39.	45.	-6.	-2253.	0.800
391- 420	35.	37.	-2.	-811.	0.108
421- 450	34.	31.	3.	1306.	0.290
451- 480	20.	26.	-6.	-2793.	1.385
481- 510	29.	22.	7.	3468.	2.227
511- 540	14.	18.	-4.	-2102.	0.889
541- 570	8.	15.	-7.	-3888.	3.267
571- 600	9.	13.	-4.	-2342.	1.231
601- 700	29.	31.	-2.	-1301.	0.129
701- 800	18.	18.	0.	0.	0.000
801- 900	20.	11.	9.	7654.	7.364
901-1000	6.	6.	0.	0.	0.000
1001-1100	4.	4.	0.	0.	
1101-1200	4.	2.	2.	2301.	0.000
1201-1300	1.	1.	0.	0.	
1301-1400	3.	1.	2.	2701.	
1401-1500	1.	1.	0.	0.	
1501-1600	0.	0.	0.	0.	
1601-1700	1.	0.	1.	1650.	
1701-1800	0.	0.	0.	0.	
1801-1900	0.	0.	0.	0.	
1901-2000	1.	0.	1.	1950.	
2001-2100	0.	0.	0.	0.	
2101-2200	0.	0.	0.	0.	
2201-2300	0.	0.	0.	0.	
2301-2400	1.	0.	1.	2350.	9.800
TOTAL	3045.	3043.		3651.	69.879

D.F. = 22

$P < 0.001$

TOTAL EXP. LOSS
--------------- = 0.8 %
TOTAL ACT. COST

Table (4.3.2)

*** 3-PARAMETER WEIBULL DIS. ***

74/1ST QUARTER DATA

ESTIMATION BY MULTINOMIAL MAX. LIKELIHOOD METHOD :-
C= 16.64 A= 119.529 B= 0.819

MEAN= 149.887 S.D.= 163.820

AMOUNT £	ACTUAL CL. NO.	EXPECTED CL. NO.	ACTUAL-EXPECTED	EXPECTED LOSS	(A-E)**2/E
1- 30	381.	384.	-3.	-46.	0.023
31- 60	428.	485.	-57.	-2593.	6.699
61- 90	351.	328.	23.	1736.	1.613
91- 120	334.	243.	91.	9600.	34.078
121- 150	211.	186.	25.	3387.	3.360
151- 180	133.	146.	-13.	-2151.	1.158
181- 210	98.	116.	-18.	-3519.	2.793
211- 240	82.	94.	-12.	-2706.	1.532
241- 270	54.	76.	-22.	-5621.	6.368
271- 300	52.	62.	-10.	-2855.	1.613
301- 330	53.	51.	2.	631.	0.078
331- 360	36.	42.	-6.	-2073.	0.857
361- 390	29.	35.	-6.	-2253.	1.029
391- 420	26.	29.	-3.	-1216.	0.310
421- 450	22.	25.	-3.	-1306.	0.360
451- 480	22.	21.	1.	466.	0.048
481- 510	17.	17.	0.	0.	0.000
511- 540	10.	15.	-5.	-2627.	1.667
541- 570	19.	12.	7.	3888.	4.083
571- 600	4.	10.	-6.	-3513.	3.600
601- 700	26.	25.	1.	651.	0.040
701- 800	21.	15.	6.	4503.	2.400
801- 900	11.	9.	2.	1701.	0.444
901-1000	10.	5.	5.	4752.	5.000
1001-1100	5.	3.	2.	2101.	
1101-1200	2.	2.	0.	0.	
1201-1300	1.	1.	0.	0.	
1301-1400	2.	1.	1.	1350.	
1401-1500	0.	1.	-1.	-1450.	
1501-1600	0.	0.	0.	0.	
1601-1700	0.	0.	0.	0.	
1701-1800	1.	0.	1.	1751.	1.125
TOTAL	2441.	2439.		2586.	80.279

D.F. =21

TOTAL EXP. LOSS / TOTAL ACT. COST = 0.7 %

P <0.001

Table (4.3.3)

*** 3-PARAMETER WEIBULL DIS. ***

74/2ND QUARTER DATA

ESTIMATION BY MULTINOMIAL MAX. LIKELIHOOD METHOD :-
C= 15.32 A= 126.376 B= 0.862

MEAN= 151.603 S.D.= 158.593

AMOUNT £	ACTUAL CL. NO.	EXPECTED CL. NO.	ACTUAL-EXPECTED	EXPECTED LOSS	(A-E)**2/E
1- 30	351.	354.	-3.	-46.	0.025
31- 60	380.	451.	-71.	-3230.	11.177
61- 90	382.	321.	61.	4605.	11.592
91- 120	295.	243.	52.	5486.	11.128
121- 150	211.	189.	22.	2981.	2.561
151- 180	142.	150.	-8.	-1324.	0.427
181- 210	114.	120.	-6.	-1173.	0.300
211- 240	101.	97.	4.	902.	0.165
241- 270	57.	79.	-22.	-5621.	6.127
271- 300	51.	64.	-13.	-3711.	2.641
301- 330	39.	53.	-14.	-4417.	3.698
331- 360	36.	43.	-7.	-2418.	1.140
361- 390	25.	36.	-11.	-4130.	3.361
391- 420	24.	30.	-6.	-2433.	1.200
421- 450	27.	25.	2.	871.	0.160
451- 480	18.	21.	-3.	-1396.	0.429
481- 510	21.	17.	4.	1982.	0.941
511- 540	17.	14.	3.	1576.	0.643
541- 570	12.	12.	0.	0.	0.000
571- 600	11.	10.	1.	585.	0.100
601- 700	30.	24.	6.	3903.	1.500
701- 800	13.	13.	0.	0.	0.000
801- 900	11.	8.	3.	2551.	1.125
901-1000	4.	5.	-1.	-951.	0.200
1001-1100	7.	3.	4.	4202.	
1101-1200	0.	2.	-2.	-2301.	
1201-1300	1.	1.	0.	0.	
1301-1400	1.	1.	0.	0.	
1401-1500	0.	0.	0.	0.	
1501-1600	1.	0.	1.	1550.	
1601-1700	0.	0.	0.	0.	
1701-1800	0.	0.	0.	0.	
1801-1900	1.	0.	1.	1850.	2.286
TOTAL	2383.	2386.		-106.	63.124

D.F. =21

P<0.001

TOTAL EXP. LOSS
---------------- = -0.0 %
TOTAL ACT. COST

Table (4.3.4)

*** 3-PARAMETER WEIBULL DIS. ***

74/3RD QUARTER DATA

ESTIMATION BY MULTINOMIAL MAX. LIKELIHOOD METHOD :-
C= 16.07 A= 139.472 B= 0.867

MEAN= 166.012 S.D.= 173.566

AMOUNT £	ACTUAL CL. NO.	EXPECTED CL. NO.	ACTUAL-EXPECTED	EXPECTED LOSS	(A-E)**2/E
1- 30	362.	366.	-4.	-62.	0.044
31- 60	427.	502.	-75.	-3412.	11.205
61- 90	383.	365.	18.	1359.	0.888
91- 120	356.	281.	75.	7912.	20.018
121- 150	283.	223.	60.	8130.	16.143
151- 180	194.	179.	15.	2482.	1.257
181- 210	137.	146.	-9.	-1759.	0.555
211- 240	97.	119.	-22.	-4961.	4.067
241- 270	86.	98.	-12.	-3066.	1.469
271- 300	71.	82.	-11.	-3140.	1.476
301- 330	64.	68.	-4.	-1262.	0.235
331- 360	45.	57.	-12.	-4146.	2.526
361- 390	44.	48.	-4.	-1502.	0.333
391- 420	25.	40.	-15.	-6082.	5.625
421- 450	26.	34.	-8.	-3484.	1.882
451- 480	22.	28.	-6.	-2793.	1.286
481- 510	25.	24.	1.	496.	0.042
511- 540	14.	20.	-6.	-3153.	1.800
541- 570	14.	17.	-3.	-1666.	0.529
571- 600	17.	15.	2.	1171.	0.267
601- 700	32.	35.	-3.	-1951.	0.257
701- 800	34.	21.	13.	9756.	8.048
801- 900	17.	12.	5.	4252.	2.083
901-1000	4.	8.	-4.	-3802.	2.000
1001-1100	9.	5.	4.	4202.	3.200
1101-1200	4.	3.	1.	1150.	
1201-1300	0.	2.	-2.	-2501.	
1301-1400	1.	1.	0.	0.	
1401-1500	3.	1.	2.	2901.	
1501-1600	0.	0.	0.	0.	
1601-1700	1.	0.	1.	1650.	
1701-1800	0.	0.	0.	0.	
1801-1900	0.	0.	0.	0.	
1901-2000	0.	0.	0.	0.	
2001-2100	0.	0.	0.	0.	
2101-2200	1.	0.	1.	2150.	
2201-2300	0.	0.	0.	0.	
2301-2400	0.	0.	0.	0.	
2401-2500	0.	0.	0.	0.	
2501-2600	1.	0.	1.	2550.	2.286
TOTAL	2799.	2800.		1419.	92.722

D.F. = 22

$P < 0.001$

TOTAL EXP. LOSS / TOTAL ACT. COST = 0.3 %

Table (4.3.5)

*** 3-PARAMETER WEIBULL DIS. ***

74/4TH QUARTER DATA

ESTIMATION BY MULTINOMIAL MAX. LIKELIHOOD METHOD :-

C= 15.15 A= 148.275 B= 0.871

MEAN= 174.111 S.D.= 183.102

AMOUNT £	ACTUAL CL. NO.	EXPECTED CL. NO.	ACTUAL-EXPECTED	EXPECTED LOSS	(A-E)**2/E
1- 30	394.	397.	-3.	-46.	0.023
31- 60	452.	522.	-70.	-3185.	9.387
61- 90	426.	386.	40.	3020.	4.145
91- 120	348.	302.	46.	4853.	7.007
121- 150	272.	241.	31.	4200.	3.988
151- 180	219.	196.	23.	3806.	2.699
181- 210	154.	161.	-7.	-1368.	0.304
211- 240	124.	133.	-9.	-2029.	0.609
241- 270	105.	111.	-6.	-1533.	0.324
271- 300	78.	93.	-15.	-4282.	2.419
301- 330	75.	78.	-3.	-947.	0.115
331- 360	58.	65.	-7.	-2418.	0.754
361- 390	55.	55.	0.	0.	0.000
391- 420	29.	47.	-18.	-7299.	6.894
421- 450	43.	40.	3.	1306.	0.225
451- 480	24.	34.	-10.	-4655.	2.941
481- 510	22.	29.	-7.	-3468.	1.690
511- 540	24.	25.	-1.	-525.	0.040
541- 570	19.	21.	-2.	-1111.	0.190
571- 600	14.	18.	-4.	-2342.	0.889
601- 700	42.	43.	-1.	-651.	0.023
701- 800	28.	26.	2.	1501.	0.154
801- 900	20.	16.	4.	3402.	1.000
901-1000	17.	10.	7.	6653.	4.900
1001-1100	8.	6.	2.	2101.	0.667
1101-1200	5.	4.	1.	1150.	0.250
1201-1300	1.	2.	-1.	-1250.	
1301-1400	5.	2.	3.	4051.	
1401-1500	0.	1.	-1.	-1450.	
1501-1600	1.	1.	0.	0.	
1601-1700	0.	0.	0.	0.	
1701-1800	0.	0.	0.	0.	
1801-1900	0.	0.	0.	0.	
1901-2000	0.	0.	0.	0.	
2001-2100	1.	0.	1.	2050.	
2101-2200	0.	0.	0.	0.	
2201-2300	0.	0.	0.	0.	
2301-2400	1.	0.	1.	2350.	1.500
TOTAL	3064.	3065.		1884.	53.137

D.F. = 23

TOTAL EXP. LOSS / TOTAL ACT. COST = 0.4 %

$P < 0.001$

Table (4.3.6)

75/1ST QUARTER DATA

ESTIMATION BY MULTINOMIAL MAX. LIKELIHOOD METHOD :-

C= 14.89 A= 149.659 B= 0.890

MEAN= 173.325 S.D.= 178.362

AMOUNT £	ACTUAL CL. NO.	EXPECTED CL. NO.	ACTUAL-EXPECTED	EXPECTED LOSS	(A-E)**2/E
1- 30	324.	326.	-2.	-31.	0.012
31- 60	387.	439.	-52.	-2366.	6.159
61- 90	345.	330.	15.	1132.	0.682
91- 120	289.	260.	29.	3059.	3.235
121- 150	253.	209.	44.	5962.	9.263
151- 180	187.	170.	17.	2813.	1.700
181- 210	138.	140.	-2.	-391.	0.029
211- 240	114.	116.	-2.	-451.	0.034
241- 270	93.	96.	-3.	-767.	0.094
271- 300	67.	80.	-13.	-3711.	2.112
301- 330	63.	67.	-4.	-1262.	0.239
331- 360	44.	56.	-12.	-4146.	2.571
361- 390	44.	47.	-3.	-1126.	0.191
391- 420	35.	40.	-5.	-2027.	0.625
421- 450	25.	34.	-9.	-3919.	2.382
451- 480	26.	29.	-3.	-1396.	0.310
481- 510	18.	24.	-6.	-2973.	1.500
511- 540	18.	21.	-3.	-1576.	0.429
541- 570	22.	18.	4.	2222.	0.889
571- 600	17.	15.	2.	1171.	0.267
601- 700	39.	36.	3.	1951.	0.250
701- 800	19.	21.	-2.	-1501.	0.190
801- 900	18.	13.	5.	4252.	1.923
901-1000	12.	8.	4.	3802.	2.000
1001-1100	3.	5.	-2.	-2101.	0.800
1101-1200	1.	3.	-2.	-2301.	
1201-1300	1.	2.	-1.	-1250.	
1301-1400	0.	1.	-1.	-1350.	
1401-1500	1.	1.	0.	0.	
1501-1600	0.	0.	0.	0.	
1601-1700	1.	0.	1.	1650.	
1701-1800	0.	0.	0.	0.	
1801-1900	0.	0.	0.	0.	
1901-2000	0.	0.	0.	0.	
2001-2100	1.	0.	1.	2050.	
2101-2200	1.	0.	1.	2150.	
2201-2300	0.	0.	0.	0.	
2301-2400	0.	0.	0.	0.	
2401-2500	0.	0.	0.	0.	
2501-2600	0.	0.	0.	0.	
2601-2700	0.	0.	0.	0.	
2701-2800	0.	0.	0.	0.	
2801-2900	0.	0.	0.	0.	
2901-3000	0.	0.	0.	0.	
3001-3100	0.	0.	0.	0.	
3101-3200	0.	0.	0.	0.	
3201-3300	0.	0.	0.	0.	
3301-3400	0.	0.	0.	0.	
3401-3500	0.	0.	0.	0.	
3501-3600	1.	0.	1.	3550.	0.000
TOTAL	2607.	2607.		1120.	38.680

TOTAL EXP. LOSS / TOTAL ACT. COST = 0.2 %

D.F. =22

P <0.02

Table (4.3.7)

*** 3-PARAMETER WEIBULL DIS. ***

75/2ND QUARTER DATA

ESTIMATION BY MULTINOMIAL MAX. LIKELIHOOD METHOD :-

C= 16.52 A= 150.912 B= 0.856

MEAN= 179.934 S.D.= 191.534

AMOUNT £	ACTUAL CL. NO.	EXPECTED CL. NO.	ACTUAL-EXPECTED	EXPECTED LOSS	(A-E)**2/E
1- 30	302.	305.	-3.	-46.	0.030
31- 60	374.	428.	-54.	-2457.	6.813
61- 90	332.	312.	20.	1510.	1.282
91- 120	277.	243.	34.	3587.	4.757
121- 150	235.	195.	40.	5420.	8.205
151- 180	187.	159.	28.	4634.	4.931
181- 210	122.	131.	-9.	-1759.	0.618
211- 240	110.	108.	2.	451.	0.037
241- 270	80.	90.	-10.	-2555.	1.111
271- 300	72.	76.	-4.	-1142.	0.211
301- 330	47.	64.	-17.	-5363.	4.516
331- 360	39.	54.	-15.	-5182.	4.167
361- 390	40.	46.	-6.	-2253.	0.783
391- 420	38.	39.	-1.	-406.	0.026
421- 450	29.	33.	-4.	-1742.	0.485
451- 480	21.	29.	-8.	-3724.	2.207
481- 510	30.	24.	6.	2973.	1.500
511- 540	19.	21.	-2.	-1051.	0.190
541- 570	17.	18.	-1.	-555.	0.056
571- 600	11.	16.	-5.	-2927.	1.562
601- 700	36.	38.	-2.	-1301.	0.105
701- 800	22.	24.	-2.	-1501.	0.167
801- 900	22.	15.	7.	5953.	3.267
901-1000	11.	9.	2.	1901.	0.444
1001-1100	4.	6.	-2.	-2101.	0.667
1101-1200	3.	4.	-1.	-1150.	0.250
1201-1300	3.	2.	1.	1250.	
1301-1400	6.	2.	4.	5402.	
1401-1500	2.	1.	1.	1450.	
1501-1600	1.	1.	0.	0.	
1601-1700	0.	0.	0.	0.	
1701-1800	1.	0.	1.	1751.	
1801-1900	0.	0.	0.	0.	
1901-2000	1.	0.	1.	1950.	
2001-2100	0.	0.	0.	0.	
2101-2200	0.	0.	0.	0.	
2201-2300	1.	0.	1.	2250.	13.500
TOTAL	2495.	2493.		3266.	61.885

D.F. =23

P<0.001

TOTAL EXP. LOSS / TOTAL ACT. COST = 0.7 %

CHAPTER 5

The Inverse Gaussian Distribution

5.1 Introduction

This is a unimodal and positively skewed distribution which in recent years has been shown to represent the distribution of random outcomes of various phenomena. The name inverse Gaussian (or inverse normal) was proposed by Tweedie (1947) who noticed the inverse relationship between the cumulant generating function of this distribution and that of the normal distribution. Tweedie (1957) investigated the statistical properties of this distribution and indicated analogies between its statistical analysis and that of the normal distribution. In recent years this distribution has been studied by many authors. Johnson and Kotz (1970) provide a bibliography, and Chhikara and Folks (1978) give a review of the development since 1915 of the inverse Gaussian distribution and of the statistical methods based upon it.

The frequency curve of the inverse Gaussian distribution is similar in shape to that of the other skewed distributions like the lognormal, the gamma or the Weibull. It is, therefore, important to consider the application of this distribution in the fields where one, or some, of the others have been successfully used as a model. Chhikara and Folks (1978) give examples of sets of data equally well fitted by the lognormal and inverse Gaussian distributions. They point out that when more than one distribution fits a set of data equally well and there is no evidence (based on the physical considerations of the problem) in favour of a particular model then it is best to use that which is more convenient to work with. An advantage of the inverse Gaussian distribution over the lognormal is that its statistical properties and inference procedures are well developed for small sample situations.

Applications of the inverse Gaussian distribution in various fields have been reported by many authors. It has been successfully used in life-testing and reliability studies by Chhikara and Folks (1977). Hasofer (1964) considered it as a model for emptiness of a dam. Lancaster (1972) used it as a model for the duration of strikes. We have not come across the use of this distribution in insurance, and it is therefore considered important to study it as a model for the distribution of claim amounts.

In the literature, only the 2-parameter inverse Gaussian distribution has been considered. We will, however, introduce a location parameter into the model and will name it the 3-parameter inverse Gaussian distribution. Professor J. L. Folks, in a personal communication with the author (1978), acknowledges that this is an original idea and that this form of the distribution (i.e. the 3-parameter) has not been previously considered. We will later show that the 3-parameter distribution provides a better model for the accidental damage data.

In this chapter we will initially define the 2 and 3 parameter inverse Gaussian distributions and will mention some of their properties. The problem of estimation from grouped data is next considered. In section 5.5 the inverse Gaussian models will be fitted to our seven samples of accidental damage data and the results will be discussed. The effect of inflation on the parameters of the model will be studied in section 5.6 and, for the accidental damage data, the distribution of claim amounts during a future period will be derived and compared with the actual data. The findings of this chapter will be summarized in section 5.7. The tables will be presented in section 5.8.

5.2 Definition

A random variable X is said to have a (2-parameter) inverse

Gaussian distribution if its probability density function (p.d.f.), denoted by $f_{IG}(x; \mu,\lambda)$, is of the form

$$f_{IG}(x; \mu,\lambda) = (\lambda/2\pi x^3)^{1/2} \exp(-\lambda(x-\mu)^2/2\mu^2 x) \; ; x > 0$$
$$= 0 \quad \text{otherwise} \qquad (5.2\text{-}1)$$

where μ and λ are two independent and positive parameters. This distribution is thus a member of the exponential family. λ is a shape parameter while μ is only partially interpretable as a location parameter.

Shuster (1968) has shown that the inverse Gaussian distribution function, denoted by $IG(x; \mu,\lambda)$, can be expressed in terms of the standard normal distribution function, $N(z; 0,1)$, as :

$$IG(x; \mu,\lambda) = N(z_1; 0,1) + e^{2\lambda/\mu} N(z_2; 0,1) \qquad (5.2\text{-}2)$$

where

$$z_1 = (\lambda/x)^{1/2}(-1 + x/\mu)$$

$$z_2 = -(\lambda/x)^{1/2}(1 + x/\mu)$$

We will show later in this chapter that the 2-parameter inverse Gaussian distribution does not provide a satisfactory model for the accidental damage claim amounts. This fact, together with the concepts of 3-parameter lognormal and Weibull distributions, led us to consider a 3-parameter inverse Gaussian distribution where one of the parameters is a location (threshold) parameter below which the values of the random variable X are not realized. Hence we introduce the unknown location parameter c into the model and assume that X - c, rather than X, is distributed as in (5.2-1).

We say that a random variable X has a 3-parameter inverse Gaussian distribution if its p.d.f., denoted by $f_{IG}(x; c,\mu,\lambda)$, is of the form

$$f_{IG}(x; c,\mu,\lambda) = [\lambda/2\pi(x-c)^3]^{1/2} \exp[-\lambda(x-c-\mu)^2/2\mu^2(x-c)] \quad ; x > c$$
$$= 0 \quad \text{otherwise} \qquad (5.2\text{-}3)$$

The transformation $x \to x-c$ is only a translation along the x-axis. Therefore, the distribution function of X(i.e. the 3-parameter inverse Gaussian random variable) is obtained from (5.2-2) by replacing x by x-c.

In our research into the literature on the inverse Gaussian distribution, we have not come across the 3-parameter distribution of the form (5.2-3). In fact, as we mentioned earlier, Professor Folks in a personal communication with the author confirms that the idea of a 3-parameter inverse Gaussian distribution, with one of the parameters being an unknown location parameter, has not been considered before. This new form suggests, as an area for research, the study of the statistical properties and the estimation and inference problems for the 3-parameter distribution. We shall deal with these problems in this chapter as far as is relevant to the present work. It will be shown that the 3-parameter distribution provides an adequate model for our accidental damage claim amounts.

5.3 Properties of the inverse Gaussian Distribution

In this section, we first outline briefly some properties of the 2-parameter distribution with p.d.f. of the form given by (5.2-1). for more details and proofs reference should be made to Johnson and Kotz (1970) or Tweedie (1957).

Moments of all orders exist and in particular the rth positive integral moment of X about 0 is given by Tweedie (1957) as :

$$E(X^r) = \mu^r \sum_{s=0}^{r-1} \frac{(r-1+s)!}{s!(r-1-s)!} \left(\frac{\mu}{2\lambda}\right)^s \qquad (5.3\text{-}1)$$

In particular $E(X) = \mu$ (5.3-2)

and $E(X^2) = \mu^2 + \mu^3\lambda^{-1}$ (5.3-3)

hence $\mathrm{var}(X) = \dfrac{\mu^3}{\lambda}$ (5.3-4)

Also $$E(X^3) = \mu^3 + 3\mu^4\lambda^{-1} + 3\mu^5\lambda^{-2} \qquad (5.3\text{-}5)$$

and $$E(X^4) = \mu^4 + 6\mu^5\lambda^{-1} + 15\mu^6\lambda^{-2} + 15\mu^7\lambda^{-3} \qquad (5.3\text{-}6)$$

Parameter μ is, therefore, the population mean and is a measure of location. This model, unlike for instance the lognormal model, possesses the useful characteristic that one of its parameters is the population mean.

The coefficient of variation for this distribution is $\phi^{-\frac{1}{2}}$ where

$$\phi = \frac{\lambda}{\mu} > 0 \qquad (5.3\text{-}7)$$

The coefficients of skewness and kurtosis, $\sqrt{\beta_1}$ and β_2 respectively, are

$$\sqrt{\beta_1} = 3\sqrt{\mu/\lambda} = 3\phi^{-\frac{1}{2}} \qquad (5.3\text{-}8)$$

$$\beta_2 = 3 + \frac{15\mu}{\lambda} = 3 + 15\phi^{-1} \qquad (5.3\text{-}9)$$

The shape of the distribution depends on ϕ only and, since $\phi > 0$, its frequency curve is skewed to the right and more peaked than the normal frequency curve. The distribution becomes more and more nearly normal when ϕ is increased.

The distribution has a single mode which is located at

$$x_{mode} = \mu\left\{(1 + \frac{9\mu^2}{4\lambda^2})^{\frac{1}{2}} - \frac{3\mu}{2\lambda}\right\} \qquad (5.3\text{-}10)$$

or

$$x_{mode} = \mu\left\{(1 + \frac{9}{4\phi^2})^{\frac{1}{2}} - \frac{3}{2\phi}\right\} \qquad (5.3\text{-}11)$$

The mode of the distribution is in general located to the left of its mean because

$$\frac{x_{mode}}{\mu} < 1 \quad ,$$

but when ϕ is increased to infinity x_{mode} will converge to μ. It is possible to express the inverse Gaussian p.d.f. in terms of the pair of parameters (μ,ϕ) or (λ,ϕ), but the form given by (5.2-1) is generally considered as the standard form and is most usually used in the literature. The one parameter form which is obtained by putting $\mu = 1$ in (5.2-1) is known as the standard Wald distribution.

The properties of the 3-parameter inverse Gaussian distribution, whose p.d.f. was given by (5.2-3), can be deduced from those of the 2-parameter. The transformation $x \to x - c$ is one of translation along the x-axis only and, therefore, the measures of location will be increased by c while the measures of dispersion and the moments about the mean will be unchanged. Hence, for the 3-parameter distribution we will have :

$$E(X) = c + \mu \tag{5.3-12}$$

$$\text{var}(X) = \frac{\mu^3}{\lambda} \tag{5.3-13}$$

$$x_{mode} = c + \mu\left\{\left(1 + \frac{9\mu^2}{4\lambda^2}\right)^{1/2} - \frac{3\mu}{2\lambda}\right\} \tag{5.3-14}$$

and

$$\text{Coeff. of variation} = \frac{\mu^{3/2}}{[\lambda(c+\mu)]^{1/2}} \tag{5.3-15}$$

which is a function of all three parameters.

5.4 Estimation of the Parameters from Grouped Data

The problem of estimation from a sample of data containing the values of individual observations has been dealt with by Schrodinger and his maximum likelihood estimators are reported in Chikkara and Folks (1978). Here we concentrate on the estimation from a sample of grouped data when all the parameters are assumed unknown.

Let us assume that we have a sample of grouped data where n independent random observations on a random variable X (in our case the claim amount) have been grouped according to their size into k mutually exclusive intervals. Further, let n_i be the number of observations (claims) in the class interval (x_{i-1}, x_i), for $i = 1, 2, \ldots, k$, and such that

$$n = \sum_{i=1}^{k} n_i$$

Also let us assume that X has a 2-parameter inverse Gaussian distribution of the form (5.2-1). The method of moments can be used to estimate the parameters. This involves putting as many sample moments as there are unknown parameters equal to their corresponding population values and solving the resulting equations for the unknown parameters. For the 2-parameter case the sample mean and coefficient of variation may be used. The sample mean is the moment estimator of μ and, when the values of individual observations are known, is equal to its maximum likelihood estimator (see Johnson and Kotz (1970)). Therefore the sample mean is an unbiased and efficient estimator of μ. In calculating the sample moments from grouped data the usual assumption about the concentration of all the observations in each interval at the mid-point of that interval may be made. In the 3-parameter case, as well as the sample mean and coefficient of variation, the third central moment, which is independent of c, should be used to yield sufficient equations in the unknown parameters.

It was observed in the previous chapters that the method of multinomial maximum likelihood yielded the best fits of the models to actual data. Here we consider this method for the 3-parameter inverse Gaussian. The 2-parameter case can be similarly dealt with by putting $c = 0$ in the following exposition.

Let us assume that we have a sample of grouped data, as defined earlier, and that random variable X has a 3-parameter inverse Gaussian distribution with p.d.f. of the form (5.2-3) and distribution function given by : (see section 5.2)

$$\begin{aligned} IG(x;\, c,\mu,\lambda) &= N(z_1;\, 0,1) + e^{2\lambda/\mu}\, N(z_2;\, 0,1) \text{ for } x > c \\ &= 0 \quad \text{otherwise} \end{aligned} \qquad (5.4\text{-}1)$$

where $z_1 = (\lambda/(x-c))^{\frac{1}{2}} (-1 + (x-c)/\mu)$

and

$$z_2 = -(\lambda/(x-c))^{\frac{1}{2}} (1 + (x-c)/\mu)$$

Let p_i be the probability that an observation (claim, in our case) occurs in the interval (x_{i-1}, x_i) , for $i = 1, 2, \ldots, k$, where

$$p_i = IG(x_i ; c,\mu,\lambda) - IG(x_{i-1}; c,\mu,\lambda) \qquad (5.4\text{-}2)$$

which can be expressed in terms of the standard normal distribution function by using (5.4-1). The likelihood of the sample will be proportional to L where

$$L = \prod_{i=1}^{k} p_i^{n_i}$$

The loglikelihood equation, log L, can thus be expressed as

$$\log L = \sum_{i=1}^{k} n_i \log p_i$$

The maximum likelihood estimates are the set of parameter values $(\hat{c}, \hat{\mu}, \hat{\lambda})$ which maximizes log L. The iterative computing technique which was described in section 3.7.7 of chapter 3 can be used to maximize log L and yield the set $(\hat{c}, \hat{\mu}, \hat{\lambda})$. This technique requires initial values for the parameters in order to start the iteration procedure. The estimates of the parameters obtained by the method of moments can be used for this purpose.

5.5 Application of the Inverse Gaussian Model to AD Data

We wrote computer program P18 to estimate the parameters of the 2-parameter inverse Gaussian distribution by the multinomial maximum likelihood , (MML) , method. This program was run with the samples of accidental damage data which were presented in tables (1.1) to (1.7).

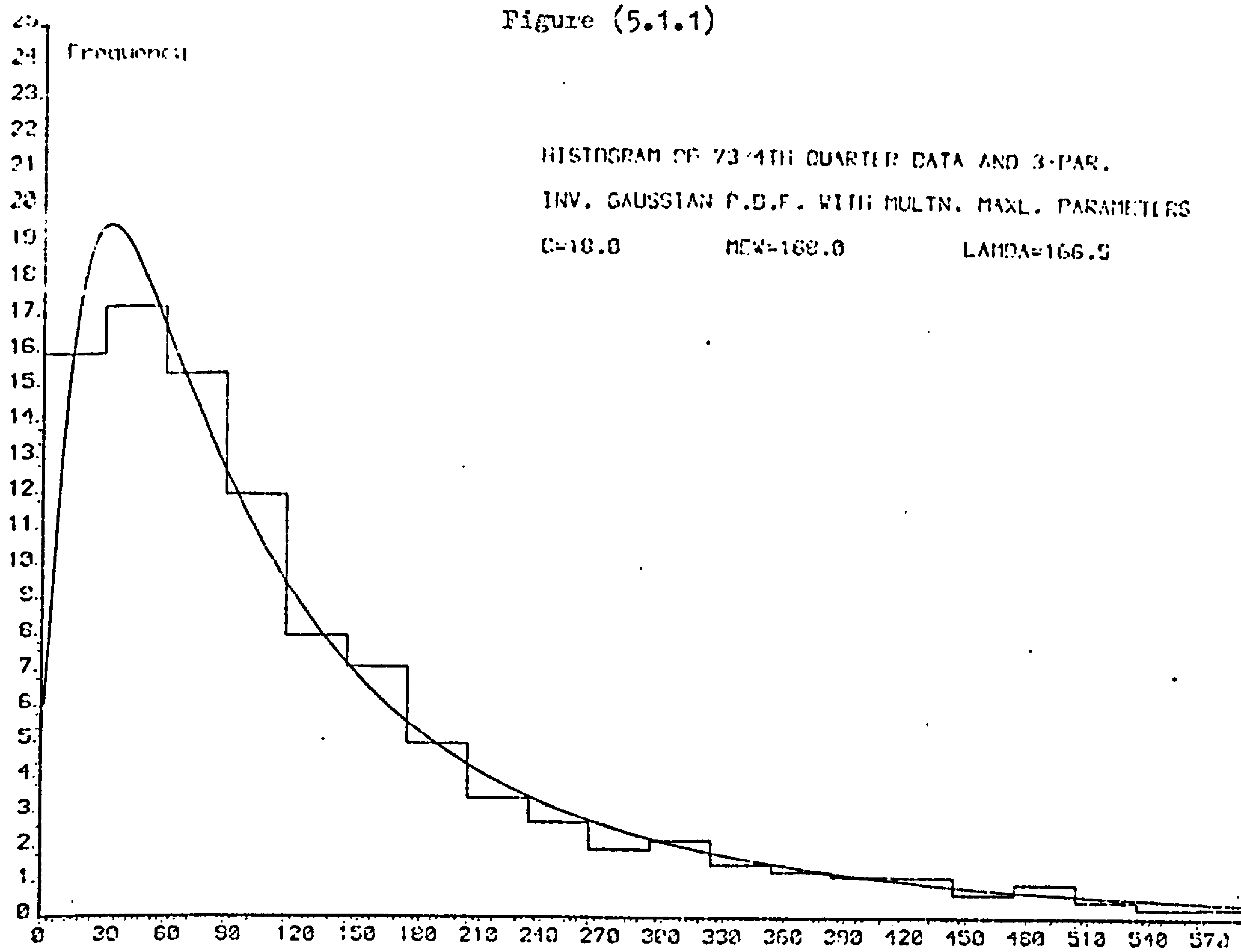
For each sample the estimates of the parameters μ and λ were found by the method of moments. The statistics given at the bottom of each of the tables (1.1) to (1.7) made the above task very easy. The moment estimates were supplied to program P18 as starting values for the iteration process. The program produced an extensive table of results for each sample. These are summarized in table (5.1). The estimate of $\hat{\mu}$ in each case is practically equal to the sample mean (refer to tables (1.1) to (1.7)). The values of μ and λ show a generally increasing trend over time. The ratio of the total expected loss statistic, T , to the total actual cost is small in each case. This indicates an overall agreement between the model and actual data. The Chi-square statistics are, however, large and indicate significant differences between the model and actual sample values. The examination of the components of this statistic, for each sample, showed that very large contributions are made to the value of X^2 by only one or two intervals in the lower tail of the distribution. For example the first and second intervals account for some 60% of the X^2 values. Therefore, the agreement between the model and actual sample values is reasonable except in the lower tail. In terms of the Chi-square statistic the 2-parameter lognormal model provided a better fit to the actual sample values than the present model has.

The frequency curve of the inverse Gaussian distribution is very similar to that of the lognormal distribution. To improve the fit of the model to the actual data, we considered the idea of a 3-parameter inverse Gaussian distribution similar to the idea of a 3-parameter lognormal distribution. We, therefore, assumed that X + c, and not X, is distributed as an inverse Gaussian distribution, with parameters μ and λ , where c is an unknown parameter which can be interpreted as the voluntary excess on the policy. This assumption simply means that the total amount of claim, X + c, (which consists of the amount X paid by

the insurance company and the amount c borne by the policy holder) has an inverse Gaussian distribution. The remarks we made about parameter c of the lognormal model, in section 3.10, are valid here as well and it is useful to refer back to them at this stage.
We, therefore, assume that c is fixed for all policies but that its value is unknown, and proceed to estimate it, along with μ and λ from the data. The computer program P19 estimates the 3 parameters from a sample of grouped data and prints an extensive table of results.
For the accidental damage data of tables (1.1) to (1.7) the results are presented in tables (5.2.1) to (5.2.7).
The estimates of c show an increase over time and are generally greater than their corresponding values for the 3-parameter lognormal model. The mean claim amount of the model is, in each case, practically equal to the sample mean. The standard deviations are also close to their corresponding actual values and in each case are smaller than those obtained by the 3-parameter lognormal model. The Chi-square statistics are smaller than in the 2-parameter case and (except in two or three cases where large contributions from one or two cells have resulted in large values for X^2) do not indicate any significant differences between the model and the actual sample values. The Chi-square statistics are practically equal to their corresponding values for the 3-parameter lognormal model. The total expected loss statistics, T, are also small and are at most 1% of the total actual cost. This indicates a general agreement between the model and the sample values.

To see how closely the frequency curve of the model agrees with the histogram of the sample, we wrote computer program P20 to plot them. The MML estimates of c, μ and λ, given in tables (5.2.1) to (5.2.7), were used to plot the frequency curve of the fitted model for each sample. The graphs are presented in figures (5.1.1) to (5.1.7). We observe an overall agreement between the curve and the histogram for each sample. In particular, the curve has portrayed the distinct

Figure (5.1.1)


Frequency
HISTOGRAM OF 73/4TH QUARTER DATA AND 3-PAR.
INV. GAUSSIAN P.D.F. WITH MULTN. MAXL. PARAMETERS
C=18.0 MEW=168.0 LAMDA=166.5
0 30 60 90 120 150 180 210 240 270 300 330 360 390 420 450 480 510 540 570

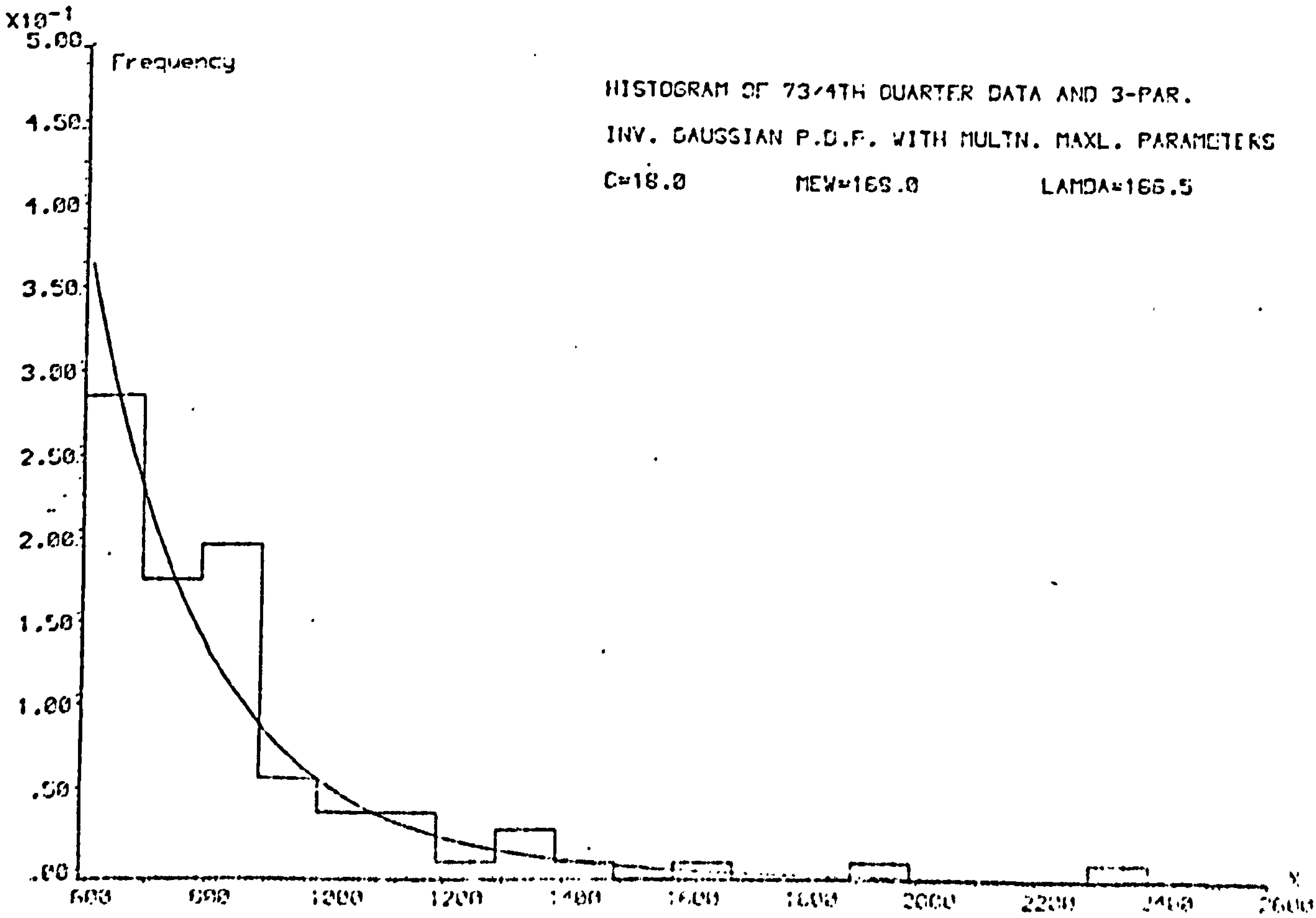
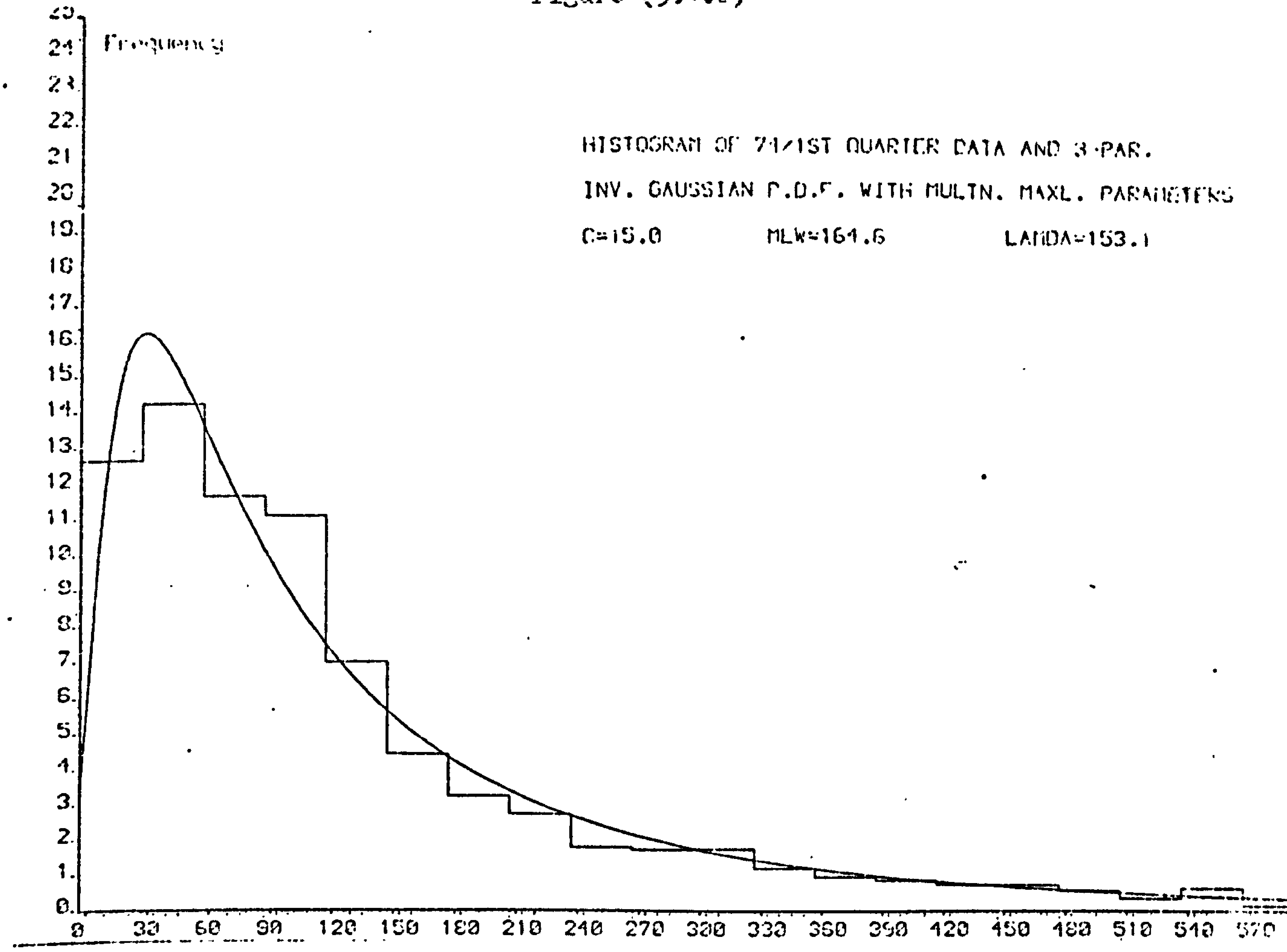

X10^-1
Frequency
HISTOGRAM OF 73/4TH QUARTER DATA AND 3-PAR.
INV. GAUSSIAN P.D.F. WITH MULTN. MAXL. PARAMETERS
C=18.0 MEW=168.0 LAMDA=166.5
5.00 4.50 4.00 3.50 3.00 2.50 2.00 1.50 1.00 .50 .00
X

Figure (5.1.2)


Frequency
HISTOGRAM OF 74/1ST QUARTER DATA AND 3-PAR.
INV. GAUSSIAN P.D.F. WITH MULTN. MAXL. PARAMETERS
C=15.0 MEW=164.6 LAMDA=153.1
0 30 60 90 120 150 180 210 240 270 300 330 360 390 420 450 480 510 540 570

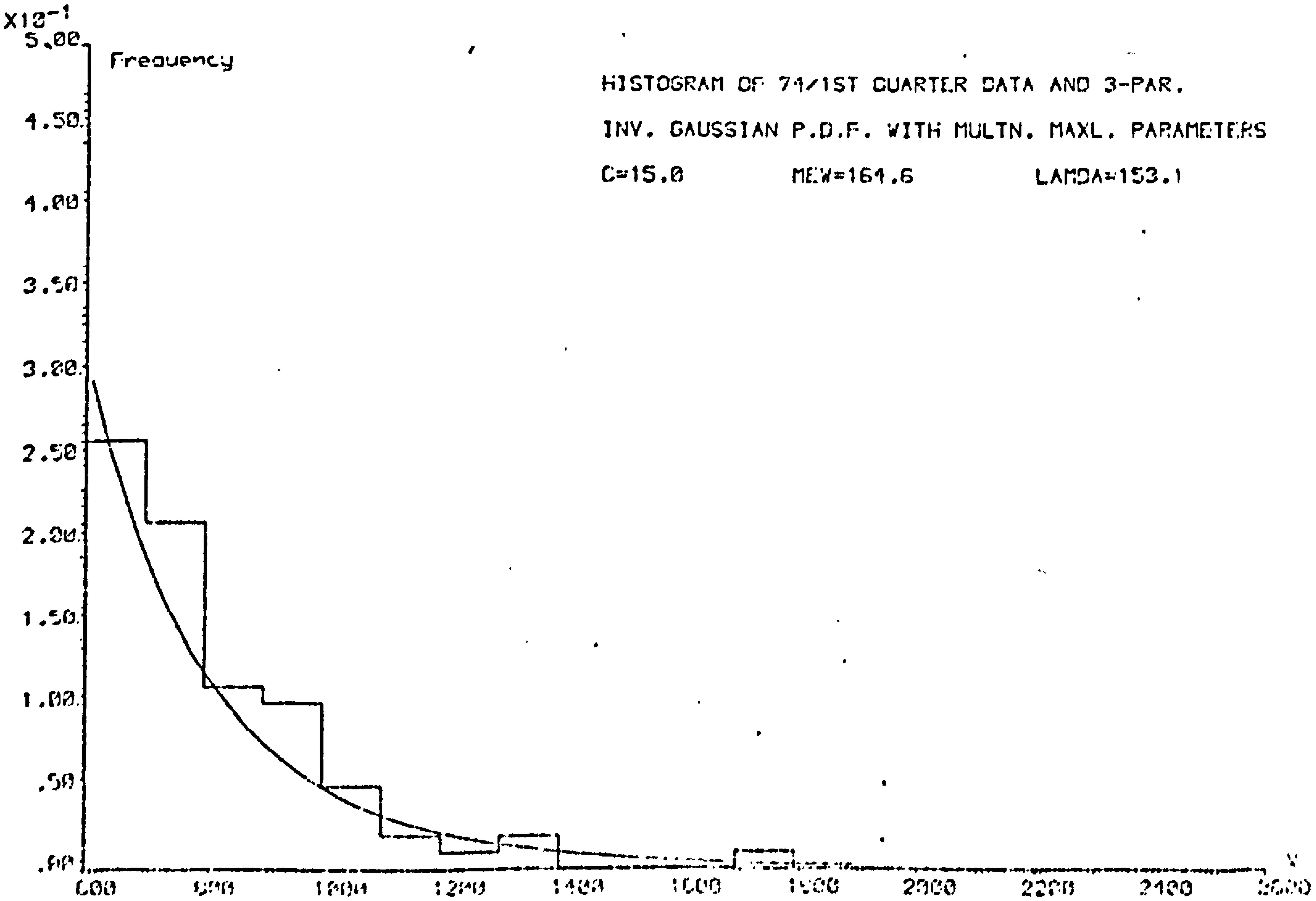
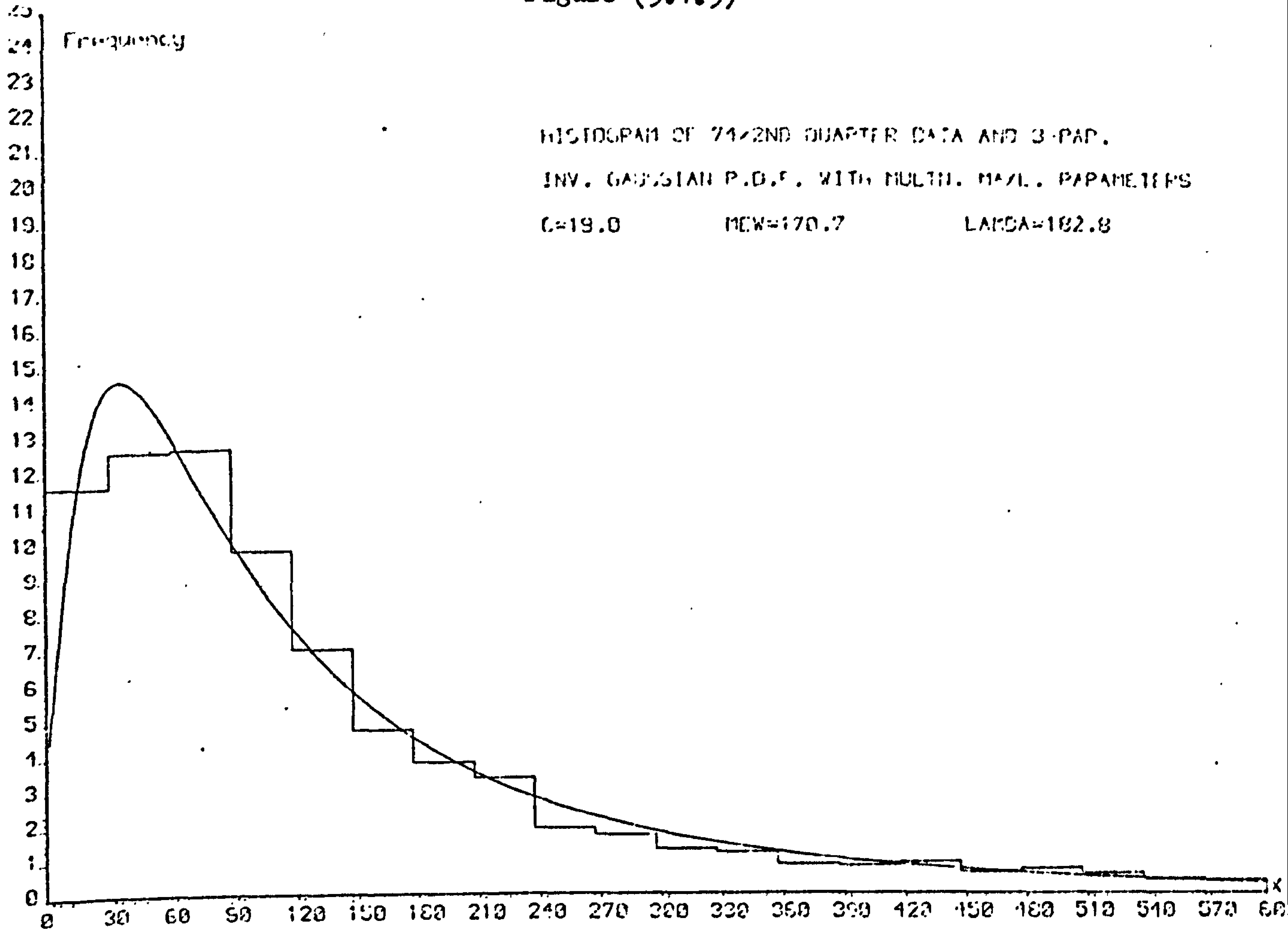
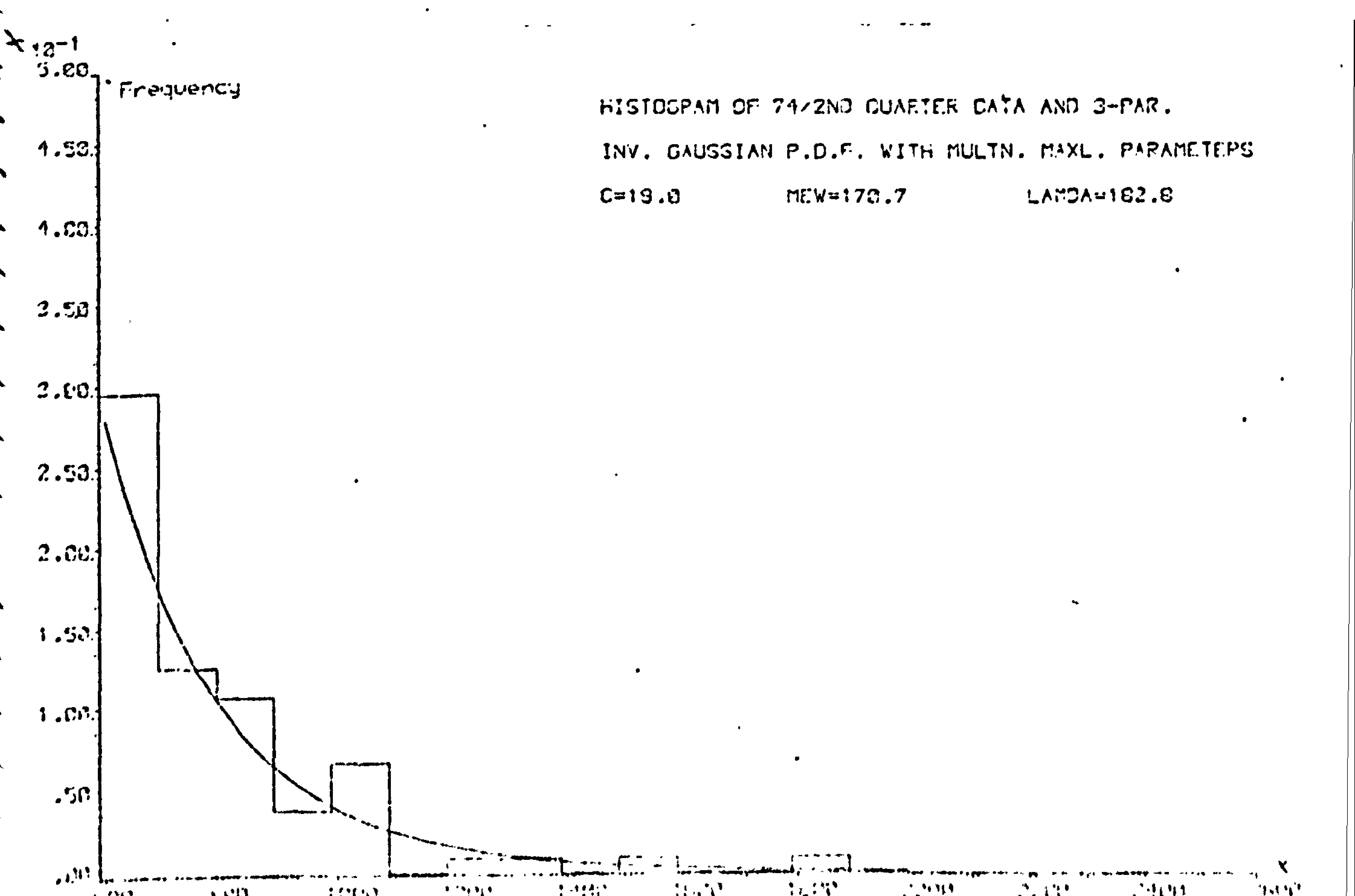

