FRESHWATER
BIOLOGICAL ASSOCIATION
OF THE
BRITISH EMPIRE

Scientific Publication No. 10

ON STATISTICAL TREATMENT OF THE RESULTS OF PARALLEL TRIALS WITH SPECIAL REFERENCE TO FISHERY RESEARCH

BY

H. J. BUCHANAN-WOLLASTON

(Principal Naturalist on the staff of the Ministry of Agriculture and Fisheries, working at the Laboratories of the Freshwater Biological Association, Wray Castle, Ambleside, Westmorland)

PRICE TO NON-MEMBERS

2s. 6d.

1945

FRESHWATER BIOLOGICAL ASSOCIATION
WRAY CASTLE, AMBLESIDE, WESTMORLAND

PUBLICATIONS

SCIENTIFIC PUBLICATIONS

(Nos. 1 to 6 and No. 8 post free 1s. 7d. each. Nos. 7, 9 and 10 post free 2s. 7d.)

No. 1. A key to the British Species of Corixidae (Hemiptera-Heteroptera) with notes on their distribution, by T. T. Macan.

No. 2. A key to the British Species of Plecoptera (Stoneflies) with notes on their ecology, by H. B. N. Hynes.

No. 3. The Food of Coarse Fish, by P. H. T. Hartley.

No. 4. A key to the British Water Bugs (Hemiptera-Heteroptera excluding Corixidae) with notes on their ecology, by T. T. Macan.

No. 5. A key to the British Species of Freshwater Cladocera, with notes on their ecology, by D. J. Scourfield and J. P. Harding.

No. 6. The Production of Freshwater Fish for Food, by T. T. Macan, C. H. Mortimer and E. B. Worthington.

No. 7. Keys to the British Species of Ephemeroptera, with keys to the genera of the nymphs, by D. E. Kimmins.

No. 8. Keys to the British Species of Aquatic Megaloptera and Neuroptera, by D. E. Kimmins.

No. 9. The British Simuliidæ, with Keys to the Species in the Adult, Pupal and Larval stages, by John Smart.

No. 10. On Statistical Treatment of the Results of Parallel Trials with special reference to Fishery Research, by H. J. Buchanan-Wollaston.

ANNUAL REPORTS. These summarize the scientific work undertaken in each year.

Nos. 1–6 (some out of print, post free 1s. 2d. each).

Nos. 7–12, for the years 1939 to 1944 (post free 1s. 8d. each).

RESULTS OF RESEARCH. These are published in scientific journals. A limited number of reprints is available to members on request.

MEMBERSHIP

Membership, which includes free publications, the right to work at the Laboratories, etc., is open to any who wish to give their support to the Association, the minimum annual subscription being £1. 0s. 0d.

All communications concerning publications and membership should be addressed to the Director (Dr E. B. Worthington).

ON STATISTICAL TREATMENT OF THE RESULTS OF PARALLEL TRIALS WITH SPECIAL REFERENCE TO FISHERY RESEARCH

By H. J. BUCHANAN-WOLLASTON

(Principal Naturalist on the staff of the Ministry of Agriculture and Fisheries, working at the Laboratories of the Freshwater Biological Association, Wray Castle, Ambleside, Westmorland)

CONTENTS

PREFACE

In this paper all the reliable methods of testing for significance the results of parallel trials of a certain type are described fully. Some sections relate to exact, others to approximate tests. The only advantage in the use of the latter lies in the fact that they are often the more expeditious. Apart from this it is always preferable to use exact methods. These, too, have the advantage that their theory is based on simple laws of chance which are comprehensible by those who have had no training whatever in statistical theory. These laws are developed *ab initio* in the first two subsections of Section 14, and it is recommended that these subsections be read and mastered before the rest of the paper is read. It must not be expected that understanding of statistical theory will come easily to those unused to reasoning logically and mathematically. It is a difficult subject and considerable concentration may be necessary. Assuming that I am speaking to people with no previous knowledge of statistics, I should recommend that, after the two subsections mentioned, the other sections should be read in the following order: Introduction, 1, 13, 2, 10 (omitting paragraph two), 4A, 6, 11, 12, 15. These sections should then be read again at least once, and the examples included worked out, independently of the detailed descriptions, by the methods given. The reader should then be in a position to apply all the exact tests and the most useful approximate test to any case of parallel trials which may arise, except for those in which a set of heterogeneous results has to be dealt with. In using the exact tests it will be found unnecessary to refer to any other statistical work except the tables mentioned in Section 1.

The use of the approximate tests described and the application of the very useful methods of Sections 8 and 9 necessitate some previous knowledge of statistical theory, including that applying to the normal distribution and the χ^2-distribution. These are treated fully in modern statistical text-books. Independent parallel trials, on the other hand, receive very scanty treatment in these while in some of them they are not even mentioned. Out of eight representative modern text-books on statistics which I have examined, only one, Fisher's *Statistical Methods for Research Workers*, includes a description of the exact test appropriate to the fourfold table; in

three no treatment of the subject is included, while no mention is made, in any one of the eight, of the methods of our Sections 8, 9 and 11, which appear to be new.

Workers using the methods described in this paper will undoubtedly come across numerous cases in which results of parallel trials have been tested for significance by incorrect methods giving misleading conclusions. To describe these incorrect methods and to say why they are incorrect would take up too much space. It is strongly recommended that anybody coming across reports on research in which he is interested and in which methods of parallel trial are used should not accept the results given without applying the tests given in this paper to see whether such results are reliable.

INTRODUCTION

Parallel trials form a most important part of the technique of scientific experimentation. Such trials may be divided into two categories. In the first the results are comparable measurements of one kind or another. In the second the data consist of records of the number of times a certain 'event' has occurred in the two sets of trials compared. Only trials of the second category are dealt with here. Statisticians will recognize the appropriate technique as that applicable either to the fourfold table, or to the binomial distribution.

Whatever kind of experimental technique be used in parallel trials it is necessary to apply statistical tests for significance to the results if these are liable to chance variation. The object of an experiment of the kind in question is to find out whether a difference in treatment of the experimental subjects has an effect of a particular kind. The experiment cannot prove that there is no effect but it can, to all intents and purposes, prove its reality if it does exist and if the correct allowance be made for chance variation. If no such allowance, provided by a statistical test for significance, be made, it is extremely likely that an apparent effect will be taken to be real whereas it may very well have been due to chance and not to the difference in treatment. The philosophical basis of tests for significance is discussed in Section 13.

Parallel trials have hitherto been treated in statistical works as a special case of the very wide class known as contingency tables. In this paper the result of two sets of parallel trials is throughout treated as a difference between the numbers of times a certain event of interest has occurred in the two sets of trials, the relative chance of occurrence having been assumed for the purpose of the test to be the same for each set. This assumption is the assumption of the truth of what R. A. Fisher has called the 'null hypothesis', the purpose of the test being to find out whether that hypothesis is acceptable or not. This line of approach is logically simpler than the approach by way of the contingency table and can be understood quite easily even by those unused to mathematical reasoning. The null hypothesis must always be acceptable *a priori*. Only if the results of the experiment show it to be unacceptable can the reality of the effect studied be considered to be proven.

It is important to draw the distinction between independent parallel trials and those which may be termed mutually dependent or interdependent parallel trials. In the former the occurrence of the event of interest in one member of any pair of trials does not influence in any way the occurrence of that event in the other member. In any single pair of trials the event may occur in neither member, in both members, or in one but not in the other. In interdependent parallel trials the event *must* occur in one member or the other. It is not always easy to ensure that an experiment takes one form or the other, but the unambiguous formulation of the appropriate null hypothesis will always settle the matter. For example, in comparing the catch of two eel traps through which all the water of a river has to flow, every eel attempting to run down the river must be caught in one trap or the other. The events, capture of an eel in trap A, capture of an eel in trap B, are mutually exclusive, and the appropriate statistical test for the significance of the difference between the two catches would be by way of the binomial $(0{\cdot}5+0{\cdot}5)^n$, where n is the total number of eels caught—as in tossing n coins to see if there is any bias towards heads or tails.* Here the null hypothesis, that each eel running is equally likely to be caught by trap A or trap B is *a priori* reasonable. If, now, we arrange two traps side by side but only an unknown fractional part of the water of the river flows through the traps we have no knowledge of the number of eels exposed to risk of capture. The null hypothesis that of eels *caught* each has the same chance of going into trap A as it has of going into trap B is, however, *a priori* reasonable; the binomial test is again applicable. If, again, it has been found experimentally that trap A, over a long period, has taken three-quarters of the total catch and we wish to find out if a certain inhibitive stimulus has an appreciable effect in preventing the entry of eels into a trap we can apply this stimulus in the case of one trap or the other and test the result by way of the binomial $(0{\cdot}75+0{\cdot}25)^n$. The purpose of the test is simply to find out whether the application of the stimulus has upset the relative catches of the traps and no interpretation of this effect is implied. That is the investigator's business. An alternative method of investigating the effect of the stimulus would be by independent parallel trials, the same trap being used for all trials. Here the null hypothesis would be: The trap takes the same proportion of the eels running whether the stimulus be applied or not. The

* See Section 10.

experimental technique would be to apply and omit the stimulus for alternate periods of time preferably equal in length. With this method it would be absolutely necessary to have an accurate estimate of the number of eels running during each period or, alternatively, of the ratios between the numbers if these are very large compared with the numbers of eels caught in the trap. Yet another method, which eliminates the necessity for knowledge of the number of eels running would be as follows: Two traps are arranged, one some distance behind the other and on the same side of the river. For the first period the inhibitive stimulus is applied to A and not to B, for the next period to B and not to A, and so on. The null hypothesis would take the form: The proportion of the total catch which is taken by A is the same, to whichever trap the stimulus be applied. This case is discussed in Section 2.

It will be seen how very important it is, in applying statistical tests, to formulate the appropriate null hypothesis without ambiguity and to make certain that it is reasonable *a priori*. It may be considered to be an axiom that every sound experiment of which the results are liable to chance variation is backed by a null hypothesis for the testing of the acceptability of which the experiment is designed to furnish all necessary information. This may have a great effect on the design of experiments.

The main object of this paper is to make easy the exact allowance for the chance element in interpreting the results of experiments. The usual practice of employing fixed criteria for 'significance' and 'high significance', though necessary when the calculation of the exact effect of chance in a variety of cases is a matter of great difficulty, cannot be considered satisfactory in the case of parallel trials, since the random sampling distribution of the difference between numbers of occurrences in these is generally easy to calculate.

The use of approximate methods of allowing for the influence of chance in experiments is not recommended except for preliminary examination. The approximate methods described in Sections 3 A and 3 B are very useful for this. In experiments in which the numbers of subjects are large and also in observational work the application of exact methods may be very laborious. Here approximate methods may be necessary, and luckily with large numbers these give much more reliable results than when numbers are small. Yates (1) has treated the question very fully. In Section 3 B a description of one of Yates's methods which is applicable to parallel

trials is given, and in Section 11 a method is described which is most useful and may perhaps be considered as superseding Yates's method.

In experimental work the case in which independent parallel trials are made with sets of subjects equal in number occurs much more often than that in which these differ in number. The former is here treated in detail, and tables are given from which the significance of a result can be estimated at a glance. These tables cover a considerable range of experimental numbers. A table is also provided by the aid of which the range of the exact test may be extended very easily, and it is thought that few cases will arise in laboratory experiments which are not covered by the tables. These have been very carefully checked, and are believed to be correct to within ± 1 in the last figure. In Section 6 it is explained how to apply the exact test for significance in cases beyond the range of any of the tables.

All the tests described are applicable in a very wide field beyond that which may be strictly termed experimental. Some examples are given in Section 15.

SECTION 1. *The exact test for the significance of the result of two sets of independent parallel trials in the case in which the numbers of trials in the sets are equal to one another* ($n_1 = n_2 = n$)

The appropriate technique is most easily presented by way of examples. Fortunately, experiments made at Wray Castle and other places by the F.B.A. provide examples suitable for the application of most of the necessary methods of statistical treatment. Though some of the actual figures obtained will be used, any necessary modifications will be made in the data to render them suitable for demonstration of statistical treatment.

In one experiment a set of 10 fish, all of the same species, were subjected to a certain stimulus, A, and the numbers which had reacted to the stimulus after periods of 5 and 30 min. respectively were recorded. The experiment was repeated after an interval with the same fish, but this time another stimulus, B, was applied simultaneously with A, the purpose of the experiment being to find out whether B had an inhibiting effect on the reaction of the fish to A. Paired trials of the same kind were then carried out on four other sets of 10 fish, since it was found that the result from one set was not sufficient to answer the inquiry definitely. Table 1 shows the results of the trials.

In the case of the first set of 10 fish 6 had reacted to stimulus A in 5 min. in the absence of stimulus B, while only 2 had reacted when B was also applied. The difference in number of reacting fish was 4, while the total of reactions in the two trials was 8. Now it is clear that, even supposing that each fish had exactly the same chance of reacting when stimulus B was present as in its absence, there would be a chance that a difference of 4 in the numbers reacting would occur sometimes. The hypothesis that the fish had the same chance of reacting in the two trials is the null hypothesis,

Table 1. *Number of fish in each set of* 10 *which reacted to stimulus A in parallel trials*

	Without B					
Fish set ...	1	2	3	4	5	Total
Reacted in first 5 min.	6	8	3	4	7	28
Did not react in first 5 min.	4	2	7	6	3	22
Reacted between 5 and 30 min.	4	2	7	6	3	22
Did not react between 5 and 30 min.	0	0	0	0	0	0
No. of fish in experiment	10	10	10	10	10	50
	With B					
Fish set ...	1	2	3	4	5	Total
Reacted in first 5 min.	2	5	2	1	4	14
Did not react in first 5 min.	8	5	8	9	6	36
Reacted between 5 and 30 min.	6	4	6	8	4	28
Did not react between 5 and 30 min.	2	1	2	1	2	8
No. of fish in experiment	10	10	10	10	10	50

and the purpose of a statistical test for the significance of the difference, 4, is to see whether this is likely to have been due to chance. For carrying out the test we must know the 'random sampling distribution' of the difference between two numbers arising from the *same* chance of occurrence. The occurrence of 6 reactions in 10 trials gives 0·6 as the estimated value of p, the chance of 1 reaction in the absence of stimulus B, while 0·2 is the estimate of p in the presence of B. If the hypothesis that these two values were estimates of the same chance be true, then the best estimate of this chance is 0·4, since there are 8 reactions in 20 trials, all of which, according to the hypothesis, gave the same chance of a reaction. By the rules applying to 'degrees of freedom', when

calculating the sampling distribution of differences between two numbers such as these, we restrict ourselves to numbers which give the same total, namely, 8. This is only common sense, for any pair of numbers giving a different total would give a different value for the hypothetical chance from that we have obtained. The distribution required is therefore that giving the probability of any given difference between two numbers of which the total is 8 and which have arisen by the same chance from 10 trials in each case. Such a probability may be written $p(d)$, where d is the difference of which p is the probability. Since it is not justifiable *a priori* to assume that the difference is of particular sign, the probability considered is $p(\pm d)$.

