BEHRENS-FISHER PROBLEM

Qureshi, Z. H M.Sc.

Mathematics Department

ABSTRACT

The relative merits of different statistics available for the classical Behrens-Fisher problem are considered in this thesis. The two means problem is treated from significance test and confidence interval aspects. Looking at the problem from significance test approach, various methods are applied to approximate the true unknown distribution of Behrens-Fisher statistic. The general case of testing several population mean values is taken into account and an approximate test, based on F-distribution, is constructed which has more practical usefulness.

Confidence intervals, for the difference in population means, are set up in terms of the sample values and their optimality, under certain conditions, is shown.

Bayes' solution of the problem is also considered under the provision of a priori knowledge for population variances. An approximation, based on the existing Student t-tables, is given which seems to be adequate for the routine tests for practical research workers.

BEHRENS -FISHER PROBLEM

by

Zia-ul-Haq Qureshi

A thesis submitted to the Faculty of

Graduate Studies and Research in partial fulfilment

of the requirements for the degree of

Master of Science.

Department of Mathematics,

McGill University,

Montreal, P. Q.

April, 1968.

ACKNOWLEDGEMENT

I wish to express my sincere gratitude and deep appreciation to Professor Mathai Arakaparampil Mathai for the encouragement and direction in writing this thesis.

PREFACE

The historical well known Behrens-Fisher problem is simple to put but difficult to answer. The available immense literature shows that it has often been discussed.

I have gone through the different research papers discussing its various aspects. This study reflects the way I have looked into the problem. This work is of expository nature. I have classified the problem and considered it from various approaches.

From significance tests view point, some methods, already used by various authors, are applied to approximate the unknown distribution of Behrens-Fisher statistic. Their theoretical limitations and practical utility have been pointed out. Comparisons of some test procedures, under the given conditions, are made in Chapter 2, to know the advantages of one test procedure over that of the other.

The confidence intervals, for the difference between the population means, have been set up by numerous authors. These are discussed in chapter 3. Two tests, given in chapter 3, are compared on the basis of their expected lengths of confidence intervals to arrive at a criterion, from which it may be possible to study the approximate relative test efficiency.

Fiducial approach to solve the problem is considered within the frame-work of group transformation model. The frequency interpretation of fiducial probability is made in chater 4. An approximate test for practising biometricians is considered which can be useful for drawing the scientific inferences.

INDEX OF SYMBOLS

Symbol	Defined as	Page
$\overset{k}{\underset{i=1}{S}} a_i$	$a_1 + a_2 + \ldots + a_k$	4
Σ_i	$\overset{n_i}{\underset{j=1}{S}} (x_{ij} - \bar{x}_i)^2, (i=1,2;\ j=1,2,\ldots,n_i)$	4
d.f	Degrees of freedom	4
Pr(•)or P(•)	Probability of the event respresented within br-ackets	10
p.d.f	Probability density function	16
L.T	Laplace Transform	16
$\Longrightarrow$	Logical Implication	31
$Z \in N(0,1)$	Z follows a standard normal distribution	92

CONTENTS

CHAPTER 0

SUMMARY:

This thesis treats the Behrens-Fisher problem from the stand-point of tests of significance and confidence intervals. It is assumed that the simple random samples of different sizes are drawn from independent normal populations. All parameters, involved, are supposed to be unknown. The question is posed: Are the data consistent with the hypothesis that population means are equal or differ by a given constant?

The answer to this is put forward by Behrens (1929) and latter Fisher (1935) gave a test based on the fiducial distribution. The justification of the use of this test is considered under the framework of group transformation, within which, the fiducial probability has a frequency interpretation.

In the situation when variance ratio is unknown, the usual procedure is to approximate with the standard normal distribution when the samples are large, but for the small sample sizes this approximation fails. The approximate distribution of two statistics, given in § 1. , is obtained in chapter 1, from which it is possible to study how far a Student t - test is valid when sample sizes are unequal and small. For the case when sample sizes are odd an exact distribution is obtained in § 1.2. Finally, an approximation procedure to solve the problem of comparing several mean values

is considered.

In chapter 2, two types of statistics are taken into account. One called unilateral, which controls the type I error, if it is known a priori that the variance of one specified population is greater than that of the other. The other is called bilateral, which controls the size of the test when there is no a priori knowledge of the population variances. Some procedures are also considered in which a preliminary test on the observed data is performed to ascertain whether the population variances may be regarded as equal or not. On the basis of the outcome of the preliminary test, one proceeds to test for the equality of population means.

Chapter 3 examines the problem of estimating the difference between the population means from the confidence interval point of view. The general case, when sample sizes are small and unequal, is discussed. An approximate confidence interval for a linear function of the population means (with known coefficients) is constructed in terms of sample estimates. The power function of the test in §3.1 is compared with the power function of the corresponding most powerful test in which variance ratio is assumed to be known.

The logical requirements for the fiducial method of inference, are considered briefly in chapter 4. The justification of Fisher (1935) test, based on fiducial distribution, is discussed within the group transformation model. Bayes' approach to solve the problem is also considered when somehow a priori distribution of population variances can be specified.

CHAPTER 1

DISTRIBUTIONS

1. **Introduction:**

Let X_1 and X_2 be independent normal random variables with means μ_1 and μ_2 and variances σ_1^2 and σ_2^2 respectively. Samples of sizes n_1 and n_2, drawn from the corresponding populations, are denoted by $x_{ij}(i=1,2;\ j=1,2,\ldots,n_i)$. The sample means and variances are

$$\bar{x}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} x_{ij} \ ; \quad s_i^2 = \Sigma_i/(n_i-1) \ ,$$

where $\Sigma_i = \sum_{j=1}^{n_i} (x_{ij}-\bar{x}_i)^2$, $(i=1,2\ ;\ j=1,2,\ldots,n_i)$.

It is required to test the hypothesis H_o: $\mu_1=\mu_2$. Under H_o, two obvious cases are (i) $\sigma_1^2 = \sigma_2^2$ and (ii) $\sigma_1^2 \neq \sigma_2^2$. In case (i) the most appropriate test is made by identifying

$$U = (\bar{x}_1-\bar{x}_2) \Big/ \left[\frac{\Sigma_1 + \Sigma_2}{(n_1+n_2-2)} \left(\frac{1}{n_1} + \frac{1}{n_2}\right)\right]^{\frac{1}{2}}$$

with a Student t - distribution with $f=(n_1+n_2-2)$ degrees of freedom (d.f). In the second case, if $R = \sigma_1^2/\sigma_2^2$ is known, the statistic

$$U_R^2 = (\bar{x}_1-\bar{x}_2)^2 \Big/ \left[\frac{\Sigma_1/R + \Sigma_2}{n_1+n_2-2}\,(R/n_1 + \frac{1}{n_2})\right] ,$$

can be used to test H_o. When $\mu_1=\mu_2$, U_R^2 is distributed as $t^2_{n_1+n_2-2}$, where $t_{n_1+n_2-2}$ is a Student t - distribution with (n_1+n_2-2) d.f.

If, however R is unknown, an alternative criterion often employed is

$$V^2 = \frac{(\bar{x}_1-\bar{x}_2)^2}{\left(\frac{\Sigma_1}{n_1(n_1-1)} + \frac{\Sigma_2}{n_2(n_2-1)}\right)} .$$

The statistic V follows an approximate standard normal distribution if sample sizes are large. But for small samples, this test is not appropriate. When $n_1=n_2$, U and V are identical. The validity, in the sense of controlling type I error satisfactorily, of U(when $\sigma_1^2 \neq \sigma_2^2$) and referring V to a t - distribution with $f=(n_1+n_2-2)$ d.f is investigated.

It is obvious that U in general, is not distributed as a t - distribution. The variance of difference of sample means is $(\sigma_1^2/n_1 + \sigma_2^2/n_2)$, and S^2 is an unbiased estimate of σ^2 when $\sigma_1^2 = \sigma_2^2$, where $S^2 = \dfrac{\Sigma_1 + \Sigma_2}{n_1+n_2-2}$, and $\sigma_1^2 = \sigma_2^2 = \sigma^2$. The statistic V does not suffer from this restriction but its distribution is also not independent of R.

Other criteria of the form $(\bar{x}_1-\bar{x}_2)/\sqrt{d\ \Sigma_1 + e\ \Sigma_2}$, where d and e are some positive constants, may be more appropriate than V. For instance, if n_1 and n_2 are both greater than 3, we might expect

$$Z = (\bar{x}_1-\bar{x}_2)/\sqrt{\frac{\Sigma_1}{n_1(n_1-3)} + \frac{\Sigma_2}{n_2(n_2-3)}}$$

to be such a criterion. The reason for these particular values of d and e is that they give to σ_z^2 the same value when $R \to 0$ or when $R \to \infty$, which means that the probability of rejection, under H_0, departs from a preassigned value, less for Z than for either of the criteria U and V.

1.0 Approximate Distributions of U and V:

Under H_0, we may write

$$\frac{(\bar{x}_1-\bar{x}_2)^2}{\sigma_1^2/n_1+\sigma_2^2/n_2} = \chi'^2 \ ; \quad \Sigma_1/\sigma_1^2 = \chi_1^2 \ ; \quad \Sigma_2/\sigma_2^2 = \chi_2^2 ,$$

where χ'^2, χ_1^2 and χ_2^2 are independently distributed as χ^2 with 1, (n_1-1) and (n_2-1) degrees of freedom respectively. Both U and V can be expressed in the form $Y(a,b) = \chi'/\sqrt{W}$, where $W = a\,\chi_1^2 + b\,\chi_2^2$ and a, b are some positive constants depending upon the sample sizes and the two variances, W is always distributed independently of χ'. When a=b er when either a or b is zero, W is distributed as χ^2 multiplied by some constant. In these cases the distribution of Y(a,b) will be some constant multiple of t. For other values of a and b the distribution of Y is obtained by approximating the distribution of W by a Pearson Type III($=\chi^2$) Curve. This approximation can also be made by representing U(and V) by a Pearson Type VII(=Student) curve after correcting its second and fourth moments.

The probability law of W, as approximated by χ^2 is given by Welch(1936). His method of approximating the distribution of Y(a,b) is as follows.

The probability function of a type III curve is

$$p(W) = \frac{1}{(2g)^{\frac{f}{2}}\,\Gamma\left(\frac{f}{2}\right)}\,(W)^{(f/2)-1}\,e^{-\frac{W}{2g}}\,,\quad W > 0\,,$$

where f and g are so chosen that the first two moments of this curve and the true moments of W are the same. First two moments of the curve are given by

$$\text{mean} = gf \quad ; \quad \mu_2 = 2\, g^2 f.$$

The true moments of W are

$$\text{mean} = (af_1+bf_2) \text{ and } \mu_2 = 2(a^2f_1+b^2f_2),$$

where $f_i = (n_i - 1)$, $i = 1, 2.$

Equating the first two moments of the type III curve with those of W, we get

$$g = \frac{a^2f_1+b^2f_2}{af_1 + bf_2} \quad ; \quad f = \frac{(af_1+bf_2)^2}{a^2f_1+b^2f_2} .$$

with these values of f and g we see that $\frac{W}{g}$ is approximately distributed as χ^2 with f degrees of freedom.

Hence $\quad \bar{x}/\sqrt{\frac{W}{fg}}$ is approximately distributed as t - distribution with f degrees of freedom.

Therefore $\quad Y(a,b) \approx C\, t_f$,

where $C = \frac{1}{\sqrt{af_1+bf_2}}$ and t_f is distributed approximately as Student t - distribution with f d.f. For U, it will be seen that

$$a = \frac{\sigma_1^2\,(n_1+n_2)}{(n_1+n_2-2)(n_1\sigma_2^2+n_2\,\sigma_1^2)} \quad ; \quad b = \frac{\sigma_2^2(n_1+n_2)}{(n_1+n_2-2)(n_1\sigma_2^2+n_2\sigma_1^2)} \qquad (1.1)$$

Considering $g, f^*, C = \frac{1}{\sqrt{gf}}$, and the values of a and b, as in (1.1), we have

$$U \approx C\, t_f\,,$$

where

$$C = \sqrt{\frac{(n_1+n_2-2)(n_1\sigma_2^2+n_2\sigma_1^2)}{(n_1+n_2)(f_1\sigma_1^2+f_2\sigma_2^2)}} \quad ; \quad f = \frac{(f_1\sigma_1^2+f_2\sigma_2^2)^2}{(f_1\sigma_1^4+f_2\sigma_2^4)}\,. \qquad (1.2)$$

Similarly the values of a and b for V are

$$a = \frac{n_2\,\sigma_1^2}{(n_1-1)(n_1\sigma_2^2+n_2\sigma_1^2)} \quad \text{and} \quad b = \frac{n_1\,\sigma_2^2}{(n_2-1)(n_2\sigma_1^2+n_1\,\sigma_2^2)}\,. \qquad (1.3)$$

We can, therefore, write $V \approx C t_f$,

$$\text{where} \quad f = \frac{(n_2\sigma_1^2+n_1\sigma_2^2)^2}{\left(\frac{n_2^2\,\sigma_1^4}{f_1}+\frac{n_1^2\,\sigma_2^4}{f_2}\right)} \quad ; \quad C = 1 \qquad (1.4)$$

* f is not necessarily an integer, but may be regarded as a number of d. f for approximation.

1.1 The Validity of U and V:

Suppose that the statistic U is to be used to test H_o at some prescribed level α . If it is assumed that $\sigma_1^2 = \sigma_2^2$, then for a t - distribution with (n_1+n_2-2) d.f it is possible to choose U_o, such that

$$Pr\ (-U_o < U < U_o) = 1 - \alpha\ .$$

If $\sigma_1^2 \neq \sigma_2^2$, and U_o is cho-sen as above, then the test which rejects H_o will be biased, we have

$$Pr\ (\ |U| > U_o) \approx Pr\ (\ |t_f| > \frac{U_o}{C})\ ,$$

where C and f are given in (1.2).
The distribution function of t - distribution with f d.f may be written as

$$2F(t_o)-1 = \frac{2}{\beta(f/2,1/2)} \int_0^{t_o} \frac{1}{\sqrt{f}}\ (1+t^2/f)^{-\frac{1}{2}(f+1)}\ dt.$$

By making transformation $Z = \dfrac{f}{f+t^2}$, we get

$$2F(t_o)-1 = \frac{1}{\beta(f/2,1/2)} \int_Z^1 Z^{(f/2)-1}\ (1-Z)^{\frac{1}{2}-1}\ dZ$$

$$= 1 - I_Z\ (f/2,\ 1/2).$$

Hence $F(t_o) = 1 - \frac{1}{2} \; I_Z(f/2, 1/2)$.

Thus the values of distribution function of Student t - distribution may be obtained from Incomplete Beta - function Tables. It is, therefore, possible to write

$$\Pr(\; |U| > U_o) \approx I_Z(f/2, 1/2) , \qquad (1.5)$$

where
$$Z = \frac{f}{(f + U_o^2/c^2 \;)} .$$

Which shows that for given Sample sizes, c and f depend only on R and it is possible to obtain, for any value of R, the probability of rejecting H_o when H_o is true. The level of significance α is preassigned. The value of U_o, appropriate to the preassigned α, when R=1, is seen from a Student t - table corresponding to (n_1+n_2-2) d.f. If $n_1=n_2=n$, then c will always be unity and

$$f = \frac{(R+1)^2 \; (n-1)}{(R^2+1)} .$$

$\Pr(\; |U| > U_o)$, for different values of R, can be obtained by using Incomplete Beta-function Tables and the relation (1.5). The extent of bias of statistic U can be studied from the graph drawn between the probabilities of rejection and the different values of R. Similarly the case when $n_1 \neq n_2$, (C and f will take values as given in (1.2)), can also be studied.

The values of C and f for the relation $V \approx C\, t_f$, are given in (1.4). Therefore, validity of the test statistic V, by identifying it with a t - distribution with (n_1+n_2-2) d. f may be investigated in the same way.

If it is known that R=1, then certainly U is the exact test statistic. When there exists a possibility that $R \neq 1$, then U will yield wrong conclusions. In this case it is appropriate to use V rather than U. Sice V controls type I error more satisfactorily than U. If there is no information about R, the statistic Z controls type I error more satisfactorily than both U and V, provided sample sizes are greater than 3. Under H_o, the probability of rejecting H_o deviates less from the preassigned level of significance, for the statistic Z than for the statistics U or V.

1.11 Approximation by K - statistics:

An other method of approximating the distribution, of the form y(a,b), is given by Gronow (1951). He has obtained the approximate moments of the distributions of U and V by Fisher's K-statistics. Both U and V are expressed in the form $\dfrac{b(k_1 - k_1')}{(k_2 + ak_2')^{\frac{1}{2}}}$, where a, b are some constants. The quantities k_1, k_2' and k_2, k_2' are the first and second cumulants of the samples. The values of a and b, in general, will differ for U and V except in the case, when $n_1 = n_2$, where U will be identically equal to V.

Let $$z = \frac{(k_1 - k_1')}{(k_2 + ak_2')^{\frac{1}{2}}}.$$ Expanding z, by Taylor's Theorem about the point (K_1, K_1', K_2, K_2'), where K_1, K_1' and K_2, K_2' are the first and second cumulants of the two populations being sampled, and taking expected value, we have, to the order $(n-1)^{-2}$:

$$E(z) = \frac{\delta}{T}\Big[1 + \frac{1}{n_1-1}\,\frac{3K_2^2}{4T^4} + \frac{1}{n_2-1}\cdot\frac{3a^2K_2'^2}{4T^4} + \frac{1}{(n_1-1)^2}$$

$$\cdot\frac{5K_2^3}{2T^6}\left(-1 + \frac{21K_2}{16T^2}\right) + \frac{1}{(n_1-1)(n_2-1)}\cdot\frac{105a^2K_2^2K_2'^2}{16T^8}$$

$$+ \frac{1}{(n_2-1)^2}\,\frac{5a^3K_2'^3}{2T^6}\left(-1 + \frac{21aK_2'}{16T^2}\right)\Big].$$

Here $\delta = K_1 - K_1' = \mu_1 - \mu_2$ and $T = (K_2 + aK_2')^{\frac{1}{2}}$.