X10^-1
Frequency
HISTOGRAM OF 74/1ST QUARTER DATA AND 3-PAR.
INV. GAUSSIAN P.D.F. WITH MULTN. MAXL. PARAMETERS
C=15.0 MEW=164.6 LAMDA=153.1
5.00 4.50 4.00 3.50 3.00 2.50 2.00 1.50 1.00 .50 .00
1000 1200 1400 1600 2000 2200 2400

Figure (5.1.3)

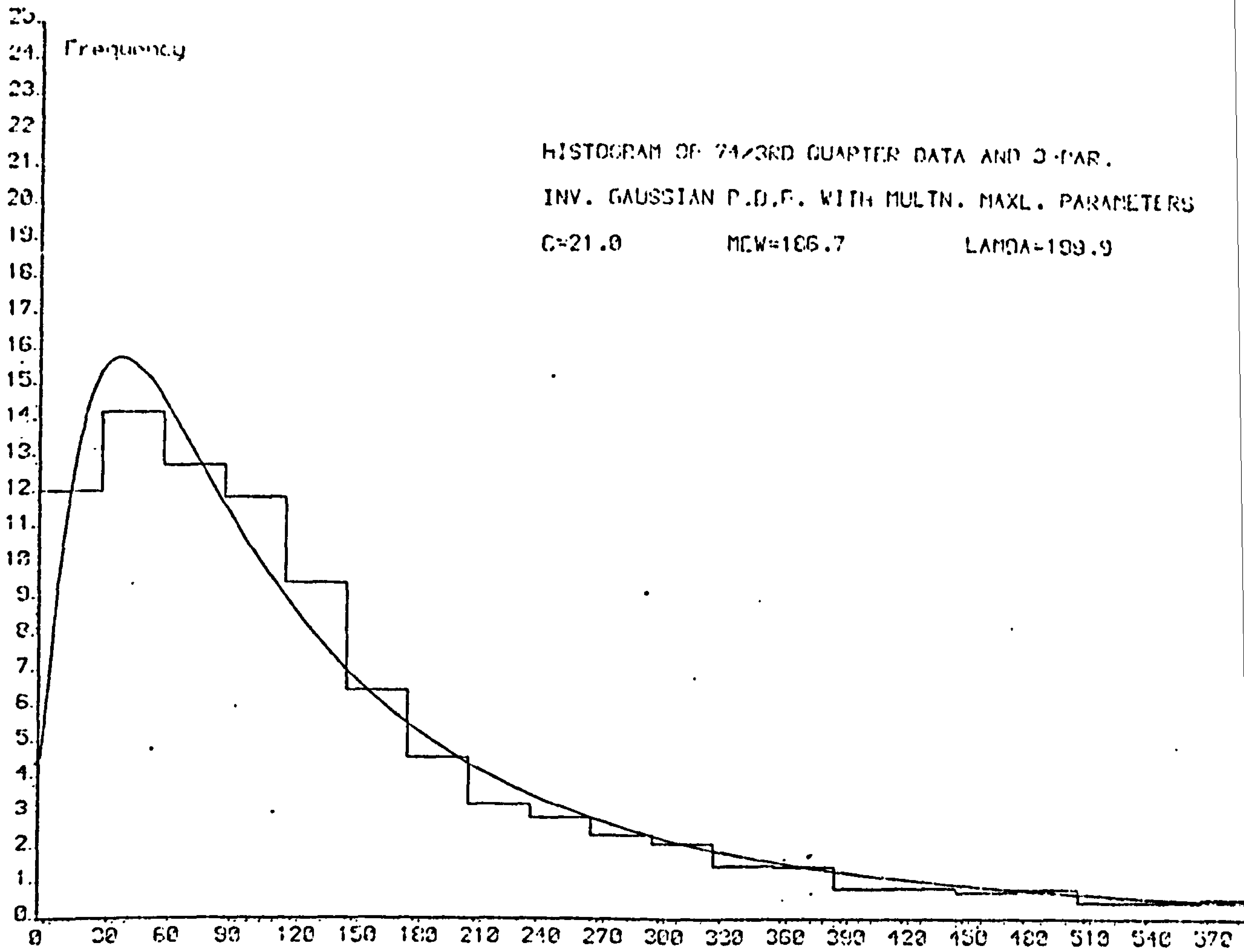


Frequency
HISTOGRAM OF 74/2ND QUARTER DATA AND 3-PAR.
INV. GAUSSIAN P.D.F. WITH MULTN. MAXL. PARAMETERS
C=19.0 MEW=170.7 LAMDA=182.8
0 30 60 90 120 150 180 210 240 270 300 330 360 390 420 450 480 510 540 570



x10^-1
Frequency
HISTOGRAM OF 74/2ND QUARTER DATA AND 3-PAR.
INV. GAUSSIAN P.D.F. WITH MULTN. MAXL. PARAMETERS
C=19.0 MEW=170.7 LAMDA=182.8
3.00 4.50 4.00 3.50 3.00 2.50 2.00 1.50 1.00 .50

Figure (5.1.4)


Frequency
HISTOGRAM OF 74/3RD QUARTER DATA AND 3-PAR.
INV. GAUSSIAN P.D.F. WITH MULTN. MAXL. PARAMETERS
C=21.0
MEW=166.7
LAMDA=199.9

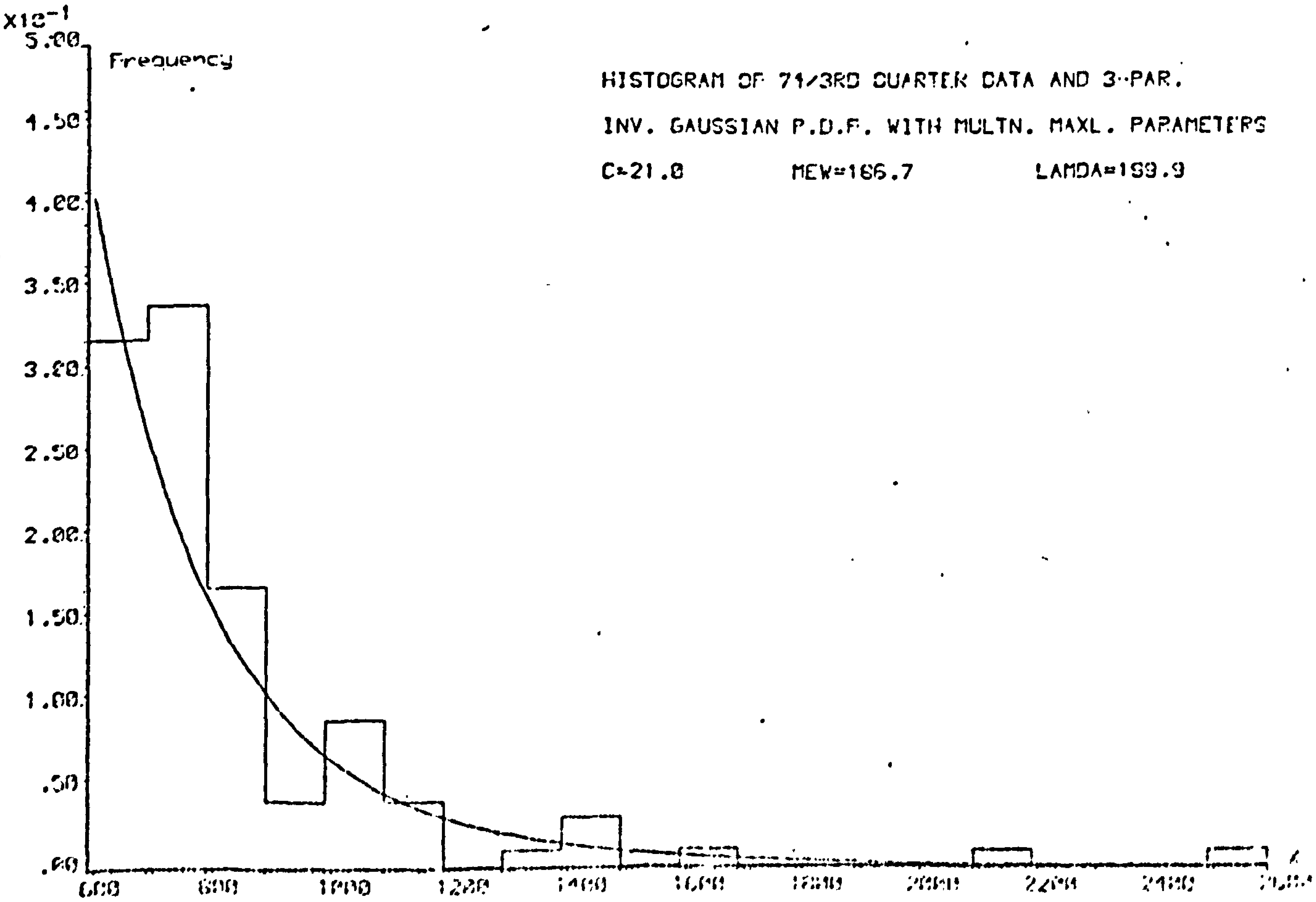
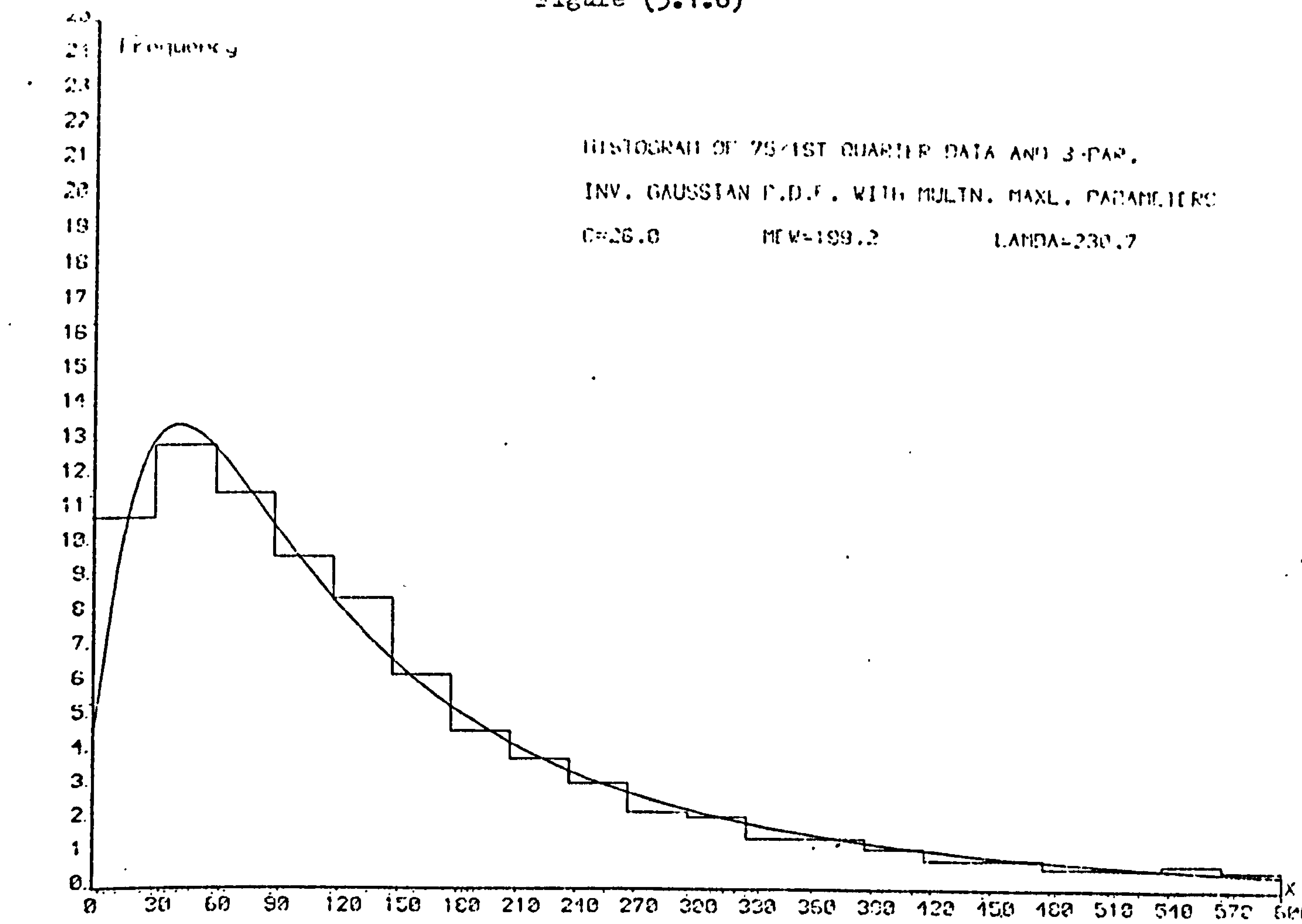

x10⁻¹
Frequency
HISTOGRAM OF 74/3RD QUARTER DATA AND 3-PAR.
INV. GAUSSIAN P.D.F. WITH MULTN. MAXL. PARAMETERS
C=21.0
MEW=166.7
LAMDA=199.9

Figure (5.1.5)


Frequency
HISTOGRAM OF 71/4TH QUARTER DATA AND 3-PAR.
INV. GAUSSIAN P.D.F. WITH MULTN. MAXL. PARAMETERS
C=24.0
MEW=197.7
0 30 60 90 120 150 180 210 240 270 300 330 360 390 420 450 480 510 540 570
0. 1. 2. 3. 4. 5. 6. 7. 8. 9. 10. 11. 12. 13. 14. 15. 16. 17. 18. 19. 20. 21. 22. 23.
X10^-1
Frequency
HISTOGRAM OF 71/4TH QUARTER DATA AND 3-PAR.
INV. GAUSSIAN P.D.F. WITH MULTN. MAXL. PARAMETERS
C=24.0
MEW=197.7
LAMDA=211.8
5.00 4.50 4.00 3.50 3.00 2.50 2.00 1.50 1.00 .50
1200
2000
X

Figure (5.1.6)


Frequency
HISTOGRAM OF 75/1ST QUARTER DATA AND 3-PAR.
INV. GAUSSIAN P.D.F. WITH MULTN. MAXL. PARAMETERS
C=26.0
MEW=199.2
LAMDA=230.7
X

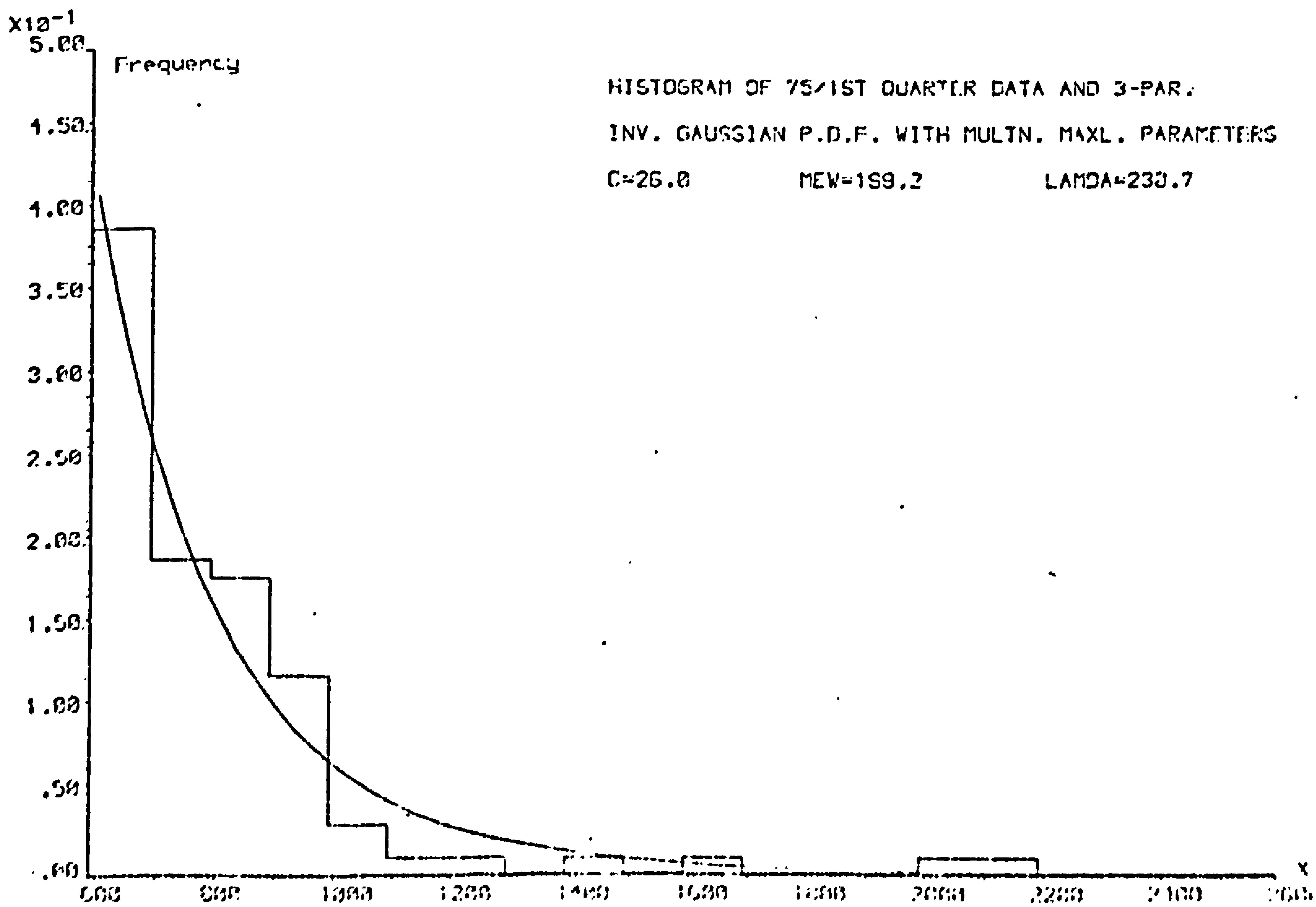
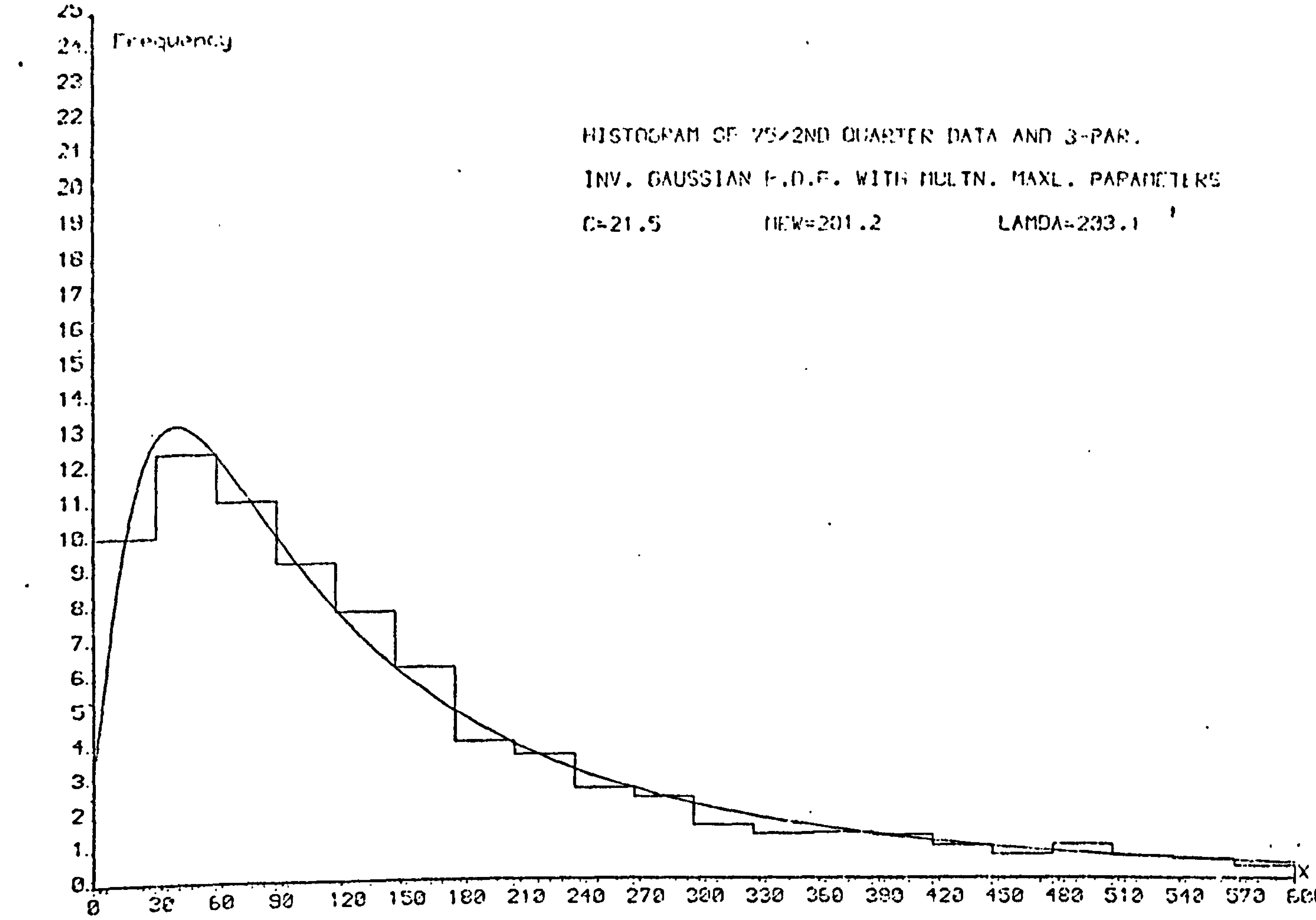
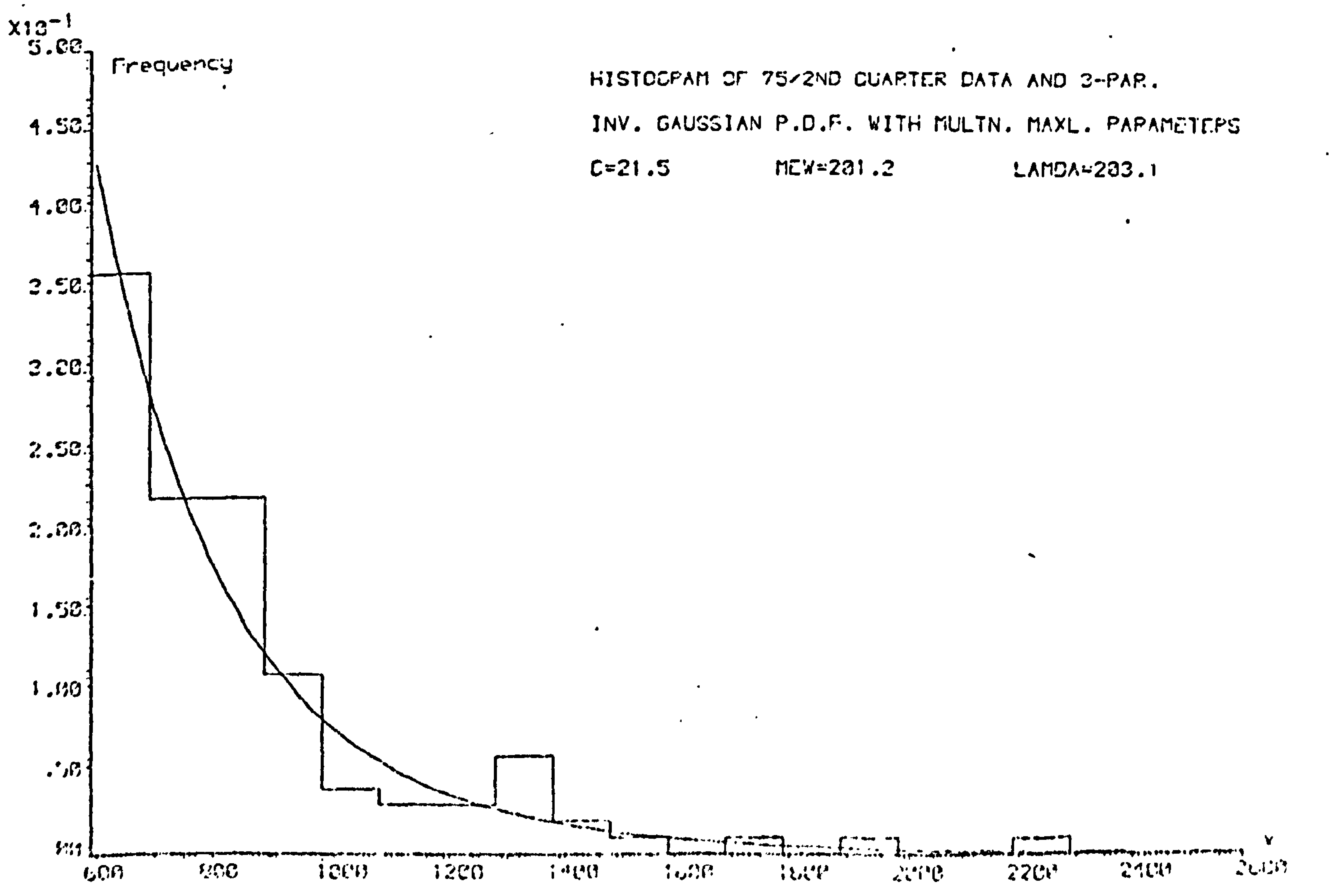

X10^-1
Frequency
HISTOGRAM OF 75/1ST QUARTER DATA AND 3-PAR.
INV. GAUSSIAN P.D.F. WITH MULTN. MAXL. PARAMETERS
C=26.0
MEW=199.2
LAMDA=230.7
5.00
4.50
4.00
3.50
3.00
2.50
2.00
1.50
1.00
.50
X

Figure (5.1.7)


Frequency
HISTOGRAM OF 75/2ND QUARTER DATA AND 3-PAR.
INV. GAUSSIAN P.D.F. WITH MULTN. MAXL. PARAMETERS
C=21.5
MEW=201.2
LAMDA=233.1
X


X10⁻¹
Frequency
HISTOGRAM OF 75/2ND QUARTER DATA AND 3-PAR.
INV. GAUSSIAN P.D.F. WITH MULTN. MAXL. PARAMETERS
C=21.5
MEW=201.2
LAMDA=233.1

mode and the long tail of the histogram very well. The curves look remarkably similar to those for the 3-parameter lognormal distribution given in figures (3.9.1) to (3.9.7). However, the latter curves have a slightly fatter and longer tail.

The above analysis shows that the 3-parameter inverse Gaussian distribution is an appropriate model for the accidental damage claim amounts. This model produces results very similar to those of the 3-parameter lognormal distribution.

5.6 Prediction of the Claim Amount Distribution

The importance of predicting the future cost of claims was mentioned in section 3.9 (chapter 3). Here, we will adopt techniques similar to those of that section for predicting the distribution of claim amounts, during any future period, when the underlying model is the inverse Gaussian distribution. To save us repeating some of the arguments of that section, we suggest referring back to it at this stage.

5.6.1 The Effects of Inflation on the Parameters of the Model

Let us again assume that the effect of inflation on a claim of amount X is to increase it, over a certain period, to $U = X(1 + i)$ where i is the rate of inflation according to some appropriate index for that period. If X is assumed to be distributed as the two-parameter inverse Gaussian, $IG(x; \mu,\lambda)$ then, by a transformation of variables, we can show that $U = X(1 + i)$ will be distributed as $IG(u; r\mu, r\lambda)$ where $r = 1 + i$. If X is assumed to have the 3-parameter inverse Gaussian distribution $IG(x; c,\mu,\lambda)$ then similarly it can be shown that is distributed as $IG(u; rc, r\mu, r\lambda)$. Therefore when the distribution of X is known we can modify its parameters to obtain the distribution of claim amounts in a future period. The foregoing shows that inflation increases all the parameters of this model over time. This is reflected in the estimates of the parameters from different samples presented in table (5.1), for the 2-parameter distribution, and in tables (5.2.1) to (5.2.7) for the 3-parameter case.

The data analysis of section 5.5 showed that the 2-parameter inverse Gaussian distribution is not an adequate model for the accidental damage claim amounts. Therefore, we will not consider this model for prediction purposes. The 3-parameter distribution, however, was shown to be a satisfactory model and hence we suggest using the technique of section 3.9.2 for predicting the distribution of accidental damage claim amounts during a future period. In chapter 3 it was shown that the appropriate index for calculating the rate of inflation for accidental damage claims is the General Index of Retail Prices. We will use the same index here again.

5.6.2 Prediction for the AD Data

To test the prediction technique on the Accidental Damage (AD) data we wrote computer program P21 for the 3-parameter inverse Gaussian model. For reasons presented in section 3.9.2 we used the MML estimates of the parameters of the distributions for 73/4th, 74/1st and 74/2nd quarters to predict the distributions of claim amounts in 74/4th, 75/1st and 75/2nd quarters respectively. The rate of inflation in each case was calculated from the General Index of Retail Prices as described in section 3.9.2. The results for the AD data are presented in tables (5.3.1) to (5.3.3). These are in the form of extensive tables which allow comparisons between the actual and predicted distributions. In all three cases the Chi-square statistics are relatively small, considering the number of degrees of freedom, and do not indicate any significant differences between the predicted distributions and the actual sample values. The total expected loss statistics are small and show an over-prediction of the total cost, by at most 3·9%, which is more acceptable than any under-prediction. The Kolmogorov-Smirnov statistics were calculated from the tables and in each case were compared with the significance points of table (2.1). They showed that the differences between the actual and predicted distributions were not significant for 74/4th and 75/1st quarters but were almost significant for 75/2nd quarter.

Our prediction technique, therefore, provides a satisfactory means of predicting the future distribution of claim amounts when the underlying model is the 3-parameter inverse Gaussian distribution. For the

AD data the values of the goodness-of-fit test statistics are very similar to those obtained from the 3-parameter lognormal distributions in tables (3.45) to (3.47). This is only to be expected since, as we mentioned earlier, these two models are very similar to each other. The lognormal model, however, resulted in smaller values for the total expected loss statistics and hence overpredicted the total actual cost by a smaller amount.

5.7 Conclusions

In this chapter the 2-parameter inverse Gaussian distribution was initially defined and then modified to a three parameter distribution by the introduction of a location parameter. The properties of these distributions were considered and their problems of parameter estimation from grouped data were dealt with. The multinomial maximum likelihood method was suggested for these purposes. The two and three parameter distributions were next considered as models for the distribution of accidental damage claim amounts. It was shown that the 2-parameter distribution is not an adequate model, while our proposed 3-parameter distribution provides a very satisfactory model which is as good as the 3-parameter lognormal. The effects of inflation on the parameters of the model were studied,and the future distribution of claim amounts was predicted. It was shown that our prediction technique produces satisfactory results when the underlying model is the 3-parameter inverse Gaussian distribution.

The similar features of the 3-parameter inverse Gaussian and lognormal models were pointed out at various stages. Although the latter distribution has a slightly longer tail, and produced a slightly better fit to the data, either of the two models may be satisfactorily used. The MML method can be used for estimation of parameters in both cases with equal labour. The distribution functions of both models are expressed in terms of the standard normal distribution function. It is, however, possible to test if a given sample of data is from the lognormal distribution, while similar tests do not exist for the inverse Gaussian model. This is an advantage of the former model since, if the test indicates that the sample is not very likely to be from the distribution under consideration, then further analysis can be avoided. The existence of a theoretical justification for the emergence of the lognormal model as the distribution of claim amounts is another strong point in favour of that model. On the question of convenience in use, we believe that, with the wide availability of computers, either model is equally suitable.

5.8 Tables

Table (5.1)

Two-parameter inverse Gaussian model fitted to

the accidental damage data by

the multinomial maximum likelihood method

Period of Accident	$\hat{\mu}$	$\hat{\lambda}$	R%	X^2 (D.F.)
73 4th Quarter	150.60	96.77	0.6	96.3 (25)
74 1st "	150.07	96.19	0.7	84.1 (25)
74 2nd "	151.82	103.06	0.5	90.2 (25)
74 3rd "	166.33	111.48	0.3	123.5 (25)
74 4th "	174.40	111.03	1.0	116.4 (26)
75 1st "	173.62	114.91	1.1	113.5 (26)
75 2nd "	180.27	113.72	0.5	96.1 (26)

R : The ratio of the total expected loss statistic, T , to the total actual cost of claims

Table (5.2.1)

*** 3-PARAMETER INV. GAUSSIAN DIS. ***

73/4TH QUARTER DATA

ESTIMATION BY MULTINOMIAL MAX. LIKELIHOOD METHOD :-

C= 18.05 MEW= 168.034 LAMDA= 166.536

MEAN= 149.984 S.D.= 168.789

AMOUNT £	ACTUAL CL. NO.	EXPECTED CL. NO.	ACTUAL-EXPECTED	EXPECTED LOSS	(A-E)**2/E
1- 30	478.	474.	4.	62.	0.034
31- 60	518.	554.	-36.	-1638.	2.339
61- 90	461.	437.	24.	1812.	1.318
91- 120	359.	328.	31.	3270.	2.930
121- 150	239.	247.	-8.	-1084.	0.259
151- 180	213.	189.	24.	3972.	3.048
181- 210	148.	147.	1.	195.	0.007
211- 240	102.	116.	-14.	-3157.	1.690
241- 270	81.	92.	-11.	-2810.	1.315
271- 300	58.	75.	-17.	-4853.	3.853
301- 330	66.	61.	5.	1577.	0.410
331- 360	45.	50.	-5.	-1727.	0.500
361- 390	39.	41.	-2.	-751.	0.098
391- 420	35.	34.	1.	406.	0.029
421- 450	34.	29.	5.	2177.	0.862
451- 480	20.	24.	-4.	-1862.	0.667
481- 510	29.	20.	9.	4459.	4.050
511- 540	14.	17.	-3.	-1576.	0.529
541- 570	8.	15.	-7.	-3888.	3.267
571- 600	9.	13.	-4.	-2342.	1.231
601- 700	29.	30.	-1.	-651.	0.033
701- 800	18.	19.	-1.	-751.	0.053
801- 900	20.	12.	8.	6804.	5.333
901-1000	6.	7.	-1.	-951.	0.143
1001-1100	4.	5.	-1.	-1050.	0.200
1101-1200	4.	3.	1.	1150.	
1201-1300	1.	2.	-1.	-1250.	
1301-1400	3.	1.	2.	2701.	
1401-1500	1.	1.	0.	0.	
1501-1600	0.	1.	-1.	-1550.	
1601-1700	1.	0.	1.	1650.	
1701-1800	0.	0.	0.	0.	
1801-1900	0.	0.	0.	0.	
1901-2000	1.	0.	1.	1950.	
2001-2100	0.	0.	0.	0.	
2101-2200	0.	0.	0.	0.	
2201-2300	0.	0.	0.	0.	
2301-2400	1.	0.	1.	2350.	2.000
TOTAL	3045.	3044.		2645.	36.397

D.F. = 22

TOTAL EXP. LOSS / TOTAL ACT. COST = 0.6 %

P = 0.025

Table (5.2.2)

*** 3-PARAMETER INV. GAUSSIAN DIS. ***

74/1ST QUARTER DATA

ESTIMATION BY MULTINOMIAL MAX. LIKELIHOOD METHOD :-

C= 15.08 MEW= 164.632 LAMDA= 153.065

MEAN= 149.553 S.D.= 170.739

AMOUNT £	ACTUAL CL. NO.	EXPECTED CL. NO.	ACTUAL-EXPECTED	EXPECTED LOSS	(A-E)**2/E
1- 30	381.	377.	4.	62.	0.042
31- 60	428.	458.	-30.	-1365.	1.965
61- 90	351.	354.	-3.	-227.	0.025
91- 120	334.	262.	72.	7596.	19.786
121- 150	211.	196.	15.	2032.	1.148
151- 180	133.	149.	-16.	-2648.	1.718
181- 210	98.	115.	-17.	-3323.	2.513
211- 240	82.	91.	-9.	-2029.	0.890
241- 270	54.	72.	-18.	-4599.	4.500
271- 300	52.	58.	-6.	-1713.	0.621
301- 330	53.	48.	5.	1577.	0.521
331- 360	36.	39.	-3.	-1036.	0.231
361- 390	29.	32.	-3.	-1126.	0.281
391- 420	26.	27.	-1.	-406.	0.037
421- 450	22.	23.	-1.	-436.	0.043
451- 480	22.	19.	3.	1396.	0.474
481- 510	17.	16.	1.	496.	0.063
511- 540	10.	14.	-4.	-2102.	1.143
541- 570	19.	12.	7.	3888.	4.083
571- 600	4.	10.	-6.	-3513.	3.600
601- 700	26.	24.	2.	1301.	0.167
701- 800	21.	15.	6.	4503.	2.400
801- 900	11.	10.	1.	851.	0.100
901-1000	10.	6.	4.	3802.	2.667
1001-1100	5.	4.	1.	1050.	
1101-1200	2.	3.	-1.	-1150.	0.000
1201-1300	1.	2.	-1.	-1250.	
1301-1400	2.	1.	1.	1350.	
1401-1500	0.	1.	-1.	-1450.	
1501-1600	0.	1.	-1.	-1550.	
1601-1700	0.	0.	0.	0.	
1701-1800	1.	0.	1.	1751.	0.200
TOTAL	2441.	2439.		1731.	49.218

D.F. =22

$P < 0.001$

TOTAL EXP. LOSS / TOTAL ACT. COST = 0.5 %

Table (5.2.3)

*** 3-PARAMETER INV. GAUSSIAN DIS. ***

74/2ND QUARTER DATA

ESTIMATION BY MULTINOMIAL MAX. LIKELIHOOD METHOD :-
C= 19.46 MEW= 170.666 LAMDA= 182.839

MEAN= 151.208 S.D.= 164.887

AMOUNT £	ACTUAL CL. NO.	EXPECTED CL. NO.	ACTUAL-EXPECTED	EXPECTED LOSS	(A-E)**2/E
1- 30	351.	346.	5.	78.	0.072
31- 60	380.	424.	-44.	-2002.	4.566
61- 90	382.	345.	37.	2793.	3.968
91- 120	295.	263.	32.	3376.	3.894
121- 150	211.	200.	11.	1490.	0.605
151- 180	142.	153.	-11.	-1820.	0.791
181- 210	114.	119.	-5.	-978.	0.210
211- 240	101.	94.	7.	1578.	0.521
241- 270	57.	75.	-18.	-4599.	4.320
271- 300	51.	60.	-9.	-2569.	1.350
301- 330	39.	49.	-10.	-3155.	2.041
331- 360	36.	40.	-4.	-1382.	0.400
361- 390	25.	33.	-8.	-3004.	1.939
391- 420	24.	27.	-3.	-1216.	0.333
421- 450	27.	23.	4.	1742.	0.696
451- 480	18.	19.	-1.	-466.	0.053
481- 510	21.	16.	5.	2477.	1.562
511- 540	17.	14.	3.	1576.	0.643
541- 570	12.	11.	1.	555.	0.091
571- 600	11.	10.	1.	585.	0.100
601- 700	30.	23.	7.	4553.	2.130
701- 800	13.	14.	-1.	-751.	0.071
801- 900	11.	9.	2.	1701.	0.444
901-1000	4.	5.	-1.	-951.	0.200
1001-1100	7.	3.	4.	4202.	
1101-1200	0.	2.	-2.	-2301.	
1201-1300	1.	1.	0.	0.	
1301-1400	1.	1.	0.	0.	
1401-1500	0.	1.	-1.	-1450.	
1501-1600	1.	0.	1.	1550.	
1601-1700	0.	0.	0.	0.	
1701-1800	0.	0.	0.	0.	
1801-1900	1.	0.	1.	1850.	1.125
TOTAL	2383.	2380.		3466.	32.127

D.F. = 21

P = 0.05

TOTAL EXP. LOSS / TOTAL ACT. COST = 1.0 %

Table (5.2.4)

*** 3-PARAMETER INV. GAUSSIAN DIS. ***

74/3RD QUARTER DATA

ESTIMATION BY MULTINOMIAL MAX. LIKELIHOOD METHOD :-

C= 20.92 MEW= 186.695 LAMDA= 199.955

MEAN= 165.775 S.D.= 180.398

AMOUNT £	ACTUAL CL. NO.	EXPECTED CL. NO.	ACTUAL-EXPECTED	EXPECTED LOSS	(A-E)**2/E
1- 30	362.	356.	6.	93.	0.101
31- 60	427.	462.	-35.	-1592.	2.652
61- 90	383.	393.	-10.	-755.	0.254
91- 120	356.	308.	48.	5064.	7.481
121- 150	283.	239.	44.	5962.	8.100
151- 180	194.	187.	7.	1158.	0.262
181- 210	137.	147.	-10.	-1955.	0.680
211- 240	97.	118.	-21.	-4735.	3.737
241- 270	86.	95.	-9.	-2299.	0.853
271- 300	71.	77.	-6.	-1713.	0.468
301- 330	64.	64.	0.	0.	0.000
331- 360	45.	53.	-8.	-2764.	1.208
361- 390	44.	44.	0.	0.	0.000
391- 420	25.	37.	-12.	-4866.	3.892
421- 450	26.	31.	-5.	-2177.	0.806
451- 480	22.	26.	-4.	-1862.	0.615
481- 510	25.	22.	3.	1486.	0.409
511- 540	14.	19.	-5.	-2627.	1.316
541- 570	14.	16.	-2.	-1111.	0.250
571- 600	17.	14.	3.	1756.	0.643
601- 700	32.	34.	-2.	-1301.	0.118
701- 800	34.	21.	13.	9756.	8.048
801- 900	17.	13.	4.	3402.	1.231
901-1000	4.	9.	-5.	-4752.	2.778
1001-1100	9.	6.	3.	3151.	1.500
1101-1200	4.	4.	0.	0.	0.000
1201-1300	0.	2.	-2.	-2501.	
1301-1400	1.	2.	-1.	-1350.	
1401-1500	3.	1.	2.	2901.	
1501-1600	0.	1.	-1.	-1550.	
1601-1700	1.	1.	0.	0.	
1701-1800	0.	0.	0.	0.	
1801-1900	0.	0.	0.	0.	
1901-2000	0.	0.	0.	0.	
2001-2100	0.	0.	0.	0.	
2101-2200	1.	0.	1.	2150.	
2201-2300	0.	0.	0.	0.	
2301-2400	0.	0.	0.	0.	
2401-2500	0.	0.	0.	0.	
2501-2600	1.	0.	1.	2550.	0.000
TOTAL	2799.	2802.		-481.	47.401

D.F. = 25

P = 0.002

TOTAL EXP. LOSS / TOTAL ACT. COST = -0.1 %

Table (5.2.5)

*** 3-PARAMETER INV. GAUSSIAN DIS. ***

74/4TH QUARTER DATA

ESTIMATION BY MULTINOMIAL MAX. LIKELIHOOD METHOD :-
C= 23.99 MEW= 197.730 LAMDA= 211.827

MEAN= 173.740 S.D.= 191.038

AMOUNT £	ACTUAL CL. NO.	EXPECTED CL. NO.	ACTUAL-EXPECTED	EXPECTED LOSS	(A-E)**2/E
1- 30	394.	390.	4.	62.	0.041
31- 60	452.	479.	-27.	-1228.	1.522
61- 90	426.	413.	13.	982.	0.409
91- 120	348.	329.	19.	2004.	1.097
121- 150	272.	259.	13.	1761.	0.653
151- 180	219.	205.	14.	2317.	0.956
181- 210	154.	164.	-10.	-1955.	0.610
211- 240	124.	132.	-8.	-1804.	0.485
241- 270	105.	107.	-2.	-511.	0.037
271- 300	78.	88.	-10.	-2855.	1.136
301- 330	75.	73.	2.	631.	0.055
331- 360	58.	61.	-3.	-1036.	0.148
361- 390	55.	51.	4.	1502.	0.314
391- 420	29.	43.	-14.	-5677.	4.558
421- 450	43.	36.	7.	3048.	1.361
451- 480	24.	31.	-7.	-3258.	1.581
481- 510	22.	26.	-4.	-1982.	0.615
511- 540	24.	23.	1.	525.	0.043
541- 570	19.	19.	0.	0.	0.000
571- 600	14.	17.	-3.	-1756.	0.529
601- 700	42.	41.	1.	651.	0.024
701- 800	28.	26.	2.	1501.	0.154
801- 900	20.	17.	3.	2551.	0.529
901-1000	17.	11.	6.	5703.	3.273
1001-1100	8.	7.	1.	1050.	0.143
1101-1200	5.	5.	0.	0.	0.000
1201-1300	1.	3.	-2.	-2501.	
1301-1400	5.	2.	3.	4051.	0.200
1401-1500	0.	2.	-2.	-2901.	
1501-1600	1.	1.	0.	0.	
1601-1700	0.	1.	-1.	-1650.	
1701-1800	0.	1.	-1.	-1751.	
1801-1900	0.	0.	0.	0.	
1901-2000	0.	0.	0.	0.	
2001-2100	1.	0.	1.	2050.	
2101-2200	0.	0.	0.	0.	
2201-2300	0.	0.	0.	0.	
2301-2400	1.	0.	1.	2350.	0.800
TOTAL	3064.	3063.		1875.	21.274

D.F. = 24

$\frac{\text{TOTAL EXP. LOSS}}{\text{TOTAL ACT. COST}}$ 0.4 %

P > 0.10

ESTIMATION BY MULTINOMIAL MAX. LIKELIHOOD METHOD :-
C= 26.28 MEW= 199.227 LAMDA= 230.747

MEAN= 172.943 S.D.= 185.120

AMOUNT £	ACTUAL CL. NO.	EXPECTED CL. NO.	ACTUAL-EXPECTED	EXPECTED LOSS	(A-E)**2/E
1- 30	324.	322.	2.	31.	0.012
31- 60	387.	399.	-12.	-546.	0.361
61- 90	345.	351.	-6.	-453.	0.103
91- 120	289.	284.	5.	528.	0.088
121- 150	253.	225.	28.	3794.	3.484
151- 180	187.	179.	8.	1324.	0.358
181- 210	138.	143.	-5.	-978.	0.175
211- 240	114.	115.	-1.	-225.	0.009
241- 270	93.	94.	-1.	-256.	0.011
271- 300	67.	77.	-10.	-2855.	1.299
301- 330	63.	63.	0.	0.	0.000
331- 360	44.	52.	-8.	-2764.	1.231
361- 390	44.	44.	0.	0.	0.000
391- 420	35.	37.	-2.	-811.	0.108
421- 450	25.	31.	-6.	-2613.	1.161
451- 480	26.	26.	0.	0.	0.000
481- 510	18.	22.	-4.	-1982.	0.727
511- 540	18.	19.	-1.	-525.	0.053
541- 570	22.	16.	6.	3333.	2.250
571- 600	17.	14.	3.	1756.	0.643
601- 700	39.	34.	5.	3252.	0.735
701- 800	19.	21.	-2.	-1501.	0.190
801- 900	18.	13.	5.	4252.	1.923
901-1000	12.	9.	3.	2851.	1.000
1001-1100	3.	6.	-3.	-3151.	1.500
1101-1200	1.	4.	-3.	-3451.	
1201-1300	1.	2.	-1.	-1250.	2.667
1301-1400	0.	2.	-2.	-2701.	
1401-1500	1.	1.	0.	0.	
1501-1600	0.	1.	-1.	-1550.	
1601-1700	1.	1.	0.	0.	
1701-1800	0.	0.	0.	0.	
1801-1900	0.	0.	0.	0.	
1901-2000	0.	0.	0.	0.	
2001-2100	1.	0.	1.	2050.	
2101-2200	1.	0.	1.	2150.	
2201-2300	0.	0.	0.	0.	
2301-2400	0.	0.	0.	0.	
2401-2500	0.	0.	0.	0.	
2501-2600	0.	0.	0.	0.	
2601-2700	0.	0.	0.	0.	
2701-2800	0.	0.	0.	0.	
2801-2900	0.	0.	0.	0.	
2901-3000	0.	0.	0.	0.	
3001-3100	0.	0.	0.	0.	
3101-3200	0.	0.	0.	0.	
3201-3300	0.	0.	0.	0.	
3301-3400	0.	0.	0.	0.	
3401-3500	0.	0.	0.	0.	
3501-3600	1.	0.	1.	3550.	0.000
TOTAL	2607.	2607.		1260.	20.082

D.F. = 23

P > 0.10

TOTAL EXP. LOSS / TOTAL ACT. COST = 0.3 %

Table (5.2.7)

*** 3-PARAMETER INV. GAUSSIAN DIS. ***

75/2ND QUARTER DATA

ESTIMATION BY MULTINOMIAL MAX. LIKELIHOOD METHOD :-

C= 21.49 MEW= 201.211 LAMDA= 203.077

MEAN= 179.717 S.D.= 200.285

AMOUNT £	ACTUAL CL. NO.	EXPECTED CL. NO.	ACTUAL-EXPECTED	EXPECTED LOSS	(A-E)**2/E
1- 30	302.	299.	3.	46.	0.030
31- 60	374.	390.	-16.	-728.	0.656
61- 90	332.	336.	-4.	-302.	0.048
91- 120	277.	267.	10.	1055.	0.375
121- 150	235.	210.	25.	3387.	2.976
151- 180	187.	166.	21.	3475.	2.657
181- 210	122.	133.	-11.	-2150.	0.910
211- 240	110.	108.	2.	451.	0.037
241- 270	80.	88.	-8.	-2044.	0.727
271- 300	72.	72.	0.	0.	0.000
301- 330	47.	60.	-13.	-4101.	2.817
331- 360	39.	50.	-11.	-3800.	2.420
361- 390	40.	42.	-2.	-751.	0.095
391- 420	38.	36.	2.	811.	0.111
421- 450	29.	30.	-1.	-436.	0.033
451- 480	21.	26.	-5.	-2327.	0.962
481- 510	30.	22.	8.	3964.	2.909
511- 540	19.	19.	0.	0.	0.000
541- 570	17.	17.	0.	0.	0.000
571- 600	11.	14.	-3.	-1756.	0.643
601- 700	36.	36.	0.	0.	0.000
701- 800	22.	23.	-1.	-751.	0.043
801- 900	22.	15.	7.	5953.	3.267
901-1000	11.	10.	1.	951.	0.100
1001-1100	4.	7.	-3.	-3151.	1.286
1101-1200	3.	5.	-2.	-2301.	0.800
1201-1300	3.	3.	0.	0.	
1301-1400	6.	2.	4.	5402.	3.200
1401-1500	2.	2.	0.	0.	
1501-1600	1.	1.	0.	0.	
1601-1700	0.	1.	-1.	-1650.	
1701-1800	1.	1.	0.	0.	
1801-1900	0.	0.	0.	0.	
1901-2000	1.	0.	1.	1950.	
2001-2100	0.	0.	0.	0.	
2101-2200	0.	0.	0.	0.	
2201-2300	1.	0.	1.	2250.	0.200
TOTAL	2495.	2491.		3447.	28.101

D.F. = 24

TOTAL EXP. LOSS / TOTAL ACT. COST = 0.8 %

P > 0.10

Table (5.3.1)

*** 3-PARAMETER INV. GAUSSIAN DIS. ***

PREDICTION OF 74/4TH QUARTER CLAIMS COST
USING 73/4TH QUARTER MULT. MAXLIK. PARAMETERS :

C= 18.1 MEW= 168.034 LAMDA= 166.536
INFLATION RATE I=18.2% CALCULATED FROM :
GENERAL INDEX OF RETAIL PRICES

PREDICTION PARAMETERS ARE :- C=21.4 MEW=198.616 LAMDA=156.846
MEAN CLAIM AMOUNT= 177.22 S.D.=199.51
ACTUAL 74/4TH PARAMETERS :- C=24.0 MEW=197.730 LAMDA=211.827
MEAN CLAIM AMOUNT= 173.73 S.D.=191.04

AMOUNT £	ACT.NO.	EXP.NO.	A-E	EXP.LOSS	(A-E)**2/E
1- 30	394.	386.	8.	124.	0.166
31- 60	452.	486.	-34.	-1547.	2.379
61- 90	426.	413.	13.	982.	0.409
91- 120	348.	326.	22.	2321.	1.485
121- 150	272.	255.	17.	2304.	1.133
151- 180	219.	202.	17.	2813.	1.431
181- 210	154.	161.	-7.	-1368.	0.304
211- 240	124.	130.	-6.	-1353.	0.277
241- 270	105.	106.	-1.	-256.	0.009
271- 300	78.	87.	-9.	-2569.	0.931
301- 330	75.	72.	3.	947.	0.125
331- 360	58.	60.	-2.	-691.	0.067
361- 390	55.	51.	4.	1502.	0.314
391- 420	29.	43.	-14.	-5677.	4.558
421- 450	43.	37.	6.	2613.	0.973
451- 480	24.	31.	-7.	-3258.	1.581
481- 510	22.	27.	-5.	-2477.	0.926
511- 540	24.	23.	1.	525.	0.043
541- 570	19.	20.	-1.	-555.	0.050
571- 600	14.	17.	-3.	-1756.	0.529
601- 700	42.	43.	-1.	-651.	0.023
701- 800	28.	28.	0.	0.	0.000
801- 900	20.	18.	2.	1701.	0.222
901-1000	17.	12.	5.	4752.	2.083
1001-1100	8.	8.	0.	0.	0.000
1101-1200	5.	6.	-1.	-1150.	0.167
1201-1300	1.	4.	-3.	-3751.	
1301-1400	5.	3.	2.	2701.	
1401-1500	0.	2.	-2.	-2901.	
1501-1600	1.	1.	0.	0.	
1601-1700	0.	1.	-1.	-1650.	
1701-1800	0.	1.	-1.	-1751.	
1801-1900	0.	1.	-1.	-1850.	
1901-2000	0.	0.	0.	0.	
2001-2100	1.	0.	1.	2050.	
2101-2200	0.	0.	0.	0.	
2201-2300	0.	0.	0.	0.	
2301-2400	1.	0.	1.	2350.	
TOTAL	3064	3061		-7528.	20.185

CHI SQ. STAT. = 21.825 , D.F.= 27 , P>0.10

TOTAL ACTUAL COST = 533707.
TOTAL EXPECTED COST = 541236.