It is clear that, as the number of trials is increased, the probability of getting any *particular* difference becomes smaller and smaller since more and more differences become possible. Thus, in a statistical test, in order that all cases under test may be comparable one with another, it is customary to consider, not $p(\pm d)$, the probability of $\pm d$, but $P(\pm d)$, the probability of a difference *at least* as great as d in either direction. The series to be used in the test therefore must give for each difference a term which is the sum of the probability of that difference and the probabilities of all possible greater differences. In other words P is the sum of the terms in the d-distribution beyond and including the terms $p(+d)$, $p(-d)$. The distribution has to answer, for each value of d, the question: Given s reactions and $2n-s$ non-reactions in $2n$ trials, what fraction of the total number of ways in which the reactions and non-reactions can be arranged in two sets of n in each gives the difference d reactions between the two sets? The calculation of this distribution is not difficult, but, since it varies both with variation in n, the number of paired trials, and in s, the sum of the two numbers giving the difference d, it is not practicable to tabulate all distributions of the kind which may arise in experimental work. Distributions covering a fairly wide range of cases are tabulated in Tables 2, 3 and 4 of this paper. Table 2 applies to pairs of trials from 2 to 15 in number, while Tables 3 and 4 apply to trials of 20 and 30 pairs respectively. If it is desired therefore to make exact tests of the results of paired trials the number of these should at first be one of those included in the tables. Any really important effect would probably show up with trials of 30 or fewer pairs of subjects. In some cases, however, it will be necessary to increase the number of trials. Methods of dealing with these will be described later.

Returning to our experimental result it is found, on reference to Table 2, that $P(\pm 4)=0{\cdot}169802$, when $n=10$ and $s=8$. A difference as great as 4 on either side would occur by chance about once in 6 trials if there were no difference between the chance of a reaction to A in the presence of B and that holding when B was absent. Clearly this is not sufficient to settle the question whether B has been *proved* to have an inhibitive effect. On repeating the experiment with another set of 10 fish it was found that the number of reactions was 5 with B against 8 without B. If we are justified in assuming that all the fish used in the experiments were homogeneous as to their reactivity to A, we may add together the results of experiments and consider the combination as one experiment. If this be done, we now have a total number of paired trials, 20, and $s=21$, $d=7$. In Tables 2, 3 and 4 no value of s occurs greater than n, but $P(d)$ when $s=s$ is equal to $P(d)$ when $s=2n-s$. Thus for entering Table 3, $n=20$, $s=19$, and $P(\pm 7)$ is found to be equal to 0·05616. The generally employed criterion for significance is $P=0{\cdot}05$. Where, however, an experiment can be repeated easily it is more important to notice whether or not P tends to decrease with increase in n rather than to adhere rigidly to a given value of P as a criterion. If there is a real difference between the chances of an occurrence in paired trials the value of P will always tend to decrease indefinitely as the number of trials is increased. If, on the other hand, there is no real difference between the chances, P will tend towards the value 0·5, varying in a random way about that value. Two values of P are not, however, sufficient to indicate a tendency. On further repetition of the experiment with another set of 10 fish, 3 reactions occurred without B against 2 with B. The value of n is now 30, $s=26$, $d=8$. From Table 4 it is found that $P=0{\cdot}06729$. This is rather greater than the corresponding value with n equal to 20, but such slight increases must be expected sometimes. The necessity for further trial is indicated. The next trial, taken in conjunction with those made previously, gave the values $n=40$, $s=31$, $d=11$. Since a value of 40 for n is beyond the range of the tables the value of $P(11)$ must be calculated. Table 5 is included to facilitate this. We have to find the probability of a difference of 11 and of differences greater than 11 positive or negative. The numbers, r, in Table 5 summing up to 31 and having differences of 11 or more are 21 and 10, 22 and 9, 23 and 8 and so on, the larger of any pair being equal to $(s+d)/2$. It will be seen that, when s is an odd number, only odd differences are possible.

The calculation of the logarithm of any value $p(\pm d)$ consists of adding together the entries in Table 5 for the numbers, r, and subtracting from the sum the logarithm of $\frac{(2n)!}{(2n-s)!s!2}$, or the entry for $n=2n$, $r=s$, if that be not beyond the range of Table 5, log 2 being subtracted from the value found. Logarithms of factorials will be found in Table LXIX of Pearson(2) or Table XXX of Fisher and Yates(3). These tables are quite necessary for anyone carrying out statistical tests. For calculations such as those with which we are dealing, which consist almost entirely of additions and subtractions, a small pocket adding machine will be found almost as convenient as a much more elaborate calculating machine. All the numerical work in the present paper was done with a small pocket adder costing about 6*s*. A set of 4-figure mathematical tables is also necessary. For our present calculation we find from Pearson's Table LXIX that

$$\log(80!)-\log(49!)-\log(31!)-\log 2=21{\cdot}8546.$$

It will be found that Table 5 only gives entries up to $r=(1/2)n$, but, when r is greater than this, the entry for $n-r$ is used instead, the entry for $n-r-1$ being used if n is an odd number. We have, therefore,

$$\begin{aligned}
\log p(11) &= 11{\cdot}1182+8{\cdot}9282-21{\cdot}8546=\bar{2}{\cdot}1918,\\
p(11) &= 0{\cdot}01555,\\
\log p(13) &= 11{\cdot}0545+8{\cdot}4369-21{\cdot}8546=\bar{3}{\cdot}6268,\\
p(13) &= 0{\cdot}004333,\\
\log p(15) &= 10{\cdot}9481+7{\cdot}8860-21{\cdot}8546=\bar{4}{\cdot}9795,\\
p(15) &= 0{\cdot}0009539,\\
\log p(17) &= 10{\cdot}7983+7{\cdot}2705-21{\cdot}8546=\bar{4}{\cdot}2142,\\
p(17) &= 0{\cdot}0001638.
\end{aligned}$$

It is not necessary to carry the calculation further. The value of $P(\pm 11)$ is given accurately enough by summing the above values of p. $P(\pm 11)$ is thus almost exactly equal to 0·021. In such calculations as this the greatest probability should always be calculated first so that the process may be stopped when sufficient precision has been reached. Though not really necessary, since the significance of the result of the experiment may be considered to have been proved by the test applied to the 40 pairs of trials, a further trial with 10 more fish was made. Adding the data to those already obtained we have $n=50$, $s=42$, $d=14$, and the values of r

Table 2

		Probability (P) of at least as great a difference, positive or negative, as						
n	s	2 or 3	4 or 5	6 or 7	8 or 9	10 or 11	12 or 13	14 or 15
2	2	·3						
3	2	·4						
3	3	·1						
4	2	·428571						
4	3	·142857						
4	4	·4571428	·0285714					
5	2	·4						
5	3	·16						
5	4	·523809	·047619					
5	5	·206349	·00793651					
6	2	·45						
6	3	·18						
6	4	·54	·06					
6	5	·24	·015					
6	6	·567100	·0800866	·00216450				
7	2	·461538						
7	3	·192308						
7	4	·559441	·0699301					
7	5	·265734	·0209790					
7	6	·592074	·102564	·00466200				
7	7	·286131	·0291376	·0(3)582751				
8	2	·46						
8	3	·2						
8	4	·569231	·076923					
8	5	·282051	·0256410					
8	6	·608392	·118881	·00699301				
8	7	·314685	·0405594	·00139860				
8	8	·619270	·131935	·01	·0(3)155400			
9	2	·470588						
9	3	·205882						
9	4	·576471	·0823529					
9	5	·294118	·0294118					
9	6	·619909	·131222	·00904977				
9	7	·334842	·0497737	·00226244				
9	8	·637186	·153435	·0152200	·0(3)411353			
9	9	·346935	·0566845	·00337310	·0(4)411353			
10	2	·473684						
10	3	·210526						
10	4	·528043	·0866873					
10	5	·303405	·0325077					
10	6	·628483	·140867	·0108359				
10	7	·349845	·0572756	·00309597				
10	8	·649916	·169802	·0197666	·0(3)714456			
10	9	·369850	·0697785	·00547750	·0(3)119076			
10	10	·656282	·178895	·0230141	·00109333	·0(4)108251		
11	2	·476190						
11	3	·2142857						
11	4	·586466	·0902256					
11	5	·310777	·0350877					
11	6	·635117	·148607	·0123839				
11	7	·361455	·0634675	·00386997				
11	8	·659443	·182662	·0237358	·00103199			
11	9	·386997	·0804953	·00751880	·0(3)221141			
11	10	·669921	·198380	·0299731	·00190522	·0(4)340217		
11	11	·394856	·0861089	·00892219	·0(3)345887	·0(5)283514		

		Probability (P) of at least as great a difference, positive or negative, as						
n	s	2 or 3	4 or 5	6 or 7	8 or 9	10 or 11	12 or 13	14 or 15
12	2	·478261						
12	3	·217391						
12	4	·590062	·0931677					
12	5	·316770	·0372670					
12	6	·640405	·154953	·0137300				
12	7	·370709	·0686499	·00457666				
12	8	·666846	·193027	·0271907	·00134608			
12	9	·400323	·0893794	·00942253	·0(3)336519			
12	10	·680172	·213757	·0360749	·00275946	·0(4)673038		
12	11	·413649	·0995327	·0122781	·0(3)644193	·0(5)961483		
12	12	·684273	·220347	·0391257	·00332895	·0(3)107242	·0(6)739602	
13	2	·480000						
13	3	·220000						
13	4	·593043	·0956521					
13	5	·321739	·0391304					
13	6	·644720	·160248	·0149068				
13	7	·378261	·0730435	·00521739				
13	8	·672769	·201556	·0302059	·00164760			
13	9	·410984	·0968421	·0111670	·0(3)457666			
13	10	·688168	·226195	·0414053	·00360749	·0(3)107686		
13	11	·428308	·110701	·0154193	·0(3)982636	·0(4)201911		
13	12	·695097	·237744	·0471800	·00483241	·0(3)212680	·0(5)269215	
13	13	·433753	·115238	·0169317	·00120262	·0(4)326904	·0(6)192297	
14	2	·481						
14	3	·2						
14	4	·595	·097					
14	5	·3259	·0407					
14	6	·648309	·164734	·0159420				
14	7	·384541	·0768116	·00579710				
14	8	·677617	·208696	·0328502	·00193237			
14	9	·419710	·103188	·0127536	·0(3)579710			
14	10	·694584	·236461	·0460717	·00442410	·0(3)152555		
14	11	·440071	·120112	·0183066	·00133910	·0(4)339012		
14	12	·703567	·251860	·0542379	·00632955	·0(3)341006	·0(5)598256	
14	13	·449482	·128346	·0213009	·00183814	·0(4)687995	·0(6)747820	
14	14	·706390	·256800	·0569826	·00702821	·0(3)422688	·0(5)982148	·0(7)498547
15	2	·482759						
15	3	·224138						
15	4	·597701	·0996169					
15	5	·329502	·0421456					
15	6	·651341	·168583	·0168583				
15	7	·389847	·0800767	·00632184				
15	8	·681659	·214759	·0351824	·00219890			
15	9	·426987	·108646	·0141929	·0(3)699650			
15	10	·699850	·245077	·0501749	·00519740	·0(3)199900		
15	11	·449725	·128136	·0209395	·00169916	·0(4)499751		
15	12	·710372	·263531	·0604224	·00776454	·0(3)483969	·0(4)105210	
15	13	·462137	·139420	·0253277	·00250927	·0(3)115732	·0(5)175351	
15	14	·715249	·272304	·0655955	·00922057	·0(3)678917	·0(4)218673	·0(6)206295
15	15	·466092	·143111	·0268377	·00281433	·0(3)145064	·0(5)291389	·0(7)128934

Notes on Tables 2, 3 and 4

When s is an odd number only odd differences are possible. When s is an even number only even differences are possible.
The tables include values of P for values of s up to n. If s is greater than n take the value of $2n-s$ as the value of s for entering the tables.
The figures in brackets in the tables give the number of ciphers preceding the first significant figure. Thus ·0(4)411353 stands for ·0000411353.

Tables 3 and 4. ($n=20$), ($n=30$)

	Probability (P) of at least as great a difference, positive or negative, as														
s	2 or 3	4 or 5	6 or 7	8 or 9	10 or 11	12 or 13	14 or 15	16 or 17	18 or 19	20 or 21	22 or 23	24 or 25	26 or 27	28 or 29	30
2	·4872														
3	·2308														
4	·6049	·1060													
5	·3416	·04713													
6	·6614	·1817	·02019												
7	·4076	·09148	·008316												
8	·6948	·2351	·04360	·003276											
9	·4506	·1274	·01966	·001228											
10	·7163	·2732	·06484	·008361	·0004359										
11	·4802	·1553	·03096	·003342	·0001453										
12	·7312	·3009	·08237	·01382	·001247	·00004510									
13	·5006	·1760	·04080	·005796	·0004316	·00001288									
14	·7411	·3203	·09584	·01870	·002200	·0001369	·0(5)3340								
15	·5145	·1908	·04837	·007913	·0007716	·00003931	·0(6)7709								
16	·7475	·3333	·1054	·02248	·003056	·0002444	·00001002	·0(6)1542							
17	·5232	·2004	·05355	·009530	·001065	·00006861	·0(5)2210	·0(7)2569							
18	·7513	·3407	·1110	·02484	·003642	·0003285	·00001665	·0(6)4056	·0(8)3352						
19	·5272	·2050	·05616	·01039	·001233	·00008750	·0(5)3358	·0(7)5819	·0(9)3047						
20	·7523	·3430	·1128	·02565	·003848	·0003600	·00001939	·0(6)5296	·0(8)5817	·0(10)1451					
2	·4916														
3	·2372														
4	·6119	·1124													
5	·3532	·05218													
6	·6707	·1945	·02372												
7	·4238	·1028	·01054												
8	·7064	·2542	·05231	·004575											
9	·4717	·1455	·02569	·001936											
10	·7307	·2990	·07973	·01219	·0007971										
11	·5062	·1806	·04191	·005579	·00031188										
12	·7481	·3335	·1042	·02115	·002466	·0001236									
13	·5321	·2093	·05747	·01025	·001050	·00004635									
14	·7611	·3604	·1253	·03034	·004769	·0004310	·00001677								
15	·5522	·2326	·07161	·01533	·002129	·0001698	·0(5)5833								
16	·7711	·3817	·1432	·03910	·007410	·0009099	·00006416	·0(5)1944							
17	·5675	·2516	·08403	·02037	·003421	·0003717	·00002315	·0(6)6186							
18	·7787	·3984	·1580	·04702	·01012	·001506	·0001448	·0(5)7954	·0(6)1870						
19	·5795	·2668	·09462	·02506	·004790	·0006310	·00005354	·0(5)2591	·0(7)5343						
20	·7846	·4117	·1702	·05389	·01269	·002152	·0002507	·00001875	·0(6)7961	·0(7)1433					
21	·5889	·2789	·1033	·02919	·006106	·0009162	·00009414	·0(5)6182	·0(6)2294	·0(8)3584					
22	·7891	·4219	·1798	·05960	·01498	·002779	·0003681	·00003324	·0(5)1908	·0(7)6147	·0(9)8269				
23	·5959	·2882	·1103	·03259	·007275	·001193	·0001390	·00001098	·0(6)5475	·0(7)1519	·0(9)1741				
24	·7925	·4296	·1872	·06407	·01685	·003330	·0004808	·00004904	·0(5)3367	·0(6)1447	·0(8)3421	·0(10)3295			
25	·6010	·2950	·1154	·03522	·008212	·001430	·0001810	·00001605	·0(6)9502	·0(7)3480	·0(9)6918	·0(11)5489			
26	·7947	·4348	·1923	·06729	·01825	·003756	·0005738	·00006317	·0(5)4834	·0(6)2439	·0(8)7515	·0(9)1231	·0(12)7841		
27	·6042	·2993	·1188	·03700	·008870	·001604	·0002135	·00002027	·0(5)1324	·0(7)5622	·0(8)1428	·0(10)1877	·0(13)9226		
28	·7962	·4379	·1954	·06923	·01911	·004026	·0006346	·00007296	·0(5)5919	·0(6)3252	·0(7)1138	·0(9)2323	·0(11)2356	·0(14)8389	
29	·6058	·3015	·1205	·03789	·009207	·001694	·0002309	·00002268	·0(5)1549	·0(7)7021	·0(8)1975	·0(10)3109	·0(12)2285	·0(15)5242	
30	·7966	·4390	·1965	·06986	·01939	·004118	·0006558	·00007640	·0(5)6321	·0(6)3564	·0(7)1298	·0(9)2819	·0(11)3215	·0(13)1524	·0(16)1691

