Application of Multiple Range Tests to the Analysis of Fish Catch Data

G.A. Motte

Fisheries and Marine Technology
NOAA Sea Grant
University of Rhode Island
Marine Technical Report 50

APPLICATION OF MULTIPLE RANGE TESTS
TO THE ANALYSIS OF FISH CATCH DATA

University of Rhode Island
Marine Technical Report 50

CONTENTS

TABLES

DIAGRAMS

1. INTRODUCTION

Multiple range testing, used to detect principal contributors toward heterogeneity within a set of treatment means, has long been regarded as a useful and practical experimental tool. As new and different approaches toward solving the multiple range testing problem have evolved, however, so the controversy as to what constitutes the most equitable base test standard has increased. Researchers wish to know which of the multiple range testing techniques is best suited to their particular experimental situation as reflected by the onus placed on the null and alternative hypotheses in question.

In this paper, the "per comparison," "per experiment," and "experimentwise" approaches of Student, Fisher, and Tukey, respectively, are examined and defined. The "improved comparisonwise" approaches of Newman-Keuls and Duncan are contrasted with the three tests mentioned above, and the concept of the sequential approach to multiple range testing is investigated in some detail.

2. APPLICATOON OF FIXED RANGE HYPOTHESES TESTING TECHNIQUES

The basic F-test, following a conventional analysis of variance or analysis of covariance, determines whether there may be considered equality of several treatment means. If the findings of the F-test favor the alternate hypothesis, very little of a specific nature may be determined concerning sources contributing to the decision. It is often desirable to know why the null hypothesis of equal treatment effects has been rejected. Indeed, this may often be the crucial point of the entire analysis, and the contrast examination afforded by multiple range hypotheses testing techniques may be employed to augment the F-test by providing an answer to this question.

The difference in value between the highest and lowest observations, within a set of quantitative data representing the outcome of some experiment, has probably always been the most obvious general measure of the variability of that data when compared with some calculated standard range acceptable for homogeneity. If the observed difference should exceed the standard calculated test range, then the hypothesis of homogeneity for the data examined would be discarded in favor of one of heterogeneity.

The problem of establishing reliable parameters for the standard test range is one that may be interpreted in both a mathematical and heuristic fashion. Some of the resultant tests commonly used by statisticians are compared as to their effect upon the outcome of a hypothetical, but ideal, fishing gear experiment in Example A in this section.

Range tests are designed to test the significance of the range of p out of n ordered means of samples of size N, where p = 2,3 ....., (n-1), n. Typically, one commences by testing the significance of the total range of all n means by comparing it with the critical range designated by the appropriate table cell for the particular level of significance deemed most suitable. If the overall range of all the means is determined as significant, the range number p issdecreased by one, and the range of (n-1) successive means is tested for significance by dropping the largest and then the smallest mean out of the test. It should be noted that there is no priority to the order in which this reduction of the p value is carried out. If the (n-1) successive mean tests prove significant, the p value is again decreased by 1 to (n-2) and the resulting experimental mean range is compared with the appropriate critical range value for detection of significance.

This reduction and testing process continues until a certain range of a subgroup of means is demonstrated to be non-significant, thus indicating that all the means within that subgroup are of homogeneous nature and that no further investigation of subgroups smaller than the value p need be tested within that particular range of means.

For the same base level of significance various multiple range hypotheses testing techniques will yield differing standard critical ranges for the same values of p and n. The difference is mainly a feature of the two separate and definite approaches to the designation of significant levels as p, the number of means contained within a range, increases. Traditionally, a rigid unchanging level of significance is assigned to all values of p, (1), (7), and (10) but this is considered by some as a rather conservative approach and so the more modern, though as yet mathematically inexact, approach favors a more liberal changing scale of significance levels based on the value of p, (2), (3), and (9). Some of the more popular fixed range tests are defined in this section, whereas the sequential testing approach is discussed in the following section.

In the 1935 first edition of his universally accepted treatise on "Design of Experiments," Fisher expressed the idea that it would be unwise to apply a fixed range test of the multiple "t" type if the overall between-treatment F-test indicated homogeneity. If the F-test indicates overall non-significance, then all sub-differences contained in the overall range are to be declared not significant.

