NOTES ON THE SUBJECT MATTER OF A PREVIOUS PAPER[1]

BY H. J. BUCHANAN-WOLLASTON

THE theory detailed in Section 7 (p. 159) needs further explanation.

TEST OF THE MEAN OF A SET OF DIFFERENCES IN N.F.T.

The meaning of the footnote on p. 162 is somewhat ambiguous and the statement comprised in it is not true for all values of n, the number of differences in N.F.T. of which the significance of the mean is tested.

The statement that, by the use of a certain approximate distribution, significance of an observation is underestimated means that the estimated value of P in the neighbourhood of the significance level is greater than the true value. Assuming that the 0·05 level be used for testing the significance of the mean of a set of differences expressed on the normal scale, it will be found generally that, when n is very small, say, 4 or less, P will be *underestimated* towards the tails of the distribution. This is due to the fact that the number of possible means is comparatively small, each therefore having a fairly high probability, p. The error due to discontinuity of the true distribution is therefore appreciable. With very small increase in n, however, the number of possible means becomes greatly increased and the error due to discontinuity unimportant. The error due to deviation of the distribution compounded of the letter classes from the normal form then begins to show up. This deviation almost always has the characteristic of a concentration of frequency in the 'shoulders' of the curve at the expense of the tails, P in the critical region being slightly overestimated and significance therefore underestimated. This error soon becomes evanescent with increasing n, the distribution of the mean normal equivalent difference rapidly approaching the normal form. If the test be not used for sets of less than 4 differences there seems to be no need to make allowance for either of the types of error referred to.

I have made an experiment in random sampling to investigate the errors mentioned in the previous paragraph. The set of letter classes chosen was the same as in Example 1 of Table 16 in my previous paper, one draw being made from each of the classes of the example, drawing being made by means of supposedly random numbers, the set of 6 draws being repeated 100 times. The expected and observed distributions were as given in Table 1 below.

The χ^2-test was applied to the letter classes separately since the total frequency in each class was limited to 100 and the frequency classes were not the same throughout. The sum of all the values of χ^2 was 21·7827, the number

[1] "On tests of the significance of differences in degrees of pollution by coliform bacteria and on the estimation of such differences," Vol. **41**, pp. 139–68.

of degrees of freedom 18 and the value of P about 0·25. The hypothesis that the numbers from which the draws were made were random numbers is therefore acceptable. For the differences in N.F.T. the equivalent normal values, x, were then substituted and the total value of each set of 6 then determined. The expected distribution of the totals was then arranged in terms of the standard deviation of a total of 6, namely $\sqrt{6}$. Table 2 below shows the expected and observed distributions and the value of χ^2.

Thus, with n equal to only 6 there is no significant deviation from the expected distribution. The observed tail frequencies seem somewhat too low, however, and, if the expected distribution had been used to estimate the significance of a total or mean near the critical value, significance would have been rather underestimated.

Table 1

Difference in N.F.T.	−6	−5	−4	−3	−2	−1	0
Expected frequency	0·5	3·0	11·6	38·1	49·0	158·9	77·8
Observed frequency	1	1	12	30	55	157	61
Difference in N.F.T.	+1	+2	+3	+4	+5	+6	—
Expected frequency	158·9	49·0	38·1	11·6	3·0	0·5	—
Observed frequency	174	54	36	15	1	2	—

Table 2

Centre of frequency class	$->2\sigma$	$-1\frac{2}{3}\sigma$	-1σ	$-\frac{1}{3}\sigma$	$+\frac{1}{3}\sigma$	$+1\sigma$	$+1\frac{2}{3}\sigma$	$+>2\sigma$
Expected	(2·3	6·9)	16·1	24·8	24·8	16·1	(6·9	2·3)
Observed	(0	4)	16	26	32	18	(3	1)

Total $\chi^2=8{\cdot}250$. D.F. $=5$. $P=$ about 0·15.