Table 5. *Logarithms of binomial coefficients*

n \ r	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
2	0·3010																							
3	0·4771																							
4	0·6021	0·7782																						
5	0·6990	1·0000																						
6	0·7782	1·1761	1·3010																					
7	0·8451	1·3222	1·5441																					
8	0·9031	1·4472	1·7482	1·8451																				
9	0·9542	1·5563	1·9243	2·1004																				
10	1·0000	1·6532	2·0792	2·3222	2·4014																			
11	1·0414	1·7404	2·2175	2·5185	2·6646																			
12	1·0792	1·8195	2·3424	2·6946	2·8987	2·9657																		
13	1·1139	1·8921	2·4564	2·8543	3·1096	3·2345																		
14	1·1461	1·9590	2·5611	3·0004	3·3015	3·4776	3·5356																	
15	1·1761	2·0212	2·6580	3·1351	3·4776	3·6994	3·8086																	
16	1·2041	2·0792	2·7482	3·2601	3·6403	3·9035	4·0584	4·1096																
17	1·2304	2·1335	2·8325	3·3766	3·7916	4·0926	4·2889	4·3858																
18	1·2553	2·1847	2·9117	3·4857	3·9329	4·2687	4·5028	4·6411	4·6868															
19	1·2788	2·2330	2·9863	3·5884	4·0655	4·4335	4·7023	4·8784	4·9656															
20	1·3010	2·2788	3·0569	3·6853	4·1905	4·5884	4·8894	5·1003	5·2252	5·2666														
21	1·3222	2·3222	3·1239	3·7771	4·3086	4·7345	5·0655	5·3086	5·4683	5·5474														
22	1·3424	2·3636	3·1875	3·8642	4·4205	4·8728	5·2318	5·5048	5·6967	5·8107	5·8485													
23	1·3617	2·4031	3·2482	3·9472	4·5270	5·0041	5·3894	5·6905	5·9123	6·0585	6·1310													
24	1·3802	2·4409	3·3062	4·0264	4·6284	5·1290	5·5392	5·8666	6·1165	6·2925	6·3973	6·4320												
25	1·3979	2·4771	3·3617	4·1021	4·7254	5·2482	5·6819	6·0341	6·3103	6·5144	6·6491	6·7160												
26	1·4150	2·5119	3·4150	4·1746	4·8181	5·3622	5·8181	6·1938	6·4948	6·7252	6·8880	6·9849	7·0171											
27	1·4314	2·5453	3·4661	4·2443	4·9070	5·4713	5·9484	6·3464	6·6709	6·9262	7·1152	7·2401	7·3023											
28	1·4472	2·5775	3·5153	4·3112	4·9925	5·5761	6·0734	6·4925	6·8393	7·1180	7·3319	7·4832	7·5734	7·6033										
29	1·4624	2·6085	3·5628	4·3757	5·0747	5·6767	6·1933	6·6327	7·0007	7·3017	7·5391	7·7151	7·8316	7·8896										
30	1·4771	2·6385	3·6085	4·4378	5·1538	5·7737	6·3087	6·7674	7·1556	7·4778	7·7374	7·9370	8·0783	8·1626	8·1907									
31	1·4914	2·6675	3·6527	4·4978	5·2302	5·8671	6·4199	6·8970	7·3045	7·6469	7·9277	8·1496	8·3144	8·4235	8·4779									
32	1·5051	2·6955	3·6954	4·5558	5·3040	5·9572	6·5271	7·0219	7·4479	7·8097	8·1107	8·3537	8·5408	8·6734	8·7526	8·7789								
33	1·5185	2·7226	3·7368	4·6119	5·3754	6·0444	6·6306	7·1425	7·5862	7·9664	8·2868	8·5500	8·7583	8·9132	9·0158	9·0370								
34	1·5315	2·7490	3·7769	4·6663	5·4444	6·1287	6·7308	7·2590	7·7198	8·1177	8·4565	8·7391	8·9675	9·1436	9·2686	9·3422	9·3680							
35	1·5441	2·7745	3·8159	4·7190	5·5114	6·2104	6·8277	7·3717	7·8489	8·2638	8·6204	8·9214	9·1692	9·3655	9·5116	9·6085	9·6568							
36	1·5563	2·7993	3·8537	4·7702	5·5763	6·2896	6·9216	7·4809	7·9738	8·4052	8·7787	9·0975	9·3638	9·5794	9·7457	9·8638	9·9344	9·9579						
37	1·5682	2·8235	3·8904	4·8199	5·6394	6·3664	7·0127	7·5867	8·0948	8·5420	8·9320	9·2678	9·5518	9·7858	9·9715	10·1098	10·2015	10·2473						
38	1·5798	2·8470	3·9261	4·8682	5·7007	6·4410	7·1011	7·6893	8·2122	8·6746	9·0804	9·4326	9·7336	9·9854	10·1895	10·3471	10·4591	10·5261	10·5483					
39	1·5911	2·8698	3·9609	4·9151	5·7602	6·5136	7·1870	7·7890	8·3262	8·8033	9·2243	9·5923	9·9097	10·1785	10·4004	10·5765	10·7077	10·7949	10·8384					
40	1·6021	2·8921	3·9948	4·9609	5·8182	6·5842	7·2705	7·8860	8·4369	8·9282	9·3639	9·7472	10·0804	10·3656	10·6045	10·7983	10·9481	11·0545	11·1182	11·1394				
41	1·6128	2·9138	4·0278	5·0055	5·8747	6·6529	7·3518	7·9802	8·5445	9·0497	9·4996	9·8976	10·2460	10·5470	10·8023	11·0132	11·1807	11·3056	11·3886	11·4300				
42	1·6232	2·9350	4·0599	5·0490	5·9298	6·7198	7·4310	8·0720	8·6492	9·1678	9·6315	10·0437	10·4069	10·7231	10·9942	11·2214	11·4060	11·5486	11·6501	11·7108	11·7310			
43	1·6335	2·9557	4·0913	5·0914	5·9835	6·7851	7·5082	8·1614	8·7512	9·2827	9·7598	10·1858	10·5632	10·8942	11·1805	11·4236	11·6245	11·7842	11·9034	11·9825	12·0221			
44	1·6435	2·9760	4·1220	5·1328	6·0359	6·8488	7·5834	8·2485	8·8506	9·3947	9·8848	10·3241	10·7153	11·0605	11·3616	11·6198	11·8366	12·0126	12·1489	12·2458	12·3038	12·3231		
45	1·6532	2·9957	4·1520	5·1732	6·0870	6·9109	7·6569	8·3336	8·9475	9·5038	10·0065	10·4588	10·8634	11·2224	11·5377	11·8107	12·0426	12·2345	12·3871	12·5011	12·5768	12·6145		
46	1·6628	3·0150	4·1813	5·2127	6·1370	6·9716	7·7286	8·4165	9·0421	9·6103	10·1252	10·5901	11·0076	11·3800	11·7091	11·9963	12·2430	12·4501	12·6185	12·7488	12·8416	12·8971	12·9156	
47	1·6721	3·0339	4·2099	5·2513	6·1858	7·0310	7·7986	8·4976	9·1344	9·7142	10·2410	10·7181	11·1482	11·5336	11·8760	12·1770	12·4379	12·6598	12·8434	12·9896	13·0987	13·1713	13·2075	
48	1·6812	3·0523	4·2379	5·2891	6·2336	7·0889	7·8671	8·5767	9·2246	9·8156	10·3540	10·8430	11·2854	11·6833	12·0387	12·3531	12·6278	12·8639	13·0623	13·2237	13·3486	13·4375	13·4908	13·5085
49	1·6902	3·0704	4·2653	5·3261	6·2804	7·1456	7·9340	8·6542	9·3127	9·9148	10·4644	10·9650	11·4193	11·8295	12·1974	12·5248	12·8129	13·0628	13·2754	13·4515	13·5916	13·6964	13·7660	13·8008
50	1·6990	3·0881	4·2923	5·3623	6·3261	7·2012	7·9995	8·7299	9·3989	10·0117	10·5723	11·0842	11·5501	11·9721	12·3523	12·6923	12·9933	13·2566	13·4830	13·6733	13·8282	13·9482	14·0336	14·0848
																							$r=25$	14·1018

Notes on Table 5

The entries in the table are referred to in the text of the paper as $F(n, r)$, in which n is the index or exponent of the binomial and r is the ordinal number of the term in the expanded binomial, the first term having the ordinal number, 0. $F(n, 0)=0$. The table includes only values of $F(n, r)$ for values of r up to $n/2$. If r is greater than $n/2$ take value of $F(n, n-r)$ if n be even; that of $F(n, n-r-1)$ if n be odd.

are 28 and 14. For the subtrahend term in the necessary calculation we have log (100!) − log (58!) − log (42!) − log 2, which is equal to 28·1501. Using the last row of entries in Table 5 and finding the value of p for differences from 14 to 22 inclusive it was found that the value of $P(\pm 14)$ was approximately 0·0101. There is thus no reasonable doubt that the value of P tended to decrease as n was increased and that the reality of the difference 14 is proved, this implying that the stimulus B really had an inhibitive effect.

It must be emphasized that the 'significance' of a difference gives no estimate of the size or of the importance of the difference between two chances since the value of P depends so greatly on the number of trials. If for the present argument it is assumed that the tendency to react is the same in all of the fish, it must be assumed that the differences found are estimates of the same real difference, whether 10, 20, 30, 40 or 50 fish are used in the experiment. This difference has been shown to be real beyond all reasonable doubt by the test for significance. The *value* of the difference is best expressed by the ratio of one chance to the other. This ratio or fraction, B/not B, is variously estimated as 2/6, 7/14, 9/17, 10/21 and 14/28, according to the number of trials considered, and the chance of a reaction in the absence of B is fairly accurately estimated as twice the chance of a reaction in the presence of B.

If, in either member of a pair of n parallel trials, there are no occurrences, although a perfectly sound test for significance of an observed difference, d, may be made, yet no valid estimate of the ratio of chances is possible. If there are no occurrences it must not be assumed that there is really no chance of an occurrence. The chance should be assumed to be undetermined. A similar argument holds when one value of r is equal to n, for then there are no non-occurrences in one member. It is of no help to use the difference instead of the ratio and it may be very misleading.

Section 2. *The exact test of Section 1 but applicable to cases in which* $n_1 \neq n_2$

It is best, in experimental work employing parallel trials, when the experimental subjects are under control, to employ equal numbers of subjects in the members of each comparable set, since in that case an effect has equal chances of showing up, whether it be positive or negative. Such equality in numbers cannot always be obtained even in the laboratory and very rarely in field experiments.

The method of testing for the significance of a result then needs modification. As an example let us consider the question whether the inhibitive effect of stimulus *B*, which has been shown to be significant for the short period of 5 min., is a lasting effect. For this we may take the numbers of fish which had not reacted to *A* in 5 min. and compare the numbers which had reacted, with and without *B* respectively, by the end of the experiment which lasted for half an hour for each set of fish. The number of fish, when *B* was not applied, which had not reacted after 5 min. was 22, and of these all had reacted by the end of the experiment. In the comparable set 36 had not reacted after 5 min., but 28 of these had reacted by the end of the experiment. The values of n are not now equal to one another, and we have $n_1=36$, $n_2=22$, $s=8$, $d=8$. The results are conveniently shown in a fourfold table (Table 6). The

Table 6

	Without *B*	With *B*	Total
Reacted in 30 min.	22	28	50
Failed to react	0	8	8
Total trials	22	36	58

result is the same whether we test the difference between reactions or that between non-reactions, but it is the easier to see the number of terms in the distribution if the difference in the same row as the smaller subtotal on the right be tested, for then it is at once seen that there are only nine terms, extending from -8 to $+8$ and including zero difference. In the fourfold table all the subtotals are fixed and the range of the distribution is therefore fixed by the smallest subtotal. The term giving the probability of $+8$ is at one end of the distribution and only this term need be calculated. If we denote the entry in Table 5 for n and r by $F(n, r)$, the logarithm of any value $p(+d)$, when r_1 and r_2 are the numbers giving the difference, $+d$, is obtained by adding together $F(n_1, r_1)$ and $F(n_2, r_2)$ and subtracting

$$\log \frac{(n_1+n_2)!}{(n_1+n_2-s)!\,s!}$$

or $F(n_1+n_2, s)$, if that be within the range of Table 5. In the present case

$$F(n_1, r_1)=F(36, 8), \quad F(n_2, r_2)=F(22, 0)$$

or 7·4809 and 0 respectively. The subtrahend term is equal to $\log 58!-\log 50!-\log 8!$ or 9·2826. Log $p(+8)=\bar{2}\cdot1983$ and $p(+8)=0\cdot01579$. It should be noticed that a particular sign has been given to the difference, d. This is because the random sampling distribution of the difference is not symmetrical when $n_1\neq n_2$, except in the particular case when $n_1+n_2=2s$, and therefore only one end of it, which we have arbitrarily termed the positive end, is being considered.* For the same reason log 2 does not occur in the subtrahend term when calculating $p(d)$. The whole distribution is shown in Table 7.

Table 7

Difference	−8	−6	−4
p	0·0001668	0·003202	0·02452
Difference	−2	0	+2
p	0·09811	0·2248	0·3028
Difference	+4	+6	+8
p	0·2347	0·09579	0·01579

It will be seen that, though the expected difference between occurrences or between non-occurrences is zero when $n_1=n_2$, this is not the case when $n_1\neq n_2$. For instance, in the example given in Table 6, the expected difference between non-occurrences is $\frac{8}{58}\times36-\frac{8}{58}\times22$ or $\frac{8}{58}$ of 14, which is equal to 1·932. This gives the mode or point of greatest frequency in the distribution, and since our observed difference, +8, is nearer the mode than is −8 it will have the greater probability. The difference, +8, is on the shorter tail, the difference, −8, on the longer tail of the distribution. Zero difference is on the longer tail, so, if we wished to find the value of P for zero difference, which would occur with 4 non-occurrences in each member, we should need the value of

$$p(0)+p(-2)+p(-4)+p(-6)+p(-8).$$

The position of the observed difference relative to the mode of the distribution always indicates which tail has to be summed, and in case of doubt on this point the expected difference should be calculated. It is a useful convention to apply the positive sign to the expected difference and to differences on the shorter tail of the distribution when only one pair of sets of trials is being dealt with. If the differences arising from several paired sets are added together for a test of the total difference, the question of similarity or

* See last paragraph of Section 3B.

dissimilarity in sign of the components will of course be decided by the nature of the experiment in question. The simplest way to decide the question of sign of the difference between numbers of events is to apply the positive sign to differences due to excess of events in the greater set of trials. If non-events are being tested for significant difference the sign rule is applied to them.

The question of the significance of our result may be investigated further by the method already explained, namely, that of continuously adding the results of further trials and looking for a tendency in the value of P either to decrease or to become stabilized. The appropriate data are given in Table 8. We have

$$\log P(+2) = 0 + 1{\cdot}4472 - F(12, 2) = \bar{1}{\cdot}6277,$$
$$P(+2) = 0{\cdot}4243,$$
$$\log P(+3) = 2{\cdot}4564 - F(19, 3) = \bar{1}{\cdot}4701,$$
$$P(+3) = 0{\cdot}2952,$$
$$\log P(+5) = 4{\cdot}3086 - F(34, 5) = \bar{2}{\cdot}8641,$$
$$P(+5) = 0{\cdot}07313,$$
$$\log P(+6) = 5{\cdot}7737 - F(49, 6) = \bar{2}{\cdot}6280,$$
$$P(+6) = 0{\cdot}04246,$$

and
$$P(+8) = 0{\cdot}01579.$$

There seems no doubt therefore that P tends to decrease with increase in $n_1 + n_2$, and the result of the experiment is undoubtedly significant. If a fixed criterion be used for one end of a distribution it should have half the value of the corresponding criterion as used for both ends of a symmetrical distribution. Thus in the present case $P = 0{\cdot}025$ would be that generally employed to test for significance.

