CHAPTER SIXTEEN

Exact Sampling Distributions-II (t, F and z Distributions)

LEARNING OBJECTIVES. Upon completion of this chapter, you should be able to :

1. Define Student's *t*, Fisher's *t*, F and z statistics and derive their probability distributions.
2. Obtain the various constants of t and F distributions and discuss their important properties.
3. Understand and appreciate various applications of t and F distributions in Statistics.
4. Understand and derive the relationship between t, F and chi-square distributions.
5. Derive the sampling distribution of correlation coefficient in a random sample from an uncorrelated bivariate normal population.
6. Define Fisher's Z-transformation for correlation coefficient and discuss its various applications in Statistics.

CHAPTER OUTLINE

16·1. INTRODUCTION

The entire large sample theory was based on the application of "*Normal Test*" (*cf.* § 14·9). However, if the sample size n is small, the distribution of the various statistics, *e.g.*, $Z = \dfrac{\bar{x} - \mu}{\sigma/\sqrt{n}}$ or $Z = (X - nP)/\sqrt{nPQ}$ etc., are far from normality and as such '*normal test*' cannot be applied if n is small. In such cases exact sample tests, pioneered by W.S. Gosset (1908) who wrote under the pen name of Student, and later on developed and extended by Prof. R.A. Fisher (1926), are used. In the following sections we shall discuss: (*i*) *t-test*, (*ii*) *F-test*, and (*iii*) *Fisher's z-transformation*.

The exact sample tests can, however, be applied to large samples also though the converse is not true. In all the exact sample tests, the basic assumption is that "*the population(s) from which sample(s) is (are) drawn is (are) normal, i.e., the parent population(s) is (are) normally distributed.*"

16·2. STUDENT'S 't' DISTRIBUTION

Let x_i $(i = 1, 2, \ldots, n)$ be a random sample of size n from a normal population with mean μ and variance σ^2. Then Student's t is defined by the statistic :

$$t = \frac{\bar{x} - \mu}{S/\sqrt{n}} \qquad \ldots(16\cdot1)$$

where $\bar{x} = \dfrac{1}{n}\sum_{i=1}^{n} x_i$, is the sample mean and $S^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$, ...(16·1a)

is an unbiased estimate of the population variance σ^2, and it follows Student's t distribution with $\nu = (n - 1)$ *d.f.* with probability density function :

$$f(t) = \frac{1}{\sqrt{\nu}\, B\left(\frac{1}{2}, \frac{\nu}{2}\right)} \cdot \frac{1}{\left(1 + \frac{t^2}{\nu}\right)^{(\nu+1)/2}} ; -\infty < t < \infty \qquad \ldots(16\cdot 2)$$

Remarks 1. A statistic t following Student's t-distribution with n $d.f.$ will be abbreviated as $t \sim t_n$.

2. If we take $\nu = 1$ in (16·2), we get :

$$f(t) = \frac{1}{B\left(\frac{1}{2}, \frac{1}{2}\right)} \cdot \frac{1}{(1+t^2)} = \frac{1}{\pi} \cdot \frac{1}{(1+t^2)} ; -\infty < t < \infty \qquad [\because \Gamma(1/2) = \sqrt{\pi}]$$

which is the $p.d.f.$ of standard Cauchy distribution. Hence, when $\nu = 1$, Student's t distribution reduces to Cauchy distribution.

16·2·1. Derivation of Student's t-distribution. The expression (16·1) can be re-written as :

$$t^2 = \frac{n(\bar{x}-\mu)^2}{S^2} = \frac{n(\bar{x}-\mu)^2}{ns^2/(n-1)} \Rightarrow \frac{t^2}{(n-1)} = \frac{(\bar{x}-\mu)^2}{\sigma^2/n} \cdot \frac{1}{ns^2/\sigma^2} = \frac{(\bar{x}-\mu)^2/(\sigma^2/n)}{ns^2/\sigma^2}$$

Since x_i $(i = 1, 2, \ldots, n)$ is a random sample from the normal population with mean μ and variance σ^2, $\quad \bar{x} \sim N(\mu, \sigma^2/n) \quad \Rightarrow \quad \dfrac{(\bar{x}-\mu)}{\sigma/\sqrt{n}} \sim N(0, 1)$

Hence $\dfrac{(\bar{x}-\mu)^2}{\sigma^2/n}$, being the square of a standard normal variate is a chi-square variate with 1 $d.f.$

Also $\dfrac{ns^2}{\sigma^2}$ is a χ^2-variate with $(n-1)$ $d.f.$ (*c.f. Theorem 15·5*).

Further since $\bar{x}$ and s^2 are independently distributed (*c.f. Theorem* 15·5), $\dfrac{t^2}{n-1}$ being the ratio of two independent χ^2-variates with 1 and $(n-1)$ $d.f.$ respectively, is a $\beta_2\left(\frac{1}{2}, \frac{n-1}{2}\right)$ variate and its distribution is given by :

$$dF(t) = \frac{1}{B\left(\frac{1}{2}, \frac{\nu}{2}\right)} \cdot \frac{(t^2/\nu)^{\frac{1}{2}-1}}{\left(1 + \frac{t^2}{\nu}\right)^{(\nu+1)/2}} d(t^2/\nu), 0 \le t^2 < \infty \qquad [\text{where } \nu = (n-1)]$$

$$= \frac{1}{\sqrt{\nu}\, B\left(\frac{1}{2}, \frac{\nu}{2}\right)} \cdot \frac{1}{\left(1 + \frac{t^2}{\nu}\right)^{(\nu+1)/2}} dt ; -\infty < t < \infty$$

the factor 2 disappearing since the integral from $-\infty$ to ∞ must be unity. This is the required probability density function as given in (16·2) of Student's t-distribution with $\nu = (n-1)$ $d.f.$

Remarks on Student's 't'. 1. ***Importance of Student's t-distribution in Statistics.*** W.S. Gosset, who wrote under pseudonym (pen-name) of Student defined his t in a slightly different way, *viz.*, $t = (\bar{x}-\mu)/s$ and investigated its sampling distribution, somewhat empirically, in a paper entitled *'The Probable Error of the Mean'*, published in 1908. Prof. R.A. Fisher, later on defined his own 't' and gave a rigorous proof for its sampling distribution in 1926. The salient feature of 't' is that both the statistic and its sampling distribution are functionally independent of σ, the population standard deviation.

The discovery of 't' is regarded as a landmark in the history of statistical inference. Before Student gave his 't', it was customary to replace σ^2 in $Z = \dfrac{\bar{x} - \mu}{\sigma/\sqrt{n}}$, by its unbiased estimate S^2 to give $t = \dfrac{\bar{x} - \mu}{S/\sqrt{n}}$ and then normal test was applied even for small samples. It has been found that although the distribution of t is asymptotically normal for large n (c.f. § 16·2·5), it is far from normality for small samples. The Student's t ushered in an era of exact sample distributions (and tests) and since its discovery many important contributions have been made towards the development and extension of small (exact) sample theory.

2. *Confidence or Fiducial Limits for* μ. If $t_{0.05}$ is the tabulated value of t for $\upsilon = (n-1)$ *d.f.* at 5% level of significance, *i.e.*, $P(\,|\,t\,|>t_{0.05}) = 0.05 \quad \Rightarrow \quad P(\,|\,t\,| \le t_{0.05}) = 0.95,$

the 95% confidence limits for μ are given by :

$$|\,t\,| \le t_{0.05}, \quad i.e., \quad \left|\frac{\bar{x} - \mu}{S/\sqrt{n}}\right| \le t_{0.05} \quad \Rightarrow \quad \bar{x} - t_{0.05}\cdot\frac{S}{\sqrt{n}} \le \mu \le \bar{x} + t_{0.05}\,\frac{S}{\sqrt{n}}$$

Thus, 95% confidence limits for μ are : $\bar{x} \pm t_{0.05}\cdot(S/\sqrt{n})$ ...(16·2a)

Similarly, 99% confidence limits for μ are : $\bar{x} \pm t_{0.01}\,(S/\sqrt{n})$ ...(16·2b)

where $t_{0.01}$ is the tabulated value of t for $v = (n-1)$ *d.f.* at 1% level of significance.

16·2·2. Fisher's 't' (*Definition*). It is the ratio of a standard normal variate to the square root of an independent chi-square variate divided by its degrees of freedom. If ξ is a $N(0, 1)$ and χ^2 is an independent chi-square variate with n *d.f.*, then Fisher's t is given by :

$$t = \frac{\xi}{\sqrt{\chi^2/n}} \qquad ...(16·3)$$

and it follows Student's 't' distribution with n degrees of freedom.

16·2·3. Distribution of Fisher's 't'. Since ξ and χ^2 are independent, their joint probability differential is given by :

$$dF(\xi, \chi^2) = \frac{1}{\sqrt{2\pi}}\exp(-\xi^2/2)\,\frac{\exp(-\chi^2/2)\,(\chi^2)^{\frac{n}{2}-1}}{2^{n/2}\,\Gamma(n/2)}\,d\xi\,d\chi^2$$

Let us transform to new variates t and u by the substitution :

$$t = \frac{\xi}{\sqrt{\chi^2/n}} \quad \text{and} \quad u = \chi^2 \quad \Rightarrow \quad \xi = t\sqrt{u/n} \text{ and } \chi^2 = u$$

Jacobian of transformation J is given by :

$$J = \frac{\partial(\xi, \chi^2)}{\partial(t, u)} = \begin{vmatrix} \sqrt{u/n} & t/(2\sqrt{un}) \\ 0 & 1 \end{vmatrix} = \sqrt{\frac{u}{n}}$$

The joint $p.d.f$ $g(t, u)$ of t and u becomes :

$$g(t, u) = \frac{1}{\sqrt{2\pi}\;2^{n/2}\,\Gamma(n/2)\sqrt{n}}\exp\left\{-\frac{u}{2}\left(1 + \frac{t^2}{n}\right)\right\}u^{\frac{n}{2}-\frac{1}{2}}\,du$$

Since $\psi^2 \ge 0$ and $-\infty < \xi < \infty$, $u \ge 0$ and $-\infty < t < \infty$.

Integrating *w.r. to* 'u' over the range 0 to ∞, the marginal *p.d.f.* $g_1(.)$ of t becomes :

$$g_1(t) = \frac{1}{\sqrt{2\pi}\;2^{n/2}\,\Gamma(n/2)\sqrt{n}}\left[\int_0^\infty \exp\left\{-\frac{u}{2}\left(1 + \frac{t^2}{n}\right)\right\}u^{(n-1)/2}\,du\right]$$

$$= \frac{1}{\sqrt{2\pi}\, 2^{n/2}\, \Gamma(n/2)\sqrt{n}} \cdot \frac{\Gamma[(n+1)/2]}{\left[\frac{1}{2}\left(1+\frac{t^2}{n}\right)\right]^{(n+1)/2}}$$

$$= \frac{\Gamma[(n+1)/2]}{\sqrt{n}\,\Gamma(n/2)\Gamma(1/2)} \cdot \frac{1}{\left(1+\frac{t^2}{n}\right)^{(n+1)/2}}, -\infty < t < \infty$$

$$= \frac{1}{\sqrt{n}\, B\left(\frac{1}{2}, \frac{n}{2}\right)\left(1+\frac{t^2}{n}\right)^{(n+1)/2}}, \quad -\infty < t < \infty$$

which is the probability density function of Student's t-distribution with n *d.f.*

Remarks 1. In Fisher's 't' the *d.f.* is the same as the *d.f.* of chi-square variate.

2. Student's 't' may be regarded as a particular case of Fisher's 't' as explained below.

Since $\bar{x} \sim N(\mu, \sigma^2/n)$, $\quad \xi = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \sim N(0, 1) \ldots (*)$ and $\chi^2 = \frac{ns^2}{\sigma^2} = \sum_{i=1}^{n} (x_i - \bar{x})^2/\sigma^2 \ldots (**)$

is independently distributed as chi-square variate with $(n-1)$ *d.f.* Hence Fisher's t is given by :

$$t = \frac{\xi}{\sqrt{\chi^2/(n-1)}} = \frac{\sqrt{n}\,(\bar{x}-\mu)}{\sigma} \cdot \frac{\sigma}{\sqrt{\Sigma(x_i - \bar{x})^2/(n-1)}} = \frac{\sqrt{n}\,(\bar{x}-\mu)}{S} = \frac{\bar{x}-\mu}{S/\sqrt{n}} \quad \ldots (***)$$

and it follows Student's t-distribution with $(n-1)$ *d.f.* (*c.f.* Remark 1 above.)

Now, (***) is same as Student's 't' defined in (16·1). Hence Student's 't' is a particular case of Fisher's 't'.

16·2·4. Constants of t-distribution. Since $f(t)$ is symmetrical about the line $t = 0$, all the moments of odd order about origin vanish, *i.e.*,

$$\mu'_{2r+1} \text{ (about origin)} = 0\,;\, r = 0, 1, 2, \ldots$$

In particular, μ_1' (about origin) $= 0 =$ Mean

Hence central moments coincide with moments about origin.

$$\therefore \quad \mu_{2r+1} = 0, \quad (r = 0, 1, 2, \ldots) \quad \ldots (16\cdot4)$$

The moments of even order are given by :

$$\mu_{2r} = \mu'_{2r} \text{ (about origin)} = \int_{-\infty}^{\infty} t^{2r} f(t)\, dt = 2\int_0^{\infty} t^{2r} f(t)\, dt$$

$$= 2 \cdot \frac{1}{B\left(\frac{1}{2}, \frac{n}{2}\right)\sqrt{n}} \int_0^{\infty} \frac{t^{2r}}{\left(1+\frac{t^2}{n}\right)^{(n+1)/2}}\, dt$$

This integral is absolutely convergent if $2r < n$.

Put $\quad 1 + \frac{t^2}{n} = \frac{1}{y} \quad \Rightarrow \quad t^2 = \frac{n(1-y)}{y} \quad \Rightarrow \quad 2t\,dt = -\frac{n}{y^2}\, dy$

When $t = 0, y = 1$ and when $t = \infty, y = 0$. Therefore,

$$\mu_{2r} = \frac{2}{\sqrt{n}\, B\left(\frac{1}{2}, \frac{n}{2}\right)} \int_1^0 \frac{t^{2r}}{(1/y)^{(n+1)/2}} \cdot \frac{-n}{2ty^2}\, dy$$

$$= \frac{n}{\sqrt{n}\, B\left(\frac{1}{2}, \frac{n}{2}\right)} \int_0^1 (t^2)^{(2r-1)/2}\, y^{[(n+1)/2]-2}\, dy$$

$$= \frac{\sqrt{n}}{B\left(\frac{1}{2}, \frac{n}{2}\right)} \int_0^1 \left[n \left(\frac{1-y}{y} \right) \right]^{r - \frac{1}{2}} y^{[(n+1)/2]-2} \, dy$$

$$= \frac{n^r}{B\left(\frac{1}{2}, \frac{n}{2}\right)} \int_0^1 y^{\frac{n}{2} - r - 1} (1-y)^{r - \frac{1}{2}} \, dy = \frac{n^r}{B\left(\frac{1}{2}, \frac{n}{2}\right)} \cdot B\left(\frac{n}{2} - r, r + \frac{1}{2}\right), n > 2r$$

$$= n^r \frac{\Gamma[(n/2) - r] \, \Gamma(r + \frac{1}{2})}{\Gamma(1/2) \, \Gamma(n/2)} \quad \ldots (16·4$$

$$= n^r \frac{(r - \frac{1}{2})(r - \frac{3}{2}) \ldots \frac{3}{2} \frac{1}{2} \Gamma(\frac{1}{2}) \Gamma(\frac{n}{2} - r)}{\Gamma(1/2) [(n/2) - 1][(n/2) - 2] \ldots [(n/2) - r] \Gamma[(n/2) - r]}$$

$$= n^r \frac{(2r-1)(2r-3)\ldots 3·1}{(n-2)(n-4)\ldots(n-2r)}, \ \frac{n}{2} > r \quad \ldots (16·4$$

In particular

$$\mu_2 = n \cdot \frac{1}{(n-2)} = \frac{n}{n-2}, \ (n > 2) \quad \ldots (16·4$$

and

$$\mu_4 = n^2 \frac{3·1}{(n-2)(n-4)} = \frac{3n^2}{(n-2)(n-4)}, \ (n > 4) \quad \ldots (16·4$$

Hence

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3} = 0 \quad \text{and} \quad \beta_2 = \frac{\mu_4}{\mu_2^2} = 3\left(\frac{n-2}{n-4}\right); \ (n > 4).$$

Remarks 1. As $n \to \infty$, $\beta_1 = 0$ and $\beta_2 = \lim\limits_{n \to \infty} 3\left(\frac{n-2}{n-4}\right) = 3 \lim\limits_{n \to \infty} \left[\frac{1-(2/n)}{1-(4/n)}\right] = 3$ ...(16·4

2. Changing r to $(r - 1)$ in [14·4(*b*)], dividing and simplifying, we shall get the recurrenc
relation for the moments as $\frac{\mu_{2r}}{\mu_{2r-2}} = \frac{n(2r-1)}{(n-2r)}, \ \frac{n}{2} > r$ ...(16·4

3. ***Moment Generating Function of t-distribution.*** From [16·4(*b*)] we observe that i
$t \sim t_n$, then all the moments of order $2r < n$ exist but the moments of order $2r \geq n$ do not exis
Hence the *m.g.f.* of *t*-distribution does not exist.

Example 16·1. *Express the constants y_0, a and m of the distribution :*

$$dF(x) = y_0 \left(1 - \frac{x^2}{a^2}\right)^m dx, \ -a \leq x \leq a \quad \ldots(*$$

in terms of its μ_2 and β_2.

Show that if x is related to a variable t by the equation :

$$x = \frac{at}{\{2(m+1) + t^2\}^{1/2}}, \quad \ldots(**$$

then t has Student's distribution with $2(m+1)$ degrees of freedom. Use the transformation t
calculate the probability that $t \geq 2$ when the degrees of freedom are 2 and also when 4.