Table 3

Centre of frequency class	$->1\frac{2}{3}\sigma$	$-1\frac{1}{3}\sigma$	$-\frac{2}{3}\sigma$	0	$+\frac{2}{3}\sigma$	$+1\frac{1}{3}\sigma$	$>+1\frac{2}{3}\sigma$
Expected	(2·39	5·54)	10·55	13·04	10·55	(5·54	2·39)
Observed	(0	5)	12	14	9	(8	2)

Total $\chi^2=2{\cdot}1203$. D.F. $=4$. $P=$ about 0·7.

To investigate the approach of the random sample distribution to normal form with increase in n, 50 sets of 12 differences were made up from the 100 sets of 6 by combining them in pairs. The resulting comparison is shown in Table 3.

The fit between hypothetical and observed distributions is now very good and, with increase in n from 6 to 12, the error in estimation of the tail frequencies has almost if not quite disappeared. The apparent discrepancy may very well be due to the error inherent in random sampling, the error which would be expected if the two distributions were really exactly of the same form.

Test of the mean error or of the standard error from zero

The test proposed in Section 7 was something of a compromise. The distribution of the standard error in random samples from a normal universe is not approximately normal unless n be large. The calculation of an estimate of the standard error of a set of differences expressed by normal equivalents, by squaring the values of the mean value of x for each class, results in an under-estimation and it was considered that this would to a certain extent counter-balance the effect of using a normal to represent a non-normal distribution. A method, more nearly theoretically correct, is to employ the χ^2-distribution in a test of a set of mean square normal equivalent differences. A comparison of the expected and observed distributions of χ^2 was made, using the random numbers referred to, with sets of 6 and of 12 differences respectively. The agreement was not, however, at all good, this being very largely due to the error of discontinuity which has a much more marked influence in the case of the mean error than in the case of the mean difference. It seems preferable, therefore, to apply Fisher's test of a set by the combination of probabilities (Fisher, 1934, p. 103).

In Fisher's test, for each difference we have

$$-2 \log_e P = \chi^2,$$

in which P is taken direct from Table 3 of my previous paper. The resulting values of χ^2 for the differences of the set are summed and the χ^2-table is entered with degrees of freedom equal to twice the number of differences in the set. Class a should be omitted when counting degrees of freedom, since there can be no error in that class when account is not taken of sign. Examples taken from Table 16 of my previous paper are given in Table 4.

Table 4

d	Class	P	χ^2
2	m	0·6926	0·9252
1	a	1·0000	0·0000
5	i	0·0238	7·4754
0	k	1·0000	0·0000
3	g	0·2222	3·0084
1	l	1·0000	0·0000
			11·4090

D.F. = 10. P = about 0·33.

d	Class	P	χ^2
5	y	0·0555	5·7828
5	p	0·0635	5·5134
8	H	0·0035	11·3158
2	d	0·5238	1·2934
3	n	0·3386	2·1658
3	l	0·3148	2·3118
			28·3830

D.F. = 12. P = less than 0·003238.

In the first example the mean error from zero is nearly that most likely to occur by chance if this mean error were really zero. The significance of the mean difference is thus due mainly to consistency in sign of the differences. In the second example the differences themselves are very much larger than would be expected by chance and there is also marked consistency in sign. The values of χ^2 are taken from Table 6 of the present paper.