Second, third and fourth moments of z can be evaluated in the same way. Taking specific numerical values, an appropriate Pearson Type Curve (with same mean and variance) can be used to approximate the unknown true distribution of U and V.

This method fails to approximate the moments of U and V if the 1st two cumulants, of the populations being sampled, do not exist or they have the same value.

1.2 Exact Distribution of V:

Consider the samples of odd sizes $(2n_1+1)$ and $(2n_2+1)$, drawn from two independent normal populations having variances σ_1^2 and σ_2^2, respectively. Let $\bar{x}_1$, $\bar{x}_2$, s_1^2, s_2^2 be the estimates of the parameters, based on the sample values. If $\mu_1-\mu_2 = \delta$, then V can be written as

$$V = \frac{(\bar{x}_1-\bar{x}_2)-\delta}{\sqrt{\frac{s_1^2}{2n_1+1} + \frac{s_2^2}{2n_2+1}}} .$$

Dividing numerator and denominator by $\left(\frac{\sigma_1^2}{2n_1+1} + \frac{\sigma_2^2}{2n_2+1}\right)^{\frac{1}{2}}$,

we obtain V in the form $\frac{x}{y^{\frac{1}{2}}}$, (1.6)

where $x = \frac{((\bar{x}_1-\bar{x}_2)-\delta)}{\left(\frac{\sigma_1^2}{2n_1+1} + \frac{\sigma_2^2}{2n_2+1}\right)^{\frac{1}{2}}}$ and $y^{\frac{1}{2}} = \left[\frac{\frac{s_1^2}{2n_1+1} + \frac{s_2^2}{2n_2+1}}{\frac{\sigma_1^2}{2n_1+1} + \frac{\sigma_2^2}{2n_2+1}}\right]^{\frac{1}{2}}$.

Obviously, x follows a standard normal distribution and y is distributed as a weighted sum of the χ^2 variates, with $2n_1$ and $2n_2$ d.f, i.e.

$$a\,\chi^2_{2n_1} + b\,\chi^2_{2n_2} ,$$

$$\text{where} \quad a = \frac{\sigma_1^2\,(2n_2+1)}{2n_1\left[\,(2n_2+1)\,\sigma_1^2+(2n_1+1)\,\sigma_2^2\,\right]}$$

$$\text{and} \quad b = \frac{\sigma_2^2(2n_1+1)}{2n_2\left[\,(2n_2+1)\,\sigma_1^2+(2n_1+1)\,\sigma_2^2\right]},$$

are constants with the condition $2n_1a + 2n_2b = 1$.

1.21 Distribution of y:

Box (1954) has given a theorem for the linear combination of χ^2 variables with even d.f, from which the exact distribution of weighted sum of two χ^2 variates can be obtained. Satterthwaite (1941) has also obtained the distribution of weighted sum of two χ^2 variables. The present solution, based on the Laplace Transform (L.T), of the probability density function (p.d.f) of a random variable, is due to Ray and Pitman (1961). They obtained the probability function of y, as follows.

Let p(y) be the p.d.f of y, where

$y = a\,\chi^2_{2n_1} + b\,\chi^2_{2n_2}$, $\chi^2_{2n_1}$ and $\chi^2_{2n_2}$ are independent χ^2 random variables based on $2n_1$ and $2n_2$ d.f. The L. T of p.d.f of a random variable, distributed as χ^2 with even d.f is

$$L(p(\chi^2_{2n_1})) = E(e^{-s\chi^2}) = \int_0^\infty \frac{e^{-s\chi^2}(\chi^2)^{n_1-1}\cdot e^{-\frac{1}{2}\chi^2}}{2^{n_1}\,\Gamma n_1}\,d\chi^2 = (1+2s)^{-n_1},$$

By definition, the L.T of p(y), therefore, is

$$L(p(y)) = \int_0^\infty d\,\chi^2_{2n_1} \int_0^\infty e^{-s(a\chi^2_{2n_1}+b\chi^2_{2n_2})}\,p(\chi^2_{2n_1},\chi^2_{2n_2})\,d\chi^2_{2n_2},$$

where $p(\chi^2_{2n_1}, \chi^2_{2n_2})$ is the joint p.d.f of $\chi^2_{2n_1}$ and $\chi^2_{2n_2}$.

$\chi^2_{2n_1}$ and $\chi^2_{2n_2}$ are independent, therefore, we may write

$$L(p(y)) = (1+2as)^{-n_1}(1+2bs)^{-n_2} = P(s), \text{ say.} \qquad (1.7)$$

We know

$$L^{-1}(P(s)) = \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} P(s)e^{sy}\,ds.$$

This integral can be evaluated if $P(s)e^{sy}$ is a meromorphic function with known poles at $s = s_k$ and infinity, by the method of residues.

Therefore

$$p(y) = S_{k=1} \operatorname{Res} \left(P(s) e^{sy}\right)_{s=s_k} . \qquad (1.8)$$

From (1.7), we may write

$$L^{-1}\left((1+2as)^{-n_1}(1+2bs)^{-n_2}\right) = (2a)^{-n_1}(2b)^{-n_2} L^{-1}\left(\left(\frac{1}{2a}+s\right)^{-n_1}\left(\frac{1}{2b}+s\right)^{-n_2}\right),$$

and $\left(s+\dfrac{1}{2a}\right)^{-n_1} \left(s+\dfrac{1}{2b}\right)^{-n_2}$

has poles at $s=(2a)^{-1}$ of order n_1, at $s=-(2b)^{-1}$ of order n_2, and at infinity.

It is known that

$$R_{-\frac{1}{(2a)}} = \frac{1}{(n_1-1)!} \frac{d^{n_1-1}}{ds^{n_1-1}} \left(e^{sy} \left(s+ \frac{1}{2b}\right)^{-n_2}\right)_{s=-\frac{1}{(2a)}} ,$$

and using Leibnitz's theorem for the multi-order differential of a product, we get

$$R_{-1/(2a)} = \left(\frac{1}{2b}-\frac{1}{2a}\right)^{-(n_1+n_2-1)} e^{-y/(2a)}$$

$$\cdot \sum_{r=0}^{n_1-1} (-1)^{n_1-r-1} \frac{\left(\left(\frac{1}{2b}-\frac{1}{2a}\right)y\right)^r}{r!} \, {}^{(n_1+n_2-r-2)}C_{(n_1-r-1)} \, .$$

$R_{-1/(2b)}$ can also be written by symmetry. Hence from (1.8), we obtain

$$p(y)=\frac{1}{(2a)^{n_1}} \frac{1}{(2b)^{n_2}} \left(R_{-1/(2a)}+R_{-1/(2b)}\right),$$

which on substituting the expressions for $R_{-1/(2a)}$ and $R_{-1/(2b)}$, becomes

$$p(y)=e^{-y/(2a)} \sum_{r=0}^{n_1-1} \alpha_r y^r + e^{-y/(2b)} \sum_{r=0}^{n_2-1} \beta_r y^r, \quad 0 < y < \infty \, , \qquad (1.9)$$

where

$$\alpha_r = \frac{(-1)^{n_1-r-1}\left(\frac{1}{2b}-\frac{1}{2a}\right)^r \, {}^{n_1+n_2-r-2}C_{n_1-r-1}}{(2a)^{n_1}(2b)^{n_2}\left(\frac{1}{2b}-\frac{1}{2a}\right)^{n_1+n_2-1} r!}$$

and
$$\beta_r = \frac{(-1)^{n_2-r-1}\left(\frac{1}{2a}-\frac{1}{2b}\right)^r \; {}^{n_1+n_2-r-2}C_{n_2-r-1}}{(2a)^{n_1}(2b)^{n_2}\left(\frac{1}{2a}-\frac{1}{2b}\right)^{n_1+n_2-1} r!} .$$

1.22 The Distribution of V:

V is distributed as $\frac{x}{y^{\frac{1}{2}}}$, where x follows standard normal distribution and the p.d.f of y is given in (1.9). x and y are independent random variables, their joint p.d.f, therefore, is given by

$$f(x,y) = \frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}x^2} . \left(e^{-y/(2a)} \sum_{r=0}^{n_1-1} \alpha_r y^r + e^{-y/(2b)} \sum_{r=0}^{n_2-1} \beta_r y^r\right),$$

$$-\infty < x < \infty \quad , \quad 0 < y < \infty .$$

Making the transformation $V = \frac{x}{y^{\frac{1}{2}}}$, $\theta = y$, we obtain

$$p(V,\theta) = \frac{1}{\sqrt{2\pi}}\, e^{-V^2\frac{\theta}{2}} \left(e^{-\theta/2a} \sum_{r=0}^{n_1-1} \alpha_r \theta^r + e^{-\theta/2b} \sum_{r=0}^{n_2-1} \beta_r \theta^r\right) . \theta^{\frac{1}{2}},$$

$$-\infty < V < \infty \quad , \quad \theta > 0 .$$

Integrating with respect to θ, this becomes

$$p(V)=\frac{1}{\sqrt{2\pi}}\int_0^\infty e^{-\frac{1}{2}\theta V^2}\theta^{\frac{1}{2}}\left(e^{-\theta/2a}\sum_{r=0}^{n_1-1}\alpha_r\theta^r+e^{-\theta/(2b)}\sum_{r=0}^{n_2-1}\beta_r\theta^r\right)d\theta. \tag{1.10}$$

By the use Gamma - function, (1.10) reduces to

$$p(V)=(2\pi)^{-\frac{1}{2}}\left(\sum_{r=0}^{n_1-1}\alpha_r\Gamma(r+3/2)\left(1/2a+\frac{V^2}{2}\right)^{-(r+3/2)}\right.$$

$$\left.+\sum_{r=0}^{n_2-1}\beta_r\,\Gamma(r+3/2)\left(\frac{1}{2b}+\frac{V^2}{2}\right)^{-(r+3/2)}\right),\quad -\infty<V<\infty\ . \tag{1.11}$$

Considering the r^{th} term of the series

$\frac{1}{\sqrt{2\pi}}\alpha_r\Gamma(r+3/2)\left(\frac{1}{2a}+\frac{V^2}{2}\right)^{-(r+3/2)}$, and transforming

it to a new variable $t_{(2r+2)}=((2r+2)a)^{\frac{1}{2}}V$,

we obtain

$$\frac{2^{-\frac{1}{2}}}{\sqrt{2}}\alpha_r\Gamma(r+3/2)\left(\frac{1}{2a}+\frac{t^2}{2(2r+2)a}\right)^{-(r+3/2)}\cdot\frac{1}{((2r+2)a)^{\frac{1}{2}}}$$

$$=\alpha_r\frac{(2a)^{r+1}\Gamma(r+3/2)}{\Gamma(\frac{1}{2})\Gamma(2r+2)}\left(1+\frac{t^2}{(2r+2)}\right)^{-(r+3/2)}.$$

It becomes

$$(2a)^{r+1} \quad \alpha_r \Gamma(r+1)\, p(t_{(2r+2)}), \qquad (1.12)$$

where $p(t_{(2r+2)})$ is the p.d.f of Student t-distribution with $(2r+2)$ d.f. Substituting the value of α_r in (1.12), we get

$$(-1)^{(n_1-r-1)}(2a)^{n_2}(2b)^{(n_1-r-1)}(2a-2b)^{-(n_1+n_2-r-1)}$$

$$p(t_{(2r+2)}) \quad {}^{(n_1+n_2-r-2)}C_{(n_1-r-1)} \quad . \qquad (1.13)$$

The r^{th} term of the second series in (1.11) may, similarly, be written as

$$(-1)^{(n_2-r-1)}(2a)^{(n_2-r-1)}(2b)^{n_1}(2b-2a)_{x}^{-(n_1+n_2-r-1)}$$

$$p(t_{(2r+2)}) \quad {}^{(n_1+n_2-r-2)}C_{(n_2-r-1)} \quad , \qquad (1.14)$$

where

$$t_{(2r+2)} = ((2r+2)b)^{\frac{1}{2}} \; V.$$

The distribution of V, thus, is the weighted sum of t-distributions.

The percentage points for V can be calculated from the p.d.f of V by using the relation, $Pr(\; |V| < V_\alpha \;) = 1 - \alpha$, where α is preassigned significance level $(0 < \alpha < 1)$. From (1.13), (1.14) and making use of the t-distribution, we get

$$(1-\alpha) = \mathop{S}_{r=0}^{n_1-1} (-1)^{n_1-r-1} (2a)^{n_2} (2b)^{n_1-r-1} (2a-2b)^{-(n_1+n_2-r-1)}$$

$$\Pr\left(|t|_{2r+2} \leq ((2r+2)a)^{\frac{1}{2}} V_\alpha \right). \; {}^{(n_1+n_2-r-2)}C_{(n_1-r-1)}$$

$$+ \mathop{S}_{r=0}^{n_2-1} (-1)^{n_2-r-1} (2a)^{n_2-r-1} (2b)^{n_1} (2b-2a)^{-(n_1+n_2-r-1)}$$

$${}^{(n_1+n_2-r-2)}C_{(n_2-r-1)} \cdot \Pr\left(|t|_{2r+2} \leq ((2r+2)b)^{\frac{1}{2}} V_\alpha \right). \qquad (1.15)$$

The expression (1.15) can be written in more compact form by substituting

$\gamma = 2an_1$; $1-\gamma = 2bn_2$, in it, we have, then

$$(1-\alpha) = \mathop{S}_{r=0}^{n_1-1} (-1)^{n_1-r-1} \left(\frac{\gamma}{n_1}\right)^{n_2} \left(\frac{1-\gamma}{n_2}\right)^{n_1-r-1} \left(\frac{\gamma}{n_1}-\frac{1-\gamma}{n_2}\right)^{-(n_1+n_2-r-1)}$$

$${}^{(n_1+n_2-r-2)}C_{(n_1-r-1)} \cdot \Pr\left(|t|_{2r+2} \leqslant \left(\frac{(r+1)\gamma}{n_1}\right)^{\frac{1}{2}} \cdot V_\alpha \right) +$$

$$\mathop{S}_{r=0}^{n_2-1} (-1)^{n_2-r-1} \left(\frac{\gamma}{n_1}\right)^{n_2-r-1} \left(\frac{1-\gamma}{n_2}\right)^{n_2} \left(\frac{1-\gamma}{n_2}-\frac{\gamma}{n_1}\right)^{-(n_1+n_2-r-1)}$$

$${}^{(n_1+n_2-r-2)}C_{(n_2-r-1)} \cdot \Pr\left(|t|_{2r+2} \leqslant \left(\frac{(r+1)(1-\gamma)}{n_2}\right)^{\frac{1}{2}} V_\alpha \right). \quad (1.16)$$

By specifying α, n_1, n_2 and γ, we may determine V_α. The only condition of normality is not sufficient for the general solution, and the unrestricted distribution of V, thus obtained, has infact no practical use in the two means problem, since it involves R which is unknown.

1.3 General Approach:

Let x_i $(i=1,2, \ldots, k)$ be stochastic variables normally and independently distributed with means μ_i and variances $\lambda_i \sigma_i^2$ respectively, with known positive constants λ_i, but μ_i and σ_i^2 being unknown. Suppose s_i^2 yield estimates of σ_i^2 which follow distribution as $\chi_i^2 \sigma_i^2 / f_i$, where f_i is the d.f of χ_i^2. The quantities s_i^2 $(i=1,2,\ldots,k)$, are supposed to be independently distributed. The aim is to test whether the data are consistent with the hypothesis $H_0 : \mu_i = \mu$. A particular case is, when x_i are the means $\bar{x}_i$ of samples of sizes n_i, drawn from k independent normal populations having true means μ_i and variances σ_i^2. Since variance of $\bar{x}_i$ is σ_i^2/n_i, so $\lambda_i = \frac{1}{n_i}$.

The hypothesis is as follows: whether the k populations being sampled may be considered to have the same mean without imposing any condition on variances.

James (1951) has considered a statistic

$$\mathop{S}_{i=1}^{k} w_i (x_i - \hat{x})^2 \text{, where } w_i = \frac{1}{\lambda_i s_i^2} \text{ and } \hat{x} = \mathop{S}_{i=1}^{k} \frac{w_i x_i}{\mathop{S}_{i=1}^{k} w_i}.$$

Under H_0, this statistic, in LARGE samples, follows approximately χ^2 distribution with (k-1) d.f. It is then possible to make a statement of the form

$$\Pr \left[\mathop{S}_{i=1}^{k} w_i (x_i - \hat{x})^2 > \chi_\alpha^2 \right] = \alpha \quad .$$

For small samples, in order to make this type of statement, he obtained a function $h(w_1, w_2, \ldots, w_k, \alpha)$, having the property

$$\Pr(\mathop{S}_{i=1}^{k} w_i (x_i - \hat{x})^2 > h(w_1, w_2, \ldots, w_k, \alpha)) = \alpha \quad ,$$

and developed a method of arriving at successive approximations to this function in terms of the orders $\frac{1}{f_i}$. The exact function of this nature was evaluated by Student (1908) for a single mean problem. Later on Welch (1947) gave iterative methods of calculating it for general case. The series given by Welch (1947) is only asymptotic and suffers from convergence difficulties. James (1951), for instance, obtained results to the order $\frac{1}{f_i}$ as

$$h(w_1, w_2, \ldots, w_k, \alpha)$$

$$= \chi_\alpha^2 \left[1 + \frac{3\chi_\alpha^2 + (k+1)}{2(k^2-1)} \mathop{S}_{i=1}^{k} \frac{1}{f_i} \Big(1 - \frac{w_i}{\mathop{S}_{i=1}^{k} w_i}\Big)^2 \right] . \quad (1.17)$$

For approximating this function involving higher orders of $\frac{1}{f_i}$, he pointed out its limited practical utility.

Welch (1951) has, to the order $\frac{1}{f_i}$, obtained the same result as given by James(1951) in (1.17), by an alternative method. He developed an approximation which involves the use of Variance Ratio Tables rather than χ^2 - Tables and has more pratical utility. Welch(1951) method of approximating the function $h(w_1,w_2,\ldots,w_k,\alpha)$ by the use of cumulant - generating function of the statistic

$\overset{k}{\underset{i=1}{S}} w_i(x_i-\hat{x})^2$ is as follows.

