UNIVERSITY OF CAPE TOWN

DEPARTMENT OF MATHEMATICAL STATISTICS

* * * * * * *

APPROXIMATIONS TO NONCENTRAL DISTRIBUTIONS AND THEIR APPLICATIONS

by

P.C. Carter-Johnson

* * * * * * *

A thesis prepared under the supervision of Dr. J.M. Juritz in fulfilment of the requirements for the degree of Master of Science in Mathematical Statistics

* * * * * * *

To my parents

and brother

ACKNOWLEDGEMENTS

I wish to express my deepest appreciation to Dr. J.M. Juritz for the patient way in which she guided me in the preparation of this thesis and for her encouragement. I also wish to thank Mrs. E.B. Prosser for the tremendously neat and skilful typing of the manuscript. My thanks also go to the C.S.I.R. who awarded me a Post-Graduate bursary.

CONTENTS

Introduction

In testing hypotheses involving noncentral distributions percentage points are not always readily available and, if they are available,are not very well tabulated except, perhaps, for smaller degrees of freedom and noncentralities. Consequently, for values that are not tabulated, interpolation, or, more likely, extrapolation of some kind is necessary and the process can become tedious.

In the case of calculations involving the power of the test, charts of the power of the F-test and t-test are available, but readings taken from these charts may be accurate to only one decimal place.

In situations like the above, and in other cases, approximations are very useful and are sometimes as accurate, if not more so, than values obtained by interpolation (or extrapolation) or values read from charts.

This thesis is chiefly concerned with applications in which approximations to the noncentral χ^2, F, t and R distributions can be used. The approximations themselves, in most cases, are dealt with in a fair amount of detail to show the reader how they were obtained.

Chapter 1 defines certain terms with which the reader may be unfamiliar, which are used in subsequent chapters.

Chapters 2-5 deal with the approximations and their applications. Each of these chapters is set out in the same way, section I defining the noncentral distribution, section II dealing with the approximations, section III comparing the accuracy of the approximations with the exact values and section IV showing in which situations the approximations can be used.

Chapter 1 : Definitions

Introduction

A number of terms are to be used in the following chapters which will need some explanation. This chapter is just a collection of definitions to which one may refer.

§1. Mathematical Functions

(i) The Incomplete Gamma Function

This is given by

$$\Gamma_T(\alpha) = \int_o^T t^{\alpha-1} e^{-t} dt. \qquad (1.1)$$

However, the ratio $\Gamma_T(\alpha)/\Gamma(\alpha)$ is used more than (1.1), where

$$\Gamma(\alpha) = \int_o^\infty t^{\alpha-1} e^{-t} dt \qquad (\alpha > o) .$$

This ratio is called the *Incomplete gamma function ratio* (the word "ratio" is usually omitted) and has been tabulated by Pearson (1922) as

$$I(u,p) = \Gamma_{u\sqrt{p+1}}(p+1)/\Gamma(p+1). \qquad (1.2)$$

(ii) The Incomplete Beta Function

This is given by

$$B_T(\alpha,\beta) = \int_o^T t^{\alpha-1} (1-t)^{\beta-1} dt \qquad (o < T < 1). \qquad (1.3)$$

Again, the *Incomplete beta function ratio* ("ratio" usually omitted) is more often used. This ratio is

$$I_T(\alpha,\beta) = B_T(\alpha,\beta)/B(\alpha,\beta), \quad (1.4)$$

where

$$B(\alpha,\beta) = \int_o^1 t^{\alpha-1}(1-t)^{\beta-1}\,dt \qquad (\alpha,\beta > o).$$

Tables of $I_T(\alpha,\beta)$ can be found in Pearson and Hartley (1962).

§2. Matrices

(i) A square matrix A is called *singular* if $|A| = o$ and *nonsingular* if $|A| \neq o$.

(ii) The *rank* of A ≡ rank (A) is the maximum number of linearly independent vectors in the set $\{\alpha_1,\ldots, \alpha_m\}$, where $\alpha_i\,(i = 1,\ldots,m)$ are the columns of $A : n\times m$.

(iii) The *trace* of a square matrix A ≡ tr(A) is the sum of the diagonal elements of A.

(iv) The symmetric matrix A and the quadratic form X'AX are called *positive definite* if all the characteristic roots of A are $>o$ and *positive semidefinite* if all the characteristic roots of A are $\geq o$, where the characteristic roots are the roots of the determinant $|A-\lambda I|$.

§3. Miscellaneous

(i) Cumulants

The logarithm of the moment generating function of X is the *cumulant generating function* of X, and is defined by

$$\psi_X(t) \quad = \quad \log_e E(e^{tX}). \tag{3.1}$$

The *rth cumulant* of X is the coefficient of $\frac{t^r}{r!}$ in the Taylor series expansion of $\psi_X(t)$, and is denoted by κ_r .

(ii) Difference Operators

Δ is the *forward difference operator* and is defined by

$$\Delta f(x) \quad = \quad f(x+1) \quad - \quad f(x). \tag{3.2}$$

If n is an integer then

$$\Delta^n f(x) \quad = \quad \sum_{j=o}^{n} \binom{n}{j} (-1)^j f(x+n-j). \tag{3.3}$$

$\Delta^n f(x)$ is known as the *nth forward difference* of f(x).

(iii) Variance-stabilizing Transformations

Theorem (see Laubscher (1960))

If X is asymptotically normally distributed about μ_1, with asymptotic variance $\sigma^2(\mu_1)$, then any function $\psi = \psi(X)$, with continuous first derivative in some neighbourhood of μ_1, is asymptotically normally distributed with mean $\psi(\mu_1)$ and variance $\sigma^2(\mu_1) \left(\frac{d\psi}{d\mu_1}\right)^2$, where $\left(\frac{d\psi}{d\mu_1}\right)$ denotes the derivative of $\psi(X)$ with respect to X, evaluated at the point μ_1.

Corollary (Laubscher (1960))

The random variable

$$\psi(X) = c \int_{K}^{X} \frac{d\mu_1}{\sigma(\mu_1)} , \qquad (3.4)$$

where $o < x < \infty$ *, and where* K *is an arbitrary constant, has a variance which is stabilized asymptotically at* c^2.

Transformations such as (3.4) not only often work well for stabilizing non-asymptotic variances, but also often serve as well to normalize non-normal distributions.

(iv) Mixture Distributions

If $\{F_j(x_1, x_2, \cdots, x_n)\}$ $(j = \cdots, -1, 0, 1, 2, \cdots)$ represents a set of cumulative distribution functions and if $a_j \geq o$ and $\sum_{j=-\infty}^{\infty} a_j = 1$, then

$$F(x_1, \cdots, x_n) = \sum_{j=-\infty}^{\infty} a_j F_j(x_1, \cdots, x_n) \qquad (3.5)$$

is also a cumulative distribution function. This is called a *mixture* of the distributions $\{F_j\}$.

§4. Systems of Frequency Curves

These systems are families of distributions which have been constructed to provide approximations to known distributions.

(i) The Pearson System

For every curve in the system, the probability density function p(x) satisfies a differential equation of the form

$$\frac{1}{p}\frac{dp}{dx} = -\frac{a+x}{c_o + c_1x + c_2x^2} . \qquad (4.1)$$

The shape of the distribution depends on the values of the parameters a, c_o, c_1 and c_2 .

Pearson classified the different shapes into a number of types. Only two of these types concern us.

Type I. This type occurs when both roots of the equation $c_o + c_1x + c_2x^2 = o$ are real and of opposite sign. We shall denote the roots by a_1 , a_2 with

$$a_1 < o < a_2 .$$

We have, therefore,

$$c_o + c_1x + c_2x^2 = -c_2(x-a_1)(a_2-x),$$

and equation (4.1) can be written as

$$\frac{d \log p(x)}{dx} = \frac{x+a}{c_2(x-a_1)(a_2-x)} = \frac{1}{c_2(a_2-a_1)}\left\{\frac{a+a_1}{x-a_1} + \frac{a+a_2}{a_2-x}\right\} .$$

Therefore,

$$p(x) = K(x-a_1)^{m_1}(a_2-x)^{m_2} , \qquad (4.2)$$

where

$$m_1 = \frac{a+a_1}{c_2(a_2-a_1)} \quad \text{and} \quad m_2 = -\frac{a+a_2}{c_2(a_2-a_1)} .$$

For both $(x-a_1)$ and (a_2-x) to be positive we must have $a_1 < x < a_2$. Equation (4.2) can represent a proper probability density function provided that $m_1 > -1$ and $m_2 > -1$.

This type I distribution is a general form of a *beta* distribution.

Type IV. These occur when the roots of the equation

$$c_o + c_1 x + c_2 x^2 = o$$

are not real.

Let

$$c_o + c_1 x + c_2 x^2 = C_o + c_2 (x + C_1)^2$$

with

$$C_o = c_o - \tfrac{1}{4} c_1^2 c_2^{-1} \text{ and } C_1 = \tfrac{1}{2} c_1 c_2^{-1} .$$

Therefore, (4.1) becomes

$$\frac{d \log p(x)}{dx} = \frac{-(x+C_1) - (a-C_1)}{C_o + c_2 (x+C_1)^2} .$$

It follows that

$$p(x) = K[C_o + c_2 (x + C_1)^2]^{-(2c_2)^{-1}} \tag{4.3}$$

$$\times \exp\left(- \frac{a - C_1}{\sqrt{c_2 C_o}} \tan^{-1} \frac{x + C_1}{\sqrt{C_o / c_2}}\right) .$$

(ii) Expansions

Gram-Charlier Expansions

For many continuous distributions one can change the values of the cumulants by applying an operator to the probability density function, i.e., it is possible to obtain useful approximations to a distribution with known cumulants in terms of a known distribution f(x).

If f(x) is a probability density function with cumulants κ_1 , κ_2, $\cdots$, then the function

$$g(x) \quad = \quad \exp\left[\sum_{j=1}^{\infty} e_j\{(-D)^j/j!\}\right] f(x) \tag{4.4}$$

will have cumulants $\kappa_1 + e_1$, $\kappa_2 + e_2$, ... The exponential in (4.4) must be expanded as

$$\sum_{i=o}^{\infty} \left[\sum_{j=1}^{\infty} e_j\{(-D)^j/j!\}\right]^i \Big/ i\,!$$

and then applied to f(x).

[Note: D is the differentiation operator, and

$$D^j f(x) \quad = \quad d^j f(x)/dx^j .].$$

Now g(x) may not satisfy the condition that $g(x) \geq o$ for all x, i.e., g(x) is not necessarily a probability density function.

The Gram-Charlier series arises when the initial family of distributions f(x) is chosen to be the normal distribution.

From (4.4) we find

$$\begin{aligned} g(x) \quad = \quad & f(x) - e_1 Df(x) + \tfrac{1}{2}(e_1^2 + e_2)D^2 f(x) \\ & - \frac{1}{6}(e_1^3 + 3e_1e_2 + e_3)D^3 f(x) \\ & + \frac{1}{24}(e_1^4 + 6e_1^2e_2 + 4e_1e_3 + e_4)D^4 f(x) + \cdots\cdots \end{aligned} \tag{4.5}$$

and, therefore,

$$\int_{-\infty}^{x} g(t)dt \quad = \quad \int_{-\infty}^{x} f(t)dt - e_1 f(x) + \tfrac{1}{2}(e_1^2 + e_2)Df(x) - \cdots\cdots . \tag{4.6}$$

When f(x) is a normal probability density function

$$D^j f(x) \quad = \quad P_j(x) f(x) ,$$

where $P_j(x)$ is a polynomial of degree j in x, so that (4.5) can be written in the form

$$g(x) = [1 - e_1 P_1(x) + \tfrac{1}{2}(e_1^2 + e_2)P_2(x) - \frac{1}{6}(e_1^3 + 3e_1 e_2 + e_3)P_3(x)$$

$$+ \frac{1}{24}(e_1^4 + 6e_1^2 e_2 + 4e_1 e_2 + e_4)P_4(x) - \cdots]f(x). \tag{4.7}$$

If the expected values and variances of f(x) and g(x) have been made to agree then $e_1 = e_2 = o$ and (4.7) becomes

$$g(x) = [1 - \frac{1}{6} e_3 P_3(x) + \frac{1}{24} e_4 P_4(x) - \cdots]f(x). \tag{4.8}$$

A simple way of making sure that the expected values and variances agree is to use standardized variables (variables with mean zero and variance 1), and to choose f(x) so that the corresponding distribution is standardized.

Let us take $f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$ (standardized normal), then $(-1)^j P_j(x)$ is the *Hermite polynomial* $H_j(x)$.

[Note: The rth Hermite polynomial is defined as

$$H_r(x) = (-1)^r e^{\frac{1}{2}x^2} D^r e^{-\frac{1}{2}x^2} \quad (r = 0,1,2, \ldots).]$$

Since the cumulants $\kappa_r = o$ $(r > 2)$ for the normal distribution, then e_3 , e_4 , ... are equal to the corresponding cumulants of the distribution we wish to approximate, and since this function is standardized we have

$$e_3 = \sqrt{\beta_1} \text{ and } e_4 = \beta_2 - 3,$$

where $\beta_1 = \frac{\mu_3^2}{\mu_2^3}$, $\beta_2 = \frac{\mu_4}{\mu_2^2}$

and μ_i (i = 1,2,3,4) are the first four moments.

Therefore, we have

$$g(x) = [1 + \frac{1}{6}\sqrt{\beta_1}\, H_3(x) + \frac{1}{24}(\beta_2-3)H_4(x) + \cdots]\, (\sqrt{2\pi})^{-1}\, e^{-\frac{1}{2}x^2} \tag{4.9}$$

and

$$\int_{-\infty}^{x} g(t)\,dt = \Phi(x) - [-\frac{1}{6}\sqrt{\beta_1}\, H_2(x) + \frac{1}{24}(\beta_2-3)H_3(x) + \cdots]Z(x)$$

$$= \Phi(x) - \frac{1}{6}\sqrt{\beta_1}\,(x^2-1)Z(x) - \frac{1}{24}(\beta_2-3)(x^3-3x)Z(x) + \cdots, \tag{4.10}$$

where

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{1}{2}t^2}\,dt \text{ and } Z(x) = \frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}x^2}.$$

Equations (4.9) and (4.10) are known as Gram-Charlier expansions.

<u>Edgeworth Expansions</u>

Sometimes a different ordering in (4.9) and (4.10) is used which is based on the fact that for a sum of n independent, identically distributed standardized random variables, the rth cumulant is proportional to $n^{1-r/2}$ $(r \geq 2)$ $\left(\text{i.e., } e_r \propto n^{1-r/2}\right)$. Collecting terms of equal order in $n^{-\frac{1}{2}}$, and arranging in ascending order, gives an *Edgeworth expansion*.

Therefore,

$$g(x) = [1 + \frac{1}{6}\sqrt{\beta_1}\, H_3(x) + \frac{1}{24}(\beta_2-3)H_4(x) + \frac{1}{72}\beta_1\, H_6(x) + \cdots]Z(x) \tag{4.11}$$

and

$$\int_{-\infty}^{x} g(t)dt = \Phi(x) - \frac{1}{6}\sqrt{\beta_1}(x^2-1)Z(x) - \frac{1}{24}(\beta_2-3)(x^3-3x)Z(x)$$

$$- \frac{1}{72}\beta_1(x^5-10x^3+15x)Z(x) + \cdots . \qquad (4.12)$$

In most applications only the first four moments are used. Therefore, the expressions for the Gram-Charlier and Edgeworth series are

$$g(x) = [1 + \frac{1}{6}\sqrt{\beta_1}(x^3-3x) + \frac{1}{24}(\beta_2-3)(x^4-6x^2+3)]Z(x) \qquad (4.13)$$

(Gram-Charlier)

and

$$g(x) = [1 + \frac{1}{6}\sqrt{\beta_1}(x^3-3x) + \frac{1}{24}(\beta_2-3)(x^4-6x^2+3)$$

$$+ \frac{1}{72}\beta_1(x^6-15x^4+45x^2-45)]Z(x). \qquad (4.14)$$

(Edgeworth)

Laguerre Expansions

If f(x) in (4.7) is a standard gamma probability density function, then expansions in terms of *Laguerre polynomials* are obtained. The rth generalized Laguerre polynomial of order a, $L_r^{(a)}(x)$ is

$$L_r^{(a)}(x) = \sum_{j=o}^{r} (-1)^j \binom{r+a}{r-j} \frac{x^j}{j!} . \qquad (4.15)$$

Cornish-Fisher Expansions

If any distribution is fitted by equating the first r moments of the actual and fitted distributions, one can calculate quantiles of the fitted distribution. These quantities of the fitted distribution will then be approximations to the quantiles of the actual distribution.

In the case of Gram-Charlier and Edgeworth expansions one can obtain expansions for standardized quantiles as functions of corresponding quantiles of the unit normal distributions.

From (4.1) we have

$$\int_{-\infty}^{x} g(t)\,dt = D^{-1} \exp\left\{\sum_{j=1}^{\infty} e_j (-D)^j/j!\right\} f(x) \quad . \qquad (4.16)$$

Let $f(x) = (\sqrt{2\pi})^{-1} e^{-\frac{1}{2}x^2} = Z(x)$, then $D^j f(x) = (-1)^j H_j(x) Z(x)$.

Define X_α and U_α by

$$\int_{-\infty}^{X_\alpha} g(x)\,dx = \alpha = \int_{-\infty}^{U_\alpha} Z(x)\,dx \quad .$$

Using (4.16) we can obtain the identity

$$\left[\sum_{i=0}^{\infty} D^{-1} \left\{\sum_{j=1}^{\infty} e_j (-D)^j/j!\right\}^i/i!\right] Z(X_\alpha) \qquad (4.17)$$

$$= \left[\sum_{j=1}^{\infty} \left\{(X_\alpha - U_\alpha)^j/j!\right\} H_j(X_\alpha)\right] Z(X_\alpha) \quad .$$

Expanding the left-hand side and dividing both sides by $Z(X_\alpha)$ gives an identity, of polynomial form, between $(X_\alpha - U_\alpha)$ and X_α.

It is possible to rearrange (4.17) to give (a) U_α as a function of X_α or (b) X_α as a function of U_α .

Cornish and Fisher collected terms according to Edgeworth's system and obtained

$$
\begin{aligned}
X(U_\alpha) = U_\alpha &+ \frac{1}{6}(U_\alpha^2-1)\kappa_3 + \frac{1}{24}(U_\alpha^3-3U_\alpha)\kappa_4 \\
&- \frac{1}{36}(2U_\alpha^3-5U_\alpha)\kappa_3^2 + \frac{1}{120}(U_\alpha^4-6U_\alpha^2+3)\kappa_5 \\
&- \frac{1}{24}(U_\alpha^4-5U_\alpha^2+2)\kappa_3\kappa_4 + \frac{1}{324}(12U_\alpha^4-53U_\alpha^2+17)\kappa_3^3 \\
&+ \frac{1}{720}(U_\alpha^5-10U_\alpha^3+15U_\alpha)\kappa_6 \\
&- \cdots\cdots ,
\end{aligned}
\tag{4.18}
$$

and

$$
\begin{aligned}
U(X_\alpha) = X_\alpha &- \frac{1}{6}(X_\alpha^2-1)\kappa_3 - \frac{1}{24}(X_\alpha^3-3X_\alpha)\kappa_4 + \frac{1}{36}(4X_\alpha^3-7X_\alpha)\kappa_3^2 \\
&- \frac{1}{120}(X_\alpha^4-6X_\alpha^2+3)\kappa_5 + \frac{1}{144}(11X_\alpha^4-42X_\alpha^2+15)\kappa_3\kappa_4 \\
&- \frac{1}{648}(69X_\alpha^4-187X_\alpha^2+52)\kappa_3^3 - \frac{1}{720}(X_\alpha^5-10X_\alpha^3+15X_\alpha)\kappa_6 \\
&+ \cdots\cdots .
\end{aligned}
\tag{4.19}
$$

The function $X(\cdot)$ expresses the quantiles of the (standardized) distribution of X as a function of corresponding quantiles of the unit normal distribution.

(iii) Johnson's S_U distribution

Johnson (see Johnson and Kotz (1970b), pages 22,23) has described a set of three transformations to a normally distributed variable, which, when combined, provide one distribution corresponding to each pair of values $\sqrt{\beta_1}$ and β_2 .

The transformations are

$$Z = \gamma + \delta \log (X-\xi) \qquad (X \geq \xi) \qquad (4.20)$$

$$Z = \gamma + \delta \log \{(X-\xi)/(\xi+\lambda-X)\} \quad (\xi \leq X \leq \xi + \lambda) \qquad (4.21)$$

$$Z = \gamma + \delta \sinh^{-1}\{(X-\xi)/\lambda\} , \qquad (4.22)$$

where $Z \sim N(0,1)$. The symbols γ, δ , ξ , λ represent parameters. The value of λ must be positive, and, by convention, the sign of δ is also positive.

Transformation (4.20) corresponds to the family of log-normal distributions. The family of distributions corresponding to (4.21) is denoted by S_B .

In (4.22), the range of X is unbounded, and the family of distributions is denoted by S_U .

Chapter 2 : The Noncentral χ^2-distribution

I. Introduction

If U_i $(i = 1,2,\ldots,\nu)$ are independent $N(0,1)$ variables, and δ_i $(i = 1,\ldots,\nu)$ are constants, then $\sum_{i=1}^{\nu}(U_i + \delta_i)^2$ is distributed as *noncentral*-χ^2 with ν *degrees of freedom* and *noncentrality parameter* $\lambda = \sum_{i=1}^{\nu}\delta_i^2$, and is denoted by $\chi_\nu'^2(\lambda)$ or $\chi_\nu'^2$.

The probability density function is given by

$$p(x) = p(\chi'^2|\lambda) = e^{-\frac{1}{2}\lambda}\sum_{i=o}^{\infty}\frac{(\frac{1}{2}\lambda)^i}{i!}\;\frac{x^{\frac{1}{2}\nu+i-1}\quad e^{-\frac{1}{2}x}}{2^{\frac{1}{2}\nu+i}\quad\Gamma(\frac{1}{2}\nu+i)}$$

$$(o \lesssim x < \infty)$$

II. Approximations

§1. Patnaik

Patnaik (1949) suggested four approximations to the noncentral-χ^2 distribution.

(i) The first is a simple approximation in which $\chi_\nu'^2$ is replaced by a multiple of the central-χ^2, say $\rho\chi_f^2$, i.e.

$$\chi_\nu'^2 \doteqdot \rho\chi_f^2 , \tag{1.1}$$

ρ and f being chosen such that the first two moments of $\chi_\nu'^2$ are equal to the first two moments of $\rho\chi_f^2$. The values of ρ and f which satisfy this condition are

$$\rho = \frac{\nu + 2\lambda}{\nu + \lambda} \quad \text{and} \quad f = \frac{(\nu + \lambda)^2}{\nu + 2\lambda} \, .$$

Since f is, in general, a fraction, interpolation is necessary if standard χ^2 tables are used. Now let $x = \chi_\nu'^2$ and $y = {}^x/\rho$. Therefore,

$$\int_0^x p(x)dx = \int_0^y p(y)dy \doteqdot \int_0^y f(y)dy, \tag{1.1a}$$

where p(x) is the true density of $\chi_\nu'^2$ and f(x) is the approximation to p(x) obtained by assuming that ${}^x/\rho = y \stackrel{\bullet}{\sim} \chi_f^2$.

(ii) Patnaik's second approximation is that

$$\sqrt{2\chi_\nu'^2(\nu + \lambda)/(\nu + 2\lambda)} \stackrel{\bullet}{\sim} N\left(\sqrt{\frac{2(\nu + \lambda)^2}{\nu + 2\lambda} - 1},\ 1\right) \tag{1.2}$$

for large ν or λ.

Patnaik then suggests two closer approximations. The first of these is as follows:

(iii) Letting p(x) and f(x) be defined as above, the r^{th} difference between the cumulants of p(x) and f(x) is denoted by c_r $(r > 2)$ and the r^{th} difference between the cumulants of p(y) and f(y) is ${}^{c}r/_{\rho}r$ $(r > 2)$, where $y = {}^x/\rho$. The *Edgeworth operator* is then applied to f(y). Taking the probability integral $\int_0^y p(y)dy$ and leaving out all derivatives higher than the third Patnaik arrives at the approximation

$$\int_0^y p(y)dy \doteqdot \int_0^y f(y)dy - \frac{c_3}{6\rho^3}\frac{d^3}{dy^3}\int_0^y f(y)dy. \tag{1.3}$$

This can be written as

$$\int_0^y p(y)dy \doteq I(u,p) - \frac{c_3}{6\rho^3[\sqrt{2\nu}]^3} \frac{d^3 I}{du^3} ,$$

where $u = \frac{y}{\sqrt{2f}}$, $p = \frac{f}{2} - 1$, f is defined in (i) and I(u,p) is the *Incomplete* Γ*-function.*

(iv) For his second closer approximation Patnaik employs terms of the *Gram-Charlier series.* He standardizes $x = \chi_\nu'^2$ by

$$\xi = \frac{x - (\nu+\lambda)}{\sqrt{2(\nu+2\lambda)}} . \qquad (1.4)$$

The cumulants of $p(\xi)$ can be found to be 0,1, $\kappa_3/\kappa_2^{3/2}$, κ_4/κ_2^2 , ..., where $\kappa_i\,(i \geq 2)$ are the cumulants of p(x). Since f(x) has the same mean and standard deviation as p(x), the cumulants of $f(\xi)$ are 0,1, $\kappa_3^*/\kappa_2^{3/2}$, κ_4^*/κ_2^2,...., where $\kappa_i^*\,(i \geq 3)$ are the cumulants of f(x). Let $Z(\xi) = e^{-\frac{1}{2}\xi^2}/\sqrt{2\pi}$ and let $H_3(\xi)$, $H_4(\xi)$, ... be the *Hermite polynomials* of orders 3,4, ... Arranging terms in order of magnitude of ν we obtain

$$p(\xi) = [1 + \frac{1}{6}\frac{\kappa_3}{\kappa_2^{3/2}} H_3(\xi) + \frac{1}{24}\frac{\kappa_4}{\kappa_2^2} H_4(\xi) + \frac{1}{72}\frac{\kappa_3^2}{\kappa_2^3} H_6(\xi) + \cdots]Z(\xi). \qquad (1.5)$$

The expression for $f(\xi)$ is similar to (1.5) with $\kappa_r\,(r > 2)$ replaced by κ_r^* .