Of the range tests that employ a single fixed test standard (against which all possible pairwise ranges are tested for significance), perhaps the L.S.D.,i.e., "Least Significant Difference" test, together with the somewhat modified F.S.D. and T.S.D. tests established by Fisher and Tukey, respectively, are the ones featured most prominently in available literature on the subject of range testing, (1), (2), (4), (7), (9), (11), and (14). These three fixed tests designate a standard difference between any two out of n means of samples of size N which, when exceeded, indicates significant difference at the particular α level (for Type 1 error of erroneously accepting alternate hypothesis) upon which the test is based, Diagram 1. These pairwise standard ranges for the three fixed tests mentioned above are as defined in the following:

1. "Student" t, L.S.D. $= t(\alpha, v)S_{\bar{d}} = q(\alpha, 2, v)S_{\bar{x}}$ (1)

Where v is the number of degrees of freedom for the error, t (α, v) is the two tailed level of Student's t with v degrees of freedom

$S_{\bar{x}}$ is the standard error of the mean, $S_{\bar{x}} = (S^2/N)^{1/2}$

$S_{\bar{d}}$ is the standard error of the difference between the means,

$$S_{\bar{d}} = (S^2/N)^{1/2} = \sqrt{2}S_{\bar{x}}$$

q $(\alpha, 2, v)$ is the upper α point of the studentised range of 2 observations with v degrees of freedom for S.

2. Fisher's Test, F.S.D. $= q(\alpha/nC2,2,v)\ S_{\bar{x}}$ (2)

In the F.S.D. test, the α level of significance for type 1 error that was assigned to each pairwise range test in the L.S.D. test is distributed equally between the nC2 pairwise comparisons of the F.S.D. test to give an overall α level. The L.S.D. approach to the application of the α level of significance may be referred to as the "per comparison" approach, while the division of α between the individual comparisons featured by the F.S.D. test may be termed the "per experiment" approach.

3. Tukey's Test, T.S.D. = $q(\alpha, n, v)S_{\bar{x}}$ (3)

where n = total number of treatment means in the range. In this case the α level of significance is awarded at the n level overall range, and is effectively reduced as the number of means tested for significance of range is decreased. This approach is generally regarded as unnecessarily conservative and would appear to give an unrealistically expanded zone of acceptance within individual subranges which could be primary contributors to heterogeneity. The Tukey method of awarding α is supposedly sensitive to the overall experiment and is therefore sometimes referred to as the "experimentwise" approach to Type 1 error designation.

Thus, for the three fixed range tests examined, there are three vastly different modes of dispersing the base level of significance which results in three somewhat different standard test ranges for the same α significance of the test. By far the most simple approach to the problem of multiple range testing is effected by extending the symmetric two-tail t-test (commonly used to decide the equality, superiority, or inferiority of one treatment to another in the two-treatment t case, Diagram 1) to the multiple treatment problem of answering the same question for each and every pair of treatments concerned. This α level multiple t rule, as this simultaneous multiple t-testing procedure is often called, isolates each pair of treatments for a marginal test. The fact that the difference between each pair is compared with the same level t standard, regardless of how many treatment means intervene in the range of ordered means, has caused a certain amount of intuitive dissatisfaction with the so-called "α level multiple t range test" and has resulted in the establishment of various alternatives.

The importance of deciding upon a realistic value for the level of significance which is sensitive to the nature of the experiment is discussed further in the next section as rationalized by the sequential approach to range testing, but the following three definitions by W.T. Federer, (7), appear to be relevant at this stage.

1) Error rate per comparison = $\dfrac{\text{number of erroneous inferences}}{\text{number of inferences attempted}}$ which is the proportion of all comparison decisions expected to be erroneous when the null hypothesis is true.

2) Error rate per experiment = $\dfrac{\text{number of erroneous inferences}}{\text{number of experiments}}$

which is the expected number of erroneous statements per experiment when the null hypothesis is true.

3) Experimentwise error rate = $\dfrac{\text{number of experiments with one or more erroneous inferences}}{\text{number of experiments}}$

which is the expected proportion of experiments with one or more erroneous statement, when the null hypothesis is true.

Furthermore, to avoid confusion when referring to error and power of tests, the following convention is adopted for use herein:

PROTECTION = $\dfrac{\text{No. of correct decisions made under the Null Hypothesis} \times 100}{\text{Total no. of decisions under the Null Hypothesis}}$

TYPE 1 ERROR (co-protection) = $\dfrac{\text{No. of incorrect decisions made under the Null Hypothesis} \times 100}{\text{Total no. of decisions under the Null Hypothesis}}$

POWER = $\dfrac{\text{No. of correct decisions made under the Alternate Hypothesis} \times 100}{\text{Total no. of decisions under the Alternate Hypothesis}}$

TYPE 2 ERROR (co-power) = $\dfrac{\text{No. of incorrect decisions made under the Alternate Hypothesis} \times 100}{\text{Total no. of decisions under the Alternate Hypothesis}}$

The relationship between these four test factors is most readily appreciated in its most basic form with just two means considered. The effect on each of the factors caused by an alteration in the level of significance can easily be construed from Diagram 1.