KOL - SMIRNOV D= 0.015
P>0.20

TOTAL EXP. LOSS / TOTAL ACT. COST = -1.41 %

*** 3-PARAMETER INV. GAUSSIAN DIS. ***

Table (5.3.2)

PREDICTION OF 75/1ST QUARTER CLAIMS COST
USING 74/1ST QUARTER MULT. MAXLIK. PARAMETERS :

C= 15.1 MEW= 164.632 LAMDA= 153.065
INFLATION RATE I=20.3% CALCULATED FROM :
GENERAL INDEX OF RETAIL PRICES

PREDICTION PARAMETERS ARE :- C=18.2 MEW=198.052 LAMDA=184.137
MEAN CLAIM AMOUNT= 179.89 S.D.=205.40
ACTUAL 75/1ST PARAMETERS :- C=27.6 MEW=199.260 LAMDA=239.515
MEAN CLAIM AMOUNT= 171.66 S.D.=181.75

AMOUNT £	ACT.NO.	EXP.NO.	A-E	EXP.LOSS	(A-E)**2/E
1- 30	324.	314.	10.	155.	0.318
31- 60	387.	423.	-36.	-1638.	3.064
61- 90	345.	356.	-11.	-831.	0.340
91- 120	289.	278.	11.	1160.	0.435
121- 150	253.	216.	37.	5013.	6.338
151- 180	187.	170.	17.	2813.	1.700
181- 210	138.	136.	2.	391.	0.029
211- 240	114.	109.	5.	1127.	0.229
241- 270	93.	89.	4.	1022.	0.180
271- 300	67.	74.	-7.	-1998.	0.662
301- 330	63.	61.	2.	631.	0.066
331- 360	44.	51.	-7.	-2418.	0.961
361- 390	44.	43.	1.	376.	0.023
391- 420	35.	37.	-2.	-811.	0.108
421- 450	25.	31.	-6.	-2613.	1.161
451- 480	26.	27.	-1.	-466.	0.037
481- 510	18.	23.	-5.	-2477.	1.087
511- 540	18.	20.	-2.	-1051.	0.200
541- 570	22.	17.	5.	2777.	1.471
571- 600	17.	15.	2.	1171.	0.267
601- 700	39.	38.	1.	651.	0.026
701- 800	19.	25.	-6.	-4503.	1.440
801- 900	18.	16.	2.	1701.	0.250
901-1000	12.	11.	1.	951.	0.091
1001-1100	3.	8.	-5.	-5252.	3.125
1101-1200	1.	5.	-4.	-4602.	3.200
1201-1300	1.	4.	-3.	-3751.	
1301-1400	0.	3.	-3.	-4051.	
1401-1500	1.	2.	-1.	-1450.	
1501-1600	0.	1.	-1.	-1550.	
1601-1700	1.	1.	0.	0.	
1701-1800	0.	1.	-1.	-1751.	
1801-1900	0.	1.	-1.	-1850.	
1901-2000	0.	0.	0.	0.	
2001-2100	1.	0.	1.	2050.	
2101-2200	1.	0.	1.	2150.	
2201-2300	0.	0.	0.	0.	
2301-2400	0.	0.	0.	0.	
2401-2500	0.	0.	0.	0.	
2501-2600	0.	0.	0.	0.	
2601-2700	0.	0.	0.	0.	
2701-2800	0.	0.	0.	0.	
2801-2900	0.	0.	0.	0.	
2901-3000	0.	0.	0.	0.	
3001-3100	0.	0.	0.	0.	
3101-3200	0.	0.	0.	0.	
3201-3300	0.	0.	0.	0.	
3301-3400	0.	0.	0.	0.	
3401-3500	0.	0.	0.	0.	
3501-3600	1.	0.	1.	3550.	
TOTAL	2607	2606		-15375.	26.809

Table (5.3.2) - continued

CHI SQ. STAT. =32.619 , D.F.=27 , P> 0.10

TOTAL ACTUAL COST = 452059.
TOTAL EXPECTED COST = 467433.

KOL - SMIRNOV D = 0.15
P> 0.20

TOTAL EXP. LOSS
---------------- = -3.40 %
TOTAL ACT. COST

Table (5.3.3)

*** 3-PARAMETER INV. GAUSSIAN DIS. ***

PREDICTION OF 75/2ND QUARTER CLAIMS COST
USING 74/2ND QUARTER MULT. MAXLIK. PARAMETERS :

C= 19.5 MEW= 170.666 LAMDA= 182.839
INFLATION RATE I=24.3% CALCULATED FROM :
GENERAL INDEX OF RETAIL PRICES .

PREDICTION PARAMETERS ARE :- C=24.2 MEW=212.138 LAMDA=227.269
MEAN CLAIM AMOUNT= 187.90 S.D.=204.95
ACTUAL 75/2ND PARAMETERS :- C=21.5 MEW=201.211 LAMDA=203.077
MEAN CLAIM AMOUNT= 179.71 S.D.=200.28

AMOUNT £	ACT.NO.	EXP.NO.	A-E	EXP.LOSS	(A-E)**2/E
1- 30	302.	273.	29.	450.	3.081
31- 60	374.	366.	8.	364.	0.175
61- 90	332.	328.	4.	302.	0.049
91- 120	277.	267.	10.	1055.	0.375
121- 150	235.	214.	21.	2845.	2.061
151- 180	187.	171.	16.	2648.	1.497
181- 210	122.	138.	-16.	-3128.	1.855
211- 240	110.	113.	-3.	-677.	0.080
241- 270	80.	93.	-13.	-3321.	1.817
271- 300	72.	77.	-5.	-1427.	0.325
301- 330	47.	64.	-17.	-5363.	4.516
331- 360	39.	54.	-15.	-5182.	4.167
361- 390	40.	45.	-5.	-1877.	0.556
391- 420	38.	38.	0.	0.	0.000
421- 450	29.	33.	-4.	-1742.	0.485
451- 480	21.	28.	-7.	-3258.	1.750
481- 510	30.	24.	6.	2973.	1.500
511- 540	19.	21.	-2.	-1051.	0.190
541- 570	17.	18.	-1.	-555.	0.056
571- 600	11.	16.	-5.	-2927.	1.562
601- 700	36.	39.	-3.	-1951.	0.231
701- 800	22.	25.	-3.	-2251.	0.360
801- 900	22.	16.	6.	5103.	2.250
901-1000	11.	11.	0.	0.	0.000
1001-1100	4.	7.	-3.	-3151.	1.286
1101-1200	3.	5.	-2.	-2301.	0.800
1201-1300	3.	4.	-1.	-1250.	
1301-1400	6.	2.	4.	5402.	
1401-1500	2.	2.	0.	0.	
1501-1600	1.	1.	0.	0.	
1601-1700	0.	1.	-1.	-1650.	
1701-1800	1.	1.	0.	0.	
1801-1900	0.	0.	0.	0.	
1901-2000	1.	0.	1.	1950.	
2001-2100	0.	0.	0.	0.	
2101-2200	0.	0.	0.	0.	
2201-2300	1.	0.	1.	2250.	
TOTAL	2495	2495		-17725.	31.021

CHI SQ. STAT. = 32.721 , D.F.= 27 , P> 0.10

TOTAL ACTUAL COST = 449308.
TOTAL EXPECTED COST = 467033.

KOL - SMIRNOV D = 0.035
P <0.01

TOTAL EXP. LOSS
--------------- = -3.94 %
TOTAL ACT. COST

CHAPTER 6

The Pareto Distribution

6.1 Introduction

The Pareto distribution, which is positively skewed and has a long tail, was originally proposed by Vilfredo Pareto as a model for the distribution of incomes over a population. It has also been successfully used to model empirical distributions of claim amounts in general insurance and, in particular, in fire insurance. Benckert and Sternberg (1957), Anderson (1971) and Benckert and Jung (1974) have used this distribution as a model for fire insurance claim amounts.

In this chapter we will study this model and will consider its application to the accidental damage claim amounts. The Pareto distributions of the first and second kind will be initially defined and some of their properties will be studied. A graphical test will be discussed for the purpose of determining whether a given sample is likely to be from a Pareto population. We will then apply this test to our accidental damage data and will, in the event, show that the Pareto distribution is not a satisfactory model. The estimation problem is next dealt with and the effects of inflation on the parameters of the model will be discussed.

6.2 Definition

A random variable X is said to have a Pareto distribution of the first kind if its distribution function, denoted by P1(x; A,B), is of the form :

$$P1(x;\, A,B) \;=\; 1 - \left(\frac{B}{x}\right)^{A} \qquad A,B > 0,\; x \geq B \qquad (6.2\text{-}1)$$

This is a special form of Pearson Type V1 distribution. The probability density function (p.d.f.) of X is

$$f_{P1}(x;\, A,B) \;=\; \frac{A}{B}\left(\frac{B}{x}\right)^{A+1} \qquad A,B > 0,\; x \geq B \qquad (6.2\text{-}2)$$

where A and B are the shape and scale parameters respectively.

A random variable X is said to have a Pareto distribution of the second kind if its distribution function, denoted by P2(x; C,A,B), is of the form :

$$P2(x;\ C,A,B) = 1 - \left(\frac{B}{x+C}\right)^{A} \qquad A,B > 0,\ x \geq B - C \qquad (6.2\text{-}3)$$

This is sometimes called the Lomax distribution and is also a Pearson Type Vl. It differs from P1(x; A,B) by having the location parameter C which acts as a threshold below which the values of X are not realized. The p.d.f. of X is given by

$$f_{P2}(x;\ C,A,B) = \frac{A}{B}\left(\frac{B}{x+C}\right)^{A+1} \qquad A,B > 0,\ x \geq B - C \qquad (6.2\text{-}4)$$

6.3 Properties of the Pareto Distribution

In this section we will mention some of the properties of the Pareto distribution of the first kind. For more details reference should be made to, for instance, Johnson and Kotz (1970).

For the distribution P1(x; A,B) it can be shown that the rth moment of X about zero is

$$E(X^r) = \frac{AB^r}{A-r} \quad (\text{if } A > r) \qquad (6.3\text{-}1)$$

Hence the mean of the distribution of X is

$$E(X) = \frac{AB}{A-1} \qquad (\text{if } A > 1) \qquad (6.3\text{-}2)$$

The variance of X can be shown to be

$$\text{var}(X) = \frac{AB^2}{(A-1)^2(A-2)} \qquad (\text{if } A > 2) \qquad (6.3\text{-}3)$$

Therefore the mean and variance of the distribution exist when $A > 1$ and $A > 2$ respectively.

The coefficient of variation of X is λ where

$$\lambda^2 = \frac{1}{A(A-2)} \qquad (\text{if } A > 2) \qquad (6.3\text{-}4)$$

which is a function of A, the shape parameter, only. As the mean of the distribution does not always exist, the single mode of the distribution which is located at B may be used in its place.

The median of the distribution is at

$$x_{median} = B(2)^{1/A} \qquad (6.3\text{-}5)$$

In general the quantile of order q is x_q where

$$x_q = B(1 - q)^{-1/A} \qquad 0 < q < 1 \qquad (6.3\text{-}6)$$

The properties of the Pareto distribution of the second kind can be obtained from those of the first kind since the former distribution is derived from the latter by the transformation $x \to x + C$. This is a translation which leaves the shape of the frequency curve of $P1(x; A,B)$ unchanged and only shifts it by an amount $-C$ along the x-axis. The measures of central tendency are, therefore, decreased by C while the moments about the mean and the measures of dispersion remain the same as for the first kind.

6.4 The Graphical Test for the Pareto Distribution

It is possible to test graphically if a given sample is likely to be from a Pareto population. For the distribution of the first kind given by (6.2-1) we can show that

$$\log(1-P1) = -A \log x + A \log B \qquad (6.4\text{-}1)$$

where $P1 = P1(x; A,B)$.

Therefore the locus of the points $(\log x, \log(1 - P1))$ is a straight line whose gradient and intercept are $-A$ and $A \log B$ respectively. Let us assume that we are given a sample of n independent random observations from a Pareto distribution. If we define the sample empirical distribution function as

$$F(x) = \text{the proportion of observations} \leq x \qquad (6.4\text{-}2)$$

then we can use F(x) as an estimate of $P1(x; A,B)$ at point x. Hence if we plot the points $(\log x, \log(1 - F(x)))$ on a rectangular system of co-ordinate axes, or plot$(x,1-F(x))$ on logarithmic axes, we should find that the points lie approximately on a straight line. If the points do not appear to lie approximately on a straight line, we conclude that the sample is not from a Pareto population of the first kind.

For the Pareto distribution of the second kind, equation (6.4-1) becomes

$$\log(1 - P2) = -A\log(x + C) + A\log B \qquad (6.4\text{-}3)$$

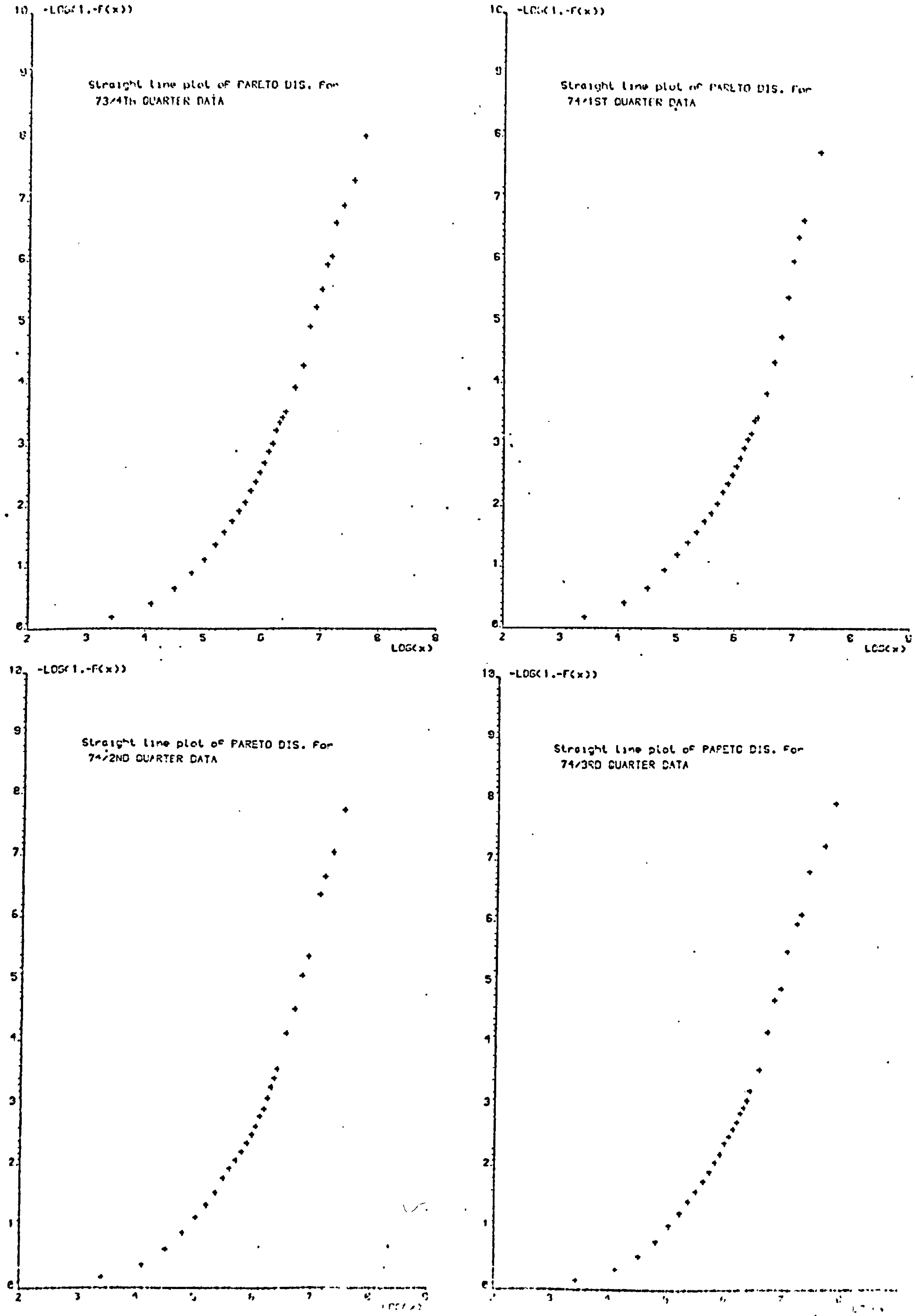
where $P2 = P2(x; C,A,B)$

Hence the locus of the points $(\log(x + C), \log(1 - P2))$ is a straight line. When C is unknown the first step is to assume $C = 0$ and to plot the resulting points (as in the case of the 3-parameter lognormal and Weibull distributions). If the sample is from a population with C not equal to zero, judging by the curvature of the resulting curve, we should choose another value for C such that the points are rectified to lie approximately on a straight line. If such a C exists, we conclude that the sample is from a Pareto population of the second kind and may use this value of C as a graphical estimate of the unknown parameter C. This technique was described, in detail, in sections 3.11 and 4.4 for the 3-parameter lognormal and Weibull distributions respectively.

6.5 The Pareto Graphical Test on AD Data

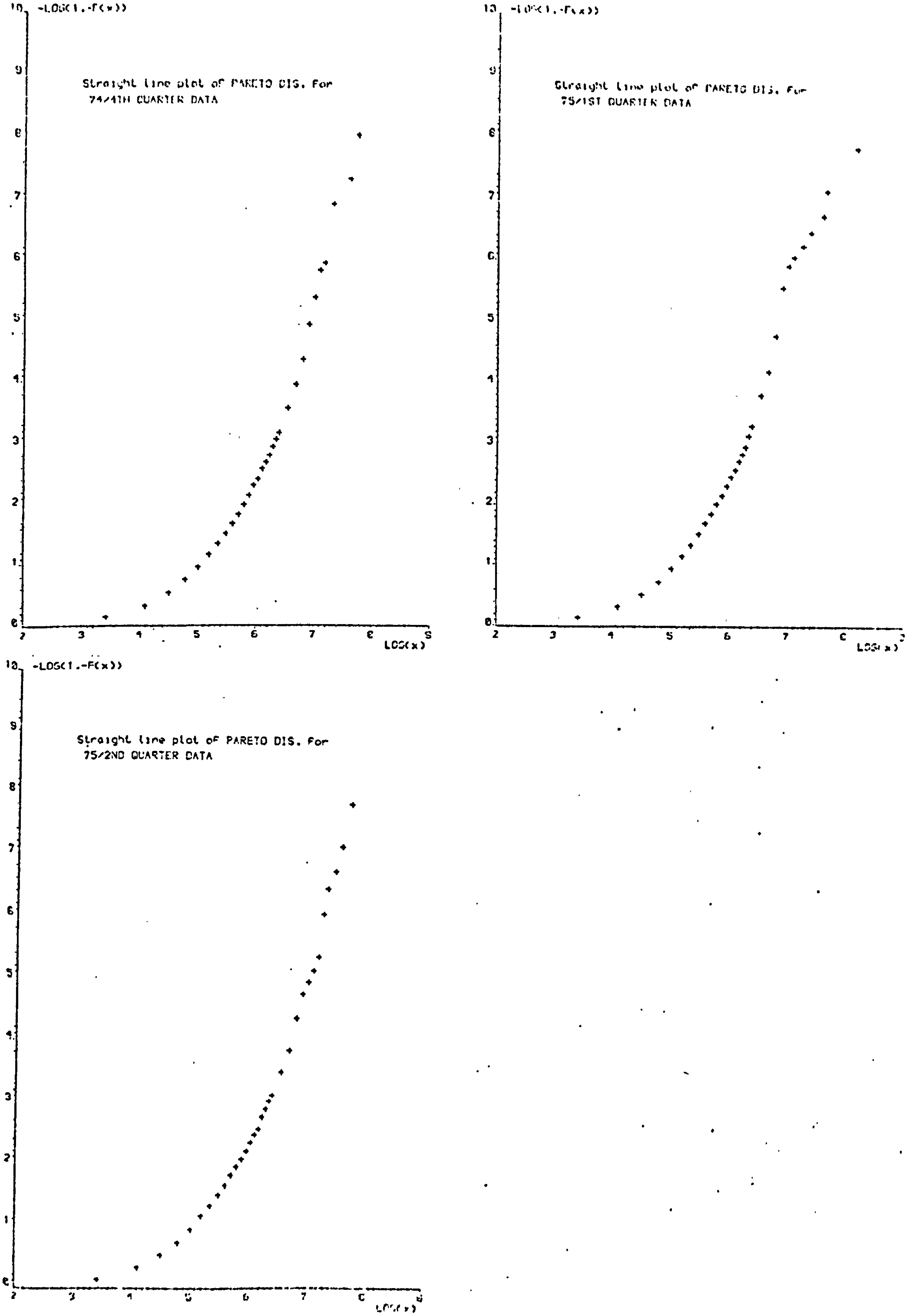
To test if the Pareto distribution can be used as a model for the accidental damage (AD) data we wrote computer program P22. From a given sample of data, the program calculates the empirical distribution function F(x) at each sample value x and then plots the points (log x, log(1 - F(x)). For the seven samples of AD data which were given in tables (1.1) to (1.7) the graphs are presented in figures (6.1-a) and (6.1-b). It is observed that for each sample the points lie approximately on a curve which is far from a straight line. The points in the upper tail of the sample values lie almost on a straight line, but the curvature in the lower tail values is so marked that even an addition of C to the values of x (and hence a Pareto distribution of the second kind) does not seem likely to rectify the points enough to lie on a straight line. We, therefore, conclude that the Pareto distribution is not a possible model for the distribution of our AD claim amounts. However, as indicated by the graphs, such a model may very well fit the distribution of the larger claims in the sample.

Figure (6.1-a)



-LOG(1.-F(x))
Straight line plot of PARETO DIS. For
73/4Th QUARTER DATA
LOG(x)
-LOG(1.-F(x))
Straight line plot of PARETO DIS. For
74/1ST QUARTER DATA
-LOG(1.-F(x))
Straight line plot of PARETO DIS. For
74/2ND QUARTER DATA
-LOG(1.-F(x))
Straight line plot of PARETO DIS. For
74/3RD QUARTER DATA

Figure (6.1-b)



-LOG(1.-F(x))
Straight line plot of PARETO DIS. For
74/4TH QUARTER DATA
LOG(x)
-LOG(1.-F(x))
Straight line plot of PARETO DIS. For
75/1ST QUARTER DATA
LOG(x)
-LOG(1.-F(x))
Straight line plot of PARETO DIS. For
75/2ND QUARTER DATA
LOG(x)

6.6 Estimation of the Parameters of the Pareto Distribution

Although the results of the previous section were not encouraging, for the sake of completeness of the present work, we deal with the problem of estimation from grouped data. Several methods of estimation such as the methods of moments, quantiles, graphical, least squares and maximum likelihood have been studied by various authors and a review is provided in Johnson and Kotz (1970). Some of these methods are directly applicable for estimation from grouped data while others (like the moments, quantiles and maximum likelihood) can be modified for this purpose in the same way as we described for the 2-parameter lognormal distribution. In this section, we explain briefly the method of least squares and then concentrate on the multinomial maximum likelihood (MML) method.

The least squares method is based on the relationship (6.4-1). Instead of fitting a straight line to the points $(\log x, \log(1-F(x)))$ by eye, we use the least squares technique to find the unknown coefficients of the line. For P1(x; A,B) this is straightforward. For P2(x; C,A,B) when C is unknown, in order to avoid solving a system of non-linear equations in the parameters, we can use the computing technique which we suggested for the 3-parameter lognormal and Weibull distributions in sections 3.12.1 and 4.6.1 respectively.

The MML is a very suitable method of parameter estimation from grouped data and it produces efficient estimators. Estimation of the parameters of the Pareto distribution of the second kind, P2(x; C,A,B), will be dealt with here but the same argument applies for P1(x; A,B) by putting C = 0 in the following exposition.

Let us assume that we have a sample of grouped data where n independent random observations on a random variable X (in our case the claim amount) have been grouped according to their size into k mutually exclusive intervals. Further, let n_i be the number of observations (claims) in the class interval (x_{i-1}, x_i) for i = 1, 2, ... , k , such that

$$n = \sum_{i=1}^{k} n_i$$

If X is distributed as P2(x; C,A,B), given by (6.2-3) , i.e.

$$P2(x;\, C,A,B) = 1 - \left(\frac{B}{x+C}\right)^A \qquad (6.6\text{-}1)$$

then let p_i be the probability that an observation (claim) occurs in the interval (x_{i-1}, x_i), i.e.

$$p_i = P2(x_i;\, C,A,B) - P2(x_{i-1};\, C,A,B) \qquad (6.6\text{-}2)$$

for i = 1, 2, ... , k

from (6.6-1) it follows that

$$p_i = \left(\frac{B}{x_{i-1}+C}\right)^A - \left(\frac{B}{x_i+C}\right)^A \qquad (6.6\text{-}3)$$

therefore the sample likelihood function will be proportional to L where

$$L = \prod_{i=1}^{k} p_i^{\,n_i} \qquad (6.6\text{-}4)$$

and the loglikelihood function will be

$$\log L = \sum_{i=1}^{k} n_i \log p_i \qquad (6.6\text{-}5)$$

from (6.6-3) it follows that

$$\log L = nA\log B + \sum_{i=1}^{k} n_i \log\left[\left(\frac{1}{x_{i-1}+C}\right)^A - \left(\frac{1}{x_i+C}\right)^A\right] \qquad (6.6\text{-}6)$$

The maximum likelihood estimates, $(\hat{C}, \hat{A}, \hat{B})$, of the parameters are obtained by maximizing log L simultaneously with respect to A, B and C. To avoid solving non-linear equations (which is a laborious task when k, the number of intervals, is large) we suggest the computing technique described in section 3.7.7 for maximising log L with respect to A, B and C. To start the iteration process we can use a set of estimates found by one of the simpler, but less accurate, methods. For instance, the moments, quantiles or graphical estimates (or their combination) may be used. The set $(\hat{C}, \hat{A}, \hat{B})$ which maximizes log L will be the required MML estimates.

The graphical tests showed earlier that the Pareto distribution does not represent our AD data. Therefore, we will not attempt to fit this model to the complete histogram of the data. However, figures (6.1-a) and (6.1-b) showed that the points corresponding to the upper tail of the distribution of claim amounts lie approximately on a straight line. In chapter 7 we will fit this model to the upper tail of the AD histograms and will compare its fit with that of the truncated lognormal distribution.

6.7 The Effects of Inflation on the Parameters

Following the argument of chapter 3 about the effects of inflation on the parameters of a model, it is considered important to study these effects theoretically for the Pareto distribution. We will need to allow for such effects when predicting the future distribution of claim amounts for a class of general insurance business in which the Pareto model has been found to represent the distribution of claim amounts.

Let us again assume that the effect of inflation is to increase a claim of amount X to $U = X(1 + i)$ over a period where i is the effective rate of inflation for that period. If X is distributed as a Pareto distribution of the second kind with p.d.f. $f_{P2}(x; C,A,B)$ then by a simple transformation of variables we can show that U is distributed as a Pareto distribution of the second kind with p.d.f. $f_{P2}(u;\ rC,A,\ rB)$ where $r = 1 + i$. Therefore inflation affects the parameters C and B but leaves the shape parameter A unaltered. The results for the Pareto distribution of the first kind can be deduced from the above by putting $C = 0$.

6.8 Conclusions

The Pareto distribution was studied in this chapter as a model for the distribution of claim amounts in general insurance and, in particular, for the accidental damage claim amounts.
The Pareto distribution, as well as other positively skewed distributions like the lognormal, has been shown to represent empirical distributions of many socio-economic and other naturally occurring quantities. In section 6.1 reference was made to the uses of Pareto distribution in fire insurance. A graphical procedure was described which may be used to test if a sample is likely to be from a Pareto population. Our AD data proved not to be from such a population. However, points corresponding to the upper tails of the histograms appeared to lie approximately on a straight line as shown in figures (6.1-a) and (6.1-b). This indicates that the tail of the distribution of claim amounts may well be modelled by the Pareto distribution. As a matter of fact, Johnson and Kotz (1970) mention that lognormal distribution fits well the distribution of income over a large part of the income range but diverges markedly at the extremities. On the other hand, it has been observed that the fit of the Pareto curve is rather good at the extremities of the income range but the fit over the whole range is often rather poor. If we consider that income and claim amount are to some extent correlated as represented by the sum at risk, then the above conclusion from the graphical tests about the tail of the distribution of our AD claim amounts becomes plausible.

We proposed the MML method of parameter estimation from grouped data but did not apply it to the AD data in this chapter. In chapter 7 this method will be used to fit the Pareto model to the tails of our AD samples.

The effects of inflation on the parameters were studied and it was shown that both C and B are increased while A, the shape parameter, remains unchanged.

CHAPTER 7

Truncated Lognormal Distribution

7.1 Introduction

Often, in practice, data only of claim amounts above a particular sum are available. For example, the excess of loss reinsurer has only access to data of claim amounts above the retention point, E. Such data, for a particular class of business, may be represented by the incomplete (or truncated) histogram in figure (7.1).

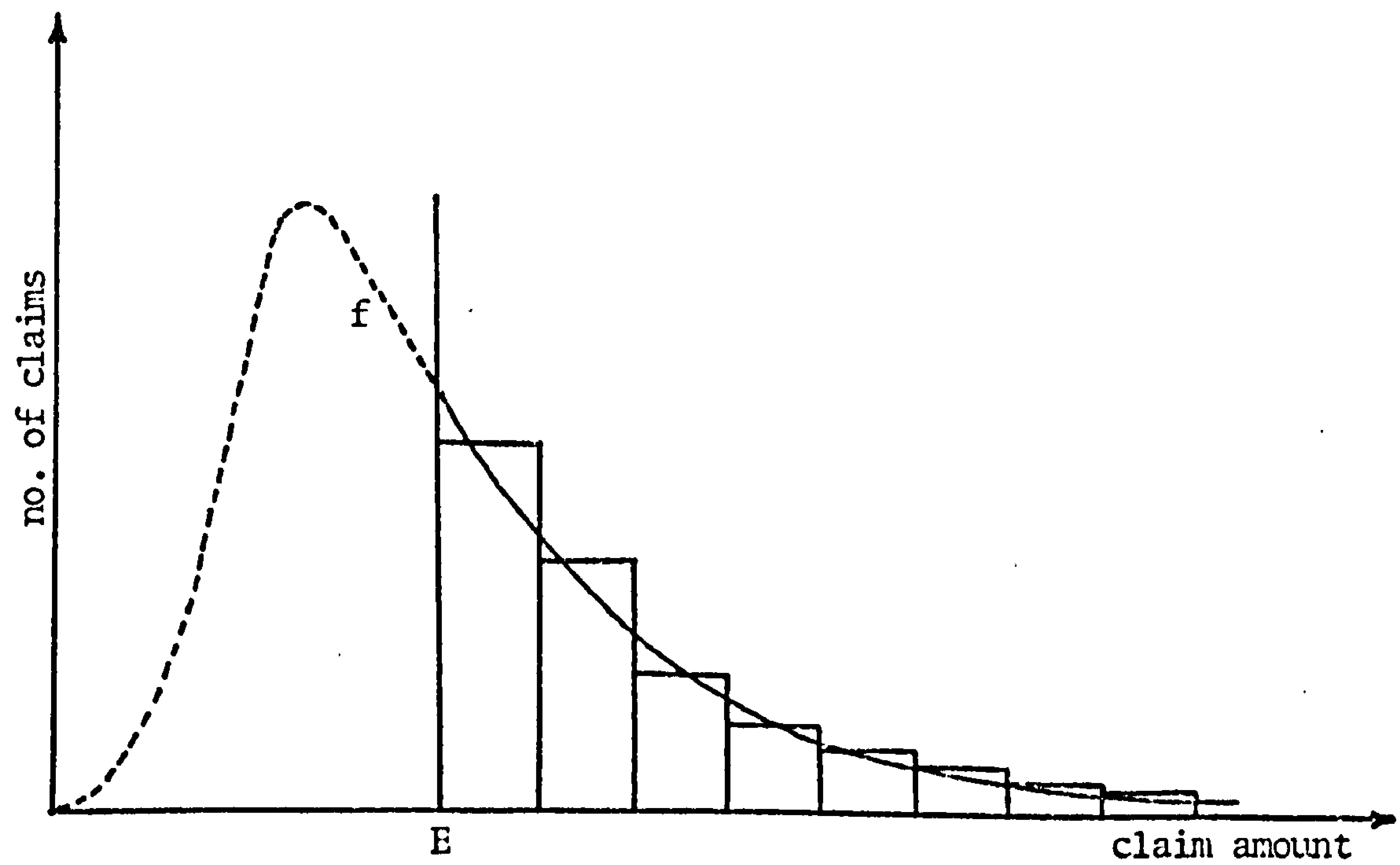


Figure (7.1) - A truncated distribution fitted to a histogram of incomplete data

As another example, the above histogram may represent claim amounts data, available to an insurance company, in respect of a class of insurance business where the policy holder pays any loss up to the 'voluntary excess' amount E. The histogram in figure (7.1) can, therefore, represent the empirical claim amount distribution when it has been singly truncated from below at the 'point of truncation' E.

If we know that a particular statistical distribution represents the complete distribution of claim amounts, then we may be able to fit a truncated form of that statistical distribution to the incomplete data as represented in figure (7.1). Essentially, we find a model with probability density function f whose tail fits the incomplete empirical distribution (see figure (7.1)). That part of the curve f which lies to the right of the truncation point E is the frequency curve of the truncated distribution. Harding (1968) considers the truncated lognormal model for the distribution of excess of loss reinsurance claim amounts of motor business.

The studies of the previous chapters of the present work showed that the lognormal model represents the distribution of our accidental damage (AD) claim amounts. The 2-parameter lognormal produced a 'better' fit (in the sense of smaller Chi-square statistics) to the actual data than the 2-parameter inverse Gaussian distribution. Therefore, in this chapter we shall consider the truncated lognormal distribution. As we are only interested in the tail of the distribution, only the 2-parameter model will be dealt with. Initially, the truncated lognormal distribution will be defined and some of its properties will be mentioned. Then the problem of estimation of the parameters from grouped data will be considered and the multinomial maximum likelihood method will be proposed. By using this method, two problems related to our AD data will then be studied ; these are discussed below :-

1 - In chapter 3 we noticed that the significant differences between the 2-parameter lognormal distribution and the actual sample values were caused by large contributions to the Chi-square statistic by one or two of the intervals in the lower tail values. In that chapter we indicated that the disagreement could be due to the inaccuracies and incompleteness

of the data in the lower tail intervals because in order to remain entitled to their substantial no-claim discounts some policy holders do not claim small amounts. As the claims in the lower tail are financially less important, it was suggested that the data should be truncated at the upper boundary of these intervals to see if a more exact fit to the rest of the data would be obtained. Therefore, in this chapter, we will fit the truncated lognormal distribution to samples of our AD data, as presented in tables (1.1) to (1.7), but truncated at £30 or £60 or £90. The results for different points of truncation will be examined and compared with those of chapter 3 for the complete samples.

2 - The larger claims form the tail of the claim amount distribution and because of their financial importance a higher degree of agreement between the model and the empirical distribution is required in this region. Therefore, it is important to find out if a truncated distribution fitted to the tail of the histogram of sample values produces a 'better' fit. For our AD data we truncate the samples at £600 and fit the truncated lognormal model to the tail of the actual sample values. The results will then be compared with the tails of the fitted lognormal distributions in chapter 3 to see if, in fact, truncation does produce a 'better' fit in this region.

The histograms of our AD samples in figures (3.5.1) to (3.5.7) indicate that the Pareto distribution with its mode at £600 is likely to fit the empirical distribution of claim amounts greater than £600. The graphical tests of chapter 6 also implied a Pareto model for the tail of our AD histograms. The Pareto distribution is much simpler to fit and develop mathematically than the truncated lognormal distribution. Therefore, to compare the performance of these two models we will, in this chapter, also fit the Pareto distribution to samples of claim amounts greater than £600.

As was mentioned previously, the excess of loss reinsurance claim amounts data are available only in the truncated form and hence the treatment of the present problem is applicable in that context.

The findings of this chapter will be summarized in the conclusion section and the tables will be presented in section 7.6.

7.2 Definition and Some Properties of the Truncated Lognormal Distribution

We say a random variable X has a truncated lognormal distribution if its distribution function, denoted by $TLN(x,E;\mu,\sigma^2)$ is of the form :

$$TLN(x,E;\mu,\sigma^2) = \frac{LN(x;\mu,\sigma^2) - LN(E;\mu,\sigma^2)}{1 - LN(E;\mu,\sigma^2)} \quad \text{for } x > E \qquad (7.2\text{-}1)$$

$$= 0 \ \ldots\ldots\ldots\ldots\ldots\ldots \quad \text{otherwise}$$

where E is the point of truncation and $LN(\cdot\,;\mu,\sigma^2)$ is the distribution function of a lognormal random variable, with parameters μ and σ^2, as defined in chapter 3 by (3.2-4). Hence the probability density function (p.d.f.) of the truncated lognormal distribution, which we denote by $f_{TLN}(x,E\,;\mu,\sigma^2)$, is :

$$f_{TLN}(x,E;\mu,\sigma^2) = \frac{f_{LN}(x;\mu,\sigma^2)}{1-LN(E;\mu,\sigma^2)} \quad \text{for } x > E \qquad (7.2\text{-}2)$$

where $f_{LN}(\cdot\,;\mu,\sigma^2)$ is the p.d.f. of a lognormal random variable, with parameters μ, σ^2, as given by (3.2-3).

The rth moment of X about zero can be shown (see for instance, Aitchison and Brown (1957)) to be :

$$E(X^r) = \exp(r\mu + \tfrac{1}{2}r^2\sigma^2)\,\frac{1 - LN(x;\ \mu + r\sigma^2,\sigma^2)}{1 - LN(E\,;\mu,\sigma^2)} \qquad (7.2\text{-}3)$$

From (3.3-3) we know that $\exp(r\mu + \frac{1}{2} r^2 \sigma^2)$ is the rth moment about zero of the 2-parameter lognormal distribution. The remaining expression on the right hand side of (7.2-3) can be expressed in terms of the standard normal distribution function by using relationship (3.2-4). Hence Quensel (1945) expresses (7.2-3) as :

$$E(X^r) = (\text{rth moment if not truncated}) \frac{1 - N(U - r\sigma ; 0,1)}{1 - N(U ; 0,1)} \qquad (7.2\text{-}4)$$

where $U = \frac{\log E - \mu}{\sigma}$ and $N(\cdot ; 0,1)$ is the standard normal distribution function.

The mean of the distribution of X is, therefore,

$$E(X) = \exp(\mu + \tfrac{1}{2}\sigma^2) \frac{1 - N(U - \sigma; 0,1)}{1 - N(U ; 0,1)} \qquad (7.2\text{-}5)$$

and the variance of X can be shown to be

$$\text{Var}(X) = \left[\frac{\exp(\mu + \frac{1}{2}\sigma^2)}{1 - N(U;0,1)}\right]^2 \Big\{ \exp(\sigma^2)[1-N(U-2\sigma;0,1)][1-N(U;0,1)] - [1 - N(U - \sigma;0,1)]^2 \Big\} \qquad (7.2\text{-}6)$$

7.3 Estimation of the Parameters

In this section we study the problem of estimating the parameters μ and σ^2 for a TLN$(x,E ; \mu,\sigma^2)$ distribution with the known point of truncation E. This problem has been dealt with by several authors. Aitchison and Brown (1957) describe the maximum likelihood method when values of the individual observations in the sample are available. They suggest using the tables given by Hald (1949) for carrying out the inverse interpolation required in this method. Harding (1968) explains this method for fitting the truncated lognormal model to the third party excess of loss reinsurance claim amounts of motor insurance business.

The quantiles and graphical methods are not applicable since an estimate of $LN(E\ ;\ \mu,\sigma^2)$ (i.e., the area under the complete frequency curve of $LN(x\ ;\ \mu,\sigma^2)$ and to the left of the truncation point) is required but there is no information on this.

The method of moments may be used which involves putting the first and second sample moments equal to their corresponding theoretical values given by (7.2-3). However, there is no easy way of solving the resulting equations for μ and σ^2 except by trial and error. Grundy (1952) has considered estimation, from grouped data, by the method of moments for the truncated normal distribution. His method may be used for the truncated lognormal distribution since $\log x$ is normally distributed. He gives expressions for adjusting the first and second sample moments about zero when they are calculated from grouped truncated data.

Tallis and Young (1962) have considered estimation from grouped data by the multinomial maximum likelihood, (MML), method but their technique involves solving iteratively a system of non-linear equations in the parameters.

We will adopt the MML method but will apply the computing technique, described in section 3.7.7, of maximizing the loglikelihood function iteratively with respect to the parameters.

Let us assume that we have a sample of grouped data where n independent random observations (claims) on a random variable X (claim amount) have been grouped according to their size into k mutually exclusive intervals. Further, let us assume that each claim is for an amount greater than E (the truncation point) and that n_i is the number of claims in the class interval (x_{i-1}, x_i), for $i = 1, 2, \dots, k$ and with $x_0 = E$, such that

$$n = \sum_{i=1}^{k} n_i$$

If X is distributed as a $TLN(x,E \; ; \mu,\sigma^2)$ distribution given by (7.2-1), then let p_i be the probability that the size of a random observation (claim) will fall in the interval (x_{i-1}, x_i), i.e.

$$p_i = TLN(x_i, E \; ; \mu,\sigma^2) - TLN(x_{i-1}, E \; ; \mu,\sigma^2) \tag{7.3-1}$$

From (7.2-1) it follows that

$$p_i = \frac{LN(x_i; \mu,\sigma^2) - LN(x_{i-1} \; ; \mu,\sigma^2)}{1 - LN(E \; ; \mu,\sigma^2)} \tag{7.3-2}$$

Therefore, the sample likelihood function will be proportional to L where

$$L = \prod_{i=1}^{k} p_i^{n_i} \tag{7.3-3}$$

and the loglikelihood function will be

$$\log L = \sum_{i=1}^{k} n_i \log p_i \tag{7.3-4}$$

From (7.3-2) and remembering that from (3.2-4)

$$LN(x \; ; \mu,\sigma^2) = N\left(\frac{\log x - \mu}{\sigma} \; ; \; 0,1\right)$$

we can write (7.3-4) as :

$$\log L = -n \log\left(N\left(\frac{-\log E + \mu}{\sigma} \; ; \; 0,1\right)\right) + \sum_{i=1}^{k} n_i \log\left[N\left(\frac{\log x_i - \mu}{\sigma} \; ; 0,1\right) - N\left(\frac{\log x_{i-1} - \mu}{\sigma} \; ; 0,1\right)\right] \tag{7.3-5}$$

The maximum likelihood estimates $(\hat{\mu},\hat{\sigma}^2)$ of the parameters are obtained by maximizing $\log L$ simultaneously with respect to μ and σ^2. To avoid solving non-linear equations in the parameters we suggest the computing technique described in section 3.7.7 for maximizing $\log L$ with respect to

μ and σ^2. To start the iteration process we can use the estimates of μ and σ^2 which we found for the complete samples in chapter 3. If such a knowledge about the possible values of $\hat{\mu}$ and $\hat{\sigma}^2$ is not available then the method of moments should be used to work out a set of ($\hat{\mu},\hat{\sigma}^2$) for starting the iteration process. In any case, the penalty for choosing a very inaccurate set of $(\hat{\mu},\hat{\sigma}^2)$ as starting values is a few more iterations, which is not really a problem when the computer is used.

7.4 Applications to the AD Data

Computer program P23 was written to estimate the parameters of the truncated lognormal distribution by multinomial maximum likelihood, (MML), method. The complete samples of the AD data which were presented in tables (1.1) to (1.7) are input to the program and a truncation point is specified. In chapter 3 we showed that for our data the estimates of the parameters of the complete 2-parameter lognormal distribution were $\hat{\mu} \doteqdot 4.5$ and $\hat{\sigma}^2 \doteqdot 1$. These values are specified as initial values of the parameters to start the iteration process.

We first dealt with problem 1 put forward in section 7.1, i.e., truncation in the lower tail sample values. Each sample of AD data was truncated respectively at E = 30, E = 60 and E = 90, and the parameters of the truncated model were estimated. In each case the computer program gave the estimates of μ and σ^2 and an extensive table of relevant statistics. As an example, for the 73/4th quarter data the results of the computer runs are given in tables (7.1), (7.2) and (7.3). The results for all the samples have been summarized in tables (7.4) to (7.7). The correspon-ding results for complete samples, which were given in tables (3.11) to (3.17), are also reproduced for the sake of comparison. Table (7.4) gives the estimates of μ. We notice that truncation has given rise to larger values of $\hat{\mu}$ compared with the corresponding values for complete samples. For $\hat{\sigma}^2$ we can see, from table (7.5), that truncation has resulted in smaller values. Tables (7.1) to (7.3) show that the contribution to the Chi-square statistic, X^2, by different intervals is generally small. The values of X^2 statistic for complete and truncated samples are presented in table (7.6). For complete samples the X^2 values showed significant differences, at 0.05 level, between the $LN(x\ ;\ \mu,\sigma^2)$ model and actual sample values for every period of accident. The values of these statistics for truncated samples are considerably smaller and, except for 74/1st and 74/3rd quarters, do not show any significant differences, at 0.05 level, between the actual sample values and the truncated fitted models. Table (7.7) gives the ratio of the total expected loss statistic, T, to the total actual cost of claims. We observe that smaller values have been obtained for truncated samples indicating a better overall agreement between the model and actual sample values. The above analysis shows that truncation in the lower tail sample values of the AD data gives rise to a more exact fitting statistical model compared with the 2-parameter lognormal distribution for the complete sample. However, the fit does not seem to improve when truncation is carried on beyond the first interval (E = 30). In terms of the Chi-square goodness-of-fit test statistics, the results for the truncated (E = 30) and 3-parameter lognormal (see tables (3.38) to (3.44)) distributions are very similar. The latter distribution is, however, easier to work with and to develop mathematically. For the AD data, we, therefore, recommend the use of the 3-parameter lognormal distribution.

The second problem we put forward in section 7.1 was the application of the truncated lognormal as a model for the distribution of large AD claim amounts. We, therefore, used program P23 with our samples of AD data. Each sample was truncated at E = £600 and ($\hat{\mu}$ = 4.5, $\hat{\sigma}^2$ = 1) were specified as starting values for the iteration process. For each sample the computer program produced the estimates of μ and σ^2 as well as a table of relevant statistics. The results are presented in tables (7.8.1) to (7.8.7). We can see that $\hat{\mu}$ is larger than for complete samples while $\hat{\sigma}^2$ is considerably smaller. The actual means and standard deviations of the truncated samples are given in table (7.11). From tables (7.8.1) to (7.8.7) we can see that the mean claim amounts of the fitted models are approximately equal to their corresponding sample values. The standard deviations of the models are greater than their actual values but this could be because we are fitting a curve with an infinite range to a sample which has a finite range. From the tables of results it is apparent that the contributions to the X^2 statistic are generally small and that the overall X^2 statistics, except for 74/3rd quarter, do not indicate any significant differences between the model and actual sample values. The total expected loss statistics are small and are at most 5% of the total actual cost of claims. This indicates an overall satisfactory fit of the model to the sample values.

To see how the truncated lognormal model performs against the, much simpler to use, Pareto distribution, we wrote computer program P24 to estimate, from a truncated sample, the parameters of the Pareto distribution of the first kind. This program uses the MML method of estimation. The program was then modified to estimate the parameters of the Pareto distribution of the second kind. In each case the estimates of the parameters by the method of moments were input as starting values for the iteration process. These programs were run with truncated samples (at B = 600) of our AD data. For a particular sample, 73/4th quarter, the results are given in tables (7.9) and (7.10). A summary of the results for all samples are presented in tables (7.12) and (7.13). For the Pareto distribution of the first kind the mean claim amount from each model is approximately equal to its corresponding sample mean. The standard deviations of the models are greater than their corresponding sample values as we are fitting a curve with an infinite range to a sample with a finite range. The X^2 statistics show that, with the exception of 74/3rd quarter, there are no significant differences, at 0.05 level, between the model and actual sample values. The ratios of the total expected loss statistics to the total actual cost of claims are larger than their corresponding values for the truncated lognormal distribution. For the Pareto distribution of the second kind, P2 (x ; C, A, B), parameter C can be interpreted as the 'voluntary excess' on the policy. Estimates of C for each sample are given in table (7.13). They vary considerably from each period of accident to the next. The mean claim amounts are much closer to their corresponding sample values. The standard deviations of the models are larger than their actual values but are smaller than their corresponding values according to the Pareto distribution of the first kind. The X^2 statistics are smaller than their corresponding values in table (7.12) and do not indicate any significant differences between the model and actual sample values. The ratios of the total expected loss statistics to the total actual cost of claims are generally smaller than those given for P1 (x ; A, B) in table (7.12) but are larger than their respective values for the truncated lognormal distribution. In terms of the X^2 goodness-of-fit test statistic, P2(x ; C, A, B) has provided a more exact fit to the AD data than P1(x ; A, B) distribution. In fact, the fit is almost as good as that of the truncated lognormal distribution. One obvious reason for this is that P1(x ; A, B) has two parameters while P2(x ; C, A, B) and TLN(x, E ; μ,σ^2) each have three parameters. Therefore, given good computer facilities, we would prefer the truncated lognormal model, for the distribution of large AD claim amounts, though just as good a fit may be obtained by P2(x ; C, A, B) or even P1 (x ; A, B) distributions.

7.5 Conclusions

In this chapter we considered the truncated lognormal distribution as a model for the incomplete distribution of accidental damage claim amounts. The method of MML was described for the estimation of parameters from grouped data. The model was then fitted to samples of AD data truncated at lower tail values. It was shown that by ignoring the first claim amount interval this model would satisfactorily fit the AD samples. In terms of the Chi-square goodness-of-fit statistic the fit was as good as that of the complete 3-parameter lognormal distribution. Because this latter distribution is relatively simpler to use and to develop mathematically its application, in place of the truncated lognormal distribution, was recommended.

Next, the truncated lognormal distribution was considered as a model for the distribution of large claim amounts. Our AD samples were truncated at £600 and the model was fitted to the incomplete samples. It was shown that a very satisfactory fit is provided by the truncated lognormal distribution. For the sake of comparison, the Pareto distributions of the first and second kind were also fitted to the same truncated samples. The P2(x ; C, A, B) provided a fit as good as the truncated lognormal distribution. It was suggested that when computing is not a problem the latter model should be used. Otherwise P2(x ; C, A, B) or even the Pareto distribution of the first kind may be used as models for large AD claim amounts.

The effects of inflation on the parameters of the truncated distribution were not studied, but because of the relationship between the complete and truncated lognormal distributions we can easily arrive at these effects by following the exposition of section 3.9. It may, in fact, be shown that inflation increases parameter μ but leaves σ^2 unchanged.

