THE EFFICIENCY AND ROBUSTNESS OF SCREENING PROCEDURES

by

J. C. YOUNG

Ph.D. Thesis

Department of Statistics

University of Edinburgh

October, 1968

Summary of Thesis

Frequently one is faced with the problem of determining the one or more best entities in a group presented for study. An example of such a situation is the choice, from a group of chemical compounds that may be active against cancer, of those worth further intensive study, or even adoption for general use. The design of the selection (or screening) procedure to be used in such a situation has considerable influence on the efficiency with which the best entities are chosen. The efficiency of various selection procedures, and the robustness of these procedures to a variety of operating conditions will be studied in this thesis for selection from both infinite and finite populations.

Part 1 of the thesis describes the problem in detail and gives a few of the many possible applications. The appropriate notation and terminology are given in chapter 1.3 and, at the same time, the many assumptions involved are described. Chapter 1.4 is the literature survey. The first section of the chapter surveys the field, emphasizing the literature relevant to the formulation of the problem used in this thesis; the last two sections describe and discuss two of the many other formulations.

Part 2 of the thesis is concerned with various aspects of selection from an infinite population. After an introduction to the topic in chapter 2.1, exact expressions are derived in chapter 2.2 for the mean and variance of the distribution of the true yield of the selected varieties after any number of stages of selection from an infinite normal distribution. Methods for using these expressions and for calculating higher order moments are then described. The chapter closes with a description of an alternative formulation of the problem that might simplify the analysis. The next chapter

is a detailed study of the influence of various factors on the infinite gain; emphasis is on the value of the inclusion of all previous results on a specific variety in the estimation of its true yield. Various methods for assessing the merit of different screening programs are discussed in the final section of the chapter. Chapter 2.4 looks briefly at selection schemes in which acceptance is allowed prior to the final stage of the program. A complicated scheme in which second stage replication is based on the first stage results is also discussed. An investigation into the effect on gain of the presence of interaction is made in the next chapter. Finally, in chapter 2.6, the robustness of the recommended selection scheme to departures from normality is studied by selecting from various symmetric, skewed, and bimodal distributions.

The last part of the thesis is concerned with selection from finite populations. After a few theoretical considerations and general comments on the importance of study of the finite case, the simulation technique is described. Chapter 3.4 consists of a comparison, for a variety of selection parameters, of the finite case with the infinite. The comparison is based on the acceptance (at every stage) of fixed proportions of the varieties being studied; emphasis is on the value of history. The next chapter investigates the use of cut-off points rather than fixed proportions in the definition of the acceptance rules. Combinations of the two types of acceptance rules are also studied. The chapter closes with a brief investigation of the value of accepting some varieties prior to the final stage.

Acknowledgements

Contents

PART 1. INTRODUCTION

1.1 Screening

A problem often encountered is that of choosing from a number of
entities presented for study those most suitable for some specific purpose.
Usually one wishes to choose those entities having the largest value of
some quantitative measurement of merit. Since in most instances the merit
of the individual in question is not measurable exactly but has associated
with it an experimental error, the design of schemes for best achieving the
desired result is statistical in nature.

As pointed out by a number of authors (Yates, 1950; Bechhofer, 1954;
Finney, 1960a, 1964), the traditional tests of significance are not adequate
to make the decisions that are necessary in such problems. Significance
tests have in fact little relevance since, in a screening program, one
generally knows in advance that the items under study are different; all
that is necessary to show significance is very intensive experimentation.
This of course may be very wasteful. Since the total amount of experimentation
is usually limited in some manner beyond the control of the experimenter,
it is necessary to balance the desire to achieve statistical significance
with the desire to test as many new entities as possible. An extreme
example of what can happen is the interesting result by Curnow (1959) that,
under certain conditions in 1-stage selection from a normally distributed
population, it never pays to select from fewer than five individuals. This
agrees with a result of Dunnett (1960) and Finney (1958a). In screening
work there is a real risk of missing a breakthrough by concentrating too
much on certain items and, as a result, not getting around to testing
others (Dunnett, 1961). What is really desired then is the most efficient

method of identifying the required number of suitable items without regard to statistical significance. This is not to belittle the role of experimental design in screening. On the contrary, blocking, replication, and other error reducing techniques are of prime importance.

In addition to the design of the experiment, the statistician is concerned with the number of entities that should be studied and the best use of the resources at the experimenter's disposal. He must advise whether to perform one large experiment or a series of smaller experiments at each of which the number of entities is decreased and the replication increased. This in turn raises the question of how many entities to retain at each stage and how much replication to use. These questions will be studied here.

It is not my purpose to produce a fixed technique that will replace the intuition of an experienced experimenter, but rather to outline a broad general approach that will guide him in his work. The robustness of the suggested approach to certain initial conditions and to the assumption of an infinite population will also be studied. It is felt that, although the terminology is drawn from plant breeding, the ideas will be applicable to a much wider range of problems.

1.2 Some Applications

In order to indicate where the results of this thesis may be useful, a few examples of fields in which screening of one sort or another has been performed in the past will be given in this section. An indication of the very wide interest in screening procedures is given by the large number of different types of journals cited in Federer's (1963) survey of the field.

One of the most widely publicized applications of screening procedures is drug screening. In this field the experimenter is presented with a large number of chemical compounds from which he must choose those active (or showing the most promise of being active) against some disease. Davies (1958) gave an example of screening for drugs active against tuberculosis. In a later paper (1964), he described the use of screening in improving the yield of antibiotics. Dunnett (1961) gave an example of procedures useful in isolating anti-cancer drugs. He pointed out that about 3000 compounds were handled per year in his company's laboratories and that very few had the desired activity.

A rather different application of screening was given by Nissen-Meyer (1964). He was concerned with the effective use of tests in those instances in medical diagnosis where preliminary tests are used to determine those people who should undergo more definitive tests that are both more expensive and more distressing to the person involved. It was Nissen-Meyer's desire to balance the extra cost and inconvenience with the very important goal of detecting as many people with the disease in question as possible. Since it is important in medical diagnosis to correctly classify all

individuals, special care is needed; medical diagnosis has, however, many characteristics in common with other types of screening.

Perhaps the most familiar of all screening processes is that which is performed on students in our educational system. The large number of tests, or screening stages, involved, and the fact that it is important that each individual be placed in a scholastic environment appropriate to his own particular talent, make this a very complex system. Finney (1962b) used a 2-stage model (a modified form of the model used in this thesis) and some approximate correlations to focus attention on the nature of the problem and the assumptions (both statistical and educational) involved. Both the qualitative and quantitative aspects of his paper and the discussion that followed it are of interest and point to a number of challenging problems for the statistician.

Cronbach and Gleser (1965) considered the related problem of screening for employment and other purposes by use of psychological tests. The nature of these screening programs differs from educational screening in that a large series of tests can be given rapidly. Although the approach taken by Cronbach and Gleser is very different from the other examples cited in this section, the problem is most certainly related.

A major difficulty in the previous two examples is the definition of what is meant by inherent ability and, subsequently, the calculation of the magnitude of the correlation between the test results and this ability. Finney (1962b) emphasized the caution necessary in such studies.

Although industrial problems are frequently a matter of optimizing on a continuous scale, it is also common to have to decide on the most

desirable of a number of distinct items. The result is a screening problem. The sort of thing one meets varies from the choice of the best of a number of different formulations of soap for marketing, to the choice of the best individual technique to give maximum yield in some manufacturing process. The examples of drug screening already mentioned give a good idea of the sort of problem often met, particularly in the chemical industry. In general, the basic purpose will be to maximize some aspect of profit.

Another rather similar application is the choice of the best (one or more) of a number (usually large) of varieties put forward by a plant breeder for testing. Finney (1958b) discussed in detail the sort of problem one meets in this field. Sprague and Federer (1951) described the screening of new varieties of corn. As this thesis is written in terms of plant selection no more will be said here about this particular application.

Obviously, with such a broad spectrum of applications, no one mathematical theory can hope to be completely adequate to them all. In any specific instance the experience of the experimenter will play a very important part in modifying the rules appropriately. It is felt, however, particularly in view of the robustness of the suggested plans to a wide range of parent distributions, that the guidelines suggested in this dissertation will apply to a wide range of applications.

1.3 Terminology - Notation - Assumptions

As was indicated in the last section, the terminology used in this thesis has, for the main part, been borrowed from the plant selection problem as formulated by Finney (see, for example, Finney, 1958a). A detailed outline of this terminology and the appropriate symbolic notation will be given in this section. In addition, where appropriate, the assumptions involved will be indicated and discussed.

The term VARIETY will be used to refer to a specific entity undergoing testing. Although originally intended to refer to one specific strain of the plant being studied, there is no reason for its not being extended to represent one formulation of a new drug, one animal, or even one specific manufacturing process. It will be assumed that each variety can be represented by a single property x that does not change over the history of the variety; we will call this property the TRUE YIELD. The purpose of the screening program will be assumed to be the selection of the variety (or varieties) with the maximum true yield(s). The modifications necessary if the minimum yield is desired are very simple. Modifications necessary for certain other purposes will be suggested at appropriate places in the main body of the thesis.

It should perhaps be pointed out here that in the sort of screening being considered individuality is not regarded as important. As long as we get the required number of good varieties it does not matter that we have discarded other varieties just as good as, or possibly even slightly better than those chosen. On the other hand, as has been mentioned earlier, it is very important when dealing with human beings to categorize each

individual correctly. Refusing the advantage of higher education to a genius is not only a loss to the community but also very unfair to the genius; speaking of being "unfair" to an individual variety of wheat is to miss the real purpose of the investigation.

Returning to the discussion at hand, it will be assumed that varieties with such basic faults as susceptibility to disease or insect infestation have been discovered and discarded during preliminary investigations prior to the major screening program being considered here. It will also be assumed that there is no other prior information on individual varieties. Under these conditions, yield will often be completely determined by a single characteristic such as the weight per unit area of the chief product of the variety. There is, however, no reason why x should not refer to some overall assessment of the variety derived from a number of individual characteristics. In an industrial context, the cost of a catalyst and the increased output resulting from its use would be compounded to give the "yield" of the catalyst. In instances where many character-istics are involved it will usually be necessary to use some sort of index. The use of indices in selection programs was discussed by Cochran (1951) and Finney (1964). Complications will also arise if a characteristic of interest is qualitative. In any case, it will be assumed here that we can somehow represent the true merit of our variety by a single value x called its true yield.

Our screening process will consist of k ($1 \leq k < \infty$) sequential experiments called STAGES. In a specific stage i ($1 \leq i \leq k$), the true yield

of each variety is estimated by the average

$$y_i = x + \epsilon_i$$

of the actual yield of all replications of that variety in that stage. Associated with the value y_i is the normally and randomly distributed error ϵ_i where

$$\epsilon_i \sim N(0, e_i^2).$$

This error is assumed to be independent of both the value of x for the variety in question and the errors of all the other varieties.

At stage i the decision to accept or reject a variety will be based on an estimate z_i of its true yield x. The variance of z_i is taken to be n_i^2. In the simplest case, where only information from the most recent stage is used to estimate x,

$$z_i = y_i,$$

and

$$n_i^2 = e_i^2.$$

When a number of stages are involved it will, however, be best to include all available information from previous stages in any estimate of the true yield x. If, at stage i, the weight attached to stage j is $a_{ij}(j \prec i)$ then

$$z_i = \sum_{j=1}^{i} a_{ij} y_j,$$

and

$$n_i^2 = \sum_{j=1}^{i} a_{ij}^2 e_j^2;$$

it is assumed that, at stage i,

$$\sum_{j=1}^{i} a_{ij} = 1.$$

Any two estimates z_i and z_j of x from two different stages i and j will usually be correlated; this correlation will be a function of the weights used. The use of information from previous stages, with appropriate weights, will be called the use of HISTORICAL INFORMATION and will be discussed in more detail in part 2.

The purpose of the screening process will be taken to be the selection of the most promising n of the N_0 varieties submitted for study;

$$\pi = \frac{n}{N_0}$$

may be termed the selection fraction. The N_0 varieties are considered to be a random sample from a continuously distributed PARENT POPULATION with mean zero (this can be achieved without loss of generality by a linear location transformation) and variance σ^2. In this study we are not concerned with the order of the n selected varieties.

The selection of the fixed fraction π of the varieties often meets with criticism as being too rigid. In any theoretical comparison such a restriction is necessary if different screening procedures are to be compared on a fair basis. More flexible schemes will be suggested later. A practical justification of this constraint is the fact that there will often be a limit to the number of varieties desired (or capable of being handled) in the final phase of selection in which detailed study is made and stocks are multiplied for commercial use.

The choice of the final n varieties is achieved as follows: prior to experimentation a random sample is made in which N_1 varieties are chosen from the N_0 originally present; at each subsequent stage i ($1 \leq i \leq k$)

an experiment is performed on the N_i varieties present, the N_{i+1} varieties with the largest estimated yields being passed on to the $(i+1)^{th}$ stage for more intensive study, and the remaining (N_i-N_{i+1}) varieties being discarded completely from further study. After the kth stage we will have the required $N_{k+1}=n$ varieties. The values N_i define a series of selection fractions

$$P_i = N_{i+1}/N_i \qquad i = 0,1,\ldots,k$$

where

$$\pi = P_0 P_1 P_2 \ldots P_k .$$

At first glance it may seem strange that we should randomly reject $(1-P_0)N_0$ varieties prior to the first experimental stage. On the other hand, unless the total amount of experimentation we can do is unlimited, some balance must be achieved between maximizing the probability of there being an exceptionally good item in the original population (by having N_1 as large as possible) and maximizing the probability of finding the best variety present (by more accurate assessment of fewer varieties). Normally, of course, the discard will not be completely random in that any hints of poor performance will be taken into account. This practical consideration cannot, however, be formulated mathematically. $P_0<1$ does, of course, point to an inefficiency in the system; certainly $(1-P_0)N_0$ of the varieties have been developed by the plant breeders in vain. There is also the possibility that $P_0>1$ indicating that, to make most efficient use of the experimental resources, it would be better to have more varieties than the N_0 with which we have been provided. In practice a small

variation from $P_0 = 1$ will be of little value since it will probably create difficulties in balancing the number of replicates with the number of varieties.

The decision to accept a fixed number of varieties from each stage is also questionable. Often it may be preferred to accept all varieties which yield above a certain fixed value or CUT-OFF point η_i at stage i. When considering selection from an infinite population this amounts to accepting a fixed proportion from each stage so there is no difference in the two approaches. In the finite case, however, differences do arise; these will be considered in part 3. As has been said before, the rules are meant to be modified depending on an experienced assessment of the promise, or lack of promise, of a particular group of varieties.

Another major assumption is that the total area of land available for experimentation, A, is fixed. We will refer to A as our RESOURCES. These resources must be split up between stages in some manner so we will consider the proportion

$$\alpha_i = A_i/A$$

of the total to be allocated to stage i; A_i ($\sum_{i=1}^{k} A_i = A$) represents the resources allocated to stage i. The assumption of a fixed value of A is again restrictive but necessary if it is desired to compare different systems on a fair basis. In practice an experimenter must often work within this sort of framework due to a decision which is beyond his control. In a continuing program there will often be k streams or COHORTS under test, one at each stage of selection. Under these circumstances, especially when there is a fixed total area of land

available, it may be necessary to fix the area of land allotted to each stage and to fix the number of varieties in each stage. The advantages of different systems of allocation of resources will be considered in part 3.

Having defined the framework of our screening system, it is now necessary to consider how the error variance e_i^2 changes with variations in A_i and N_i. Since an increase in the number of varieties for a fixed A_i will mean less experimentation on each variety, we might expect e_i^2 to be a strictly increasing function of N_i. In the same manner, we might expect e_i^2 to be a strictly decreasing function of A_i. If it is further assumed that all changes in A_i and N_i affect only replication, and that the change in replication is spread evenly over all varieties, it seems reasonable to write

$$e_i^2 = V \frac{A N_i}{N_0 A_i} \sigma^2.$$

V is a dimensionless constant called the VARIANCE FACTOR; it connects the experimental error with the variance of the parent population. The constants A and N_0 are included so as to enable us to simplify the expression to

$$e_i^2 = \frac{V \pi_{i-1}}{\alpha_i} \sigma^2$$

where

$$\pi_i = P_0 P_1 \ldots P_i.$$

The constants also serve, when considering a finite number of varieties, to remind us that the expression is consistent for different overall schemes only if they start with the same number of varieties, N_0, and the

same total resources A. This method of calculating the error variance
was suggested by Finney (1958a).

The magnitude of the variance factor plays a very important role in
a screening program. All other things being equal, very close determination
of the best n varieties will be possible when a small V is appropriate;
the choice will, on the other hand, be very nearly random when a large V
is appropriate. The actual magnitude of V will be affected by the
efficiency of our experimental design. This aspect of selection will not
be considered here.

In most of the work in this thesis it will be assumed that there is no
variety by stage (that is to say variety by year) or variety by experimental
site interaction. This assumption will be discussed in more detail in
part 2.

Finally we must decide how we are to compare screening programs using
different values of α_i and P_i. Obviously, the merit of any screening
procedure must be based on a comparison of those varieties selected with
those originally put forward for consideration. Since the value of an
item will usually be more or less linearly proportional to its yield, we
will take as our criterion the mean value of the true yields of the n
selected varieties. In other words, we will attempt to maximize the GAIN
or expected value of x, $E(x)$, in the selected population. In certain
cases, higher order moments of x may be of interest. In particular, we
may wish to minimize the variance of x or, in order to increase the chance
of finding a really spectacular variety, we may wish to make both the
variance and the positive skew of the selected varieties as large as

possible. The difficulty of interpretation of higher order moments makes criteria of this nature extremely complicated to handle. Even the variance (for non-normal distributions) tells us little about our distribution other than that it is (or is not) very spread out; it does not tell us how or where the spread takes place.

Other authors have of course used other criteria; Davies (1958) suggested maximizing the proportion of "good" varieties selected. Since Davies was considering his varieties to be either "good" or "bad", his criterion is equivalent to ours. That this is so can be seen by letting the proportion of good varieties selected be q and transforming (this can be done without loss of generality) to let the true yield of the bad varieties be 0; the true yield of the "good" varieties can be any fixed value g; the gain can be seen to be gq. Since g is fixed the maximization of q will also maximize the gain. If, on the other hand, there are degrees of "goodness" it would be unusual not to want to take this fact into account. It is difficult to imagine a criterion that will be as widely applicable as the maximization of gain so it will be used here.

1.4 The Literature

1.4.1 General survey

In 1963 Federer published a very broad survey of work done on statistical screening. His paper included an extensive bibliography not only of papers specifically concerned with screening, but also with a number of associated subjects such as truncated distributions. Emphasis was placed on the importance of interchange of ideas between workers in different areas of application. Federer also mentioned some experimental designs useful in selection programs.

A survey of work more directly related to this thesis was given by Finney (1964). After a detailed description of the general screening problem, Finney discussed, with examples, variations which are met in practice. He commented on the methods used by various other authors in investigating the screening problem and suggested techniques which might be useful in extending the theory. Although presented in the terminology of varietal selection, Finney's basic approach and computations would seem to apply to a more general class of screening problems.

The first paper within whose framework the bulk of the work in this thesis falls is that of Cochran (1951). In his paper, Cochran set as his objective the maximization of the mean value of true yield subject, for the sake of comparison of different screening plans, to a fixed total outlay of resources and a fixed average proportion of varieties accepted. Cochran showed that if the cumulative distribution function of the regression $r(y)$ of x on y (using the notation of this thesis) is assumed to be continuous and strongly monotone, the optimum selection rule is to select all varieties for which $r(y) > \eta$ where η is chosen to satisfy the desired

frequency of selection. Assuming a normal parent population, he calculated exact expressions for the mean of the selected population after 1-stage and (using properly weighted historical information) 2-stage selection. He used these expressions to calculate a few examples. Although hindered in his 2-stage calculations by the limited nature of tables of the bivariate normal distribution, Cochran found that the specifications for maximum gain did not change very much for various values of the ratio $\frac{e^2}{\sigma^2}$. This, and the fact that maximum gain occurred near $P_1 = \sqrt{\pi}$, foreshadowed Finney's suggested symmetrical specification for optimal selection. Cochran went on to discuss at some length the use of selection indices in screening programs and the complications that arise when they are derived from a sample.

The selection problem was attacked in a different manner in a series of papers which began with Finney's 1956 publication. In this paper, Finney derived the cumulants for the distribution of the true yield x after 1-stage selection from a normal distribution with normally distributed error. These cumulants were expressed in terms of the proportion selected. He also derived the first few terms of the infinite series for each of the first four moments of x after selection with normal error from a general distribution. These expansions were in terms of the cumulants of the distribution of x prior to selection.

In the next paper in the series, Finney (1958a) gave a detailed description of his formulation of the problem (the same formulation used in this thesis) and the assumptions involved. Using the formulae of his 1956 paper, he studied the effect of variation of the parameters for 1-stage and 2-stage selection from a normal parent population. In 1-stage selection,

Finney found that an initial random discard ($P_0 < 1$) could considerably increase the expected gain, especially for either fairly intensive selection or a large value of V. Based on the small variation of gain in the optimum region he found in 2-stage selection schemes, he suggested a general rule of taking

$$P_0 = 1,$$
$$P_i = (\pi)^{\frac{1}{k}} \quad \text{and}$$
$$A_i = \frac{A}{k}$$

as the approximate location of maximum gain in a k-stage program.

Finney's generalization in 1961 of his earlier (1956) results enabled calculation, for any error distribution, of the first four moments after selection; this was done in terms of the preselection cumulants (assuming they exist). Two series were derived for each of these moments: one in terms of an arbitrary cut-off value η and the other in terms of the proportion selected. Repeated applications of the appropriate series for the moments enabled the study of multiple-stage screening. An investigation of this sort would, of course, be limited to the first four moments and to the assumption, at each stage, that higher order moments were of negligible magnitude.

Curnow (1960) derived, as functions of the cut-off point η, exact expressions for the moments of the distribution of true yield after 1-stage selection, with normally distributed measurement error, from certain specific non-normal distributions. These expressions were given in terms of tabulated functions. They were useful in that they enabled the robustness

of screening programs to the assumption of normality of the parent population to be tested. They could also be used, in conjunction with Finney's results, to attempt a study of multi-stage screening from non-normal distributions.

Following the notation used by Finney (1956) for the first four moments after selection, Curnow (1961) gave an expression for the fifth moment. He proceeded to use the resulting increase in accuracy to study 1-stage and 2-stage selection from specific non-normal distributions, and 2-stage and 3-stage selection from normal distributions. Because of the increasing magnitude of higher order moments with increasing skew, the study of non-normal distributions was limited to distributions closely resembling the normal. In this region, Curnow found that the symmetrical selection scheme suggested by Finney for the normal distribution still resulted in gains near the maximum. In 2-stage and 3-stage selection from a normal parent, Finney's symmetrical scheme also gave results very near the maximum gain. Curnow pointed out that, in Cochran's example (Cochran, 1951), the use of first stage information in the second stage did not result in very large increases in gain.

In addition to a presentation of most of the results of the previous two papers, Curnow (1959) derived finite series for the exact cumulants after 2-stage selection from a normal distribution. A similar method is used in this thesis to derive exact expressions for the mean and variance of the distribution of the true yield after k-stage selection. Curnow also discussed the problems arising in the assessment of more than one character in a screening program.

Extension of methods for calculating cumulants after truncation was made to multi-variate distributions by Finney (1962a, 1963).

The question of how selection programs deduced from infinite population considerations applied to finite populations was investigated by Finney (1966). In this paper, by use of order statistics, exact results were obtained for 1-stage selection with error from finite normal populations. Multi-stage selection was studied by numerical simulation. Historical information was not used. Although, as expected, the infinite case considerably overestimated the gain actually achieved, it appeared to be a good guide to the sort of variation of gain with operating conditions found in the finite case. In particular, the symmetrical specifications suggested by Finney for near-maximum gain in the infinite case also resulted in near-maximum finite gain. As a result of the work in this paper Finney also suggested that (with fixed total resources) little would be gained by going beyond three or four stages.

Davies (1958) presented a slightly different approach to the study of selection. He suggested that a good approximation to the distribution of true yield in experiments in the drug industry was a 2-point distribution; inactivity being represented by zero mean and activity by a positive mean. Davies suggested criteria based on the maximization of the number of good drugs in those accepted. The analogy between screening and acceptance sampling was indicated and use was made of the operating characteristic curve of acceptance sampling in the comparison of different screening programs. Davies suggested that multiple-stage screening would be advantageous.

In a later paper, Davies (1964) studied selection from a distribution whose shape was derived from past results. Upper and lower confidence limits to the distribution were introduced in order that the dependence of results on the distribution shape might be studied. A simulation study of multi-stage selection from a finite population distributed in this non-normal manner led to optimum screening specifications very much in accord with those suggested by Curnow and Finney for normal populations. The author felt that only rarely would it be advantageous to go beyond 2-stage selection. He also found that the optimum location changed very little when a double-exponential error distribution was used.

Dunnett (1961) elaborated on the approach suggested by Davies (1958). Graphs were provided which gave the required cut-off points for 2-stage and 3-stage selection from a 2-point distribution for specified values of γ (the probability of accepting a bad drug) and β (the risk of rejecting a good drug). Equal replication was assumed at each stage. Formulae were provided for calculation of the expected number of stages required to reach a decision. Although γ and β specify the entire operating characteristic for a 2-point distribution, Dunnett pointed out that their choice depends solely on a rather arbitrary personal decision. To get around this weakness, he suggested a number of alternative criteria based on economic considerations.

Although possibly valid for a 2-point distribution, it may be dangerous to base a scheme for selection from a continuous distribution on two points of the operating characteristic; these curves often vary considerably between the two fixed points (Finney, 1964) and, as a result, there may be

considerable variation in the nature of the screening achieved.

King (1963, 1964), using restrictions similar to those used by Finney and Curnow, investigated conditions that would maximize the "proportion interesting" (meaning the proportion of varieties with activity above a certain level) in those varieties accepted from a drug screening program. Although selecting from distributions considerably removed from those studied by Finney and Curnow (the 2-point distribution and the negative exponential), King's conclusions were very similar. The largest discrepancy was in the allocation of resources. He, too, suggested that the advantage of using more than three stages would likely be very slight.

1.4.2 External economy

In the previous sections, and indeed throughout this thesis, we are concerned with what Finney (in the discussion following Grundy, Healy and Rees, 1956) has termed the "internal economy" of selection. Very briefly, this term refers to the optimization of a selection program, or any other procedure, within a fixed framework. Since, in our case, we are attempting to maximize the gain for a fixed total amount of experimental resources our work falls into this category.

An extended scheme in which the object is to decide upon the total experimentation required to maximize the net gain to society, or equivalently to minimize the net risk to society, has been advocated by a number of authors. Finney has called this approach the "external economy" of selection. Many of the authors whose papers were reviewed in the previous section have considered this broader formulation of the problem. Due to the greater

generality of this approach, in that it takes into account the costs of the program as well as the gains from it, some of the pertinent literature will be discussed in this section.

In a very interesting and useful paper, Grundy, Healy and Rees (1956) calculated the amount of experimentation necessary to minimize the economic risk associated with each individual variety in the screen. The fact that they considered the varieties individually makes their approach rather different than ours but, with certain simple modifications (many of which were mentioned by the authors), their approach can be usefully employed in a study of the external economics of a problem similar to ours. This, and the fact that they were essentially concerned with the finite case, makes consideration of their paper profitable.

Grundy, Healy and Rees assumed an initial experiment of n_1 replicates resulting in an estimate x_1 (with variance σ^2/n_1) of the true but unknown mean $\circ$ of the variety in question. At this point the economic parameters k, the cost per unit of additional experimentation, and k', the gain (or loss) due to unit increase (or decrease) in yield, were introduced. This permitted calculation of the risk associated with accepting a variety with a specific yield and specific second stage replication n_2. The "integral risk" was then found by integrating the risk over the fiducial distribution of $\circ$ based on x_1 and σ^2/n_1. The amount of second stage replication $n_2 \geq 0$ was chosen to minimize this integral risk.

The authors also indicated the rather simple modifications necessary when k was a function of n_2 or when k' was a function of the amount of delay involved when a second stage of experimentation was recommended.

The value $n_2 = 0$ was taken to indicate that the variety was to be rejected or accepted immediately based on x_1 and the associated risk.

An interesting fact arising out of the calculation of n_2 was that a small amount of additional experimentation was never advisable. This was due to the fact that a certain minimum of information was necessary to alter the decision that would have been made with no additional experiment- ation. The robustness of the procedure was indicated by the fact that small changes in n_2 near the optimum altered the integral risk only slightly. It is interesting to note that the ratio of maximum second stage to first stage replication n_2/n_1 is roughly of the same order of magnitude as the ratios recommended by Finney and Curnow in their formulation of the problem.

Although the authors based their decision to accept or reject a variety on the sign of the mean yield

$$\frac{(n_1 x_1 + n_2 x_2)}{(n_1 + n_2)},$$

they pointed out that other economic factors such as changeover costs could, without difficulty, be included by replacing x_1 and x_2 by $y_1 = x_1 - c$, and $y_2 = x_2 - c$ respectively throughout their calculations. Presumably the cost of initial development and the cost of the first stage of experimentation could also be included in this manner. Although Grundy, Healy and Rees's method has been criticized for not including n_1 as an economic variable, it seems reasonable in the absence of any prior knowledge of $^\circ$ to perform a small standard sized experiment on all proposed varieties and to include it as a fixed cost by increasing c. An advantage of the absence of a prior

assumption as to the nature of α is that n_2 is based entirely on the experimental results as is the resultant estimate of α.

In order to test the variation of the recommended value of n_2 under rather different circumstances, Grundy, Healy and Rees calculated its value for the case in which there is a prior distribution of α with mean 0 and variance $\alpha^2 \sigma^2 / n_1$. The resultant values of n_2 for $\sim = 1$ were very close to those found when no prior distribution was used. As in all previous modifications the nomogram used by the authors for calculation of n_2 in the simplest case was used, with slight modifications to the definition of the parameters, to calculate n_2 for the normal prior distribution of α.

One difficulty in applying Grundy, Healy and Rees's scheme to selection from a population of varieties is the fact that there is no upper limit to the number of varieties recommended for acceptance; what is usually desired is to accept the variety with the minimum risk, or possibly to accept a fixed maximum number of the varieties which promise to be most profitable. As long as we have the required number there is no loss (or gain) associated with rejecting (or accepting) a variety which is good but less profitable than the rest. The necessity in Grundy, Healy and Rees's scheme of differentiating between various profitable varieties raises another question: how are we to decide between two varieties when their estimated means are very similar but when one estimated mean has associated with it a much larger error variance than the other? Special rules depending on the experimenter's requirements would need to be introduced to solve these problems.

A useful screening scheme might possibly result from a combination

of the approach taken by Grundy, Healy and Rees with that studied in this thesis. Finney (1958a) pointed out that his approach is concerned with the second of three phases of selection; the third phase will usually involve a very intensive study of a very few varieties. A recommendation as to the size of this third phase could be made by using the results from the second phase as though they were the preliminary results of a program like that studied by Grundy, Healy and Rees. That this is appropriate can be seen from the fact that our program is of fixed total size and results in a few varieties each with an estimated yield and variance. Since the economic aspects of a screening program are very difficult to specify until something is known about the varieties in question, and since resources for initial investigations of a large number of unknown varieties are likely to be limited and beyond the control of the experimenter, this should prove a very practical combination of the two approaches.

The external economy of the type of procedure studied in this thesis was investigated by Finney (1960b) and extended by Curnow (1961). In these papers a study was made of the balance between the total cost of experimentation and the net gain to the national interest of the resulting increase in yield. The costs included the cost of breeding new varieties for study (varying N) and the cost of changing the experimental area A. Both these variables affect the statistical variability of the program and hence the potential gain. Finney calculated the optimum choice of variables for 1-stage selection for various assumed monetary values of cost and gain.

Curnow, in a similar study of 2-stage selection, found the total expected gains to be 15 to 20 percent higher than for 1-stage selection.

Optimal A was the same for both 1-stage and 2-stage programs but optimal N was much larger in the latter case. Curnow suggested that more than two stages would be unlikely to produce much additional gain. He discussed in detail some of the disadvantages of having a large number of stages.

The difficulty of evaluating the real monetary effect of the factors involved is the reason for our not considering the external economy of selection in more detail in this thesis. In practice an external study is always internal to a larger problem. Eventually one runs up against some limit on funds whether it be the budget of one's department, the research budget of the company or the total income of the company. These factors and many others are related and obviously the relationship is a very complicated one. In any situation it is necessary to consider how the money could otherwise have been spent had it not been allocated to the program under study; without this consideration the total resources of the economy would soon be entirely used up.

Another problem is the fact that the exact value of an increase in yield is difficult to assess. Any change in a marketable product in a competitive society changes the entire market, making the situation decidedly non-linear. The same problem in another context is very effectively stated by King (1963) when he says, "The problem of attaching a numerical value to the discovery of a cure for a disease is exceedingly nasty".

The above is not to say that we should not use the external economy approach; rather that we should put very great care into any such formulation as there are a lot of factors involved and the results of decisions can be far reaching. The immense difficulty of the problem does not let the

statistician off the hook; no problem of this nature will ever be truly solved until we come to grips with these broader issues. Any investigation is of great value even if it only brings to light the large number of implicit assumptions involved in the economics of research projects.

1.4.3 Another approach to screening

Another approach to screening, although primarily concerned with correctly ranking (according to some predetermined system) all individuals, and hence somewhat different in its goal from this thesis, has been recommended for similar situations and considered by a number of authors therefore it will be discussed briefly here. Since Bechhofer (1954) described the theme in some detail and included a number of interesting variations, this discussion will be based on the formulation and notation of his paper.

Bechhofer was primarily concerned with dividing, in one stage, k variables (varieties) into s (s $\leq$ k) classes. In Bechhofer's procedure the experimenter was asked to specify in advance the number of variables, k_j ($\sum_{j=1}^{s} k_j = k$), to go into each class. The procedure then enabled calculation of the number of observations N_i to be made on each variable i. The result was an estimate $\bar{x}_i$ of the true (but unknown) population mean u_i of the variable i. It was assumed that

$$\bar{x}_i \sim N(u_i, \sigma_i^2/N_i).$$

The variables were then ranked according to the estimates of their true means, the lowest k_1 being put into the first class, the next k_2 into the

second and so on until exactly k_s were left for uppermost class. No importance was placed on order within a class. A summary of Bechhofer's notation along with our equivalent (when one exists) is given in table 1.4.1.

<div align="center">TABLE 1.4.1</div>

<div align="center">A comparison of the notation used in this thesis with that
used by Bechhofer (1954).</div>

	This thesis (finite case)	Bechhofer (1954)
True yield	x	μ_i
Estimated yield	y	$\bar{x}_i$
Variance of estimated yield	$\dfrac{V\sigma^2 \pi_{i-1}}{\gamma_i}$	$\dfrac{\sigma^2}{N_i}$
Population size	$P_0 N_0$	k
No. stages	$k \geq 1$	1
No. categories	2 (good and bad)	$s \leq k$
No. accepted into top category	n	k_s

Although a large number of situations (unequal variances σ_i^2, two-way classifications, more than two classes) were mentioned by Bechhofer, they result in so many difficulties and are so different to the problem of this thesis that this discussion will be limited to the division of k variables with equal variances into 2 classes ($s = 2$). This case corresponds closely with what we are trying to do and is the case which Bechhofer studied in

most detail.

A number of differences arise between Finney's approach and Bechhofer's in the method of tackling the problem. Bechhofer set as his goal the determination of the minimum sample size N required to give a minimum probability P^* of placing the varieties into the correct classes when the true means of the varieties in the two classes differ by at least $\delta^* \geq 0$. Both P^* and δ^* were assumed to have been specified by the experimenter in advance. In order to calculate the sample size, some assumption had to be made about the configuration of the true means. Bechhofer assumed a "least favourable" configuration of population means; that is to say he chose k_1 (the number of varieties to go in the lower class) of the means equal to zero and the remaining k_2 equal to $+\delta^*$. Based on this assumption (rather like a minimax procedure), he gave a table of the values of $\sqrt{N}\,\delta^*/\sigma$ required to achieve various values of P^* for certain values of k_2 and populations as large as k = 14. Since the μ_i will usually be much more widely scattered, Bechhofer's procedure will recommend a larger sample size than is actually required to achieve the value P^* for a given δ^*. In this sense it is inefficient.

Partly as a result of this difficulty, Dunnett (1960) suggested a number of modifications for choosing the "best" of k varieties. He devised a procedure in which a prior estimate U_i of the population mean u_i was used. It was assumed that

$$U_i \sim N(u_i, \sigma_0^2).$$

In addition to studying the procedures suggested by Bechhofer, Dunnett

relaxed his conditions slightly in that he required P^* to be the probability of choosing a population whose true mean is within δ^* of the largest population mean. The degree of the resulting multivariate integral was such that, with existing tables, calculations could be made only for k = 2. In order to overcome the rather arbitrary nature of P^* and δ^*, Dunnett also introduced methods for calculating them based on economic considerations.

It would be interesting to study Dunnett's procedures for larger values of k, especially if the U_i were obtained from a small first stage experiment. The advantage, in such a situation, of letting the recommended value of N_i vary according to the value of U_i for each individual variety would also be an interesting study. The paper by Grundy, Healy and Rees (1956) discussed in the last section was concerned with this sort of thing. An even simpler case would be to have N_i = 0 for all varieties with U_i less than some specified amount, and all other values of N_i equal to some constant N. Since in practice we are often interested in the absolute gain, or at least the gain relative to existing varieties or procedures, it might be useful to add a control.

The advantage of Dunnett's and Bechhofer's approach is that it is concerned with the finite case. On the other hand, the algebra is just as intractable as that met by Finney (1966) in his study of the finite case and progress will be very difficult.

One of the difficulties with the preceding work is that it is only concerned with 1-stage selection while a number of authors (e.g. Davies, 1958; Finney, 1958a, 1966) have shown that 2-stage or even 3-stage work

makes more efficient use of resources. Bechhofer (1958) suggested
2-stage and sequential procedures based on the same formulation as his
1-stage work. These procedures did not allow for rejection of uninteresting
items at intermediate stages. In addition, although guaranteed of stopping,
his sequential procedure could take a considerable number of stages
(Bechhofer 1966) to reach a decision. The large number of uninteresting
varieties usually involved, and the long wait before experimental results
from a given stage become available, make the use of Bechhofer's multi-stage
procedures inadvisable for the sort of problems considered here.

PART 2. THE INFINITE CASE

2.1 Introduction

In this part of the thesis, we will study selection from a population assumed large enough to be represented by a continuous frequency function. In particular, we must assume that both $P_0 N_0$, the number of varieties initially present, and n, the number of varieties accepted after the final stage, are large. Just how large they must be will be investigated in part 3 as part of a general comparison of the finite case with the infinite.

Although exact expressions have been derived for the mean (Cochran, 1951; Curnow, 1959) and the other moments (Curnow, 1959) of the distribution of x after 2-stage selection from an infinite normal population, their use has been made difficult by the lack of adequate tables for the bivariate normal distribution; prior to this study, no simple exact expressions were available for the gain from more than two stages of selection from a normal population, or from more than one stage from non-normal populations. As a result, detailed study (Finney, 1958a; Curnow, 1961) of selection in more than one stage has been limited to the use, at each stage of the selection program, of series expansions for the first four (Finney, 1956, 1961) or five (Curnow, 1961) moments of the distribution of x.

The dependence of screening studies on these series has involved a number of sources of error of unknown magnitude. Since there is, in general, no justification for assuming that high order cumulants will decrease in magnitude, there is certainly a real danger in ignoring them. Another aspect of the danger of ignoring them is illustrated by a theorem of Marcinkiewicz to which Curnow (1961) drew attention. This theorem states

that there are no distributions other than the normal that have only a finite number of non-zero cumulants. Since we are dealing, after the first stage, with the tails of distributions, where it is known that cumulants do not, in general, decrease quickly, this theorem most certainly applies. To add to the difficulties, the exact cumulants were known only for the distribution of x after one stage of selection from a normal distribution. This meant that accuracy suffered not only from the fact that high order cumulants were ignored, but also from inaccuracies in the lower order cumulants. In addition, Curnow pointed out that convergence would be particularly bad when the intensity of selection is very high and when the variance factor V is very small. The dependence, in screening studies, of the moments of the distribution of x (and hence its cumulants) at any stage on its cumulants at the previous stage means that the errors just discussed will accumulate as the number of stages increases.

The accuracy of the results found when using cumulant expansions to calculate the gain when selecting from non-normal distributions was even more questionable; in that case the exact cumulants were usually not even known for the distribution of x after one stage of selection. Additional problems were also introduced by the fact that higher order cumulants increase in importance the further the distribution departs from normality. That these problems exist on a practical scale can be seen from the fact that Curnow (1961) ran into convergence problems in 2-stage selection as soon as his parent distributions departed very far from symmetry.

The main advantage of the work in this part of the thesis is the fact that exact expressions for the gain and variance, after any number of stages

of selection, have been derived and subsequently used (with the aid of numerical integration and a computer) for a wide number of selection schemes with both normal (up to four stages) and non-normal (up to three stages) parent distributions. Another useful feature is that these expressions permit the use, at any stage, of all previous information on a variety in the estimate of its yield. This means that the decision to accept or reject can be based on all available information on the variety in question. History has not been taken into account in any of the previous detailed studies.

