stribution and the values of statistic are always considered the best estim
s of the population from which the sample is drawn. But it is always n
of small size is drawn, then instead of using central limit theorem in wh
iately a normal distribution a new technique is available, which involves t
m, which is explain in the upcoming section.

Number of Degrees of Freedom

$_2, \ldots, x_n$ be a sample of size n then the degrees of freedom (d.f.) is (n
$x_3 + x_4 = 20$, we can assign any arbitrary values to $x_1 + x_2 + x_3$ and esti
e, in this case $x_1 + x_2 + x_3$ are free and independent choices to find the va
e, these are degrees of freedom, and here $d.f.$ is 3. So, we can say the n
is nothing but is the number of values in a sample which are assigned arbi
om sample of size n from a population, then to calculate sample mean
is used therefore to estimate the population variance based on this samp
are left.

STUDENT'S t-DISTRIBUTION

with mean μ & variance σ^2 of $\bar{x}$

$x_2, x_3, \ldots, x_n$ be a random sample of small size n from a normal population
s^2 be the sample mean and sample variance, respectively, then the statistic
owing:

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}, \text{ where } \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \text{ and } s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \ldots$$

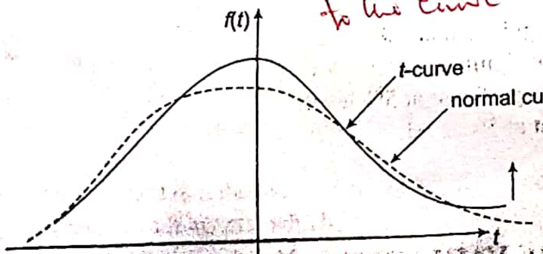
has t-distribution with $(n-1)$ $d.f.$ and the distribution of t is given by

$$f(t) = \frac{A}{\left(1 + \frac{t^2}{v}\right)^{\frac{v+1}{2}}}; -\infty < t < \infty$$

here $v = n - 1$ and A is a constant such that the area under the curve is uni

2.1 Properties of t-Distribution

1. Since $f(-t) = f(t)$, the probability curve is symmetrical about the line $t=$ … decreases rapidly and tends to zero as $t \to \infty$, so that t-axis is an asymp… to the curve



…um value of t-curve also at $t = 0$, hence, the mode and mean of t-distribution are same.

…, then $f(t) = \frac{A \cdot e^{-\frac{1}{2}t^2}}{\sqrt{2\pi}}, -\infty < t < \infty$, i.e., for large values of n, t-distribution
…ndard normal distribution. Therefore for large values of n instead of using t-table,
normal table.

$\int f(t)\,dt$

…of t_α have been calculated for various values of p and different values of $v = n - 1$
1 to 30.

…symmetrical about $t = 0$, then all the moments of odd order about origin/and mean

$$\mu'_{2k+1} = 0, k = 0, 1, 2, \ldots = \mu_{2k+1}$$

…mean $= 0$

…s of even order are given by

$$\mu_{2k} = \frac{nk(2k-1)(2k-3)\cdots 3.1}{(n-2)(n-4)\cdots(n-2k)}, \frac{n}{2} > k$$

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

$$\mu_4 = \frac{n^2 \cdot 3.1}{(n-2)(n-4)} = \frac{3n^2}{(n-2)(n-4)}, n > 4$$

$= 0$ and $\beta_2 = \frac{\mu_4}{\mu_2^2} = 3\frac{(n-2)}{(n-4)}, n > 4$

$= 0$ and $\beta_2 = 3$

> 1 i.e., the variance of t-distribution is greater than the variance of standard
…bution.

> 3, therefore, t-distribution is more flat on the top than normal curve.

$= \alpha$

$t_{\nu,\alpha}] = 1 - \alpha$

…tion is generally used to test the hypothesis about the mean when the variance of
…n is unknown.
…om which sample is drawn is always taken normal.

…ence Test of a Sample Mean

…random sample of size n from a normal population with a specified population

The maximum value of t-curve also at $t=0$, hence, the mode and mean of t-distribution are same.

2. When $n \to \infty$, then $f(t) = \frac{A \cdot e^{-\frac{1}{2}t^2}}{\sqrt{2\pi}}; -\infty < t < \infty$, i.e., for large values of n, t-distribution tends to standard normal distribution. Therefore for large values of n instead of using t-table, we can use normal table.

3. $p[T > t_\alpha] = \int_\alpha^\infty f(t)\,dt$.

The values of t_α have been calculated for various values of p and different values of $v = n-1$ (d.f.) from 1 to 30.

