16·5. F-DISTRIBUTION
16·5·1 Derivation of Snedecor's F-distribution
16·5·2 Constants of F-distribution
16·5·3 Mode and Points of inflexion of F-Distribution
16·6. APPLICATIONS OF F-DISTRIBUTION
16·6·1 F-test for Equality of Two Population Variances
16·6·2 F-test for Testing the Significance of an Observed Multiple Correlation Coefficient
16·6·3 F-test for Testing the Significance of an Observed Sample Correlation Ratio
16·6·4 F-test for Testing the Linearity of Regression
16·6·5 F-test for Equality of Several Means
16·7 RELATION BETWEEN t AND F DISTRIBUTIONS
16·8 RELATION BETWEEN F AND χ^2 DISTRIBUTIONS
16·9 FISHER'S z-DISTRIBUTION
16·10 FISHER'S Z-TRANSFORMATION
16·10·1 Applications of Z-transformation
DISCUSSION & REVIEW QUESTIONS/
ASSORTED REVIEW PROBLEMS FOR SELF-ASSESSMENT

16·1. INTRODUCTION

The entire large sample theory was based on the application of "*Normal Test*" (c.f. § 14·9). However, if the sample size n is small, the distribution of the various statistics, e.g., $Z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$ 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 parent 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 :

$$t = \frac{\bar{x} - \mu}{S/\sqrt{n}} \qquad \ldots(16\text{·}1)$$

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

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

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, $\bar{x} \sim N(\mu, \sigma^2/n) \Rightarrow \frac{(\bar{x} - \mu)}{\sigma/\sqrt{n}} \sim N(0, 1)$

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

Also $\frac{ns^2}{\sigma^2}$ 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*), $\frac{t^2}{n-1}$ 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 :

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

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

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.*, $P(|t| > t_{0.05}) = 0.05 \Rightarrow P(|t| \le t_{0.05}) = 0.95$, 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·

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

where $t_{0.01}$ is the tabulated value of t for $\upsilon = (n-1)$ *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 :

$$t = \frac{\xi}{\sqrt{\chi^2/n}} \quad \text{...(16}$$

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 :

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

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.* $g_1(\cdot)$ of t becomes

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

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

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

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

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 $\bar{x} \sim N(\mu, \sigma^2/n)$, $\quad \xi = \dfrac{\bar{x}-\mu}{\sigma/\sqrt{n}} \sim N(0, 1)$ ...(*) and $\chi^2 = \dfrac{ns^2}{\sigma^2} = \sum\limits_{i=1}^{n}(x_i - \bar{x})^2/\sigma^2$ ...(**)

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

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

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.

$$\therefore \quad \mu_{2r+1} = 0, \quad (r = 1, 2, \ldots) \quad \text{...(16·4)}$$

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

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

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

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

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

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

$$= n^r \frac{\Gamma[(n/2)-r]\,\Gamma(r+\frac{1}{2})}{\Gamma(1/2)\,\Gamma(n/2)} \qquad \ldots (16\cdot4a)$$

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

$$= n^r \frac{(2r-1)(2r-3)\ldots3\cdot1}{(n-2)(n-4)\ldots(n-2r)}, \frac{n}{2} > r \qquad \ldots (16\cdot4b)$$

In particular

$$\mu_2 = n \cdot \frac{1}{(n-2)} = \frac{n}{n-2}, (n > 2) \qquad \ldots (16\cdot4c)$$

and $$\mu_4 = n^2 \frac{3\cdot1}{(n-2)(n-4)} = \frac{3n^2}{(n-2)(n-4)}, (n > 4) \qquad \ldots (16\cdot4d)$$

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 $\beta_2 = \lim_{n \to \infty} 3\left(\frac{n-2}{n-4}\right) = 3 \lim_{n \to \infty} \left[\frac{1-(2/n)}{1-(4/n)}\right] = 3$ ...(16·4e)

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

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

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

in terms of its μ_2 and β_2.

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

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

then t has Student's distribution with 2(m+1) degrees of freedom. Use the transformation to 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 consideration 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, $2\int_0^{\pi/2} \sin^p\theta \cos^q\theta\, d\theta = B\left(\frac{p+1}{2}, \frac{q+1}{2}\right)$ ...(1)

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

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

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

The moments of even order are given by :

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

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

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

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

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

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

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

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

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

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

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

Hence the *p.d.f.* of X transforms to

$$dF(t) = y_0 \frac{1}{\left(1+\frac{t^2}{n}\right)^m} \cdot \frac{a}{\sqrt{n}} \cdot \frac{dt}{\left(1+\frac{t^2}{n}\right)^{3/2}}$$

$$= \frac{1}{a\,B\left(m+1, \frac{1}{2}\right)} \cdot \frac{a}{\sqrt{n}} \cdot \frac{dt}{\left(1+\frac{t^2}{n}\right)^{m+(3/2)}}$$

$$= \frac{1}{\sqrt{n}\,B\left(\frac{n}{2}, \frac{1}{2}\right)} \cdot \frac{dt}{\left(1+\frac{t^2}{n}\right)^{(n+1)/2}}, -\infty < t < \infty \qquad ...(5)$$

which is the probability differential of Student's t-distribution with $n = 2(m+1)$ d.f. Hence the result.