Subtracting the series $f(\xi)$ from the series $p(\xi)$ we obtain

$$p(\xi) = f(\xi) + Z(\xi)\left[\frac{1}{6}\frac{c_3}{\kappa_2^{3/2}} H_3(\xi) + \frac{1}{24}\frac{c_4}{\kappa_2^2} H_4(\xi) + \cdots\right]. \tag{1.6}$$

This infinite series is not uniformly convergent. If we integrate (1.6) term by term and make use of the first few terms in brackets we obtain a better approximation than that given by the integral of $f(\xi)$ alone. Retaining terms up to $O(n^{-3/2})$ we derive an approximation to the probability integral

$$\int_0^x p(x)dx = \int_0^\xi p(\xi)d\xi.$$

That is,

$$\int_0^x p(x)dx \doteqdot \int_0^\xi f(\xi)d\xi - Z(\xi)\left\{\frac{1}{6}\frac{c_3}{\kappa_2^{3/2}} H_2(\xi)\right.$$

$$+ \frac{1}{24}\frac{c_4}{\kappa_2^2} H_3(\xi) + \frac{1}{72}\frac{c_{33}}{\kappa_2^3} H_5(\xi) + \frac{1}{120}\frac{c_5}{\kappa_2^{5/2}} H_4(\xi) \tag{1.7}$$

$$\left. + \frac{1}{144}\frac{c_{34}}{\kappa_2^{7/2}} H_6(\xi) + \frac{1}{1296}\frac{c_{333}}{\kappa_2^{9/2}} H_8(\xi)\right\}$$

where $c_{33} = \kappa_3^2 - \kappa_3^{*2}$, $c_{34} = \kappa_3\kappa_4 - \kappa_3^*\kappa_4^*$ and $c_{333} = \kappa_3^3 - \kappa_3^{*3}$.

The first term in (1.7) is just the approximation (1.1a).

§2. Pearson

Pearson (1959) improved on Patnaik's approximation (1.1) by introducing an additional constant b, and choosing b, c and f so that the first three moments of the noncentral-χ^2 distribution, $\chi_\nu'^2(\lambda)$, are equal to the first three moments of $c\chi_f^2 + b$. These values were found to be

$$b = \frac{-\lambda^2}{\nu+3\lambda} \quad , \quad c = \frac{\nu+3\lambda}{\nu+2\lambda} \quad , \quad f = \frac{(\nu+2\lambda)^3}{(\nu+3\lambda)^2} .$$

That is,

$$\chi_\nu'^2(\lambda) \mathrel{\dot\sim} c\chi_f^2 + b. \tag{2.1}$$

§3. Johnson

(i) Let $x = \sum_{i=1}^{\nu} (u_i + \delta_i)^2$, where $u_i \sim N(0,1)$ $(i = 1,\cdots,\nu)$

and δ_i $(i = 1,\cdots,\nu)$ are constants with $\sum_{i=1}^{\nu} \delta_i^2 = \lambda$. Then $x \sim \chi_\nu'^2(\lambda)$.

If ν is even, then integrating with respect to x repeatedly by parts, Johnson (1959) found that

$$\Pr\{x < X\} = \sum_{i=o}^{\infty} \frac{e^{-\frac{1}{2}\lambda}(\frac{1}{2}\lambda)^i}{i!} e^{-\frac{1}{2}X}\left\{\frac{(\frac{1}{2}X)^{\frac{1}{2}\nu+j}}{(\frac{1}{2}\nu+j)!} + \frac{(\frac{1}{2}X)^{\frac{1}{2}\nu+j+1}}{(\frac{1}{2}\nu+j+1)!} + \cdots\right\}.$$

This can be written as

$$\Pr\{x < X\} = \sum_{i=o}^{\infty} \Pr\{\ell = i\} \Pr\{\xi \geq \tfrac{1}{2}\nu+i\} ,$$

where ξ is a Poisson variable with expected value $\frac{1}{2}X$ and ℓ is a Poisson variable with expected value $\frac{1}{2}\lambda$.

Assuming that ℓ and ξ are independent we obtain

$$\Pr\{x < X\} = \Pr\{\xi - \ell \geq \tfrac{1}{2}\nu\}. \tag{3.1}$$

By applying a Normal approximation to the right-hand side of (3.1) we obtain

$$\Pr\{x < X\} \doteqdot \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-\frac{1}{2}y^2} dy, \tag{3.2}$$

where $t = (X - \lambda - \nu + 1)/\sqrt{2(X+\lambda)}$.

(ii) Johnson also approximated the noncentral-χ^2 by a *mixture* of central-χ^2's. Let $\chi_\nu'^2(\lambda)$ be a mixture of two central-χ^2's with ν_1 and ν_2 degrees of freedom in the ratio $c : 1-c$. The values of ν_1, ν_2 and c obtained are

$$\begin{aligned} \nu_1 &= \nu + \lambda + 1 + \sqrt{2\lambda + 1} \\ \nu_2 &= \nu + \lambda + 1 - \sqrt{2\lambda + 1} \\ c &= \tfrac{1}{2}[1 - (2\lambda + 1)^{-\frac{1}{2}}] . \end{aligned}$$

Therefore, we have the approximation

$$\chi_\nu'^2(\lambda) \doteqdot c\chi_{\nu_1}^2 + (1 - c)\chi_{\nu_2}^2 . \tag{3.3}$$

§4. Abdel-Aty

Abdel-Aty (1954) makes the transformation

$$y = \left(\frac{\chi_\nu'^2}{\nu + \lambda}\right)^h , \text{ where } h \text{ is a positive constant.}$$

To determine h, he examines the cumulants of y expressed in terms of the cumulants of $\chi_\nu'^2$ as power series in r^{-s}. He finds that, after expansion, $h = \frac{1}{3}$ is an appropriate value to choose for minimizing the coefficient of r^{-2} in $\kappa_2(y)$, and reducing the higher cumulants, $\kappa_s(y)$ for $s > 2$, to a minimum. Abdel-Aty obtains the normal approximation

$$y \stackrel{\cdot}{\sim} N\left[1 - \frac{2}{9}\left\{\frac{1+b}{r}\right\} \quad , \quad \frac{2}{9}\left\{\frac{1+b}{r}\right\}\right] , \tag{4.1}$$

where $b = \frac{\lambda}{\nu+\lambda}$ and $r = \nu + \lambda$.

This approximation can be arrived at much more quickly by using, directly, the Wilson-Hilferty [see Johnson and Kotz (1970b), Page 176] approximation to the central-χ_f^2, i.e., $\chi_f^2 \doteqdot f\left(U\sqrt{\frac{2}{9f}} + 1 - \frac{2}{9f}\right)^3$ where $U \sim N(0,1)$ and applying it to Patnaik's two-moment central-χ^2 approximation (1.1),

i.e. $\chi_\nu'^2 \doteqdot \rho\chi_f^2 \doteqdot \rho f\left(U\sqrt{\frac{2}{9f}} + 1 - \frac{2}{9f}\right)^3$

where $\rho = (\nu+2\lambda)/(\nu+\lambda)$ and $f = (\nu+\lambda)^2/(\nu+2\lambda)$.

§5. Sankaran

(i) Sankaran (1959) modified Abdel-Aty's method by taking $y = \left(\frac{\chi_\nu'^2}{\nu+\lambda}\right)^h$ with h no longer fixed at $\frac{1}{3}$ but taken as that function of ν and λ which caused the leading term in the third cumulant of y to vanish.

He obtains

$$\left(\frac{\chi_\nu'^2}{\nu+\lambda}\right)^h$$

approximately normally distributed with expected value

$$1 + \frac{h(h-1)(\nu+2\lambda)}{(\nu+\lambda)^2} - \frac{h(h-1)(2-h)(1-3h)(\nu+2\lambda)^2}{2(\nu+\lambda)^4}$$

and variance (5.1)

$$\frac{2h^2(\nu+2\lambda)}{(\nu+\lambda)^2}\left[1 - \frac{(1-h)(1-3h)(\nu+2\lambda)}{(\nu+\lambda)^2}\right]$$

with

$$h = 1 - \frac{2}{3}\,\frac{(\nu+\lambda)(\nu+3\lambda)}{(\nu+2\lambda)^2} .$$

(ii) Sankaran (1963) discusses two other approximations. He makes the transformation

$$Z = \left(\frac{(\chi_\nu'^2 - b)}{(\nu+\lambda)}\right)^{\frac{1}{2}} .$$

Taking $b = \frac{1}{2}(\nu-1)$ he obtains the approximation

$$[\chi_\nu'^2 - \tfrac{1}{2}(\nu-1)]^{\frac{1}{2}} \mathrel{\dot{\sim}} N[\{\lambda+\tfrac{1}{2}(\nu-1)\}^{\frac{1}{2}}, 1] , \tag{5.2}$$

and taking $b = \frac{1}{3}(\nu - 1)$ he obtains

$$\{[\chi_\nu'^2 - \tfrac{1}{3}(\nu - 1)]/(\nu+\lambda)\}^{\frac{1}{2}} \mathrel{\dot{\sim}} N\left[\{1 - \tfrac{1}{3}\,\frac{(\nu+2)}{(\nu+\lambda)}\}^{\frac{1}{2}},\ \frac{1}{(\nu+\lambda)}\right] \tag{5.3}$$

or

$$[\chi_\nu'^2 - \tfrac{1}{3}(\nu-1)]^{\frac{1}{2}} \mathrel{\dot{\sim}} N[\{\lambda + \tfrac{2}{3}(\nu-1)\}^{\frac{1}{2}}, 1] . \tag{5.3a}$$

A better approximation, using $b = \frac{1}{3}(\nu-1)$, would be

$$[\chi_\nu'^2 - \tfrac{1}{3}(\nu-1)]^{\frac{1}{2}} \mathrel{\dot{\sim}} N\left[\{\lambda + \tfrac{2}{3}(\nu-1)\}^{\frac{1}{2}},\ 1 - \frac{\nu-1}{6(\nu+\lambda)}\right] \tag{5.4}$$

by adding a further term of $\kappa_2(Z_2)$, the second cumulant of Z_2, where $\kappa_2(Z_2)$ and Z_2 are defined in Sankaran (1963).

§6. Roy and Mohamad

Roy and Mohamad (1964) make the transformation

$$y = \chi_\nu'^2/2\rho, \text{ where } \rho \text{ is a constant.}$$

Deriving a *Laguerre series* expansion for p(y), the density function of y, they obtained two additional corrective terms to Patnaik's approximation (1.1) as follows:

Let us denote the Laguerre polynomials as

$$L_r^{(m)}(y) = \sum_{t=o}^{\infty} c_{r,t}^{(m)}(-y)^t/t! \quad , \tag{6.1}$$

where

$$c_{r,t}^{(m)} = (m+t)(m+t+1)\cdots(m+r-1)/(r-t)! \quad \text{for } t=0,1,\cdots,r-1$$
$$= 1 \quad \text{for } t = r. \tag{6.2}$$

The formal expansion of p(y) is

$$p(y) = p_m(y) \sum_{r=o}^{\infty} a_r^{(m)} L_r^{(m)}(y) \quad , \tag{6.3}$$

where

$$a_r^{(m)} = \int_o^{\infty} \frac{L_r^{(m)}(y)\ p(y)}{c_{r,o}^{(m)}} dy \tag{6.4}$$

and
$$p_m(y) = \frac{1}{\Gamma(m)} e^{-y} y^{m-1} \qquad o \leq y < \infty$$
$$= o \qquad \text{otherwise} .$$

The first four cumulants of y and the fourth and fifth coefficients $a_3^{(m)}$ and $a_4^{(m)}$ are calculated ($a_o^{(m)} = 1$ and $a_1^{(m)}$ and $a_2^{(m)}$ are set equal to 0).

Using the first five terms of (6.3) they obtained the approximation

$$p(y) \doteq [1 + a_3^{(m)} L_3^{(m)}(y) + a_4^{(m)} L_4^{(m)}(y)]\, p_m(y). \qquad (6.5)$$

p(y) can be put into an alternate form, i.e.,

$$p(y) \doteq p_m(y) + b_3^{(m)} [p_m(y) - 3p_{m+1}(y) + 3p_{m+2}(y) - p_{m+3}(y)]$$
$$+ b_4^{(m)} [p_m(y) - 4p_{m+1}(y) + 6p_{m+2}(y) - 4p_{m+3}(y) + p_{m+4}(y)], \qquad (6.6)$$

where

$$b_3^{(m)} = \frac{\lambda^2 m}{3(\nu+2\lambda)^2}, \quad b_4^{(m)} = \frac{\lambda^2 m(\nu+4\lambda)}{4(\nu+2\lambda)^3}, \quad m = \frac{(\nu+\lambda)^2}{2(\nu+2\lambda)}$$

and $\rho = \frac{(\nu+2\lambda)}{\nu+\lambda}$.

Let $F_{\nu,\lambda}(x) = \Pr\{\chi'^2_\nu \leq x\}$.

Then $F_{\nu,\lambda}(x) = \int_0^y p(t)dt$, where $y = \frac{x}{2\rho}$,

and

$$F_{\nu,\lambda}(x) \doteq P_m(y) + b_3^{(m)}[P_m(y) - 3P_{m+1}(y) + 3P_{m+2}(y) - P_{m+3}(y)]$$
$$+ b_4^{(m)}[P_m(y) - 4P_{m+1}(y) + 6P_{m+2}(y) - 4P_{m+3}(y) + P_{m+4}(y)], \qquad (6.7)$$

where $P_m(y) = \int_0^y p_m(t)dt$.

Putting $u = 2y = {}^x/\rho$ and $n = 2m = \frac{(\nu+\lambda)^2}{\nu+2\lambda}$, (6.7) becomes

$$F_{\nu,\lambda}(x) \doteq F_n(u) + b_3^{(m)}[F_n(u) - 3F_{n+2}(u) + 3F_{n+4}(u) - F_{n+6}(u)]$$
$$+ b_4^{(m)} [F_n(u) - 4F_{n+2}(u) + 6F_{n+4}(u) - 4F_{n+6}(u) + F_{n+8}(u)], \qquad (6.8)$$

where $F_\nu(y) = \int_0^y f_\nu(t)dt$ is the distribution of the central chi-square statistic with ν degrees of freedom.

The first term in (6.8) is just Patnaik's approximation (1.1a).

§7. Tiku

Tiku (1965) derived *Laguerre series forms* for three transformations, namely,

$$X = \tfrac{1}{2}\chi_\nu'^2 \quad , \quad Y = \frac{1}{2\rho}\chi_\nu'^2 \quad , \quad Z = \frac{1}{2\rho'}(\chi_\nu'^2+d).$$

(i) Let the noncentrality parameter of $\chi_\nu'^2$ be λ and let $m = \tfrac{1}{2}\nu$.

He obtains the approximation, using the transformation X,

$$P_m'(x_o) \doteqdot P_m(x_o) + \sum_{r=1}^{5} P_{(r)}(x_o), \tag{7.1}$$

where $P_m'(x_o) = \int_{x_o}^{\infty} p(X)dX$, $P_m(x_o) = \int_{x_o}^{\infty} p_m(X)dX$

and $p_m(X)$ and $p(X)$ are defined as in §6 above. The coefficients $a_r^{(m)}$ are found to be

$$a_o^{(m)} = 1, \quad a_r^{(m)} = \frac{\Gamma(m)}{\Gamma(m+r)}(-\tfrac{1}{2}\lambda)^r \qquad (r \geqslant 1)$$

and

$$P_{(r)}(x_o) = \frac{-1}{r(r-1)(m+r-1)}\Big\{(\tfrac{1}{2}\lambda)(r-1)(m+2r-3-x_o)P_{(r-1)}(x_o)$$
$$+ (\tfrac{1}{2}\lambda)^2(r-2)P_{(r-2)}(x_o)\Big\}$$

for $r = 2,3,4,\cdots\cdots$,

where $P_{(o)}(x_o) = o$ and $P_{(1)}(x_o) = (\tfrac{1}{2}\lambda)\dfrac{e^{-x_o}x_o^m}{\Gamma(m+1)}$.

(ii) Let

$$p(Y) = p_n(Y) \sum_{r=o}^{\infty} \beta_r L_r^{(n)}(Y) . \qquad (7.2)$$

We must find n and ρ such that $Y = \frac{1}{2}\chi_\nu'^2/\rho$ has its first two moments equal to those of $p_n(Y)$. We obtain

$$\rho = \frac{\nu+2\lambda}{\nu+\lambda} \quad \text{and} \quad n = \frac{(\nu+\lambda)^2}{2(\nu+2\lambda)}$$

and the coefficients β_r are found to be

$$\beta_1' = \beta_2' = o \; ; \quad \beta_3' = 2\lambda^2/(\nu+2\lambda)^2$$

$$\beta_4' = 6\lambda^2(\nu+4\lambda)/(\nu+2\lambda)^3 \qquad (7.2a)$$

$$\beta_5' = 24\lambda^2(\nu^2+6\nu\lambda + 11\lambda^2)/(\nu+2\lambda)^4$$

with $\beta_r' = \dfrac{\Gamma(n+r)}{\Gamma(n+1)} \beta_r$.

From the first four non-zero terms of series (7.2) we obtain the approximation

$$P_m'(x_o) \doteqdot P_n(y_o) + \sum_{r=3}^{5} Q_{(r)}(y_o), \qquad (7.3)$$

where

$$P_n(y_o) = \int_{y_o}^{\infty} p_n(Y)\,dY$$

and

$$Q_{(r)}(y_o) = \frac{(-1)^r}{r!} \beta_r'(\Delta^{r-1}N_1) \left[e^{-y_o} \frac{y_o^n}{\Gamma(n)}\right] \quad (r \geq 3)$$

where $N_1 = 1$, $N_r = \left(\dfrac{y_o}{n+r-1}\right) N_{r-1} \quad (r = 2,3, \cdots)$

and $\Delta^{r-1}N_1$ is the (r-1)th *forward difference* of N_1 .

[Note: $p_n(Y)$ is Patnaik's two-moment central-χ^2 approximation (1.1)]

(iii) Expanding the distribution of $Z = \frac{1}{2\rho'} (\chi_\nu'^2 + d)$ in terms of Laguerre polynomials, we obtain

$$f(Z) = p_k(Z) \sum_{r=o}^{\infty} \gamma_r L_r^{(k)}(Z), \tag{7.4}$$

where k, d and ρ' are determined such that Z has its first three moments equal to those of $p_k(Z)$. We obtain

$$k = \frac{(\nu+2\lambda)^3}{2(\nu+3\lambda)^2}, \quad d = \frac{\lambda^2}{\nu+3\lambda}, \quad \rho' = \frac{\nu+3\lambda}{\nu+2\lambda}$$

and the coefficients are

$$\gamma_0' = 1, \quad \gamma_1' = \gamma_2' = \gamma_3' = o$$

$$\gamma_4' = -6\lambda^2/(\nu+3\lambda)^2, \quad \gamma_5' = -48\lambda^2(\nu+4\lambda)/(\nu+3\lambda)^3,$$

where $\gamma_r' = \frac{\Gamma(k+r)}{\Gamma(k+1)} \gamma_r$.

From the first three non-zero terms of (7.4) we obtain the approximation

$$P_m'(x_o) \doteqdot P_k(z_o) + \sum_{r=4}^{5} R_{(r)}(z_o), \tag{7.5}$$

where

$$P_k(z_o) = \int_{z_o}^{\infty} p_k(Z)dZ$$

and $R_{(r)}(z_o)$ is the same as $Q_{(r)}(y_o)$ with β_r' , y_o and n replaced by γ_r' , z_o and k, respectively.

[Note: $p_k(Z)$ is Pearson's 3-moment central-χ^2 approximation (2.1).]

§8. Germond and Hastings

Germond and Hastings [see Johnson and Kotz (1970c), Page 143] give the approximation

$$\Pr\{\chi_2'^2(\lambda) \leq R^2\} \doteqdot \frac{R^2}{2 + R^2/2} \exp\left[-\frac{\lambda}{2 + R^2/2}\right] \qquad (8.1)$$

which is correct to 4 decimal places for $R \leq 0{,}4$. For $R > 5$, they give the approximation

$$\Pr\{\chi_2'^2(\lambda) \leq R^2\} \doteqdot \frac{1}{\sqrt{2\pi}} \int_b^\infty e^{-\frac{1}{2}t^2} dt \, , \qquad (8.2)$$

where $b = \sqrt{\lambda} - \sqrt{R^2-1}$.

§9. Tukey

Tukey (1957) obtained a simple empirical approximation to the upper 5% point of the noncentral-χ^2. The approximation is

$$\chi'^2_{0,05} \doteqdot \left[1{,}6449 + \sqrt{\lambda} + 0{,}51\,\frac{(\nu-1)}{\sqrt{\lambda}+1} - 0{,}024\,\frac{(\nu-5)(\nu-1)}{\sqrt{\lambda}(\sqrt{\lambda}+1)}\right]^2$$

(9.1)

§10. Another approximation

This approximation makes use of Pearson's 3-moment central-χ^2 approximation (2.1), namely,

$$\chi_\nu'^2 \doteqdot c\chi_f^2 + b \, , \qquad (10.1a)$$

where

$$c = \frac{\nu+3\lambda}{\nu+2\lambda}, \quad f = \frac{(\nu+2\lambda)^3}{(\nu+3\lambda)^2}$$

and $$b = \frac{-\lambda^2}{\nu+3\lambda} .$$

We now approximate χ_f^2 in (10.1a) by the Wilson-Hilferty normal approximation [see Johnson and Kotz (1970b), Page 176], namely,

$$\chi_f^2 \doteqdot f\left(U\sqrt{\frac{2}{9f}} + 1 - \frac{2}{9f}\right)^3, \text{ where } U \sim N(0,1). \quad (10.1b)$$

Therefore, combining approximations (10.1a) and (10.1b) we obtain the approximation

$$\chi_\nu'^2(\lambda) \doteqdot cf\left(U\sqrt{\frac{2}{9f}} + 1 - \frac{2}{9f}\right)^3 + b, \quad (10.2)$$

where c, f, b and U have been defined above.

§11. Two central-χ^2 approximations

(i) It would be convenient if a good central-χ^2 approximation could be found where the degrees of freedom of the central- and noncentral-χ^2 are the same. We, therefore, make the approximation

$$\chi_\nu'^2(\lambda) \doteqdot k\chi_\nu^2 + a ,$$

where k and a are chosen such that the first two central moments of $\chi_\nu'^2(\lambda)$ and the first two central moments of $k\chi_\nu^2 + a$ are equal. To do this we equate $E(\chi_\nu'^2)$ with $E(k\chi_\nu^2 + a)$ and $var(\chi_\nu'^2)$ with $var(k\chi_\nu^2 + a)$, i.e.,

$$\nu + \lambda = k\nu + a$$

and $\quad 2(\nu+2\lambda) = 2k^2\nu.$

Solving for k and a we obtain

$$k = +\sqrt{\frac{\nu+2\lambda}{\nu}} \quad \text{and} \quad a = \nu\left[1 - \sqrt{\frac{\nu+2\lambda}{\nu}}\right] + \lambda .$$

The approximation is, therefore,

$$\chi_\nu'^2(\lambda) \doteqdot \sqrt{\frac{\nu+2\lambda}{\nu}}\,\chi_\nu^2 + \nu\left[1 - \sqrt{\frac{\nu+2\lambda}{\nu}}\right] + \lambda . \qquad (11.1)$$

However, this approximation is only accurate for small λ as can be seen from Table I, Section III.

(ii) For the second approximation we use Pearson's 3-moment central-χ^2 approximation (2.1), namely,

$$\chi_\nu'^2(\lambda) \doteqdot c\chi_f^2 + b,$$

where $c = (\nu+3\lambda)/(\nu+2\lambda)$, $f = (\nu+2\lambda)^3/(\nu+3\lambda)^2$

and $b = (-\lambda^2)/(\nu+3\lambda)$.

It is obviously more convenient to have integer degrees of freedom for the central-χ^2. We, therefore, let f^* be the integer closest to f and equate the first two central moments of $\chi_\nu'^2(\lambda)$ with the first two central moments of $c^*\chi_{f^*}^2 + b^*$ as in (i) above.

We obtain

$$\nu + \lambda = c^*f^* + b^*$$

$$\text{and } 2(\nu+2\lambda) = 2c^{*2}f^*.$$

Solving for c^* and b^* (remembering that f^* is now fixed) we find

$$c^* = +\sqrt{\frac{\nu+2\lambda}{f^*}} \quad \text{and} \quad b^* = \nu + \lambda - f^*\sqrt{\frac{\nu+2\lambda}{f^*}} .$$

The approximation is, therefore,

$$\chi_\nu'^2(\lambda) \doteq \sqrt{\frac{\nu+2\lambda}{f^*}}\ \chi_{f^*}^2 + \nu + \lambda - f^* \sqrt{\frac{\nu+2\lambda}{f^*}} \qquad (11.2)$$

The accuracy of this approximation is shown in Section III, Tables I and II.

III. Comparisons of Accuracy

Tables I and II give comparisons between the exact and approximate values of the upper and lower 5% points of the noncentral-χ^2 distribution. The values in the columns marked with an asterisk (*) are taken from Johnson and Kotz (1970c), page 142. The exact values were taken from Johnson and Pearson (1969). Note that Germond and Hasting's approximation (8.2) is valid only for $\nu = 2$. Note also that Tukey's empirical approximation (9.1) is valid only for the *upper* 5% points.

It can be seen from the tables that approximations (2.1), (5.1), (10.2) and (11.2) are quite reliable over a wide range of values of λ. Approximations (1.1), (4.1) and (11.1) deteriorate as λ increases while others improve. It would seem that approximation (5.4) is the best to use due to the ease of application and its relative accuracy for larger λ.

Approximation (11.1) is much more accurate for small λ ($\lambda \leqslant 1$) especially for the upper 5% points. This is because that, as λ approaches zero, $\chi_\nu'^2(\lambda)$ approaches χ_ν^2 . Approximation (11.2) is almost as accurate as approximation (2.1) and is much easier to apply due to its integerized degrees of freedom.

Approximations (1.7), (6.8) and (7.3) are just improvements on Patnaik's approximation (1.1) but computations are tedious. Approximation (7.5) is an improvement on approximation (2.1) but is, again, tedious to compute. Approximations (1.2) and (3.2) would be more useful for higher degrees of freedom ν and noncentrality λ.

Note: It seems that the values in column (23.3) of Table 1, page 142 of Johnson and Kotz (1970c) are incorrect using the approximation (23.3) of page 140 of the same reference. Perhaps the values were obtained by using terms up to $O(r^{-4})$ in the first two cumulants of Z_2, $\kappa_1(Z_2)$ and $\kappa_2(Z_2)$, in §3 of Sankaran (1963), where $r = \nu+\lambda$ and $Z_2 = \{\chi'^2_\nu - \frac{1}{3}(\nu-1)\}^{\frac{1}{2}}/\sqrt{\nu+\lambda}$.

If one wishes to find a λ (or ν) that would yield a chance $1-\beta$ of establishing significance at level α, for given ν (λ), then the easiest approximations to use are (3.2) and (5.2) with (5.2) being much more accurate than (3.2).

For power calculations the approximations which make use of the standardized normal variable are the easiest to apply. These approximations are (1.2), (3.2), (4.1), (5.2), (5.3), (5.4), (8.2) for $\nu = 2$, and (10.2), approximations (5.4) and (10.2) being the most accurate.