Solution. First of all, we shall determine the constant y_0 from the consideratio
that total probability is unity.

$$\therefore \quad y_0 \int_{-a}^{a} \left(1 - \frac{x^2}{a^2}\right)^m dx = 1 \quad \Rightarrow \quad 2y_0 \int_0^a \left(1 - \frac{x^2}{a^2}\right)^m dx = 1$$

($\because$ Integrand is an even function of *x*

$$\Rightarrow \quad 2y_0 \int_0^{\pi/2} \cos^{2m}\theta \,.\, a\cos\theta\, d\theta = 1, \qquad (x = a\sin\theta)$$

$$\Rightarrow \quad 2ay_0 \int_0^{\pi/2} \cos^{2m+1}\theta\, d\theta = 1$$

But we have the Beta integral, $\quad 2\int_0^{\pi/2} \sin^p\theta \cos^q\theta\, d\theta = B\left(\frac{p+1}{2}, \frac{q+1}{2}\right)$ ...(1)

$$\therefore \quad ay_0 . 2\int_0^{\pi/2} \cos^{2m+1}\theta \sin^0\theta\, d\theta = 1 \quad \Rightarrow \quad ay_0\, B(m+1, \tfrac{1}{2}) = 1 \qquad \text{[Using (1)]}$$

$$\Rightarrow \quad y_0 = \frac{1}{a\,B\left(m+1 \cdot \frac{1}{2}\right)} \qquad ...(2)$$

Since the given probability function is symmetrical about the line $x = 0$, we have as in § 16·2·4, $\quad \mu_{2r+1} = \mu'_{2r+1} = 0\,;\, r = 0, 1, 2, \ldots$ $\quad$ [$\because$ Mean = Origin]

The moments of even order are given by :

$$\mu_{2r} = \mu_{2r}' \text{ (about origin)} = \int_{-a}^{a} x^{2r} f(x)\, dx = y_0 \int_{-a}^{a} x^{2r}\left(1 - \frac{x^2}{a^2}\right)^m dx$$

$$= 2y_0 \int_0^a x^{2r}\left(1 - \frac{x^2}{a^2}\right)^m dx = 2y_0 \int_0^{\pi/2} (a\sin\theta)^{2r} \cos^{2m}\theta .\; a\cos\theta\, d\theta, \qquad (x = a\sin\theta)$$

$$= y_0\, a^{2r+1} . \,2\int_0^{\pi/2} \sin^{2r}\theta \,.\, \cos^{2m+1}\theta\, d\theta = y_0\, a^{2r+1}\, B(r + \tfrac{1}{2}, m+1) \qquad \text{[Using (1)]}$$

$$= a^{2r}\frac{B(r+\frac{1}{2}, m+1)}{B(m+1, \frac{1}{2})} = a^{2r} . \frac{\Gamma\left(r+\frac{1}{2}\right)\Gamma\left(m+\frac{3}{2}\right)}{\Gamma\left(m+r+\frac{3}{2}\right)\Gamma\left(\frac{1}{2}\right)} \qquad ...(***)$$

In particular, $\quad \mu_2 = a^2 . \dfrac{\Gamma\{m+(3/2)\} . \frac{1}{2}\Gamma(1/2)}{\{m+(3/2)\}\,\Gamma\{m+(3/2)\}\,\Gamma(1/2)} = \dfrac{a^2}{2m+3}$

$$\Rightarrow \quad a^2 = (2m+3)\mu_2 \qquad ...(3)$$

Also $\quad \mu_4 = a^4 \dfrac{\Gamma(5/2)}{\Gamma\{m+(7/2)\}} \times \dfrac{\Gamma\{m+(3/2)\}}{\Gamma(1/2)} = \dfrac{3a^4}{(2m+5)(2m+3)}$ $\quad$ (On simplification)

$$\therefore \quad \beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{3(2m+3)}{(2m+5)} \quad \Rightarrow \quad m = \frac{9-5\beta_2}{2(\beta_2 - 3)} \qquad \text{(On simplification)} \ldots (4)$$

Equations (2), (3) and (4) express the constants y_0, a and m in terms of μ_2 and β_2.

$$x = \frac{at}{[2(m+1)+t^2]^{1/2}} \quad \Rightarrow \quad \frac{x^2}{a^2} = \frac{t^2}{2(m+1)+t^2}$$

i.e., $\quad 1 - \dfrac{x^2}{a^2} = \dfrac{2(m+1)}{2(m+1)+t^2} = \left(1 + \dfrac{t^2}{n}\right)^{-1}, \qquad (n = 2m+2)$

Also $\quad dx = a\left[\dfrac{dt}{(n+t^2)^{1/2}} - t . \dfrac{1}{2}\dfrac{2t\,dt}{(n+t^2)^{3/2}}\right] = a\dfrac{1}{(n+t^2)^{1/2}}\left(1 - \dfrac{t^2}{n+t^2}\right)dt$

$$= \frac{an}{(n+t^2)^{3/2}}\,dt = \frac{a}{\sqrt{n}} \cdot \frac{dt}{[1 + (t^2/n)]^{3/2}}$$

$$= 1 - \frac{1}{2B\left(\frac{1}{2}, \frac{n}{2}\right)} \int_0^x z^{(n/2)-1}(1-z)^{(1/2)-1}\,dz \qquad \left[\text{where } x = \left(1 + \frac{t^2}{n}\right)^{-1}\right]$$

$$= 1 - \frac{1}{2} I_x\left(\frac{n}{2}, \frac{1}{2}\right), \qquad \left[x = \left(1 + \frac{t^2}{n}\right)^{-1}\right]$$

where $I_x(p, q)$ is defined in (*).

Example 16·4. *Show that for t-distribution with n d.f., mean deviation about mean is given by :*

$$\sqrt{n}\,\Gamma[(n-1)/2] \,/\, \sqrt{\pi}\,\Gamma(n/2)$$

Solution. $E(t) = 0$.

$$\text{M.D. (about mean)} = \int_{-\infty}^{\infty} |\,t\,|\,f(t)\,dt = \frac{1}{\sqrt{n}\,B\left(\frac{1}{2}, \frac{n}{2}\right)} \int_{-\infty}^{\infty} \frac{|\,t\,|\,dt}{\left(1 + \frac{t^2}{n}\right)^{(n+1)/2}}$$

$$= \frac{2}{\sqrt{n}\,B\left(\frac{1}{2}, \frac{n}{2}\right)} \int_0^{\infty} \frac{t\,dt}{\left(1 + \frac{t^2}{n}\right)^{(n+1)/2}} = \frac{\sqrt{n}}{B\left(\frac{1}{2}, \frac{n}{2}\right)} \int_0^{\infty} \frac{dy}{(1+y)^{(n+1)/2}}, \qquad \left(\frac{t^2}{n} = y\right)$$

$$= \frac{\sqrt{n}}{B\left(\frac{1}{2}, \frac{n}{2}\right)} \int_0^{\infty} \frac{y^{1-1}}{(1+y)^{\frac{n-1}{2}+1}}\,dy = \frac{\sqrt{n}}{B\left(\frac{1}{2}, \frac{n}{2}\right)} \cdot B\left(\frac{n-1}{2}, 1\right) = \frac{\sqrt{n}\,\Gamma[(n-1)/2]}{\sqrt{\pi}\,\Gamma(n/2)}$$

16·2·5. Limiting Form of t-distribution. *As $n \to \infty$, the p.d.f. of t-distribution with n d.f. viz.,*

$$f(t) = \frac{1}{\sqrt{n}\,B\left(\frac{1}{2}, \frac{n}{2}\right)} \left(1 + \frac{t^2}{n}\right)^{-(n+1)/2} \to \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}t^2\right), -\infty < t < \infty$$

Proof.
$$\lim_{n \to \infty} \frac{1}{\sqrt{n}\,B\left(\frac{1}{2}, \frac{n}{2}\right)} = \lim_{n \to \infty} \frac{1}{\sqrt{n}} \frac{\Gamma[(n+1)/2]}{\Gamma(1/2)\,\Gamma(n/2)} = \frac{1}{\sqrt{n}} \cdot \frac{1}{\sqrt{\pi}} \left(\frac{n}{2}\right)^{\frac{1}{2}} = \frac{1}{\sqrt{2\pi}}$$

$$\left[\because \Gamma(1/2) = \sqrt{\pi} \text{ and } \lim_{n \to \infty} \frac{\Gamma(n+k)}{\Gamma(n)} = n^k, \text{ (c.f. Remark to § 16·8)}\right]$$

$$\therefore \quad \lim_{n \to \infty} f(t) = \lim_{n \to \infty} \frac{1}{\sqrt{n}\,B\left(\frac{1}{2}, \frac{n}{2}\right)} \cdot \lim_{n \to \infty} \left[\left(1 + \frac{t^2}{n}\right)^n\right]^{-\frac{1}{2}} \times \lim_{n \to \infty} \left(1 + \frac{t^2}{n}\right)^{-\frac{1}{2}}$$

$$= \frac{1}{\sqrt{2\pi}} \exp(-t^2/2), \; -\infty < t < \infty$$

Hence for large d.f. t-distribution tends to standard normal distribution.

16·2·6. Graph of t-distribution. The *p.d.f.* of t-distribution with n *d.f.* is :

$$f(t) = C.\left(1 + \frac{t^2}{n}\right)^{-(n+1)/2}, -\infty < t < \infty$$

Since $f(-t) = f(t)$, the probability curve is symmetrical about the line $t = 0$. As t increases, $f(t)$ decreases rapidly and tends to zero as $t \to \infty$, so that t-axis is an asymptote to the curve. We have shown that

$$\mu_2 = \frac{n}{n-2}, n > 2 \;;\; \beta_2 = \frac{3(n-2)}{(n-4)}, n > 4$$

Hence for $n > 2$, $\mu_2 > 1$ *i.e.*, the variance of t-distribution is greater than that of standard normal distribution and for $n > 4$, $\beta_2 > 3$ and thus t-distribution is more flat on the top than the normal curve. In fact, for small n, we have

$$P(|t| \geq t_0) \geq P(|Z| \geq t_0), \quad Z \sim N(0, 1)$$

i.e., the tails of the t-distribution have a greater probability (area) than the tails of standard normal distribution. Moreover we have also seen [§ 16·2·5], that for large n (*d.f.*), t-distribution tends to standard normal distribution.

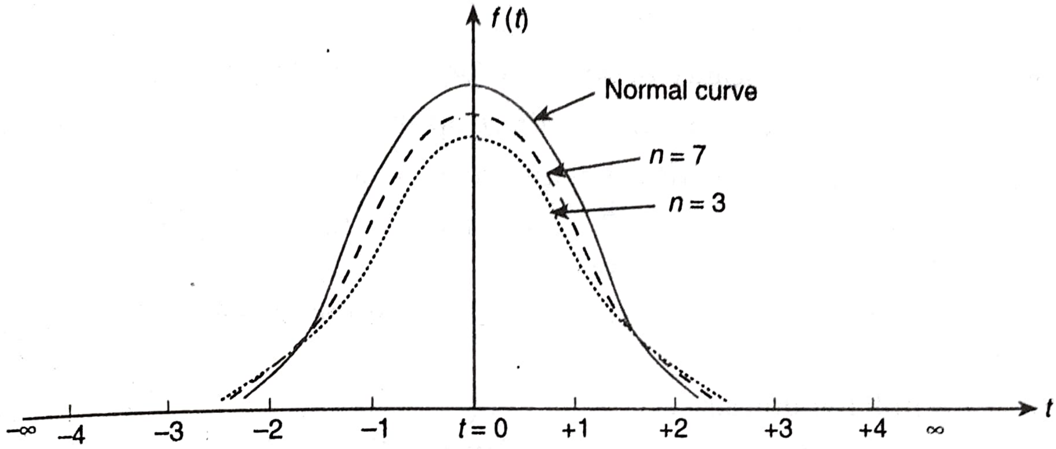

Fig. 16·1 :Graph of t-distribution

16·2·7. Critical Values of t. The critical (or significant) values of t at level of significance α and *d.f.* v for *two-tailed test* are given by the equation :

$$P[|t| > t_v(\alpha)] = \alpha \qquad \ldots(16\cdot5)$$

$$\Rightarrow \quad P[|t| \leq t_v(\alpha)] = 1 - \alpha \qquad \ldots(16\cdot5a)$$

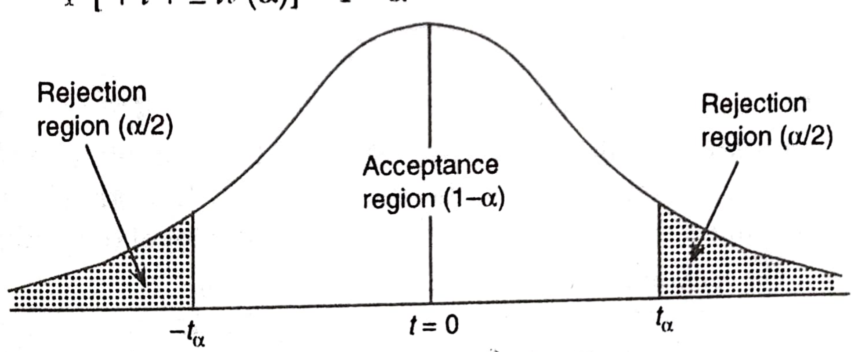

Fig. 16·2 : Critical values of t-distribution

The values $t_v(\alpha)$ have been tabulated in Fisher and Yates' Tables, for different values of α and v and are given in *Table I* at the end of the chapter.

Since t-distribution is symmetric about $t = 0$, we get from (16·5)

$$P(t > t_v(\alpha)] + P[t < -t_v(\alpha)] = \alpha \quad \Rightarrow \quad 2P[t > t_v(\alpha)] = \alpha$$

$$\Rightarrow \quad P[t > t_v(\alpha)] = \alpha/2 \qquad \therefore \quad P[t > t_v(2\alpha)] = \alpha \qquad \ldots(16\cdot5b)$$

$t_v(2\alpha)$ (from the Tables at the end of the chapter) gives the significant value of t for a single-tail test [Right-tail or Left-tail-since the distribution is symmetrical], at level of significance α and v *d.f.*

Hence the significant values of t at level of significance 'α' for a single-tailed test can be obtained from those of two-tailed test by looking the values at level of significance 2α.

For example,

$t_8(0\cdot05)$ for single-tail test $= t_8(0\cdot10)$ for two-tail test $= 1\cdot86$

$t_{15}(0\cdot01)$ for single-tail test $= t_{15}(0\cdot02)$ for two-tail test $= 2\cdot60$.

16·3. APPLICATIONS OF t-DISTRIBUTION

The t-distribution has a wide number of applications in Statistics, some of which are enumerated below.

(i) To test if the sample mean ($\bar{x}$) differs significantly from the hypothetical value μ of the population mean :

(ii) To test the significance of the difference between two sample means.

(iii) To test the significance of an observed sample correlation coefficient and sample regression coefficient.

(iv) To test the significance of observed partial correlation coefficient.

In the following sections we will discuss these applications in detail, one by one.

16·3·1. t-Test for Single Mean. Suppose we want to test :

(i) if a random sample x_i ($i = 1, 2, \ldots, n$) of size n has been drawn from a normal population with a specified mean, say μ_0, or

(ii) if the sample mean differs significantly from the hypothetical value μ_0 of the population mean.

Under the null hypothesis, H_0 :

(i) *The sample has been drawn from the population with mean μ_0 or*

(ii) *there is no significant difference between the sample mean $\bar{x}$ and the population mean μ_0,*

the statistic
$$t = \frac{\bar{x} - \mu_0}{S/\sqrt{n}}, \qquad \ldots(16\cdot6)$$

where
$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \quad \text{and} \quad S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2, \qquad \ldots[16\cdot6(a)]$$

follows Student's t-distribution with $(n - 1)$ $d.f.$

We now compare the calculated value of t with the tabulated value at certain level of significance. If calculated $|t|$ > tabulated t, null hypothesis is rejected and if calculated $|t|$ < tabulated t, H_0 may be accepted at the level of significance adopted.

Remarks 1. ***On computation of S^2 for numerical problems.*** If $\bar{x}$ comes out in integers, the formula (16·6a) can be conveniently used for computing S^2. However, if $\bar{x}$ comes in fractions then the formula (16·6a) for computing S^2 is very cumbersome and is not recommended. In that case, step deviation method, given below, is quite useful.

If we take $d_i = x_i - A$, where A is any arbitrary number, then

$$S^2 = \frac{1}{n-1}\left[\Sigma(x_i - \bar{x})^2\right] = \frac{1}{n-1}\left[\Sigma x_i^2 - \frac{(\Sigma x_i)^2}{n}\right] \qquad \ldots(16\cdot6b)$$

$$= \frac{1}{n-1}\left[\Sigma d_i^2 - \frac{(\Sigma d_i)^2}{n}\right], \text{ since variance is independent of change of origin.} \qquad \ldots(16\cdot6c)$$

Also, in this case
$$\bar{x} = A + \frac{\Sigma d_i}{n}. \qquad \ldots(16\cdot6d)$$

2. We know, the sample variance : $s^2 = \frac{1}{n}\sum_i(x_i - \bar{x})^2 \Rightarrow ns^2 = (n - 1)\,S^2$

$$\therefore \quad \frac{S^2}{n} = \frac{s^2}{n-1} \qquad \ldots(16\cdot6e)$$

Hence for numerical problems test statistic (16·6) on using [16·6(e)] becomes

$$t = \frac{\bar{x} - \mu_0}{\sqrt{s^2/(n-1)}} \sim t_{n-1} \qquad \ldots(16\cdot6f)$$

3. Assumption for Student's t-test. The following assumptions are made in the Student's t-test :

(i) The parent population from which the sample is drawn is normal.
(ii) The sample observations are independent, *i.e.*, the sample is random.
(iii) The population standard deviation σ is unknown.

Example 16·5. *A machinist is making engine parts with axle diameters of 0·700 inch. A random sample of 10 parts shows a mean diameter of 0·742 inch with a standard deviation of 0·040 inch. Compute the statistic you would use to test whether the work is meeting the specifications. Also state how you would proceed further.*

Solution. Here we are given :

$$\mu = 0{\cdot}700 \text{ inche}, \quad \bar{x} = 0{\cdot}742 \text{ inche}, \quad s = 0{\cdot}040 \text{ inche} \quad \text{and} \quad n = 10$$

Null Hypothesis, $H_0 : \mu = 0{\cdot}700$, *i.e., the product is conforming to specifications.*

Alternative Hypothesis, $H_1 : \mu \neq 0{\cdot}700$

Test Statistic. Under H_0, the test statistic is : $t = \dfrac{\bar{x} - \mu}{\sqrt{S^2/n}} = \dfrac{\bar{x} - \mu}{\sqrt{s^2/(n-1)}} \sim t_{(n-1)}$

$$\therefore \quad t = \frac{\sqrt{9}\,(0{\cdot}742 - 0{\cdot}700)}{0{\cdot}040} = 3{\cdot}15$$

How to proceed further. Here the test statistic 't' follows Student's t-distribution with $10 - 1 = 9$ *d.f.* We will now compare this calculated value with the tabulated value of t for 9 *d.f.* and at certain level of significance, say 5%. Let this tabulated value be denoted by t_0.

(i) If calculated 't', *viz.*, $3{\cdot}15 > t_0$, we say that the value of t is significant. This implies that $\bar{x}$ differs significantly from μ and H_0 is rejected at this level of significance and we conclude that the product is not meeting the specifications.

(ii) If calculated $t < t_0$, we say that the value of t is not significant, *i.e.*, there is no significant difference between $\bar{x}$ and μ. In other words, the deviation $(\bar{x} - \mu)$ is just due to fluctuations of sampling and null hypothesis H_0 may be retained at 5% level of significance, *i.e.*, we may take the product conforming to specifications.