For 2 d.f., *i.e.*, $n = 2$, we get $2(m+1) = 2 \Rightarrow m = 0$. Hence from (**), we get (for $m = 0$)

$$x = \frac{at}{(2+t^2)^{1/2}} \Rightarrow x = \frac{\sqrt{2}}{\sqrt{3}}a, \text{ when } t = 2.$$

$$\therefore \quad P(t \geq 2) = P\left[X \geq \sqrt{(2/3)}\,a\right] = \int_{a\sqrt{(2/3)}}^{a} dF(x) = \int_{a\sqrt{(2/3)}}^{a} \frac{1}{a\,B(1, \frac{1}{2})}\,dx$$

[From (*), since $m = 0$]

$$= \frac{1}{2a}\left(a - \frac{\sqrt{2}}{\sqrt{3}}a\right) = \frac{\sqrt{3}-\sqrt{2}}{2\sqrt{3}} \qquad \left[\because B\left(1, \frac{1}{2}\right) = \frac{\Gamma 1\,\Gamma(1/2)}{\Gamma(3/2)} = \frac{\Gamma(1/2)}{(1/2)\,\Gamma(1/2)} = 2\right]$$

For 4 d.f., *i.e.*, $n = 4$, we get $m = 1$. Proceeding exactly similarly we shall obtain

$$P(t \geq 2) = \frac{1}{2} - \frac{5\sqrt{2}}{16}.$$

Examlpe 16·2. *If the random variables X_1 and X_2 are independent and follow chi-square distribution with n d.f., show that $\sqrt{n}\,(X_1 - X_2)/2\sqrt{X_1X_2}$ is distributed as Student's t with n d.f., independently of $X_1 + X_2$.*

Solution. Since X_1 and X_2 are independent chi-square variates each with n d.f., their joint *p.d.f.* is given by :

$$p(x_1, x_2) = p_1(x_1) \times p_2(x_2)$$

$$= \frac{1}{2^n\,[\Gamma(n/2)]^2} \cdot e^{-(x_1+x_2)/2}\, x_1^{(n/2)-1}\, x_2^{(n/2)-1};\ 0 \leq x_1 < \infty, 0 \leq x_2 < \infty$$

Put $\quad u = \dfrac{\sqrt{n}\,(x_1 - x_2)}{2\sqrt{x_1 x_2}} \quad$ and $\quad v = x_1 + x_2$

$$\Rightarrow \quad x_1 = \frac{v}{2}\left[1 + \frac{1}{\sqrt{\left(1+\frac{n}{u^2}\right)}}\right], \quad x_2 = \frac{v}{2}\left[1 - \frac{1}{\sqrt{\left(1+\frac{n}{u^2}\right)}}\right]$$

Jacobian of transformation is : $\quad J = \dfrac{\partial(x_1, x_2)}{\partial(u, v)} = \dfrac{v}{2\sqrt{n}\left(1+\frac{u^2}{n}\right)^{3/2}}$

The joint *p.d.f.* of U and V becomes

$$g(u, v) = p(x_1, x_2)\ |\,J\,| = \frac{1}{2^{2n-1}\,\Gamma(n/2)\,\Gamma(n/2)\,\sqrt{n}} \cdot \frac{e^{-v/2}\,v^{n-1}}{\left(1+\frac{u^2}{n}\right)^{(n+1)/2}};\ -\infty < u < \infty, 0 \leq v < \infty$$

Using Legender's duplication formula, *viz.*,

$$\Gamma n = 2^{n-1}\,\Gamma(n/2)\,\Gamma\left(\frac{n+1}{2}\right)\Big/\sqrt{\pi} \Rightarrow \Gamma(n/2) = \frac{\Gamma n\,\sqrt{\pi}}{2^{n-1}\,\Gamma\left(\frac{n+1}{2}\right)}, \text{ we get}$$

$$2^{2n-1}\,\Gamma(n/2)\,\Gamma(n/2)\,\sqrt{n} = \frac{2^{2n-1}.\,\Gamma n\,\sqrt{\pi}}{2^{n-1}\,\Gamma\left(\frac{n+1}{2}\right)}\,\Gamma\left(\frac{n}{2}\right)\sqrt{n} = 2^n\,\Gamma n\,\sqrt{n}\,B\left(\frac{1}{2}, \frac{n}{2}\right) \left[\because \sqrt{\pi} = \Gamma\left(\frac{1}{2}\right)\right]$$

$$\therefore \quad g(u, v) = \left(\frac{1}{2^n\,\Gamma n}\,e^{-v/2}\,v^{n-1}\right)\left[\frac{1}{\sqrt{n}\,B\left(\frac{1}{2}, \frac{n}{2}\right)} \cdot \frac{1}{\left(1+\frac{u^2}{n}\right)^{(n+1)/2}}\right];\ 0 < v < \infty,\ -\infty < u < \infty.$$

$$\Rightarrow \quad g(u, v) = g_1(u)\,g_2(v), \qquad ...(i)$$

where $\quad g_1(u) = \dfrac{1}{\sqrt{n}\,\beta\left(\frac{1}{2}, \frac{n}{2}\right)} \cdot \dfrac{1}{\left(1+\frac{u^2}{n}\right)^{(n+1)/2}} \cdot -\infty < u < \infty \qquad ...(ii)$

and $\quad g_2(v) = \dfrac{1}{2^n\,\Gamma n}\,e^{-v/2}\,v^{n-1}, 0 < v < \infty \qquad ...(iii)$