7.6 Tables

Table (7.1)

*** TRUNCATED 2-PARAMETER LOGNORMAL DIS. ***

73/4TH QUARTER DATA

ESTIMATION BY MULTINOMIAL MAX. LIKELIHOOD METHOD :-

MEW= 4.654 SIGMA SQ.= 0.862

MEAN= 175.678 S.D.= 231.181

AMOUNT £	ACTUAL CL. NO.	EXPECTED CL. NO.	ACTUAL-EXPECTED	EXPECTED LOSS	(A-E)**2/E
31- 60	518.	522.	-4.	-182.	0.031
61- 90	461.	452.	9.	680.	0.179
91- 120	359.	346.	13.	1371.	0.488
121- 150	239.	260.	-21.	-2845.	1.696
151- 180	213.	196.	17.	2813.	1.474
181- 210	148.	149.	-1.	-195.	0.007
211- 240	102.	115.	-13.	-2931.	1.470
241- 270	81.	90.	-9.	-2299.	0.900
271- 300	58.	72.	-14.	-3997.	2.722
301- 330	66.	57.	9.	2839.	1.421
331- 360	45.	46.	-1.	-346.	0.022
361- 390	39.	38.	1.	376.	0.026
391- 420	35.	31.	4.	1622.	0.516
421- 450	34.	26.	8.	3484.	2.462
451- 480	20.	22.	-2.	-931.	0.182
481- 510	29.	18.	11.	5450.	6.722
511- 540	14.	15.	-1.	-525.	0.067
541- 570	8.	13.	-5.	-2777.	1.923
571- 600	9.	11.	-2.	-1171.	0.364
601- 700	29.	27.	2.	1301.	0.148
701- 800	18.	17.	1.	751.	0.059
801- 900	20.	11.	9.	7654.	7.364
901-1000	6.	8.	-2.	-1901.	0.500
1001-1100	4.	5.	-1.	-1050.	0.200
1101-1200	4.	4.	0.	0.	0.000
1201-1300	1.	3.	-2.	-2501.	
1301-1400	3.	2.	1.	1350.	0.200
1401-1500	1.	2.	-1.	-1450.	
1501-1600	0.	1.	-1.	-1550.	
1601-1700	1.	1.	0.	0.	
1701-1800	0.	1.	-1.	-1751.	
1801-1900	0.	1.	-1.	-1850.	
1901-2000	1.	0.	1.	1950.	
2001-2100	0.	0.	0.	0.	
2101-2200	0.	0.	0.	0.	
2201-2300	0.	0.	0.	0.	
2301-2400	1.	0.	1.	2350.	0.667
TOTAL	2567.	2562.		3737.	31.809

D.F. =24

TOTAL EXP. LOSS / TOTAL ACT. COST = 0.8 %

P > 0.10

Table (7.2)

*** TRUNCATED 2-PARAMETER LOGNORMAL DIS. ***

73/4TH QUARTER DATA

ESTIMATION BY MULTINOMIAL MAX. LIKELIHOOD METHOD :-
MEW= 4.633 SIGMA SQ.= 0.881

MEAN= 208.871 S.D.= 202.444

AMOUNT £	ACTUAL CL. NO.	EXPECTED CL. NO.	ACTUAL-EXPECTED	EXPECTED LOSS	(A-E)**2/E
61- 90	461.	459.	2.	151.	0.009
91- 120	359.	348.	11.	1160.	0.348
121- 150	239.	260.	-21.	-2845.	1.696
151- 180	213.	195.	18.	2979.	1.662
181- 210	148.	149.	-1.	-195.	0.007
211- 240	102.	115.	-13.	-2931.	1.470
241- 270	81.	90.	-9.	-2299.	0.900
271- 300	58.	71.	-13.	-3711.	2.380
301- 330	66.	57.	9.	2839.	1.421
331- 360	45.	46.	-1.	-346.	0.022
361- 390	39.	38.	1.	376.	0.026
391- 420	35.	31.	4.	1622.	0.516
421- 450	34.	26.	8.	3484.	2.462
451- 480	20.	22.	-2.	-931.	0.182
481- 510	29.	18.	11.	5450.	6.722
511- 540	14.	15.	-1.	-525.	0.067
541- 570	8.	13.	-5.	-2777.	1.923
571- 600	9.	11.	-2.	-1171.	0.364
601- 700	29.	27.	2.	1301.	0.148
701- 800	18.	17.	1.	751.	0.059
801- 900	20.	11.	9.	7654.	7.364
901-1000	6.	8.	-2.	-1901.	0.500
1001-1100	4.	5.	-1.	-1050.	0.200
1101-1200	4.	4.	0.	0.	0.000
1201-1300	1.	3.	-2.	-2501.	
1301-1400	3.	2.	1.	1350.	0.200
1401-1500	1.	2.	-1.	-1450.	
1501-1600	0.	1.	-1.	-1550.	
1601-1700	1.	1.	0.	0.	
1701-1800	0.	1.	-1.	-1751.	
1801-1900	0.	1.	-1.	-1850.	
1901-2000	1.	0.	1.	1950.	
2001-2100	0.	0.	0.	0.	
2101-2200	0.	0.	0.	0.	
2201-2300	0.	0.	0.	0.	
2301-2400	1.	0.	1.	2350.	0.667
TOTAL	2049.	2047.		3631.	51.312

D.F. =23

P > 0.10

TOTAL EXP. LOSS / TOTAL ACT. COST = 0.9 %

Table (7.3)

*** TRUNCATED 2-PARAMETER LOGNORMAL DIS. ***

73/4TH QUARTER DATA

ESTIMATION BY MULTINOMIAL MAX. LIKELIHOOD METHOD :-

MEW= 4.648 SIGMA SQ.= 0.870

MEAN= 247.649 S.D.= 186.312

AMOUNT £	ACTUAL CL. NO.	EXPECTED CL. NO.	ACTUAL-EXPECTED	EXPECTED LOSS	(A-E)**2/E
91- 120	359.	345.	14.	1477.	0.568
121- 150	239.	259.	-20.	-2710.	1.544
151- 180	213.	195.	18.	2979.	1.662
181- 210	148.	149.	-1.	-195.	0.007
211- 240	102.	115.	-13.	-2931.	1.470
241- 270	81.	90.	-9.	-2299.	0.900
271- 300	58.	71.	-13.	-3711.	2.380
301- 330	66.	57.	9.	2839.	1.421
331- 360	45.	46.	-1.	-346.	0.022
361- 390	39.	38.	1.	376.	0.026
391- 420	35.	31.	4.	1622.	0.516
421- 450	34.	26.	8.	3484.	2.462
451- 480	20.	22.	-2.	-931.	0.182
481- 510	29.	18.	11.	5450.	6.722
511- 540	14.	15.	-1.	-525.	0.067
541- 570	8.	13.	-5.	-2777.	1.923
571- 600	9.	11.	-2.	-1171.	0.364
601- 700	29.	27.	2.	1301.	0.148
701- 800	18.	17.	1.	751.	0.059
801- 900	20.	11.	9.	7654.	7.364
901-1000	6.	8.	-2.	-1901.	0.500
1001-1100	4.	5.	-1.	-1050.	0.200
1101-1200	4.	4.	0.	0.	0.000
1201-1300	1.	3.	-2.	-2501.	
1301-1400	3.	2.	1.	1350.	0.200
1401-1500	1.	2.	-1.	-1450.	
1501-1600	0.	1.	-1.	-1550.	
1601-1700	1.	1.	0.	0.	
1701-1800	0.	1.	-1.	-1751.	
1801-1900	0.	1.	-1.	-1850.	
1901-2000	1.	0.	1.	1950.	
2001-2100	0.	0.	0.	0.	
2101-2200	0.	0.	0.	0.	
2201-2300	0.	0.	0.	0.	
2301-2400	1.	0.	1.	2350.	0.667
TOTAL	1588.	1584.		3932.	31.372

D.F. =22

TOTAL EXP. LOSS / TOTAL ACT. COST = 1.0 %

P > 0.05

Table (7.4)

MML estimates of μ
for complete and truncated AD samples

Period of accident	Complete sample	E=30	E=60	E=90
73/4th quarter	4.516	4.654	4.633	4.648
74/1st "	4.509	4.624	4.581	4.296
74/2nd "	4.546	4.685	4.592	4.604
74/3rd "	4.637	4.765	4.758	4.581
74/4th "	4.672	4.807	4.823	4.851
75/1st "	4.684	4.821	4.881	4.891
75/2nd "	4.701	4.815	4.849	4.830

Table (7.5)

MML estimates of σ^2
for complete and truncated AD samples

Period of accident	Complete sample	E=30	E=60	E=90
73/4th quarter	1.055	0.862	0.881	0.870
74/1st "	1.057	0.896	0.935	1.142
74/2nd "	1.013	0.815	0.901	0.892
74/3rd "	1.011	0.821	0.828	0.960
74/4th "	1.056	0.851	0.836	0.813
75/1st "	1.024	0.815	0.754	0.747
75/2nd "	1.056	0.880	0.846	0.861

Table (7.6)

χ^2 statistic for complete and truncated AD samples*

Period of accident	Complete sample		E = 30		E = 60		E = 90	
73/4th quarter	53.4	(25)	31.8	(24)	31.3	(23)	31.3	(22)
74/1st "	56.9	(24)	49.1	(23)	47.8	(22)	38.9	(21)
74/2nd "	51.8	(24)	26.3	(22)	22.8	(22)	23.0	(21)
74/3rd "	69.3	(26)	41.6	(24)	41.6	(23)	35.8	(23)
74/4th "	56.4	(27)	21.6	(25)	22.0	(24)	22.0	(23)
75/1st "	59.7	(26)	23.3	(24)	21.9	(23)	22.9	(22)
75/2nd "	44.6	(27)	24.7	(25)	25.0	(24)	24.1	(23)

* the number in ( ) is the degrees of freedom

Table (7.7)

Ratio of the total expected loss statistic, T , to the total actual cost of claims for complete and truncated AD samples (in percentage)

Period of accident	Complete sample	E = 30	E = 60	E = 90
73/4th quarter	- 1.5	0.8	0.9	1.0
74/1st "	0.6	1.5	1.8	1.2
74/2nd "	- 1.0	1.1	0.6	0.7
74/3rd "	- 1.0	1.1	1.3	0.7
74/4th "	- 1.9	0.6	0.4	0.6
75/1st "	- 2.1	0.5	0.6	0.9
75/2nd "	- 0.9	0.7	0.4	0.7

Table (7.8.1)

*** TRUNCATED 2-PARAMETER LOGNORMAL DIS. ***

73/4TH QUARTER DATA

ESTIMATION BY MULTINOMIAL MAX. LIKELIHOOD METHOD :-

MEW= 5.827 SIGMA SQ.= 0.350

MEAN= 856.558 S.D.= 490.808

AMOUNT £	ACTUAL CL. NO.	EXPECTED CL. NO.	ACTUAL-EXPECTED	EXPECTED LOSS	(A-E)**2/E
601- 700	29.	30.	-1.	-651.	0.033
701- 800	18.	20.	-2.	-1501.	0.200
801- 900	20.	13.	7.	5953.	3.769
901-1000	6.	8.	-2.	-1901.	0.500
1001-1100	4.	6.	-2.	-2101.	0.667
1101-1200	4.	4.	0.	0.	
1201-1300	1.	3.	-2.	-2501.	
1301-1400	3.	2.	1.	1350.	0.571
1401-1500	1.	1.	0.	0.	
1501-1600	0.	1.	-1.	-1550.	
1601-1700	1.	1.	0.	0.	
1701-1800	0.	0.	0.	0.	
1801-1900	0.	0.	0.	0.	
1901-2000	1.	0.	1.	1950.	
2001-2100	0.	0.	0.	0.	
2101-2200	0.	0.	0.	0.	
2201-2300	0.	0.	0.	0.	
2301-2400	1.	0.	1.	2350.	0.800
TOTAL	89.	89.		1400.	6.540

D.F. = 4

P > 0.10

TOTAL EXP. LOSS / TOTAL ACT. COST = 1.8 %

Table (7.8.2)

*** TRUNCATED 2-PARAMETER LOGNORMAL DIS. ***

74/1ST QUARTER DATA

ESTIMATION BY MULTINOMIAL MAX. LIKELIHOOD METHOD :-
MEW= 6.385 SIGMA SQ.= 0.130

MEAN= 816.831 S.D.= 723.298

AMOUNT £	ACTUAL CL. NO.	EXPECTED CL. NO.	ACTUAL-EXPECTED	EXPECTED LOSS	(A-E)**2/E
601- 700	26.	27.	-1.	-651.	0.037
701- 800	21.	19.	2.	1501.	0.211
801- 900	11.	13.	-2.	-1701.	0.308
901-1000	10.	8.	2.	1901.	0.500
1001-1100	5.	5.	0.	0.	0.000
1101-1200	2.	3.	-1.	-1150.	
1201-1300	1.	2.	-1.	-1250.	
1301-1400	2.	1.	1.	1350.	
1401-1500	0.	1.	-1.	-1450.	
1501-1600	0.	0.	0.	0.	
1601-1700	0.	0.	0.	0.	
1701-1800	1.	0.	1.	1751.	0.143
TOTAL	79.	79.		300.	1.198

D.F. = 3

$$\frac{\text{TOTAL EXP. LOSS}}{\text{TOTAL ACT. COST}} = 0.5\ \%$$

$P > 0.50$

Table (7.8.3)

*** TRUNCATED 2-PARAMETER LOGNORMAL DIS. ***

74/2ND QUARTER DATA

ESTIMATION BY MULTINOMIAL MAX. LIKELIHOOD METHOD :-

MEW= 5.265 SIGMA SQ.= 0.419

MEAN= 804.178 S.D.= 368.034

AMOUNT £	ACTUAL CL. NO.	EXPECTED CL. NO.	ACTUAL-EXPECTED	EXPECTED LOSS	(A-E)**2/E
601- 700	30.	29.	1.	651.	0.034
701- 800	13.	16.	-3.	-2251.	0.563
801- 900	11.	9.	2.	1701.	0.444
901-1000	4.	5.	-1.	-951.	0.200
1001-1100	7.	3.	4.	4202.	
1101-1200	0.	2.	-2.	-2301.	
1201-1300	1.	1.	0.	0.	
1301-1400	1.	1.	0.	0.	
1401-1500	0.	1.	-1.	-1450.	
1501-1600	1.	0.	1.	1550.	
1601-1700	0.	0.	0.	0.	
1701-1800	0.	0.	0.	0.	
1801-1900	1.	0.	1.	1850.	1.125
TOTAL	69.	67.		3001.	2.366

D.F. = 2

P =0.30

TOTAL EXP. LOSS / TOTAL ACT. COST = 5.4 %

Table (7.8.4)

*** TRUNCATED 2-PARAMETER LOGNORMAL DIS. ***

74/3RD QUARTER DATA

ESTIMATION BY MULTINOMIAL MAX. LIKELIHOOD METHOD :-

MEW= 5.596 SIGMA SQ.= 0.399

MEAN= 842.542 S.D.= 421.929

AMOUNT £	ACTUAL CL. NO.	EXPECTED CL. NO.	ACTUAL-EXPECTED	EXPECTED LOSS	(A-E)**2/E
601- 700	32.	39.	-7.	-4553.	1.256
701- 800	34.	24.	10.	7505.	4.167
801- 900	17.	15.	2.	1701.	0.267
901-1000	4.	10.	-6.	-5703.	3.600
1001-1100	9.	6.	3.	3151.	1.500
1101-1200	4.	4.	0.	0.	
1201-1300	0.	3.	-3.	-3751.	1.286
1301-1400	1.	2.	-1.	-1350.	
1401-1500	3.	1.	2.	2901.	
1501-1600	0.	1.	-1.	-1550.	
1601-1700	1.	1.	0.	0.	
1701-1800	0.	0.	0.	0.	
1801-1900	0.	0.	0.	0.	
1901-2000	0.	0.	0.	0.	
2001-2100	0.	0.	0.	0.	
2101-2200	1.	0.	1.	2150.	
2201-2300	0.	0.	0.	0.	
2301-2400	0.	0.	0.	0.	
2401-2500	0.	0.	0.	0.	
2501-2600	1.	0.	1.	2550.	0.800
TOTAL	107.	106.		3050.	12.876

D.F. = 4

TOTAL EXP. LOSS / TOTAL ACT. COST = 3.4 %

P = 0.01

Table (7.8.5)

*** TRUNCATED 2-PARAMETER LOGNORMAL DIS. ***

74/4TH QUARTER DATA

ESTIMATION BY MULTINOMIAL MAX. LIKELIHOOD METHOD :-
MEW= 6.183 SIGMA SQ.= 0.223

MEAN= 845.861 S.D.= 634.330

AMOUNT £	ACTUAL CL. NO.	EXPECTED CL. NO.	ACTUAL-EXPECTED	EXPECTED LOSS	(A-E)**2/E
601- 700	42.	43.	-1.	-651.	0.023
701- 800	28.	29.	-1.	-751.	0.034
801- 900	20.	20.	0.	0.	0.000
901-1000	17.	13.	4.	3802.	1.231
1001-1100	8.	8.	0.	0.	0.000
1101-1200	5.	5.	0.	0.	0.000
1201-1300	1.	4.	-3.	-3751.	
1301-1400	5.	2.	3.	4051.	0.000
1401-1500	0.	2.	-2.	-2901.	
1501-1600	1.	1.	0.	0.	
1601-1700	0.	1.	-1.	-1650.	
1701-1800	0.	0.	0.	0.	
1801-1900	0.	0.	0.	0.	
1901-2000	0.	0.	0.	0.	
2001-2100	1.	0.	1.	2050.	
2101-2200	0.	0.	0.	0.	
2201-2300	0.	0.	0.	0.	
2301-2400	1.	0.	1.	2350.	0.200
TOTAL	129.	128.		2550.	1.539

D.F. = 5

$$\frac{\text{TOTAL EXP. LOSS}}{\text{TOTAL ACT. COST}} = 2.3\ \%$$

P > 0.90

Table (7.8.6)

*** TRUNCATED 2-PARAMETER LOGNORMAL DIS. ***

75/1ST QUARTER DATA

ESTIMATION BY MULTINOMIAL MAX. LIKELIHOOD METHOD :-

MEW= 1.278 SIGMA SQ.= 1.614

MEAN= 834.771 S.D.= 15.667

AMOUNT £	ACTUAL CL. NO.	EXPECTED CL. NO.	ACTUAL-EXPECTED	EXPECTED LOSS	(A-E)**2/E
601- 700	39.	40.	-1.	-651.	0.025
701- 800	19.	22.	-3.	-2251.	0.409
801- 900	18.	12.	6.	5103.	3.000
901-1000	12.	8.	4.	3802.	2.000
1001-1100	3.	5.	-2.	-2101.	0.800
1101-1200	1.	3.	-2.	-2301.	
1201-1300	1.	2.	-1.	-1250.	
1301-1400	0.	2.	-2.	-2701.	1.800
1401-1500	1.	1.	0.	0.	
1501-1600	0.	1.	-1.	-1550.	
1601-1700	1.	1.	0.	0.	
1701-1800	0.	0.	0.	0.	
1801-1900	0.	0.	0.	0.	
1901-2000	0.	0.	0.	0.	
2001-2100	1.	0.	1.	2050.	
2101-2200	1.	0.	1.	2150.	
2201-2300	0.	0.	0.	0.	
2301-2400	0.	0.	0.	0.	
2401-2500	0.	0.	0.	0.	
2501-2600	0.	0.	0.	0.	
2601-2700	0.	0.	0.	0.	
2701-2800	0.	0.	0.	0.	
2801-2900	0.	0.	0.	0.	
2901-3000	0.	0.	0.	0.	
3001-3100	0.	0.	0.	0.	
3101-3200	0.	0.	0.	0.	
3201-3300	0.	0.	0.	0.	
3301-3400	0.	0.	0.	0.	
3401-3500	0.	0.	0.	0.	
3501-3600	1.	0.	1.	3550.	0.000
TOTAL	98.	97.		3850.	8.834

D.F. = 4

TOTAL EXP. LOSS / TOTAL ACT. COST = 4.7 %

P > 0.05

Table (7.8.7)

*** TRUNCATED 2-PARAMETER LOGNORMAL DIS. ***

75/2ND QUARTER DATA

ESTIMATION BY MULTINOMIAL MAX. LIKELIHOOD METHOD :-

MEW= 5.982 SIGMA SQ.= 0.324

MEAN= 873.216 S.D.= 543.923

AMOUNT £	ACTUAL CL. NO.	EXPECTED CL. NO.	ACTUAL-EXPECTED	EXPECTED LOSS	(A-E)**2/E
601- 700	36.	36.	0.	0.	0.000
701- 800	22.	24.	-2.	-1501.	0.167
801- 900	22.	16.	6.	5103.	2.250
901-1000	11.	11.	0.	0.	0.000
1001-1100	4.	8.	-4.	-4202.	2.000
1101-1200	3.	5.	-2.	-2301.	0.800
1201-1300	3.	4.	-1.	-1250.	
1301-1400	6.	2.	4.	5402.	1.500
1401-1500	2.	2.	0.	0.	
1501-1600	1.	1.	0.	0.	
1601-1700	0.	1.	-1.	-1650.	
1701-1800	1.	1.	0.	0.	
1801-1900	0.	0.	0.	0.	
1901-2000	1.	0.	1.	1950.	
2001-2100	0.	0.	0.	0.	
2101-2200	0.	0.	0.	0.	
2201-2300	1.	0.	1.	2250.	0.200
TOTAL	113.	111.		3801.	6.917

D.F. = 5

TOTAL EXP. LOSS / TOTAL ACT. COST = 3.8 %

P ≈ 0.20

Table (7.9)

*** PARETO DIS. OF THE 1ST KIND ***

73/4TH QUARTER DATA

ESTIMATION BY MULTINOMIAL MAX. LIKELIHOOD METHOD :-
A= 3.188

MEAN= 875.009 S.D.= 449.738

AMOUNT £	ACTUAL CL. NO.	EXPECTED CL. NO.	ACTUAL-EXPECTED	EXPECTED LOSS	(A-E)**2/E
601- 700	29.	35.	-6.	-3903.	1.029
701- 800	18.	19.	-1.	-751.	0.053
801- 900	20.	11.	9.	7654.	7.364
901-1000	6.	7.	-1.	-951.	0.143
1001-1100	4.	5.	-1.	-1050.	0.200
1101-1200	4.	3.	1.	1150.	
1201-1300	1.	2.	-1.	-1250.	0.000
1301-1400	3.	2.	1.	1350.	
1401-1500	1.	1.	0.	0.	
1501-1600	0.	1.	-1.	-1550.	
1601-1700	1.	1.	0.	0.	
1701-1800	0.	1.	-1.	-1751.	
1801-1900	0.	0.	0.	0.	
1901-2000	1.	0.	1.	1950.	
2001-2100	0.	0.	0.	0.	
2101-2200	0.	0.	0.	0.	
2201-2300	0.	0.	0.	0.	
2301-2400	1.	0.	1.	2350.	0.167
TOTAL	89.	88.		3251.	9.200

D.F. = 5

TOTAL EXP. LOSS / TOTAL ACT. COST = 4.2 %

P = 0.10

Table (7.10)

*** PARETO DIS. OF THE 2ND KIND ***

73/4TH QUARTER DATA

ESTIMATION BY MULTINOMIAL MAX. LIKELIHOOD METHOD :-

A= 3.493 C= 27.94

MEAN= 869.306 S.D.= 368.417

AMOUNT £	ACTUAL CL. NO.	EXPECTED CL. NO.	ACTUAL-EXPECTED	EXPECTED LOSS	(A-E)**2/E
601- 700	29.	29.	0.	0.	0.000
701- 800	18.	23.	-5.	-3752.	1.087
801- 900	20.	13.	7.	5953.	3.769
901-1000	6.	8.	-2.	-1901.	0.500
1001-1100	4.	5.	-1.	-1050.	0.200
1101-1200	4.	3.	1.	1150.	
1201-1300	1.	2.	-1.	-1250.	0.000
1301-1400	3.	1.	2.	2701.	
1401-1500	1.	1.	0.	0.	
1501-1600	0.	1.	-1.	-1550.	
1601-1700	1.	1.	0.	0.	
1701-1800	0.	0.	0.	0.	
1801-1900	0.	0.	0.	0.	
1901-2000	1.	0.	1.	1950.	
2001-2100	0.	0.	0.	0.	
2101-2200	0.	0.	0.	0.	
2201-2300	0.	0.	0.	0.	
2301-2400	1.	0.	1.	2350.	2.250
TOTAL	89.	87.		4601.	8.006

D.F. = 4

P = 0.10

TOTAL EXP. LOSS / TOTAL ACT. COST = 6.0 %

Table (7.11)

Actual mean claim amount and standard deviation for AD samples truncated at E = 600

Period of accident	73/4	74/1	74/2	74/3	74/4	75/1	75/2
Mean	860.6	820.1	808.5	847.7	849.7	844.4	876.2
S.D.	286.0	199.0	223.0	288.0	256.0	373.0	287.0

Table (7.12)

Pareto distribution of the first kind fitted to the tail of the AD samples. (B = 600)

Period of accident	$\hat{A}$	Mean	S.D.	X^2	(D.F.)	R*
73/4th	3.188	875.0	449.7	9.2	(5)	4.2
74/1st	3.568	834.3	352.7	5.8	(4)	4.6
74/2nd	3.843	811.7	305.0	3.0	(3)	5.9
74/3rd	3.344	856.7	404.1	17.0	(6)	5.9
74/4th	3.259	866.3	427.8	9.6	(6)	3.8
75/1st	3.544	836.6	357.7	10.5	(5)	7.6
75/2nd	3.026	897.0	509.1	10.9	(6)	6.2

Table (7.13)

Pareto distribution of the second kind fitted to the tail of the AD samples (B = 600)

Period of accident	$\hat{C}$	$\hat{A}$	Mean	S.D.	X^2	(D.F.)	R*
73/4th quarter	28.0	3.493	869.3	368.4	8.0	(4)	6.0
74/1st "	39.0	4.163	829.6	263.4	3.4	(3)	6.4
74/2nd "	5.4	3.918	811.7	294.2	2.5	(3)	4.2
74/3rd "	43.0	3.938	848.3	291.4	5.4	(4)	3.6
74/4th "	30.4	3.612	860.8	344.2	5.2	(5)	4.9
75/1st "	11.5	3.686	835.6	330.6	8.2	(4)	5.1
75/2nd "	24.5	3.260	890.7	427.5	8.5	(5)	4.4

* R is the ratio, in percentage, of the total expected loss statistic, T, to the total actual cost of claims.

CHAPTER 8

The Gamma Distribution

8.1 Introduction

The gamma distribution is another positively skewed distribution with a long tail which can assume a variety of shapes. It has, therefore, been a most appropriate statistical model in applied science work. Many sets of empirical data on positive valued random outcomes of various random phenomena in science and engineering have been represented by this distribution. Johnson and Kotz (1970) and Bury (1975) give bibliographies on the applications of this model and the formers' list includes references to the actuarial field as well.

Beard (1978) argues that "the amount of damage (physical) will arise from impact and thus be related to the energy involved, that is, to the square of the velocity". If X is the amount of loss, and m and V are the mass and the velocity of the car at the time of impact, respectively, then the above statement means that

$$X \text{ is proportional to } \frac{1}{2} mV^2 \qquad (8.1\text{-}1)$$

and hence $$Z = \sqrt{X} \text{ is proportional to } V \qquad (8.1\text{-}2)$$

Beard (1978) then goes on to say that "velocity should have a reasonable shape" and in Beard (1978, personal communication) suggests the gamma distribution as a model for the distribution of velocity and hence of the square root of the claim amount. Therefore, the problem we shall consider in this chapter is the application of the two-parameter gamma distribution as a model for the square root of the claim amount. It is immediately appreciated that data on the square root of the claim amount, and not on the claim amount itself, are required. When a grouped frequency distribution of the claim amount is available, for instance like our AD data of tables (1.1) to (1.7), then the square roots of the boundaries of the claim amount intervals may be used to obtain a frequency distribution for the square root of the claim amount. This is, however, not satisfactory since the resulting intervals will be of different lengths and the useful assumption which is often made about the concentration of observations (claims) within each interval at the mid-point of that interval is unlikely to hold in reality.

Therefore the insurance company provided us with data on the square root of claim amounts. The claims are for accidental damage and are from the same portfolio of comprehensively insured private motor cars which we mentioned in chapter 1. Data combined for all age groups, vehicle groups, districts and types of use were made available. Six samples relating to six periods of accident were provided. Each period consists of one quarter of the calendar year and all claims which occured during that particular quarter constitute a sample. The six periods of accident were 1976/3rd quarter to 1977/4th quarter. For each sample the number of claims grouped according to the square root of their size, i.e. Z as in (8.1-2), were provided for z up to 38. The grouping was in equal bands of length 2. Such a format for a particular sample would look like the following histogram

square root of the claim amount	number of claims
up to 2	n_1
2.01 - 4	n_2
4.01 - 6	n_3
....	...
....	...
....	...
36.01 - 38	n_{19}

For larger claims details of individual claims were provided. We grouped these into further bands of length 2 until the band containing the largest claim (or claims) was reached. Because the square root of claim amount is not readily understood in monetary terms we squared the boundaries of each interval for presentation of the data. The frequency distributions of the six samples of AD data can, therefore, be seen as the first two columns of tables (8.1.1) to (8.1.6). Both open and settled claims have been included in each sample. For 1976/3rd quarter to 1977/2nd quarter the data were collected more than six months after the end of each period of accident and hence no more, or only few, claims are still to be reported. For 1977/3rd and 1977/4th quarters, as we can see from tables (8.1.5) and (8.1.6), the total number of claims are less than for other periods. This is because for these two periods of accident the data were collected, respectively, three months after the end and at the end of the periods. Therefore many more claims are still expected to be reported to complete the histogram in each case. We will not make any allowances for IBNR claims of these two periods but will judge the results of the model fitting in the light of the incompleteness of the data.

Initially, in this chapter, we will define the gamma distribution and will mention some of its properties. The estimation problem will be discussed next and the method of multinomial maximum likelihood (MML) will be described for estimation of the parameters from grouped data. This method will then be used to fit the gamma distribution to the AD square root data mentioned above. The results will be analysed and the goodness-of-fit tests will indicate the adequacy of this model. The effects of inflation on the parameters of the model will be considered in section 8.5. These will then be used to predict the future distribution of the square root of claim amounts from past data. The actual and predicted distributions will be compared. The findings of this chapter will be summarized in section 8.7. All the tables are presented in section 8.8.

8.2 Definition and Some Properties of the Gamma Distribution

A random variable Z is said to have a (two-parameter) gamma distribution if its probability density function, (p.d.f.), which we denote by $f_G(z\ ;\ \alpha,\beta)$, is of the form :

$$f_G(z\ ;\alpha,\beta) = \frac{z^{\alpha-1}e^{-z/\beta}}{\beta^{\alpha}\Gamma(\alpha)} \qquad (\alpha,\beta,\ z > 0) \qquad (8.2\text{-}1)$$

α and β are the shape and scale parameters respectively. $\Gamma(\alpha)$ is called "the gamma function" and is defined by :

$$\Gamma(\alpha) = \int_0^{\infty} t^{\alpha-1}e^{-t}\,dt \qquad (8.2\text{-}2)$$

which, by integrating by parts, can be shown to be

$$\Gamma(\alpha) = (\alpha-1)\Gamma(\alpha-1) \qquad (\text{for } \alpha > 0) \qquad (8.2\text{-}3)$$

When α is a positive integer $\Gamma(\alpha)$ is the factorial function, i.e.,

$$\Gamma(\alpha) = (\alpha-1)! \qquad (8.2\text{-}4)$$

If in (8.2-1) we put $\beta = 1$ then $f_G(z\ ;\ \alpha,1)$ will be the p.d.f. of the standard gamma distribution.

For $\alpha = \beta = 1$, (8.2-1) will give the p.d.f. of the exponential distribution.

If $\beta = 2$ and $\alpha = \frac{\nu}{2}$ then (8.2-1) will be the p.d.f. of the Chi-square distribution with ν degrees of freedom.

If we make a transformation, on the random variable Z, of the form $z \to z-c$ which is a translation along the z-axis then (8.2-1) will be the p.d.f. of the 3-parameter gamma distribution whose properties can be easily derived from those of $f_G(z;\alpha,\beta)$. Therefore it will suffice to consider only the 2-parameter gamma distribution here. Further reference can be made to Bury (1975) and Johnson and Kotz (1970).

We denote the distribution function of $f_G(z;\alpha,\beta)$ by $G(z;\alpha,\beta)$. For the standard gamma distribution we have

$$G(z;\alpha,1) = \frac{1}{\Gamma(\alpha)} \int_0^z z^{\alpha-1} e^{-z}\, dz \qquad (8.2\text{-}5)$$

where the integral is called "the incomplete gamma function" and is denoted by $\Gamma_z(\alpha)$. Therefore

$$G(z;\alpha,1) = \frac{\Gamma_z(\alpha)}{\Gamma(\alpha)} \qquad (8.2\text{-}6)$$

Since β is a scale parameter we can show that

$$G(z;\alpha,\beta) = G\left(\frac{z}{\beta};\alpha,1\right) \qquad (8.2\text{-}7)$$

hence the two parameter gamma distribution function is

$$G(z;\alpha,\beta) = \frac{\Gamma_{z/\beta}(\alpha)}{\Gamma(\alpha)} \qquad (8.2\text{-}8)$$

Let $\chi^2_\nu(u)$ denote the distribution function of a Chi-square random variable, U, with ν degrees of freedom. As we mentioned earlier, $f_G(z;\alpha,2)$ is the p.d.f. of a Chi-square random variable with $\nu = 2\alpha$ degrees of freedom. Hence the relationship between the gamma and the Chi-square distribution functions will be

$$G(z;\alpha,\beta) = G\left(\frac{2z}{\beta};\alpha,2\right) = \chi^2_{2\alpha}\left(\frac{2z}{\beta}\right) \qquad (8.2\text{-}9)$$ ✓

Therefore tables of the Chi-square cumulative distribution function may be used to calculate $G(z;\alpha,\beta)$. Pearson (1922) has, however, tabulated $G(z;\alpha,1)$, to seven decimal places, for $\alpha = 0\,(.05)1(0.1)6(0.2)51$. More accurate tabulations of $G(z;\alpha,1)$, to nine decimal places, have been given by Harter (1964) for $\alpha = 0.5(0.5)75(1)165$. Pearson (1922) has shown that,

$$G(z\ ;\ \alpha, \beta) = e^{-y} \sum_{j=0}^{\infty} [y^{\alpha+j}/\Gamma(\alpha+j+1)] \qquad (8.2\text{-}10)$$

where $y = \frac{z}{\beta} > 0$.

For computing purposes, a good approximation to the value of $G(z\ ;\ \alpha, \beta)$ can be obtained by taking sufficient terms in the above series. We shall use this formula in our estimation programs and other calculations involving the gamma cumulative distribution function. The values of the complete gamma function $\Gamma(\cdot)$, as required in (8.2-10) are computed by a special routine which exists in the NAG[1] library of computer routines. This routine is based on the Chebyshev expansion of the complete gamma function which is given in Abramowitz and Stegun (1968).

The rth moment of Z about zero can be directly obtained from (8.2-1) and (8.2-2) as :

$$E(Z^r) = \beta^r \frac{\Gamma(\alpha+r)}{\Gamma(\alpha)} \qquad (8.2\text{-}11)$$

hence the expected value of Z is

$$E(Z) = \alpha\beta \qquad (8.2\text{-}12)$$

and the variance of Z can be shown to be

$$Var(Z) = \alpha\beta^2 \qquad (8.2\text{-}13)$$

The coefficient of variation of Z is, therefore, equal to $\alpha^{-\frac{1}{2}}$ which is positive and depends on the shape parameter only. The 3rd and 4th central moments of Z, denoted by θ_3 and θ_4 respectively, can be shown, by using (8.2-11), to be

$$\theta_3 = 2\alpha\beta^3 \qquad (8.2\text{-}14)$$

and

$$\theta_4 = 3\alpha\beta^4(\alpha+2) \qquad (8.2\text{-}15)$$

---

1 - Nottingham Algorithmic Group

The coefficient of skewness, γ_1, is, therefore,

$$\gamma_1 = 2\alpha^{-\frac{1}{2}} > 0 \qquad (8.2\text{-}16)$$

hence all gamma models are positively skewed. The coefficient of kurtosis is γ_2 where

$$\gamma_2 = 3 + 6\alpha^{-1} > 3 \qquad (8.2\text{-}17)$$

which implies that all gamma densities are more peaked than normal densities.

For large values of α, γ_1 and γ_2 approach their normal values 0 and 3 respectively. Thus for large α the central portion of the corresponding gamma p.d.f. should be well approximated by a normal p.d.f.

When $\alpha \geq 1$ the gamma distribution has a single mode which is located at

$$z_{mode} = \beta(\alpha-1) \qquad (8.2\text{-}18)$$

For values of $\alpha < 1$ the gamma p.d.f. has a singularity at the origin and $f_G(z\ ;\ \alpha,\beta) \to \infty$ as $z \to 0$.

The median, and in general the quantile of order q, of the gamma distribution cannot be expressed in terms of its parameters but may be obtained, by interpolation, from tables of the gamma distribution function.

As we mentioned in section 8.1 we intend to use the gamma distribution as a model for the distribution of Z, the square root of claim amount X. The mean and standard deviation of Z do not have any significance in monetary terms. Therefore, it is important to obtain the mean and standard deviation of X in terms of α and β, the parameters of the gamma model. The expected value of X is

$$E(X) = E(Z^2) \qquad \text{since} \quad Z = \sqrt{X}$$

but from (8.2-11) and (8.2-3) it follows that

$$E(X) = \beta^2\alpha(\alpha+1) \qquad (8.2\text{-}19)$$

The variance of X is by definition

$$\mathrm{Var}(X) = E(X^2) - (E(X))^2 = E(Z^4) - (E(X))^2$$

On using (8.2-11), (8.2-3) and (8.2-19), respectively, it follows that

$$\mathrm{Var}(X) = \beta^4 \alpha^2 (4\alpha + 10 + \frac{6}{\alpha}) \qquad (8.2\text{-}20)$$

8.3 Estimation of the Parameters

Johnson and Kotz (1970) deal with the estimation problem when values of individual observations in the sample are known. They consider the usual methods of moments and maximum likelihood. We are, however, interested in estimation from grouped data. Therefore, in this section, we will consider the method of multinomial maximum likelihood (MML). Let us assume that we have a sample of grouped data where n independent random observations (claims) on a random variable Z (in our case, the square root of the claim amount X) have been grouped according to their size into k mutually exclusive intervals. Further, let n_i be the number of observations (claims) in the class interval (z_{i-1}, z_i), for $i = 1, 2, \ldots, k$, such that

$$n = \sum_{i=1}^{k} n_i .$$

If Z has a gamma distribution given by $G(z ; \alpha,\beta)$ in (8.2-7), then let p_i be the probability that an observation (claim) occurs in the interval (z_{i-1},z_i), i.e.

$$p_i = G(z_i ; \alpha, \beta) - G(z_{i-1} ; \alpha,\beta) \qquad (8.3\text{-}1)$$

for $i = 2, 3, \ldots, k$

For $i = 1$, taking $z_0 = 0$,

$$p_1 = G(z_1 ; \alpha, \beta) \qquad (8.3\text{-}2)$$

Therefore, the sample likelihood function will be proportional to L where

$$L = \prod_{i=1}^{k} p_i^{n_i} \qquad (8.3\text{-}3)$$

and the loglikelihood function will be

$$\log L = \sum_{i=1}^{k} n_i \log p_i \qquad (8.3\text{-}4)$$

where p_i is as defined above. The values of $G(z_i\ ; \alpha, \beta)$ are calculated from (8.2-10).

The maximum likelihood estimates of the parameters, $(\hat{\alpha}, \hat{\beta})$, are obtained by maximizing log L simultaneously with respect to α and β. To avoid solving non-linear equations (which is a laborious task when k, the number of intervals, is large) we suggest the computing technique described in section 3.7.7 for maximizing log L directly with respect to the unknown parameters. To start the iteration process we can use the moment estimates of α and β. For grouped data the sample moments can be calculated by assuming that in each interval (z_{i-1}, z_i), all the n_i observations are concentrated at the mid-point of the interval, i.e., at $(z_{i-1} + z_i)/2$. The sample first and second moments when calculated can be put equal to their corresponding population values to yield the moment estimates of α and β.

8.4 Application of the Gamma Distribution to the AD Square Root Data

To examine the adequacy of the gamma distribution as a model for the distribution of the square root of the AD claim amount we wrote computer program P25 for estimating parameters α and β from a sample of grouped data. The ML method described in section 8.3 was used. The samples of the square root AD data, which were described in section 8.1, were input to the program. For each sample the program computed and printed out the ML estimates of α and β, the shape and scale parameters respectively, of the gamma distribution. The mean and standard deviation of the claim amounts according to the fitted model were also produced. The program also provided an extensive table of results, giving in particular the values of the expected claim numbers, the expected loss and the contribution to the total X^2 statistic for each claim amount interval. The results for the square root AD data are presented in tables (8.1.1) to (8.1.6). The original grouping of the data was in bands of length 2 but, for convenience in interpretation, we have squared the boundaries of each interval and, therefore, the grouping in the tables is in terms of the amount of claim and not its square root. From these tables we can see that $\hat{\alpha}$ is greater than 1 for all samples. Hence the model portrays the mode of the sample histogram. The actual minus expected claim numbers are not small and in few of the intervals the contributions to the X^2 statistic are large. The X^2 statistics are significant, at 0.05 level, for all the samples. These indicate significant differences between the model and actual sample values.

The expected loss statistic for each interval was calculated by multiplying the actual minus expected claim number by the square of the mid-point of the original interval in terms of Z, the square root of the claim amount. The values of this statistic show considerable disagreement, in monetary terms, in the middle range of the claim amount distribution. Our analysis shows that the 2-parameter gamma distribution is not an appropriate model for the distribution of the square root of AD claim amounts. The fit of this model to the AD square root data is worse than that of the 2-parameter lognormal distribution to the AD data. Hence it seems that collection of data on the square root of the claim amount is an unnecessary task. To ponder on why the fit of the model should be so poor we should examine the two assumptions which we made in section 8.1. These were :

(i) - Z, the square root of the claim amount X, is proportional to the velocity of the car at the time of impact (since X is proportional to $\frac{1}{2} mV^2$, the kinetic energy of the car at the time of impact).

(ii) - The velocity of the car at the time of impact has a gamma distribution

As regards assumption (i), it may be more reasonable to assume that the claim amount X is proportional to mV, the momentum of the car at the time of impact, and not to its kinetic energy. Hence, if assumption (ii) holds, we should expect the claim amount X to have a gamma distribution. To examine this on our AD data samples, presented in tables (1.1) to (1.7), we used the method of moments to estimate the parameters α and β. The estimates of α were less than 1 for each sample, hence indicating that the gamma model had no mode to portray that of the histogram of the sample values. The model also had a much shorter tail than the sample histogram. The values of the X^2 statistic were calculated from each sample and its corresponding gamma model.. These were consistently large and indicated significant differences between the model and actual sample values. Again the fit was considerably poorer than that of the 2-parameter lognormal. This is not totally unexpected since the lognormal model has a mode and a longer tail than the gamma and is, therefore, more capable of representing the mode and the long tail of the claim amount histogram.

There are no statistics available on the velocity of cars at the time of impact to allow the examination of assumption (ii). Therefore, even if assumption (i) or its revised form (X proportional to mV) is considered reasonable, there is no reason why Z, the square root of claim amount X, (or X itself) should have a gamma distribution since it is possible for V to have some other distribution.

8.5 The Effects of Inflation on the Parameters of the Model

Although the gamma model proved not to be appropriate for the distribution of the square root of the AD claim amounts, however, to make this chapter complete, we consider the effects of inflation on the parameters of the model. We follow the exposition of section 3.9 and investigate these effects theoretically. As was mentioned in that section, we would need to allow for such effects when a prediction of the future distribution of the square root of claim amounts is required. Let us, therefore, again assume that the effect of inflation is to increase a claim of amount X to $U = X(1 + i)$ over a period of time where i is the effective rate of inflation, according to some suitable index of wages or prices, for that period. If $Z = \sqrt{X}$ has a gamma distribution with p.d.f. $f_G(z ; \alpha, \beta)$, as given by (8.2-1), then by a transformation of variables it can be shown that

$$\sqrt{U} = rZ \quad \text{where } r = (1 + i)^{\frac{1}{2}}$$

will have a gamma distribution with parameters α and $r\beta$. Therefore inflation affects β and leaves α, the shape parameter, unchanged. This means that when a gamma model is found appropriate for the distribution of Z, the square root of the claim amount from a particular class of insurance business, we can "update" the scale parameter β, with respect to inflation, and hence obtain the distribution of Z in a future period.

8.6 Prediction for the AD Square Root Data

For the purpose of demonstrating our prediction technique for the distribution of square roots of claim amounts (despite the inadequacy of the gamma model for the AD data) we wrote computer program P26. The ML estimates $\hat{\alpha}$ and $\hat{\beta}$ for a particular period of accident and the rate of inflation, calculated from the General Index of Retail Prices, are input to the program. The parameters of the gamma model for the future period are then calculated and the program prints out an extensive table of results for comparing the actual and predicted distributions of the square root of the claim amount. To see the result of predicting one year ahead we used 1976/3rd and 1976/4th quarters to predict the distributions for 1977/3rd and 1977/4th quarters respectively. The results are presented in tables (8.2.1) and (8.2.2). The tables are self-explanatory, and we can see that the actual minus expected number of claims is large for some intervals and that the Chi-square statistics are very large and indicate highly significant differences between the predicted and actual sample values. We did not expect to obtain satisfactory results, since the fit of the model to the actual data of 1976/3rd and 4th quarters was so poor in both cases.

8.7 Conclusions

As we mentioned in the introduction to this chapter, encouraged by a remark by Beard (1978), we studied the gamma distribution as a model for the distribution of the square root of the claim amount. The properties of the gamma distribution were studied and the ML method was considered for estimation of its parameters from grouped data. In order to show the application of this model in practice, we obtained six samples of the square root of AD claim amounts. The model was fitted to each sample by ML method and in each case it exhibited a sharper peak and a shorter tail than the histogram of the sample values. The Chi-square goodness-of-fit test statistic indicated highly significant differences between the model and actual sample values. Judged by this statistic, the fit in each case was worse than that of the 2-parameter lognormal model to the AD claim amounts as shown in chapter 3. In section 8.4, based on the assumption that the amount of claim should be proportional to the momentum of the car at the time of impact, we examined the gamma distribution as a model for the distribution of the claim amount itself. Hence we fitted this model to the AD samples of tables (1.1) to (1.7). The values of α, the shape parameter, were less than 1 in each case. This indicated that the model ignored the distinct mode of the actual distribution of claim amounts. The model also exhibited a much shorter tail than the histograms of the samples. Judging by X^2 statistic, a very poor fit was obtained. In each case, the fit was worse than that of the 2-parameter lognormal distribution as shown in chapter 3. This was not surprising since the lognormal distribution has a longer tail than the gamma and also always possesses a mode.

For the sake of a more complete study, the effects of inflation on the parameters of the gamma model, when used for the distribution of the square root of the claim amount, were considered. It was shown that inflation increases β, the scale parameter, but leaves α, the shape parameter, unchanged. Predictions were made, using two samples of our square root AD data and, as expected from the poor fit of the gamma model to the sample values, the predicted distributions were shown, by the X^2 statistic, to differ significantly from the actual ones.

We therefore reject the gamma distribution as a model for the distribution of AD claim amounts or the square roots of claim amounts. For the latter, we believe that the task of collecting data on the square root of claim amount is unnecessary when more adequate models, like the lognormal, exist for the distribution of claim amount itself.

8.8 Tables

Table (8.1.1)

*** TWO-PARAMETER GAMMA DIS. ***

76/3RD QUARTER DATA

ESTIMATION BY MULTINOMIAL MAX.LIKELIHOOD METHOD :-

ALPHA = 3.474 BETA= 3.866

MEAN= 232.223 S.D.= 262.746

AMOUNT £	ACTUAL CL. NO.	EXPECTED CL. NO.	ACTUAL-EXPECTED	EXPECTED LOSS	(A-E)**2/E
UPTO 4	14.	13.	1.	1.	0.077
4- 16	90.	86.	4.	36.	0.186
16- 36	147.	178.	-31.	-777.	5.399
36- 64	240.	243.	-3.	-147.	0.037
64- 100	297.	271.	26.	2108.	2.494
100- 144	290.	265.	25.	3028.	2.358
144- 196	260.	239.	21.	3552.	1.845
196- 256	188.	204.	-16.	-3602.	1.255
256- 324	175.	166.	9.	2603.	0.488
324- 400	119.	130.	-11.	-3973.	0.931
400- 484	76.	99.	-23.	-10148.	5.343
484- 576	57.	74.	-17.	-8997.	3.905
576- 676	49.	54.	-5.	-3126.	0.463
676- 784	45.	39.	6.	4376.	0.923
784- 900	28.	28.	0.	0.	0.000
900-1024	21.	20.	1.	961.	0.050
1024-1156	18.	14.	4.	4357.	1.143
1156-1296	17.	9.	8.	9803.	7.111
1296-1444	4.	6.	-2.	-2739.	0.667
1444-1600	2.	4.	-2.	-3043.	
1600-1764	8.	3.	5.	8407.	1.296
1764-1936	1.	2.	-1.	-1849.	
1936-2116	1.	1.	0.	0.	
2116-2304	2.	1.	1.	2209.	
2304-2500	2.	1.	1.	2401.	
2500-2704	0.	0.	0.	0.	
2704-2916	0.	0.	0.	0.	
2916-3136	0.	0.	0.	0.	
3136-3364	0.	0.	0.	0.	
3364-3600	0.	0.	0.	0.	
3600-3844	0.	0.	0.	0.	
3844-4096	0.	0.	0.	0.	
4096-4356	0.	0.	0.	0.	
4356-4624	0.	0.	0.	0.	
4624-4900	0.	0.	0.	0.	
4900-5184	0.	0.	0.	0.	
5184-5476	0.	0.	0.	0.	
5476-5776	1.	0.	1.	5626.	0.800
TOTAL	2152.	2150.		11067.	36.762

D.F. = 16

TOTAL EXP. LOSS / TOTAL ACT. COST = 2.2 %

P < .01

Table (8.1.2)

*** TWO-PARAMETER GAMMA DIS. ***

76/4TH QUARTER DATA

ESTIMATION BY MULTINOMIAL MAX.LIKELIHOOD METHOD :-
ALPHA = 3.733 BETA= 3.888

MEAN= 266.975 S.D.= 290.604

AMOUNT £	ACTUAL CL. NO.	EXPECTED CL. NO.	ACTUAL-EXPECTED	EXPECTED LOSS	(A-E)**2/E
UPTO 4	11.	10.	1.	1.	0.100
4- 16	107.	82.	25.	226.	7.622
16- 36	145.	192.	-47.	-1177.	11.505
36- 64	266.	287.	-21.	-1030.	1.537
64- 100	361.	342.	19.	1541.	1.056
100- 144	361.	354.	7.	848.	0.138
144- 196	359.	334.	25.	4228.	1.871
196- 256	325.	296.	29.	6529.	2.841
256- 324	252.	249.	3.	868.	0.036
324- 400	200.	202.	-2.	-722.	0.020
400- 484	137.	159.	-22.	-9707.	3.044
484- 576	112.	122.	-10.	-5292.	0.820
576- 676	80.	92.	-12.	-7503.	1.565
676- 784	61.	68.	-7.	-5105.	0.721
784- 900	51.	49.	2.	1683.	0.082
900-1024	31.	35.	-4.	-3845.	0.457
1024-1156	28.	25.	3.	3268.	0.360
1156-1296	24.	18.	6.	7352.	2.000
1296-1444	13.	12.	1.	1369.	0.083
1444-1600	8.	8.	0.	0.	0.000
1600-1764	4.	6.	-2.	-3363.	0.667
1764-1936	9.	4.	5.	9247.	
1936-2116	1.	3.	-2.	-4051.	1.286
2116-2304	2.	2.	0.	0.	
2304-2500	2.	1.	1.	2401.	
2500-2704	1.	1.	0.	0.	
2704-2916	0.	1.	-1.	-2810.	
2916-3136	0.	0.	0.	0.	
3136-3364	0.	0.	0.	0.	
3364-3600	1.	0.	1.	3482.	
3600-3844	1.	0.	1.	3722.	
3844-4096	0.	0.	0.	0.	
4096-4356	0.	0.	0.	0.	
4356-4624	1.	0.	1.	4490.	1.800
TOTAL	2954.	2954.		6648.	39.610

D.F. = 20

$$\frac{\text{TOTAL EXP. LOSS}}{\text{TOTAL ACT. COST}} = 0.8\ \%$$

P < .01

Table (8.1.3)

*** TWO-PARAMETER GAMMA DIS. ***

77/1ST QUARTER DATA

ESTIMATION BY MULTINOMIAL MAX.LIKELIHOOD METHOD :-
ALPHA = 3.822 BETA= 3.790

MEAN= 264.715 S.D.= 284.505

AMOUNT £	ACTUAL CL. NO.	EXPECTED CL. NO.	ACTUAL-EXPECTED	EXPECTED LOSS	(A-E)**2/E
UPTO 4	7.	7.	0.	0.	0.000
4- 16	85.	62.	23.	208.	8.532
16- 36	107.	151.	-44.	-1102.	12.821
36- 64	206.	229.	-23.	-1129.	2.310
64- 100	280.	275.	5.	405.	0.091
100- 144	328.	286.	42.	5087.	6.168
144- 196	299.	271.	28.	4736.	2.893
196- 256	262.	239.	23.	5178.	2.213
256- 324	196.	201.	-5.	-1446.	0.124
324- 400	151.	163.	-12.	-4334.	0.883
400- 484	104.	127.	-23.	-10148.	4.165
484- 576	79.	97.	-18.	-9526.	3.340
576- 676	62.	73.	-11.	-6878.	1.658
676- 784	50.	53.	-3.	-2188.	0.170
784- 900	45.	38.	7.	5889.	1.289
900-1024	27.	27.	0.	0.	0.000
1024-1156	22.	19.	3.	3268.	0.474
1156-1296	8.	13.	-5.	-6127.	1.923
1296-1444	16.	9.	7.	9586.	5.444
1444-1600	8.	6.	2.	3043.	0.667
1600-1764	6.	4.	2.	3363.	
1764-1936	4.	3.	1.	1849.	1.286
1936-2116	5.	2.	3.	6076.	
2116-2304	2.	1.	1.	2209.	
2304-2500	0.	1.	-1.	-2401.	
2500-2704	0.	1.	-1.	-2602.	
2704-2916	0.	0.	0.	0.	
2916-3136	1.	0.	1.	3026.	
3136-3364	0.	0.	0.	0.	
3364-3600	0.	0.	0.	0.	
3600-3844	0.	0.	0.	0.	
3844-4096	0.	0.	0.	0.	
4096-4356	1.	0.	1.	4226.	3.200
TOTAL	2361.	2358.		10268.	59.654

TOTAL EXP. LOSS / TOTAL ACT. COST = 1.6 %

D.F. = 19

P < .01

Table (8.1.4)

*** TWO-PARAMETER GAMMA DIS. ***

77/2ND QUARTER DATA

ESTIMATION BY MULTINOMIAL MAX.LIKELIHOOD METHOD :-
ALPHA = 3.746 BETA= 3.941

MEAN= 276.198 S.D.= 300.056

AMOUNT £	ACTUAL CL. NO.	EXPECTED CL. NO.	ACTUAL-EXPECTED	EXPECTED LOSS	(A-E)**2/E
UPTO 4	5.	6.	-1.	-1.	0.167
4- 16	68.	51.	17.	154.	5.667
16- 36	93.	120.	-27.	-676.	6.075
36- 64	162.	182.	-20.	-981.	2.198
64- 100	236.	219.	17.	1379.	1.320
100- 144	237.	229.	8.	969.	0.279
144- 196	240.	218.	22.	3721.	2.220
196- 256	209.	195.	14.	3152.	1.005
256- 324	153.	165.	-12.	-3470.	0.873
324- 400	137.	135.	2.	722.	0.030
400- 484	99.	107.	-8.	-3530.	0.598
484- 576	71.	83.	-12.	-6351.	1.735
576- 676	58.	63.	-5.	-3126.	0.397
676- 784	51.	47.	4.	2917.	0.340
784- 900	31.	34.	-3.	-2524.	0.265
900-1024	22.	25.	-3.	-2884.	0.360
1024-1156	14.	18.	-4.	-4357.	0.889
1156-1296	15.	13.	2.	2451.	0.308
1296-1444	10.	9.	1.	1369.	0.111
1444-1600	10.	6.	4.	6086.	2.667
1600-1764	2.	4.	-2.	-3363.	
1764-1936	5.	3.	2.	3699.	0.000
1936-2116	5.	2.	3.	6076.	
2116-2304	0.	1.	-1.	-2209.	
2304-2500	1.	1.	0.	0.	
2500-2704	2.	1.	1.	2602.	
2704-2916	0.	0.	0.	0.	
2916-3136	1.	0.	1.	3026.	3.200
TOTAL	1937.	1937.		4848.	30.702

D.F. = 19

TOTAL EXP. LOSS / TOTAL ACT. COST = 0.9 %

P < .05

Table (8.1.5)

*** TWO-PARAMETER GAMMA DIS. ***

77/38D QUARTER DATA

ESTIMATION BY MULTINOMIAL MAX.LIKELIHOOD METHOD :-
ALPHA = 3.331 BETA= 4.372

MEAN= 275.774 S.D.= 319.158

AMOUNT £	ACTUAL CL. NO.	EXPECTED CL. NO.	ACTUAL-EXPECTED	EXPECTED LOSS	(A-E)**2/E
UPTO 4	9.	10.	-1.	-1.	0.100
4- 16	96.	59.	37.	334.	23.203
16- 36	67.	122.	-55.	-1378.	24.795
36- 64	159.	169.	-10.	-491.	0.592
64- 100	191.	192.	-1.	-81.	0.005
100- 144	189.	195.	-6.	-727.	0.185
144- 196	210.	182.	28.	4736.	4.308
196- 256	174.	161.	13.	2927.	1.050
256- 324	158.	136.	22.	6362.	3.559
324- 400	107.	112.	-5.	-1806.	0.223
400- 484	86.	90.	-4.	-1765.	0.178
484- 576	74.	70.	4.	2117.	0.229
576- 676	56.	54.	2.	1251.	0.074
676- 784	32.	41.	-9.	-6563.	1.976
784- 900	21.	31.	-10.	-8413.	3.226
900-1024	13.	23.	-10.	-9613.	4.348
1024-1156	13.	17.	-4.	-4357.	0.941
1156-1296	11.	12.	-1.	-1225.	0.083
1296-1444	14.	9.	5.	6847.	2.778
1444-1600	5.	6.	-1.	-1521.	0.167
1600-1764	4.	4.	0.	0.	0.000
1764-1936	4.	3.	1.	1849.	
1936-2116	2.	2.	0.	0.	0.200
2116-2304	2.	2.	0.	0.	
2304-2500	3.	1.	2.	4803.	
2500-2704	1.	1.	0.	0.	
2704-2916	0.	1.	-1.	-2810.	
2916-3136	1.	0.	1.	3026.	
3136-3364	0.	0.	0.	0.	
3364-3600	0.	0.	0.	0.	
3600-3844	0.	0.	0.	0.	
3844-4096	0.	0.	0.	0.	
4096-4356	0.	0.	0.	0.	
4356-4624	2.	0.	2.	8979.	3.200
TOTAL	1704.	1705.		2479.	75.419

D.F. = 20

TOTAL EXP. LOSS
---------------- = 0.5 %
TOTAL ACT. COST

P <.01

Table (8.1.6)

*** TWO-PARAMETER GAMMA DIS. ***

77/4TH QUARTER DATA

ESTIMATION BY MULTINOMIAL MAX.LIKELIHOOD METHOD :-
ALPHA = 2.698 BETA= 5.387

MEAN= 289.606 S.D.= 375.684

AMOUNT £	ACTUAL CL. NO.	EXPECTED CL. NO.	ACTUAL-EXPECTED	EXPECTED LOSS	(A-E)**2/E
UPTO 4	3.	6.	-3.	-3.	1.500
4- 16	50.	25.	25.	226.	25.000
16- 36	19.	42.	-23.	-576.	12.595
36- 64	31.	51.	-20.	-981.	7.843
64- 100	57.	54.	3.	243.	0.167
100- 144	54.	52.	2.	242.	0.077
144- 196	67.	48.	19.	3213.	7.521
196- 256	42.	42.	0.	0.	0.000
256- 324	53.	36.	17.	4916.	8.028
324- 400	24.	30.	-6.	-2167.	1.200
400- 484	17.	25.	-8.	-3530.	2.560
484- 576	13.	20.	-7.	-3705.	2.450
576- 676	22.	16.	6.	3752.	2.250
676- 784	10.	12.	-2.	-1459.	0.333
784- 900	9.	10.	-1.	-841.	0.100
900-1024	13.	7.	6.	5768.	5.143
1024-1156	0.	6.	-6.	-6536.	6.000
1156-1296	0.	4.	-4.	-4901.	
1296-1444	0.	3.	-3.	-4108.	7.000
1444-1600	3.	2.	1.	1521.	
1600-1764	1.	2.	-1.	-1681.	0.000
1764-1936	5.	1.	4.	7398.	
1936-2116	2.	1.	1.	2025.	
2116-2304	3.	1.	2.	4419.	
2304-2500	0.	1.	-1.	-2401.	
2500-2704	1.	0.	1.	2602.	
2704-2916	1.	0.	1.	2810.	16.000
TOTAL	500.	497.		6244.	105.767

D.F. = 17

TOTAL EXP. LOSS
--------------- = 4.3 %
TOTAL ACT. COST

P <.01

Table (8.2.1)

*** TWO-PARAMETER GAMMA DIS. ***

PREDICTION OF 77/3RD QUARTER CLAIMS COST
USING 76/3RD QUARTER MULT. MAXL. PARAMETERS :

ALPHA= 3.474 BETA= 3.866
INFLATION RATE I=16.5% CALCULATED FROM :
GENERAL INDEX OF RETAIL PRICES.