Example 16·6. *The mean weekly sales of soap bars in departmental stores was 146·3 bars per store. After an advertising campaign the mean weekly sales in 22 stores for a typical week increased to 153·7 and showed a standard deviation of 17·2. Was the advertising campaign successful ?*

Solution. We are given : $n = 22, \bar{x} = 153{\cdot}7, s = 17{\cdot}2$.

Null Hypothesis. The advertising campaign is not successful, *i.e.*, $H_0 : \mu = 146{\cdot}3$

Alternative Hypothesis, $H_1 : \mu > 146{\cdot}3$ (Right-tail).

Test Statistic. Under H_0, the test statistic is : $t = \dfrac{\bar{x} - \mu}{\sqrt{s^2/(n-1)}} \sim t_{22-1} = t_{21}$

$$\therefore \quad t = \frac{153{\cdot}7 - 146{\cdot}3}{\sqrt{(17{\cdot}2)^2/21}} = \frac{7{\cdot}4 \times \sqrt{21}}{17{\cdot}2} = 9{\cdot}03$$

Conclusion. Tabulated value of t for 21 *d.f.* at 5% level of significance for single-tailed test is 1·72. Since calculated value is much greater than the tabulated value, it is

3. Assumption for Student's t-test. The following assumptions are made in the Student's t-test :

(i) The parent population from which the sample is drawn is normal.
(ii) The sample observations are independent, *i.e.*, the sample is random.
(iii) The population standard deviation σ is unknown.

Example 16·5. *A machinist is making engine parts with axle diameters of 0·700 inch. A random sample of 10 parts shows a mean diameter of 0·742 inch with a standard deviation of 0·040 inch. Compute the statistic you would use to test whether the work is meeting the specifications. Also state how you would proceed further.*

Solution. Here we are given :

$$\mu = 0{\cdot}700 \text{ inche}, \quad \bar{x} = 0{\cdot}742 \text{ inche}, \quad s = 0{\cdot}040 \text{ inche} \quad \text{and} \quad n = 10$$

Null Hypothesis, $H_0 : \mu = 0{\cdot}700$, *i.e., the product is conforming to specifications.*

Alternative Hypothesis, $H_1 : \mu \neq 0{\cdot}700$

Test Statistic. Under H_0, the test statistic is : $t = \dfrac{\bar{x} - \mu}{\sqrt{S^2/n}} = \dfrac{\bar{x} - \mu}{\sqrt{s^2/(n-1)}} \sim t_{(n-1)}$

$$\therefore \qquad t = \frac{\sqrt{9}\,(0{\cdot}742 - 0{\cdot}700)}{0{\cdot}040} = 3{\cdot}15$$

How to proceed further. Here the test statistic 't' follows Student's t-distribution with $10 - 1 = 9$ *d.f.* We will now compare this calculated value with the tabulated value of t for 9 *d.f.* and at certain level of significance, say 5%. Let this tabulated value be denoted by t_0.

(*i*) If calculated 't', *viz.*, $3{\cdot}15 > t_0$, we say that the value of t is significant. This implies that $\bar{x}$ differs significantly from μ and H_0 is rejected at this level of significance and we conclude that the product is not meeting the specifications.

(*ii*) If calculated $t < t_0$, we say that the value of t is not significant, *i.e.*, there is no significant difference between $\bar{x}$ and μ. In other words, the deviation $(\bar{x} - \mu)$ is just due to fluctuations of sampling and null hypothesis H_0 may be retained at 5% level of significance, *i.e.*, we may take the product conforming to specifications.

Example 16·6. *The mean weekly sales of soap bars in departmental stores was 146·3 bars per store. After an advertising campaign the mean weekly sales in 22 stores for a typical week increased to 153·7 and showed a standard deviation of 17·2. Was the advertising campaign successful ?*

Solution. We are given : $n = 22$, $\bar{x} = 153{\cdot}7$, $s = 17{\cdot}2$.

Null Hypothesis. The advertising campaign is not successful, *i.e.*, $H_0 : \mu = 146{\cdot}3$

Alternative Hypothesis, $H_1 : \mu > 146{\cdot}3$ (Right-tail).

Test Statistic. Under H_0, the test statistic is : $t = \dfrac{\bar{x} - \mu}{\sqrt{s^2/(n-1)}} \sim t_{22-1} = t_{21}$

$$\therefore \qquad t = \frac{153{\cdot}7 - 146{\cdot}3}{\sqrt{(17{\cdot}2)^2/21}} = \frac{7{\cdot}4 \times \sqrt{21}}{17{\cdot}2} = 9{\cdot}03$$

Conclusion. Tabulated value of t for 21 *d.f.* at 5% level of significance for *single-tailed test* is 1·72. Since calculated value is much greater than the tabulated value, it is highly significant. Hence we reject the null hypothesis and conclude that the advertising campaign was definitely successful in promoting sales.

Example 16·7. *A random sample of 10 boys had the following I.Q.'s : 70, 120, 110, 101, 88, 83, 95, 98, 107, 100. Do these data support the assumption of a population mean I.Q. of 100 ? Find a reasonable range in which most of the mean I.Q. values of samples of 10 boys lie.*

Solution. Null hypothesis, H_0 : *The data are consistent with the assumption of a mean I.Q. of 100 in the population, i.e.,* $\mu = 100$.

Alternative hypothesis, $H_1 : \mu \neq 100$.

Test Statistic. Under H_0, the test statistic is : $t = \dfrac{(\bar{x} - \mu)}{\sqrt{S^2/n}} \sim t_{(n-1)}$,

where $\bar{x}$ and S^2 are to be computed from the sample values of I.Q.'s.

TABLE 16·1 : CALCULATIONS FOR SAMPLE MEAN AND S.D.

x	$(x - \bar{x})$	$(x - \bar{x})^2$
70	− 27·2	739·84
120	22·8	519·84
110	12·8	163·84
101	3·8	14·44
88	− 9·2	84·64
83	− 14·2	201·64
95	− 2·2	4·84
98	0·8	0·64
107	9·8	96·04
100	2·8	7·84
Total 972		1833·60

Here $n = 10$, $\bar{x} = \dfrac{972}{10} = 97·2$ and $S^2 = \dfrac{1833·60}{9} = 203·73$

$$\therefore \quad |t| = \frac{|97·2 - 100|}{\sqrt{203·73/10}} = \frac{2·8}{\sqrt{20·37}} = \frac{2·8}{4·514} = 0·62$$

Tabulated $t_{0·05}$ for (10 − 1), *i.e.,* 9 *d.f.* for two-tailed test is 2·262.

Conclusion. Since calculated t is less than tabulated $t_{0·05}$ for 9 *d.f.*, H_0 may be accepted at 5% level of significance and we may conclude that the data are consistent with the assumption of mean I.Q. of 100 in the population.

The 95% confidence limits within which the mean I.Q. values of samples of 10 boys will lie are given by :

$$\bar{x} \pm t_{0·05}\, S/\sqrt{n} = 97·2 \pm 2·262 \times 4·514 = 97·2 \pm 10·21 = 107·41 \text{ and } 86·99$$

Hence the required 95% confidence interval is [86·99, 107·41].

Remark. Aliter for computing $\bar{x}$ ***and*** S^2. Here we see that $\bar{x}$ comes in fractions and as such the computation of $(x - \bar{x})^2$ is quite laborious and time consuming. In this case we use the method of step deviations to compute $\bar{x}$ and S^2, as given below.

x	$d = x - 90$	d^2
70	−20	400
120	30	900
110	20	400
101	11	121
88	−2	4
83	−7	49
95	5	25
98	8	64
107	17	289
100	10	100
Total	$\Sigma d = 72$	$\Sigma d^2 = 2{,}352$

Here $d = x - A$, where $A = 90$. Therefore

$$\bar{x} = A + \frac{1}{n}\Sigma d = 90 + \frac{72}{10} = 97{\cdot}2 \text{ and } S^2 = \frac{1}{n-1}\left[\Sigma d^2 - \frac{(\Sigma d)^2}{n}\right] = \frac{1}{9}\left[2352 - \frac{(72)^2}{10}\right] = 203{\cdot}73.$$

Example 16·8. *The heights of 10 males of a given locality are found to be 70, 67, 62, 68, 61, 68, 70, 64, 64, 66 inches. Is it reasonable to believe that the average height is greater than 64 inches ? Test at 5% significance level assuming that for 9 degrees of freedom $P(t > 1{\cdot}83) = 0{\cdot}05$.*

Solution. *Null Hypothesis,* $H_0 : \mu = 64$ inches

Alternative Hypothesis, $H_1 : \mu > 64$ inches

TABLE 16·2 : CALCULATIONS FOR SAMPLE MEAN AND S.D.

x	70	67	62	68	61	68	70	64	64	66	Total 660
$x - \bar{x}$	4	1	−4	2	−5	2	4	−2	−2	0	0
$(x - \bar{x})^2$	16	1	16	4	25	4	16	4	4	0	90

$$\bar{x} = \frac{\Sigma x}{n} = \frac{660}{10} = 66; \quad S^2 = \frac{1}{n-1}\Sigma(x - \bar{x})^2 = \frac{90}{9} = 10$$

Test Statistic. Under H_0, the test statistic is :

$$t = \frac{\bar{x} - \mu}{\sqrt{S^2/n}} = \frac{66 - 64}{\sqrt{10/10}} = 2,$$

which follows Student's t-distribution with $10 - 1 = 9$ *d.f.*

Tabulated value of t for 9 *d.f.* at 5% level of significance for single (right) tail-test is 1·833. (This is the value $t_{0{\cdot}10}$ for 9 *d.f.* in the two-tailed tables given at the end of the chapter.)

Conclusion. Since calculated value of t is greater than the tabulated value, it is significant. Hence H_0 is rejected at 5% level of significance and we conclude that the average height is greater than 60 inches.

Example 16·9. *A random sample of 16 values from a normal population showed a mean of 41·5 inches and the sum of squares of deviations from this mean equal to 135 square inches. Show that the assumption of a mean of 43·5 inches for the population is not reasonable. Obtain 95 per cent and 99 per cent fiducial limits for the same.*

You may use the following information from statistical tables :

$$v = 15, \begin{cases} P = 0{\cdot}05, t = 2{\cdot}131 \\ P = 0{\cdot}01, t = 2{\cdot}947 \end{cases}$$

Solution. We are given $n = 16, \bar{x} = 41{\cdot}5$ inches and $\Sigma(x - \bar{x})^2 = 135$ sq. inches.

$$\therefore \quad S^2 = \frac{1}{n-1}\Sigma(x-\bar{x})^2 = \frac{135}{15} = 9 \quad \Rightarrow \quad S = 3$$

Null Hypothesis, $H_0 : \mu = 43{\cdot}5$ inches, *i.e.*, the data are consistent with the assumption that the mean height in the population is 43·5 inches.

Alternative Hypothesis, $H_1 : \mu \neq 43{\cdot}5$ inches.

Test Statistic. Under H_0, the test statistic is : $t = \dfrac{\bar{x} - \mu}{S/\sqrt{n}} \sim t_{(n-1)}$

$$\therefore \quad |t| = \frac{|41{\cdot}5 - 43{\cdot}5|}{3/4} = \frac{8}{3} = 2{\cdot}667$$

Here number of degrees of freedom is $(16 - 1) = 15$.

We are given : $t_{0{\cdot}05}$ for 15 *d.f.* = 2·131 and $t_{0{\cdot}01}$ for 15 *d.f.* = 2·947.

Conclusion. Since calculated $|t|$ is greater than 2·131, null hypothesis is rejected at 5% level of significance and we conclude that the assumption of mean of 43·5 inches for the population is not reasonable.

Remark. Since calculated $|t|$ is less than 2·947, null hypothesis ($\mu = 43{\cdot}5$) may be accepted at 1% level of significance.

95% fiducial limits for μ : (*d.f.* = 15)

$$\bar{x} \pm t_{0{\cdot}05} \times \frac{S}{\sqrt{n}} = 41{\cdot}5 \pm 2{\cdot}131 \times \frac{3}{4} = 41{\cdot}5 \pm 1{\cdot}598 \quad \Rightarrow \quad 39{\cdot}902 < \mu < 43{\cdot}098$$

99% fiducial limits for μ : (*d.f.* = 15)

$$\bar{x} \pm t_{0{\cdot}01} \times \frac{S}{\sqrt{n}} = 41{\cdot}5 \pm 2{\cdot}947 \times \frac{3}{4} = 43{\cdot}71 \text{ and } 39{\cdot}29 \quad \Rightarrow \quad 39{\cdot}29 < \mu < 43{\cdot}71$$

16·3·2. t-Test for Difference of Means. Suppose we want to test if two independent samples x_i $(i = 1, 2, \ldots, n_1)$ and y_j, $(j = 1, 2, \ldots, n_2)$ of sizes n_1 and n_2 have been drawn from two normal populations with means μ_X and μ_Y respectively.

Under the null hypothesis (H_0) *that the samples have been drawn from the normal populations with means* μ_X *and* μ_Y *and under the assumption that the population variance are equal, i.e.,* $\sigma_X^2 = \sigma_Y^2 = \sigma^2$ (say), the statistic

$$t = \frac{(\bar{x} - \bar{y}) - (\mu_X - \mu_Y)}{S\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \qquad \ldots(16{\cdot}7)$$

where $\quad \bar{x} = \dfrac{1}{n_1}\sum_{i=1}^{n_1} x_i, \quad \bar{y} = \dfrac{1}{n_2}\sum_{j=1}^{n_2} y_j$

and $\quad S^2 = \dfrac{1}{n_1 + n_2 - 2}\left[\sum_i (x_i - \bar{x})^2 + \sum_j (y_j - \bar{y})^2\right] \qquad \ldots(16{\cdot}7a)$

is an unbiased estimate of the common population variance σ^2, follows Student's *t*-distribution with $(n_1 + n_2 - 2)$ *d.f.*

Proof. *Distribution of t defined in (16·7).*

$$\xi = \frac{(\bar{x}-\bar{y})-E(\bar{x}-\bar{y})}{\sqrt{V(\bar{x}-\bar{y})}} \sim N(0, 1)$$

But $\quad E(\bar{x}-\bar{y}) = E(\bar{x})-E(\bar{y}) = \mu_X - \mu_Y$

$$V(\bar{x}-\bar{y}) = V(\bar{x})+V(\bar{y}) = \frac{\sigma_X^2}{n_1}+\frac{\sigma_Y^2}{n_2} = \sigma^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right) \qquad \text{(By assumption)}$$

[The covariance term vanishes since samples are independent.]

$$\therefore \quad \xi = \frac{(\bar{x}-\bar{y})-(\mu_X-\mu_Y)}{\sqrt{\sigma^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} \sim N(0, 1) \qquad \ldots(*)$$

Let
$$\chi^2 = \frac{1}{\sigma^2}\left[\sum_{i=1}^{n_1}(x_i-\bar{x})^2+\sum_{j=1}^{n_2}(y_j-\bar{y})^2\right]$$

$$= \left[\sum_i (x_i-\bar{x})^2/\sigma^2\right]+\left[\sum_j (y_j-\bar{y})^2/\sigma^2\right] = \frac{n_1 s_X^2}{\sigma^2}+\frac{n_2 s_Y^2}{\sigma^2} \qquad \ldots(**)$$

Since $n_1 s_X^2/\sigma^2$ and $n_2 s_Y^2/\sigma^2$ are independent χ^2-variates with (n_1-1) and (n_2-1) *d.f.* respectively, by the additive property of chi-square distribution, χ^2 defined in (**) is a χ^2-variate with $(n_1-1)+(n_2-1)$, *i.e.*, n_1+n_2-2 *d.f.* Further, since sample mean and sample variance are independently distributed, ξ and χ^2 are independent random variables. Hence Fisher's t statistic is given by

$$t = \frac{\xi}{\sqrt{\dfrac{\chi^2}{n_1+n_2-2}}}$$

$$= \frac{(\bar{x}-\bar{y})-(\mu_X-\mu_Y)}{\sqrt{\sigma^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} \times \frac{1}{\left[\dfrac{1}{n_1+n_2-2}\left\{\sum_i (x_i-\bar{x})^2+\sum_j (y_j-\bar{y})^2\right\}/\sigma^2\right]^{1/2}}$$

$$= \frac{(\bar{x}-\bar{y})-(\mu_X-\mu_Y)}{S\sqrt{\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}, \text{ where } S^2 = \frac{1}{n_1+n_2-2}\left[\sum_i (x_i-\bar{x})^2+\sum_j (y_j-\bar{y})^2\right]$$

and it follows Student's t-distribution with (n_1+n_2-2) *d.f.* (*c.f.* Remark 1, § 16·2·3).

Remarks 1. S^2, defined in (16·7*a*) is an unbiased estimate of the common population variance σ^2, since

$$E(S^2) = \frac{1}{n_1+n_2-2}E\left[\sum_i (x_i-\bar{x})^2+\sum_j (y_j-\bar{y})^2\right] = \frac{1}{n_1+n_2-2}E\left[(n_1-1)S_X^2+(n_2-1)S_Y^2\right]$$

$$= \frac{1}{n_1+n_2-2}\left[(n_1-1)E(S_X^2)+(n_2-1)E(S_Y^2)\right] = \frac{1}{n_1+n_2-2}\left[(n_1-1)\sigma^2+(n_2-1)\sigma^2\right] = \sigma^2$$

2. An important deduction which is of much practical utility is discussed below :

Suppose we want to test if : (*a*) two independent samples x_i $(i = 1, 2, \ldots, n_1)$, and y_j $(j = 1, 2, \ldots, n_2)$, have been drawn from the populations with same means, or (*b*) the two sample means $\bar{x}$ and $\bar{y}$ differ significantly or not.

Under the null hypothesis, H_0 that *(a) samples have been drawn from the populations with same means, i.e., $\mu_X = \mu_Y$, or (b) the sample means $\bar{x}$ and $\bar{y}$ do not differ significantly,* the statistic :

$$t = \frac{\bar{x} - \bar{y}}{S\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \qquad [\because \mu_X = \mu_Y, \text{ under } H_0] \qquad \ldots(16\cdot8)$$

where symbols are defined in (16·7*a*), follows Student's t-distribution with $(n_1 + n_2 - 2)$ *d.f.*

3. *On the assumption of t-test for difference of means.* Here we make the following three fundamental assumptions :

(*i*) Parent populations, from which the samples have been drawn are normally distributed.

(*ii*) The population variances are equal and unknown, *i.e.*, $\sigma_X^2 = \sigma_Y^2 = \sigma^2$ (say), where σ^2 is unknown.

(*iii*) The two samples are random and independent of each other.

Thus before applying t-test for testing the equality of means it is theoretically desirable to test the equality of population variances by applying F-test. (*c.f* § 16·6·1) If the variances do not come out to be equal then t-test becomes invalid and in that case Behren's '*d*'-test based on fiducial intervals is used. For practical problems, however, the assumptions (*i*) and (*ii*) are taken for granted.