(*i*) $\Rightarrow U = \sqrt{n}\,(X_1 - X_2)/2\sqrt{X_1X_2}$ and $V = X_1 + X_2$ are independently distributed.

(*ii*) $\Rightarrow U = \sqrt{n}\,(X_1 - X_2)/2\sqrt{X_1X_2} \sim t_n$, and

(*iii*) $\Rightarrow V = X_1 + X_2 \sim \gamma(a = \frac{1}{2}, n)$

Example 6·3. *If $I_x(p, q)$ represents the incomplete Beta function defined by :*

$$I_x(p, q) = \frac{1}{B(p, q)} \int_0^x t^{p-1}\,(1-t)^{q-1}\,dt\ ;\ p > 0, q > 0 \qquad ...(*)$$

show that the distribution function F(.) of Student's t-distribution is given by :

$$F(t) = 1 - \frac{1}{2} I_x\left(\frac{n}{2}, \frac{1}{2}\right), \text{ where } x = \left(1+\frac{t^2}{n}\right)^{-1}.$$

Solution. If $f(.)$ is *p.d.f.* of Student's t-distribution with n *d.f.*, then

$$F(t) = \int_{-\infty}^{t} f(u)\,du = 1 - \int_t^{\infty} f(u)\,du = 1 - \frac{1}{\sqrt{n}\,B\left(\frac{1}{2}, \frac{n}{2}\right)} \int_t^{\infty} \left(1+\frac{u^2}{n}\right)^{-(n+1)/2} du \qquad ...(**)$$

Put $\quad 1 + \dfrac{u^2}{n} = \dfrac{1}{z} \quad \Rightarrow \quad u = \sqrt{n}\left(\dfrac{1-z}{z}\right)$

Also $\quad \dfrac{2u\,du}{n} = \dfrac{-dz}{z^2} \quad \Rightarrow \quad du = -\dfrac{n\,dz}{2uz^2} = -\dfrac{\sqrt{n}}{2}\left(\dfrac{z}{1-z}\right)^{1/2} \cdot \dfrac{dz}{z^2}$

Substituting in (**), we get :

$$F(t) = 1 - \frac{1}{\sqrt{n}B\left(\frac{1}{2}, \frac{n}{2}\right)} \int_{\left(1+\frac{t^2}{n}\right)^{-1}}^{0} z^{(n+1)/2}\left\{-\frac{\sqrt{n}}{2}\,z^{-3/2}\,(1-z)^{-1/2}\right\} dz$$

$$= 1 + \frac{1}{2\,B\left(\frac{1}{2}, \frac{n}{2}\right)} \int_{\left(1+\frac{t^2}{n}\right)^{-1}}^{0} z^{(n/2)-1}\,(1-z)^{-1/2}\,dz$$

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

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

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

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

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.

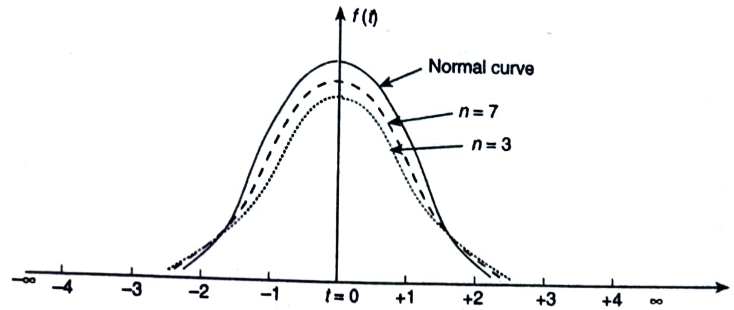

Fig. 16·1 : Graph of t-distribution

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 :

$$P[\,|t| > t_v(\alpha)] = \alpha \qquad \ldots(16·5)$$

$$\Rightarrow \quad P[\,|t| \le t_v(\alpha)] = 1 - \alpha \qquad \ldots(16·5a)$$

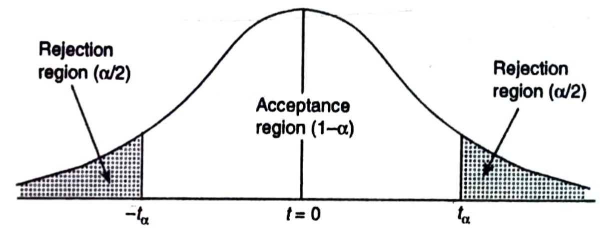

Fig. 16·2 : Critical values of t-distribution

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)

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

$$\Rightarrow \quad P[t > t_v(\alpha)] = \alpha/2 \qquad \therefore\; P[t > t_v(2\alpha)] = \alpha \qquad \ldots(16·5b)$$

$t_v(2\alpha)$ (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 (0·05) for single-tail test $= t_8$ (0·10) for two-tail test $= 1·86$

t_{15}(0·01) for single-tail test $= t_{15}$ (0·02) for two-tail test $= 2·60$.