For $k=2$, the statistic $\overset{2}{\underset{i=1}{S}} w_i(x_i-\hat{x})^2$ reduces to

$$\frac{(x_1-x_2)^2}{\left(\frac{1}{w_1}+\frac{1}{w_2}\right)}\text{, i.e. } \overset{2}{\underset{i=1}{S}} w_i(x_i-\hat{x})^2=V^2,$$

where $V = \frac{(x_1-x_2)}{\sqrt{\lambda_1 s_1^2+\lambda_2 s_2^2}}$.

This statistic seperates into a function of x_i divided by a function of s_i^2, when $k=2$. But for $k>2$, such separation is not possible, which shows that the approximation for distribution of $\underset{i=1}{\overset{k}{S}} w_i(x_i-\hat{x})^2$ is to be made independent of s_i^2. Assuming that the moment - generating function of this statistic exists, we write

$$M(u) = E_1\, E_2\, e^{\left[u\, S_{i=1}^{k} w_i(x_i-\hat{x})^2\right]} ,$$

where E_1, E_2 denote averaging over the joint distribution of x_i and s_i^2 respectively. Recalligg $w_i = \frac{1}{\lambda_i s_i^2}$ and equating $W_i = \frac{1}{\lambda_i \sigma_i^2}$, the moment - generating function is

$$M(u) = (1-2u)^{-\frac{1}{2}(k-1)}\Big[\, 1 + \big(2u(1-2u)^{-1} + 3u^2(1-2u)^{-2}\big)$$

$$\Big(\underset{i=1}{\overset{k}{S}} \frac{1}{f_i}\big(1 - \frac{W_i}{S_{i=1}^{k} W_i}\big)^2\Big)\Big] .$$

The corresponding cumulant - generating function, to the order $\frac{1}{f_i}$, therefore is

$$K(u) = -\tfrac{1}{2}(k-1)\log(1-2u) + \left(2u(1-2u)^{-1} + 3u^2(1-2u)^{-2}\right) \cdot$$

$$\left(S_{i=1}^{k} \frac{1}{f_i} \left(1 - \frac{W_i}{S_i^k W_i}\right)^2 \right). \qquad (1.18)$$

1.31 Approximation by F-distribution:

The F - distribution is more convenient to use than t - distribution while comparing several mean values when population variances are known to be equal. Consider the moment - generating function of the F - distribution. We have

$$F = \left(\frac{\chi_1^2}{\hat{f}_1}\right) \Big/ (\chi_2^2/\hat{f}_2),$$

where χ_1^2 and χ_2^2 are distributed independently as χ^2, with $\hat{f}_1$ and $\hat{f}_2$ d.f respectively. Now

$$E\, e^{u\, \chi_1^2/\hat{f}_1} = (1 - 2u/\hat{f}_1)^{-\frac{1}{2}\hat{f}_1} \quad .$$

For given χ_2^2, we have

$$E\ e^{uF} = \left(1 - \frac{2U\ \hat{f}_2}{\hat{f}_1\ \chi_2^2}\right)^{-\frac{1}{2}\hat{f}_1}.$$

The moment - generating function, $M_F(u)$, is then given by averaging it over χ_2^2 distribution. Writing, to the order $\frac{1}{\hat{f}_2}$, we get

$$M_F(u)=(1-\frac{2u}{\hat{f}_1})^{-\frac{1}{2}\hat{f}}\left[\ 1+\frac{2u}{\hat{f}_2}(1-2u/\hat{f}_1)^{-1}+\frac{(\hat{f}_1+2)}{\hat{f}_1\hat{f}_2}\ u^2(1-2u/\hat{f}_1)^{-2}\right].$$

(1.19)

By substituting $\hat{f}_1=(k-1)$ and $G=\left[\ (k-1)+A/\hat{f}_2\ \right]F$ in (1.19), we obtain

$$M_G(u)=(1-2u)^{-\frac{1}{2}(k-1)}\left[1+\frac{A+a(k-1)}{\hat{f}_2}\ u(1-2u)^{-1}\right.$$

$$\left.+\ \frac{(k^2-1)}{\hat{f}_2}\ u^2(1-2u)^{-2}\right]\ .$$

An equivalent expression corresponding to cumulant-generating function of $\underset{i=1}{\overset{k}{S}}\ w_i(x_i-\hat{x})^2$, therefore, is

$$K_G(u) = -\tfrac{1}{2}(k-1)\log(1-2u) + \frac{1}{\hat{f}_2}(A+2(k-1))\,u(1-2u)^{-1}$$

$$+\frac{(k^2-1)}{\hat{f}_2}\,u^2(1-2u)^{-2}\,. \qquad (1.20)$$

Comparing (1.18) and (1.20), it can be easily seen that

$$\left.\begin{aligned} \frac{A+2(k-1)}{\hat{f}_2} &= 2\sum_{i=1}^{k}\frac{1}{f_i}\Big(1-\frac{W_i}{\sum_{i=1}^{k}W_i}\Big)^2 \\ \text{and}\quad \frac{(k^2-1)}{\hat{f}_2} &= 3\sum_{i=1}^{k}\frac{1}{f_i}\Big(1-\frac{W_i}{\sum_{i=1}^{k}W_i}\Big)^2 \end{aligned}\right\}\Rightarrow$$

$$\frac{A}{\hat{f}_2} = 2\sum_{i}^{k}\frac{1}{f_i}\Big(1-\frac{W_i}{\sum_{i=1}^{k}W_i}\Big)^2 - \frac{2(k-1)}{\hat{f}_2}\,, \qquad (1.21)$$

$$\frac{1}{\hat{f}_2} = \frac{3}{(k^2-1)}\sum_{i=1}^{k}\frac{1}{f_i}\Big(1-\frac{W_i}{\sum_{i=1}^{k}W_i}\Big)^2\,. \qquad (1.22)$$

After substituting the value of $\frac{1}{\hat{f}_2}$ from (1.22), into (1.21), we have

$$\frac{A}{\hat{f}_2} = \frac{2(k-2)}{(k+1)}\sum_{i=1}^{k}\frac{1}{f_i}\Big(1-\frac{W_i}{\sum_{i=1}^{k}W_i}\Big)^2\,. \qquad (1.23)$$

Which means, to order $\frac{1}{f_i}$, the quantity $S_{i=1}^{k} w_i (x_i - \hat{x})^2$

is distributed as

$((k-1) + \frac{A}{\hat{f}_2})$ times F, where $\frac{A}{\hat{f}_2}$ and $\hat{f}_2$ are given by

(1.23) and (1.22).

1.32 Practical Application:

We define

$$V^2 = \frac{S_{i=1}^{k} w_i (x_i - \hat{x})^2 / (k-1)}{\left[1 + \frac{2(k-2)}{(k^2-1)} \underset{i=1}{\overset{k}{S}} \frac{1}{f_i} \left(1 - \frac{W_i}{S_{i=1}^{k} W_i}\right)^2 \right]} .$$

Let F_α be the tabulated value of the Variance Ratio F - Table, corresponding to the significance α with d.f

$$\hat{f}_1 = (k-1) \text{ and } \hat{f}_2 = \left[\frac{3}{(k^2-1)} \underset{i=1}{\overset{k}{S}} \frac{1}{f_i} \left(1 - \frac{W_i}{S_{i=1}^{k} W_i}\right)^2 \right]^{-1} .$$

Under H_0, we can say approximately (to order $\frac{1}{f_i}$), then

$$Pr(V^2 > F_\alpha) = \alpha . \qquad (1.24)$$

V^2 involves the unknown W_i and sample values w_i, and W_i also enter into $\hat{f}_2$. We, therefore, cannot use (1.24).

However, as W_i enter only into expressions of order $\frac{1}{f_i}$, with the substitution of w_i for W_i, we can make approximate probability statement like (1.24).

The approximate test procedure, therefore, is:

(i) Calculate
$$V^2 = \frac{\mathop{S}\limits_{i=1}^{k} w_i(x_i-\hat{x})^2/(k-1)}{1+\frac{2(k-2)}{(k^2-1)}\mathop{S}\limits_{i=1}^{k}\frac{1}{f_i}\left(1-\frac{w_i}{\mathop{S}\limits_{i=1}^{k} w_i}\right)^2},$$

where $\hat{f}_1=(k-1)$, $\hat{f}_2=\left[\frac{3}{(k^2-1)}\mathop{S}\limits_{i=1}^{k}\frac{1}{f_i}\left(1-\frac{w_i}{\mathop{S}\limits_{i=1}^{k} w_i}\right)^2\right]^{-1}$.

(ii) Refer V^2 to F-Table with $\hat{f}_1$ and $\hat{f}_2$ d.f.

1.4 References:

Barnard, G.A. (1950), The Fisher - Behrens Test, Biometrika, Vol 37, 203-207.

Box, G.E.P. (1954), Some theorems on quadratic forms applied in the study of analysis of variance problem, 1. Effect of inequality of variance in the one way classification, Ann. Math. Stat. Vol 25, 290-302.

Gro-now, D.G.C. (1951), Significance test for difference between means, Biometrika, Vol 38, 252-256.

James, G.S.(1951), The comparison of several groups of observations when the ratio of the pop variances are unknown, Biometrika, Vol 38, 324-329.

James, G.S. (1959), Behrens-Fisher distribution and weighted means, J. Roy. Stat. Soc. Series B, Vol 21, 73-90.

Kabe, D. G. (1966), On exact distribution of Behrens-Fisher Welch Statistic, Matrika, Vol 10, 13-15.

Ray, W.D. and Pitman, A.E.N.T. (1961), An exact distribution of the Fisher-Behrens-Welch statistic for testing the difference between the means of two normal populations with unknown variances, J. Roy. Stat. Soc. Series B, Vol 23, 377-383.

Student (1908), The probable error of a mean, Biometrika, Vol 6, 1-25.

Satterthwaite, F.E. (1941), Synthesis of variance, Psychometrika, Vol 6, 309-316.

Welch, B.L.(1936), Specification of rules for rejecting too variable a product with particular reference to an electric lamp problem, J. Roy. Stat. Soc. Supp III, Vol, 47.

Welch, B.L. (1947), The generalisation of Student's problem when several different population variances are in-valved, Biometrika, Vol 34, 29-35.

Welch, B.L. (1951), On the comparison of several mean values an alternative approach, Biometrika, Vol 38, 330-335.

Wilks, S.S. (1940), On the problem of two samples from normal populations with unequal variances, Ann. Math. Stat. Vol 11, 475.

CHAPTER 2

UNILATERAL & BILATERAL STATISTICS

2.0 Introduction:

The statistics U, V and Z, as given in §(1.), are considered by Welch (1937, 1947). Aspin (1948, 1949) has extended the results and investigated the numerical behaviour of the series developed by Welch (1947). Ura (1955) has obtained the power function of Welch(1947) test, for the case when two population variances are equal and compared it with that of Student t-test. In Welch (1947) a function of $s_i^2(i=1,2)$, and α, with the property

$$Pr\left((\bar{x}_1-\bar{x}_2)-\delta < h(s_i^2, \alpha)\right) = \alpha ,$$

is obtained. The corresponding critical points considered by him are not fixed and are the functions of the sample variances. Wald (1955) gave a statistic for equal sample sizes but has also used a random critical point.

From §(1.), a statistic of the general form

$$Y(r_1, r_2) = \frac{(\bar{x}_1-\bar{x}_2)^2}{r_1\Sigma_1 + r_2\Sigma_2} , \qquad (2.1)$$

may be obtained, where r_1 and r_2 are positive constants, depending upon the sample sizes. This statistic might be considered to control the size of a test for various values of R. It is, therefore, possible to choose r_1 and r_2 such that, the hypothesis H_o, is rejected only when $Y(r_1, r_2) > 1$.

The distribution of $Y(r_1, r_2)$, under H_o, may be given by

$$Y(r_1, r_2) = \frac{\chi_1^2}{a\chi_{f_1}^2 + b\chi_{f_2}^2} ,$$

where $f_i = (n_i - 1), (i = 1, 2)$, is the d.f of χ^2- variates and

$$a = \frac{\sigma_1^2\ r_1}{(\sigma_1^2/n_1 + \sigma_2^2/n_2} ; \quad b = \frac{\sigma_2^2\ r_2}{(\sigma_1^2/n_1 + \sigma_2^2/n_2} .$$

All χ^2 variables are independently distributed. Under $H_1 : \mu_1 \neq \mu_2$, the distribution of $Y(r_1, r_2)$, is given by

$$Y(r_1, r_2) = \frac{\chi_{1,h}^2}{a\ \chi_{f_1}^2 + b\chi_{f_2}^2} .$$

The numerator is non-central χ^2 with one d.f and non-centrality parameter h, is given by

$$h = \frac{(\mu_1 - \mu_2)^2}{\sigma_1^2/n_1 + \sigma_2^2/n_2} .$$

The class of statistics $Y(r_1, r_2)$, as suggested by Welch(1937), is considered by Gurland et al (1960), and the size α of the test, is examined by them for a fixed point (unity). They have treated two kinds of statistics separately. The first kind, called unilateral, keeps the size of the test less than or equal to a given fixed value over the range $R \geq 1$, if it is known a priori that the variance of one population is greater than the other. The second is called bilateral, which keeps the size of the test less than or equal to a preassigned value over the whole range of R, if there is no apriori knowledge of population variances. For the bilateral case, they have shown that Student t - tables may be used to find the appropriate statistic for any pair of sample sizes. Their method of finding the size of the test, for both unilateral and bilateral cases may be described as below.

2.1 Size of the statistic $Y(r_1, r_2)$:

In order to calculate the size of a test using the statistic $Y(r_1, r_2)$, it is required to evaluate the probabilities of the form.

$$Pr(Y(r_1, r_2) > 1 \mid R), \tag{2.2}$$

where R is an unknown constant. Under H_0, and for the specific values of n_1, n_2 and r_1, r_2, the statistic $Y(r_1, r_2)$ is distributed as

$$\frac{\chi_1^2}{(a\chi_{f_1}^2 + b\chi_{f_2}^2)} .$$

In two extreme cases the statistic $Y(r_1, r_2)$, takes the following form.

$$\frac{\chi_1^2}{n_2\, r_2\, \chi_{f_2}^2} \quad \text{when} \quad \sigma_1^2 \longrightarrow 0 , \tag{2.3}$$

and

$$\frac{\chi_1^2}{n_1\, r_1\, \chi_{f_1}^2} \quad \text{when} \quad \sigma_2^2 \longrightarrow 0 . \tag{2.4}$$

Various probabilities in (2.2) can be plotted against the different values of R. The graph so obtained will approach horizontal asymptotes as $R \rightarrow 0$ and $R \longrightarrow \infty$.

The probabilities in the extreme cases may be calculated by the use of Incomplete Beta - function, due to its relation with Student t - distribution. These can also be evaluated as follows.

Let

$$G_{1,f_1}(c) = \Pr\left((\chi_1^2/\chi_{f_1}^2) > c\right) = \alpha \ .$$

Since $\dfrac{\chi_1^2}{\chi_{f_1}^2} f_1$ is distributed as t^2 - distribution with f_1 d.f. obviously, then

$$c = (t_{f_1}^{\alpha})^2/f_1 \ . \tag{2.5}$$

Here $t_{f_1}^{\alpha}$ is the two sided α % points of Student t - distribution with f_1 d.f.

To compute probabilities in between the extreme cases of the general form (2.2), the following theorem, obtained from a theorem of the linear combination of χ^2 variates, due to Box (1954), is used. Theorem: Let Y be a random variable with the form

$$Y = \frac{\chi_n^2}{(a_1\chi_{f_1}^2 + a_2\chi_{f_2}^2)} \ ,$$

where χ^2 - variables are independent, the $f_j = 2g_j$ are even integers, and a_i are positive constants. Then the distribution function of Y is given by

$$F_Y(y) = \sum_{j=1}^{2} \sum_{s=1}^{g_j} \alpha_{js} F_{n,2s}(a_j, y), \tag{2.6}$$

where $F_{n,m}(x) = 1 - G_{n,m}(x)$, the constants α_{js} are given by

$$\alpha_{1s} = (-1)^{g_1-s} \frac{\Gamma(g_1+g_2-s)}{\Gamma g_2 \; \Gamma(g_1-s+1)} \cdot \frac{a_1^{g_2} \; a_2^{g_1-s}}{(a_1-a_2)^{g_1+g_2-s}} \quad (s=1,2,\ldots,g_1).$$

$$\alpha_{2s} = (-1)^{g_2-s} \frac{\Gamma(g_1+g_2-s)}{\Gamma g_1 \; \Gamma(g_2-s+1)} \cdot \frac{a_1^{g_2-s} \; a_2^{g_1}}{(a_2-a_1)^{g_1+g_2-s}} \quad (s=1,2,\ldots,g_2).$$

2.11 Optimality of Unilateral and Bilateral Statistics:

The statistic Z, as given in §(1.), can be useful, if it is known that $\sigma_1^2 > \sigma_2^2$. Since for particular values of r_1 and r_2, the size for the statistic Z has an asymptotic value α, as $R \to \infty$, and is only slightly less than α for entire range $1 < R < \infty$. For instance, in the case when populations might consist of measurements made by two different techniques, a particular one of which is known to be more precise than the other. This information is utilised by the statistic Z, in keeping the size of the test practically constant over the relevant range $R > 1$. Such a statistic may be called unilateral statistic and can be looked upon as optimal within the class $Y(r_1, r_2)$, because no other statistic in this class keeps the size as nearly constant and less than or equal to α, over the range $R \geq 1$.