The three fixed range tests by Student, Fisher, and Tukey provide standard significant differences, and so the logical method of comparison would seem to be through these standard differences for a given base level of significance. This is attempted by means of the following example with an individual confidence interval worked for a per comparison, per experiment and experimentwise approach, using the L.S.D., F.S.D., and T.S.D. tests, respectively.

EXAMPLE A

A hundred fishing vessels were divided randomly into five groups of twenty vessels each. Five trawl nets of different design were to evaluated for fishing efficiency as reflected by catching rate and 20 trawls of one design provided for each group of fishing vessels, one for each boat.

If a convenient standard error of the sample mean of unity is adopted, the above test standards are to be compared with all pairwise ranges and if exceeded, will indicate heterogeneity between the particular pair of treatments, (in this case trawl nets) tested. Thus the pairwise 95 percent confidence interval not to be exceeded, if the null hypothesis of homogeneity for a given pair of treatments is to be accepted, will vary from a low of ± 2.82 for the L.S.D. test, through ± 3.94 for the T.S.D. test to a high of ± 4.05 for the F.S.D. test. If the base level of significance is lowered to .01 to give a 99 percent confidence interval, the resulting test ranges for the example will be extended to a low of ± 3.72 for the L.S.D. test, through ± 4.74 for the T.S.D. test, to a high of ± 4.80 for the F.S.D. test. The standard test values for t or q at various degrees of freedom and levels of significance are contained in refs. (12) and (13) or any complete text on experimental statistics such as ref. (6). Approximate values may be obtained from interpolation of Table 4.

The difference in the values of the test ranges is a direct function of the manner in which the basic error rate is dispersed, and the approaches shown here clearly display this difference. However, a base level of significance of .5 could be selected for the per experiment approach, and therefore equation (2) would yield:

$$\text{F.S.D.} = \pm\, q(^{.5}/5C2, 2, .95)\, S_{\bar{x}} = \pm\, t(.05, .95)\sqrt{2}\, S_{\bar{x}}$$

and so the confidence interval resulting would have been the same as for a per comparison test, using the L.S.D. test with a .05 level of significance but the two adopted levels of significance would have been vastly different. Thus the difference between the per comparison and per experiment test range is entirely due to the manner in which the adopted level of significance is applied.

The difference between the per comparison approach of the L.S.D. test and the experimentwise approach of the T.S.D. test is due to the adoption of the α level of significance at the $p = 2$ level in the L.S.D.

Thus total degrees of freedom = $nN - 1 = (5.20) - 1 = 99$
Treatment degrees of freedom between nets = $n - 1 = 5 - 1 = 4$
Error degrees of freedom between boats = $nK - n = (5.20) - 5 = 95$

A preliminary F-test indicated a very significant difference between nets or treatments. The ordered sample means were = $\bar{x}_a = 8.27$, $\bar{x}_b = 11.74$, $\bar{x}_c = 13.06$, $\bar{x}_d = 17.10$, $\bar{x}_e = 23.60$

(a) Error rate base per comparison using L.S.D. for a 95 percent confidence interval, equation (1)

$$\bar{x}_i - \bar{x}_j = \pm q(\alpha, 2, v)\, S_{\bar{x}}$$

$$\bar{x}_i - \bar{x}_j = \pm q(.05, 2, 95) S_{\bar{x}}$$

$$\bar{x}_i - \bar{x}_j = \pm t(.05, 95)\, \sqrt{2}\, S_{\bar{x}}$$

$$\bar{x}_i - \bar{x}_j = \pm 2.00(1.414) S_{\bar{x}}$$

$$\bar{x}_i - \bar{x}_j = \pm 2.82\, S_{\bar{x}}$$

(b) Error rate base per experiment using F.S.D. for 95 percent confidence interval, equation (2)

$$\bar{x}_i - \bar{x}_j = \pm q(\alpha/nC2, 2, v) S_{\bar{x}}$$

$$\bar{x}_i - \bar{x}_j = \pm q(\alpha/5C2, 2, 95) S_{\bar{x}}$$

$$\bar{x}_i - \bar{x}_j = \pm q(\alpha/10, 2, 95) S_{\bar{x}}$$

$$\bar{x}_i - \bar{x}_j = \pm t(.005, 95)\, \sqrt{2}\, S_{\bar{x}}$$

$$\bar{x}_i - \bar{x}_j = \pm 2.95(1.414) S_{\bar{x}}$$

$$\bar{x}_i - \bar{x}_j = \pm 4.05\, S_{\bar{x}}$$

(c) Experimentwise error rate base using T.S.D. for 95 percent confidence interval, equation (3)