PREDICTION PARAMETERS ARE :- ALPHA= 3.474 BETA= 4.173
MEAN CLAIM AMOUNT= 270.63 S.D.=306.19
ACTUAL 77/3RD PARAMETERS :- ALPHA= 3.331 BETA= 4.372
MEAN CLAIM AMOUNT= 275.75 S.D.=319.15

AMOUNT £	ACT.NO.	EXP.NO.	A-E	EXP.LOSS	(A-E)**2/E
UPTO 4	9.	8.	1.	1.	0.125
4- 16	96.	55.	41.	370.	30.564
16- 36	67.	119.	-52.	-1303.	22.723
36- 64	159.	169.	-10.	-491.	0.592
64- 100	191.	195.	-4.	-324.	0.082
100- 144	189.	199.	-10.	-1211.	0.503
144- 196	210.	186.	24.	4059.	3.097
196- 256	174.	164.	10.	2252.	0.610
256- 324	158.	139.	19.	5494.	2.597
324- 400	107.	113.	-6.	-2167.	0.319
400- 484	86.	90.	-4.	-1765.	0.178
484- 576	74.	70.	4.	2117.	0.229
576- 676	56.	53.	3.	1876.	0.170
676- 784	32.	40.	-8.	-5834.	1.600
784- 900	21.	29.	-8.	-6730.	2.207
900-1024	13.	22.	-9.	-8652.	3.682
1024-1156	13.	16.	-3.	-3268.	0.563
1156-1296	11.	11.	0.	0.	0.000
1296-1444	14.	8.	6.	8216.	4.500
1444-1600	5.	6.	-1.	-1521.	0.167
1600-1764	4.	4.	0.	0.	
1764-1936	4.	3.	1.	1849.	
1936-2116	2.	2.	0.	0.	
2116-2304	2.	1.	1.	2209.	
2304-2500	3.	1.	2.	4803.	
2500-2704	1.	1.	0.	0.	
2704-2916	0.	0.	0.	0.	
2916-3136	1.	0.	1.	3026.	
3136-3364	0.	0.	0.	0.	
3364-3600	0.	0.	0.	0.	
3600-3844	0.	0.	0.	0.	
3844-4096	0.	0.	0.	0.	
4096-4356	0.	0.	0.	0.	
4356-4624	2.	0.	2.	8979.	
TOTAL	1704	1704		11985.	74.504

CHI SQ. STAT. = 81.8 , D.F.=21 P <.01

TOTAL ACTUAL COST = 472208.
TOTAL EXPECTED COST = 460223.

TOTAL EXP. LOSS
--------------- = 2.54 %
TOTAL ACT. COST

Table (8.2.2)

*** TWO-PARAMETER GAMMA DIS. ***

PREDICTION OF 77/4TH QUARTER CLAIMS COST
USING 76/4TH QUARTER MULT. MAXL. PARAMETERS :

ALPHA= 3.733 BETA= 3.888
INFLATION RATE I=13.0% CALCULATED FROM :
GENERAL INDEX OF RETAIL PRICES.

PREDICTION PARAMETERS ARE :- ALPHA= 3.733 BETA= 4.133
MEAN CLAIM AMOUNT= 301.80 S.D.=328.50
ACTUAL 77/4TH PARAMETERS :- ALPHA= 2.698 BETA= 5.387
MEAN CLAIM AMOUNT= 289.54 S.D.=375.62

AMOUNT £	ACT.NO.	EXP.NO.	A-E	EXP.LOSS	(A-E)**2/E
0- 4	3.	1.	2.	2.	
4- 16	50.	12.	38.	343.	120.333
16- 36	19.	28.	-9.	-225.	2.893
36- 64	31.	43.	-12.	-589.	3.349
64- 100	57.	53.	4.	324.	0.302
100- 144	54.	56.	-2.	-242.	0.071
144- 196	67.	55.	12.	2030.	2.618
196- 256	42.	50.	-8.	-1801.	1.280
256- 324	53.	44.	9.	2603.	1.841
324- 400	24.	36.	-12.	-4334.	4.000
400- 484	17.	29.	-12.	-5295.	4.966
484- 576	13.	23.	-10.	-5292.	4.348
576- 676	22.	18.	4.	2501.	0.889
676- 784	10.	14.	-4.	-2917.	1.143
784- 900	9.	10.	-1.	-841.	0.100
900-1024	13.	8.	5.	4807.	3.125
1024-1156	0.	6.	-6.	-6536.	6.000
1156-1296	0.	4.	-4.	-4901.	
1296-1444	0.	3.	-3.	-4108.	
1444-1600	3.	2.	1.	1521.	
1600-1764	1.	1.	0.	0.	
1764-1936	5.	1.	4.	7398.	
1936-2116	2.	1.	1.	2025.	
2116-2304	3.	0.	3.	6628.	
2304-2500	0.	0.	0.	0.	
2500-2704	1.	0.	1.	2602.	
2704-2916	1.	0.	1.	2810.	
TOTAL	500	498		-1489.	157.258

CHI SQ. STAT. = 191.6 , D.F.=17 P <.01

TOTAL ACTUAL COST = 145477.
TOTAL EXPECTED COST = 146966.

TOTAL EXP. LOSS / TOTAL ACT. COST = -1.02 %

CHAPTER 9

Summary of Conclusions

In this work several statistical distributions were examined as models for the distribution of claim amounts in general insurance. A chapter was devoted to the study of each. The results of examining the applicability of different distributions were summarized in the concluding sections of the various chapters. In those sections, where applicable, recommendations were made as to procedures to be adopted when testing whether a sample can be regarded as drawn from a particular distribution. The preferred method (methods) of parameter estimation for the model was also suggested and its performance on the accidental damage (AD) data was reviewed. In addition, the effects of inflation on the parameters of the model were summarized and, with appropriate models, the performance of the prediction technique on the AD data was reported. To avoid unnecessary repitition we suggest that for a summary of the findings on each model reference should be made to the relevant conclusions at the end of each chapter. The purpose of this chapter is to summarize those aspects of the present work which are common to all the models considered and which may be useful in fitting distributions to general insurance data.

When fitting statistical models to empirical data we have to deal with two problems. Firstly the estimation of the parameters of the model from the sample and secondly the testing of the agreement between the model and actual sample values. Due to the large number of claims, general insurance claim amounts data are usually in grouped frequency format. Therefore the efficient method of multinomial maximum likelihood (MML) is recommended for estimation from grouped data. Without the aid of computers this method is not always practicable. However, as computers are now commonly available the MML method should present no problems when our proposed computing technique is used to find the estimates of the parameters. Applications of this method were demonstrated on the AD data for various distributions. With some models, like the 3-parameter lognormal or Weibull, it is possible to use the least squares (LS) method of estimation which is simpler than the MML method but less efficient. In such circumstances our LS computing technique is recommended in place of solving a system of non-linear equations. When no computer is at hand one of the other simpler but less efficient methods of estimation may be adopted.

Due to the importance of testing the agreement between a fitted model and actual sample values we devoted Chapter 2 to the study of some goodness-of-fit tests. The Chi-square is the one test which is widely used by actuaries and others. As it is not possible to interpret the value of this statistic in monetary terms, we recommended supplementing it by our proposed total expected loss statistic T. The T statistic is a measure of overall agreement in monetary terms between the model and the actual sample values. Another goodness-of-fit test which is usually more 'powerful' than the Chi-square is the Kolmogorov-Smirnov test. This test does not appear to have been exploited in actuarial work. It is, therefore, dealt with in Chapter 2 where the application procedure for its use is also described. In practice, we recommend that, when conditions allow (see Chapter 2), all three statistics be used since in some cases they may tell a slightly different story. In our analysis of the AD data whenever possible we used all these statistics for examining the goodness-of-fit of the models to the actual sample values.

One of the purposes of model fitting is the prediction of the future distribution of claim amounts and hence the future cost of claims to the insurer. Inflation is believed to be the main cause of increase in the claim amount over time. Therefore, its effects on the parameters of each model were studied. A prediction technique was proposed which consisted of updating the parameters of the past model with respect to a given rate of inflation, as calculated from a suitable index of wages or prices. These updated parameters were then used in the model for the distribution of future claim amounts. Whenever a suitable model was found for the AD claim amounts this technique was deomonstrated on the past data and shown to perform very well (i.e. in terms of the goodness-of-fit tests, no significant differences were observed between the predicted and actual distributions). It was also shown that for AD claim amounts the appropriate index, for the calculation of the rate of inflation, is the General Index of Retail Prices. The effectiveness of our proposed total expected loss statistic, T, is clearly demonstrated when judging the agreement between the predicted and actual distributions since in that case it shows by how much, in monetary terms, we have overpredicted or underpredicted the total cost of claims. The advantage of using a statistical model was made apparent in section 3.14.2 by the fact that it gave a smaller over-prediction of the total cost of claims than the simple method of updating the sample mean with respect to inflation. Therefore, the distribution theory approach has the merit of allowing the setting up of smaller, but accurate, reserves.

It is not possible to recommend any one distribution as the best model for the distribution of general insurance claim amounts since different models may fit data from different classes of business.

Care must be taken to adopt a model which represents the special features of the data. For example, if the actual samples consistently exhibit a mode near the origin, we must use a model like the lognormal or the inverse Gaussian rather than the Weibull or the gamma which have no mode when the value of their shape parameter is less than 1. If the histograms of the samples have very long tails then the lognormal, or the inverse Gaussian, would be more appropriate than the Weibull or the gamma since the former models have a longer tail. If we are interested in the distribution of large claims, the truncated lognormal or the Pareto distributions should be found more suitable. If two models are found to fit the same data equally well, then the one which has a thicker tail, and which may overpredict the total cost of claims slightly more, should be preferred. This is because it is wiser to overpredict slightly rather than underpredict the total cost of claims. We recommend the 3-parameter lognormal distribution as the model for the distribution of the AD claim amounts. The location parameter, c, of this model was interpreted as the amount of voluntary excess on the policy. This distribution clearly exhibited the mode and the long tail of the histograms of the samples. The agreements between the model and the sample values and between the predicted models and the sample values were satisfactory. Our proposed 3-parameter inverse Gaussian distribution also presented the above features and gave almost as good a fit to the data. However, the lognormal distribution seems more desirable because we have a theoretical justification for it and because tests of lognormality are available. In addition a slightly better fit to the AD data was provided by this distribution. For the distribution of large AD claim amounts we recommend the truncated lognormal on the basis of its good fit to the data, but almost as good results may be obtained by the simpler to use Pareto distribution.

We believe that with the wide availability of computers all models are relatively easy to use. Therefore, it must be emphasised that the particular features of the data rather than the ease of use should be the criteria in selecting a model.