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 ($i = 1, 2, ..., n$) 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
$$t = \frac{\bar{x} - \mu_0}{S/\sqrt{n}}, \qquad ...(16·6$$

where
$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \quad \text{and} \quad S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2, \qquad ...[16·6$$

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

$$S^2 = \frac{1}{n-1}\left[\Sigma(x_i - \bar{x})^2\right] = \frac{1}{n-1}\left[\Sigma x_i^2 - \frac{(\Sigma x_i)^2}{n}\right] \qquad ...(16·6$$

$$= \frac{1}{n-1}\left[\Sigma d_i^2 - \frac{(\Sigma d_i)^2}{n}\right], \text{ since variance is independent of change of origin.} \qquad ...(16·6$$

Also, in this case
$$\bar{x} = A + \frac{\Sigma d_i}{n}. \qquad ...(16·6$$

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

$$\therefore \quad \frac{S^2}{n} = \frac{s^2}{n-1} \qquad ...(16·6$$

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

$$t = \frac{\bar{x} - \mu_0}{\sqrt{S^2/n}} = \frac{\bar{x} - \mu_0}{\sqrt{s^2/(n-1)}} \sim t_{n-1} \qquad ...(16·6$$

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 :

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

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

Alternative Hypothesis, $H_1 : \mu \neq 0·700$

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

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

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.*, $3·15 > t_0$, 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 : $n = 22$, $\bar{x} = 153·7$, $s = 17·2$.

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

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

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

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

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

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

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

Conclusion. Since calculated t is less than tabulated $t_{0{\cdot}05}$ 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 :

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

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

Remark. Alter 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. inch[es]

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

Null Hypothesis, $H_0 : \mu = 43{\cdot}5$ inches, *i.e.*, the data are consistent with 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 reje[cted] at 5% level of significance and we conclude that the assumption of mean of 43·5 in[ches] for the population is not reasonable.

Remark. Since calculated $|t|$ is less than 2·947, null hypothesis ($\mu = 43{\cdot}5$) may be acce[pted] 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 $(i = 1, 2, ..., n_1)$ and y_j, $(j = 1, 2, ..., n_2)$ of sizes n_1 and n_2 h[ave] 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 nor[mal] populations with means* μ_X *and* μ_Y *and under the assumption that the population varian[ces are] equal, i.e.,* $\sigma_X^2 = \sigma_Y^2 = \sigma^2$ (say), the statistic

$$t = \frac{(\bar{x}-\bar{y}) - (\mu_X - \mu_Y)}{S\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \quad \text{...(16}$$

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

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

is an unbiased estimate of the common population variance σ^2, follows Student['s] 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 $E(\bar{x}-\bar{y}) = E(\bar{x}) - E(\bar{y}) = \mu_X - \mu_Y$

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

[The covariance term vanishes since samples are independent.]

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

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

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

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 $(i = 1, 2, ..., n_1)$, and y_j $(j = 1, 2, ..., n_2)$, 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 the same means, i.e.,* $\mu_X = \mu_Y$, *or (b) the sample means* $\bar{x}$ *and* $\bar{y}$ *do not differ significantly,* the statistic :

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

where symbols are defined in (16·7a), 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 th... fundamental assumptions :

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

(ii) The population variances are equal and unknown, *i.e.,* $\sigma_X^2 = \sigma_Y^2 = \sigma^2$ (say), where ... 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 ... test the equality of population variances by applying *F*-test. (*c.f* § 16·6·1) If the variances do ... come out to be equal then *t*-test becomes invalid and in that case Behren's '*d*'-test based ... fiducial intervals is used. For practical problems, however, the assumptions (*i*) and (*ii*) are ta... for granted.

16·3·3. Paired t-test for Difference of Means. Let us now consider the c... when (*i*) the sample sizes are equal, *i.e.,* $n_1 = n_2 = n$ (say), and (*ii*) the two samples ... not independent but the sample observations are paired together, *i.e.,* the pair... observations (x_i, y_i), $(i = 1, 2, \ldots, n)$ corresponds to the same (*i*th) sample unit. T... 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, i... inducing sleep. Let x_i and y_i $(i = 1, 2, \ldots, n)$ be the readings, in hours of sleep, on the i... individual, before and after the drug is given respectively. Here instead of applyi... the difference of the means test discussed in § 16·3·2, we apply the paired *t*-test gi... 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... the drug is not responsible for these increments,* the statistic :

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

where

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

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 die... 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 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) :

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

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

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 ...% level of significance and we may conclude that the two diets do not differ ...gnificantly 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 ...omputing $\Sigma(x - \bar{x})^2$ and $\Sigma(y - \bar{y})^2$ is used. In case $\bar{x}$ and (or) $\bar{y}$ comes out to be fractional, then ...e 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 ...nd 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 ...garding length of life ?