Table I : Approximate and Exact Values for the Upper 5% points of $\chi_{\nu}'^{2}(\lambda)$

ν	λ	EXACT	PATNAIK		PEARSON	JOHNSON	ABDEL-ATY	SANKARAN			TUKEY	GERMOND AND HASTINGS			
			(1.1)*	(1.2)	(2.1)*	(3.2)*	(4.1)*	(5.1)*	(5.2)*	(5.4)	(9.1)	(8.2)	(10.2)	(11.2)	(11.1)
2	1	8,64	8,63	8,24	8,60	9,56	8,56	8,58	8,73	8,68	8,62	7,99	8,54	8,56	8,64
	4	14,64	14,72	14,25	14,58	15,19	14,66	14,63	14,68	14,68	14,65	14,29	14,54	14,58	14,92
	16	33,05	33,35	32,81	33,02	33,33	33,32	33,07	33,07	33,05	33,07	32,86	33,00	33,01	34,45
	25	45,31	45,66	45,12	45,28	45,54	45,64	45,31	45,32	45,30	45,32	45,15	45,27	45,27	47,35
4	1	11,71	11,72	11,37	11,69	12,59	11,67	11,69	11,91	11,84	11,87		11,66	11,52	11,72
	4	17,31	17,38	16,95	17,27	17,88	17,34	17,28	17,42	17,34	17,36		17,24	17,26	17,51
	16	35,43	35,69	35,17	35,40	35,73	35,66	35,43	35,47	35,42	35,46		35,38	35,41	36,46
	25	47,61	47,94	47,41	47,59	47,86	47,91	47,62	47,64	47,61	47,64		47,58	47,59	49,16
7	1	16,00	16,01	15,69	15,99	16,83	15,98	15,98	16,28	16,25	16,25		15,98	16,01	16,01
	4	21,22	21,28	20,89	21,21	21,82	21,25	21,26	21,41	21,33	21,32		21,19	21,22	21,35
	16	38,98	39,16	38,70	38,95	39,30	39,16	38,96	39,05	38,98	38,97		38,93	38,92	39,68
	25	51,07	51,34	50,83	51,04	51,32	51,33	51,06	51,11	51,06	51,06		51,04	51,03	52,16

Table II : Approximate and Exact Values for Lower 5% Points of $\chi'^2_{\nu}(\lambda)$

ν	λ	EXACT	PATNAIK		PEARSON	JOHNSON	ABDEL-ATY	SANKARAN					
			(1.1)*	(1.2)	(2.1)*	(3.2)*	(4.1)*	(5.1)*	(5.2)*	(5.4)	(10.2)	(11.2)	(11.1)
2	1	0,17	0,20	0,03	0,08	-	0,17	0,08	0,67	0,43	0,05	0,21	0,32
	4	0,65	0,94	0,60	0,53	0,22	0,89	0,66	0,73	0,62	0,50	0,55	1,76
	16	6,32	6,89	6,41	6,30	6,07	6,87	6,34	6,34	6,31	6,29	6,28	10,18
	25	12,08	12,68	12,17	12,07	11,87	12,67	12,11	12,09	12,07	12,06	12,06	17,33
4	1	0,91	0,93	0,68	0,88	0,84	0,91	0,87	1,50	1,03	0,86	0,77	0,97
	4	1,77	1,95	1,61	1,71	1,53	1,94	1,74	2,00	1,73	1,69	1,67	2,30
	16	7,88	8,36	7,89	7,86	7,68	8,35	7,90	7,94	7,86	7,86	7,88	10,13
	25	13,73	14,26	13,77	13,72	13,56	14,26	13,78	13,77	13,71	13,72	13,73	15,92
7	1	2,49	2,51	2,23	2,49	2,59	2,49	2,47	3,13	2,49	2,47	2,52	2,52
	4	3,66	3,78	3,44	3,64	3,59	3,76	3,65	4,00	3,59	3,63	3,67	3,93
	16	10,26	10,64	10,19	10,25	10,11	10,63	10,27	10,37	10,20	10,24	10,22	11,59
	25	16,22	16,68	16,20	16,22	16,09	16,67	16,25	16,30	16,18	16,21	16,20	18,21

IV. Applications

§1. Power functions

One of the most important uses of non-central distributions is the evaluation of *power functions*. The power function is used to detect the extent of departures from the null hypothesis which will be significant at a prescribed level with a given probability. The power function is also used to determine in advance the sample size necessary to make sure that a worthwhile difference will be established as significant, if it exists.

Suppose $x_1, x_2, \cdots\cdots, x_\nu$ are ν independent observations in a sample. Let the null hypothesis be

$$H_0 \; : \; x_i \sim N(0,1) \qquad (i = 1,2, \cdot\cdot,\nu).$$

If H_0 is true, then $\sum_{i=1}^{\nu} x_i^2$ is distributed as a central-χ^2 with ν degrees of freedom and $\Pr\{\chi_\nu^2 > \chi_\alpha^2\} = \Pr\{\sum_{i=1}^{\nu} x_i^2 > \chi_\alpha^2\} = \alpha$, where χ_α^2 is the α-significance point of the χ^2-distribution.

The power of the χ^2-test is given by the probability that $\sum_{i=1}^{\nu} x_i^2$ is greater than χ_α^2 under an alternative hypothesis H_1. Let the alternative hypothesis be

$$H_1 \; : \; x_i \sim N(\mu_i,1) \qquad (i = 1,2, \cdots,\nu).$$

We now have that $\sum_{i=1}^{\nu} x_i^2$ is distributed as a noncentral-χ^2 with ν degrees of freedom and noncentrality parameter $\lambda = \sum_{i=1}^{\nu} \mu_i^2$. The power function is given by

$$\int_{\chi_\alpha^2}^{\infty} p(\chi_\nu'^2 | \lambda) d\chi'^2 = 1 - \beta.$$

Patnaik (1949) has evaluated the power function at level of significance $\alpha = 0{,}05$ for values of $\lambda = 2(2)20$ and $\nu = 2(1)10(2)20$ using his approximation (1.7).

There are three types of problems that can arise. These are: (a) For given λ and ν , what is the probability of significance at level α ? (b) For given ν, how large must λ be to have a chance $1-\beta$ of establishing significance at level α when a real difference in the μ_i exists? (c) For given λ, how large a sample must be taken to have a chance $1-\beta$ of establishing significance?

Example: (a) Suppose ν is given and $\lambda = 16$.

What is the probability of significance at level $\alpha = 0{,}05$?

The following table gives values of the power $1-\beta$ for $\nu = 2$ and 20 for seven different approximations numbered as in section II of this chapter.

		(1.2)	(1.7)	(3.2)	(4.1)	(5.2)	(5.4)	(8.2)	Exact
ν	2	0,958	0,956	0,952	0,969	0,957	0,962	0,961	0,956
	20	0,653	0,648	0,644	0,647	0,644	0,642	-	0,648

(b) Suppose ν is given and $\alpha = 0{,}05$. How large must λ be to yield power $1-\beta$?

The following table gives values of λ for $\nu = 2$ and 20 with power $1-\beta = 0{,}983$ and 0,648,respectively.

		(1.7)	(3.2)	(5.2)	(5.3)	(8.2)	Exact
ν	2	20,00	20,39	19,42	19,28	18,96	20
	20	16,00	16,10	16,10	16,35	-	16

(Note: In both of these tables in this example, (8.2) is valid for $\nu = 2$ only.)

§2. χ^2-test for the goodness-of-fit

The χ^2-test for goodness-of-fit is a test which compares the observed frequencies with the expected frequencies under a given hypothesis. Divide the data into k groups and denote the observed frequencies by n_i $(i = 1,2, \cdots,k)$ and the expected frequencies by $N\pi_i$ $(i = 1, \cdot\cdot,k)$, where N is the total number of observations. Then we have

$$\sum_{i=1}^{k} n_i = \sum_{i=1}^{k} N\pi_i = N.$$

It can be shown [Patnaik (1949)] that

$$\phi^2 = \sum_{i=1}^{k} \frac{(n_i - N\pi_i)^2}{N\pi_i} \tag{2.1}$$

is approximately distributed as central-χ^2 with k-1 degrees of freedom when the $N\pi_i$ are the true expectations.

However, suppose that Np_i $(i = 1,\cdots,k)$ are the true expectations. Patnaik (1949) shows that

$$\phi'^2 = \sum_{i=1}^{k} \frac{(n_i - N\pi_i)^2}{Np_i} \tag{2.2}$$

is approximately distributed as a noncentral-χ^2 with k-1 degrees of freedom and noncentrality parameter

$$\lambda' = N \sum_{i=1}^{k} \frac{(p_i - \pi_i)^2}{p_i} . \qquad (2.2a)$$

The sums of squares needed is the ϕ^2 of (2.1) and not the ϕ'^2 of (2.2). If we introduce a further approximation by replacing Np_i with $N\pi_i$ then, under the alternative hypothesis,

$$\phi^2 = \sum_{i=1}^{k} \frac{(n_i - N\pi_i)^2}{N\pi_i} \qquad (2.3)$$

is distributed as a noncentral-χ^2 with k-1 degrees of freedom and noncentrality parameter

$$\lambda = N \sum_{i=1}^{k} \frac{(p_i - \pi_i)^2}{\pi_i} = N \left\{ \sum_{i=1}^{k} \frac{p_i^2}{\pi_i} - 1 \right\} . \qquad (2.3a)$$

Patnaik (1949) shows that this further approximation should not be serious if the differences $N\pi_i - Np_i$ are small compared to $N\pi_i$.

With this result we are now able to determine the power of the goodness-of-fit test of any simple hypothesis H_o (specifying probabilities π_i) with respect to a simple alternative hypothesis H_1 (specifying probabilities p_i).

The power function can be used to solve three problems connected with the goodness-of-fit test. These are: (a) For a given sample size N and number of groups k, what is the chance of establishing the inadequacy of the hypothesis H_o using a given significance level? (b) For given k, how many observations are necessary to give a chance 1-β of establishing significance at level α? (c) For given k and N, how large will λ be with given chance 1-β ?

<u>Examples</u>: We shall use Patnaik's example [see Patnaik (1949)] from genetics. In this example, the offspring are of four types with frequencies in the proportions 9,3,3,1. We shall test whether these proportions are true against the alternative that the frequencies should be in the proportions 9,3,3r,r ($r < 1$). The expected frequencies are

$$\pi_i \;:\; \frac{9}{16}\,,\;\frac{3}{16}\,,\;\frac{3}{16}\,,\;\frac{1}{16}$$

and $$p_i \;:\; \frac{9}{4(3+r)}\,,\;\frac{3}{4(3+r)}\,,\;\frac{3r}{4(3+r)}\,,\;\frac{r}{4(3+r)}\;.$$

Therefore, $$\lambda = N\left(\frac{4(3+r^2)}{(3+r)^2} - 1\right).$$

(a) Suppose $N = 100$ and $\alpha = 0{,}05$. We shall test $H_0 : r = 1$ against $H_1 : r = \frac{1}{2}$ at level α. We want to find power $1-\beta$.

Now, $k - 1 = 3$ and $\lambda = {}^{300}/49$

Patnaik's approximation (1.7) yields $1-\beta = 0{,}52$. [see Patnaik (1949), Table 6].

Other approximations yield the following values:

$1-\beta = 0{,}54$ for (1.2),

$1-\beta = 0{,}52$ for (3.2),

$1-\beta = 0{,}52$ for (4.1),

$1-\beta = 0{,}53$ for (5.2),

$1-\beta = 0{,}52$ for (5,3),

$1-\beta = 0{,}53$ for (5.4).

(b) Suppose we want a 90% chance of detecting that $r = \frac{1}{2}$ using the 5% level. How large must N be?

Patnaik's values are $\lambda = 14{,}1$ and, therefore,

$$N = \frac{49}{3}\lambda = 230{,}3.$$

Therefore, we shall need a sample of size N = 230.

Other approximations yield:

$N = 234$		for (3.2),
$N = 231$		for (5.2),
$N = 233$		for (5.3).

(c) If $N = 100$, $\alpha = 0{,}05$, how small must r be to give a 50 : 50 chance for establishing significance?

We have $1-\beta = 0{,}50$. Patnaik's values are $\lambda = 5{,}8$ and, therefore,

$$r = 0{,}51.$$

Other approximations yield:

$r = 0{,}52$	for (1.2),
$r = 0{,}51$	for (3.2),
$r = 0{,}51$	for (5.2),
$r = 0{,}51$	for (5.3).

§3. <u>The distribution of a sample variance from a normal population with unstable expected value</u>

(i) We have $S = \sum_{j=1}^{n} (X_j - \overline{X})^2$, where $\overline{X} = \frac{1}{n}\sum_{j=1}^{n} X_j$,

$$X_j \sim N(\mu_j, \sigma^2) \qquad (j = 1,2, \cdots,n)$$

and all the X_j's are independent.

Let $U_j' = \dfrac{X_j - \mu_j}{\sigma} \sim N(0,1) \qquad (j = 1,2, \cdots,n)$,

then $X_j = \mu_j + \sigma U_j'$.

Therefore,

$$S = \sigma^2 \sum_{j=1}^{n} [U_j' + \mu_j \sigma^{-1} - (\bar{U}' + \bar{\mu}\, \sigma^{-1})]^2 ,$$

where $\bar{U}' = \frac{1}{n} \sum_{j=1}^{n} U_j'$ and $\bar{\mu} = \frac{1}{n} \sum_{j=1}^{n} \mu_j$.

Transforming U_1' , U_2' , ..., U_n' to U_1 , U_2 , ..., $U_{n-1}, \bar{U}'$

such that $\sum_{j=1}^{n} (U_j' - \bar{U}')^2 = \sum_{j=1}^{n-1} U_j^2$,

where U_1 , ... , U_{n-1} are each distributed independently as $N(0,1)$ variables, we see that

$$S = \sigma^2 \sum_{j=1}^{n-1} (U_j + \delta_j)^2 ,$$

where the δ_j's are linear functions of the μ_j's and the U_j's are linear functions of the U_j''s. Putting $U_j' = o$ for all j, it follows that $U_j = o$ for all j and, therefore,

$$\sum_{j=1}^{n-1} \delta_j^2 = \sum_{j=1}^{n} (\mu_j - \bar{\mu})^2/\sigma^2 .$$

Hence, we have that S is distributed as σ^2 times a noncentral-χ^2 with n-1 degrees of freedom and noncentrality parameter

$$\lambda = \sum_{j=1}^{n} (\mu_j - \bar{\mu})^2/ \sigma^2,$$

i.e., $$S \sim \sigma^2 \chi'^2_{n-1} (\lambda) .$$

(ii) The standard hypothesis testing situations concerning the variance σ^2 are

$$H_o : \sigma^2 \geq \sigma_o^2 \quad \text{against} \quad H_1 : \sigma^2 < \sigma_o^2, \tag{3.1}$$

$$H_o : \sigma^2 \leq \sigma_o^2 \quad \text{against} \quad H_1 : \sigma^2 > \sigma_o^2 \quad \text{and} \tag{3.2}$$

$$H_o : \sigma^2 = \sigma_o^2 \quad \text{against} \quad H_1 : \sigma^2 \neq \sigma_o^2. \tag{3.3}$$

If $X_1, X_2, \ldots, X_n$ is a random sample from a normal population with $X_j \sim N(\mu,\sigma^2)$ $(j = 1,2,\ldots,n)$, then the statistic used to test the above hypotheses is

$$Y_{n-1} = \frac{\sum_{j=1}^{n} (X_j-\bar{X})^2}{\sigma_o^2} = \frac{S}{\sigma_o^2}$$

which is distributed as central-χ^2 with n-1 degrees of freedom.

However, if $X_1, X_2, \ldots, X_n$ is a random sample from a normal population with $X_j \sim N(\mu_j,\sigma^2)$ $(j = 1,2,..,n)$, then Y_{n-1} is distributed as noncentral-χ^2 with n-1 degrees of freedom and noncentrality parameter $\lambda_o = \sum_{j=1}^{n} (\mu_j-\bar{\mu})^2/\sigma_o^2$.

Suppose we wish to test the hypothesis (3.1). The critical region for level of significance α is

$$Y_{n-1} = \frac{S}{\sigma_o^2} < \chi'^2_{n-1;\alpha}(\lambda_o) .$$

<u>Example</u>: Let us test the hypothesis $H_o : \sigma^2 \geq \sigma_o^2 = 0{,}00027$ against $H_1 : \sigma^2 < \sigma_o^2 = 0{,}00027$ in a sample of size $n = 4$ at level of significance $\alpha = 0{,}05$. Let us suppose that $\sum_{j=1}^{4} (\mu_j-\bar{\mu})^2 = 4\sigma_o^2$, and that $\sum_{j=1}^{4} (X_j-\bar{X})^2 = 0{,}004$.

Therefore, we have that $\lambda_o = 4$ and $Y_{n-1} = \frac{S}{\sigma_o^2} = \frac{0{,}004}{0{,}00027} =$

14,815.

Using Johnson and Pearson's (1969) tables of percentage points of the noncentral-χ we find that $\chi'^2_{3;0,05}(4) = 15,984$. Therefore, we have $Y_{n-1} < \chi'^2_{3;0,05}(4)$ and, so, we reject H_o . Approximation (9.1) yields $\chi'^2_{3;0,05}(4) = 16,007$ and approximation (10.2) yields $\chi'^2_{3;0,05}(4) = 15,897$. Using these approximations we would reject H_o . For the calculation of 5% points of the noncentral-χ^2 distribution approximation (9.1) is the most convenient to use.

(iii) Suppose we wish to find the power of the test for hypothesis (3.1) above when $\sigma^2 = \sigma_1^2 < \sigma_o^2$. The power will be

$$\Pr\left(\frac{S}{\sigma_o^2} < \chi'^2_{n-1;\alpha}(\lambda_o)\right) = \Pr\left(\frac{S}{\sigma_1^2} < \frac{\sigma_o^2}{\sigma_1^2}\chi'^2_{n-1;\alpha}(\lambda_o)\right)$$
$$= \Pr\left(Y_{n-1} < \frac{\sigma_o^2}{\sigma_1^2}\chi'^2_{n-1,\alpha}(\lambda_o)\right) . \qquad (3.4)$$

(iv) Suppose we wish to find the minimum sample size that will yield a probability 1-β of establishing significance at level α (the minimum sample size problem is discussed in more detail in Chapter 4).

If we want (3.4) to be at least 1-β , then we must have

$$\frac{\sigma_o^2}{\sigma_1^2}\chi'^2_{n-1;\alpha}(\lambda_o) \geq \chi'^2_{n-1;1-\beta}(\lambda_1)$$

or

$$\frac{\chi'^2_{n-1;1-\beta}(\lambda_1)}{\chi'^2_{n-1;\alpha}(\lambda_o)} \leq \frac{\sigma_o^2}{\sigma_1^2} , \qquad (3.5)$$

where $\lambda_1 = \sum_{j=1}^{n} (\mu_j - \bar{\mu})^2/\sigma_1^2$.

The minimum n which will satisfy inequality (3.5) is the solution. This can be found quite easily if good noncentral-χ^2 tables are available.

<u>Example</u>: Suppose we have $\sigma_0^2 = 25\ \sigma_1^2$, i.e., $\frac{\sigma_0^2}{\sigma_1^2} = 25$ and $\lambda_0 = 1$, $\lambda_1 = 25$.

We must find the minimum n that will satisfy the inequality (3.5) with power $1-\beta=0{,}95$ and at level of significance $\alpha = 0{,}05$. Therefore, we obtain, using the Johnson and Pearson (1969) tables

$$\frac{\chi'^2_{6;0,95}(25)}{\chi'^2_{6;0,05}(1)} = 25{,}91 \quad \text{and} \quad \frac{\chi'^2_{7;0,95}(25)}{\chi'^2_{7;0,05}(1)} = 20{,}51.$$

The exact solution is, therefore, n-1 = 7 or n = 8.

Using approximation (11.2) we obtain

$$\frac{\chi'^2_{6;0,95}(25)}{\chi'^2_{6;0,05}(1)} = 25{,}44 \quad \text{and} \quad \frac{\chi'^2_{7;0,95}(25)}{\chi'^2_{7;0,05}(1)} = 20{,}25.$$

Approximation (11.1) yields the values

$$\frac{\chi'^2_{6;0,95}(25)}{\chi'^2_{6;0,05}(1)} = 26{,}09 \quad \text{and} \quad \frac{\chi'^2_{7;0,95}(25)}{\chi'^2_{7;0,05}(1)} = 20{,}70.$$

Both of these approximations yield the correct result, in this case. Approximation (11.1) yields the correct result because the noncentrality in the denominator is small. Approximation

(11.2) was used here because of its consistent accuracy and its ease of application.

(v) The type of situation in which there is an unstable mean could arise when sampling from a population in which the mean exhibits some kind of secular trend. This puts the statistic $Y_{n-1} = \frac{S}{\sigma_o^2}$ at a disadvantage because the usual mean square estimator of variance, $S/(n-1) = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$ will tend to be larger than the true population value, i.e., $S/(n-1)$ will be a biased estimator of σ^2. When the variation in the mean is gradual, so that a trend (not necessarily linear) is shifting the mean of the population, a simple method of minimizing the effect of the trend on dispersion is to estimate the variance from differences. Von Neumann *et al.* (1941) suggested the *mean square successive difference*,

$$D^2 = \frac{1}{2(n-1)} \sum_{i=1}^{n} (X_{i+1} - X_i)^2 ,$$

which is an unbiased estimator of σ^2, i.e., $E(D^2) = \sigma^2$. They also show, that, for large n,

$$\frac{D^2}{2} \mathrel{\dot\sim} N\left(2\sigma^2, \frac{4(3n-4)}{(n-1)^2}\sigma^4\right) .$$

§4. Quadratic forms

(i) Definition

Suppose $X' = (X_1, X_2, \ldots., X_n)$ is a random vector following a multivariate normal distribution with expected value vector $\mu' = (\mu_1, \mu_2, \ldots, \mu_n)$ and covariance matrix $\sum$.

The *quadratic form* $Q(X_1, X_2, \ldots, X_n)$ associated with the symmetric matrix A is defined as

$$Q(X) = Q(X_1, X_2, \ldots, X_n) = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} X_i X_j ,$$

where a_{ij} is the (i,j)th element of A.

(ii) <u>Conditions for a quadratic form to be distributed as noncentral-χ^2</u>

<u>Theorem</u>. *(Styan (1970))*

Let $X' = (X_1, X_2, \ldots, X_n)$ be a random normal vector with mean vector μ and covariance matrix Σ (not necessarily nonsingular). Let A be an n×n real symmetric matrix. A set of necessary and sufficient conditions for Q(x) to follow a noncentral-χ^2 distribution with r degrees of freedom and noncentrality parameter λ is

$$\Sigma A \Sigma A \Sigma = \Sigma A \Sigma ,$$
$$\mathrm{rank}\ (\Sigma A \Sigma) = \mathrm{tr}\ (A\Sigma) = r,$$
$$\Sigma A \Sigma A \mu = \Sigma A \mu$$

and

$$\lambda = \mu' A \Sigma A \mu = \mu' A \mu .$$

<u>Corollary</u>. *(Styan (1970))*

A necessary and sufficient condition for Q(X) to follow a noncentral-χ^2 distribution is

$$A \Sigma A \Sigma = A \Sigma ,$$
$$\lambda = \mu' A \Sigma A \mu = \mu' A \mu$$

if and only if $rank(A\Sigma) = tr(A\Sigma) = r$ *or*

$$rank(A\Sigma) = rank(\Sigma A\Sigma) = r.$$

The applications of quadratic forms have been summarized in Johnson and Kotz (1970c), pages 180-183, and under the conditions stated above we could apply the approximations to the noncentral-χ^2. The five applications summarized are

(1) Weapons and Optimal Control Problems,

(2) Estimation of Power Spectra,

(3) Distribution of Quadratic forms in Variate Differences,

(4) χ^2-tests, and

(5) Analysis of Variance.

§5. The probability of hitting a target

The probability that a random point $(X_1, X_2, \ldots, X_\nu)$, with the X's mutually independent normal variables, each having expected value 0 and variance σ^2, falls within an *offset hypersphere* $\sum_{i=1}^{\nu} (X_i-\mu_i)^2 \leq R^2$ is

$$\Pr[\chi'^2_\nu(\lambda) \leq (R/\sigma)^2] ,$$

where

$$\lambda = \frac{1}{\sigma^2} \sum_{i=1}^{\nu} \mu_i^2 .$$

(see Johnson and Kotz (1970c), page 144).

When $\nu = 2$, we have the case where we wish to find the probability that a random point (x,y) falls within the offset circle

$$(x-\mu_1)^2 + (y-\mu_2)^2 \leq R^2 .$$

Example:

If $\mu_2 = 0$, then values of the probability that a random point (x,y) falls *outside* an offset circle of radius R can be read from Owen (1962), Table 8.2.

Suppose we have $\mu_1 = 2\sigma$, i.e., $\mu_1/\sigma = 2$. Then

$$\frac{R}{\sigma} - \frac{\mu_1}{\sigma} = \frac{R}{\sigma} - 2 \quad \text{and} \quad \lambda = \frac{\mu_1^2 + \mu_2^2}{\sigma^2} = \frac{\mu_1^2}{\sigma^2} = 4.$$

Let $(R-\mu_1)/\sigma = 1$.

Therefore, $R/\sigma = 3$.

From Owen's (1962) tables we find that, for $\mu_1/\sigma = 2$ and $(R-\mu_1)/\sigma = 1$, the probability that a random point falls *outside* an offset circle of radius $R = 3\sigma$ is 0,214.

Therefore, the probability that the random point falls *within* the offset circle is

$$1 - 0{,}214 = 0{,}786.$$

Various approximations yield the following results.

Approximation	(1.2)	(3.2)	(4.1)	(5.4)	(8.2)	(10.2)
Probability	0,787	0,783	0,797	0,786	0,797	0,789

These approximations were used because of their ease of application in finding probabilities.

§6. Approximations to the noncentral distributions of test criteria in Multivariate Analysis of Variance

Multivariate Analysis of Variance is concerned with the problem of testing the equality of mean vectors of k p-variate normal populations with a common covariance matrix.

Various criteria for testing $H_o : \underset{p\times m}{\mu} = 0$ have been proposed, all of which are functions of the matrix

$$\underset{p\times p}{L} = [(A+B)^{-\frac{1}{2}}]'B[(A+B)^{-\frac{1}{2}}]$$

where A is distributed as a central Wishart and B is distributed as a noncentral Wishart.

Let $\ell_1 \leqslant \ell_2 \leqslant \cdots\cdots \leqslant \ell_p$ be the characteristic roots of L.

The criteria are

(i) Wilk's Likelihood Ratio,

$$W = \frac{|A|}{|A+B|} = |I-L| = \prod_{i=1}^{p} (1-\ell_i),$$

(ii) Hotelling-Lawly T_o^2 criterion,

$$T_o^2 = NtrBA^{-1} = NtrL(I-L)^{-1} ,$$

(iii) Pillai's $V^{(p)}$ criterion,

$$V^{(p)} = trL = \sum_{i=1}^{p} \ell_i ,$$

(iv) Pillai's $Q^{(p)}$ criterion,

$$Q^{(p)} = tr(I-L)L^{-1} , \quad \text{and}$$

(v) Roy's largest root criterion,

$$\ell_p \equiv \text{the largest characteristic root of } L.$$

These criteria can also be used to test the independence between a q set and an r set of vectors in a sample from a single multivariate normal population.