For all sample sizes, unilateral statistics can be found from the following two conditions.

$$\Pr(Y (r_1, r_2) > 1 \mid R=1) = \alpha ,$$

and

$$\Pr (Y (r_1, r_2) > 1 \mid R\to\infty) = \alpha .$$

The second condition, by (2.2) and (2.4), may be written as

$$\Pr\left((\chi_1^2/\chi_{f_1}^2) > r_1 n_1\right) = \alpha . \tag{2.7}$$

The parameter r_1 from (2.5) and (2.7) is given by

$$r_1 = (t_{f_1}^{\alpha})^2/(f_1 n_1) .$$

The parameter $r_2 = \phi/(n_2 f_2)$, where ϕ is tabulated by Gurland et al (1960), for some particular values of α and various sample sizes, by making use of the relation (2.6).

In case when it is not possible to assume that one population variance is greater than the other, all values of R must be accounted for in constructing a statistic in order to control the type I error. Owing to practical limitations the size of the test is kept $\leq \alpha$. It can be seen that the statistic whose parameters are defined by

$$\Pr\ (Y(r_1,\ r_2) > 1 \mid R \to 0) = \alpha \quad ,$$

and (2.8)

$$\Pr\ (Y(r_1,\ r_2) > 1 \mid R \to \infty) = \alpha \quad ,$$

is the most optimal within the class of statistics considered, when R is unknown. Such a statistic may be called bilateral statistic.

From (2.8), we find r_1 and r_2 for all values of n_1 and n_2 and for a preassigned value of α , by the use of Student t - distribution as

$$r_1 = (t^{\alpha}_{f_1})^2/n_1f_1 \ ,$$

and (2.9)

$$r_2 = (t^{\alpha}_{f_2})^2/n_2f_2 \ .$$

It is obvious that if the information, $R > 1$, is ignored it is possible to arrive at a different conclusion than if this information is utilised.

2.2 Proposed Procedures:

The problem of testing the difference in means of two normal populations, without assuming the equality of the variances involved, is treated from the point of view of employing a preliminary test for the population variances. The use of a preliminary test in testing a statistical hypothesis has been considered by Bancraft (1944) and Bozivich (1956) in various contexts. Chand (1950) has studied the behaviour of type I error in repeated sampling from populations with a fixed value of unknown variance ratio by utilising an approximate knowledge about the unknown variance ratio. The preliminary hypothesis, H_{00} : $\sigma_1^2 = \sigma_2^2$, is tested by using a test statistic $\Sigma_1/\Sigma_2 = \gamma$. The effect of departure of R from unity on the size of some tests for H_0, has been investigated by Gurland et al (1962). Their aim is that if the size of a test cannot be made constant for all values of R, then it should be kept as close to a constant value as possible and this should not be accomplished on the expense of decreasing the power of the test under consideration. Their method of calculating the size of some tests, proposed by them, is as follows.

2.21 Unilateral Case:

In preliminary part, the hypothesis, H_{oo}: $\sigma_1^2 = \sigma_2^2$, is tested by using γ. If we denote the critical point by a and the significance level of this preliminary test by α_o, then

$$\Pr(\gamma > a \mid R = 1) = \alpha_o . \qquad (2.10)$$

We make use of the statistic U^2, as given in § (1.), for testing H_o, if H_{oo} is not rejected. But if H_{oo} is rejected then the statistic would be some constant times U^2. In case H_{oo} is rejected, it will be equivalent to retaining the statistic U^2 but changing the critical point. The description of this test procedure I (say), is as follows.

Procedure I:

If $\gamma \leq a$, reject H_o if $U^2 > c$ but accept H_o if $U^2 \leq c$.

If $\gamma > a$, reject H_o if $U^2 > c'$ but accept H_o if $U^2 \leq c'$.

The size of the test, using procedure I is given by

$$\Pr(\gamma \leq a;\ U^2 > c) + \Pr(\gamma > a;\ U^2 > c').$$

This is obtained by the expression

$$\Pr\left(R\,\frac{\chi^2_{f_1}}{\chi^2_{f_2}} \leq a \,;\; \frac{\chi^2_1}{R\,\chi^2_{f_1}+\chi^2_{f_2}} > cd\right)$$

$$+ \Pr\left(\frac{\chi^2_{f_1}}{\chi^2_{f_2}} > a/R \,;\; \frac{\chi^2_1}{R\,\chi^2_{f_1}+\chi^2_{f_2}} > c'd\right), \qquad (2.11)$$

where d is a constant and is equal to $\dfrac{n_1+n_2}{(f_1+f_2)(n_2R+n_1)}$.

χ^2 variates in (2.11) are independent random variables with $f_i=n_i-1, (i=1,2)$, d.f. The complete specification of this procedure requires the values of the constants c, c' and α to be given. The value of a can be determined from (2.10).

Substituting $r=r_1/r_2$ in (2.1), we get

$$Y(r)=\frac{(\bar{x}_1-\bar{x}_2)^2}{r\,\Sigma_1+\Sigma_2}=r_2\,Y(r_1,\,r_2). \qquad (2.12)$$

It is possible to test H_0 by defining a test procedure II (say). Using critical points c and c' in conjunction with U^2 and $Y(r)$ respectively, we can write the statement for procedure II as follows.

Procedure II:

If $\gamma \leq a$, reject H_o if $u^2 > c$ but accept H_o if $u^2 \leq c$.

If $\gamma > a$, reject H_o if $Y(r) > c'$ but accept H_o if $Y(r) \leq c'$.

The size for the test procedure II is obtained from

$$\Pr\left(\frac{\chi^2_{f_1}}{\chi^2_{f_2}} \leq a/R;\ \frac{\chi^2_1}{R\,\chi^2_{f_1}+\chi^2_{f_2}} > cd\right)+\Pr\left(\frac{\chi^2_{f_1}}{\chi^2_{f_2}} > a/R;\ \frac{\chi^2_1}{Rr\chi^2_{f_1}+\chi^2_{f_2}} > c'g\right), \tag{2.13}$$

where the constant $g = \dfrac{n_1 n_2}{n_1+Rn_2}$.

The values of a, c and c', used in procedure II, would in general differ from the values used in the procedure I.

Second term of the expression (2.13), when r=1, reduces to

$$\Pr\left(\frac{\chi^2_{f_1}}{\chi^2_{f_2}} > a/R\ ;\ \frac{\chi^2_1}{R\chi^2_{f_1}+\chi^2_{f_2}} > g\,c'\right),$$

which is same as the second term of expression (2.11).

2.22 Bilateral Case:

In considering this case all values of $R(=R > 0)$ can be used. The modified test, for testing H_{00}, will involve two critical points, a_1 and a_2, such that

$$Pr(\gamma \leq a_1 | R = 1) + Pr(\gamma > a_2 | R = 1) = \alpha \quad . \qquad (2.14)$$

If equal tail areas for this test are considered, then

$$Pr(\gamma \leq a_1 | R = 1) = Pr(\gamma > a_2 | R = 1) = \frac{\alpha}{2} \quad .$$

This test procedure III(say), is analogous to the procedure I. The formal statement for procedure III may be given as follows.

Procedure III:

If $\gamma \leq a_1$, reject H_0 if $U^2 > c_1'$ but accept H_0 if $U^2 \leq c_1'$.

If $\gamma > a_2$, reject H_0 if $U^2 > c_2'$ but accept H_0 if $U^2 \leq c_2'$.

If $a_1 < \gamma \leq a_2$, reject H_0 if $U^2 > c$ but accept H_0 if $U^2 \leq c$.

Applying procedure III, the size of the test is obtained by

$$\Pr\Big(\frac{\chi^2_{f_1}}{\chi^2_{f_2}} \leqslant a_1/R;\ \frac{\chi^2_1}{R\chi^2_{f_1}+\chi^2_{f_2}} > c_1'd\Big)+\Pr\Big(\frac{\chi^2_{f_1}}{\chi^2_{f_2}} > \frac{a_2}{R};\ \frac{\chi^2_1}{R\chi^2_{f_1}+\chi^2_{f_2}}\ \ c_2'd\Big)$$

$$+\Pr\Big(\frac{\chi^2_{f_1}}{\chi^2_{f_2}} \leqslant \frac{a_2}{R};\ \frac{\chi^2_1}{R\chi^2_{f_1}+\chi^2_{f_2}} > cd\Big) - \Pr\Big(\frac{\chi^2_{f_1}}{\chi^2_{f_2}} \leqslant \frac{a_1}{R};\ \frac{\chi^2_1}{R\chi^2_{f_1}+\chi^2_{f_2}} > cd\Big). \tag{2.15}$$

Complete specification of procedure III involves the constants α, c_1', c_2' and c. The critical points a_1, a_2 are determined by the relation (2.14).

2.23 Size of Tests:

In order to calculate the sizes of the tests for different procedures considered, we require the evaluation of the expressions,

$$\Pr\Big(\frac{\chi^2_{f_1}}{\chi^2_{f_2}} \leqslant a/R;\ \frac{\chi^2_1}{R\chi^2_{f_1}+\chi^2_{f_2}} > cd\Big) \tag{2.16}$$

and

$$\Pr\Big(\frac{\chi^2_{f_1}}{\chi^2_{f_2}} > a/R;\ \frac{\chi^2_1}{Rr\,\chi^2_{f_1}+\chi^2_{f_2}} > c'g\Big). \tag{2.17}$$

Let $X = \frac{1}{2}\chi_1^2$; $W = \frac{1}{2}\chi_{f_1}^2$ and $M = \frac{1}{2}\chi_{f_2}^2$.

Applying, that the p.d.f of $\frac{\chi_k^2}{n}$ is

$$f\left(\frac{\chi_k^2}{n}\right) = \frac{\left(\frac{n}{2}\right)^{\frac{k}{2}}}{\Gamma\left(\frac{k}{2}\right)} g^{\frac{k}{2}-1} e^{-ng/2} , \quad g > 0 ,$$

where g is distributed as χ^2 with k d.f, we can write the joint probability density function of X, W and M as

$$p(x,w,m) = \frac{1}{\Gamma\frac{1}{2}\,\Gamma(\frac{1}{2}f_1)\,\Gamma(\frac{1}{2}f_2)} x^{-\frac{1}{2}} w^{(\frac{1}{2}f_1)-1} m^{(\frac{1}{2}f_2)-1} e^{-x-w-m}, (x,w,m > 0).$$

The expressions (2.16) and (2.17), then, reduce to the form $\Pr(RW-aM \leqslant 0;\ cd\, RW + cdM < X)$,

and

$$\Pr(RW - aM > 0;\ Rrc'gW + c'gM < X).$$

By making use of Pearson (1934) Incomplete Beta - function tables, the computations can be simplified for the cases when $R \rightarrow 0$ or ∞ and $R = 1$.

2.3 References:

Aspin, A.A. (1948), An examination and further development of a formula arising in the problem of comparing two mean values, Biometrika, Vol 35, 88-96.

Aspin, A.A. (1949), Tables for use in comparison whose accuracy involves two variances, separately estimated, Biometrika, Vol 36, 290-296.

Bancroft, T.A. (1944), On the biases in estimation due to the use of preliminary tests of significance, Ann. Math. Statist. Vol 15, 190-204.

Box, G.E.P. (1954), Some theorems on quadratic forms applied in the study of analysis of variance problem, 1. Effect of inequality of variance in the one way classification, Ann. Math. Statist. Vol 25, 290-302.

Bozivich, H., Bancroft, T.A. and Hartley, H. O.(1956), Power of analysis of variance test procedures for certain incompletely specified models, Ann. Math. Statist. Vol 27, 1017-43.

Chand, U. (1950), Distributions related to comparison of two means and two regression coefficients, Ann. Math. Statist. Vol 21, 507-22.

Gurland, J. and McCullough, R. (1962), Testing equality of means after preliminary test of equality of variances, Biometrika, Vol 49, 402-417.

McCullough, R.S., Gurland, J. and Rosenberg, L.(1960), Small sample behaviour of certain tests of hypothesis of equal means under variance hetrogeneity, Biometrika, Vol 47, 345-53.

Pearson, K. (1934), Tables of Incomplete Beta Function, Cambridge University Press.

Wald, A. (1955), Testing the difference between the means of two normal populations with unknown standard deviations, Selected papers in statistics and probability; by A. Wald, N. York: McGraw Hill.

Welch, B.L. (1937), The significance of difference between two means when the population variances are unequal, Biometrika, Vol 29, 350-62.

Welch, B.L. (1947), The generalisation of Student, problem when several different population variances are involved, Biometrika, Vol 34, 28-35.

CHAPTER 3

CONFIDENCE INTERVALS

3.0 Introduction:

The Behrens - Fisher problem is treated from confidence intervals point of view. Neyman (1941) has made a simplified but less general statement of the result, obtained by unpublished solution of Bartlett, which is also briefly mentioned by Welch (1938). Neyman (1941) result is based on successive differences of the two sample observations and may be obtained in the following way.

Let $(x_1,\ldots,x_{n_1})$ and $(y_1,\ldots,y_{n_2})$ be the random samples drawn from two independent normal populations with mean values μ_1 and μ_2 and variances σ_1^2 and σ_2^2, respectively. Suppose $n_1 \leqslant n_2$ and $\delta = \mu_1 - \mu_2$. Select randomly a subset of n_1 from n_2 variates of the second sample and calculate n_1 differences, $\ell_i = y_i - x_i$ $(i=1,2,\ldots,n_1)$, neglecting $n_2 - n_1$ observations of y. The standard error of ℓ_i is $(\sigma_1^2+\sigma_2^2)^{\frac{1}{2}}$. The differences ℓ_i will be normally and independently distributed. The problem is then reduced to that of estimating the mean of ℓ_i.

The confidence interval for which can be given by

$$\bar{\ell} - st_\alpha \leqslant \quad \delta \quad \leqslant \quad \bar{\ell} \quad + st_\alpha \,, \tag{3.1}$$

where $s^2 = \mathop{S}\limits_{i=1}^{n_1} (\ell_i - \bar{\ell})^2/n_1(n_1-1)$, and t_α is to be taken with (n_1-1) d.f. Whatever be the values of μ's, δ and σ_1^2, σ_2^2, the proportion of cases in which the statement of the form (3.1) to be true, will approximately be equal to α.

The unsatisfactory aspect of this solution lies in the fact that n_2-n_1 observations of a sample are discarded. Moreover it does not indicate whether it is possible to construct intervals which would be, in some sense, shorter than those of the form (3.1). The answer to this question is given by Scheffe (1943). His solution shares the obvious advantages of the solution mentioned by Neyman (1941), and is also free from the objection of the case when $n_1 \neq n_2$. Scheffe (1943) obtained his results in the following way.

3.1 Solution in Simple Case:

Let d_i $(i=1,2,\ldots,n_1)$, are independently and normally distributed random variables with mean δ and variance σ^2. Define P and θ by

$$P = \mathop{S}\limits_{i=1}^{n_1} \frac{d_i}{n_1}\ ; \quad \theta = \mathop{S}\limits_{i=1}^{n_1} (d_i - p)^2 .$$

Then $(P - \delta)/\sigma/\sqrt{n_1}$ will follow a standard normal distribution and θ/σ^2 will be χ^2 distributed with (n_1-1) d.f, both being independently distributed, the quantity $\dfrac{\sqrt{n_1}(P - \delta)}{\sqrt{\theta/(n_1-1)}}$, will then be distributed as a Student's t- distribution with $K = (n_1-1)$ d.f.

Let $\quad \Pr (-t_{k,\alpha} \leq t_k \leq t_{k,\alpha}) = \alpha$.

A set of confidence intervals for δ with a confidence coefficient α is

$$|P - \delta| \leq t_{(n_1-1),\alpha} \sqrt{\frac{\theta}{n_1(n_1-1)}} \quad . \qquad (3.2)$$

If $E(p)$ be the expected length of the confidence interval (3.2), then

$$E(p) = 2 \cdot t_{(n_1-1),\alpha} \left(n_1(n_1-1)\right)^{-\frac{1}{2}} \sigma \cdot E\,(\theta/\sigma^2)^{\frac{1}{2}}$$

$$= t_{n_1-1,\alpha} \cdot C_{n_1-1} \cdot \frac{\sigma}{\sqrt{n_1}} , \qquad (3.3)$$

where $C_{n_1-1} = \dfrac{\sqrt{(8/k)}\ \Gamma(k/2 + 1/2)}{\Gamma(k/2)}$.

The symmetrical choice of α will minimise $E(p)$.

Consider a linear function

$$d_i = X_i - \mathop{S}_{j=1}^{n_2} C_{ij} Y_j \quad (i=1,2,\ldots,n_1) , \qquad (3.4)$$

then d_i will have a multivariate normal distribution. The necessary and sufficient conditions that all d_i have means δ and variances σ^2 and covariance zero are,

$$\mathop{S}_{j=1}^{n_2} C_{ij} = 1, \quad \text{and} \quad \mathop{S}_{k=1}^{n_2} C_{ik}\, C_{kj} = C^2\, \delta_{ij} \quad ,$$

where $$\delta_{ij} = \begin{cases} 1 & \text{when } i=j \\ 0 & \text{when } i \neq j. \end{cases}$$

If a linear function d_i, as defined in (3.4), is used in finding the set of confidence intervals, then the expected length of the confidence interval $E(p)$, will be given by (3.3) with $\sigma^2 = \sigma_1^2 + C^2\,\sigma_2^2$. In order to minimise $E(p)$ we must find a matrix $n_1 \times n_2 = (C_{ij})$, satisfying the necessary and sufficient conditions and for which C^2 is minimum. The minimum value of C^2 is n_1/n_2 *.

Writing $$P = \bar{x} - \bar{y}\ , \quad \theta = \mathop{S}_{i=1}^{n_1} (z_i - \bar{z})^2 \ ,$$

where $\bar{x}$, $\bar{y}$ are the means of two samples, $z_i = x_i - \left(\frac{n_1}{n_2}\right)^{\frac{1}{2}} y_i$,

and $\bar{z} = \mathop{S}_{i=1}^{n_1} z_i/n_1$.