Table 5

	Sums			Equivalent mean normal abscissa for total difference in N.F.T. equal to							
	S_1	S_2	S_3	0 or 1	2 or 3	4 or 5	6 or 7	8 or 9	10 or 11	12 or 13	14 or 15
a	0	0	1	0·7979							
b	0	0	2	0·3650	1·340						
c	0	0	3	0·5872	1·835						
d	0	0	4	0·3077	1·131	2·373					
e	0	0	5	0·5534	1·690	2·985					
f	0	1	1	0·3247	1·271						
g	0	1	2	0·5393	1·703						
h	0	1	3	0·2666	1·015	2·151					
i	0	1	4	0·4845	1·486	2·625					
j	0	1	5	0·2530	0·9640	1·987	3·177				
k	0	2	2	0·2620	0·9888	2·055					
l	0	2	3	0·4630	1·396	2·502					
m	0	2	4	0·2359	0·9018	1·833	2·846				
n	0	2	5	0·4440	1·344	2·293	3·381				
o	0	3	3	0·2316	0·8810	1·786	2·787				
p	0	3	4	0·4243	1·294	2·165	3·178				
q	0	3	5	0·2222	0·8470	1·713	2·568	3·609			
r	0	4	4	0·2182	0·8316	1·672	2·564	3·500			
s	0	4	5	0·4123	1·241	2·075	2·954	4·000			
t	0	5	5	0·2120	0·8148	1·655	2·462	3·398	4·425		
u	1	1	1	0·5150	1·647						
v	1	1	2	0·2477	0·9521	2·016					
w	1	1	3	0·4466	1·370	2·422					
x	1	1	4	0·2288	0·8754	1·776	2·837				
y	1	1	5	0·4293	1·310	2·277	3·367				
z	1	2	2	0·4370	1·332	2·368					
A	1	2	3	0·2148	0·8416	1·716	2·660				
B	1	2	4	0·4073	1·236	2·086	3·072				
C	1	2	5	0·2095	0·8119	1·642	2·539	3·545			
D	1	3	3	0·3995	1·207	2·074	3·002				
E	1	3	4	0·2036	0·7830	1·578	2·450	3·352			
F	1	3	5	0·3882	1·170	1·972	2·818	3·844			
G	1	4	4	0·3819	1·145	1·960	2·771	3·706			
H	1	4	5	0·1971	0·7603	1·529	2·344	3·187	4·184		
I	1	5	5	0·3762	1·131	1·916	2·743	3·552	4·481		
J	2	2	2	0·2138	0·8235	1·669	2·637				
K	2	2	3	0·3919	1·192	2·006	2·968				
L	2	2	4	0·1994	0·7683	1·553	2·406	3·343			
M	2	2	5	0·3801	1·148	1·950	2·857	3·753			
N	2	3	3	0·1951	0·7571	1·525	2·350	3·330			
O	2	3	4	0·3709	1·114	1·871	2·699	3·545			
P	2	3	5	0·1904	0·7340	1·484	2·257	3·093	4·042		
Q	2	4	4	0·1881	0·7219	1·465	2·212	3·040	3·938		
R	2	4	5	0·3606	1·083	1·808	2·587	3·366	4·406		
S	2	5	5	0·1856	0·7155	1·438	2·179	2·915	3·833	4·583	
T	3	3	3	0·3645	1·100	1·850	2·604	3·498			
U	3	3	4	0·1857	0·7154	1·437	2·175	2·967	3·844		
V	3	3	5	0·3542	1·072	1·780	2·542	3·335	4·315		
W	3	4	4	0·3501	1·052	1·760	2·519	3·271	4·092		
X	3	4	5	0·1804	0·6972	1·402	2·098	2·796	3·729	4·472	
Y	3	5	5	0·3461	1·042	1·747	2·456	3·207	4·083	4·862	
Z	4	4	4	0·1775	0·6857	1·391	2·066	2·788	3·590	4·442	
α	4	4	5	0·3418	1·028	1·723	2·384	3·118	3·987	4·695	
β	4	5	5	0·1746	0·6789	1·371	2·036	2·736	3·485	4·270	5·078
γ	5	5	5	0·3373	1·014	1·680	2·297	3·088	3·904	4·625	5·469

Note. The odd differences apply to letter classes in which $S_1+S_2+S_3$ = an odd number.