16·3·3. Paired t-test for Difference of Means. Let us now consider the case when (*i*) the sample sizes are equal, *i.e.*, $n_1 = n_2 = n$ (say), and (*ii*) the two samples are not independent but the sample observations are paired together, *i.e.*, the pair of observations (x_i, y_i), $(i = 1, 2, \ldots, n)$ corresponds to the same (*i*th) sample unit. The problem is to test if the sample means differ significantly or not.

For example, suppose we want to test the efficacy of a particular drug, say, for inducing sleep. Let x_i and y_i $(i = 1, 2, \ldots, n)$ be the readings, in hours of sleep, on the ith individual, before and after the drug is given respectively. Here instead of applying the difference of the means test discussed in § 16·3·2, we apply the paired t-test given below.

Here we consider the increments, $d_i = x_i - y_i$, $(i = 1, 2, \ldots, n)$.

Under the null hypothesis, H_0 that *increments are due to fluctuations of sampling, i.e., the drug is not responsible for these increments,* the statistic :

$$t = \frac{\bar{d}}{S/\sqrt{n}} \qquad \ldots(16\cdot9)$$

where

$$\bar{d} = \frac{1}{n}\sum_{i=1}^{n} d_i \text{ and } S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(d_i - \bar{d})^2 \qquad \ldots(16\cdot9a)$$

follows Student's t-distribution with $(n - 1)$ *d.f.*

Example 16·10. *Below are given the gain in weights (in kgs.) of pigs fed on two diets A and B.*

Gain in weight

Diet A : 25, 32, 30, 34, 24, 14, 32, 24, 30, 31, 35, 25

Diet B : 44, 34, 22, 10, 47, 31, 40, 30, 32, 35, 18, 21, 35, 29, 22

Test, if the two diets differ significantly as regards their effect on increase in weight.

Solution. *Null hypothesis, $H_0 : \mu_X = \mu_Y$, i.e., there is no significant difference between the mean increase in weight due to diets A and B.*

Alternative hypothesis, H_1: $\mu_X \neq \mu_Y$ (two-tailed).

Diet A			Diet B		
x	$x-\bar{x}$	$(x-\bar{x})^2$	y	$y-\bar{y}$	$(y-\bar{y})^2$
25	−3	9	44	14	196
32	4	16	34	4	16
30	2	4	22	−8	64
34	6	36	10	−20	400
24	−4	16	47	17	289
14	−14	196	31	1	1
32	4	16	40	10	100
24	−4	16	30	0	0
30	2	4	32	2	4
31	3	9	35	5	25
35	7	49	18	−12	144
25	−3	9	21	−9	81
			35	5	25
			29	−1	1
			22	−8	64
$\Sigma x = 336$	$\Sigma(x-\bar{x}) = 0$	$\Sigma(x-\bar{x})^2 = 380$	$\Sigma y = 450$	$\Sigma(y-\bar{y}) = 0$	$\Sigma(y-\bar{y})^2 = 1{,}410$

$$\bar{x} = \frac{336}{12} = 28,\ \bar{y} = \frac{450}{15} = 30,\ S^2 = \frac{1}{n_1 + n_2 - 2}\left[\Sigma(x-\bar{x})^2 + \Sigma(y-\bar{y})^2\right] = 71{\cdot}6$$

and $n_1 = 12$, $n_2 = 15$

Under null hypothesis (H_0) :

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{S^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \sim t_{n_1 + n_2 - 2}$$

$$\therefore \qquad t = \frac{28 - 30}{\sqrt{71{\cdot}6\left(\frac{1}{12} + \frac{1}{15}\right)}} = \frac{-2}{\sqrt{10{\cdot}74}} = -0{\cdot}609$$

Tabulated $t_{0{\cdot}05}$ for $(12 + 15 - 2) = 25$ *d.f.* is 2·06.

Conclusion. Since calculated $|t|$ is less than tabulated t, H_0 may be accepted at 5% level of significance and we may conclude that the two diets do not differ significantly as regards their effect on increase in weight.

Remark. Here $\bar{x}$ and $\bar{y}$ come out to be integral values and hence the direct method of computing $\Sigma(x-\bar{x})^2$ and $\Sigma(y-\bar{y})^2$ is used. In case $\bar{x}$ and (or) $\bar{y}$ comes out to be fractional, then the step deviation method is recommended for computation of $\Sigma(x-\bar{x})^2$ and $\Sigma(y-\bar{y})^2$.

Example 16·11. *Samples of two types of electric light bulbs were tested for length of life and following data were obtained :*

	Type I	*Type II*
Sample No.	$n_1 = 8$	$n_2 = 7$
Sample Means	$\bar{x}_1 = 1{,}234$ *hrs.*	$\bar{x}_2 = 1{,}036$ *hrs.*
Sample S.D.'s	$s_1 = 36$ *hrs.*	$s_2 = 40$ *hrs.*

Is the difference in the means sufficient to warrant that type I is superior to type II regarding length of life ?

Solution. *Null Hypothesis,* $H_0 : \mu_X = \mu_Y$, *i.e.,* the two types I and II of electric bulbs are indentical.

Alternative Hypothesis, $H_1 : \mu_X > \mu_Y$, *i.e.,* type I is superior to type II.

Test Statistic. Under H_0, the test statistic is :

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{S^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \sim t_{n_1 + n_2 - 2} = t_{13},$$

where
$$S^2 = \frac{1}{n_1 + n_2 - 2}\left[\Sigma(x_1 - \bar{x}_1)^2 + \Sigma(x_2 - \bar{x}_2)^2\right]$$
$$= \frac{1}{n_1 + n_2 - 2}\left(n_1 s_1^2 + n_2 s_2^2\right) = \frac{1}{13}\left[8 \times (36)^2 + 7 \times (40)^2\right] = 1,659{\cdot}08$$

$$\therefore \quad t = \frac{1234 - 1036}{\sqrt{1659{\cdot}08\left(\frac{1}{8} + \frac{1}{7}\right)}} = \frac{198}{\sqrt{1659{\cdot}08 \times 0{\cdot}2679}} = 9{\cdot}39$$

Tabulated value of t for 13 *d.f.* at 5% level of significance for right (single)-tailed test is 1·77. [This is the value of $t_{0{\cdot}10}$ for 13 *d.f.* from two-tail tables given at the end of the chapter.]

Conclusion. Since calculated 't' is much greater than tabulated 't', it is highly significant and H_0 is rejected. Hence the two types of electric bulbs differ significantly. Further, since $\bar{x}_1$ is much greater than $\bar{x}_2$, we conclude that type I is definitely superior to type II.

Example 16·12. *The heights of six randomly chosen sailors are (in inches) : 63, 65, 68, 69, 71, and 72. Those of 10 randomly chosen soldiers are 61, 62, 65, 66, 69, 69, 70, 71, 72 and 73. Discuss, the light that these data throw on the suggestion that sailors are on the average taller than soldiers.*

Solution. If the heights of sailors and soldiers be represented by the variables X and Y respectively then the *Null Hypothesis* is, $H_0 : \mu_X = \mu_Y$, *i.e., the sailors are not on the average taller than the soldiers.*

Alternative Hypothesis, $H_1 : \mu_X > \mu_Y$ (Right-tailed).

Under H_0, the test statistic is : $t = \dfrac{\bar{x} - \bar{y}}{\sqrt{S^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \sim t_{n_1 + n_2 - 2} = t_{14}$

Sailors			*Soldiers*		
X	$d = X - A$ $= X - 68$	d^2	Y	$D = Y - B$ $= Y - 66$	D^2
63	−5	25	61	−5	25
65	−3	9	62	−4	16
68	0	0	65	−1	1
69	1	1	66	0	0
71	3	9	69	3	9
72	4	16	69	3	9
			70	4	16
			71	5	25
Total	0	60	72	6	36
			73	7	49
			Total	18	186

$$\bar{x} = A + \frac{\Sigma d}{n_1} = 68 + 0 = 68 \qquad \bar{y} = B + \frac{\Sigma D}{n_2} = 66 + \frac{18}{10} = 67{\cdot}8$$

$\therefore$ and $\Sigma(x - \bar{x})^2 = \Sigma d^2 - \frac{(\Sigma d)^2}{n_1} = 60 - 0 = 60$ and $\Sigma(y - \bar{y})^2 = \Sigma D^2 - \frac{(\Sigma D)^2}{n_2} = 186 - \frac{324}{10} = 153{\cdot}6$

$$S^2 = \frac{1}{n_1 + n_2 - 2}\left[\Sigma(x - \bar{x})^2 + \Sigma(y - \bar{y})^2\right] = \frac{1}{14}(60 + 153{\cdot}6) = 15{\cdot}2571$$

$$\therefore \quad t = \frac{68 - 67{\cdot}8}{\sqrt{15{\cdot}2571}\left(\frac{1}{6} + \frac{1}{10}\right)^{1/2}} = \frac{0{\cdot}2}{\sqrt{15{\cdot}2571 \times 0{\cdot}2667}} = 0{\cdot}099$$

Tabulated $t_{0.05}$ for 14 *d.f.* for single-tail test is 1·76.

Conclusion. Since calculated t is much less than 1·76, it is not at all significant at 5% levels of significance. Hence null hypothesis may be retained at 5% level of significance and we conclude that the data are inconsistent with the suggestion that the sailors are on the average taller than soldiers.

Example 16·13. *To test the claim that the resistance of electric wire can be reduced by at least 0·05 ohm by alloying, 25 values obtained for each alloyed wire and standard wire produced the following results :*

	Mean	*Standard deviation*
Alloyed wire	*0·083 ohm*	*0·003 ohm*
Standard wire	*0·136 ohm*	*0·002 ohm*

Test at 5% level whether or not the claim is substantiated.

Solution. *Null Hypothesis* $H_0 : \mu_1 - \mu_2 \geq 0{\cdot}05$, [*i.e.*, the claim is substantiated]

Alternative Hypothesis $H_1 : \mu_1 - \mu_2 < 0{\cdot}05$ (Left-tailed, test)

Test Statistic. Under H_0, the test statistic is :

$$t = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{S^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \sim N(0, 1)$$

where $S^2 = \frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2} = \frac{25 \times (0{\cdot}003)^2 + 25 \times (0{\cdot}002)^2}{25 + 25 - 2} = \frac{0{\cdot}000225 + 0{\cdot}0001}{48} = 0{\cdot}0000067$

$$\therefore \quad t = \frac{(0{\cdot}083 - 0{\cdot}136) - 0{\cdot}05}{\sqrt{\left\{0{\cdot}0000067\left(\frac{1}{25} + \frac{1}{25}\right)\right\}}} = -\frac{0{\cdot}103}{0{\cdot}00071} = -145{\cdot}07$$

The (critical) tabulated value of t for 48 *d.f.*, at 5% level of significance for left-tailed test is $-$ 1·645.

Conclusion. Since calculated value of t is much less than tabulated value of t, it falls in the rejection region. We, therefore, reject the null hypothesis and conclude that the claim is not substantiated.

Example 16·14. *A certain stimulus administered to each of the 12 patients resulted in the following increase of blood pressure :*

5, 2, 8, –1, 3, 0, –2, 1, 5, 0, 4 and 6

Can it be concluded that the stimulus will, in general, be accompanied by an increase in blood pressure ?

Solution. Here we are given the increments in blood pressure, *i.e.*, $d_i (= x_i - y_i)$.

Null Hypothesis, $H_0 : \mu_X = \mu_Y$, *i.e.*, there is no significant difference in the blood pressure readings of the patients before and after the drug. In other words, the given increments are just by chance (fluctuations of sampling) and not due to the stimulus.

Alternative Hypothesis, $H_1 : \mu_X < \mu_Y$, *i.e.*, the stimulus results in an increase in blood pressure.

Test Statistic. Under H_0, the test statistic is : $t = \dfrac{\bar{d}}{S/\sqrt{n}} \sim t_{(n-1)}$

d	5	2	8	–1	3	0	–2	1	5	0	4	6	31
d^2	25	4	64	1	9	0	4	1	25	0	16	36	185

$$\bar{d} = \frac{1}{n}\sum d = 2{\cdot}58 \quad \text{and} \quad S^2 = \frac{1}{n-1}\left[\sum d^2 - \frac{(\sum d)^2}{n}\right] = \frac{1}{11}\left[185 - \frac{(31)^2}{12}\right] = 9{\cdot}5382$$

$$\therefore \qquad t = \frac{\bar{d}}{S/\sqrt{n}} = \frac{2{\cdot}58 \times \sqrt{12}}{\sqrt{9{\cdot}5382}} = \frac{2{\cdot}58 \times 3{\cdot}464}{3{\cdot}09} = 2{\cdot}89$$

Tabulated $t_{0{\cdot}05}$ for 11 *d.f.* for single-tail test is 1·80. [This is the value of $t_{0{\cdot}10}$ for 11 *d.f.* in the table for two-tail test given at the end of the chapter.]

Conclusion. Since calculated $t > t_{0{\cdot}05}$, H_0 is rejected at 5% level of significance. Hence we conclude that the stimulus will, in general, be accompanied by an increase in blood pressure.

Example 16·15. *In a certain experiment to compare two types of animal foods A and B, the following results of increase in weights were observed in animals :*

Animal number		1	2	3	4	5	6	7	8	*Total*
Increase weight in lb	*Food A*	49	53	51	52	47	50	52	53	407
	Food B	52	55	52	53	50	54	54	53	423

(i) *Assuming that the two samples of animals are independent, can we conclude that food B is better than food A ?*

(ii) *Also examine the case when the same set of eight animals were used in both the foods.*

Solution. *Null Hypothesis*, H_0 : If the increase in weights due to foods A and B are denoted by X and Y respectively, then $H_0 : \mu_X = \mu_Y$, *i.e.*, there is no significant difference in increase in weights due to diets A and B.

Alternative Hypothesis, $H_1 : \mu_X < \mu_Y$ (Left-tailed).

(i) If the two samples of animals be assumed to be independent, then we will apply t-test for difference of means to test H_0.

Test Statistic. Under $H_0 : \mu_X = \mu_Y$, the test criterion is :

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{S^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \sim t_{n_1 + n_2 - 2}$$

Food A			Food B		
X	$d = X - 50$	d^2	Y	$D = Y - 52$	D^2
49	–1	1	52	0	0
53	3	9	55	3	9
51	1	1	52	0	0
52	2	4	53	1	1
47	–3	9	50	–2	4
50	0	0	54	2	4
52	2	4	54	2	4
53	3	9	53	1	1
Total	7	37		7	23

$$\therefore \quad \bar{x} = 50 + \frac{7}{8} = 50{\cdot}875 \qquad \bar{y} = 52 + \frac{7}{8} = 52{\cdot}875$$

and
$$\Sigma(x - \bar{x})^2 = \Sigma d^2 - \frac{(\Sigma d)^2}{n_1} = 37 - \frac{49}{8} = 30{\cdot}875$$

$$\Sigma(y - \bar{y})^2 = \Sigma D^2 - \frac{(\Sigma D)^2}{n_2} = 23 - \frac{49}{8} = 16{\cdot}875$$

$$S^2 = \frac{1}{n_1 + n_2 - 2}\left[\Sigma(x - \bar{x})^2 + \Sigma(y - \bar{y})^2\right] = \frac{1}{14}(30{\cdot}875 + 16{\cdot}875) = 3{\cdot}41$$

$$\therefore \quad t = \frac{\bar{x} - \bar{y}}{\sqrt{S^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{50{\cdot}875 - 52{\cdot}875}{\sqrt{3{\cdot}41\left(\frac{1}{8} + \frac{1}{8}\right)}} = -2{\cdot}17$$

Tabulated $t_{0.05}$ for $(8 + 8 - 2) = 14$ *d.f.* for one-tail test is 1·76.

Conclusion. The critical region for the left-tail test is $t < -1{\cdot}76$. Since calculated t is less than –1·76, H_0 is rejected at 5% level of significance. Hence we conclude that the foods A and B differ significantly as regards their effect on increase in weight. Further, since $\bar{y} > \bar{x}$, food B is superior to food A.

(*ii*) If the same set of animals is used in both the cases, then the readings X and Y are not independent but they are paired together and we apply the paired t-test for testing H_0.

Under $H_0 : \mu_X = \mu_Y$, the statistic is : $t = \dfrac{\bar{d}}{S/\sqrt{n}} \sim t_{(n-1)}$

X	49	53	51	52	47	50	52	53	Total
Y	52	55	52	53	50	54	54	53	
$d = X - Y$	–3	–2	–1	–1	–3	– 4	–2	0	–16
d^2	9	4	1	1	9	16	4	0	44

$$\bar{d} = \frac{\Sigma d}{n} = \frac{-16}{8} = -2 \quad \text{and} \quad S^2 = \frac{1}{n-1}\left[\Sigma d^2 - \frac{(\Sigma d)^2}{n}\right] = \frac{1}{7}\left(44 - \frac{256}{8}\right) = 1{\cdot}714$$

$$\therefore \quad |t| = \frac{|\bar{d}|}{\sqrt{S^2/n}} = \frac{2}{\sqrt{1{\cdot}7143/8}} = \frac{2}{0{\cdot}4629} = 4{\cdot}32$$

Tabulated $t_{0{\cdot}05}$ for $(8-1) = 7$ *d.f.* for one-tail test is 1·90.

Conclusion. Here also the observed value of 't' is significant at 5% level of significance and we conclude that food B is superior to food A.

Example 16·16. *Two laboratories carry out independent estimates of a particular chemicals in a medicine produced by a certain firm. A sample is taken from each batch, halved and the separate halves sent to the two laboratories. The following data is obtained :*

No. of samples	10
Mean value of the difference of estimates	0·6
Sum of the squares of the differences (from their means)	20

Is the difference significant ? (Value of t at 5% level for 9 d.f. is 2·262.)

Solution. Let d stand for the difference between the estimates of the chemical between the two halves of each batch, and $\bar{d}$ the mean value of the difference of estimates. In usual notations, we are given :

$$n = 10, \ \bar{d} = 0{\cdot}6, \ \Sigma(d - \bar{d})^2 = 20$$

Null hypothesis, $H_0 : \mu_1 = \mu_2$, *i.e.*, the difference is insignificant.

Alternative hypothesis, $H_1 : \mu_1 \neq \mu_2$.

Test Statistic. Under H_0, the test statistic is : $t = \dfrac{\bar{d}}{\sqrt{S^2/n}} \sim t_{10-1}$

where $\quad S^2 = \dfrac{1}{n-1}\Sigma(d - \bar{d})^2 = \dfrac{20}{9} = 2{\cdot}22 \quad \therefore \quad t = \dfrac{0{\cdot}6}{\sqrt{2{\cdot}22/10}} = \dfrac{0{\cdot}6}{0{\cdot}471} = 1{\cdot}274.$

The tabulated value of t at 5% level for 9 *d.f.*, is 2·262 (given).

Conclusion. Since calculated value of t is less than tabulated value of t, it is not significant. Hence, we may accept the null hypothesis and conclude that the difference is not significant.