$$\bar{x}_i - \bar{x}_j = \pm q(\alpha, n, v) S_{\bar{x}}$$

$$\bar{x}_i - \bar{x}_j = \pm q(.05, 5, 95) S_{\bar{x}}$$

$$\bar{x}_i - \bar{x}_j = \pm 3.94 S_{\bar{x}}$$

test and the $p = n$ level in the T.S.D. test. The difference between the per experiment and experimentwise approaches of the F.S.D. and T.S.D. range tests is also due to the manner in which the level of significance is set and dispersed. These two tests are compared in the most directly contrasting way possible by awarding the per comparison level of significance at the common .05 level for a two-fold L.S.D. test and noting the resultant per experiment levels of significance necessary to effect this at higher levels under the F.S.D. test. These results are shown in Table 1 together with the resultant levels of significance caused by assuming a .05 level for the experimentwise T.S.D. test, thus contrasting the resulting intermediate levels of significance and test protection levels due to the most liberal and conservative designation of the common .05 level of significance from the F.S.D. and T.S.D. tests respectively.

n for F.S.D. p for T.S.D.		2	3	4	5	10	20	30
F	α_n	.05	.122	.203	.286	.627	.918	.984
S	P_n	.95	.878	.797	.714	.373	.082	.016
D								
T	α_p	.0002	.0005	.001	.002	.007	.025	.05
S	P_p	.9993	.9995	.999	.998	.993	.975	.95
D								

Table 1 Resultant Per Experiment Error Rate $\alpha_n \alpha_p$ and Protection Levels P_n, P_p From A .05 Level L.S.D. and T.S.D. Test, Respectively

The vase difference between the three approaches may be clearly discerned from Table 1, where for the parameters adopted it requires a per comparison level of only .0002 under the T.S.D. test to give a resultant experimentwise error of .05. It is this reduction of the level of significance at lower levels that suggests apparent conservatism as a criticism of the Tukey T.S.D. test, but judicial adjustment of the overall experimentwise error rate can result in a more reasonable set of values for lower levels of p. This principle is also applicable when comparing the F.S.D. and L.S.D. or per experiment to per comparison approaches, as previously indicated.

3. APPLICATION OF SEQUENTIAL RANGE HYPOTHESES TESTING TECHNIQUES

It can readily be determined from Table 1 that the difference between the levels of significance by the per comparison and experimentwise approach to range testing increases as the number of treatment means involved in an experiment increases. This feature has been strongly condemned by advocates of the sequential approach to multiple range testing, who argue that an acceptable level of significance for the experiment by the experimentwise approach often renders the resultant level of significance for comparisons unrealistically small, and thus prone to Type 2 errors, Diagram 1. Furthermore, an acceptable level of significance for comparisons by the per comparison approach often renders the overall experimentwise level of significance too large to be of practical use. Hence this unacceptable relationship between the per comparison and experimentwise error rates leads to the foundation of sequential tests providing levels of significance following some "middle ground" between the per comparison and experimentwise levels, being sensitive to the number of treatment means involved in a given test.

The New Multiple Range Test by D.B. Duncan, (2), (3), (4), and (5), attempts to utilize some of the more favorable characteristics of the aforementioned fixed range tests while employing a special protection levels system, based on degrees of freedom. This protection is achieved by considering every possible alternative to the hypothesis of homogeneity as an individual case, defined by its own power characteristics. For example, the most basic test of homogeneity is when p = 2 and just two sample populations are to be equated. Duncan's Multiple Range Test considers each of the three decisions possible:

1) $\bar{x}_1$ and $\bar{x}_2$ are not significantly different

2) $\bar{x}_1$ is significantly smaller than $\bar{x}_2$

3) $\bar{x}_2$ is significantly smaller than $\bar{x}_1$

Whereas some testing procedures consider 2) and 3) above grouped to form a two decision test:

1) $\bar{x}_1$ and $\bar{x}_2$ are not significantly different

2) $\bar{x}_1$ and $\bar{x}_2$ are significantly different

As Duncan stated, (3), "function which combines probabilities of correct decisions with probabilities of serious errors in this way is of not value in measuring desirable or undesirable properties."

If the number of means featured in an experiment is increased to three or more, the number of possible alternatives to the hypothesis of homogeneity rises sharply. For example, when three means are involved, there are 19 possible decisions that could be made when investigating the ranking status of those means: when n = 6, there are 57 possible decisions. A detailed treatment of the power structure and definition of the alternative situations for n > 2 is provided in refs. (3) and (11).