Table 6

	Sums			Value of χ^2, for two degrees of freedom, equivalent to total difference in N.F.T. of						
	S_1	S_2	S_3	2 or 3	4 or 5	6 or 7	8 or 9	10 or 11	12 or 13	14 or 15
b	0	0	2	1·6220						
c	0	0	3	3·5832						
d	0	0	4	1·2934	6·0892					
e	0	0	5	3·1558	9·6718					
f	0	1	1	1·3864						
g	0	1	2	3·0084						
h	0	1	3	1·0782	4·9706					
i	0	1	4	2·5056	7·4754					
j	0	1	5	1·0112	4·4678	11·0580				
k	0	2	2	1·0466	4·6302					
l	0	2	3	2·3118	6·5916					
m	0	2	4	0·9252	3·8914	9·0976				
n	0	2	5	2·1658	5·9812	12·6850				
o	0	3	3	0·8960	3·7574	8·5528				
p	0	3	4	2·0372	5·5138	11·0580				
q	0	3	5	0·8530	3·5138	7·7758	14·6284			
r	0	4	4	0·8350	3·4022	7·4762	13·5890			
s	0	4	5	1·9300	5·1356	9·9818	17·1368			
t	0	5	5	0·8150	3·3070	7·1814	12·8744	20·8286		
u	1	1	1	2·7726						
v	1	1	2	0·9850	4·3746					
w	1	1	3	2·1974	6·3560					
x	1	1	4	0·8836	3·7312	8·8624				
y	1	1	5	2·0708	5·7824	12·4902				
z	1	2	2	2·1244	6·0166					
A	1	2	3	0·8390	3·4758	7·9778				
B	1	2	4	1·9028	5·1352	10·4840				
C	1	2	5	0·8030	3·2758	7·2986	14·0710			
D	1	3	3	1·8538	4·9696	9·9382				
E	1	3	4	0·7696	3·0996	6·7794	12·4494			
F	1	3	5	1·7700	4·6754	9·0994	16·0328			
G	1	4	4	1·7322	4·5422	8·7706	14·9750			
H	1	4	5	0·7408	2·9496	6·3522	11·3158	18·5232		
I	1	5	5	1·6954	4·4232	8·4548	14·2124	22·1172		
J	2	2	2	0·8192	3·3576	7·6380				
K	2	2	3	1·8058	4·8040	9·6000				
L	2	2	4	0·7536	3·0138	6·5606	12·1068			
M	2	2	5	1·7278	4·5346	8·8212	15·6986			
N	2	3	3	0·7380	2·9400	6·3560	11·5592			
O	2	3	4	1·6556	4·3014	8·2308	14·0710			
P	2	3	5	0·7122	2·8130	5·9960	10·6290	17·6096		
Q	2	4	4	0·7002	2·7480	5·8362	10·2624	16·5880		
R	2	4	5	1·5930	4·0986	7·7398	12·8508	20·1448		
S	2	5	5	0·6884	2·6920	5·6878	9·9152	15·8040	23·8370	
T	3	3	3	1·6216	4·1962	7·9768	13·5188			
U	3	3	4	0·6878	2·6896	5·6882	9·9410	16·0328		
V	3	3	5	1·5622	4·0064	7·5264	12·4394	19·6168		
W	3	4	4	1·5346	3·9180	7·3304	12·0234	18·5588		
X	3	4	5	0·6662	2·5848	5·4194	9·3534	14·6770	22·1072	
Y	3	5	5	1·5078	3·8374	7·1440	11·6384	17·7476	25·7984	
Z	4	4	4	0·6562	2·5360	5·2966	9·1014	14·1642	20·6568	
α	4	4	5	1·4826	3·7588	6·9730	11·2874	16·9368	24·6118	
β	4	5	5	0·6464	2·4902	5·1830	8·8654	13·7372	20·0386	28·3036
γ	5	5	5	1·4584	3·6866	6·8112	10·9780	16·4340	23·1648	31·8864

Note. The odd differences apply to letter classes in which $S_1+S_2+S_3=$ an odd number. For a difference of 0 or 1, $\chi^2=0$.