The distributions of these statistics under an alternative hypothesis $H_1 : \mu \neq 0$, which are used for power calculations and also for comparisons between the criteria, are extremely complicated and approximations or asymptotic distributions are necessary to enable them to be used in practical applications.

A first approximation to the distribution functions of these criteria is the noncentral-χ^2 distribution. For greater accuracy more terms can be added which are distribution functions of the noncentral-χ^2 with increasing degrees of freedom. A review of these approximations and the references to them have been summarized in an unpublished technical report by Juritz (1973).

Chapter 3 : The Noncentral F distribution

I. Introduction

The noncentral F distribution is defined as the distribution of

$$F'_{\nu_1,\nu_2}(\lambda) = \frac{\chi'^2_{\nu_1}(\lambda)/\nu_1}{\chi^2_{\nu_2}/\nu_2} = F' ,$$

where $\chi'^2_{\nu_1}(\lambda)$ is a noncentral-χ^2 with ν_1 degrees of freedom and noncentrality parameter λ and $\chi^2_{\nu_2}$ is a central-χ^2 with ν_2 degrees of freedom. The probability density function is given by

$$p(f) = \sum_{i=o}^{\infty} \frac{\Gamma\left(\frac{2i+\nu_1+\nu_2}{2}\right)}{\Gamma\left(\frac{\nu_2}{2}\right)\Gamma\left(\frac{2i+\nu_1}{2}\right)} \frac{e^{-\frac{1}{2}\lambda}(\frac{1}{2}\lambda)^i}{i!} \frac{\left(\frac{\nu_1}{\nu_2}\right)^{\frac{1}{2}(2i+\nu_1)} f^{\frac{1}{2}(2i+\nu_1-2)}}{\left(1+\frac{\nu_1}{\nu_2}f\right)^{\frac{1}{2}(2i+\nu_1+\nu_2)}}$$

$$o \leq f < \infty .$$

II. Approximations

§1. Patnaik

Patnaik (1949) follows the procedure he adopted in the case of the noncentral-χ^2 approximation (1.1) (see Chapter 2). He regards F'/k as following a central F distribution with ν and ν_2 degrees of freedom, i.e.,

$$F'_{\nu_1,\nu_2} \overset{\bullet}{\sim} kF_{\nu,\nu_2} \tag{1.1}$$

and then equates the expressions for the first two moments of

F_{ν_1,ν_2} and kF_{ν,ν_2} to obtain

$$k = \frac{\nu_1+\lambda}{\nu_1} \quad \text{and} \quad \nu = \frac{(\nu_1+\lambda)^2}{\nu_1+2\lambda} \; .$$

Using this approximation, the probability integral $\int_0^{F'} p(f)df$ is approximately equal to $\int_0^{F'/k} p(g)dg$, where $p(g)$ is the probability density function of the central F with ν and ν_2 degrees of freedom. This can be expressed in the form on an Incomplete B-function, $I_x\left(\frac{\nu}{2}, \frac{\nu_2}{2}\right)$,

where $$x = \frac{\nu F'/k}{\nu_2 + \nu F'/k} \; .$$

§2. Johnson

Let $v = \chi^2_{\nu_2}$. Then, using equation (3.1) of Chapter 2, the conditional probability that F' does not exceed $\chi'^2_{\nu_1}/\nu_1$, given the value of v, is

$$\begin{aligned} \Pr\{F' < X/\nu_1 \mid v\} &= \Pr\{x < Xv/\nu_2\} \\ &= \Pr\{\xi_v - \ell \geq \tfrac{1}{2}\nu_1\} \qquad (\nu \text{ even}), \end{aligned} \tag{2.1}$$

where ξ_v is a Poisson variable independent of ℓ, with expected value $\frac{1}{2}Xv/\nu_2$, ℓ is a Poisson variable with expected value $\frac{1}{2}\lambda$ and $x = \sum_{i=1}^{\nu} (u_i + \delta_i)^2 \sim \chi'^2_\nu(\lambda)$, where $u_i \sim N(0,1)$ and $\lambda = \sum_{i=1}^{\nu} \delta_i^2$. Averaging (2.1) over the distribution of v, we find the unconditional probability,

$$\Pr\{F' < X/\nu_1\} = \Pr\{\xi' - \ell > \tfrac{1}{2}\nu_1\}, \tag{2.2}$$

where ξ' is a negative binomial random variable, independent of ℓ, with expected value $\frac{1}{2}X$ and variance $\frac{1}{2}X(1 + X/\nu_2)$.

Therefore, the expected value of $\xi'-\ell$ is $\frac{1}{2}(X-\lambda)$ and variance of $\xi'-\ell$ is $\frac{1}{2}(X + \lambda + X^2/\nu_2)$. Applying a normal approximation to the right-hand side of (2.2), Johnson (1959) obtains the approximation,

$$\Pr\{F' < X/\nu_1\} \doteqdot \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{U} e^{-\frac{1}{2}x^2}\, dx \quad , \tag{2.3}$$

where $U = (X - \lambda - \nu_1 + 1)/\sqrt{2(X + \lambda + X^2/\nu_2)} \sim N(0,1)$.

Johnson (1959) also obtains another approximation to the probability integral of the noncentral F based on a mixture of central F's. The approximation is

$$\Pr\{F'_{\nu_1,\nu_2} < X/\nu_1\} \doteqdot \tfrac{1}{2}[1 - (2\lambda+1)^{-\frac{1}{2}}]\Pr\{F_{\nu_0,\nu_2} < X/\nu_0\}$$
$$+ \tfrac{1}{2}[1 + (2\lambda+1)^{-\frac{1}{2}}]\Pr\{F_{\nu,\nu_2} < X/\nu\} \; , \tag{2.4}$$

where $\nu_0 = \nu_1 + \lambda + 1 + \sqrt{2\lambda+1}$

and $\nu = \nu_1 + \lambda + 1 - \sqrt{2\lambda+1}$.

§3. Severo and Zelen

We had, in §1, a two-moment central F approximation to the noncentral F, namely,

$$F'_{\nu_1,\nu_2} \doteqdot kF_{\nu,\nu_2} \; ,$$

where

$$k = \frac{\nu_1 + \lambda}{\nu_1} \quad \text{and} \quad \nu = \frac{(\nu_1 + \lambda)^2}{\nu_1 + 2\lambda} \; .$$

Paulson suggested an approximation to F_{ν,ν_2} which is based on the Wilson-Hilferty approximation to the distribution of a central-χ^2 (see Johnson and Kotz (1970c), page 83), namely,

$$F^{\frac{1}{3}}_{\nu,\nu_2} \doteqdot \frac{1 - \frac{2}{9\nu} + U_1\sqrt{\frac{2}{9\nu}}}{1 - \frac{2}{9\nu_2} + U_2\sqrt{\frac{2}{9\nu_2}}} , \qquad (3.1)$$

where U_1 and U_2 are independent unit normal variables. Using approximation (3.1) above, Severo and Zelen (1960) were led to suggest the approximation

$$\frac{\left(1 - \frac{2}{9\nu_2}\right)\left(\frac{\nu_1 F'}{\nu_1+\lambda}\right)^{\frac{1}{3}} - \left(1 - \frac{2}{9}\frac{(\nu_1+2\lambda)}{(\nu_1+\lambda)^2}\right)}{\left(\frac{2}{9}\frac{(\nu_1+2\lambda)}{(\nu_1+\lambda)^2} + \frac{2}{9\nu_2}\left(\frac{\nu_1 F'}{\nu_1+\lambda}\right)^{\frac{2}{3}}\right)^{\frac{1}{2}}} \;\dot{\sim}\; N(0,1). \qquad (3.2)$$

§4. Barton, David and O'Neill

Barton *et al.* (1960) considered the variate

$$Z^* = \tfrac{1}{2}\log_e F' ,$$

which is analogous to the Fisher statistic

$$Z = \tfrac{1}{2}\log_e F .$$

They obtained two different series expansions of the distribution of Z^*. Their series I represents the amount by which $\Pr\{Z < z_\alpha\}$ is exceeded in the Z^* distribution and their series II is the sum of the first few terms of an Edgeworth expansion of the distribution of a standardized variate

$$x^* = \frac{2Z - \kappa_1(2Z)}{\sqrt{\kappa_2(2Z)}} , \text{ where } \kappa_1(2Z) \text{ and } \kappa_2(2Z)$$

are the first and second cumulants of $2Z = \log_e F$.

§5. Tiku

(i) Using series (7.2) of §7, Chapter 2, Tiku (1965) obtains an approximation to the probability integral of the noncentral F distribution as

$$P(F_o) \doteq I_{x_o}\left(\frac{\nu_2}{2}, \frac{\nu}{2}\right) + \sum_{r=3}^{\infty} P_r(x_o) , \qquad (5.1)$$

where

$$\nu = \frac{(\nu_1+\lambda)^2}{(\nu_1+2\lambda)} , \quad x_o = 1/\left[1 + \frac{\nu_1(\nu_1+\lambda)}{\nu_2(\nu_1+2\lambda)} F_o\right] ,$$

$$P(F_o) = \int_{F_o}^{\infty} p(f)df$$

and

$$P_r(x_o) = \frac{(-1)^r}{r!} \beta'_r (\Delta^{r-1}T_1)\{x_o^{\frac{1}{2}\nu_2} (1 - x_o)^{\frac{1}{2}\nu}/B(\tfrac{1}{2}\nu_2, \tfrac{1}{2}\nu)\}$$

$$(r = 3, 4, \cdots),$$

where the β'_r are given by (7.2a) of §7, Chapter 2 and the T_r are given by

$$T_1 = 1, \; T_2 = T_1 \frac{(\nu_1+\nu_2)}{(\nu_1+2)} (1 - u_o), \; T_3 = T_2 \frac{(\nu_1+\nu_2+2)}{(\nu_1+4)} (1-u_o), \cdots$$

and $u_o = 1/[1 + \frac{\nu_1}{\nu_2} F_o]$.

(ii) Tiku (1965) also gives a three-moment central F approximation, namely,

$$P(F_o) \doteq I_{y_o}\left(\frac{\nu_2}{2}, \frac{b}{2}\right), \text{ where } y_o = 1\Big/\left\{1 + \frac{b}{\nu_2}\,\frac{(F_o+c)}{h}\right\} \tag{5.2}$$

by letting $F'_{\nu_1,\nu_2}(\lambda) \doteq hF_{b,\nu_2} - c$.

Equating the first three moments of $F'_{\nu_1,\nu_2}(\lambda)$ and $hF_{b,\nu_2} - c$ he obtains

$$b = \tfrac{1}{2}(\nu_2 - 2)\left\{\sqrt{\frac{E}{E-4}} - 1\right\},$$

$$h = \left(\frac{b}{\nu_1}\right)\frac{1}{(2b + \nu_2 - 2)}\,\frac{H}{K}$$

and

$$c = \frac{\nu_2}{(\nu_2 - 2)}\left\{h - \frac{\nu_1+\lambda}{\nu_1}\right\},$$

where

$$H = 2(\nu_1+\lambda)^3+3(\nu_1+\lambda)(\nu_1+2\lambda)(\nu_2-2) + (\nu_1+3\lambda)(\nu_2-2)^2,$$

$$K = (\nu_1+\lambda)^2 + (\nu_2-2)(\nu_1+2\lambda),$$

and

$$E = H^2/K^3 .$$

§6. Pearson

Pearson (1960) obtained an approximation to the probability integral of the noncentral F distribution based on Johnson's S_u curve (see Johnson and Kotz (1970b), page 22),

$$P(F_o) = \Pr(F' > F_o) \doteq \int_{Z_o}^{\infty} \frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}Z^2}\, dZ, \tag{6.1}$$

where

$$Z_o = \gamma + \delta \sinh^{-1}\{(F_o - \xi)/\lambda\}$$

and γ, δ , ξ and λ are determined by equating the first four moments of $\frac{1}{2}\log_e F'$ and $\gamma + \delta \sinh^{-1}\{(F' - \xi)/\lambda\}$.

§7. Another Approximation

Fisher's approximation to the central F is

$$Z_{\nu,\nu_2} = \tfrac{1}{2}\log_e F_{\nu,\nu_2} \overset{\bullet}{\sim} N\left(\tfrac{1}{2}\left(\frac{1}{\nu_2} - \frac{1}{\nu}\right), \tfrac{1}{2}\left(\frac{1}{\nu} + \frac{1}{\nu_2}\right)\right) \tag{7.1a}$$

for ν and ν_2 large (see Johnson and Kotz (1970c), page 81).

Patnaik's two-moment central F approximation (1.1) is

$$F'_{\nu_1,\nu_2} \doteqdot kF_{\nu,\nu_2}$$

where $k = \dfrac{\nu_1 + \lambda}{\nu_1}$ and $\nu = \dfrac{(\nu_1 + \lambda)^2}{\nu_1 + 2\lambda}$.

Therefore,

$$\tfrac{1}{2}\log_e F'_{\nu_1,\nu_2} \doteqdot \tfrac{1}{2}\log_e k + \tfrac{1}{2}\log_e F_{\nu,\nu_2}$$

and approximating $\frac{1}{2}\log_e F_{\nu,\nu_2}$ by (7.1a) we obtain

$$\tfrac{1}{2}\log_e F'_{\nu_1,\nu_2} \overset{\bullet}{\sim} N\left(\tfrac{1}{2}\log_e k + \tfrac{1}{2}\left(\frac{1}{\nu_2} - \frac{1}{\nu}\right), \tfrac{1}{2}\left(\frac{1}{\nu} + \frac{1}{\nu_2}\right)\right). \tag{7.1b}$$

Fisher suggested that replacement of $\frac{1}{\nu}$, $\frac{1}{\nu_2}$ by $\frac{1}{\nu-1}$, $\frac{1}{\nu_2-1}$ might improve accuracy (see Johnson and Kotz (1970c), page 82) so that we now have

$$\tfrac{1}{2}\log_e F'_{\nu_1,\nu_2} \overset{\bullet}{\sim} N\left(\tfrac{1}{2}\log_e k + \tfrac{1}{2}\left(\frac{1}{\nu_2-1} - \frac{1}{\nu-1}\right), \tfrac{1}{2}\left(\frac{1}{\nu-1} + \frac{1}{\nu_2-1}\right)\right). \tag{7.2}$$

§8. A Two-moment Central F Approximation

It would be convenient if we could approximate the non-central F by a central F where the degrees of freedom, ν_1 and ν_2 , remain unchanged. We, therefore, make the approximation

$$F'_{\nu_1,\nu_2}(\lambda) \doteqdot cF_{\nu_1,\nu_2} + b \,, \tag{8.1}$$

where c and b are obtained by equating the first two central moments of $F'_{\nu_1,\nu_2}(\lambda)$ with the first two central moments of $cF_{\nu_1,\nu_2} + b$. The two equations will be

$$\frac{\nu_2}{\nu_1} \; \frac{(\nu_1 + \lambda)}{(\nu_2 - 2)} = c\,\frac{\nu_2}{\nu_2-2} + b \qquad (\nu_2 > 2)$$

and

$$2\left(\frac{\nu_2}{\nu_1}\right)^2 \frac{(\nu_1+\lambda)^2 + (\nu_1+2\lambda)(\nu_2-2)}{(\nu_2-2)^2(\nu_2-4)}$$
$$= 2c^2\,\frac{\nu_2^2}{\nu_1}\;\frac{(\nu_1 + \nu_2 - 2)}{(\nu_2-2)^2(\nu_2-4)} \qquad (\nu_2 > 4)$$

from which we obtain

$$c = +\sqrt{\frac{(\nu_1+\lambda)^2 + (\nu_1+2\lambda)(\nu_2-2)}{\nu_1(\nu_1+\nu_2-2)}}$$

and

$$b = \frac{\nu_2}{\nu_2 - 2}\left(\frac{\nu_1+\lambda}{\nu_1} - c\right) .$$

§9. Khatri

Khatri (1966) proposed the approximation

$$F'_{\nu_1,\nu_2}(\lambda) \doteqdot \frac{\nu_2}{\nu_1}\,(1 - \bar{R}^2)\;\omega\,\frac{R^2}{1-R^2} \,, \tag{9.1}$$

where

$$\lambda = \tfrac{1}{2}[\nu_2(\nu_1+\nu_2)]^{\frac{1}{2}}\bar{R}^2\omega$$

$$\omega = [\nu_1+\{\nu_2+\sqrt{\nu_2(\nu_1+\nu_2)}\}\bar{R}^2]/[\nu_1+\nu_2(2-\bar{R}^2)\bar{R}^2]$$

$R \equiv$ sample multiple correlation coefficient

$\bar{R} \equiv$ population multiple correlation coefficient

(see Chapter 5, section II, §1).

III. Comparisons of Accuracy

In Table I, all values except those for approximations (2.3) and (7.2) were taken from Tiku (1966). The first of the values for each λ is for $\alpha = 0,01$ and the second for $\alpha = 0,05$. The asterisks denote values that are unobtainable using Pearson's approximation (6.1).

It can be seen from Table I that Tiku's two approximations (5.1) and (5.2) are the most accurate. Calculations, however, are laborious.

Patnaik's (1.1) and Severo and Zelen's (3.2) approximations seem to be of equal accuracy. Severo and Zelen's (3.2) approximation, involving the use of standard normal tables, is quite easy to use for power calculations and percentage points. However, Patnaik's approximation (1.1) uses interpolation in tables of the F distribution due to its noninteger degrees of freedom.

Johnson's (2.3) approximation is easy to apply but is not very accurate compared to the others. Pearson's (6.1) approximation involves so much labour in evaluating its parameters that it can perhaps be ruled out.

Approximation (7.2) improves as both ν_1 and ν_2 get large and is as easy to apply as Severo and Zelen's (3.2) approximation in the case of power calculations and percentage points. However, if one wishes to obtain the value of λ (or ν_1), given ν_1 (or λ), that will yield a specified power, then Johnson's approximation (2.3) is, by far, the easiest to use and yields quite reasonable results.

Barton *et al.* approximations mentioned in §4, section II of this chapter makes use of tables which appear in Barton *et al.* (1960). This makes the use of the approximations quite involved and, consequently, no values for β appear in Table I for these approximations.

Approximation (8.1), values of which do not appear in the Table I, seems to be of value only for percentage points.

Khatri's approximation (9.1) is difficult to apply because the noncentral F distribution is complicated and has not been well tabulated.

Table I. Values of $\beta = 1 - \int_{F_\alpha}^{\infty} p(f)df$ for α = 0,01 and 0,05

ν_1	ν_2	λ	Exact	Patnaik (1.1)	Johnson (2.3)	Severo and Zelen (3.2)	Tiku (5.1)	Tiku (5.2)	Pearson (6.1)	(7.2)
3	12	4	0,909	0,910	0,886	0,911	0,909	0,910	*	0,903
		4	0,731	0,736	0,744	0,737	0,731	0,731	*	0,752
		16	0,463	0,466	0,495	0,466	0,463	0,462	0,385	0,471
		16	0,178	0,173	0,184	0,173	0,179	0,177	0,154	0,179
3	20	4	0,887	0,889	0,873	0,889	0,886	0,888	*	0,883
		4	0,700	0,706	0,710	0,707	0,699	0,700	*	0,735
		16	0,347	0,350	0,364	0,349	0,348	0,346	0,341	0,366
		16	0,126	0,119	0,129	0,119	0,128	0,125	0,121	0,122
15	12	4	0,975	0,975	0,941	0,974	0,975	0,975	0,975	0,978
		4	0,895	0,896	0,874	0,896	0,895	0,895	0,901	0,894
		16	0,881	0,881	0,863	0,881	0,881	0,881	0,887	0,879
		16	0,655	0,656	0,686	0,657	0,655	0,655	0,655	0,645
15	20	4	0,969	0,969	0,944	0,969	0,969	0,969	0,969	0,968
		4	0,879	0,879	0,866	0,879	0,879	0,879	0,880	0,876
		16	0,812	0,812	0,815	0,813	0,812	0,812	0,814	0,811
		16	0,554	0,556	0,583	0,556	0,554	0,554	0,554	0,554

IV. Applications

§1. Power functions in Analysis of Variance Tests

The noncentral F distribution is used in the calculation of the power functions of tests of general linear hypotheses. These include standard tests used in the analysis of variance. The general linear model may be formulated in the following way.

If y is a linear function of the independent variables $x_2, x_3, \cdots, x_k$, plus error, then the ith observation y_i can be written as

$$y_i = \beta_1 + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} + e_i \qquad (i = 1,2,\cdots,n), \tag{1.1}$$

where $\beta_1, \beta_2, \cdots, \beta_k$ are unknown constants estimated from a sample of size n. In matrix and vector notation (1.1) becomes

$$Y = X\beta + e$$

or

$$\begin{pmatrix} y_1 \\ y_2 \\ \cdot \\ \cdot \\ \cdot \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & x_{12} & \cdots & x_{1k} \\ 1 & x_{22} & \cdots & x_{2k} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ 1 & x_{n2} & \cdots & x_{nk} \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \cdot \\ \cdot \\ \cdot \\ \beta_k \end{pmatrix} + \begin{pmatrix} e_1 \\ e_2 \\ \cdot \\ \cdot \\ \cdot \\ e_n \end{pmatrix} .$$

We make the assumption that the e_i $(i = 1,2,\cdots,n)$ are independent and $e \sim N(0,\sigma^2 I)$, where I is the identity matrix.

The unbiased maximum likelihood estimates of β and σ^2 can be found to be

$$\hat{\beta} = (X'X)^{-1}X'Y$$

and
$$S^2 = \frac{1}{n-k} (Y-X\hat{\beta})'(Y-X\hat{\beta}).$$

If we wish to test the hypothesis

$$H_o : \beta = \beta^*$$

then it can be shown that the ratio

$$\frac{(n-k)}{k} \frac{(\hat{\beta}-\beta)'X'X(\hat{\beta}-\beta)}{(Y-X\hat{\beta})'(Y-X\hat{\beta})} = \frac{(\hat{\beta}-\beta)'X'X(\hat{\beta}-\beta)}{kS^2} \tag{1.2}$$

has a central F distribution with degrees of freedom k and n-k.

However, under an alternative hypothesis the ratio (1.2) has a noncentral F distribution with k and n-k degrees of freedom and noncentrality parameter λ. Therefore, from this we obtain the F-test of the analysis of variance and the power function of this test with respect to an alternative hypothesis as an F'-integral.

We shall now consider the question of evaluating the power of the analysis of variance test for the one-way classification. The set-up may be written as

$$y_{ij} = \mu + \alpha_i + e_{ij} \qquad (i = 1,2,\cdots,k; \quad j = 1,2,\cdots,n)$$

(i.e. there are k groups and n observations in each group), where μ is the general mean,

α_i is the main effect due to the ith group $(i = 1,\cdots,k)$, and the $\{e_{ij}\}$ are distributed independently as $N(0,\sigma^2)$.

We make the assumption that $\sum_{i=1}^{k} \alpha_i = 0.$

The null hypothesis is

$$H_o \ : \ \alpha_1 = \alpha_2 = \cdots\cdot = \alpha_n = 0. \qquad (1.3)$$

The ANOVA (Analysis of Variance) Table is shown below.

Note: The following notation is used here and later on in this section.

$$y_{i\cdot} = \frac{1}{n} \sum_j y_{ij} \ , \quad y_{\cdot j} = \frac{1}{k} \sum_i y_{ij} \ , \quad y_{\cdot\cdot} = \frac{1}{nk} \sum_i \sum_j y_{ij}$$

Source of Variation	Sums of Squares	Degrees of freedom	Mean Squares
Between groups	$SS_1 = \sum_i n(y_{i\cdot} - y_{\cdot\cdot})^2$	k-1	$MS_1 = SS_1/(k-1)$
Within groups	$SS_2 = \sum_{ij}\sum (y_{ij} - y_{i\cdot})^2$	k(n-1)	$MS_2 = SS_2/k(n-1)$

It is easy to show that MS_1/MS_2 is distributed as non-central F with k-1 and k(n-1) degrees of freedom and non-centrality parameter λ.

To find λ we use Rule 1 on page 39 of Scheffé (1959). The rule states that if we replace each observation in the sums of squares of the numerator of the F-statistic by its expected value then the result is $\lambda\sigma^2$. Therefore, we have $\sigma^2\lambda = n\sum_i [E(y_{i\cdot}) - E(y_{\cdot\cdot})]^2$. Now $E(y_{i\cdot}) = \mu + \alpha_i$ and

$E(y_{..}) = \mu$. Therefore, we have $\sigma^2\lambda = n\sum_i \alpha_i^2$ or $\lambda = (n\sum_i \alpha_i^2)/\sigma^2$.

We shall now define ϕ by $\phi = \sqrt{\frac{\lambda}{\nu_1+1}} = \sqrt{\frac{n\sum_i \alpha_i^2}{k\sigma^2}}$, where ν_1 is the degrees of freedom of the numerator of the F-statistic. Under H_o , MS_1/MS_2 is distributed as central F with (k-1) and k(n-1) degrees of freedom. Therefore, the test of hypothesis (1.3) is based on the critical region $MS_1/MS_2 > F_\alpha$, where F_α is the α-significance point of the F-distribution.

Let us consider an alternative hypothesis H_1 : α_i's not all zero.

The power function will be $Pr\{F'_{k-1,k(n-1)}(\lambda) > F_\alpha\}$ which depends only on $\phi = \sqrt{\lambda/k}$.

Two questions may be asked in connection with the test for differences between groups. They are

(a) What is the extent of departure from the null hypothesis that could be detected with a given chance? and

(b) How many observations are we to take in each group so that we could detect a given configuration of the α_i with a prescribed chance?

Example (a)

This example is based on an example given by Patnaik (1949).

Suppose 13 samples of 15 c.c. of water are taken from a pond on the first day of each of five months. We want to test whether there is significant variation in the density of a certain type of alga from month to month. We apply an analysis of variance test at the 5% level.

(i) Suppose we wish to know how large the ratio $\sum_i \alpha_i^2/k\sigma^2$ should be so that we could detect it with a 90% chance. We have, therefore,

$$\nu_1 = k-1 = 5-1 = 4 \text{ and } \nu_2 = k(n-1) = 5(12) = 60.$$

For this type of inverse problem, Patnaik has made graphical representation of the relationship between ν_1, ν_2, λ, $\alpha = 0{,}05$ and power $1-\beta = 0{,}90$ and $0{,}50$ using his approximation (1.1).

Using these charts we find $\lambda = 16{,}3$ and, therefore,

$$\frac{\sum_i \alpha_i^2}{k\sigma^2} = \frac{\lambda}{nk} = \frac{16{,}3}{13\times 4} = 0{,}313.$$

Johnson's (2.3) approximation is the only approximation that can be *conveniently* used to find λ. We want to find a λ such that

$$\Pr[F' > F_\alpha] = 1-\beta = 0{,}90 ,$$

where F_α is the α-percentage point of the central F distribution. We are testing at the 5% level so that $F_\alpha = F_{0,05} = 2{,}53$.