APPENDIX

Computer Programs

1 - For each program the section number refers to the relevant section in the text where the use of that particular computer program is mentioned for the first time.

```
      MASTER CUM-DIS-AND-RELEVANT-STATS-CALCULATION

C                  *** PROGRAM P1 ***
C
C***** A PROGRAM TO CALCULATE CUMULATIVE PERCENTAGES AND RELEVANT
C***** STATISTICS AND PRINT RESULTS. ACTUAL DATA IN HISTOGRAM
C***** FORM IS ALSO PRINTED.

      DIMENSION X(100),F(100),CSUM(100),NCLMGX(100),NAME(20)
C***** F HOLDS CLAIM NO.S ,CSUM HOLDS CUM. %S,N IS NO. OF RUNS
C***** M IS NO. OF CLAIM AMOUNT BANDS,NAME HOLDS A CAPTION.
C***** X HOLDS THE MID-POINTS OF THE INTERVALS, NCLMGX HOLDS THE
C***** NO. OF CLAIMS WHOSE VALUES ARE GREATER THAN THE LOWER
C***** BOUND OF THE I TH INTERVAL.

      REAL MEDIAN,MODE,MEW3,MEW4
      READ(5,100)N
  100 FORMAT(I0)
      DO 50 J=1,N
      READ(5,110)NAME,M
  110 FORMAT(20A4/I0)
      WRITE(6,900)NAME
  900 FORMAT(//////////,23X,20A4,
     $///,10X,'  AMOUNT £',6X,'NO. OF CLAIMS',6X,'CUM. %',4X,
     $'NCLM > LB. AMOUNT')
C***** READ IN THE NO. OF CLAIMS PER BAND.
      READ(5,120)(F(I) ,I=1,M)
  120 FORMAT(F0.0)
C***** FIND CUM. DIS. OF CLAIM.
      CSUM(1)=F(1)
      DO 1 I=2,M
      CSUM(I)=CSUM(I-1)+F(I)
    1 CONTINUE
      S=CSUM(M)
      IS=S
      NCLMGX(1)=IS
      DO 11 I=2,M
      NCLMGX(I)=IS-CSUM(I-1)
   11 CONTINUE
C***** FIND CUM. %S.
      DO 5 I=1,M
      CSUM(I)=100.*CSUM(I)/S
      NFI=F(I)
C***** CALCULATE BOUNDS FOR THE INTERVALS.
      IF(I-20)2,2,3
    2 IUB=30*I
      LB=IUB-29
      GOTO 4
    3 IUB=(I-14)*100
      LB=IUB-99
    4 WRITE(6,910)LB,IUB,NFI,CSUM(I),NCLMGX(I)
  910 FORMAT(10X,I5,'-',I4,10X,I5,10X,F6.2,8X,I6)
    5 CONTINUE
C***** FORM THE MID-VALUES OF THE INTERVALS.
      X(1)=15.5
      DO 6 I=2,20
      X(I)=30.+X(I-1)
    6 CONTINUE
      X(21)=650.5
      DO 7 I=22,M
      X(I)=100.+X(I-1)
    7 CONTINUE
C***** CALCULATE MOMENTS. S1,S2,S3 AND S4 ARE MOMENTS ABOUT ZERO.
C***** VAR,MEW3 AND MEW4 ARE MOMENTS ABOUT THE MEAN.
C***** COFVAR IS COEFFICIENT OF VARIATION.SQRTB1 IS SKEWNESS.
C***** B2 IS KURTOSIS.
      S1=0.
      S2=0.
      S3=0.
      S4=0.
      SLOG1=0.
      SLOG2=0.
      SLOG3=0.
      SLOG4=0.
      DO 8 I=1,M
      XI=X(I)
      XLOG=ALOG(XI)
      FI=F(I)
      FXI=FI*XI
      FLOGXI=FI*XLOG
      S1=S1+FXI
      S2=S2+FXI*XI
      S3=S3+FXI*XI*XI
      S4=S4+FXI*XI**3
      SLOG1=SLOG1+FLOGXI
      SLOG2=SLOG2+FLOGXI*XLOG
      SLOG3=SLOG3+FLOGXI*XLOG*XLOG
      SLOG4=SLOG4+FLOGXI*XLOG**3
    8 CONTINUE
      XBAR=S1/S
      VAR=S2/S-XBAR*XBAR
      SD=SQRT(VAR)
      COFVAR=SD/XBAR
      MEW3=S3/S-3*XBAR*VAR-XBAR**3
      MEW4=S4/S-4*XBAR*MEW3-6*XBAR*XBAR*VAR-XBAR**4
      SQRTB1=MEW3/(VAR*SD)
      B2=MEW4/(VAR*VAR)
      XLGMN=SLOG1/S
      XLGVAR=SLOG2/S-XLGMN*XLGMN
      XLGM3=SLOG3/S-3.*XLGMN*XLGVAR-XLGMN**3
      XLGM4=SLOG4/S-4.*XLGMN*XLGM3-6.*XLGMN*XLGMN*XLGVAR-XLGMN**4
      G1=XLGM3/XLGVAR**1.5
      G2=XLGM4/(XLGVAR*XLGVAR)-3.
C***** CALCULATE THE MODE OF THE DIS.
      FMAX=0.0
      DO 9 I=1,5
      IF(F(I) .LT. FMAX)GOTO 9
      J=I
      FMAX=F(J)
    9 CONTINUE
      H1=F(J)-F(J-1)
      H2=F(J)-F(J+1)
      MODE=X(J)-15.+30.*H1/(H1+H2)
C***** CALCULATE THE MEDIAN.
      DO 10 I=1,10
      IF(CSUM(I) .GT. 50.)GOTO 12
   10 CONTINUE
   12 D=CSUM(I)-CSUM(I-1)
      C=50.-CSUM(I-1)
      MEDIAN=X(I)-15.+30.*C/D
      WRITE(6,920)IS,XBAR,SD,COFVAR,MEDIAN,MODE,SQRTB1,B2,G1,G2
  920 FORMAT(////,11X,'TOTAL NO. OF CLAIMS=',I5/
     $11X,'MEAN CLAIM AMOUNT  =',F6.2/
     $11X,'STANDARD DEVIATION =',F6.2/
     $11X,'COEFF. OF VARIATION=',F5.2/
     $11X,'MEDIAN ............=',F6.2/
     $11X,'MODE ..............=',F6.2/
     $11X,'SQRTB1 -SKEWNESS   =',F6.2/
     $11X,'B2 ... -KURTOSIS   =',F6.2/
     $11X,'G1 ................=',F6.3/
     $11X,'G2 ................=',F6.3///
     $11X,'G1 IS COEFF. OF SKEWNESS FOR LOG OF CL. AMOUNTS'/
     $11X,'G2 IS EXCESS KURTOSIS FOR LOG OF CL. AMOUNTS.'//////)
   50 CONTINUE
      STOP
      END
      FINISH
```

```
LIBRARY(SUBGROUPNAGF)
LIST(LP)
PROGRAM(P2)
INPUT 5=CR0
OUTPUT 6=LP0
TRACE 1
COMPACT
COMPRESS INTEGER AND LOGICAL
END

      MASTER LOGNORMAL SIMULATION
C
C                         *** PROGRAM P2 ***
C
C     THIS PROGRAM GENERATES N SAMPLES EACH CONTAINING M
C     RANDOM VARIATES FROM LOGNORMAL DIS.(MEW=A , SIGMA=B)
C     Z IS AN INITIALIZATION REAL NUMBER(0<Z<1) WHICH IS REQUIRED
C     BY NAG ROUTINE G05BAF . FOR A GIVEN Z THE SAME CHAIN OF
C     RANDOM VARIATES WILL ALWAYS BE PRODUCED.
C
      DIMENSION L(100),CSUM(100)
      READ(5,100)N
  100 FORMAT(I0)
      DO 10 J=1,N
      READ(5,101)Z,A,B,M
  101 FORMAT(3F0.0,I0)
      WRITE(6,900)M,A,B,Z
  900 FORMAT(//////////,' *******DISTRIBUTION OF ',I5,' LOGNORMAL',
     $' VARIATES WITH',/,' PARAMETERS (MEW=',F6.3,
     $' ;SIGMA=',F6.3,')',/,' INITIALIZATION PARAMETER FOR SIMU',
     $'LATION =',F6.4,////,'    AMOUNT     FREQUENCY     CUM. DIS.')
      DO 1 I=1,60
    1 L(I)=0
C     G05BAF IS A NAG ROUTINE. IT FIXES THE STARTING POINT
C     OF THE CHAIN OF UNIFORM RANDOM VARIATES ACCORDING TO
C     A GIVEN Z.
      CALL G05BAF(Z)
C     KMAX WILL BE THE NO. OF INTERVALS IN THE END
      KMAX=0
      DO 4 I=1,M
C     G05AFF IS NAG ROUTINE WHICH GENERATES A LOGNORMAL RANDOM
C     VARIATE X FROM DIS. WITH PARAMETERS (MEW=A ,SIGMA=B)
      X=G05AFF(A,B)
C     GROUP THE SIMULATED VARIATES ACCORDING TO FORMAT OF AD DATA
C     L(I) IS FREQUENCY IN INTERVAL I
C     CSUM(I) IS CUM. DIS. FUNCTION AT UPPER BOUNDARY OF INT. I
      K=1+X/30.
      IF(K-20)2,2,3
    2 L(K)=L(K)+1
      IF(K .GT. KMAX)KMAX=K
      GOTO 4
    3 K=15+X/100.
      L(K)=L(K)+1
      IF(K .GT. KMAX)KMAX=K
    4 CONTINUE
      CSUM(1)=L(1)
      DO 5 I=2,KMAX
      CSUM(I)=CSUM(I-1)+L(I)
    5 CONTINUE
      DO 9 I=1,KMAX
      IF(I-20)6,6,7
    6 K=(I-1)*30
      N=K+30
      GOTO 8
    7 K=(I-15)*100
      N=K+100
    8 CSUM(I)=100.*CSUM(I)/FLOAT(M)
      WRITE(6,901)K,N,L(I),CSUM(I)
  901 FORMAT(I5,'-',I4,6X,I5,9X,F6.2)
    9 CONTINUE
   10 CONTINUE
      STOP
      END
      FINISH
```

```
LIBRARY(SUBGROUPNAGF)
LIST(LP)
PROGRAM(P3)
INPUT 5=CR0
OUTPUT 6=LP0
TRACE 1
COMPACT
COMPRESS INTEGER AND LOGICAL
END

      MASTER SIMULATION OF 100 LOGNORMAL SAMPLES
C
C                         *** PROGRAM P3 ***
C
C     THIS PROGRAM GENERATES 100 SAMPLES FROM THE LOGNORMAL
C     POP. (MEW=4.5 , SIGMA=1). EACH SAMPLE IS OF SIZE 2500
C
      DIMENSION F(100)
      DO 10 J=1,100
C     READ A REAL NO. Z (0<Z<1) WHICH IS REQUIRED BY NAG ROUTINE
C     G05BAF. FOR A GIVEN Z THE SAME CHAIN OF RANDOM VARIATES
C     IS ALWAYS PRODUCED.
      READ(5,100)Z
  100 FORMAT(F0.0)
      WRITE(6,900)J,Z
  900 FORMAT(15X,'SAMPLE NO. ',I3,5X,'Z=',F9.6)
C     F(I) WILL CONTAIN FREQ. OF INTERVAL I . INITIALIZE F TO ZERO.
      DO 1 I=1,100
    1 F(I)=0.0
C     G05BAF IS NAG ROUTINE WHICH FIXES THE STARTING POINT IN THE
C     CHAIN OF UNIFORM RANDOM NUMBERS ACC. TO GIVEN Z
      CALL G05BAF(Z)
C     KMAX IN THE END WILL BE THE NO. OF INTERVALS
      KMAX=0
      DO 4 I=1,2500
C     G05AFF IS NAG ROUTINE WHICH GENERATES X THE LOGNORMAL
C     RANDOM VARIATE FROM DIS. WITH (MEW=4.5 , SIGMA=1)
      X=G05AFF(4.5,1.0)
C     GROUP THE SIMULATED VARIATES ACC. TO FORMAT OF AD DATA
      K=1+X/30.
      IF(K-20)2,2,3
    2 F(K)=F(K)+1
      IF(K .GT. KMAX)KMAX=K
      GOTO 4
    3 K=15+X/100.
      F(K)=F(K)+1
      IF(K .GT. KMAX)KMAX=K
    4 CONTINUE
      WRITE(6,901)KMAX
  901 FORMAT(2X,I5)
      WRITE(6,902)(F(I),I=1,KMAX)
  902 FORMAT(10(2X,F4.0))
   10 CONTINUE
      STOP
      END
      FINISH
```

```
      LIBRARY(SUBGROUPNAGF)
      LIST(LP)
      PROGRAM(P4)
      INPUT 5=CR0
      OUTPUT 6=LP0
      TRACE 0
      COMPACT
      COMPRESS INTEGER AND LOGICAL
      END

      MASTER EFFICIENCY FOR LOGNORMAL
C
C                   *** PROGRAM P4 ***
C
C     THIS PROGRAM MEASURES THE EFFICIENCY OF VARIOUS METHODS OF
C     ESTIMATION FOR THE 2-PAR LOGNORMAL DISTRIBUTION BY
C     APPLYING THEM TO  SIMULATED SAMPLES
C
      DIMENSION XMID(100),F(100),NAME(20),CSUM(100),CF(100)
      DIMENSION XEND(100),XLOG(100),LB(100),IUB(100),RESULT(21)
      DIMENSION A(6,5),PMEW(6,105),SIG2(6,105)
      DIMENSION ALFA(6,105),BETA(6,105),ARAY(105)
      COMMON A,IND,XMID,CF
C     READ IN THE NO. OF SIMULATED SAMPLES
      READ(5,100)NS
  100 FORMAT(I0)
      DO 60 IRUN =1,NS
C     READ DESCRIPTION OF EACH SAMPLE AND NO. OF ITS INTERVALS
      READ(5,101)NAME,M
  101 FORMAT(20A4/I0)
C     READ IN THE SAMPLE FREQ. DIS.
      READ(5,102)(F(I), I=1,M)
  102 FORMAT(10F0.0)
      CSUM(1)=F(1)
      DO 1 I=2,M
      CSUM(I)=CSUM(I-1)+F(I)
    1 CONTINUE
      TOTAL=CSUM(M)
C     XMID(I) IS MID-POINT OF INT I,XEND IS UPPER BNDRY OF INT. I
C     IUB IS UPPER END OF INT I.  LB IS LOWER  END OF INT I
      XMID(1)=15.5
      XEND(1)=30.5
      IUB(1)=30
      LB(1)=1
      DO 10 I=2,20
      XMID(I)=30.+XMID(I-1)
      XEND(I)=30.+XEND(I-1)
      IUB(I)=XEND(I)
      LB(I)=IUB(I)-29
   10 CONTINUE
      XMID(21)=650.5
      XEND(21)=700.5
      IUB(21)=700
      LB(21)=601
      DO 20 I=22,M
      XMID(I)=100.+XMID(I-1)
      XEND(I)=100.+XEND(I-1)
      IUB(I)=XEND(I)
      LB(I)=IUB(I)-99
   20 CONTINUE
      S1=0.
      S2=0.
      SLOG1=0.
      SLOG2=0.
      DO 30 I=1,M
      FI=F(I)
      XI=XMID(I)
      FXI=FI*XI
      ALOGXI=ALOG(XI)
      FLOGXI=FI*ALOGXI
      S1=S1+FXI
      S2=S2+FXI*XI
      SLOG1=SLOG1+FLOGXI
      SLOG2=SLOG2+FLOGXI*ALOGXI
      CF(I)=100.*CSUM(I)/TOTAL
   30 CONTINUE
      XMEAN=S1/TOTAL
      SS=S2/TOTAL
      HATM1=2.*ALOG(XMEAN)-.5*ALOG(SS)
      S2HAT1=ALOG(SS)-2.*ALOG(XMEAN)
      CALL STATS(HATM1,S2HAT1,1)
      HATM2=SLOG1/TOTAL
      S2HAT2=SLOG2/TOTAL-HATM2*HATM2
      CALL STATS(HATM2,S2HAT2,2)
      XA=QUAN(7.,1)
      XB=QUAN(93.,IND)
      CALL QCAL(XA,XB,93,HATM3,S2HAT3)
      CALL STATS(HATM3,S2HAT3,3)
      XA=QUAN(27.,1)
      XB=QUAN(73.,IND)
      CALL QCAL(XA,XB,73,HATM4,S2HAT4)
      CALL STATS(HATM4,S2HAT4,4)
      HATM5=ALOG(QUAN(50.,1))
      S2HAT5=ALOG(SS/(XMEAN*XMEAN))
      CALL STATS(HATM5,S2HAT5,5)
C     K WILL BE THE NO. OF INT.S WITH NON-ZERO FREQUENCY
      K=0
      DO 40 I=1,M-1
      N2=CSUM(I+1)
      N1=CSUM(I)
      IF(N2 .EQ. N1)GOTO 40
      K=K+1
      XLOG(K)=ALOG(XEND(I))
      T=1.-CF(I)/100.
      Z(K)=HAS(T)
   40 CONTINUE
      XM=0.
      YM=99999.
      IFAIL=0
C     G02CCF IS NAG ROUTINE FOR L.S. REGRESSION
      CALL G02CCF(K,XLOG,Z,XM,YM,RESULT,IFAIL)
      B=RESULT(6)
      C=RESULT(7)
      SHAT6=1./B
      HATM6=-C/B
      S2HAT6=SHAT6*SHAT6
      CALL STATS(HATM6,S2HAT6,6)
      DO 50 J=1,6
      PMEW(J,IRUN)=A(J,1)
      SIG2(J,IRUN)=A(J,2)
      ALFA(J,IRUN)=A(J,3)
      BETA(J,IRUN)=A(J,4)
   50 CONTINUE
   60 CONTINUE
      ALFA0=EXP(5.0)
      BETA0=ALFA0*SQRT(EXP(1.0)-1.)
      CALL EFICAL(PMEW,NS,4.5)
      CALL EFICAL(SIG2,NS,1.0)
      CALL EFICAL(ALFA,NS,ALFA0)
      CALL EFICAL(BETA,NS,BETA0)
      WRITE(6,900)
  900 FORMAT(////10X,'PARAMETER ESTIMATION OF SIM. LOG. SAMPLES',//)
      WRITE(6,910)
  910 FORMAT(//20X,'MEW = 4.5')
      WRITE(6,920)
  920 FORMAT(5X,'SAMPLE NO.   MOMNTS  MAXL.  7093%3S  27273%3S  ',
     $'MEDCOFVAR   REG.')
      DO 70 I=1,NS
      WRITE(6,930)I,(PMEW(J,I),J=1,6)
  930 FORMAT(7X,I3,3X,6(F7.3,2X))
```

```
 70   CONTINUE
      WRITE(6,940)((PNEW(J,NS+I),J=1,6),I=1,3)
 940  FORMAT(//10X,'MEAN      ',6(F7.3,2X)/
     $10X,'S.D.      ',6(F7.3,2X)/
     $10X,'EFFI      ',6(F7.3,2X)///)
      WRITE(6,950)
 950  FORMAT(//20X,'SIGMA SQ. =1.0'//)
      WRITE(6,920)
      DO 90 I=1,NS
      WRITE(6,930)I,(SIG2(J,I),J=1,6)
 90   CONTINUE
      WRITE(6,940)((SIG2(J,NS+I),J=1,6),I=1,3)
      WRITE(6,955)ALFA0
 955  FORMAT(//20X,'ALPHA=',F7.3,//)
      WRITE(6,920)
      DO 130 I=1,NS
      WRITE(6,930)I,(ALFA(J,I),J=1,6)
 130  CONTINUE
      WRITE(6,940)((ALFA(J,NS+I),J=1,6),I=1,3)
      WRITE(6,960)BETA0
 960  FORMAT(//20X,'BETA=',F7.3,//)
      WRITE(6,920)
      DO 150 I=1,NS
      WRITE(6,930)I,(BETA(J,I),J=1,6)
 150  CONTINUE
      WRITE(6,940)((BETA(J,NS+I),J=1,6),I=1,3)
      STOP
      END
C
      SUBROUTINE STATS(HATM,S2HAT,J)
C
C     CALCULATES MEAN AND S.D. FROM A SET OF PARAMETERS
      DIMENSION A(6,5)
      COMMON A
      A(J,1)=HATM
      A(J,2)=S2HAT
      A(J,3)=EXP(HATM+.5*S2HAT)
      E2=EXP(S2HAT)-1.
      E=SQRT(E2)
      A(J,4)=E*A(J,3)
      RETURN
      END
C
      FUNCTION QUAN(Q,L)
C
C     FINDS SAMPLE QUANTILE OF ORDER Q
      DIMENSION CF(100),XMID(100),A(6,5)
      COMMON A,IND,XMID,CF
      DO 1 I=L,20
      IF(CF(I) .LT. Q)GOTO 1
      IND=I
      IF(I-1)2,2,3
 2    D=CF(I)
      QUAN=0.5+30.*Q/D
      GOTO 4
 3    D=CF(I)-CF(I-1)
      C=Q-CF(I-1)
      QUAN=XMID(I)-15.+30.*C/D
      GOTO 4
 1    CONTINUE
 4    RETURN
      END
C
      SUBROUTINE QCAL(XA,XB,IQ,HATM,S2HAT)
C
C     FINDS QUANTILE ESTIMATES
      IF(IQ .EQ. 93)V=1.475
      IF(IQ .EQ. 73)V=0.514
      HATM=.5*(ALOG(XA)+ALOG(XB))
      SHAT=.5*(ALOG(XB)-ALOG(XA))/V
      S2HAT=SHAT*SHAT
      RETURN
      END
C
      FUNCTION HAS(Q)
C
C     GIVES STANDARD NORMAL VARIATE CORRESP. TO AREA Q UNDER
C     THE UPPER TAIL OF ITS FREQ. CURVE, 0<Q<0.5
      IF(Q)6,5,6
 5    Q=0.000001
 6    P=Q
      IF(Q-.5)1,1,2
 2    P=1.-Q
 1    Z=ALOG(1./(P*P))
      Y=SQRT(Z)
      HAS=Y-((2.515517+0.802853*Y+.010328*Z)/(1.+1.432788*Y+
     $0.189269*Z+0.001308*Y*Z))
      IF(Q-.5)3,3,4
 4    HAS=-HAS
 3    RETURN
      END
C
      SUBROUTINE EFICAL(ARAY,NS,X0)
C
C     FINDS THE EFFICIENCY FOR EACH PARAMETER
      DIMENSION ARAY(6,105)
      DO 2 J=1,6
      S=0.
      S2=0.
      DO 1 I=1,NS
      X=ARAY(J,I)
      S=S+X
      S2=S2+X*X
 1    CONTINUE
      XBAR=S/NS
      ARAY(J,NS+1)=XBAR
      VAR=S2/NS-XBAR*XBAR
      ARAY(J,NS+2)=SQRT(VAR)
      EFI2=(S2-2.*X0*S)/FLOAT(NS)+X0*X0
      ARAY(J,NS+3)=SQRT(EFI2)
 2    CONTINUE
      RETURN
      END
      FINISH
```

```
LIST(LP)
LIBRARY(ED,SUBGROUPNAGF)
PROGRAM(P5)
INPUT 5=CR0
OUTPUT 6=LP0
TRACE 1
COMPACT
COMPRESS INTEGER AND LOGICAL
END

      MASTER TWO-PARAMETER LOGNORMAL ESTIMATION
C
C                        *** PROGRAM P5 ***
C
C     THIS PROGRAM ESTIMATES THE PARAMETERS BY 6 DIFFERENT METHODS
C
      DIMENSION XMID(100),F(100),NAME(20),CSUM(100),CF(100),Z(100)
      DIMENSION XEND(100),XLOG(100),LB(100),IUB(100),RESULT(21)
      DIMENSION FEXP(6,100),AME(6,100),CHISQ(6,100),ELOSS(6,100)
      DIMENSION A1(9),A2(9),A3(9),A4(9),A5(9),A6(9)
      DIMENSION ACOST(100),CLCOST(6,100)
      COMMON IND,XMID,CF,M,TOTAL,XEND,F,FEXP,AME,CHISQ,ELOSS,CLCOST
C     READ IN THE DESCRIPTION OF THE DATA AND NO. OF INTERVALS
      READ(5,100)NAME,M
  100 FORMAT(20A4/I0)
      WRITE(6,900)NAME
  900 FORMAT(///////////////////////////////////,22X,20A4///
     $10X,'FITTING THE TWO PARAMETER LOGNORMAL DIS. BY DIFF. METHODS
     $//15X,'TABLE OF ESTIMATED PARAMETERS AND STATISTICS'
     $///4X,'METHOD--> ',10X,'MAXIMUM    7&93 %    27&73 % ',/
     $2X,'             MOMNTS LIKELIHOOD QUANTILES QUANTILES',
     $' GRAPHICAL REGRESSION'/4X,'...................',
     $'..........................................'/
     $3X,'STATISTICS'/)
C     READ IN THE DATA AND FIND THE CUM. DIS. FUNCTION
      READ(5,101)(F(I), I=1,M)
  101 FORMAT(F0.0)
      CSUM(1)=F(1)
      DO 5 I=2,M
      CSUM(I)=CSUM(I-1)+F(I)
    5 CONTINUE
      TOTAL=CSUM(M)
      XMID(1)=15.5
      XEND(1)=30.5
      IUB(1)=30
      LB(1)=1
      DO 10 I=2,20
      XMID(I)=30.+XMID(I-1)
      XEND(I)=30.+XEND(I-1)
      IUB(I)=XEND(I)
      LB(I)=IUB(I)-29
   10 CONTINUE
      XMID(21)=650.5
      XEND(21)=700.5
      IUB(21)=700
      LB(21)=601
      DO 20 I=22,M
      XMID(I)=100.+XMID(I-1)
      XEND(I)=100.+XEND(I-1)
      IUB(I)=XEND(I)
      LB(I)=IUB(I)-99
   20 CONTINUE
      S1=0.
      S2=0.
      SLOG1=0.
      SLOG2=0.
      ACOST(M+1)=0.
      DO 30 I=1,M
      FI=F(I)
      XI=XMID(I)
      FXI=FI*XI
      ALOGXI=ALOG(XI)
      FLOGXI=FI*ALOGXI
      S1=S1+FXI
      S2=S2+FXI*XI
      SLOG1=SLOG1+FLOGXI
      SLOG2=SLOG2+FLOGXI*ALOGXI
      CF(I)=100.*CSUM(I)/TOTAL
      ACOST(I)=XI*FI
      ACOST(M+1)=ACOST(M+1)+ACOST(I)
   30 CONTINUE
      XMEAN=S1/TOTAL
      SS=S2/TOTAL
      HATM1=2.*ALOG(XMEAN)-.5*ALOG(SS)
      S2HAT1=ALOG(SS)-2.*ALOG(XMEAN)
      CALL STATS(HATM1,S2HAT1,A1)
      HATM2=SLOG1/TOTAL
      S2HAT2=SLOG2/TOTAL-HATM2*HATM2
      CALL STATS(HATM2,S2HAT2,A2)
      XA=QUAN(7.,1)
      XB=QUAN(93.,IND)
      CALL QCAL(XA,XB,93,HATM3,S2HAT3)
      CALL STATS(HATM3,S2HAT3,A3)
      XA=QUAN(27.,1)
      XB=QUAN(73.,IND)
      CALL QCAL(XA,XB,73,HATM4,S2HAT4)
      CALL STATS(HATM4,S2HAT4,A4)
      READ(5,102)E16,E50,E84
  102 FORMAT(3F0.0)
      HATM5=ALOG(E50)
      SHAT5=ALOG(.5*(E50/E16+E84/E50))
      S2HAT5=SHAT5*SHAT5
      CALL STATS(HATM5,S2HAT5,A5)
      K=0
      DO 60 I=1,M-1
      N2=CSUM(I+1)
      N1=CSUM(I)
      IF(N2 .EQ. N1)GOTO 60
      K=K+1
      XLOG(K)=ALOG(XEND(I))
      T=1.-CF(I)/100.
      Z(K)=HAS(T)
   60 CONTINUE
      XM=0.
      YM=99999.
      IFAIL=0
      CALL G02CCF(K,XLOG,Z,XM,YM,RESULT,IFAIL)
      B=RESULT(6)
      A=RESULT(7)
      SHAT6=1./B
      HATM5=-A/B
      S2HAT6=SHAT6*SHAT6
      CALL STATS(HATM6,S2HAT6,A6)
      WRITE(6,905)A1(1),A2(1),A3(1),A4(1),A5(1),A6(1)
  905 FORMAT(4X,'   MEWHAT   ',F7.3,5(3X,F7.3))
      WRITE(6,910)A1(2),A2(2),A3(2),A4(2),A5(2),A6(2)
  910 FORMAT(/4X,'SIGMA2HAT  ',F7.3,5(3X,F7.3))
      WRITE(6,915)A1(3),A2(3),A3(3),A4(3),A5(3),A6(3)
  915 FORMAT(/3X,'ALPHA-MEAN  ',F7.3,5(3X,F7.3))
      WRITE(6,920)A1(4),A2(4),A3(4),A4(4),A5(4),A6(4)
  920 FORMAT(/4X,'BETA-S.D.  ',F7.3,5(3X,F7.3))
      WRITE(6,925)A1(5),A2(5),A3(5),A4(5),A5(5),A6(5)
  925 FORMAT(/3X,'COF OF VAR',1X,F7.3,5(3X,F7.3))
      WRITE(6,930)A1(6),A2(6),A3(6),A4(6),A5(6),A6(6)
  930 FORMAT(/4X,'   MEDIAN   ',F7.3,5(3X,F7.3))
      WRITE(6,935)A1(7),A2(7),A3(7),A4(7),A5(7),A5(7)
  935 FORMAT(/4X,'    MODE    ',F7.3,5(3X,F7.3))
      WRITE(6,940)A1(8),A2(8),A3(8),A4(8),A5(8),A6(8)
  940 FORMAT(/4X,' SKEWNESS  ',F7.3,5(3X,F7.3))
      WRITE(6,945)A1(9),A2(9),A3(9),A4(9),A5(9),A6(9)
```

```
 945  FORMAT(/4X,' KURTOSIS ',F7.3,5(3X,F7.3))
      CALL XPCTED(HATM1,S2HAT1,1)
      CALL XPCTED(HATM2,S2HAT2,2)
      CALL XPCTED(HATM3,S2HAT3,3)
      CALL XPCTED(HATM4,S2HAT4,4)
      CALL XPCTED(HATM5,S2HAT5,5)
      CALL XPCTED(HATM6,S2HAT6,6)
      WRITE(6,950)NAME
 950  FORMAT(/////////////////////////////////////22X,20A4//
     $10X,'EXPECTED NO. OF CLAIMS BY DIFF. METHODS OF ESTIMATION'/
     $/24X,'TWO PARAMETER LOGNORMAL DIS. '//)
      WRITE(6,955)
 955  FORMAT(5X,'  AMOUNT £  ACT. NO.  MOMNTS  MAX.LIK. ',
     £' 7693%US  27673%US GRAPH REG.')
      DO 80 I=1,M
      WRITE(6,960)LB(I),IUB(I),F(I),(FEXP(J,I),J=1,6)
 960  FORMAT(5X,I5,'-',I4,2X,F6.0,3X,F6.0,2X,F6.0,5X,F6.0,3X,F6.0,
     $1X,F6.0,F6.0)
  80  CONTINUE
      WRITE(6,965)TOTAL,(FEXP(J,M+1) ,J=1,6)
 965  FORMAT(5X,'.......................................',
     $'..........................'/6X,'TOTAL     ',
     $F7.0,3X,F6.0,1X,F7.0,4X,F7.0,2X,F7.0,F7.0,F6.0)
      WRITE(6,970)NAME
 970  FORMAT(/////////////////////////////////////////22X,20A4//
     $15X,'(ACTUAL - EXPECTED) CLAIM NO.S BY DIFF. METHODS'/
     $/24X,'TWO-PARAMETER LOGNORMAL DIS. '//)
      WRITE(6,975)
 975  FORMAT(5X,'  AMOUNT £    MOMENTS  MAX. LIK. ',
     £' 7693%US  27673%US  GRAPHCL  REG.')
      DO 90  I=1,M
      WRITE(6,980)LB(I),IUB(I),(AME(J,I) , J=1,6)
 980  FORMAT(5X,I5,'-',I4,4X,F6.0,3X,F6.0,6X,F6.0,2X,F6.0,4X,F6.0,
     $2X,F6.0)
  90  CONTINUE
      WRITE(6,985)NAME
 985  FORMAT(///////////////////////////////////////,22X,20A4,
     $//15X,'CHI SQ. GOODNESS OF FIT TEST FOR DIFF. METHODS :'//
     $24X,'TWO-PARAMETER LOGNORMAL DIS. '//)
      WRITE(6,990)
 990  FORMAT(5X,'  AMOUNT $  MOMNTS   MAX.LIK.  7693%US',
     £'   27673*US  GRAPHCL    REG.'/15X,'..................',
     £' (A-E)**2/E ...........................')
      DO 150 I=1,24
      WRITE(6,995)LB(I),IUB(I),(CHISQ(J,I) ,J=1,6)
 995  FORMAT(5X,I5,'-',I4,1X,F8.3,2X,F7.3,1X,2(3X,F7.3),2X,F7.3,3X,F
7.3)
 150  CONTINUE
      WRITE(6,1000)(CHISQ(J,M+1) ,J=1,6)
1000  FORMAT(///////////////////////////////////////
     $6X,'TOTAL CHISQ',6(1X,F7.3))
      WRITE(6,1005)NAME
1005  FORMAT(/////////////////////////////////////////22X,20A4,
     £//15X,'ACTUAL AND EXPECTED CLAIMS COST BY DIFF. METHODS',
     $//24X,'TWO PARAMETER LOGNORMAL DIS.'//)
      WRITE(6,1010)
1010  FORMAT(5X,'  AMOUNT £   ACTUAL  MOMNTS MAXLIK. 7693%US ',
     $'27673*US GRAPHCL   REG.')
      DO 155 I=1,M
      WRITE(6,1015)LB(I),IUB(I),ACOST(I),(CLCOST(J,I) , J=1,6)
1015  FORMAT(5X,I5,'-',I4,1X,5(1X,F7.0),1X,2(1X,F7.0))
 155  CONTINUE
      WRITE(6,1020)ACOST(M+1),(CLCOST(J,M+1),J=1,6)
1020  FORMAT(4X,'..............................................',
     $'.............................',/3X,'TOTAL CL COST',
     $5(1X,F7.0),1X,2(1X,F7.0)///////////////////)
      WRITE(6,1025)NAME
1025  FORMAT(/////////////////////////////////////////22X,20A4,
     $//12X,'EXPECTED LOSS BY DIFF. METHODS: (A-E)*(CLAIM AMOUNT)'/
     $/24X,'TWO-PARAMETER LOGNORMAL DIS.'//)
      WRITE(6,1030)
1030  FORMAT(5X,'  AMOUNT £   MOMNTS  MAX.LIK.  ',
     $'7693%US  27673%US  GRAPHCL  REG.')
      DO 160 I=1,M
      WRITE(6,1035)LB(I),IUB(I),(ELOSS(J,I) ,J=1,6)
1035  FORMAT(5X,I5,'-',I4,2(3X,F6.0),3X,F7.0,2(3X,F6.0),2X,F6.0)
 160  CONTINUE
      WRITE(6,1040)(ELOSS(J,M+1) ,J=1,6)
1040  FORMAT(5X,'......................................',
     $'...........................'/6X,'TOTAL ',
     $3X,2(2X,F7.0),3X,F7.0,2X,F7.0,2X,F7.0,1X,F7.0)
      DO 165 J=1,6
      ELOSS(J,M+3)=100.*ELOSS(J,M+1)/ACOST(M+1)
      ELOSS(J,M+1)=ELOSS(J,M+1)/SQRT(ELOSS(J,M+2))
 165  CONTINUE
      WRITE(6,1045)(ELOSS(J,M+3) , J=1,6)
1045  FORMAT(/2X,'TOTAL EXP. LOSS'/2X,'----------------',
     $F5.2,'%',3X,F5.2,'%',1X,3(3X,F5.2,'%'),2X,F5.2,'%'/
     $2X,'TOTAL ACT. COST')
      WRITE(6,1050)(ELOSS(J,M+2),J=1,6),(ELOSS(J,M+1),J=1,6)
1050  FORMAT(///////////////////3X,'TOTAL FI**2(EXPI)',
     $6(2X,F10.0)//3X,'SD. NORMAL DEV.',6(3X,F6.3)///////)
      STOP
      END
C
      SUBROUTINE STATS(HATM,S2HAT,A)
C      CALCULATES  RELEVANT STATS. FROM FITTED MODEL
      DIMENSION A(9)
      A(1)=HATM
      A(2)=S2HAT
      A(3)=EXP(HATM+.5*S2HAT)
      E2=EXP(S2HAT)-1.
      E=SQRT(E2)
      A(4)=E*A(3)
      A(5)=E
      A(6)=EXP(HATM)
      A(7)=EXP(HATM-S2HAT)
      A(8)=E*(E2+3.)
      A(9)=3.+16.*E2+15.*E2*E2+6.*E2*E2*E2+E2**4
      RETURN
      END
C
      FUNCTION QUAN(Q,L)
C     FINDS SAMPLE QUANTILE OF ORDER Q
      DIMENSION CF(100),XMID(100)
      COMMON IND,XMID,CF
      DO 1 I=L,20
      IF(CF(I) .LT. Q)GOTO 1
      IND=I
      IF(I-1)2,2,3
    2 D=CF(I)
      QUAN=0.5+30.*Q/D
      GOTO 4
    3 D=CF(I)-CF(I-1)
      C=Q-CF(I-1)
      QUAN=XMID(I)-15.+30.*C/D
      GOTO 4
    1 CONTINUE
    4 RETURN
      END
C
      SUBROUTINE QCAL(XA,XB,IQ,HATM,S2HAT)
C     FINDS  THE QUANTILE ESTIMATE
C     FINDS SAMPLE QUANTILE OF ORDER Q
      IF(IQ .EQ. 93)V=1.475
      IF(IQ .EQ. 73)V=0.614
      HATM=.5*(ALOG(XA)+ALOG(XB))
      SHAT=.5*(ALOG(XB)-ALOG(XA))/V
      S2HAT=SHAT*SHAT
      RETURN
      END
```

```
C
      SUBROUTINE XPCTED(HATM,S2HAT,J)
C     FINDS E, A-E, EXP LOSS, (A-E)**2/E,A AND THEIR SUM
      DIMENSION XEND(100),XMID(100),F(100),CF(100),FEXP(6,100)
      DIMENSION AME(6,100),CHISQ(6,100),ELOSS(6,100),CLCOST(6,100)
      COMMON IND,XMID,CF,M,TOTAL,XEND,F,FEXP,AME,CHISQ,ELOSS,CLCOST
      SHAT=SQRT(S2HAT)
      FEXP(J,M+1)=0.
      CHISQ(J,M+1)=0.
      ELOSS(J,M+1)=0.
      ELOSS(J,M+2)=0.
      CLCOST(J,M+1)=0.
      S=0.
      DO 10 I=1,M
      X=XEND(I)
      Y=(ALOG(X)-HATM)/SHAT
C     S15ABF IS NAG ROUTINE FOR STAND. NORMAL CUM. DIS. FUNCTION
      P=S15ABF(Y,0)
      Q=TOTAL*P
      FEXP(J,I)=Q-S
      NEXP=FEXP(J,I)+.5
      FEXP(J,M+1)=FEXP(J,M+1)+NEXP
      AME(J,I)=F(I)-NEXP
      IF(NEXP .LT. 5)GOTO 1
      CHISQ(J,I)=AME(J,I)**2/NEXP
      CHISQ(J,M+1)=CHISQ(J,M+1)+CHISQ(J,I)
    1 ELOSS(J,I)=XMID(I)*AME(J,I)
      ELOSS(J,M+1)=ELOSS(J,M+1)+ELOSS(J,I)
      ELOSS(J,M+2)=ELOSS(J,M+2)+NEXP*XMID(I)**2
      CLCOST(J,I)=XMID(I)*NEXP
      CLCOST(J,M+1)=CLCOST(J,M+1)+CLCOST(J,I)
      S=Q
   10 CONTINUE
      RETURN
      END
C
      FUNCTION HAS(Q)
C     GIVES  STAN. NORMAL VARIATE CORRESP. TO Q THE AREA UNDER
C     THE UPPER TAIL OF STAN. NORMAL FREQ. CURVE, 0<Q<0.5
      IF(Q)6,5,6
    5 Q=0.000001
    6 P=Q
      IF(Q-.5)1,1,2
    2 P=1.-Q
    1 Z=ALOG(1./(P*P))
      Y=SQRT(Z)
      HAS=Y-((2.515517+0.802853*Y+.010328*Z)/(1.+1.432788*Y+
     $0.189269*Z+0.001308*Y*Z))
      IF(Q-.5)3,3,4
    4 HAS=-HAS
    3 RETURN
      END
      FINISH
```

```
      LIBRARY(SUBGROUPNAGF)
      LIST(LP)
      PROGRAM(P6)
      COMPACT
      INPUT 1,5=CR0
      OUTPUT 2,6=LP0
      TRACE 2,100
      COMPRESS INTEGER AND LOGICAL
      END

      MASTER ESTIMATION BY MML METHOD FOR 2-PAR LOGNORMAL
C
C                     *** PROGRAM P6 ***
C
C     THIS PROGRAM FINDS ESTIMATES OF THE PARAMETERS OF THE
C     2-PAR LOGNORMAL DIS. BY THE MML METHOD.
C
      DIMENSION X(60),XLOG(60),Y(60),AME2(60),CSUM(60),W(40),YY(60)
      DIMENSION XX(60),XMID(100),FEXP(60),NAME(20),P(2),ELOSS(100)
      COMMON K,XX,YY
      REAL MEW
C     READ DESCRIPTION OF DATA AND NO. OF INTERVALS
      READ(5,101)NAME,M
  101 FORMAT(20A4/I0)
      WRITE(6,900)NAME
  900 FORMAT(////' *********',20A4)
C     READ STARTING VALUES FOR THE PARAMETERS
      READ(5,102)(P(I),I=1,2)
  102 FORMAT(2F0.0)
      S=0.0
C     READ IN THE DATA AND FIND  SAMPLE CUM. DIS. FUNCTION
      DO  3 I=1,M
      READ(5,103)Y(I)
  103 FORMAT(F0.0)
      S=S+Y(I)
      IF(I-20)1,1,2
    1 X(I)=30.*I+0.5
      XMID(I)=X(I)-15.
      GOTO 3
    2 X(I)=(I-14)*100.+0.5
      XMID(I)=X(I)-50.
    3 CONTINUE
C     FIND K THE NO. OF INTERVALS WITH NON-ZERO FREQUENCY
      K=0
      DO 5 I=1,M
      NI=Y(I)
      IF(NI .EQ. 0)GOTO 5
      K=K+1
      XX(K)=X(I)
      YY(K)=Y(I)/S
    5 CONTINUE
      IFAIL=1
C     E04CEF IS NAG ROUTINE FOR MINIMIZING (OR MAX. ) A GENERAL
C     FUNCTION OF SEVERAL VARIABLES.
      CALL E04CEF(2,P,F,W,40,IFAIL)
      IF(IFAIL .NE. 0)WRITE(6,901)IFAIL
  901 FORMAT('  IFAIL=',I2////////)
C     MML ESTIMATES OF THE PARAMETERS ARE :-
      SIGMA=P(2)
      SIGMA2=SIGMA*SIGMA
      MEW=P(1)
C     FIND MEAN AND S.D. OF THE LOGNORMAL MODEL
      ALPHA=EXP(MEW+.5*SIGMA2)
      ETA=SQRT(EXP(SIGMA2)-1.0)
      BETA=ETA*ALPHA
      WRITE(6,902)NAME,MEW,SIGMA2,ALPHA,BETA
  902 FORMAT(////20X,'*** TWO-PARAMETER LOGNORMAL DIS. ***'//
     $27X,20A4//12X,'ESTIMATION BY MULTINOMIAL MAX.',
     $'LIKELIHOOD METHOD :-',
     $/16X,'MEW =',F7.4,'          SIGMA2=',F7.4/,
     $/16X,'MEAN=',F8.3,'          S.D.=',F8.3)
      WRITE(6,903)
  903 FORMAT(///6X,'AMOUNT £    ACTUAL     EXPECTED    ACTUAL-',
```

```
$'    EXPECTED',/17X,'CL. NO.    CL. NO.    EXPECTED',
$'      LOSS        (A-E)**2/E ')
C      COMPUTE THE TABLE OF RESULTS
       SEXP=0.0
       SS=0.
       ACOST=0.0
       ELOSS(M+1)=0.0
       Q1=0.0
       DO 7 I=1,M
       V=(ALOG(X(I))-MEW)/SIGMA
C      S15ABF IS NAG ROUTINE FOR STANDARD NORMAL CUM. DIS. FUN.
       Q2=S15ABF(V,0)
       FEXP(I)=(Q2-Q1)*S
       NEXP=FEXP(I)+.5
       SEXP=SEXP+NEXP
       CSUM(I)=Y(I)-NEXP
       ACOST=ACOST+XMID(I)*Y(I)
       ELOSS(I)=XMID(I)*CSUM(I)
       ELOSS(M+1)=ELOSS(M+1)+ELOSS(I)
       IF(NEXP .LT. 5)GOTO 20
       AVE2(I)=CSUM(I)*CSUM(I)/NEXP
       SS=SS+AVE2(I)
   20  Q1=Q2
    7  CONTINUE
       DO 11 I=1,M
       IF(I-20)8,8,9
    8  LB=X(I)-29.
       GOTO 10
    9  LB=X(I)-99.
   10  LU=X(I)-0.5
       IF(FEXP(I) .LT. 5)GOTO 21
       WRITE(6,904)LB,LU,Y(I),FEXP(I),CSUM(I),ELOSS(I),AVE2(I)
  904  FORMAT(4X,I5,'-',I4,F9.0,F10.0,F11.0,4X,F8.0,3X,F8.3)
       GOTO 11
   21  WRITE(6,905)LB,LU,Y(I),FEXP(I),CSUM(I),ELOSS(I)
  905  FORMAT(4X,I5,'-',I4,F9.0,F10.0,F11.0,4X,F8.0)
   11  CONTINUE
       RATIO=100.*ELOSS(M+1)/ACOST
       WRITE(6,906)S,SEXP,ELOSS(M+1),SS,RATIO
  906  FORMAT('     ----------------------------------------',
      $'----------------------'/,'       TOTAL     ',F9.0,1X,F9.0,15X,
      $F8.0,3X,F8.3/'     ----------------------------------------',
      $'-------------------------------'/57X,'D.F. = ',/10X,
      $'TOTAL EXP. LOSS'/10X,'-----------------',F5.1,' %',/
      $10X,'TOTAL ACT. COST'///////////////////////////////)
       STOP
       END
C
       SUBROUTINE FUNCT1(NP,PC,FC)
C
C      FINDS THE VALUE OF THE LOGLIKELIHOOD FUNCTION FOR A GIVEN
C      SET OF PARAMETERS. THIS IS REQUIRED BY E04CEF ROUTINE.
       DIMENSION PC(NP),XX(60),YY(60)
       COMMON K,XX,YY
       REAL MEW
       MEW=PC(1)
       SIGMA=PC(2)
       Z=(ALOG(XX(1))-MEW)/SIGMA
       D1=S15ABF(Z,0)
       S=YY(1)*ALOG(D1)
       DO 1 I=2,K
       IF(I .LT. 21)GOTO 2
       Z=(ALOG(XX(I)-100)-MEW)/SIGMA
       D1=S15ABF(Z,0)
    2  Z=(ALOG(XX(I))-MEW)/SIGMA
       D2=S15ABF(Z,0)
       S=S+YY(I)*ALOG(D2-D1)
       D1=D2
    1  CONTINUE
       FC=-S
       RETURN
       END
       FINISH
```

```
       LIBRARY(SUBGROUPGRAF)
       LIBRARY(SUBGROUPGINO)
       LIBRARY(SUBGROUPNAGF)
       LIST(LP)
       PROGRAM(P7)
       COMPACT
       INPUT 1,5=CR0
       OUTPUT 2,6=LP0
       TRACE 0
       COMPRESS INTEGER AND LOGICAL
       END

       MASTER HISTOGRAM AND LOGNORMAL FREQUENCY CURVE PLOT
C
C                  *** PROGRAM P7 ***
C
C      THIS PROGRAM PLOTS THE HISTOGRAM OF A SAMPLE OF AD DATA
C      ALONG WITH THE FREQ. CURVE OF THE 2-PAR LOGNORMAL DIS.
C      FITTED TO IT. 3 PLOTS ARE PRODUCED. ONE FOR ALL CLAIM
C      AMOUNTS, ONE FOR CLAIMS TO £600 AND ONE FOR CLAIMS >£600.
C
       DIMENSION X(70),Y(70),NAME(20),A(70),XMID(70),F(300),Z(300)
       DIMENSION ITITLE(20),IPAR(20),H(70),FF(300),ZZ(300)
       COMMON SFACX,SFACY,NAME,ITITLE,IPAR
       DATA PI/3.1415926536/
C      READ DESCRIPTION OF DATA(NAME),THE METHOD OF ESTIMATION(
C      ITITLE),THE DESCRIPTION OF PARAMETERS TO BE PRINTED ON THE
C      GRAPHS(IPAR), NO. OF INTERVALS(M),EST. OF MEW(HATMEW),
C      EST. OF SIGMA SQ.(SIGMA2)
       READ(5,101)NAME,ITITLE,IPAR,M,HATMEW,SIGMA2
  101  FORMAT(20A4/20A4/20A4/I0/2F0.0)
       WRITE(6,900)NAME,ITITLE
  900  FORMAT(20A4//20A4)
       X(1)=0.0
       DO 1 I=1,20
       X(I+1)=X(I)+30.
       XMID(I)=(X(I+1)+X(I))/2.
    1  CONTINUE
       DO 2 I=21,M
       X(I+1)=X(I)+100.
       XMID(I)=(X(I+1)+X(I))/2.
    2  CONTINUE
       TOTAL=0.
C      READ IN THE DATA
       DO 3 I=1,M
       READ(5,102)Y(I)
  102  FORMAT(F0.0)
       TOTAL=TOTAL+Y(I)
    3  CONTINUE
       XMPLS1=X(M+1)
       IF(XMPLS1 .LT. 2599)XMPLS1=XMPLS1+100.
       LENTHX=XMPLS1/10.+36
       NFF=LENTHX-95
       T=1./SIGMA2
       CONST=TOTAL*SQRT(T/(2.*PI))
       DO 30 I=1,LENTHX
       IF(I-39)6,6,7
    6  R=I
       GOTO 8
    7  R=(I-36)*10.
    8  Z(I)=R
       F(I)=CONST/R*EXP(-0.5*T*(ALOG(R)-HATMEW)**2)
   30  CONTINUE
       K=NGT25(F)
       N95=95-K
       LENTHX=LENTHX-K
       DO 31 I=1,LENTHX
       Z(I)=Z(I+K)
       F(I)=F(I+K)
   31  CONTINUE
```

```
      J=0
      DO 32 J=1,NFF
      IF(F(N95+J) .LT. 0.5)GOTO 33
      J=J+1
   32 CONTINUE
   33 NFF=NFF-J
      DO 40 I=1,NFF
      FF(I)=F(I+N95+J)
      ZZ(I)=Z(I+N95+J)
   40 CONTINUE
      A(1)=X(2)-X(1)
      Y(1)=Y(1)/A(1)
      DO 4 I=1,M-1
      B=X(I+2)-X(I+1)
      Y(I+1)=Y(I+1)/B
      H(I)=Y(I+1)-Y(I)
      A(I+1)=B
    4 CONTINUE
      H(M)=Y(M)
      SFACX=0.1
      SFACY=8.
C     CALL GINO ROUTINES TO PLOT THE GRAPHS
      CALL M19346
      CALL DEVPAP(700.,700.,0)
      CALL CHASWI(1)
      CALL CHASIZ(3.0,3.0)
      CALL FRAME(1)
      CALL AXIS1
      CALL GRACUR(Z,F,LENTHX)
      CALL MOVTO2(50.,37.)
      Y1=Y(1)*SFACY
      CALL MOVBY2(0.,Y1)
      DO 5 I=1,M-1
      C=A(I)
      D=H(I)
      CALL HIS(C,D)
    5 CONTINUE
      C=A(M)
      D=-H(M)
      CALL HIS(C,D)
      CALL FRAME(2)
      CALL AXIS2
      N96=N95+1
      CALL GRACUR(Z,F,N96)
      SFACX=0.46
      SFACY=8.
      CALL MOVTO2(45.,257.)
      CALL MOVBY2(0.,Y1)
      DO 9 I=1,19
      C=A(I)
      D=H(I)
      CALL HIS(C,D)
    9 CONTINUE
      C=A(20)
      D=-Y(20)
      CALL HIS(C,D)
      CALL FRAME(3)
      CALL AXIS3
      CALL GRACUR(ZZ,FF,NFF)
      SFACX=0.13
      SFACY=360.
      Y1=Y(21)*SFACY
      CALL MOVTO2(50.,490.)
      CALL MOVBY2(0.,Y1)
      DO 11 I=21,M-1
      C=A(I)
      D=H(I)
      CALL HIS(C,D)
   11 CONTINUE
      CALL HIS(A(M),-H(M))
      CALL DEVEND
      STOP
      END
C
      SUBROUTINE FRAME(J)
C
C     DRAWS AN A4 FRAME FOR PLOT NO; J=1 OR 2 OR 3
      IF(J-2)1,2,3
    1 CALL MOVTO2(30.,30.)
      GOTO 4
    2 CALL MOVTO2(30.,250.)
      GOTO 4
    3 CALL MOVTO2(30.,470.)
    4 CALL LINBY2(297.,0.)
      CALL LINBY2(0.,210.)
      CALL LINBY2(-297.,0.)
      CALL LINBY2(0.,-210.)
      RETURN
      END
C
      SUBROUTINE AXIS1
C
C     DRAWS THE AXES FOR PLOT NO. 1
      CALL AXIPOS(1,50.,37.,270.,1)
      CALL AXIPOS(1,50.,37.,200.,2)
      CALL AXISCA(3,27,0.,2700.,1)
      CALL AXISCA(3,25,0.,25.,2)
      CALL AXIDRA(2,1,1)
      CALL AXIDRA(-2,-1,2)
      CALL DES(50.,37.,323.,230.)
      RETURN
      END
C
      SUBROUTINE AXIS2
C
C     DRAWS THE AXES FOR PLOT NO. 2
      CALL AXIPOS(1,45.,257.,276.,1)
      CALL AXIPOS(1,45.,257.,200.,2)
      CALL AXISCA(3,20,0.,600.,1)
      CALL AXISCA(3,25,0.,25.,2)
      CALL AXIDRA(2,1,1)
      CALL AXIDRA(-2,-1,2)
      CALL DES(45.,257.,322.,450.)
      RETURN
      END
C
      SUBROUTINE AXIS3
C
C     DRAWS THE AXES FOR PLOT NO. 3
      CALL AXIPOS(1,50.,490.,260.,1)
      CALL AXIPOS(1,50.,490.,180.,2)
      CALL AXISCA(3,20,600.,2600.,1)
      CALL AXISCA(3,10,0.,.50,2)
      CALL AXIDRA(2,1,1)
      CALL AXIDRA(-2,-1,2)
      CALL DES(50.,490.,315.,665.)
      RETURN
      END
C
      SUBROUTINE DES(X0,Y0,X,Y)
C
C     WRITES THE DESCRIPTIONS OF THE DATA,THE METHOD OF ESTIMATION
C     AND THE VALUES OF THE PARAMETERS ON THE PLOTS
      DIMENSION NAME(20),ITITLE(20),IPAR(20)
      COMMON SFACX,SFACY,NAME,ITITLE,IPAR
      CALL MOVTO2(X,Y0)
      CALL CHAHOL(3HX*.)
      X0=X0+5.
      CALL MOVTO2(X0,Y)
      CALL CHAHOL(13HF*LREQUENCY*.)
```

```
      X0=X0+105.
      Y0=Y0+170.
      CALL MOVTO2(X0,Y0)
      CALL CHAARR(NAME,20,4)
      Y0=Y0-10.
      CALL MOVTO2(X0,Y0)
      CALL CHAARR(ITITLE,20,4)
      Y0=Y0-10.
      CALL MOVTO2(X0,Y0)
      CALL CHAARR(IPAR,20,4)
      RETURN
      END
C
      FUNCTION NGT25(F)
C
C     FINDS THE NO. OF POINTS ON THE FREQ. CURVE WHOSE FREQ.
C     IS GREATER THAN 25 AND  WHICH ARE FOR AMOUNTS >£600.
      DIMENSION F(300)
      NGT25=0
    1 IF(F(NGT25+1) .LT. 25.)GOTO 2
      NGT25=NGT25+1
      GOTO 1
    2 RETURN
      END
C
      SUBROUTINE HIS(C,D)
C
C     DRAWS ONE COLUMN OF THE HISTOGRAM.
      COMMON SFACX,SFACY
      C=C*SFACX
      D=D*SFACY
      CALL LINBY2(C,0.)
      CALL LINBY2(0.,D)
      RETURN
      END
      FINISH
```

```
      LIBRARY(SUBGROUPNACF)
      LIST(LP)
      PROGRAM(P8)
      INPUT 5=CR0
      OUTPUT 6=LP0
      TRACE 1
      COMPACT
      COMPRESS INTEGER AND LOGICAL
      END

      MASTER PREDICTION TWO-PARAMETER LOGNORMAL
C
C                   *** PROGRAM P8 ***
C
C     THIS PROGRAM FINDS THE DIS. OF CLAIM AMOUNTS IN A FUTURE
C     PERIOD ACCORDING TO 2-PAR LOGNORMAL MODEL
C
      DIMENSION F(100),USDQTR(8),INDEX(20)
      COMMON ADJMEW,SIGMA,S,NCLM
      REAL MEW
C     READ RESPECTIVELY THE DES. OF PERIOD BEING PREDICTED.
C     THE DES. OF PERIOD USED FOR PREDICTION,PARAMETERS OF MODEL
C     FOR PERIOD  USED ,RATE OF INFLATION,DES. OF INDEX USED.
C     NO. OF CLAIMS IN PERIOD BEING PREDICTED. NO. OF INTERVALS.
C     THE ACTUAL PARAMETERS OF THE MODEL  FOR THE PERIOD
C     BEING PREDICTED.
      READ(5,100)PRDQTR,USDQTR,MEW,SIGMA2,RATINF,INDEX,NCLM,M,
     $AMEW,ASGMA2
  100 FORMAT(A8/8A8/2F0.0/F0.0/20A4/I0/I0/2F0.0)
      DELTA=ALOG(1.+RATINF/100.)
C     FIND THE PARAMETERS AND MEAN AND S.D.  OF PREDICTION MODEL
C     BY ADJUSTING FOR INFLATION
      ADJMEW=MEW+DELTA
      PALPHA=EXP(ADJMEW+.5*SIGMA2)
      PE=SQRT(EXP(SIGMA2)-1.0)
      PSD=PALPHA*PE
      ALPHA=EXP(AMEW+.5*ASGMA2)
      E=SQRT(EXP(ASGMA2)-1.0)
      SD=ALPHA*E
      WRITE(6,900)PRDQTR,USDQTR,MEW,SIGMA2,RATINF,INDEX,
     $ADJMEW,SIGMA2,PALPHA,PSD,PRDQTR,AMEW,ASGMA2,ALPHA,SD
  900 FORMAT(//////////////////////20X,
     $'*** TWO-PARAMETER LOGNORMAL DIS. ***'//8X,
     $'PREDICTION OF ',A8,'  QUARTER CLAIMS COST '/8X,'USING ',8A8//
     $12X,'MEW=',F6.3,16X,'SIGMA2=',F5.3/12X,'INFLATION RATE I=',
     $F4.1,'% CALCULATED  FROM  :'/12X,
     $20A4//8X,'PREDICTION PARAMETERS ARE  :-  MEW=',F6.3,6X,
     $'SIGMA2=',F5.3/25X,'MEAN CLAIM AMOUNT=',F7.2,7X,'S.D.=',
     $F6.2/8X,'ACTUAL',A8,'PARAMETERS    :-  MEW=',F6.3,6X,
     $'SIGMA2=',F5.3/25X,'MEAN CLAIM AMOUNT=',F7.2,7X,'S.D.=',
     $F6.2///8X,'  AMOUNT £  ACT.NO.  EXP.NO.   ',
     $'A-E   EXP.LOSS  (A-E)**2/E')
      SIGMA=SQRT(SIGMA2)
C     READ IN THE ACTUAL FREQ. DIS. FOR THE PREDICTED PERIOD.
      READ(5,101)(F(I),I=1,M)
  101 FORMAT(F0.0)
      S=0.
      SELOSS=0.
      ISNEXP=0.
      SS=0.
      SACOST=0.
      SF2E=0.
      I=0
   10 I=I+1
      IF(I-20)1,1,2
    1 XEND=30.*I+0.5
      IUB=XEND
      LB=IUB-29
      XMID=XEND-15.
      GOTO 3
```

```
      LIBRARY(SUBGROUPGRAF)
      LIBRARY(SUBGROUPGINO)
      LIBRARY(SUBGROUPNAGF)
      LIST(LP)
      PROGRAM(P9)
      COMPACT
      INPUT 1,5=CR0
      OUTPUT 2,6=LP0
      TRACE 1
      COMPRESS INTEGER AND LOGICAL
      END

      MASTER LOGNORMAL GRAPHICAL TEST
C
C                     *** PROGRAM P9 ***
C
C     THIS PROGRAM PLOTS (LOG(X+C), Z) FOR DIFF. VALUES OF C
C     FOR THE PURPOSE OF TESTING LOGNORMALITY
C
      DIMENSION NAME(20), X(60),XLOG(60),Y(60),F(60),CSUM(60)
      COMMON K,X,XLOG
C     READ DESCRIPTION OF DATA AND NO. OF INTERVALS
      READ(5,101)NAME,M
  101 FORMAT(20A4/I0)
      WRITE(6,900)NAME
  900 FORMAT(////' *********',20A4)
C     READ THE DATA
      DO  3 I=1,M
      READ(5,102)Y(I)
  102 FORMAT(F0.0)
      IF(I-20)1,1,2
    1 X(I)=30.*I+0.5
      GOTO 3
    2 X(I)=(I-14)*100.+0.5
    3 CONTINUE
C     FIND SAMPLE CUM. DIS. FUNCTION
      CSUM(1)=Y(1)
      DO 4 I=2,M
    4 CSUM(I)=CSUM(I-1)+Y(I)
      S=CSUM(M)
C     FIND K THE NUMBER OF INTERVALS WITH NON-ZERO FREQUENCY
      K=0
      DO 5 I=1,M-1
      N2=CSUM(I+1)
      N1=CSUM(I)
      IF(N2 .EQ. N1)GOTO 5
      K=K+1
      X(K)=X(I)
      F(K)=CSUM(I)/S
    5 CONTINUE
      DO 6 I=1,K
      F(I)=HAS(1.-F(I))
    6 CONTINUE
C     CALL GINO ROUTINES TO PLOT THE GRAPHS
      CALL M19346
      CALL DEVPAP(700.,700.,0)
      CALL CHASWI(1)
      CALL CHASIZ(3.0,3.0)

      CALL MOVTO2(30.,30.)
      CALL LINBY2(210.,0.)
      CALL LINBY2(0.,297.)
      CALL LINBY2(-210.,0.)
      CALL LINBY2(0.,-297.)
      CALL AXIPOS(1,50.,185.,180.,1)
      CALL AXIPOS(0,50.,165.,270.,2)
      CALL AXISCA(1,7,2.,9.,1)
      CALL AXISCA(1,8,-4.,4.,2)
      CALL AXIDRA(2,1,1)
      CALL AXIDRA(-2,-1,2)
      CALL MOVTO2(205.,187.)
```

```
    2 XEND=(I-14)*100.+0.5
      IUB=XEND
      LB=IUB-99
      XMID=XEND-50.
    3 P=EXPNO(XEND)
      NEXP=P+0.5
      IF(P-0.5)4,4,5
    4 IF(I-M)6,6,7
    6 P=0.0
      GOTO 11
    7 GOTO 20
    5 IF(I-M)11,11,9
    9 F(I)=0.0
   11 AMINE=F(I)-NEXP
      SACOST=SACOST+XMID*F(I)
      ELOSS=XMID*AMINE
      SELOSS=SELOSS+ELOSS
      SF2E=SF2E+XMID*XMID*NEXP
      IF(NEXP .LT. 5)GOTO 12
      Z=AMINE*AMINE/NEXP
      SS=SS+Z
      WRITE(6,901)LB,IUB,F(I),P,AMINE,ELOSS,Z
  901 FORMAT(8X,I5,'-',I4,2X,F6.0,3X,F6.0,3X,F5.0,2X,F7.0,4X,F7.3)
      GOTO 13
   12 WRITE(6,902)LB,IUB,F(I),P,AMINE,ELOSS
  902 FORMAT(8X,I5,'-',I4,2X,F6.0,3X,F6.0,3X,F5.0,2X,F7.0)
   13 ISNEXP=ISNEXP+NEXP
      GOTO 10
   20 WRITE(6,903)NCLM,ISNEXP,SELOSS,SS
  903 FORMAT(8X,'-------------------------------------------',
     $'-----------'/8X,'   TOTAL     ',I5,4X,I5,11X,F7.0,4X,F7.3/
     $8X,'-----------------------------------------',
     $'-----------------')
      SECOST=SACOST-SELOSS
      Z=100.*SELOSS/SACOST
      WRITE(6,904)SACOST,SECOST,Z
  904 FORMAT(/10X,'CHI SQ. STAT. =        , D.F.= '//10X,
     $'TOTAL ACTUAL COST   =',F8.0/10X,'TOTAL EXPECTED COST =',F8.0/
/
     $10X,'TOTAL EXP. LOSS'/10X,'----------------',F6.2,' %'/10X,
     $'TOTAL ACT. COST'//////////////////)
      STOP
      END
C
      FUNCTION EXPNO(X)
C
C     GIVES THE EXPECTED NO. OF CLAIMS IN THE INTERVAL WITH
C     UPPER BOUNDARY X.
      COMMON ADJMEW,SIGMA,S,NCLM
      Z=(ALOG(X)-ADJMEW)/SIGMA
C     S15ABF IS NAG ROUTINE FOR STANDARD NORMAL CUM. DIS. FUN.
      W=NCLM*S15ABF(Z,0)
      EXPNO=W-S
      S=W
      RETURN
      END
      FINISH
```

```
      LIBRARY(SUBGROUPNAGF)
      LIST(LP)
      PROGRAM(P10)
      COMPACT
      INPUT 1,5=CR0
      OUTPUT 2,6=LP0
      TRACE 0
      COMPRESS INTEGER AND LOGICAL
      END

      MASTER LEAST SQUARES ESTIMATION FOR 3-PAR LOGNORMAL
C
C                   *** PROGRAM P10 ***
C
C     THIS PROGRAM FINDS ESTIMATES OF THE PARAMETERS OF THE
C     3-PAR LOGNORMAL DIS. BY LEAST SQUARES METHOD
C
      DIMENSION X(100),XLOG(100),Y(100),CSUM(100),XMID(100)
      DIMENSION XX(100),NAME(20),Z(100)
      COMMON M,K,S,NAME,X,XMID,CSUM,Y,Z
      REAL MEW
C     READ DESCRIPTION OF DATA AND NO. OF INTERVALS
      READ(5,101)NAME,M
  101 FORMAT(20A4/I0)
C     READ IN THE DATA
      DO  3 I=1,M
      READ(5,102)Y(I)
  102 FORMAT(F0.0)
      IF(I-20)1,1,2
    1 X(I)=30.*I+0.5
      XMID(I)=X(I)-15.0
      GOTO 3
    2 X(I)=(I-14)*100.+0.5
      XMID(I)=X(I)-50.
    3 CONTINUE
      CSUM(1)=Y(1)
      DO 4 I=2,M
    4 CSUM(I)=CSUM(I-1)+Y(I)
      S=CSUM(M)
      K=0
C     FIND K THE NO. OF INTERVALS WITH NON-ZERO FREQUENCY
      DO 5 I=1,M-1
      N2=CSUM(I+1)
      N1=CSUM(I)
      IF(N2 .EQ. N1)GOTO 5
      K=K+1
      XX(K)=X(I)
      Z(K)=CSUM(I)/S
    5 CONTINUE
C     FIND THE VALUES OF Z (THE STANDARD NORMAL VARIATE) CORRESP.
C     TO EACH SAMPLE QUANTILE
      DO 6 I=1,K
      T=1.0-Z(I)
      Z(I)=HAS(T)
    6 CONTINUE
C     READ IN C AND IPRINT . IPRINT=0 MEANS THAT ONLY SSD
C     CORRESP. TO C IS REQUIRED. IPRINT=1 WHEN AN EXTENSIVE
C     TABLE IS REQUIRED.
   13 READ(5,103)C,IPRINT
  103 FORMAT(F0.0,I0)
C     IF C>999 IS SPECIFIED THE PROGRAM STOPS
      IF(C .GT. 999)GOTO 111
      DO 14 I=1,K
      XLOG(I)=ALOG(XX(I)+C)
   14 CONTINUE
      CALL LISTER(XLOG,C,IPRINT)
      GOTO 13
```

```
      CALL CHAHOL(15H *ULOG*L(X+C)*.)
      CALL MOVTO2(55.,315.)
      CALL CHAHOL(30H*ULOG(LOG(1./(1.-F(*LX*U))))*.)
      CALL MOVTO2(110.,87.)
      CALL CHAHOL(48HS*LTRAIGHT LINE PLOT OF *ULOGNORMAL DIS. *LFOR*
.)
      CALL MOVTO2(110.,80.)
      CALL CHAARR(NAME,20,4)
      CALL MOVTO2(140.,71.)
      CALL SYMBOL(7)
      CALL MOVTO2(140.,70.)
      CALL CHAHOL(10H *L: C=0*.)
      CALL MOVTO2(140.,64.)
      CALL SYMBOL(5)
      CALL MOVTO2(140.,63.)
      CALL CHAHOL(11H *L: C=10*.)
      CALL MOVTO2(140.,59.)
      CALL SYMBOL(3)
      CALL MOVTO2(140.,58.)
      CALL CHAHOL(11H *L: C=15*.)
      CALL MOVTO2(140.,54.)
      CALL SYMBOL(1)
      CALL MOVTO2(140.,53.)
      CALL CHAHOL(11H *L: C=20*.)
      CALL MOVTO2(140. ,49.)
      CALL SYMBOL(8)
      CALL MOVTO2(140.,48.)
      CALL CHAHOL(11H *L: C=25*.)
      CALL CONVRT(C.0)
      CALL GRASYM(XLOG,F,K,7,0)
      CALL CONVRT(10.0)
      CALL GRASYM(XLOG,F,K,5,0)
      CALL CONVRT(15.0)
      CALL GRASYM(XLOG,F,K,3,0)
      CALL CONVRT(20.0)
      CALL GRASYM(XLOG,F,K,1,0)
      CALL CONVRT(25.0)
      CALL GRASYM(XLOG,F,K,8,0)
      CALL DEVEND
      STOP
      END
C
      SUBROUTINE CONVRT(C)
C
C     GIVES ARRAY XLOG WHICH CONTAINS VALUES OF LOG(X+C)
      DIMENSION X(60),XLOG(60)
      COMMON K,X,XLOG
      DO 1 I=1,K
      XLOG(I)=ALOG(X(I)+C)
    1 CONTINUE
      RETURN
      END
C
      FUNCTION HAS(Q)
C
C     FINDS VARIATE CORRESPONDING TO THE AREA Q UNDER THE
C     UPPER TAIL OF STANDARD NORMAL FREQ. CURVE, 0<Q<0.5
      IF(Q)6,5,6
    5 Q=0.000001
    6 P=Q
      IF(Q-.5)1,1,2
    2 P=1.-Q
    1 Z=ALOG(1./(P*P))
      Y=SQRT(Z)
      HAS=Y-((2.515517+0.802853*Y+.010328*Z)/(1.+1.432788*Y+
     $0.189269*Z+0.001308*Y*Z))
      IF(Q-.5)3,3,4
    4 HAS=-HAS
    3 RETURN
      END
      FINISH
```

```
  111 STOP
      END
C
      SUBROUTINE LISTS2(G,C,IPRINT)
C
C     FINDS ESTIMATES OF MEW AND SIGMA AS WELL AS SSD CORRESPONDING
C     TO A GIVEN VALUE OF C. IF REQUIRED WILL PRINT A TABLE
      DIMENSION X(100),RESLT(21),FEXP(100),G(100),CSUM(100),Z(100)
      DIMENSION Y(100),NAME(20),AME2(100),XMID(100),ELOSS(100)
      COMMON M,K,S,NAME,X,XMID,CSUM,Y,Z
      REAL MEW
      XM=0.
      YM=99999.
      IFAIL=0
C     G02CCF IS THE NAG ROUTINE FOR FINDING THE LS REGRESSION
C     LINE FOR A GIVEN SET OF POINTS
      CALL G02CCF(K,G,Z,XM,YM,RESLT,IFAIL)
C     IF IPRINT=0 JUST PRINT C AND SSD ELSE IF IPRINT=1 PRINT AN
C     EXTENSIVE TABLE OF RESULTS
      IF(IPRINT .EQ. 1)GOTO 112
      WRITE(6,900)C,RESLT(16)
  900 FORMAT(/30X,'C=',F6.2,5X,'SSD=',F7.5)
      GOTO 113
  112 B=RESLT(6)
      A=RESLT(7)
      SIGMA=1.0/B
      SIGMA2=SIGMA*SIGMA
      MEW=-A/B
      ALPHA=-C+EXP(MEW+0.5*SIGMA2)
      ETA=SQRT(EXP(SIGMA2)-1.)
      BETA=ETA*(ALPHA+C)
      WRITE(6,901)NAME,C,MEW,SIGMA2,ALPHA,BETA,RESLT(16)
  901 FORMAT(////25X,20A4////'          C=',F5.2,
     $/'        PARAMETERS OF THE LOGNORMAL DIS.',
     $' ESTIMATED BY L.S. REGRESSION :-',/,
     $10X,'MEW =',F6.3,7X,'          SIGMA SQ.=',F6.3/,
     $/10X,'MEAN=',F8.3,12X,'  S.D.=',F8.3//10X,'SSD=',F5.3)
      WRITE(6,902)
  902 FORMAT(///6X,'AMOUNT £   ACTUAL     EXPECTED   ACTUAL-',
     $'    EXPECTED',/17X,'CL. NO.    CL. NO.   EXPECTED',
     $'    LOSS      (A-E)**2/E ')
      SEXP=0.0
      SS=0.
      Q1=0.0
      ACOST=0.
      ELOSS(M+1)=0.
      DO 7 I=1,M
      V=(ALOG(X(I)+C)-MEW)/SIGMA
      Q2=S15ABF(V,G)
      FEXP(I)=(Q2-Q1)*S
      NEXP=FEXP(I)+.5
      SEXP=SEXP+NEXP
      CSUM(I)=Y(I)-NEXP
      ELOSS(I)=XMID(I)*CSUM(I)
      ELOSS(M+1)=ELOSS(M+1)+ELOSS(I)
      ACOST=ACOST+XMID(I)*Y(I)
      IF(NEXP .LT. 5)GOTO 20
      AME2(I)=CSUM(I)*CSUM(I)/NEXP
      SS=SS+AME2(I)
   20 Q1=Q2
    7 CONTINUE
      DO 11 I=1,M
      IF(I-20)8,8,9
    8 LB=X(I)-29.
      GOTO 10
    9 LB=X(I)-99.
   10 LU=X(I)
      IF(FEXP(I) .LT. 5)GOTO 21
      WRITE(6,903)LB,LU,Y(I),FEXP(I),CSUM(I),ELOSS(I),AME2(I)
  903 FORMAT(4X,I5,'-',I4,F9.0,F10.0,F11.0,4X,F8.0,3X,F8.3)
      GOTO 11
   21 WRITE(6,904)LB,LU,Y(I),FEXP(I),CSUM(I),ELOSS(I)
  904 FORMAT(4X,I5,'-',I4,F9.0,F10.0,F11.0,4X,F8.0)
   11 CONTINUE
      RATIO=100.*ELOSS(M+1)/ACOST
      WRITE(6,905)S,SEXP,ELOSS(M+1),SS,RATIO
  905 FORMAT('   ---------------------------------------------------
     $'--------------',/,'      TOTAL    ',F9.0,F10.0,15X,F8.0,
     $3X,F8.3/'    ------------------------------------------------',
     $'-------------------'/57X,'D.F. = '/10X,
     $'TOTAL EXP. LOSS'/10X,'----------------',F5.1,' %'/
     $10X,'TOTAL ACT. COST'//////////////////////)
  113 RETURN
      END
C
      FUNCTION HAS(Q)
C
C     GIVES THE STANDARD NORMAL VARIATE CORRESPONDING TO THE
C     AREA Q UNDER THE UPPER TAIL OF ITS FREQ. CURVE, 0<Q<0.5
      IF(Q)6,5,6
    5 Q=0.000001
    6 P=Q
      IF(Q-.5)1,1,2
    2 P=1.-Q
    1 Z=ALOG(1./(P*P))
      Y=SQRT(Z)
      HAS=Y-((2.515517+0.802853*Y+.010328*Z)/(1.+1.432788*Y+
     $0.189269*Z+0.001308*Y*Z))
      IF(Q-.5)3,3,4
    4 HAS=-HAS
    3 RETURN
      END
      FINISH
```

```
      LIBRARY(SUBGROUPGRAF)
      LIBRARY(SUBGROUPGINO)
      LIBRARY(SUBGROUPNAGF)
      LIST(LP)
      PROGRAM(P11)
      COMPACT
      INPUT 1,5=CR0
      OUTPUT 2,6=LP0
      TRACE 2,100
      COMPRESS INTEGER AND LOGICAL
      END

      MASTER LOGNORMAL LEAST SQUARES LINE PLOT
C
C                        *** PROGRAM P11 ***
C
C     THIS PROGRAM PLOTS THE SAMPLE POINTS ACCORDING TO THE
C     3-PARAMETER LOGNORMAL DIS. ALONG WITH THE L.S. LINE
C     FITTED TO THEM.
C
      DIMENSION NAME(20),IPAR(20), X(60),XLOG(60),Y(60),F(60),CSUM(6
0)
      REAL MEW
C     READ IN DESCRIPTION OF DATA,NO. OF INTERVALS, VALUES OF
C     THE PARAMETERS ,AND IPAR WHICH IS AN ARRAY HOLDING
C     THE NAMES AND VALUES OF THE PARAMETERS TO BE PRINTED ON THE
C     GRAPH.
      READ(5,101)NAME,M,MEW,SIGMA2,C,IPAR
  101 FORMAT(20A4/I0/3F0.0/20A4)
      WRITE(6,900)NAME
  900 FORMAT(////' *********',20A4)
C     READ IN THE DATA
      DO  3 I=1,M
      READ(5,102)Y(I)
  102 FORMAT(F0.0)
      IF(I-20)1,1,2
    1 X(I)=30.*I+0.5
      GOTO 3
    2 X(I)=(I-14)*100.+0.5
    3 CONTINUE
      CSUM(1)=Y(1)
      DO 4 I=2,M
    4 CSUM(I)=CSUM(I-1)+Y(I)
      S=CSUM(M)
      K=0
      DO 5 I=1,M-1
      N2=CSUM(I+1)
      N1=CSUM(I)
      IF(N2 .EQ. N1)GOTO 5
      K=K+1
      X(K)=X(I)
      F(K)=CSUM(I)/S
    5 CONTINUE
      DO 6 I=1,K
      F(I)=HAS(1.-F(I))
      XLOG(I)=ALOG(X(I)+C)
    6 CONTINUE
      G=1./SQRT(SIGMA2)
      D=-MEW*G
      T1=XLOG(1)-0.5
      Z1=G*T1+D
      TK=XLOG(K)+0.5
      DT=TK-T1
      DZ=G*DT
      P1=G*XLOG(1)+D
      P2=G*XLOG(K)+D
C     CALL GINO ROUTINES FOR PLOTTING THE GRAPH
      CALL M19346
      CALL DEVPAP(700.,700.,0)
      CALL CHASWI(1)
      CALL CHASIZ(3.0,3.0)
      CALL MOVTO2(30.,30.)
      CALL LINBY2(210.,0.)
      CALL LINBY2(0.,297.)
      CALL LINBY2(-210.,0.)
      CALL LINBY2(0.,-297.)
      CALL AXIPOS(1,50.,185.,180.,1)
      CALL AXIPOS(0,50.,185.,270.,2)
      CALL AXISCA(1,7,2.,9.,1)
      CALL AXISCA(1,8,-4.,4.,2)
      CALL AXIDRA(2,1,1)
      CALL AXIDRA(-2,-1,2)
      CALL MOVTO2(205.,187.)
      CALL CHAHOL(15H *ULOG*L(X+C)*.)
      CALL MOVTO2(55.,315.)
      CALL CHAHOL(5H*LZ*.)
      CALL MOVTO2(110.,87.)
      CALL CHAHOL(48HS*LTRAIGHT LINE PLOT OF *ULOGNORMAL DIS. *LFOR*
.)
      CALL MOVTO2(110.,80.)
      CALL CHAARR(NAME,20,4)
      CALL MOVTO2(140.,71.)
      CALL SYMBOL(3)
      CALL MOVTO2(140.,70.)
      CALL CHAHOL(9H *L: C=*.)
      IC=C
      CALL CHAINT(IC,2)
      CALL MOVTO2(135.,67.)
      CALL LINBY2(5.,0.)
      CALL MOVTO2(140.,65.)
      CALL CHAHOL(29H *L: *UL*LEAST SQUARES LINE*.)
      CALL MOVTO2(145.,60.)
      CALL CHAARR(IPAR,20,4)
      CALL GRASYM(XLOG,F,K,3,0)
      CALL GRAMOV(XLOG(1),P1)
      CALL GRALIN(XLOG(K),P2)
      CALL DEVEND
      STOP
      END
C
      FUNCTION HAS(Q)
C
C     GIVES THE VARIATE CORRESPONDING TO AREA Q UNDER THE UPPER
C     TAIL OF THE STANDARD NORMAL FREQ. CURVE, 0<Q<0.5
      IF(Q)6,5,6
    5 Q=0.000001
    6 P=Q
      IF(Q-.5)1,1,2
    2 P=1.-Q
    1 Z=ALOG(1./(P*P))
      Y=SQRT(Z)
      HAS=Y-((2.515517+0.802853*Y+.010328*Z)/(1.+1.432788*Y+
     $0.189269*Z+0.001308*Y*Z))
      IF(Q-.5)3,3,4
    4 HAS=-HAS
    3 RETURN
      END
      FINISH
```

```
      LIBRARY(SUBGROUPNAGF)
      LIST(LP)
      PROGRAM(P12)
      COMPACT
      INPUT 1,5=CR0
      OUTPUT 2,6=LP0
      TRACE 0
      COMPRESS INTEGER AND LOGICAL
      END

      MASTER MULTINOMIAL MAXLIKE METHOD FOR 3-PAR LOGNORMAL
C
C                    *** PROGRAM P12 ***
C
C     THIS PROGRAM ESTIMATES THE PARAMETERS BY MML METHOD
C
      DIMENSION X(100),XLOG(100),Y(100),AME2(100),CSUM(100),W(40)
      DIMENSION XX(100),FEXP(100),NAME(20),P(3),XTOL(5),YY(100)
      DIMENSION ELOSS(100),XMID(100)
      COMMON K,XX,YY
      EXTERNAL FUNCT,MONIT
      REAL MEW
C     READ IN DESCRIPTION OF DATA AND NO. OF INTERVALS
      READ(5,101)NAME,M
  101 FORMAT(20A4/I0)
      WRITE(6,900)NAME
  900 FORMAT(////' *********',20A4)
C     READ IN STARTING VALUES FOR THE PARAMETERS
      READ(5,102)(P(I),I=1,3)
  102 FORMAT(3F0.0)
      WRITE(6,905)(P(I),I=1,3)
  905 FORMAT(//'    INITIAL VALUES=',3(3X,F8.3)//)
      S=0.0
C     READ IN THE DATA
      DO  3 I=1,M
      READ(5,103)Y(I)
  103 FORMAT(F0.0)
      S=S+Y(I)
      IF(I-20)1,1,2
    1 X(I)=30.*I+0.5
      XMID(I)=X(I)-15.0
      GOTO 3
    2 X(I)=(I-14)*100.+0.5
      XMID(I)=X(I)-50.0
    3 CONTINUE
      K=0
      DO 5 I=1,M
      NI=Y(I)
      IF(NI .EQ. 0)GOTO 5
      K=K+1
      XX(K)=X(I)
      YY(K)=Y(I)/S
    5 CONTINUE
C     SET THE VALUES FOR THE PARAMETERS OF THE E04CAF NAG
C     SUBROUTINE. THIS  SUBROUTINE FINDS THE MAX. OF A GENERAL
C     FUNCTION OF SEVERAL VARIABLES.
      XTOL(1)=.1
      XTOL(2)=.001
      XTOL(3)=.001
      STEPMX=1000.0
      ISTOP=2
      IW=40
      IFAIL=0
      MAXCAL=200
      IPRINT=4
      CALL E04CAF(3,P,F,XTOL,STEPMX,ISTOP,W,IW,FUNCT,MONIT,
     $IPRINT,MAXCAL,IFAIL)
      IF(IFAIL .NE. 0)WRITE(6,910)IFAIL
  910 FORMAT('  IFAIL=',I2///)
      F=-F
      SIGMA=P(3)
      SIGMA2=P(3)*P(3)
      MEW=P(2)
      C=P(1)
C     FIND THE MEAN AND S.D. OF THE FITTED MODEL
      ALPHA=-C+EXP(MEW+.5*SIGMA2)
      ETA=SQRT(EXP(SIGMA2)-1.0)
      BETA=ETA*(ALPHA+C)
      WRITE(6,915)NAME,C,MEW,SIGMA2,ALPHA,BETA
  915 FORMAT(///////////20X,'*** 3-PARAMETER LOGNORMAL DIS. ***'//
     $27X,20A4//12X,'ESTIMATION BY MULTINOMIAL MAX. ',
     $'LIKELIHOOD METHOD :-',/16X,'C=',F6.2,5X,'MEW=',
     $F8.3,5X,'SIGMA SQ.=',F6.3//16X,'MEAN=',F8.3,17X,'S.D.=',F8.3)
      WRITE(6,920)
  920 FORMAT(///6X,'AMOUNT E   ACTUAL    EXPECTED   ACTUAL-',
     $'    EXPECTED',/17X,'CL. NO.    CL. NO.   EXPECTED',
     $'     LOSS      (A-E)**2/E ')
C     FIND E, A-E, EXP. LOSS AND (A-E)**2/E
      SEXP=0.0
      SS=0.
      Q1=0.0
      ACOST=0.
      ELOSS(M+1)=0.
      DO 7 I=1,M
      V=(ALOG(X(I)+C)-MEW)/SIGMA
      Q2=S15ABF(V,0)
      FEXP(I)=(Q2-Q1)*S
      NEXP=FEXP(I)+.5
      SEXP=SEXP+NEXP
      CSUM(I)=Y(I)-NEXP
      ELOSS(I)=XMID(I)*CSUM(I)
      ELOSS(M+1)=ELOSS(M+1)+ELOSS(I)
      ACOST=ACOST+XMID(I)*Y(I)
      IF(NEXP .LT. 5)GOTO 20
      AME2(I)=CSUM(I)*CSUM(I)/NEXP
      SS=SS+AME2(I)
   20 Q1=Q2
    7 CONTINUE
      DO 11 I=1,M
      IF(I-20)8,8,9
    8 LB=X(I)-29.
      GOTO 10
    9 LB=X(I)-99.
   10 LU=X(I)
      IF(FEXP(I) .LT. 5)GOTO 21
      WRITE(6,925)LB,LU,Y(I),FEXP(I),CSUM(I),ELOSS(I),AME2(I)
  925 FORMAT(4X,I5,'-',I4,F9.0,F10.0,F11.0,4X,F8.0,3X,F6.3)
      GOTO 11
   21 WRITE(6,930)LB,LU,Y(I),FEXP(I),CSUM(I),ELOSS(I)
  930 FORMAT(4X,I5,'-',I4,F9.0,F10.0,F11.0,4X,F8.0)
   11 CONTINUE
      RATIO=100.*ELOSS(M+1)/ACOST
      WRITE(6,935)S,SEXP,ELOSS(M+1),SS,RATIO
  935 FORMAT('     ------------------------------------------------------------
     $'--------------',/,'       TOTAL    ',F9.0,F10.0,15X,F8.0,
     $3X,F8.3/'      -----------------------------------------------',
     $'---------------------'/57X,'D.F. = '/13X,
     $'TOTAL EXP. LOSS'/10X,'----------------',F5.1,' %'/
     $10X,'TOTAL ACT. COST'/////////////////////)
      STOP
      END
C
      SUBROUTINE FUNCT(NP,PC,FC)
C     THIS SUBROUTINE IS PART OF E04CAF ROUTINE AND FINDS THE VALUE
C     OF THE FUNCTION FOR A SET OF PARAMETERS.
```

```
      DIMENSION PC(NP),XX(100),YY(100)
      COMMON K,XX,YY
      REAL MEW
      C=PC(1)
      MEW=PC(2)
      SIGMA=PC(3)
      Z=(ALOG(XX(1)+C)-MEW)/SIGMA
      D1=S15ABF(Z,0)
      S=YY(1)*ALOG(D1)
      DO 1 I=2,K
      IF(I .LT. 21)GOTO 2
      Z=(ALOG(XX(I)-100.0+C)-MEW)/SIGMA
      D1=S15ABF(Z,0)
    2 Z=(ALOG(XX(I)+C)-MEW)/SIGMA
      D2=S15ABF(Z,0)
      S=S+YY(I)*ALOG(D2-D1)
      D1=D2
    1 CONTINUE
      FC=-S
      RETURN
      END
C
      SUBROUTINE MONIT(N,PC,FC,NCALL)
C     THIS IS PART OF E04CAF ROUTINE AND WRITES THE VALUES
C     OF THE FUNCTION AND ITS PARAMETERS AT DIFF. ITERATIONS.
      DIMENSION PC(N)
      WRITE(6,940)NCALL,FC
  940 FORMAT('   AFTER',I5,'  FUNCTION CALLS THE VALUE IS',3X,F10.6)
      SIGMA2=PC(3)*PC(3)
      WRITE(6,945)(PC(I),I=1,N),SIGMA2
  945 FORMAT('   AT THE POINT  C=',F6.3,'   MEW=',F7.5,'  SIGMA=',
     £F7.5,'   S2=',F7.5)
      RETURN
      END
      FINISH
```

```
      LIBRARY(SUBGROUPGRAF)
      LIBRARY(SUBGROUPGINO)
      LIBRARY(SUBGROUPNAGF)
      LIST(LP)
      PROGRAM(P13)
      COMPACT
      INPUT 1,5=CR0
      OUTPUT 2,6=LP0
      TRACE 1
      COMPRESS INTEGER AND LOGICAL
      END

      MASTER WEIBULL GRAPHICAL TEST
C
C                    *** PROGRAM P13 ***
C
C     THIS PROGRAM PLOTS THE SAMPLE POINTS FOR DIFF. VALUES
C     OF C IN ORDER TO PERFORM WEIBULL TEST
C
      DIMENSION NAME(20), X(60),XLOG(60),Y(60),F(60),CSUM(60)
      COMMON K,X,XLOG
C     READ DESCRIPTION OF DATA AND NO. OF INTERVALS
      READ(5,101)NAME,M
  101 FORMAT(20A4/I0)
      WRITE(6,900)NAME
  900 FORMAT(////' *********',20A4)
C     READ IN THE DATA
      DO  3 I=1,M
      READ(5,102)Y(I)
  102 FORMAT(F0.0)
      IF(I-20)1,1,2
    1 X(I)=30.*I
      GOTO 3
    2 X(I)=(I-14)*100.
    3 CONTINUE
C     FIND CUM. DIS. FUNCTION
      CSUM(1)=Y(1)
      DO 4 I=2,M
    4 CSUM(I)=CSUM(I-1)+Y(I)
      S=CSUM(M)
      K=0
      DO 5 I=1,M-1
      N2=CSUM(I+1)
      N1=CSUM(I)
      IF(N2 .EQ. N1)GOTO 5
      K=K+1
      X(K)=X(I)
      F(K)=CSUM(I)/S
    5 CONTINUE
C     FIND K THE NO. OF INTERVALS WITH NON-ZERO FREQUENCY
      DO 6 I=1,K
      T=1./(1.-F(I))
      T=ALOG(T)
      F(I)=ALOG(T)
    6 CONTINUE
C     CALL GINO  ROUTINES TO PLOT THE GRAPHS
      CALL M19346
      CALL DEVPAP(700.,700.,0)
      CALL CHASWI(1)
      CALL CHASIZ(3.0,3.0)

      CALL MOVTO2(30.,30.)
      CALL LINBY2(210.,0.)
      CALL LINBY2(0.,297.)
      CALL LINBY2(-210.,0.)
      CALL LINBY2(0.,-297.)
      CALL AXIPOS(1,50.,185.,180.,1)
```

```
      CALL AXIPOS(0,50.,185.,270.,2)
      CALL AXISCA(1,7,2.,9.,1)
      CALL AXISCA(1,6,-3.,3.,2)
      CALL AXIDRA(2,1,1)
      CALL AXIDRA(-2,-1,2)
      CALL MOVTO2(210.,187.)
      CALL CHAHOL(15H *ULOC*L(X-C)*.)
      CALL MOVTO2(55.,315.)
      CALL CHAHOL(30H*ULOG(LOG(1./(1.-F(*LX*U))))*.)
      CALL MOVTO2(110.,87.)
      CALL CHAHOL(46HS*LTRAIGHT LINE PLOT OF *UWEIBULL DIS. *LFOR*.)
      CALL MOVTO2(110.,80.)
      CALL CHAARR(NAME,20,4)
      CALL MOVTO2(140.,71.)
      CALL SYMBOL(7)
      CALL MOVTO2(140.,70.)
      CALL CHAHOL(10H *L: C=0*.)
      CALL MOVTO2(140.,64.)
      CALL SYMBOL(5)
      CALL MOVTO2(140.,63.)
      CALL CHAHOL(11H *L: C=10*.)
      CALL MOVTO2(140.,59.)
      CALL SYMBOL(3)
      CALL MOVTO2(140.,58.)
      CALL CHAHOL(11H *L: C=15*.)
      CALL MOVTO2(140.,54.)
      CALL SYMBOL(1)