4. Since $f(t)$ is symmetrical about $t=0$, then all the moments of odd order about origin/and mean are zero, i.e.,

$$\mu'_{2k+1} = 0, k = 0, 1, 2, \ldots = \mu_{2k+1},$$

i.e. $\mu'_1 = \text{mean} = 0$

The moments of even order are given by

$$\mu_{2k} = \frac{nk(2k-1)(2k-3)\cdots 3.1}{(n-2)(n-4)\cdots(n-2k)}, \frac{n}{2} > k$$

In particular

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

$$\mu_4 = \frac{n^2 \cdot 3.1}{(n-2)(n-4)} = \frac{3n^2}{(n-2)(n-4)}, n > 4$$

$$\therefore \beta_1 = \frac{\mu_3^2}{\mu_2^3} = 0 \text{ and } \beta_2 = \frac{\mu_4}{\mu_2^2} = 3\frac{(n-2)}{(n-4)}, n > 4$$

Remark:

1. As $n \to \infty$, $\beta_1 = 0$ and $\beta_2 = 3$.
2. For $n > 2$, $\mu_2 > 1$ i.e., the variance of t-distribution is greater than the variance of standard normal distribution.
3. For $n > 4$, $\beta_2 = 3$, therefore, t-distribution is more flat on the top than normal curve.
4. $p\left[|t| > t_{v,\alpha}\right] = \alpha$

   $\Rightarrow \quad p\left[|t| \le t_{v,\alpha}\right] = 1 - \alpha$
5. The t-distribution is generally used to test the hypothesis about the mean when the variance of the population is unknown.
6. Population from which sample is drawn is always taken normal.

7.22.2 Significance Test of a Sample Mean

Let $x_1, x_2, \ldots, x_n$ be a random sample of size n from a normal population with a specified population mean μ_0. To test the hypothesis $\mu = \mu_0$ or there is no significance difference between sample mean $\bar{x}$ and population mean μ_0.

Scanned by CamScanner

(i) Compute the statistic

$$t = \frac{\bar{x}-\mu_0}{s/\sqrt{n}}, \text{ where } s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2$$

(ii) Find the value of P for the given degrees of freedom from the table.

(a) if $t_{cal} > t_{.05}$, then the difference between sample mean $\bar{x}$ and specified population mean μ_0 is significant at 5% level of significance; otherwise not significant.

(b) if $t_{cal} > t_{.01}$, then the difference between sample mean $\bar{x}$ and specified population mean μ_0 is significant at 1% level of significance otherwise not significant.

(c) In general if $t_{cal} > t_\alpha$, then the difference between sample mean $\bar{x}$ and population mean μ_0 is significant at level of significance α.

Example 14 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 at 5% level of significance.

Solution Here, $n = 22, \bar{x} = 153.7, s = 17.2, \mu_0 = 146.3$

$$t = \frac{\bar{x}-\mu_0}{s/\sqrt{\nu}}, \text{ where } \nu = n-1 = 21$$

$$= \frac{153.7-146.3}{17.2/\sqrt{21}} = 9.03$$

t_{tab} at $\nu = 21 = 2.08$

$t_{cal} = 9.03 > 2.08 \Rightarrow$ It is highly significant. We conclude that the advertising campaign was highly successful.

Example 15 The heights of 10 males of a given locality are found to be 70, 67, 62, 68, 61, 68, 70, 64, 64 and 66 inches. Is it reasonable to believe that the average height is greater than 64 inches? Test at 5% level of significance assuming that for 9 degrees of freedom $P(t > 1.83) = 0.05$.

Solution

$$n = 10, \bar{x} = \frac{\sum x}{n} = \frac{660}{10} = 66, \mu_0 = 64$$

$$s^2 = \frac{1}{n-1}\sum(x_i-\bar{x})^2 = \frac{90}{9} = 10$$

$$\therefore \quad t = \frac{\bar{x}-\mu_0}{s/\sqrt{n}} = \frac{66-64}{\sqrt{10/10}} = 2.$$

For 9 d.f. $t_{tab}, .05 = 2.26$

$\therefore$ $t_{cal} < 2.26$. Average height is not greater than 6 inches.

Example 16 A random sample of 10 boys had the following IQs: 70, 120, 110, 101, 88, 83, 95, 98, 107 and 100. Do these data support the assumption of a population mean IQ of 100? Find a reasonable range in which most of the means IQ values of samples of 60 boys lie.

Solution Here $n = 10$, $\mu_0 = 100$

$$\bar{x} = \sum \frac{x_i}{n} = 97.2$$

$$s^2 = \frac{1}{n-1}\sum (x_i - \bar{x})^2 = 203.73$$

$$\therefore \quad t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{97.2 - 100}{\sqrt{203.73/10}} = 0.62$$

$|t| = 0.62$

t_{tab} at 5% = 2.262 > 0.62 ⇒ the data is consistent with the assumption of mean IQ of 100 in the population.