16·3·4. t-test for Testing the Significance of an Observed Sample Correlation Coefficient. If r is the observed correlation coefficient in a sample of n pairs of observations from a bivariate normal population, then Prof. Fisher proved that under the null hypothesis, $H_0 : \rho = 0$, *i.e., population correlation coefficient is zero,* the statistic :

$$t = \frac{r}{\sqrt{(1-r^2)}}\sqrt{(n-2)} \qquad \ldots(16{\cdot}10)$$

follows Student's t-distribution with $(n-2)$ *d.f.* (*c.f. Remark to* § 16·4).

If the value of t comes out to be significant, we reject H_0 at the level of significance adopted and conclude that $\rho \neq 0$, *i.e.*, 'r' is significant of correlation in the population.

If t comes out to be non-significant, then H_0 may be accepted and we conclude that variables may be regarded as uncorrelated in the population.

16·5. F-DISTRIBUTION

Definition. *If X and Y are two independent chi-square variates with υ_1 and υ_2 d.f. respectively, then F-statistic is defined by*

$$F = \frac{X/\nu_1}{Y/\nu_2} \qquad \ldots(16\cdot13)$$

In other words, F is defined as the ratio of two independent chi-square variates divided by their respective degrees of freedom and it follows Snedecor's F-distribution with (υ_1, υ_2) *d.f.* with probability function given by :

$$f(F) = \frac{\left(\frac{\nu_1}{\nu_2}\right)^{\frac{\nu_1}{2}}}{B\left(\frac{\nu_1}{2}, \frac{\nu_2}{2}\right)} \cdot \frac{F^{\frac{\nu_1}{2}-1}}{\left(1+\frac{\nu_1}{\nu_2}F\right)^{(\nu_1+\nu_2)/2}}, 0 \le F < \infty \qquad \ldots[16\cdot13(a)]$$

Remarks 1. The sampling distribution of F-statistic does not involve any population parameters and depends only on the degrees of freedom υ_1 and υ_2.

2. A statistic F following Snedecor's F-distribution with (υ_1, υ_2) *d.f.* will be denoted by $F \sim F(\upsilon_1, \upsilon_2)$.

16·5·1 Derivation of Snedecor's F-distribution. Since X and Y are independent chi-square variates with ν_1 and ν_2 *d.f.* respectively, their joint probability density function is given by :

$$f(x, y) = \left\{\frac{1}{2^{\nu_1/2}\Gamma(\nu_1/2)} \exp(-x/2)\, x^{(\nu_1/2)-1}\right\} \times \left\{\frac{1}{2^{\nu_2/2}\Gamma(\nu_2/2)} \exp(-y/2)\, y^{(\nu_2/2)-1}\right\}$$

$$= \frac{1}{2^{(\nu_1+\nu_2)/2}\Gamma(\nu_1/2)\Gamma(\nu_2/2)} \exp\{-(x+y)/2\} \times x^{(\nu_1/2)-1}\, y^{(\nu_2/2)-1}, 0 \le (x, y) < \infty$$

Let us make the following transformation of variables :

$F = \dfrac{x/\nu_1}{y/\nu_2}$ and $u = y$, so that $0 \le F < \infty, 0 < u < \infty$ $\qquad \therefore \quad x = \dfrac{\nu_1}{\nu_2}Fu$ and $y = u$

Jacobian of transformation J is given by :

$$J = \frac{\partial(x, y)}{\partial(F, u)} = \begin{vmatrix} \frac{\nu_1}{\nu_2}u & 0 \\ \frac{\nu_1}{\nu_2}F & 1 \end{vmatrix} = \frac{\nu_1 u}{\nu_2}$$

Thus the joint *p.d.f.* of the transformed variables is :

$$g(F, u) = \frac{1}{2^{(\nu_1+\nu_2)/2}\Gamma(\nu_1/2)\Gamma(\nu_2/2)} \exp\left\{-\frac{u}{2}\left(1+\frac{\nu_1}{\nu_2}F\right)\right\} \times \left(\frac{\nu_1}{\nu_2}Fu\right)^{(\nu_1/2)-1} u^{(\nu_2/2)-1} |J|$$

$$= \frac{(\nu_1/\nu_2)^{\nu_1/2}}{2^{(\nu_1+\nu_2)/2}\Gamma(\nu_1/2)\Gamma(\nu_2/2)} \exp\left\{-\frac{u}{2}\left(1+\frac{\nu_1}{\nu_2}F\right)\right\} \times u^{[(\nu_1+\nu_2)/2]-1} F^{(\nu_1/2)-1}; \; 0 < u < \infty, 0 \le F < \infty$$

Integrating *w.r. to u* over the range 0 to ∞, the *p.d.f.* of F becomes :

$$g_1(F) = \frac{(\nu_1/\nu_2)^{(\nu_1/2)} F^{(\nu_1/2)-1}}{2^{(\nu_1+\nu_2/2)}\Gamma(\nu_1/2)\,\Gamma(\nu_2/2)} \times \left[\int_0^\infty \exp\left\{-\frac{u}{2}\left(1+\frac{\nu_1}{\nu_2}F\right)\right\} u^{[(\nu_1+\nu_2)/2]-1}\,du\right]$$

$$= \frac{(\nu_1/\nu_2)^{(\nu_1/2)} F^{(\nu_1/2)-1}}{2^{(\nu_1+\nu_2)/2}\Gamma(\nu_1/2)\,\Gamma(\nu_2/2)} \times \frac{\Gamma[(\nu_1+\nu_2)/2]}{\left[\frac{1}{2}\left(1+\frac{\nu_1}{\nu_2}F\right)\right]^{(\nu_1+\nu_2)/2}}$$

$$\therefore \qquad g_1(F) = \frac{(\nu_1/\nu_2)^{\nu_1/2}}{B\left(\frac{\nu_1}{2},\frac{\nu_2}{2}\right)} \cdot \frac{F^{(\nu_1/2)-1}}{\left(1+\frac{\nu_1}{\nu_2}F\right)^{(\nu_1+\nu_2)/2}}, \quad 0 \le F < \infty$$

which is the required probability function of F-distribution with (ν_1, ν_2) *d.f.*

Aliter. $$F = \frac{x/\nu_1}{y/\nu_2}$$

$\therefore \quad \dfrac{\nu_1}{\nu_2}F = \dfrac{x}{y}$, being the ratio of two independent chi-square variates with υ_1 and υ_2 *d.f.* respectively is a $\beta_2\left(\frac{\nu_1}{2}, \frac{\nu_2}{2}\right)$ variate. Hence the probability function of F is given by :

$$dP(F) = \frac{1}{B\left(\frac{\nu_1}{2},\frac{\nu_2}{2}\right)} \cdot \frac{\left(\frac{\nu_1}{\nu_2}F\right)^{(\nu_1/2)-1}}{\left(1+\frac{\nu_1}{\nu_2}F\right)^{(\nu_1+\nu_2)/2}}\, d\left(\frac{\nu_1}{\nu_2}F\right)$$

$$\Rightarrow \qquad f(F) = \frac{\left(\frac{\nu_1}{\nu_2}\right)^{\nu_1/2}}{B\left(\frac{\nu_1}{2},\frac{\nu_2}{2}\right)} \cdot \frac{F^{(\nu_1/2)-1}}{\left(1+\frac{\nu_1}{\nu_2}F\right)^{(\nu_1+\nu_2)/2}}, \quad 0 \le F < \infty$$

16·5·2. Constants of F-distribution.

$$\mu'_r \text{ (about origin)} = E(F^r) = \int_0^\infty F^r f(F)\, dF$$

$$= \frac{(\nu_1/\nu_2)^{\nu_1/2}}{B\left(\frac{\nu_1}{2},\frac{\nu_2}{2}\right)} \int_0^\infty F^r \frac{F^{(\nu_1/2)-1}}{\left(1+\frac{\nu_1}{\nu_2}F\right)^{(\nu_1+\nu_2)/2}}\, dF \qquad ...(*)$$

To evaluate the integral, put : $\dfrac{\nu_1}{\nu_2}F = y$, so that $dF = \dfrac{\nu_2}{\nu_1}dy$

$$\therefore \qquad \mu_r' = \frac{[\nu_1/\nu_2]^{\nu_1/2}}{B\left(\frac{\nu_1}{2},\frac{\nu_2}{2}\right)} \int_0^\infty \frac{\left(\frac{\nu_2}{\nu_1}y\right)^{r+(\nu_1/2)-1}}{(1+y)^{(\nu_1+\nu_2)/2}} \left(\frac{\nu_2}{\nu_1}\right) dy$$

$$= \frac{\left(\frac{\nu_2}{\nu_1}\right)^r}{B\left(\frac{\nu_1}{2}, \frac{\nu_2}{2}\right)} \int_0^\infty \frac{y^{r+(\nu_1/2)-1}}{(1+y)^{(\nu_1/2)+r+[(\nu_2/2)-r]}}\, dy$$

$$= \left(\frac{\nu_2}{\nu_1}\right)^r \cdot \frac{1}{B\left(\frac{\nu_1}{2}, \frac{\nu_2}{2}\right)} \cdot B\left(r + \frac{\nu_1}{2}, \frac{\nu_2}{2} - r\right),\ \nu_2 > 2r \qquad \ldots(16\cdot14)$$

Aliter for (16·14). (16·14) could also be obtained by substituting $\frac{\nu_1}{\nu_2} F = \tan^2\theta$ in (*) and using the Beta integral :

$$2\int_0^{\pi/2} \sin^p\theta \cos^q\theta\, d\theta = B\left(\frac{p+1}{2}, \frac{q+1}{2}\right)$$

$$\therefore \quad \mu_r' = \left(\frac{\nu_2}{\nu_1}\right)^r \cdot \frac{\Gamma[r+(\nu_1/2)]\,\Gamma[(\nu_2/2)-r]}{\Gamma(\nu_1/2)\,\Gamma(\nu_2/2)}\ ;\ r < \frac{\nu_2}{2} \quad \Rightarrow \quad \nu_2 > 2r \qquad \ldots(16\cdot15)$$

In particular

$$\mu_1' = \frac{\nu_2}{\nu_1} \cdot \frac{\Gamma[1+(\nu_1/2)]\,\Gamma[(\nu_2/2)-1]}{\Gamma(\nu_1/2)\,\Gamma(\nu_2/2)} = \frac{\nu_2}{\nu_2 - 2},\ \nu_2 > 2 \qquad \ldots(16\cdot15a)$$

$$[\because \Gamma(r) = (r-1)\,\Gamma(r-1)]$$

Thus the mean of F-distribution is independent of ν_1.

$$\mu_2' = \left(\frac{\nu_2}{\nu_1}\right)^2 \cdot \frac{\Gamma[(\nu_1/2)+2]\,\Gamma[(\nu_2/2)-2]}{\Gamma(\nu_1/2)\,\Gamma(\nu_2/2)}$$

$$= \left(\frac{\nu_2}{\nu_1}\right)^2 \cdot \frac{[(\nu_1/2)+1]\,(\nu_1/2)}{[(\nu_2/2)-1]\,[(\nu_2/2)-2]} = \frac{\nu_2^2(\nu_1+2)}{\nu_1(\nu_2-2)(\nu_2-4)},\ \nu_2 > 4.$$

$$\therefore \quad \mu_2 = \mu_2' - \mu_1'^2 = \frac{\nu_2^2(\nu_1+2)}{\nu_1(\nu_2-2)(\nu_2-4)} - \frac{\nu_2^2}{(\nu_2-2)^2} = \frac{2\nu_2^2(\nu_2+\nu_1-2)}{\nu_1(\nu_2-2)^2(\nu_2-4)},\ \nu_2 > 4 \qquad \ldots(16\cdot15b)$$

Similarly, on putting $r = 3$ and 4 in μ_r', we get μ_3' and μ_4' respectively, from which the central moments μ_3 and μ_4 can be obtained.

Remark. It has been proved that for large degrees of freedom, ν_1 and ν_2, F tends to $N[1, 2\{(1/\nu_1)+(1/\nu_2)\}]$ variate.

16·5·3. Mode and Points of Inflexion of F-distribution. We have

$$\log f(F) = C + \{(\nu_1/2) - 1\}\log F - \left(\frac{\nu_1+\nu_2}{2}\right)\log\{1 + (\nu_1/\nu_2)F\},$$

where C is a constant independent of F.

$$\frac{\partial}{\partial F}[\log f(F)] = \left(\frac{\nu_1}{2} - 1\right) \cdot \frac{1}{F} - \frac{(\nu_1+\nu_2)}{2} \cdot \frac{1}{\left(1 + \frac{\nu_1}{\nu_2}F\right)} \cdot \frac{\nu_1}{\nu_2}$$

$$f'(F) = \frac{\partial}{\partial F} f(F) = 0 \quad \Rightarrow \quad \frac{\nu_1 - 2}{2F} - \frac{\nu_1(\nu_1+\nu_2)}{2(\nu_2 + \nu_1 F)} = 0$$

Hence

$$F = \frac{\nu_2(\nu_1-2)}{\nu_1(\nu_2+2)}$$

It can be easily verified that at this point $f''(F) < 0$. Hence mode $= \frac{\nu_2(\nu_1-2)}{\nu_1(\nu_2+2)}$ $\qquad \ldots(16\cdot16)$

***Remarks* 1.** Since $F > 0$, mode exists if and only if $\nu_1 > 2$.

2. $$\text{Mode} = \left(\frac{\nu_2}{\nu_2+2}\right)\cdot\left(\frac{\nu_1-2}{\nu_1}\right)$$

Hence mode of F-distribution is always less than unity.

3. The points of inflexion of F-distribution exist for $\upsilon_1 > 4$ and are equidistant from mode.

Proof. We have $$\frac{\nu_1}{\nu_2}F = \frac{X}{Y} \sim \beta_2(l, m), \quad \text{...(*)}$$

where $l = \upsilon_1/2$ and $m = \upsilon_2/2$. We now find the points of inflexion of Beta distribution of second kind with parameters l and m. If $X \sim \beta_2(l, m)$, its *p.d.f.* is :

$$f(x) = \frac{1}{\beta(l, m)}\cdot\frac{x^{l-1}}{(1+x)^{l+m}}\,;0 \le x < \infty \quad \text{...(**)}$$

Points of inflexion are the solution of $f''(x) = 0$ and $f'''(x) \ne 0$

From (**), $\log f(x) = -\log\beta(l, m) + (l-1)\log x - (l+m)\log(1+x)$

Differentiating twice *w.r. to* x, we get

$$\frac{f'(x)}{f(x)} = \frac{l-1}{x} - \frac{l+m}{1+x} \quad \text{...(***)}$$

and $$\frac{f(x)f''(x) - [f'(x)]^2}{[f(x)]^2} = -\left(\frac{l-1}{x^2}\right) + \frac{l+m}{(1+x)^2}$$

If $f''(x) = 0$, then $$-\left[\frac{f'(x)}{f(x)}\right]^2 = -\left(\frac{l-1}{x^2}\right) + \frac{l+m}{(1+x)^2} \Rightarrow -\left[\frac{l-1}{x} - \frac{l+m}{1+x}\right]^2 = -\left(\frac{l-1}{x^2}\right) + \frac{l+m}{(1+x)^2}$$

[On using (***)]

$$\Rightarrow \quad \frac{l-1}{x^2}\left(l-1-1\right) - 2\frac{(l-1)(l+m)}{x(1+x)} + \frac{l+m}{(1+x)^2}\times(l+m+1) = 0$$

$$\Rightarrow (l-1)(l-2)(1+x)^2 - 2x(1+x)(l-1)(l+m) + x^2(l+m)(l+m+1) = 0 \quad \text{...(****)}$$

which is a quadratic in x. It can be easily verified that at these values of x, $f'''(x) \ne 0$, if $l > 2$.

The roots of (****) give the points of inflexion of $\beta_2(l, m)$ distribution. The sum of the points of inflexion is equal to the sum of roots of (****) and is given by :

$$-\left[\frac{\text{Coeff. of } x \text{ in (****)}}{\text{Coeff. of } x^2 \text{ in (****)}}\right] = -\left[\frac{2(l-1)(l-2) - 2(l-1)(l+m)}{(l-1)(l-2) - 2(l-1)(l+m) + (l+m)(l+m+1)}\right]$$

$$= \frac{2(l-1)[(l+m)-(l-2)]}{(l-1)(l-2) - (l-1)(l+m) - (l-1)(l+m) + (l+m)(l+m+1)}$$

$$= \frac{2(l-1)(m+2)}{(l-1)[(l-2-l-m)] + (l+m)(l+m+1-l+1)}$$

$$= \frac{2(l-1)(m+2)}{-(l-1)(m+2) + (l+m)(m+2)} = \frac{2(l-1)}{l+m-l+1} = \frac{2(l-1)}{(m+1)}$$

$\therefore$ Sum of points of inflexion of $\left(\frac{\nu_1}{\nu_2}F\right)$ distribution $= \dfrac{2(l-1)}{(m+1)} = \dfrac{2\left(\frac{\nu_1}{2}-1\right)}{\left(\frac{\nu_2}{2}+1\right)} = \dfrac{2(\nu_1-2)}{(\nu_2+2)}$

$\Rightarrow$ Sum of points of inflexion of $F(\upsilon_1, \upsilon_2)$ distribution

$$= \frac{\nu_2}{\nu_1}\cdot\frac{2(\nu_1-2)}{(\nu_2+2)}, \text{ provided } l = \frac{\nu_1}{2} > 2 = 2\frac{\nu_2(\nu_1-2)}{\nu_1(\nu_2+2)} = 2\text{ Mode, provided } \nu_1 > 4$$

Hence the points of inflexion of $F(\upsilon_1, \upsilon_2)$ distribution, when they exist, (*i.e.*, when $\upsilon_1 > 4$) are equidistant from the mode.

4. Karl Pearson's coefficient of skewness is given by : $S_k = \dfrac{\text{Mean} - \text{Mode}}{\sigma} > 0$, since mean > 1 and mode < 1. Hence F-distribution is highly positively skewed.

5. The probability $f(F)$ increases steadily at first until it reaches its peak (corresponding to the modal value which is less than 1) and then decreases slowly so as to become tangential at $F = \infty$, i.e., F-axis is an asymptote to the right tail.