The effective difference between the most commonly accepted and used multiple range tests is a function of the method employed to arrive at the collective significance level set for each subrange of p means within the total set of n means involved. For example, if the protection levels of the two mean

tests for each of the Multiple t, Newman Keuls (7) and (8) and Duncan's New Multiple Range Test are all set at 95 percent then, because of the different considerations used in allowing for increase in alternatives as the number of means tested together increases, the resultant protection levels at higher ranges will vary accordingly. For the total null hypothesis concerning three means, that is, when n = 3, the resultant protection levels for the three tests mentioned are: 87.8 percent for the Multiple t-test, Table 1, 95 percent for the Newman-Keuls Test and 90.25 percent for Duncan's New Multiple Range Test, (12) and (13).

The differences between these assigned protection levels, upon which the test parameters are based, are a result of the particular type of protection sought by the persons responsible for constructing these tests. The relatively high level of significance in the case of just three means for the Multiple t-Test where α = 12.2 would seem in many cases to present an unrealistic risk. In this case, the chance of rejecting the Null Hypothesis, when it is indeed true, could be as high as 12.2 percent. However, this test does provide good protection against Type 2 errors, and for this reason (when optimization of the alternative situation is sought) may in some circumstances provide the most attractive test conditions.

To the other extreme, the Newman-Keuls Test maintains a constant protection level with a relatively low level of significance of five percent, thus allowing a greater probability for Type 2 errors, Diagram 1. For this reason the Newman-Keuls Test could be considered generally insensitive to the increasing number of alternatives to the Null Hypothesis as the number of means involved increases. However, this test base may be attractive for circumstances that render a conservative approach to the Null Hypothesis a desirable feature. The New Multiple Range Test by Duncan establishes standard conditions with the intention of the Null Hypothesis being protected, and yet remaining sensitive in the ability to detect Type 2 errors at each level of p from 1 to n.

The apparent insensitivity of fixed range tests at intermediate values for p between 1 and n is countered in two different ways by the Newman-Keuls and Duncan sequential range tests. The Newman-Keuls or N.K.M. standard test range for p out of n ordered means of samples each of size N is based on the studentised range of p observations instead of on the studentised range of n observations as in the T.S.D. test. Thus when p = n the test standards of the N.K.M. and the T.S.D. are the same, and when p = 2 the test standards of the N.K.M. and the L.S.D. are the same. For other values of p between 2 and n the resultant N.K.M. test standard will be established at some intermediate level between the fixed test standards of the L.S.D. and T.S.D. tests. The p level choice for the sequential range test by Newman-Keuls is as follows:

$$\text{N.K.M.}: \quad \alpha_p = \alpha \qquad (4)$$

The New Multiple Range Test proposed by Duncan approaches the problem in a similar sequential manner, but based upon the number of degrees of freedom available. Duncan reasons that one has (p - 1) degrees of freedom for testing p means and it is therefore possible to make (p - 1) independent tests, each with its own (1-α) protection level. Thus the joint protection level for the range of p means is $(1-\alpha)^{p-1}$, which is the probability of finding no significant differences in making (p-1) independent tests (each with its own level of significance) under the hypothesis that all p population means are equal. The p level choice for the sequential range test by Duncan is as follows:

D.N.M.: $\alpha_p = 1 - (1 - \alpha)^{p-1}$

Thus the difference between the sequential multiple range tests of Newman-Keuls and Duncan lies in the choice of α_p, p = 2,3,..., (n - 1), (n).

These two sequential range tests are compared and contrasted with the three fixed range tests, (previously defined in Section 2) in the following section.

4. COMPARISON OF MULTIPLE RANGE HYPOTHESES TESTING TECHNIQUES

The resultant factors by which $S_{\bar{x}}$ must be multipled to obtain the critical test ranges for the fixed and sequential methods discussed in the previous two sections are contained in Table 2. A convenient $S_{\bar{x}}$ of unity has been assumed for the data contained in example A in order to afford direct comparison of the factors shown in the Table with the ranges between the sample means of the example.

TEST STANDARDS for ν= 95, p = 2,3,4,5,$S_{\bar{x}}$ = 1.0

TEST CODE	p = 2	p = 3	p = 4	p = 5
L.S.D. α = .05	2.81	2.81	2.81	2.81
D.N.M. α = .05	2.81	2.96	3.05	3.13
L.S.D. α = .01	3.73	3.73	3.73	3.73
N.K.M. α = .05	2.81	3.37	3.70	3.94
T.S.D. α = .05	3.95	3.95	3.95	3.95
F.S.D. α = .05	4.05	4.05	4.05	4.05
D.N.M. α = .01	3.73	3.89	3.99	4.07
N.K.M. α = .01	3.72	4.22	4.53	4.74
T.S.D. α = .01	4.76	4.76	4.76	4.76
F.S.D. α = .01	4.80	4.80	4.80	4.80

TABLE 2 TEST STANDARDS FOR FIVE MULTIPLE RANGE TESTS ON EXAMPLE A

The ten different test standards, for the five different approaches to the multiple range testing problem at two base levels of significance, produce the inferences suggested by Table 3 when applied to the data in Example A. The various test standards are compared to the ordered sample mean ranges and non-significant ranges due to each test are underlined. Hence, for each test procedure, only those means which are not connected by an unbroken line in Table 3 are judged to be significantly different.