The inclusion of historical information in our calculations raises the question of how much weight to attach to the results of each stage. In the case in which the variances at each stage are known exactly, a well known result is that the weighted mean will have minimum variance when it is calculated by weighting each result on a variety inversely to its error variance. This means that we must choose the weight applied at the pth stage, to the information from the ith stage, to be

$$a_{pi} = \frac{1/e_i^2}{\sum_{j=1}^{p} 1/e_j^2} . \qquad (2.1.1)$$

The resulting estimate of x is then

$$z_p = \frac{\sum_{j=1}^{p} y_j/e_j^2}{\sum_{j=1}^{p} 1/e_j^2} , \qquad (2.1.2)$$

and its error variance is

$$e_p^2 = \frac{1}{\sum_{j=1}^{p} 1/e_j^2} , \qquad (2.1.3)$$

Since the exclusion of properly weighted information on the yield of a variety at previous stages seems to mean that valuable information is being ignored, and since, when the variances at each stage are known exactly, the inclusion of this information will obviously improve the expected yield, it will, unless specifically stated, always be included in calculations in this thesis. The practical value of this procedure will be discussed in detail as we go along.

Practical difficulties arise in using historical information when the error variances (or at least their relative values) are not known exactly. Yates and Cochran (1938) pointed out that the weighted mean loses greatly in efficiency when the error variances are estimated from a small number of degrees of freedom. They also pointed out that further losses in the efficiency of the weighted mean will occur when there is variation in the true yield from stage to stage. Such stage-to-stage variation in true yield will occur when there is a variety by year inter-action. The procedure to follow in such a situation will depend very much on what is desired of the varieties selected. The question of appropriate weights will be discussed further in chapter 2.5 during consideration of an idealized version of the operation of a selection program when interaction is present.

2.2   The Derivation and Calculation of the Exact Moments

2.2.1   The derivation of the moments

In the derivation in this section, direct use will be made of the cut-off points η_i rather than the proportions P_i used by Finney in his work. This is done out of necessity and not necessarily out of preference. Fortunately, selecting all varieties yielding above a certain fixed cut-off point η_i is, in the infinite case, exactly the same as accepting a fixed proportion of the varieties. In fact, even though the expressions are derived in terms of the η_i, the P_i are still, for the sake of ease of comparison of different systems, the basic parameters; the η_i are calculated from the P_i. The effect of various other acceptance rules on the finite gain will be studied in part 3.

At stage i we will accept a variety if our estimate z_i of its true yield is greater than the cut-off point η_i, otherwise the variety will be completely discarded from further consideration. As a result, a variety will reach a specific stage j if and only if

$$z_i > \eta_i$$

for all $i \leq j$. This means that the probability of a variety's surviving the entire k stages of the program is

$$S(x) = \Pr(z_1 > \eta_1, z_2 > \eta_2, \dots, z_k > \eta_k \mid x),$$

where x is the variety's true yield. $S(x)$ has been called the selection function (Curnow, 1961) and is closely related to the operating characteristic of acceptance sampling. Since, for specified values of the parameters P_0, V, σ_i and η_i $(1 \leq i \leq k)$, $S(x)$ completely describes the screening program

it could, in a sense, be thought of as the "operating characteristic" of the screening program. It can be seen that when it is desired to accept all varieties whose true yield is greater than some constant η_0, perfect selection will occur when

$$S(x) = 1 \quad (x > \eta_0),$$
$$S(x) = 0 \quad (x \leq \eta_0).$$

When using $S(x)$ it must be kept in mind that, for specified values of the P_i, the values of the η_i are functions of the shape of the assumed parent population. The result is that $S(x)$ is not independent of the shape of the parent distribution. The properties of $S(x)$ will be discussed further, and put to use in the assessment of the merit of different screening programs, at the end of the next chapter.

If we make use of the fact that

$$z \sim N(x, \theta^2),$$

and make the transformation

$$u_i = \frac{z_i - x}{\theta_i},$$

we can express the selection function in the form

$$S(x) = I(\frac{\eta_1 - x}{\theta_1}, \frac{\eta_2 - x}{\theta_2}, \ldots, \frac{\eta_k - x}{\theta_k}; R_0)$$

$$= I(\frac{\eta - x}{\theta}; R_0), \tag{2.2.1}$$

where I denotes the k-variate normal probability integral

$$I(\frac{\eta - x}{\theta}; R_0) = \int_{\frac{\eta_1 - x}{\theta_1}}^{\infty} \ldots \int_{\frac{\eta_k - x}{\theta_k}}^{\infty} \exp(- \tfrac{1}{2} u R_0^{-1} u') / [(2\pi)^k |R_0|]^{\frac{1}{2}} du_1 \ldots du_k. \tag{2.2.2}$$

The matrix R_0 is the $k \times k$ variance-covariance matrix of the variables u_i of the row vector u; R_0 has elements r_{ij}. In the case in which only the current information on a variety is used in the decision to accept or reject it, we have

$$z_i = y_i,$$

and

$$\eta_i = e_i.$$

This means that

$$r_{ii} = 1 \quad (i = j),$$

and

$$r_{ij} = 0 \quad (i \neq j). \qquad (2.2.3)$$

This is the situation studied by Finney (1956, 1958) and Curnow (1959, 1961). When properly weighted historical information is used (see equations 2.1.1, 2.1.2 and 2.1.3), we have

$$r_{ij} = \frac{\eta_i}{\eta_j} \quad (i > j),$$

$$r_{ij} = \frac{\eta_j}{\eta_i} \quad (i < j), \qquad (2.2.4)$$

and

$$r_{ii} = 1 \;;$$

since the minimum variance estimate of x is being used, $\eta_{i+1} \leq \eta_i$ and the above correlation will always be ≤ 1.

Curnow (1960) and, with different notation, Cochran (1951) studied the use of history briefly for 2-stage selection.

The correlations just stated are correct only if the errors in measuring the yield of a variety at one stage are uncorrelated with those at any other stage. In practice a number of factors may invalidate this assumption: the samples chosen to represent a particular chemical compound (variety) may, particularly if they are from one particular supplier, be unusually uniform; the inability to choose years (stages) truly at random may result in unknown correlation; the use of the same sites to represent a variety at different stages may cause unknown correlation. The action of these factors is difficult to formulate mathematically but the comments in chapter 2.5 may have some relevance to their effect.

When it is desired to select a proportion P of an infinite population of varieties that are initially distributed according to a frequency function $h(x)$, the frequency function $g(x)$ of the varieties left after the kth stage of selection will be given by

$$g(x) = S(x)h(x)/P \qquad (2.2.5)$$

where

$$P = \int_{-\infty}^{+\infty} S(x)h(x)dx.$$

This is true for any distribution $h(x)$ so, in theory, the problem of calculating the moments of the distribution of x after selection is now solved in that all moments can be calculated directly from the relationship

$$u_n = \int_{-\infty}^{+\infty} x^n g(x)dx.$$

This integral is not usually easily integrable; even when using numerical integration a large number of points and hence a lot of labour, or computer

time, will be required. For lack of a better method it will, however, be used in special cases in chapters 2.4 and 2.6 when a simpler method is not available.

Fortunately considerable simplification is possible when selecting from the normal distribution. From this point on in this chapter it will be assumed that selection is being performed on the normally distributed parent population

$$h(x) = \phi(x) = \frac{1}{\sqrt{2\pi}} \exp(-\tfrac{1}{2}x^2). \qquad (2.2.6)$$

The assumed mean of zero and variance of one can always be easily achieved, without loss of generality, by a location and scale transformation. The advantage of making this transformation is that the gain is automatically calculated as a fraction of the parent population standard deviation and so is scale invariant under any linear transformation of the units of measurement.

The following method of deriving the gain and the variance of x after k stages of selection is similar to that used by Curnow (1960) when he calculated the exact moments of the distribution of x after two stages of selection from a normal population. From 2.2.1, 2.2.5 and 2.2.6 we obtain the moment generating function of x

$$M_x(t) = \int_{-\infty}^{+\infty} e^{xt} \phi(x) I(\tfrac{\eta-x}{\rho}; R_0) \, dx / P$$

$$= e^{\frac{1}{2}t^2} \int_{-\infty}^{+\infty} \phi(x-t) I(\tfrac{T-x}{\rho}; R_0) \, dx / P .$$

Putting $x = t+s$ we get

$$M_x(t) = e^{\frac{1}{2}t^2} \int_{-\infty}^{+\infty} \emptyset(s) I\left(\frac{\eta-t-s}{\rho};R_0\right) ds/P \ .$$

If we now transform our variables from u_i to v_i such that

$$v_i = \frac{\rho_i u_i + s}{\omega_i} \ ,$$

where

$$\omega_i^2 = 1 + \rho_i^2 \ ,$$

we get

$$M_x(t) = e^{\frac{1}{2}t^2} \int_{-\infty}^{+\infty} \emptyset(s) I(\eta(t);R) ds/P \ ,$$

with the result that

$$M_x(t) = e^{\frac{1}{2}t^2} I(\eta(t);R)/P; \qquad\qquad (2.2.7)$$

$\eta(t)$ is considered to be a k-element row vector with elements

$$\eta_i(t) = \frac{\eta_i - t}{\omega_i} \ .$$

I always, unless specifically stated to the contrary, is taken to represent
a k-variate normal probability integral of the form of 2.2.2. The
elements of the k×k variance covariance matrix R are

$$\rho_{ij} = \frac{\rho_i \rho_j r_{ij} + 1}{\omega_i \omega_j} \qquad (i \neq j).$$

When no historical information is used, they reduce to

$$\rho_{ij} = \frac{1}{\omega_i \omega_j}$$

and, when optimally weighted historical information is used, to

$$\rho_{ij} = \frac{\omega_i}{\omega_j} \qquad (i > j),$$

and

$$\rho_{ij} = \frac{\omega_j}{\omega_i} \qquad (i < j).$$

In both cases the variance is given by $\rho_{ii} = 1$. An immediate result of 2.2.7 is that

$$P = \overline{M_x(\theta)} = I(T/\omega;R). \qquad (2.2.8)$$

The gain after k stages can now be calculated, but first some new notation must be introduced. In 2.2.7 the row vector

$$v = (v_1, v_2, \ldots, v_k),$$

where k is as usual the number of stages, is implicit in the integral I. We now introduce the modified k-1 element vector

$$v^m = (v_1^m, v_2^m, \ldots, v_{m-1}^m, v_{m+1}^m, \ldots, v_k^m),$$

where v_i^m represents the transformation

$$v_i^m = \frac{v_i - \rho_{im} v_m}{\sqrt{1 - \rho_{im}^2}} \qquad (1 \le i \le k, \text{ but } i \ne m); \qquad (2.2.9)$$

v_i and v_m are elements of the vector v and ρ_{im} their correlation. The vector v^m is similar to the vector v except that the elements are transformed according to 2.2.9 and the mth element is missing. The correlation coefficient of two transformed variables v_i^m and v_j^m is given by

$$\rho_{ij}^m = \frac{\rho_{ij} - \rho_{im} \rho_{jm}}{\sqrt{1 - \rho_{im}^2} \sqrt{1 - \rho_{jm}^2}} \qquad (i \ne j), \qquad (2.2.10)$$

and

$$\rho_{ii}^{m} = 1$$

where $1 \leq i,j \leq k$, but $i,j \neq m$. It can be seen the ρ_{ij}^{m} is equal to $\rho_{ij \cdot m}$, the partial correlation of the variables i and j after m has been eliminated. R^{m} is the corresponding $(k-1) \times (k-1)$ variance-covariance matrix. As with v^{m} and v, R^{m} is similar to R except that it has elements ρ_{ij}^{m} instead of ρ_{ij} and the mth row and column have been removed. Finally we define the $k-1$ variate normal integral

$$I^{m}(\eta^{m}(t);R^{m}) = I^{m}(\eta_{1}^{m}(t),\eta_{2}^{m}(t),\ldots,\eta_{m-1}^{m}(t),\eta_{m+1}^{m}(t),\ldots,\eta_{k}^{m}(t);R^{m}), \quad (2.2.11)$$

where

$$\eta_{i}^{m}(t) = \frac{\eta_{i}m_{i} - \rho_{im}\eta_{m}m_{i} - t(m_{m} - \rho_{im}m_{i})}{m_{i}m_{j}\sqrt{1-\rho_{im}^{2}}} .$$

We can now calculate the gain from the relationship

$$E(x) = \frac{d}{dt} M_{x}(t)\Big|_{t=0}$$

and equation 2.2.7. The result is

$$E(x) = \frac{1}{P}[te^{\frac{1}{2}t^{2}}I(\eta(t);R)+e^{\frac{1}{2}t^{2}}\frac{d}{dt}I(\eta(t);R)]\Big|_{t=0}. \quad (2.2.12)$$

Calculating the derivative in the above expression according to the method of appendix 1 (equation A3) and substituting $t=0$ we get

$$E(x) = \sum_{i=1}^{k}\phi(T_{i})I^{i}(\eta^{i}(0);R^{i})/(m_{i}P), \quad (2.2.13)$$

where $T_{i} = \eta_{i}/m_{i}$. Any vagueness in the notation arising when $k=1$ will be cleared up immediately after the following derivation of $E(x^{2})$.

In order to derive the second moment of the distribution of x in the selected population in a concise form, some logical extensions must be made to the preceding notation. Now, keeping in mind that $1 \leq i,j \leq k$ and $i,j \neq m,n$, we let

$$v_i^{mn} = \frac{v_i^m - \rho_{in}^m v_n^m}{\sqrt{1-(\rho_{in}^m)^2}} \tag{2.2.14}$$

represent the ith of the k-2 elements of the vector v^{mn} of transformed variables. As a result of this transformation

$$\rho_{ij}^{mn} = \frac{\rho_{ij}^m - \rho_{in}^m \rho_{jn}^m}{\sqrt{1-(\rho_{in}^m)^2}\sqrt{1-(\rho_{jn}^m)^2}} \qquad (i \neq j)$$

will be the $(ij)^{th}$ correlation coefficient in the $(k-2)\times(k-2)$ variance covariance matrix R^{mn} of the variables v_i^{mn}; as usual $\rho_{ii} = 1$. Still following the previous notation we define the k-2 variate normal integral

$$I^{mn}(\eta^{mn}(t);R^{mn}) = I(\eta_1^{mn}(t),\ldots,\eta_{m-1}^{mn}(t),\eta_{m+1}^{mn}(t),\ldots,\eta_{n-1}^{mn}(t),\eta_{n+1}^{mn}(t),$$

$$\ldots,\eta_k^{mn}(t);R^{mn}), \tag{2.2.15}$$

where

$$\eta_i^{mn}(t) = \frac{\eta_i^m(t) - \rho_{in}^m \eta_n^m(t)}{\sqrt{1-(\rho_{in}^m)^2}} \quad .$$

The pattern developing in the notation is obvious and extension to calculation of any moment is, as far as notation is concerned, easy; for the nth moment the largest number of superscripts involved will be equal to n.

While on the topic of superscripts it should be pointed out that if

the number of superscripts is equal to the number of stages our vectors have k-k = 0 elements; in this case we define

$$I^{i_1, i_2, \ldots i_k}(\ldots) = 1.$$

When k = 1 we have, for example,

$$I(\eta(t); R) = \int_{\eta_1(t)}^{\infty} \emptyset(v) dv = \Phi(\eta_1(t))$$

and

$$I^1(\eta^1(t); R) = 1;$$

for the case k = 2, $I(\eta(t); R)$ is the bivariate normal distribution and

$$I^{1,2}(\eta^{1,2}(t); R) = 1.$$

When the number of superscripts is greater than the number of stages, I is defined to be equal to zero.

Returning to the problem at hand, we now have adequate notation to calculate the second moment of x in the selected population. From the relationship

$$E(x^2) = \frac{d}{dt}[\frac{d}{dt}M_x(t)]\Big|_{t=0},$$

and 2.2.12, it can be seen that

$$E(x^2) = \frac{1}{P} \frac{d}{dt} \left(e^{\frac{1}{2}t^2}[tI(\eta(t); R) + \sum_{i=1}^{k} \emptyset(\eta_i(t)) I^i(\eta^i(t); R^i)/\eta_i] \right)\Big|_{t=0}. \tag{2.2.16}$$

Now, using equation A6 from appendix 1 and setting $t = 0$, we have

$$F(x^2) = 1 + \frac{1}{P}\left\{\sum_{i=1}^{k}\emptyset(T_i)/m_i\left[T_i I^i(\eta^i(0);R^i)/m_i\right.\right.$$

$$\left.\left. + \sum_{\substack{j=1 \\ j\neq i}}^{k}(m_i - m_j\rho_{ij})/(m_i m_j\sqrt{1-\rho_{ij}^2})\emptyset(\eta_j^i(0))I^{ij}(\eta^{ij}(0);R^{ij})\right]\right\}.$$

Therefore

$$E(x^2) = 1 + \frac{1}{P}\left[\sum_{i=1}^{k}T_i\emptyset(T_i)I^i(\eta^i(0);R^i)/m_i^2\right.$$

$$\left. + \sum_{i<j}^{k}(2m_i m_j - \rho_{ij}(m_i^2 + m_j^2))/(m_i^2 m_j^2\sqrt{2\pi}\sqrt{1-\rho_{ij}^2})\emptyset\left\{\left(\frac{T_i^2 + T_j^2 - 2\rho_{ij}T_i T_j}{1-\rho_{ij}^2}\right)^{\frac{1}{2}}\right\}I^{ij}(\eta^{ij}(0);R^{ij})\right],$$

$$(2.2.17)$$

where $\sum_{i<j}^{k}$ represents summation over all $\binom{k}{2}$ pairs of i and j. The variance of x in the selected population can, of course, be calculated from $\mathrm{Var}(x) = E(x^2) - E(x)^2$.

The following specific expressions for $E(x)$ and $E(x^2)$ after one to three stages of selection have been derived from equations 2.2.13 and 2.2.17:

one stage,

$$\text{gain} = E(x) = \frac{1}{Pm_1}\emptyset(T_1), \qquad (2.2.18)$$

$$E(x^2) = 1 + \frac{T_1}{Pm_1^2}\emptyset(T_1); \qquad (2.2.19)$$

- 47 -

two stages,

$$E(x) = \frac{1}{P}\left[\frac{1}{m_1}\phi(T_1)\Phi\left(\frac{T_2-\rho_{12}T_1}{\sqrt{1-\rho_{12}^2}}\right) + \frac{1}{m_2}\phi(T_2)\Phi\left(\frac{T_1-\rho_{12}T_2}{\sqrt{1-\rho_{12}^2}}\right)\right], \tag{2.2.20}$$

$$E(x^2) = 1 + \frac{1}{P}\left\{\frac{T_1}{m_1^2}\phi(T_1)\Phi\left(\frac{T_2-\rho_{12}T_1}{\sqrt{1-\rho_{12}^2}}\right) + \frac{T_2}{m_2^2}\phi(T_2)\Phi\left(\frac{T_1-\rho_{12}T_2}{\sqrt{1-\rho_{12}^2}}\right)\right.$$

$$\left. + \frac{2m_1m_2-\rho_{12}(m_1^2+m_2^2)}{m_1^2 m_2^2 \sqrt{2\pi}\sqrt{1-\rho_{12}^2}}\phi\left(\left[\frac{T_1^2+T_2^2-2\rho_{12}T_1T_2}{1-\rho_{12}^2}\right]^{\frac{1}{2}}\right)\right\}; \tag{2.2.21}$$

three stages,

$$E(x) = \frac{1}{P}\left[\frac{1}{m_1}\phi(T_1)I\left\{\frac{T_2-\rho_{12}T_1}{\sqrt{1-\rho_{12}^2}}, \frac{T_3-\rho_{13}T_1}{\sqrt{1-\rho_{13}^2}}; \begin{pmatrix}1 & \rho_{23\cdot1} \\ \rho_{23\cdot1} & 1\end{pmatrix}\right\}\right.$$

$$+ \frac{1}{m_2}\phi(T_2)I\left\{\frac{T_1-\rho_{12}T_2}{\sqrt{1-\rho_{12}^2}}, \frac{T_3-\rho_{23}T_2}{\sqrt{1-\rho_{23}^2}}; \begin{pmatrix}1 & \rho_{13\cdot2} \\ \rho_{13\cdot2} & 1\end{pmatrix}\right\}$$

$$\left. + \frac{1}{m_3}\phi(T_3)I\left\{\frac{T_1-\rho_{13}T_3}{\sqrt{1-\rho_{13}^2}}, \frac{T_2-\rho_{23}T_3}{\sqrt{1-\rho_{23}^2}}; \begin{pmatrix}1 & \rho_{12\cdot3} \\ \rho_{12\cdot3} & 1\end{pmatrix}\right\}\right] \tag{2.2.22}$$

$$E(x^2) = 1 + \frac{1}{P}\left[\frac{T_1}{m_1^2}\phi(T_1)I\left\{\frac{T_2-\rho_{12}T_1}{\sqrt{1-\rho_{12}^2}}, \frac{T_3-\rho_{13}T_1}{\sqrt{1-\rho_{13}^2}}; \begin{pmatrix}1 & \rho_{23\cdot1} \\ \rho_{23\cdot1} & 1\end{pmatrix}\right\}\right.$$

$$+ \frac{T_2}{m_2^2}\phi(T_2)I\left\{\frac{T_1-\rho_{12}T_2}{\sqrt{1-\rho_{12}^2}}, \frac{T_3-\rho_{23}T_2}{\sqrt{1-\rho_{23}^2}}; \begin{pmatrix}1 & \rho_{13\cdot2} \\ \rho_{13\cdot2} & 1\end{pmatrix}\right\}$$

$$+ \frac{T_3}{m_3^2} \phi(T_3) I \left\{ \frac{T_1 - \rho_{13} T_3}{\sqrt{1-\rho_{13}^2}}, \frac{T_2 - \rho_{23} T_3}{\sqrt{1-\rho_{23}^2}}; \begin{pmatrix} 1 & \rho_{12 \cdot 3} \\ \rho_{12 \cdot 3} & 1 \end{pmatrix} \right\} $$

$$+ \frac{2 m_1 m_2 - \rho_{12}(m_1^2 + m_2^2)}{m_1^2 m_2^2 \sqrt{2\pi} \sqrt{1-\rho_{12}^2}} \phi \left\{ \left(\frac{T_1^2 + T_2^2 - 2\rho_{12} T_1 T_2}{1-\rho_{12}^2} \right)^{\frac{1}{2}} \right\} \Phi \left(\frac{m_1 m_2 m_3 \left(T_3 \sqrt{1-\rho_{12}^2} - \rho_{23} T_2 - \rho_{13} T_1 + \rho_{12}(\rho_{23} T_1 + \rho_{13} T_2) \right)}{|R|^{\frac{1}{2}} \sqrt{1-\rho_{12}^2}} \right)$$

$$+ \frac{2 m_1 m_3 - \rho_{13}(m_1^2 + m_3^2)}{m_1^2 m_3^2 \sqrt{2\pi} \sqrt{1-\rho_{13}^2}} \phi \left\{ \left(\frac{T_1^2 + T_3^2 - 2\rho_{13} T_1 T_3}{1-\rho_{13}^2} \right)^{\frac{1}{2}} \right\} \Phi \left(\frac{m_1 m_2 m_3 \left(T_2 \sqrt{1-\rho_{13}^2} - \rho_{12} T_1 - \rho_{23} T_3 + \rho_{13}(\rho_{23} T_1 + \rho_{12} T_3) \right)}{|R|^{\frac{1}{2}} \sqrt{1-\rho_{13}^2}} \right)$$

$$+ \frac{2 m_2 m_3 - \rho_{23}(m_2^2 + m_3^2)}{m_2^2 m_3^2 \sqrt{2\pi} \sqrt{1-\rho_{23}^2}} \phi \left\{ \left(\frac{T_2^2 + T_3^2 - 2\rho_{23} T_2 T_3}{1-\rho_{23}^2} \right)^{\frac{1}{2}} \right\} \Phi \left(\frac{m_1 m_2 m_3 \left(T_1 \sqrt{1-\rho_{23}^2} - \rho_{12} T_2 - \rho_{13} T_3 + \rho_{23}(\rho_{12} T_3 + \rho_{13} T_2) \right)}{|R|^{\frac{1}{2}} \sqrt{1-\rho_{23}^2}} \right) \Bigg]$$

$$(2.2.23)$$

The preceding examples not only illustrate the use of the notation, but also the pattern developing in both the expression for $E(x)$, where the only change required for additional stages is an increase in the degree of the integral I and one additional term per stage, and in the expression for $E(x^2)$, where we have a change similar to the above in the first part of the expression and an increase in the degree of I and in the number of terms (to $\binom{k}{2}$) in the second part. It is clear that the expressions for $E(x)$ and $E(x^2)$ after any number of stages can be easily obtained. In this thesis, the only equation needed in addition to those already derived is that for the gain after four stages; because of its length and its obvious form, it will not be given here.

The expression for $E(x)$ after one stage agrees with that found by

various authors previously. It is more interesting to note that the expressions for both $E(x)$ and $E(x^2)$ after two stages of selection agree with those derived by Curnow (1959) even though, once he found $M_x(t)$, he used a different method of derivation. I have not found exact expressions for the moments after three or more stages of selection anywhere else in the literature.

By using the obvious notation,

$$v^{mnp} = \frac{v_i^{mn} - o_{ip}^{mn} v_p^{mn}}{\sqrt{1 - (o_{ip}^{mn})^2}} \, ,$$

$$\rho_{ij}^{mnp} = \frac{o_{ij}^{mn} - o_{ip}^{mn} o_{jp}^{mn}}{\sqrt{1 - (o_{ip}^{mn})^2}\,\sqrt{1 - (o_{jp}^{mn})^2}} \, ,$$

$$\eta_i^{mnp}(t) = \frac{\eta_i^{mn}(t) - o_{ip}^{mn}\eta_p^{mn}(t)}{\sqrt{1 - (o_{ip}^{mn})^2}} \, ,$$

and

$$I^{mnp}(\eta^{mnp}(t); R^{mnp}) = I(\eta_1^{mnp}(t), \ldots, \eta_{m-1}^{mnp}(t), \eta_{m+1}^{mnp}(t), \ldots, \eta_{n-1}^{mnp}(t), \eta_{n+1}^{mnp}(t), \ldots$$

$$, \eta_{p-1}^{mnp}(t), \eta_{p+1}^{mnp}(t), \ldots \eta_k^{mnp}(t); R^{mnp}),$$

the third moment, $E(x^3)$, can be obtained by taking the derivative of 2.2.16 by the method of appendix 1. Higher order moments can be obtained in a similar manner. Although $\eta_i^{mnp}(t)$ and o_{ij}^{mnp} become rather complicated by the time they are simplified into expressions containing only the original values $\eta_i(t)$ and o_{ij} ($1 \le i, j \le k$), they are fairly simple in logical construction and routines for handling them could easily be written for an

electronic computer. The same is true of the expressions for higher order moments. This means that as long as $E(x)$ can be calculated, as many higher order moments as desired can also be calculated. The algebra involved in taking the required derivatives is tedious, but no more so than that involved in calculating cumulant expansions for the same moments; the resulting expressions have the advantage of being exact.

2.2.2 Calculation of the moments

Since untabulated high order integrals are involved, the calculation of the moments after selection is not free from difficulty. In order to calculate the nth moment after selection in k stages it is, for k>n, necessary to calculate a (k-n)-variate normal probability integral; fortunately if $k \leq n$ a function of $\emptyset(x)$ is all that needs to be evaluated. In addition, if

$$P = I(T_1, T_2, \ldots, T_k; R)$$

is not known for the desired cut-off points, it must also be calculated. In this thesis I have used Owen's (1956) expression for evaluation of the bivariate normal integral and Steck's (1958) expression for the trivariate normal integral.

A slight difficulty was encountered in using Steck's method in conjunction with Owen's in that the expression

$$\frac{B-A \cdot C}{A \sqrt{1-C^2}}$$

occurs in both methods but is treated slightly differently in each. The difference arises in the definition of the sign of the expression when A=0.

Steck always used a positive sign while Owen used the sign of B. Once one is aware of this difference it presents no problem. Since C is a correlation coefficient and since the troublesome expression was not used when $C = \pm 1$, the difficulty did not arise there.

Although Steck suggested a method of calculating a 4-variate normal integral, he did not give any results. For this reason the relatively inefficient method of calculating

$$I(T_1,T_2,T_3,T_4;R) = \int_{T_4}^{\infty} \emptyset(v_4) I(\frac{T_1 - \rho_{14}v_4}{\sqrt{1-\rho_{14}^2}}, \frac{T_2 - \rho_{24}v_4}{\sqrt{1-\rho_{24}^2}}, \frac{T_3 - \rho_{34}v_4}{\sqrt{1-\rho_{34}^2}}; R^4) dv_4$$

by means of Gaussian quadrature was used in the calculation of the 4-stage gain. Since only a few calculations were made for four stages this method proved adequate. Higher order integrals, although easily found by an extension of the above method or, more efficiently, by an extension of Steck's method, would be very time consuming to calculate and were not attempted.

Gaussian quadrature was always used when numerical integration was required in this thesis. This was done since it is the most efficient of the common methods of numerical integration. It is exact for a polynomial of degree 2n-1 (n being the number of points or nodes used); other methods such as Simpson's rule are exact only for a polynomial of degree n-1. Krylov (1962) gave an excellent study of this integration technique. Extensive tables, for various weighting functions, were provided by Stroud and Secrest (1966). Although one of Stroud and Secrest's tables was applicable for infinite limits of integration and a

weighting function equal to $\phi(x)$, integration using the points in this table was found to be less efficient (i.e. more points were required) than using the points for finite limits of integration and a weighting function of 1, even though slightly more calculation was involved per point. Since one or both limits of integration were in fact infinite on a number of occasions, it was necessary to choose finite upper and lower limits which were far enough from zero that the answer was, to the desired accuracy, unaffected. Comparison with known results indicated accuracy to six decimals for probabilities and at least four figures for gains. An additional advantage of Gaussian quadrature is that, when known values are not available for comparison, there are methods of calculating error bounds.

2.2.3 Calculation of the cut-off points for given values of P_i

It can be seen from the preceding discussion that, given the values T_i, it is fairly straightforward to calculate the moments after selection from a normal distribution in four stages or less. The catch is the fact that comparisons of different screening programs are only possible if both schemes end up accepting the same overall proportion of the population. This necessitates the study of screening programs that accept known proportions at each stage rather than unknown proportions based on cut-off points. Although the two methods are equivalent in the infinite case, the nature of the problem forces us to calculate the cut-off points T_i corresponding to fixed proportions P_i rather than the other way round. The absence of adequate tables for calculating T_i as a function

of P_i (except for the normal distribution) leaves us with a problem. This problem was overcome by using Newton's method to iterate to the value of η_i corresponding to a given P_i.

Let P_1 and $P_1 P_2$ be the desired proportions of the population left after one and two stages respectively. We then let

$$f_1(\eta_e) = I(\eta_e; R),$$

where η_e is an estimate of the value η_1 required to give $f_1(\eta_1) = P_1$. Similarly, we let

$$f_2(\eta_e | \eta_1) = I(\eta_1, \eta_e; R_2);$$

η_e is now an estimate of the value η_2 required to give $f_2(\eta_2 | \eta_1) = P_1 P_2$; η_1 is assumed to have been calculated by iterative methods from the earlier expression. Using the notation

$$f_1'(\eta_e) = \frac{d}{d\eta_e} I(\eta_e; R_1)$$

and

$$f_2'(\eta_e | \eta_1) = \frac{d}{d\eta_e} I(\eta_1, \eta_e; R_2),$$

we can get an improved estimate η_e' of η_1 from

$$\eta_e' = \eta_e - \frac{f_1(\eta_e) - P_1}{f_1'(\eta_e)}$$

in the usual manner. Once we have found η_1 to the desired accuracy, η_2 can be found from

$$\eta_e' = \eta_e - \frac{f_2(\eta_e | \eta_1) - P_1 P_2}{f_2'(\eta_e | \eta_1)}.$$

If desired, η_3 and η_4 can be found from similar expressions.

Since the correlations are usually relatively small in the situations studied in this thesis, the initial estimate was taken to be the easily calculated value of η_i required when all correlations between variable i and the remaining variables are zero. Prior to iteration this estimate was tested against the sufficient condition

$$\left| \frac{f(\eta_1)f''(\eta_1)}{[f'(\eta_1)]^2} \right| < 1$$

(see page 203, Scarborough, 1958) for the convergence of Newton's method; satisfaction of this condition led to an immediate commencement of iteration by Newton's method; otherwise a simple bisection procedure was used until a value satisfying the condition was found; Newton's method was then used. Although more efficient methods of iteration exist, their convergence is more sensitive to deviations of the initial estimate from the final result and, considering the relatively low accuracy required, they were not considered further.

Using this technique, the gain for a 2-stage process can be found on the KDF-9 in $\frac{1}{2}$ second, that for a 3-stage process in 6 seconds, and for a 4-stage process in 3-4 minutes (all accurate to four figures). One redeeming factor of using this rather tedious technique is that, once the cut-off points are known, moments of as high order as desired can be found very quickly. The reason for this is that the most time consuming aspect of the program is the calculation of the cut-off points. Another advantage is that the higher the order of the desired moment the lower the order of

the integrals involved and the faster the calculation.

2.2.4 More complicated selection schemes

In certain cases it may be desired to truncate at upper limits as well as lower limits at some or all of the stages. This would happen if, for instance, we wanted to select for some specific value of gain rather than the maximum value. Another example would be the case in which we wanted to accept varieties completely, not only after the kth stage, but after some intermediate stages as well. The modifications necessary when using upper cut-off points all follow naturally if $S(x)$ in 2.2.1 (and 2.2.2) is replaced by

$$S(x) = \int_{\frac{\eta_1-x}{\sigma_1}}^{\frac{\lambda_1-x}{\sigma_1}} \cdots \int_{\frac{\eta_k-x}{\sigma_k}}^{\frac{\lambda_k-x}{\sigma_k}} \exp(-\tfrac{1}{2}uR_0^{-1}u')/[(2\pi)^k|R_0|]^{\frac{1}{2}}du_1\ldots du_k;$$

the λ_i represent the upper cut-off points. The modifications necessary when using upper acceptance points are similar and will be outlined in chapter 2.5 where, in addition, some calculations will be made to indicate the advantage of such a system.

2.2.5 A functional statement of a generalized selection problem

In view of the necessity of a computer when studying more than one stage of selection, and especially of the time required on the computer for the calculation of the required values of η_i, a more straightforward formulation of the selection problem would be very useful. A general functional statement of the possible line a simplified formulation might

take will be given in this section.

If we define the distribution of x after the ith stage of selection to be $g_i(x)$, $g_0(x)$ being the original distribution of x prior to any selection, and if we define the selection function of x between the $(i-1)^{th}$ and the ith stages to be $S_i(x)$, we then have, following the method outlined earlier in this chapter, the relationship

$$g_i(x) = S_i(x)g_{i-1}(x)/P_i \; .$$

In the previous derivation we were immediately in trouble since the distribution of x after the kth stage was given by the complex expression

$$g_k(x) = \prod_{i=1}^{k} S_i(x)g_0(x)/P_i \; ,$$

and very little simplification was possible. If, however, we were able to find a family of curves approximately satisfying the general characteristics of our population (e.g. continuous, monotone increasing from zero to a hump in the middle and then monotone decreasing back to zero) such that $g_i(x)$ was simply a more extreme member of the family than $g_{i-1}(x)$, the analysis would be much simpler, especially if the family used were tabulated, or if tables could easily be calculated on a computer. An appropriate notation would be to take

$$g_i(x) = h(x; \lambda^i) \; ,$$

where λ^i represents one or more parameters defining the particular member of the family of curves appropriate to stage i. What is desired then is to choose this family so that

$$h(x; \lambda^i) = S_i(x)h(x; \lambda^{i-1})/P_i \; .$$

Although they would not be likely to work in this problem, families like the χ^2-distribution (where λ would equal the degrees of freedom) and the β-distribution (where λ would represent the two parameters p and q) are the sort of families that are needed. Apart from the fact that the values of the relevant parameters λ^i depend in some way on the values λ^{i-1} and the selection function $S_i(x)$, the distribution $h(x;\lambda^i)$ would be independent of the results at earlier stages. In all other ways the scheme would be similar to the one studied in this thesis.

If it would make calculation of $h(x;\lambda^i) = S_i(x)h(x;\lambda^{i-1})/P_i$ easier, and especially if it meant the availability of a good family of curves, a reasonable non-normal error distribution might be worth considering. In view of Davies' (1964) results with a double exponential error distribution, a non-normal error might prove to be a good guide to the situation when there is a normally distributed error. Further study of this is needed.

A solution to this more general formulation of the screening problem would not only simplify the calculations, but also the stage-by-stage study of the problem. More generally, in view of the robustness (found later in the thesis) of the area of maximum gain to the operating conditions, a study of the sort outlined in this section should be a very good guide to the results of selecting from many other distributions.

2.3 The Investigation of Selection from a Normal Distribution

2.3.1 General comments

Ideally, now that we have an expression for the gain after any number of stages, we would proceed to derive expressions for the values of the parameters γ_i and P_i necessary to achieve maximum gain for specified values of π and V. This is, however, obviously impossible with the expression derived in the last section. On the other hand, it would not be too difficult to calculate the appropriate values numerically, especially when routines have already been written for the necessary multivariate normal integrals. For given π and V we would have to find the values $\eta_1,\ldots,\eta_k$, and $\gamma_1,\ldots,\gamma_{k-1}$ (keeping in mind that $\gamma_k = 1 - \sum_{i=1}^{k-1}\gamma_i$) that, subject to the restriction that

$$\pi = I(\frac{\eta_1}{\gamma_1},\ldots,\frac{\eta_k}{\gamma_k};R),$$

would result in maximum gain. This maximization could easily be achieved by using Lagrange multipliers and a standard iterative technique such as that used by Fisher (1959) to calculate maximum likelihood estimates of unknown parameters. The appropriate values of the γ_i and the P_i could, for two or three stages of selection, be calculated for specified V and π in at most two or three minutes of computing time. The algebra necessary to write the program would be tedious but not impossible.

The many assumptions involved in the formulation of the problem (the distribution shape, the value of V, the value of π) make the calculation of the exact optimum selection scheme of relatively little value; as long as we have an idea of the approximate location of the optimum scheme, our

main concern is with the nature of the variation of gain as the α_i and
the P_i vary in the region of the optimum location. This is the approach
used by Finney (1958a, 1966) and Curnow (1961) in their studies and it
will be used here. Particular emphasis will be placed on how close
Finney's suggested symmetrical operating conditions come to achieving
maximum gain when history is used.

Prior to a detailed study of the factors that have a specific
influence on gain, it is worth noting that, when using history, the
gain is very flat in the region of optimum selection. Finney and Curnow
have already shown that this is true for the gain when historical information
is not used. The flatness of the optimum area when history is used is
illustrated in figures F1, F2 and F3* in which contours of constant gain
are plotted against α_1 and P_1 for two stage selection from a normal
distribution with $\pi = 0.01$ and $V = 0.1$, 1.0 and 10 respectively. In these
figures the gain is expressed as a percentage of the target of 2.665σ
(the maximum possible gain). The fact that the surface is even flatter
when historical information is used is indicated by the relatively large
gain as $\alpha_1 \rightarrow 1$ and $P_1 \rightarrow 1$; the large amount of information ignored in the same
area when historical information is not used results in very low gains.
As with the graph presented by Finney (1958a), the contours resemble
ellipses whose major axis is tilted upwards towards the corner that
represents light selection in stage 1 ($P_1 \rightarrow 1$) and correspondingly light

---

* Tables and figures with numbers preceded by a letter (table T1, figure F1)
  are found at the rear of the thesis; all other tables and figures are located
  in the text close to where they are cited.

replication $(\gamma_1 \rightarrow 0)$.

One other general observation worth mentioning at this point is the fact that the minimum variance of x in the selected population is usually very close in magnitude to the variance both at symmetry and at the location of maximum gain. Table T1 gives the variance of x for $\pi = 0.01$ and $V = 1$. The minimum variance of $0.227\sigma^2$ occurs at $\gamma_1 = 0.4$ and $P_1 = 0.050$. The variance at symmetry is only $0.258\sigma^2$ and that at the approximate position $(\gamma_1 = 0.6 \; P_1 = 0.07)$ of maximum gain is $0.249\sigma^2$. Obviously the small difference between these three values is of little practical consequence. Because this same general pattern was found for all values of V, π and k, and because of the difficulty of interpreting the variance of a skewed distribution, no further comments will be made.

One interesting point in passing is that the 143 values in table T1 and the corresponding gains both with and without historical information were calculated in 2 minutes. Calculation of the next three moments would not take more than a few additional seconds.

2.3.2 The influence of V on the gain

For $V = 1$ and $\pi = 0.01$, maximum gain in 2-stage selection from a normal distribution is approximately 2.353σ or 88.3% of the target of 2.665σ; this gain occurs at $\gamma_1 = 0.575$ and $P_1 = 0.08$. Since the gain at the symmetrical specifications is 2.350σ or only about 0.1% lower*, it

---

* Percentage increases are given by $100(L-S)/S$ and percentage decreases by $100(L-S)/L$, where L represents the larger value and S the smaller.

can be seen that little will be lost in using symmetry as an approximation
to the optimum specifications. Figure F2 illustrates the wide range of
operating conditions resulting in a gain very close to the maximum
attainable gain. It can, for instance, be seen from this graph that
operation in the region $0.3 < \gamma_1 < 0.8$ and $0.04 < P_1 < 0.2$ results in a gain of
more than 86% of target or within about 2% of the maximum. It can also be
seen that, if it is necessary to move from symmetry, it is advisable
(when $V = 1$) to move in the direction of both slightly heavier replication
and slightly heavier selection in stage one. Figure F1 for $V = 0.1$ and
$\pi = 0.01$ displays the same characteristics and an even flatter surface.
In particular, the maximum gain of 2.647σ occurs at $\gamma_1 = 0.55$ and
$P_1 = 0.05$ and is only 0.2% greater than the gain at symmetry. Obviously
as $V \to 0$ the surface will become perfectly flat with a gain of 100% of
target.