      CALL MOVTO2(140.,53.)
      CALL CHAHOL(11H *L: C=20*.)
      CALL MOVTO2(140. ,49.)
      CALL SYMBOL(8)
      CALL MOVTO2(140.,48.)
      CALL CHAHOL(11H *L: C=25*.)
      CALL CONVRT(0.0)
      CALL GRASYM(XLOG,F,K,7,0)
      CALL CONVRT(10.0)
      CALL GRASYM(XLOG,F,K,5,0)
      CALL CONVRT(15.0)
      CALL GRASYM(XLOG,F,K,3,0)
      CALL CONVRT(20.0)
      CALL GRASYM(XLOG,F,K,1,0)
      CALL CONVRT(25.0)
      CALL GRASYM(XLOG,F,K,8,0)
      CALL DEVEND
      STOP
      END
      SUBROUTINE CONVRT(C)
      DIMENSION X(60),XLOG(60)
      COMMON K,X,XLOG
      DO 1 I=1,K
      XLOG(I)=ALOG(X(I)-C)
    1 CONTINUE
      RETURN
      END
      FINISH
```

```
      COMPACT
      INPUT 1,5=CR0
      OUTPUT 2,6=LP0
      TRACE 1
      COMPRESS INTEGER AND LOGICAL
      END

      MASTER ESTIMATION 3-PAR WEIBULL LS METHOD
C
C                  *** PROGRAM P14 ***
C
C     THIS PROGRAM ESTIMATES THE PARAMETERS OF THE 3-PARAMETER
C     WEIBULL DIS. BY LEAST SQUARES METHOD
C
      DIMENSION X(100),XLOG(100),Y(100),CSUM(100),XMID(100)
      DIMENSION XX(100),NAME(20),Z(100)
      COMMON M,K,S,NAME,X,XMID,CSUM,Y,Z
C     READ  THE DESCRIPTION OF DATA AND NO. OF INTERVALS
      READ(5,101)NAME,M
  101 FORMAT(20A4/I0)
C     READ IN THE DATA
      DO  3 I=1,M
      READ(5,102)Y(I)
  102 FORMAT(F0.0)
      IF(I-20)1,1,2
    1 X(I)=30.*I+0.5
      XMID(I)=X(I)-15.0
      GOTO 3
    2 X(I)=(I-14)*100.+0.5
      XMID(I)=X(I)-50.
    3 CONTINUE
      CSUM(1)=Y(1)
      DO 4 I=2,M
    4 CSUM(I)=CSUM(I-1)+Y(I)
      S=CSUM(M)
C     FIND K THE NO. OF INTERVALS WITH NON ZERO FREQUENCY
      K=0
      DO 5 I=1,M-1
      N2=CSUM(I+1)
      N1=CSUM(I)
      IF(N2 .EQ. N1)GOTO 5
      K=K+1
      XX(K)=X(I)
      Z(K)=CSUM(I)/S
    5 CONTINUE
C     CALCULATE THE VALUES OF LOG(LOG(1./(1.-F(X)))) FROM THE SAMPLE
      DO 6 I=1,K
      T=1./(1.0-Z(I))
      T=ALOG(T)
      Z(I)=ALOG(T)
    6 CONTINUE
C     READ IN C AND IPRINT . IPRINT=0 MEANS  THAT AN EXTENSIVE
C     TABLE OF RESULTS IS NOT REQUIRED. IPRINT=1 IS THE OPPOSITE.
   13 READ(5,103)C,IPRINT
  103 FORMAT(F0.0,I0)
C     IF A VALUE OF C > 999 IS SPECIFIED THEN PROGRAM STOPS
      IF(C .GT. 999)GOTO 111
      DO 14 I=1,K
      XLOG(I)=ALOG(XX(I)-C)
   14 CONTINUE
      CALL LISTSQ(XLOG,C,IPRINT)
      GOTO 13
  111 STOP
      END
C
      SUBROUTINE LISTSQ(G,C,IPRINT)
C
C     FINDS ESTIMATES OF A & B CORRESPONDING TO GIVEN C AND IF
C     REQUIRED PRINTS A TABLE  OF RESULTS
C
      DIMENSION X(100),RESLT(21),FEXP(100),G(100),CSUM(100),Z(100)
      DIMENSION Y(100),NAME(20),AME2(100),XMID(100),ELOGS(100)
      COMMON M,K,S,NAME,X,XMID,CSUM,Y,Z
      REAL MEW
```

```
      IFAIL=0
C     G02CCF IS A NAG ROUTINE FOR FINDING THE L.S. REGRESSION
C     LINE FOR A SET OF GIVEN POINTS
      CALL G02CCF(K,G,Z,XM,YM,RESLT,IFAIL)
C     IF IPRINT=0 THEN JUST PRINT C AND ITS CORRESP. SSD
      IF(IPRINT .EQ. 1)GOTO 112
      WRITE(6,900)C,RESLT(16)
 900  FORMAT(/30X,'C=',F6.2,5X,'SSD=',F7.5)
      GOTO 113
 112  B=RESLT(6)
      A=RESLT(7)
      A=EXP(-A/B)
      Z1=1.+1./B
      Z2=1.+2./B
C     S14AAF IS THE NAG ROUTINE FOR CALCULATING VALUES OF
C     THE COMPLETE GAMMA FUNCTION
      GAM1=S14AAF(Z1,0)
      GAM2=S14AAF(Z2,0)
      XMEAN=C+A*GAM1
      SDX=A*SQRT(GAM2-GAM1*GAM1)
      WRITE(6,901)NAME,C,A,B,XMEAN,SDX,RESLT(16)
 901  FORMAT(////25X,20A4////'           C=',F5.2,
     $/'        PARAMETERS OF THE WEIBULL DIS.',
     $' ESTIMATED BY L.S. REGRESSION :-',/,
     $10X,'A =',F8.3,15X,'             B=',F6.3/,
     $/10X,'MEAN=',F8.3,20X,'  S.D.=',F8.3//10X,'SSD=',F5.3)
      WRITE(6,902)
 902  FORMAT(///6X,'AMOUNT E   ACTUAL    EXPECTED   ACTUAL-',
     $'    EXPECTED',/17X,'CL. NO.    CL. NO.   EXPECTED',
     $'     LOSS      (A-E)**2/E ')
      SEXP=0.0
      SS=0.
      Q1=0.0
      ACOST=0.
      ELOSS(M+1)=0.
      DO 7 I=1,M
      V=-((X(I)-C)/A)**B
      Q2=1.-EXP(V)
      FEXP(I)=(Q2-Q1)*S
      NEXP=FEXP(I)+.5
      SEXP=SEXP+NEXP
      CSUM(I)=Y(I)-NEXP
      ELOSS(I)=XMID(I)*CSUM(I)
      ELOSS(M+1)=ELOSS(M+1)+ELOSS(I)
      ACOST=ACOST+XMID(I)*Y(I)
      IF(NEXP .LT. 5)GOTO 20
      AVE2(I)=CSUM(I)*CSUM(I)/NEXP
      SS=SS+AVE2(I)
 20   Q1=Q2
 7    CONTINUE
      DO 11 I=1,M
      IF(I-20)8,8,9
 8    LB=X(I)-29.
      GOTO 10
 9    LB=X(I)-99.
 10   LU=X(I)
      IF(FEXP(I) .LT. 5)GOTO 21
      WRITE(6,903)LB,LU,Y(I),FEXP(I),CSUM(I),ELOSS(I),AVE2(I)
 903  FORMAT(4X,I5,'-',I4,F9.0,F10.0,F11.0,4X,F8.0,3X,F8.3)
      GOTO 11
 21   WRITE(6,904)LB,LU,Y(I),FEXP(I),CSUM(I),ELOSS(I)
 904  FORMAT(4X,I5,'-',I4,F9.0,F10.0,F11.0,4X,F8.0)
 11   CONTINUE
      RATIO=100.*ELOSS(M+1)/ACOST
      WRITE(6,905)S,SEXP,ELOSS(M+1),SS,RATIO
 905  FORMAT('         ------------------------------------------------
     $'---------------','/,'      TOTAL    ',F9.0,F10.0,15X,F8.0,
     $3X,F8.3/'      ----------------------------------------',
     $'-------------------'/57X,'D.F. = '/10X,
     $'TOTAL EXP. LOSS'/10X,'----------------',F5.1,' %'/
     $10X,'TOTAL ACT. COST'//////////////////////)
 113  RETURN
      END
```

```
      LIBRARY(SUBGROUPGRAF)
      LIBRARY(SUBGROUPGINC)
      LIBRARY(SUBGROUPNAGF)
      LIST(LP)
      PROGRAM(P15)
      COMPACT
      INPUT 1,5=CR0
      OUTPUT 2,6=LP0
      TRACE 2,100
      COMPRESS INTEGER AND LOGICAL
      END

      MASTER WEIBULL LEAST SQUARES LINE  PLOT
C
C                     *** PROGRAM P15 ***
C
C     THIS PROGRAM PLOTS THE SAMPLE POINTS AND THE L.S. LINE
C     FITTED TO THEM
C
      DIMENSION NAME(20),IPAR(20), X(60),XLOG(60),Y(60),F(60),CSUM(6
0)
C     READ IN DESCRIPTION OF THE DATA, THE NO. OF INTERVALS,
C     THE PARAMETERS AND THE ARRAY IPAR WHICH HAS THE
C     DESCRIPTION OF THE PARAMETERS WHICH IS TO BE PRINTED
C     ON THE GRAPH.
      READ(5,101)NAME,M,A,B,C,IPAR
 101  FORMAT(20A4/I0/3F0.0/20A4)
      WRITE(6,900)NAME
 900  FORMAT(////' *********',20A4)
C     READ IN DATA AND FIND SAMPLE POINTS AND FORM THE EQUATION
C     OF THE L.S. LINE.
      DO  3 I=1,M
      READ(5,102)Y(I)
 102  FORMAT(F0.0)
      IF(I-20)1,1,2
 1    X(I)=30.*I+0.5
      GOTO 3
 2    X(I)=(I-14)*100.+0.5
 3    CONTINUE
      CSUM(1)=Y(1)
      DO 4 I=2,M
 4    CSUM(I)=CSUM(I-1)+Y(I)
      S=CSUM(M)
      K=0
      DO 5 I=1,M-1
      N2=CSUM(I+1)
      N1=CSUM(I)
      IF(N2 .EQ. N1)GOTO 5
      K=K+1
      X(K)=X(I)
      F(K)=CSUM(I)/S
 5    CONTINUE
      DO 6 I=1,K
      T=1./(1.-F(I))
      T=ALOG(T)
      F(I)=ALOG(T)
      XLOG(I)=ALOG(X(I)-C)
 6    CONTINUE
      D=-B*ALOG(A)
      T1=XLOG(1)-0.5
      Z1=B*T1+D
      TK=XLOG(K)+0.5
      DT=TK-T1
      DZ=B*DT
      P1=B*XLOG(1)+D
      P2=B*XLOG(K)+D
C     CALL GINO PLOTTING ROUTINES TO DRAW THE GRAPH.
      CALL H19346
```

```
      CALL DEVPAP(700.,700.,0)
      CALL CHASWI(1)
      CALL CHASIZ(3.0,3.0)
      CALL MOVTO2(30.,30.)
      CALL LINBY2(210.,0.)
      CALL LINBY2(0.,297.)
      CALL LINBY2(-210.,0.)
      CALL LINBY2(0.,-297.)
      CALL AXIPOS(1,50.,105.,180.,1)
      CALL AXIPOS(0,50.,105.,270.,2)
      CALL AXISCA(1,7,2.,9.,1)
      CALL AXISCA(1,6,-3.,3.,2)
      CALL AXIDRA(2,1,1)
      CALL AXIDRA(-2,-1,2)
      CALL MOVTO2(210.,197.)
      CALL CHAHOL(15H *ULOG*L(X-C)*.)
      CALL MOVTO2(55.,315.)
      CALL CHAHOL(30H*ULOG(LOG(1./(1.-F(*LX*U))))*.)
      CALL MOVTO2(110.,87.)
      CALL CHAHOL(46HS*LTRAIGHT LINE PLOT OF *UWEIBULL DIS. *LFOR*.)
      CALL MOVTO2(110.,80.)
      CALL CHAARR(NAME,20,4)
      CALL MOVTO2(140.,71.)
      CALL SYMBOL(3)
      CALL MOVTO2(140.,70.)
      CALL CHAHOL(9H *L: C=*.)
      IC=C
      CALL CHAINT(IC,2)
      CALL MOVTO2(135.,67.)
      CALL LINBY2(5.,0.)
      CALL MOVTO2(140.,65.)
      CALL CHAHOL(29H *L: *UL*LEAST SQUARES LINE*.)
      CALL MOVTO2(145.,60.)
      CALL CHAARR(IPAR,20,4)
      CALL GRASYM(XLOG,F,K,5,0)
      CALL GRAMOV(XLOG(1),P1)
      CALL GRALIN(XLOG(K),P2)
      CALL DEVEND
      STOP
      END
      FINISH
```



```
      LIBRARY(SUBGROUPNAGF)
      LIST(LP)
      PROGRAM(P16)
      COMPACT
      INPUT 1,5=CR0
      OUTPUT 2,6=LP0
      TRACE 1
      COMPRESS INTEGER AND LOGICAL
      END

      MASTER ESTIMATION BY MML METHOD FOR 2-PAR WEIBULL
C
C                      *** PROGRAM P16 ***
C
C     THIS PROGRAM FINDS ESTIMATES OF THE PARAMETERS OF THE
C     2-PAR WEIBULL DIS. BY THE MML METHOD.
C
      DIMENSION X(100),XLOG(100),Y(100),AME2(100),CSUM(100),#(40)
      DIMENSION YY(100),XX(100),XMID(100),FEXP(100),P(2),ELOSS(100)
      DIMENSION NAME(20)
      COMMON K,XX,YY
C     READ DESCRIPTION OF DATA AND NO. OF INTERVALS
      READ(5,101)NAME,M
  101 FORMAT(20A4/I0)
      WRITE(6,900)NAME
  900 FORMAT(////' *********',20A4////////)
C     READ IN THE STARTING VALUES FOR THE PARAMETERS
      READ(5,102)(P(I),I=1,2)
  102 FORMAT(2F0.0)
      S=0.0
C     READ IN THE DATA AND COMPUTE THE SAMPLE CUM. DIS. FUN.
      DO  3 I=1,M
      READ(5,103)Y(I)
  103 FORMAT(F0.0)
      S=S+Y(I)
      IF(I-20)1,1,2
    1 X(I)=30.*I+0.5
      XMID(I)=X(I)-15.
      GOTO 3
    2 X(I)=(I-14)*100.+0.5
      XMID(I)=X(I)-50.
    3 CONTINUE
C     FIND K THE NO. OF NON-ZERO INTERVALS.
      K=0
      DO 5 I=1,M
      NI=Y(I)
      IF(NI .EQ. 0)GOTO 5
      K=K+1
      XX(K)=X(I)
      YY(K)=Y(I)/S
    5 CONTINUE
      IFAIL=1
C     E04CEF IS THE NAG ROUTINE FOR MINIMIZING (OR MAX.) A
C     GENERAL FUN. OF SEVERAL VARIABLES.
      CALL E04CEF(2,P,F,W,40,IFAIL)
      IF(IFAIL .NE. 0)WRITE(6,901)IFAIL
  901 FORMAT('  IFAIL=',I2////////)
C     MML ESTIMATES ARE :-
      A=P(1)
      B=P(2)
C     FIND MEAN AND S.D. OF THE  WEIBULL MODEL. S14AAF IS THE
C     NAG ROUTINE FOR  COMPLETE GAMMA FUNCTION.
      Z1=1.+1./B
      Z2=1.+2./B
      GAM1=S14AAF(Z1,0)
      GAM2=S14AAF(Z2,0)
      XMEAN=A*GAM1
      V=GAM2-GAM1*GAM1
      SDX=A*SQRT(V)
      WRITE(6,902)NAME,A,B,XMEAN,SDX
  902 FORMAT(////20X,'*** TWO-PARAMETER WEIBULL DIS. ***'//
     $27X,20A4//12X,'ESTIMATION BY MULTINOMIAL MAX.',
     $'LIKELIHOOD METHOD :-',
     $/16X,'A =',F9.4,'           B=',F7.4/,
```

```
      LIBRARY(SUBGROUPNAGF)
      LIST(LP)
      PROGRAM(P17)
      COMPACT
      INPUT 1,5=CR0
      OUTPUT 2,6=LP0
      TRACE 0
      COMPRESS INTEGER AND LOGICAL
      END

      MASTER MULTINOMIAL MAXLIKE METHOD FOR 3-PAR WEIBULL DIS
C
C                     *** PROGRAM P17 ***
C
C     THIS PROGRAM FINDS THE ESTIMATES OF THE 3-PARAMETER WEIBULL
C     DIS. BY MML METHOD.
C
      DIMENSION X(100),XLOG(100),Y(100),AME2(100),CSUM(100),W(40)
      DIMENSION XX(100),FEXP(100),NAME(20),P(3),XTOL(5),YY(100)
      DIMENSION ELOSS(100),XMID(100)
      COMMON K,XX,YY
      EXTERNAL FUNCT,MONIT
C     READ IN THE DESCRIPTION OF THE DATA AND NO. OF INTERVALS.
      READ(5,101)NAME,M
  101 FORMAT(20A4/I0)
      WRITE(6,900)NAME
  900 FORMAT(////' *********',20A4)
C     READ IN THE STARTING VALUES FOR THE PARAMETERS
      READ(5,102)(P(I),I=1,3)
  102 FORMAT(3F0.0)
      WRITE(6,905)(P(I),I=1,3)
  905 FORMAT(//'   INITIAL VALUES=',3(3X,F8.3)//)
      S=0.0
C     READ IN THE DATA
      DO  3 I=1,M
      READ(5,103)Y(I)
  103 FORMAT(F0.0)
      S=S+Y(I)
      IF(I-20)1,1,2
    1 X(I)=30.*I+0.5
      XMID(I)=X(I)-15.0
      GOTO 3
    2 X(I)=(I-14)*100.+0.5
      XMID(I)=X(I)-50.0
    3 CONTINUE
      K=0
      DO 5 I=1,M
      NI=Y(I)
      IF(NI .EQ. 0)GOTO 5
      K=K+1
      XX(K)=X(I)
      YY(K)=Y(I)/S
    5 CONTINUE
C     SET THE PARAMETERS OF E04CAF NAG ROUTINE FOR MAXIMIZING
C     A GENERAL FUNCTION OF SEVERAL VARIABLES
      XTOL(1)=.5
      XTOL(2)=.5
      XTOL(3)=.005
      STEPMX=10.0
      ISTOP=2
      IW=49
      IFAIL=0
      MAXCAL=300
      IPRINT=4
      CALL E04CAF(3,P,F,XTOL,STEPMX,ISTOP,W,IW,FUNCT,MONIT,
     $IPRINT,MAXCAL,IFAIL)

     $/16X,'MEAN=',F8.3,'          S.D.=',F8.3/
      WRITE(6,903)
  903 FORMAT(///6X,'AMOUNT C   ACTUAL     EXPECTED   ACTUAL-',
     $'   EXPECTED',/17X,'CL. NO.    CL. NO.   EXPECTED',
     $'    LOSS     (A-E)**2/E ')
C     COMPUTE THE TABLE OF RESULTS
      SEXP=0.0
      SS=0.
      ACOST=0.0
      ELOSS(M+1)=0.0
      Q1=0.0
      DO 7 I=1,M
      V=-(X(I)/A)**B
      Q2=1.-EXP(V)
      FEXP(I)=(Q2-Q1)*S
      NEXP=FEXP(I)+.5
      SEXP=SEXP+NEXP
      CSUM(I)=Y(I)-NEXP
      ACOST=ACOST+XMID(I)*Y(I)
      ELOSS(I)=XMID(I)*CSUM(I)
      ELOSS(M+1)=ELOSS(M+1)+ELOSS(I)
      IF(NEXP .LT. 5)GOTO 20
      AME2(I)=CSUM(I)*CSUM(I)/NEXP
      SS=SS+AME2(I)
   20 Q1=Q2
    7 CONTINUE
      DO 11 I=1,M
      IF(I-20)8,8,9
    8 LB=X(I)-29.
      GOTO 10
    9 LB=X(I)-99.
   10 LU=X(I)-0.5
      IF(FEXP(I) .LT. 5)GOTO 21
      WRITE(6,904)LB,LU,Y(I),FEXP(I),CSUM(I),ELOSS(I),AME2(I)
  904 FORMAT(4X,I5,'-',I4,F9.0,F10.0,F11.0,4X,F8.0,3X,F8.3)
      GOTO 11
   21 WRITE(6,905)LB,LU,Y(I),FEXP(I),CSUM(I),ELOSS(I)
  905 FORMAT(4X,I5,'-',I4,F9.0,F10.0,F11.0,4X,F8.0)
   11 CONTINUE
      RATIO=100.*ELOSS(M+1)/ACOST
      WRITE(6,906)S,SEXP,ELOSS(M+1),SS,RATIO
  906 FORMAT('          ------------------------------------------',
     $'----------------------'/,'       TOTAL    ',F9.0,1X,F9.0,15X,
     $F8.0,3X,F8.3/'    ------------------------------------------',
     $'-----------------------------'/57X,'D.F. = ',/10X,
     $'TOTAL EXP. LOSS'/10X,'-----------------',F5.1,' %',/
     $10X,'TOTAL ACT. COST'////////////////////////////////)
      STOP
      END
C
      SUBROUTINE FUNCT1(NP,PC,FC)
C
C     FINDS THE VALUE OF THE LOGLIKELIHOOD FUNCTION FOR A GIVEN
C     SET OF PARAMETERS. THIS IS REQUIRED BY E04CEF ROUTINE.
      DIMENSION PC(NP),XX(100),YY(100)
      COMMON K,XX,YY
      A=PC(1)
      B=PC(2)
      Z=-(XX(1)/A)**B
      D1=1.-EXP(Z)
      S=YY(1)*ALOG(D1)
      DO 1 I=2,K
      IF(I .LT. 21)GOTO 2
      Z=-((XX(I)-100.)/A)**B
      D1=1.-EXP(Z)
    2 Z=-(XX(I)/A)**B
      D2=1.-EXP(Z)
      S=S+YY(I)*ALOG(D2-D1)
      D1=D2
    1 CONTINUE
      FC=-S
      RETURN
      END
      FINISH
```

```
      IF(IFAIL .NE. 0)WRITE(6,910)IFAIL
  910 FORMAT('  IFAIL=',I2///)
      C=P(1)
      A=P(2)
      B=P(3)
C     FIND THE MEAN AND S.D. OF THE FITTED MODEL
      Z1=1.+1./B
      Z2=1.+2./B
      GAM1=S14AAF(Z1,0)
      GAM2=S14AAF(Z2,0)
      XMEAN=C+A*GAM1
      SDX=A*SQRT(GAM2-GAM1*GAM1)
      WRITE(6,915)NAME,C,A,B,XMEAN,SDX
  915 FORMAT(///////////20X,'*** 3-PARAMETER WEIBULL DIS. ***'//
     $27X,20A4//12X,'ESTIMATION BY MULTINOMIAL MAX. ',
     $'LIKELIHOOD METHOD :-',/16X,'C=',F6.2,7X,'A=',
     $F8.3,5X,'B=',F6.3//16X,'MEAN=',F8.3,17X,'S.D.=',F8.3)
      WRITE(6,920)
  920 FORMAT(///6X,'AMOUNT £   ACTUAL    EXPECTED   ACTUAL-',
     $'    EXPECTED',/17X,'CL. NO.     CL. NO.    EXPECTED',
     $'     LOSS      (A-E)**2/E  ')
C     FIND VALUES OF E, A-E, EXP. LOSS, (A-E)**2/E
      SEXP=0.0
      SS=0.
      Q1=0.0
      ACOST=0.
      ELOSS(M+1)=0.
      D7 7 I=1,M
      V=((X(I)-C)/A)**B
      Q2=1.-EXP(-V)
      FEXP(I)=(Q2-Q1)*S
      NEXP=FEXP(I)+.5
      SEXP=SEXP+NEXP
      CSUM(I)=Y(I)-NEXP
      ELOSS(I)=XMID(I)*CSUM(I)
      ELOSS(M+1)=ELOSS(M+1)+ELOSS(I)
      ACOST=ACOST+XMID(I)*Y(I)
      IF(NEXP .LT. 5)GOTO 20
      AME2(I)=CSUM(I)*CSUM(I)/NEXP
      SS=SS+AME2(I)
   20 Q1=Q2
    7 CONTINUE
      DO 11 I=1,M
      IF(I-20)8,8,9
    8 LB=X(I)-29.
      GOTO 10
    9 LB=X(I)-99.
   10 LU=X(I)
      IF(FEXP(I) .LT. 5)GOTO 21
      WRITE(6,925)LB,LU,Y(I),FEXP(I),CSUM(I),ELOSS(I),AME2(I)
  925 FORMAT(4X,I5,'-',I4,F9.0,F10.0,F11.0,4X,F8.0,3X,F8.3)
      GOTO 11
   21 WRITE(6,930)LB,LU,Y(I),FEXP(I),CSUM(I),ELOSS(I)
  930 FORMAT(4X,I5,'-',I4,F9.0,F10.0,F11.0,4X,F8.0)
   11 CONTINUE
      RATIO=100.*ELOSS(M+1)/ACOST
      WRITE(6,935)S,SEXP,ELOSS(M+1),SS,RATIO
  935 FORMAT('    ------------------------------------------------
     $'-------------',/,'      TOTAL    ',F9.0,F10.0,15X,F8.0,
     $3X,F8.3/'     --------------------------------------',
     $'-------------------'/57X,'D.F. = '/10X,
     $'TOTAL EXP. LOSS'/10X,'---------------',F5.1,' %'/
     $10X,'TOTAL ACT. COST'/////////////////////)
      STOP
      END
C
      SUBROUTINE FUNCT(NP,PC,FC)
C
C     FINDS THE VALUE OF THE FUNCTION FOR A SET OF PARAMETERS.
C     IT IS REQUIRED BY E04CAF ROUTINE.
      DIMENSION PC(NP),XX(100),YY(100)
      COMMON K,XX,YY
      C=PC(1)
      A=PC(2)
      B=PC(3)
      V=((XX(1)-C)/A)**B
      D1=1.-EXP(-V)
      S=YY(1)*ALOG(D1)
      DO 1 I=2,K
      IF(I .LT. 21)GOTO 2
      V=((XX(I)-100.-C)/A)**B
      D1=1.-EXP(-V)
    2 V=((XX(I)-C)/A)**B
      D2=1.-EXP(-V)
      S=S+YY(I)*ALOG(D2-D1)
      D1=D2
    1 CONTINUE
      FC=-S
      RETURN
      END
C
      SUBROUTINE MONIT(N,PC,FC,NCALL)
C
C     WRITES THE VALUE OF THE FUNCTION AND ITS CORRESP. PARAMETERS
C     AT VARIOUS ITERATIONS. IT IS REQUIRED BY E04CAF ROUTINE.
      DIMENSION PC(N)
      WRITE(6,940)NCALL,FC
  940 FORMAT('   AFTER',I5,'  FUNCTION CALLS THE VALUE IS',3X,F10.6)
      WRITE(6,945)(PC(I),I=1,N)
  945 FORMAT('    AT THE POINT  C=',F6.3,'    A=',F10.5,'  B=',
     $F7.5)
      RETURN
      END
      FINISH
```

```
      LIST(LP)
      PROGRAM(P1B)
      COMPACT
      INPUT 1,5=CRO
      OUTPUT 2,6=LPO
      TRACE 1
      COMPRESS INTEGER AND LOGICAL
      END

      MASTER ESTIMATION BY MML METHOD FOR 2-PAR INV GAUSSIAN
C
C                    *** PROGRAM P1B ***
C
C     THIS PROGRAM FINDS ESTIMATES OF THE PARAMETERS OF THE
C     2-PAR INVERSE GAUSSIAN DIS. BY THE MML METHOD.
C
      DIMENSION X(100),XLOG(100),Y(100),AME2(100),CSUM(100),W(40)
      DIMENSION YY(100),XX(100),XMID(100),FEXP(100),P(2),ELOSS(100)
      DIMENSION NAME(20)
      COMMON K,XX,YY
      REAL MEW,LAMDA
C     READ DESCRIPTION OF DATA AND NO. OF INTERVALS
      READ(5,101)NAME,M
  101 FORMAT(20A4/I0)
      WRITE(6,900)NAME
  900 FORMAT(////' *********',20A4////////)
C     READ IN STARTING VALUES FOR THE PARAMETERS
      READ(5,102)(P(I),I=1,2)
  102 FORMAT(2F0.0)
      S=0.0
C     READ IN THE DATA AND COMPUTE THE SAMPLE CUM. DIS. FUN.
      DO  3 I=1,M
      READ(5,103)Y(I)
  103 FORMAT(F0.0)
      S=S+Y(I)
      IF(I-20)1,1,2
    1 X(I)=30.*I+0.5
      XMID(I)=X(I)-15.
      GOTO 3
    2 X(I)=(I-14)*100.+0.5
      XMID(I)=X(I)-50.
    3 CONTINUE
C     FIND K THE NO. OF INTERVALS WITH NON-ZERO FREQUENCY
      K=0
      DO 5 I=1,M
      NI=Y(I)
      IF(NI .EQ. 0)GOTO 5
      K=K+1
      XX(K)=X(I)
      YY(K)=Y(I)/S
    5 CONTINUE
      IFAIL=1
C     E04CEF IS NAG ROUTINE FOR MINIMIZING (OR MAX. ) A GENERAL
C     FUNCTION OF SEVERAL VARIABLES.
      CALL E04CEF(2,P,F,W,40,IFAIL)
      IF(IFAIL .NE. 0)WRITE(6,901)IFAIL
  901 FORMAT('  IFAIL=',I2////////)
C     MML ESTIMATES ARE:-
      MEW=P(1)
      LAMDA=P(2)
C     FIND S.D. OF THE MODEL
      SDX=MEW*SQRT(MEW/LAMDA)
      WRITE(6,902)NAME,MEW,LAMDA,MEW,SDX
  902 FORMAT(////16X,'*** TWO-PARAMETER INVERSE GAUSSIAN DIS. ***'/
     $/27X,20A4//12X,'ESTIMATION BY MULTINOMIAL MAX.',
     $'LIKELIHOOD METHOD :-',
     $/16X,'MEW =',F9.4,'             LAMDA=',F9.4/,
     $/16X,'MEAN=',F8.3,'             S.D.=',F8.3)
      WRITE(6,903)
  903 FORMAT(///6X,'AMOUNT £   ACTUAL     EXPECTED   ACTUAL-',
     $'    EXPECTED',/17X,'CL. NO.     CL. NO.   EXPECTED',
     $'     LOSS      (A-E)**2/E ')
C     COMPUTE THE TABLE OF RESULTS
      SEXP=0.0
      ACOST=0.0
      ELOSS(M+1)=0.0
      Q1=0.0
      DO 7 I=1,M
      Q2=SINVG(X(I),MEW,LAMDA)
      FEXP(I)=(Q2-Q1)*S
      NEXP=FEXP(I)+.5
      SEXP=SEXP+NEXP
      CSUM(I)=Y(I)-NEXP
      ACOST=ACOST+XMID(I)*Y(I)
      ELOSS(I)=XMID(I)*CSUM(I)
      ELOSS(M+1)=ELOSS(M+1)+ELOSS(I)
      IF(NEXP .LT. 5)GOTO 20
      AME2(I)=CSUM(I)*CSUM(I)/NEXP
      SS=SS+AME2(I)
   20 Q1=Q2
    7 CONTINUE
      DO 11 I=1,M
      IF(I-20)8,8,9
    8 LB=X(I)-29.
      GOTO 10
    9 LB=X(I)-99.
   10 LU=X(I)-0.5
      IF(FEXP(I) .LT. 5)GOTO 21
      WRITE(6,904)LB,LU,Y(I),FEXP(I),CSUM(I),ELOSS(I),AME2(I)
  904 FORMAT(4X,I5,'-',I4,F9.0,F10.0,F11.0,4X,F8.0,3X,F8.3)
      GOTO 11
   21 WRITE(6,905)LB,LU,Y(I),FEXP(I),CSUM(I),ELOSS(I)
  905 FORMAT(4X,I5,'-',I4,F9.0,F10.0,F11.0,4X,F8.0)
   11 CONTINUE
      RATIO=100.*ELOSS(M+1)/ACOST
      WRITE(6,906)S,SEXP,ELOSS(M+1),SS,RATIO
  906 FORMAT('       ----------------------------------------',
     $'---------------------'/,'       TOTAL     ',F9.0,1X,F9.0,15X,
     $F8.0,3X,F8.3/'        ----------------------------------------',
     $'----------------------------'/57X,'D.F. = ',/10X,
     $'TOTAL EXP. LOSS'/10X,'----------------',F5.1,' %',/
     $10X,'TOTAL ACT. COST'//////////////////////////////)
      STOP
      END
C
      SUBROUTINE FUNCT1(NP,PC,FC)
C
C     FINDS THE VALUE OF THE LOGLIKELIHOOD FUNCTION FOR A GIVEN
C     SET OF PARAMETERS. THIS IS REQUIRED BY E04CEF ROUTINE.
      DIMENSION PC(NP),XX(100),YY(100)
      COMMON K,XX,YY
      REAL MEW,LAMDA
      MEW=PC(1)
      LAMDA=PC(2)
      D1=SINVG(XX(1),MEW,LAMDA)
      S=YY(1)*ALOG(D1)
      DO 1 I=2,K
      IF(I .LT. 21)GOTO 2
      D1=SINVG(XX(I)-100.0,MEW,LAMDA)
    2 D2=SINVG(XX(I),MEW,LAMDA)
      S=S+YY(I)*ALOG(D2-D1)
      D1=D2
    1 CONTINUE
      FC=-S
      RETURN
      END
      FUNCTION SINVG(X,MEW,LAMDA)
      REAL MEW,LAMDA
      R1=SQRT(LAMDA/X)
      R2=X/MEW
      R3=EXP(2.*LAMDA/MEW)
      V1=R1*(R2-1.)
      V2=-R1*(R2+1.)
C     S15ABF IS NAG ROUTINE FOR STANDARD NORMAL CUM. DIS. FUN.
      SINVG=S15ABF(V1,0)+R3*S15ABF(V2,0)
      RETURN
      END
      FINISH
```

```
      LIBRARY(SUBGROUPNAGF)
      LIST(LP)
      PROGRAM(P19)
      COMPACT
      INPUT 1,5=CR0
      OUTPUT 2,6=LP0
      TRACE 0
      COMPRESS INTEGER AND LOGICAL
      END

      MASTER MULTINOMIAL MAXLIKE METHOD FOR 3-PAR INV. GAUSSIAN DIS
C
C                      *** PROGRAM P19 ***
C
C     THIS PROGRAM  ESSTIMATES PARAMETERS OF THE 3-PARAMETER INV.
C     GAUSSIAN DIS. BY MML METHOD.
C
      DIMENSION X(100),XLOG(100),Y(100),AME2(100),CSUM(100),W(40)
      DIMENSION XX(100),FEXP(100),NAME(20),P(3),XTOL(5),YY(100)
      DIMENSION ELOSS(100),XMID(100)
      COMMON K,XX,YY
      EXTERNAL FUNCT,MONIT
      REAL MEW,LAMDA
      REAL MEW
C     READ DESCRIPTION OF DATA AND THE NO. OF INTERVALS
      READ(5,101)NAME,M
  101 FORMAT(20A4/I0)
      WRITE(6,900)NAME
  900 FORMAT(////' *********',20A4)
C     READ STARTING VALUES FOR THE PARAMETERS
      READ(5,102)(P(I),I=1,3)
  102 FORMAT(3F0.0)
      WRITE(6,905)(P(I),I=1,3)
  905 FORMAT(//'    INITIAL VALUES=',3(3X,F8.3)//)
      S=0.0
C     READ IN THE DATA
      DO  3 I=1,M
      READ(5,103)Y(I)
  103 FORMAT(F0.0)
      S=S+Y(I)
      IF(I-20)1,1,2
    1 X(I)=30.*I+0.5
      XMID(I)=X(I)-15.0
      GOTO 3
    2 X(I)=(I-14)*100.+0.5
      XMID(I)=X(I)-50.0
    3 CONTINUE
      K=0
      DO 5 I=1,M
      NI=Y(I)
      IF(NI .EQ. 0)GOTO 5
      K=K+1
      XX(K)=X(I)
      YY(K)=Y(I)/S
    5 CONTINUE
C     SET THE PARAMETERS OF E04CAF ROUTINE
      XTOL(1)=.5
      XTOL(2)=0.5
      XTOL(3)=0.5
      STEPMX=10.0
      ISTOP=2
      IW=40
      IFAIL=0
      MAXCAL=300
      IPRINT=4
C     E04CAF IS THE NAG ROUTINE FOR MAXIMISING A GENERAL FUN.
      CALL E04CAF(3,P,F,XTOL,STEPMX,ISTOP,W,IW,FUNCT,MONIT,
     $IPRINT,MAXCAL,IFAIL)
      IF(IFAIL .NE. 0)WRITE(6,910)IFAIL
  910 FORMAT('  IFAIL=',I2///)
C     THE MML ESTIMATES ARE
      C=P(1)
      MEW=P(2)
      LAMDA=P(3)
C     FIND MEAN AND S.D. OF THE FITTED MODEL
      XMEAN=-C+MEW
      SDX=MEW*SQRT(MEW/LAMDA)
      WRITE(6,915)NAME,C,MEW,LAMDA,XMEAN,SDX
  915 FORMAT(///////////20X,'*** 3-PARAMETER INV. GAUSSIAN DIS. ***
'//
     $27X,20A4//12X,'ESTIMATION BY MULTINOMIAL MAX. ',
     $'LIKELIHOOD METHOD :-',/16X,'C=',F6.2,5X,'MEW=',
     $F8.3,5X,'LAMDA=',F8.3//16X,'MEAN=',F8.3,17X,'S.D.=',F8.3)
      WRITE(6,920)
  920 FORMAT(///6X,'AMOUNT £    ACTUAL     EXPECTED    ACTUAL-',
     $'     EXPECTED',/17X,'CL. NO.     CL. NO.    EXPECTED',
     $'      LOSS      (A-E)**2/E ')
      SEXP=0.0
      SS=0.
      Q1=0.0
      ACOST=0.
      ELOSS(M+1)=0.
      DO 7 I=1,M
      Q2=SINVG(X(I),C,MEW,LAMDA)
      FEXP(I)=(Q2-Q1)*S
      NEXP=FEXP(I)+.5
      SEXP=SEXP+NEXP
      CSUM(I)=Y(I)-NEXP
      ELOSS(I)=XMID(I)*CSUM(I)
      ELOSS(M+1)=ELOSS(M+1)+ELOSS(I)
      ACOST=ACOST+XMID(I)*Y(I)
      IF(NEXP .LT. 5)GOTO 20
      AME2(I)=CSUM(I)*CSUM(I)/NEXP
      SS=SS+AME2(I)
   20 Q1=Q2
    7 CONTINUE
      DO 11 I=1,M
      IF(I-20)8,8,9
    8 LB=X(I)-29.
      GOTO 10
    9 LB=X(I)-99.
   10 LU=X(I)
      IF(FEXP(I) .LT. 5)GOTO 21
      WRITE(6,925)LB,LU,Y(I),FEXP(I),CSUM(I),ELOSS(I),AME2(I)
  925 FORMAT(4X,I5,'-',I4,F9.0,F10.0,F11.0,4X,F8.0,3X,F8.3)
      GOTO 11
   21 WRITE(6,930)LB,LU,Y(I),FEXP(I),CSUM(I),ELOSS(I)
  930 FORMAT(4X,I5,'-',I4,F9.0,F10.0,F11.0,4X,F8.0)
   11 CONTINUE
      RATIO=100.*ELOSS(M+1)/ACOST
      WRITE(6,935)S,SEXP,ELOSS(M+1),SS,RATIO
  935 FORMAT('     ----------------------------------------------
',
     $'--------------',/,'       TOTAL    ',F9.0,F10.0,15X,F8.0,
     $3X,F8.3/'     ------------------------------------------',
     $'---------------------'/57X,'D.F. = '/10X,
     $'TOTAL EXP. LOSS'/10X,'----------------',F5.1,' %'/
     $10X,'TOTAL ACT. COST'/////////////////////)
      STOP
      END
C
      SUBROUTINE FUNCT(NP,PC,FC)
C
C     FINDS THE VALUE OF THE FUNCTION FOR A SET OF PARAMETERS
```

```
      DIMENSION PC(NP),XX(100),YY(100)
      COMMON K,XX,YY
      REAL MEW,LAMDA
      C=PC(1)
      MEW=PC(2)
      LAMDA=PC(3)
      D1=SINVG(XX(1),C,MEW,LAMDA)
      S=YY(1)*ALOG(D1)
      DO 1 I=2,K
      IF(I .LT. 21)GOTO 2
      D1=SINVG(XX(I)-100.0,C,MEW,LAMDA)
    2 D2=SINVG(XX(I),C,MEW,LAMDA)
      S=S+YY(I)*ALOG(D2-D1)
      D1=D2
    1 CONTINUE
      FC=-S
      RETURN
      END
C
      FUNCTION SINVG(X,C,MEW,LAMDA)
C
C     GIVES THE CUM. DIS. FUN. FOR INV. GAUSSIAN
      REAL MEW,LAMDA
      R1=SQRT(LAMDA/(X+C))
      R2=(X+C)/MEW
      R3=EXP(2.*LAMDA/MEW)
      V1=R1*(R2-1.)
      V2=-R1*(R2+1.)
      SINVG=S15ABF(V1,0)+R3*S15ABF(V2,0)
      RETURN
      END
C
      SUBROUTINE MONIT(N,PC,FC,NCALL)
C
C     PRINTS VALUE OF THE FUN. AND ITS PARAMETERS AT DIFF.
C     ITERATIONS. THIS IS REQUIRED BY E04CAF NAG ROUTINE
      DIMENSION PC(N)
      REAL MEW,LAMDA
      WRITE(6,940)NCALL,FC
  940 FORMAT('   AFTER',I5,'  FUNCTION CALLS THE VALUE IS',3X,F10.6)
      WRITE(6,945)(PC(I),I=1,N)
  945 FORMAT('   AT THE POINT  C=',F6.3,'   MEW=',F10.5,'  LAMDA=',
     SF7.5)
      RETURN
      END
      FINISH
```

```
      LIBRARY(SUBGROUPGRAF)
      LIBRARY(SUBGROUPGINO)
      LIBRARY(SUBGROUPNAGF)
      LIST(LP)
      PROGRAM(P20)
      COMPACT
      INPUT 1,5=CR0
      OUTPUT 2,6=LP0
      TRACE 0
      COMPRESS INTEGER AND LOGICAL
      END

      MASTER HISTOGRAM AND INV. GAUSSIAN FREQ. PLOT
C
C                  *** PROGRAM P20 ***
C
C     THIS PROGRAM PLOTS THE HISTOGRAM OF A SAMPLE OF AD DATA
C     ALONG WITH THE FREQ. CURVE OF THE 3-PAR INV. GAUSSIAN
C     DIS. FITTED TO IT. 3 PLOTS ARE PRODUCED. ONE FOR ALL
C     CLAIM AMOUNTS, ONE FOR CLAIM<£600 AND ONE FOR CLAIMS>£600.
C
      DIMENSION X(70),Y(70),NAME(20),A(70),XMID(70),F(300),Z(300)
      DIMENSION ITITLE(20),IPAR(20),H(70),FF(300),ZZ(300)
      COMMON SFACX,SFACY,NAME,ITITLE,IPAR
      REAL MEW,LAMDA
      DATA PI/3.1415926536/
C     READ DESCRIPTION OF DATA(NAME),THE METHOD OF ESTIMATION(
C     ITITLE),THE DES. OF PARAMETERS TO BE PRINTED ON THE PLOTS(
C     IPAR),NO. OF INTERVALS(M),VALUES OF THE PARAMETERS.
      READ(5,101)NAME,ITITLE,IPAR,M,C,MEW,LAMDA
  101 FORMAT(20A4/20A4/20A4/I0/3F0.0)
      WRITE(6,900)NAME,ITITLE
  900 FORMAT(20A4//20A4)
      X(1)=0.0
      DO 1 I=1,20
      X(I+1)=X(I)+30.
      XMID(I)=(X(I+1)+X(I))/2.
    1 CONTINUE
      DO 2 I=21,M
      X(I+1)=X(I)+100.
      XMID(I)=(X(I+1)+X(I))/2.
    2 CONTINUE
      TOTAL=0.
C     READ IN THE DATA
      DO 3 I=1,M
      READ(5,102)Y(I)
  102 FORMAT(F0.0)
      TOTAL=TOTAL+Y(I)
    3 CONTINUE
      XMPLS1=X(M+1)
      IF(XMPLS1 .LT. 2599)XMPLS1=XMPLS1+100.
      LENTHX=XMPLS1/10.+36
      NFF=LENTHX-95
      C1=TOTAL*SQRT(LAMDA/(2.*PI))
      C2=LAMDA/(2.*MEW*MEW)
      DO 30 I=1,LENTHX
      IF(I-39)6,6,7
    6 R=I
      GOTO 8
    7 R=(I-36)*10.
    8 Z(I)=R
      R=R+C
      F(I)=C1/(R*SQRT(R))*EXP(-C2*(R-MEW)*(1.-MEW/R))
   30 CONTINUE
      K=NGT25(F)
      N95=95-K
      LENTHX=LENTHX-K
```

```
      DO 31 I=1,LENTHX
      Z(I)=Z(I+K)
      F(I)=F(I+K)
   31 CONTINUE
      J=0
      DO 32 J=1,NFF
      IF(F(N95+J) .LT. 0.5)GOTO 33
      J=J+1
   32 CONTINUE
   33 NFF=NFF-J
      DO 40 I=1,NFF
      FF(I)=F(I+N95+J)
      ZZ(I)=Z(I+N95+J)
   40 CONTINUE
      A(1)=X(2)-X(1)
      Y(1)=Y(1)/A(1)
      DO 4 I=1,M-1
      B=X(I+2)-X(I+1)
      Y(I+1)=Y(I+1)/B
      H(I)=Y(I+1)-Y(I)
      A(I+1)=B
    4 CONTINUE
      H(M)=Y(M)
      SFACX=0.1
      SFACY=8.
C     CALL GINO ROUTINES TO PLOT THE GRAPHS
      CALL M19346
      CALL DEVPAP(700.,700.,0)
      CALL CHASWI(1)
      CALL CHASIZ(3.0,3.0)
      CALL FRAME(1)
      CALL AXIS1
      CALL GRACUR(Z,F,LENTHX)
      CALL MOVTO2(50.,37.)
      Y1=Y(1)*SFACY
      CALL MOVBY2(0.,Y1)
      DO 5 I=1,M-1
      C=A(I)
      D=H(I)
      CALL HIS(C,D)
    5 CONTINUE
      C=A(M)
      D=-H(M)
      CALL HIS(C,D)
      CALL FRAME(2)
      CALL AXIS2
      N96=N95+1
      CALL GRACUR(Z,F,N96)
      SFACX=0.46
      SFACY=8.
      CALL MOVTO2(45.,257.)
      CALL MOVBY2(0.,Y1)
      DO 9 I=1,19
      C=A(I)
      D=H(I)
      CALL HIS(C,D)
    9 CONTINUE
      C=A(20)
      D=-Y(20)
      CALL HIS(C,D)
      CALL FRAME(3)
      CALL AXIS3
      CALL GRACUR(ZZ,FF,NFF)
      SFACX=0.13
      SFACY=350.
      Y1=Y(21)*SFACY
      CALL MOVTO2(50.,450.)
      CALL MOVBY2(0.,Y1)
      DO 11 I=21,M-1
      C=A(I)
      D=H(I)
      CALL HIS(C,D)
   11 CONTINUE
      CALL HIS(A(M),-H(M))
      CALL DEVEND
      STOP
      END
C
      SUBROUTINE FRAME(J)
C
C     DRAWS AN A4 FRAME FOR PLOT NO. J=1 OR 2OR 3
      IF(J-2)1,2,3
    1 CALL MOVTO2(30.,30.)
      GOTO 4
    2 CALL MOVTO2(30.,250.)
      GOTO 4
    3 CALL MOVTO2(30.,470.)
    4 CALL LINBY2(297.,0.)
      CALL LINBY2(0.,210.)
      CALL LINBY2(-297.,0.)
      CALL LINBY2(0.,-210.)
      RETURN
      END
C
      SUBROUTINE AXIS1
C
C     DRAWS THE AXES FOR PLOT NO. 1
      CALL AXIPOS(1,50.,37.,270.,1)
      CALL AXIPOS(1,50.,37.,200.,2)
      CALL AXISCA(3,27,0.,2700.,1)
      CALL AXISCA(3,25,0.,25.,2)
      CALL AXIDRA(2,1,1)
      CALL AXIDRA(-2,-1,2)
      CALL DES(50.,37.,323.,230.)
      RETURN
      END
C
      SUBROUTINE AXIS2
C
C     DRAWS THE AXES FOR PLOT NO. 2
      CALL AXIPOS(1,45.,257.,276.,1)
      CALL AXIPOS(1,45.,257.,200.,2)
      CALL AXISCA(3,20,0.,600.,1)
      CALL AXISCA(3,25,0.,25.,2)
      CALL AXIDRA(2,1,1)
      CALL AXIDRA(-2,-1,2)
      CALL DES(45.,257.,322.,450.)
      RETURN
      END
C
      SUBROUTINE AXIS3
C
C     DRAWS THE AXES FOR PLOT NO. 3
      CALL AXIPOS(1,50.,490.,260.,1)
      CALL AXIPOS(1,50.,490.,180.,2)
      CALL AXISCA(3,20,600.,2600.,1)
      CALL AXISCA(3,10,0.,.50,2)
      CALL AXIDRA(2,1,1)
      CALL AXIDRA(-2,-1,2)
      CALL DES(50.,490.,315.,665.)
      RETURN
      END
C
```

```
      SUBROUTINE DES(X0,Y0,X,Y)
C
C     PRINTS THE METHOD OF EST. AND THE PARAMETERS ON THE GRAPHS
      DIMENSION NAME(20),ITITLE(20),IPAR(20)
      COMMON SFACX,SFACY,NAME,ITITLE,IPAR
      CALL MOVTO2(X,Y0)
      CALL CHAHOL(3HX*.)
      X0=X0+5.
      CALL MOVTO2(X0,Y)
      CALL CHAHOL(13HF*LREQUENCY*.)
      X0=X0+105.
      Y0=Y0+170.
      CALL MOVTO2(X0,Y0)
      CALL CHAARR(NAME,20,4)
      Y0=Y0-10.
      CALL MOVTO2(X0,Y0)
      CALL CHAARR(ITITLE,20,4)
      Y0=Y0-10.
      CALL MOVTO2(X0,Y0)
      CALL CHAARR(IPAR,20,4)
      RETURN
      END
C
      FUNCTION NGT25(F)
C
C     FINDS THE NO. OF POINTS IN THE ARRAY F WHICH HAVE FREQ.
C     > 25 AND WHICH HAVE TO BE EXCLUDED SO AS NOT TO
C     OVER-SHOOT FROM THE FRAME OF THE GRAPH(PLOT 3)
      DIMENSION F(300)
      NGT25=0
    1 IF(F(NGT25+1) .LT. 25.)GOTO 2
      NGT25=NGT25+1
      GOTO 1
    2 RETURN
      END
C
      SUBROUTINE HIS(C,D)
C
C     DRAWS ONE COLUMN OF THE HISTOGRAM.
      COMMON SFACX,SFACY
      C=C*SFACX
      D=D*SFACY
      CALL LINBY2(C,0.)
      CALL LINBY2(0.,D)
      RETURN
      END
      FINISH
```

```
      LIBRARY(SUBGROUPNAGF)
      LIST(LP)
      PROGRAM(P21)
      INPUT 5=CR0
      OUTPUT 6=LP0
      TRACE 1
      COMPACT
      COMPRESS INTEGER AND LOGICAL
      END

      MASTER PREDICTION 3-PARAMETER INV. GAUSSIAN
C
C                   *** PROGRAM P21 ***
C
C     THIS PROGRAM FINDS THE DIS. OF CLAIM AMOUNTS IN A FUTURE
C     PERIOD ACCORDING TO 3-PAR INVERSE GAUSSIAN MODEL
C
      DIMENSION F(100),USDQTR(8),INDEX(20)
      COMMON ADJC,ADJMEW,ADJLAM,S,NCLM
      REAL MEW,LAMDA
C     READ RESPECTIVELY THE DES. OF PERIOD BEING PREDICTED,
C     THE DES. OF PERIOD USED FOR PREDICTION,PARAMETERS OF MODEL
C     FOR PERIOD  USED ,RATE OF INFLATION,DES. OF INDEX USED,
C     NO. OF CLAIMS IN PERIOD BEING PREDICTED, NO. OF INTERVALS,
C     THE ACTUAL PARAMETERS OF THE MODEL  FOR THE PERIOD
C     BEING PREDICTED.
      READ(5,100)PRDQTR,USDQTR,C,MEW,LAMDA,RATINF,INDEX,NCLM,M,
     $AC,AMEW,ALAMDA
  100 FORMAT(A8/8A8/3F0.0/F0.0/20A4/I0/I0/3F0.0)
      R=1.+RATINF/100.
C     FIND THE PARAMETERS AND MEAN AND S.D.  OF PREDICTION MODEL
C     BY ADJUSTING FOR INFLATION
      ADJC=R*C
      ADJMEW=R*MEW
      ADJLAM=R*LAMDA
      PMEAN=-ADJC+ADJMEW
      PSD=ADJMEW*SQRT(ADJMEW/ADJLAM)
      AMEAN=-AC+AMEW
      ASD=AMEW*SQRT(AMEW/ALAMDA)
      WRITE(6,900)PRDQTR,USDQTR,C,MEW,LAMDA,RATINF,INDEX,
     $ADJC,ADJMEW,ADJLAM,PMEAN,PSD,PRDQTR,AC,AMEW,ALAMDA,AMEAN,ASD
  900 FORMAT(/////////////////////////20X,
     $'*** 3-PARAMETER INV. GAUSSIAN DIS. ***'//8X,
     $'PREDICTION OF ',A8,'  QUARTER CLAIMS COST '/3X,'USING ',8A8//
     $12X,'C=',F5.1,6X,'MEW=',F8.3,6X,'LAMDA=',F8.3/12X,
     $'INFLATION RATE I=',F4.1,'% CALCULATED  FROM  :'/12X,
     $20A4//8X,'PREDICTION PARAMETERS ARE  :-  C=',F4.1,2X,'MEW=',
     $F7.3,2X,'LAMDA=',F7.3/25X,'MEAN CLAIM AMOUNT=',F7.2,7X,'S.D.='
     $F6.2/8X,'ACTUAL',A8,'PARAMETERS   :-  C=',F4.1,2X,'MEW=',F7.3,
2X,
     $'LAMDA=',F7.3/25X,'MEAN CLAIM AMOUNT=',F7.2,7X,'S.D.=',
     $F6.2//8X,'  AMOUNT £  ACT.NO.  EXP.NO.   ',
     $'A-E   EXP.LOSS  (A-E)**2/E')
C     READ IN THE ACTUAL FREQ. DIS. FOR THE PREDICTED PERIOD.
      READ(5,101)(F(I),I=1,M)
  101 FORMAT(F0.0)
      S=0.
      SELOSS=0.
      ISNEXP=0.
      SS=0.
      SACOST=0.
      I=0
   10 I=I+1
      IF(I-20)1,1,2
    1 XEND=30.*I+0.5
      IUB=XEND
      LB=IUB-29
      XMID=XEND-15.
```

```
      LIBRARY(SUBGROUPGRAF)
      LIBRARY(SUBGROUPGINO)
      LIBRARY(SUBGROUPNAGF)
      LIST(LP)
      PROGRAM(P22)
      COMPACT
      INPUT 1,5=CR0
      OUTPUT 2,6=LP0
      TRACE 1
      COMPRESS INTEGER AND LOGICAL
      END

      MASTER PARETO STRAIGHT LINE GRAPHICAL TEST
C
C                  *** PROGRAM P22 ***
C
C     THIS PROGRAM PLOTS SAMPLE POINTS ACCORDING TO PARETO
C     DIS. FOR THE PURPOSE OF GRAPHICAL TESTING.
C
      DIMENSION NAME(20), X(60),XLOG(60),Y(60),F(60),CSUM(60)
C     READ IN DESCRIPTION OF DATA AND NO. OF INTERVALS
      READ(5,101)NAME,M
  101 FORMAT(20A4/I0)
      WRITE(6,900)NAME
  900 FORMAT(////' *********',20A4)
C     READ IN DATA
      DO  3 I=1,M
      READ(5,102)Y(I)
  102 FORMAT(F0.0)
      IF(I-20)1,1,2
    1 X(I)=30.*I+0.5
      GOTO 3
    2 X(I)=(I-14)*100.+0.5
    3 CONTINUE
      CSUM(1)=Y(1)
      DO 4 I=2,M
    4 CSUM(I)=CSUM(I-1)+Y(I)
      S=CSUM(M)
      K=0
      DO 5 I=1,M-1
      N2=CSUM(I+1)
      N1=CSUM(I)
      IF(N2 .EQ. N1)GOTO 5
      K=K+1
      X(K)=X(I)
      F(K)=CSUM(I)/S
    5 CONTINUE
      DO 6 I=1,K
      T=1.-F(I)
      F(I)=-ALOG(T)
      XLOG(I)=ALOG(X(I))
    6 CONTINUE
C     CALL GINO PLOTTING ROUTINES TO PLOT THE GRAPH
      CALL M19346
      CALL DEVPAP(700.,700.,0)
      CALL CHASWI(1)
      CALL CHASIZ(3.0,3.0)
      CALL MOVTO2(30.,30.)
      CALL LINBY2(210.,0.)
      CALL LINBY2(0.,297.)
      CALL LINBY2(-210.,0.)
      CALL LINBY2(0.,-297.)
      CALL AXIPOS(1,50.,50.,180.,1)
```

```
      GOTO 3
    2 XEND=(I-14)*100.+0.5
      IUB=XEND
      LB=IUB-99
      XMID=XEND-50.
    3 P=EXPNO(XEND)
      NEXP=P+0.5
      IF(P-0.5)4,4,5
    4 IF(I-M)6,6,7
    6 P=0.0
      GOTO 11
    7 GOTO 20
    5 IF(I-M)11,11,9
    9 F(I)=0.0
   11 AMINE=F(I)-NEXP
      SACOST=SACOST+XMID*F(I)
      ELOSS=XMID*AMINE
      SELOSS=SELOSS+ELOSS
      IF(NEXP .LT. 5)GOTO 12
      Z=AMINE*AMINE/NEXP
      SS=SS+Z
      WRITE(6,901)LB,IUB,F(I),P,AMINE,ELOSS,Z
  901 FORMAT(8X,I5,'-',I4,2X,F6.0,3X,F6.0,3X,F5.0,2X,F7.0,4X,F7.3)
      GOTO 13
   12 WRITE(6,902)LB,IUB,F(I),P,AMINE,ELOSS
  902 FORMAT(8X,I5,'-',I4,2X,F6.0,3X,F6.0,3X,F5.0,2X,F7.0)
   13 ISNEXP=ISNEXP+NEXP
      GOTO 10
   20 WRITE(6,903)NCLM,ISNEXP,SELOSS,SS
  903 FORMAT(8X,'--------------------------------------------',
     $'------------'/8X,'   TOTAL     ',I5,4X,I5,11X,F7.0,4X,F7.3/
     $8X,'--------------------------------------------',
     $'-----------------')
      SECOST=SACOST-SELOSS
      Z=100.*SELOSS/SACOST
      WRITE(6,904)SACOST,SECOST,Z
  904 FORMAT(/10X,'CHI SQ. STAT. =         , D.F.= '//10X,
     $'TOTAL ACTUAL COST     =',F8.0/10X,'TOTAL EXPECTED COST =',F8.0/
/
     $10X,'TOTAL EXP. LOSS'/10X,'----------------',F6.2,' %'/10X,
     $'TOTAL ACT. COST'//////////////////)
      STOP
      END
C
      FUNCTION EXPNO(X)
C
C     GIVES THE EXPECTED NO. OF CLAIMS IN THE INTERVAL WITH
C     UPPER BOUNDARY X.
      COMMON ADJC,ADJMEW,ADJLAM,S,NCLM
      R1=SQRT(ADJLAM/(X+ADJC))
      R2=(X+ADJC)/ADJMEW
      R3=EXP(2.*ADJLAM/ADJMEW)
      V1=R1*(R2-1.)
      V2=-R1*(R2+1.)
C     S15ABF IS NAG ROUTINE FOR STANDARD NORMAL CUM. DIS. FUN.
C     SINVG GIVES THE INV. GAUSSIAN DIS. FUNCTION (3 PAR. CASE)
      SINVG=S15ABF(V1,0)+R3*S15ABF(V2,0)
      Q=NCLM*SINVG
      EXPNO=Q-S
      S=Q
      RETURN
      END
      FINISH
```

```
      CALL AXIPOS(0,50.,50.,270.,2)
      CALL AXISCA(1,7,2.,9.,1)
      CALL AXISCA(1,10,0.,10.,2)
      CALL AXIDRA(2,1,1)
      CALL AXIDRA(-2,-1,2)
      CALL MOVTO2(208.,40.)
      CALL CHAHOL(13H *ULOG*L(X)*.)
      CALL MOVTO2(55.,320.)
      CALL CHAHOL(21H*U-LOG(1.-F(*LX*U))*.)
      CALL MOVTO2(75.,287.)
      CALL CHAHOL(45HS*LTHAIGHT LINE PLOT OF *UPARETO DIS. *LFOR*.)
      CALL MOVTO2(75.,280.)
      CALL CHAARR(NAME,20,4)
      CALL CRASYM(XLOG,F,K,3,0)
      CALL DEVEND
      STOP
      END
      FINISH
```

```
      LIBRARY(SUBGROUPNAGF)
      LIST(LP)
      PROGRAM(P23)
      COMPACT
      INPUT 1,5=CR0
      OUTPUT 2,6=LP0
      TRACE 2,100
      COMPRESS INTEGER AND LOGICAL
      END

      MASTER ESTIMATION FOR TRUNCATED 2-PAR LOGNORMAL BY MML METHOD
C
C                  *** PROGRAM P23 ***
C
C     THIS PROGRAM ESTIMATES THE PARAMETERS OF THE TRUNCATED
C     2-PAR LOGNORMAL DIS. BY THE MML METHOD.
C
      DIMENSION X(100),XLOG(100),Y(100),AME2(100),CSUM(100),W(40)
      DIMENSION XX(100),FEXP(100),NAME(20),P(3),XTOL(5),YY(100)
      DIMENSION ELOSS(100),XMID(100)
      COMMON K,ITP,XT,XX,YY
      EXTERNAL FUNCT,MONIT
      REAL MEW
C     READ DESC. OF DATA, (M) NO. OF INTERVALS AND (ITP) THE
C     INTERVAL NO. WHOSE UPPER BOUNDARY IS TO BE TAKEN AS
C     TRUNCATION POINT.
      READ(5,101)NAME,M,ITP
  101 FORMAT(20A4/2I0)
      WRITE(6,900)NAME
  900 FORMAT(////' *********',20A4)
C     READ THE STARTING VALUES OF THE PARAMETERS.
      READ(5,102)(P(I),I=1,2)
  102 FORMAT(2F0.0)
C     READ IN THE DATA
      DO  3 I=1,M
      READ(5,103)Y(I)
  103 FORMAT(F0.0)
      IF(I-20)1,1,2
    1 X(I)=30.*I+0.5
      XMID(I)=X(I)-15.0
      GOTO 3
    2 X(I)=(I-14)*100.+0.5
      XMID(I)=X(I)-50.0
    3 CONTINUE
      S=0.0
      DO 4 I=ITP+1,M
      S=S+Y(I)
    4 CONTINUE
C     FIND K THE NO. OF INTERVALS WITH NON-ZERO FREQUENCY
      K=0
C     XT IS THE TRUNCATION POINT
      XT=X(ITP)
      DO 5 I=ITP+1,M
      NI=Y(I)
      IF(NI .EQ. 0)GOTO 5
      K=K+1
      XX(K)=X(I)
      YY(K)=Y(I)/S
    5 CONTINUE
C     SET THE PARAMETERS OF NAG ROUTINE E04CAF
      XTOL(1)=.01
      XTOL(2)=.01
      STEPMX=10.0
      ISTOP=2
      IW=40
      IFAIL=1
      MAXCAL=200
```

```
      IPRINT=4
C     E04CAF IS THE NAG ROUTINE FOR MINIMIZING(OR MAX.) A GENERAL
C     FUNCTION OF SEVERAL VARIABLES.
      CALL E04CAF(2,P,F,XTOL,STEPMX,ISTOP,W,IW,FUNCT,MONIT,
     $IPRINT,MAXCAL,IFAIL)
      IF(IFAIL .NE. 0)WRITE(6,910)IFAIL
  910 FORMAT('  IFAIL=',I2///)
C     THE MML ESTIMATES ARE:-
      SIGMA=P(2)
      SIGMA2=P(2)*P(2)
      MEW=P(1)
C     FIND THE MEAN AND S.D. OF THE TRUNCATED MODEL
      ALPHA=EXP(MEW+.5*SIGMA2)
      E=XT
      Z=ALOG(E)
      V=(Z-MEW)/SIGMA
      D=S15ABF(V,0)
      CD=1.-D
      V=(Z-MEW-SIGMA2)/SIGMA
      D1=S15ABF(V,0)
      V=(Z-MEW-2.*SIGMA2)/SIGMA
      D2=S15ABF(V,0)
      XMEAN=ALPHA*(1.-D1)/CD
      SDX=ALPHA*SQRT(EXP(SIGMA2)*(1.-D2))*CD/(1.-D1)**2
      WRITE(6,915)NAME,MEW,SIGMA2,XMEAN,SDX
  915 FORMAT(////////////////////////////////
     $/////////15X,'*** TRUNCATED 2-PARAMETER LOGNORMAL DIS. ***'
     $//27X,20A4//12X,'ESTIMATION BY MULTINOMIAL MAX. ',
     $'LIKELIHOOD METHOD :-',/16X,'MEW=',
     $F6.2,20X,'SIGMA SQ.=',F6.3//16X,'MEAN=',F8.3,17X,'S.D.=',F8.3)
      WRITE(6,920)
  920 FORMAT(///6X,'AMOUNT &   ACTUAL     EXPECTED   ACTUAL-',
     $'    EXPECTED',/17X,'CL. NO.    CL. NO.   EXPECTED',
     $'       LOSS      (A-E)**2/E ')
C     COMPUTE THE TABLE OF RESULTS
      SEXP=0.0
      SS=0.
      G1=0
      ACOST=0.
      ELOSS(M+1)=0.
      DO 7 I=ITP+1,M
      V=(ALOG(Y(I))-MEW)/SIGMA
      G2=S15ABF(V,0)
      FEXP(I)=(G2-G1)*S/CD
      NEXP=FEXP(I)+.5
      SEXP=SEXP+NEXP
      CSUM(I)=Y(I)-NEXP
      ELOSS(I)=XMID(I)*CSUM(I)
      ELOSS(M+1)=ELOSS(M+1)+ELOSS(I)
      ACOST=ACOST+XMID(I)*Y(I)
      IF(NEXP .LT. 5)GOTO 20
      AME2(I)=CSUM(I)*CSUM(I)/NEXP
      SS=SS+AME2(I)
   20 G1=G2
    7 CONTINUE
      DO 11 I=ITP+1,M
      IF(I-20)8,8,9
    8 LB=X(I)-29.
      GOTO 10
    9 LB=X(I)-99.
   10 LU=X(I)
      IF(FEXP(I) .LT. 5)GOTO 21
      WRITE(6,925)LB,LU,Y(I),FEXP(I),CSUM(I),ELOSS(I),AME2(I)
  925 FORMAT(4X,I5,'-',I4,F9.0,F10.0,F11.0,4X,F8.0,3X,F8.3)
      GOTO 11
   21 WRITE(6,930)LB,LU,Y(I),FEXP(I),CSUM(I),ELOSS(I)
  930 FORMAT(4X,I5,'-',I4,F9.0,F10.0,F11.0,4X,F8.0)
   11 CONTINUE
      RATIO=100.*ELOSS(M+1)/ACOST
      WRITE(6,935)S,SEXP,ELOSS(M+1),SS,RATIO
  935 FORMAT('    --------------------------------------------------
'
     $'---------------',/,'      TOTAL    ',F9.0,F10.0,15X,F8.0,
     $3X,F8.3/'     ---------------------------------------------',
     $'----------------------'/57X,'D.F. = '/10X,
     $'TOTAL EXP. LOSS'/10X,'---------------',F5.1,' %'/
     $10X,'TOTAL ACT. COST'////////////////////)
      STOP
      END
C
      SUBROUTINE FUNCT(NP,PC,FC)
C
C     FINDS THE VALUE OF THE LOGLIKELIHOOD FUNC. FOR A SET OF
C     PARAMETERS. THIS IS REQUIRED BY E04CAF ROUTINE.
      DIMENSION PC(NP),XX(100),YY(100)
      COMMON K,ITP,XT,XX,YY
      REAL MEW
      MEW=PC(1)
      SIGMA=PC(2)
      Z=(ALOG(XT)-MEW)/SIGMA
      DT=S15ABF(Z,0)
      S=0.0
      DO 4 I=1,K
      IF(I+ITP-20)1,1,2
    1 Z=(ALOG(XX(I)-30.0)-MEW)/SIGMA
      D1=S15ABF(Z,0)
      GOTO 3
    2 Z=(ALOG(XX(I)-100.0)-MEW)/SIGMA
      D1=S15ABF(Z,0)
    3 Z=(ALOG(XX(I))-MEW)/SIGMA
      D2=S15ABF(Z,0)
      S=S+YY(I)*ALOG(D2-D1)
    4 CONTINUE
      S=S-ALOG(1.-DT)
      FC=-S
      RETURN
      END
C
      SUBROUTINE MONIT(N,PC,FC,NCALL)
C
C     PRINTS THE PARAMETERS AND VALUE OF THE LOGLIKELIHOOD
C     FUNCTION AT VARIOUS ITERATIONS. REQUIRED BY E04CAF.
      DIMENSION PC(N)
      WRITE(6,940)NCALL,FC
  940 FORMAT('   AFTER',I5,'  FUNCTION CALLS THE VALUE IS',3X,F15.12
)
      SIGMA2=PC(2)*PC(2)
      WRITE(6,945)(PC(I),I=1,N),SIGMA2
  945 FORMAT('   AT THE POINT      MEW=',F7.5,'  SIGMA=',
     $F7.5,'   S2=',F7.5)
      RETURN
      END
      FINISH
```

```
      LIBRARY(SUBGROUPNAGF)
      LIST(LP)
      PROGRAM(P24)
      COMPACT
      INPUT 1,5=CR0
      OUTPUT 2,6=LP0
      TRACE 2,100
      COMPRESS INTEGER AND LOGICAL
      END

      MASTER MULTN MAXLIK METHOD FOR PARETO DIS OF THE 1ST KIND
C
C                    *** PROGRAM P24 ***
C
C     THIS PROGRAM FINDS ESTIMATE OF PARAMETER A BY MML
C     METHOD WHEN PARAMETER B IS TAKEN AS THE LOWEST BOUNDARY
C     OF SAMPLE VALUES.
C
      DIMENSION X(100),XLOG(100),Y(100),AME2(100),CSUM(100),W(40)
      DIMENSION XX(100),FEXP(100),NAME(20),P(3),XTOL(5),YY(100)
      DIMENSION ELOSS(100),XMID(100)
      COMMON K,XT,XX,YY
      EXTERNAL FUNCT,MONIT
C     READ IN DESCRIPTION OF DATA, NO. OF INTERVALS AND ITP
C     WHICH THE INTERVAL NO. WHERE SAMPLE IS TO BE TRUNCATED.
      READ(5,101)NAME,M,ITP
  101 FORMAT(20A4/2I0)
      WRITE(6,900)NAME
  900 FORMAT(////' *********',20A4)
C     READ IN THE STARTING VALUE OF A FOR ITERATION PROCESS
      READ(5,102)P(1)
  102 FORMAT(F0.0)
C     READ IN THE DATA
      DO 3 I=1,M
      READ(5,103)Y(I)
  103 FORMAT(F0.0)
      IF(I-20)1,1,2
    1 X(I)=30.*I+0.5
      XMID(I)=X(I)-15.0
      GOTO 3
    2 X(I)=(I-14)*100.+0.5
      XMID(I)=X(I)-50.0
    3 CONTINUE
      S=0.0
      DO 4 I=ITP,M
      S=S+Y(I)
    4 CONTINUE
      K=0
      XT=X(ITP-1)
      DO 5 I=ITP,M
      NI=Y(I)
      IF(NI .EQ. 0)GOTO 5
      K=K+1
      XX(K)=X(I)
      YY(K)=Y(I)/S
    5 CONTINUE
C     SET THE PARAMETERS OF E04CAF ROUTINE FOR MAXIMIZING A
C     GENERAL FUNCTION OF SEVERAL VARIABLES.
      XTOL(1)=0.01
      STEPMX=10.0
      ISTOP=2
      IW=40
      IFAIL=1
      MAXCAL=400
      IPRINT=4
      CALL E04CAF(1,P,F,XTOL,STEPMX,ISTOP,W,IW,FUNCT,MONIT,
     $IPRINT,MAXCAL,IFAIL)
      IF(IFAIL .NE. 0)WRITE(6,910)IFAIL
  910 FORMAT('  IFAIL=',I2///)
      A=P(1)
C     FIND THE MEAN AND S.D. OF THE FITTED MODEL
      E=XT
      XMEAN=A*E/(A-1.)
      SDX=E*SQRT(A*(A-2.))/((A-1.)*(A-2.))
      WRITE(6,915)NAME,A,XMEAN,SDX
  915 FORMAT(//////////////////////////////////////
     $15X,'*** PARETO DIS. OF THE 1ST KIND ***'
     $//23X,20A4//12X,'ESTIMATION BY MULTINOMIAL MAX. ',
     $'LIKELIHOOD METHOD :-',/16X,'A=',
     $F6.3//16X,'MEAN=',F8.3,17X,'S.D.=',F8.3)
      WRITE(6,920)
  920 FORMAT(///6X,'AMOUNT £   ACTUAL    EXPECTED   ACTUAL-',
     $'    EXPECTED',/17X,'CL. NO.    CL. NO.   EXPECTED',
     $'     LOSS      (A-E)**2/E ')
C     FIND E, A-E, EXP. LOSS, (A-E)**2/E
      SEXP=0.0
      SS=0.
      Q1=0.0
      ACOST=0.
      ELOSS(M+1)=0.
      DO 7 I=ITP,M
      V=E/X(I)
      Q2=1.0-V**A
      FEXP(I)=(Q2-Q1)*S
      NEXP=FEXP(I)+.5
      SEXP=SEXP+NEXP
      CSUM(I)=Y(I)-NEXP
      ELOSS(I)=XMID(I)*CSUM(I)
      ELOSS(M+1)=ELOSS(M+1)+ELOSS(I)
      ACOST=ACOST+XMID(I)*Y(I)
      IF(NEXP .LT. 5)GOTO 20
      AME2(I)=CSUM(I)*CSUM(I)/NEXP
      SS=SS+AME2(I)
   20 Q1=Q2
    7 CONTINUE
      DO 11 I=ITP,M
      IF(I-20)8,8,9
    8 LB=X(I)-29.
      GOTO 10
    9 LB=X(I)-99.
   10 LU=X(I)
      IF(FEXP(I) .LT. 5)GOTO 21
      WRITE(6,925)LB,LU,Y(I),FEXP(I),CSUM(I),ELOSS(I),AME2(I)
  925 FORMAT(4X,I5,'-',I4,F9.0,F10.0,F11.0,4X,F8.0,3X,F8.3)
      GOTO 11
   21 WRITE(6,930)LB,LU,Y(I),FEXP(I),CSUM(I),ELOSS(I)
  930 FORMAT(4X,I5,'-',I4,F9.0,F10.0,F11.0,4X,F8.0)
   11 CONTINUE
      RATIO=100.*ELOSS(M+1)/ACOST
      WRITE(6,935)S,SEXP,ELOSS(M+1),SS,RATIO
  935 FORMAT('     -----------------------------------------------
     $'--------------',/,'      TOTAL    ',F9.0,F10.0,15X,F8.0,
     $3X,F8.3/'     ----------------------------------------',
     $'--------------------'/57X,'D.F. = '/10X,
     $'TOTAL EXP. LOSS'/10X,'----------------',F5.1,' %'/
     $10X,'TOTAL ACT. COST'/////////////////////////)
  777 STOP
      END
C
      SUBROUTINE FUNCT(NP,PC,FC)
C
C     FINDS THE VALUE OF THE FUNCTION FOR A GIVEN VALUE A
C     THIS IS REQUIRED BY E04CAF ROUTINE.
```

```
      DIMENSION PC(NP),XX(100),YY(100)
      COMMON K,XT,XX,YY
      A=PC(1)
      E=XT
      Z=E/XX(1)
      DT=1.-Z**A
      S=YY(1)*ALOG(DT)
      DO 1 I=2,K
      Z=E/(XX(I)-100.)
      D1=1.-Z**A
      Z=E/XX(I)
      D2=1.-Z**A
      S=S+YY(I)*ALOG(D2-D1)
    1 CONTINUE
      FC=-S
      RETURN
      END
C
      SUBROUTINE MONIT(N,PC,FC,NCALL)
C
C     WRITES THE VALUE OF THE FUNCTION AND THE CORRESP. PARAMETER
C     AT VARIOUS STAGES IN THE ITERATION PROCESS. THIS IS REQUIRED
C     BY E04CAF ROUTINE.
      DIMENSION PC(N)
      WRITE(6,940)NCALL,FC
  940 FORMAT('    AFTER',I5,'  FUNCTION CALLS THE VALUE IS',3X,F16.13
     )
      WRITE(6,945)PC(1)
  945 FORMAT('    AT THE POINT      A=',F10.5)
      RETURN
      END
      FINISH
```

```
      LIBRARY(SUBGROUPNAGF)
      LIST(LP)
      PROGRAM(P25)
      COMPACT
      INPUT 1,5=CR0
      OUTPUT 2,6=LP0
      TRACE 0
      COMPRESS INTEGER AND LOGICAL
      END

      MASTER ESTIMATION FOR 2-PAR GAMMA BY MML METHOD
C
C                  *** PROGRAM P25 ***
C
C     THIS PROGRAM FINDS ESTIMATES OF THE PARAMETERS OF THE
C     2-PAR GAMMA DIS. BY MML METHOD.
C
      DIMENSION X(100),XLOG(100),Y(100),AME2(100),CSUM(100),W(40)
      DIMENSION YY(100),XX(100),XMID(100),FEXP(100),P(2),ELOSS(100)
      DIMENSION NAME(20),XTOL(2)
      COMMON K,XX,YY
      EXTERNAL FUNCT,MONIT
C     READ DESCRIPTION OF THE DATA AND NO. OF INTERVALS
      READ(5,101)NAME,M
  101 FORMAT(20A4/I0)
      WRITE(6,900)NAME
  900 FORMAT(////' *********',20A4////////)
C     READ STARTING VALUES FOR THE PARAMETERS
      READ(5,102)(P(I),I=1,2)
  102 FORMAT(2F0.0)
      READ(5,103)(Y(I),I=1,M)
  103 FORMAT(10F0.0)
      S=0.0
C     READ IN THE DATA AND FIND SAMPLE CUM. DIS. FUNCTION
C     SAMPLE FREQ. FUN. ACCORDING TO AD SQ. ROOT DATA IS SET UP
      DO 3 I=1,M
      S=S+Y(I)
      X(I)=2.*I+.005
      XMID(I)=X(I)-1.0
    3 CONTINUE
C     FIND K THE NO. OF NON-ZERO INTERVALS
      K=0
      DO 5 I=1,M
      NI=Y(I)
      IF(NI .EQ. 0)GOTO 5
      K=K+1
      XX(K)=X(I)
      YY(K)=Y(I)/S
    5 CONTINUE
C     SET THE PARAMETERS OF THE E04CAF ROUTINE
      XTOL(1)=.01
      XTOL(2)=.01
      STEPMX=30.0
      ISTOP=2
      IW=40
      IFAIL=0
      MAXCAL=200
      IPRINT=4
C     E04CAF IS NAG ROUTINE FOR MINIMIZING ( OR MAX. ) A
C     GENERAL FUNCTION OF SEVERAL VARIABLES.
      CALL E04CAF(2,P,F,XTOL,STEPMX,ISTOP,W,IW,FUNCT,MONIT,
     $IPRINT,MAXCAL,IFAIL)
      IF(IFAIL .NE. 0)WRITE(6,911)IFAIL
  911 FORMAT('  IFAIL=',I2///)
C     MML ESTIMATES OF THE PARAMETERS ARE :-
      ALPHA=P(1)
      BETA=P(2)
```

```
C     FIND MEAN CLAIM AMOUNT AND S.D. ACCORDING TO GAMMA MODEL
C     FOR SQ. ROOT  OF CLAIM AMOUNTS.
      C=ALPHA*BETA*BETA
      XMEAN=C*(1.+ALPHA)
      SDX=C*SQRT(4.*ALPHA+10.+6./ALPHA)
      WRITE(6,902)NAME,ALPHA,BETA,XMEAN,SDX
  902 FORMAT(////20X,'*** TWO-PARAMETER  GAMMA  DIS. ***'/
     $/27X,20A4//12X,'ESTIMATION BY MULTINOMIAL MAX.',
     $'LIKELIHOOD METHOD :-',
     $/16X,'ALPHA =',F6.3,'              BETA=',F6.3/,
     $/16X,'MEAN=',F8.3,'              S.D.=',F8.3)
      WRITE(6,903)
  903 FORMAT(///6X,'AMOUNT £   ACTUAL    EXPECTED   ACTUAL-',
     $'    EXPECTED',/17X,'CL. NO.    CL. NO.   EXPECTED',
     $'     LOSS      (A-E)**2/E ')
C     COMPUTE THE TABLE OF RESULTS
      SEXP=0.0
      SS=0.
      ACOST=0.0
      ELOSS(M+1)=0.0
      D1=0.0
      DO 7 I=1,M
      D2=SGAMMA(X(I),ALPHA,BETA,0.)
      FEXP(I)=(D2-D1)*S
      NEXP=FEXP(I)+.5
      SEXP=SEXP+NEXP
      CSUM(I)=Y(I)-NEXP
      C=XMID(I)**2
      ACOST=ACOST+C*Y(I)
      ELOSS(I)=C*CSUM(I)
      ELOSS(M+1)=ELOSS(M+1)+ELOSS(I)
      IF(NEXP .LT. 5)GOTO 20
      AVE2(I)=CSUM(I)*CSUM(I)/NEXP
      SS=SS+AVE2(I)
   20 D1=D2
    7 CONTINUE
      LU=X(1)**2
      WRITE(6,910)LU,Y(1),FEXP(1),CSUM(1),ELOSS(1),AVE2(1)
  910 FORMAT(6X,'UPTO',I4,F9.0,F10.0,F11.0,4X,F8.0,3X,F8.3)
      DO 11 I=2,M
      LB=LU
      LU=X(I)**2
      IF(FEXP(I) .LT. 5)GOTO 21
      WRITE(6,904)LB,LU,Y(I),FEXP(I),CSUM(I),ELOSS(I),AVE2(I)
  904 FORMAT(4X,I5,'-',I4,F9.0,F10.0,F11.0,4X,F8.0,3X,F8.3)
      GOTO 11
   21 WRITE(6,905)LB,LU,Y(I),FEXP(I),CSUM(I),ELOSS(I)
  905 FORMAT(4X,I5,'-',I4,F9.0,F10.0,F11.0,4X,F8.0)
   11 CONTINUE
      RATIO=100.*ELOSS(M+1)/ACOST
      WRITE(6,906)S,SEXP,ELOSS(M+1),SS,RATIO
  906 FORMAT('     ------------------------------------------',
     $'----------------------'/,'        TOTAL    ',F9.0,1X,F9.0,15X,
     $F8.0,3X,F8.3/'     --------------------------------------------',
     $'-----------------------------'/57X,'D.F. = ',/10X,
     $'TOTAL EXP. LOSS'/10X,'-----------------',F5.1,' %',/
     $10X,'TOTAL ACT. COST'///////////////////////////////)
      STOP
      END
C
      SUBROUTINE FUNCT(NP,PC,FC)
C
C     FINDS THE VALUE OF THE LOGLIKELIHOOD FUNCTION FOR A GIVEN
C     SET OF PARAMETERS. THIS IS REQUIRED BY E04CAF ROUTINE.
      DIMENSION PC(NP),XX(100),YY(100)
      COMMON K,XX,YY
      ALPHA=PC(1)
      BETA=PC(2)
      D1=SGAMMA(XX(1),ALPHA,BETA,0.)
      S=YY(1)*ALOG(D1)
      DO 1 I=2,K
      D2=SGAMMA(XX(I),ALPHA,BETA,0.)
      S=S+YY(I)*ALOG(D2-D1)
      D1=D2
    1 CONTINUE
      FC=-S
      RETURN
      END
C
      SUBROUTINE MONIT(N,PC,FC,NCALL)
C
C     PRINTS THE VALUES OF LOGLIKELIHOOD FUN. AND THE PARAMETERS
C     AT DIFF. ITERATIONS.  THIS IS REQUIRED BY E04CAF ROUTINE.
      DIMENSION PC(N)
      WRITE(6,940)NCALL,FC
  940 FORMAT('    AFTER',I5,'  FUNCTION CALLS THE VALUE IS',3X,F10.6)
      WRITE(6,945)(PC(I),I=1,N)
  945 FORMAT('     AT THE POINT  ALPHA=',F6.3,'    BETA=',F6.3)
      RETURN
      END
C
      FUNCTION SGAMMA(X,A,B,G)
C
C     GIVES THE 3-PAR GAMMA CUM. DIS. FUNCTION AT POINT X.
C     A,B,G ARE PARAMETERS OF DIS.   FOR 2-PAR CASE G=0
      Y=(X-G)/B
      C=EXP(-Y)
      R=1./(100000.*C)
      S=0.0
      J=0
C     S14AAF IS NAG ROUTINE WHICH GIVES VALUES OF THE
C     COMPLETE GAMMA FUNCTION.
    1 T=Y**(A+J)/S14AAF(A+J+1,0)
      S=S+T
      IF(T .LT. R)GOTO 2
      J=J+1
      GOTO 1
    2 SGAMMA=C*S
      RETURN
      END
      FINISH
```

```
      LIST(LP)
      LIBRARY(SUBGROUPNAGF)
      PROGRAM(P26)
      INPUT 5=CR0
      OUTPUT 6=LP0
      TRACE 2,100
      COMPACT
      COMPRESS INTEGER AND LOGICAL
      END

      MASTER PREDICTION TWO-PARAMETER GAMMA
C
C                        *** PROGRAM P26 ***
C
C     THIS PROGRAM FINDS THE DIS. OF CLAIM AMOUNTS IN A FUTURE
C     PERIOD ACCORDING TO 2-PAR GAMMA MODEL FOR THE SQ.
C     ROOT OF THE CLAIM AMOUNT.
C
      DIMENSION F(100),USDQTR(8),INDEX(20)
      COMMON ALPHA,ADJBTA,S,NCLM
C     READ RESPECTIVELY THE DES. OF PERIOD BEING PREDICTED,
C     THE DES. OF PERIOD USED FOR PREDICTION ,PARAMETERS OF MODEL
C     FOR PERIOD USED, RATE OF INFLATION, DES. OF INDEX USED,
C     NO. OF CLAIMS IN PREDICTED PERIOD,NO. OF INTERVALS,
C     THE ACTUAL PARAMETERS OF THE MODEL FOR PREDICTED PERIOD.
      READ(5,100)PRDQTR,USDQTR,ALPHA,BETA,RATINF,INDEX,NCLM,M,
     $AALPHA,ABETA
  100 FORMAT(A8/8A8/2F0.0/F0.0/20A4/I0/I0/2F0.0)
C     FIND THE PARAMETERS OF PREDICTED MODEL BY ADJUSTING FOR
C     INFLATION. ALSO FIND MEAN CLAIM AMOUNT AND S.D.  ACCORDING
C     TO THE GAMMA MODEL FOR SQ. ROOT OF CLAIM AMOUNT.
      ADJBTA=BETA*SQRT(1.+RATINF/100.)
      CONS=ALPHA*ADJBTA*ADJBTA
      PMEAN=CONS*(1.+ALPHA)
      PSD=CONS*SQRT(4.*ALPHA+10.+6./ALPHA)
      ACONS=AALPHA*ABETA*ABETA
      AMEAN=ACONS*(1.+AALPHA)
      ASD=ACONS*SQRT(4.*AALPHA+10.+6./AALPHA)
      WRITE(6,899)PRDQTR,USDQTR,ALPHA,BETA,RATINF,INDEX,
     $ALPHA,ADJBTA,PMEAN,PSD,PRDQTR,AALPHA,ABETA,AMEAN,ASD
  899 FORMAT(//////////////////////////////20X,
     $'*** TWO-PARAMETER GAMMA DIS. ***'//8X,
     $'PREDICTION OF ',A8,' QUARTER CLAIMS COST '/8X,'USING ',8A8//
     $12X,'ALPHA=',F6.3,16X,'BETA=',F6.3/12X,'INFLATION RATE I=',
     $F4.1,'% CALCULATED  FROM :'/12X,
     $20A4//8X,'PREDICTION PARAMETERS ARE  :-  ALPHA=',F6.3,6X,
     $'BETA=',F6.3/25X,'MEAN CLAIM AMOUNT=',F7.2,7X,'S.D.=',
     $F6.2/8X,'ACTUAL',A8,'PARAMETERS   :-  ALPHA=',F6.3,6X,
     $'BETA=',F6.3/25X,'MEAN CLAIM AMOUNT=',F7.2,7X,'S.D.=',
     $F6.2//8X,'  AMOUNT £  ACT.NO.  EXP.NO.   ',
     $'A-E   EXP.LOSS  (A-E)**2/E')
C     READ IN THE ACTUAL FREQ. DIS. FOR PREDICTED PERIOD
      READ(5,101)(F(I),I=1,M)
  101 FORMAT(10F0.0)
      S=0.
      SELOSS=0.
      ISNEXP=0.
      SS=0.
      SACOST=0.
      I=0
   10 I=I+1
      XEND=2.*I+0.005
      LB=LU
      LU=XEND**2
      XMID=XEND-1.0
    1 P=EXPNO(XEND)
      NEXP=P+0.5
      IF(P-4.5)2,2,3
    2 IF(I-M)4,4,5
    4 P=0.0
      GOTO 9
    5 GOTO 20
    3 IF(I-M)9,9,7
    7 F(I)=0.0
    9 AMINE=F(I)-NEXP
      X2=XMID**2
      SACOST=SACOST+X2*F(I)
      ELOSS=X2*AMINE
      SELOSS=SELOSS+ELOSS
      IF(NEXP .LT. 5)GOTO 12
      Z=AMINE*AMINE/NEXP
      SS=SS+Z
      IF(I .GT. 1)GOTO 11
      WRITE(6,900)LU,F(1),P,AMINE,ELOSS,Z
  900 FORMAT(10X,'UPTO',I4,2X,F6.0,3X,F6.0,3X,F5.0,2X,F7.0,4X,F7.3)
      GOTO 13
   11 WRITE(6,901)LB,LU,F(I),P,AMINE,ELOSS,Z
  901 FORMAT(8X,I5,'-',I4,2X,F6.0,3X,F6.0,3X,F5.0,2X,F7.0,4X,F7.3)
      GOTO 13
   12 WRITE(6,902)LB,LU,F(I),P,AMINE,ELOSS
  902 FORMAT(8X,I5,'-',I4,2X,F6.0,3X,F6.0,3X,F5.0,2X,F7.0)
   13 ISNEXP=ISNEXP+NEXP
      GOTO 10
   20 WRITE(6,903)NCLM,ISNEXP,SELOSS,SS
  903 FORMAT(8X,'-----------------------------------------------',
     $'-----------'/8X,'   TOTAL     ',I5,4X,I5,11X,F7.0,4X,F7.3/
     $8X,'-----------------------------------------',
     $'-----------------')
      SECOST=SACOST-SELOSS
      Z=100.*SELOSS/SACOST
      WRITE(6,904)SACOST,SECOST,Z
  904 FORMAT(//10X,'CHI SQ. STAT. =         , D.F.= '//10X,
     $'TOTAL ACTUAL COST    =',F8.0/10X,'TOTAL EXPECTED COST =',F8.0/
/
     $10X,'TOTAL EXP. LOSS'/10X,'----------------',F6.2,' %'/10X,
     $'TOTAL ACT. COST'///////////////////)
      STOP
      END
C
      FUNCTION EXPNO(X)
C
C     GIVES EXPECTED NO. OF CLAIMS IN THE INTERVAL WITH
C     UPPER BOUNDARY X.
      COMMON ALPHA,ADJBTA,S,NCLM
      Q=NCLM*SGAMMA(X,ALPHA,ADJBTA,0)
      EXPNO=Q-S
      S=Q
      RETURN
      END
C
      FUNCTION SGAMMA(X,A,B,G)
C
C     GIVES THE 3-PAR GAMMA CUM. DIS. FUNCTION AT POINT X.
C     A,B,G ARE THE PARAMETERS OF DIS.  FOR 2-PAR CASE G=0
      Y=(X-G)/B
      C=EXP(-Y)
      R=1./(1000000.*C)
      S=0.0
      J=0
C     S14AAF IS  NAG ROUTINE WHICH GIVES VALUES  OF THE
C     COMPLETE GAMMA FUNCTION.
    1 T=Y**(A+J)/S14AAF(A+J+1,0)
      S=S+T
      IF(T .LT. R)GOTO 2
      J=J+1
      GOTO 1
    2 SGAMMA=C*S
      RETURN
      END
      FINISH
```