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

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

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

$$= \frac{1}{n_1+n_2-2}\left(n_1 s_1^2 + n_2 s_2^2\right) = \frac{1}{13}\left[8\times(36)^2 + 7\times(40)^2\right] = 1{,}659\ldots$$

$$\therefore \quad t = \frac{1234-1036}{\sqrt{1659.08\left(\frac{1}{8}+\frac{1}{7}\right)}} = \frac{198}{\sqrt{1659.08\times 0.2679}} = 9.39$$

Tabulated value of t for 13 *d.f.* at 5% level of significance for right (single)-tail test is 1·77. [This is the value of $t_{0.10}$ 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, 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 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
			72	6	36
			73	7	49
Total	0	60	Total	18	186

$$\therefore \quad \bar{x} = A + \frac{\Sigma d}{n_1} = 68 + 0 = 68 \qquad \bar{y} = B + \frac{\Sigma D}{n_2} = 66 + \frac{18}{10} = 67.8$$

$$\text{and } \Sigma(x-\bar{x})^2 = \Sigma d^2 - \frac{(\Sigma d)^2}{n_1} = 60 - 0 = 60 \qquad \text{and } \Sigma(y-\bar{y})^2 = \Sigma D^2 - \frac{(\Sigma D)^2}{n_2} = 186 - \frac{324}{10} = 153.6$$

$$S^2 = \frac{1}{n_1+n_2-2}\left[\Sigma(x-\bar{x})^2 + \Sigma(y-\bar{y})^2\right] = \frac{1}{14}(60+153.6) = 15.2571$$

$$\therefore \quad t = \frac{68-67.8}{\sqrt{15.2571}\left(\frac{1}{6}+\frac{1}{10}\right)^{1/2}} = \frac{0.2}{\sqrt{15.2571\times 0.2667}} = 0.099$$

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* $H_0 : \mu_1 - \mu_2 \geq 0.05$, [*i.e.*, the claim is substantiated]

Alternative Hypothesis $H_1 : \mu_1 - \mu_2 < 0.05$ (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 $$S^2 = \frac{n_1 s_1^2 + n_2 s_2^2}{n_1+n_2-2} = \frac{25\times(0.003)^2 + 25\times(0.002)^2}{25+25-2} = \frac{0.000225+0.0001}{48} = 0.0000067$$

$$\therefore \quad t = \frac{(0.083-0.136)-0.05}{\sqrt{\left\{0.0000067\left(\frac{1}{25}+\frac{1}{25}\right)\right\}}} = -\frac{0.103}{0.00071} = -145.07$$

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 ...

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

Alternative Hypothesis, $H_1 : \mu_X < \mu_Y$, *i.e.*, the stimulus results in an increase ... 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
d^2	25	4	64	1	9	0	4	1	25	0	16	36

$$\bar{d} = \frac{1}{n}\Sigma d = 2{\cdot}58 \quad \text{and} \quad S^2 = \frac{1}{n-1}\left[\Sigma d^2 - \frac{(\Sigma d)^2}{n}\right] = \frac{1}{11}\left[185 - \frac{(31)^2}{12}\right] = 9{\cdot}5...$$

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

Tabulated $t_{0{\cdot}05}$ for 11 *d.f.* for single-tail test is 1·80. [This is the value of $t_{0{\cdot}1}$ ... *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 signi... Hence we conclude that the stimulus will, in general, be accompanied by an i... in blood pressure.

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

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

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

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

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

$$\Sigma(x - \bar{x})^2 = \Sigma d^2 - \frac{(\Sigma d)^2}{n_1} = 37 - \frac{49}{8} = 30{\cdot}875 \qquad \Sigma(y - \bar{y})^2 = \Sigma D^2 - \frac{(\Sigma D)^2}{n_2} = 23 - \frac{49}{8} = 16{\cdot}875$$

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

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

Tabulated $t_{0{\cdot}05}$ 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 $-1{\cdot}76$, 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 $t_{0.05}$ 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

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

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.

Example 16·17. (*a*) *A random sample of 27 pairs of observations from a normal population gave a correlation coefficient of 0·6. Is this significant of correlation in the population ?*

(*b*) *Find the least value of r in a sample of 18 pairs of observations from a bi-variate normal population, significant at 5% level of significance.*

Solution. (*a*) We set up the null hypothesis, $H_0 : \rho = 0$, *i.e.,* the *observed sample correlation coefficient is not significant of any correlation in the population.*

Under H_0 : $t = \dfrac{r\sqrt{(n-2)}}{\sqrt{(1-r^2)}} \sim t_{(n-2)} \quad \therefore \quad t = \dfrac{0{\cdot}6\sqrt{27-2}}{\sqrt{(1-0{\cdot}36)}} = \dfrac{3}{\sqrt{0{\cdot}64}} = 3{\cdot}75.$

Tabulated $t_{0.05}$ for $(27-2) = 25$ $d.f.$ is 2·06.