* For proof see appendix

The confidence interval, therefore, as in (3.2), is given by

$$|\bar{x} - \bar{y} - \delta| \leq t_{k,\alpha} \left[\frac{\theta}{n_1(n_1-1)}\right]^{\frac{1}{2}} . \qquad (3.5)$$

The expected length of the confidence interval, when $n_1 = n_2$ and $n_1 < n_2$ will be

$$t_{k,\alpha} \frac{\sigma}{\sqrt{n_1}} \cdot \frac{(\frac{8}{k})^{\frac{1}{2}} \Gamma(k/2 + 1/2)}{\Gamma(k/2)} , \qquad (3.6)$$

with $\sigma^2 = \sigma_1^2 + \sigma_2^2$ and $\sigma^2 = \sigma_1^2 + (\frac{n_1}{n_2}) \sigma_2^2$ respectively.

3.11 General Case:

Let P be a linear and θ be a quadratic form of the variates $(x_1, x_2, \ldots, x_{n_1})$ and $(y_1, y_2, \ldots, y_{n_2})$, with coefficients independent of the parameters. For some constant f, independent of the parameters and some function R of the parameters, the quantity $\frac{(P-\delta)}{R/f}$ will be distributed as a standard normal distribution and $\frac{\theta}{R^2}$ will follow a χ^2 law with $(k-1)$ d.f. Both are independently distributed, therefore the quotient

$$\frac{f(P-\delta)/R}{(\theta/(k-1)\ R^2)^{\frac{1}{2}}} , \qquad (3.7)$$

will be distributed as a Student's t - distribution with k-1 d.f. The sufficient condition for (3.7) to be a symmetric t - distribution with (k-1) d.f. is also discussed by Scheffe (1944). Obviously E(p) must be equal to δ and

$$\frac{f^2 E\,(P-\delta)^2}{R^2} = 1 \; . \tag{3.8}$$

The t - distribution of (3.7) leads to the confidence intervals

$$\left| P - \delta \right| \leqslant t_{k-1,\alpha} \, (\theta/(k-1))^{\frac{1}{2}} /f \; . \tag{3.9}$$

The expected length of (3.9) is then given by

$$E(p) = t_{k-1,\alpha} \, R/f \, . \, 2(k-1)^{-\frac{1}{2}} \, E\,(\theta/R^2)^{\frac{1}{2}}$$

$$= t_{k-1,\alpha} \, R/f \, . \, C_{k-1} \; , \tag{3.10}$$

where $\quad C_{k-1} = 2(k-1)^{-\frac{1}{2}} \, E(\theta/R^2)^{\frac{1}{2}} \; .$

If we write, $\quad P = \underset{i=1}{\overset{n_1}{S}} a_i x_i - \underset{i=1}{\overset{n_2}{S}} b_i y_i \; ,$

then $\quad E(p) = \mu_1 \underset{i=1}{\overset{n_1}{S}} a_i - \mu_2 \underset{i=1}{\overset{n_2}{S}} b_i \; . \qquad (3.11)$

From (3.8), $E(p) = \delta$, and a_i , b_i are independent of the parameters, therefore,

$$\underset{i=1}{\overset{n_1}{S}} a_i = \underset{i=1}{\overset{n_2}{S}} b_i = 1 \qquad (3.12)$$

Let $X_i = x_i - \mu_1$, $Y_i = y_i - \mu_2$,

then

$$P - \delta = \underset{i=1}{\overset{n_1}{S}} a_i X_i - \underset{i=1}{\overset{n_2}{S}} b_i Y_i \qquad (3.13)$$

and

$$E(P-\delta)^2 = \sigma_1^2 \underset{i=1}{\overset{n_1}{S}} a_i^2 + \sigma_2^2 \underset{i=1}{\overset{n_2}{S}} b_i^2 \ . \qquad (3.14)$$

From (3.8) and (3.14) we can write

$$R^2 = f^2 \sigma_1^2 \underset{i=1}{\overset{n_1}{S}} a_i^2 + f^2 \sigma_2^2 \underset{i=1}{\overset{n_2}{S}} b_i^2 \qquad (3.15)$$

$$= \sigma_1^2 a^2 + \sigma_2^2 b^2 ,$$

where $a^2 = f^2 \underset{i=1}{\overset{n_1}{S}} a_i^2$, $b^2 = f^2 \underset{i=1}{\overset{n_2}{S}} b_i^2$ are independent of the parameters.

By (3.15), the relation (3.10) may be written as

$$E(p) = t_{k-1,\alpha} C_{k-1} (\sigma_1^2 \underset{i=1}{\overset{n_1}{S}} a_i^2 + \sigma_2^2 \underset{i=1}{\overset{n_2}{S}} b_i^2)^{\frac{1}{2}}. \qquad (3.16)$$

3.12 Minimum Expected Length:

Amongst all confidence intervals of the form

$$|P - \delta| \leq t_{k-1,\alpha} \, (e/(k-1))^{\frac{1}{2}}/f \text{ with } k=n_1 \text{ , the}$$

confidence intervals (3.5), have the minimum expected length. The coefficients a_i and b_i in (3.16) are subject to the restriction

$$\mathop{S}_{i=1}^{n_1} a_i = \mathop{S}_{i=1}^{n_2} b_i = 1 \text{ , therefore}$$

$$\mathop{S}_{i=1}^{n_1} a_i^2 \geq \frac{1}{n_1} \quad , \quad \mathop{S}_{i=1}^{n_2} b_i \geq \frac{1}{n_2} \quad . \qquad (3.17)$$

From (3.17) and (3.16) we get,

$$E(p) \geq t_{n_1-1,\alpha} \, C_{n_1-1} \frac{(\sigma_1^2 + (\frac{n_1}{n_2}) \, \sigma_2^2)^{\frac{1}{2}}}{(n_1^{\frac{1}{2}})}$$

or

$$E(p) \geq t_{n_1-1,\alpha} \, C_{n_1-1} \frac{\sigma}{\sqrt{n_1}} \quad . \qquad (3.18)$$

The relation (3.18) proves the assertion that within all the confidence intervals of the form (3.9), the expected length of the confidence intervals (3.5) is minimum.

3.13 Asymptotic Shortness of the Confidence Intervals (3.5).

Let $\Sigma_1 = \mathop{S}\limits_{i=1}^{n_1} (x_i-\bar{x})^2$, $\Sigma_2 = \mathop{S}\limits_{i=1}^{n_2} (y_i-\bar{y})^2$,

$P = \bar{x} - \bar{y}$, and $\sigma_p^2 = \sigma_1^2/n_1 + \sigma_2^2/n_2$.

The quantities $(P-\delta)/\sigma_p$, Σ_1/σ_1^2 and Σ_2/σ_2^2 are mutually independent and are distributed as $N(0,1)$, $\chi^2_{(n_1-1)}$ and $\chi^2_{(n_2-1)}$ respectively. Therefore

$$t_{(n_1+n_2-2)} = \frac{(P-\delta)\quad(n_1+n_2-2)^{\frac{1}{2}}}{\left[\sigma_p^2\;(\Sigma_1/\sigma_1^2 + \Sigma_2/\sigma_2^2)\right]^{1/2}} \tag{3.19}$$

The efficient and shortest confidence interval available, is given by

$$|P-\delta| \leq t_{(n_1+n_2-2),\alpha}\;(n_1+n_2-2)^{-\frac{1}{2}}\;\left[(\sigma_1^2/n_1+\sigma_2^2/n_2)\chi^2_{n_1+n_2-2}\right]^{\frac{1}{2}} , \tag{3.20}$$

with the expected length

$$E(p) = t_{(n_1+n_2-2),\alpha} \cdot \frac{(\sigma_1^2 + (\frac{n_1}{n_2})\;\sigma_2^2)^{\frac{1}{2}}}{(n_1^{\frac{1}{2}})} \cdot C_{(n_1+n_2-2)} , \tag{3.21}$$

where $$C_{(n_1+n_2-2)} = \frac{2E(\chi_{n_1+n_2-2})}{\sqrt{n_1+n_2-2}} .$$

The ratio L, of the expected lengths of the confidence intervals (3.5) and (3.20), is then

$$L = \left(t_{(n_1-1),\alpha} \, C_{(n_1-1)} \right) / \left(t_{(n_1+n_2-2),\alpha} \, C_{(n_1+n_2-2)} \right). \tag{3.22}$$

The behaviour of L, for specific values of n_1, n_2 and α can be studied from the relation (3.22). The percentage by which the expected length of confidence interval (3.5) is greater than the available optimum confidence interval length (3.21), can also be calculated. It can safely be concluded from the fact, $L \longrightarrow 1$ when $n_1 \longrightarrow \infty$, that the confidence intervals (3.5) are at least asymptotically efficient.

3.2 Confidence Interval for a Linear Function of Population Means:

Banerjee (1960) obtained a confidence interval for a linear function of the population means based on the sample estimates and the Student' t - table values. He indicated the method for the case of two samples, and extended his results in Banerjee (1961) and comparing them with the existing Fisher (1935) and Welch (1947) solutions. Banerjee (1960, 61) results are based on a property of the convex function and have been obtained as follows.

3.21 Two Samples Case:

Let $(\bar{x}_1+\bar{x}_2-\mu_1-\mu_2)^2 \leqslant \frac{t_1^2\, s_1^2}{n_1} + \frac{t_2^2\, s_2^2}{n_2}$ be an event,

and $\Pr(-t_i \leqslant t \leqslant t_i) = \alpha$, $(i=1,2)$, where t_1 and t_2 are Student's t - table values with (n_i-1) d.f. and confidence coefficient α. If P is the probability of the event considered, we can then write

$$\Pr\left(\frac{(\bar{x}_1+\bar{x}_2 - \mu_1-\mu_2)^2}{\sigma_1^2/n_1 + \sigma_2^2/n_2} \leqslant \frac{\frac{t_1^2 s_1^2}{n_1} + \frac{t_2^2 s_2^2}{n_2}}{\sigma_1^2/n_1 + \sigma_2^2/n_2} \right) = P\,. \tag{3.23}$$

For fixed s_1^2 and s_2^2, the quantity $\frac{(\bar{x}_1+\bar{x}_2 - \mu_1-\mu_2)^2}{\sigma_1^2/n_1 + \sigma_2^2/n_2}$ is distributed as a χ^2 with 1 d.f. The probability P is given by,

$$P = \int_0^\infty\!\!\int_0^\infty f_1(s_1^2,\, \sigma_1^2,\, n_1)\, g(s_1^2,\, s_2^2)\, d\, s_1^2\;\; f_2(s_2^2,\, \sigma_2^2,\, n_2)\, d\, s_2^2\,, \tag{3.24}$$

where $f_1(s_1^2, \sigma_1^2, n_1)$, $f_2(s_2^2, \sigma_2^2, n_2)$ are the probability density functions of s_1^2 and s_2^2, and

$$g(s_1^2, s_2^2) = \int_0^A \frac{1}{2\Gamma(1/2)} (\chi^2/2)^{\frac{1}{2}-1} e^{-\chi^2/2} d\chi^2 .$$

Here
$$A = \frac{\sum_{i=1}^{2} t_i^2 s_i^2/n_i}{\sum_{i=1}^{2} \sigma_i^2/n_i}, \quad (i=1,2) .$$

Since $\int_0^A f(\chi^2)\, d\chi^2$ is a convex function in A, therefore,

$$\int_0^A f(\chi^2)\, d\chi^2 = g(s_1^2, s_2^2) \geqslant \frac{\sigma_1^2/n_1}{\sigma_1^2/n_1 + \sigma_2^2/n_2} \int_0^{t_1^2 s_1^2/\sigma_1^2} f(\chi^2).d\chi^2$$

$$+ \frac{\sigma_2^2/n_2}{\sigma_1^2/n_1 + \sigma_2^2/n_2} \int_0^{t_2^2 s_2^2/\sigma_2^2} f(\chi^2)\, d\chi^2 . \tag{3.25}$$

We have

$$\Pr(-t_i \leqslant t \leqslant t_i) = \alpha = \int_0^{t_i^2 s_i^2/\sigma_i^2} f(\chi^2)\, d\chi^2 , \quad (i=1,2). \tag{3.26}$$

From (3.25) and (3.26) it follows

$$\int_0^A f(\chi^2)\, d\chi^2 \geqslant \frac{\sigma_1^2/n_1}{\sigma_1^2/n_1 + \sigma_2^2/n_2}(\alpha) + \frac{\sigma_2^2/n_2}{\sigma_1^2/n_1 + \sigma_2^2/n_2}(\alpha) ,$$

or $$\int_0^A f(\chi^2)\, d\chi^2 \geqslant \alpha . \qquad (3.27)$$

The relations (3.23), (3.25) and (3.27) $\Rightarrow$

$$\Pr\left((\bar{x}_1+\bar{x}_2 - \mu_1-\mu_2)^2 \leqslant \mathop{S}_{i=1}^{2} \frac{t_i^2\, s_i^2}{n_i}\right) \geqslant \alpha . \qquad (3.28)$$

If a_1 and a_2 are some known constants, (3.28) can then be written as follows

$$\Pr\left(\mathop{S}_{i=1}^{2} a_i\bar{x}_i - \sqrt{\mathop{S}_{i=1}^{2} \frac{t_i^2 s_i^2}{n_i}} \leqslant \mathop{S}_{i=1}^{2} a_i\mu_i \leqslant \mathop{S}_{i=1}^{2} a_i\bar{x}_i + \sqrt{\mathop{S}_{i=1}^{2} \frac{t_i^2 s_i^2}{n_i}}\right) \geqslant \alpha.$$

Assume $a_1=1$, $a_2=\pm 1$, obviously (3.28) reduces to

$$\Pr\left(\left| (\bar{x}_1-\bar{x}_2) - (\mu_1-\mu_2) \right| \leqslant \sqrt{\frac{t_1^2 s_1^2}{n_1} + \frac{t_2^2 s_2^2}{n_2}} \right) \geqslant \alpha. \qquad (3.29)$$

The expression (3.29) is a confidence interval for the difference in two population means in terms of their sample estimates and Student's t - distributions.

3.22 General Solution:

<u>Theorem</u>: - If z be a standard normal variable and $\chi^2_{v_i}$ $(i=1,2,\ldots,R)$ be distributed as χ^2 variates, mutually independent and independent of z, with $v_i (i=1,2,\ldots,R)$ d.f. and $W_i (i=1,2,\ldots,R)$ be a set of arbitrary weights with the condition

$$\mathop{S}_{i=1}^{R} W_i = 1, \quad W_i \geqslant 0,$$

then

$$\Pr\left(z^2 \leq \mathop{S}_{i=1}^{R} \frac{t_i^2}{v_i} W_i \chi^2_{v_i}\right) \geqslant \alpha ,$$

where t_i are the Student's t - table values with v_i d.f. following $\Pr(-t_i \leqslant t \leqslant t_i) = \alpha$, $(i=1,2,\ldots,R)$.

<u>Proof</u>: The probability of the event

$$z^2 \leq \mathop{S}_{i=1}^{R} \frac{t_i^2}{v_i} W_i \chi^2_{v_i} ,$$

is given as follows.

$$\Pr\left(z^2 \leq \mathop{S}_{i=1}^{R} \frac{t_i^2}{v_i} W_i \chi^2_{v_i}\right)$$

$$= \int_0^\infty \ldots\ldots \int_0^\infty \prod_{i=1}^{R} f(\chi^2_{v_i})\, d\chi^2_{v_1} \ldots\ldots d\chi^2_{v_R} \left(\int_0^A f(\chi^2)\, d\chi^2 \right) , \tag{3.30}$$

where $A = \mathop{S}\limits_{i=1}^{R} (t_i^2/v_i) W_i \chi^2_{v_i}$ and $f(\chi^2)$ is the probability density function of χ^2 variate with one d.f.

Since $\int_0^A f(\chi^2)\, d\,\chi^2$ is a convex function in A,

therefore,

$$\int_0^A f(\chi^2)\, d\,\chi^2 \geqslant \mathop{S}\limits_{i=1}^{R} W_i \int_0^{(t_i^2/v_i)\chi^2_{v_i}} f(\chi^2)\, d\,\chi^2 . \tag{3.31}$$

we have

$$Pr(-t_i \leqslant t \leqslant t_i) = \alpha = \int_0^{(t_i^2/v_i)\chi^2_{v_i}} f(\chi^2)\, d\chi^2 , \quad (i=1,2,..R). \tag{3.32}$$

From (3.31) and (3.32) we obtain

$$\int_0^A f(\chi^2)\, d\,\chi^2 \geqslant \mathop{S}\limits_{i=1}^{R} W_i\, \alpha .$$

or

$$\geqslant \alpha , \quad \text{Since } \mathop{S}\limits_{i=1}^{R} W_i = 1. \tag{3.33}$$

The relation (3.30) and (3.33) $\Rightarrow$

$$Pr\left(z^2 \leqslant \mathop{S}\limits_{i=1}^{R} (t_i^2/v_i) W_i \chi^2_{v_i} \right) \geqslant \alpha. \tag{3.34}$$

<u>Theorem</u>: Let X be a random variable following a normal probability law with mean value δ and variance

$$\overset{R}{\underset{i=1}{S}} a_i \sigma_i^2 + \overset{\ell}{\underset{j=1}{S}} b_j \sigma_j^2 \text{, where}$$

$$a_i \text{ , } b_j \quad (i=1,2,\ldots,R;\ j=1,2,\ldots,\ell),$$

are some known positive constants. If s_i^2 are the estimates of σ_i^2, where $v_i s_i^2/\sigma_i^2$ are distributed as χ^2 with v_i d.f. and are mutually independent and also independent of X and if σ_j^2 be known, then

$$\Pr\left((X-\delta)^2 \leq \overset{R}{\underset{i=1}{S}} t_i^2\, a_i\, s_i^2 + \overset{\ell}{\underset{j=1}{S}} d^2\, b_j\, \sigma_j^2\right) \geq \alpha,$$

where t_i and d are the values of Student's t - distribution with v_i d.f. and a standard normal deviate with the confidence coefficient α, respectively.