It is only as V gets large that the maximum gain departs much in
magnitude from the gain at symmetry. Figure F3 showing the gain for $V = 10$
and $\pi = 0.01$ exhibits the characteristically flat surface, but this time
the ridge of maximum gain occurs for lower values of γ_1 and slightly lower
values of P_1. In addition, the surface has been pushed down considerably
in magnitude: the maximum is 88.3% for $V = 1$ and only 53.5% for $V = 10$.
The maximum gain now occurs at $\gamma_1 = 0.35$ and $P_1 = 0.07$ and is 4.2% higher
than the gain at the symmetrical specification.

The entries in table 2.3.1 indicate that the same general pattern of
behaviour with variation in V occurs over a wide range of values of π.
The figures in the table suggest that the percentage advantage of maximum

gain over symmetrical gain increases slightly as V decreases from 1, but it is only for increasing V that there is a marked advantage.

TABLE 2.3.1

A comparison of the 2-stage gain at the symmetrical operating conditions with the maximum gain for various V and π.

π	V	G/σ at symmetry $\gamma_1 = 0.5 \; P_1 = \sqrt{\pi}$	Approximate maximum G/σ		
			% increase over symmetry	Location γ_1	P_1
0.0625	0	1.968	–	–	–
	1	1.605	0.1	0.5	0.23
	5	1.039	1.1	0.5	0.21
	60	0.349	4.8	0.3	0.2
0.01	0	2.665	–	–	–
	1	2.350	0.1	0.575	0.08
	5	1.706	0.8	0.4	0.1
	60	0.647	17.5	0.2	0.044
0.001	0	3.367	–	–	–
	1	3.123	0.5	0.65	0.03
	5	2.321	0.0	0.5	0.032
	60	1.229	20.3	0.15	0.014
0.0001	0	3.959	–	–	–
	1	3.753	1.3	0.7	0.014
	5	3.230	0.8	0.55	0.014
	60	2.029	9.0	0.2	0.006

In view of all the assumptions made in the formulation of the problem, the increases in gain of one percent or less achieved by moving from symmetry when $V \leq 5$ will not be worth considering in practice so we should usually be very safe in operating at the symmetrical specifications; only for very large values of V will it be necessary to consider other operating conditions. This agrees with a suggestion made by Curnow (1961) for the

case when historical information is not used. For large V it appears that we should use smaller than symmetrical values of both γ_1 and P_1.

Figures F4, F5 and F6 illustrate the dependence of gain (G/σ) on $\log(V)$ $(0.1 \leq V \leq 10,000)$ at the symmetrical specifications for $\pi = 0.0625$, 0.01, and 0.0001 respectively. It can be seen that very nearly perfect selection is achieved for all values of π when $V = 0.1$. This is followed by an almost linear drop in mid range to very nearly no gain when $V = 10,000$. Obviously the gain will be 0 in the limit as $V \to \infty$ and 100% of target in the limit as $V \to 0$.

If it is assumed that all additional resources go to increasing the number of replicates per variety, these graphs can be used to study the effect of increasing or decreasing the total resources A while maintaining the same values γ_i $(1 \leq i \leq k)$. Doubling the value of A would have the effect of doubling the replication and halving V. It can be seen from figure F4 for $\pi = 0.0625$ that, in moving from $V = 2$ to $V = 1$, we get an increase in gain from approximately 1.38σ to 1.6σ or about 16%; doubling replication should result in approximately this increase in gain. This technique should be useful when trying to balance the cost of additional experimentation with the benefit of that experimentation.

2.3.3 The influence of π on the gain

Returning to figure F1 and table 2.3.1, it can be seen that the surface of gain remains very flat for various values of π; it is only when V is large that the maximum gain departs very much in magnitude from the gain at symmetry. For a given V it appears that there may be a slight reduction

in this effect as π decreases to very small values: for $V = 60$ the optimum is only 9.6% better than the symmetrical gain for $\pi = 0.0001$ while it is 17.5% and 20.3% better for $\pi = 0.01$ and $\pi = 0.001$ respectively. In the more realistic range $0.2 \leq V \leq 5$ the differences between optimum and symmetrical gain are much smaller. At $\pi = 0.25$, for example, the symmetrical gain is within 0.1% of the maximum and is located very close to symmetry. As π decreases, the location of the optimum begins to move slowly to values of α_1 and P_1 larger than the symmetrical values. At $\pi = 0.0001$ and $V = 1$, for instance, the maximum gain occurs at $\alpha_1 = 0.7$ and $P_1 = 0.014$; the surface is still, however, very flat and the symmetrical gain is within 1.3% of the maximum. Clearly, the gain at the symmetrical specifications is very close to the maximum gain over a wide range of values of π.

Since, for a given number of stages, a specific value of V ensures that the first stage replication is constant independent of π we have, since $\alpha_1 = \frac{1}{k}$ at symmetry,

$$e_1^2 = Vk\sigma^2$$

independent of π. At a later stage (stage j say) we have

$$e_j^2 = Vk\sigma^2\pi_{j-1} \; ,$$

where $\pi_{j-1} = (\pi)^{(j-1)/k}$. It can be seen that, in all stages but the first, the variance becomes smaller as the selection becomes more intense. The reason for this is the fact that, as a result of the very intense selection, there will be relatively fewer varieties around at each stage after the first and, since the resources are fixed independent of π, there will be more resources per variety. As a result of the increased accuracy in the

estimation of the true yields of the individual varieties for small values of π, we can expect, for a given V, gains closer to target. This increased accuracy may also be the reason for the observed increase in the optimum values of both α_1 and P_1 over the values found for higher values of π. It can be seen from the above equation for e_j' that, as π decreases, the later the stage the greater the increase in accuracy; it seems logical that it might be preferable to spread this increase in accuracy more evenly over all stages, and to keep more varieties around longer before discarding them; this can be done by increasing α_i and P_i in the early stages. As π decreases for constant V we might then expect operation at the location

$$P_1 > P_2 > \ldots > P_k$$

and

$$\alpha_1 > \alpha_2 > \ldots > \alpha_k$$

to result in improved gain. This effect can be seen to occur for low values of V in table 2.3.1. At higher values of V the effect is less obvious because of the overpowering effect of V.

The improved precision for small values of π is illustrated in figures F4, F5 and F6 where it can be seen that, for a fixed V, the percent of target achieved increases as π decreases. For 2-stage selection at V = 60, for instance, we achieve 4.4% of target for π = 0.0625 and 16.2% of target for π = 0.0001. Obviously in the limit as $\pi \to 1$ we will approach 100% of target again since the target approaches zero. The same effect is illustrated more vividly in figures F7, F8 and F9 where the gain is plotted against π

for V = 0.2, 1, and 5 respectively. From figure F8 we have, for 2-stage selection, 79.5% of target achieved at π = 0.1, 88.2% at π = 0.01, 92.8% at π = 0.001 and 94.8% at π = 0.0001.

As can be seen from the graphs, the effect described above does not occur for 1-stage selection. In fact, for 1-stage selection with a given value of V and no initial random discard, the gain achieved is a fixed percentage of the target independent of π. This can be seen by calculating the gain from equation 2.2.18 and dividing by the target (given by the same equation with V = 0). The result is that, for one stage of selection and P_0 = 1,

$$\frac{\text{Gain}}{\text{Target}} \;=\; \frac{1}{m_1} \;=\; \frac{1}{\sqrt{1+V\sigma^2}} \quad .$$

This is a function only of V. When $P_0 \neq 1$ the simplification is not possible.

2.3.4 Multiple-stage screening and its effect on gain

In the preceding sections very little reference has been made to the number of stages involved; the conclusions on the flatness of the optimum area and the adequacy of the symmetrical specifications have, in fact, been based on only 2-stage selection. In view of this, of the remark by Finney (1958a), "In the practice of plant selection reliance on a single stage of selection is unlikely", and of the remarks of various authors (see the comments of Finney (1966), King (1963) and Davies (1958) in section 1.4.1) to the effect that screening in more than two or three stages is unlikely to be of much value, it will be interesting to look at multiple-stage

screening in some detail.

The flatness of the optimal area for 2-stage selection has been established. Curnow (1961), using 53 combinations of the α_i's and P_i's around the symmetrical specifications, found the surface to be very flat for 3-stage selection without history for $\pi = 0.001$ and $V = 1, 4,$ and 10. In each of these three cases the maximum gain was less than 0.7% greater than the symmetrical gain.

As Curnow pointed out, it is very difficult to discuss the shape of the surface of the gain for three stages of selection. Solution for the exact location of the maximum gain would be more valuable here (because of the complexity of the surface) but would, although possible, involve numerical evaluation both of the five parameters $\pi_1, \pi_2, \pi_3, \alpha_1$ and α_2, and of the Lagrange multiplier required to incorporate the restriction on the proportion accepted.

Because of the difficulty of exact evaluation, a number of surveys of a similar nature to those made by Curnow have been performed for 3-stage selection when history is included. Table T2 gives the gain for 3-stage selection with $V = 5$ for both $\pi = 0.0001$ (the upper table) and $\pi = 0.01$ (the lower table). These tables are set up in a manner similar to that used by Curnow in that selection becomes more intense in the early stages as you move down in the table, and replication becomes heavier in the early stages as you move from left to right across the table. In addition, the average yield for each set of the P's or α's is included at the end of the appropriate row or column. Since the parameters range over a wide area, and since the minimum tabulated gain is, in both cases, more than 85% of

the maximum tabulated gain (85.6% for $\pi = 0.01$ and 86.7% for $\pi = 0.0001$),
it can be seen that the surface is quite flat in the optimum area. Since
the maximum gain occurs at symmetry for $\pi = 0.01$, and at $P_1 = P_2 = P_3 = \pi^{1/3}$
and $\gamma_1 = \frac{1}{2}$, $\gamma_2 = \gamma_3 = \frac{1}{4}$ for $\pi = 0.0001$ (resulting in an increase over
symmetrical gain of only 0.9%), the symmetrical specifications again seem

TABLE 2.3.2

Values of G/σ for 4-stage symmetrical selection from a normal
distribution.

π	Target	Values of V		
		0.2	1	5
0.0001	3.959	3.958	3.877	3.601
0.001	3.367	3.359	3.258	2.886
0.01	2.665	2.640	2.472	1.946
0.0625	1.968	1.913	1.677	1.138

to be adequate. These conclusions are in keeping with Curnow's results and
other tables (V = 1 with $\pi = 0.0625$, 0.01 and 0.001) calculated by myself
but not, for the sake of brevity, included here. Since the best results
occur in the upper right hand corner for $\pi = 0.0001$ (in keeping with the
comments made for very intensive selection in section 2.3.3) and in the
centre to lower left hand corner for $\pi = 0.01$, no general conclusion can
be drawn on the movement of the optimum area as a function of k.

Because it is far more time consuming to find the 4-stage gain than
the 3-stage gain, far fewer results have been calculated. Some of these
results are reproduced in table 2.3.2. We are so close to perfect
selection at V = 0.2 that little can be expected to be gained by moving

from the symmetrical specifications. This is not so obvious for $V \geq 1$.

In order to get a (very) rough idea whether or not the gain acts in the same manner in 4-stage selection as it did in 2-stage and 3-stage selection, a few values of the parameters other than the symmetrical specifications

TABLE 2.3.3

Values of G/σ for 4-stage selection from a normal distribution with $V = 5$.

π	α_1	α_2	α_3	α_4	P_1	P_2	P_3	P_4	G/σ
0.0001	0.333	0.280	0.230	0.157	0.120	0.111	0.090	0.083	3.644
	0.300	0.277	0.233	0.190	0.120	0.111	0.090	0.083	3.639
	0.333	0.280	0.230	0.157	0.100	0.100	0.100	0.100	3.632
	0.300	0.277	0.233	0.190	0.100	0.100	0.100	0.100	3.624
	0.250	0.250	0.250	0.250	0.100	0.100	0.100	0.100	3.601
	0.200	0.233	0.277	0.290	0.083	0.090	0.111	0.120	3.515
0.01	0.300	0.277	0.233	0.190	0.450	0.380	0.263	0.222	1.857
	0.250	0.250	0.250	0.250	0.215	0.215	0.215	0.215	1.946
	0.200	0.233	0.277	0.290	0.222	0.263	0.380	0.450	1.934

were tried for $\pi = 0.01$ and $\pi = 0.0001$ with $V = 5$. These values are presented in table 2.3.3. Obviously no general conclusions can be drawn from these results but it is interesting to note that, as with 2-stage and 3-stage procedures when selection is very intense, better gains are obtained for $\pi = 0.0001$ when there is heavier than symmetrical replication and slightly less intense than symmetrical selection in the early stages. The results also suggest that the maximum gain for $\pi = 0.01$ will occur at symmetry or with slightly lighter replication and possibly slightly heavier selection in the early stages. This is again very similar to the result for 3-stage

selection. There is certainly nothing to suggest radical departure from the results of 3-stage selection, nor is there any general pattern emerging that would indicate any variation in the location of maximum gain with k; what changes do occur seem to be the usual functions of V and π (as discussed in section 2.3.2 and 2.3.3). In general, unless V is very large or π very small, symmetry still seems to be an extremely good guide to obtaining nearly maximum gain for 4-stage selection.

TABLE 2.3.4

The percent improvement over one less stage of selection for various values of π and V.

π: Number of stages:	0.0625			0.01			0.001			0.0001		
	2	3	4	2	3	4	2	3	4	2	3	4
V												
0.2	5.2	0.9	0.4	8.2	0.8	0.3	9.2	0.4	0.1	9.4	0.2	0.1
1.0	15.3	3.2	1.3	29.4	3.2	1.4	35.5	3.2	1.0	37.4	2.5	0.8
5.0	29.3	6.5	2.8	72.3	9.8	3.8	103.2	9.9	3.3	117.4	8.7	2.5

It can be seen from the graphs already studied (figures F4 - F9) that the pattern of variation of gain at symmetry with V and π for 3-stage and 4-stage screening is the same as that for 1-stage and 2-stage screening; only the actual value of using additional stages varies. Table 2.3.4 indicates the percentage improvement over one less stage of selection for various values of π and V at symmetry. From this table and from the graphs just mentioned it can be seen that, as is to be expected, the value of each additional stage decreases as k increases. It seems unlikely that it will be beneficial to use values of k much greater than 4. Although the

percentage increase due to additional stages increases with V in the range
shown, the actual magnitude of the increase can be seen in figures F4, F5
and F6 to start to decrease at around 100<V<1000 depending on the value of
π. For V<1, 4 stages and probably even 3 stages will rarely be worthwhile.
Information on the magnitude of the gain due to additional stages for the
range 1>π>0.0001 can be obtained from figures F7, F8 and F9. More than
2-stage selection can be seen to be rarely necessary for π>0.1; it is in
the range π<0.1 that larger values of k will be most useful, especially if
V is large.

In all this discussion it should, however, be remembered that the
real value of additional stages will depend on the economics of the situation;
since a profitable return is often a matter of a one or two percent improvement,
extra stages may often be exactly what is required. On the other hand, the
delay involved may be prohibitive.

2.3.5 The value of the use of historical information

The work done in the previous three sections has, with the exception
of the 4-stage calculations, been very similar to the studies done by Finney
(1958a) and Curnow (1961). There is one major difference: historical
information has been used throughout. Little, however, has been said about
the effect of historical information on the size and variation of the gain.
Obviously, since the conclusions are very similar to those made by Finney
and Curnow, the gain changes very little in the optimum area.

Table T3, giving the percentage increase in gain when selecting in two
stages with π = 0.01 and V = 1, indicates that, in the optimum area, history

is of little value (0.2% at symmetry); it is only when α_1 and P_1 are
large that history is of much value (27.3% when α_1 = 0.8 and P_1 = 0.5).
The reason for this is obvious on looking at the ratio of the current
weight to the weight applied to the information from previous stages.
In the case being considered here we have, from equation 2.1.1,

$$\frac{a_{22}}{a_{21}} = \frac{e_1^2}{e_2^2} = \frac{\alpha_2}{\alpha_1 P_1} \; ;$$

since π = 0.01 this means that, at symmetry, a_{22}/a_{21} = 10.0. When the
estimate of yield from the second stage is 10 times as accurate as that
from the first, it is obvious that little improvement will result from
the inclusion of the first stage information in the estimate z_2 of x.
At α_1 = 0.8 and P_1 = 0.5, however, a_{22}/a_{21} = 0.5 and, as has already been
seen, a 27.3% increase in yield results from the use of the first stage
information.

Table 2.3.5 gives a summary of the value of history for 3-stage
selection with V = 1 for both π = 0.001 and π = 0.0625. The lower third
of the table gives, for π = 0.0625, the ratio of the weight a_{33} (applied
in the third stage to the third stage information), to the weights a_{32} and
a_{31} (applied in the same stage to the second and first stage information
respectively). The same behavior found with 2-stage selection can be seen
to occur: heavy replication and light selection in early stages cause more
nearly equal weights and a corresponding increase in the importance of
historical information. When history is used in this region the gain levels
off for all values of V and π and the surface becomes fairly flat. This
can be contrasted with the fact that, when history is not used, the gain

approaches zero as $P_1 \to 1$ and $\alpha_1 \to 1$.

<u>TABLE 2.3.5</u>

The percent increase in gain due to the use of historical information when selecting in 3 stages with V = 1. The ratios of the weights applied to historical information in the third stage are (for π = 0.0625) given in the lower third of the table.

	Values of				Values of $(\alpha_1 \alpha_2 \alpha_3)$		
π	P_1	P_2	P_3	% increase in gain	(1/4 1/4 1/2)	(1/3 1/3 1/3)	(2/5 2/5 1/5)
0.001	0.14	0.14	0.051		0.1	0.3	0.7
	0.10	0.10	0.10		0.05	0.1	0.3
	0.071	0.071	0.196		0.03	0.04	0.1
0.0625	0.5	0.5	0.25		2.4	4.5	8.5
	0.391	0.4	0.4		1.4	2.2	3.9
	0.281	0.389	0.571		0.8	1.2	2.1

				ratio of weights	$\dfrac{a_{33}}{a_{31}}$ $\dfrac{a_{33}}{a_{32}}$	$\dfrac{a_{33}}{a_{31}}$ $\dfrac{a_{33}}{a_{32}}$	$\dfrac{a_{33}}{a_{31}}$ $\dfrac{a_{33}}{a_{32}}$
0.0625	0.5	0.5	0.25		8.0 4.0	4.0 2.0	2.0 1.0
	0.391	0.4	0.4		12.8 5.1	6.4 2.6	3.2 1.3
	0.281	0.389	0.571		18.3 7.1	9.2 3.6	4.6 1.8

The value of the inclusion of history varies considerably with π. When operating at symmetry, the ratio of the weight applied to the yield of the final stage k, to the weight applied to the yield of any other stage j, is given by

$$\frac{a_{kk}}{a_{kj}} = \frac{e_j^2}{e_k^2} = \frac{1}{(\pi)^{(k-j)/k}} \quad,$$

where, of course, k > j. For any possible value of j this ratio increases rapidly as selection becomes more intense (π decreases) and, as a result,

history decreases rapidly in importance. This effect can be seen for V = 5 in figure F10. In 3-stage selection this figure shows history to increase the gain by 5.1% for π = 0.1 but only 0.3% for π = 0.001.

The above expression also illustrates the fact that the weights applied to intermediate stages are more nearly equal for large values of k. As a result, at symmetry, the importance of historical information increases as the number of stages increases. The resultant increase in the value of history with k can be seen for fixed V in figure F10, and for fixed π in figure F11. At V = 5 and π = 0.1, for example, only 2.3% is gained in 2 stages while 5.1% is gained in 3 stages.

The importance of history also increases as V increases. Figure F11 for π = 0.0625 shows that its value increases rapidly in the range 0.01<V<10 and then flattens out as V increases further. The same effect is present, to a lesser degree, for smaller values of π.

Since the inclusion of history always gives some additional gain it should, if the correct weights are known accurately and unless the administrative cost is high, always be included in estimating x. When the weights are not known, or if they can only be estimated with a few degrees of freedom, problems arise. Yates and Cochran have suggested that equal weighting may be preferable if this is the case. Because of the large increase in accuracy in later stages in the sort of screening procedures considered here it will, however, probably be preferable to ignore historical information completely rather than weight it equally. Further comments on weighting will be made in section 2.5.1.

2.3.6 The use of an initial random discard

Finney (1958a) introduced the idea of an initial random discard represented by $P_0 < 1$. Although this resulted in large increases in gain for $k = 1$, he found much less improvement when $k = 2$. In that case he suggested that an initial random discard would not be worthwhile unless V was very large or π very small.

The reason for the advantage, in some instances, of an initial random discard is that, by discarding some varieties without experimentation, more resources are made available for each of the remaining varieties with the result that their yields can be determined more accurately. This increased accuracy will often more than offset the fact that a few good varieties have been discarded and, when this happens, increased gains will occur. Mathematically the effect is to change V to $V' = P_0 V$ and π to $\pi' = \dfrac{\pi}{P_0}$. If additional varieties can be assumed to be available and drawn randomly from the same parent distribution, this idea can be extended to take into account $P_0 > 1$. This would represent a situation in which the resources available could be used more efficiently by experimenting less intensely on more varieties. In this case, the increased probability of there being really high yielding varieties in the population more than offsets the decrease in accuracy.

When increased gain can be achieved by randomly discarding some varieties (optimum $P_0 < 1$) there is obviously an inefficiency in the system in that $1 - P_0$ of the varieties have been developed in vain. It may, on the other hand, be impossible for economic or other practical reasons to come up with more varieties in order to take advantage of the increase in gain

that will result when optimum $P_0>1$. Despite these difficulties, a brief discussion of this aspect of screening will, for sake of completeness, be presented here.

Since the only change is in V and π, the earlier remarks on flatness and optimality will apply. As a result, all calculations in this section will be done at symmetry. This means that we will use $P_i' = (\pi')^{1/k} = (\frac{\pi}{P_0})^{1/k}$.

Figure F12 gives, as a function of πV, the value of P_0 that will result in maximum gain. It can, for example, be seen that if $\pi = 0.002$ and $V = 5$, maximum gain occurs for $P_0 = 0.1$, 0.35 and 0.7 for 1-, 2- and 3-stage selection respectively; for $V = 1$ and $0.0001<\pi<1$, maximum gain occurs for $P_0>1$ for both 3-stage and 2-stage screening. The actual gains achieved for $P_0 \neq 1$ for more than 1-stage screening can be seen, from table 2.3.6, to be small unless π is large. Since the expected gain is 0 when $\pi = 1$, taking $P_0>1$ and $\pi' = \frac{\pi}{P_0}$ will always result in an infinite increase in gain there.

The actual gain achieved when using optimum P_0 is given in figure F13 as a function of πV. The great advantage of varying the initial number of varieties ($P_0 \neq 1$) in 1-stage selection is emphasized by how much closer the curve for $k = 1$ is to the curves for $k = 2$ and $k = 3$ in figure F13 than it was for $P_0 = 1$ in, for instance, figure F8.

As Finney suggested, it is for $k = 1$ that increasing or decreasing the initial number of varieties will be of most value. The difficulty of balancing plots and varieties for small changes in the number of varieties, and the difficulty of developing new varieties will quickly negate any advantage that might arise in most instances when $k \geq 2$. The larger gains

achieved by using an initial random discard when $k \geq 2$ and V is very large can be explained by the fact that, as we have already seen, the optimum moves from symmetry both in location and in magnitude when V is large. Little increased gain results from an initial random discard if we move to the region of optimum gain when V is large.

TABLE 2.3.6

The recommended value of P_0 and the resulting increase in gain for 1-, 2- and 3-stage symmetrical selection with V = 1 and V = 5.

Stages:		1		2		3	
Value of V:		1	5	1	5	1	5
π 0.0001	P_0	0.2	0.4	1	0.3	6.3	1
	% inc.	23	89	0	3	1.6	0
0.001	P_0	0.2	0.07	1.3	0.3	4.6	1
	% inc.	20	67	0.2	3	2.9	0
0.01	P_0	0.45	0.16	1.7	0.4	3.7	1
	% inc.	8.8	33	1	2.6	4.1	0
0.1	P_0	1	0.5	2.5	1.1	4.3	1.6
	% inc.	0	5	6	0.3	10.3	3

2.3.7 Other descriptive and comparative techniques

In the preceding sections we have been concerned with the gain, in particular with its maximum value and its variation in the area of that maximum. We now turn to other descriptive and comparative techniques that can usefully be employed in the study of screening programs. Finney (1964)

used both operating characteristic (OC) curves (giving the probability of
selection of a variety as a function of its true yield x) and the frequency
distribution of the selected varieties, to illustrate the operation of
various screening systems. Figure F14 (top) gives both these curves for
selection in two and three stages from a normal parent population at the
symmetrical specifications for V = 1 and π = 0.001. Perfect selection for
π = 0.001 would mean selection of all varieties with true yield greater than
x = 3.090σ; this value is represented by a small vertical arrow in the figures.

The operating characteristic S(x) (see equation 2.2.1) is plotted in
the upper left hand corner of figure F14. From it we can see that both
2-stage and 3-stage screening result in a very steep rise in the probability
of selection in the area of x = 3.09σ, but that 3-stage screening results
in a rather steeper rise. This is illustrated by the fact that at x = 2.7σ
the 2-stage procedure will accept 6% of all varieties and the 3-stage only
2%, while at x = 3.5σ the 2-stage procedure accepts only 57% of all varieties
while the 3-stage accepts 70%. It is interesting to note that at about
x = 6.25σ the probability of selection under the 3-stage system drops below
that in the 2-stage system. Since both systems accept over 98% of
all varieties with true yields in this region this drop will not usually
be of much practical importance. Finney found the same rather curious
result in comparing 1-stage (P_0 = 1) and 2-stage systems. Inspection of the
frequency distribution (g(x)) after selection (right hand corner of figure
F14) confirms the advantages of 3-stage screening over 2-stage screening.
With 3-stage selection g(x) is more highly peaked and it falls off much
more rapidly as x decreases from 3.09σ.

Both the graphs described in the previous paragraph are very useful if the experimenter is particularly interested in a special range of values of x. From the OC-curve it can be seen that we are better off in the region $3\sigma < x < 5\sigma$ with 3-stage screening; if, however, we are particularly interested in values of $x > 5\sigma$, and if there is a real possibility of getting such high values, it may not be worth our going to 3 stages; if we only want varieties for which $x > 6.25\sigma$, 3-stage screening is actually the poorer of the two. Referring to the frequency distribution $g(x)$, we can see that if, for instance, items yielding less than 2.5σ result in a loss, we will be much better off using 3-stage screening. Obviously these two graphical techniques are very useful in describing a given system.

A somewhat more quantitative method of comparing two systems is to take the ratio

$$ROC = \frac{g_3(x)}{g_2(x)} = \frac{S_3(x)h(x)}{S_2(x)h(x)} = \frac{S_3(x)}{S_2(x)}$$

(the subscripts refer to the number of stages used). This relative operating characteristic (ROC) curve provides, for a given parent population $h(x)$, the relative probability of selecting a variety with a given true yield x. If the ROC value is greater than 1, the 3-stage system (in this case) is more likely to accept a variety with the yield in question; if less than 1, it is less likely to accept it. A graph of this function on a log scale for the situation under consideration ($\pi = 0.001$, $V = 1$, at symmetry) is presented in the lower part of figure F14. Since neither this graph nor, for that matter, the OC-curve specifically take into account the parent distribution of the varieties under consideration, a graph of the relevant

tail of the normal distribution is plotted in the same figure against the same scale of x/σ.

It can be seen from this graph that the 3-stage system is only slightly better, relatively speaking, than the 2-stage at selecting varieties with yields greater than 3.09σ. At $x = 3.2\sigma$, for example, the 3-stage system is only 1.33 times more likely to accept a variety than the 2-stage system. It is also slightly more likely to accept the presumably undesirable items in the range $2.93\sigma < x < 3.09\sigma$. It is the much greater probability of rejection by the 3-stage system in the range $x < 2.93\sigma$ that makes it the better of the two. At $x = 2.75\sigma$ the 3-stage system is twice as likely to reject an item ~~(half as likely to accept it)~~. The extremely rapid increase in the relative probability of rejection as x decreases is emphasized by the fact (not shown on the graph) that at $x = 2.3\sigma$ the 3-stage system is 200 times as likely to reject an item. The large probability of there being a variety in the region $x \leq 2.3\sigma$ (relative to $\Pr(x > 2.9\sigma)$), as illustrated by the graph of $N(0,1)$ on the same scale, emphasizes the desirability of this characteristic. The sharp drop in $g(x)$ after three stages as x decreases from 3.09σ is a result of this rapid decrease in the relative probability of selection. This aspect of the difference between the two systems is not as well illustrated by just the OC-curve. In fact, the reliance on the vertical difference between the two OC-curves for comparison may be misleading because of the very small vertical difference in the important lower range of x and the large vertical difference in the upper.

For the sake of better illustrating the nature of the improvement due

to history, the ROC-curve of the ratio of the probability of selection when history is included versus the probability when it is not is given in figure F15 for 3-stage selection at symmetry when $\pi = 0.0625$ and $V = 1$. The relevant tail of the normal distribution is again, for convenience, plotted on the same scale of x/σ. The relatively small value of the inclusion of history is well illustrated by the slow increase in the relative probability of rejection for x less than the target of 1.534σ. For this comparison the ROC-curve is particularly useful since the OC-curves and the two curves of $g(x)$ are almost indistinguishable on a reasonably sized graph.

It has already been mentioned that it is necessary to consider the distribution of the parent population when discussing OC-curves and ROC-curves and, for this reason, a graph of the parent population is included with both ROC-curves. The necessity for this consideration is aptly illustrated by the fact that a selection system which is 100 times as good as another for $x \geq x_0$ is of little value if $P(x \geq x_0) = 10^{-10}$. Another slightly more subtle difficulty arises from the fact that the cut-off points used in evaluating the OC-curves are calculated to accept fixed proportions from a specified parent distribution and will vary considerably as a function of the parent distribution. As a result, the OC-curve, which at first glance appears to depend only on the error distribution, is a function of the parent distribution. If, on the other hand, one has determined desirable cut-off points for some other reason, or wants to find cut-off points to give an almost vertical rise in the probability of selection in the region of some specified value of x, these curves are particularly useful in that, for specified η_i, they describe the operation

of the selection scheme independently of the parent population. The proportions accepted will, of course, vary as a function of the cut-off points and the parent population. In general though, whether or not the distribution of the parent population is known, the use of these three curves will add greatly to the understanding of the operation of a selection program, especially if more knowledge is desired than just the gain.

2.4 Modified Selection Schemes

2.4.1 Acceptance prior to the final stage

One of the conclusions of the previous chapter is that two or more stages are usually necessary for effective screening. One obvious drawback of multistage varietal screening is the long delay involved in passing through the k stages of the program. A possible improvement over the basic system would be to accept a few of the most promising varieties immediately after the first stage of experimentation rather than wait until the final stage. This would not only allow certain varieties to get into general use earlier but would also allow more intensive study of the remaining varieties. As has already been pointed out in the discussion of OC-curves, the probability (when $V = 1$ and $\pi = 0.01$) of selecting a variety with a very high yield ($x \geq 5.9\sigma$; Finney, 1964) is greater for 1-stage selection than for 2-stage. The unrealistically high value of x at which this higher probability of selection occurs is not too promising but it does indicate the theoretical possibility of an advantage in gain arising out of acceptance of unusually high yielding varieties prior to the final stage. It should also be remembered that getting an improved variety into general usage one or more years early may balance even a small decrease in the theoretically expected overall gain.

In order to investigate the result of early acceptance, we will look briefly at the result of accepting a small proportion of varieties immediately after the first of two stages of selection from a normal distribution. If we take η_u as the first stage acceptance point and if we reject, as before, all varieties yielding below η_1 in the first stage or

η_2 in the second we have

$$S(x) = \Pr(y_1 > \eta_u | x) + \Pr(\eta_u \geq y_1 > \eta_1, y_2 > \eta_2 | x) \ .$$

Following exactly the same procedure used in chapter 2.2, and letting

$$I(\eta_1 : \eta_u, \eta_2 ; \rho) = \int_{\eta_1}^{\eta_u} \int_{\eta_2}^{\infty} \exp(-\tfrac{1}{2}(u^2 + v^2 - 2\rho uv)/(1-\rho^2))/(2\pi \sqrt{1-\rho^2}) \, du \, dv \ ,$$

we have

$$E(x) = \frac{d}{dt} \left[\frac{e^{\frac{1}{2}t^2}}{P} \left\{ \int_{\frac{\eta_u - t}{\omega_1}}^{\infty} \phi(u) \, du + I\left(\frac{\eta_1 - t}{\omega_1} : \frac{\eta_u - t}{\omega_1}, \frac{\eta_2 - t}{\omega_2} ; \rho \right) \right\} \right] \Bigg|_{t=0} \ ;$$

taking the derivative and simplifying we have

$$E(x) = \frac{1}{P} \left[\frac{1}{\omega_1} \phi(T_1) \Phi[(T_2 - \rho T_1)/\sqrt{1-\rho^2}] + \frac{1}{\omega_2} \phi(T_2) \Phi[(T_1 - \rho T_2)/\sqrt{1-\rho^2}] \right.$$

$$\left. + \frac{1}{\omega_1} \phi(T_u) \Phi[(\rho T_u - T_2)/\sqrt{1-\rho^2}] - \frac{1}{\omega_2} \phi(T_2) \Phi[(T_u - \rho T_2)/\sqrt{1-\rho^2}] \right], \quad (2.4.1)$$

where $T_u = \eta_u / \omega_1$. At first glance, this expression appears to be the gain after two stages (for a given value of T_2 and without an upper acceptance point) plus a correction factor. That this is not true arises from the fact that it is now unnecessary to study some of the varieties (all those yielding above η_u) that would normally have been studied in the second stage. This causes more resources to be available for each of the remaining varieties in stage 2 and e_2^2 is reduced to

$$e_2^2 = \frac{V\sigma^2}{\alpha_2} (P_1 - P_u) \ ,$$

where $P_u = \Phi(T_u)$. Obviously the system only makes sense if $P_u < \pi$ (note that

$P_1 > \pi$). Since e_2^g changes, $T_2 = \eta_2/m_2$ must also change if it is still desired to accept a fixed proportion of the varieties. As a result of these changes, the first two terms in the above expression are quite different from the corresponding terms without upper acceptance.

In order to study the advantage of early acceptance, the gain has been calculated for a number of values of π using various values of η_u calculated

TABLE 2.4.1

Values of G/σ for 2-stage symmetrical selection showing the advantage of accepting some varieties at stage 1.

V	π	G/σ at $P_u = 0$	Maximum gain P_u	G/σ	Maximum P_u without loss P_u	G/σ
1	0.5	0.573	0.3	0.586	0.375	0.573
1	0.25	0.956	0.05	0.962	0.1	0.956
2	0.125	1.563	0.03	1.565	0.04	1.563
1	0.125	1.292	0.003	1.293	0.003	1.293
5	0.125	0.803	0	0.803	0.003	0.803
1	0.0625	1.605	0	1.605	0.001	1.605
1	0.01	2.350	loss in gain even at $P_u = 0.0001$			

for specific values of P_u. The calculations were made for the same value of η_1 appropriate for symmetrical screening without early acceptance; other values might give slightly higher gains but these were not studied. The results of these computations are given in table 2.4.1. In this table both the gain for $P_u = 0$ (no acceptance prior to stage 2) and the maximum gain are given. The largest value of P_u for which the gain is still approximately as large as the symmetrical gain for $P_u = 0$ is also given.

It can be seen from the table that it is only for very large values of π that any increase in gain is achieved by early acceptance and even then the maximum improvement is only about 2.3% at $\pi = 0.5$. Virtually no improvement occurs at $\pi = 0.125$.

The calculations for $V = 0.2$ at $\pi = 0.125$ suggest that a slight improvement may occur as V decreases.

Since the first and second stage error variances are very nearly equal for large values of π, the relative increase in the information on a variety due to the second stage is much less than for intense selection. It is therefore less likely that a variety with an exceptionally high first stage yield will be rejected at the second stage. This no doubt explains the increased value of accepting some varieties immediately after the first stage ($P_u > 0$) for large π. Clearly, however, the main advantage of early acceptance lies in whatever advantage may arise from skipping the final stage rather than from an increase in gain. It would also appear that if for some reason it were necessary to experiment with equal intensity at all stages, the advantage of early acceptance would be greater.

2.4.2 Variable replication

In the previous section we have used the value of y_1 to decide whether to reject, to accept, or to experiment further on a given variety. Obviously this concept can be extended in such a manner that the amount of resources allocated in the second stage to a given variety is a continuous function $A_2(y_1)$ of its first stage yield y_1. In view of the results of the previous section, such an elaborate scheme cannot be expected to result

in much of an increase in gain. It does, however, point to an interesting
general theoretical problem that probably has wider application than just
to screening. In general terms, the problem is to find a functional form
$f(u)$ that, subject to certain conditions, maximizes or minimizes a second
function involving $f(u)$. In order to make the nature of this general
problem clearer it will be illustrated in this chapter in terms of the
screening problem.

In the terminology of this thesis, the problem is to find the function
$A_2(y_1)$ that maximizes the gain. Since the first stage resources A_1 are
distributed equally over all varieties, the total resources A are given by

$$A = A_1 + \int_{-\infty}^{+\infty} A_2(y_1)\,dy_1 \ ,$$

or, in terms of the proportion of the total resources allocated to each
stage, by

$$\alpha_1 + \int_{-\infty}^{+\infty} \alpha_2(y_1)\,dy_1 = 1 , \tag{2.4.3}$$

where $0 < \alpha_1 \leq 1$. Since, as pointed out by Grundy, Healy and Rees (1956), a
small amount of second stage experimentation is never economically
advisable, we can expect there to be values y_1 and y_u such that $\alpha_2(y_1) = 0$
for $y_1 < y_1$ and $y_1 > y_u$. Rejection at stage 1 is implied by $y_1 < y_1$ and
acceptance by $y_1 > y_u$. If, as usual,

$$e_1^2 = \frac{V\sigma^2}{\alpha_1} \ ,$$

the proportion of varieties accepted into the second stage is, as in the
previous section, given by

$$P_1 = \int_{y_1}^{y_u} \exp(-\tfrac{1}{2}y_1^2/w_1^2)/(w_1\sqrt{2\pi})\,dy_1 \; .$$

The resulting value of e_2^2 is

$$e_2^2 = \frac{V\sigma^2 P_1}{\alpha_2(y_1)} \; .$$

In addition to 2.4.3, we can expect the function $\alpha_2(y_1)$ to satisfy other conditions. It will, for instance, probably be continuous with possibly a discontinuous slope. There will, of course, be a jump to zero if either or both of the values y_u and y_1 are appropriate. The function will probably also be monotone increasing to a hump in the middle and then monotone decreasing, but will not necessarily be symmetrical.

Since the problem of finding the function $\alpha_2(y_1)$ (and its parameters α_1, y_1 and y_u) to maximize the expected value of x in the selected population is extremely difficult, and probably only solvable numerically, it is not, considering the assumptions involved and the size of the likely gains, worth pursuing very far. The corresponding general problem of a functional answer to an optimization problem is, on the other hand, of both theoretical and practical interest and is certainly worth investigation. The general problem is, however, beyond the scope of this thesis and will not be considered further.

It is of interest at this point to mention that Grundy, Healy and Rees (1956) have solved a problem very similar to the one just formulated. They found a nomogram that gives the amount of second stage replication required to minimize the economic value of the "integral risk" in a selection program in which the first stage replication is taken as given (see section

1.4.2 for a more complete discussion). In a situation in which the external economy is known, their paper is extremely useful; it may also give clues to the solution of the problem stated in this section.

Out of curiosity, the result of using a rather arbitrary replication technique is investigated in this chapter. The technique consists of replicating each variety in proportion to the probability, based on the first stage results, of making an "incorrect decision" on that variety. An incorrect decision is, in this study, defined as rejecting (or accepting) a variety when it is (or is not) among the top π percent of the varieties being considered. In order to calculate the probability that x is in a specified interval, the linear regression of x on y, with regression coefficient

$$\beta = \frac{\sigma^2}{\sigma^2 + e_1^2} = \frac{1}{\omega_1^2}$$

and variance βe_1^2 (see, for example, Finney, 1956), is used to construct the function

$$g(x \mid y_1) dx = \frac{\omega_1}{e_1 \sqrt{2\pi}} \exp[- \tfrac{1}{2}(x - y_1/\omega_1^2)^2/(e_1^2/\omega_1^2)] dx \; ;$$

for a particular value of y_1, this function gives the probability that the true value of x is in the interval dx. The value x_0 where

$$\pi = \int_{x_0}^{\infty} \phi(u) du,$$

is taken as the division between desirable $(x \geq x_0)$ and undesirable $(x < x_0)$ varieties. Another arbitrary decision is made in using $y_0 = \eta_2$ (the second stage cut-off point) as the break-point for y_1: for $y_1 < \eta_2$ replication

is made proportional to $\Pr(x \geq x_0 \mid y=y_1)$, and for $y_1 \geq \eta_2$ to $\Pr(x < x_0 \mid y=y_1)$. In order to prevent a rather illogical jump in the amount of replication as y moves from below η_2 to above it, the replication for $y \leq \eta_2$ is multiplied by

$$Q = \frac{\Pr(x < x_0 \mid y=\eta_2)}{\Pr(x \geq x_0 \mid y=\eta_2)} .$$

Varieties yielding below $y_1 = \eta_1$ are rejected, without replication, immediately after the first stage. Finally, the function $\alpha_2(y_1)$ is obtained by dividing the recommended proportion of replication by

$$\bar{R} = \int_{\eta_1}^{y_0} Q\Pr(x \geq x_0 \mid y_1)\,dy_1 + \int_{y_0}^{\infty} \Pr(x < x_0 \mid y_1)\,dy_1 ,$$

in order to satisfy the condition 2.4.3. The resulting function $\alpha_2(y_1)$ is given by

$$\alpha_2(y_1) = 0 \quad (y_1 < \eta_1),$$

$$\alpha_2(y_1) = (1-\alpha_1) \int_{x_0}^{\infty} Q g(x \mid y_1)\,dx / \bar{R} \quad (\eta_1 < y_1 < y_0),$$

and

$$\alpha_2(y_1) = (1-\alpha_1) \int_{-\infty}^{x_0} g(x \mid y_1)\,dx / \bar{R} \quad (y_1 > y_0) .$$

Obviously this is all rather arbitrary but the graph of recommended replication as a function of y_1 does have roughly the same shape as that recommended by Grundy, Healy and Rees and, even more interestingly, it has the effect of flattening the graph of the expected value of G/σ versus α_1 and P_1. To illustrate this point, the gain achieved under this modified

system is given for nine pairs of values of α_1 and P_1 in table 2.4.2. The increase in gain over that for constant replication ($\alpha_2 = 1-\alpha_1$) is given in brackets beside each value of G/σ. The only major difference between the two systems can be seen to be a fairly large increase in gain for large values of P_1. The resulting flattening of the surface and broadening

TABLE 2.4.2

Values of G/σ for variable replication in the second stage ($\alpha_2 = \alpha_2(y_1)$) for V = 1 and π = 0.0625 (bracketed figures represent the loss or gain over the case when α_2 is a constant equal to $1-\alpha_1$).