Conclusion. Since calculated t is much greater than the tabulated t, it is significant and hence H_0 is discredited at 5% level of significance. Thus we conclude that the variables are correlated in the population.

(*b*) Here $n = 18$. From the tables $t_{0.05}$ for $(18-2) = 16$ $d.f.$ is 2·12.

Under $H_0 : \rho = 0, \quad t = \dfrac{r\sqrt{(n-2)}}{\sqrt{(1-r^2)}} \sim t_{(n-2)} = t_{16}$

In order that the calculated value of t is significant at 5% level of significance, we should have

$$\left|\frac{r\sqrt{(n-2)}}{\sqrt{(1-r^2)}}\right| > t_{0.05} \qquad \Rightarrow \left|\frac{r\sqrt{16}}{\sqrt{(1-r^2)}}\right| > 2{\cdot}12$$

$$\Rightarrow 16r^2 > (2{\cdot}12)^2(1-r^2) \quad \text{or} \quad 20{\cdot}493r^2 > 4{\cdot}493 \quad \text{or} \quad r^2 > \frac{4{\cdot}493}{20{\cdot}493} = 0{\cdot}2192$$

Hence $\qquad |r| > 0{\cdot}4682.$

Example 16·18. *A coefficient of correlation of 0·2 is derived from a random sample of 625 pairs of observations. (i) Is this value of r significant ? (ii) What are the 95% and 99% confidence limits to the correlation coefficient in the population ?*

Solution. *Under the null hypothesis* $H_0 : \rho = 0$, *i.e., the value of r* = 0·2 *is not significant, the test statistic is :*

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} \sim t_{n-2}$$

$$\therefore \qquad t = \frac{0{\cdot}2 \times \sqrt{(625-2)}}{\sqrt{(1-0{\cdot}04)}} = 5{\cdot}09$$

Since $d.f. = 625 - 2 = 623$, the significant values of t are same as in the case of normal distribution, *viz.,* $t_{0.05} = 1{\cdot}96$ and $t_{0.01} = 2{\cdot}58$. Since calculated t is much greater than these values; it is highly significant. Hence $H_0 : \rho = 0$ is rejected and we conclude that the sample correlation is significant of correlation in the population.

95% Confidence Limits for ρ *(population correlation coefficient)* are :

$$r \pm 1{\cdot}96 \text{ S.E. } (r) = r \pm 1{\cdot}96\,(1-r^2)/\sqrt{n} \qquad [\text{Since } n \text{ large}]$$
$$= 0{\cdot}2 \pm (1{\cdot}96 \times 0{\cdot}96/\sqrt{625}) = 0{\cdot}2 \pm 0{\cdot}075 = (0{\cdot}125, 0{\cdot}275)$$

99% Confidence Limits for ρ are :

$$0{\cdot}2 \pm 2{\cdot}58 \times 0{\cdot}0384 = 0{\cdot}2 \pm 0{\cdot}099 = (0{\cdot}101, 0{\cdot}299)$$

16·3·5. t-test for Testing the Significance of an Observed Regression Coefficient. Here the problem is to test if a random sample (x_i, y_i), $(i = 1, 2, \ldots, n)$ has been drawn from a bivariate normal population in which regression coefficient of Y on X is β.

$$\Rightarrow \quad dt = \sqrt{(n-2)}\ \frac{dr}{\sqrt{(1-r^2)}}\left(1+\frac{r^2}{1-r^2}\right) \quad \Rightarrow \quad dr = \frac{1}{\sqrt{n-2}}\ (1-r^2)^{3/2}\ dt$$

$$\Rightarrow \quad dt = \sqrt{(n-2)} \times \frac{dr}{(1-r^2)^{3/2}}$$

As r ranges from -1 to 1, from (*i*), t ranges from $-\infty$ to ∞.

When $\rho = 0$, the *p.d.f.* of 'r' is given by (16·12) and it transforms to

$$dG(t) = \frac{1}{B\left(\frac{1}{2}, \frac{n-2}{2}\right)}(1-r^2)^{(n-4)/2}\frac{1}{\sqrt{(n-2)}}(1-r^2)^{3/2}\ dt$$

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

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

which is the *p.d.f.* of *t*-distribution with $(n-2)$ *d.f.*

Hence $\quad t = \frac{r}{\sqrt{(1-r^2)}} \cdot \sqrt{(n-2)} \sim t_{(n-2)}$ [From

Example 16·19. (*a*) *If* (x_i, y_i) *is a random sample drawn from an uncorrelated bivariate normal population, derive the distribution of :*

$$r = \frac{\Sigma(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\Sigma(x_i - \bar{x})^2\ \Sigma(y_i - \bar{y})^2}}$$

(*b*) *Further, when* $n = 5$ *and if* $P(\,|\,r\,|\geq C) = \alpha$, *show that* C *is a root of the equation*

$$C\sqrt{(1-C^2)} + \sin^{-1} C + \frac{\pi(\alpha-1)}{2} = 0$$

Solution. (*a*) *c.f.* § 16·4, page 16·26.