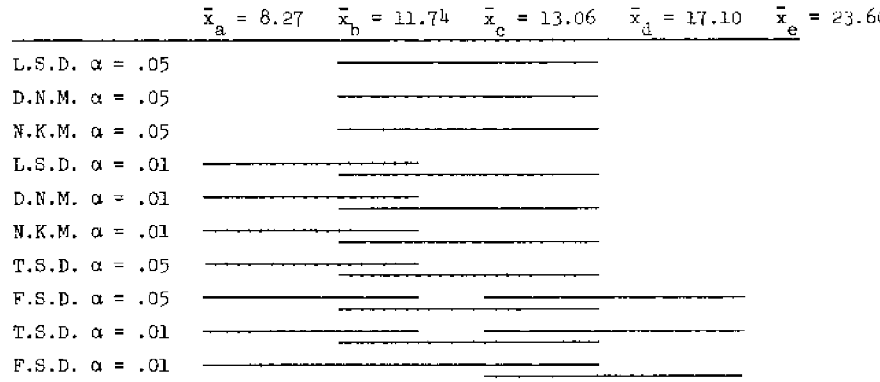

TABLE 3 NONSIGNIFICANT RANGES FOR TEN RANGE TESTS ON EXAMPLE A

The fundamental principles of range testing nullify the range testing of observed subranges contained within a larger nonsignificant range. Thus the comparisons between the observed and test standard ranges are carried out in a routine and methodical manner commencing with the largest value of p (p = 5 in Example A) and reducing the p factor by 1 until p = 2 or until a range of nonsignificance is reached. The procedure is somewhat quicker for the fixed range tests which require only one test standard to be computed for comparison with the observed ranges at all levels of p.

The test standards presented in Table 2 are reproduced in graph form in Diagram 2 in an attempt to illustrate the trend of sequential range tests in a continous manner. The test values appear on the y axis and may be selected for the corresponding values of p which constitute the x axis.

5. CONCLUSION

From the preceding Table and Diagram it may be deduced that for the ten test standards computed, four different situations of homogeneity were demonstrated. These situations vary from the rather liberal approach to protection of the homogeneity hypothesis of the L.S.D., α = .05 test to the more conservative approach of the F.S.D., α = .01 test. That is from just ($\bar{x}_b = \bar{x}_c$) for the L.S.D., α = .05 test to ($\bar{x}_a = \bar{x}_b = \bar{x}_c$ and $\bar{x}_c = \bar{x}_d$) at the α = .01 level for the F.S.D. test.

However, it is clearly illustrated by Table 3 that the trawl nets of d and e design were proven significantly more efficient catchers than the other 3 nets (for the sample trials of Example A) by seven out of the ten range tests. Furthermore, the e type trawl net was proven significantly better than all the other net designs by all ten of the multiple range tests applied. In experiments of this nature, it would appear that selection of a single range test standard is best governed by the physical considerations of the particular experiment. If the primary concern is for protection of the null hypothesis of homogeneity, a conservative approach is dictated, such as afforded by a low level F.S.D. of T.S.D. test. Should the experimental emphasis be towards optimization of the power of the alternative hypothesis then the higher level, less protective approach of the L.S.D. test can be applied. From Diagram 2 it can be seen that judicial selection of the test level will provide a range standard geared towards a less biased approach, and certainly the data can easily be subjected to two or three fairly representative multiple range test comparisons before conclusive results on the various hypotheses are finalized.