α_1:	0.2		0.5		0.8	
P_1						
0.125	1.355	(0%)	1.540	(+0.1%)	1.564	(+0.1%)
0.250	1.541	(-0.1%)	1.602	(-0.2%)	1.544	(+0.5%)
0.875	1.473	(+3.2%)	1.528	(+7.9%)	1.504	(+7.4%)

of the optimum area is certainly desirable. The reason for this effect is the low replication recommended for the least promising varieties coming from stage one.

The calculations for table 2.4.2 were made directly from S(x) and an equation of the form 2.2.5; π_1 and π_2 were calculated by iteration and the gain by numerical integration. The values of G/σ are accurate to the number of figures given.

The example just described gives some idea of the sort of replication system that might be investigated in attempting to improve the gain. The

use of some sort of "hill climbing" technique in conjunction with variation of y_0, x_0, and the variance of the function $g(x)$ would probably result in still further improvement. Small changes in x_0 and y_0 did, in fact, result in slightly improved gains. In view of the small expected increase in gain, no further work was done. It is, however, hoped that this brief outline has indicated the nature of the more interesting general problem mentioned earlier in this section.

2.5 Two Assumptions Affecting Error Variance

2.5.1 Interaction

In all the previous work we have assumed no interaction effect; as stated by Curnow (1961), this is the most unrealistic of all our assumptions. That interaction is important in crop screening was pointed out by Sprague and Federer (1951) who, using actual results from corn yield trials, discussed in detail the practical problems resulting from variety by year (vxy) and variety by site (vxs) interactions. For convenience, this study is in the terminology of a variety x year interaction.

A rough idea of the effect of interaction on the average true yield or gain after selection can be obtained theoretically. To do this we must assume that we have a random sample of years and that the interaction is normally distributed with mean zero and variance σ^2_{vy}, independent of the variety in question. Obviously both these assumptions are dubious. Two or three successive years do not form a random sample and, in addition, the size of the interaction is not likely to be independent of the true yield of the variety in question or, for that matter, normally distributed. Investigation using these assumptions should, however, result in some idea of the magnitude of the reduction in gain due to the effect of interaction, and of the direction in which operating conditions should be moved to minimize this effect.

If the assumptions are accepted, the only difference from the situation studied in previous chapters is the addition of a constant term σ^2_{vy}, invariant from stage to stage, to the error variance e_i^2. As a result, we define the modified "error" variance $e_i'^2$ at stage i to be

$$e_i'^2 = e_i^2 + \sigma_{vy}^2 \ .$$

It should be mentioned that the term σ_{vy}^2 need not, if all the assumptions are satisfied, apply only to a vxy interaction; it may either contain components corresponding to other interactions or correspond to an entirely different interaction.

The expression for gain derived previously can now be used, with the modified error variance, to calculate the gain when (idealized) interaction is present. There are, however, important differences between this situation and the one that occurs when there is no interaction. The major difference is that the error variance $e_i'^2$ is now bounded below by σ_{vy}^2 and cannot decrease indefinitely as it could in the previous simplified model. As a result, we can expect the magnitude and probably even the location of the optimum gain to be affected in a manner dependent on the magnitude of σ_{vy}^2. In addition, the weights necessary to give the minimum variance estimate of x change. Since, under the idealized model, the interaction acts simply as an increased error term, the minimum variance estimate of the mean yield can be obtained by weighting the yield from each stage inversely according to the modified error variance $e_i'^2$. This however, even if the idealized model is appropriate, is usually not advisable since it is rare to have an estimate of σ_{vy}^2 that is based on more than a few degrees of freedom; Yates and Cochran (1938) have cautioned that, under these circumstances, the weighted mean loses greatly in efficiency. This problem is not likely to occur when there is no interaction present since, in that case, we usually know the number of replicates per stage, and the observations can be weighted accordingly (as long as plot size remains constant).

In the study done for this chapter, three systems are used for the weighting of historical information: firstly, $a_{ij} = 1/(e_j^2 + \sigma_{vy}^2)$ is used since, when accurate information is available, it will give the minimum variance estimate; secondly, $a_{ij} = 1/i$ is used since it can be expected to be a reasonably good estimate, especially if replication does not vary too much from stage to stage or if the interaction is large; lastly, the gain is calculated using $a_{ij} = 0$ ($i \neq j$) and $a_{ii} = 1$, thus ignoring all historical information.

As far as the magnitude of the interaction studied is concerned, various values of σ_{vy}^2 in the range $0.01 \leq \sigma_{vy}^2 \leq 100$ are used since this range nicely spans the change from a rather insignificant to a rather overpowering interaction effect. It also brackets both $\sigma_{vy}^2/\sigma^2 = 1$ ($\sigma^2 = 1$ is, as usual, the variance of the parent population) and the range suggested by the values of the ratios σ_{vy}^2/e^2 and σ^2/e^2 given by Sprague and Federer (1951).

It can be seen from table T4 for 3-stage symmetrical selection from a normal distribution that, even for very low values of σ_{vy}^2, interaction has a marked effect on gain. By the time $\sigma_{vy}^2 \doteq 1$ we are, when using equal weighting, losing between 14.4% (for $\pi = 0.0625$) and 20.5% (for $\pi = 0.001$) of the gain achieved when $\sigma_{vy}^2 = 0$. It is at this point ($\sigma_{vy}^2 = 1$) that equal weighting starts to show an improvement over using only the information from the most recent stages. For very large values of σ_{vy}^2 ($\sigma_{vy}^2 = 100$) we are losing about 80% of the gain achieved when there is no interaction. At this point equal and inverse variance weighting are equally good (to the accuracy of the figures shown). The reason for this is the fact that the

interaction is so large at this point that it dominates and the inverse variance weights are, in fact, very nearly equal. For such a large value of σ^2_{vy}, equal weighting gives a 7.7% ($\pi = 0.001$) to 11.7% ($\pi = 0.0625$) increase over the use of only the information from the most recent stage. All through the table it can be seen that inverse variance weighting gives, as expected, maximum gain.

Table T4 provides a good idea both of the approximate effect of interaction and of the advisability of different weighting systems when operating at symmetry. It can, however, be seen from the survey of gains in table T5 that maximum gain is not, at least for $\pi = 0.01$ and equal weighting, achieved at symmetry. In all these cases ($\sigma^2_{vy} = 0.1$, 1.0, 10.0) larger gains are achieved by using heavier than symmetrical replication ($\alpha_1 > \alpha_2 > \alpha_3$) in early stages. The values of optimum P_1, P_2 and P_3 vary from lighter than symmetrical in early stages for $\sigma^2_{vy} = 10$, to heavier than symmetrical in early stages for $\sigma^2_{vy} = 0.1$. When $\pi = 0.01$, $V = 1$ and $\sigma^2_{vy} = 1$ the approximate maximum gain occurs at $P_1 = P_2 = P_3$ and $\alpha_1 = 0.6$, $\alpha_2 = 0.3$ and $\alpha_3 = 0.1$. The improvement over the gain at symmetry is 3.3%. As can be expected because of the increased replication in early stages and the presence of a reasonably large value of σ^2_{vy}, the use of equal weighting gives a better (by 4.5%) result than the use of only the information from the most recent stage. It is also reassuring to note that equal weighting gives a gain only 1.2% lower than the maximum gain found for inverse variance weighting (at $P_1 = P_2 = P_3$ and $\alpha_1 = 0.5$, $\alpha_2 = 0.333$ and $\alpha_3 = 0.167$). It can be seen that σ^2_{vy} is already beginning to dominate. The same characteristics are found when $\sigma^2_{vy} = 10$, only to a greater degree (13.5% better than no history

and only 0.1% lower than the maximum gain with inverse variance weighting).
When σ^2_{vy} is small (σ^2_{vy} = 0.1) the maximum gain for equal weighting is 2.0%
lower than the maximum gain found when history is ignored. Since equal
weights are far removed from the optimum weights for such a small interaction,
this is not a surprise.

TABLE 2.5.1

Values of G/σ for equal weighting when interaction is present in 3-stage
selection from a normal distribution with V = 1 and $P_1 = P_2 = P_3 = \pi^{1/3}$

		Values of $(\alpha_1 \alpha_2 \alpha_3)$			
π	σ^2_{vy}	$(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$	$(\frac{11}{20}, \frac{3}{10}, \frac{3}{20})$	$(\frac{6}{10}, \frac{3}{10}, \frac{1}{10})$	$(\frac{13}{20}, \frac{5}{20}, \frac{1}{10})$
0.0625	0**	1.656	1.618	1.594	1.585
	0.1	1.560*	1.591	1.594	1.567
	1.0	1.417	1.444	1.433	1.428
	10.0	0.860	0.869	0.867	0.866
0.001	0**	3.226	3.236	3.216	3.216
	0.1	2.839*	2.974*	2.989*	2.998*
	1.0	2.566	2.664	2.677	2.683
	10.0	1.518	1.541	1.544	1.546

* indicates that the gain for $a_{ij} = 0$ ($i \neq j$), $a_{ii} = 1$ is higher.

** $a_{ij} = \frac{1}{e^2_j} / \sum_{j=1}^{i} \frac{1}{e^2_j}$ (inverse variance) weighting used for $\sigma^2_{vy} = 0$.

In order to test the theory that larger gains are obtained for heavier
replication in earlier stages, the entries in table 2.5.1 were calculated.
Although not sufficient to prove a general theory, this table and table T5
both suggest that when equal weighting is appropriate, increased gain can
be obtained over a wide range of values of π by increasing replication in
the early stages. It also appears that the more intense the selection the

further one should move in this direction. This agrees with comments made earlier (section 2.3.2) on intense selection. It can also be seen that equal weighting should probably not be used when σ^2_{vy} is much below 1. In this range the interaction has relatively little effect and it may, as suggested earlier for the case when there was no interaction, be better to ignore historical information unless the optimum weights are known accurately. A few similar calculations were made for $V = 0.2$ and $V = 5$. They did not indicate any major differences in behaviour other than those normally expected for variation in V.

On account of all the assumptions involved, caution is necessary in the interpretation of these results. Two conclusions can, however, be made with safety: the optimality of the symmetrical operating conditions certainly depends on the absence of interaction, and historical information is much more important than usual when $\sigma^2_{vy} > \sigma^2$. The fact that the optimum weights are nearly equal when $\sigma^2_{vy} > \sigma^2$ means that equal weighting is (as born out by the calculations in this section) better than ignoring historical information; it is, in fact, nearly as good as optimum weighting.

2.5.2 Plot size

So far it has been assumed that any change in our experimental resources (A) can be directly translated into a change in replication. In practice, however, a small change in A will rarely be easily translated into the exact number of plots needed to balance replication. As a result, it may not be possible to operate at symmetry; it may not even be possible to achieve the desired plot size. In experiments of the type considered in this thesis

it will often be necessary to change plot size as we progress through the stages of the experiment. For $\pi = 0.001$ and symmetrical 3-stage selection we are, in effect, assuming (since $e_1^2 = V\sigma^2 \cdot 3$ and $e_3^2 = V\sigma^2 \cdot 3 \cdot 0.01$) that we have 100 times as much replication in stage 3 as in stage 1; this seems rather unlikely. The very large number of varieties in stage 1 will probably force us to use a plot size that is uneconomically small; much of the additional area per variety in later stages will then have to go to increasing the plot to an economical size rather than to replication.

Finney (1966) suggested using a rule originated by Smith (1938) to handle problems arising out of changes in plot size. Smith, based on empirical evidence from 39 experiments, suggested the approximate rule that

$$\epsilon_x^2 = \frac{\epsilon_u^2}{\left(\frac{x}{u}\right)^p} ,$$

where ϵ_x^2 is the error variance of the yield for a plot of size x, and ϵ_u^2 is the error variance for a plot of size u. Smith found the power p to vary in the range $\frac{1}{20} \leq p \leq \frac{17}{20}$. Obviously, according to this rule, it will be best to take as small plots as are consistent with other requirements.

In order to apply this rule to screening it is necessary to use the variance under certain specific circumstances as a standard. If we let the variance

$$e_j^2 = \frac{V\sigma^2\pi^{j-1}}{\alpha_j}$$

at stage j be the standard, and if we assume that there are n_j replicates at stage j, the variance at any other stage will be given by

$$e_i^2 = \frac{V\sigma^2 \pi_{j-1}}{\alpha_j} \cdot \frac{1}{\dfrac{n_i}{n_j}} \cdot \frac{1}{\left[\dfrac{\alpha_i \pi_{j-1} n_j}{\alpha_j \pi_{i-1} n_i}\right]^p} \quad .$$

That this is so can be seen from the fact that, in Smith's terminology, u is given by

$$u = \frac{A_j}{N_j n_j} \quad ,$$

and x by

$$x = \frac{A_i}{N_i n_i} \quad .$$

The ratio $\frac{x}{u}$ is then

$$\frac{x}{u} = \frac{A_i N_j n_j}{A_j N_i n_i}$$

$$= \frac{\alpha_i \pi_{j-1} n_j}{\alpha_j \pi_{i-1} n_i} \quad .$$

If we take as standard the variance in a 1-stage program that studies all the varieties and uses all the resources $(e^2 = V\sigma^2)$, and if changes in area brought about by additional stages affect only plot size and not replication, this rule reduces to that studied by Finney (1966). Specifically, we have

$$e_i^2 = V\sigma^2 \left[\frac{\pi_{i-1}}{\alpha_i}\right]^p \quad .$$

Clearly, the use of the modified variance rule involves so many assumptions about the experimental situation that detailed calculations would be of little value without some specific situation in mind. Finney (1966) suggested that the modified rule would not affect the optimum very much either in location or in magnitude.

2.6 Non-Normal Parent Distributions

2.6.1 General comments

A major criticism that applies to all the work done so far in this thesis is that it is entirely based on the assumption of a normal distribution of the parent population; this will rarely be exactly true; in many instances the normal distribution will not even approximate the true parent distribution. Cochran (1951) pointed out that even if the main body of the distribution looks very normal the tail shape may be very different and, when selecting small fractions of the total population, it is the tail that is important. In this chapter, selection from a wide range of parent populations will be studied in order to determine the robustness of the suggested optimum selection programs to departures from normality and, in particular, to a wide variety of shapes of tails.

Curnow (1961), using Finney's cumulant series, made some calculations for 2-stage selection from non-normal distributions but encountered convergence difficulties when the departure from normality was at all pronounced. Curnow's convergence difficulties are avoided in this chapter by the use of numerical integration in conjunction with the expression

$$E(x) = \int_{-\infty}^{+\infty} xS(x)h(x)dx. \qquad (2.6.1)$$

As long as enough integration points are used, this method enables exact calculation of the desired value of gain for a very wide variety of parent distributions, skewed and otherwise. Unfortunately 2.6.1 can rarely be simplified.

Four types of parent distribution are studied: the family of curves used by Box and Tiao (1962) is taken as representative of a wide range of symmetrical distributions; the ρ-distribution is used to study the effect of positive skew (long tail towards positive values of x) and negative skew (long tail towards negative values of x); the sum of two normal distributions with differing mean values is studied to determine the effect of bimodality; and, finally, the 2-point distribution suggested by Davies (1958) is studied to complete the survey. It is felt that the wide variety of tail shapes covered by the various members of the families of curves just mentioned should give a good idea of the robustness of the symmetrical selection scheme. Further details on the characteristics of the various families of curves are given in the appropriate sections.

2.6.2 Symmetrical distributions: the Box and Tiao family of curves

In their 1962 paper, Box and Tiao studied the robustness of a certain situation (that need not concern us) by means of varying the parameter β in the distribution

$$h(x;\beta) = \frac{1}{\Gamma(\varphi)2^{\varphi}\emptyset} \, \exp\left[-\tfrac{1}{2} |(x-\theta)/\emptyset|^{2/(1+\beta)} \right] \,,$$

where

$$-1 < \beta \leq 1,$$

$$\varphi = 1 + \tfrac{1}{2}(\beta+1),$$

and the function $\Gamma(\varphi)$ is the gamma function. For consistency with other work, the mean will be chosen to be zero and $\emptyset$ will be chosen to give a

variance of 1 for all values of β studied. Although the family was not originated by Box and Tiao it will, for the sake of convenience, be called the Box and Tiao distribution. Its usefulness lies in the fact that, as a function of β, it varies through not only a wide variety of symmetrical distributions but also a correspondingly wide variety of tail shapes.

In this section, four representative curves are chosen from the family. They cover the range from a very long (exponential) tail at β = +1 to an almost rectangular tail at β = -0.9999. In addition to the extremes, the curves corresponding to β = ±0.5 are also studied since they result in curves midway between the extremes and normality. Information on the normal distribution (β = 0) from previous chapters is also included for reference.

Table T6 summarizes, for a variety of situations, the variation in gain for π = 0.01, V = 1.0 and selection for the five values of β just mentioned. The most interesting information in this table is that giving the location and increase over symmetry of the approximate maximum gain. Symmetry still seems to be a very good approximation both to the location and magnitude of maximum gain; it is only when selecting from the (rather unrealistic) rectangular parent population corresponding to β = -0.9999 that there is much difference in location (P_1 = 0.04 as opposed to P_1 = 0.1 at symmetry) but even then the increase in gain is only 1.3%. Although the recommended values of P_1 are consistently only slightly lower than symmetry, the recommended values of α_1 seem to decrease gradually as β decreases (the tail becomes shorter); the corresponding increases in gain are, however, negligible. Figure F16 for β = 0.5, V = 1 and π = 0.01 gives

the dependence of the gain (as a percent of target) on α_1 and P_1. Although this chart is for a distribution with longer and more gradually decreasing tail than the normal distribution, it can be seen (upon comparison with figure F2) that the only difference in the surface is that it is slightly flatter and slightly higher.

TABLE 2.6.1

The variation of gain (as a percent of target) with V for symmetrical 2-stage selection from the Box and Tiao family with $\pi = 0.01$.

β:	-0.5 (short tail)	0	0.5 (long tail)
V			
0.2	95.8	98.0	98.9
1	84.6	88.2	91.6
5	65.6	64.0	65.6
Target:	2.189	2.665	3.093

Further perusal of table T6 indicates that the value of history and of additional stages is similar to that found in the normal distribution throughout the entire range of β. There is, however, some variation in the recommended amount of initial random discard. It appears that with tails longer than normal ($\beta > 0$), little (1-stage selection) or no (2-stage selection) initial random discard is required (possibly $P_0 > 1$ may be useful in this region). On the other hand, a fairly heavy discard is recommended as $\beta \to -1$ even for 2-stage selection. Since the increase in gain for 2 stages is only 1.2% (less than the increase achieved by operating at optimum), $P_0 < 1$ is not, however, likely to be of value unless selection is being performed in one stage.

Table 2.6.1 gives the variation of gain (as a percent of target) with V for three values of β and 2-stage selection with $\pi = 0.01$. Once again there is no indication of any major departure from the behaviour found when selecting from the normal distribution. Similar results for other values of β and other types of selection ($P_0 \neq 0$, 1-stage and 3-stage selection) did nothing to weaken this belief. The only apparent differences

TABLE 2.6.2

The variation of G/σ with π for symmetrical 2-stage selection from the Box and Tiao family with V = 1 (bracketed figures give percent of target).

β:	-0.5 (short tail)	0	0.5 (long tail)
π			
0.0625	1.445 (80.4)	1.605 (81.6)	1.733 (82.8)
0.01	1.851 (84.6)	2.350 (88.2)	2.833 (91.6)
0.001	2.236 (88.8)	3.125 (92.8)	4.074 (91.3)

were those minor ones already mentioned in connection with table T6.

The same conclusion applies when the intensity of selection is varied. Table 2.6.2 giving G/σ and the corresponding percent of target achieved bears this out. The only possible conclusion that can be made from tables 2.6.1 and 2.6.2 is that there may be a slight increase in the percent of target achieved as β increases from low values to high values and even then the results for $\beta = -0.9999$ with V = 1 (table T6) and $\beta = 0$ with V = 5 (table 2.6.1) show that there are exceptions to this rule.

In conclusion, it certainly seems that the results of screening from a normal distribution will be a very good guide to selection from a wide

range of symmetrical distributions.

2.6.3 Skewed distributions: the β-distribution

In order to determine the effect of departure from symmetrical parent populations on the gain and other selection properties, selection from various members of the β-distribution

$$\beta(x;p,q) = \frac{1}{\beta(p,q)} \, x^{p-1}(1-x)^{q-1} \qquad (0 \leq x \leq 1),$$

will be studied in this section. Since the β-distribution is defined on a limited range $(0 \leq x \leq 1)$, and since it can be either very strongly positively or negatively skewed, it also provides a good contrast to the shape of the tail area of the symmetrical Box and Tiao family which, with the exception of the curve corresponding to β = -0.9999, covered an infinite range.

The nine members of the beta family that will be studied in this section vary from a positively skewed curve (p = 10, q = 30) through a symmetrical curve (p = q = 10) to a number of curves with varying degrees of negative skew. Since a number of examples of distributions with long upper tails (normal and Box and Tiao with β ≥ 0) have already been studied and since, as will be shown even more clearly in this section, the selection properties of a distribution depend mainly on its tail shape I have, to save computer time, limited this study to only the one example of positive skew (p = 10, q = 30).

All calculations in this section were made using equation 2.6.1. All the results are given in terms of standard deviation units from the mean of the parent distribution (i.e. gain = $(E(x) - \mu)/\sigma$ where μ is the

mean and σ the variance of the parent β-distribution).

A summary of the results for $V = 1$ and $\pi = 0.01$ is given in table T7. Once again symmetry is seen to give very nearly maximum gain. Only when the upper tail drops off very quickly ($p = 10$, 30 or 50 with $q = 3$) is any additional gain achieved by moving from symmetry and then it is only about 1 to 1.5 percent. It is worth noting that, in agreement with the results

TABLE 2.6.3

The variation of gain (as a percent of target) with V for symmetrical 2-stage selection from the β-distribution with $\beta = 0.01$.

| | Values of (p,q) | | |
V	(30, 10) (short upper tail)	(10, 10)	(10, 30) (long upper tail)
0.2	96.4	97.5	98.6
1	84.5	87.2	90.6
5	62.5	64.5	66.9
Target:	2.227	2.523	2.964

of the previous section, the optimum value of α_1 gradually decreases (0.6 to 0.4) as we move from a long upper tail through increasingly shorter and more rapidly decreasing upper tails. Optimum P_1, although always less (more intense selection) than symmetry also tends to decrease slightly. This same general effect also occurs for various other values of π and V. The fact that the gain is still very flat in the optimum area, even when selecting from a parent distribution with a much shorter tail than the normal, is illustrated in figure F17 for $p = 50$, $q = 3$, $\pi = 0.01$

and V = 1. Apart from the slightly greater departure of the optimum from symmetry already described and a slightly lower percent of target achieved, the figure is very similar to those already discussed for the normal (F2) and the Box and Tiao (F16) distributions with the same values of π and V.

Returning to table T7 we see that, as usual, history is of little value. It is interesting to note that, to four figure accuracy in the

TABLE 2.6.4

The variation of $(E(x)-\mu)/\sigma$ with π for symmetrical 2-stage selection from the β-distribution with V = 1 (bracketed figures give percent of target).

	Values of (p,q)		
π	(30, 10) (short upper tail)	(10, 10)	(10, 30) (long upper tail)
0.0625	1.388 (78.7)	1.571 (81.2)	1.788 (83.9)
0.01	1.880 (84.5)	2.199 (87.2)	2.686 (90.6)
0.001	2.336 (89.4)	2.781 (91.6)	3.611 (94.8)

gains, this value does not vary with p or q. In all cases the second stage gives an increase of more than 20% over the first, and the third an increase of approximately 4% over the second. This is very similar to the behaviour for the normal and Box and Tiao distributions. As was the case with the Box and Tiao distribution, the value of an initial random discard seems to increase as the upper tail becomes shorter. It is only for very negatively skewed distributions (short upper tails) that $P_0 < 1$ gives any increase in gain for 2-stage selection and even then the increase is only 1 to 1.5%.

The variation of gain with V (as a percent of target) is given for

negatively skewed (p = 30, q = 10) symmetrical (p = q = 10) and positively
skewed (p = 10, q = 30) members of the beta family in table 2.6.3 (note
that as you move from left to right the changes in upper tail shape for
table 2.6.3 and 2.6.4 are similar to those for tables 2.6.1 and 2.6.2 for
the Box and Tiao distribution). Although the actual gain varies slightly
in magnitude from the entries in the corresponding table (2.6.1) for the
Box and Tiao distribution, the gain expressed as a percent of target is
surprisingly similar in magnitude. The same can be seen to be true as π is
varied in table 2.6.4. As in the equivalent tables in the previous section,
there seems to be a slight increase in the percent of target achieved in
both tables as the upper tail increases in length. Similar results for
1-stage and 3-stage selection and variation in P_0 only served to emphasize
these points.

The great similarity between the behaviour under selection of the
beta and Box and Tiao families, even though they are very different in
shape, suggests that the basic shape of the upper tail is far more important
in determining the location and magnitude of the gain than the overall
distribution shape. This is emphasized by the fact that the location of
the optimum for the one relatively long tailed distribution (p = 10 and
q = 30) studied in this chapter is almost exactly the same as the location
for the normal. The relatively minor changes over a wide variation in tail
shape also suggests that the procedures are fairly robust even to major
changes in tail shape. For these reasons no further examples of unimodal
parent populations will be studied.

In conclusion, especially considering the variation expected in the

finite case (Finney, 1966), symmetrical operating conditions should give very nearly optimum results for a very wide range of unimodal parent populations.

2.6.4 Bimodal distributions: the sum of two normals

In this section, the effect of screening from a distribution with a lump in the upper tail will be studied. Barnard, in a comment to Curnow's 1961 paper, suggested the possibility of a bimodal parent population. He felt that consideration of history might be more important in such a situation. The calculations in this section were made not only with this in mind, but also to test the operation of screening programs when the tail of the parent population departs radically from the smooth monotone decreasing tails already studied.

For simplicity, and on account of the robustness of the system to individual tail shapes, a bimodal distribution consisting of a weighted sum of two normal distributions is used. The lower distribution (with mean 0 and variance 1) is given weight W and the upper (with mean D and variance V_u^2) weight $1-W$. The resulting parent population has the form

$$h(x) = \frac{1}{\sqrt{2\pi}} \left[W\exp(- \tfrac{1}{2}x^2) + (1-W)\exp(- \tfrac{1}{2}(x-D)^2/V_u^2)/V_u \right] .$$

To calculate the gain, the cut-off points η_i corresponding to a given set of P_i values are found iteratively in the usual manner. Once these points have been found, the proportions of varieties chosen from each of the two normal distributions making up the parent population are easily found. Using the notation of chapter 2.2 and equation 2.2.8 it can be seen that π_1,

the proportion of varieties chosen from the lower population, is given by

$$\pi_1 = I(\frac{\eta_1}{\omega_1}, \dots \frac{\eta_k}{\omega_k} ; R) \ .$$

Using equation 2.2.13, the mean value u_1 of the varieties selected from the lower population can be seen to be

$$u_1 = \frac{1}{\pi_1} \sum_{i=1}^{k} \phi(T_i) I^i(\eta^i(0); R^i)/\omega_i \ .$$

Since the overall proportion of varieties selected is fixed, the proportion of varieties π_u chosen from the upper population can be calculated from the expression

$$\pi = W\pi_1 + (1-W)\pi_u \ .$$

If we consider the varieties from the upper population to be represented by the variable $u \sim N(D, V_u^2)$, and if we make the transformation

$$u' = (u-D)/V_u \ ,$$

we can calculate u_u, the mean value of the selected u' varieties, from equation 2.2.13 in the same manner as u_1 was calculated. The weighted average

$$G = \pi_1 W u_1 + \pi_u (1-W)(u_u V_u + D)$$

gives the overall gain. For consistency in presentation all gains discussed in this section are standardized to the form

$$G' = \frac{G-u}{\sigma}$$

where

$$u = (1-W)D$$

is the mean of $h(x)$, and

$$\sigma^2 = \int_{-\infty}^{+\infty} x^2 h(x)\,dx - [(1-W)D]^2$$

$$= W + (1-W)V_u^2 + (1-W)WD^2$$

its variance.

The results of the bimodal calculations are given in table T8. This table is divided into six sections: section A gives the results for selection from a normal distribution and sections B to F the results for selection from a bimodal. In each of these latter sections one of the relevant parameters D, W, V_u^2, V and π is varied; for ease in comparison, the middle entry in each section repeats the case taken as standard (D = 4, W = 0.99, V_u^2 = 0.1, V = 1 and π = 0.01).

The reason for the choice of such a large value of D as standard can be seen in a comparison of selection from a normal distribution with selection from a bimodal with D = 3 (section A with the top row of section B). In all respects (advantage of additional stages, recommended P_0, value of history) these two situations are very similar; even the percentage of target achieved in 2-stage symmetrical selection for the bimodal (91.9%) is only slightly greater than that for the normal (88.2%). The similarities are even greater for smaller values of D, reaching equality when D = 0. Inspection of a graph of $h(x)$ for D = 3 shows the reason for the similarity

in the two cases just described: the hump is barely visible even on a
very large scale. The hump becomes much more obvious as D is increased
and, as this happens, we move very rapidly to almost perfect selection
(for $W = 0.99$ and $V_u^2 = 0.1$). At $D = 5$, for instance, we achieve 89.9%
of target in only one stage and 98.9% in two stages (at symmetry).
Obviously a value of D much greater than 4 would be just as unacceptable
as a standard as a value much smaller than 4.

The only major difference between the standard case and the normal
distribution is that 96.2% of target is achieved in 2-stage symmetrical
selection (versus 88.2%). Other minor differences include a higher
recommended value of P_0 in 1-stage selection (0.95 v.s. 0.5) and a
slightly higher recommended value of α_1 (0.7 v.s. 0.575) for optimum
second stage gain. Since the improvement of optimum over symmetrical gain
is only 0.3%, the slight change in optimum location is of little interest.
Comparison of optimum gain with symmetrical gain once again reaffirms our
belief in the flatness of the surface and the adequacy of the symmetrical
specifications. It is also interesting to note that $P_0 < 1$ is not, for
2-stage selection, advisable in any of the cases studied. History, although
slightly more valuable for $D = 5$ is, as usual, hardly worth including in
2-stage selection.

One reason for the behaviour mentioned above is the choice of the
weighting factor W. $W = 0.99$ is used mainly because we are choosing a
proportion $\pi = 0.01$ of the population: by giving the weight $1-W = 0.01$
to the upper population, perfect selection is nearly equivalent to choosing
all the upper population and rejecting all the lower. Section C of the

table shows the effect of varying W. In all situations studied for both $W = 0.9$ and $W = 0.999$ the target, the absolute gain, and the percent of target achieved decrease from the corresponding values for $W = 0.99$. Obviously as $W \rightarrow 1$ we will approach the results for selection from the normal distribution. This is already becoming evident when $W = 0.999$. The same happens (as far as percent of target achieved is concerned) in the limit as $W \rightarrow 0$. That this is happening, although rather slower than in the other direction, can be seen from the figures for $W = 0.9$. The major difference between the figures for $W = 0.9$ and those for both normality and $W = 0.99$ is the lower recommended value of α_1 ($\alpha_1 = 0.4$). The increase in gain over symmetry is still, however, only 0.8%.

Changes in V_u^2 can, from section D, be seen to have little effect in any of the situations studied. Sections E and F for variations in V and π respectively exhibit only the behaviour usually found for variation in these two parameters.

Obviously, a study in which the various pertinent parameters are varied together in the manner of a factorial experiment would be better than the above study; there would, however, be considerable difficulty in presenting and interpreting the results; statistical significance, since the results are exact, would not be relevant. In view of both these difficulties, and especially because of the required computer time, this study was limited to the information in table T8. The results in table T8 do, however, indicate that it is very unlikely that anything would turn up to cast much doubt on the conclusions already reached in previous sections on the behaviour of the optimum area.

2.6.5 The two-point distribution

In order to round out this study and particularly in view of the studies made by Davies (see, for example, Davies, 1958), selection from a 2-point distribution will be studied in this section. Davies suggested this distribution in connection with problems in the drug industry. In this study the distribution is defined to have two possible yields: the lower, with yield O, occurs with probability W and the upper, with yield 1, occurs with probability $1-W$.

Following the notation of chapter 2.2, the proportion of varieties selected can be seen to be

$$\pi = W \cdot I(\eta/e;R_0)+(1-W) \cdot I((\eta-1)/e;R_0) \ .$$

The values of η can be found by iteration in the usual manner. If no history is used, the matrix R_0 is the unit matrix and the integrals in the above expression are just products of normal frequency functions. In this simplified case, the values of η can easily be iteratively found from existing tables. Once the η's have been found the gain is, since the lower mean is O, automatically given by

$$E(x) = 1 \cdot (1-W) \cdot I(\frac{\eta-1}{e};R_0)/\pi \ .$$

Standardization of the gain is, since there are only two possible values of x in the parent population, felt to be confusing and is not used. In addition, the variance of the parent population is dropped from the expression for the error variance so that, in this section only,

$$e_i^2 = \frac{V\pi_{i-1}}{\eta_i} \ .$$

For reasons given in the previous section, the basic weight is again chosen to be $W = 0.99$. The basic value of the variance factor is $V = 0.2$. This value is used (somewhat arbitrarily) because it gives a medium-sized gain.

The main results of the study are presented in table T9 in three sections: section A showing the effect of varying W, section B of varying V and section C of varying π. As in table T8, the middle entry in each section is a repeat of the case ($\pi = 0.01$, $V = 0.2$ and $W = 0.99$) taken as standard.

There are a number of very interesting differences between the results in this table and all the preceding results. The most striking of these differences is the fact, apparent from the 2-stage results, that the symmetrical specifications no longer result in gains within a percent or so of the corresponding maximum 2-stage gains. This trait, except where 2-stage symmetrical gain is already nearly 100% of target, is present throughout the table: at $\pi = 0.01$, $V = 0.2$ and $W = 0.99$ the optimum gain is 19% higher than the symmetrical gain; at $\pi = 0.0625$, $V = 0.2$ and $W = 0.99$ it is 7.6% higher. In all instances where a marked increase occurs, it occurs at approximately $\alpha_1 = 0.6$ and at a P_1 which is 1.5 to 2 times as large as the symmetrical value. The only similarity between this location and the results in previous chapters is the larger than symmetrical value of α_1: approximately the same value of α_1 was recommended in nearly all cases in the study of selection from the (somewhat similar) bimodal distributions in section 2.6.4.

Figure F18, showing the variation in gain (as a percent of target) for

2-stage selection from a 2-point distribution with W = 0.99, V = 0.2 and
π = 0.01, strikingly illustrates the differences just mentioned. Comparison
of this figure with any of the earlier figures (F2, F16, etc.) shows that
it has narrower and longer contours in the optimum area and that this
optimum area is shifted in the direction of heavier first stage replication
and lighter first stage selection.

TABLE 2.6.5

An indication of the improvement in gain resulting from increased
early replication (α_1 = 0.433, α_2 = 0.333 and α_3 = 0.234) and
lighter than symmetrical early selection when selecting from a
2-point distribution.

W	π	V	P_1	P_2	P_3	Gain	% increase over max. 2-stage	% increase over 3-stage symmetrical
0.99	0.01	0.4	0.316	0.215	0.146	0.547	19.7	15.2
0.99	0.01	0.2	0.316	0.215	0.146	0.787	9.8	17.8
0.99	0.01	0.1	0.316	0.215	0.146	0.939	1.3	11.4
0.99	0.0625	0.2	0.450	0.397	0.350	0.139	9.4	7.8
0.999	0.01	0.2	0.316	0.215	0.146	0.0798	3.5	18.0

The fact that the symmetrical 3-stage gain is frequently lower than
the maximum 2-stage gain (6.8% lower at π = 0.01, V = 0.2 and W = 0.99)
suggests that the same sort of behaviour continues on into 3-stage selection.
In order to test this theory, the values in table 2.6.5 were calculated.
Although the locations of the test cases in the table were more or less
arbitrarily chosen in the expected optimum area, all cases result in
considerable increases over the symmetrical 3-stage results. Except for
π = 0.01, V = 0.1 and W = 0.99 (where we are already at 94% of target),

they also result in larger than usual increases over the maximum 2-stage results. A brief survey of the gain at various locations for $\pi = 0.01$, $V = 0.2$ and $W = 0.99$ resulted in a further increase of 1.6% (at $\alpha_1 = 0.5$, $\alpha_2 = 0.35$, $\alpha_3 = 0.15$ and the same set of values of P as in table 2.6.5). No doubt surveys for the other parent distributions would also lead to additional increases. Since it is obvious from the results of the table that this general area is preferable, no further surveys were made.

In addition to the change in the manner in which gain varies with the α_i and P_i values, and the larger than usual improvement due to the use of a third stage, there are also other differences for selection from a 2-point distribution.

Section B of table T9 indicates that the optimum 2-stage gain changes much more rapidly than usual as V is varied, moving from 46% of target at $V = 0.1$ to 72% at $V = 0.2$ and 93% at $V = 0.4$. Although optimum P_1 decreases slightly as V increases, optimum α_1 does not change.

The importance of the difference between $1-W$ and π is indicated by the results in section A for $\pi = 0.01$. When there are far more varieties with true yield equal to 1 than are required $(1-W > \pi)$, we get very good results. At $1-W = 0.1$ and $\pi = 0.01$, for instance, perfect selection occurs everywhere except for 1-stage selection with $P_0 = 1$. On the other hand, when $1-W \leq \pi$, the maximum 2-stage gain varies from 77% of target at $1-W = 0.001$ to only 72% at $1-W = 0.01$ (both for $\pi = 0.01$). The same sort of behaviour is found in section C of the table when π is varied for constant $1-W = 0.01$: when $\pi \geq 1-W$ the gains are in the range 72% ($\pi = 0.01$) to 86% ($\pi = 0.0625$) of target; when $\pi = 0.001$, it is only

when V is increased to 1.0 (from 0.2) that perfect selection is not achieved in two stages and even then optimum 2-stage gain is 98% of target. Although the results quoted in table T9 are not sufficient to prove a general rule, it appears that the gain (expressed as a percent of target) must reach a minimum somewhere in the region of $\pi/(1-W) = 1$ and then increase both when $\pi/(1-W) \rightarrow 0$ and when $\pi/(1-W) \rightarrow \infty$. Obviously 100% of target must be achieved in the limit in both directions.

Once again, history is seen to be of little value; the same is true of $P_0 < 1$ (except for 1-stage selection). Apart from the general increased sensitivity of gain to changes in any of the parameters, no other important effects are evident.

In view of the results of this section, it would obviously be a mistake to automatically use symmetrical operating conditions when selecting from a 2-point distribution. If no method of analysis is available, operation in the area of the 2-stage and 3-stage optima found in this section should, because of their consistency in location, result in an improvement over symmetry. Since history is of little or no value, and since the optimum occurs in very nearly the same location both with and without using historical information, a rough study of the optimum area should be possible using only tables of the normal distribution. The reason for this is the fact, mentioned earlier, that both the values of η_i and the gain can be calculated, for any number of stages, in terms of only the univariate normal distribution when history is not used.

2.7 Concluding Remarks

In closing, it is worth mentioning once again that the symmetrical operating conditions usually give very nearly optimum results. It is only when selecting from a very extreme distribution shape like the 2-point distribution that any marked increase in gain occurs for non-symmetrical operation. If, on the other hand, it is impossible to operate at symmetry, or if it is suspected that the distribution resembles the 2-point distribution, operation at values of P_i close to the symmetrical values but with increased replication in the early stages ($\alpha_1 \doteq 0.6$) should, unless V is very large, result in optimum or near optimum gain for a very wide range of parent populations, even the 2-point.

One other general comment worth mention is the wide variation in absolute gain ($(G-\mu)/\sigma$) that occurs for the different parent distributions. The fact that the gain is fairly constant when expressed as a percent of target could tend to be a bit misleading on this point. In view of the many distribution shapes studied, this variation is not very surprising. It does however suggest that, for given values of V and σ^2, the theoretically expected value of gain resulting from a screening program will, if based on a fixed break-even point calculated from a specific assumed parent distribution, depend heavily on that distribution. Tables T6, T7 and T8 indicate the large differences that occur in absolute gain when a fixed proportion of varieties is accepted, even though, in many cases, there are only subtle differences in tail shape; in a program involving long term profit (or other forms of gain) the observed differences could be disastrous. Careful study of the robustness of the

results of economically based studies to various parent distributions is therefore advisable. On the other hand, we have seen that the gain reacts very consistently to variation in the operating conditions, the use of history, the use of additional stages and other selection parameters for a wide range of parent distributions. The conclusions on optimum operation should, therefore, be a good guide to the mechanics of setting up any selection program. In view of the small increases in gain found in many instances, it should also be mentioned that the difference between success and failure in any operation is frequently as small or smaller.