Another interesting example of the use of the exact method of testing for significance when $n_1 \neq n_2$ is afforded by an experiment carried out by the F.B.A. on Cunsey Beck in 1941. The object of the experiment, which was one of several of the same kind, was to find out whether the application of artificial light had the effect of diverting silver eels from the most direct line of run during their migration to the sea. In the experiment two eel traps, A and B, were arranged side by side and barriers made such that all the water of the beck had to flow through the traps. Thus every eel running was caught by one trap or the other. Light was applied in the case of trap A for two periods of 2 hr. each and in the case of trap B

for one period of 1 hr. and 20 min. It was intended that the light should be applied for equal intervals in the case of both traps, but the experiment was spoilt by circumstances beyond the experimenters' control. To test the result seems at first sight rather hopeless, but it will be interesting to see if it is possible to devise a sound test. The catch of each trap at the end of each period was recorded, the total catch of eels being only 14. It was suspected

Table 8

Of those not reacting to A within 5 min.	Fish sets: 1		1 and 2		1, 2 and 3		1, 2, 3 and 4		All sets	
	B	Not B	B	Not B	B	Not B	B	Not B	B	Not B
No. reacting before 30 min.	6	4	10	6	16	13	24	19	28	22
No. not reacting before 30 min.	2	0	3	0	5	0	6	0	8	0
Total	8	4	13	6	21	13	30	19	36	22
Difference B – not B	2		3		5		6		8	

Table 9

	A lit B not	B lit A not	
Trap A	2	5	7
Trap B	6	1	7
	8	6	14

that the flow of water through trap A was greater than that through trap B, so that the null hypothesis that there were equal chances that an eel would enter trap A or trap B, independently of the lighting, was not reasonable *a priori*. The test of the effect of the lighting cannot be made by way of the binomial $(0{\cdot}5 + 0{\cdot}5)^{14}$. A sound null hypothesis can, however, be framed to cover the case. The data from the experiment were arranged as shown in Table 9. It will be seen that *under the conditions of the experiment* equal numbers of eels entered trap A and trap B. By the null hypothesis the two differing ratios, catch of A : catch of B, are estimates of the mean ratio unity, the arrangement of the lights having had no effect on these ratios. Taking the catches of trap B for testing we

find the difference to be $+5$, a positive difference since that expected is $7/14\times(8-6)$ or $+1$. We must now find the values of $p(+5)$ and $p(+7)$, no greater difference being possible since the marginal totals are all fixed. We have then from Table 5:

$$\log p(+5)=F(8,6)+F(6,1)-F(14,7)=\bar{2}\cdot6898,$$
$$p(+5)=0\cdot04896,$$
$$\log p(+7)=F(8,7)+F(6,0)-F(14,7)=\bar{3}\cdot3675,$$
$$p(+7)=0\cdot002331,$$
$$P(+5)=0\cdot05129.$$

The result is not significant with the criterion, $P=0\cdot025$, but only a very strong effect would give significant results with so small a catch, particularly since the periods of lighting A and B were not equal to one another and therefore the effect of the lights was not given the fullest possible chance of showing if it existed. If the test had shown the result of the experiment to be significant, before interpreting the result as showing a significant effect of the lighting we should have had to make sure that the proportionate flow of water through A and B did not vary during the experiment. If, for instance, the bias towards A was greater when B was lit than when A was lit a difference between the ratios in Table 9 would occur from this cause alone, quite apart from the lighting. The test for significance is here, as always, simply and solely a test of the difference between two ratios. Interpretation of the meaning of the difference, if proved to be real, is not part of the function of a statistical test. To imagine beforehand all reasonable interpretations of a real difference, and to eliminate those causes which are not of interest is a necessity in designing a fruitful experiment.

If, in a fourfold table, the members of either pair of marginal subtotals are equal to one another, the distribution is symmetrical about the expected difference. If the equal subtotals are s and n_1+n_2-s these may be considered as n_1 and n_2, n_1 and n_2 in the fourfold table being considered as s and n_1+n_2-s. The table may then be treated by the method of Section 1 for an exact test or by that of Section 3A for an approximate test. For example, in the last case discussed, $s=2n-s=7$. The distribution is symmetrical about the difference, $+1$, and the value of $P(+5$ or $-3)$ is equal to that of $P(\pm4)$ when $n_1=n_2=7$, $s=6$. The value of $P(\pm4)$ will be found from Table 2 to be $0\cdot10256$. The value of $P(+5)$ is half this, or $0\cdot05128$.

SECTION 3A. *An approximate test applicable to cases in which the exact test of Section 1 is applicable*

The standard deviation of the difference between two numbers distributed binomially may be expected, from analogy with the normal distribution, to be approximately equal to $\sqrt{2}$ times that of either component. These, by the null hypothesis, having been assumed to be samples of n from the binomial distribution, $\left(\frac{2n-s}{2n}+\frac{s}{2n}\right)^{2n}$, and each therefore having a standard deviation equal to $\sqrt{\frac{(2n-s)s}{4n}}$, the standard deviation of the difference between them may be expected to be approximately equal to $\sqrt{\frac{(2n-s)\,s}{2n}}$. This is found to be the case. The distribution is sufficiently near the normal in form for a useful approximate test, based on normal theory, to be applicable, provided that a very simple correction be applied to allow for the fact that the distribution of the difference is discontinuous while the normal distribution is continuous. This correction, which performs exactly the same function as the correction for continuity described in Yates (1), consists merely in subtracting 1 from any difference under test. The test gives results quite sufficiently accurate for a preliminary trial in cases beyond the range of Tables 2, 3 and 4, if, for instance, we wish to know whether further experiments are necessary to show significance of a result. As an example let us consider the results of our 5 min. experiment. It would not be worth while to use the approximate method for the results for the first three paired sets of 10 fish, since the significance of these may be estimated by reference to the tables. On taking in the result from the fourth set we had $n=40$, $s=31$, $d=11$, and the correct value of P was found to be $0\cdot021$. For the approximate test we have

$$\sigma=\sqrt{\frac{(2n-s)\,s}{2n}}=\sqrt{\frac{49\times31}{80}}=4\cdot357.$$

If correction be made for continuity,

$$\frac{x}{\sigma}=\frac{10}{4\cdot357}=2\cdot295.$$

The corresponding value of P is found by doubling the value of $\frac{1}{2}(1+\alpha)$ as given in Table II of Pearson (2) for x/σ equal to $2\cdot295$

and subtracting this from 2. P is found to be equal to 0·022 to two significant figures, an extremely good approximation. Taking in the result from the fifth pair we have

$$\sigma = \sqrt{\frac{58 \times 42}{100}} = 4{\cdot}935,$$

$$\frac{x}{\sigma} = \frac{13}{4{\cdot}935} = 2{\cdot}634,$$

$$P = 0{\cdot}00844.$$

Significance of the result is thus somewhat overestimated by the approximate method since the true value of P is 0·0101. As it is not possible to say beforehand whether significance will be overestimated or underestimated by the approximate method, it is preferable to use the exact method in published papers when time allows. It cannot be considered satisfactory to publish figures which are known to be wrong even if the errors are not likely to be large.

A method of interpolation is described, in a later section, which is useful for obtaining quickly a value of P very near to the true value in certain cases.

Section 3 B. *An approximate test applicable to cases in which the exact test of Section 2 is applicable. Yates's test*

The simplest way to apply Yates's test for those who have got used to the methods of the previous sections is as follows. The example of Table 6 will serve as an illustration.

Denote by n the smallest of the four marginal subtotals. This may be either in the pair on the right or in the pair at the bottom. Whichever pair it occurs in, denote by n' the smaller of the other pair.

Denote by N the sum of either pair of subtotals.

Denote by m the smallest expected value in the body of the table. This is equal to nn'/N.

Denote by p the value of m/n.

Proceed as follows for the example

$$m = \frac{nn'}{N} = \frac{8 \times 22}{58} = 3{\cdot}035.$$

Find the next smallest expected value by subtracting m from n. This equals $8 - 3{\cdot}035 = 4{\cdot}965$.

The expected difference $= 4{\cdot}965 - 3{\cdot}035 = 1{\cdot}930$. This gives the mode of the distribution. The difference, $+8$, is equal to $+6{\cdot}070$ when measured from the mode. It is therefore on the shorter tail.

Calculate x/σ or χ' as follows:

$$\sigma^2 = \frac{50 \times 8 \times 22 \times 36}{58 \times (29)^2} = 6{\cdot}495.$$

That is to say, σ^2 is equal to the product of the marginal subtotals divided by the product of the grand total and the square of half the grand total. The value of x is 5·070, being one less than the value of d measured from the mode of the distribution. This subtraction of unity from the observed difference is the correction for continuity;

$$\frac{x}{\sigma} = \chi' = \frac{5{\cdot}070}{2{\cdot}549} = 1{\cdot}990.$$

From Pearson's Table II the value of $P(\chi')$ is found to be 0·02330. This is much higher than the true value which is 0·01579. Table VIII in Fisher and Yates(3) may be used, however, to find out whether the true value of P, which has been estimated from the normal x/σ, is less than 0·025 or less than 0·005. For this we require the value of p, which is equal to 3·035/8 or 0·3793. Yates's table gives the limiting values of χ' which correspond to the 0·025 point and the 0·005 point of the true distribution, for certain values of m and p. Our value of χ' is 1·990 and is on the shorter tail, $m = 3{\cdot}035$, $p = 0{\cdot}3793$. It will be seen from the table that the value of χ' corresponding to the true 2·5% point lies somewhere between 1·73 and 1·94. Our value, 1·99, is therefore beyond the 0·025 point and the difference is 'significant' if judged by the criterion, $P = 0{\cdot}025$. The values of χ' corresponding to the 0·005 point of the true distribution lie somewhere between 2·18 and 2·50 and therefore the true value of P is between 0·025 and 0·005.

Yates's method saves a great deal of time in cases in which many values of p have to be calculated to find the true value of P. In cases within the range of Table 5, however, it is probably just as expeditious and much more satisfactory to calculate the true value, particularly since it may happen that the value of χ' lies between the limits given in Yates's table. In that case the exact value will

have to be calculated. Instead of applying Yates's method in cases beyond the range of Table 5 it is preferable to use the method of Section 11 since in most cases this will settle the question of significance.

SECTION 4A. *An approximate test applicable to cases in which the exact test of Section 1 is applicable, but in which s is very small in comparison with n*

If n, the number of paired trials, is at least 30 times as great as s, the distribution of the difference, d, is expressed closely enough by the binomial $(0{\cdot}5+0{\cdot}5)^s$. Thus, if the number of reactions to stimulus A in the presence of B be 2, when 300 fish are used in the experiment, the number of reactions in the absence of B being 8, the chance of the difference, $+6$, is given approximately by the binomial term ${}_{10}C_2\,(0{\cdot}5)^{10}$. The chance of the difference, ± 6, is twice this, that is to say, the index is $s-1$ or 9 instead of 10. To obtain the value of $P(\pm 6)$ the terms nearer the tails of the distribution must be added to the term for $r=2$. The values of the terms to be summed may be obtained quickly from Table 5 when s lies between 2 and 50 inclusive.* In the present case the antilogs of the entries for $n=10$, $r=2$, 1, 0, are summed and the sum divided by 2^9 to give $P(\pm 6)$. Its value is found to be 0·1093. The method is approximate and only gives exact results when $n=\infty$. The true value of $P(\pm 6)$, when $n=300$, $s=10$, is equal to 0·1064. The true value of $P(\pm 8)$ is 0·020466, the binomial approximation 0·02148 and the corresponding values of $P(\pm 10)$ are 0·0018096 and 0·00195 respectively. The binomial approximations rapidly approach the true values as n/s is increased, but even when this is only equal to 30 the binomial approximations are nearer the true values than are those given by the methods of Section 3A.

SECTION 4B. *The approximate test of Section 4A but applicable to cases in which* $n_1 \neq n_2$

If the numbers of subjects in a set of parallel trials are both very large in comparison with s but not equal to one another the distribution of the difference, d, is not that of the binomial $(0{\cdot}5+0{\cdot}5)^s$ but that of the binomial $(q+p)^s$, in which

$$q=\frac{n_1}{n_1+n_2}, \quad p=\frac{n_2}{n_1+n_2}.$$

* For treatment of cases beyond the range of Table 5, see Sections 10 and 11.

As an example we may take the numbers of fish shown in Table 6 but shall assume that $n_1=3600$, $n_2=2200$, instead of 36 and 22 respectively. The binomial is therefore

$$\left(\frac{36}{58}+\frac{22}{58}\right)^8 \quad \text{or} \quad (0{\cdot}6207+0{\cdot}3793)^8,$$

and $p(+8)$, which is the same as $P(+8)$, since it is the end-term, is equal to $(0{\cdot}6207)^8$ or 0·02204. The likelihood that the difference, $+8$, would have arisen by chance is thus greater than it was in the actual experiment. It is interesting to see that so great a difference as $+8$ would be distinctly unlikely to arise by chance however great be the numbers of fish in the trials.

In calculating the binomial terms the coefficients are obtained from Table 5 but the fractional parts must be calculated. This is a simple process since the fractional part of each term is obtained from the next by multiplying by p/q or q/p according to the direction in which the terms are taken. If s is a large number and many terms are required the calculation may take a long time. This case is not so likely to arise in experimental work as in observational. When it does arise, the method described in Section 11 should be tried first as it may be proved to be unnecessary to calculate the binomial terms.

When using the methods of Sections 4A and 4B, unless the values of n are known it is quite necessary to provide some means of determining that these are very large and equal to one another or, alternatively, for determining the ratio of one to the other, means independent of the experiment in which the methods of testing are used. It is easy to see that a false assumption of the equality of the values of n or a false estimate of the ratio between them may entirely vitiate the results of a test.

SECTION 5. *On a method of finding the value of* $P(\pm d)$ *by interpolation in cases in which* $n_1=n_2=n$

For any given value of s the value of $P(\pm d)$ varies very smoothly with change in $1/n$. If for any difference, d, we take $1/n$ as abscissa and $P(\pm d)$ as ordinate for values of $1/n$ for which the values of $P(\pm d)$ have been tabulated, then the value of $P(\pm d)$ for any intermediate value of $1/n$ may be obtained very quickly and with very considerable accuracy by interpolation, graphical or numerical. The end-point, when $n=\infty$, may be found quickly by way of the

binomial $(0{\cdot}5+0{\cdot}5)^s$. As an example let us take the case dealt with in Section 4A in which n was equal to 300, s equal to 10, d to 6. We obtain the values of $P(\pm 6)$ when $n=20$, $n=30$ from the tables, these values being respectively 0·06484 and 0·07973, while the end-value, that for $n=\infty$, is equal to 0·10937, the corresponding values of $1/n$ being respectively 0·05, 0·0$\dot{3}$ and 0. The ordinates will be seen to lie so near to a straight line that linear interpolation between $1/n=0{\cdot}0\dot{3}$ and $1/n=0$ will meet the case. Since for the interpoland $1/n=0{\cdot}00\dot{3}$, the required value of $P(\pm d)$ is $0{\cdot}10937-0{\cdot}1(0{\cdot}10937-0{\cdot}07973)$ which is equal to 0·10641, a result correct to the fourth significant figure. The value given by linear interpolation for $P(\pm 10)$ is 0·001835 against the true figure 0·0018096. If we use the tabulated values of $P(\pm 10)$ for $n=15$ and $n=30$ instead of those for $n=20$ and $n=30$, the values of $1/n$ are then equidistant and ordinary 3-point interpolation by finite differences may be used. That process gives 0·001810 for the value of $P(\pm 10)$ which is correct to the fourth significant figure.