In Table 5 are given approximate values of the mean normal equivalent abscissa for each total difference in N.F.T. Examples of the use of these values are given in Tables 15 and 16 of my original paper. The values in Table 5 were more carefully calculated than were those of Tables 15 and 16 and do not exactly agree with them. Even those in the first table are approximate only but the errors in them are unbiased and it may be accepted that the sum of several of them would be correct enough for all practical purposes.

It may be noticed that

$$\frac{\bar{x}}{\sigma_{\bar{x}}} = \frac{S(x)}{n} \div \frac{1}{\sqrt{n}} = \frac{S(x)}{\sqrt{n}},$$

and that it is therefore unnecessary to find the mean of a set of normal abscissae when using Table 7. It is only necessary to sum the appropriate entries, giving each its correct sign and to divide the sum by the square root of the number of differences, including zero differences.

The value of $\frac{S(x)}{\sqrt{n}}$ obtained from a series of differences in N.F.T. is used *as the value of x for entering a table of the normal error function, the two* tails of the normal distribution being added together to obtain P. For the convenience of workers who may not have access to such a table I give here a short table, Table 7, which gives certain values of P with the corresponding values of x. Using this table we find that, for the value, $x = 3{\cdot}864$, from Table 15 of my previous paper, P is slightly greater than 0·0001 and that, for the values, $x = 1{\cdot}988$, and $x = 1{\cdot}965$, from the first and second examples in Table 16, respectively, P is slightly less than 0·05.

Table 7. *Value of the normal abscissa for certain values of P*

P	x	P	x
1·00	0·000	10^{-3}	3·291
0·50	0·675	10^{-4}	3·890
0·10	1·645	10^{-5}	4·417
0·05	1·960	10^{-6}	4·893
0·01	2·576	10^{-7}	5·328
—	—	10^{-8}	5·730

Table 6 gives the values of χ^2 for each total difference in N.F.T. This table is used, as already detailed, for testing the significance of a combination of differences without regard to sign, that is to say, the significance of the mean *size* of a difference from zero. Whether the mean *value* of a difference is significant or not, if, for the mean difference without regard to sign, the value of P_{χ^2} is unduly low, it should be assumed that the differences are not such as might be expected often to arise by chance if, in all cases, the water samples were equally polluted by coliform bacteria. The level of significance, $P = 0{\cdot}05$ is generally used for this kind of test.

When using Table 6 the values of χ^2 for all differences in the set examined are added together and the table of χ^2 is entered with number of degrees of freedom equal to twice the number of differences in the set.

The χ^2-table cannot be given conveniently in compressed form and therefore recourse must be had to the tables published in the works of Pearson or of Fisher referred to in my previous paper.

FURTHER USES FOR THE TESTS OF THE MEAN DIFFERENCE, ETC.

In the examples given in my previous paper the observations were already paired by the nature of the research. This is not always convenient or even possible. A test of the significance of the differential effect of Dr Clegg's methods A and B could have been made by treating a large number of sets of tubes according to the method A and another large number of sets according to the method B, all samples having been drawn from the same source. In such a case the correct procedure would be to draw a set of tubes at random from series A and to compare its N.F.T. with that of another set of tubes drawn at random from series B, the process being continued until all tubes have been drawn. The result would not have the same importance as that derived from comparison when many differently polluted waters have been drawn upon, but the estimate of the effect might be expected to be more precise.

Fisher's combined probability test applied by way of Table 6 may be applied to test whether all the members in a set of samples may be regarded as equally polluted. In this case the members are paired at random. If they are sufficient in number the method detailed in Section 3 of my paper becomes applicable. These methods are useful *inter alia* for finding out how large a volume of water or soil taken at a particular place may be regarded as showing equal pollution throughout. If there were 'patchiness' in degree of pollution this would be demonstrated.

(*MS. received for publication* 15. XII. 41—Ed.)