Example 16·20. *When $\nu_1 = 2$, show that the significance level of F corresponding to a significant probability p is : $F = \dfrac{\nu_2}{2}\left[p^{-(2/\nu_2)} - 1\right]$ where ν_1 an ν_2 have their usual meanings.*

Solution. When $\nu_1 = 2$,

$$f(F) = \frac{1}{B\left(1, \frac{\nu_2}{2}\right)} \cdot \frac{2}{\nu_2} \cdot \frac{1}{\left(1 + \frac{2}{\nu_2} F\right)^{(\nu_2/2)+1}}$$

$$= \frac{\Gamma(\frac{\nu_2}{2} + 1)}{\Gamma(1)\Gamma(\nu_2/2)} \times \frac{2/\nu_2}{\left(\frac{2}{\nu_2}\right)^{(\nu_2/2)+1}\left(F + \frac{\nu_2}{2}\right)^{(\nu_2/2)+1}} = \frac{\left(\frac{\nu_2}{2}\right)^{(\nu_2/2)+1}}{\left(F + \frac{\nu_2}{2}\right)^{(\nu_2/2)+1}}$$

Hence $$p = \int_F^\infty f(F)\, dF = \left[\frac{\nu_2}{2}\right]^{(\nu_2/2)+1} \times \int_F^\infty \frac{dF}{\left(F + \frac{\nu_2}{2}\right)^{(\nu_2/2)+1}}$$

$$= \left(\frac{\nu_2}{2}\right)^{(\nu_2/2)+1} \times \left|\frac{\left(F + \frac{\nu_2}{2}\right)^{-(\nu_2/2)}}{-\frac{\nu_2}{2}}\right|_F^\infty = \left[\frac{\left(\frac{\nu_2}{2}\right)}{F + \frac{\nu_2}{2}}\right]^{\nu_2/2} = \frac{1}{\left(1 + \frac{2}{\nu_2} F\right)^{\nu_2/2}}$$

$$\Rightarrow \quad p^{-(2/\nu_2)} = 1 + \frac{2F}{\nu_2} \quad \Rightarrow \quad F = \frac{\nu_2}{2}\left[p^{-(2/\nu_2)} - 1\right].$$

Example 16·21. *X is a binomial variate with parameters n and p and F_{ν_1, ν_2} is an F-statistic with ν_1 and ν_2 d.f. Prove that :*

$$P(X \le k - 1) = P\left[F_{2k,\, 2(n-k+1)} > \frac{n-k+1}{k} \cdot \frac{p}{1-p}\right].$$

Solution. If $X \sim B(n, p)$, then we have

$$P(X \le k-1) = (n - k + 1).\binom{n}{k-1} \int_0^q t^{n-k}(1-t)^{k-1}\, dt$$

$$= \frac{1}{\beta(n-k+1, k)} \int_0^q t^{n-k}(1-t)^{k-1}\, dt \qquad \ldots (*)$$

[See Example 8·31.]

$$P = P_r\left[F_{2k, 2(n-k+1)} > \frac{n-k+1}{k}\left(\frac{p}{1-p}\right)\right] = \int_{\frac{n-k+1}{k} \cdot \frac{p}{q}}^\infty p[F_{2k,\, 2(n-k+1)}]\, dF$$

$$= \frac{1}{B(k, n-k+1)} \int_{\frac{n-k+1}{k} \cdot \frac{p}{q}}^\infty \frac{[k/(n-k+1)]^k \cdot F^{k-1}\, dF}{\left(1 + \frac{kF}{n-k+1}\right)^{n+1}} \qquad \ldots (**)$$

Put $1 + \frac{kF}{n-k+1} = \frac{1}{y} \Rightarrow F = \frac{n-k+1}{k}\left(\frac{1-y}{y}\right)$ and $dF = \frac{n-k+1}{k} \cdot \frac{-dy}{y^2}$

$F = \infty \Rightarrow y = 0$ and $F = \frac{n-k+1}{k} \cdot \frac{p}{q} \Rightarrow \frac{1}{y} = \frac{q+p}{q} = \frac{1}{q} \Rightarrow y = q$

Substituting in (**), we get :

$$P = \frac{1}{B(k, n-k+1)} \int_q^0 \left(\frac{1-y}{y}\right)^{k-1} \cdot y^{n+1} \left(\frac{-dy}{y^2}\right)$$

$$= \frac{1}{B(k,n-k+1)} \int_0^q y^{n-k} (1-y)^{k-1} dy \qquad \ldots (***)$$

From (*), (**) and (***), we get the result.

Example 16·22. *If $F(n_1, n_2)$ represents an F-variate with n_1 and n_2 d.f., prove that $F(n_2, n_1)$ is distributed as $1/F(n_1, n_2)$ variate. Deduce that*

$$P[F(n_1, n_2) \geq c] = P\left[F(n_2, n_1) \leq \frac{1}{c}\right]$$

Or

Show how probability points of $F(n_2, n_1)$ can be obtained from those of $F(n_1, n_2)$.

Solution. Let X and Y be independent chi-square variates with n_1 and n_2 d.f. respectively. Then by definition, we have

$$F = \frac{(X/n_1)}{(Y/n_2)} \sim F(n_1, n_2) \quad \text{and} \quad \frac{1}{F} = \frac{(Y/n_2)}{(X/n_1)} \sim F(n_2, n_1) \qquad \ldots(*)$$

Hence the result.

We have : $P\left[F(n_1, n_2) \geq c\right] = P\left[\frac{1}{F(n_1, n_2)} \leq \frac{1}{c}\right] = P\left[F(n_2, n_1) \leq \frac{1}{c}\right]$ [From (*)]

Remark. Probability Points of $F(n_2, n_1)$ from Those of $F(n_1, n_2)$ Distribution.

Let $P\left[F(n_1, n_2) \geq c\right] = \alpha$, *i.e.*, let c be the upper α-significant point of $F(n_1, n_2)$ distribution.

$$\therefore \quad 1 - \alpha = 1 - P\left[F(n_1, n_2) \geq c\right] = 1 - P\left[\frac{1}{F(n_1, n_2)} \leq \frac{1}{c}\right]$$

$$\Rightarrow \quad \alpha = P\left[F(n_2, n_1) \leq \frac{1}{c}\right] = 1 - P\left[F(n_2, n_1) \geq \frac{1}{c}\right] \qquad \text{[From (*)]}$$

$$\therefore \quad P\left[F(n_2, n_1) \geq \frac{1}{c}\right] = 1 - \alpha$$

Thus $(1-\alpha)$ significant points of $F(n_2, n_1)$ distribution are the reciprocal of α-significant points of $F(n_1, n_2)$ distributions, *e.g.*, $F_{8,4}(0{\cdot}05) = 6{\cdot}04 \Rightarrow F_{4,8}(0{\cdot}95) = \frac{1}{6{\cdot}04}$.

Example 16·23. *Prove that if $n_1 = n_2$, the median of F-distribution is at $F = 1$ and that the quartiles Q_1 and Q_3 satisfy the condition $Q_1Q_3 = 1$.*

Solution. Since $n_1 = n_2 = n$, (say), the median (M) of $F(n_1, n_2) = F(n, n)$ distribution is given by :

$$P\left[F(n, n) \leq M\right] = 0{\cdot}5 \qquad \ldots(*)$$

(c) Since $X_1 \sim N(0, 1)$ and $X_2 \sim N(0, 1)$ are independent, $X_1^2 \sim \chi^2_{(1)}$ and $X_2^2 \sim \chi^2_{(1)}$, are also independent. Hence by definition of F-statistic,

$$\frac{X_1^2/1}{X_2^2/1} \sim F_{(1, 1)} \quad \Rightarrow \quad \frac{X_1^2}{X_2^2} \sim F_{(1, 1)}$$

(d) X_1/X_2, being the ratio of two independent standard normal variates is a standard Cauchy variate.

16·6. APPLICATIONS OF F-DISTRIBUTION

F-distribution has the following applications in statistical theory.

16·6·1. F-test for Equality of Two Population Variances. Suppose we want to test (i) whether two independent samples x_i, $(i = 1, 2, \ldots, n_1)$ and y_j, $(j = 1, 2, \ldots, n_2)$ have been drawn from the normal populations with the same variance σ^2 (say), or (ii) whether the two independent estimates of the population variance are homogeneous or not.

Under the null hypothesis (H_0) that (i) $\sigma_X^2 = \sigma_Y^2 = \sigma^2$, *i.e., the population variances are equal, or (ii) Two independent estimates of the population variance are homogeneous*, the statistic F is given by :

$$F = \frac{S_X^2}{S_Y^2} \qquad \ldots(16·17)$$

where $S_X^2 = \dfrac{1}{n_1 - 1} \sum_{i=1}^{n_1} (x_i - \bar{x})^2$ and $S_Y^2 = \dfrac{1}{n_2 - 1} \sum_{j=1}^{n_2} (y_j - \bar{y})^2$ ...(16·17a)

are unbiased estimates of the common population variance σ^2 obtained from two independent samples and it follows Snedecor's F-distribution with (ν_1, ν_2) *d.f.* where $\nu_1 = n_1 - 1$ and $\nu_2 = n_2 - 1$.

Proof. $F = \dfrac{S_X^2}{S_Y^2} = \left[\dfrac{n_1}{n_1 - 1} s_X^2\right] \Big/ \left[\dfrac{n_2}{n_2 - 1} s_Y^2\right]$

$$= \left[\frac{n_1 s_X^2}{\sigma_X^2} \cdot \frac{1}{(n_1 - 1)}\right] \Big/ \left[\frac{n_2 s_Y^2}{\sigma_Y^2} \cdot \frac{1}{(n_2 - 1)}\right] \qquad (\because \sigma_X^2 = \sigma_Y^2 = \sigma^2, \text{ under } H_0)$$

Since $\dfrac{n_1 s_X^2}{\sigma_X^2}$ and $\dfrac{n_2 s_Y^2}{\sigma_Y^2}$ are independent chi-square variates with $(n_1 - 1)$ and $(n_2 - 1)$ *d.f.* respectively, F follows Snedecor's F-distribution with $(n_1 - 1, n_2 - 1)$ *d.f.* (*c.f.* § 16·5).

Remarks 1. In (16·17), greater of the two variances S_X^2 and S_Y^2 is to be taken in the numerator and n_1 corresponds to the greater variance.

By comparing the calculated value of F obtained by using (16·17) for the two given samples, with the tabulated value of F for (n_1, n_2) *d.f.* at certain level of significance (5% or 1%), H_0 is either rejected or accepted.

2. *Critical values of F-distribution.* The available F-tables (given in Table II-A and II-B at the end of the chapter) give the critical values of F for the right-tailed test, *i.e.*, the critical region is determined by the right-tail areas. Thus the significant value $F_\alpha(n_1, n_2)$ at level of significance α and (n_1, n_2) *d.f.* is determined by : $P[F > F_\alpha(n_1, n_2)] = \alpha$, ...(*)
as shown in the diagram on page 16·37.

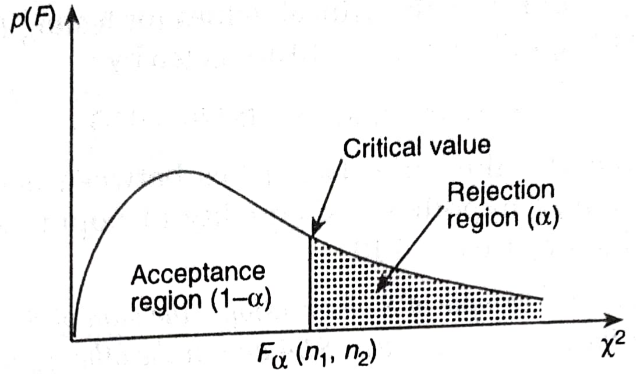

Fig. 16·3 : Critical Values of F-Distribution

From the *Remark to Example 16·23*, we have the following reciprocal relation between the upper and lower α-significant points of *F*-distribution :

$$F_\alpha(n_1, n_2) = \frac{1}{F_{1-\alpha}(n_2, n_1)} \quad \Rightarrow \quad F_\alpha(n_1, n_2) \times F_{1-\alpha}(n_2, n_1) = 1 \qquad \ldots(**)$$

The critical values of F for left tail–test $H_0 : \sigma_1^2 = \sigma_2^2$ against $H_1 : \sigma_1^2 < \sigma_2^2$ are given by $F < F_{n_1-1, n_2-1}(1-\alpha)$, and for the two – tailed test, $H_0 : \sigma_1^2 = \sigma_2^2$ against $H_1 : \sigma_1^2 \neq \sigma_2^2$ are given by $F > F_{n_1-1, n_2-1}(\alpha/2)$ and $F < F_{n_1-1, n_2-1}(1-\alpha/2)$ [For details, see Chapter Eighteen.]

Example 16·25. *Pumpkins were grown under two experimental conditions. Two random samples of 11 and 9 pumpkins show the sample standard deviations of their weights as 0·8 and 0·5 respectively. Assuming that the weight distributions are normal, test the hypothesis that the true variances are equal, against the alternative that they are not, at the 10% level. [Assume that $P(F_{10,8} \geq 3{\cdot}35) = 0{\cdot}05$ and $P(F_{8,10} \geq 3{\cdot}07) = 0{\cdot}05$.]*

Solution. We want to test *Null Hypothesis*, $H_0 : \sigma_X^2 = \sigma_Y^2$ against the *Alternative Hypothesis*, $H_1 : \sigma_X^2 \neq \sigma_Y^2$ (Two-tailed).

We are given : $n_1 = 11, n_2 = 9, s_X = 0{\cdot}8$ and $s_Y = 0{\cdot}5$.

Under the null hypothesis, $H_0 : \sigma^2_X = \sigma^2_Y$, the statistic :

$$F = \frac{S_X^2}{S_Y^2} \text{ follows } F \text{ distribution with } (n_1 - 1, n_2 - 1) \text{ d.f.}$$

$$n_1 s_X^2 = (n_1 - 1) S_X^2 \Rightarrow S_X^2 = \left(\frac{n_1}{n_1 - 1}\right) s_X^2 = \left(\frac{11}{10}\right) \times (0{\cdot}8)^2 = 0{\cdot}704$$

Similarly, $$S_Y^2 = \left(\frac{n_2}{n_2 - 1}\right) s_Y^2 = \left(\frac{9}{8}\right) \times (0{\cdot}5)^2 = 0{\cdot}28125$$

$$\therefore \quad F = \frac{0{\cdot}704}{0{\cdot}28125} = 2{\cdot}5$$

The significant values of F for two-tailed test at level of significance $\alpha = 0{\cdot}10$ are :

$$\left.\begin{array}{ll} F > & F_{10,8}(\alpha/2) = F_{10,8}(0{\cdot}05) \\ \text{and} \quad F < & F_{10,8}(1 - \alpha/2) = F_{10,8}(0{\cdot}95) \end{array}\right\} \qquad \ldots(*)$$

We are given the tabulated (significant) values :

$$P(F_{10,8} \geq 3{\cdot}35) = 0{\cdot}05 \quad \Rightarrow \quad F_{10,8}(0{\cdot}05) = 3{\cdot}35 \qquad \ldots(**)$$

Also $$P(F_{8,10} \geq 3{\cdot}07) = 0{\cdot}05 \quad \Rightarrow \quad P\left(\frac{1}{F_{8,10}} \leq \frac{1}{3{\cdot}07}\right) = 0{\cdot}05$$

$$\Rightarrow \quad P(F_{10,8} \leq 0{\cdot}326) = 0{\cdot}05 \quad \Rightarrow \quad P(F_{10,8} \geq 0{\cdot}326) = 0{\cdot}95 \qquad \ldots(***)$$

Hence from (*), (**) and (***), the critical values for testing $H_0 : \sigma_X^2 = \sigma_Y^2$, against $H_1 : \sigma_X^2 \neq \sigma_Y^2$ at level of significance $\alpha = 0{\cdot}10$ are given by :

$$F > 3{\cdot}35 \quad \text{and} \quad F < 0{\cdot}326 = 0{\cdot}33$$

Since, the calculated value of F (=2·5) lies between 0·33 and 3·35, it is not significant and hence null hypothesis of equality of population variances may be accepted at level of significance $\alpha = 0{\cdot}10$.

Example 16·26. *In one sample of 8 observations, the sum of the squares of deviations of the sample values from the sample mean was 84·4 and in the other sample of 10 observations it was 102·6. Test whether this difference is significant at 5 per cent level, given that the 5 per cent point of F for $n_1 = 7$ and $n_2 = 9$ degrees of freedom is 3·29.*

Solution. Here $n_1 = 8, n_2 = 10$ and $\Sigma(x-\bar{x})^2 = 84{\cdot}4,\ \Sigma(y-\bar{y})^2 = 102{\cdot}6$

$$\therefore \qquad S_X^2 = \frac{1}{n_1 - 1}\Sigma(x-\bar{x})^2 = \frac{84{\cdot}4}{7} = 12{\cdot}057$$

$$S_Y^2 = \frac{1}{n_2-1}\Sigma(y-\bar{y})^2 = \frac{102{\cdot}6}{9} = 11{\cdot}4$$

Under $H_0 : \sigma_X^2 = \sigma_Y^2 = \sigma^2$, *i.e.*, the estimates of σ^2 given by the samples are homogeneous, the test statistic is :

$$F = \frac{S_X^2}{S_Y^2} = \frac{12{\cdot}057}{11{\cdot}4} = 1{\cdot}057$$

Tabulated $F_{0.05}$ for (7, 9) *d.f.* is 3·29.

Since calculated $F < F_{0.05}$, H_0 may be accepted at 5% level of significance.

Example 16·27. *Two random samples gave the following results :*

Sample	*Size*	*Sample mean*	*Sum of squares of deviations from the mean*
1	10	15	90
2	12	14	108

Test whether the samples come from the same normal population at 5% level of significance.

[*Given* : $F_{0.05}(9, 11) = 2{\cdot}90$, $F_{0.05}(11, 9) = 3{\cdot}10$ (approx.) and $t_{0.05}(20) = 2{\cdot}086$, $t_{0.05}(22) = 2{\cdot}07$]

Solution. A normal population has two parameters, *viz.*, mean μ and variance σ^2. To test if two independent samples have been drawn from the same normal population, we have to test (*i*) the equality of population means, and (*ii*) the equality of population variances.

Null Hypothesis : The two samples have been drawn from the same normal population, *i.e.*,

$$H_0 : \mu_1 = \mu_2 \quad \text{and} \quad \sigma_1^2 = \sigma_2^2.$$

Equality of means will be tested by applying *t*-test and equality of variances will be tested by applying *F*-test. Since *t*-test assumes $\sigma_1^2 = \sigma_2^2$, we shall first apply *F*-test and then *t*-test. In usual notations, we are given :

$$n_1 = 10, \quad n_2 = 12; \quad \bar{x}_1 = 15, \quad \bar{x}_2 = 14, \quad \Sigma(x_1 - \bar{x}_1)^2 = 90, \quad \Sigma(x_2-\bar{x}_2)^2 = 108.$$

F-test : Here

$$S_1^2 = \frac{1}{n_1 - 1} \Sigma(x_1 - \bar{x}_1)^2 = \frac{90}{9} = 10, \quad S_2^2 = \frac{1}{n_2 - 1} \Sigma(x_2 - \bar{x}_2)^2 = \frac{108}{11} = 9{\cdot}82$$

Since $S_1^2 > S_2^2$, under $H_0 : \sigma_1^2 = \sigma_2^2$, the test statistic is

$$F = \frac{S_1^2}{S_2^2} \sim F(n_1 - 1, n_2 - 1) = F(9, 11)$$

$$\therefore \quad F = \frac{S_1^2}{S_2^2} = \frac{10}{9{\cdot}82} = 1{\cdot}018$$

Tabulated $F_{0{\cdot}05}$ (9, 11) = 2·90. Since calculated F is less than tabulated F, it is not significant. Hence null hypothesis of equality of population variances may be accepted.