Johnson's approximation (2.3) is

$$\Pr[F' > X/\nu_1] \doteq \frac{1}{\sqrt{2\pi}} \int_U^\infty e^{-\frac{1}{2}x^2}\, dx ,$$

where

$$U = \frac{(X-\lambda-\nu_1+1)}{\sqrt{2(X+\lambda+X^2/\nu_2)}} \sim N(0,1). \qquad (1.4)$$

The appropriate value of U which yields

$$\Pr[F' > X/\nu_1] = 0{,}90 \text{ is } U_{0,90} = -1{,}28.$$

Now $X/\nu_1 = F_\alpha = 2{,}53$ and so $X = 10{,}12$. Solving for λ in (1.4) above we find

$$\lambda = (X-\nu_1+1+U_{0,90}^2) \pm U_{0,90}\sqrt{U_{0,90}^2 + 2(2X-\nu_1+1+X^2/\nu_2)}$$

which yields $\lambda = 16{,}81$ or $0{,}71$. Using the appropriate value of λ, i.e., $\lambda = 16{,}81$, we obtain

$$\frac{\sum_i \alpha_i^2}{k\sigma^2} = 0{,}323.$$ From (1.4) one can

see that when $U < 0$ then $X-\lambda-\nu_1+1 < 0$. This means that the appropriate value of λ is the λ that satisfies the inequality $\lambda > X-\nu_1+1$. In this example $X-\nu_1+1 = 7{,}12$.

The Pearson and Hartley (1951) charts of the power function of the analysis of variance tests yields

$$\lambda = 16{,}56 \quad \text{and} \quad \frac{\sum_i \alpha_i^2}{k\sigma^2} = 0{,}318.$$

In this case, Johnson's approximation is as accurate as Patnaik's.

(ii) Suppose we wish the power to be $1-\beta = 0{,}50$.

Patnaik's chart yields $\lambda = 7{,}0$ and $\dfrac{\sum_i \alpha_i^2}{k\sigma^2} = 0{,}135$.

Johnson's approximation (2.3) yields $\lambda = 7{,}12$ and $\dfrac{\sum_i \alpha_i^2}{k\sigma^2} = 0{,}137$.

The Pearson and Hartley charts yield $\lambda = 6{,}96$ and $\dfrac{\sum_i \alpha_i^2}{k\sigma^2}$ $= 0{,}134$.

This example shows that Johnson's approximation yields pretty accurate results.

<u>Example (b)</u>

Suppose we are given $\dfrac{\sum_i \alpha_i^2}{k\sigma^2}$, i.e., we do not know λ (or ϕ), and we wish to know how large n must be to have a chance $1-\beta$ of establishing significance. This is the minimum sample size problem which is described in more detail in Chapter 4, Section IV, §8.

(i) Suppose we test the hypothesis (1.3) above with $k = 4$ means and significance level $\alpha = 0{,}05$. We wish to find the minimum n such that the power is at least 0,95 when $\mu_1 = \mu_2 = \mu_3$ and $\mu_4 = \mu_1 + \sigma$, where $\mu_i = \mu + \alpha_i$ $(i = 1,2,3,4)$ are the means. Therefore, $\mu = \mu_1 + \sigma/4$, where μ is the general mean. We have, therefore,

$$\sum_{i=1}^{4} \alpha_i^2 = \sum_{i=1}^{4} (\mu_i-\mu)^2 = \frac{3\sigma^2}{4}$$

and $\phi = \sqrt{\dfrac{n}{4\sigma^2}\left(\dfrac{3\sigma^2}{4}\right)} = \sqrt{\dfrac{3n}{16}}$ and, so, $n = \dfrac{16}{3}\phi^2$.

With $\nu_1 = k-1 = 3$ and $\nu_2 = k(n-1) = \infty$ $(n = \infty)$ the Pearson and Hartley graphs show that the power is 0,95 if $\phi = 2{,}07$ and, therefore

$$n = (16/3)(2{,}07)^2 = 22{,}85.$$

With $\nu_2 = 60$ $(n = 16)$, the power is 0,95 if $\phi = 2{,}14$ and, so,

$$n = 24{,}42.$$

The minimum sample size would, therefore, lie between 22,85 and 24,42 and 23 or 24 would be satisfactory solutions.

Using Johnson's approximation (2.3) we find that for $\nu_2 = \infty$, n = 23,32 and for $\nu_2 = 60$, n = 24,78. The actual minimum can be found to be 24 and, so, Johnson's approximation, in this case, yields the correct solution without any further calculations being necessary.

(ii) Suppose we want to test the hypothesis (1.3) with k = 9, $\alpha = 0{,}05$, $1-\beta = 0{,}95$ and $\mu_1 = \mu_2 = \cdots = \mu_8$ and $\mu_9 = \mu_1 + \sigma$. We have

$$n = \frac{729}{89}\phi^2 .$$

With $\nu_1 = k-1 = 8$ and $\nu_2 = \infty$ $(n = \infty)$, the Pearson and Hartley graphs yield n = 20,71 and for $\nu_2 = 60$ (n = 7,66), n = 23,39. Johnson's approximation (2.3) yields

$$n = 20{,}88 \qquad \text{for} \qquad \nu_2 = \infty$$

and

$$n = 23{,}48 \qquad \text{for} \qquad \nu_1 = 60$$

which will again give the correct solution.

Notice how the accuracy of Johnson's approximation improved noticeably as ν_1 increased from 3 to 8 in this example.

We shall now find the noncentrality parameter for certain other special cases of the analysis of variance.

(i) <u>The two-way classification with one observation per cell</u>

The model is

$$y_{ij} = \mu + \alpha_i + \beta_j + e_{ij} \qquad (i=1,\cdots,h;\ j=1,\cdots,k),$$

where μ is the general mean,

α_i is the effect due to the ith level of factor A,

β_j is the effect due to the jth level of factor B and the e_{ij} are independent and each distributed as $N(0,\sigma^2)$. We shall assume

$$\sum_i \alpha_i = 0 = \sum_j \beta_j .$$

If we wish to test whether an α-effect exists, then the null hypothesis is

$$H_o : \alpha_i = 0 \qquad (i = 1,\cdots,h).$$

The ANOVA table is

Source of Variation	Sums of Squares	Degrees of Freedom	Mean Squares
α-effects	$SS_A = k\sum_i (y_{i\cdot}-y_{\cdot\cdot})^2$	h-1	$SS_A/(h-1)=MS_A$
β- effects	$SS_B = h\sum_j (y_{\cdot j}-y_{\cdot\cdot})^2$	k-1	$SS_B/(k-1)=MS_B$
Residual	$SS_e = \sum_{ij}\sum (y_{ij}-y_{i\cdot}-y_{\cdot j}+y_{\cdot\cdot})^2$	(h-1)(k-1)	$SS_e/(h-1)(k-1)=MS_e$

If H_o is not true, then $MS_A/MS_e \sim F'_{h-1,(h-1)(k-1)}(\lambda)$. Using Rule 1 on page 39 of Scheffé (1959) we find

$$\lambda = \frac{k\sum_i \alpha_i^2}{\sigma^2} \quad \text{or} \quad \phi = \sqrt{\frac{\lambda}{h}} = \sqrt{\frac{k\sum_i \alpha_i^2}{h\sigma^2}} .$$

(ii) <u>The two-way classification with n observations per cell</u>

The model is

$$y_{ij\ell} = \mu + \alpha_i + \beta_j + \gamma_{ij} + e_{ij\ell}$$

$$(i=1,\cdots h;\ j=1,\cdots,k;\ \ell=1,\cdots,n),$$

where μ,α_i and β_j are as previously defined,

γ_{ij} is the interaction between the ith level of factor A and the jth level of factor B and the $e_{ij\ell}$ are independent and each distributed as $N(0,\sigma^2)$.

We shall assume

$$\sum_i \alpha_i = \sum_j \beta_j = \sum_i \gamma_{ij} = \sum_j \gamma_{ij} = 0$$

$$(i=1,\cdots,h;\quad j=1,\cdots,k)\ .$$

The ANOVA table is

Source of Variation	Sums of Squares	Degrees of Freedom	Mean Squares
α main effects	$SS_A = kn \sum_i (y_{i..}-y_{...})^2$	h-1	$MS_A = SS_A/(h-1)$
β main effects	$SS_B = hn \sum_j (y_{.j.}-y_{...})^2$	k-1	$MS_B = SS_B/(k-1)$
γ interactions	$SS_{AB} = n\sum_i\sum_j (y_{ij.}-y_{i..}-y_{.j.}+y_{...})^2$	(h-1)(k-1)	$MS_{AB} = SS_{AB}/(h-1)(k-1)$
Error	$SS_e = \sum\sum\sum_{ij\ell} (y_{ij\ell}-y_{ij.})^2$	hk(n-1)	$MS_e = SS_e/hk(n-1)$

If we test for the presence of a main effect, say

$$H_o \; : \; \alpha_i \; = \; 0 \qquad (i = 1,\cdots,h)$$

then $MS_A/MS_e \sim F'_{h-1,hk(n-1)}(\lambda)$

and the noncentrality is

$$\phi = \sqrt{\frac{kn \sum_i \alpha_i^2}{h\sigma^2}} = \sqrt{\frac{\lambda}{h}} \; .$$

If we test for the presence of interaction, then the null hypothesis is

$$H_o \; : \; \gamma_{ij} \; = \; 0 \qquad (\forall i,j).$$

If H_o is not true, then

$$MS_{AB}/MS_e \sim F'_{(h-1)(k-1),hk(n-1)}(\lambda) \; .$$

The noncentrality is

$$\phi = \sqrt{\frac{n \sum_i\sum_j \gamma_{ij}^2}{\sigma^2[(h-1)(k-1)+1]}} = \sqrt{\frac{\lambda}{(h-1)(k-1)+1}} \; .$$

(iii) The k×k Latin Square

The model is

$$y_{ij\ell} = \mu + \alpha_i + \beta_j + \gamma_\ell + e_{ij\ell}$$

$$(i=1,\cdots,k; \; j=1,\cdots,k; \; \ell=1,\cdots,k).$$

Let α_i and β_j represent the row and column terms and γ_ℓ the possible treatment effects.

We assume

$$\sum_i \alpha_i = \sum_i \beta_j = \sum_\ell \gamma_\ell = 0.$$

The ANOVA table is:

Source of Variation	Sums of Squares	Degrees of Freedom	Mean Squares
α	$SS_A = k\sum_i (y_{i..}-y_{...})^2$	k-1	$MS_A = SS_A/(k-1)$
β	$SS_B = k\sum_j (y_{.j.}-y_{...})^2$	k-1	$MS_B = SS_B/(k-1)$
γ	$SS_C = k\sum_\ell (y_{..\ell}-y_{...})^2$	k-1	$MS_C = SS_C/(k-1)$
Residual	$SS_e = \sum_{ij}\sum (y_{ij\ell}-y_{i..}-y_{.j.}-y_{..\ell}+2y_{..})^2$	(k-1)(k-2)	$MS_e=SS_e/(k-1)(k-2)$

Let the null hypothesis be $H_o : \gamma_\ell = 0 \quad (\ell=1,2,\cdots,k)$.

If H_o is not true, then

$$MS_C/MS_e \sim F'_{k-1,(k-1)(k-2)}(\lambda),$$

where the noncentrality is given by

$$\phi = \sqrt{\frac{k\sum_\ell \gamma_\ell^2}{k\sigma^2}} = \sqrt{\frac{\lambda}{k}}$$

and $\sigma^2 = \text{var}(e_{ij\ell})$ $(\forall\ i,j,\ell)$.

Example:

Let us suppose we have a certain product. There are 3 assistants regularly employed on the testing of this product and 3 labourers who are used for mixing the test samples. To investigate the possible effects of personal factors among the men, we can plan a simple 3×3 experiment in which n samples are tested for each of the 9 combinations of tester and mixer. Let α_i (i = 1,2,3) represent the mixer effects and β_j (j = 1,2,3) represent the tester effects. Using the model (ii) above, in testing for the presence of either of these effects, we have

$$\nu_1 = 2 \text{ and } \nu_2 = 9(n-1).$$

Let us take n = 4 ($\nu_2 = 27$).

(a) Using the Pearson and Hartley (1951) charts, the probability will be at least 0,90 of establishing significance for a mixer (or tester) effect at the 5% level if

$$\phi \gtrsim 2{,}16.$$

Johnson's approximation (2.3) yields

$$\phi \gtrsim 2{,}19.$$

(b) At the 1% level, we must have

$$\phi \gtrsim 2{,}63.$$

Johnson's approximation (2.3) yields

$$\phi \gtrsim 2{,}64.$$

(c) If, however, the noncentrality is smaller, say

$$\phi = 1{,}73,$$

there would be a probability of only

(i) 0,72 at the 5% level, and

(ii) 0,45 at the 1% level

of establishing significance.

Other approximations yield

(i) 0,711 using (2.3),

0,723 using (3.2),

0,686 using (7.2)

at the 5% level, and

(ii) 0,443 using (2.3)

0,447 using (3.2),

0,408 using (7.2)

at the 1% level.

These three approximations were used because of the ease with which the probabilities could be found.

2. Hotelling's T^2-statistic

(i) Hotelling's T^2-statistic is the multivariate analogue of the square of the Student t-statistic and is given by

$$T^2 = N(\bar{x}-\mu)'S^{-1}(\bar{x}-\mu),$$

where

$\bar{x}$ is the mean vector of a sample of size N,

$\mu' = (\mu_1, \mu_2, \cdots, \mu_p)$ is the mean vector of the population

and

S is the sample covariance matrix.

The statistic T^2 is used for testing hypotheses about the mean vector μ, e.g. $H_o : \mu = \mu_o$.

Under an alternative hypothesis

$$\frac{T^2}{(N-1)} \frac{(N-p)}{p} = \frac{N(\bar{x}-\mu_o)' S^{-1} (\bar{x}-\mu_o)(N-p)}{(N-1)\ p}$$ has the

noncentral F distribution with p and N-p degrees of freedom and noncentrality parameter

$$\lambda = N(\mu-\mu_o)' \Sigma^{-1} (\mu-\mu_o).$$

Example:

Suppose we wish to find the probability of rejecting $H_o : \mu = 0$ at level of significance α when $p = 3$, $N = 23$ and $\phi = 2$. We have, therefore, $N-p = 20$ and $\lambda = (p+1)\ \phi^2 = 16$.

Using the Pearson and Hartley (1951) charts we find that the probability of rejecting H_o at level $\alpha = 0{,}01$ is $1-\beta = 0{,}66$. The exact value is $1-\beta = 0{,}653$ (see Table 1, section III of this chapter) which is found by subtracting the exact value $\beta = 0{,}347$ from 1. The values of the approximations can be found from the same table by subtracting the values from 1. For example, approximation (2.3) yields the value $1 - 0{,}364 = 0{,}636$ and approximation (3.2) yields the value $1 - 0{,}349 = 0{,}651$.

The values for significance level $\alpha = 0{,}05$ can also be read from the Table 1, section III of this chapter in the same way. The value taken from the Pearson and Hartley (1951) charts is 0,876. The exact value is 0,874. Johnson's approximation (2.3) yields the value 1 - 0,129 = 0,871.

(ii) Let $\{x_\alpha^{(i)}\}$ $(\alpha = 1,\cdots, N_i;\ i = 1,\cdots,q)$ be samples from $N(\mu^{(i)}, \Sigma_i)$ $(i = 1,\cdots,q)$, respectively. Consider testing the hypothesis

$$H_o \; : \; \sum_{i=1}^{q} \beta_i \mu^{(i)} = \mu ,$$

where $\beta_1,\cdots,\beta_q$ are given scalars and μ is a given vector. If the N_i are not equal, let N_1 be the smallest. Let

$$y_\alpha = \beta_1 x_\alpha^{(1)} + \sum_{i=2}^{q} \beta_i \sqrt{\frac{N_1}{N_i}} \left(x_\alpha^{(i)} - \frac{1}{N_1} \sum_{\beta=1}^{N_1} x_\beta^{(i)} + \frac{1}{\sqrt{N_1 N_i}} \sum_{\gamma=1}^{N_i} x_\gamma^{(i)} \right) .$$

Then

$$E(y_\alpha) = \beta_1 \mu^{(1)} + \sum_{i=2}^{q} \beta_i \sqrt{\frac{N_1}{N_i}} \left(\mu^{(i)} - \frac{1}{N_1} N_1 \mu^{(i)} + \frac{N_i}{\sqrt{N_1 N_i}} \mu^{(i)} \right)$$

$$= \sum_{i=1}^{q} \beta_i \mu^{(i)}$$

and

$$E[y_\alpha - E(y_\alpha)][y_\beta - E(y_\beta)]' = \delta_{\alpha\beta} \left(\sum_{i=1}^{q} \frac{\beta_i^2 N_1}{N_i} \Sigma_i \right) ,$$

where $\delta_{\alpha\beta} = 1$ for $\alpha = \beta$

$\qquad = 0$ for $\alpha \neq \beta$.

Let $\bar{y}$ and S be defined by

$$\bar{y} = \frac{1}{N_1}\sum_{\alpha=1}^{N_1} y_\alpha = \sum_{i=1}^{q} \beta_i \bar{x}^{(i)} ,$$

where $$\bar{x}^{(i)} = \frac{1}{N_i}\sum_{\beta=1}^{N_i} x_\beta^{(i)} ,$$

$$(N_1-1)S = \sum_{\alpha=1}^{N_1} (y_\alpha-\bar{y})(y_\alpha-\bar{y})' .$$

Then

$$T^2 = N_1(\bar{y}-\mu)'S^{-1}(\bar{y}-\mu)$$

is suitable for testing H_o and when an alternative hypothesis is true the statistic $\frac{T^2}{(N_1-1)} \; \frac{(N_1-q)}{q}$ is distributed as $F'_{q,N_1-q}(\lambda)$, where the noncentrality parameter λ is given by

$$\lambda = \left(\sum_{i=1}^{q} \beta_i \mu^{(i)} - \mu\right)' \left(\sum_{i=1}^{q} \frac{\beta_i^2}{N_i} \Sigma_i\right)^{-1} \left(\sum_{i=1}^{q} \beta_i \mu^{(i)} - \mu\right).$$

(iii) We shall now compare the power probabilities of the T^2-statistic with the two-way *mixed* model analysis of variance (ANOVA)-statistic under the hypothesis $H_A : \mu_1 = \mu_2 = \cdots = \mu_p$ when the covariance matrix of the observations y_{ij} $(i = 1,2,\cdots,p;\ j = 1,2,\cdots,N)$ is

$$\Sigma = \begin{pmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & & \vdots \\ \vdots & & \ddots & \vdots \\ \rho & \cdots & \cdots & 1 \end{pmatrix} , \qquad (2.1)$$

where $\underset{p\times N}{Y} \sim N(\mu,\Sigma)$ with $\mu' = (\mu_1,\mu_2,\cdots,\mu_p)$.

Let MS_A denote the mean square due to the fixed effects and let MS_{AB} denote the mean square due to the interaction between the fixed and random effects. Under the alternative hypothesis of a general mean vector μ the statistic MS_A/MS_{AB} has the noncentral F distribution with (p-1) and (N-1)(p-1) degrees of freedom and noncentrality parameter

$$\lambda = N[\sigma^2(1-\rho)]^{-1} \sum_{i=1}^{p} (\mu_i - \bar{\mu}_.)^2 , \qquad (2.2)$$

where $\bar{\mu}_. = \frac{1}{p} \sum_{i=1}^{p} \mu_i$.

If, instead of the ANOVA-statistic, we use the scaled T^2-statistic to test H_A then, under an alternative hypothesis,

$\frac{(N-p+1)T^2}{(N-1)(p-1)}$ has the noncentral F distribution with (p-1) and (N-p+1) degrees of freedom and noncentrality parameter λ given in (2.2).

The exact power can be found using both the T^2- and the ANOVA-statistics. The T^2-statistic is exact for any positive definite covariance matrix Σ but the ANOVA-statistic is exact only for the case (2.1) above.

Morrison (1972) shows that higher power probabilities are obtained using the ANOVA-statistic. Some of the power values he gives are as follows: when $p = 4$, $N = 10$, $\alpha = 0{,}05$ and $\lambda = 10{,}24$ then the ANOVA power is 0,71 and the T^2 power is 0,52; when $p = 4$, $N = 10$, $\alpha=0{,}01$ and $\lambda = 23{,}04$ the ANOVA and T^2 powers are 0,94 and 0,49, respectively. The conclusion is that if symmetric Σ (2.1) model holds, the power of the

ANOVA test is appreciably better than that of the T^2 test for small samples.

§3. Testing Variances of Two Distributions

Let σ_1^2 and σ_2^2 be the variances of two normal distributions. The standard hypothesis testing situations are

$$H_o : \sigma_1^2 \geq \sigma_2^2 \quad \text{against} \quad H_1: \sigma_1^2 < \sigma_2^2 \tag{3.1}$$

$$H_o : \sigma_1^2 \leq \sigma_2^2 \quad \text{against} \quad H_1: \sigma_1^2 > \sigma_2^2 \tag{3.2}$$

$$H_o : \sigma_1^2 = \sigma_2^2 \quad \text{against} \quad H_1: \sigma_1^2 \neq \sigma_2^2 . \tag{3.3}$$

If $X_{11}, X_{21}, \cdots, X_{n1}$ and $X_{12}, X_{22}, \cdots, X_{n2}$ are random samples of size n from the two normal distributions with variances σ_1^2 and σ_2^2 respectively, then the statistic

$$F_{n-1,n-1} = \frac{S_1^2}{S_2^2} , \tag{3.4}$$

where S_1^2 and S_2^2 are the two sample variances, is used to test the above hypotheses. When $\sigma_1^2 = \sigma_2^2$, then $F_{n-1,n-1}$ has a central F distribution with n-1 and n-1 degrees of freedom.

If, however, the sample $X_{11}, X_{21}, \cdots, X_{n1}$ has an unstable mean, i.e., say

$$X_{i1} \sim N(\mu_{i1}, \sigma_1^2) \qquad (i = 1, \cdots, n)$$

then the statistic

$$F'_{n-1,n-1} = \frac{S_1^2}{S_2^2} \tag{3.5}$$

is used to test the above hypotheses, as before. However, when $\sigma_1^2 = \sigma_2^2$, then $F'_{n-1-n-1}$ of (3.5) is distributed as a noncentral F with n-1 and n-1 degrees of freedom and non-centrality parameter,

$$\lambda = \sum_{i=1}^{n} (\mu_{i1}-\bar{\mu}_1)^2/\ \sigma_1^2, \text{ where } \bar{\mu}_1 = \frac{1}{n} \sum_{i=1}^{n} \mu_{i1} .$$

The critical regions of size α are obtained by rejecting when

$$F'_{n-1,n-1} < F'_{n-1,n-1;\alpha} \quad \text{for} \quad (3.1),$$

$$F'_{n-1,n-1} > F'_{n-1,n-1;1-\alpha} \quad \text{for} \quad (3.2)$$

and

$$F'_{n-1,n-1} < F'_{n-1,n-1;\alpha/2} \text{ and } F'_{n-1,n-1} > F'_{n-1,n-1;1-\alpha/2}$$

$$\text{for} \quad (3.3).$$

The above noncentral F percentage points are easily approximated by approximations (2.3), (3.2), (7.2) and (8.1) of section II of this chapter.

§4. Minimum Sample Size Problem

In the test (3.1) in §3 above, if $\sigma_1^2 \neq \sigma_1^2$, say $\sigma_1^2/\sigma_2^2 = R^2$, then the statistic which has the F' distribution is

$$F'_{n-1,n-1} = \frac{S_1^2}{S_2^2}\,\frac{1}{R^2}\,. \tag{4.1}$$

The power of the test (3.1) is given by

$$\Pr\left(\frac{S_1^2}{S_2^2} < F'_{n-1,n-1;\alpha}\right) = \Pr\left(\frac{S_1^2}{S_2^2}\,\frac{1}{R^2} < \frac{1}{R^2}\,F'_{n-1,n-1;\alpha}\right)$$

$$= \Pr\left(F'_{n-1,n-1} < \frac{1}{R^2}\,F'_{n-1,n-1;\alpha}\right) ,$$

where $F'_{n-1,n-1}$ is given by (4.1).

If we want the power to be at least $1-\beta$, then we must have

$$\frac{1}{R^2}\,F'_{n-1,n-1;\alpha} \geq F'_{n-1,n-1;1-\beta}$$

or

$$\frac{F'_{n-1,n-1;1-\beta}}{F'_{n-1,n-1;\alpha}} \leq \frac{1}{R^2}\,. \tag{4.2}$$

Given R^2, we can find the minimum n that satisfied the inequality (4.2) by substituting an approximation to the noncentral F. A quick and convenient approximation to use would be approximation (2.3).

Chapter 4 : The Noncentral t distribution

I. Introduction

The distribution of the ratio

$$t'_\nu(\delta) = (U + \delta)/\sqrt{\chi^2_\nu/\nu} \quad ,$$

where U and χ^2_ν are independent random variables distributed as N(0,1) and chi-square with ν degrees of freedom respectively, and δ is a constant, is called the noncentral t distribution with ν degrees of freedom and noncentrality parameter δ.

The probability density function is given by

$$p_{t'_\nu}(t) = \frac{e^{-\frac{1}{2}\delta^2}\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\pi\nu}\;\Gamma\left(\frac{\nu}{2}\right)}\left(\frac{\nu}{\nu+t^2}\right)^{\frac{1}{2}(\nu+1)}\sum_{j=o}^{\infty}\frac{\Gamma\left(\frac{\nu+j+1}{2}\right)}{j!\;\Gamma\left(\frac{\nu+1}{2}\right)}\left(\frac{t\delta\sqrt{2}}{\sqrt{\nu+t^2}}\right)^j .$$

II. Approximations

§1. Jennett and Welch

Jennett and Welch (1939) used the approximate normality of $(U - t\chi_\nu\nu^{-\frac{1}{2}})$ in the equation

$$\Pr[t'_\nu \leqslant t] = \Pr[U - t\chi_\nu\nu^{-\frac{1}{2}} \leqslant -\delta]$$

to obtain

$$\Pr[t'_\nu \leqslant t] \doteqdot \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{1}{2}u^2}\,du \quad ,$$

where

$$X = \left(1 + \frac{t^2}{\nu} \ \mathrm{var}(\chi_\nu)\right)^{-\frac{1}{2}} \left(-\delta + \frac{t}{\sqrt{\nu}} \ E(\chi_\nu)\right) \sim N(0,1)$$

(1.1)

with

$$E(\chi_\nu) = \frac{\sqrt{2} \ \Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)}$$

and

$$\mathrm{var}(\chi_\nu) = E(\chi_\nu^2) - [E(\chi_\nu)]^2$$

$$= \nu - [E(\chi_\nu)]^2 \ .$$

The larger ν becomes, the more difficult it becomes to calculate $E(\chi_\nu)$. Owen (1968) gives an approximation to

$$c_{11} = \frac{\sqrt{\frac{f}{2}} \ \Gamma\left(\frac{f-1}{2}\right)}{\Gamma\left(\frac{f}{2}\right)}$$

which is correct to five decimal places for $f \geq 9$. The approximation is

$$c_{11} \doteqdot 1 + \frac{3}{4(f - 1{,}042)} \ .$$

Letting $f = \nu+1$, we obtain

$$c_{11} = \frac{\sqrt{\frac{\nu+1}{2}} \ \Gamma\left(\frac{\nu}{2}\right)}{\Gamma\left(\frac{\nu+1}{2}\right)} \doteqdot 1 + \frac{3}{4(\nu - 0{,}042)} = c.$$

Hence, $E(\chi_\nu) = \dfrac{\sqrt{\nu+1}}{c_{11}}$

and, therefore, $E(\chi_\nu) \doteqdot \dfrac{\sqrt{\nu+1}}{c}$.