SECTION 6. *On the calculation of the exact value of $P(d)$ in independent parallel trials in which either n_1 or n_2 is greater than* 50

Consider the fourfold table (Table 10). Here the numbers of experimental subjects or paired trials are unequal and greater than 50, and are thus beyond the range of Table 5. In such a case the best procedure is as follows.

Test for significance the difference between the pair of numbers, in the same row or column of the table, which have the smallest total. These are 20 and 30.

Calculate the smallest expected number and the expected difference.

The smallest expected number is equal to the product of the two smallest subtotals divided by the grand total. Thus

$$m=50\times\tfrac{60}{150}=20.$$

Find the other member of the pair by subtracting 20 from 50, the total of the pair, giving 30 as the other member. These are shown in brackets in Table 10. The expected difference is therefore +10. The observed difference is −10.

For the value of $P(-10)$ we require the values of $p(-10)$, $p(-12)$ and so on.

The most convenient form of the equation for calculating p in cases beyond the range of Table 5 is that given by Yates (1). This gives the following rule.

The value of p for any set of entries in a fourfold table is equal to the product of the factorials of the marginal subtotals divided by the product of the factorials of the cell frequencies and of the grand total. Thus

$$\begin{aligned}\log p(-10)&=\log(50!)+\log(100!)+\log(60!)+\log(90!)-\log(30!)\\&\quad-\log(20!)-\log(30!)-\log(70!)-\log(150!)\\&=\bar{4}{\cdot}4764.\end{aligned}$$

Table 10

20, (30)	30, (20)	50
70	30	100
90	60	150

The logarithm of the probability of the next difference, going towards the tail of the distribution, is calculated from that already found by adding to it the logarithms of the lesser pair of numbers on a diagonal of the table corresponding to the first difference and subtracting from it the logarithms of one more than each number on the other diagonal. Thus

$$\begin{aligned}\log p(-12)&=\bar{4}{\cdot}4764+\log(20)+\log(30)-\log(31)-\log(71)\\&=\bar{5}{\cdot}9118,\\\log p(-14)&=\bar{5}{\cdot}9118+\log(19)+\log(29)-\log(32)-\log(72)\\&=\bar{5}{\cdot}2906.\end{aligned}$$

The multipliers each decrease by 1 and the divisors each increase by 1 each time. The same rules hold good for any part of the distribution, except that when going towards the centre the multipliers are the larger pair of numbers. By 'lesser pair' is meant the pair having the smaller product.

SECTION 7. *Limitations of the tests discussed in previous sections*

The test, by the foregoing methods, of a total difference made up of several component differences, is justifiable only if every fish in the experiment has the same chance of reacting to stimulus A. This assumption is necessary not only in the case of all the fish in each set but also for all the sets. Furthermore, the summing up of

the various differences implies that, should their sum prove to be significant, the effect producing each difference is throughout of the same kind. If this be not so the result of adding the differences together, each with its particular sign, is meaningless. Heterogeneity among the experimental subjects may render the test of a total difference unsound, since it is implied in the null hypothesis that all fish have the same chance of reacting to stimulus A and therefore that hypothesis is unreasonable *a priori* if it is known that the chance varies. Since all the subjects in any set will almost certainly have been treated as far as possible in exactly the same way and will have been chosen with an eye to homogeneity it is probably reasonable to assume that there is no significant heterogeneity among the subjects in any set. If there is doubt whether there is homogeneity between the sets it is a simple matter to test whether it is reasonable to assume this. The test is fully described in Section 21 of Fisher (4), and it is therefore unnecessary to give an account of it here. In applying the test s in each set corresponds to a in Fisher (p. 90), $2n-s$ to a', total s to n, total $2n-s$ to n'. If heterogeneity be found it is preferable to use the methods of the next section.

The methods of Sections 8 and 9 do not give reliable results except when, in the case of independent parallel trials, $n_1=n_2$ in each component distribution and when, in interdependent parallel trials, each component is of the form, $(0{\cdot}5+0{\cdot}5)^n$.

Section 8. *On testing for significance the result of independent or interdependent parallel trials with sets of subjects heterogeneous as to their chance of affording an event*

If the members of a set of differences be additive in nature owing to their investigated cause being the same throughout, but if the sets of experimental subjects be heterogeneous as to the variate in which differences are measured for statistical testing, the null hypothesis takes a form rather different from that applying to homogeneous variation in that variate. We have to consider a series of independent hypotheses each having the form: The difference is really zero and any difference arising from random sampling is equally likely to be positive or negative; but it is not implied that all the component differences have the same random sampling distribution. To make *one* comprehensive test which is applicable to such a case each observed difference must be given its sign and each must be graded according to the probability that

a difference at least as great will arise by chance if the null hypothesis be true. The only practicable way of grading them according to the value of P in each case is to put them on the 'normal scale', that is to say, to allot to each of them the value of x/σ, in the normal distribution, which gives the same value of P. The values of x/σ, each with its appropriate sign, may then be summed for a composite test of the truth of all the null hypotheses. Though the distribution of this sum is not exactly normal in form yet it rapidly approaches that form as the number of values in the sum is increased. To correct for continuity the total X/σ is multiplied by $(D-1)/D$, where D is the total difference.

According to the most accurate method the value of X/σ for an aggregate difference is calculated from the true values of P and p for each of the component differences. What is found for each is the mean normal equivalent abscissa, x/σ, for the probability interval, p, corresponding to each difference. It may be shown that, for any difference, d, with particular sign,

$$\text{Mean } \frac{x}{\sigma}=\frac{z_d-z_{(d+2)}}{p(d)},$$

where z_d and $z_{(d+2)}$ are respectively the normal ordinates at the inner and outer ends of the probability interval $p(d)$ in the normal distribution. The entries in Tables 2, 3 and 4 are the values of $P(\pm d)$. For the example used in the previous method the procedure is as follows, using the values of P given in Table 2 for $n=10$:

(1) $$d=+4, \quad s=8,$$

$1-\frac{1}{2}P(\pm 4)=0{\cdot}9151=\frac{1}{2}(1+\alpha)$ in Table II of Pearson (2).

Taking the nearest tabulated value of $\frac{1}{2}(1+\alpha)$ in that table, the corresponding value of z is found to be 0·1561:

$$1-\tfrac{1}{2}P(\pm 6)=0{\cdot}9901, \quad z_6=0{\cdot}02643.$$

The value of $p(d)$ is preferably taken as the difference between the *tabulated* values of $\frac{1}{2}(1+\alpha)$ corresponding to the values of z taken from Pearson's Table II. This process renders it unnecessary to interpolate between tabulated values of z. We have therefore

$$\text{Mean } \frac{x}{\sigma}=\frac{0{\cdot}1561-0{\cdot}02643}{0{\cdot}07544}=1{\cdot}720.$$

(2) $1-\frac{1}{2}P(\pm 3)=0{\cdot}8250, \quad z_3=0{\cdot}2589,$

$1-\frac{1}{2}P(\pm 5)=0{\cdot}9713, \quad z_5=0{\cdot}06562.$

$$\text{Mean } \frac{x}{\sigma}=\frac{0{\cdot}1933}{0{\cdot}1475}=1{\cdot}311.$$

(3) $1-\frac{1}{2}P(\pm 1)=0{\cdot}5, \quad z_1=0{\cdot}3989,$

$1-\frac{1}{2}P(\pm 3)=0{\cdot}8483, \quad z_3=0{\cdot}2347.$

$$\text{Mean } \frac{x}{\sigma}=\frac{0{\cdot}1642}{0{\cdot}3485}=0{\cdot}4711.$$

(4) $$\text{Mean } \frac{x}{\sigma}=1{\cdot}436.$$

(5) $$\text{Mean } \frac{x}{\sigma}=1{\cdot}267.$$

The sum of the mean values of x/σ is equal to 6·2051. The standard deviation of the sum is equal to $\sqrt{5}$, since there are five component differences. Thus for the sum, 14, corrected for continuity,

$$\frac{X}{\sigma}=\frac{6{\cdot}2051}{2{\cdot}236}\times\frac{13}{14}=2{\cdot}576.$$

From Pearson's Table II the value of $P(\pm 2{\cdot}576)$ is found to be 0·009996. The fact that this result agrees very nearly with the value, 0·0101, found by the exact method of Section 1, indicates that the method of the mean normal equivalent is likely to give reliable results, since there is little doubt that the experimental subjects in this case were homogeneous in their response to stimuli and that therefore the application of the exact method to the total difference was justified.

The procedure last described is preferable when tabulated values of $P(d)$ are available. In other cases the approximate method of Section 3A may be used to obtain the necessary values of mean χ. Thus, if Table 2 had not been available, for the first set of 10 fish in the experiment of Table 1 we should have proceeded as follows:

$$\frac{3}{\sigma}=1{\cdot}37, \quad z_4=0{\cdot}1561,$$

$$\frac{5}{\sigma}=2{\cdot}28, \quad z_6=0{\cdot}02965.$$

The value of $p(4)$ is the difference between the values of $\frac{1}{2}(1+\alpha)$ corresponding to the x-values of 1·37 and 2·28, respectively, in Pearson's Table II. We have, therefore,

$$\text{Mean } \chi=\frac{0{\cdot}1561-0{\cdot}02965}{0{\cdot}07404}=1{\cdot}710.$$

The total χ for the five component experiments was found by this method to be equal to 2·565 when corrected for continuity by multiplication by $\frac{13}{14}$. The corresponding value of P is 0·01032, this result agreeing very well with that given by the exact test of the total difference.

In applying the methods of this section, though it is necessary that $n_1=n_2=n$ in each component, it is not necessary that n be the same in all the components: n may be any number. The case in which $n=\infty$, the distribution then being binomial, is discussed later in this section. Mean values of χ obtained from various experiments in which n varies from one to another and may be infinite in some of them may justifiably be summed for a composite test for significance provided that the sign of the difference in each component is relevant and that a definite meaning may be attached to the total difference. This would be the case, for instance, if the investigated cause of the differences were the same throughout.

The following is a hypothetical case in which the method of the mean normal equivalent would be very useful. The perch trapping in Windermere may be expected to change the length frequency distribution of the remaining stock since the method of fishing is selective of the smaller fish. To measure large samples of fish from every fishing beat is out of the question with the present small staff of measurers. On measuring a few fish, however, it soon becomes clear that the range of length is about 10–20 cm. One way then to spot a change in the proportion of smaller to larger fish over the whole area fished would be to take small random samples, say 20 fish of each sex, from each fishing beat and, for each sample, to find the number of fish over some median length, say 15 cm. The aggregate of the differences between these proportions in one year and in the following year could then be tested for significance by allotting to each difference the corresponding value of x/σ and testing the sum for significance. Of course the ratios compared in the two years must apply to the same beat in each year and also

approximately to the same date. Changes in sex ratio could be investigated in a similar way. In this kind of work it would save a great deal of time if the number of fish in each sample were one of the values of n included in our tables.

The methods of this section are also applicable to testing for significance a series of differences between numbers of occurrences in *interdependent* parallel trials in cases in which the chance of an occurrence varies widely from set to set of trials. For instance, in a series of experiments to find out whether a line of lights has a diverting effect on migrating eels, experiments in which all migrating eels are caught in one or other of two traps, A and B, it may be found that the proportion of eels caught in trap A varies widely. It would not, in this case, be justifiable to add together all the numbers of eels caught in trap A and then to test the ratio between that total and the total caught in trap B by way of the binomial $(0{\cdot}5+0{\cdot}5)^n$. It would be preferable, in each component experiment, to find the mean value of x/σ or χ corresponding to each difference between A and B, and to test the total χ for significance. It is just as simple a matter to find the required values of χ in a case of this kind as it was in our previous example. Let us suppose, for instance, that in four experiments the numbers of eels in trap A and in trap B were, respectively, 1, 5; 4, 9; 3, 2; 5, 15. By the method explained in Section 10 the values of $P(\pm 4)$, $P(\pm 5)$, $P(\pm 1)$, and $P(\pm 10)$ are found to be, respectively, 0·2188, 0·2669, 1 and 0·04139, while the corresponding values of $P(d+2)$ are, respectively, 0·03125, 0·09018, 0·375 and 0·01179. The value of $p(+4)$ is equal to $\frac{1}{2}(0{\cdot}2188-0{\cdot}03125)$ or 0·09377. The required values of p will have been found, however, during the process of calculating those of P. For the first difference we have, therefore,

$$1-\tfrac{1}{2}P(\pm 4)=0{\cdot}8906=\tfrac{1}{2}(1+\alpha), \quad z_4=0{\cdot}1872,$$

$$1-\tfrac{1}{2}P(\pm 6)=0{\cdot}9844, \quad z_6=0{\cdot}03955.$$

$$\text{Mean } \chi=\frac{0{\cdot}1872-0{\cdot}03955}{0{\cdot}09357}=1{\cdot}579.$$

The values of mean χ are summed, multiplied by $\frac{17}{18}$ to correct for continuity, and the result divided by $\sqrt{4}$ or 2, the value of P for total χ being obtained from Pearson's Table II as in our previous example.

The method of mean χ is also applicable to testing for significance of differences between numbers of occurrences in paired

equal samples if these samples form a very small proportion of the sampled field. In this case, as is explained in Section 14(3), the distribution of the difference between the numbers in any pair of samples is expressed by the binomial $(0{\cdot}5+0{\cdot}5)^s$, where s is the total number of occurrences in the two samples. For example, let us suppose that a farmer who is also a statistician wishes to find out whether infestation of a particular piece of land by wireworms has increased or decreased significantly from one year to the next. In the first year he takes a bucketful of soil from each of ten different sites on the land and counts the wireworms in each bucketful. On some sites, however, he finds that he has to take two bucketfuls to obtain any wireworms. In the next year he repeats the process on the same sites, taking of course the same number of bucketfuls of soil as he did on the same site the year before. The test for significance of the change in infestation is the same as in our previous example, in each component experiment the difference between numbers of wireworms being considered as a term of the binomial $(0{\cdot}5+0{\cdot}5)^s$, in which s is the total number counted at the same site in the two years.

The test just explained is applicable in a very wide field. Here are some examples: (1) comparing the catch of perch traps in different years over large areas; (2) comparing the catch of planktonic organisms from a large number of different places or periods; (3) uniformity trials, pairs being chosen at random from a large number of experimental takings extending over the period or the area for which the question of uniformity in distribution of organisms or other subjects is of interest.

Section 9. *On testing for significance a set of heterogeneous differences arising from parallel trials*

Sometimes it may be of interest to test for significance the aggregate of a series of differences of which the sign is not taken into account. For example, an experiment similar to that discussed in Section 1 might be made in which, however, the fish in each set of 10 were of species different from those in the other sets. Here the question of interest might be whether the application of stimulus B caused a change in the reaction of the fish to A, whether or not the change was in the same sense in each case. The question would be, therefore, whether the aggregate of such differences as were observed, without regard to sign, in the proportion reacting to A,

were such as would be likely to arise by chance on the assumption that in each component experiment the effect of B was nil. The appropriate test here is a form of the well-known χ^2-test, the normal equivalent, x/σ or χ, for each difference being used in its squared form instead of in its first power. Each difference is therefore replaced by the equivalent mean χ^2. This is not the same thing as the square of the mean χ. Symbolically, mean $\chi^2 \neq (\text{mean } \chi)^2$. It may be shown that, for any difference, $\pm d$,

$$\text{Mean } \chi^2 = 1 + \frac{(zx)_d - (zx)_{d+2}}{p(d)},$$

in which z and $p(d)$ have the same meaning as in Section 8 and x is the value of x corresponding to that of z in Pearson's Table II. We shall take the example of Section 1 to illustrate the necessary procedure in calculating total χ^2.