Since $\sigma_1^2 = \sigma_2^2$, we can now apply t test for testing $H_0 : \mu_1 = \mu_2$.

t-test : Under $H_0' : \mu_1 = \mu_2$, against alternative hypothesis, $H_1' : \mu_1 \neq \mu_2$, the test statistic is :

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{S^2 \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \sim t_{n_1 + n_2 - 2} = t_{20}$$

where $$S^2 = \frac{1}{n_1 + n_2 - 2}\left[\Sigma(x_1 - \bar{x}_1)^2 + \Sigma(x_2 - \bar{x}_2)^2\right] = \frac{1}{20}(90 + 108) = 9{\cdot}9$$

$$\therefore \quad t = \frac{15 - 14}{\sqrt{9{\cdot}9\left(\frac{1}{10} + \frac{1}{12}\right)}} = \frac{1}{\sqrt{9{\cdot}9 \times \frac{11}{60}}} = \frac{1}{\sqrt{1{\cdot}815}} = 0{\cdot}742)$$

Tabulated $t_{0{\cdot}05}$ for 20 $d.f.$ = 2·086. Since $|t| < t_{0{\cdot}05}$, it is not significant. Hence the hypothesis $H_0' : \mu_1 = \mu_2$ may be accepted. Since both the hypotheses, *i.e.*, $H_0' : \mu_1 = \mu_2$ and $H_0 : \sigma_1^2 = \sigma_2^2$ are accepted, we may regard that the given samples have been drawn from the same normal population.

16·6·2. F-test for Testing the Significance of an Observed Multiple Correlation Coefficient. If R is the observed multiple correlation coefficient of a variate with k other variates in a random sample of size n from a $(k + 1)$ variate population, then Prof. R.A. Fisher proved that under the null hypothesis (H_0) *that the multiple correlation coefficient in the population is zero, the statistic* :

$$F = \frac{R^2}{1 - R^2} \cdot \frac{n - k - 1}{k} \qquad \ldots(16{\cdot}18)$$

conforms to F-distribution with $(k, n - k - 1)$ $d.f.$

16·6·3. F-test for Testing the Significance of an Observed Sample Correlation Ratio η_{YX}. Under the null hypothesis that *population correlation ratio is zero*, the test statistic is :

$$F = \frac{\eta^2}{1 - \eta^2} \cdot \frac{N - h}{h - 1} \sim F(h - 1, N - h) \qquad \ldots(16{\cdot}19)$$

where N is the size of the sample (from a bi-variate normal population) arranged in h arrays.

16·6·4. F-test for Testing the Linearity of Regression. For a sample of size N arranged in h arrays, from a bi-variate normal population, the test statistic for testing the hypothesis of linearity of regression is :

$$F = \frac{\eta^2 - r^2}{1 - \eta^2} \cdot \frac{N - h}{h - 2} \sim F(h - 2, N - h) \qquad \ldots(16{\cdot}20)$$

16·6·5. F-test for Equality of Several Means. This test is carried out by the technique of Analysis of Variance, which plays a very important and fundamental role in Design of Experiments in Agricultural Statistics.

[For a detailed discussion of the Analysis of Variance Technique, see Fundamentals of Applied Statistics by the same authors.]

16·7. RELATION BETWEEN t AND F DISTRIBUTIONS

In F-distribution with (υ_1, υ_2) *d.f.* [*c.f.* 16·5 (*a*)], take $\upsilon_1 = 1$, $\upsilon_2 = \upsilon$ and $t^2 = F$, *i.e.*, $dF = 2t\,dt$. Thus the probability differential of F transforms to :

$$dG(t) = \frac{(1/\nu)^{1/2}}{B\left(\frac{1}{2}, \frac{\nu}{2}\right)} \cdot \frac{(t^2)^{\frac{1}{2}-1}}{\left(1 + \frac{t^2}{\nu}\right)^{(\nu+1)/2}} \, 2t\,dt, \; 0 \le t^2 < \infty$$

$$= \frac{1}{\sqrt{\nu}\,B\left(\frac{1}{2}, \frac{\nu}{2}\right)} \cdot \frac{1}{\left(1 + \frac{t^2}{\nu}\right)^{(\nu+1)/2}} \, dt, \; -\infty < t < \infty$$

the factor 2 disappearing since the total probability in the range $(-\infty, \infty)$ is unity. This is the probability function of Student's t-distribution with υ *d.f.* Hence we have the following relation between t and F distributions.

If a statistic t follows Student's t distribution with n d.f., then t^2 follows Snedecor's F-distribution with (1, n) d.f. Symbolically,

$$\left.\begin{array}{ll} \text{if} & t \sim t_{(n)} \\ \text{then} & t^2 \sim F_{(1, n)} \end{array}\right\} \qquad \ldots(16{\cdot}21)$$

Aliter Proof of (16·21). If $\xi \sim N(0, 1)$ and $X \sim \chi^2_{(n)}$ are independent *r.v.*'s, then :

$$U = \xi^2 \sim \chi^2_{(1)} \text{ [Square of a } S.N.V.]$$

and

$$t = \frac{\xi}{\sqrt{X/n}} \sim t_{(n)} \quad \Rightarrow \quad t^2 = \frac{\xi^2}{(X/n)} = \frac{(\xi^2/1)}{(X/n)},$$

being the ratio of two independent chi-square variates divided by their respective degrees of freedom is $F(1, n)$ variate. Hence $t^2 \sim F(1, n)$.

With the help of relation (16·21), all the uses of t-distribution can be regarded as the applications of F-distribution also, *e.g.*, for test for a single mean, instead of computing $t = \dfrac{\bar{x} - \mu}{S/\sqrt{n}}$, we may compute $F = t^2 = \dfrac{n(\bar{x} - \mu)^2}{S^2}$ and then apply F-test with $(1, n)$ *d.f.*, and so on.

Similarly, we can write the test statistic F from § 16·3·4, § 16·3·5 and § 16·3·6 for testing the significance of an observed sample correlation coefficient, regression coefficient and partial correlation coefficient respectively.

Example 16·28. *Given* : $P[F(10, 12) > 2{\cdot}753] = 0{\cdot}05 = P[F(1, 12) > 4{\cdot}747]$,

find $P[F(12, 10) > (2{\cdot}753)^{-1}]$, *and* $P(-\sqrt{4{\cdot}747} < t_{12} < \sqrt{4{\cdot}747})$

Solution.
$$P[F(12, 10) > (2{\cdot}753)^{-1}] = P\left[\frac{1}{F(12, 10)} < 2{\cdot}753\right] = P[F(10, 12) < 2{\cdot}753]$$
$$= 1 - P[F(10, 12) > 2{\cdot}753] = 1 - 0{\cdot}05 = 0{\cdot}95$$
$$P(-\sqrt{4{\cdot}747} < t_{12} < \sqrt{4{\cdot}747}) = P(t^2_{12} < 4{\cdot}747) = P[F(1, 12) < 4{\cdot}747]$$
$$= 1 - P[F(1, 12) > 4{\cdot}747] = 1 - 0{\cdot}05 = 0{\cdot}95$$

16·8. RELATION BETWEEN F AND χ^2 DISTRIBUTION

In $F(n_1, n_2)$ distribution if we let $n_2 \to \infty$, then $\chi^2 = n_1 F$ follows χ^2-distribution with n_1 d.f. ... (16·22)

Proof. We have

$$f(F) = \frac{(n_1/n_2)^{n_1/2} F^{(n_1/2)-1}}{\Gamma(n_1/2)\,\Gamma(n_2/2)} \cdot \frac{\Gamma[(n_1+n_2)/2]}{\left(1+\frac{n_1}{n_2}F\right)^{(n_1+n_2)/2}}, \; 0 < F < \infty$$

In the limit as $n_2 \to \infty$, we have

$$\frac{\Gamma[(n_1+n_2)/2]}{n_2^{n_1/2}\,\Gamma(n_2/2)} \to \frac{(n_2/2)^{n_1/2}}{n_2^{n_1/2}} = \frac{1}{2^{n_1/2}} \qquad \ldots (*)$$

$$\left[\because \frac{\Gamma(n+k)}{\Gamma(n)} \to n^k \text{ as } n \to \infty. \text{ (c.f. Remark below)}\right]$$

Also
$$\lim_{n_2\to\infty}\left(1+\frac{n_1}{n_2}F\right)^{(n_1+n_2)/2} = \lim_{n_2\to\infty}\left[\left(1+\frac{n_1}{n_2}F\right)^{n_2}\right]^{1/2} \times \lim_{n_2\to\infty}\left(1+\frac{n_1}{n_2}F\right)^{n_1/2}$$
$$= \exp(n_1 F/2) = \exp(\chi^2/2) \qquad (\because\; n_1 F = \chi^2) \qquad \ldots (**)$$

Hence in the limit, on using (*) and (**), the *p.d.f.* of $\chi^2 = n_1 F$ becomes :

$$dP(\chi^2) = \frac{(n_1/2)^{n_1/2}\, e^{-\chi^2/2}}{\Gamma(n_1/2)} \cdot \left(\frac{\chi^2}{n_1}\right)^{(n_1/2)-1} d\left(\frac{\chi^2}{n_1}\right)$$
$$= \frac{1}{2^{n_1/2}\,\Gamma(n_1/2)} \cdot e^{-\chi^2/2}\,(\chi^2)^{(n_1/2)-1}\, d\chi^2, \; 0 < \chi^2 < \infty$$

which is the p.d.f. of chi-square distribution with n_1 *d.f.*

Remark.
$$\lim_{n\to\infty}\frac{\Gamma(n+k)}{\Gamma(n)} = \lim_{n\to\infty}\frac{(n+k-1)\,!}{(n-1)\,!}, \text{ (for large } n) \simeq \lim_{n\to\infty}\frac{\sqrt{2\pi}\, e^{-(n+k-1)}\,(n+k-1)^{n+k-(1/2)}}{\sqrt{2\pi}\, e^{-(n-1)}\,(n-1)^{n-(1/2)}}$$

(On using Stirling's approximation for $n\,!$ as $n \to \infty$.)

$$= e^{-k}\lim_{n\to\infty}\frac{n^{n+k-\frac{1}{2}}\left(1+\frac{k-1}{n}\right)^{n+k-\frac{1}{2}}}{n^{n-\frac{1}{2}}\left(1-\frac{1}{n}\right)^{n-\frac{1}{2}}} = e^{-k}\, n^k\, \frac{\lim_{n\to\infty}\left(1+\frac{k-1}{n}\right)^n \lim_{n\to\infty}\left(1+\frac{k-1}{n}\right)^{k-\frac{1}{2}}}{\lim_{n\to\infty}\left(1-\frac{1}{n}\right)^n \lim_{n\to\infty}\left(1-\frac{1}{n}\right)^{-\frac{1}{2}}}$$

$$= e^{-k}\, n^k\left[\frac{e^{(k-1)}}{e^{-1}} \cdot \frac{1}{1}\right] = n^k$$

$$\therefore \qquad \lim_{n\to\infty}\frac{\Gamma(n+k)}{\Gamma n} = n^k \qquad \ldots (16{\cdot}22a)$$

16·9. FISHER'S z-DISTRIBUTION

In G.W. Snedecor's F-distribution with (ν_1, ν_2) *d.f.*, if we put

$$F = \exp(2Z) \quad \Rightarrow \quad Z = \frac{1}{2}\log_e F, \qquad \ldots(16\cdot23)$$

the distribution of Z becomes

$$g(z) = p(F).\left|\frac{dF}{dz}\right| = \frac{(\nu_1/\nu_2)^{\nu_1/2}}{B\left(\frac{\nu_1}{2}, \frac{\nu_2}{2}\right)} \cdot \frac{(e^{2z})^{(\nu_1/2)-1}\, 2e^{2z}}{\left[1 + \frac{\nu_1}{\nu_2}e^{2z}\right]^{(\nu_1+\nu_2)/2}}$$

$$= 2\,\frac{(\nu_1/\nu_2)^{\nu_1/2}}{B\left(\frac{\nu_1}{2}, \frac{\nu_2}{2}\right)} \cdot \frac{e^{\nu_1 z}}{\left[1 + \frac{\nu_1}{\nu_2}e^{2z}\right]^{(\nu_1+\nu_2)/2}}\,; -\infty < z < \infty \qquad \ldots(16\cdot24)$$

which is the probability function of Fisher's z-distribution with (ν_1, ν_2) *d.f.* The tables of significant values z_0 of z which will be exceeded in random sampling with probabilities 0·05 and 0·01, *i.e.*, $P(z > z_0) = 0\cdot05$ and $P(z > z_0') = 0\cdot01$ corresponding to various *d.f.* (ν_1, ν_2) were published by Fisher (*c.f.* Statistical Methods for Research Workers) in 1925. From these tables, Snedecor (1934-38) by using (16·23) deduced the tables of significant values of the variance ratio which he denoted by F in honour of Prof. R.A. Fisher.

Remark. With the help of relation (16·23), all the applications of F-distribution may be regarded as the applications of z-distribution also.

16·9·1. Moment Generating Function of z-distribution.

$$M_Z(t) = E(e^{tZ}) = \int_{-\infty}^{\infty} e^{tz} g(z)\, dz = \int_0^{\infty} F^{t/2} f(F)\, dF \qquad [\because e^{2z} = F]$$

Since μ_r' (about origin) for F-distribution is $\int_0^\infty F^r f(F)\, dF$, we can find *m.g.f.* of the z-distribution by putting $r = t/2$ in the expression for μ_r' for F-distribution.

Hence $$M_Z(t) = \left(\frac{\nu_2}{\nu_1}\right)^{t/2} \cdot \frac{\Gamma\{(\nu_1+t)/2\}\,\Gamma\{(\nu_2-t)/2\}}{\Gamma(\nu_1/2)\,\Gamma(\nu_2/2)} \qquad [\textit{c.f.} \text{ Equation } (16\cdot15)]$$

$$\Rightarrow \quad K_Z(t) = \log M_Z(t)$$

$$= \frac{t}{2}(\log \nu_2 - \log \nu_1) + \log \Gamma\{(\nu_1+t)/2\} + \log \Gamma\{(\nu_2-t)/2\} - \log \Gamma(\nu_1/2) - \log \Gamma(\nu_2/2)$$

Using Stirling's approximation for $n!$, when n is large, *viz.*,

$$\lim_{n\to\infty} \Gamma(n+1) = \lim_{n\to\infty} n! \approx \sqrt{2\pi}\, e^{-n}\, n^{n+\frac{1}{2}}$$

$$\Rightarrow \quad \log \Gamma(n+1) = (n + \tfrac{1}{2}) \log n - n + \log\sqrt{2\pi}\text{, we get}$$

$$\kappa_1 = \mu_1' = \frac{1}{2}\left(\frac{1}{\nu_2} - \frac{1}{\nu_1}\right), \qquad \kappa_2 = \mu_2 = \frac{1}{2}\left(\frac{1}{\nu_2} + \frac{1}{\nu_1} + \frac{1}{\nu_1^2} + \frac{1}{\nu_2^2}\right)$$

$$\kappa_3 = \mu_3 = \frac{1}{2}\left[\left(\frac{1}{\nu_2^2} - \frac{1}{\nu_1^2}\right) + \left(\frac{1}{\nu_2^3} + \frac{1}{\nu_1^3}\right)\right]$$

$$\kappa_4 = \mu_4 - 3\mu_2^2 = \frac{1}{\nu_1^3} + \frac{1}{\nu_2^3} + 3\left(\frac{1}{\nu_1^4} + \frac{1}{\nu_2^4}\right),$$

whence β_1 and β_2 can be found.

Remark. z-distribution tends to normal distribution with mean $\frac{1}{2}\left(\frac{1}{\nu_2}-\frac{1}{\nu_1}\right)$ and variance $\frac{1}{2}\left(\frac{1}{\nu_1}+\frac{1}{\nu_2}\right)$, as υ_1 and υ_2 become large.

16·10. FISHER'S Z-TRANSFORMATION

To test the significance of an observed sample correlation coefficient from an uncorrelated bivariate normal population, *t*-test is used. But in random sample of size *n* from a bivariate normal population in which $\rho \neq 0$, Prof. R.A. Fisher proved that the distribution of 'r' is by no means normal and in the neighbourhood of $\rho = \pm 1$, its probability curve is extremely skewed even for large n. If $\rho \neq 0$, Fisher suggested the following transformation :

$$Z = \frac{1}{2}\log_e \frac{1+r}{1-r} = \tanh^{-1} r \qquad \ldots(16{\cdot}25)$$

and proved that even for small samples, the distribution of Z is approximately

normal with mean : $$\xi = \frac{1}{2}\log_e \frac{1+\rho}{1-\rho} = \tanh^{-1}\rho \qquad \ldots(16{\cdot}25a)$$

and variance $1/(n-3)$ and for large values of n, say > 50, the approximation is fairly good.

Remark. The values of Z have been tabulated for different values of r and are given in Table III at the end of the chapter.

16·10·1. Applications of Z-Transformation.

Z-transformation has the following applications in Statistics.

(1) To test if an observed value of 'r' differs significantly from a hypothetical value ρ of the population correlation coefficient.

H_0 : *There is no significant difference between r and ρ. In other words, the given sample has been drawn from a bivariate normal population with correlation coefficient ρ.* If we take

$$Z = \tfrac{1}{2}\log_e \{(1+r)/(1-r)\} \text{ and } \xi = \tfrac{1}{2}\log_e \{(1+\rho)/(1-\rho)\}, \text{ then under } H_0,$$

$$Z \sim N\left(\xi, \frac{1}{n-3}\right) \quad \Rightarrow \quad \frac{Z-\xi}{\sqrt{1/(n-3)}} \sim N(0, 1)$$

Thus if $(Z-\xi)\sqrt{(n-3)} > 1{\cdot}96$, H_0 is rejected at 5% level of significance and if it is greater than 2·58, H_0 is rejected at 1% level of significance, where Z and ξ are defined in (16·25) and (16·25a).

Remark. Z defined in equation (16·25) should not be confused with the Z used in Fisher's z-distribution (*c.f.* § 16·9).

Example 16·29. *A correlation coefficient of 0·72 is obtained from a sample of 29 pairs of observations.*

(i) Can the sample be regarded as drawn from a bivariate normal population in which true correlation coefficient is 0·8 ?

(ii) Obtain 95% confidence limits for ρ in the light of the information provided by the sample.