This leads to X in equation (1.1) becoming

$$X \doteqdot \left[1 + \frac{t^2}{\nu}\left(\nu - \frac{\nu+1}{c^2}\right)\right]^{-\frac{1}{2}} \left(-\delta + \frac{t}{\sqrt{\nu}} \frac{\sqrt{\nu+1}}{c}\right) . \qquad (1.2)$$

An approximation to the percentage point $t'_{\nu,\alpha}(\delta)$, defined by

$$\Pr[t'_\nu(\delta) \leqslant t'_{\nu,\alpha}(\delta)] = \alpha ,$$

solving (1.1) for t and putting $X = U_\alpha$, is

$$t'_{\nu,\alpha}(\delta) \doteqdot \frac{\delta b_\nu + U_\alpha \sqrt{b_\nu^2 + (1-b_\nu^2)(\delta^2 - U_\alpha^2)}}{b_\nu^2 - U_\alpha^2 (1 - b_\nu^2)} , \qquad (1.1a)$$

where $b_\nu = E(\chi_\nu)/\sqrt{\nu}$ and U_α is the α-percentage point of the standard normal distribution.

However, the approximation (1.1a) can only be applied if

$$b_\nu^2 + (1 - b_\nu^2)(\delta^2 - U_\alpha^2) \geqslant 0 \qquad (1.1b)$$

or $$U_\alpha^2 \leqslant b_\nu^2 (1 - b_\nu^2)^{-1} + \delta^2 .$$

§2. Johnson and Welch

Johnson and Welch (1944) discuss three approcimations.

(i) In large samples $t'_\nu(\delta)$ becomes approximately normally distributed with mean δ and standard deviation $(1 + \delta^2/2\nu)^{\frac{1}{2}}$, i.e.,

$$t'_\nu(\delta) \stackrel{\bullet}{\sim} N(\delta, 1 + \delta^2/2\nu) \text{ for large } \nu . \qquad (2.1)$$

The approximation to the percentage point $t'_{\nu,\alpha}(\delta)$, defined by

$$\Pr[t'_\nu(\delta) \lesssim t'_{\nu,\alpha}(\delta)] = \alpha ,$$

is

$$t'_{\nu,\alpha}(\delta) \doteqdot \delta + U_\alpha \sqrt{1 + \delta^2/2\nu} . \quad (2.1a)$$

(ii) The probability that $t'_\nu(\delta)$ exceeds a given value t_o is

$$\Pr[(U + \delta)/\sqrt{\chi^2_\nu/\nu} > t_o]$$

and this is equivalent to

$$\Pr[-U + t_o \sqrt{\chi^2_\nu/\nu} < \delta] .$$

Since (-U) is a unit normal deviate, and $\sqrt{\chi^2_\nu/\nu}$ is very nearly normally distributed even for small ν, and U and $\sqrt{\chi^2_\nu/\nu}$ are independent, then $(-U + t_o \sqrt{\chi^2_\nu/\nu})$ must be more nearly normally distributed than $\sqrt{\chi^2_\nu/\nu}$.

Let $E[\sqrt{\chi^2_\nu/\nu}] = a$
and $\mathrm{var}[\sqrt{\chi^2_\nu/\nu}] = b/\sqrt{2\nu}$.
Then,

$$(-U + t_o \sqrt{\chi^2_\nu/\nu}) \mathrel{\dot{\sim}} N\left(at_o, 1 + \frac{b^2 t_o^2}{2\nu}\right) \text{ for large } \nu. \quad (2.2)$$

(iii) Introducing the further approximation that, for large ν , $\mathrm{var}(\chi_\nu) \doteqdot \frac{1}{2}$ and $E(\chi_\nu) \doteqdot \sqrt{\nu}$, Johnson and Welch arrived at the approximation

$$(-U + t_o \sqrt{\chi^2_\nu/\nu}) \mathrel{\dot{\sim}} N\left(t_o , 1 + \frac{t_o^2}{2\nu}\right) \text{ for large } \nu. \quad (2.3)$$

This leads to the approximation for the percentage point

$t'_{\nu,\alpha}(\delta)$, given by

$$t'_{\nu,\alpha}(\delta) \doteq \frac{\delta + U_\alpha \sqrt{1 + (\delta^2 - U_\alpha^2)/2\nu}}{1 - (U_\alpha^2/2\nu)} \quad . \tag{2.3a}$$

(2.3a) is the same as (1.1a) with $E(\chi_\nu) \doteq \sqrt{\nu}$ and $var(\chi_\nu) \doteq \frac{1}{2}$.

Approximation (2.3a) holds only if

$$1 + (\delta^2 - U_\alpha^2)/2\nu \gtrsim 0 \tag{2.3b}$$

or $\quad U_\alpha^2 \lesssim 2\nu + \delta^2$.

§3. Harley

Harley (1957) shows that, if the population correlation coefficient ρ is not zero, then

$$v = r(1-r)^{-\frac{1}{2}} \nu^{\frac{1}{2}} g\{\rho\} \stackrel{\bullet}{\sim} t'_\nu(\delta), \tag{3.1}$$

where r is the sample correlation coefficient based on a sample of size $(\nu+2)$ from a bivariate normal population.

$g\{\rho\}$ and δ were obtained by finding the first three moments about the origin of v, i.e., $\mu'_i(v)$ $(i = 1,2,3)$ and the first three moments about the origin of the noncentral t distribution, i.e., $\mu'_j(t)$ $(j = 1,2,3)$ and then putting

$$\frac{\mu'_3(t)}{\mu'_1(t)} = \frac{\mu'_3(v)}{\mu'_1(v)} \quad \text{and} \quad \mu'_2(t) = \mu'_2(v) \quad .$$

$g\{\rho\}$ and δ were found to be

$$g\{p\} = [2(1-\rho^2)/(2-\rho^2)]^{\frac{1}{2}}$$

$$\delta = [(2\nu+1)\rho^2/(2-\rho^2)]^{\frac{1}{2}} \quad . \qquad (3.2)$$

From (3.2) we obtain

$$\rho = [2\delta^2/(2\nu+1+\delta^2)]^{\frac{1}{2}}.$$

The percentage point $t'_{\nu,\alpha}(\delta)$ is approximated by

$$t'_{\nu,\alpha}(\delta) \doteqdot \left(\frac{2\nu(1-\rho^2)}{2-\rho^2}\right)^{\frac{1}{2}} \frac{r_\alpha}{(1-r_\alpha)^{\frac{1}{2}}} , \qquad (3.3)$$

where r_α is given by

$$\Pr\{r > r_\alpha\} = \alpha$$

and can be found from David's (1938) tables.

§4. Merrington and Pearson

Merrington and Pearson (1958) found that a good approximation to the noncentral t distribution is obtained by fitting it to a *Pearson Type IV* distribution, making the first four moments agree with the first four moments of the noncentral t distribution.

§5. Halperin

Halperin (1963) proposed two inequalities, namely,

$$\Pr\left\{t'_\nu \leqslant \frac{\delta\sqrt{\nu}}{\chi_{\nu,1-\alpha}} + t_{\nu,\alpha} \middle| \delta \gtrsim 0\right\} \geqslant 1-\alpha \quad (\alpha \lesssim 0{,}5) \tag{5.1}$$

and

$$\Pr\left\{t'_\nu \leqslant \frac{\delta\sqrt{\nu}}{\chi_{\nu,\alpha}} - t_{\nu,\alpha} \middle| \delta \gtrsim 0\right\} \leqslant \alpha \quad (\alpha \lesssim 0{,}43), \tag{5.2}$$

where $\chi_{\nu,1-\alpha}$ and $\chi_{\nu,\alpha}$ are the $(1-\alpha)$- and α-percentage points of the central chi-distribution, respectively.

§6. Hogben, Pinkham and Wilk

Hogben, Pinkham and Wilk (1964) approximate the distribution of

$$Q = t'_\nu(\nu+t'^2_\nu)^{-\frac{1}{2}} = W/(W^2+Z^2)^{\frac{1}{2}} ,$$

where $W \sim N(\delta,1)$ is independent of $Z^2 \sim \chi^2_\nu$.

Thus,

$$t'_\nu(\delta) = \sqrt{\nu}\, Q/(1-Q^2)^{\frac{1}{2}} . \tag{6.1}$$

A Beta (Pearson Type I) distribution is fitted to Q.

Let X have a Beta distribution with parameters a and b. The approximation is made by equating the first two moments of Q and $Y = 2X - 1$ to obtain a and b. This yields

$$a = (1+\mu)(1-\mu^2-\sigma^2)/2\sigma^2 \tag{6.2}$$

$$b = (1-\mu)(1-\mu^2-\sigma^2)/2\sigma^2 , \tag{6.3}$$

where $\mu = E(Q)$ and $\sigma^2 = \mathrm{var}(Q)$. Therefore, we have

$$Q \overset{\bullet}{\sim} \beta(a,b) \quad , \tag{6.4}$$

where a, b are given in (6.2) and (6.3).

If $\Pr\{Q > k\} = \alpha$

then, by (6.1),

$$\begin{aligned} \alpha &= \Pr\{t'_\nu > \sqrt{\nu}k/(1-k^2)^{\frac{1}{2}}\} \\ &= \Pr\{t'_\nu > t_o\} \ . \end{aligned} \tag{6.5}$$

§7. Laubscher

Laubscher (1960) proposed a *variance-stabilizing* transformation. The first three central moments of the noncentral t, $t'_\nu(\delta)$, are

$$\mu_1 = (\tfrac{1}{2}\nu)^{\frac{1}{2}} \delta\Gamma(\tfrac{1}{2}\nu-\tfrac{1}{2})/\Gamma(\tfrac{1}{2}\nu) ,$$

$$\mu_2 = [\nu(1+\delta^2)/(\nu-2)] - \mu_1^2 ,$$

and

$$\mu_3 = \mu_1\{\nu(\delta^2+2\nu-3)/[(\nu-2)(\nu-3)] - 2\mu_2\} .$$

Eliminating δ between μ_1 and μ_2 he found that

$$\mu_2 = a^2 + b^2\mu_1^2 \quad ,$$

where

$$a = [\nu/(\nu-2)]^{\frac{1}{2}}$$

and

$$b = \{2\Gamma^2(\tfrac{1}{2}\nu)/[(\nu-2)\Gamma^2(\tfrac{1}{2}\nu-\tfrac{1}{2})] - 1\}^{\frac{1}{2}}$$

which is a positive real number for $\nu \geq 4$.

Referring to Chapter 1, equation (3.4), with K = o and c = 1, the variance-stabilizing transformation, $\xi(t'_\nu)$, of noncentral t is

$$\xi(t'_\nu) = \int_o^{t'_\nu} (a^2 + b^2\mu_1^2)^{-\frac{1}{2}} \, d\mu_1$$

$$= \alpha \sinh^{-1} (\beta t'_\nu) ,$$

where $\alpha = b^{-1}$ and $\beta = b/a$.

The random variable

$$\xi_1(t'_\nu) = \xi(t'_\nu) - \alpha \sinh^{-1}(\beta\mu)$$

will be, approximately, distributed as N(O,1). That is,

$$\alpha \sinh^{-1}(\beta t'_\nu) - \alpha \sinh^{-1}(\beta\mu) \mathrel{\dot{\sim}} N(O,1). \tag{7.1}$$

§8. Constance van Eeden

Constance van Eeden (see Johnson and Kotz (1970c), page 207) applied a *Cornish-Fisher* expansion to the distribution of $t'_\nu(\delta)$ to obtain the approximate expansion (up to and including terms in ν^{-2})

$$t'_{\nu,\alpha}(\delta) \doteqdot U_\alpha + \delta + \tfrac{1}{4}\,[U_\alpha^2 + U_\alpha + (2U_\alpha^2 + 1)\delta + U_\alpha\delta^2]\nu^{-1}$$

$$+ \frac{1}{96}\,[5U_\alpha^5 + 16U_\alpha^3 + 3U_\alpha + 3(4U_\alpha^4 + 12U_\alpha^2 + 1)\delta \tag{8.1}$$

$$+ 6(U_\alpha^3 + 4U_\alpha)\delta^2 - 4(U_\alpha^2-1)\delta^3 - 3U_\alpha\delta^4]\nu^{-2}.$$

Putting δ = o in (8.1) yields approximate central t_ν α-percentage points,

$$t_{\nu,\alpha} \doteqdot U_\alpha + \tfrac{1}{4}[U_\alpha^2 + U_\alpha]\nu^{-1} + \frac{1}{96}[5U_\alpha^5 + 16U_\alpha^3 + 3U_\alpha]\nu^{-2}. \tag{8.2}$$

Using approximation (8.2), (8.1) becomes

$$\begin{aligned} t'_{\nu,\alpha}(\delta) \doteqdot t_{\nu,\alpha} &+ \delta + \tfrac{1}{4}\delta[2U_\alpha^2 + 1 + \delta U_\alpha]\nu^{-1} \\ &+ \frac{1}{96}\delta[3(4U_\alpha^4 + 12U_\alpha^2 + 1) + 6\delta(U_\alpha^3 + 4U_\alpha) \\ &- 4\delta^2(U_\alpha^2 - 1) - 3U_\alpha\delta^3]\nu^{-2}. \end{aligned} \tag{8.3}$$

III. Comparisons of Accuracy

Tables 1 and 2 show comparisons between Jennett and Welch's approximation (1.1a), Johnson and Welch's approximations (2.1a) and (2.3a) and Halperin's approximations with the exact values of the upper and lower 5% points of the noncentral t distribution which were taken from Merrington and Pearson (1958), Table 1. It can be seen that the approximations (1.1a), (2.1a) and (2.3a) improve as ν increases and deteriorate for larger λ while Halperin's approximations only improve as λ decreases.

Of the other approximations not tabled Harley's approximation (3.3) involves the use of David's (1938) tables which are not always readily available. The approximation is, however, remarkably accurate. Merrington and Pearson's approximation makes use of tables of standardized deviates of the Type IV distribution and Hogben *et al.* approximation (6.5) uses the tables of the Incomplete Γ-function. Laubscher (1960) shows that approximation (7.1) is only good for large ν and small δ. Johnson and Kotz (1970c), page 209 mention that, for $\delta > o$, approximation (8.1) gives better results for lower percentage points ($\alpha < 0,5$), while (8.3) is better for $\alpha > 0,5$.

Tables comparing the accuracy of (1.1a) and (2.3a) are also to be found in section IV, §8 of this chapter. Since these two approximations make use only of percentage points of the standard normal distribution and as long as the sample size is not too small (say < 12) and δ is not too large, then they seem to be very useful. Approximation (2.3a) has the advantage that it can be used over a wider range of values than (1.1a) due to the conditions (1.1b) and (2.3b).

Table 1. Upper 5% points of the noncentral t distribution

ν	δ	Exact	Jennett & Welch	Johnson and Welch		Halperin
			(1.1a)	(2.1a)	(2.3a)	(5.1)
12	2,432 3,737	4,79 6,55	4,83 6,62	4,27 5,81	4,71 6,46	5,47 7,45
20	3,091 14,161	5,28 19,60	5,29 19,77	4,92 18,20	5,15 19,49	5,92 20,95
49	4,769 21,851	6,79 26,63	6,79 26,71	6,60 25,84	6,76 26,56	7,41 27,94

Table 2. Lower 5% points of the noncentral t distribution

ν	δ	Exact	Jennett & Welch	Johnson and Welch		Halperin
			(1.1a)	(2.1a)	(2.3a)	(5.2)
12	2,432 3,737	0,78 2,00	0,78 2,00	0,59 1,67	0,77 1,96	0,06 1,04
20	3,091 14,161	1,42 10,96	1,42 11,01	1,26 10,13	1,41 10,89	0,74 9,58
49	4,769 21,851	3,06 18,43	3,06 18,47	2,94 17,87	3,05 18,38	2,42 17,11

IV. Applications

Unless otherwise stated, in the following applications we shall assume a random sample $x_1, x_2, \cdots\cdots, x_n$ from a normal population with unknown mean μ and variance σ^2. The statistics used will be $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ and

$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$, where $\bar{x}$ and s^2 are the mean and variance, respectively, of the sample.

§1. One-sided tolerance limits for a normal distribution

Let proportion $P = \Pr\{X \leq \mu + K_P\sigma\}$, where K_P is the deviate corresponding to P for a standardized normal distribution, i.e.,

$$G(K_P) = P = \int_{-\infty}^{K_P} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}Z^2} dZ. \tag{1.1}$$

If μ and σ^2 are known it is possible to say that exactly a proportion P of the normal population is below the *upper tolerance limit*, $\mu + K_P\sigma$, for a normal random variable X.

Usually, however, μ and σ^2 are unknown and they must be estimated from a random sample. Then, a tolerance limit of the form $\bar{x} + ks$ may be used. Since $\bar{x}$ and s are random variables a given probability must be attached to the tolerance limit statement.

The problem now becomes one of finding k such that the probability is γ that at least a proportion P of the population is below $\bar{x} + ks$, i.e., find k such that

$$\Pr\{\Pr[X \leqslant \bar{x} + ks] \geqslant P\} = \gamma .$$

But $\Pr\{\Pr[X \leqslant \bar{x} + ks] \geqslant P\}$ (1.2)

$$= \Pr\{\Pr[X \leqslant \bar{x} + ks] \geqslant \Pr[X \leqslant \mu + K_P\sigma]\}$$

$$= \Pr\{\bar{x} + ks \geqslant \mu + K_P\sigma\}$$

$$= \Pr\{(\bar{x} + ks - \mu)/\sigma \geqslant K_P\}$$

$$= \Pr\left\{ \frac{\left(\frac{\bar{x}-\mu}{\sigma}\right)\sqrt{n} - K_P\sqrt{n}}{s/\sigma} \geqslant -k\sqrt{n} \right\} = \gamma .$$

This is now in the form of the noncentral t distribution with $\nu = n-1$ degrees of freedom and noncentrality parameter $\delta = -K_P\sqrt{n}$ and can be equivalently written as

$$\Pr\{t'_{n-1}(\delta) \leqslant k\sqrt{n}\,|\,\delta = K_P\sqrt{n}\} = \gamma.$$

Example. (Owen (1968))

Suppose k is required for n = 10, P = 0,975 and γ = 0,95.

Therefore $K_P = 1{,}96$ and $\delta = K_P\sqrt{n} = 6{,}19806$. Owen's tables (1968) using the Johnson and Welch (1944) procedure give values of λ for each η to give $\Pr\{t'_{n-1}(\delta) \leqslant t_o|\delta\} = 0{,}95$, where

$$\eta = \delta/\sqrt{2\nu+\delta^2} \quad \text{and } \lambda \text{ is given by}$$

$$k = \frac{K_P + \lambda\left(\frac{1}{n} + \frac{K_P^2}{2\nu} - \frac{\lambda^2}{2\nu n}\right)^{\frac{1}{2}}}{\left(1 - \frac{\lambda^2}{2\nu}\right)} . \qquad (1.3)$$

Equation (1.3) gives the result

$$k = 3{,}402.$$

Various approximations, numbered as in section II of this chapter, yield the following values for k.

	(1.1a)	(2.1a)	(2.3a)	(5.1)	(7.2)
k	3,478	2,878	3,357	3,802	2,937

Approximations (2.1a) and (5.1) are the easiest to apply but are very inaccurate. With a little bit more labour (1.1a) and (2.3a) show a vast improvement in accuracy. However, for a much larger n (2.1a) would improve in accuracy. Similarly, if δ were smaller (5.1) and (7.2) would improve in accuracy.

§2. One-sided variables sampling plans based on the Normal distribution

Definition 1 : A *variables sampling plan* is a particular procedure for accepting or rejecting a lot of material based on a single measurement of a single variable (assumed to be normally distributed) on a random selection of items from the lot.

Definition 2 : A *one-sided variables sampling plan* is one which guarantees that at least a proportion P of the population being measured is greater than a value L with a given probability γ , or which guarantees that at least a proportion P of the population measured is less than a value U with a given probability γ.

The procedure is to compute $\bar{x}$ and s and to accept the lot if $\bar{x} + ks \leq U$ for an upper limit and reject the lot otherwise or accept the lot if $\bar{x} - ks \geq L$ for a lower limit and reject the lot otherwise. One can say that one is 100 γ% confident that at least 100 P% of the lot is below U for an upper limit or is above L for a lower limit.

We must find k such that

$$\Pr\{\bar{x} + ks \leq U\} = 1 - \gamma.$$

Using the same procedure as in equation (1.2) above this leads to

$$\Pr\{t'_{n-1}(\delta) \leq -k\sqrt{n} \mid \delta = -\left(\frac{U-\mu}{\sigma}\right)\sqrt{n}\} = 1 - \gamma$$

or

$$\Pr\{t'_{n-1}(\delta) \leq k\sqrt{n} \mid \delta = \left(\frac{U-\mu}{\sigma}\right)\sqrt{n}\} = \gamma .$$

This equation is identical to the one given for the tolerance limit problem with K_P replaced by $\left(\frac{U-\mu}{\sigma}\right)$.

<u>Example</u>: (Owen (1968))

Suppose a sampling plan is chosen so that there is a 10% chance of accepting a lot with as much as 5% of the measured variable above U, based on a sample of size 20. This means that we have $\gamma = 0,90$, $P = 0,95$ and $n = 20$. The exact value of k is given by Owen (1968) as $k = 2,208$. The procedure is then to accept the lot if $\bar{x} + 2,208\ s \leq U$ and to reject otherwise.

Jennett and Welch's approximation (1.1a) yields the result k = 2,206. Johnson and Welch's approximations (2.1a) and (2.3a) yield the values k = 2,085 and k = 2,175, respectively. The accuracy of these three approximations has improved markedly compared to the example in §1 due to the increase in the sample size.

Suppose it is necessary to find out what percentage of the normal population above the value U will give us 95% assurance of accepting the population. That is, it is necessary to solve the following equation for $(U-\mu)/\sigma$:

$$\Pr\left\{ \frac{\bar{x} + ks - \mu}{\sigma} \geq \frac{U-\mu}{\sigma} \right\} = 0{,}05, \text{ where } k = 2{,}208.$$

Owen's (1968) exact value for $(U-\mu)/\sigma$ is

$$(U-\mu)/\sigma = 2{,}883.$$

This is the value K_P. The proportion P of the normal population above U can be found by looking up the quantity K_P in a table of the normal probability distribution. P is found to be 0,0020. That is, if only 0,20% of the normal population is above U, then the probability of accepting the population is 0,95.

With k = 2,206, using Jennett and Welch's approximation (1.2) we obtain $(U-\mu)/\sigma = 2{,}866$ which, in turn, yields the proportion P = 0,0021 or 0,21%.

§3. Confidence limits on a One-sided Quantile

We wish to find confidence limits on $\mu + \sigma K_P$ where $P = \Pr\{X \leq \mu + \sigma K_P\}$. That is, confidence limits are required on the value of the quantile for which exactly a proportion P of the population is smaller than that quantile.

An upper one-sided confidence limit on $\mu + \sigma K_P$ is given by $\bar{x} + ks$, where k can be found as in §1 and §2 of this section for values of P, γ and n, where γ is the confidence required in the confidence limit. A lower one-sided confidence limit on $\mu + K_P\sigma$ is given by $\bar{x} + ks$ where the corresponding values needed are P, $1-\gamma$ and n.

Two-sided confidence limits on $\mu + K_P\sigma$ are given by $\bar{x} + k_{(1-\gamma)/2}s$ and $\bar{x} + k_{(1+\gamma)/2}s$. That is,

$$\Pr\{\bar{x} + k_{(1-\gamma)/2}s \leq \mu + K_P\sigma \leq \bar{x} + k_{(1+\gamma)/2}s\} = \gamma .$$

§4. Confidence limits on a One-Sided Proportion P

Suppose one would like to know a lower confidence limit on the proportion P, which is defined by $P = \Pr\{X \leq x^*\}$, where x^* is some constant. Define k^* by $x^* = \bar{x} + k^*s$. This is similar to the one-sided tolerance limit problem except that instead of knowing P we are trying to find a value P_L which is a lower confidence limit on P.

We have to find $\delta = K_{P_L}\sqrt{n}$ given that

$$\Pr\{t'_{n-1}(\delta) \leq k^*\sqrt{n} \mid \delta = K_{P_L}\sqrt{n}\} = \gamma.$$

Example: (Owen (1968))

Suppose $\bar{x} = 15$ and $s = 3$ in a random sample of 20 items and we wish to know a *lower* confidence limit on the proportion less than $x^* = 21$. Now

$$t = k^*\sqrt{n} = \frac{(x^* - \bar{x})}{s}\sqrt{n} = 8{,}944272.$$

Let $\gamma = 0{,}95$. Owen's (1968) exact value is

$$K_{P_L} = 1{,}339246$$

and, therefore, $P_L = 0{,}90975$.
Therefore, we are 95% sure that at least a proportion 0,91 of the population is below $x^* = 21$. Using Jennett and Welch's approximation (1.2) we find

$$K_{P_L} = 1{,}33065$$

which yields $P_L = 0{,}9082$.

Suppose an *upper* 95% limit on the proportion below $x^* = 21$ is required. t will now be negative.

Owen's (1968) exact values are

$$K_{P_U} = 2{,}629415$$

and $$P_U = 0{,}99575.$$

Jennett and Welch's approximation (1.2) yields

$$K_{P_U} = 2{,}617087$$

and $$P_U = 0{,}9956.$$

Therefore, using the exact and approximate values we can say that we are 95% sure that at most a proportion 0,996 of the population is below $x^* = 21$.

Jennett and Welch's approximation (1.2) is again used here because of its accuracy and the ease with which it can be applied.

§5. The distribution of the sample coefficient of variation

The population *coefficient of variation* is defined by $V = \sigma/\mu$.

An estimate of V is provided by the sample coefficient of variation, $v = s/\bar{x}$. In practical situations, where V is an appropriate measure of variability, the variable x is usually necessarily positive. For a normal population V has to be of the order of 1/3 or less for the chance of a negative x to be negligible. Therefore, the sample should not be assumed normal if $v > 1/3$.

Now

$$\sqrt{n}/v = \sqrt{n}\,\bar{x}/s = \left(\frac{\sqrt{n}(\bar{x}-\mu)}{\sigma} + \frac{\sqrt{n}\mu}{\sigma}\right)\Bigg/\frac{s}{\sigma} \, . \qquad (5.1)$$

Therefore, we have

$\frac{\sqrt{n}}{v} \sim t'_{n-1}(\delta)$, where the noncentrality parameter is

$$\delta = \frac{\sqrt{n}\mu}{\sigma} = \frac{\sqrt{n}}{V} \, .$$

There are three problems relating to V that can be solved.