REFERENCES

ABRAMOWITZ, M. and STEGUN, I.A. (1968)
" Handbook of Mathematical Functions "
Dover Publications.

AITCHISON, J. and BROWN, J.A.C. (1957)
" The Lognormal Distribution"
Cambridge University Press, Cambridge.

ANDERSON, H. (1971)
" An analysis of the development of the fire losses in Northern Countries after the Second World War "
The ASTIN Bulletin, Vol. 6, pp. 25-30.

BARTON, D.E. and DENNIS, K.E.R. (1952)
" The conditions under which Gram-Charlier and Edgeworth curves are positive definite and Unimodal "
Biometrika, Vol. 39, pp. 425-427.

BEARD, R.E. (1955)
" Some statistical aspects of non-life insurance "
Journal of the Institute of Actuaries Students' Society, Vol. 13, part 3.

BEARD, R.E. (1957)
" Analytical expressions of the risks involved in general insurance "
Transactions of the 15th International Congress of Actuaries, New York.

BEARD, R.E. (1964)
" Some statistical problems arising from the transaction of motor insurance business "
Journal of the Institute of Actuaries Students' Society, Vol. 17, part 4.

BEARD, R.E. (1974)
" Some historical, theoretical and practical aspects"
An article in (Claims Provisions for Non-Life Insurance Business); The Institute of Mathematics and its Applications.

BEARD, R.E. (1978)
Personal communication with the author.

BEARD, R.E. (1978)
" Claims inflation"
GIRO Bulletin, Number 17, pp. 7-10.

BENCKERT, L-G. (1962)
" The lognormal model for the distribution of one claim "
The ASTIN Bulletin, Vol. 2, part 1.

BENCKERT, L-G. and JUNG, J. (1974)
" Statistical models of claim distributions in fire insurance "
ASTIN Bulletin, Vol. 8, pp. 1-25.

BENCKERT, L-G. and STERNBERG, I. (1957)
" An attempt to find an expression for the distribution of fire damage amount "
Transactions of the 15th International Congress of Actuaries

BENJAMIN, B. (1977)
" General Insurance "
Heinemann, London.

BENJAMIN, B. and HAYCOCKS, H.W. (1970)
" The Analysis of Mortality and Other Actuarial Statistics "
Cambridge University Press, Cambridge.

BICKERSTAFF, D.R. (1972)
" Automobile collision deductibles and repair cost groups: The lognormal model "
Proceedings of the Casuality Actuarial Society, Vol. 59, pp. 68-102.

BIRNBAUM, Z.W. (1952)
" Numerical tabulation of the distribution of Kolmogorov's statistic for finite sample size "
Journal of the American Statistical Association, Vol. 47, pp. 425-441.

BURY, K.V. (1975)
" Statistical Models in Applied Science "
John Wiley & Sons Inc., New York.

CHHIKARA, R.S. and FOLKS, J.L. (1977)
" The inverse Gaussian distribution as a lifetime model "
Technometrics, Vol. 19, pp. 461-468.

CHHIKARA, R.S. and FOLKS, J.L. (1978)
" The inverse Gaussian distribution and its statistical application - a review "
Journal of the Royal Statistical Society: series B, Vol. 40, pp. 263-289.

COCHRAN, W.G. (1942)
" The χ^2 correction for continuity "
Iowa State College Journal of Science, Vol. 16, pp. 421-436.

COCHRAN, W.G. (1952)
" The χ^2 test of goodness of fit "
Annals of Mathematical Statistics, Vol. 23, pp. 315-345.

CRAMER, H. (1946)
" Mathematical Methods of Statistics "
Princeton University Press, Princeton, New Jersey.

DAVID, F.N. (1947)
" A χ^2 'Smooth' test for goodness of fit "
Biometrika, Vol. 34, pp. 299-310.

DUBEY, S.D. (1963)
" Asymptotic efficiencies of the moment estimators for the parameters of the Weibull laws "
Annals of Mathematical Statistics; Vol. 34, pp. 680-681.

DUBEY, S.D. (1967)
" Some percentile estimators for Weibull parameters "
Technometrics, Vol. 9, pp. 119-129.

FERRARA, G. (1971)
" Distributions des sinistres incendie selon leur cout "
The ASTIN Bulletin, Vol. 6, pp. 31-41.

FINGER, R.J. (1976)
" Estimating pure premiums by layer- an approach "
Proceedings of the Casualty Actuarial Society, Vol. 63, pp. 34-52.

FISHER, R.A. (1924)
" The conditions under which Chi-square measures the discrepancy between observation and hypothesis "
Journal of the Royal Statistical Society, Vol. 87, pp. 442-450.

FOLKS, J. L. (1978)
Personal communication on the inverse Gaussian distribution

FRASER, D.A.S. (1950)
" Note on the χ^2 smooth test "
Biometrika, Vol. 37, pp. 447-448.

GEARY, R.C. and PEARSON, E.S. (1938)
" Tests of Normality "
Biometrika Office, University College London & Cambridge University Press.

GILL, P.E. and MURRAY, W. (1972)
" Quasi-Newton methods for unconstrained optimization "
Journal of the Institute of Mathematics and its Applications, Vol. 9, pp. 91-108.

GJEDDEBAEK, N.F. (1949)
" Constributions to the study of grouped observations. Application of the method of maximum likelihood in case of normally distributed observations "
Skandinavisk Aktuarietidskrift, Vol. 32, p. 135.

GRUNDY, P.M. (1952)
" The fitting of grouped truncated and grouped censored normal distributions "
Biometrika, Vol. 39, pp. 252-259.

GUMBEL, E.J. (1943)
" On the reliability of the classical χ^2 test "
Annals of Mathematical Statistics, Vol. 14, pp. 253-263.

HAAN, C.T. and BEER, C.E. (1967)
" Determination of maximum likelihood estimators for the three parameter Weibull distribution "
Iowa State Journal of Science, Vol. 42, pp. 37-42.

HALD, A. (1949)
" Maximum likelihood estimation of the parameters of a normal distribution which is truncated at a known point ".
Skandinavisk Aktuarietidskrift, Vol. 32; p. 119.

HARDING, V. (1968)
" The calculation of premiums for excess of loss reinsurances of motor business "
Transactions of the 18th International Congress of Actuaries.

HARTER, H.L. (1964)
" New Tables of the Incomplete Gamma Function Ratio and of Percentage Points of the Chi-square and Beta Distributions "
U.S. Government Printing Office, Washington, D.C.

HASOFER, A.M. (1964)
" A dam with inverse Gaussian input "
Proceedings of the Cambridge Philosophical Society, Vol. 60, pp. 931-933.

HASTINGS, C. (1955)
" Approximations For Digital Computers "
Princeton University Press.

HEYDE, C.C. (1963)
" On a property of the lognormal distribution "
Journal of the Royal Statistical Society, Series B, Vol. 25, pp. 392-393.

JOHNSON, N.L. and KOTZ, S. (1970)
" Continuous Univariate Distributions - 1 "
Houghton Mifflin Company, Boston.

KAO, J.H.K. (1956)
" A new life-quality measure for election tubes "
IRE Transactions on Reliability and Quality Control, PGRQC-7, pp. 1-11.

KAO, J.H.K. (1958)
" Computer methods for estimating Weibull parameters in reliability studies "
IRE Transactions on Reliability and Quality Control PGRQC-13, pp. 15-22.

KAO, J.H.K. (1959)
" A graphical estimation of mixed Weibull parameters in life-testing of electron tubes "
Technometrics, Vol. 1, pp. 389-407.

KEMSLEY, W.F.F. (1952)
" Body weight at different ages and heights "
Annals of Eugenics, London, Vol. 16, p. 316.

KENDALL, M.G. and STUART, A. (1973)
" The Advanced Theory of Statistics - Volume II "
Charles Griffin & Company Limited, London.

LANCASTER, A. (1972)
" A stochastic model for the duration of a strike "
Journal of the Royal Statistical Society,
Series A, Vol. 135, pp. 257-271.

LAURENT, A.G. (1963)
" The lognormal distribution and the translation method: description and estimation problems "
Journal of the American Statistical Association,
Vol. 58, pp. 231-235.

LILLIEFORS, H.W. (1967)
" On the Kolmogorov-Smirnov test for normality with mean and variance unknown "
Journal of the American Statistical Association,
Vol. 62, pp. 399-402.

LILLIEFORS, H.W. (1969)
" On the Kolmogorov-Smirnov test for the exponential distribution with mean unknown "
Journal of the American Statistical Association,
Vol. 64, pp. 387-389.

MANN, H.B. and WALD, A. (1942)
" On the choice of the number of class intervals in the application of the Chi square test ".
Annals of Mathematical Statistics, Vol. 13,
pp. 306-317.

MANN, N.R., SCHAFER, R.E. and SINGPURWALLA, N.D. (1974)
" Methods for Statistical Analysis of Reliability and Life Data "
John Wiley, New York.

MASSEY, F.J., Jr. (1950-a)
" A note on the stimation of a cumulative distribution function by confidence intervals "
Annals of Mathematical Statistics, Vol. 21,
pp. 116-119.

MASSEY, F.J., Jr. (1950-b)
" A note on the power of a non-parametric test "
Annals of Mathematical Statistics, Vol. 21,
pp. 440-443.

MASSEY, F.J., Jr. (1951)
" The Kolmogorov-Smirnov test for goodness of fit "
Journal of the American Statistical Association,
Vol. 46, pp. 68-78.

PEARSON, K. (1922)
" Tables of the Incomplete Γ-Function "
Cambridge University Press, Cambridge.

QUENSEL, C-E. (1945)
" Studies of the logarithmic normal curve "
Skandinavisk Aktuarietidskrift, Vol. 28, pp. 141-153.

RAMACHANDRAN, G. (1970)
" Some possible applications of the theory of extreme values for the analysis of fire loss data "
Fire Research Note no. 837, Fire Research Station.

SAVAGE, I.R. (1953)
" Bibliography of non-parametric statistics and related topics "
Journal of the American Statistical Association, Vol. 48, pp. 844-846.

SHUSTER, J.J. (1968)
" On the inverse Gaussian distribution function "
Journal of the American Statistical Association, Vol. 63, pp. 1514-1516.

SMIRNOV, N. (1948)
" Table for estimating the goodness of fit of empirical distributions "
Annals of Mathematical Statistics, Vol. 19, pp. 279-281.

TALLIS, G.M. and YOUNG, S.S.Y. (1962)
" Maximum likelihood estimation of parameters of the normal, the lognormal, truncated normal and bivariate normal distributions from grouped data "
The Australian Journal of Statistics, Vol. 4, pp. 49-54.

TWEEDIE, M.C.K. (1947)
" Functions of a statistical variate with given means, with special reference to Laplacian distributions "
Proceedings of the Cambridge Philosophical Society, Vol. 43, pp. 41-49.

TWEEDIE, M.C.K. (1957)
" Statistical properties of inverse Gaussian distributions, I "
Annals of Mathematical Statistics, Vol. 28, pp. 362-377.

WEIBULL, W. (1951)
" A statistical distribution function of wide applicability "
Journal of Applied Mechanics, Vol. 18, pp. 293-297.