Table 4 DISTRIBUTION OF THE STUDENT t VALUES BY THE PROBABILITY AND THE NUMBER OF "DEGREES OF FREEDOM" (Hoel, 1947)[1]

Number of "Degrees of Freedom"	Probability to become larger than t					
	0.005	0.01	0.025	0.05	0.1	0.15
1	63.657	31.821	12.706	6.314	3.078	1.963
2	9.925	6.965	4.303	2.920	1.886	1.386
3	5.841	4.541	3.182	2.353	1.638	1.250
4	4.604	3.747	2.776	2.132	1.533	1.190
5	4.032	3.365	2.571	2.015	1.476	1.156
6	3.707	3.143	2.447	1.943	1.440	1.134
7	3.499	2.998	2.365	1.895	1.415	1.119
8	3.355	2.896	2.306	1.860	1.397	1.108
9	3.250	2.821	2.262	1.833	1.383	1.100
10	3.169	2.764	2.228	1.812	1.372	1.093
11	3.106	2.718	2.201	1.796	1.363	1.088
12	3.055	2.681	2.179	1.782	1.356	1.083
13	3.012	2.650	2.160	1.771	1.350	1.079
14	2.977	2.624	2.145	1.761	1.345	1.076
15	2.947	2.602	2.131	1.753	1.341	1.074
16	2.921	2.583	2.120	1.746	1.337	1.071
17	2.898	2.567	2.110	1.740	1.333	1.069
18	2.878	2.552	2.101	1.734	1.330	1.067
19	2.861	2.539	2.093	1.729	1.328	1.066
20	2.845	2.528	2.086	1.725	1.325	1.064
21	2.831	2.518	2.080	1.721	1.323	1.063
22	2.819	2.508	2.074	1.717	1.321	1.061
23	2.807	2.500	2.069	1.714	1.319	1.060
24	2.797	2.492	2.064	1.711	1.318	1.059
25	2.787	2.485	2.060	1.708	1.316	1.058
26	2.779	2.479	2.056	1.706	1.315	1.058
27	2.771	2.473	2.052	1.703	1.314	1.057
28	2.763	2.467	2.048	1.701	1.313	1.056
29	2.756	2.462	2.045	1.699	1.311	1.055
30	2.750	2.457	2.042	1.697	1.310	1.055
∞	2.576	2.326	1.960	1.645	1.282	1.036

Note: When the absolute value is used, the probability must be doubled.

[1] Paul G. Hoel; Introduction to Mathematical Statistics, John Wiley and Sons, Inc., 1947.