(i) Suppose we wish to test the hypothesis $H_o : V \lesssim V_o$ and we decide to reject H_o when $v \gtrsim v_o$ where v_o is chosen so that $\Pr\{v > v_o | V = V_o\} = \alpha$. Then v_o will be given by

$\frac{\sqrt{n}}{v_o} = t'_{n-1,1-\alpha}(\delta)$, where $\delta = \frac{\sqrt{n}}{V_o}$ and $t'_{n-1,1-\alpha}(\delta)$ is the $(1-\alpha)$-percentage point of the noncentral t distribution.

(ii) Suppose we decide to reject a sample as unsatisfactory when v is greater than a given v_o , and we require to know how low V should be kept to ensure that the probability of rejection will not exceed a given α. That is, we require V_o such that

$$\Pr\{v > v_o | V = V_o\} = \alpha .$$

That is,

$$\Pr\left\{\frac{\sqrt{n}}{v} > \frac{\sqrt{n}}{v_o} \middle| \delta = \frac{\sqrt{n}}{V_o}\right\} = 1 - \alpha, \qquad (5.2)$$

where

$$\frac{\sqrt{n}}{v_o} = t'_{n-1,1-\alpha}(\delta) .$$

Hence, the problem is one of finding the noncentrality parameter δ such that (5.2) holds.

(iii) Suppose that a value of v is observed and an upper confidence limit of V is required so that the chance is α of this limit being exceeded, i.e., a lower confidence limit of $\sqrt{n}/V$ is required. Since

$$\Pr\left\{\sqrt{n}/v > t'_{n-1,\alpha}\left(\frac{\sqrt{n}}{V}\right)\right\} = \alpha$$

and since the inequality

$$\sqrt{n}/v > t'_{n-1,\alpha}\left(\frac{\sqrt{n}}{V}\right)$$

is equivalent to the inequality

$$\sqrt{n}/V < \delta(n-1, \frac{\sqrt{n}}{v}, \alpha),$$

therefore, the required upper confidence limit of V is

$$V_L = \sqrt{n}/\delta(n-1, \frac{\sqrt{n}}{v}, \alpha) \; . \qquad (5.3)$$

[δ(a,b,c) means that δ depends on a,b, and c.]

<u>Example</u>: (Johnson and Welch (1944))

In a sample of size n = 25 a coefficient of variation v = 2,6 is observed. We wish to obtain a 50% confidence limit for V. From (5.3) above the required limit is

$$V_L = \sqrt{25}/\delta(24, \frac{\sqrt{n}}{v}, 0{,}5) \; ,$$

where

$$\sqrt{n}/v = 5/2{,}6 = 1{,}9231.$$

Using Johnson and Welch's (1944), Table IV we find the exact value of δ required is

$$\delta = 1{,}9027 \quad \text{which yields}$$

$$V_L = 2{,}628.$$

Using Jennett and Welch's approximation (1.2) we obtain

$$\delta = 1{,}9032$$

and, therefore,

$$V_L = 2{,}627 \; .$$

Johnson and Welch's approximation (2.3a) yields

$$\delta = \sqrt{n}/v = 1{,}9231$$

and $V_L = 2{,}6.$

This arises because for $\alpha = 0,5$, $U_\alpha = o$ and, therefore,

$$\delta = t = \sqrt{n}/v .$$

Halperin's approximation yields

$$\delta = 1,8964$$

and $$V_L = 2,637.$$

The improved accuracy of Halperin's approximation compared to the example in §1, section IV of this chapter is because $t_{\nu;0,5} = o$.

Note: This example violates the assumption that normality be assumed only if $v \leq 1/3$.

§6. Expected Coverage Tolerance Limits

Expected coverage tolerance limits are limits of the form $\bar{x} + ks$ such that the expected value of

$$\frac{1}{\sqrt{2\pi}\,\sigma} \int_{-\infty}^{\bar{x}+ks} \exp\left(-\tfrac{1}{2} \left(\frac{y-\mu}{\sigma}\right)^2 \right) dy \quad \text{is } P.$$

The tolerance limit which corresponds to the expected value 1-P is $\bar{x} - ks$. In words, expected coverage tolerance limits are limits of the form $\bar{x} + ks$ such that, *on the average*, the proportion P of the population is less than $\bar{x} + ks$.

The formula for k is

$$k = t_P \sqrt{\frac{n+1}{n}} ,$$ where t_P is a critical value of the central t distribution.

If μ is unknown and σ is known, then

$$k = K_P \sqrt{\frac{n+1}{n}} .$$

If μ is known and σ is unknown, then

$$k = t_P .$$

Alternatively, we could find k such that the expected value of $\bar{x} + ks$ is equal to $\mu + K_P\sigma$. The expected value of $\bar{x} + ks$ is

$$E(\bar{x} + ks) = \mu + \frac{k \sqrt{\frac{2}{n-1}}\, \Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)} \sigma$$

and, hence, k is

$$k = \frac{(n-1)}{(n-2)} \frac{K_P}{c_{11}} ,$$

where c_{11} is defined in section II, §1 of this chapter.

One question we may ask is this. What sample size should be used to obtain estimates of P using the formula $k = t_P\sqrt{(n+1/n}$? If we want to know that $\Pr\{\hat{P} < P + d\} \geq \alpha$, then we must solve the inequality

$$\Pr\left\{G\left(\frac{\bar{x} + ks - \mu}{\sigma}\right) \leq P + d\right\} \geq \alpha$$

for n. $G(\cdot)$ is given by (1.1), section IV of this chapter. This leads to the equation

$$\Pr\{t'_{n-1}(\delta) \leq t_P\sqrt{n+1} \,|\, \delta = K_{P+d}\sqrt{n}\} = 1 - \alpha.$$

This equation may be solved by trial and error to give the required value of n.

Similarly, if $\Pr\{\hat{P} > P - d\} \geq \alpha$, we arrive at the equation

$$\Pr\{t'_{n-1}(\delta) \leq t_p \sqrt{n+1} \mid \delta = K_{P-d}\sqrt{n}\} = \alpha.$$

§7. <u>The Power of Student's t-test</u>

(i) The statistic $\sqrt{n}(\bar{x}-\mu_o)/s$ can be used for testing hypotheses about the mean, μ, of a normal population in three standard ways, namely

$$H_o \;:\; \mu \geq \mu_o \quad \text{against} \quad H_1 \;:\; \mu < \mu_o \tag{7.1}$$

$$H_o \;:\; \mu \leq \mu_o \quad \text{against} \quad H_1 \;:\; \mu > \mu_o \tag{7.2}$$

$$H_o \;:\; \mu = \mu_o \quad \text{against} \quad H_1 \;:\; \mu \neq \mu_o \,. \tag{7.3}$$

When $\mu = \mu_o$, then $\sqrt{n}(\bar{x}-\mu_o)/s$ has a central t distribution with n-1 degrees of freedom. If $\mu = \mu_1$, then $\sqrt{n}(\bar{x}-\mu_o)/s$ has a noncentral t distribution with noncentrality parameter $\delta = \sqrt{n}(\mu_1 - \mu_o)/\sigma$ and the power of the test is calculated as a partial integral of the probability density function of this noncentral t distribution, i.e., if one is testing H_o in (7.2) above at level of significance α then the power of the test, $1-\beta$, is given by

$$\Pr\{t'_{n-1}(\delta) \geq t_{n-1,\alpha}\} = \int_{t_{n-1,\alpha}}^{\infty} p_{t'_\nu}(t)\,dt = 1-\beta \,,$$

where $\nu = n-1$.

(ii) Let μ_1 , μ_2 be the means of two normal populations. Then the three standard hypotheses to be tested are

$$H_o \ : \ \mu_1 \leq \mu_2 \quad \text{against} \quad H_1 \ : \ \mu_1 > \mu_2 \tag{7.4}$$

$$H_o \ : \ \mu_1 \geq \mu_2 \quad \text{against} \quad H_1 \ : \ \mu_1 < \mu_2 \tag{7.5}$$

$$H_o \ : \ \mu_1 = \mu_2 \quad \text{against} \quad H_1 \ : \ \mu_1 \neq \mu_2 . \tag{7.6}$$

If x_{11} , x_{21} ,...., $x_{n_1 1}$ and x_{12} , x_{22} ,....., $x_{n_2 2}$ are random samples from the two normal populations, then the statistic

$$\sqrt{n_1 n_2 (n_1+n_2)^{-1}}(\bar{x}_1-\bar{x}_2)/\sqrt{(n_1+n_2-2)^{-1}[(n_1-1)s_1^2+(n_2-1)s_2^2]},$$

where $\bar{x}_1$, $\bar{x}_2$ are the two sample means and s_1^2 , s_2^2 are the two sample variances, is used to test the above hypotheses. Assumin common variances, when $\mu_1 = \mu_2$ the above statistic has a central t distribution with $(n_1+ n_2-2)$ degrees of freedom.

When $\mu_1 \neq \mu_2$ then the statistic has a noncentral t distribution with (n_1+n_2-2) degrees of freedom and noncentrality paramete

$$\delta = \frac{(\mu_1-\mu_2)}{\sigma} \sqrt{\frac{n_1 n_2}{n_1+n_2}} .$$

For level of significance α , the power of the test (7.4) above will be

$$\Pr\{t'_{n_1+n_2-2}(\delta) \geq t_{n_1+n_2-2,\alpha}\} = \int_{t_{n_1+n_2-2,\alpha}}^{\infty} p_{t'_\nu}(t)dt = 1-\beta \quad ,$$

where $\nu = n_1+n_2-2$.

§8. The Minimum Sample Size Problem

(i) Suppose we wish to obtain a solution to the sample size problem in §7, (i) above. That is, we wish to determine the minimum sample size such that two power conditions are satisfied. These two conditions are that

Power $\leq \alpha$ when an hypothesis is true

$\geq 1-\beta$ when an alternative is true.

We have $\delta = \sqrt{n}(\mu_1 - \mu_0)/\sigma$ and, therefore,

$$n = \left(\frac{\sigma\delta}{\mu_1 - \mu_0}\right)^2 . \tag{8.1}$$

Using the Owen (1962) graphs of the power of Student's t-test, one can find the δ which yields a power $1-\beta$ for $n-1 = \infty$ and then use (8.1) to compute an n, say n_c. Since this δ is smaller than would be required to achieve a power $1-\beta$ with $n-1$ degrees of freedom, n_c will be smaller than the minimum actually required. That is, n_c is a lower bound for the value that is sought. Furthermore, if we use a δ which yields power $1-\beta$ for an $(n-1)$ that is too small, this δ will be larger than necessary and (8.1) will yield an n_c which may be larger than the minimum needed to meet requirements. Hence $(n-1)$ should be decreased from ∞ to 24, to 12, and so on (these are the values of $n-1$ for which curves are available) until the computed n, i.e., n_c, is greater than the chosen n. This n_c is an upper bound for the minimum n while the computed n obtained before this n_c is a lower bound. These Owen (1962) graphs only given an approximate solution to the problem.

A more efficient method to solve the sample size problem is by using Owen's (1965) tables of the power of Student's t-test.

Example:

Suppose we test an hypothesis about the mean with a one-sided alternative. How large must n be to yield power at least 0,99 when $\mu = \mu_1$ if $|\mu_1 - \mu_o| = \sigma/2$, at significance level α.

With $|\mu_1 - \mu_o| = \sigma/2$, (8.1) yields $n = (2\delta)^2$. Using the Owen (1962) graphs, for $n = \infty$ we obtain $\delta = 3{,}97$ and, therefore, $n_c = 4(3{,}97)^2 = 63{,}04$.

With n = 25 we find $\delta = 4{,}10$ and $n_c = 67{,}24$. This $n_c = 67{,}24$ is greater than the chosen n = 25 so that we know that the minimum n is between 63,04 and 67,24. We might take n = 65 as a practical solution.

Using Jennett and Welch's approximation (1.2) of section II we find, for $n = \infty$, $\delta = 3{,}965$ and $n_c = 62{,}89$ and we obtain for $n = 25$, $\delta = 4{,}086$ and $n_c = 66{,}78$.

From Owen's (1965) tables (exact to five decimal places) we obtain, for $n = \infty$, $\delta = 3{,}97120$ and $n_c = 63{,}08$ and

for $n = 25$, $\delta = 4{,}09018$ and $n_c = 66{,}92$.

Using interpolation in Owen's tables, we find that,

for $n = 64$, $\delta = 4{,}01439$ and $n_c = 64{,}46$ and

for $n = 65$, $\delta = 4{,}01393$ and $n_c = 64{,}45$. Therefore, n = 65 is the exact minimum sample size. [see Guenther (1973)]

Jennett and Welch's approximation (1.2) yields,

for $n = 64, \delta = 4{,}00808$ and $n_c = 64{,}26$ and

for $n = 65, \delta = 4{,}00748$ and $n_c = 64{,}24$.

Thus, we see that Jennett and Welch's approximation yields the same result that n = 65 is the minimum sample size.

In Table I, Johnson and Welch's approximation (2.3a) and Jennett and Welch's approximation (1.1a) have been compared with the exact values from Owen's (1965) tables for different values of the chosen n. The meanings of the symbols at the column heads are listed below.

δ_O ≡ noncentrality parameter taken from Owen's (1965) table. (The values in brackets (•) are approximate values read from Owen's (1962) graphs).

δ_{JW} ≡ noncentrality parameter due to Jennett and Welch's approximation (1.1).

δ_{JWA} ≡ noncentrality parameter due to Jennett and Welch's approximation (1.2).

δ_{JOW} ≡ noncentrality parameter due to Johnson and Welch's approximation (2.3).

n_O , n_{JW} , n_{JWA} and n_{JOW} are the computed values of n $= (2\delta)^2$ using the noncentrality parameters δ_O , δ_{JW} , δ_{JWA} and δ_{JOW} , respectively.

From Table I it can be seen that Johnson and Welch's approximation is more accurate, for $\nu > 24$, than Jennett and Welch's approximation and there is less computation involved.

Tables II, III and IV compare values of $n_c = (2\delta)^2$ using Jennett and Welch's (JW) approximation (1.2) and Johnson and Welch's (JOW) approximation (2.3) with the exact values using Owen's (1965) tables (O) for $\nu = n-1 = 3,24$ and 100, powers $1-\beta = 0,20; 0,50; 0,90; 0,95$ and levels of significance $\alpha = 0,005; 0,01; 0,025; 0,05$.

Table I ($1-\beta$ = 0,01; α = 0,05)

chosen n	degrees of freedom $\nu = n-1$	α = 0,05 significance limit of t-distribution $t_{\nu;\alpha}$	δ_o	δ_{JW}	δ_{JWA}	δ_{JOW}	computed $n = n_c$: n_o	n_{JW}	n_{JWA}	n_{JOW}
∞	∞	1,645	(3,97) 3,97120	3,965		3,965	(63,04) 63,08	62,89	62,89	62,89
201	200	1,6525	3,98473	3,97740		3,97946	63,51	63,28	63,28	63,34
101	100	1,6602	3,99845	3,99229		3,99644	63,95	63,75	63,75	63,89
65	64	1,6690	4,01393	4,00748		4,01452	64,45	64,24	64,24	64,47
64	63	1,6694	4,01439	4,00808		4,01492	64,46	64,26	64,26	64,48
41	40	1,6839	4,04085	4,03285		4,04334	65,31	65,06	65,06	65,39
25	24	1,7109	(4,10) 4,09018	4,08563		4,1005	(67,24) 66,92	66,77	66,77	67,26
13	12	1,7823	(4,21) 4,22447	4,21106		4,25078	(70,90) 71,39	70,93	70,93	72,28
7	6	1,9432	(4,57) 4,54674	4,50905	4,50903	4,60424	(83,54) 82,69	81,33	81,33	84,80
5	4	2,1318	(4,95) 4,95463	4,87370	4,87358	5,03644	(98,01) 98,19	95,01	95,01	101,46
4	3	2,3534	(5,44) 5,46629	5,31276	5,31374	5,57124	(118,37) 119,52	112,90	112,94	124,16
3	2	2,9200	(6,85) 6,88234	6,49002	6,49499	7,0264	(187,69) 189,47	168,48	168,74	197,48

Table II

Values of n_c for chosen $\nu = n-1 = 3$

Power = $1-\beta$		α: 0,050	0,025	0,010	0,005
0,95	O	79,44	127,87	237,61	378,44
	JW	77,16	122,83	225,68	357,35
	JOW	85,68	137,85	255,78	406,58
0,90	O	61,72	100,55	188,40	301,09
	JW	60,96	98,75	183,92	292,98
	JOW	68,19	111,53	209,55	334,96
0,50	O	18,20	32,91	66,22	108,90
	JW	18,80	34,39	70,01	115,84
	JOW	22,15	40,51	82,48	136,46
0,20	O	4,15	9,91	23,75	41,78
	JW	4,24	10,18	24,59	43,47
	JOW	5,65	13,04	30,72	53,84

Table III

Values of n_c for chosen $\nu = n-1 = 24$

Power = $1-\beta$		α: 0,050	0,025	0,010	0,005
0,95	O	45,90	56,52	71,07	82,55
	JW	45,76	56,34	70,83	82,24
	JOW	46,24	56,99	71,70	83,35
0,90	O	36,31	45,69	58,62	68,89
	JW	36,28	45,63	58,54	68,77
	JOW	36,71	46,21	59,33	69,78
0,50	O	11,46	16,68	24,31	30,62
	JW	11,47	16,69	24,33	30,65
	JOW	11,71	17,04	24,84	31,29
0,20	O	2,73	5,43	9,89	13,85
	JW	2,74	5,44	9,91	13,89
	JOW	2,86	5,64	10,24	14,31

Table IV

Values of n_c for chosen $\nu = n-1 = 100$

Power = 1-β			α 0,050	0,025	0,010	0,005
	0,95	O	43,88	53,00	64,84	73,70
		JW	43,76	52,87	64,68	73,53
		JOW	43,87	53,01	64,87	73,75
	0,90	O	34,73	42,85	53,51	61,55
		JW	34,69	42,82	53,47	61,50
		JOW	34,79	42,95	53,64	61,71
	0,50	O	10,97	15,67	22,25	27,44
		JW	10,97	15,67	22,24	27,45
		JOW	11,03	15,76	22,35	2,58
	0,20	O	2,62	5,10	9,06	12,44
		JW	2,63	5,11	9,08	12,46
		JOW	2,65	5,16	9,15	12,56

It can be seen that the approximations improve as n gets larger. The approximations improve as α gets larger and as power 1-β gets smaller. In general, Jennett and Welch's approximation is the more accurate of the two.

(ii) Suppose we now wish to obtain a solution to the sample size problem in the case of two normal populations.

Let $n = n_1 = n_2$. Therefore,

$$\delta = \frac{(\mu_1 - \mu_2)}{\sigma} \sqrt{\frac{n}{2}}$$

and

$$n = 2 \left(\frac{\sigma\delta}{\mu_1 - \mu_2} \right)^2$$

and the degrees of freedom will now be $\nu = 2n-2$.

<u>Example</u>:

Suppose we test hypothesis (7.6) with $\alpha = 0{,}05$. How large must n be to yield power at least 0,80 if $|\mu_1 - \mu_2| = \sigma$?

We have $n = 2\delta^2$. Because it is a two-sided test we must enter Owen (1962) graphs with $\alpha/2 = 0{,}025$. For $n = \infty$ we find that $\delta = 2{,}80$ yields $n_c = 15{,}68$. For $n = 13$ ($\nu = 2n-2 = 24$) we obtain $\delta = 2{,}92$ and $n_c = 17{,}05$. Since $n_c = 17{,}05 > n = 13$ we know that the minimum n must lie between 15,68 and 17,05. We could have $n = 16$ or $n = 17$.

Using Owen's tables, with $n = 16$ ($\nu = 2n-2 = 30$) we find that $\delta = 2{,}89536$ and $n_c = 16{,}77$. For $n = 17$ ($2n-2 = 32$) we find $\delta = 2{,}88980$ [see Guenther (1973)] and $n_c = 16{,}70$. Therefore, the exact minimum sample size is $n = 17$.

Using Jennett and Welch's approximation we find that,

for $n = \infty$, $\delta = 2{,}80$ and $n_c = 15{,}68$ and

for $n = 13$, $\delta = 2{,}91864$ and $n_c = 17{,}04$.

Using Johnson and Welch's approximation we obtain,

for $n = \infty$, $\delta = 2{,}83276$ and $n_c = 16{,}05$ and

for $n = 13$, $\delta = 2{,}94002$ and $n_c = 17{,}29$.

Again, using Jennett and Welch's approximation we find that,

for $n = 16$, $\delta = 2{,}89391$ and $n_c = 16{,}75$ and

for $n = 17$, $\delta = 2{,}88803$ and $n_c = 16{,}68$.

Also, for Johnson and Welch's approximation we find that,

for $n = 16$, $\delta = 2{,}91086$ and $n_c = 16{,}95$ and

for $n = 17$, $\delta = 2{,}90388$ and $n_c = 16{,}87$.

It can be seen that, in this example, Jennett and Welch's approximation is much more accurate than Johnson and Welch's approximation and, in fact, is almost exact.

In tables III and IV one can divide each of the values n_c by 2 to obtain the values of n_c for this example for $\alpha/2 = 0{,}050$; $0{,}025$; $0{,}010$; $0{,}005$ and $1-\beta = 0{,}95$; $0{,}90$; $0{,}50$; and $0{,}20$, and for $2n-2 = 24;100$ (i.e. $n = 13;51$).

Chapter 5 : The Multiple Correlation Coefficient

I. Introduction

The multiple correlation coefficient, $\bar{R}$, between a random variable X_o (the dependent variable) and variables $X_1, X_2, \ldots, X_k$ (the independent variables) with $k \geq 2$, is defined as the maximum correlation between X_o and the linear combination $\sum_{j=1}^{k} a_j X_j = a'X^{(2)}$, where $a' = (a_1, a_2, \ldots, a_k)$ and $X^{(2)'} = (X_1, X_2, \ldots, X_k)$, i.e.,

$$\bar{R} = \underset{a_1, a_2, \ldots, a_k}{\text{maximum}} \text{ correlation } (X_o, \sum_{j=1}^{k} a_j X_j).$$

If the covariance matrix of $X_o, X_1, \ldots, X_k$ is

$$V = \begin{pmatrix} \sigma_o & \sigma'_{(o)} \\ \sigma_{(o)} & \Sigma_{11} \end{pmatrix},$$

where Σ_{11} is the covariance matrix of $X_1, X_2, \ldots, X_k$, then the correlation between X_o and $\sum_{j=1}^{k} a_j X_j$ is

$$\rho(X_o, \sum_{j=1}^{k} a_j X_j) = (\sigma'_{(o)} a)/\{(a'\Sigma_{11} a)\sigma_o\}^{\frac{1}{2}}.$$

By an appropriate choice of the signs of the a_j's, it can always be arranged that $\rho(\cdot)$ is not negative. One can, therefore, choose a to maximize the square

$$(a'\sigma_{(o)}\sigma'_{(o)} a)/(a'\Sigma_{11} a).$$

The maximized value of the square of the correlation coefficient is

$$\sigma'_{(o)} \Sigma_{11}^{-1} \sigma_{(o)} / \sigma_o \quad ,$$

so the multiple correlation coefficient is

$$\bar{R} = \sqrt{\sigma'_{(o)} \Sigma_{11}^{-1} \sigma_{(o)} / \sigma_o} \quad .$$

Suppose now that X_o, X_1, ..., X_k have a joint multinormal distribution and that we have available values of n independent sets of these variables. If the elements of V are replaced in $\bar{R}$ by their maximum likelihood estimators, we obtain the sample multiple correlation coefficient, R. Hodgson (1968) has shown that, for $n > k+1$, $\frac{R^2}{1-R^2}$ is distributed exactly as

$$[\chi^2_{k-1} + \{U + \bar{R}(1 - \bar{R}^2)^{-\frac{1}{2}} \chi_{n-1}\}^2]/\chi^2_{n-k-1} \quad ,$$

where the χ^2's and the unit normal variable U are independent.

The distribution of R^2 was originally obtained by Fisher. The formula he obtained is

$$p_{R^2}(r^2) = \frac{\Gamma\left(\frac{n}{2}\right) (1-\bar{R}^2)^{\frac{1}{2}(n-1)}}{\pi \, \Gamma\left(\frac{k-1}{2}\right) \Gamma\left(\frac{n-k}{2}\right)} (r^2)^{\frac{1}{2}k-1} (1-r^2)^{\frac{1}{2}(n-k)-1}$$

$$\times \int_o^{\pi} \int_{-\infty}^{\infty} \frac{\sin^{k-2}\theta}{(\cosh\phi - \bar{R}r\cos\theta)^{n-1}} \, d\phi d\theta \qquad (r^2 > o).$$

II. Approximations

§1. Khatri

Khatri (1966) proposed two approximations.

(i) The Noncentral F approximation.

This approximation is

$$\frac{1}{k}\ (n-k-1)(1-\bar{R}^2)\omega\ \frac{R^2}{(1-R^2)}\ \dot{\sim}\ F'_{2k,2(n-k-1)}(\lambda)\ , \qquad (1.1)$$

where $\lambda = [(n-k-1)(n-1)]^{\frac{1}{2}}\bar{R}^2\omega$

and

$$\omega = [k + \{n-k-1 + \sqrt{(n-1)(n-k-1)}\}\bar{R}^2]/[k + (n-k-1)(2-\bar{R}^2)\bar{R}^2].$$

(ii) The Central F approximation

This approximation is

$$(n-k-1)(1-\bar{R}^2)[(n-k-1)\bar{R}^2+k]^{-1}\ \frac{R^2}{1-R^2}\ \dot{\sim}\ F_{a,b}\ , \qquad (1.2)$$

where

$$a = [(n-k-1)\bar{R}^2+k]^2/[(n-k-1)\bar{R}^2(2-\bar{R}^2)+k]$$

and

$$b = n-k-1.$$

§2. Gurland

Gurland (1968) proposed a simple approximation, namely,

$$\frac{R^2}{1-R^2}\ \dot{\sim}\ g\ G_{f,n-k-1} = g\ \frac{\chi^2_f}{\chi^2_{n-k-1}}\ , \qquad (2.1)$$

where g and f are found by equating the first two moments of $g\chi_f^2$ with the first two moments of

$$Y_1 = \frac{a'_{(o)} \ A_{11}^{-1} \ a_{(o)}}{\sigma_o - \sigma'_{(o)} \ \Sigma_{11}^{-1} \ \sigma_{(o)}} .$$

$a_{(o)}$ and A_{11} are the maximum likelihood estimates of $\sigma_{(o)}$ and Σ_{11}, respectively.

g and f are found to be

$$g = \{(n-1)\theta(\theta+2)+k\}/[(n-1)\theta+k]$$

and $$f = [(n-1)\theta+k]^2/\{(n-1)\theta(\theta+2)+k\} ,$$

where $$\theta = \bar{R}^2/(1-\bar{R}^2).$$

We can rewrite (2.1) as

$$\frac{R^2}{1-R^2} \stackrel{\cdot}{\sim} \frac{f}{n-k-1} \ gF_{f,n-k-1} = \frac{(n-1)\theta+k}{n-k-1} \ F_{f,n-k-1} . \tag{2.2}$$

(Approximation (2.2) is the same as Khatri's approximation (1.2))

§3. <u>Yoong-Sin Lee</u>

Yoong-Sin Lee (1971) proposed two approximations.