PART 3. THE FINITE CASE

3.1 Introduction

As is frequently the case in statistical work, the preceding discussion suffers from the weakness of being only asymptotically correct. To make the most of the information in part 2 of the thesis it is therefore important to determine how closely the finite case follows the infinite, in particular, to determine how large N_1 and n (remember n = N_{k+1}) must be for the finite and infinite cases to be, for practical purposes, indistinguishable. In general, for all values of N_1 and n, it is also important to know whether optimum gain is still located at symmetry and whether it is robust to changes in operating conditions.

Finney (1966) studied this subject in detail for selection when historical information is not used, and when fixed proportions are accepted at each stage. He found that, although overestimating the gain, the infinite case is in all other respects a very good guide to the finite case. In chapter 3.4 a similar study is made for the case when history is used. In chapter 3.5, we study the effect on gain of schemes that are, in a sense, somewhat more flexible than those studied by Finney. The increased flexibility is achieved by replacing schemes based on accepting fixed proportions at each stage with schemes that accept all varieties yielding above fixed cut-off points at each stage; combinations of these two acceptance rules are also studied.

A major difference between the finite and infinite cases is the variation in quality between different cohorts. It is, in fact, in order to determine whether we can take advantage of this variation that we

study the above-mentioned "flexible" selection schemes. The reason for the variation in quality is the assumption, necessary for a study of the finite case, that N_1 is a random sample from an infinite continuously distributed population. Obviously, in practice, the overall average true yield will vary from cohort to cohort and so the assumption is, in part at least, reasonable. Another and perhaps less reasonable aspect of the assumption is that the parent distribution is $N(0,1)$. It has already been indicated that the mean of 0 and the variance of 1 can be achieved without loss of generality; if the robustness (found in chapter 2.6) of the infinite screening results to variations in parent population carries over to the finite case, the reliance on a normal parent should not be limiting either. I see no reason why this should not be the case, especially in view of the close similarity found between the infinite and finite cases in chapters 3.4 and 3.5; study of finite selection from non-normal parent distributions is, however, necessary.

3.2 Theoretical Considerations

The difference, arising in the finite case, between using cut-off points and using fixed proportions to define our acceptance rules has already been mentioned; an even more important difference, as far as finding the gain is concerned, is that we are now forced to base all our calculations on order statistics rather than on tails of distributions.

TABLE 3.2.1

A comparison of the infinite target (normal tail) with the finite target for $\pi = 1/16$ and various values of $N_1 = N_0$.

$N_1 = N_0$	n	Target
16	1	1.766σ
32	2	1.858σ
64	4	1.911σ
128	8	1.939σ
256	16	1.953σ
Limit (normal tail)		1.968σ

When accepting fixed proportions this means that the finite target is the mean of the top n values of the order statistics appropriate to a sample of N_1 numbers from the normal distribution. In contrast, the infinite target is the mean value of the corresponding proportion of the upper tail of the normal distribution. The effect on the finite target that results from using cut-off points is indicated in chapter 3.4.

Some indication of the difference between the finite case and the infinite case is given in table 3.2.1. This table is reproduced from Finney (1966). It shows the finite target for constant $\pi = n/N_1 = 1/16$,

but for various values of N_1 and n. It can be seen that the decrease of the finite target from the infinite is not very great: for N_1 = 16 and n = 1 it is only 10.3%; for N_1 = 256 and n = 16 it is down to a drop of only 0.8%. Obviously the approach to the infinite case is rapid.

When error is introduced into the system, the necessity of dealing with order statistics makes analysis difficult. Finney (1966) worked out expressions for the expected value of the gain ($E(G)$), and of the variance ($E[Var(G)]$) of the gain between selections from completely independent cohorts. Letting, as usual,

$$e_i^2 = \frac{VN_iA}{N_0A_i}\,\sigma^2 \ ,$$

we have, for 1-stage selection,

$$e_1^2 = \frac{VN_1}{N_0}\,\sigma^2$$

and, as a result, we get Finney's expression,

$$E(G) = [N_0/(N_0+N_1V)]^{\frac{1}{2}}\,\sigma g(N_1,N_2),$$

for the gain. The function $g(N_1,N_2)$ is the expectation of the mean of the top N_2 values in a random sample of N_1 members from a $N(0,1)$ parent population. Its value can (for $N_2 \leq N_1 \leq 400$) be found from the entries in Harter's (1960) tables. Finney also found the expression

$$E[Var(G)] = [N_0/(N_0+N_1V)]\sigma^2[h(N_1,N_2)+N_1V/N_0]$$

for the variance. The function $h(N_1,N_2)$ in this expression represents the expectation of the variance of the mean of the largest N_2 values in a random

sample of N_1 members from $N(0,1)$. Its value can be obtained from Teichroew's (1956) tables of expected squares and products of the first, second, ... largest values in random samples of $N_1 \leq 20$ from $N(0,1)$. When $P_0 = 1$, that is to say $N_0 = N_1$, the preceding two expressions reduce to

$$E(G) = (1+V)^{-\frac{1}{2}} \sigma g(N_1, N_2) \qquad (3.2.1)$$

and

$$E[Var(G)] = (1+V)^{-1} \sigma^2 [h(N_1, N_2) + V] . \qquad (3.2.2)$$

Since Finney studied the use of an initial random discard in finite 1-stage and 2-stage selection and did not find any unusual differences from the infinite case, and since the use of history has been seen to have no significant effect on the value of an initial random discard, we will use $N_0 = N_1$ throughout this study.

In moving to 2-stage selection, the above expressions can be modified to calculate the gain at the margins ($\alpha_1 = 0$ or 1, and $P_1 = 1$ or π). When history is used and $\alpha_1 = 1$, for example,

$$E(G) = (1+V)^{-\frac{1}{2}} \sigma g(N_1, N_3) ,$$

and

$$E[Var(G)] = (1+V)^{-1} \sigma^2 [h(N_1, N_3) + V] ,$$

independent of N_2. The same expressions apply when $N_2 = N_1$ independent of α_1. For either $N_2 = N_3$ or $\alpha_1 = 0$ the use of history has no effect and the expressions derived by Finney (1966) apply: for $N_2 = N_3$, $E(G)$ and $E[Var(G)]$ can be calculated from the 1-stage expressions by inserting V/α_1

for V; for $\alpha_1 = 0$ they can be calculated by replacing N_1 by N_2, and N_2 by N_3 in the 1-stage expressions. Expressions can also be calculated for similar marginal situations for higher order selection procedures but, since selection under these conditions is equivalent to 1-stage selection, it is unlikely in practice and the expressions are of little interest.

General expressions for the finite gain after any number of stages are possible, but are so complex that numerical integration would almost certainly be necessary. The expression

$$E(x_{(r)}) = \frac{n!}{(n-r)!(r-1)!} \int_{-\infty}^{+\infty} x[F(x)]^{n-r}[1-F(x)]^{r-1}f(x)dx ,$$

for the expected value of the rth highest order statistic in a sample of n members from a distribution $f(x)$ where

$$F(x) = \int_{-\infty}^{x} f(x)dx ,$$

gives some idea of what would be involved. In a screening study

$$f(x) = S(x)g(x) ,$$

where $S(x)$ is the selection function (a k-variate normal distribution function). Obviously numerical evaluation of the order statistics, under these circumstances, would be very time consuming; this approach was not attempted and, in its place, the alternative method of simulation was used.

3.3 Simulation Technique

Although the basic simulation technique used in this thesis is nearly the same as that used by Finney (1966), it is so important to the study that it is described in detail in this section. The computer used for the simulations, and indeed for all calculations in this thesis, was the University of Edinburgh KDF-9.

Pseudo-random deviates from a $N(0,1)$ population were generated on the computer in independent pairs $(\varkappa_1, \varkappa_2)$ by means of the relationships

$$\varkappa_1 = (-2 \log \xi_1)^{\frac{1}{2}}\cos(2\pi\xi_2),$$

and

$$\varkappa_2 = (-2 \log \xi_1)^{\frac{1}{2}}\sin(2\pi\xi_2) \; ;$$

ξ_1 and ξ_2 are uniform random deviates in the interval $(0,1)$. The uniform pseudo-random deviates were generated by a standard sub-routine. This method of generating random $N(0,1)$ deviates is due to Box and Muller (1958); faster methods are available but, in nearly all the cases studied, the generation of the pseudo-random numbers took only a small portion of the program time so this method was used to save program storage space. Since both the error and the parent population are normally distributed, there was no need for a separate store of deviates for each. Each new deviate was, of course, taken directly from the store as required.

In a given computer run, a large number (usually between 6000 and 7000) of pseudo-random $N(0,1)$ deviates was generated. Multiple usage of up to ten per deviate was achieved by regarding the numbers as being in a continuous loop in storage and continuing around the loop until the maximum

permissible number of usages had been made. When the required number of
runs for a given set of selection specifications was complete, the next
set of specifications was read in and studied without the necessity of
generating additional numbers unless, of course, the numbers had already
been used the permitted number of times. Similarly, if necessary, a new
set of pseudo-random numbers could be generated in the middle of a series
of runs for a given set of specifications. The chance of patterns of usage
was decreased by always using a prime number of deviates. In order to
prevent any two selections being based on nearly the same set of deviates,
the interval between the position in store of any two deviates was increased
in a manner dependent on the total number of usages made of each of the
deviates: adjacent deviates were used until every deviate had been used
once and then every second deviate until all had been used twice and so on
until new deviates were required.

In a given selection, N_1 deviates were chosen to represent the true
yields of the parent population; to these were added another N_1 deviates,
this time multiplied by the experimental error e_1; the results were taken to
be the first stage yields. From these results the highest yielding N_2
varieties were chosen and N_2 random deviates (this time multiplied by e_2)
were added to the true yields of the selected varieties. At this point,
since history was being used, the estimate z_2 of the true yield was
calculated according to equation 2.1.2, and the N_3 highest yielding varieties
were chosen for study in stage 3. The same procedure was followed until
the entire k stages had been completed. The mean value of the chosen n
varieties was then calculated and its first four powers were accumulated.

The process was then repeated with another set of deviates. At the end of the required number of runs the average gain and the variance of this gain between the independent runs (Var(G)) were calculated, as were the third and fourth cumulants of the frequency distribution of G.

At any stage i, the bulk of computer time consumed was used in sorting to find the N_{i+1} highest yielding varieties from the N_i varieties being studied. Since methods were available from the work in part 2 for calculating theoretical cut-off points corresponding to the desired proportions, it was possible to increase efficiency over the method used by Finney. The method used was to divide the varieties into two categories immediately after adding e_i and calculating z_i; those for which $z_i < \eta_i$ went into the lower category (tentatively rejected) and all the rest into the upper (tentatively accepted). If selection was based on cut-off points, the stage was then complete and it was possible to proceed directly to the next stage; obviously this resulted in a very large saving of time. If, on the other hand, exactly N_{i+1} varieties were desired, sorting was still necessary, but only of relatively few varieties. This sorting was done by first calculating $\partial = n_u - N_{i+1}$ (n_u represents the number of varieties tentatively accepted); when $\partial > 0$ (too many varieties accepted) the N_{i+1} highest yielding varieties were found by first arranging any ∂ of the n_u varieties in order; the highest yielding variety of these ∂ ordered varieties was then compared with each of the remaining $n_u - \partial = N_{i+1}$ varieties until one with a lower yield had been found; the two were interchanged and the lower yielding variety placed in its proper order among the ∂ ordered varieties; the procedure was continued from where we left off in the N_{i+1}

varieties in the same manner until the ∂ lowest yielding varieties of the n_u tentatively accepted had been placed in the ordered group. The ∂ varieties were than discarded and we were left with the required N_{i+1} highest yielding varieties. A similar procedure was followed when $\partial < 0$. When $\partial = 0$, as often happened, no sorting at all was required. This method saved considerable time, not only over sorting all the varieties, but also over sorting all members of the category that had too many varieties.

In view of the natural variation of the average parent population from one set of runs to another, it was decided to standardize the gain. The technique used was that of Finney (1966): in each run for a given set of selection specifications a "target" value (the average true yield of the 2n best varieties in the population) was calculated; when the set was completed the regression of the gain in a given run against the corresponding target in that run was calculated over all runs in the set; this regression was used to calculate the gain G_R that would have resulted had the average "target" been the average of the top 2n order statistics in a sample of N_1 members from a $N(0,1)$ population. This gain was frequently checked against the gain found by an alternative standardization technique (also suggested by Finney); agreement was always very good. Because of the slight effect of history on the gain, it was assumed that the linearity of regression found by Finney in his work would still apply; no reason arose during the study to cast doubt on this assumption. Since the regression seemed to standardize the mean very well, and since the variance of the adjusted regression gain, G_R, was considerably lower than the variance of the mean gain $\bar{G}$, it was used in all situations in this thesis in which exactly n varieties were

accepted in each run. In order to have a true measure of the variability of performance, Var(G) was not adjusted for the regression but was always taken to be the actual variance of G between the independent runs.

When a final cut-off point was used, G_R was not, however, thought to be appropriate and was not used. This decision was taken in view of the fact that, although the number of varieties in the target was, of necessity, fixed at the number originally chosen (say 2n), the number of varieties accepted in a given run varied considerably depending on the experimental error and on the quality of the varieties in the cohort. As a result, although the target moved up and down as a function of the best 2n varieties, the average gain remained relatively stable since all the varieties yielding above the fixed cut-off point were chosen: in a poor cohort with only one good variety, it would frequently be the only variety chosen and, although the target was low, the gain would be reasonably high; in a cohort in which all the varieties were very good, the target would be very high but, because many more varieties were chosen and mistakes were more frequent, the average gain was not usually particularly high. The regression was expected to be rather poor under these circumstances. In addition, it was not known how to take into account the fact that, on many occasions, all the varieties in a cohort were rejected. This, plus the difficulty of taking into account the variation in the number of varieties making up the average gain in a given selection, made it impossible to take into account all available information in calculating the regression. Although, for the above reasons, G_R was not used, the use of a final cut-off point had the effect of standardizing the gain to quite an extent and, for a given

number of runs, there was little difference between the standard errors of G_R for fixed n and of $\bar{G}$ for varying n.

In order to investigate the possibility of a further reduction in error variance, selection from a population consisting of N_1 varieties with true yields equal to the appropriate N_1 order statistics was studied. This procedure was found to be inadvisable for two reasons: first, for a given number of runs, the variance of the resulting mean gain was almost identical with the variance of the regression estimate; second, and even more important, the estimate was biased.

Both the technique and, in particular, the bias are illustrated by the following example of 1-stage selection (with V = 1) of one variety from a population of two. When selecting from a population of two varieties whose true yields are independently N(0,1) distributed, the expected 1-stage finite gain is 0.399 (see equation 3.2.1). When selecting from a population of two varieties, one with true yield +0.5642 and the other with true yield -0.5642 (the order statistics appropriate to a N(0,1) population of size two), the expected finite gain can be calculated from

$$Pr(y_1 > y_u) = Pr(y_1 - y_u > 0),$$

where y_1 is the yield (with error) of the variety with the lower true yield and y_u the yield of the other. Since V = 1, we have

$$e_1^2 = V\sigma^2 = 1$$

and

$$y_1 - y_u \sim N(-1.1284, 2).$$

This means that

$$Pr(y_1-y_u > 0) = Pr(z > 1.1284/\sqrt{2})$$

$$= 0.2125.$$

This gives the proportion of times the variety with true yield equal to
-0.5642 is chosen. Since the variety with true yield equal to +0.5642
is the only other one that can be chosen, the expected value of the
finite gain can be seen to be

$$E(G) = 0.5642(1-0.2145) - 0.5642(0.2145)$$

$$= 0.325,$$

a considerably lower value than that calculated by equation 3.2.1. This
is not to say the value is incorrect, it is just not the one we are looking
for. Out of curiosity, a simulation of 10,000 runs was made on a parent
population consisting of the two order statistics: the result was a gain
of 0.324±0.005. A similar simulation on a population consisting of two
independent values from an $N(0,1)$ distribution resulted in a gain of
0.405±0.006. It can be seen that the agreement with the theoretical
values is very close.

The negative bias was found to persist when selecting 1 variety from
16 in two stages with $V = 1$. In that case, the gain was found to be
1.339±0.007 in a simulation of 6000 runs on a population comprising the
16 order statistics and 1.401±0.004 in a simulation of 12,000 runs on
random samples of 16 from a normal distribution. In view of these results,
selection from the order statistics was abandoned and use of G_R adopted.

In closing, it should be pointed out that Var(G) is a very important
part of the results of these simulations; in most instances the experimenter

will be very interested in the variation of the gain between independent runs, perhaps even more than in the variance of the mean gain, $\overline{G}$, or the regression gain, G_R, over a very large number of runs. Any estimation technique that does not provide this value will be considerably less valuable than one that does.

3.4 The Effect of the Use of Historical Information

3.4.1 The overall gain and the approach to the infinite case

Table T11 illustrates, for V = 1 and 2-stage selection of 1 variety from a population of 16, the fact that there is very little variation in the finite gain (both with and without the use of history) in the optimum region. When history is used, it can be seen that even in the large region in the neighbourhood of symmetry covered by the 19 most accurately determined gains (marked with single asterisks) the minimum gain (1.358 at $N_2 = 4$, $\alpha_1 = 0.8$) is only 3.2% lower than the maximum (1.402 at $N_2 = 4$, $\alpha_1 = 0.7$); since $Var(G)/\sigma^2$ varies between 0.533 and 0.579 (about 40% of gain) in this region, this variation in gain will not be very important. Keeping in mind that approximately 14,000 runs have been made to calculate each of these 19 gains, and that the corresponding standard error is about ±0.004, it can be seen that operation anywhere in the region $3 \leq N_2 \leq 6$ and $0.4 \leq \alpha_1 \leq 0.7$ should give very nearly optimum results.

Another reassuring fact comes to light on comparison of the upper half of table T11 with table T10 giving the equivalent figures for the gain with history in the infinite case. The similarity in the variation of the gain is illustrated by the fact that the finite gain is consistently at a level of about 85-91 percent of the infinite gain; there is a slight falling off in the region of light replication and heavy selection in early stages but even then only one figure falls below 80% of the corresponding infinite gain. Considering the number of figures involved, and the fact that most of them have a standard error of 1.5 to 2 percent of the finite

gain, this can be seen to be very consistent behaviour. Table 3.4.1, giving the percent of infinite gain achieved (when history is used) by the finite gain when $N_1 = 100$, $N_3 = 1$ and $V = 1.0$, indicates that this same consistency exists for $\pi = 0.01$. The fact that the variation in the figures is far from smooth is explained by the fact that the standard

TABLE 3.4.1

The percent of infinite gain achieved in the finite case for
$N_1 = 100$, $N_3 = 1$ ($\pi = 0.01$), $V = 1.0$ and 2-stage selection
from a normal distribution.

N_1:	4	5	10	20	25
α_1					
0.3	92.3	93.4	91.4	92.5	92.7
0.4	91.2	91.3	90.1	92.8	94.1
0.5	88.0	90.2	91.8	92.5	93.0
0.6	90.0	92.7	94.8	93.7	95.5
0.7	90.5	91.3	94.4	96.1	95.5

errors of the finite regression gains are, in most cases, in the region ±0.035 to ±0.040, or about 1.5% of the gain. Similar calculations for $V = 0.2$ and $V = 5.0$ for both $\pi = 0.01$ and $\pi = 0.0625$ indicated the same consistent behaviour in the range of 85 to 93 percent of the infinite gain. Every indication is that the infinite case will be a very good guide to the behaviour in the finite case.

The value of the use of history in the finite case is brought out in a comparison of the upper and lower halves of table T11. As in the infinite case, history is of little value in the optimum region; the same is true in the region corresponding to light replication and heavy selection in the

first stage; only when first-stage replication is heavy and the corresponding selection light is history of any real value. As usual, the correspondence between the behaviour in the finite and infinite cases is very close. Out of curiosity, 14,000 runs were made both with and without history at the symmetrical specifications for $N_1 = 16$, $N_3 = 1$ and $V = 1$: the regression gain was 1.392σ without history and 1.401σ with history (using a completely new set of random numbers). Since the standard error of the regression mean was ± 0.005 in the former case and ± 0.004 in the latter, it can be seen that there is little to choose between the two, even after 14,000 selections. Selection with $\tau = 0.01$ and with $V = 0.2$ and 5.0 gave results similar to those already mentioned.

The variance of the gain between independent runs at the same specifications $(Var(G)/\sigma^2)$ is given in table T12. It can be seen that this variance is very nearly constant at its minimum value in the optimum region. Not only is this behaviour very similar to that found when history was not used, but the actual values of $Var(G)/\sigma^2$ are very similar: at symmetry (and after 14,000 runs each) $Var(G)/\sigma^2 = 0.555$ with history and 0.553 without. Comparison of table T12 with the corresponding table (table 2) in Finney's 1966 paper indicates that it is only in the region of light first-stage selection and heavy first-stage replication that the use of history results in a decrease in $Var(G)/\sigma^2$.

The third and fourth cumulants of the frequency function of G after 2-stage selection with $N_1 = 16$, $N_3 = 1$ and $V = 0.2$ are given in table T15. In sharp contrast with the reasonably smoothly varying values of G and $Var(G)$ in tables T11 and T12, the entries in table T15 show little or no pattern

at all. Similar behaviour was found for other values of V and π. The reason is, no doubt, the combination of a large sampling error with the relatively small values of the cumulants.

TABLE 3.4.2

The approach at the symmetrical specifications ($\alpha_1 = \alpha_2 = 0.5$, $N_2/N_1 = N_3/N_2$) of the finite gain to the infinite gain both with and without the use of history ($\pi = 0.0625$, V = 1).

N_1	N_2	N_3	G_R/σ without history	G_R/σ with history	$Var(G)/\sigma^2$ with history	$E(G)/\sigma$ with V=0.0
16	4	1	1.392±0.005	1.401±0.004	0.555	1.766
32	8	2	1.484±0.012*	1.492±0.010	0.278	1.858
64	16	4	1.517±0.008*	1.554±0.009	0.143	1.911
128	32	8	1.558±0.012*	1.575±0.008	0.074	1.933
256	64	16	1.572±0.014*	1.595±0.007	0.028	1.953
∞ limit			1.590	1.605	0.000	1.968

* Finney's 1966 results.

The manner in which the finite gain approaches the infinite gain when N_1 is increased for constant π is illustrated, and compared with Finney's (1966) results, in table 3.4.2. Although the gain with history is always slightly greater than the gain without, there is no discernible difference in the rate with which the two systems approach the infinite case. The same conclusion applies to the rate with which $E(G)/\sigma$ approaches the infinite limit when V = 0.0. The variance of the gain between runs was so similar in the two cases that the non-historical results were not included. Similar results for $\pi = 0.01$ indicated only a slight decrease (as expected) in the value of history; in all other ways they resembled

the behaviour of the results in table 3.4.2.

The results of this section certainly suggest that history has little effect on the finite gain other than that already found in the infinite case. In a broader sense, the preceding results confirm Finney's conclusions that the finite gain behaves in a very similar manner to the infinite gain and that the finite gain approaches the infinite quite rapidly.

TABLE 3.4.3

The finite gain (with and without history) as a function of V for symmetrical selection ($\alpha_1 = \alpha_2 = 0.5$, $N_2/N_1 = N_3/N_2$) from a normal population with $N_1 = 16$, $N_3 = 1$ and $V = 1.0$.

V	G_R/σ**		$Var(G)/\sigma^2$ with history	$E(G)/\sigma$ for infinite model with history
	without history	with history		
0	1.766	1.766	0	1.968
5^{-3}	1.755±0.006*	1.759±0.009	0.306	1.965
5^{-2}	1.740±0.007*	1.739±0.008	0.294	1.952
5^{-1}	1.674±0.009*	1.684±0.004	0.374	1.889
$5^{-\frac{1}{2}}$	1.564±0.013*	1.567±0.010	0.426	1.789
1.0	1.392±0.005	1.401±0.004	0.556	1.605
$5^{\frac{1}{2}}$	1.136±0.022*	1.162±0.009	0.686	1.340
5	0.915±0.025*	0.888±0.006	0.793	1.039
5^2	0.454±0.030*	0.467±0.010	0.948	0.529
5^3	0.215±0.029*	0.225±0.011	0.941	0.244
5^4	0.070±0.031*	0.095±0.012	0.990	0.110

* Finney's (1966) results.
** From at least 1000 runs each.

3.4.2 Variation in the value of history with V and π

It can be seen from table 3.4.3, showing the gain both with and without history over a wide range of values of V, that there is very little variation in the value of history with V; there is probably a slight increase in its value as V is increased but, considering the errors involved, even this is not very obvious. The error was, in fact, so large

TABLE 3.4.4

Variation of G_R/σ in the optimum region for V = 60, N_1 = 16 and N_3 = 1 (based on 3000 runs each; error of ±0.018).

N_2:	3	4	5	6
α_1				
0.3	0.303	0.270	0.303	0.309
0.4	0.285	0.316	0.289	0.269
0.5	0.267	0.295	0.301	0.317
0.6	0.314	0.318	0.269	0.281

relative to the value of history that the gain with history was often lower than the gain without; this occurred frequently throughout the simulation study. There was no discernible pattern to the difference between $Var(G)/\sigma^2$ with and without history.

Considering the errors involved, the fact that G_R/σ is consistently in the range 86-90 percent of the infinite gain over the entire range of V indicates a very close adherence to the infinite case; for this reason it is probably safe to say that the variation in the value of history as a function of V is also very close to that found in the infinite case (see figure F11).

The fact that symmetry is still, for very large values of V, a very good approximation to the location of optimum gain is illustrated in table 3.4.4 showing the gain at various locations in the region of symmetry for $\pi = 0.0625$ and $V = 60$. Considering the errors involved, operation at

TABLE 3.4.5

The finite gain (with and without history) as a function of N_1 for $N_3 = 1$ and near-symmetrical selection ($\alpha_1 = \alpha_2 = 0.5$) from a normal population with $V = 1$.

N_1	N_2	G_R/σ without history	G_R/σ with history	$\text{Var}(G)/\sigma^2$ with history	$E(G)/\sigma$ for infinite model with history	Finite target
1	1	0	0	1.0	0	0
4	2	0.719±0.013*	0.744±0.010	0.744	0.956	1.029
9	3	1.127±0.014*	1.131±0.013	0.646	1.347	1.485
16	4	1.392±0.005	1.401±0.004	0.551	1.605	1.766
32	6	1.721±0.023*	1.697±0.013	0.464	1.898	2.070
64	8	1.975±0.027*	1.969±0.014	0.395	2.179	2.344
100	10	2.128±0.014*	2.158±0.014	0.360	2.350	2.508
128	11	2.290±0.028*	2.270±0.016	0.382	2.442	2.595
256	16	2.491±0.034*	2.489±0.015	0.328	2.687	2.827
512	23	2.753±0.033*	2.767±0.011	0.273	2.918	3.044

* Finney's 1966 results.

either $\alpha_1 = 0.3$, $N_2 = 3$ (the approximate location of the infinite optimum) or at symmetry should give very nearly optimum gain. Similar studies for $\pi = 0.01$ and various values of V gave results similar to those just described.

The results in table 3.4.5 for the gain both with and without history, over a wide range of values of N_1 ($N_3 = 1$), indicate that, once again, the infinite case is a very good model. In this case, however, the percent of infinite gain achieved by the finite gain appears to vary slightly as a

function of N_1: at $N_1 = 4$ ($\pi = 1/4$) we achieve 78% of the infinite gain; at $N_1 = 512$ ($\pi = 1/512$) we achieve 95% of the infinite gain. Although masked by the simulation error in G_R/σ, this trend seems to occur consistently over the entire range of N_1. This, no doubt, is a result of the fact that, as N_1 increases, we gradually approach the infinite case even though N_3 stays constant at 1.

The variation with π of the value of history is, because of the fact that the error in G_R/σ is of the same order of magnitude, harder to determine; the closeness with which the finite case (both with and without history) follows the infinite suggests that the behaviour found in the infinite case (figure F10) is probably still an adequate guide. The fact that history results in a 3.5% increase in gain when $N_1 = 4$, $N_2 = 2$ and $N_1 = 1$ does not, at least, contradict this; little more, however, can be said.

Calculations similar to the above for $V = 0.2$ and $V = 5$ only served to add weight to the comments already made.

3.4.3 The value of history in multi-stage selection

It was seen in part 2 that the value of historical information increases with the number of stages; it will be seen in this section that the same is true in the finite case.

Table T13 gives the gain with and without history for multiple-stage finite selection with $V = 1$. It can be seen that there is a marked increase in the value of history as k increases and that the value seems to be about the same for both $N_1 = 64$ and for $N_1 = 16$ (both for $\pi = 0.0625$). Comparison

with corresponding figures for the infinite case suggests that the value

of history is as great if not greater in the finite case: in four stages

we have, in the finite case, an increase of 5.3% with N_1 = 16 and 4.6% with

TABLE 3.4.6

Values of gain and the use of history for multiple-stage
near-symmetrical screening with N_1 = 100, n = 1 and V = 1.0.

k	Intermediate N_i	G_R/σ (finite target=2.508) with history	G_R/σ (finite target=2.508) without history	% value of history	Var(G)/σ^2 when using history	% value of[**] history in infinite case
1	-	1.773	1.773	-	0.580	0
2	10	2.158±0.014	2.128±0.014*	+1.4	0.360	0.2
3	22,5	2.238±0.012	2.227±0.017*	+0.5	0.339	0.6
4	30,10,3	2.282±0.012	2.294±0.016*	-0.5	0.359	1.0
5	40,16,6,3	2.287±0.012	2.266±0.015*	+0.9	0.327	
6	46,21,10,5,2	2.303±0.011	2.241±0.011	+2.8	0.291	
8	56,32,18,10, 6,3,2	2.318±0.011	2.245±0.011	+3.3	0.312	
10	63,40,25,16, 10,6,4,3,2	2.274±0.011	2.220±0.018*	+2.4	0.283	

* Finney's (1966) results.
** Not available for k > 4.

N_1 = 64; in the infinite case the increase is down to 3.4% (for π = 0.0625).
Because of the error involved, these results are not conclusive, but the
fact that the finite values are consistently larger than the infinite does
give strong support to the argument.

Inspection of table 3.4.6 showing similar figures for π = 0.01 with
N_1 = 100 and V = 1.0, indicates that the decrease in the value of history
found for decreasing π in the infinite case carries over into the finite.

In view of the low (around 1%) value of history for $k < 5$ and the relatively higher value (2-3 percent) for $k > 5$ it appears that the expected increase in the value of history with increasing k does occur, but the values are so small that no more can be said.

Unfortunately, since infinite results are not available for the gain when using history for $k > 4$, little more can be said of the similarity between the finite and infinite cases. As far as the tables go, especially considering how close the error is to the value of additional stages, there is no reason to doubt that the similarity between the infinite and the finite cases continues. Finney (1966) calculated the infinite gain for larger values of k without the use of history; his results support this conclusion.

As far as Var(G) is concerned, neither history nor the number of stages seem to have much effect.

Obviously, since the (k-1)-stage gain can always be expressed as a special case of the k-stage gain, the optimum gain must be a non-decreasing function of k. The fact that this is true for the finite gain up to reasonably large values of k in both table 3.4.6 and table T13, and that we are getting quite close to the maximum gain (only 8% below for 8-stage selection in table 3.4.6), suggests that we are, at least, close to the optimum location. In order to test this theory, the finite 3-stage gain was calculated at nine locations each for $\pi = 0.0625$ and $V = 1$ with $N_1 = 16$ and $N_1 = 64$. The accuracy was about ±0.011 (500 runs) for $N_1 = 64$, $N_3 = 4$ and about ±0.008 (4000 runs) for $N_1 = 16$, $N_3 = 1$. The observed differences in the results at the different locations could easily have been due to the experimental errors; because of the obvious flatness in the surface and of

the time consumed to get the required accuracy, no further 3-stage work was done.

The only other survey work done was for V = 1 and 10-stage selection with N_1 = 100 and N_3 = 1. This number of stages was chosen because of the fact that symmetrical G_R/σ appears to have decreased on moving from 8 stages to 10 (2.318±0.011 to 2.274±0.011) suggesting that the optimum has moved from symmetry. The results of this brief study appear in table T14. Because of the time consumed in calculating the gain to the desired accuracy, and the very large number of possible specifications, only a few calculations were made. In view of the relatively large number of varieties (37) being discarded with very little replication at the symmetrical specifications (column 2 of table T14), the gain was calculated for heavier than symmetric early replication and lighter than symmetric early selection, that is to say for

$$\alpha_1 > \alpha_2 > \ldots > \alpha_{10} \tag{3.4.1}$$

and

$$P_1 \geq P_2 \geq \ldots \geq P_{10} \; . \tag{3.4.2}$$

Table T14 shows that the simulated gain achieved under these circumstances was always as great as, or greater than the symmetric gain; the one value of gain calculated for movement in the opposite direction (column 1) was quite a bit lower than the symmetric gain. With this in mind, it appears that movement in the direction of 3.4.1 and perhaps 3.4.2 will probably do no harm; it may even be beneficial. As usual, the surface is so flat, and the error (after a reasonable number of runs) so large with respect to the possible improvement in gain, that the results cannot prove anything;

they are, however, suggestive of the nature of the variation of gain with operating conditions.

If it were desired to locate the optimum operating conditions accurately, a technique that might be useful would be to compile stage by stage statistics on the places at which varieties with desirable true yields were being rejected; if it turned out that a disproportionate number were being rejected at one or two specific stages, replication in those stages could be increased and/or selection made less intense; by spreading the change evenly over all other stages this approach should help to locate the optimum. The use of a simulation program designed on statistical experimental design principles might also be helpful if very accurate location of the optimum finite gain were desired. In view of the obvious proximity of the symmetrical gain to the optimum, and of the computer time that would be required for such a study, no such attempt was made here.

3.5 Modified Acceptance Rules

3.5.1 Cut-off points throughout

It was pointed out in part 2 that accepting fixed proportions at each stage is, in the infinite case, equivalent to using cut-off points; in the finite case, the use of cut-off points results in variation from cohort to cohort in the number of varieties accepted at a given stage and the similarity is destroyed. Since, in the previous chapter, the finite (fixed proportion) gain was usually found to be quite a bit smaller than the infinite gain it was decided to study the advantage of the use of cut-off points at all stages. For consistency, the cut-off points were chosen to be the values of η_i that correspond to the values of P_i used in the infinite case (and in the previous chapter).

Variation in the number of varieties accepted into a given stage creates a new problem: how are we to replicate these varieties? One method of maintaining consistency is to leave A fixed at the level used in previous chapters, splitting it, as usual, among the various stages according to the values $\alpha_i = A_i/A$ and calculating the error variance for stage i from the standard formula

$$e_i^2 = \frac{VN_i}{\alpha_i N_1} \sigma^2.$$

A difference from previous work now arises: the difference is that e_i^2 is no longer fixed but fluctuates from cohort to cohort as a function of N_i. The fluctuations in N_i may also create practical difficulties in maintaining the desired plot size while at the same time using all the

available resources and replicating each variety equally. The variation
in e_i^2 causes the correlation between two specific stages to vary from
cohort to cohort; the variation in both e_i^2 and ρ_{ij}, in turn, causes
the value of n_i necessary to achieve the desired average value of P_i to
vary from cohort to cohort.

Another possible scheme, also consistent in its way, is to fix the
value of e_i^2 at the desired level (by replicating each variety at the
previous fixed proportion level) and to vary the total number of plots
according to the number of varieties being studied. This will create
fluctuation in the values of A_i (and, as a result, A) from cohort to cohort,
but e_i^2 will remain fixed at the value

$$e_i^2 = \frac{V\pi_{i-1}}{\alpha_i} \sigma^2 \; ,$$

where π_{i-1} is, as usual, the long term (infinite case) average value of
N_i/N_1. This approach will create considerable administrative difficulties
in that the experimental resources required at stage i are unknown until the
end of stage i-1; this may be a very important consideration if the experimenter
has a number of experiments (cohorts) under study and is limited by a fixed
total amount of resources; less difficulty will arise if the experiments
are rapidly completed and if there is a very large supply of the required
experimental material.

Obviously the increased gain resulting from the use of one or the other
of the above methods must be balanced with the difficulty of using it.
Both schemes are studied in this part of the thesis: the first is called
the variable replication scheme and the second the fixed replication scheme.

One comment that should be made with respect to the actual simulation of the two replication schemes is that they were, in order to save computer time, both tested on the same populations and, in addition, the first stage results were made common to both; the results in later stages were also based on the same random numbers but, of course, the error variance was modified according to whether fixed or variable replication was being used. This approach tended to make the results of the two systems more alike than if they had been based on entirely different sets of random numbers. Arguments could be put forward in support of either method of simulation but, especially in view of the saving in computer time, the one used here was felt to be adequate.

For a given set of selection specifications the variation of n from cohort to cohort forces a redefinition of G and Var(G). Since all values of $n \leq N_1$ are now possible, the mean value of all varieties accepted, $\overline{G}$, is used in place of G_R; the variance of the gain (Var(G)) is replaced by the variance (over all runs at a given set of specifications) of all the varieties accepted; and finally, letting N_t represent the total number of varieties accepted in a given set of runs, the variance of the mean gain is $Var(G)/N_t$. Although not entirely satisfactory, these redefined values, in conjunction with the average value of n ($\overline{n}$) over all runs for a given set of specifications, permit adequate comparison of the different screening systems.

It can be seen from table 3.5.1 that, for V = 1.0, the finite gain is very nearly equal to the infinite gain over a wide range of values of π when cut-off points are used at all stages. In addition, judging from the

results of the table, it appears that leaving the amount of replication
per variety (and hence e_i^2) fixed at the infinite case levels may result in
a slight additional increase in gain. It is, however, very difficult,
based only on the results of the table, to say whether either of the

TABLE 3.5.1

The variation in 2-stage finite gain ($V = 1.0$, $\alpha_1 = 0.5$) with
π (N_1) using cut-off points throughout (the cut-off points are
made to correspond to near-symmetrical specifications and $N_3 = 1$).

N_1	$E(G)/\sigma$ for infinite case	Variable replication			Fixed replication		
		$\bar{G}/\sigma$*	$\bar{n}$	$Var(G)/\sigma^2$	$\bar{G}/\sigma$*	$\bar{n}$	$Var(G)/\sigma^2$
4	0.956	0.922	0.946	0.536	0.950	0.922	0.515
9	1.347	1.328	0.972	0.510	1.371	0.936	0.491
16	1.605	1.574	0.966	0.411	1.610	0.937	0.403
32	1.898	1.869	1.065	0.334	1.891	1.045	0.318
64	2.179	2.214	0.907	0.302	2.222	0.893	0.295
128	2.442	2.369	0.940	0.210	2.386	0.929	0.200
256	2.687	2.675	1.056	0.190	2.691	1.041	0.182
512	2.918	2.935	1.000	0.161	2.930	1.012	0.160

* All values of mean gain accurate to a
standard deviation of approximately ±0.016.

above statements is true due to the fact (as usual) that the standard error
of the mean gain (approximately ±0.015) is of the order of the observed
difference; the difficulty of comparison is increased further by the fact
that $\bar{n}$ very rarely equals πN_1.

Table 3.5.2, showing the same results for various values of V with $N_1 = 16$
and $\pi = 0.0625$, suggests the same two conclusions; again, for the same reasons,
it is impossible to make them with certainty. In any case, the same general

behaviour does appear to occur over a very wide range of both π and V.

As can be expected with the errors involved, surveys of the optimum region (for π = 0.01 and 0.0625 with V = 0.2, 1.0, and 5.0) did not indicate any reason to doubt the optimality of the use of symmetrical specifications as the framework for the operation of the cut-off selection scheme.

TABLE 3.5.2

The variation in 2-stage finite gain (α_1 = 0.5) with V for N_1 = 16 and using cut-off points corresponding to N_2 = 4 and N_3 = 1.

V	$E(G)/\sigma$ for infinite case	Variable replication			Fixed replication		
		$\bar{G}/\sigma*$	Avg.n	$Var(G)/\sigma^2$	$\bar{G}/\sigma*$	Avg.n	$Var\ G/\sigma^2$
5^{-3}	1.965	1.978	1.045	0.147	1.980	1.043	0.146
5^{-2}	1.952	1.966	1.041	0.179	1.970	1.037	0.178
5^{-1}	1.889	1.898	0.988	0.230	1.903	0.988	0.227
$5^{-\frac{1}{2}}$	1.789	1.782	1.055	0.330	1.796	1.046	0.318
1	1.605	1.574	0.966	0.411	1.610	0.937	0.403
$5^{\frac{1}{2}}$	1.340	1.338	1.010	0.628	1.380	0.951	0.613
5	1.039	1.023	0.996	0.748	1.072	0.930	0.722
5^2	0.529	0.531	1.023	0.912	0.563	0.932	0.919
5^3	0.244	0.246	1.004	0.989	0.214	0.936	0.990
5^4	0.100	0.104	0.997	(1.007)	0.113	0.901	0.994

* All values of mean gain accurate to a standard deviation of approximately ±0.015.

Similar, but necessarily briefer, studies of the same scheme for 3-stage selection with N_1 = 64 and π = 0.0625, and with N_1 = 100 and π = 0.01, both for V = 1.0 suggested similar conclusions: at symmetry, the gain with constant replication was, in the former case, 1.672±0.016 versus an infinite

gain of 1.657 and, in the latter case, 2.399±0.015 versus an infinite gain
of 2.439. Once again, there is no reason to doubt that very nearly
infinite gain will be achieved by using cut-off points throughout.