(1) $$(zx)_4 = 0{\cdot}1561 \times 1{\cdot}37 = 0{\cdot}2138,$$
$$(zx)_6 = 0{\cdot}02643 \times 2{\cdot}33 = 0{\cdot}06159.$$
$$\text{Mean } \chi^2 = 1 + \frac{0{\cdot}2138 - 0{\cdot}06159}{0{\cdot}07544} = 3{\cdot}018.$$

(2) $$(zx)_3 = 0{\cdot}2589 \times 0{\cdot}93 = 0{\cdot}2407,$$
$$(zx)_5 = 0{\cdot}06562 \times 1{\cdot}90 = 0{\cdot}1247.$$
$$\text{Mean } \chi^2 = 1 + \frac{0{\cdot}2407 - 0{\cdot}1247}{0{\cdot}1475} = 1{\cdot}787.$$

(3) $$(zx)_1 = 0{\cdot}3989 \times 0 = 0,$$
$$(zx)_3 = 0{\cdot}2347 \times 1{\cdot}03 = 0{\cdot}2417.$$
$$\text{Mean } \chi^2 = 1 - \frac{0{\cdot}2417}{0{\cdot}3485} = 0{\cdot}3066.$$

(4) $$\text{Mean } \chi^2 = 2{\cdot}133.$$

(5) $$\text{Mean } \chi^2 = 1{\cdot}666.$$

The sum of these values of χ^2 is 8·9106, which, multiplied by the square of $\frac{13}{14}$ to correct for continuity, becomes 7·683. By interpolation in Table IV of Fisher and Yates (3), according to the method described in Section 21·1 of Fisher (4), the value of P, for

5 degrees of freedom, is found to be approximately 0·1738. Thus, the aggregate of the observed differences would not be considered significant were it not for consistency in sign though they are as a whole rather larger than would be expected often by chance.

If a difference of zero is one of the components in an aggregate to be tested for significance it should be noted that, though the mean χ is zero in that case, mean χ^2 is not zero. Thus, if in (1) of our example the difference had been zero we should have had

$$1 - \tfrac{1}{2}P(0) = 0{\cdot}5, \quad (zx)_0 = 0,$$
$$1 - \tfrac{1}{2}P(2) = 0{\cdot}6750, \quad (zx)_2 = 0{\cdot}3605 \times 0{\cdot}45 = 0{\cdot}1622,$$
$$p(0) = 0{\cdot}1736.$$
$$\text{Mean } \chi^2 = 1 - \frac{0{\cdot}1622}{0{\cdot}1736} = 0{\cdot}0656.$$

The value of $p(0)$ is that applying to half the distribution.

It should be mentioned that the method of testing for significance a combination of probabilities, which is described in Fisher (4, Section 21·1), is not applicable when these probabilities apply to discontinuous distributions such as that of the differences discussed here. To add together values of χ^2, each calculated as applying to the inner end of the corresponding probability interval, would result in an aggregate from which significance of the combination tested would be greatly underestimated.

The value of P found by the exact method of Section 1 for the total difference, +14, when consistency in sign was taken into account, was found to be 0·0101. The value of $P(\chi^2)$ is 0·1738 and the value of P for five heads or five tails in a toss of 5 coins is 0·0625. Thus the value of P for a set of deviations at least as great as those observed, in either direction from the expected difference, together with consistency in direction in all five cases, is approximately $0{\cdot}1738 \times 0{\cdot}0625$ or 0·01086. This is, as expected, in reasonably close agreement with the result of the direct combined test. The combination of the χ^2-test with the direct method is valuable as a test of consistency in the component differences. If, as in the present case, the combined test shows greater significance than does the χ^2-test, consistency in the components is indicated. If, on the other hand, the χ^2-test had given a smaller value of P than that given either by the exact method or by the test of the algebraic sum of the component values of χ, it would have been shown that

the sum of the differences was an aggregate of inconsistent differences. The sets of experimental subjects would have varied more than would be expected by chance in their response to the causes of the differences.

An interesting example of the application of the methods of the present section is a test of consistency in the catches of floating fish eggs made by parallel vertical hauls with a plankton net at different places and times. In a short series of such pairs of hauls the numbers of plaice eggs caught were, respectively, 2, 4; 5, 1; 6, 7; and 3, 6. The total number of plaice eggs caught in the first-made hauls of each pair was 16 and in the second hauls 18, so that it is clear that in this case there was no significant tendency for the catch of first hauls to be greater than that of second hauls or vice versa. A test for significance of total χ is not called for. It is of interest, however, to find out whether there is, on the whole, significant discrepancy between the catches of first and those of second hauls. The χ^2-test is applicable. For the first pair we have

$$1-\tfrac{1}{2}P(\pm 2)=0{\cdot}6563, \quad (zx)_2=0{\cdot}3683\times 0{\cdot}40=0{\cdot}1473,$$

$$1-\tfrac{1}{2}P(\pm 4)=0{\cdot}8906, \quad (zx)_4=0{\cdot}1872\times 1{\cdot}23=0{\cdot}2302,$$

$$p(+2)=0{\cdot}8906-0{\cdot}6563=0{\cdot}2343.$$

$$\text{Mean } \chi^2=1+\frac{0{\cdot}1473-0{\cdot}2302}{0{\cdot}2343}=0{\cdot}6462.$$

For the other three pairs the values of mean χ^2 are, respectively, 2·548, 0·0997 and 0·9707. The sum of these values is 4·2646, which, multiplied by the square of 9/10, becomes 3·453. From Table IV of Fisher and Yates(3) the value of P, for 4 degrees of freedom, is found to lie between 0·3 and 0·5. Thus no significant discrepancy is shown between the catches of first and second hauls of the net at each place. The data were, however, taken from a long series of similar data, and it must not be concluded that such discrepancy would not have been shown had all the data been included in the test. If, after examination of all the relevant data, it were to be found that there was no significant discrepancy between the catches of first and second hauls it would be shown to be reasonable to assume not only that the catching power of the net was uniform but also that variation in the quantity of the eggs under a given surface of water between the first and second hauls of each pair was only such as could be ascribed to pure chance.

Only if such level working of the net and uniform distribution of the eggs can be assumed is it justifiable to add together the catches of the two hauls at each station and to ascribe to each total the standard error, $\sqrt{\text{total}}$, this standard error being that pertaining to a number distributed according to the Poisson Series.

The question whether the component data of a set are homogeneous or not affects the choice of a method to be used in estimating the aggregate or average difference arising in parallel trials. In our example of Section 1 the total numbers of reactions to A, with and without B respectively, were added together to obtain the average ratio. Had the values of the ratio $s/2n$ varied widely from one component experiment to another, however, that procedure would not have been justified for the reason that any component having a very high value of s would have been unduly weighted. In cases of that kind the average effect should be estimated by taking the average of the ratios for the components. These ratios, in the case of Table 1, are 1/3, 5/8, 2/3, 1/4 and 4/7, their average being 0·4893, the ratio given by the totals being exactly 0·5. The two results are in close agreement. If this had not been the case the value of the average of the ratios should have been taken as the estimate of the average effect. Consideration of our hypothetical case in which change in degree of infestation of land by wireworms was discussed will show that the choice of the correct method of estimating the average ratio may be important. It can hardly be expected that the samples taken over a wide area will show anything like constant infestation. The same considerations apply to our example taken from research on distribution of fish eggs. The numbers of these vary widely from place to place. In these two cases it is a simple matter to test whether there is significant variation in numbers from place to place. The counts in each set are tested by the χ^2 method to see whether they could all have arisen from the same Poisson distribution. The mean, $\bar{x}$, of all the counts in the set is found, x being any count, and χ^2 is calculated as follows:

$$\chi^2=\frac{S(x-\bar{x})^2}{\bar{x}},$$

in which S stands for 'the sum of'. $P(\chi^2)$ is found from Table IV of Fisher and Yates(3), the number of degrees of freedom being one less than the number of counts in the set. $P<0{\cdot}05$ may be used as the criterion of significance.

In cases in which n, in either independent or interdependent parallel trials, is outside the range of our tables, it is quite safe to use the approximate methods of Section 3A or Section 10 when calculating values of mean χ^2.

Section 10. *The binomial test, the test applicable to interdependent parallel trials*

Cases constantly arise both in experimental and in field work in which the appropriate test for significance is by way of the symmetrical binomial $(0{\cdot}5+0{\cdot}5)^n$. For example, in an experiment carried out by the F.B.A. on Cunsey Beck, two eel traps were arranged one above the other, the object of the experiment being to determine whether migrating eels tend to move more in the lower than in the upper layers of the water. All migrating eels were caught in one or other of the two traps. In the upper trap 14 eels, in the lower trap 28 eels were caught. Is this result in accordance with the hypothesis that each eel is equally likely to be caught by either trap? Here $n=42$ and there are 28 'heads' and 14 'tails', or, since it would be incorrect to assume *a priori* that an excess of heads is more significant than an excess of tails, there are 28 heads (or tails) and 14 tails (or heads). For an exact test we must sum the terms of the binomial $(0{\cdot}5+0{\cdot}5)^{42}$ beyond and including that corresponding to 28 heads, 14 tails, and double the result. Since $n<51$ the logarithms of the required terms in order are obtained by subtracting from $F(42, r)$ in Table 5 the logarithm of 2^{41} which is equal to 12·3422. The value of P is found to be equal to 0·04356. Significance is shown, with the criterion $P=0{\cdot}05$.

Though it may be considered the more satisfactory to obtain the true value of P, yet the symmetrical binomial is so very near to the normal in form when $n=50$ or more that an approximate test founded on normal theory then fulfils all practical requirements. The procedure when dealing with our example is as follows:

$$m=\tfrac{1}{2}n=21,$$
$$\sigma=\tfrac{1}{2}\sqrt{n}=3{\cdot}241,$$
$$x=\text{number of heads}-m=7.$$

Yates's correction for continuity consists in subtracting 0·5 from x. Thus

$$x/\sigma=6{\cdot}5/3{\cdot}241=2{\cdot}005,$$
$$P=0{\cdot}045.$$

The error is thus very small.

Cases for which the appropriate test is by way of the asymmetrical binomial do not arise so frequently as those in which the symmetrical binomial gives the correct test. When they do arise, however, the necessary terms of the exact distribution will have to be calculated if an exact test be required, since the normal approximation is strictly applicable only to symmetrical distributions and there is no simple way of determining whether a distribution is nearly enough symmetrical for the approximate method to be safe. A simple device which is explained in Section 11 will, however, be found usually to render the laborious business of calculating a large number of binomial terms unnecessary. The necessary calculations for finding the exact value of P will be explained by way of an example.

Let us suppose that two eel traps, A and B, through which all the water of a river has to flow, are arranged side by side and that over a long period of test trap A has caught 80% of the total catch of the two, the component results from which the percentage total has been derived having shown reasonable consistency when tested by the χ^2 method. It is now wished to make an experiment to see whether a series of lights arranged in front of one of the traps has the effect of inhibiting the entry of eels into it. Trap A is chosen for lighting, since this will show up the effect in the greater degree. It is found, say, that when A is lit, of a total catch of 100 eels 45 entered trap A. We require to know whether so great a deviation from the expected ratio 80 : 20 is likely to have arisen by chance. The general term of the binomial $(q+p)^n$ is

$$\frac{n!}{(n-r)!\,r!}\,q^r p^{n-r},$$

in which p is the chance of an eel's entering trap A. The term for the chance of 45 entries is therefore equal to

$$\frac{100!}{45!\,55!}\,(0{\cdot}2)^{55}\,(0{\cdot}8)^{45}.$$

Since many terms may have to be calculated and each may be calculated from that previous it is necessary to ensure that the first is calculated very exactly. Thus it is necessary to use logarithms to at least seven figures in the calculation.

We have $$\log \frac{100!}{45!\,55!} = +28{\cdot}7885111$$
$$55 \log 0{\cdot}2 = -38{\cdot}4450000$$
$$45 \log 0{\cdot}8 = -\ 4{\cdot}3609500$$
$$-14{\cdot}0174389 = \bar{1}5{\cdot}9825611,$$
$$p(55, 45) = 0{\cdot}0(14)9924.$$

For obtaining the necessary multiplier for calculating the next term from that already calculated we have

$$\frac{100!}{44!\,56!} \div \frac{100!}{45!\,55!} = \frac{45!\,55!}{44!\,56!} = \frac{45}{56} = 0{\cdot}8035,$$

and, for the fractional part,

$$\frac{(0{\cdot}2)^{56}\,(0{\cdot}8)^{44}}{(0{\cdot}2)^{55}\,(0{\cdot}8)^{45}} = \frac{0{\cdot}2}{0{\cdot}8} = 0{\cdot}25.$$

The multiplier is therefore 0·2009.

For the next succeeding term we have the multiplier $\frac{44}{57} \times 0{\cdot}25$, and so on. The multipliers decrease each time and therefore the terms are decreasing at an increasing rate.

SECTION 11. *On a method of approximating to the value of P in parallel trials, both independent and interdependent, by way of the geometrical progression*

We can always find a value known to be greater than the sum of any given series of terms in a binomial distribution, and if this value is less than the value of P which is considered as the criterion of significance we may safely label the result to which the value applies as significant even if we do not carry out the sometimes laborious task of calculating the true value of P. This value, greater than P, may be calculated by assuming that the coefficient decreases from term to term according to the ratio shown by the first multiplier, which, in the example discussed in Section 10, was 0·2009. Since the rate of decrease of the coefficient really increases from term to term the sum of the terms calculated in this way will always be greater than the true value of P. According to the approximate method the terms are assumed to form a geometrical progression and their sum to infinity is given by

$$S = a \times \frac{1}{1-r},$$

where a is the first term and r the common ratio which, in our example, is equal to 0·2009. We know, therefore, that the true value of P is less than 0·0(14)9924 × 1/0·7991 or 0·0(13)1242 and greater than 0·0(14)9924(1 + 0·2009) or 0·0(13)1192, since that is the value of the first two terms only. Practically speaking this approximation is quite good enough and, as will be shown, the approximate value given by the geometrical progression is only very slightly greater than the true value of P.

It may happen that the observed term is in such a position in the distribution that the central term thereof is included in those to be summed. The central term has the greatest coefficient. Even here, however, though the coefficients increase in value up to the central term, their rate of increase decreases from term to term so that the product of the coefficient and the fractional part decreases at an increasing rate from term to term. Thus supposing that 60 eels were caught by trap A, we know that the true value of P is less than

$$\frac{100!}{40!\,60!}\,(0{\cdot}2)^{40}\,(0{\cdot}8)^{60} \times \frac{1}{1 - 0{\cdot}25 \times \frac{60}{41}},$$

or 0·000002313 × 1·577, which is equal to 0·000003649.

The rule for finding the coefficient factor (cf) of r is as follows. Let $p(l, k)$ be the observed term, in which l is the number of occurrences, k the number of non-occurrences.