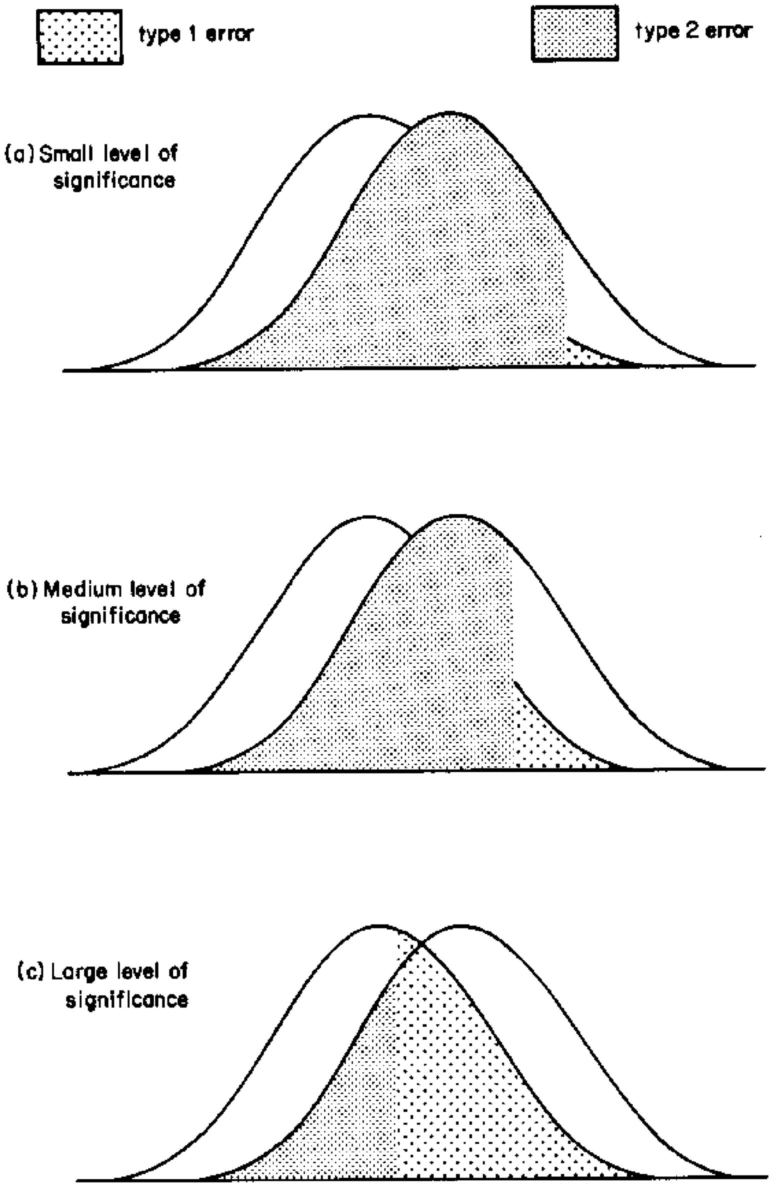

DIAGRAM 1 TYPE I AND TYPE II ERROR RELATIONSHIP FOR A TWO DECISION TEST

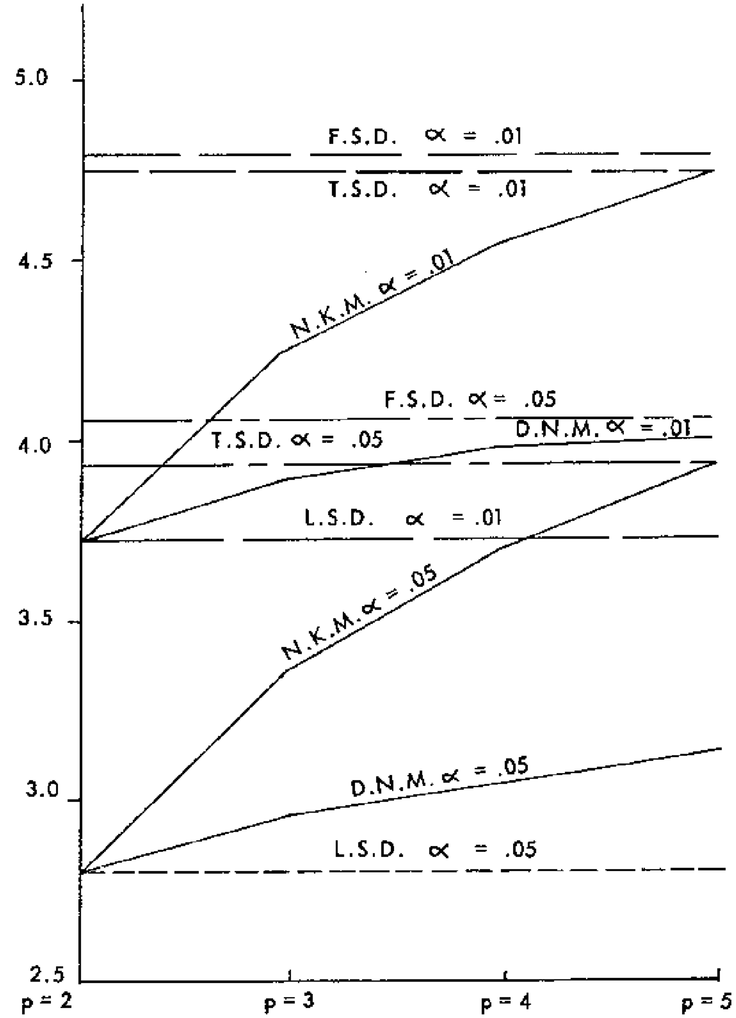

DIAGRAM 2 COMPARISON OF TEST STANDARDS FOR TRAWL NET EXAMPLE A

6. BIBLIOGRAPHY

1 Cox D.R. (1965) "A Remark on Multiple Comparison Methods," Technometrics 7.

2 Duncan D.B. (1952) "On the Properties of the Multiple Comparisons Test." Virginia J. Sci., 3.

3 Duncan D.B. (1955) "Multiple Range and Multiple F Tests," Biometrics 11.

4 Duncan D.B. (1961) "Bayes Rules for a Common Multiple Comparisons Problem and Related Student t Problems," Annals Math. Stat. 32.

5 Duncan D.B. (1965) "A Bayesian Approach to Multiple Comparisons", Technometrics 7.

6 Federer, W.T. (1955) "Experimental Design," Macmillan, N.Y.

7 Federer, W.T. (1961) "Experimental Error Rates,: Proc. AM. Soc. Hort. Sci.

8 Fryer H.C. (1968) "Concepts and Methods of Experimental Statistics," Allyn and Bacon, Inc.

9 Kramer J. (1957) "Extension of Multiple Range Tests to Group Correlated **Adjusted** Means," Biometrics 13.

10 Lord E. (1947) "The Use of Range in Place of Standard Deviation in the t Test." Biometrika 34.

11 Motte G.A. (1972) "A Survey of the Development of Multiple Range Hypothesis Testing Techniques," University of Rhode Island, U.S.A.

12 Nicholson (1943) "Statistical Tables"

13 Pearson E.S. and Hartley H.O. (1954) "Biometrika Tables for Statisticians, Vol. 1."

14 Scheffe H. (1959) "The Analysis of Variance," John Wiley, N.Y.