Although it is impossible to reach definite conclusions based only on
the preceding results, it can be seen from consideration of what is actually
happening that we must approach the infinite case. This is so because the
idea of a cohort becomes less relevant when using cut-off points since, in
a sense, varieties are no longer being assessed on their merit relative
to other varieties in a given cohort but on their merit relative to an
absolute standard. In our study, this standard is chosen to accept approximately
the fraction π of all the varieties studied in all runs at a given set of
operating conditions. The result is that, in effect, we are selecting from
one very large cohort with $N_1' =$ the number of runs x the number of varieties
studied per run and, correspondingly, $n' = \pi N_1'$. In view of the size of N_1'
and n', and of the rapid approach to the infinite results as N_1 and n are
increased (see table 3.4.2), it is obvious that we will, over the long run,
achieve nearly infinite gain when using cut-off points. We can also expect
the gain to be slightly larger when using a level of replication fixed at
the level found to be optimum in the infinite case (the symmetrical level),
than when A and A_i are left fixed and replication is varied to accommodate
the variation in N_i; the slight increase in gain will, however, rarely be
worth the trouble.

When the above considerations and the previous results are both taken
into account, it can be seen that we are safe in concluding that very
nearly infinite gain will be achieved in the finite case if cut-off points

are used instead of fixed proportions. This in turn means that, for small N_1 and especially small n, we will achieve much higher long-term finite gain than is achieved by always accepting fixed proportions. The comparisons of finite (fixed proportion) gain with infinite gain in chapter 3.4 indicate the size of the increases that should be achieved.

The use of cut-off points is, however, not without its difficulties. The fact that 30 to 40 percent of the individual runs result in the rejection of all the varieties will cause problems if it is very important to come up with something quickly. On the other hand, many runs result in the acceptance of many more than the desired fraction πN_1; this may create practical difficulties if the follow-up facilities are limited. A scheme in which a maximum of some fixed number of varieties is accepted provided each yields above some fixed cut-off point might solve some of these problems. It would, however, not usually, unless the cut-off point were chosen by some method beyond the scope of this thesis, result in the desired long term fraction π of the varieties being studied; comparison with other selection schemes would therefore be difficult.

In addition to the difficulties outlined above, it will frequently be difficult to establish the desired cut-off points. Although we have seen that the infinite gain does not, as a percent of target, change much with variation in the parent distribution, the cut-off points are very much functions of the parent distribution. This means that they will rarely be available. One exception to this is that the use of a standard, or even the experimenter's knowledge of the situation, may make a final cut-off point reasonable.

If, however, the main purpose of the selection program is to achieve overall maximum gain rather than maximum average gain for each cohort, one should, in spite of these difficulties, do one's best to establish appropriate cut-off points rather than use fixed proportions. This approach will be particularly important if varieties are presented for study one at a time. In view of the obvious difficulties in the creation and use of such schemes, the next two chapters will investigate modified schemes based partly on fixed proportions and partly on cut-off points.

3.5.2 Intermediate cut-off points with n fixed

Because of the possibility of limited follow-up facilities, selection schemes will be studied in this section in which n is fixed but which use cut-off points at all intermediate stages. Since n is fixed in every cohort, the original definitions of gain (G_R/σ) and Var(G) will be used rather than the modified definitions of the previous chapter. As in the previous chapter, both constant and variable replication will be employed.

The results in table 3.5.3 for 2-stage selection with V = 1 and various values of N_1 (n fixed at 1) suggest that, when a fixed proportion is accepted from the final stage, there is very little to choose between using a cut-off point at the end of stage one and accepting a fixed proportion at the end of stage one.

In order to determine whether this result depends on the size of n, selection was performed for π = 0.0625 and π = 0.01 with various values of N_1 for each. Once again, the results were very close to similar studies for selection when fixed proportions are accepted throughout (see table 3.4.2

for $\pi = 0.0625$).

The results for $N_1 = 16$ and $n = 1$ in table 3.5.4 suggest that the two acceptance schemes are also similar over a wide range of values of V. Surveys of the optimum area for various values of π and V only served to confirm this similarity.

TABLE 3.5.3

The variation in 2-stage finite gain ($\alpha_1 = 0.5$, $V = 1.0$) with π (N_1) for $n = 1$, but with a cut-off point (made to correspond with the value of N_2 closest to $P_1 = \sqrt{\pi}$) at stage 1.

| N_1 | $E(G)/\sigma$ for infinite case | Fixed proportions at both stages | | Cut-off at stage 1 | | | |
| | | | | variable replication | | fixed replication | |
		G_R/σ	$Var(G)/\sigma^2$	G_R/σ^*	$Var(G)/\sigma^2$	G_R/σ^*	$Var(G)/\sigma^2$
4	1.956	0.744±0.010	0.744	0.739	0.715	0.754	0.708
9	1.347	1.131±0.013	0.646	1.115	0.700	1.135	0.692
16	1.605	1.401±0.004	0.551	1.374	0.566	1.386	0.566
32	1.898	1.697±0.013	0.464	1.685	0.471	1.688	0.481
64	2.179	1.969±0.014	0.395	1.978	0.444	1.973	0.454
128	2.442	2.270±0.016	0.382	2.235	0.308	2.235	0.314
256	2.687	2.489±0.015	0.328	2.497	0.283	2.499	0.289
512	2.918	2.767±0.011	0.273	2.777	0.251	2.762	0.260

* Mean gain accurate to a standard deviation of approximately ±0.014.

In order to get some idea whether the same conclusions apply for more than two stages of selection, a brief study was made of 3-stage selection in which n is fixed and intermediate cut-off points used. The results were almost identical to the results of 3-stage selection in which fixed proportions are used throughout.

In both tables in this section (as with the two tables in the previous

section) there is little to choose between fixed and variable replication*.

Once again the gain for fixed replication does seem to be slightly higher

TABLE 3.5.4

The variation in 2-stage finite gain ($\alpha_1 = 0.5$) with V for $N_1 = 16$ and $N_3 = 1$ but with a cut-off point corresponding to $N_2 = 4$ at stage 1.

V	$E(G)/\sigma$ for infinite case	Fixed proportions at both stages		Cut-off at stage 1			
				variable replication		fixed replication	
		G_R/σ	Var(G)	$G_R/\sigma*$	Var(G)/σ^2	$G_R/\sigma*$	Var(G)/σ^2
0	1.968	1.766		1.766		1.766	
5^{-3}	1.965	1.759±0.009	0.306	1.764	0.323	1.764	0.321
5^{-2}	1.952	1.739±0.008	0.294	1.746	0.351	1.744	0.352
5^{-1}	1.889	1.684±0.004	0.374	1.680	0.399	1.686	0.396
$5^{-\frac{1}{2}}$	1.789	1.567±0.010	0.426	1.562	0.498	1.564	0.498
1	1.605	1.401±0.004	0.556	1.374	0.566	1.386	0.566
$5^{\frac{1}{2}}$	1.340	1.162±0.009	0.686	1.126	0.743	1.157	0.757
5	1.039	0.888±0.006	0.793	0.857	0.840	0.876	0.735
5^2	0.529	0.467±0.010	0.948	0.452	0.926	0.474	0.921
5^3	0.244	0.225±0.011	0.941	0.218	0.994	0.211	1.000
5^4	0.100	0.095±0.012	0.990	0.093	1.010	0.090	0.998

* Mean gain accurate to a standard deviation of approximately ±0.012.

than that for variable replication but the difference, if it does in fact

exist, is so slight that it is not worth consideration.

In keeping with all the above comments, comparison of the values of

---

* Keep in mind that, because of the system used for simulation, the gains and variances in comparisons of fixed and variable replication will always be more alike than the gains and variances in comparison of the two first-stage acceptance rules.

$Var(G)/\sigma^2$ in tables 3.5.3 and 3.5.4 does nothing to indicate any difference between the use of cut-off points and the use of fixed proportions at the end of the first stage, nor is there any apparent difference between fixed and variable replication.

In view of the results of this section it seems safe, when it is necessary to accept a fixed number of varieties from the final stage, to recommend the acceptance of a fixed number of varieties at each intermediate stage as well. The extreme difficulty of establishing appropriate intermediate cut-off points makes this conclusion a welcome one.

3.5.3 Cut-off point at the final stage only

Because of the ease of operation of systems based on accepting fixed proportions, and of the fact that it will frequently be possible, by using a standard or by some sort of economic analysis, to establish a final cut-off point, the operation of a selection system using a cut-off point at the last stage and accepting fixed proportions at all intermediate stages will be studied in this section. As in section 3.5.1, we use $\bar{G}$ rather than G_R since n varies from cohort to cohort as a result of the reliance on a cut-off point. In addition, we define the quantity $\bar{\pi} = \bar{n}/N_1$; by calculating the infinite gain corresponding to a selection fraction $\bar{\pi}$, we compensate for inflation or deflation of the finite gain resulting from values of $\bar{n}$ smaller or larger (respectively) than target. This permits comparisons to be made on a more equal basis.

Table 3.5.5, showing the gain achieved for various values of N_1 when accepting approximately $N_2 = \sqrt{N_1}$ varieties after stage 1, indicates that

close to infinite gain is achieved under this modified system. That the
same is true over a wide range of values of V (for N_1 = 16, N_2 = 4 and
σ_1 = 0.5) can be seen in table 3.5.6. Although the infinite (modified)
gain is very close to the finite gain in all cases in both tables, it
appears that the finite gain may be slightly lower; this, however, is far

TABLE 3.5.5

The effect on the 2-stage finite gain of using a cut-off point
at the final stage, for various values of N_1 (π) and V = 1 (the
cut-off point was chosen to correspond to a long term average
of $\overline{n}$ = 1).

N_1	N_2	Infinite case $E(G)/\sigma$ for $\overline{\pi} = \overline{n}/N_1$	Final cut-off			Finite fixed proportions G_R/σ for	
			$\overline{G}/\sigma$	$Var(G)/\sigma^2$	$\overline{n}$	fixed n=1	$Var(G)/\sigma^2$
4	2	0.958	0.922±0.011	0.568	0.996	0.744±0.010	0.744
9	3	1.359	1.323±0.011	0.495	0.973	1.131±0.013	0.646
16	4	1.608	1.615±0.007	0.425	0.994	1.401±0.004	0.551
32	6	1.901	1.881±0.011	0.331	0.992	1.697±0.013	0.464
64	8	2.202	2.159±0.011	0.274	0.940	1.969±0.014	0.395
128	11	2.428	2.442±0.012	0.255	1.039	2.270±0.016	0.322
256	16	2.673	2.637±0.013	0.176	1.042	2.489±0.015	0.328

from certain when the magnitude of the standard deviation of the mean
gain is considered. In any case, it is obvious from both tables that the
gain under the modified system is much larger than that achieved when
accepting fixed proportions at both stages. It can also be seen that, unless
V is large, $Var(G)/\sigma^2$ is smaller in the new system than in the original
fixed proportion system. There is, however, the drawback that 30-35 percent
of all selections result in no varieties being recommended for acceptance.

The same sort of behaviour is found in 3-stage selection when fixed proportions are accepted at the end of the first and second stages and a cut-off point used at the final stage. In this case, symmetrical selection

TABLE 3.5.6

The effect on the 2-stage finite gain of using a cut-off point at the last stage for $N_1 = 16$, $N_2 = 4$, $\alpha_1 = 0.5$ and various values of V (the cut-off point was chosen to correspond to a long term average of $\bar{n} = 1$).

V	Infinite case $E(G)/\sigma$ for $\bar{\pi} = \bar{n}/N_1$	Final cut-off $\bar{G}/\sigma$	$Var(G)/\sigma^2$	$\bar{n}$	Finite fixed proportions G_R/σ for fixed n=1	$Var(G)/\sigma^2$
5^{-3}	1.962	1.938±0.015	0.133	1.007	1.759±0.009	0.306
5^{-2}	1.988	1.945±0.013	0.165	0.919	1.739±0.008	0.294
5^{-1}	1.918	1.920±0.019	0.247	0.934	1.684±0.004	0.374
$5^{-\frac{1}{2}}$	1.791	1.775±0.014	0.280	0.997	1.567±0.010	0.426
1	1.608	1.338±0.015	0.593	0.994	1.401±0.004	0.551
$5^{\frac{1}{2}}$	1.312	1.338±0.015	0.593	1.077	1.162±0.009	0.686
5^1	1.039	0.983±0.016	0.741	0.998	0.888±0.006	0.793
5^2	0.529	0.518±0.015	0.921	1.002	0.467±0.010	0.948
5^3	0.246	0.254±0.014	0.993	0.977	0.225±0.011	0.941
5^4	0.113	0.125±0.013	0.998	0.916	0.095±0.012	0.990

for $V = 1$ and $N_1 = 100$ (target value of $\pi = 0.01$) resulted in a finite gain of 2.426±0.012 versus an infinite gain (for $\bar{\pi} = \bar{n}/N_1 = 1.004/100$) of 2.436. Other 3-stage results (for various values of V and π) also agreed very closely with the infinite case.

Comparison of the results of this section with those of sections 3.5.1

and 3.5.2 indicates that the main damage is done to the gain when a fixed
proportion of varieties is accepted from the final stage; what is done at
intermediate stages seems to be of much less importance. In view of this,
and of the difficulty that will likely be met in establishing the intermediate
cut-off points that correspond to the desired symmetrical selection specifi-
cations, the most acceptable selection procedure will probably be the system
investigated in this section. Special circumstances may, of course, dictate
the necessity of accepting some varieties from every cohort thus prohibiting
the use of this modified scheme.

3.5.4 Final acceptance prior to the final stage

In section 2.4.1 we found that little or no increase in gain occurs
in the infinite case if some varieties are accepted, without further study,
immediately after the first stage. There is however the advantage,
especially if π is small, of getting a few varieties into service, or into
the next phase of the study, without any corresponding decrease in gain.
For this reason a very brief study was made, for small values of π, of
early acceptance in finite selection.

Selection in which a relatively large number of varieties is desired
was chosen for the study since, for very small values of n, early acceptance
makes little or no sense. In any case, early acceptance from all but
extremely large populations must, if it is to be of value, be based on
cut-off points rather than fixed proportions and, when using cut-off points,
the idea of a cohort has been seen to lose its significance. In this study,
the early acceptance was based on a cut-off point while both the number of

varieties rejected at the first stage, and the total number of varieties
finally accepted were fixed.

Table 3.5.7 gives the results for N_1 = 100 (π = 0.25) and for
N_1 = 256 (π = 0.0625). In both cases, the populations are so large that
the finite gain is very nearly as large as the infinite gain. As was the

TABLE 3.5.7

The gain when some varieties are accepted immediately after
the first stage in 2-stage selection for V = 1, α_1 = 0.5.

100 → 50 → 25				256 → 64 → 16			
No.accepted early		finite	infinite	No.accepted early		finite	infinite
Targ.	Actual	gain*	gain	Targ.	Actual	gain*	gain
0	0	0.937	0.956	0	0	1.591	1.605
1	1.027	0.950	0.958	1	0.828	1.574	1.603
2	1.838	0.942	0.959	2	2.028	1.585	1.596
3	2.945	0.957	0.961	3	2.690	1.583	1.587
4	3.965	0.949	0.961	4	4.028	1.566	1.575
5	4.950	0.948	0.962	5	4.788	1.551	1.560
6	6.080	0.942	0.962	6	5.972	1.537	1.542
10	9.572	0.933	0.956	7	7.060	1.498	1.522
12	11.605	0.930	0.948	8	8.228	1.471	1.500

* Standard error of mean gain $\leq$ 0.009.

case in the infinite study when using large values of π, early acceptance
of a large number of varieties did not seem to result in much of a decrease
in gain: for π = 0.25, early acceptance of 12 varieties, or nearly half
the total desired number, resulted in a finite gain of 0.930σ, only a few
units lower than the gain (0.937σ) with no early acceptance. It can also
be seen that there may even be an increase in gain when 3-4 varieties are
accepted early. The results for π = 0.0625 illustrate the rapid decrease

in the value of early acceptance as π decreases: early acceptance of more than 2-3 varieties in this case appears to result in a decrease in gain. When the size of the standard error is considered, there is no reason to suspect that the behaviour in the finite case departs very much from the infinite case.

Various other combinations of cut-off points and fixed proportions could be studied but, in view of the results of previous sections and of the close correspondence between the finite and the infinite case in this section, nothing very startling can be expected.

3.6 Concluding Remarks

It is apparent from both the preceding work and Finney's 1966 paper that, although the gains frequently differ in magnitude, the infinite and finite cases are very similar in behaviour.

Because of the inconclusive results in chapter 3.4 on the value of history it is worth pointing out that properly weighted history will, for obvious reasons, always result in some improvement in gain. If, however, special circumstances make the use of history difficult or costly, we have seen that it can, at the optimum operating specifications, usually be ignored with little effect on the finite gain. Caution is, however, necessary in view of the increased importance of history under non-ideal circumstances (see section 2.5.1 on the effect of interaction). In general then, one should do the obvious and carefully consider any great changes in the yield of a specific variety from stage to stage; in addition, if accurate estimates of the optimum weights are available, history should be included in the estimate of x throughout.

The work of chapter 3.5 has shown that when cut-off points are used (even if only at the final stage), the infinite case is a good guide to the actual magnitude of the average finite gain. This means that there will be a large increase in the expected finite gain when N_1 and the target value of n are relatively small. This fact is most certainly worth taking into account when designing a selection program.

One general recommendation that seems very safe, especially in view of the magnitude of $Var(G)/\sigma^2$ (relative to $\bar{G}/\sigma$), is the use of the symmetrical selection specifications; they should give very nearly optimum gain in

finite selection from a normal population under all circumstances. The
close proximity found in chapter 2.6 between the results of selection from
infinite normal and (continuous) infinite non-normal distributions, and
in the last two chapters between the finite and the infinite cases,
suggests that this recommendation will also apply to a wide range of other
distributions. In view of the differences between the infinite 2-point
distribution and the other distributions studied, simulation of that case
should be done; once again the infinite case will probably be a good guide
to the finite.

BIBLIOGRAPHY

BECHHOFER, R.E. (1954): A single-sample multiple decision procedure for ranking means of normal populations with known variances. Ann. Math. Stat. $\underline{25}$, 16-39.

BECHHOFER, R.E. (1958): A sequential multiple-decision procedure for selecting the best one of several normal populations with a common unknown variance, and its use with various experimental designs. Biometrics $\underline{14}$, 408-429.

BECHHOFER, R.E. (1966): Multiple-decision procedures for ranking means of normal populations with a common variance. Unpublished paper. Cornell University.

BOX, G.E.P. and MULLER, M.E. (1958): A note on the generation of random normal deviates. Ann. Math. Stat. $\underline{29}$, 610-611.

BOX, G.E.P. and TIAO, G.C. (1962): A further look at robustness via Bayes theorem. Biometrika $\underline{49}$, 419-432.

COCHRAN, W.G. (1951): Improvement by means of selection. Proc. Second Berkley Symposium Math. Stat. and Prob., 449-470.

CRONBACH, L.J. and GLESER, G.C. (1965): Psychological tests and personnel decisions. 2nd Ed. Univ. Illinois Press, Urbana.

CURNOW, R.N. (1959): The statistical theory of selection. Unpublished Ph.D. Thesis, University of Aberdeen.

CURNOW, R.N. (1960): The consequences of errors of measurement for selection from certain non-normal distributions. Bull. de l'Inst. Int. Stat. $\underline{37}$, 291-308.

CURNOW, R.N. (1961): Optimal programmes for varietal selection. J. Roy. Stat. Soc., $\underline{B23}$, 282-318.

DAVIES, O.L. (1958): The design of screening tests in the pharmaceutical industry. Bull. de l'Inst. Int. Stat. 36, 226-241.

DAVIES, O.L. (1964): Screening for improved mutants. Biometrics 20, 576-591.

DUNNETT, C.W. (1960): On selecting the largest of k normal population means. J. Roy. Stat. Soc. B22, 1-40.

DUNNETT, C.W. (1961): Approaches to some problems in drug screening and selection. Unpublished paper presented to the Gordon Research Conference on Statistics in Chemistry and Chemical Engineering.

FEDERER, W.T. (1963): Procedures and designs useful for screening material in selection and allocation, with a bibliography. Biometrics 19, 553-587.

FINNEY, D.J. (1956): The consequences of selection for a variate subject to errors of measurement. Revue de l'Inst. Int. Stat. 24, 1-10.

FINNEY, D.J. (1958a): Statistical problems of plant selection. Bull. de l'Inst. Int. Stat. 36, 242-268.

FINNEY, D.J. (1958b): Plant selection for yield improvement. Euphytica 7, 83-106.

FINNEY, D.J. (1960a): A comment in the discussion of Dunnett (1960), page 38.

FINNEY, D.J. (1960b): A simple example of the external economy of varietal selection. Bull. de l'Inst. Int. Stat. 37, 91-106.

FINNEY, D.J. (1961): The transformation of a variate under selection. Sankhya A23, 309-324.

FINNEY, D.J. (1962a): Cumulants of truncated multinormal distributions. J. Roy. Stat. Soc. B24, 535-536.

FINNEY, D.J. (1962b): The statistical evaluation of educational allocation and selection. J. Roy. Stat. Soc. A125, 525-564.

FINNEY, D.J. (1963): Some properties of a distribution specified by its cumulants. Technometrics 5, 63-69.

FINNEY, D.J. (1964): Screening processes: Problems and illustrations. Contributions to Statistics presented to P. C. Mahalanobis, 73-100.

FINNEY, D.J. (1966): An experimental study of certain screening processes. J. Roy. Stat. Soc. B28, 88-109.

FISHER, R.A. (1959): Statistical methods and scientific inference. 2nd Ed. Oliver and Boyd. Edinburgh.

GRUNDY, P.M., HEALY, M.J.R. and REES, D.H. (1956): Economic choice of the amount of experimentation. J. Roy. Stat. Soc. B18, 32-55.

HARTER, H.L. (1960): Expected values of normal order statistics. A.R.L. Technical Report 60-292, U.S. Aeronautical Research Laboratories.

KING, E.P. (1963): A statistical design for drug screening. Biometrics 19, 429-440.

KING, E.P. (1964): Optimal replication in sequential drug screening. Biometrika 51, 1-10.

KRYLOV, V.I. (1962): Approximate calculation of integrals. MacMillan.

NISSEN-MEYER, S. (1964): Evaluation of screening tests in medical diagnosis. Biometrics 20, 730-755.

OWEN, D.B. (1956): Tables for computing bivariate normal probabilities. Ann. Math. Stat. 27, 1075-1090.

SCARBOROUGH, J.B. (1958): Numerical mathematical analysis. 4th Ed. The Johns Hopkins Press.

SMITH, H.F. (1938): An empirical law describing heterogeneity in the yields of agricultural crops. J. Agric. Sci. Camb. <u>28</u>, 1-23.

SPRAGUE, G.F. and FEDERER, W.T. (1951): A comparison of variance components in corn yield trials: II. Error, year × variety, location × variety, and variety components. Agronomy Journal <u>43</u>, 535-541.

STECK, G.P. (1958): A table for trivariate normal probabilities. Ann. Math. Stat. <u>29</u>, 780-800.

STROUD, A.H. and SECREST, D. (1966): Gaussian quadrature formulas. Prentice-Hall. Englewood Cliffs, N.J.

TEICHROEW, D. (1956): Tables of expected values of order statistics for samples of size twenty and less from the normal distribution. Ann. Math. Stat. <u>27</u>, 410-426.

YATES, F. (1950): Experimental techniques in plant improvement. Biometrics <u>6</u>, 200-207.

YATES, F. and COCHRAN, W.G. (1938): The analysis of groups of experiments. J. Agric. Sci. Camb. <u>28</u>, 556-580.

APPENDIX 1

The calculation of the derivatives of a multivariate

normal probability integral

The derivative

$$\left(\frac{d}{dt}\right)^n \int_{\frac{\eta_1-t}{\omega_1}}^{\infty} \cdots \int_{\frac{\eta_k-t}{\omega_k}}^{\infty} \exp(-\tfrac{1}{2}vR^{-1}v')/[(2\pi)^k|R|]^{\frac{1}{2}}dv_1\ldots dv_k$$

plays a **very** important part in chapter 2.2 in the derivation of the nth moment of the distribution of x after k stages of screening. Since I have not found an expression for this derivative in any standard text I will give its derivation here.

Using the notation of chapter 2.2, particularly expressions 2.2.2 and 2.2.11, it can be seen that

$$I[\eta(t);R]= \int_{\eta_m(t)}^{\infty} \phi(v_m)I^m\left(\frac{\eta_1(t)-\rho_{1m}v_m}{\sqrt{1-\rho_{1m}^2}}, \ldots, \frac{\eta_{m-1}(t)-\rho_{m-1m}v_m}{\sqrt{1-\rho_{m-1m}^2}}, \frac{\eta_{m+1}(t)-\rho_{m+1m}v_m}{\sqrt{1-\rho_{m+1m}^2}}, \right.$$
$$\left. \ldots, \frac{\eta_k(t)-\rho_{km}v_m}{\sqrt{1-\rho_{km}^2}};R^m\right)dv_m$$

and hence that,

$$\frac{dI[\eta(t);R]}{d\eta_m(t)} = -\phi(\eta_m(t))I^m(\eta^m(t);R^m). \tag{A1}$$

Keeping this in mind and remembering, from elementary calculus, that the derivative of a composite function,

$$u = u(x_1(t),\ldots,x_n(t)),$$

is given by

$$\frac{du}{dt} = \sum_{i=1}^{n} \frac{\partial u}{\partial x_i(t)} \frac{dx_i(t)}{dt} \quad , \tag{A2}$$

we have, from A1 and A2,

$$\frac{dI[\eta(t);R]}{dt} = \sum_{i=1}^{k} \frac{\partial I[\eta(t);R]}{\partial \eta_i(t)} \frac{d\eta_i(t)}{dt}$$

$$= \sum_{i=1}^{k} \phi(\eta_i(t)) I^i(\eta^i(t);R^i)/m_i. \tag{A3}$$

In order to show the pattern that is developing, the second derivative will also be found. Referring again to the terminology of 2.2, especially 2.2.14, we have

$$I^m(\eta^m(t);R^m) = \int_{\eta_n^m(t)}^{\infty} \phi(v_n^m) I^{mn}(\eta^{mn}(t;v_n^m);R^{mn}) dv_n^m \quad (n \neq m), \tag{A4}$$

where $\eta^{mn}(t;v_n^m)$ is taken to be a k-2 element vector with element i equal to $(\eta_i^m(t) - \rho_{in}^m v_n^m)/\sqrt{1-(\rho_{in}^m)^2}$ and with m^{th} and n^{th} elements missing. This time we have

$$\frac{dI^m(\eta^m(t);R^m)}{d\eta_n^m(t)} = -\phi(\eta_n^m(t)) I^{mn}(\eta^{mn}(t);R^{mn}) \quad (n \neq m). \tag{A5}$$

Making use of the previous method and equation A3, we have,

$$\frac{d^2}{dt^2} I[\eta(t);R] = \frac{d}{dt} \sum_{i=1}^{k} \phi(\eta_i(t)) I^i(\eta^i(t);R^i)/m_i$$

$$= \sum_{i=1}^{k} (\eta_i - t)\phi(\eta_i(t)) I^i(\eta^i(t);R^i)/m_i^3$$

$$+ \sum_{i=1}^{k} \phi(\eta_i(t))/m_i \sum_{\substack{j=1 \\ j \neq i}}^{k} \frac{\partial I^i(\eta^i(t);R^i)}{\partial \eta_j^i(t)} \frac{d\eta_j^i(t)}{dt}$$

$$= \sum_{i=1}^{k} \emptyset(\eta_i(t))/m_i \left[(\eta_i - t) I^i(\eta^i(t); R^i)/m_i^2 \right.$$

$$\left. + \sum_{\substack{j=1 \\ j \neq i}}^{k} \emptyset(\eta_j(t)) (\frac{1}{m_j} - \frac{\rho_{ij}}{m_i}) I^{ij}(\eta^{ij}(t); R^{ij})/ \sqrt{1 - \rho_{ij}^2} \right] . \quad (A6)$$

Keeping in mind the fact that

$$I^{i_1 i_2 \cdots i_k} \left[\eta^{i_1 i_2 \cdots i_k}(t); R^{i_1 i_2 \cdots i_k} \right] = 1,$$

in other words the fact that the result is 1 when the number of superscripts (the number of integrals left out) is equal to the order of the integral, higher order derivatives can be calculated using extensions of the basic notation. They are not used in this thesis and so will not be developed here.

FIGURE F1. Dependence of G/σ on α_1 and P_1 for 2-stage selection with $\pi = 0.01$, $V = 0.1$ and $F_o = 1$. Contours show G/σ expressed as a percentage of the maximum attainable gain (target).
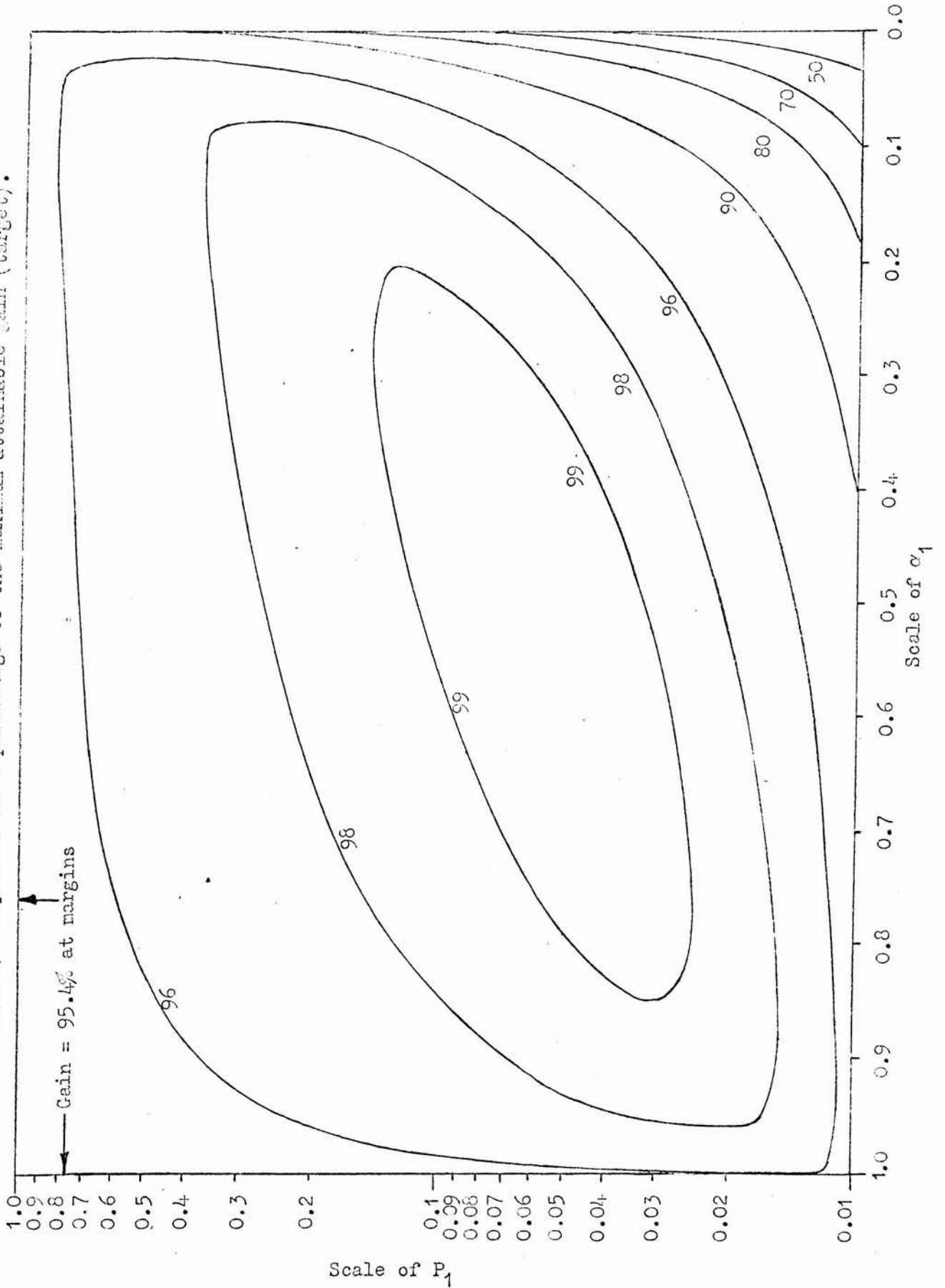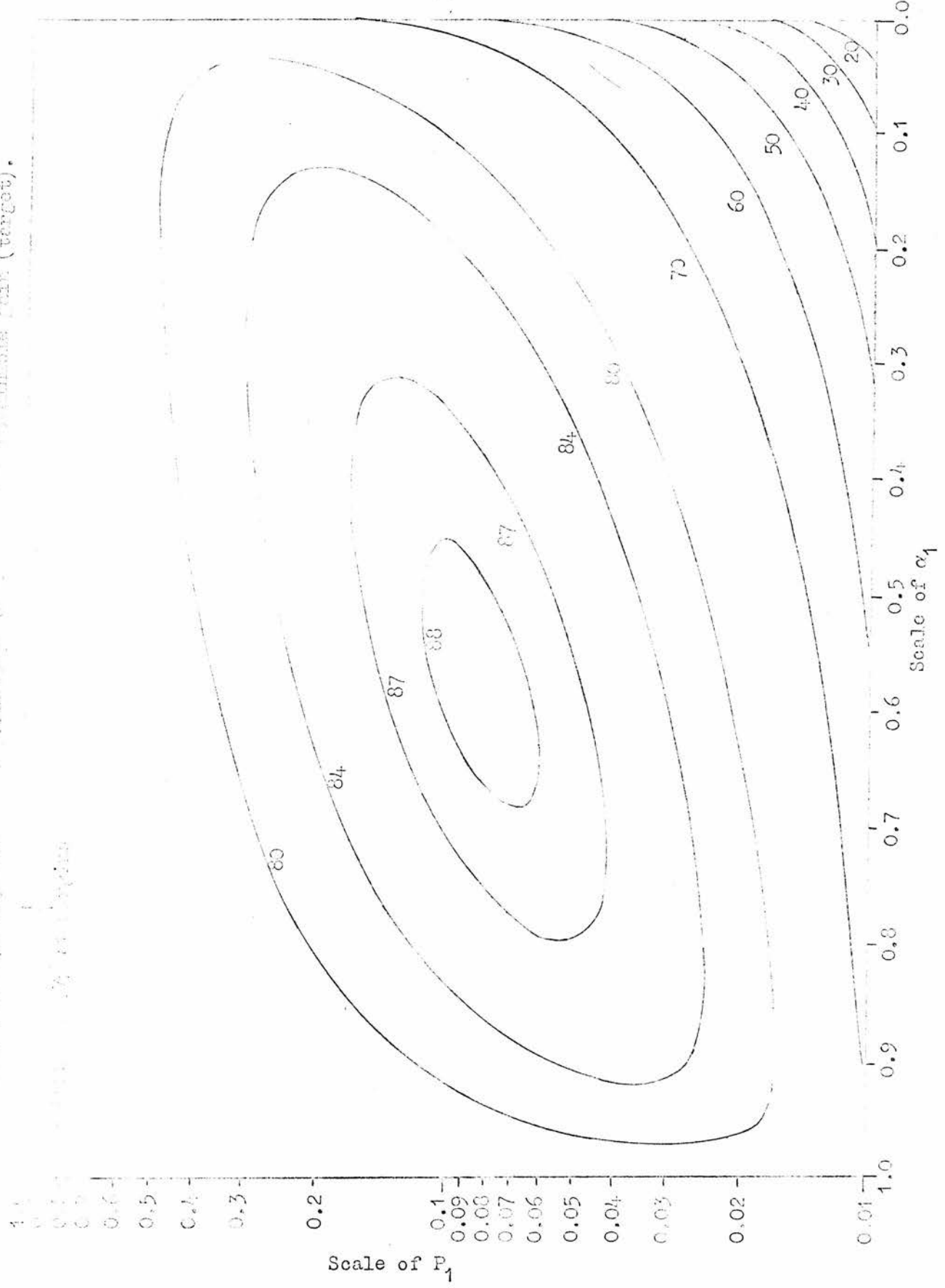
FIGURE F3. Dependence of G/σ on α_1 and P_1 for 2-stage selection with $\pi = 0.01$, $V = 10.0$ and $P_o = 1$. Contours show G/σ expressed as a percentage of the maximum attainable gain (target).

FIGURE F4. Dependence of G/σ on V for symmetrical selection from a normal distribution, π = 0.0625.

FIGURE 35. Dependence of G/σ on V for symmetrical selection from a normal distribution, π = 0.01.

C — 3-stage selection
B — 2-stage selection
A — 1-stage selection

Scale of V

Scale of G/σ
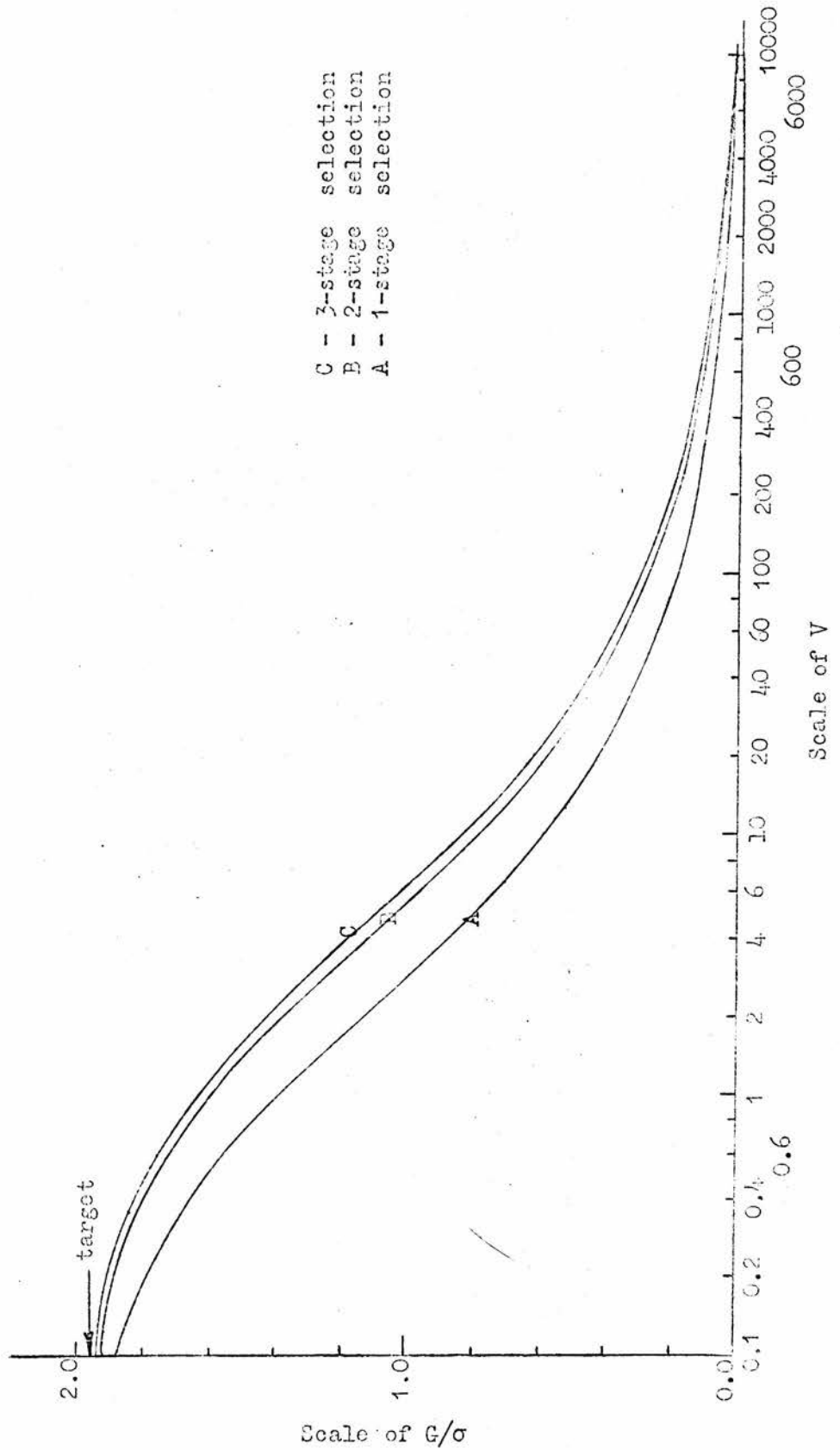
FIGURE F6. Dependence of G/σ on V for symmetrical selection from a normal distribution, π = 0.0001.