(A) The required sum of terms does not include the central term:

$$cf = l/k + 1, \quad \text{if} \quad l < k,$$
$$cf = k/l + 1, \quad \text{if} \quad k < l.$$

(B) The required sum includes the central term:

$$cf = l/k + 1, \quad \text{if} \quad l \geqslant k,$$
$$cf = k/l + 1, \quad \text{if} \quad k \geqslant l.$$

For the fractional factor of r, if, in the binomial $(q+p)^n$, p is the chance of an occurrence, the left-hand tail of the distribution representing o occurrences, and the left-hand tail is to be summed, starting with the greatest term, we have for ff, $ff = q/p$, and for the other tail, $ff = p/q$.

As an example of the application of the geometrical progression method in the case of a term on the shorter tail of the distribution, let us assume that 90 eels were caught in our trap A. The value of

$p(90, 10)$ is found to be 0·003362, $cf=\frac{10}{91}$, $ff=4$, $r=ff\times cf=0{\cdot}4396$, $Q=0{\cdot}006001$, the true value of P being 0·005694.

The method of the geometrical progression is also most useful in dealing with the results of independent parallel trials, whether $n_1=n_2$ or not. Thus, in the example shown in Table 1, a value greater than $P(\pm 11)$ and which I propose from now on to designate by Q is given by

$$Q(\pm 11)=0{\cdot}01555\times\frac{1}{1-r},$$

where

$$r=\frac{10\times 19}{22\times 31},$$

10 and 19 being the lesser pair of numbers on a diagonal of the appropriate fourfold table, 21 and 31 being one more in each case than the numbers on the other diagonal, in accordance with the rule given in Section 6. The value of r is 0·2786 and

$$Q=0{\cdot}01555\times 1{\cdot}386=0{\cdot}02157.$$

In cases in which $n_1=n_2$ it will always be found that, in the region of the distribution in which P is less than 0·1, Q, though always greater than P, is an extremely good approximation to P.

Table 11

4	16	20
4	68	72
8	84	92

As an example of the application of the method in a case in which $n_1\neq n_2$, the case shown on p. 232 of Yates(1) may be taken. The value of p for the observation shown in the fourfold table, Table 11, is 0·05355, $r=\frac{4\times 16}{5\times 69}=0{\cdot}1856$, $Q=0{\cdot}06577$. The true value of P is 0·06460 and the value given by the normal method with Yates's correction is 0·0571. Thus the geometrical progression method gives a result far closer to the true value than does the χ-test even with the correction and a result far the more useful since it is known to be greater than the true value.

If one single method be required for all cases of parallel trials, both independent and binomial, the geometrical progression method is highly recommended for this purpose as its results are never

ambiguous and no reference to any table embodying corrections or to a table of the normal integral is ever necessary. It will always be found that, in the 'critical region' of the distribution, in which the true value of P lies between 0·05 and 0·01 on one tail, and also in the whole region beyond the 0·01 point the geometrical progression method gives a very good approximation to the true value of P. Here are some further examples, in which P is the true value and $P(\chi')$ the approximation given by the normal method with Yates's correction.

(1) What is the probability of at least 15 heads in a toss of 20 coins?

$$\begin{aligned} p(15, 5)&=0{\cdot}01479,\\ P&=0{\cdot}020695,\\ P(\chi')&=0{\cdot}02211,\\ Q&=0{\cdot}02152. \end{aligned}$$

(2) Required the sum of the last 12 terms of the binomial $(0{\cdot}75+0{\cdot}25)^{20}$:

$$\begin{aligned} p(9, 11)&=0{\cdot}0271,\\ P&=0{\cdot}0410,\\ P(\chi')&=0{\cdot}0351,\\ Q&=0{\cdot}0428. \end{aligned}$$

(3) Required the sum of the first 261 terms of the binomial $(0{\cdot}\dot{3}+0{\cdot}\dot{6})^{450}$:

$$\begin{aligned} p(260, 190)&=1{\cdot}71\times 10^{-5},\\ P&=5{\cdot}17\times 10^{-5},\\ P(\chi')&=3{\cdot}91\times 10^{-5},\\ Q&=5{\cdot}35\times 10^{-5}. \end{aligned}$$

In addition to supplying a value which is known to be greater than the true value of P the geometrical progression provides a method of continued approximation, only one step of which is used in the method as described in the previous part of this section. Thus, if the greatest term in that part of a binomial or of a fourfold table distribution which is to be summed to give the value of P is designated by a', the second term in the sum by a'' and so on, the

first multiplier or the ratio of a'' to a' by r', the second multiplier by r'' and so on, we have

$$P > a'(1+r') < \frac{a'}{1-r'},$$

$$P > a' + a''(1+r'') < a' + \frac{a''}{1-r''},$$

$$P > a' + a'' + a'''(1+r''') < a' + a'' + \frac{a'''}{1-r'''},$$

and so on. Thus, in our example 3 of this section, omitting the factor 10^{-5} in the working,

$$r' = \frac{l}{k+1} \times \frac{q}{p} = \frac{260}{191} \times 0{\cdot}5 = 0{\cdot}6808,$$

$$a' = 1{\cdot}71,$$

$$P > 1{\cdot}71(1{\cdot}6808) < \frac{1{\cdot}71}{0{\cdot}3192},$$

$$P > 2{\cdot}875 < 5{\cdot}357, \quad d = 2{\cdot}482,$$

$$r'' = \frac{259}{192} \times 0{\cdot}5 = 0{\cdot}6745, \quad a'' = 2{\cdot}875 - 1{\cdot}71 = 1{\cdot}165,$$

$$P > 1{\cdot}71 + 1{\cdot}165(1{\cdot}6745) < 1{\cdot}71 + \frac{1{\cdot}165}{0{\cdot}3255},$$

$$P > 3{\cdot}662 < 5{\cdot}290, \quad d = 1{\cdot}628,$$

$$r''' = \frac{258}{193} \times 0{\cdot}5 = 0{\cdot}6683, \quad a''' = 3{\cdot}662 - 2{\cdot}875 = 0{\cdot}787,$$

$$P > 2{\cdot}875 + 0{\cdot}787(1{\cdot}6683) < 2{\cdot}875 + \frac{0{\cdot}787}{0{\cdot}3317},$$

$$P > 4{\cdot}188 < 5{\cdot}248, \quad d = 1{\cdot}060,$$

$$r'''' = \frac{257}{194} \times 0{\cdot}5 = 0{\cdot}6624, \quad a'''' = 4{\cdot}188 - 3{\cdot}662 = 0{\cdot}526,$$

$$P > 3{\cdot}662 + 0{\cdot}526(1{\cdot}6624) < 3{\cdot}662 + \frac{0{\cdot}526}{0{\cdot}3376},$$

$$P > 4{\cdot}536 < 5{\cdot}220, \quad d = 0{\cdot}684.$$

In the above calculations d stands for the difference between the limits for the value of P. The method of linear extrapolation may be used to obtain the true value of P with a good degree of accuracy from the two last calculated values of the upper limit for P and the two last values of d. The difference between the values of d is 0·376 and that between the corresponding values of P is 0·028. Thus the corresponding decrease in P for a decrease in d from 1·060 to zero is

$$\frac{0{\cdot}028}{0{\cdot}376} \times 1{\cdot}060 \quad \text{or} \quad 0{\cdot}07894.$$

Therefore our estimate of P is $5{\cdot}1691 \times 10^{-5}$.

The true value of P, found by adding up the binomial terms, is $5{\cdot}17 \times 10^{-5}$ to two significant places of decimals, and the method of linear extrapolation gives the same result. Assuming now that we had carried out only the first two steps in the approximation we should have had, for the decrease in P,

$$\frac{0{\cdot}067}{0{\cdot}854} \times 2{\cdot}482 = 0{\cdot}1946,$$

$$P = 5{\cdot}357 - 0{\cdot}1946 = 5{\cdot}1624 \times 10^{-5}.$$

Even with only two steps of approximation therefore a very good estimate of the true value of P has been obtained. It is, however, preferable to carry out at least three stages of approximation, to find out whether linear extrapolation is justifiable, if it is wished to estimate the value of P to a given degree of accuracy and not merely 'approximately'. Our estimates are 5·1624, 5·1697 and 5·1691, multiplied each by 10^{-5}, according to the step in the approximation from which the estimate is made. It is safe, therefore, to say that the true value of P is 5·17 to within ± 1 in the last figure.

SECTION 12. *Suggestions as to choice of method of testing for significance in particular cases*

(1) *Fourfold table.* $n_1 = n_2 = n$. $n < 30$s

If n be less than 16 or is equal to 20 or 30, obtain exact value of P from Tables 2, 3 or 4. If n be not included in Tables 2, 3 or 4, for preliminary test use method of Section 3A. If working by a

fixed criterion and if the approximate value of P found by that method be well below the chosen critical value, it is not necessary to carry our further tests. If exact value of P be required or if working by decrease in value of P with additional trials, find exact values of P by methods of Sections 1 or 6.

(2) *Fourfold table.* $n_1 = n_2 = n$. $n > 30s$

If $s < 51$ use method of Section 4A with Table 5. If $s > 50$ use method of Section 11 to obtain Q, a value slightly greater than the true value of P.

(3) *Fourfold table.* $n_1 \neq n_2$

For preliminary trial use method of Section 11 to obtain Q. If significance doubtful and both n_1 and n_2 are less than 51, obtain exact value of P by method of Section 2 with Table 5, but, if either n_1 or n_2 are greater than 50, find value of P by method of Section 6. If both n_1 and n_2 are very large compared with s the method of Section 4B is applicable, but it is not possible to say how large the ratio $n : s$ must be for this. The method of Section 11 will nearly always settle the question of significance. If it does not it is necessary to find the exact value of P by the method of Section 6.

(4) *Binomial cases. The symmetrical binomial,* $(0{\cdot}5 + 0{\cdot}5)^n$

For preliminary trial use the approximate method of Section 10. If significance is doubtful and if n be less than 51 use the exact method of Section 10 with Table 5. For values of n greater than 50 the normal method gives a value very near to the true value of P except when P is small ($< 0{\cdot}05$ for both tails together). In that case the method of Section 11 gives a better approximation and should be used when the approximate value given by the normal method is less than 0·06.

(5) *Binomial cases. The asymmetrical binomial*

For preliminary trial use the method of Section 11. If significance is doubtful the exact method of Section 10 must be used but it may be very laborious. The normal method is not applicable to asymmetrical binomials and may give very misleading results.

SECTION 13. *On the philosophical basis of tests for significance*

It is a help, in understanding the basic theory of statistical tests, to form an imaginary model 'population' from which an observed value—a difference for instance—may be considered to be a random draw. A suitable model would be a heap of coloured balls, each difference being represented by balls of a particular colour, the number of balls of any colour being proportional to the probability of the represented difference in the random sampling distribution. We shall assume that there are 10,000 balls altogether in the heap. Thus, for a difference, $\pm d$, represented by red balls and having the probability, p, equal to 0·05, there will be 500 balls in the heap. There is a very large number of these heaps of balls, the number of colours and also the total number of balls in each being the same, but the proportional numbers of balls of each colour varying from one heap to another. In many of the heaps the proportional numbers of balls of each colour may, however, be the same. In one and only one kind of heap, type A, the number of balls of each colour is known but it is not known what proportion the number of heaps of this particular kind bears to the total number of heaps. In other words, it is quite impossible to estimate the *probability* that any particular heap is of the kind in question. The balls in each heap are assumed to have been thoroughly mixed together. We now draw a ball from one of the heaps, find that it is a red ball, and wish to decide whether it is reasonable to assume that it came from a heap of type A. We know, perhaps, that in a heap of this type there are 500 red balls and that there are altogether 200 balls of colours representing differences further from the 'expected' difference than that represented by red balls. In the type of distribution we are concerned with greater distance from the expected difference means lesser probability, so we can say that 200 is the total of balls represented less frequently than are red balls. We now have to decide whether the *probability* of drawing, from a heap of type A, a red ball or one of those less well represented is so small that it is very *unlikely* that the heap drawn from was of type A. We know perfectly well that, in many of the other heaps, there may be a far greater proportion of these balls than in the heaps of type A, and that our ball may be therefore much more likely to have been drawn from one of these others. We are only concerned, however, in deciding whether it can be considered as *proved* that it did not come from type A. By convention this

proof is considered sufficiently strong if the number of balls in question is less than 5% of the total number of balls. If the differences, $+d$ and $-d$, have differing probabilities the observed positive difference and greater positive differences will be represented by colours unlike those corresponding to the negative differences, and in that case the 'critical' proportion of balls is 2·5%. It is open to any worker using statistical tests to employ values of P other than these as criteria of proof. It savours rather of arrogance to lay it down as a law that $P<0{\cdot}05$ shows 'significance' and that $P<0{\cdot}01$ shows 'high significance', as if these concepts appeared suddenly with a given value of P. It is far better in published work, wherever possible, to give the true value of P and to leave it to the reader to decide whether the reality of an effect is proved to his satisfaction or not.

If the critical value, $P=0{\cdot}05$, be adhered to as deciding significance the decision that the reality of an effect is proved will be wrong on the average in one case in twenty trials, *if the null hypothesis be really true.* Since, however, there is no possible way of proving its truth one must not be misled into thinking that it *will* be true once in 20 times if, in each of twenty trials, the observed difference corresponds to the value 0·05 for P.

It should be noticed that the degree of conformity, shown by an observation, with the distribution defined by the null hypothesis is spoken of as showing the degree of *likelihood* of the truth of that hypothesis. The words 'probable' and 'probability' should never be used as if they were synonymous with 'likely' and 'likelihood'. Probability has an exact meaning in statistical theory, implicit in that meaning being the fact that all possibilities are known and that their relative chances of occurrence are calculable, at least approximately. Likelihood is merely a term measuring agreement between an observation and a particular *hypothetical* cause thereof and is in no way a measure of the *probability* of that cause. The two concepts are fundamentally different, likelihood having nothing whatever to do with chance of occurrence. One can compare likelihoods, say that one hypothetical cause of a result is more likely than another, and say that a certain hypothetical cause is the most likely of all hypothetical causes of the same type or, in other words, has maximum likelihood. Such values of likelihood are, however, merely relatively comparable measures of conformity between certain chosen hypotheses and observed values and have nothing to do with the *probability* of the truth of those hypotheses, which are picked out from a completely unknown distribution of hypotheses.

Section 14. *Theoretical notes*

(1) *The origin of the binomial distribution*

The binomial is the most important of the discontinuous distributions. Its development is easy to follow and its logical basis simple. To understand the theory of the binomial is to be convinced that statistical methods are really founded on common-sense principles. The simplest case to which the binomial applies is that of a toss of a number of coins, and it will be assumed that five coins are to be used for illustrating our argument. It is necessary at first to distinguish the coins from each other, and we shall assume that they are marked with the letters A, B, C, D and E respectively. Now for any coin, for instance, A, the chances of a head or a tail are clearly equal to one another if there is absolutely no bias and the toss is carried out absolutely fairly. The probability of a head on any coin is therefore equal to $\frac{1}{2}$, the chance of a head *or* a tail on any coin being equal to unity or certainty. Considering now a toss of two coins, A and B, there are four possible results, $A(h)\,B(h)$, $A(h)\,B(t)$, $A(t)\,B(h)$, $A(t)\,B(t)$, and the probability of a combination of any two *particular* independent events, for instance, $A(h)$ and $B(t)$, for each of which the probability is $\frac{1}{2}$, is equal to $\frac{1}{2}\times\frac{1}{2}$. This is obvious from the fact that, if there are two ways of doing a thing once, there are four ways of doing it twice, eight ways of doing it three times and so on. It is an illustration of a fundamental axiom in the theory of chance, namely, that the probability of the simultaneous occurrence of n events, for each of which the probability is p, is equal to p^n, or, in general, is equal to the product of the separate probabilities. Now let us consider alternative events. In the toss of two coins the event, one head and one tail, when the coins are not distinguished from one another, can occur in two ways, $A(h)\,B(t)$ and $A(t)\,B(h)$. This is an illustration of the axiom that the probability of one or other of two independent events is equal to the sum of the separate probabilities.