(*b*) $P(\,|\,r\,| \geq C) = 1 - P(\,|\,r\,| \leq C) = 1 - P(-C \leq r \leq C)$

$$= 1 - 2\,P(0 \leq r \leq C) = 1 - 2\int_0^C f(r)\,dr$$

[$\because$ $f(r)$ is symmetrical about $r = 0$]

When $n = 5$, $\quad f(r) = \frac{1}{B\left(\frac{1}{2}, \frac{3}{2}\right)} \cdot (1-r^2)^{1/2}$ [*c.f.* Equation 16·12, page 16·26]

$$\therefore \quad P(\,|\,r\,| \geq C) = 1 - 2\frac{\Gamma(2)}{\Gamma(1/2)\Gamma(3/2)}\int_0^C (1-r^2)^{1/2}\,dr$$

$$= 1 - 2 \times \frac{1}{\frac{1}{2}\pi}\left[\frac{1}{2}r(1-r^2)^{1/2} + \frac{1}{2}\sin^{-1} r\right]_0^C \qquad \left[\because \Gamma\left(\tfrac{1}{2}\right) = \sqrt{\pi}\right]$$

$$= 1 - \frac{4}{\pi}\left[\tfrac{1}{2}C(1-C^2)^{1/2} + \tfrac{1}{2}\sin^{-1} C\right] = \alpha \qquad \text{(Given)}$$

$$\therefore \quad 1 - \frac{2}{\pi}\left[C(1-C^2)^{\frac{1}{2}} + \sin^{-1} C\right] = \alpha \quad \Rightarrow \quad C(1-C^2)^{1/2} + \sin^{-1} C + (\alpha - 1)\frac{\pi}{2} = 0.$$

16·5. F-DISTRIBUTION

Definition. *If X and Y are two independent chi-square variates with* v_1 *and* v_2 *d.f. respectively, then F-statistic is defined by*

$$F = \frac{X/v_1}{Y/v_2} \qquad \ldots(16\cdot13)$$

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 :

$$f(F) = \frac{\left(\frac{v_1}{v_2}\right)^{\frac{v_1}{2}}}{B\left(\frac{v_1}{2}, \frac{v_2}{2}\right)} \cdot \frac{F^{\frac{v_1}{2}-1}}{\left(1+\frac{v_1}{v_2}F\right)^{(v_1+v_2)/2}}, \quad 0 \leq F < \infty \qquad \ldots[16\cdot13(a)]$$

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 v_1 and v_2 *d.f.* respectively, their joint probability density function is given by :

$$f(x, y) = \left\{\frac{1}{2^{v_1/2}\,\Gamma(v_1/2)}\exp(-x/2)\,x^{(v_1/2)-1}\right\} \times \left\{\frac{1}{2^{v_2/2}\,\Gamma(v_2/2)}\exp(-y/2)\,y^{(v_2/2)-1}\right\}$$

$$= \frac{1}{2^{(v_1+v_2)/2}\,\Gamma(v_1/2)\,\Gamma(v_2/2)}\exp\{-(x+y)/2\} \times x^{(v_1/2)-1}\,y^{(v_2/2)-1}, \quad 0 \leq (x, y) < \infty$$

Let us make the following transformation of variables :

$$F = \frac{x/v_1}{y/v_2} \text{ and } u = y, \text{ so that } 0 \leq F < \infty,\ 0 < u < \infty \qquad \therefore \quad x = \frac{v_1}{v_2}Fu \text{ and } y = u$$

Jacobian of transformation J is given by :

$$J = \frac{\partial(x, y)}{\partial(F, u)} = \begin{vmatrix} \frac{v_1}{v_2}u & 0 \\ \frac{v_1}{v_2}F & 1 \end{vmatrix} = \frac{v_1 u}{v_2}$$

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

$$g(F, u) = \frac{1}{2^{(v_1+v_2)/2}\,\Gamma(v_1/2)\,\Gamma(v_2/2)}\exp\left\{-\frac{u}{2}\left(1+\frac{v_1}{v_2}F\right)\right\} \times \left(\frac{v_1}{v_2}Fu\right)^{(v_1/2)-1} u^{(v_2/2)-1}\ |\,J\,|$$

$$= \frac{(v_1/v_2)^{v_1/2}}{2^{(v_1+v_2)/2}\,\Gamma(v_1/2)\,\Gamma(v_2/2)}\exp\left\{-\frac{u}{2}\left(1+\frac{v_1}{v_2}F\right)\right\} \times u^{[(v_1+v_2)/2]-1}\,F^{(v_1/2)-1};\quad 0 < u < \infty,\ 0 \leq F < \infty$$

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

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

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

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

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

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

$\therefore \quad \frac{\nu_1}{\nu_2}F = \frac{x}{y}$, 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 :

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

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

16·5·2. Constants of F-distribution.