D — 4-stage selection
C — 3-stage selection
B — 2-stage selection
A — 1-stage selection
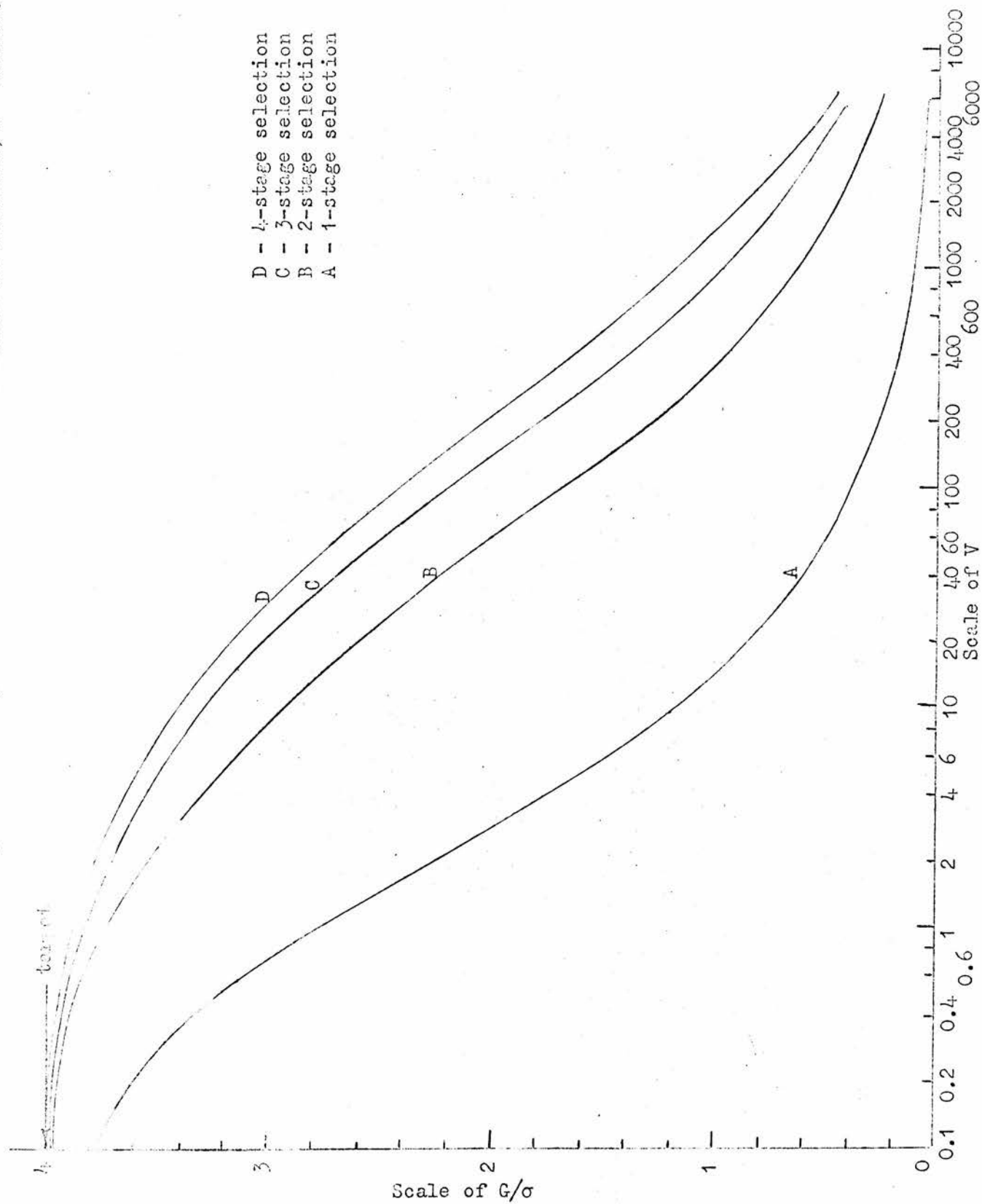
FIGURE F7. Dependence of G/σ on π for symmetrical selection from a normal distribution, V = 0.2.

D — target
C — 3-stage selection
B — 2-stage selection
A — 1-stage selection

Scale of G/σ

Scale of π
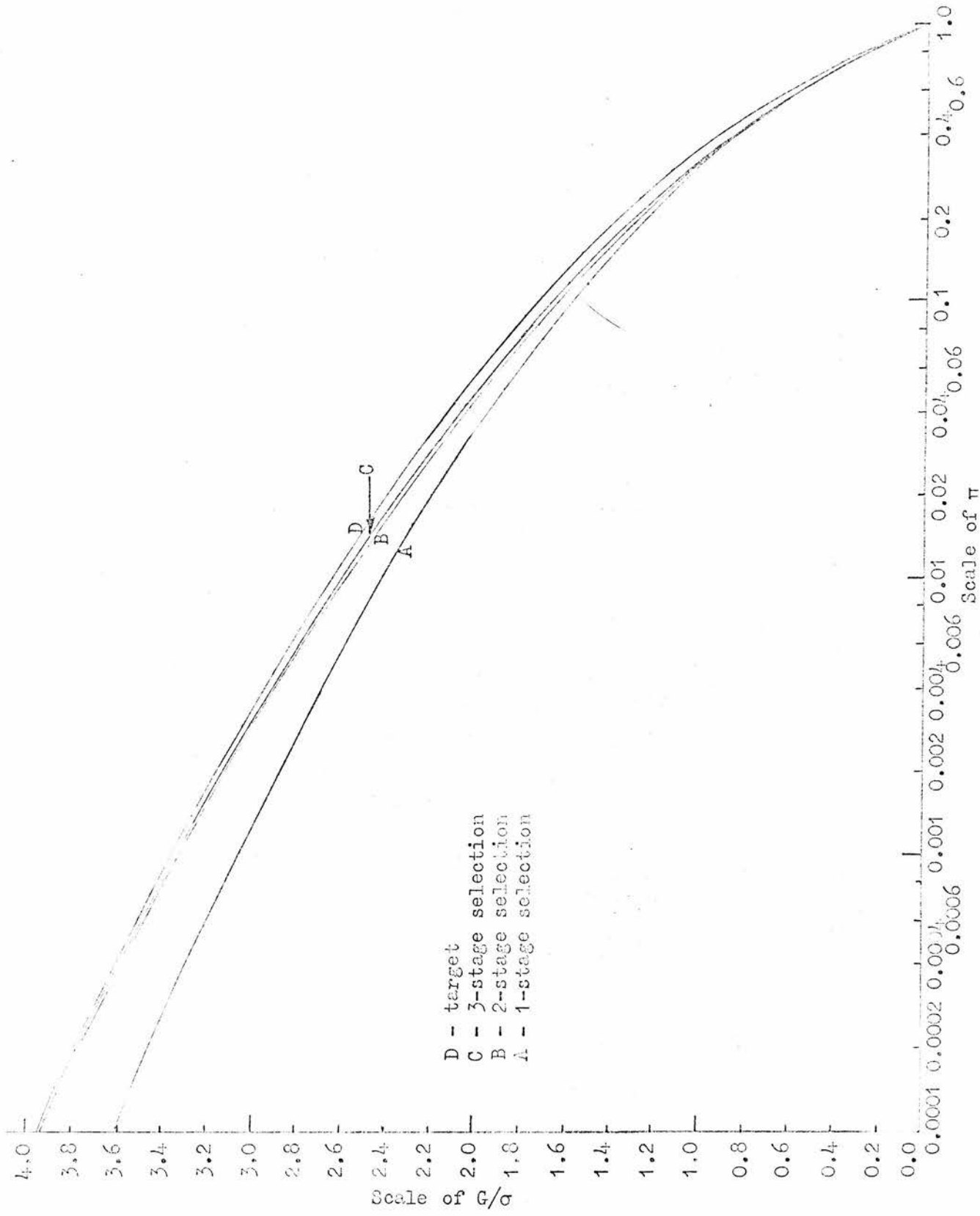
FIGURE F8. Dependence of G/σ on π for symmetrical selection from a normal distribution, V = 1.0.
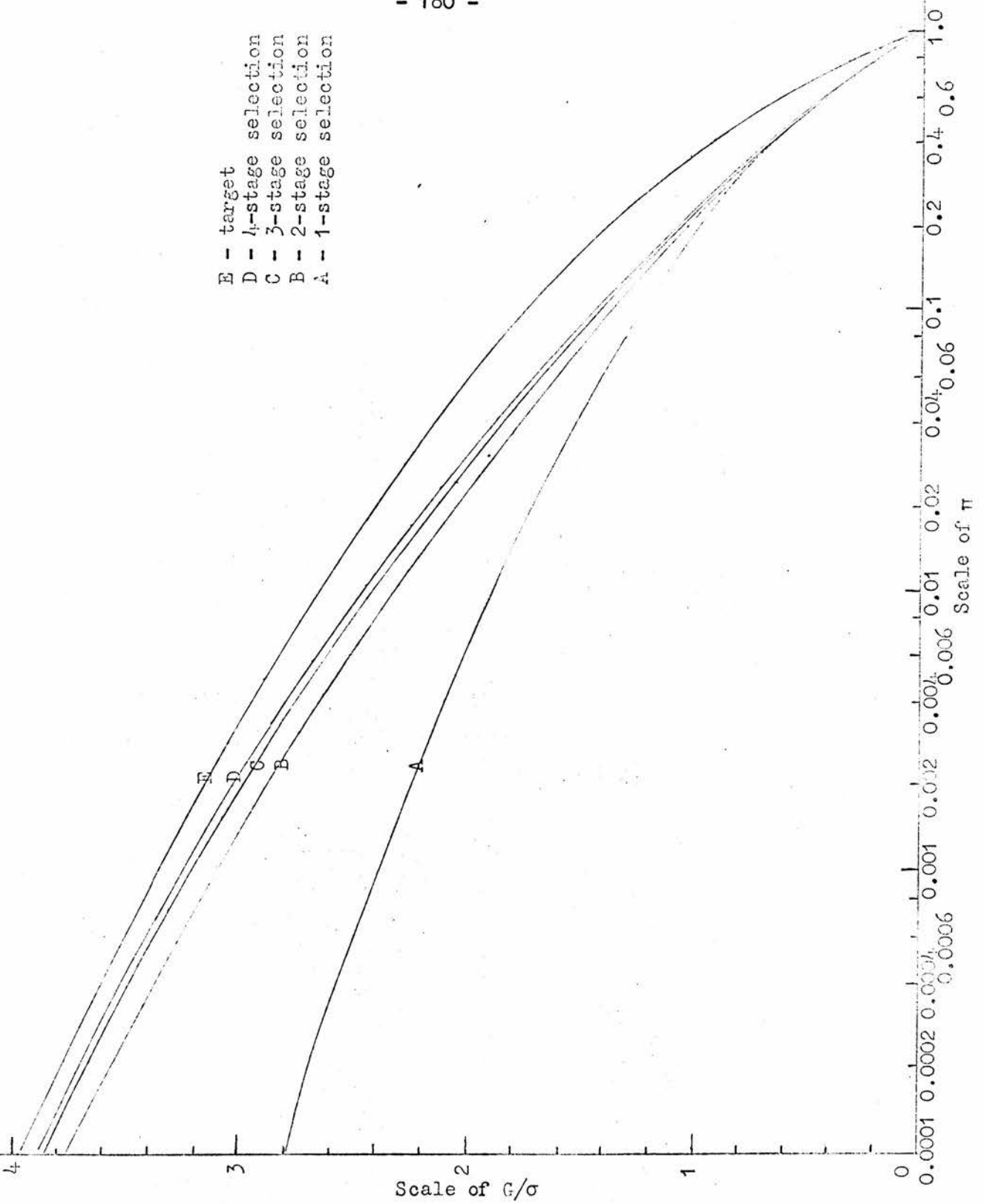
E — target
D — 4-stage selection
C — 3-stage selection
B — 2-stage selection
A — 1-stage selection

FIGURE F9. Dependence of G/σ on π for symmetrical selection from a normal distribution, V = 5.0.

E – target
D – 4-stage selection
C – 3-stage selection
B – 2-stage selection
A – 1-stage selection
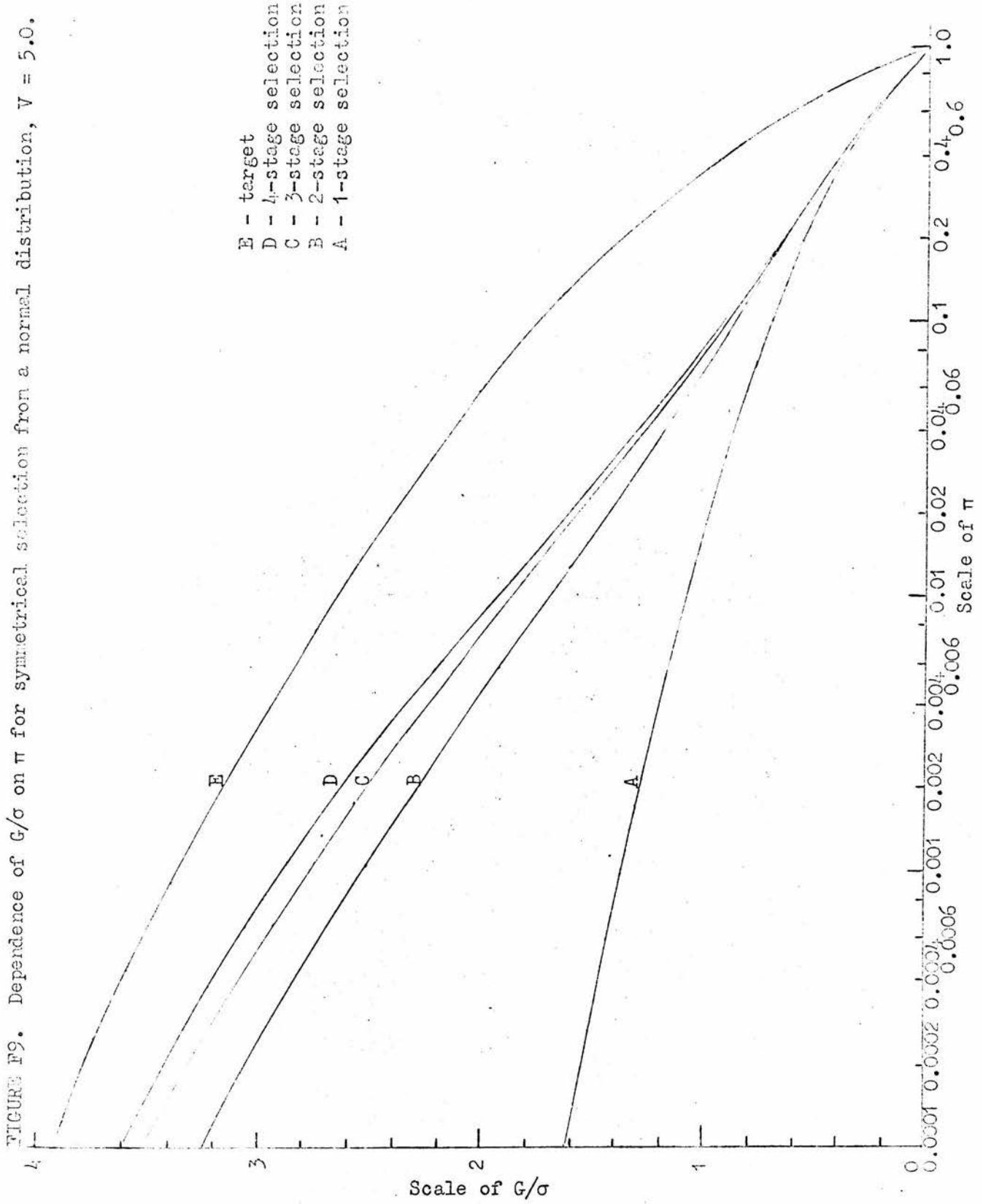
FIGURE F10. Dependence on π of percent increase in gain due to the use of
historical information for symmetrical selection from a normal distribution,
$V = 5$.
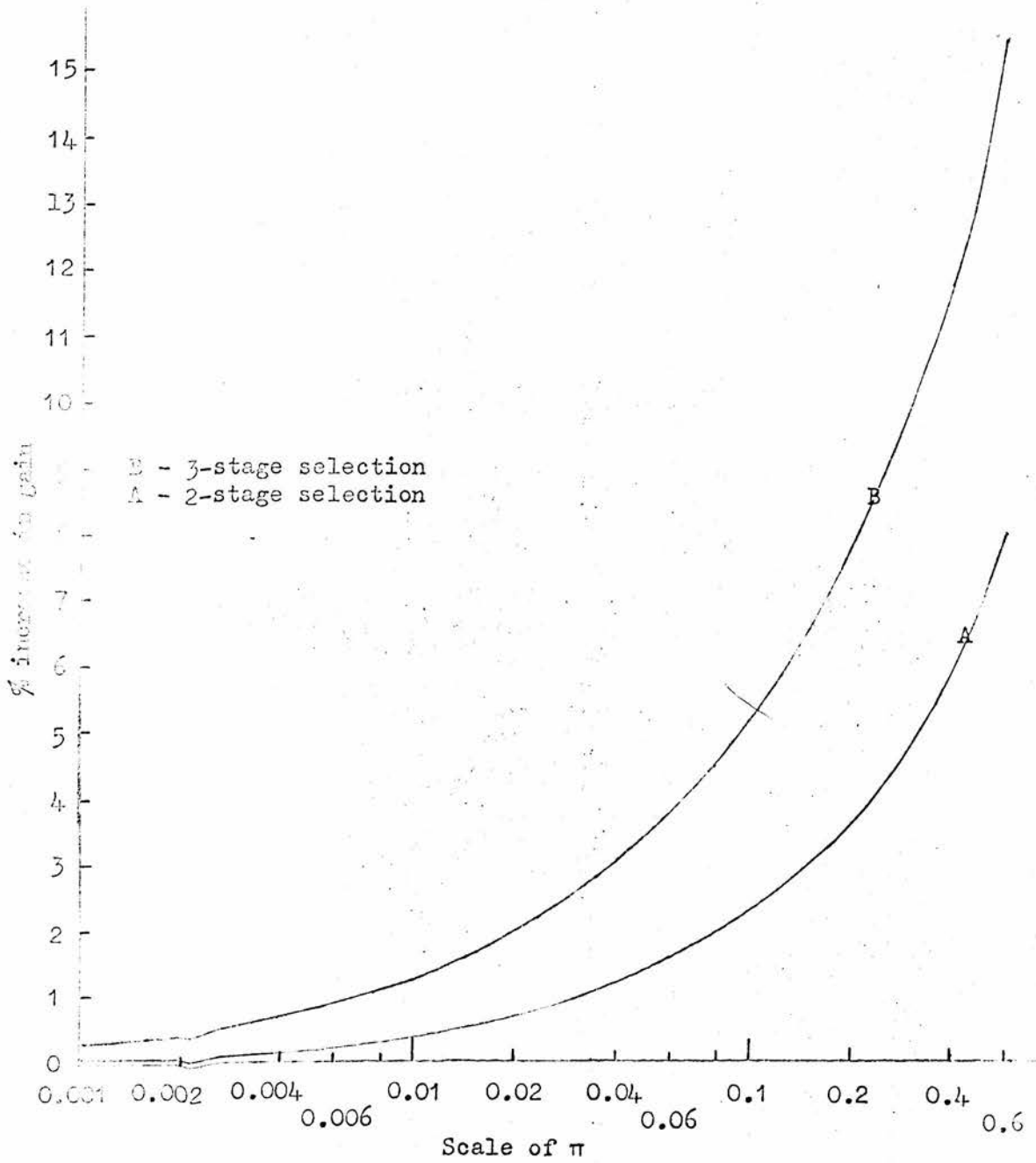
FIGURE F11. Dependence on V of percent increase in gain due to the use of
historical information for symmetrical selection from a normal distribution,
$\pi = 0.0625$.
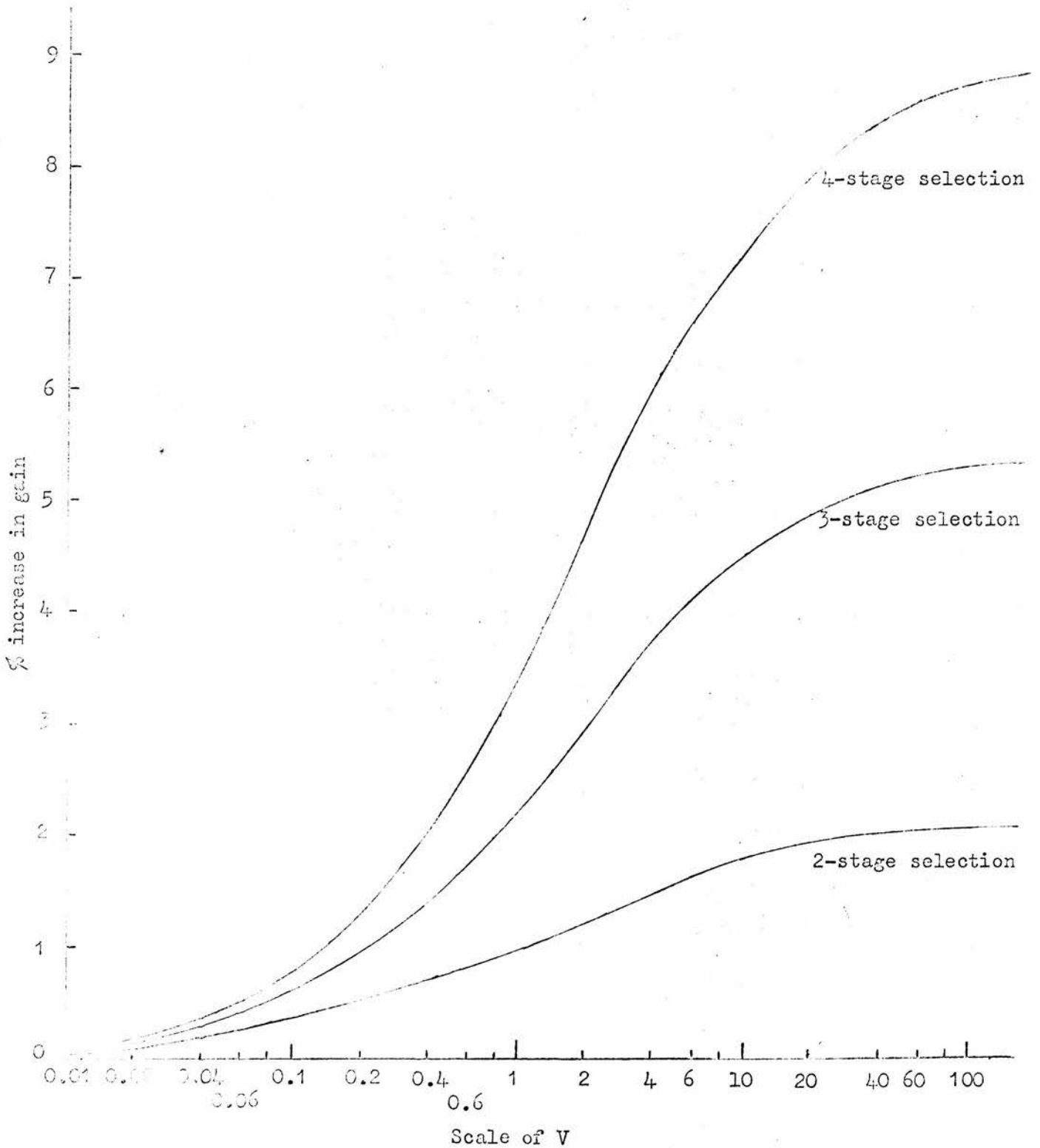
FIGURE F12. Variation of value of P_oV required to achieve maximum gain with πV for symmetrical selection from a normal distribution.
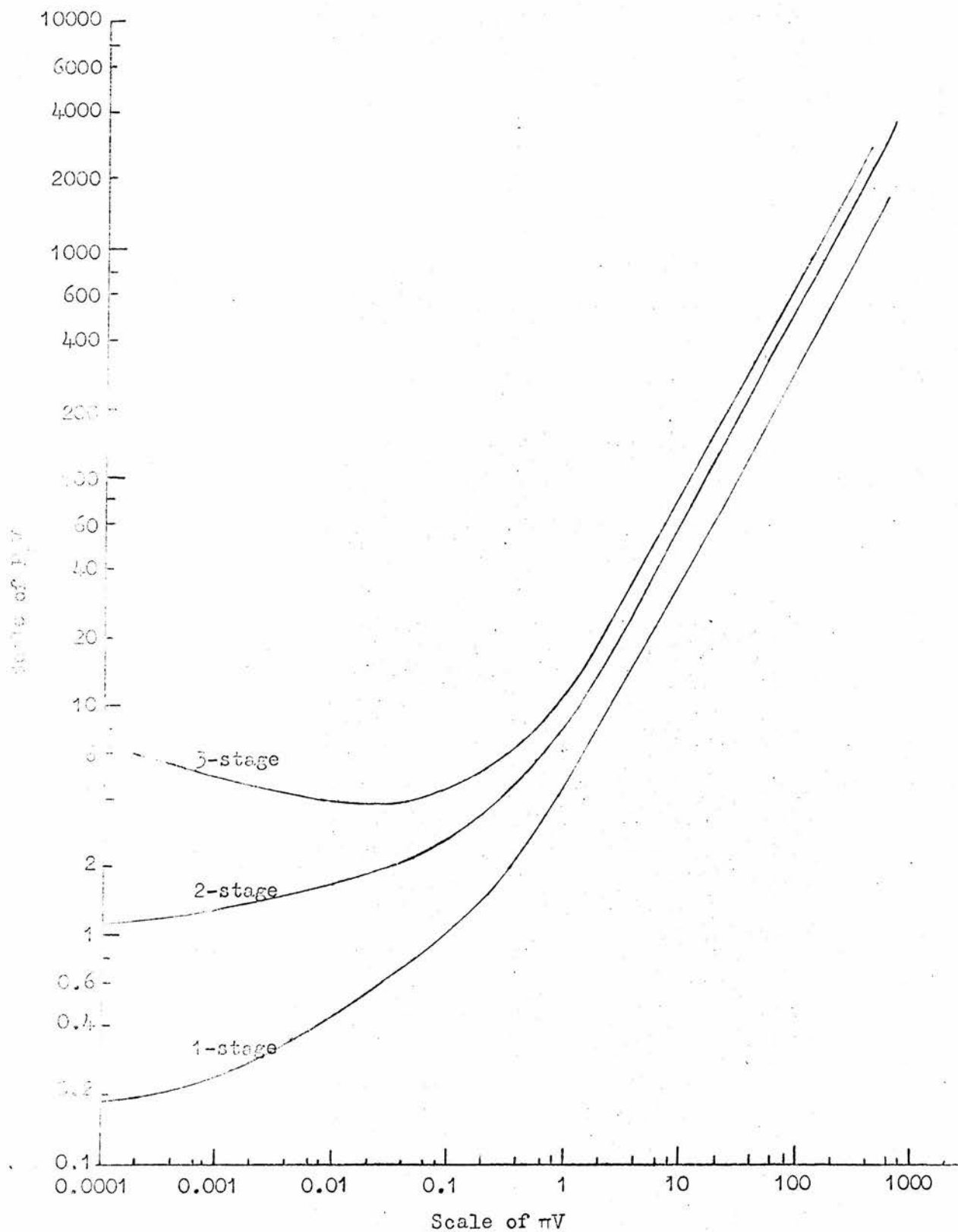
FIGURE F13. Maximum gain achieved by using $P_o \neq 1$ for symmetrical selection from a normal distribution.

C - 3-stage selection
B - 2-stage selection
A - 1-stage selection

Scale of C/c

Scale of πV
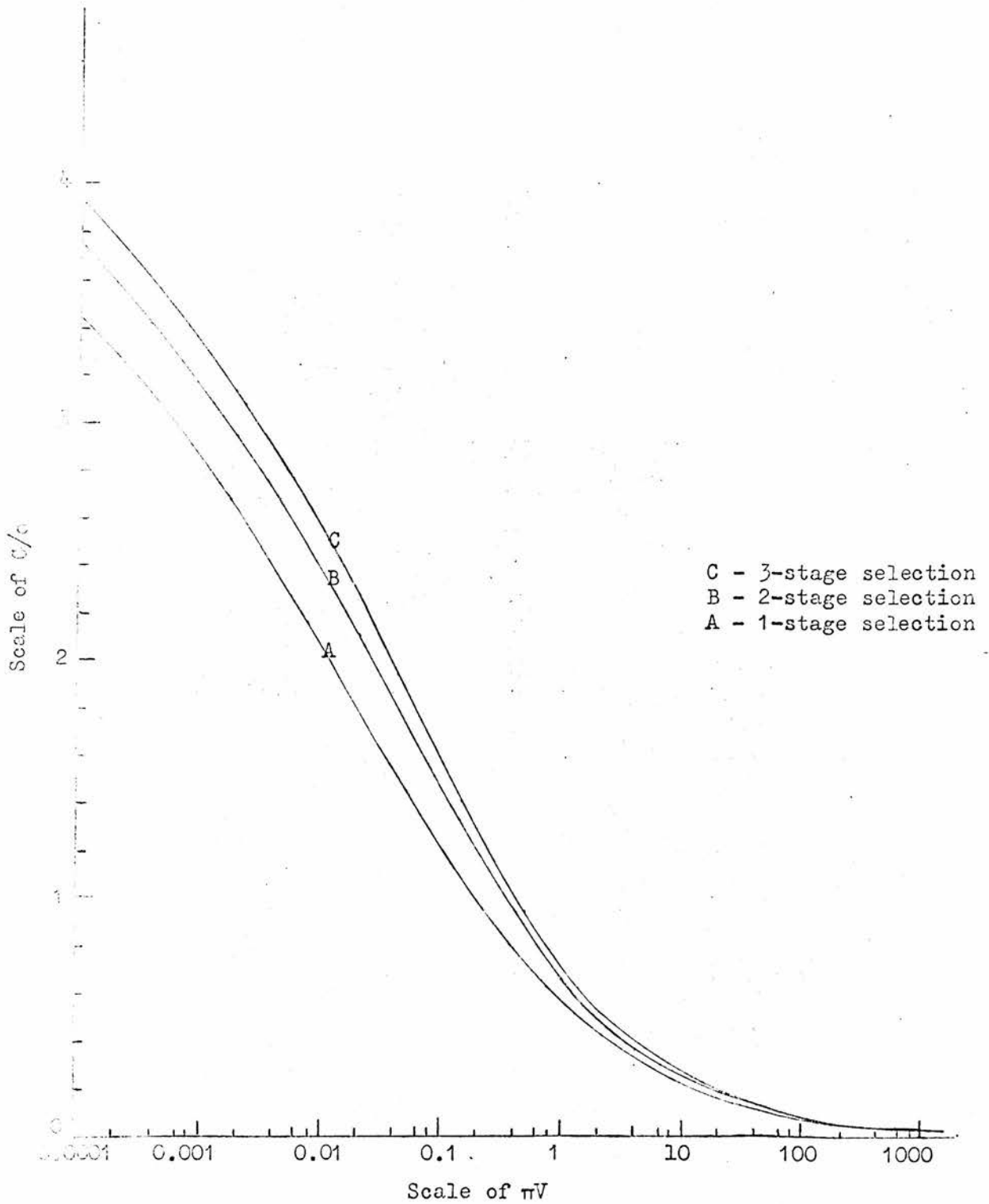
FIGURE F14. The use of various graphical techniques to compare 2-stage and 3-stage selection from a normal distribution with $\pi = 0.001$ and $V = 1$. The arrow marks $x = 3.090\sigma$ the boundary between the correct classes.
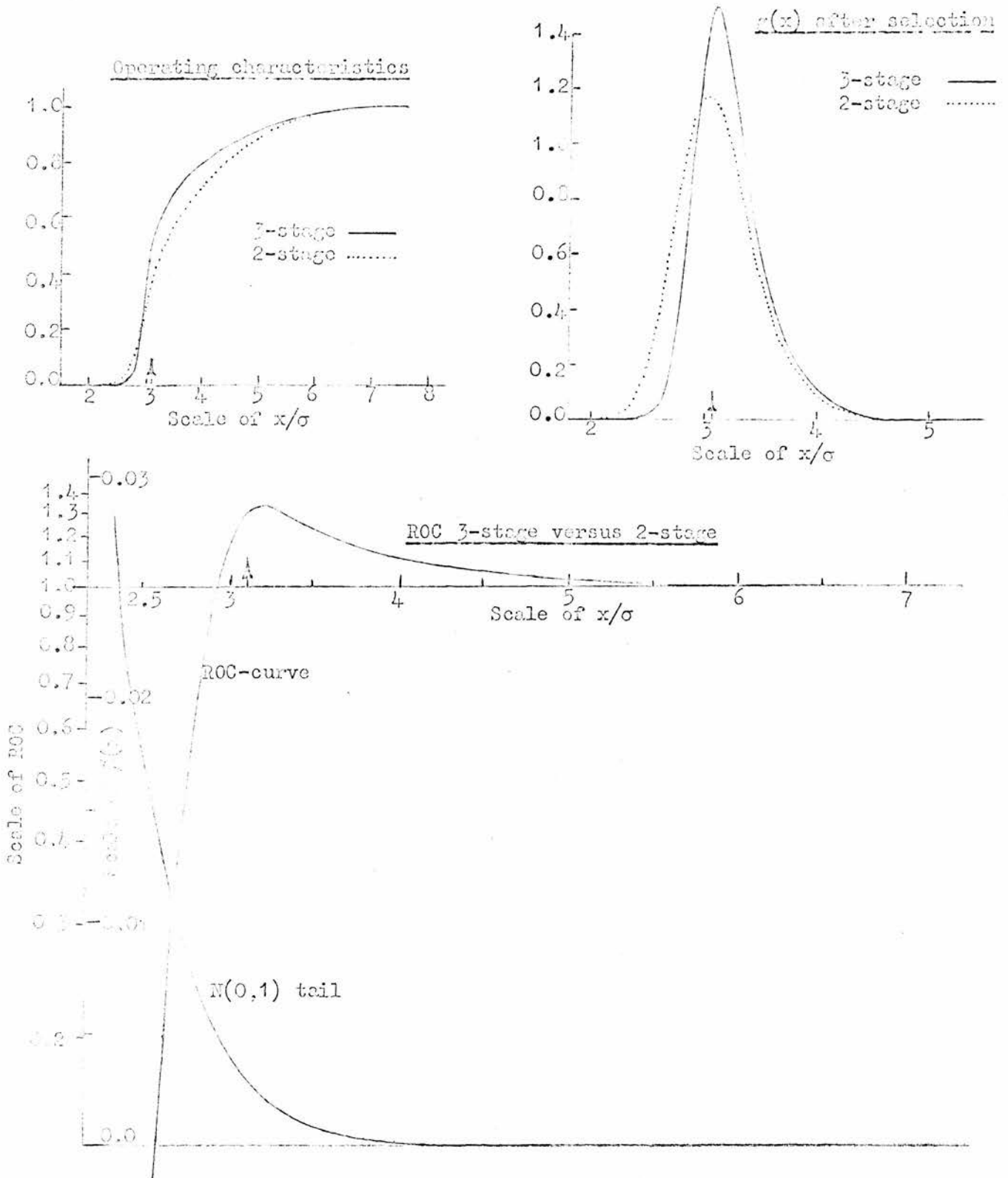
FIGURE F15. The use of an ROC-curve to compare the gain with history to that without history for 3-stage selection from a normal distribution (π = 0.0625 and V = 1). The arrow marks x = 1.534σ, the boundary between the correct classes.
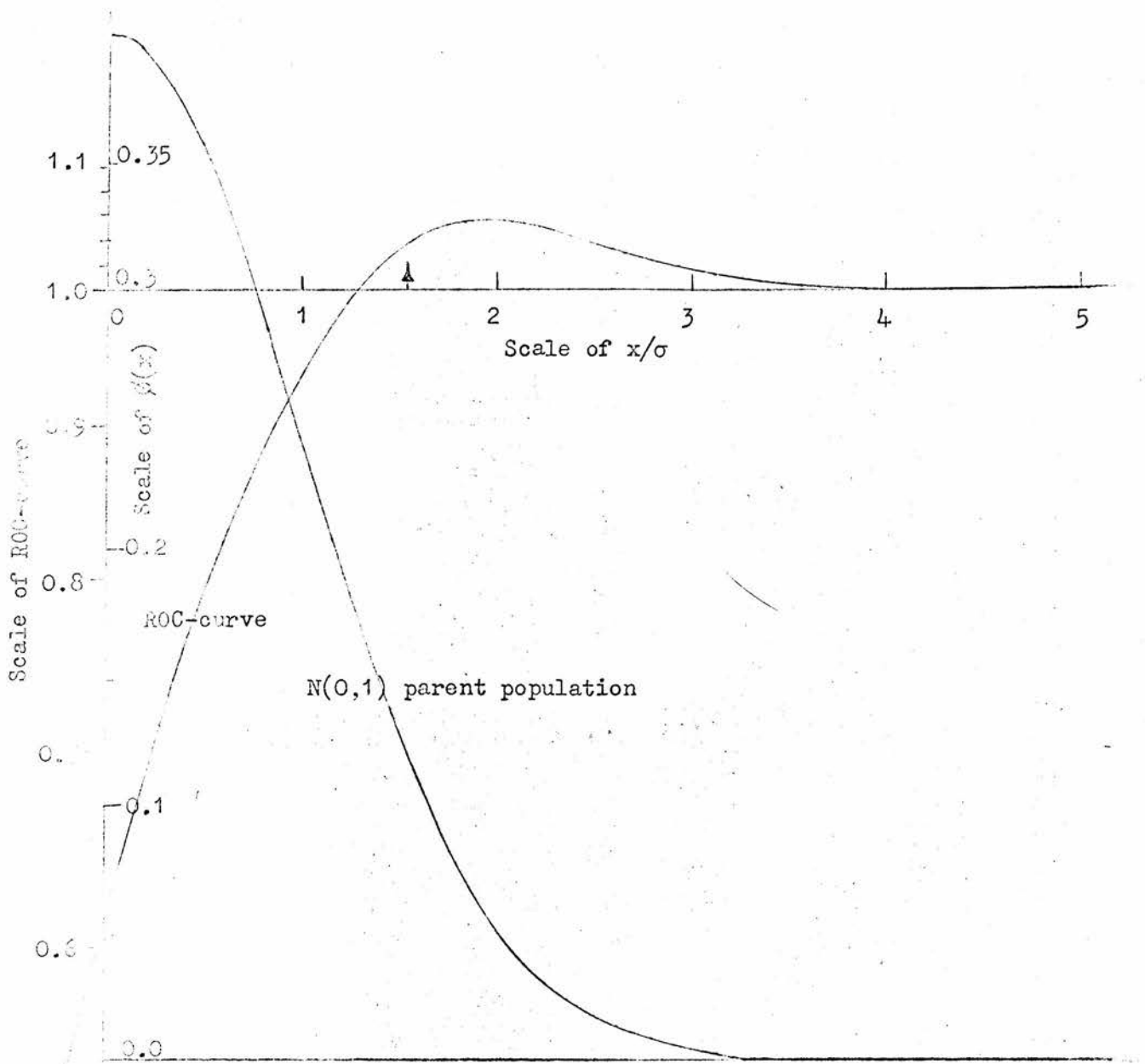
FIGURE 16. Dependence of G/σ (as a percent of target) on α_1 and P_1 for 2-stage selection from a Box and Tiao distribution with $a = +0.5$ ($\pi = 0.01$, $V = 1$).
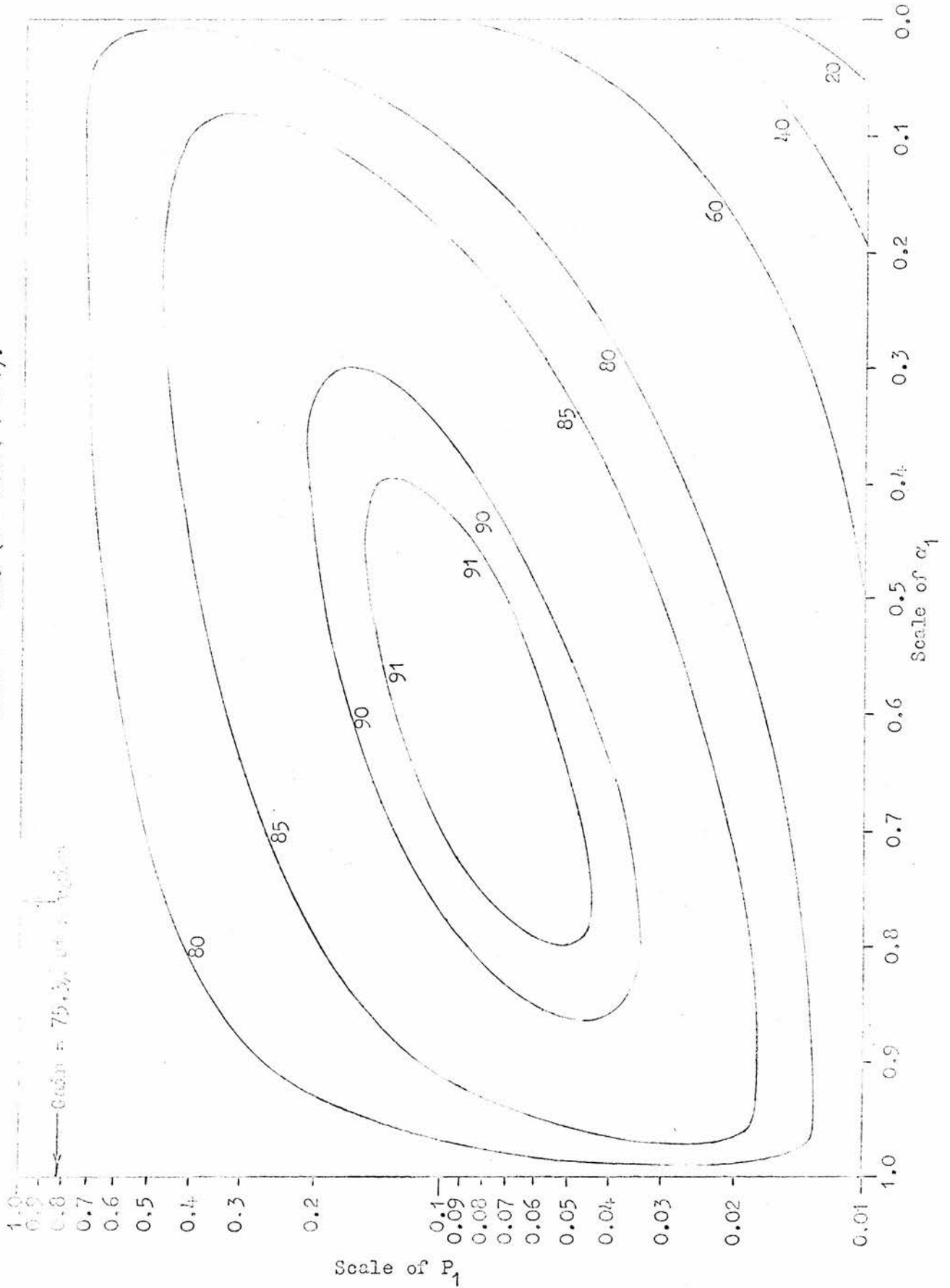
FIGURE F17. Dependence of (% of percent of target) on α_1 and P_1 for 2-stage selection from a beta distribution, with $P = 50$, $Q = 3$ ($\pi = 0.01$, $V = 1$).