<u>Proof</u>: Consider the probability of the event as given below,

$$\Pr\left[\frac{(X-\delta)^2}{(\overset{R}{\underset{i=1}{S}} a_i\sigma_i^2 + \overset{\ell}{\underset{j=1}{S}} b_j\sigma_j^2)} \leq \frac{\overset{R}{\underset{i=1}{S}} t_i^2\, a_i s_i^2 + \overset{\ell}{\underset{j=1}{S}} d^2 b_j \sigma_j^2}{(\overset{R}{\underset{i=1}{S}} a_i\sigma_i^2 + \overset{\ell}{\underset{j=1}{S}} b_j\, \sigma_j^2)}\right]$$

$$= \int_0^\infty \cdots \int_0^\infty \prod_{i=1}^{R} f(\chi_{v_i}^2)\, d\, \chi_{v_i} \int_0^A f(\chi^2)\, d\, \chi^2, \qquad (3.35)$$

where

$$A = \frac{\underset{i=1}{\overset{R}{S}} t_i^2\, a_i\, s_i^2 + \underset{j=1}{\overset{\ell}{S}} d^2\, b_j\, \sigma_j^2}{\left(\underset{i=1}{\overset{R}{S}} a_i\, \sigma_i^2 + \underset{j=1}{\overset{\ell}{S}} b_j\, \sigma_j^2 \right)} \qquad (3.36)$$

and $f(\chi^2)$ is the probability density function (p.d.f.) of χ^2 variable with one d.f.

The relation (3.36) can also be written as follows.

$$A = \underset{i=1}{\overset{R}{S}} (t_i^2/v_i)\, W_i\, \chi^2_{v_i} + d^2 \underset{j=1}{\overset{\ell}{S}} W_j\,, \qquad (3.37)$$

where $$W_i = \frac{a_i\, \sigma_i^2}{\underset{i=1}{\overset{R}{S}} a_i \sigma_i^2 + \underset{j=1}{\overset{\ell}{S}} b_j \sigma_j^2}\,,$$

$$W_j = \frac{b_j\, \sigma_j^2}{\underset{i=1}{\overset{R}{S}} a_i \sigma_i^2 + \underset{j=1}{\overset{\ell}{S}} b_j \sigma_j^2}$$

and $\underset{i=1}{\overset{R}{S}} W_i + \underset{j=1}{\overset{\ell}{S}} W_j = 1.$

Since $\int_0^A f(\chi^2)\, d\,\chi^2$ is a convex function in A,

therefore we may write

$$\int_0^A f(\chi^2)\,d\chi^2 \geqslant \mathop{S}_{i=1}^{R} W_i \int_0^{(t_i^2/v_i)\chi^2_{v_i}} f(\chi^2)\; d\chi^2 + \mathop{S}_{j=1}^{\ell} W_j \int_0^{d^2} f(\chi^2)\, d\,\chi^2. \tag{3.38}$$

We have

$$\int_0^{\left(\frac{t_i}{v_i}\right)\chi^2_{v_i}} f(\chi^2)\; d\,\chi^2 = \alpha = \int_0^{d^2} f(\chi^2)\; d\,\chi^2. \tag{3.39}$$

From (3.35), (3.38) and (3.39), it follows that

$$\Pr((X-\delta)^2 \leqslant \mathop{S}_{i=1}^{R} t_i^2\, a_i\, s_i^2 + \mathop{S}_{j=1}^{\ell} d^2\, b_j\, \sigma_j^2) \geqslant \alpha \quad .$$

3.3 **Comparison of the Power Functions of Two Tests:**

Scheffe' (1943) solution for Behrens - Fisher problem, as discussed in §(3.1), is based on a Student's t - distribution and possesses certain desirable properties. In his solution the numerator is a difference of means of the observations while the denominator is square root of the function of sample values having a χ^2 - distribution with (n_1-1) d.f. Walsh (1949) has compared the power function of Scheffe' (1943) test with the power function of a most powerful (when $\frac{\sigma_1}{\sigma_2}$ is known) t - test. His comparison is based on a modification of the normal approximation to the power function of one sided t - test, given by Johnson and Welch (1940). Walsh (1949) obtained the power efficiency of one sided t - tests. Since it is shown by Walsh (1949) that a symmetrical t - test with significance level 2α has the same power efficiency as that of a one sided t - test with significance level α. The explicit formula obtained by Walsh (1949), for calculating approximate power efficiency for some preassigned α and different values of the sample sizes, is arrived at in the following way.

A power efficiency of 100 E% means that the given test, based on n_1 and n_2 observations, has approximately the same power function as that of the corresponding most powerful test based on the sample sizes En_1 and En_2.

The problem then is to evaluate E such that, a most powerful test (under same hypothesis and significance level), based on En_1 and En_2 observations, will have approximately the same power function as that of a given t - test based on n_1 and n_2 observations. The power efficiency of the given test will then be equal to 100 E%.

Scheffe (1943) one sided t - test, and corresponding most powerful one sided t - test, have same power function when E is so choosen that under H_o : $\mu_1 = \mu_2$,

$$K_\alpha - \delta \sqrt{E} \left(1 - \frac{K_\alpha^2/2}{(n_1E + n_2E - 2)}\right)^{\frac{1}{2}} = K_\alpha - \delta \left(1 - \frac{K_\alpha^2/2}{(n_1 - 1)}\right)^{\frac{1}{2}} . \quad (3.40)$$

The expression (3.40) is obtained by using a modification to the normal approximation given by Johnson and Welch (1940). The quantity δ is a function of n_1, n_2, μ_1, μ_2 & $R = \frac{\sigma_1^2}{\sigma_2^2}$. The significance level of the tests is α and K_α is the critical value corresponding to α of standard normal distribution. The accuracy of the approximation involved, in equality of power functions of two tests, increases with the increase in n_1.

From (3.40) E can be evaluated as follows,

$$E \left(1 - \frac{K_\alpha^2/2}{(n_1E + n_2E - 2)} \right) = \left(1 - \frac{K_\alpha^2/2}{(n_1 - 1)} \right) , \quad (3.41)$$

writing $1 - \dfrac{K_\alpha^2/2}{(n_1-1)} = B$, (3.41) becomes

$$E^2(n_1+n_2) - E\,(2 + B(n_1+n_2) + K_\alpha^2/2) + 2B = 0 \qquad (3.42)$$

Solving (3.42) for E, we obtain

$$E = \frac{1}{2(n_1+n_2)}\left((2+B(n_1+n_2) + K_\alpha^2/2) + \sqrt{(2+B(n_1+n_2)+K_\alpha/2)^2 - 8(n_1+n_2)B}\right).$$

Thus the approximate percentage efficiency of Scheffe's (1943) one sided t - test compared with the Student's t - test, when R is known, therefore, is given by

$$\frac{50}{(n_1+n_2)}\left(2+B(n_1+n_2)+K_\alpha^2/2 + \sqrt{(2+B(n_1+n_2)+K_\alpha^2/2)^2 - 8(n_1+n_2)B}\right)\% \,. \qquad (3.43)$$

for appropriate values of α & n, $n_1 \leq n_2$.

3.4 Comparison of the Expected Lengths of Confidence Intervals of Two Tests:

The expected length of the confidence interval of Scheffe (1943) Solution for Behrens-Fisher problem is given at (3.6). An equivalent approximate (because of asymptotic series) expression for Welch(1947) solution has been obtained by James (1966). On the basis of these two expected lengths a criterion is developed by him from which it is possible to decide which of the two solutions is more appropriate in the prevailing situation. James (1966) procedure of measuring the relative test efficiency is based on the solution of expected value of the linear function of χ^2 variates by hypergeometric function. James (1966) obtained the criterion for measuring the approximate relative test efficiency in the following way.

The relation (3.5) gives the confidence interval for Scheffe (1943) solution. The expected length of the confidence interval (3.5), when $n_1 < n_2$, may be written as

$$E(p) = \frac{2t_{k,\alpha}}{n_1(n_1-1)} \; E\left(\frac{S_{i=1}^{n_1}(z_i-\bar{z})^2}{\sigma^2}\right)^{\frac{1}{2}} \sigma \quad , \qquad (3.44)$$

where $\sigma^2 = \sigma_1^2 + (n_1/n_2)\ \sigma_2^2$ & $k=(n-1)$.

Applying the relation $E(\chi_{n_1-1}) = \dfrac{\sqrt{2}\ \Gamma((n_1-1)/2 + 1/2)}{\Gamma((n_1-1)/2)}$,

the expression (3.44) reduces to,

$$E(p) = \frac{2\sqrt{2}\ \ t_{k,\alpha}\ \sigma_1\ \sqrt{1+(\frac{n_1}{n_2})\ \gamma^2}\ \ \Gamma(\frac{n_1}{2})}{\Gamma((n_1-1)/2)\ \sqrt{n_1(n_1-1)}}\ , \qquad (3.45)$$

where $\quad \gamma^2 = \dfrac{1}{R} = \dfrac{\sigma_2^2}{\sigma_1^2}$

Welch (1947) obtained a quantity $h(s_1^2,\ s_2^2,\ \alpha)$, in a similar way as Gosset (Student (1908)), who derived for a single sample, the expression $\Pr((\bar{x}-\mu) < t_\alpha \frac{s}{\sqrt{n}})$. The quantity $h(s_1^2,\ s_2^2,\ \alpha)$ is a function of s_1^2, s_2^2 and α, but independent of γ with the property,

$$\Pr\left((\bar{x}_1-\bar{x}_2)-(\mu_1-\mu_2) < h(s_1^2,\ s_2^2,\ \alpha)\right) = \alpha\ .$$

Welch (1947) asymptotic series for calculating h does not converge. His results yield,

$$h = \eta \sqrt{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)} \left(1 + \frac{(1+\eta^2) \left(\frac{s_1^4/n_1^2}{f_1} + \frac{s_2^4/n_2^2}{f_2}\right)}{4(s_1^2/n_1 + s_2^2/n_2)^2} \right.$$

$$\left. - \frac{(1+\eta^2)\left(\frac{s_1^4/n_1^2}{f_1^2} + \frac{s_2^4/n_2^2}{f_2^2}\right)}{2(s_1^2/n_1 + s_2^2/n_2)^2} + \cdots \right) , \quad (3.46)$$

where η is a standard normal deviate such that

$$\Pr \left(N(0,1) \leq \eta \right) = \alpha, \ \& \ f_i = (n_i - 1), i = 1, 2.$$

Substituting $s^2 = s_1^2/n_1 + s_2^2/n_2$ in (3.46), we get

$$h = \eta\, s \left[1 + \frac{(1+\eta^2)}{4} \left(\frac{(s^{-2}s_1^2n_1^{-1})^2}{(n_1-1)} + \frac{(s^{-2}s_2^2n_2^{-1})^2}{(n_2-1)} \right) \right.$$

$$\left. - \frac{(1+\eta^2)}{2} \left(\frac{(n_1^{-1}s_1^2s^{-2})^2}{(n_1-1)^2} + \frac{(n_2^{-1}s^{-2}s_2^2)^2}{(n_2-1)^2} \right) + \cdots \right] .$$

The expected length of the confidence interval for Welch (1947) procedure is then given by

$$2E(h) = 2 \left[\eta\, E(s) + (\eta/4)(1+\eta^2) \left(E\left(\frac{s_1^4}{n_1^2 s^3 (n_1-1)}\right) + E\left(\frac{s_2^4}{n_2^2 s^3 (n_2-1)}\right) \right) \right.$$

$$\left. + \cdots \quad . \right] . \qquad (3.47)$$

3.41 Procedure for Evaluating A_1:

The substitution $\mathcal{S}^2 = (s_1^2/n_1 + s_2^2/n_2)$, is of the form $(K_1 x + K_2 y)$. Let x and y be independent χ^2 - distributed variates with (n_1-1) and (n_2-1) d.f and K_1, K_2 be two positive constants.

The expected value of $(K_1 x + K_2 y)^{\frac{1}{2}}$ is given by

$$E(K_1x+K_2y)^{\frac{1}{2}} = \frac{1}{2^{(n_1+n_2-2)/2}\Gamma(\frac{n_1-1}{2})\ \Gamma(\frac{n_2-1}{2})} \times$$

$$\int_0^\infty \int_0^\infty (K_1x+K_2y)^{\frac{1}{2}}\ x^{(\frac{n_1-1}{2})-1}\ y^{(\frac{n_2-1}{2})-1}\ e^{-(\frac{x+y}{2})}\ dx\ dy\ . \qquad (3.50)$$

Transforming $z=x$, $t= \frac{x}{(x+y)^{\frac{1}{2}}}$ and applying Gamma-function the expression (3.50) reduces to

$$E(K_1x+K_2y)^{\frac{1}{2}} = \frac{\sqrt{2K_1}\ \Gamma((n_1+n_2-1)/2)}{\Gamma(\frac{n_1-1}{2})\quad \Gamma(\frac{n_2-1}{2})} \times$$

$$\int_0^1 t^{(n_1-3)/2}(1-t)^{(n_2-3)/2}\ (t + \frac{K_2}{K_1} - \frac{tK_2}{K_1})^{\frac{1}{2}}\ dt.$$

(3.51)

From the ratio of (3.47) and (3.45), a measure of the relative efficiency $\mathcal{E}$, can then be obtained as

$$\mathcal{E} = \frac{zE(h)}{E(p)}. \tag{3.48}$$

The smaller value of $\mathcal{E}$ indicates that Welch (1947) procedure is better, but a larger value favours Scheffe (1943) solution.

Writing

$$\frac{E(s)}{\sigma_1} = A_1 \,, \quad \frac{E(s_1^4/s^3)}{n_1^2(n_1-1)\sigma_1} = A_2 \,, \quad \frac{E(s_2^4/s^3)}{n_2^2(n_2-1)\sigma_1} = A_3 \,,$$

and neglecting the higher order terms in (3.47), the relative efficiency $\mathcal{E}$ may then be written in the following form,

$$\mathcal{E} \simeq \frac{\Gamma(\tfrac{1}{2}(n_1-1))\sqrt{n_1(n_1-1)}}{\sqrt{2}\; t_{(n_1-1),\alpha}\sqrt{1+(\frac{n_1}{n_2})\gamma^2}\;\Gamma(\frac{n_1}{2})}\left[\eta/_4(1+\eta^2)(A_2+A_3)+\eta A_1\right]. \tag{3.49}$$

The hypergeometric function by definition is

$$F(a;b;c;z)= \frac{S_{n=0}^{\infty}(a)_n(b)_n z^n}{(c)_n\ n\ !}, \text{ where } \quad |z| < 1, \qquad (3.52)$$

with the notations

$(a)_n = a(a+1)(a+2),\ldots,(a+n-1), \quad n \geq 1$, and $(\alpha)_0=1,\ \alpha\neq 0$.

The special case of (3.52), when a=c, b=1 yields a geometric series, $S_{n=0}^{\infty}\ z^n$.

By the property of hypergeometric function we know (Rainville, P:45) that, if $|z| < 1$ and $c > b > 0$, then (3.52) may be written as

$$F(a,b;c;z)= \frac{\Gamma c}{\Gamma b\,\Gamma(c-b)} \int_0^1 t^{b-1}(1-t)^{c-b-1}(1-tz)^{-a}\,dt\ . \qquad (3.53)$$

Making transformation $r = (1-t)$ in (3.51), we obtain,

$$\int_0^1 r^{(n_2-3)/2}(1-r)^{(n_1-3)/2}\left[1-r(1-K_2/K_1)\right]^{\frac{1}{2}} dr. \qquad (3.54)$$

We assume first $K_2 < 2K_1$, then since $K_1,\ K_2 > 0$,

$0 < K_2 < 2K_1 \Longrightarrow \left|1-\frac{K_2}{K_1}\right| < 1$, so after comparing (3.53) with (3.54), we can write

$$(a,b;c) \equiv (-\tfrac{1}{2},\ (n_2-1)/2;\ (n_1+n_2-2)/2). \qquad (3.55)$$

The expression (3.51) can also be written as follows:

$$(K_2/K_1)^{\frac{1}{2}} \int_0^1 t^{(n_1-3)/2}(1-t)^{(n_2-3)/2}\left(1-t\left(1-\frac{K_1}{K_2}\right)\right)^{\frac{1}{2}} dt.$$

(3.56)

Assume now $2K_1 \leq K_2 \Longrightarrow \left|1 - \frac{K_1}{K_2}\right| < 1$, comparing again (3.53) with (3.56) we get,

$$(a,b;c) = (-\tfrac{1}{2},\ (n_1-1)/2\ ;\ (n_1+n_2-2)/2)\ . \qquad (3.57)$$

By defining $\delta = \dfrac{\Gamma(^1/_2(n_1+n_2-1))\ \sqrt{2}}{\Gamma(^1/_2(n_1+n_2-2))}$, the expected value of $(K_1x + K_2y)$ may then be written in the following form,

$$E(K_1x+K_2y)^{\frac{1}{2}} = \begin{cases} \delta\sqrt{K_1}\ F\left(-\tfrac{1}{2},\ \dfrac{(n_2-1)}{2};\ \dfrac{(n_1+n_2-2)}{2};\ \left(1-\dfrac{K_2}{K_1}\right)\right), & \text{when } K_2 < 2K_1\ , \\ \delta\sqrt{K_2}\ F\left(-\tfrac{1}{2},\ \dfrac{(n_1-1)}{2};\ \dfrac{(n_1+n_2-2)}{2};\left(1-\dfrac{K_1}{K_2}\right)\right), & \text{when } 2K_1 \leq K_2\ . \end{cases}$$

Let $K_1 = \dfrac{\sigma_1^2}{n_1(n_1-1)}$; $K_2 = \dfrac{\sigma_2^2}{n_2(n_2-1)}$.

We have $\sigma_1 A_1 = E(\sigma)$, which is given by $E(K_1x+K_2y)^{\frac{1}{2}}$.