(i) A Noncentral F approximation.

We have already seen that

$$\frac{R^2}{1-R^2} \sim [\chi_{k-1}^2 + \{U + \bar{R}(1-\bar{R}^2)^{-\frac{1}{2}}\chi_{n-1}\}^2]/\chi_{n-k-1}^2 \tag{3.1}$$

$$= S/\chi_{n-k-1}^2$$ (see section I of this chapter).

Lee first approximates the variate S by $g\chi'^2_\nu(\lambda)$ where g, ν and λ are determined by equating the first three moments of S with the first three moments of $g\chi'^2_\nu(\lambda)$. The hth cumulant of S is found to be

$$\kappa_h = 2^{h-1}(h-1)!\{(n-1)(\gamma^{2h}-1) + k\}$$
$$= 2^{h-1}(h-1)!\phi_h ,$$

where $\gamma^{2h} = (1 - \bar{R}^2)^{-h}$. The first three cumulants of S are, therefore, ϕ_1 , $2\phi_2$ and $8\phi_3$. The first three cumulants of $g\chi'^2_\nu(\lambda)$ are found to be $g(\nu+\lambda)$, $2g^2(\nu+2\lambda)$ and $8g^3(\nu+3\lambda)$ so that the solutions for g, ν and λ are

$$g = [\phi_2 - (\phi_2^2 - \phi_1\phi_3)^{\frac{1}{2}}]/\phi_1 ,$$
$$\nu = [\phi_2 - 2\tilde{P}^2\gamma\sqrt{(n-1)(n-k-1)}]/g^2 \qquad (3.2)$$

and $\lambda = \tilde{P}^2\gamma\sqrt{(n-1)(n-k-1)}/g^2$,

where $\tilde{P}^2 = \bar{R}^2/(1 - \bar{R}^2)$.

Therefore,

$$\tilde{R}^2 = \frac{R^2}{1-R^2} \mathrel{\dot\sim} \frac{\nu g}{n-k-1} F'_{\nu,n-k-1}(\lambda) . \qquad (3.3)$$

The number ν found by (3.2) is usually a non-integer. Lee integerized ν in the following way.

Let $\nu^* = J(\nu)$,

where $J(\nu)$ is the integer nearest to ν. With ν^* fixed in this way, he assumed that

$$\tilde{R}^2 \overset{\bullet}{\sim} \frac{\nu^* g^*}{n-k-1} F'_{\nu^*,n-k-1}(\lambda^*). \tag{3.4}$$

Equating the first two moments he found that

$$g^* = \phi_2/[\phi_1 + (\phi_1^2 - \nu^*\phi_2)^{\frac{1}{2}}] \tag{3.5}$$

and $$\lambda^* = [\phi_1^2 - \nu^*\phi_2 + \phi_1(\phi_1^2 - \nu^*\phi_2)^{\frac{1}{2}}]/\phi_2 \ .$$

(ii) A Central F approximation.

Lee assumes that

$$\tilde{R}^2 \overset{\bullet}{\sim} c[F_{q,n-k-1} + a] \quad , \tag{3.6}$$

where q, a and c are determined by equating the first three moments (see Chapter 3, Section II, §5, (5.2)).

Therefore,

$$q = \tfrac{1}{2}(n-k-3)\left(\sqrt{\frac{E}{E-4}} - 1\right) \quad ,$$

$$c = \frac{q}{(n-k-1)\ (n+2q-k-3)}\ \frac{H}{K} \tag{3.7}$$

and $$a = \left(\frac{\phi_1}{c} - (n-k-1)\right)/(n-k-3),$$

where

$$H = 2\phi_1^2 + 3\phi_1\phi_2(n-k-3) + \phi_3(n-k-3)^2 \ ,$$

$$K = \phi_1^2 + \phi_2(n-k-3)$$

and $$E = H^2/K^3 \ .$$

Computations are again simplified by integerizing q. We then have

$$q^* = J(q) ,$$

$$c^* = \frac{1}{n-k-1}\left(\frac{q^*\{(n-k-3)\phi_2 + \phi_1^2\}}{n-k-3+q^*}\right)^{\frac{1}{2}} \qquad (3.8)$$

and $$a^* = \frac{(\phi_1/c^*) - (n-k-1)}{n-k-3} .$$

Therefore, the approximation is

$$\tilde{R}^2 \mathrel{\dot{\sim}} c^*\{F_{q^*,n-k-1} + a^*\} . \qquad (3.9)$$

§4. Hodgson

Hodgson (1968) suggested the following Normal approximation.

$$\left(\tilde{R}\left(n-k-\frac{3}{2}\right)^{\frac{1}{2}} - (k-1 + \{n-\frac{3}{2}\}\tilde{P}^2)^{\frac{1}{2}}\right)/[1 + \tfrac{1}{2}\tilde{R}^2 + \tfrac{1}{2}\tilde{P}^2]^{\frac{1}{2}} \qquad (4.1)$$

$$\mathrel{\dot{\sim}} N(0,1),$$

where $\tilde{R}^2 = R^2/(1-R^2)$

and $\tilde{P}^2 = \bar{R}^2/(1-\bar{R}^2)$.

§5. Gajjar

Gajjar has shown that, for n large,

$$\sqrt{n-1} \tanh^{-1} R \mathrel{\dot{\sim}} \chi'^2_k(\delta), \qquad (5.1)$$

where

$$\delta = (n-1)[\tanh^{-1} \bar{R}]^2$$

(see Johnson and Kotz (1970c), Page 245).

III. Comparisons of Accuracy

Table I below gives values of Prob[R $\leq$ x] , using the fact that

$$\text{Prob}\left(\frac{R^2}{1-R^2} \leq \frac{x^2}{1-x^2}\right) = \text{Prob}\,[R^2 \leq x^2] = \text{Prob}\,[R \leq x]\ ,$$

and various approximations are compared with the exact values. The exact values and the values of the approximations, with the exception of (4.1), are taken from Yoong-Sin Lee (1971).

Khatri's approximation (1.2) and Gurland's approximation (2.2) are the same 2-moment central F approximation.

Approximation (3.4) is difficult to apply because the noncentral F distribution is complicated and has not been well tabulated. The same applies to approximation (1.1) which does not appear in Table I.

Approximation (3.9), the integerized version of (3.6), is just as accurate to 3 decimal places as (3.6) and (1.2) and is, of course, easier to apply.

For ease of application Hodgson's approximation (4.1) takes the honours but is not anywhere near as accurate as the other approximations. The approximation (4.1) seems to be more accurate at the tails.

Approximation (5.1), which does not appear in Table I, improves as n increases.

Gurland and Milton (1970) show that Gurland's approximation (2.2) works best for smaller values of $\bar{R}$ or larger values of x and the approximation generally improves as k increases.

For the calculation of percentage points approximation

(3.9) is the best to use due to the integer degrees of freedom of the central F, but a little bit of tedious calculation is necessary.

Table I : Values of Prob[$R \leq x$]

n	$\bar{R}$	k	x	EXACT	KHATRI AND GURLAND (1.2) & (2.2)	YOONG-SIN LEE (3.4)	YOONG-SIN LEE (3.6)	YOONG-SIN LEE (3.9)	HODGSON (4.1)
15	0,5	3	0,2	0,012	0,006	0,012	0,021	0,024	0,042
			0,5	0,246	0,240	0,247	0,243	0,244	0,278
			0,8	0,902	0,903	0,902	0,902	0,902	0,892
25	0,8	5	0,5	0,001	0,000	0,001	0,001	0,001	0,001
			0,8	0,261	0,261	0,261	0,261	0,261	0,268
50	0,5	7	0,2	0,000	0,000	0,000	0,000	0,000	0,001
			0,5	0,173	0,170	0,173	0,172	0,172	0,196
			0,8	0,997	0,998	0,998	0,998	0,998	0,996

IV. Applications

§1. Power functions

$\frac{R^2}{1-R^2}$ is the quantity that arises in regression (or least squares) theory for testing the hypothesis that the regression of X_o on X_1 , X_2, ..., X_k is zero.

The common test of hypothesis which involves the population multiple correlation coefficient, $\bar{R}$, is

$$H_o \; : \; \bar{R} = o \qquad \text{against} \qquad H_1 \; : \; \bar{R} > o. \tag{1.1}$$

($\bar{R}$ is always $\geqslant$ o.)

The statistic

$$R^* = \frac{R^2/k}{(1-R^2)/n-k-1}$$

can be used to test this hypothesis, where it is well known that R^* is distributed exactly as a central F with k and (n-k-1) degrees of freedom. However, under H_1 , R^* is *not* distributed as a noncentral F but as

$$R^{*'} = \frac{k}{(n-k-1)} \left[\chi^2_{k-1} + \left\{ U + \bar{R}(1-\bar{R}^2)^{-\frac{1}{2}} \chi_{n-1} \right\}^2 \right] \bigg/ \chi^2_{n-k-1} \, , \tag{1.2}$$

where $U \sim N(O,1)$ [see Hodgson (1968)] .

We can, however, use Khatri's noncentral F approximation to $R^{*'}$, namely,

$$R^{*'} \mathrel{\dot{\sim}} [(1-\bar{R}^2)\omega]^{-1} \, F'_{2k,2(n-k-1)}(\lambda) \quad ,$$

where ω and λ are defined in section II, §1 of this chapter.

The power function for test (1.1) above is given by

$$\Pr[R^{*'} > R^{*}_{\alpha}] = 1-\beta \ , \qquad (1.3)$$

where R^{*}_{α} is the α-percentage point of R^{*} and is equivalent to $F_{k,n-k-1;\alpha}$, the α-percentage point of the central F distribution. Furthermore, if we approximate $R^{*'}$ by Khatri's approximation (1.1), (1.3) above becomes

$$\Pr[F'_{2k,2(n-k-1)}(\lambda) > (1-\bar{R}^2)\omega\, F_{k,n-k-1;\alpha}] \doteq 1-\beta \ , \qquad (1.4)$$

where $F'_{2k,2(n-k-1)}(\lambda)$ can be approximated by the approximations given in Chapter 3, section II, the most convenient approximations being (2.3) and (7.2).

§2. The Minimum Sample Size Problem

This involves finding the minimum n such that (1.3) above holds or, approximately, such that (1.4) holds. That is, using (1.4), we must find the n which satisfies the inequality

$$F_{k,n-k-1;\alpha} < (1-\bar{R}^2)\omega\, F'_{2k,2(n-k-1);1-\beta}(\lambda) \qquad (2.1)$$

most closely.

§3. The Partial Multiple Correlation Coefficient

Let $X = \begin{pmatrix} X^{(1)} \\ X^{(2)} \\ X^{(3)} \end{pmatrix}$ with q, r and s components, respectively.

Let X have a p-dimensional normal distribution, $N_p(0,\Sigma)$, where $p = q + r + s$.

Suppose we have two sets of conditional variables $(X^{(1)}|X^{(3)} = x^{(3)})$ and $(X^{(2)}|X^{(3)} = x^{(3)})$ with q and r components, respectively. We select one variable $(X_i^{(1)}|x^{(3)})$ from the set $(X^{(1)}|x^{(3)})$. The maximum correlation between $(X_i^{(1)}|x^{(3)})$ and a linear combination of the set $(X^{(2)}|x^{(3)})$ is called the *partial multiple correlation coefficient.* The population value is denoted by

$$\bar{R}_{i\cdot(r|s)} = \sqrt{\frac{\sigma_{(i\cdot3)} \Sigma_{22\cdot3}^{-1} \sigma'_{(i\cdot3)}}{\sigma_{ii\cdot3}}}, \tag{3.1}$$

where

$$\Sigma_{\cdot3} = \begin{bmatrix} \Sigma_{11\cdot3} & \Sigma_{12\cdot3} \\ \Sigma_{21\cdot3} & \Sigma_{22\cdot3} \end{bmatrix}$$ denotes the covariance

matrix of $\left(\begin{matrix} X^{(1)} \\ X^{(2)} \end{matrix} \middle| X^{(3)} = x^{(3)}\right)$ with

$$\Sigma_{ij\cdot3} = \Sigma_{ij} - \Sigma_{i3}\Sigma_{33}^{-1}\Sigma_{3j} \quad (i = 1,2; \quad j = 1,2),$$

$\sigma_{(i\cdot3)}$ denotes the ith row of $\Sigma_{12\cdot3}$, $\sigma_{ii\cdot3}$ denotes the variance of $(X_i^{(1)}|x^{(3)})$ and i locates the dependent variable.

Suppose we have a sample of size N from an $N_p(\mu,\Sigma)$ population. Let us suppose, too, that $q = 1$, $r > 1$ and $s \geq 1$. If we replace the elements of $\bar{R}_{i\cdot(r|s)}$ in (3.1) above by their maximum likelihood estimates we obtain the sample partial multiple correlation coefficient, $R_{1\cdot(r|s)}$ (noting that $q = 1$).

Furthermore, it can be shown that the distribution of the sample partial multiple correlation coefficient, $R_{1\cdot(r|s)}$, is the same as the distribution of the sample multiple correlation coefficient based on a sample size N-s from an (r+1)-dimensional normal population.

Thus, the approximations mentioned in section II of this chapter can be applied to the sample partial multiple correlation coefficient by replacing k by r and n by N-s. Therefore, if $\bar{R}_{1\cdot(r|s)} = o$, then

$$\frac{R^2_{1\cdot(r|s)} \Big/ r}{(1 - R^2_{1\cdot(r|s)}) \Big/ (N-s-r-1)} = F_{r,N-s-r-1} \quad . \qquad (3.2)$$

The partial multiple correlation coefficient can be used to determine whether or not certain variables should be excluded in a regression analysis. Often, too many variables are available in the regression analysis to give a good prediction of the criterion.

Suppose we wish to predict X_1 using either the set of s variables $X^{(3)}$ or the (r+s) variables $(X^{(2)'}, X^{(3)'})$. One measure of the accuracy of prediction of X_1 by linear combination of $X^{(3)}$ is $\mathrm{var}(X_1 - \beta_1 X^{(3)}) = \sigma_{11\cdot 3}$ and, similarly, using $X^{(2)}$ and $X^{(3)}$, the accuracy of prediction can be measured by

$$\mathrm{var}\left(X_1 - \beta_2 \begin{pmatrix} X^{(2)} \\ X^{(3)} \end{pmatrix}\right) = \sigma_{11\cdot 23} \quad .$$

If we wish to test whether there is significant difference between the two conditional variances $\sigma_{11\cdot 3}$ and $\sigma_{11\cdot 23}$, i.e., if we wish to test whether or not $X^{(2)}$ should be included in the prediction equation, then we test the hypothesis $H_o : \sigma_{11\cdot3} = \sigma_{11\cdot23}$ against the alternative hypothesis $H_1 : \sigma_{11\cdot3} > \sigma_{11\cdot23}$. It can be shown (Juritz(1969)) that this test is equivalent to testing if the partial multiple correlation coefficient is zero. That is, the test is

$$H_o : \bar{R}_{1\cdot(r|s)} = o \quad \text{against} \quad H_1 : \bar{R}_{1\cdot(r|s)} > o \tag{3.3}$$

and the test criterion is given by equation (3.2) above.

Linhart (1960) uses a different approach and his test for the same problem is

$$H_o : \bar{R}^2_{1\cdot(r|s)} \leq \frac{r}{N-s-1} \quad \text{against} \quad H_1 : \bar{R}^2_{1\cdot(r|s)} > \frac{r}{N-s-1} . \tag{3.4}$$

Another measure of the predictive accuracy is given by the expected value of the mean square error of prediction (Kerridge (1967)). If X_1 is predicted by the s variables $X^{(3)}$, the expected value of the mean square error of prediction will be given by

$$\delta^2_{(s)} = \frac{(N+1)(N-2)}{N(N-s-2)} \sigma_{11\cdot3} \qquad (N > s + 2). \tag{3.5}$$

Similarly, if X_1 is predicted by the (r+s) variables $X^{(2)}$ and $X^{(3)}$, then the expected mean square error of prediction will be

$$\delta^2_{(r+s)} = \frac{(N-2)(N+1)}{(N-r-s-1)(N-r-s-2)} \sigma_{11\cdot23} . \tag{3.6}$$

Thus, the decision whether to include $X^{(2)}$ or not, is based on the test

$$H_o : \delta^2_{(s)} \leq \delta^2_{(r+s)} \quad \text{against} \quad H_1 : \delta^2_{(s)} > \delta^2_{(r+s)} . \tag{3.7}$$

It can be shown that this test is equivalent to testing

$$H_o : \bar{R}^2_{1\cdot(r|s)} \leq \frac{r}{N-s-2} \quad \text{against} \quad H_1 : \bar{R}^2_{1\cdot(r|s)} > \frac{r}{N-s-2} . \tag{3.8}$$

Therefore, (3.4) and (3.8) must be tested using the test criterion (1.2) in §1 of this section which must again be approximated, preferably by Hodgson's approximation (4.1) of section II.

Example: (Juritz (1969))

In this example, there are nine variables $X_i (i = 1,2,\cdots,9)$ and it is assumed that the physical fitness variables $X_3 , X_4 , \cdots, X_9$ and height X_2 are correlated with weight X_1 . It is further assumed that the variables X_3 , X_4 and X_8 measure "muscular power" and the variables X_2 , X_5 , X_6 , X_7 and X_9 measure "agility". Juritz investigates whether there is any correlation between weight and "agility" when the effect of "muscular power" has been removed. The variables are partitioned as follows:

$$X^{(1)} = X_1 \qquad q = 1$$
$$X^{(2)'} = (X_2 , X_5 , X_6 , X_7 , X_9)' \qquad r = 5$$
$$X^{(3)'} = (X_3 , X_4 , X_8)' \qquad s = 3 .$$

The hypothesis $H_o : \bar{R}_{1\cdot(5|3)} = o$ is then tested and it is found that correlation exists.

Juritz then tests if the correlation could be as high as 0,7. That is, the hypothesis

$$H_o : \bar{R}_{1\cdot(r|s)} = 0,7$$
$$H_1 : \bar{R}_{1\cdot(r|s)} < 0,7$$

is set up. Hodgson's approximation (4.1) is then used to test this hypothesis with n replaced by N-s = N-3 and k by r = 5. Hodgson's statistic will then be

$$U' = \frac{[\tilde{R}(N-3-r-\frac{3}{2})^{\frac{1}{2}} - (r-1+\{N-3-\frac{3}{2}\}\tilde{P}^2)^{\frac{1}{2}}]}{[1+\frac{1}{2}\tilde{R}^2+\frac{1}{2}\tilde{P}^2]^{\frac{1}{2}}} \overset{\bullet}{\sim} N(0,1).$$

For example, for white girls of the age of 16, with N-3 = 40, we have $\tilde{P}$ = 0,9802 and $\tilde{R}$ = 1,05379 and U' = -0,3253 while the lower 5% point of the standard normal distribution is -1,64. That is, we accept H_o . (All values were obtained from Juritz (1969).)

Khatri's central F approximation (1.2), namely,

$$G = (N-3-r-1)(1-\bar{R}^2)[(N-3-r-1)\bar{R}^2+r]^{-1}\ \frac{R^2}{1-R^2} \overset{\bullet}{\sim} F_{a,b} ,$$

where $a = [(N-3-r-1)\bar{R}^2+r]^2/[(N-3-r-1)\bar{R}^2(2-\bar{R}^2)+k]$

and $b = N-3-r-1$

yields the values a = 16,28, b = 34, G = 0,7087. Now $F_{a,b;\ 0,95} = 1/F_{b,a;\ 0,05}$. Putting a $\doteqdot$ 16, we obtain $F_{b,a;\ 0,05} \doteqdot 2,17$ and, therefore, $F_{a,b;\ 0,95} \doteqdot 0,4608$. Thus, we have $G > F_{a,b;\ 0,95}$ and we accept the hypothesis.

As can be seen from section III of this chapter Khatri's approximation (1.2) is more accurate than Hodgson's approximation (4.1) and is as easy to apply. However, a is usually non-integer but if we approximate a by the integer closest to it one can guess the F value pretty accurately to two decimal places if good F tables are available.

Bibliography

ABDEL-ATY, S.H. (1954). Approximate formulae for the percentage points and the probability integral of the non-central χ^2-distribution, *Biometrika*, 41, 538-540.

ANDERSON, T.W. (1958). *An Introduction to Multivariate Statistical Analysis*, John Wiley, New York - London.

BARTON, D.E., DAVID, F.N. and O'NEILL, A.F. (1960). Some properties of the distribution of the logarithm of non-central F, *Biometrika*, 47, 417-424.

DAVID, F.N. (1938). *Tables of the Correlation Coefficient*, London; Cambridge University Press.

FIX, EVELYN. (1949). Tables of noncentral χ^2, *University of California Publications in Statistics*, 1, No. 2, 15-19.

GUENTHER, WILLIAM C. (1973). Determination of sample size for tests concerning means and variances of normal distributions, *Statistica Neerlandica*, 27, 103-113.

GURLAND, JOHN. (1968). A relatively simple form of the distribution of the multiple correlation coefficient, *Journal of the Royal Statistical Society, Series B*, 30, 276-283.

GURLAND, J. and MILTON, R. (1970). Further consideration of the distribution of the multiple correlation coefficient, *Journal of the Royal Statistical Society, Series B*, 32, 381-394.

HALPERIN, M. (1963). Approximations to the noncentral t, with applications, *Technometrics*, 5, 295-305.

HARLEY, B.I. (1957). Relation between the distributions of noncentral t and a transformed correlation coefficient, *Biometrika*, 44, 219-224.

HODGSON, V. (1968). On the sampling distribution of the multiple correlation coefficient (Abstract), *Annals of Mathematical Statistics*, 39, 307.

HOGBEN, D., PINKHAM, R.S. and WILK, M.B. (1964). An approximation to the distribution of Q (a variate related to the noncentral t), *Annals of Mathematical Statistics*, 35, 315-318.

JENNETT, W.J. and WELCH, B.L. (1939). The control of proportion defective as judged by a single quality characteristic varying on a continuous scale, *Journal of the Royal Statistical Society, Series B*, 6, 80-88.

JOHNSON, N.L. (1959). On an extension of the connexion between Poisson and χ^2-distributions, *Biometrika*, 46, 352-363.

JOHNSON, N.L. and KOTZ, S. (1970a). *Discrete Distributions*, Houghton Mifflin Company, Boston.

JOHNSON, N.L. and KOTZ, S. (1970b). *Continuous Univariate Distributions - 1*, Houghton Mifflin Company, Boston.

JOHNSON, N.L. and KOTZ, S. (1970c). *Continuous Univariate Distributions - 2*, Houghton Mifflin Company, Boston.

JOHNSON, N.L. and PEARSON, E.S. (1969). Tables of percentage points of noncentral χ, *Biometrika*, 56, 255-272.

JOHNSON, N.L. and WELCH, B.L. (1944). Applications of the noncentral t distribution, *Biometrika*, 31, 362-389.

JURITZ, J.M. (1969). The generalized partial correlation matrix, *Unpublished M.Sc. thesis of the University of Cape Town.*

JURITZ, J.M. (1973). A review of approximations to the distributions of test criteria in MANOVA, *Unpublished Technical Report of the University of Cape Town.*

KERRIDGE, D. (1967). Errors of prediction in multiple regression with stochastic regressor variables, *Techometrics*, 9, No.2, 309-311.

KHATRI, C.G. (1966). A note on the large sample distribution of a transformed multiple correlation coefficient, *Annals of the Institute of Statistical Mathematics*, 18, 375-380.

LAUBSCHER, NICO F. (1960). Normalizing the noncentral t and F distributions, *Annals of Mathematical Statistics*, 31, 1105-1112.

LEE, YOONG-SIN. (1971). Some results on the sampling distribution of the multiple correlation coefficient, *Journal of the Royal Statistical Society, Series B*, 31, 117-130.

LINHART, H. (1960). A criterion for selecting variables in a regression analysis, *Psychometrika*, 25, No. 1, 45-58.

MERRINGTON, MAXINE and PEARSON, E.S. (1958). An approximation to the distribution of noncentral t, *Biometrika*, 45, 484-491.

MORRISON, DONALD F. (1972). The analysis of a single sample of repeated measurements, *Biometrics*, 28, 55-71.

OWEN, D.B. (1962). *Handbook of Statistical Tables*, Reading, Massachusetts; Addison-Wesley Publishing Company, Inc.

OWEN, D.B. (1965). The power of Student's t-test, *Journal of the American Statistical Association*, 60, 320-333.

OWEN, D.B. (1968). A survey of properties and applications of the noncentral t-distribution, *Technometrics*, 10, 445-478.

PATNAIK, P.B. (1949). The noncentral χ^2- and F-distributions and their application, *Biometrika*, 36, 202-232.

PEARSON, E.S. (1959). Note on an approximation to the distribution of noncentral χ^2, *Biometrika*, 46, 364.

PEARSON, E.S. (1960). Editorial note, *Biometrika*, 47, 430-431.

PEARSON, E.S. and HARTLEY, H.O. (1951). Charts of the power function for analysis of variance tests, derived from the noncentral F-distribution, *Biometrika*, 38, 112-130.

PEARSON, E.S. and HARTLEY, H.O. (Editors) (1962). *Biometrika Tables for Statisticians*, London; Cambridge University Press.

PEARSON, K. (Editor) (1922). *Tables of the Incomplete Γ-function*, London; H.M. Stationery Office.

ROY, J. and MOHAMAD, J. (1964). An approximation to the noncentral chi-square distribution, *Sankhyā*, *Series A*, 26, 81-84.

SANKARAN, M. (1959). On the non-central chi-square distribution, *Biometrika*, 46, 235-237.

SANKARAN, M. (1963). Approximations to the non-central chi-square distribution, *Biometrika*, 50, 199-204.

SCHEFFÉ, H. (1959). *The Analysis of Variance*, New York; John Wiley & Sons, Inc.

SEVERO, N.C. and ZELEN, M. (1960). Normal Approximation to the Chi-square and noncentral F probability functions, *Biometrika*, 47, 411-416.

STYAN, GEORGE P.H. (1970). Notes on the distribution of quadratic forms in singular normal variables, *Biometrika*, 57, 567-572.

TIKU, M.L. (1965). Laguerre series forms of non-central χ^2 and F distributions, *Biometrika*, 52, 415-427.

TIKU, M.L. (1966). A note on approximating to the non-central F distribution, *Biometrika*, 53, 606-610.

TUKEY, JOHN W. (1957). Approximations to the upper 5% points of Fisher's B distribution and noncentral χ^2, *Biometrika*, 44, 528-530.

VON NEUMANN, J., KENT, R.H., BELLINSON, H.R. and HART, B.I. (1941). The mean square successive difference, *Annals of Mathematical Statistics*, 12, 153-162.