Gain = 66.4% on samples
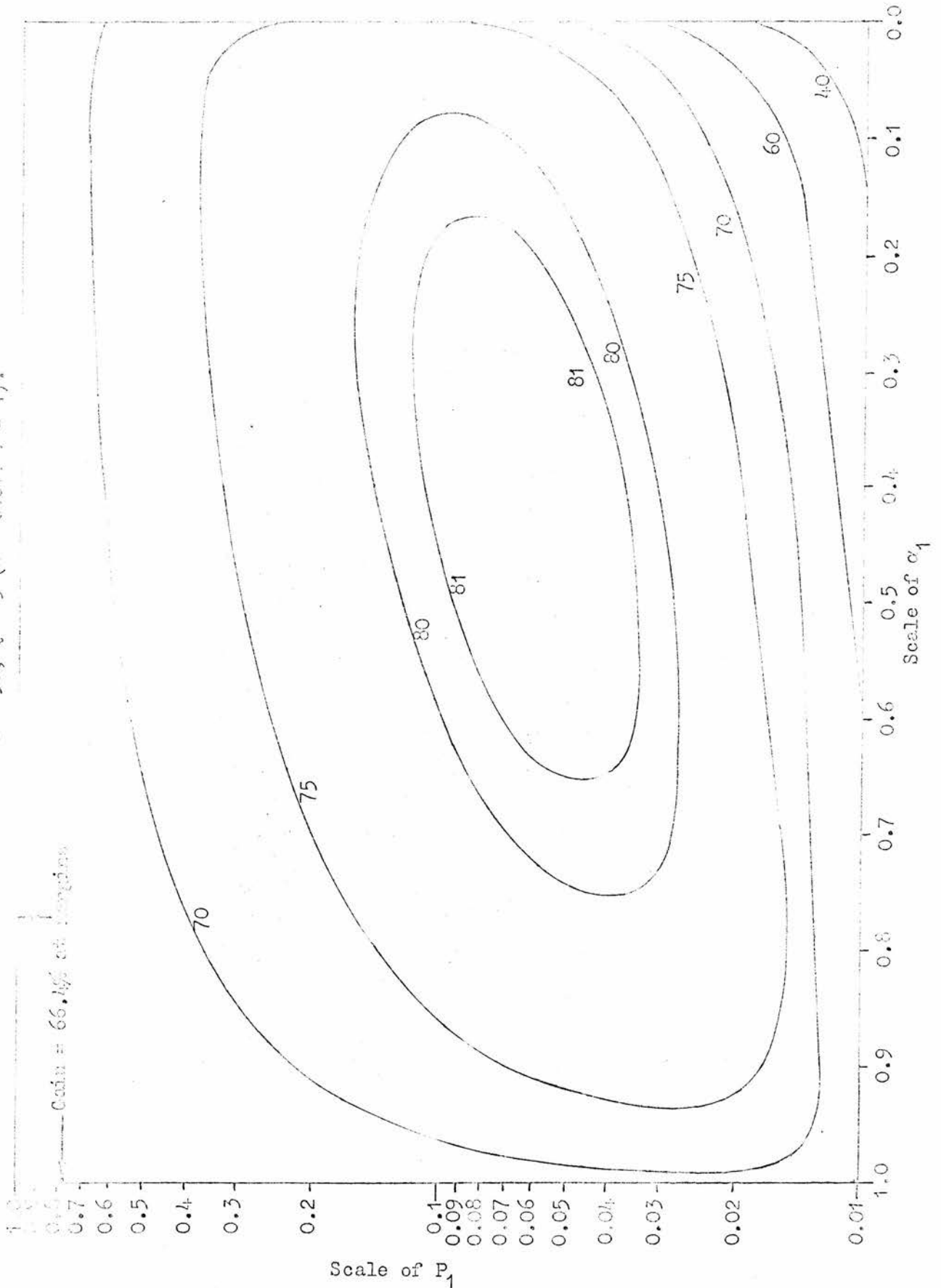
FIGURE F18. Dependence of G/σ (as a per cent of target) on α_1 and P_1 for 2-stage selection from a 2-point distribution with D = 1 and ... .99 (π = 0.01, V = 0.2).
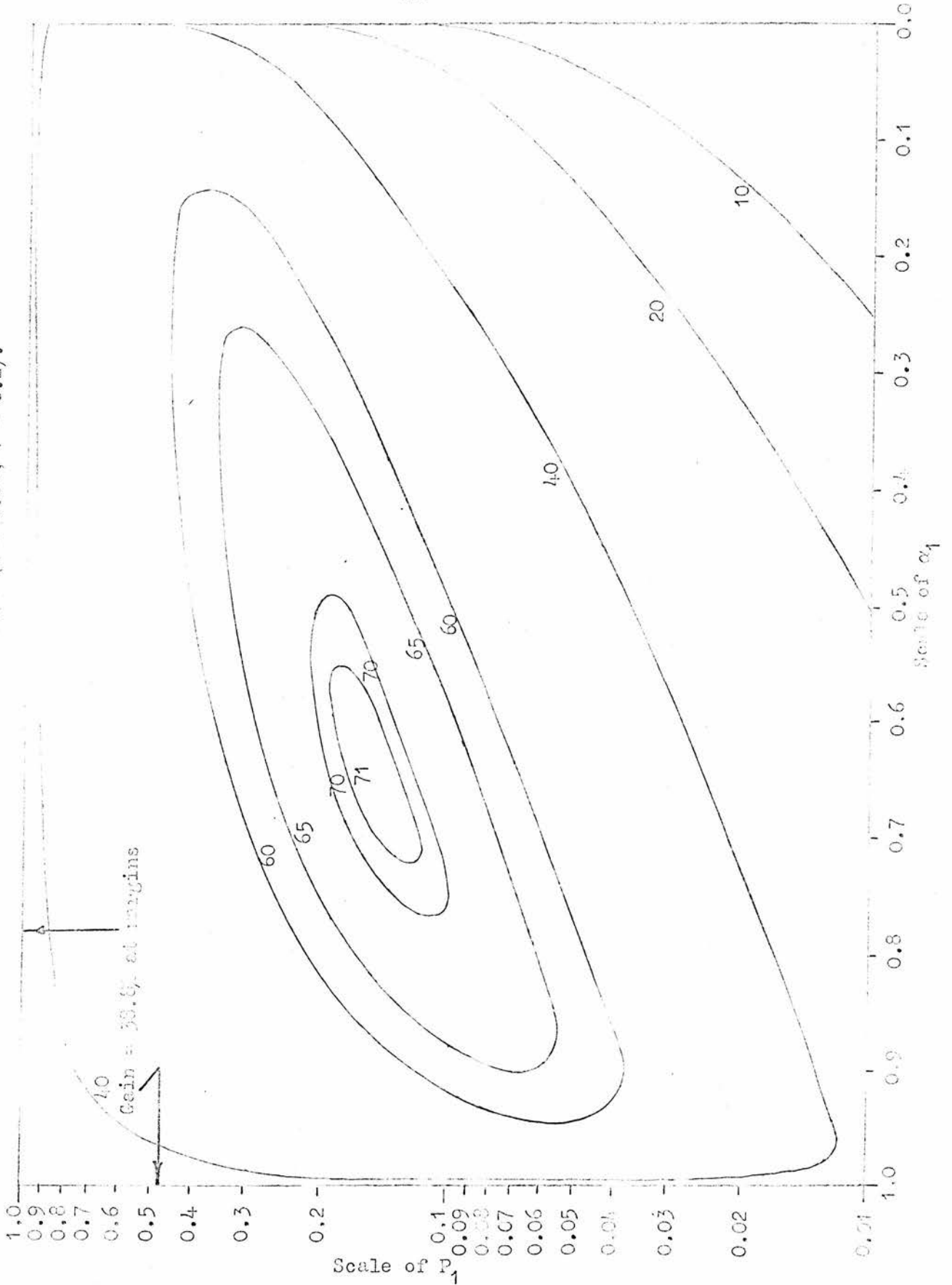
TABLE T1. Values of the variance of $\frac{x}{\sigma}$ in the selected population with 2-stage selection from a normal distribution with $V = 1$ and $\pi = 0.01$.

α_1: P_1	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
1.0	0.549	0.549	0.549	0.549	0.549	0.549	0.549	0.549	0.549	0.549	0.549
0.7	0.473	0.477	0.483	0.490	0.497	0.505	0.513	0.521	0.530	0.539	0.549
0.5	0.407	0.413	0.422	0.432	0.445	0.458	0.473	0.490	0.507	0.527	0.549
0.33	0.340	0.345	0.353	0.365	0.380	0.397	0.418	0.443	0.472	0.507	0.549
0.22	0.291	0.293	0.299	0.309	0.323	0.341	0.363	0.393	0.430	0.480	0.549
0.15	0.260	0.259	0.262	0.269	0.279	0.295	0.316	0.345	0.387	0.449	0.549
0.1	0.245	0.239	0.238	0.241	0.247	0.258	0.275	0.300	0.341	0.410	0.549
0.07	0.244	0.235	0.230	0.229	0.231	0.237	0.249	0.270	0.306	0.375	0.549
0.05	0.256	0.243	0.234	0.229	0.227	0.229	0.236	0.251	0.280	0.344	0.549
0.033	0.289	0.271	0.257	0.247	0.240	0.237	0.238	0.245	0.265	0.316	0.549
0.022	0.353	0.328	0.308	0.292	0.279	0.270	0.265	0.265	0.273	0.308	0.549
0.015	0.475	0.439	0.409	0.384	0.363	0.347	0.333	0.325	0.322	0.338	0.549
0.01	1.000	0.913	0.845	0.787	0.738	0.695	0.658	0.625	0.595	0.569	0.549

TABLE T2. Values of G/σ when using historical information in 3-stage selection from a normal distribution with V = 5.

Values of $(\alpha_1, \alpha_2, \alpha_3)$

P_1	P_2	P_3	$(\frac{1}{6},\frac{5}{12},\frac{5}{12})$	$(\frac{5}{12},\frac{1}{6},\frac{5}{12})$	$(\frac{1}{6},\frac{1}{2},\frac{1}{3})$	$(\frac{1}{3},\frac{1}{3},\frac{1}{3})$	$(\frac{1}{2},\frac{1}{6},\frac{1}{3})$	$(\frac{1}{4},\frac{1}{2},\frac{1}{4})$	$(\frac{1}{2},\frac{1}{4},\frac{1}{4})$	$(\frac{1}{3},\frac{1}{2},\frac{1}{6})$	$(\frac{1}{2},\frac{1}{3},\frac{1}{6})$	Averages
π = 0.0001												
.215	.046	.010	3.486	3.428	3.476	3.477	3.415	3.455	3.419	3.385	3.365	3.433
.046	.215	.010	3.334	3.454	3.314	3.406	3.452	3.339	3.419	3.330	3.364	3.379
.022	.215	.022	3.238	3.404	3.229	3.351	3.428	3.285	3.410	3.314	3.383	3.338
.215	.022	.022	3.482	3.377	3.502	3.480	3.386	3.508	3.438	3.482	3.443	3.455
.215	.010	.046	3.400	3.247	3.442	3.395	3.268	3.464	3.355	3.466	3.404	3.382
.046	.046	.046	3.427	3.489	3.425	3.513	3.512	3.472	3.545	3.493	3.542	3.491
.010	.215	.046	3.077	3.279	3.073	3.220	3.319	3.149	3.311	3.204	3.298	3.214
.100	.010	.100	3.410	3.266	3.448	3.433	3.291	3.489	3.403	3.513	3.470	3.413
.010	.100	.100	3.091	3.292	3.088	3.238	3.336	3.267	3.333	3.225	3.325	3.253
.046	.010	.215	3.349	3.268	3.376	3.410	3.299	3.433	3.409	3.474	3.471	3.388
.010	.046	.215	3.096	3.281	3.096	3.243	3.326	3.177	3.338	3.239	3.338	3.237
Averages			3.308	3.344	3.315	3.379	3.366	3.358	3.398	3.375	3.400	
π = 0.01												
.464	.215	.100	1.767	1.735	1.733	1.724	1.691	1.683	1.649	1.605	1.578	1.685
.215	.464	.100	1.766	1.772	1.742	1.760	1.736	1.728	1.707	1.692	1.664	1.729
.147	.464	.147	1.797	1.829	1.778	1.818	1.807	1.781	1.786	1.761	1.751	1.790
.464	.147	.147	1.819	1.776	1.796	1.783	1.744	1.753	1.714	1.678	1.650	1.746
.464	.100	.215	1.826	1.770	1.820	1.799	1.753	1.790	1.742	1.727	1.696	1.769
.215	.215	.215	1.874	1.847	1.862	1.874	1.831	1.851	1.828	1.810	1.791	1.841
.100	.464	.215	1.779	1.842	1.766	1.829	1.835	1.787	1.824	1.786	1.797	1.805
.316	.100	.316	1.837	1.760	1.846	1.826	1.758	1.839	1.778	1.801	1.763	1.801
.100	.316	.316	1.807	1.844	1.801	1.858	1.847	1.825	1.856	1.824	1.838	1.833
.215	.100	.464	1.800	1.707	1.824	1.806	1.718	1.839	1.771	1.828	1.788	1.787
.100	.215	.464	1.794	1.786	1.802	1.843	1.801	1.834	1.841	1.841	1.846	1.821
Averages			1.806	1.788	1.797	1.811	1.775	1.792	1.772	1.759	1.742	

TABLE T3. The percentage increase in gain due to the use of historical information in 2-stage ~~symmetrical~~ selection from a normal distribution with $\pi = 0.01$ and $V = 1$.

P_1:	1.0	0.7	0.5	0.3	0.17	0.1	0.07	0.05	0.03	0.02	0.01
α_1											
0.0	0	0	0	0	0	0	0	0	0	0	0
0.1	2.7	1.1	0.5	0.1	0	0	0	0	0	0	0
0.2	6.0	2.8	1.3	0.4	0.1	0	0	0	0	0	0
0.3	10.1	5.1	2.5	0.8	0.2	0.1	0	0	0	0	0
0.4	15.4	8.3	4.3	1.4	0.4	0.1	0.1	0	0	0	0
0.5	22.3	12.7	6.8	2.5	0.7	0.2	0.1	0	0	0	0
0.6	32.1	18.9	10.7	4.1	1.3	0.4	0.2	0.1	0	0	0
0.7	46.9	28.4	16.8	7.0	2.3	0.7	0.3	0.1	0	0	0
0.8	72.8	44.4	27.3	12.4	4.5	1.6	0.7	0.3	0.1	0	0
0.9	133.6	78.6	49.8	25.0	10.6	4.2	2.1	1.0	0.3	0.1	0
1.0	∞	267.7	158.5	88.7	50.8	29.8	20.0	13.1	6.0	1.8	0

TABLE T4. The effect of interaction (represented by σ^2_{vy}) on the gain in 3-stage symmetrical selection from a normal distribution with V = 1.

			Weighting system	
	σ^2_{vy}	inverse variance	equal weight	no history
$\pi = 0.0625$	0.00	1.656	1.579	1.620
	0.01	1.653	1.577	1.616
	0.03	1.647	1.573	1.610
	0.10	1.628	1.560	1.587
$\alpha_1 = \alpha_2 = \alpha_3 = 1/3$	0.30	1.578	1.524	1.528
	1.00	1.446	1.417	1.377
$P_1 = P_2 = P_3 = 0.39685$	3.00	1.214	1.204	1.130
	10.00	0.862	0.860	0.784
	30.00	0.557	0.556	0.501
	100.00	0.319	0.319	0.286
$\pi = 0.01$	0.00	2.440	2.228	2.425
	0.01	2.433	2.226	2.417
	0.03	2.420	2.220	2.402
	0.10	2.379	2.201	2.354
$\alpha_1 = \alpha_2 = \alpha_3 = 1/3$	0.30	2.285	2.149	2.242
	1.00	2.060	1.991	1.985
$P_1 = P_2 = P_3 = 0.21544$	3.00	1.700	1.679	1.604
	10.00	1.190	1.188	1.102
	30.00	0.764	0.764	0.701
	100.00	0.437	0.437	0.400
$\pi = 0.001$	0.00	3.226	2.874	3.223
	0.01	3.214	2.870	3.209
	0.03	3.193	2.863	3.184
	0.10	3.130	2.839	3.107
$\alpha_1 = \alpha_2 = \alpha_3 = 1/3$	0.30	2.995	2.772	2.942
	1.00	2.682	2.566	2.592
$P_1 = P_2 = P_3 = 0.1$	3.00	2.192	2.157	2.085
	10.00	1.523	1.518	1.428
	30.00	0.974	0.973	0.907
	100.00	0.557	0.557	0.517

TABLE T5. The variation of G/σ when interaction (σ^2_{vy}) is present in 3-stage selection from a normal distribution with $V = 1$ and $\pi = 0.01$ (when there is no interaction the maximum gain is 2.444σ and occurs at approximately $P_1 = 0.21$, $P_2 = 0.19$, $P_3 = 0.25$, $\alpha_1 = 0.45$, $\alpha_2 = 0.25$ and $\alpha_3 = 0.3$).

Values of $(\alpha_1, \alpha_2, \alpha_3)$

P_1	P_2	P_3	$(\tfrac{1}{2},\tfrac{1}{2},\tfrac{1}{2})$	$(\tfrac{1}{3},\tfrac{1}{3},\tfrac{1}{2})$	$(\tfrac{1}{3},\tfrac{1}{3},\tfrac{1}{3})$	$(\tfrac{1}{2},\tfrac{1}{3},\tfrac{2}{3})$	$(\tfrac{2}{5},\tfrac{2}{5},\tfrac{1}{5})$	$(\tfrac{1}{3},\tfrac{1}{3},\tfrac{2}{6})$	$(\tfrac{1}{2},\tfrac{2}{5},\tfrac{1}{10})$	$(\tfrac{3}{5},\tfrac{3}{10},\tfrac{1}{10})$	$(\tfrac{11}{20},\tfrac{2}{5},\tfrac{1}{20})$
$\sigma^2_{vy} = 0.1$											
0.464	0.464	0.046	1.932	2.029	2.111	2.083	2.133	2.138	2.091	2.088	1.966
0.464	0.215	0.100	1.945	2.044	2.137	2.108	2.179	2.193	2.179	2.174	2.116
0.215	0.464	0.100	1.983	2.142	2.188	2.218	2.223	2.249	2.223	2.238	2.157
0.215	0.215	0.215	1.990	2.150	2.201	2.232	2.246	2.279	2.270	2.286	2.244
0.100	0.464	0.215	1.973	2.166	2.191	2.261	2.235	2.278	2.264	2.294	2.239
0.100	0.215	0.464	1.971	2.155	2.185	2.255	2.235	2.282	2.278	2.307	2.276
$\sigma^2_{vy} = 1.0$											
0.464	0.464	0.046	1.795	1.872	1.936	1.914	1.953	1.956	1.920	1.918	1.823
0.464	0.215	0.100	1.805	1.883	1.955	1.933	1.987	1.997	1.987	1.984	1.940
0.215	0.464	0.100	1.829	1.953	1.986	2.010	2.013	2.033	2.014	2.026	1.965
0.215	0.215	0.215	1.828	1.952	1.991	2.015	2.025	2.049	2.044	2.056	2.027
0.100	0.464	0.215	1.799	1.949	1.966	2.021	2.000	2.033	2.023	2.046	2.001
0.100	0.215	0.464	1.764	1.918	1.940	1.995	1.979	2.016	2.014	2.038	2.018
$\sigma^2_{vy} = 10.0$											
0.464	0.464	0.046	1.165	1.186	1.201	1.197	1.206	1.207	1.198	1.198	1.174
0.464	0.215	0.100	1.164	1.185	1.202	1.198	1.210	1.213	1.211	1.210	1.201
0.215	0.464	0.100	1.156	1.190	1.197	1.204	1.203	1.209	1.205	1.208	1.195
0.215	0.215	0.215	1.145	1.179	1.188	1.195	1.196	1.202	1.202	1.205	1.199
0.100	0.464	0.215	1.113	1.156	1.159	1.173	1.167	1.175	1.174	1.179	1.172
0.100	0.215	0.464	1.081	1.125	1.129	1.143	1.139	1.147	1.147	1.153	1.146

TABLE T6. The variation in G/σ for selection from 5 members of the Box and Tiao family with V = 1 and π = 0.01 (all percent increases and decreases relative to 2-stage symmetrical gains).

| σ | Target | 1-stage | | | G/σ at symmetry (σ₁=0.5, P₁=0.1) | 2-stage | | | | | | 3-stage | |
| | | G/σ at P_o=1 | Approx. maximum for P_o<1 | | | Approx. optimum | location | | P_o<1 at symmetry | | % loss in G/σ without history | | |
			P_o	G/σ	P₁=0.1	% inc.	σ₁	P₁	% inc.	P_o		G/σ	% inc.
-0.9999	1.715	1.334	0.2	1.454	1.518	1.3	0.5	0.04	1.2	0.3	0.2	1.553	2.1
-0.5	2.189	1.520	0.3	1.653	1.851	0.5	0.5	0.08	0.2	0.8	0.2	1.926	4.1
0.0	2.665	1.885	0.5	1.977	2.350	0.1	0.6	0.08	0.0	1.0	0.2	2.439	3.8
0.5	3.093	2.329	0.7	2.355	2.833	0.2	0.6	0.08	0.0	1.0	0.2	2.913	2.8
1.0	3.471	2.775	1.0	2.775	3.263	0.3	0.6	0.08	0.0	1.0	0.1	3.328	2.0

TABLE T7. The variation in $(G-\mu)/\sigma$ for selection from various numbers of the beta distribution with $V = 1$ and $\pi = 0.01$ (all percent increases and decreases relative to 2-stage symmetrical gains).

| | | | 1-stage | | | | 2-stage | | | | | | 3-stage | |
| | | | $\dfrac{(G-\mu)}{\sigma}$ at $P_o=1$ | Approx. maximum for $P_o<1$ | | $\dfrac{(G-\mu)}{\sigma}$ at symmetry ($\alpha_1=0.5$, $P_1=0.1$) | Approx. optimum | | | $P_o<1$ at symmetry | | % loss in $\dfrac{(G-\mu)}{\sigma}$ without history | $\dfrac{(G-\mu)}{\sigma}$ | % inc. |
p	q	Target		P_o	$\dfrac{(G-\mu)}{\sigma}$		% inc.	location α_1	P_1	% inc. P_o	$P_o<1$			
10	30	2.964	2.203	0.6	2.261	2.686	0.2	0.6	0.075	0.0	1.00	0.2	2.770	3.1
10	10	2.523	1.767	0.4	1.885	2.199	0.1	0.5	0.085	0.0	1.00	0.2	2.286	4.0
10	6	2.238	1.526	0.3	1.668	1.900	0.4	0.5	0.075	0.1	0.85	0.2	1.981	4.3
10	3	1.795	1.212	0.2	1.363	1.478	1.2	0.4	0.060	1.1	0.50	0.2	1.538	4.1
30	10	2.227	1.497	0.3	1.640	1.880	0.3	0.5	0.075	0.1	0.95	0.2	1.964	4.5
30	3	1.636	1.089	0.2	1.234	1.327	1.3	0.4	0.055	1.5	0.40	0.2	1.381	4.1
50	30	2.511	1.740	0.4	1.855	2.181	0.1	0.5	0.085	0.0	1.00	0.2	2.270	4.1
50	10	2.152	1.433	0.3	1.579	1.802	0.4	0.5	0.075	0.1	0.90	0.2	1.883	4.5
50	3	1.601	1.063	0.2	1.207	1.294	1.5	0.4	0.060	1.6	0.40	0.2	1.347	4.1

TABLE T8. The variation in $(G-\mu)/\sigma$ for selection from the sum of two normal distributions (lower variance always equals 1 and lower mean $\mu_1 = 0$; all percent increases and decreases relative to 2-stage symmetrical gains).

| D | W | V_u^2 | V | π | Target | 1-stage $\frac{G-\mu}{\sigma}$ at $P_0=1$ | 1-stage Approx. maximum for $P_0<1$ | | 2-stage* $\frac{G-\mu}{\sigma}$ at symmetry $(\alpha_1=0.5, P_1=\sqrt{\pi})$ | 2-stage* Approx. optimum | | | | 3-stage $\frac{G-\mu}{\sigma}$ | 3-stage |
							P_0	$\frac{G-\mu}{\sigma}$		% inc.	location α_1	P_1	% loss in gain without history		% inc.
Section A															
0	1	1	1	0.01	2.665	1.885	0.5	1.977	2.350	0.1	0.6	0.08	0.2	2.439	3.8
Section B															
3	0.990	0.1	1	0.01	2.979	2.220	0.6	2.291	2.739	0.2	0.6	0.075	0.2	2.820	3.0
4					3.705	3.010	0.95	3.011	3.565	0.3	0.7	0.075	0.2	3.619	1.5
5					4.449	3.998	1.0	3.998	4.401	0.3	0.7	0.075	0.5	4.413	0.3
Section C															
4	0.999	0.1	1	0.01	2.805	2.041	0.6	2.077	2.505	0.1	0.6	0.075	0.2	2.591	3.4
	0.99				3.705	3.010	0.95	3.011	3.565	0.3	0.7	0.075	0.2	3.619	1.5
	0.9				2.711	2.451	0.5	2.464	2.559	0.8	0.4	0.055	0.2	2.589	1.2
Section D															
4	0.99	0.01	1	0.01	3.697	2.986	0.9	2.992	3.566	0.2	0.7	0.075	0.1	3.622	1.6
		0.1			3.705	3.010	0.95	3.011	3.565	0.3	0.7	0.075	0.2	3.619	1.5
		1.0			3.792	3.184	1.0	3.184	3.635	0.4	0.4	0.075	0.4	3.685	1.4
Section E															
4	0.99	0.1	0.2	0.01	3.705	3.624	1.0	3.624	3.697	0.1	0.6	0.032	0.0	3.701	0.1
			1.0		3.705	3.010	0.95	3.011	3.565	0.3	0.7	0.075	0.2	3.619	1.5
			5.0		3.705	1.557	0.3	1.884	2.547	0.2	0.6	0.100	0.4	2.807	10.2
Section F															
4	0.99	0.1	1	0.001	4.212	3.766	0.4	3.842	4.082	0.5	0.6	0.01	0.0	4.137	1.4
				0.01	3.705	3.010	0.95	3.011	3.565	0.3	0.7	0.075	0.2	3.619	1.5
				0.0625	2.156	1.669	1.0	1.669	1.857	0.1	0.6	0.21	0.7	1.900	2.3

* In 2-stage symmetrical selection $P_0<1$ always resulted in a loss.

TABLE T9. The variation in gain for selection from a distribution in which x = 0 with probability W and x = 1 with probability 1-W (all percent increases and decreases relative to 2-stage symmetrical gains).

- 199 -

W	V	π	Target	1-stage Gain at P₀=1	Approx. maximum for P₀<1 P₀	Gain	2-stage Gain at symmetry (α₁=0.5, P₁=√π)	Approx. optimum Gain	% inc.	α₁	P₁	% loss in gain without history	3-stage Gain	% inc.
Section A														
0.999	0.2	0.01	0.1	0.0452	0.8	0.0471	0.0613	0.0771	25.8	0.6	0.21	0	0.0676	10.3
0.990			1.0	0.388	0.7	0.420	0.602	0.717	19.0	0.6	0.17	0	0.668	10.9
0.900			1.0	0.979	0.2	1.000	1.000	1.000	-			0	1.000	-
Section B														
0.990	0.1	0.01	1.0	0.669	1.0	0.669	0.817	0.927	13.2	0.6	0.21	0	0.843	3.1
	0.2			0.388	0.7	0.420	0.602	0.717	19.0	0.6	0.17	0	0.668	10.9
	0.4			0.196	0.4	0.257	0.413	0.457	11.0	0.6	0.13	0.2	0.475	15.1
Section C														
0.990	0.2	0.0625	0.16	0.116	1.0	0.116	0.127	0.137	7.6	0.6	0.35	0.3	0.129	1.8
	0.2	0.01	1.0	0.388	0.7	0.420	0.602	0.717	19.0	0.6	0.17	0	0.668	10.9
	0.2	0.001	1.0	0.847	0.2	1.000	1.000	1.000	-			0	1.000	-
	1.0	0.001	1.0	0.154	0.1	0.669	0.975	0.975	0	0.5	0.032	0.1	1.000	2.5

* For 2-stage symmetrical selection P₀<1 resulted in a loss in all cases studied except π = 0.001 with V = 1 where P₀ = 0.2 resulted in a gain of 2.5%.

TABLE T10. G/σ for 2-stage selection from an infinite population

with P = 0.0625 and V = 1.0 (infinite target = 1.968).

P_1: σ_1	$\frac{1}{16}$	$\frac{1}{8}$	$\frac{3}{16}$	$\frac{1}{4}$	$\frac{5}{16}$	$\frac{3}{8}$	$\frac{1}{2}$	$\frac{5}{8}$	$\frac{3}{4}$	$\frac{7}{8}$	1
0.0	0.000	0.758	1.006	1.141	1.225	1.282	1.347	1.379	1.392	1.395	1.391
0.1	0.598	1.213	1.383	1.461	1.499	1.516	1.517	1.496	1.465	1.429	1.391
0.2	0.808	1.355	1.490	1.543	1.561	1.563	1.541	1.506	1.467	1.428	1.391
0.3	0.949	1.442	1.550	1.583	1.588	1.579	1.544	1.502	1.461	1.424	1.391
0.4	1.056	1.501	1.584	1.602	1.595	1.579	1.536	1.492	1.453	1.420	1.391
0.5	1.140	1.540	1.601	1.605	1.590	1.568	1.521	1.479	1.444	1.415	1.391
0.6	1.208	1.563	1.604	1.596	1.574	1.549	1.502	1.464	1.434	1.411	1.391
0.7	1.266	1.572	1.592	1.573	1.547	1.522	1.479	1.448	1.424	1.406	1.391
0.8	1.315	1.563	1.562	1.536	1.510	1.487	1.453	1.430	1.414	1.401	1.391
0.9	1.357	1.522	1.504	1.479	1.459	1.444	1.424	1.412	1.403	1.396	1.391
1.0	1.391	1.391	1.391	1.391	1.391	1.391	1.391	1.391	1.391	1.391	1.391

TABLE T11. G_R/σ for the finite model with $V = 1.0$, $k = 2$, $N_1 = 16$ and $N_3 = 1$ (finite target = 1.766).

N_2:	1	2	3	4	5	6	8	10	12	14	16
α_1											
with history											
0.0	0.000	0.532	0.777	0.921	1.015	1.081	1.162	1.207	1.232	1.244	1.249
0.1	0.532	0.963	1.156	1.232	1.309	1.312	1.348	1.312	1.296	1.304	1.249
0.2	0.721	1.106	1.255	1.284	1.343	1.371	1.390	1.324	1.289	1.304	1.249
0.3	0.848	1.173	1.302	1.378	1.384*	1.388*	1.373*	1.335	1.306	1.263	1.249
0.4	0.944	1.278	1.361	1.369*	1.396*	1.392*	1.377	1.334	1.283	1.270	1.249
0.5	1.020	1.277	1.383*	1.401*	1.399*	1.390*	1.380	1.340	1.308	1.217	1.249
0.6	1.081	1.323	1.384*	1.397*	1.395*	1.377*	1.332	1.293	1.273	1.260	1.249
0.7	1.133	1.357	1.400*	1.402*	1.379*	1.323	1.321	1.295	1.295	1.269	1.249
0.8	1.177	1.340	1.377*	1.358*	1.341	1.298	1.296	1.290	1.300	1.243	1.249
0.9	1.215	1.361	1.370	1.343	1.269	1.242	1.295	1.237	1.256	1.251	1.249
1.0	1.249	1.249	1.249	1.249	1.249	1.249	1.249	1.249	1.249	1.249	1.249
without history*											
0.0	0.000	0.532	0.777	0.921	1.015	1.081	1.162	1.207	1.232	1.244	1.249
0.1	0.532	0.994	1.159	1.234	1.267	1.334	1.341	1.366	1.291	1.232	1.215
0.2	0.721	1.063	1.238	1.305	1.345	1.379	1.365	1.308	1.286	1.201	1.177
0.3	0.848	1.161	1.326**	1.358**	1.393***	1.399	1.353	1.278	1.266	1.225	1.133
0.4	0.944	1.236	1.365**	1.377**	1.387***	1.404	1.341	1.287	1.218	1.155	1.081
0.5	1.020	1.317	1.390**	1.370**	1.380**	1.381	1.250	1.254	1.165	1.131	1.020
0.6	1.081	1.325	1.417**	1.402**	1.359**	1.267	1.228	1.107	1.109	0.946	0.944
0.7	1.133	1.313	1.385**	1.338**	1.306**	1.286	1.216	1.098	1.091	0.938	0.843
0.8	1.177	1.354	1.898	1.289	1.318	1.220	1.162	0.982	0.925	0.790	0.721
0.9	1.215	1.317	1.257	1.234	1.089	1.042	0.944	0.808	0.704	0.677	0.532
1.0	1.249	1.079	0.932	0.849	0.760	0.680	0.538	0.408	0.283	0.154	0.000

All entries have a standard error of approximately ±0.025 to ±0.030, except that the margins are exact,

* indicates a standard error of approximately ±0.004 and

** indicates a standard error of approximately ±0.014.

*** This half of the table is taken from Table 1 in Finney's 1966 paper.

TABLE T12. Estimates of $Var(G)/\sigma^2$ for 2-stage selection from a finite population with $N_1 = 16$, $N_3 = 1$, and $V = 1$.

N_2:	1	2	3	4	5	6	8	10	12	14	16
α_1											
0.0	1.000	0.717	0.629	0.593	0.579	0.575	0.582	0.597	0.614	0.631	0.648
0.1	0.936	0.590	0.550	0.550	0.602	0.541	0.510	0.493	0.529	0.616	0.648
0.2	0.883	0.588	0.579	0.549	0.603	0.596	0.586	0.566	0.583	0.375	0.648
0.3	0.837	0.520	0.545	0.548	0.566*	0.566*	0.556*	0.533	0.596	0.669	0.648
0.4	0.799	0.655	0.526	0.535*	0.545*	0.548*	0.626	0.576	0.648	0.581	0.648
0.5	0.765	0.575	0.547*	0.555*	0.551*	0.576*	0.679	0.591	0.615	0.646	0.648
0.6	0.736	0.608	0.556*	0.533*	0.565*	0.577*	0.628	0.571	0.660	0.718	0.648
0.7	0.710	0.605	0.563*	0.573*	0.555	0.550	0.617	0.592	0.663	0.628	0.648
0.8	0.687	0.546	0.557*	0.579*	0.522	0.551	0.569	0.570	0.655	0.685	0.648
0.9	0.666	0.580	0.560	0.611	0.553	0.614	0.562	0.663	0.656	0.622	0.648
1.0	0.648	0.648	0.648	0.648	0.648	0.648	0.648	0.648	0.648	0.648	0.648

All entries are from 500 runs except for the margins which are exact, and the items marked

* which are from approximately 14,000 runs.

TABLE T13. The value of history for near-symmetrical multiple-stage finite selection with V = 1 and π = 0.0625.

	N₁ = 16 (finite target = 1.766)				N₁ = 64 (finite target = 1.911)				
	G_R/σ		% inc. in gain due to history	$Var(G)/\sigma^2$ with history	G_R/σ		% inc. in gain due to history	$Var(G)/\sigma^2$ with history	Percent inc. due to history in infinite case**
k	without history	with history			without history	with history			
1	1.249	1.249	-	0.640	1.351	1.351	-	0.173	-
2	1.392±0.004	1.401±0.004	0.6	0.555	1.517±0.008*	1.554±0.008	2.4	0.143	1.0
3	1.388±0.016*	1.442±0.008	3.9	0.561	1.541±0.018*	1.598±0.010	3.7	0.131	2.2
4	1.380±0.017*	1.453±0.011	5.3	0.533	1.539±0.019*	1.610±0.011	4.6	0.122	3.4
5	1.354±0.017*	1.427±0.011	5.4	0.548	1.554±0.020*	1.592±0.011	2.5	0.139	
6	1.385±0.017*	1.436±0.011	5.6	0.541	1.532±0.018	1.627±0.010	6.2	0.131	
8	1.296±0.019*	1.426±0.011	2.2	0.550	1.498±0.022*	1.606±0.010	7.2	0.125	
10					1.505±0.021*	1.607±0.009	6.8	0.149	
12					1.438±0.021*	1.581±0.010	9.9	0.115	
20					1.332±0.026*	1.543±0.009	15.8	0.132	

* Finney's (1966) results.
** Exact infinite values not available for k > 4.

TABLE T14. Values of 10-stage finite gain for $N_1 = 100$, $n = 1$ and $V = 1$.

G_R/σ	2.246 ± 0.011	2.274 ± 0.011*	2.288 ± 0.010	2.289 ± 0.010	2.274 ± 0.011
N_1	100	100	100	100	100
N_2	63	63	63	63	72
N_3	40	40	40	40	47
N_4	25	25	25	25	30
N_5	16	16	16	16	19
N_6	10	10	10	10	12
N_7	6	6	6	6	7
N_8	4	4	4	4	4
N_9	3	3	3	3	3
N_{10}	2	2	2	2	2
n	1	1	1	1	1
α_1	0.05	0.1	0.15	0.17	0.17
α_2	0.0625	0.1	0.1375	0.1475	0.1475
α_3	0.075	0.1	0.125	0.135	0.135
α_4	0.0875	0.1	0.1125	0.1225	0.1225
α_5	0.1	0.1	0.1	0.1	0.1
α_6	0.1	0.1	0.1	0.1	0.1
α_7	0.1125	0.1	0.0875	0.0875	0.0875
α_8	0.125	0.1	0.075	0.065	0.065
α_9	0.1375	0.1	0.0625	0.0525	0.0525
α_{10}	0.15	0.1	0.05	0.02	0.02

* this column corresponds to near-symmetrical selection.

TABLE T15. Values of the third and fourth cumulants after simulated 2-stage selection from a finite population with $N_1 = 16$, $N_3 = 1$, and $V = 0.2$ (350 runs each except that * indicates at least 1000 runs).

N_2: α_1	1	2	3	4	5	6	8	10	12	14	16
third cumulant											
0.0	0.1463	0.1269	0.1775	0.0842	0.0623	0.1069	0.0602	0.0118	0.0623	0.0722	-0.0298
0.1	-0.0316	0.0781	0.0890	0.0166	0.0572	0.0046	0.1634	0.0215	0.0731	0.0159	0.0360
0.2	0.0668	0.0763	0.0029	0.0655	0.0921	0.0495	0.1459	-0.0032	0.0488	0.0183	0.0880
0.3	0.0387	0.0861	0.1402*	0.0510*	0.0713*	0.0687	0.1171	0.1442	0.0204	0.0276	0.0273
0.4	0.0728	0.1002	0.0499*	0.0186*	0.0876*	0.0400	0.1016	0.1400	0.1275	0.0802	0.0916
0.5	0.0602	0.0688	0.0355*	0.0493*	0.0663*	0.0710	0.0646	0.1221	0.0986	0.0258	0.0401
0.6	0.0994	0.1002	0.0433*	0.0464*	0.0127*	0.0386	0.0818	0.1421	0.1295	0.0348	0.0397
0.7	0.1508	0.0432	0.0659	0.0098*	0.0640*	0.0321	0.0776	0.0249	0.0920	0.0719	0.0308
0.8	0.1472	0.0309	0.0301	0.0816	0.0785	0.0150	0.0437	0.0735	0.0624	0.0345	0.0259
0.9	0.0811	0.0793	0.0316	0.0850	0.1078	0.0449	0.0827	0.0381	0.1166	0.0450	0.0284
1.0	0.1883	0.0578	0.0122	0.0548	0.0413	0.0351	0.0659	0.0506	0.0989	0.0083	0.0651
fourth cumulant											
0.0	-0.0067	0.0664	0.0491	0.0470	-0.0534	0.0631	0.0489	-0.0715	0.0117	0.0695	0.1022
0.1	-0.0206	0.1294	0.0689	0.0218	-0.0520	0.0115	0.1934	-0.0965	0.0311	-0.0032	-0.0211
0.2	-0.0411	0.1359	0.1082	-0.0270	-0.0095	0.0092	0.1062	-0.0730	0.0165	-0.0363	0.0251
0.3	-0.0978	0.0717	0.1583*	-0.0219*	-0.0130*	-0.0244	0.1142	0.2089	0.0118	0.0401	0.0296
0.4	0.0561	0.0961	-0.0085*	-0.0107*	0.0911*	-0.0010	0.0370	0.1475	0.0345	0.0311	0.0086
0.5	0.0274	0.0032	-0.0062*	0.0197*	0.0166*	-0.0188	0.0340	0.0914	0.0108	-0.0439	-0.0644
0.6	0.0630	0.0107	-0.0071*	0.0166*	-0.0290*	-0.0635	0.0446	0.2236	0.0628	-0.0369	-0.0726
0.7	0.2318	-0.0093	0.0103	-0.0458*	0.0098*	-0.0329	0.0162	-0.0345	0.0639	0.0265	-0.0728
0.8	0.0971	0.0101	0.0075	0.0230	0.0501	-0.0260	-0.0626	-0.0085	0.0924	0.0611	-0.0327
0.9	0.1276	-0.0793	0.0185	0.0549	0.1357	-0.0451	0.0417	-0.0185	0.0308	-0.0155	-0.0616
1.0	0.2072	0.1218	-0.0460	0.0290	0.0992	-0.0109	-0.0378	0.0098	0.0304	-0.0510	-0.0347