Similarly A_2 and A_3 are calculated with the constants

$$\delta_1 = \frac{\sqrt{2}\ \Gamma((n_1+n_2-1)1/2)\ \Gamma(\frac{1}{2}(n_1+3))}{n_1(n_1-1)^2\Gamma(\frac{1}{2}(n_1-1))\ \Gamma(\frac{(n_1+n_2+2)}{2})} ,$$

and

$$\delta_2 = \frac{\sqrt{2}\ \gamma\Gamma(\frac{1}{2}(n_1+n_2-1))\ \Gamma(\frac{1}{2}(n_2+3))}{n_2\ (n_2-1)^2\ \Gamma(\frac{1}{2}(n_2-1))\ \Gamma(\frac{1}{2}(n_1+n_2+2))} .$$

The values of A_2 and A_3 are given by the expressions

$$A_2 = \begin{cases} \delta_1(n_1(n_1-1))^{\frac{1}{2}}\ F(3/2,\ (n_1-1)/2;\ \frac{(n_1+n_2+2)}{2};\ \frac{K_1-K_2}{K_1}) , & \text{when } K_2 < 2K_1 , \\ \frac{\delta_1(n_2(n_2-1))^{\frac{3}{2}}}{n_1(n_1-1)\ \gamma^3}\ F(\frac{3}{2},\ \frac{n_1+3}{2};\ \frac{n_1+n_2+2}{2};\ (1-\frac{K_1}{K_2}), & \text{when } 2K_1 \leqslant K_2 , \end{cases}$$

and

$$
A_3 = \begin{cases} \dfrac{\delta_2(n_1(n_1-1))^{\frac{3}{2}}\gamma^3}{n_2(n_2-1)} \; F\left(3/2, \dfrac{n_2+3}{2}; \dfrac{n_1+n_2+2}{2}; \dfrac{K_1-K_2}{K_1}\right), & \text{when } K_2 < 2K_1, \\[2em] \delta_2(n_2(n_2-1))^{\frac{1}{2}} \; F\left(\dfrac{3}{2}, (n_1-1)/2; \dfrac{n_1+n_2+2}{2}; \dfrac{K_2-K_1}{K_2}\right), & \text{when } 2K_1 \leq K_2. \end{cases}
$$

The relative efficiency $\mathcal{E}$ can now be computed by assigning the specific values to n_1, n_2, γ and at some preassigned levels of significance of the tests.

For instance if $\alpha = 0.95$, then $\eta = 1.64$, and $t_{(n_1-1),\alpha}$ is the 0.95 point of one sided t - distribution with (n_1-1) d.f. By fixing them and taking some values of n_1 & n_2 and appropriately choosing γ, a comparative study of the two test considered, can be made. It will be possible, then, to ascertain which test would control type I error more efficiently than the other, under prevailing conditions.

3.5 References:

Banerjee, S.K. (1960), Approximate confidence interval for linear function of means of K - populations when variances are not equal, Sankhya, Vol 22 Part III & IV, 357-59.

Banerjee, S.K. (1961), On confidence interval for two means problem based on separate estimates of variances and tabulated values of t - table, Sankhya, Vol 23, Part IV, 359-378.

Fisher, R.A. (1935), The fiducial arguments in statistical inference, Ann, Eugen, Vol 6, 391-398.

Johnson, W.L., & Welch, B.L. (1940), Application of the non central t - distribution, Biometrika, Vol 31, 376-

James, S. (1966), A confidence interval comparison of two test procedures proposed for the Behrens - Fisher problem, J. Amm. Statist. Assoc., Vol 61, 454-465.

Neyman, J. (1941), Fiducial arguments and the theory of confidence interval, Biometrika, Vol 32, 128-140.

Rainville, Earl, D. (1960), Special functions, The Macmillian Co., New York.

Scheffe, H. (1943), On solution of the Behrens - Fisher problem, Ann. Math. Statist., Vol 14, 35-44.

Scheffe, H. (1944), A note on the Behrens - Fisher problem, Ann. Math. Statist., Vol 15, 430-32.

Student (1908), The probable error of a mean, Biometrika, Vol 6, 1-25.

Welch, B. L. (1938), The significance of the difference between two means when population variances are unequal, Biometrika, Vol 29, 350-362.

Welch, B. L. (1947), The generalisation of Student's problem when several different population variances are involved, Biometrika, Vol 34, 28-35.

Walsh, John E. (1949), On the power function of "Best" t - test solution of the Behrens-Fisher problem, Ann. Math. Statist., Vol 20, 616-618.

Walsh, John E. (1949), Some significance tests for the median which are valid under very general conditions, Ann. Math. Statist., Vol 20, 64-81.

CHAPTER 4

FIDUCIAL APPROACH

4.0 Introduction:

The concept of fiducial probability distribution (a distribution of "trust") was introduced by R. A. Fisher (1930). In his original paper entitled "Inverse probability" he discussed the importance of maximum likelihood method and then produced a fiducial distribution for a parameter. Fraser (1961a) has reviewed some problems analysised by fiducial method and obtained the results for them after putting forward a mathematical frame-work within which fiducial probability has a frequency interpretation. In his paper, Fraser (1961b) has examined the logical requirements of fiducial distributions by setting up a transformation model which generates fiducial distributions. The initial development and discussion of fiducial analysis from transformation model does not appear in Fisher's own writings, but can be found in Fraser (1961a,b) who stresses (Fraser (1963)) that there is a method of inference underlying the fiducial writings of Fisher which was only partially realised and this method can be derived from Fisher's writings for the purpose of scientific inference.

4.1 Transformation Model:

The requirements put forward by Fraser (1961b) may briefly be given in the following way. Consider a basic sample space on which there are probability distributions with a parameter θ which takes values in a parameter space Y. Let the sufficient statistic exist and take values in the derived sample space X. Let now there be a group G of transformations on the sample space. A class G of transformations is a group if

(i) $g, h \in G \Rightarrow h \circ g \in G$, where $h \circ g$ is a composite transformation.

(ii) $g \in G \Rightarrow \exists g^{-1} \in G$.

Also suppose the following properties hold for this class of transformations.

(i) The transformations on the basic sample space induce transformations on the values of sufficient statistic i.e. these transformations can be conceived as applying to the space X.

(ii) There is a unique transformation which takes any $x \in X$ into another point $x' \in X$.

(iii) A transformation g carries a variable x with a distribution θ into a variable gx having a distribution $g^* \theta \in Y$. There exists an unique transformation which takes any point $\theta \in Y$ into another point $\theta' \in Y$. As an example let $(x_1 \ldots x_n)$ be a sample from a normal population with mean μ and standard deviation σ, both are unrestricted. The parameter point is $\theta = (\mu, \sigma)$, with parameter space $Y = (-\infty, \infty) \times (0, \infty)$. The sample mean $\bar{x}$ and standard deviation s are jointly sufficient for (μ, σ) and x is equal to $(\bar{x}, s)$ and lies in the sample space X which is also upper half plane. The spaces X & Y are identical.

Let $[a,b]$ be a linear transformation which moves the origin by a and changes the scale by positive unit b. When this transformation is applied to a basic sample space, it takes the form as

$$[a, b]\,(x_1, x_2, \ldots, x_n) = (a + bx_1, \ldots, a + bx_n).$$

The corresponding induced transformation on sufficient statistic space will be

$$[a, b]\,(\bar{x}, s) = (a + b\bar{x}, bs).$$

The class G of such transformations is equal to $\{[a, b] : -\infty < a < \infty, 0 < b < \infty\}$, and will be a group because it is closed under the products and inverses with the following formulae.

$$[c, d]\,[a, b] = [c + da, db],$$

$$[a, b]^{-1} = \left[\frac{-a}{b}, \frac{1}{b}\right] \qquad (4.1)$$

Under a linear transformation $[a, b]$, a sample from a normal distribution (μ, σ) is carried into a sample from normal distribution $(a + b\mu, b\sigma)$. A class of distributions having satisfied these properties has the advantage that the sample and parameter points have a position relative to the other sample and parameter points.

4.11 Pivotal Quantity:

A pivotal quantity is a function of sufficient statistic and parameter and has a fixed distribution which is independent of the parameter value. Usually the fiducial distributions are derived by means of a pivotal quantity. A pivotal quantity may not be unique and different pivotal quantities may yield different fiducial distributions for the parameter.

Let x_0, θ_0 be arbitrary but fixed reference points in X and Y. Let g_x and h_θ be the unique elements of G which transform x_0 & θ_0 into general sample and parameter points x & θ respectively.

A transformation h_θ on Y carries θ_0 into θ. Therefore, as a transformation on X, h_θ must carry a variable with a θ_0 distribution into a variable with a θ distribution. The inverse h_θ^{-1} transforms a variable with distribution θ into a variable with a distribution θ_0. Let $x = g_x x_0$ be a variable with a distribution θ. Applying the transformation h_θ^{-1} produces a variable $h_\theta^{-1} g_x x_0$

with θ_0 distribution. This variable has a fixed distribution and is independent of the value of parameter. It is generated by the random variable $h_\theta^{-1}\, g_x$ treated as a random transformation and applied to the fixed reference point x_o. Thus $g = h_\theta^{-1}\, g_x$ takes values in G and has a fixed distribution when x is treated as a variable with a θ distribution; it is infact a pivotal quantity. As a function of x and θ it is invariant under transformation in G.

In order to prove the uniqueness of g (in the sense that any other pivotal quantity which is invariant under group transformation will be the function of g), let us assume $P(x, \theta)$ be invariant w.r. to G then,

$$P(fx;\, f\theta) = P(x, \theta) \quad \text{for} \quad f \in G.$$

$$\begin{aligned} P(x;\, \theta) &= P(g_x\, x_o;\, h_\theta\, \theta_o) \\ &= P(h_\theta^{-1}\, g_x\, x_o\, ;\, h_\theta^{-1}\, h_\theta\, \theta_o) \\ &= P(h_\theta^{-1}\, g_x\, x_o\, ;\, \theta_o) \Rightarrow \end{aligned}$$

$P(x, \theta)$ is expressed as a function of $h_\theta^{-1}\, g_x$. Hence the pivotal quantity $h_\theta^{-1}\, g_x$ is unique.

Considering the example from normal distribution and using (o, 1) as the reference point in both the sample and parameter spaces, produces

$$g_x = [\bar{x}, s]\,, \quad h_\theta = [\mu, \sigma]\,.$$

The pivotal quantity then has the form

$$g = h_\theta^{-1}\, g_x = [\mu, \sigma]^{-1}\, [\bar{x}, s]$$
$$= \left[-\frac{\mu}{\sigma}, \frac{1}{\sigma}\right] [\bar{x}, s]$$
$$= \left[\frac{(\bar{x}-\mu)}{\sigma}, \frac{s}{\sigma}\right],$$

which is the unique invariant pivotal quantity. Its distribution can be expressed by the pivotal quantity,

$$g = \left[\frac{Z}{\sqrt{n}}, \frac{\chi^2}{n-1}\right], \qquad (4.2)$$

where $Z \in N(0,1)$ and χ^2 is an independent χ^2 distribution with n-1 d.f.

It is obvious now that the frequency distribution for x produced a fixed frequency distribution for the pivotal variable g and when this fixed distribution of pivotal vatiable is used along with the pivotal equation $g = h_\theta^{-1}\, g$, completely describes the problem. The equation

$$g = h_\theta^{-1}\, g_x,$$

can be written as

$$g_x = h_\theta\, g,$$

$$x = g_x\, x_o = h_\theta\, g\, x_o,$$

which indicates that the frequency distribution for x is obtained by transformation h_θ which is applied to $g\, x_o$, where $g\, x_o$ is a variable in the sample space X obtained by applying transformation g to the reference point x_o.

4.12 Fiducial Distribution:

The method of obtaining a fiducial distribution is that the observed value of the sufficient statistic is substituted into the pivotal equation, the pivotal variable has its own frequency distribution, the parameter in the pivotal equation is treated as a free variable and the distribution of the pivotal variable is transferred to it by the pivotal equation.

Consider the pivotal equation

$$h_\theta = g_x \, g^{-1} ,$$

$$h_\theta \, \theta_o = g_x \, g^{-1} \, \theta_o$$

$$\hat{\theta} = g_x \, g^{-1} \, \theta_o \, . \qquad (4.3)$$

where g is the pivotal variable with a fixed pivotal distribution, x be the observed value of the sufficient statistic and $\hat{\theta}$ be a variable representing possible values for the parameter in the form of frequency information. The equation $\hat{\theta} = g_x \, g^{-1} \, \theta_o$, gives the fiducial distribution for $\hat{\theta}$ as obtained from the fixed frequency distribution of the pivotal variable.

In practical situations, θ has a particular but unknown value. This value of θ determines the distribution of sufficient statistic which has observed value x. In this case one should not infer that α probability statement cannot be made regarding θ.

In terms of repeated sampling from the fixed distribution of pivotal variable g, there is generated a frequency distribution $\hat{\theta}$ of possible parameter values corresponding to the observed x which is the fiducial distribution and has the frequency interpretation. This interpretation can be elaborated by the following example.

The pivotal variable as obtained in § 4.11 is

$$g = \left[\frac{z}{\sqrt{n}}, \frac{\chi}{\sqrt{n-1}}\right].$$

By applying a transformation to the reference point $(\theta,1)$ a frequency distribution is generated and the frequency function of the observable variable is then obtained by a transformation on the sample space where the transformation is determined by the parameter. By formula (4.3) and applying (4.1) the fiducial distribution is produced as below,

$$g^{-1}(\theta, 1) = \left[\frac{z}{\sqrt{n}}, \frac{\chi}{\sqrt{n-1}}\right]^{-1} (\theta, 1)$$

$$= \left(-\frac{\frac{z}{\sqrt{n}}}{\frac{\chi}{\sqrt{n-1}}}, \frac{\sqrt{n-1}}{\chi}\right)$$

$$(\hat{\mu}, \hat{\sigma}) = g_x\, g^{-1}(0,1) = [\bar{x}\,,\, s]\; g^{-1}(0,1)$$

$$= [\bar{x}\,,\, s]\left(- \frac{\frac{Z}{\sqrt{n}}}{\frac{\chi}{\sqrt{n-1}}}\,,\, \frac{\sqrt{n-1}}{\chi} \right)$$

$$= \left(\bar{x} - \frac{Z}{\frac{\chi}{\sqrt{n-1}}} \cdot \frac{s}{\sqrt{n}}\,,\, \frac{\sqrt{n-1}}{\chi} \cdot s\right) .$$

Hence $\hat{\sigma} = \frac{\sqrt{n-1}}{\chi}\, s$,

$$\hat{\mu} = \bar{x} - \frac{Z}{\frac{\chi}{\sqrt{n-1}}}\, \frac{s}{\sqrt{n}} = \bar{x} - Z\,\frac{\hat{\sigma}}{\sqrt{n}}$$

$$= \bar{x} - t\,\frac{s}{\sqrt{n}}\;, \qquad (4.4)$$

where the t variable implicitly defined is statistically dependent on the χ^2 variable. The fiducial variables $\hat{\mu}$, $\hat{\sigma}$ are obtained from the observed values of $\bar{x}$, s. More-over the fiducial distribution of $\hat{\mu}$ is centered at $\bar{x}$ and is scaled by $\frac{s}{\sqrt{n}}$ and has the form of Student's, t - distribution with (n-1) d · f.

Let a particular sample from a normal distribution give the values $\bar{x}$,s. Now imagine the possible experiments involving samples of size n from normal distributions. To make these samples comparable to the sample already available, transformation is applied on each sample to relocate and rescale so that the mean and standard deviation move to the values of $\bar{x}$ & s respectively. The transformation is conceptually applied to the mean of the distribution to yield a value which is appropriate for the comparison with the values $\bar{x}$, s. The class of these transformed means generates a frequency distribution which is the fiducial t - distribution. From this point of view the fiducial distribution is a frequency distribution of possible values for the parameter relevant to the specific observed $\bar{x}$, s. It is then in this form that the distribution is used to make probability statements in which μ, σ appear as variables.

4.13 Two Means Problem:

Let $\bar{x}_1$, $\bar{x}_2$, s_1^2, s_2^2 be the means and variances of the two samples having sizes n_1 and n_2, drawn from two independent normal populations with the unknown parameters μ_1, μ_2, σ_1^2 and σ_2^2. The problem is to make the test of significance or estimate the parameter difference $\mu_1 - \mu_2$. For the 1st system, $(\bar{x}_1, s_1)$ is the sufficient statistic for (μ_1, σ_1) and for second system, $(\bar{x}_2, s_2)$ is for (μ_2, σ_2). The relation (4.4) shows that the information concerning μ_1 and μ_2 is [illegible]in the variables described by

$$\bar{x}_1 - t_1 \frac{s_1}{\sqrt{n_1}} \quad \text{and} \quad \bar{x}_2 - t_2 \frac{s_2}{\sqrt{n_2}} ,$$

where t_1 and t_2 are Student's, t - distributions with $(n_1-1) = V_1$ and $V_2 = (n_2 - 1)$ d.f. These distributions together provide a distribution for $(\mu_1 - \mu_2)$ and are appropriate to the values $\bar{x}_1$, s_1, $\bar{x}_2$ & s_2. The frequency distribution for $\mu_1 - \mu_2$ is given by

$$(\bar{x}_1 - \bar{x}_2) - (t_1 \frac{s_1}{\sqrt{n_1}} - t_2 \frac{s_2}{\sqrt{n_2}}) ,$$

which may also be written as

$$(\bar{x}_1 - \bar{x}_2) - r(\text{Sin } \theta \, . \, t_1 - \text{Cos } \theta \, . \, t_2). \qquad (4.5)$$

The constants r and θ are evaluated from observed values s_1, s_2 by the relations

$$r = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} ,$$

$$\text{Sin } \theta = \frac{s_1}{\sqrt{n_1}} / r \quad ; \quad \text{Cos } \theta = \frac{s_2}{\sqrt{n_2}} / r .$$

The distribution of $\mu_1 - \mu_2$ is obtained from the distribution of the linear combination

$$e = t_1 \text{ Sin } \theta - t_2 \text{ Cos } \theta , \qquad (4.6)$$

of two independent Student's, t variables. It is only for convenience that, r and θ are introduced in place of s_1 and s_2. Percentage points for (4.6) have been tabulated by Sukhatme (1938). For instance 99% fiducial interval is given by

$$\bar{x}_1 - \bar{x}_2 \pm r\, e_{1\%} ,$$

where the interval $\pm e_{1\%}$ contains 99% of the probability according to Sukhatme's table.