Solution. (*i*) H_0 : There is no significant difference between $r = 0{\cdot}72$; and $\rho = 0{\cdot}80$, *i.e.*, the sample can be regarded as drawn from the bivariate normal population with $\rho = 0{\cdot}8$. Here

$$Z = \frac{1}{2}\log_e\left(\frac{1+r}{1-r}\right) = 1\cdot1513 \log_{10}\left(\frac{1+r}{1-r}\right) = 1\cdot1513 \log_{10} 6\cdot14 = 0\cdot907$$

$$\xi = \frac{1}{2}\log_e\left(\frac{1+\rho}{1-\rho}\right) = 1\cdot1513 \log_{10}\frac{(1+0\cdot8)}{(1-0\cdot8)} = 1\cdot1513 \times 0\cdot9541 = 1\cdot1$$

$$\text{S.E. }(Z) = \frac{1}{\sqrt{n-3}} = \frac{1}{\sqrt{26}} = 0\cdot196$$

Under H_0, the test statistic is : $$U = \frac{Z-\xi}{1/\sqrt{n-3}} \sim N(0,1)$$

$$\therefore \quad U = \frac{(0\cdot907 - 1\cdot100)}{0\cdot196} = -0\cdot985$$

Since $|U| < 1\cdot96$, it is not significant at 5% level of significance and H_0 may be accepted. Hence the sample may be regarded as coming from a bivariate normal population with $\rho = 0\cdot8$.

(*ii*) 95% confidence limits for ρ on the basis of the information supplied by the sample, are given by :

$$|U| \le 1\cdot96 \quad \Rightarrow \quad |Z - \xi| \le 1\cdot96 \times \frac{1}{\sqrt{n-3}} = 1\cdot96 \times 0\cdot196$$

$$\Rightarrow \quad |0\cdot907 - \xi| \le 0\cdot384 \quad \text{or} \quad 0\cdot907 - 0\cdot384 \le \xi \le 0\cdot907 + 0\cdot384$$

$$\Rightarrow \quad 0\cdot523 \le \xi \le 1\cdot291 \quad \text{or} \quad 0\cdot523 \le \frac{1}{2}\log_e\left(\frac{1+\rho}{1-\rho}\right) \le 1\cdot291$$

$$\Rightarrow 0\cdot523 \le 1\cdot1513 \log_{10}\left(\frac{1+\rho}{1-\rho}\right) \le 1\cdot291 \quad \text{or} \quad \frac{0\cdot523}{1\cdot1513} \le \log_{10}\left(\frac{1+\rho}{1-\rho}\right) \le \frac{1\cdot291}{1\cdot1513}$$

$$\therefore \quad 0\cdot4543 \le \log_{10}\left(\frac{1+\rho}{1-\rho}\right) \le 1\cdot1213 \qquad \ldots(*)$$

Now $\log_{10}\left(\frac{1+\rho}{1-\rho}\right) = 0\cdot4543$

$$\Rightarrow \frac{1+\rho}{1-\rho} = \text{Antilog }(0\cdot4543) = 2\cdot846$$

$$\therefore \quad \rho = \frac{2\cdot846-1}{2\cdot846+1} = \frac{1\cdot846}{3\cdot846} = 0\cdot4799$$

and $\log_{10}\left(\frac{1+\rho}{1-\rho}\right) = 1\cdot1213$

$$\Rightarrow \frac{1+\rho}{1-\rho} = \text{Antilog }(1\cdot1213) = 13\cdot22$$

$$\therefore \quad \rho = \frac{13\cdot22-1}{13\cdot22+1} = \frac{12\cdot22}{14\cdot22} = 0\cdot86$$

Hence, substituting in (*), we get $\quad 0\cdot48 \le \rho \le 0\cdot86$

(2) To test the significance of the difference between two independent sample correlation coefficients. Let r_1 and r_2 be the sample correlation coefficients observed in two independent samples of sizes n_1 and n_2 respectively, then,

$$Z_1 = \log_e\left(\frac{1+r_1}{1-r_1}\right) \quad \text{and} \quad Z_2 = \frac{1}{2}\log_e\left(\frac{1+r_2}{1-r_2}\right)$$

Under the null hypothesis, H_0 : *that sample correlation coefficients do not differ significantly, i.e., the samples are drawn from the same bivariate normal population or from different populations with same correlation coefficient* ρ, *(say), the statistic* :

$$Z = \frac{(Z_1 - Z_2) - E(Z_1 - Z_2)}{\text{S.E. }(Z_1 - Z_2)} \sim N(0,1)$$

$$E(Z_1 - Z_2) = E(Z_1) - E(Z_2) = \xi_1 - \xi_2 = 0$$

$$\left[\because \xi_1 = \xi_2 = \frac{1}{2}\log_e \frac{1+\rho}{1-\rho} \text{ (under } H_0)\right]$$

and $\quad \text{S.E.}(Z_1 - Z_2) = \sqrt{V(Z_1) + V(Z_2)} = \sqrt{\left\{\frac{1}{n_1 - 3} + \frac{1}{n_2 - 3}\right\}}$

[Covariance term vanishes since samples are independent.]

Under H_0, the test statistic is :

$$Z = \frac{Z_1 - Z_2}{\sqrt{\left\{\frac{1}{n_1 - 3} + \frac{1}{n_2 - 3}\right\}}} \sim N(0, 1)$$

By comparing this value with 1·96 or 2·58, H_0 may be accepted or rejected at 5% and 1% levels of significance respectively.

(3) To obtain pooled estimate of ρ. Let $r_1, r_2, \ldots, r_k$ be observed correlation coefficients in k-independent samples of sizes $n_1, n_2, \ldots, n_k$ respectively from a bivariate normal population. The problem is to combine these estimates of ρ to get a pooled estimate for the parameter. If we take

$Z_i = \frac{1}{2}\log_e\left(\frac{1+r_i}{1-r_i}\right)$; $i = 1, 2, \ldots, k$; then Z_i ; $i = 1, 2, \ldots, k$ are independent normal variates with variances $\frac{1}{(n_i - 3)}$; $i = 1, 2, \ldots, k$ and common mean $\xi = \frac{1}{2}\log_e\left(\frac{1+\rho}{1-\rho}\right)$.

The weighted mean, (say $\bar{Z}$), of these Z_i's is given by :

$$\bar{Z} = \sum_{i=1}^{k} w_i Z_i \Big/ \sum_{i=1}^{k} w_i, \text{ where } w_i \text{ is the weight of } Z_i.$$

Now $\bar{Z}$ is also an unbiased estimate of ξ, since

$$E(\bar{Z}) = \frac{1}{\Sigma w_i}\left(E \sum_{i=1}^{k} w_i Z_i\right) = \frac{1}{\Sigma w_i}\left[\sum_i w_i\, E(Z_i)\right] = \frac{1}{\Sigma w_i}\left(\sum_i w_i \xi\right) = \xi$$

and $\quad V(\bar{Z}) = \frac{1}{(\Sigma w_i)^2} V(\Sigma w_i Z_i) = \frac{1}{(\Sigma w_i)^2}\left[\Sigma w_i^2\, V(Z_i)\right] \quad \ldots (*)$

The weights w_i's, $(i = 1, 2, \ldots, n)$ are so chosen that $\bar{Z}$ has minimum variance.

In order that $V(\bar{Z})$ is minimum for variations in w_i, we should have

$$\frac{\partial}{\partial w_i} V(\bar{Z}) = 0 ; \quad i = 1, 2, \ldots, k.$$

$$\Rightarrow \quad \frac{(\Sigma w_i)^2 2 w_i V(Z_i) - \left[\Sigma w_i^2 V(Z_i)\right] 2(\Sigma w_i)}{(\Sigma w_i)^4} = 0 \quad \text{or} \quad w_i V(Z_i) = \frac{\Sigma w_i^2\, V(Z_i)}{\Sigma w_i}, \text{ a constant.}$$

$$\therefore \quad w_i \propto \frac{1}{V(Z_i)} = (n_i - 3) ; i = 1, 2, \ldots, k. \quad \ldots (**)$$

Hence the minimum variance estimate of ξ is given by :

TABLE I.

SIGNIFICANT VALUES $t_\nu(\alpha)$ of t-Distribution

(TWO-TAIL AREAS)

$P[\,|\,t\,|\, > t_\nu(\alpha)] = \alpha$

d.f. (ν)	Probability (Level of Significance)					
	0·50	0·10	0·05	0·02	0·01	0·001
1	1·00	6·31	12·71	31·82	63·66	636·62
2	0·82	2·92	4·30	6·97	6·93	31·60
3	0·77	2·35	3·18	4·54	5·84	12·94
4	0·74	2·13	2·78	3·75	4·60	8·61
5	0·73	2·02	2·57	3·37	4·03	6·86
6	0·72	1·94	2·45	3·14	3·71	5·96
7	0·71	1·90	2·37	3·00	3·50	5·41
8	0·71	1·86	2·31	2·90	3·36	5·04
9	0·70	1·83	2·26	2·82	3·25	4·78
10	0·70	1·81	2·23	2·76	3·17	4·59
11	0·70	1·80	2·20	2·72	3·11	4·44
12	0·70	1·78	2·18	2·68	3·06	4·32
13	0·69	1·77	2·16	2·65	3·01	4·22
14	0·69	1·76	2·15	2·62	2·98	4·14
15	0·69	1·75	2·13	2·60	2·95	4·07
16	0·69	1·75	2·12	2·58	2·92	4·02
17	0·69	1·74	2·11	2·57	2·90	3·97
18	0·69	1·73	2·10	2·55	2·88	3·92
19	0.69	1.73	2.09	2.54	2.86	3·88
20	0·69	1·73	2·09	2·53	2·85	3·85
21	0·69	1·72	2·08	2·52	2·83	3·83
22	0·69	1·72	2·07	2·51	2·82	3·79
23	0·69	1·71	2·07	2·50	2·81	3·77
24	0·69	1·71	2·06	5·49	2·80	3·75
25	0·68	1·71	2·06	2·49	2·79	3·73
26	0·68	1·71	2·06	2·48	2·78	3·71
27	0·68	1·70	2·05	2·47	2·77	3·69
28	0·68	1·70	2·05	2·47	2·76	3·67
29	0·68	1·70	2·05	2·46	2·76	3·66
30	0·68	1·70	2·04	2·46	2·75	3·65
∞	0·67	1·65	1·96	2·33	2·58	3·29

TABLE II-A

SIGNIFICANT VALUES OF THE VARIANCE-RATIO F-DISTRIBUTION (RIGHT TAIL AREAS) 5 PER CENT POINTS

v_2 \ v_1	1	2	3	4	5	6	8	12	24	∞
1	161.40	199.50	215.70	224.60	230.20	234.00	238.90	243.90	249.00	254.30
2	18.51	19.00	19.16	19.25	19.30	19.35	19.37	19.41	19.45	19.50
3	10.13	9.55	9.28	9.12	9.01	8.94	8.84	8.74	8.64	8.55
4	7.71	6.94	6.59	6.39	6.26	6.16	6.04	5.91	5.77	5.65
5	6.61	5.79	5.41	5.19	5.05	4.95	4.82	4.68	4.53	4.96
6	5.99	5.14	4.76	4.53	4.39	4.28	4.15	4.00	3.84	3.67
7	5.59	4.74	4.35	4.12	3.97	3.87	3.73	3.57	3.41	3.23
8	5.32	4.46	4.07	3.84	3.69	3.58	3.44	3.28	3.12	2.93
9	5.12	4.26	3.865	3.63	3.48	3.37	3.23	3.07	2.90	2.71
10	4.96	4.10	3.71	3.48	3.33	3.22	3.07	2.91	2.74	2.54
11	4.84	3.98	3.59	3.365	3.20	3.09	2.95	2.79	2.61	2.40
12	4.75	3.88	4.49	3.26	3.11	3.00	2.85	2.69	2.50	2.30
13	4.67	3.80	3.41	3.18	3.02	2.92	2.77	2.60	2.42	2.21
14	4.60	3.74	3.34	3.11	2.96	2.85	2.70	2.53	2.35	2.13
15	4.54	3.68	3.29	3.06	2.90	2.79	2.64	2.48	2.29	2.07
16	4.49	3.63	3.24	3.01	2.85	2.74	2.59	2.42	2.24	2.01
17	4.45	3.59	3.20	2.96	2.81	2.70	2.55	2.38	2.19	1.96
18	4.41	3.55	3.16	2.93	2.77	2.66	2.51	2.34	2.15	1.92
19	4.38	3.52	3.13	2.90	2.74	2.63	2.48	2.31	2.11	1.88
20	4.35	3.49	3.10	2.87	2.71	2.60	2.45	2.28	2.08	1.84
21	4.32	3.47	3.07	2.84	2.68	2.57	2.42	2.25	2.05	1.81
22	4.30	3.44	3.05	2.82	2.66	2.55	2.40	2.23	2.03	1.76
23	4.28	3.42	3.03	2.80	2.64	2.53	2.38	2.20	2.00	1.76
24	4.26	3.40	3.01	2.78	2.62	2.51	2.36	2.18	1.98	1.73
25	4.24	3.38	2.99	2.76	2.60	2.49	2.34	2.16	1.96	1.71
26	4.22	3.37	2.98	2.74	2.59	2.47	2.32	2.15	1.95	1.60
27	4.21	3.35	2.96	2.73	2.57	2.46	2.30	2.13	1.93	1.67
28	4.20	3.34	2.95	2.71	2.56	2.44	2.29	2.12	1.91	1.65
29	4.18	3.33	2.93	2.70	2.54	2.43	2.28	2.10	1.90	1.64
30	4.17	3.32	2.92	2.69	2.53	2.42	2.27	2.09	1.89	1.62
40	4.08	3.23	2.84	2.61	2.45	2.34	2.18	2.00	1.79	1.51
60	4.00	3.15	2.76	2.53	2.37	2.25	2.10	1.92	1.70	1.30
120	3.92	3.87	2.68	2.45	2.29	2.17	2.02	1.83	1.62	1.25
240	3.84	2.99	2.60	2.37	2.21	2.09	1.94	1.75	1.52	1.00

TABLE II-B

SIGNIFICANT VALUES OF THE VARIANCE RATIO F-DISTRIBUTION (RIGHT TAIL AREAS) — 1 PER CENT POINTS

v_2 \ v_1	1	2	3	4	5	6	8	12	24	∞
1	4052	4999.5	5403	5625	5764	5859	5982	6106	6235	6366
2	98.50	99.00	99.17	99.25	99.30	99.33	99.37	99.42	99.46	99.50
3	34.12	30.82	29.46	28.71	28.24	27.91	27.49	27.05	26.60	26.13
4	21.20	18.00	16.69	15.98	15.52	15.21	14.80	14.37	13.93	13.46
5	16.26	13.27	12.06	11.39	10.97	10.67	10.29	9.89	9.47	9.02
6	13.75	10.92	9.78	9.15	8.75	8.47	8.10	7.72	7.31	6.88
7	12.25	9.95	8.45	7.85	7.46	7.19	6.84	6.47	6.07	5.65
8	11.26	8.65	7.59	7.01	6.63	6.37	6.03	5.67	5.28	4.86
9	10.56	8.02	6.99	6.42	6.06	5.80	5.47	5.11	4.73	4.31
10	10.04	7.56	6.55	5.99	5.64	5.39	5.06	4.71	4.33	3.91
11	9.65	7.21	6.22	5.67	5.32	5.07	4.74	4.40	4.02	3.60
12	9.33	6.93	5.95	5.41	5.06	4.82	4.50	4.16	3.78	3.36
13	9.07	6.70	5.74	5.21	4.86	4.62	4.30	3.96	3.59	3.17
14	8.86	6.51	5.56	5.04	4.69	4.46	4.14	3.80	3.43	3.00
15	8.68	6.36	5.42	4.89	4.56	4.32	4.00	3.67	3.29	2.87
16	8.53	6.23	5.29	4.77	4.44	4.20	3.89	3.55	3.18	2.75
17	8.40	6.11	5.18	4.67	4.34	4.10	3.79	3.46	3.08	2.65
18	8.29	6.01	5.09	4.58	4.25	4.01	3.71	3.37	3.00	2.57
19	8.18	5.93	5.01	4.50	4.17	3.94	3.63	3.30	2.92	2.49
20	8.10	5.85	4.94	4.43	4.10	3.87	3.56	3.23	2.86	2.42
21	8.02	5.78	4.87	4.37	4.04	3.81	3.51	3.17	2.80	2.36
22	7.95	5.72	4.82	4.31	3.99	3.76	3.45	3.12	2.75	2.31
23	7.88	5.66	4.76	4.26	3.94	3.71	3.41	3.07	2.70	2.26
24	7.82	5.61	4.72	4.22	3.90	3.67	3.36	3.03	2.66	2.21
25	7.77	5.57	4.68	4.18	3.85	3.63	3.32	2.99	2.62	2.17
26	7.72	5.53	4.64	4.14	3.82	3.59	3.29	2.96	2.58	2.13
27	7.68	5.49	4.60	4.11	3.78	3.56	3.26	2.93	2.55	2.10
28	7.64	5.45	4.57	4.07	3.75	3.53	3.23	2.90	2.52	2.06
29	7.60	5.42	4.54	4.04	3.73	3.50	3.20	2.87	2.49	2.03
30	7.56	5.39	4.51	4.02	3.70	3.47	3.17	2.84	2.47	2.01
40	7.31	5.18	4.31	3.83	3.51	3.29	2.99	2.66	2.29	1.80
60	7.08	4.98	4.13	3.65	3.34	3.12	2.82	2.50	2.12	1.60
120	6.85	4.79	3.95	3.48	3.17	2.96	2.66	2.34	1.95	1.38
∞	6.63	4.61	3.78	3.32	3.02	2.80	2.51	2.18	1.79	1.00

TABLE III—TRANSFORMATION FROM r to $Z = \frac{1}{2}\log_e\left(\frac{1+r}{1-r}\right)$

r	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0	0.0000	0.0100	0.0200	0.0300	0.0400	0.0500	0.0601	0.0701	0.0802	0.0902
·1	0.1003	0.1104	0.1206	0.1307	0.1409	0.1511	0.1614	0.7117	0.1820	0.1923
·2	0.2027	0.2132	0.2237	0.2342	0.2448	0.2554	0.2661	0.2769	0.2877	0.2986
·3	0.3005	0.3205	0.3316	0.3428	0.3541	0.3654	0.3769	0.3884	0.4001	0.4118
·4	0.4236	0.4356	0.4477	0.4599	0.4722	0.4847	0.4973	0.5101	0.5230	0.5361
·5	0.5493	0.5627	0.5763	0.5901	0.6042	0.6184	0.6328	0.6475	0.6625	0.6777
·6	0.6931	0.7089	0.7250	0.7414	0.7582	0.7753	0.7928	0.8107	0.8291	0.8480
·7	0.8673	0.8872	0.9076	0.9287	0.9505	0.9730	0.9962	1.0203	1.0454	1.0714
·8	1.0996	1.1270	1.1568	1.1881	1.2212	1.2562	1.2933	1.3331	1.3758	1.4219
·9	1.4722	1.5275	1.5890	1.6584	1.7380	1.8318	1.9459	2.0923	2.2976	2.6467