In the binomial distribution as applied to coins these are not distinguished from each other and the terms are usually arranged according to the number of 'successes'—heads, for instance—for which the corresponding term gives the probability. Thus the different results in a toss of two coins may be arranged in a binomial distribution as follows:

Heads	0	1	2
p	1/4	$2\times1/4$	1/4

Let us suppose now that we toss three coins. The probability of 3 heads is equal to $1/2 \times 1/2 \times 1/2$, since each coin must show a head; the result, 2 heads and 1 tail, may be either $A(h)\ B(h)\ C(t)$, $A(h)\ B(t)\ C(h)$ or $A(t)\ B(h)\ C(h)$. Each of these combinations has the probability, 1/8, so the probability of any one or other of them is $3 \times 1/8$. For a toss of 5 coins the distribution is as follows:

Heads	0	1	2	3	4	5
p	$1/32$	$5 \times 1/32$	$10 \times 1/32$	$10 \times 1/32$	$5 \times 1/32$	$1/32$

Here, as before, the terms for 1, 2, 3 and 4 heads are each the sum of a number of ways of getting those results, each of which ways has the probability, $1/2 \times 1/2 \times 1/2 \times 1/2 \times 1/2$. The *whole number* in each term is known as the coefficient of that term, the other factor may be called the fractional factor. In the symmetrical binomial, in which the probability of a single component event is 1/2, the fractional factor is always equal to $(1/2)^n$, where n is the number of coins in the toss. The mathematical equation for calculating the value of the coefficient is

$$nCr = \frac{n!}{(n-r)!\,r!},$$

in which n is the number of coins and r the number of heads. The expression on the left stands for 'the number of combinations of n things taken r at a time' or 'the number of different ways of drawing r coins from n coins', while $n!$ stands for 'factorial n' which is the product of all integers up to n inclusive. Note that $0! = 1! = 1$.

The symmetrical binomial is written shortly in the form $(0{\cdot}5 + 0{\cdot}5)^n$. Our Table 5 gives the logarithm of the coefficient, from $n = 2$ to $n = 50$. Since, in tests using the symmetrical binomial, heads and tails are not usually distinguished, it is convenient to arrange the terms of the distribution so as to show the probability of any given difference between the numbers of heads and tails. The 'expected difference' is 0 or ± 1, according to whether n is even or odd.

Let us suppose now that instead of coins we use 5 dice, lettered A to E, and the chance distribution of the number of sixes is required. The chance of throwing a 6 on any particular die, A, is 1/6 and the chance of not doing so is 5/6. With two dice, A and B,

the chance of throwing a 6 on each is $(1/6)^2$, but the chance of a 6 on A and another number on B is $1/6 \times 5/6$. This differs from the result obtained with the particular coins, A and B, in which the chance of the throw, $A(h)\ B(h)$ was seen to be equal to the chance of $A(h)\ B(t)$. If we designate a throw of a number other than 6 by the letter a and a throw of 5 dice be made, we shall see, for instance, that the throws, $A(6)\ B(6)\ C(a)\ D(a)\ E(a)$, $A(a)\ B(6)\ C(a)\ D(6)\ E(a)$ or any other *particular* throw of two sixes has the probability $(1/6)^2\,(5/6)^3$. The coefficients remain the same as in the case of the symmetrical binomial but the fractional factors decrease in the ratio 5:1 from term to term.

The general term for the asymmetrical binomial, $(q+p)^n$, in which p is the chance of a 'success' is given by

$$_nC_r q^{n-r} p^r.$$

The general law of chance exemplified by the binomial may be stated as follows: If an event can happen in a number of different ways, the proportional number of times it *will* happen in any particular way will, in the long run, be equal to the proportional number of times it *can* happen in that way. The truth of this law is self-evident. If in a very long series of throws with coins or dice the observed results do not conform more and more nearly with the appropriate binomial distribution as the number of throws is increased there will be every reason to suspect either that the coins or dice are biased or that there is something wrong with the method of throwing. The question of testing whether an apparent disagreement is significant or not is outside the scope of this paper, since the test is, in this case, for the significance of a discrepancy between a complete observed distribution and a complete theoretical distribution, whereas every observation dealt with in this paper is considered in our tests as a single term from a hypothetical distribution. The purpose of our tests has been to decide whether or not this term is near enough to the most likely term for the hypothetical distribution to be acceptable as having given rise to the observation.

(2) *The origin of the distribution of the difference between numbers of events in two sets of independent parallel trials*

Let us assume that we have two sets of five dice, all of which are believed to be loaded, so that the chance of a six on any die is not $\frac{1}{6}$ as it would be in the case of true dice but is an unknown

quantity. It has been suggested that the two sets of dice came from the same maker and were made at the same time, and it is wished to test whether this is likely or not. We are not interested in finding out whether the dice are untrue as there is strong evidence that this is so although the degree of bias has not been estimated. The question to be decided is simply whether the two sets are likely to have had the same origin, as judged by similarity of bias. We therefore make six throws with each set, add up the number of sixes for each set and form a 'fourfold table' (Table 12) to show the results.

Table 12

	Set 1	Set 2	
Sixes	10	15	25
Not six	20	15	35
Total	30	30	60

There is a difference of 5 between the numbers of times a 6 is thrown in the two sets. We now make the assumption that the chance of a 6 in every one of the dice is the same. If this is the case the best estimate—indeed, the only estimate—we have of the chance of a 6 on any die is given by the total number of sixes divided by the total number of throws. This chance, which we shall designate by p, is thus $\frac{25}{60}$ or 0·416. The expected distribution of the number of sixes in either case is therefore the binomial distribution $(0{\cdot}58\dot{3}+0{\cdot}41\dot{6})^{30}$. Now the difference, 5, between the numbers of times a 6 may be thrown in a pair of sets of 30 throws may arise in many different ways. For instance, we may, as shown in Table 12, throw 10 sixes with one set, 15 sixes with the other or we may throw 11 sixes with one and 16 with the other. As, however, we have estimated the value of p from a total of 25 sixes we must not take into account any throw which would give a different value of p. That would be using imaginary information which we have not got. We must take account therefore only of the one way, 10 sixes in one set, 15 in the other. Since the probability of two independent events is equal to the product of the separate probabilities the probability of 10 sixes and 15 sixes respectively in two throws of 30 each, or of a difference of 5 *either way* would be given by twice

$$\frac{30!}{10!\,20!}(0{\cdot}41\dot{6})^{10}(0{\cdot}58\dot{3})^{20}\times\frac{30!}{15!\,15!}(0{\cdot}41\dot{6})^{15}(0{\cdot}58\dot{3})^{15},$$

were it not for the fact that in our test the total number of sixes is limited to 25. Now, if all possible pairs of throws be considered, with the total number of sixes limited to 25, and the probabilities of all such be added together, it will be found that the resulting sum is not unity but is equal to the probability of throwing a total of 25 sixes in 60 throws with p equal to $0{\cdot}41\dot{6}$. A complete sum of probabilities, if these are to be absolute and not merely relative, must always, however, be equal to unity or certainty. Therefore to render the probability of the difference, 5, absolute the unrestricted probability found must be divided by the probability of 25 sixes in 60 throws. This is equal to

$$\frac{60!}{25!\,35!}(0{\cdot}41\dot{6})^{25}(0{\cdot}58\dot{3})^{35}.$$

The result of the division is equal to twice

$$\frac{30!}{10!\,20!}\times\frac{30!}{15!\,15!}\div\frac{60!}{25!\,35!}.$$

It will be seen that the terms in p and q do not enter into this result at all. The only terms remaining are the binomial coefficients. It is the logarithms of these coefficients which are given in our Table 5, and this table therefore affords a rapid means of calculating the probability of any difference between numbers of occurrences in two sets of parallel trials when the total number of trials in each set is 50 or less. If, however, the total in the two sets is over 50 the dividing probability will have to be calculated from a table of logarithms of factorials.

If, now, the number of trials in the two sets be unequal, exactly the same kind of procedure as that described above is applicable except that the binomial coefficients to be multiplied together are those applying to different numbers, n, in Table 5. Thus, if the numbers of trials had been 20 and 30 respectively, the total number of sixes being 25 as before and the difference, 5, being due to an excess of sixes in the greater set of trials, we should have for the probability of this difference

$$\frac{30!}{15!\,15!}\times\frac{20!}{10!\,10!}\div\frac{50!}{25!\,25!}.$$

If the difference, 5, had been due to an excess of sixes in the lesser set of trials the probability of this difference would have been

$$\frac{30!}{10!\,20!}\times\frac{20!}{15!\,5!}\div\frac{50!}{25!\,25!}.$$

It will be seen therefore that the probability of the difference, 5, varies according to whether it is due to an excess in the greater or in the lesser set of trials.

The application of results such as those discussed in this section to practical tests of significance is fully described in other sections, and it is therefore unnecessary here to complete the test of our imaginary case.

(3) *Note on the method of Section 4A*

If the chance, p, of an event be very small the distribution of number of events approximates to the Poisson distribution, in which p is assumed to be infinitesimal or n, the number of trials infinitely large, p and n being so balanced that the mean number of events, m, is a finite number. The probability of any difference between two numbers independently distributed according to the Poisson series is equal to the sum of the products of the appropriate pairs of terms in that series. The estimate of m, the mean of the series which is assumed to be common to both sets of trials, is s, the sum of the numbers of events in the two sets. It is assumed for the test for significance that s is invariable. There is therefore only one pair of terms to be multiplied together to give the probability of any particular difference. If, for instance, for the two sets together, $m=s=6$, the proportional probabilities of a difference of 0, 2, 4, 6, between the numbers of events in the two sets are given by the following products in order:

$$e^{-3}.e^{-3}\left(\frac{3^3}{3!}\frac{3^3}{3!}\right)=e^{-6}\left(\frac{3^6}{3!\,3!}\right),$$

$$e^{-3}.e^{-3}\left(\frac{3^2}{2!}\frac{3^4}{4!}\right)=e^{-6}\left(\frac{3^6}{2!\,4!}\right),$$

$$e^{-3}.e^{-3}\left(\frac{3^1}{1!}\frac{3^5}{5!}\right)=e^{-6}\left(\frac{3^6}{1!\,5!}\right),$$

$$e^{-3}.e^{-3}\left(\frac{3^6}{6!}\right)=e^{-6}\left(\frac{3^6}{6!}\right).$$

To obtain the actual probabilities these terms must be divided by the probability of 6 events with a mean of 6. This is equal to

$$e^6\left(\frac{6^6}{6!}\right).$$

The resulting quotients are as follows:

$$\frac{6!}{3!\,3!}\left(\frac{1}{2}\right)^6,$$

$$\frac{6!}{2!\,4!}\left(\frac{1}{2}\right)^6,$$

$$\frac{6!}{1!\,5!}\left(\frac{1}{2}\right)^6,$$

$$\left(\frac{1}{2}\right)^6,$$

and these are the terms in the expansion of the binomial $(0{\cdot}5+0{\cdot}5)^6$. A similar process applied in the asymmetrical case of Section 4B results in the asymmetrical binomial $(q+p)^n$ in which $p\neq q$.

SECTION 15. *Some cases in which the methods described in previous sections are applicable*

(1) Bacteriological applications. Some of these are dealt with in Buchanan-Wollaston(5). Where there are discrepancies in theory in that and in the present paper the treatment in the present paper is to be preferred. Such discrepancies do not occur in relation to exact methods.

Of the following examples only the last refers to an actual, the remainder to hypothetical cases.

(2) Of two newly employed workmen one was late twice in 30 days, the other not late once. Assuming that lateness may be caused by chance missing of buses and so on, is what happened significant? In Table 4, $s=2$, $d=2$, $P=0{\cdot}4916$. In very nearly every alternate trial of 30 days the observed result would occur by chance if the two workmen had exactly the same chance of being late. Taken by itself the result is completely insignificant.

(3) One workman broke 10 drills in 20 days' work, another broke 5 drills, both having done the same amount of work. Assuming that the number of 'trials' was very large compared with the

number of breakages in each case and equal for both workmen, the method of Section 4A is applicable. $P(\pm 5)=0{\cdot}3018$. The difference is quite insignificant.

(4) Of 15 drills broken in a particular test, 12 were of one make, 3 of another make, 30 of each being used altogether. In Table 4, $s=15$, $d=9$, $P=0{\cdot}007913$. In less than one in a hundred trials would so great a difference arise by chance. It is assumed here that the trials were such as to give all drills the same chance of being broken.

(5) In a compound flower cyme there were found to be 40 sterile and 110 fertile florets. In a cyme of a related species there were 30 sterile and 110 fertile florets. Is there a significant difference between the proportions of fertile to sterile florets? Find Q by method of Section 11 and, if necessary, carry out further steps in the approximation to P. It is assumed here that the chance of being fertile is the same for all florets in a cyme.

(6) In 10 randomly chosen cymes of one species there were 453 sterile and 1152 fertile florets. In 10 cymes of another species there were 312 sterile and 1264 fertile florets. Same test as in Ex. 5.

(7) Five water samples from various sites in Lake Windermere gave counts, respectively, of 10, 15, 30, 40 and 70 diatoms in a haemocytometer cell. Five samples taken from the same sites a week later gave counts, respectively, of 12, 20, 29, 45 and 85 diatoms in the cell. Was there a significant rise or fall in the diatom population of the sampled water masses? Test the counts of each set first of all by the χ^2 method to see whether it is reasonable to assume that they have arisen from the same Poisson distribution. If so, test for significance the difference between the total, 165, for the first set and the total, 191, for the second set. The methods of Section 10 or those of Section 11 may be used. If the numbers in either set are heterogeneous, find the values of $P(d)$ and $P(d+2)$ for the first two pairs of numbers by the exact method of Section 10 and, for the last three pairs, by the approximate method of that section. Calculate mean χ for each pair and test total χ for significance by method of Section 8. For the test for homogeneity and in the test of totals dilution must be the same for all samples. For the test of total χ it is only necessary that dilution be the same for the two members of each pair.

(8) Two trawlers were fishing on the same grounds. Trawler A caught 56 soles over 25 cm. long in 7 hauls while trawler B caught 107 soles of over that length in 6 hauls. Of the soles above 25 cm. long caught by A, 23 were over 31 cm. long. Of those more than 25 cm. long caught by B, there were also 23 over 31 cm. long. Is there a significant difference between the ratios, number of soles 25–31 cm. long : number of soles over 31 cm. long in the two cases? The value of Q calculated by the method of Section 11 was found to be $0{\cdot}007866$, and therefore the difference is found to be highly significant.

LIST OF LITERATURE

(1) YATES, F. (1934). Contingency tables involving small numbers and the χ^2-test. (Supplement to *J. Roy. Statist. Soc.* vol. 1, no. 2.)

(2) PEARSON, KARL (1930). *Tables for Statisticians and Biometricians.* Part 1, Third edition. Cambridge University Press.

(3) FISHER, R. A. and YATES, F. (1938). *Statistical Tables for Biological, Agricultural, and Medical Research.* London: Oliver and Boyd.

(4) FISHER, R. A. (1934). *Statistical Methods for Research Workers.* Fifth or later edition. London: Oliver and Boyd.

(5) BUCHANAN-WOLLASTON, H. J. (1941). On tests for the significance of differences in degree of pollution by coliform bacteria and on the estimation of such differences. *J. Hyg., Camb.*, vol. XLI, no. 2.

CAMBRIDGE: PRINTED BY
WALTER LEWIS, M.A.
AT THE UNIVERSITY PRESS