4.2 The Effect of Restriction on Statistic d:

Fisher (1939) obtained the unrestricted significance level of

$$d = \frac{(\bar{x}_1-\bar{x}_2)-(\mu_1-\mu_2)}{\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}} ,$$

where $\sqrt{\frac{s_1^2}{n_1}}$, $\sqrt{\frac{s_2^2}{n_2}}$ are the estimates of standard errors of two means, by computing first the probability that d would exceed a specified value on the assumption that $K = \frac{s_1^2}{s_2^2}$ and $R = \frac{\sigma_1^2}{\sigma_2^2}$ are known. The average value of the probability so obtained is then calculated over the range $0 < R < \infty$ by assigning to R/K its approprate fiducial distribution for a known K. The fiducial distribution in this case is the F - distribution with V_1 & V_2 d.f.

The probability distribution of d, when K and R are known, is given as follows.

The quantity

$$\frac{\left[(\bar{x}_1-\bar{x}_2) - (\mu_1-\mu_2)\right] \Big/ \sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}{\left[\dfrac{V_1 \dfrac{s_1^2}{\sigma_1^2} + V_2 \dfrac{s_2^2}{\sigma_2^2}}{V_1 + V_2}\right]^{\frac{1}{2}}},$$

is distributed as a Student's t - distribution with $(V_1 + V_2)$ d.f, which on substituting the values of d, R and K reduces to

$$t_{(V_1+V_2)} = \frac{d\sqrt{\left(\frac{K}{n_1} + \frac{1}{n_2}\right)(V_1 + V_2)}}{\left[\left(V_2 + V_1 \frac{K}{R}\right)\left(\frac{R}{n_1} + \frac{1}{n_2}\right)\right]^{\frac{1}{2}}} \cdot \quad (4.7)$$

From (4.7) the probability that d is greater than a specified value, when K and R being known, can be obtained from Student's t - table corresponding to (V_1+V_2) d.f. The probability obtained from (4.7) is then averaged over the fiducial distribution of R/K from 0 to ∞ .

Let $\hat{\sigma}_1^2$ and $\hat{\sigma}_2^2$ be the variances of two populations in which $(\hat{\sigma}_1^2 / \hat{\sigma}_2^2) > 1$ is assumed to be true. If $\hat{K} = \hat{s}_1^2 / \hat{s}_2^2$ be the estimated variance ratio based on V_1 and V_2 d.f, then $\hat{K}$ will be known from the data. Assume also $\sigma_1^2 = a\ \hat{\sigma}_1^2$ and $\sigma_2^2 = b\ \hat{\sigma}_2^2$, where σ_1^2 and σ_2^2 are the variances in-volved in Behrens - Fisher problem and a, b are some known constants. The restriction

$(\hat{\sigma}_1^2 / \hat{\sigma}_2^2) > 1$ can then be written as

$$\frac{R}{K} > \frac{1}{\hat{K}} \quad ,$$

where $K = s_1^2 / s_2^2 = \frac{a}{b} \hat{K}$.

Under $\frac{R}{K} > \frac{1}{\hat{K}}$, the modification of the test (4.7), as given by Cochran (1963), would be to average the probability over the values of $(R/K) > \frac{1}{\hat{K}}$.

Writing $V = R/K$ ($= F_{V_2,V_1}$) , the Pr ($d > d_e$) in the region $V > \frac{1}{\hat{K}}$ is given in the following way.

The quantity, $Pr\left(|t|_{(V_1+V_2)} > d_\alpha \, g(V) \right)$, is the two sided probability that Student's t - variable with (V_1+V_2) d.f. is greater than $d_\alpha(g(V))$, may be given by (4.7), where

$$g(V) = \left[\frac{(1/n_2 + K/n_1)(V_1 + V_2) \, V}{(V \, V_2 + V_1)(\frac{1}{n_2} + \frac{K}{n_1} V)} \right]^{1/2} .$$

The average value of this probability over the fiducial distribution in the region $V > \frac{1}{\hat{K}}$ is then

$$\frac{\displaystyle\int_{1/\hat{K}}^{\infty} \frac{v^{\frac{V_2}{2}-1}}{(V_1+V_2\ v)^{\frac{V_1+V_2}{2}}}\ \Pr\left(\left|t_{(V_1+V_2)}\right| > d_\alpha\, g(v)\right)\ d\,v}{\displaystyle\int_{1/\hat{K}}^{\infty} \frac{v^{(V_2)/2\,-\,1}}{(V_1+V_2\ v)^{\frac{V_1+V_2}{2}}}\ d\,v} \quad .(4.8)$$

From the expression (4.8), it is possible to calculate the actual probability with which d exceeds the tabulated Behrens - Fisher significance levels in the restricted region, for some chosen values of V_1 and V_2 and taking $\hat{K}$ at its certain level of significance. The direction of the disturbance to the significance levels of d can then be examined.

4.3 Fiducial Arguments and Baye's Solution:

The problem of testing the difference in means of two normal populations, as discussed by Behrens (1964), can also be considered from the concept of reference sets on one hand and the concept of random variable on the otherhand. The former requires the definition of a chance event with reference to a particular experiment ie. the experiment and the reference set to which the probability relates are required to be stated. The problem is to find the probability P, such that

$$P = \Pr\left[(\bar{x}_1-\bar{x}_2) - (\mu_1-\mu_2) \leq t_{12} \left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2} \right)^{\frac{1}{2}} \right], \quad (4.9)$$

for a given t_{12} .

In its general solution n_1 , n_2 and t_{12} are fixed. The values μ_1 , μ_2, σ_1, σ_2, $\bar{x}_1$, $\bar{x}_2$, s_1 and s_2 are associated with a reference set, which for fixed n_1 and n_2, constitute all possible normal distributions, which changes from experiment to experiment. Looking at the problem from the latter aspect, we are interested in the probability function $f(t_{12})$ of a random variable t_{12} . In fiducial solution, the prediction about the values of parameters is not made on the basis of the previous experiments as the experiments are set up under fresh conditions. The reference set for which the probability statement $\Pr\left((\bar{x} - \mu) \leq \frac{s}{\sqrt{n}} t \right)$ holds, is that of the values μ, $\bar{x}$ and s corresponding to the same sample, for all samples of a given size of all normal populations. The choice of jointly sufficient statistics $\bar{x}$ and s, and absence of a priori knowledge about μ & σ, excludes the possibility of any subset within a general set for which a different value of the probability should hold. In the case of Bayes' solution, experiments are not conducted under different conditions and the experience gained in the previous experiments can be utilised for the new one. In other words we deal with a priori known reference set and distribution of σ.

Let

$$f(t_1)dt_1 = Pr\left(\frac{t_1 s_1}{\sqrt{n_1}} \leqslant \bar{x}_1 - \mu_1 < (t_1 + dt_1)\frac{s_1}{\sqrt{n_1}}\right), \quad (4.10)$$

and

$$f(t_2)dt_2 = Pr\left(\frac{t_2 s_2}{\sqrt{n_2}} \leqslant \bar{x}_2 - \mu_2 < (t_2 + dt_2)\frac{s_2}{\sqrt{n_2}}\right). \quad (4.11)$$

Writing $p = \dfrac{s_2\sqrt{n_1}}{s_1\sqrt{n_2}}$ and using (4.11), the relation (4.9) reduces to

$$Pr\left((\bar{x}_1-\mu_1) \leqslant (t_{12}\sqrt{(1+p^2)} + t_2 p)\frac{s_1}{\sqrt{n_1}}\right)$$

$$= F(t_{12}\sqrt{(1+p^2)} + t_2 p), \quad (4.12)$$

where t_{12}, t_2 and p are held constant.

We have $\quad p\dfrac{\sqrt{n_2}}{s_2} = \dfrac{\sqrt{n_1}}{s_1}$ and $\dfrac{t_2 s_2}{\sqrt{n_2}} = \bar{x}_2 - \mu_2$.

Therefore the conditional probability (given p) may be written from (4.12) as

$$Pr\left((\bar{x}_1-\mu_1) - (\bar{x}_2-\mu_2) \leqslant t_{12}\sqrt{(1+p^2)}\frac{s_1}{\sqrt{n_1}} \,\middle|\, p\right), \quad (4.13)$$

which becomes

$$\int_{-\infty}^{\infty} f(t_2)F(t_{12}\sqrt{1+p^2} + t_2p)\, dt_2 = F_1(t_{12}, p). \tag{4.14}$$

The equality of (4.13) and (4.14) is valid if fiducial arguments are accepted and leads to Behrens-Fisher test.

$F_1(t_{12}, p)$ can also be regarded as the conditional distributian function of random variable t_{12} given p.

Let $u = \dfrac{\bar{x}-\mu}{\sigma/\sqrt{n}}$ and $v = \dfrac{s}{\sigma}$ be two independent random variables following a standard normal law and $\dfrac{\chi_{(n-1)}}{\sqrt{(n-1)}}$ distribution respectively. The conditional probability $f(u, v \mid \sigma)$, given σ, may be written as

$$f(u,v \mid \sigma)dudv = \frac{1}{\sqrt{2\pi}}\, e^{-u^2/2}\, d\, u\; C_n\, v^{n-2}\, e^{-(n-1)\frac{v^2}{2}} dv, \tag{4.15}$$

where

$$C_n = \frac{(n-1)^{\frac{n-1}{2}}}{2^{\left(\frac{n-3}{2}\right)} \left(\frac{n-3}{2}\right)!}.$$

Transforming $u = \dfrac{ts}{\sigma}$ and $v = \dfrac{s}{\sigma}$ we get

$$f(t,s \mid \sigma)\; d\, s\; d\, t = \frac{1}{\sqrt{2\pi}}\, C_n\, e^{-\frac{s^2}{2\sigma^2}(t^2+n-1)}\, \frac{s^{n-1}}{\sigma^n}\, d\, s\; d\, t. \tag{4.16}$$

Baye's approach assumes that σ belongs to a specific reference set with certain known prior distribution $g(\sigma)$. The joint probability $f(t,s)\,dt\,ds$ is then given by

$$f(t,s)\ d\,t\ d\,s = \frac{C_n}{\sqrt{2\pi}}\ d\,t\ d\,s\ \int_0^{\infty} e^{-\frac{s^2}{2\sigma^2}(t^2+n-1)}\,\frac{s^{n-1}}{\sigma^n}\,g(\sigma)\ d\,\sigma\,. \tag{4.17}$$

The marginal probabilities of the events

$$t \leq \frac{\bar{x}-\mu}{s/\sqrt{n}} < (t + dt) \tag{4.18}$$

and

$$s \leq \sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{(n-1)} < (s + ds), \tag{4.19}$$

may be obtained from (4.17), and are given by

$$f_1(t)\ dt = dt \int_0^{\infty} f(t,s)\ ds \tag{4.20}$$

and

$$f_2(s)\ ds = ds \int_0^{\infty} f(t,s)\ dt, \tag{4.21}$$

respectively.

The conditional probability $f(t|s)\ dt$, of the event (4.18) can be given by

$$f(t|s)\ dt = \frac{f\,(t,s)\ dt}{f_2(s)}\,. \tag{4.22}$$

Infact $f_1(t)$ is the Student's probability density function $f(t_1)$, based on a random sample of size n.

With Baye's approach the comparison of the means can be made by the use of conditional density function

$f(t|s)$ and conditional distribution function $Q(t|s) = \int_{-\infty}^{t} f(t|s)dt$, instead of the Student's density and its distribution functions. The conditional probability, given s_1 and s_2, is obtained by

$$\Pr\left[\left((\bar{x}_1-\mu_1) - (\bar{x}_2-\mu_2)\right) \leq t_{12} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \,\middle|\, s_1, s_2\right]$$

$$= \int_{-\infty}^{\infty} f(t_2|s_2)\, dt_2 \;\; H(t_{12} \sqrt{1+p^2} + t_2 p \mid s_1) \qquad (4.23)$$

where $f(t_2|s_2)dt_2 = \Pr\left(t_2 \leq \frac{\bar{x}_2-\mu_2}{s_2/\sqrt{n_2}} < (t_2+dt_2) \mid s_2 \right)$,

and

$$H(t_{12} \sqrt{1+p^2} + t_2 p \mid s_1) = \Pr\left[(\bar{x}_1-\mu_1) \leq (t_{12} (1+p^2)^{1/2} + t_2 p) \frac{s_1}{\sqrt{n_1}} \,\middle|\, p, s_1\right]$$

4.4 Approximation :

The test criterion, as suggested by Fisher (1935), is given by

$$d = \frac{(\bar{x}_1-\bar{x}_2) - (\mu_1-\mu_2)}{\left[\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right]^{1/2}} .$$

He identified his solution of the two means problem with the solution given by Behrens in (1929). The computation of all percentage points is difficult by the direct application of the formula. Some approximations to the test have been suggested by Fisher (1941) and Rubin (1960) for the use of practical workers. Cochran (1964) produced an empirical approximation based on Student t-table values in the following way.

Let α_1, α_2 be the critical points of Student t-distribution with $V_1=(n_1-1)$ and $V_2=(n_2-1)$ d.f at some preassigned significance level α. The approximate critical point for d is then given by the weighted mean of α_1 and α_2 with the weights s_1^2/n_1 and s_2^2/n_2 respectively, i.e.

$$\hat{d}_{\alpha} = \frac{(s_1^2/n_1)\,\alpha_1+(s_2^2/n_2)\,\alpha_2}{s_1^2/n_1 + s_2^2/n_2} \qquad (4.24)$$

$\hat{d}_{\alpha}$ reduces to a Student t-value with V d.f when $V_1=V_2=V$, say. If $V_1\neq V_2$, it is often apparent, by observing α_1 and α_2, that $\hat{d}_{\alpha}$ will exceed both of them or not. Its advantage lies in simplicity and fair accuracy. Cochran (1964) has measured its accuracy by calculating the actual probability α' that Behrens-Fisher d exceeds the approximate $\hat{d}_{\alpha}$ with some preassigned α, by making use of asymptotic formula due to Fisher (1941), and concluded that the approximation (4.24) is adequate for routine tests between 1% and 10% levels of significance, but not for accurate calculations.

4.5 References:

Behrens, W.V. (1929), Ein Beitrag Zur Fehlerberechnung bei Wenigen Beobuchtungen, Land wirtschaftliche Jahrbrücher, 68, 807.

Behrens, W.V. (1964), The comparison of means of in-depent normal distributions with different variences, Biometrika, Vol 20, 16-27.

Cochran, W.G. (1963), The Behrens - Fisher test when range of the unknown variance ratio is restricted, Sankhȳa, series A, Vol 25, 353-362.

Cochran, W.G. (1964), Approximate Significance levels of Behrens - Fisher test, Biometrika, Vol 20, 191-195.

Fisher, R.A. (1930), Inverse probability, Proc. Camb. Philos. Soc., Vol 26, 528-535.

Fisher, R.A. (1939), The comparison of samples with possibly unequal variances, Ann of Eugenics, Vol 9, 174-180.

Fisher, R.A. (1941), The asymptotic approach to Behrens integral, Ann of Eugenics, Vol 11, 141-172.

Fraser, D.A.S. (1961a), On fiducial inference, Ann. Math. Statist., Vol 32, 661-76.

Fraser, D.A.S. (1961b), The fiducial method and invarience, Biometrika, Vol 48, 261-80.

Fraser, D.A.S. (1963), On definition of fiducial probability, Bulletin, Proc., 34th Session, Ottaw, 842-55.

Patil, G.P. (1965), Approximation to the Behrens - Fisher distributions, Biometrika, Vol 52, 267.

Rubin, H. (1960), On the distribution of the weighted difference of two independent Student' variables, J.R. Statist. Soc., Series B, Vol 22, 188-194.

Sukhatme, P.V. (1938), On Fisher and Behrens test of significance, Sankhya, Vol 4, 39-48.

Appendix:

For a $n_1 \times n_2$ matrix $C = (c_{ij})$ satisfying the necessary and sufficient conditions that all d_i have means δ and variances σ^2 are

$$S_j\, c_{ij} = 1$$

$$S\, c_{ij}^2 = c^2$$

$$S_j\, c_{ij} c_{kj} = 0 \quad \text{when} \quad i \neq k\,,$$

has minimum value of $c^2 = n_1/n_2$.

<u>Proof</u>: Writing these conditions in vector form we get

$$A_i U' = 1$$

$$A_i A_k' = \begin{cases} c^2 & \text{when} \quad i = k \\ 0 & \text{when} \quad i \neq k, \end{cases} \qquad (3.58)$$

where A_i is the i^{th} row vector of matrix (c_{ij}) and U is the $1 \times n_2$ matrix $(1,1,\ldots,1)$. Prime denotes the transpose of a matrix.

If n_1 vectors A_i satisfy (3.58), we can adjoin $(n_2 - n_1)$ vectors, satisfying the second condition of (3.58), so that the resultant set forms a basis for an n_2 - space. The matrix U can be expressed as a linear combination of n_2 A - vectors,

$$U = S_{k=1}^{n_2}\, g_k\, A_k, \qquad (3.59)$$

where g_k are scalars.

Using (3.58) and (3.59), we obtain

$$1 = A_i U' = A_i \mathop{S}_{k=1}^{n_2} g_k A_k' = S\, g_k A_i A_k' = g_i c^2. \tag{3.60}$$

Hence $g_i = \dfrac{1}{c^2}$, $i = 1, 2, \ldots n_1$.

U is a unit row vector therefore,

$$n_2 = UU' = (\mathop{S}_{k=1}^{n_2} g_k A_k)(\mathop{S}_{k=1}^{n_2} g_k A_k'),$$

which by applying (3.58) becomes

$$n_2 = \mathop{S}_{k=1}^{n_2} g_k^2 A_k A_k'$$

$$= c^2(\mathop{S}_{k=1}^{n_1} + \mathop{S}_{k=n_1+1}^{n_2}) g_k^2 . \tag{3.61}$$

By making use of (3.60), we obtain from (3.61)

$$n_2 = c^2(n_1/c^4 + \mathop{S}_{k=n_1+1}^{n_2} g_k^2)$$

$$n_2 \geqslant \frac{n_1}{c_1} \Rightarrow c_1 \geqslant \frac{n_1}{n_2} .$$

The equality sign holds whenever $g_k = 0$ for $k = n_1+1, \ldots, n_2$.