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

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

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

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

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

$$= \left(\frac{\nu_2}{\nu_1}\right)^r \cdot \frac{1}{B\left(\frac{\nu_1}{2},\frac{\nu_2}{2}\right)} \cdot B\left(r+\frac{\nu_1}{2}, \frac{\nu_2}{2}-r\right), \quad \nu_2 > 2r \quad \ldots(16·14)$$

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

$$\therefore \quad \mu_r' = \left(\frac{\nu_2}{\nu_1}\right)^r \cdot \frac{\Gamma[r+(\nu_1/2)]\,\Gamma[(\nu_2/2)-r]}{\Gamma(\nu_1/2)\,\Gamma(\nu_2/2)}; \; r < \frac{\nu_2}{2} \quad \Rightarrow \quad \upsilon_2 > 2r \quad \ldots(16·15)$$

In particular

$$\mu'_1 = \frac{\nu_2}{\nu_1} \cdot \frac{\Gamma[1+(\nu_1/2)]\,\Gamma[(\nu_2/2)-1]}{\Gamma(\nu_1/2)\,\Gamma(\nu_2/2)} = \frac{\nu_2}{\nu_2-2}, \quad \nu_2 > 2 \quad \ldots(16·15a)$$

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

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

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

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

$$\therefore \quad \mu_2 = \mu_2' - \mu_1'^2 = \frac{\nu_2^2(\nu_1+2)}{\nu_1(\nu_2-2)(\nu_2-4)} - \frac{\nu_2^2}{(\nu_2-2)^2} = \frac{2\nu_2^2\,(\nu_2+\nu_1-2)}{\nu_1(\nu_2-2)^2(\nu_2-4)}, \quad \nu_2 > 4 \quad \ldots(16·15b)$$

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 $= \dfrac{\nu_2(\nu_1-2)}{\nu_1(\nu_2+2)}$ ...(16·16)

(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 $\chi^2_{(1)}$, 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 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 w test (i) whether two independent samples x_i, $(i = 1, 2, ..., n_1)$ and y_j, $(j = 1, 2, ...$ have been drawn from the normal populations with the same variance σ^2 (sa (ii) whether the two independent estimates of the population varianc homogeneous or not.

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

$$F = \frac{S_X^2}{S_Y^2} \quad ...(1$$

where $\quad S_X^2 = \frac{1}{n_1 - 1}\sum_{i=1}^{n_1}(x_i - \bar{x})^2 \quad$ and $\quad S_Y^2 = \frac{1}{n_2 - 1}\sum_{j=1}^{n_2}(y_j - \bar{y})^2 \quad ...(1$

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

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

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

Since $\frac{n_1 s_X^2}{\sigma_X^2}$ and $\frac{n_2 s_Y^2}{\sigma_Y^2}$ are independent chi-square variates with $(n_1 - 1)$ $(n_2 - 1)$ *d.f.* respectively, F follows Snedecor's F-distribution with $(n_1 - 1, n_2 - 1)$ (*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 numerator and n_1 corresponds to the greater variance.

By comparing the calculated value of F obtained by using (16·17) for the two samples, with the tabulated value of F for (n_1, n_2) *d.f.* at certain level of significance (5% o 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 end of the chapter) give the critical values of F for the right-tailed test, *i.e.*, the critical re determined by the right-tail areas. Thus the significant value $F_\alpha(n_1, n_2)$ at level of signific 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.

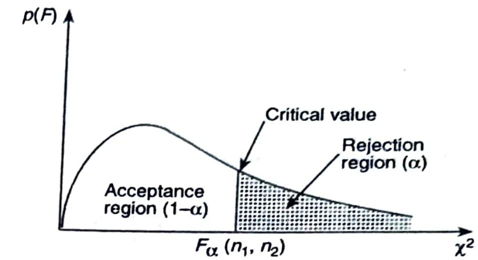

Fig. 16·3 : Critical Values of F-Distribution

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

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

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* $P(F_{10, 8} \geq 3·35) = 0·05$ *and* $P(F_{8, 10} \geq 3·07) = 0·05$.*]*

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 : $n_1 = 11$, $n_2 = 9$, $s_X = 0·8$ and $s_Y = 0·5$.

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

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

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

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

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

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

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

We are given the tabulated (significant) values :

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

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

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

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 $\Sigma(x - \bar{x})^2 = 84{\cdot}4, \quad \Sigma(y - \bar{y})^2 = 102{\cdot}6$

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

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

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 $F_{0{\cdot}05}$ for (7, 9) *d.f.* is 3·29.

Since calculated $F < F_{0{\cdot}05}$, 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* : $F_{0{\cdot}05}(9, 11) = 2{\cdot}90$, $F_{0{\cdot}05}(11, 9) = 3{\cdot}10$ (approx.) and $t_{0{\cdot}05}(20) = 2{\cdot}086$, $t_{0{\cdot}05}(22) = 2$[illegible]

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

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

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.