95% confidence limits are as follows:

$$\bar{x} \pm t_{.05}\, s/\sqrt{n} = 97.2 \pm 2.262\sqrt{\frac{203.73}{10}}$$

$$= 97.2 \pm 10.21 = 107.41 \text{ and } 86.99$$

$\therefore$ 95% C.I is [86.99, 107.41]

Example 17 A random sample of 16 values from a normal population showed a mean of 41.5 inches and the sum of squares of deviation from this mean is equal to 135 square inches. Show that the assumption of mean of 43.5 inches for the population is not reasonable. Obtain 95% fiducial limits for the same.

Solution Here $n = 16$, $\bar{x} = 41.5$ inches and $\sum_i (x_i - \bar{x})^2 = 135$ sq. inch

$$\therefore \quad s^2 = \frac{1}{n-1}\sum (x_i - \bar{x})^2 = \frac{1}{15}(135) = 9, s = 3$$

$\mu_0 = 43.5$ inches

$$\therefore \quad t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{41.5 - 43.5}{3/\sqrt{16}} = 2.669$$

$|t| = 2.667$

$t_{.05}$ for 15 d.f. = 2.131

$|t|_{cal} = 2.667 > 2.131$ and we conclude that the mean 43.5 inches for the population is not reasonable

95% fiducial limits are as follows:

$$\bar{x} \pm t_{.05} \times \frac{s}{\sqrt{n}} = 41.5 \pm 2.131\frac{s}{\sqrt{n}} = 4.15 \pm 1.598$$

$$\Rightarrow \quad 39.902 < \mu < 43.098$$

Scanned by CamScanner

7.23 SIGNIFICANT TEST OF DIFFERENCE BETWEEN TWO SAMPLES

(i) Suppose two independent samples $x_1, x_2, \ldots, x_{n_1}$ and $y_1, y_2, \ldots, y_{n_2}$ of sizes n_1 and n_2 have been drawn from two normal populations having means μ_X and μ_Y, respectively, and same variance $\sigma^2 = \sigma_X^2 = \sigma_Y^2$, that statistic is

$$t = \frac{(\bar{x} - \bar{y}) - (\mu_X - \mu_Y)}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}},$$

where $\bar{x} = \sum_i x_i / n_1$; $\bar{y} = \sum_j y_j / n_2$

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

has a t-distribution with $n_1 + n_2 - 2$ d.f.

If $t_{cal} > t_{.05}$, the difference between sample means $\bar{x}$ and $\bar{y}$ is significant at 5% level of significance, otherwise not significant.

If $t_{cal} > t_{.01}$, then the difference between sample means $\bar{x}$ and $\bar{y}$ is significant at 1% level of significance, otherwise not significant.

In general if $t_{cal} > t_\alpha$ then the difference between the sample means $\bar{x}$ and $\bar{y}$ is significant α% level of significance, otherwise not.

(ii) Paired t-test for difference between two samples.

If (a) $n_1 = n_2 = n$ (say) and

(b) Two samples are not independent but the sample observations are paired together i.e., (x_i, y_i) $(i = 1, 2, \ldots, n)$ observations corresponds to the same i^{th} sample

Then the problem is to test if sample means differ significantly or not i.e. increments are due to fluctuations of sampling. The statistic $t = \frac{\bar{d}}{s/\sqrt{n}}$,

where $\bar{d} = \frac{1}{n}\sum di = \frac{1}{n}\sum (x_i - y_i)$, where $d_i = x_i = y_i$.

and $s^2 = \frac{1}{n-1}\sum (di - \bar{d})^2$, has t-dist. with $n - 1$ d.f.

then find t_α if $t > t_\alpha$ The difference is significant, otherwise not.

Example 18 Given is the gain in weights (in kg) of pigs fed on two diet 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, 25, 29, 22

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

Solution Here $n_1 = 12$, $n_2 = 15$

$$\bar{x} = \sum \frac{x_i}{12} = \frac{336}{12} = 28, \quad \bar{y} = \sum \frac{y_j}{15} = \frac{450}{15} = 30$$

$$s^2 = \frac{1}{n_1 + n_2 - 2}\left[\sum(x_i - \bar{x})^2 + \sum(y_j - \bar{y})^2\right] = 71.6$$

$$\therefore \quad t = \frac{\bar{x} - \bar{y}}{s\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{28.30}{\sqrt{71.6\left(\frac{1}{12} + \frac{1}{15}\right)}} = -0.609$$

$$|t| = 0.609$$

T_{cal} for $(n_1 + n_2 - 2)$ d.f. $= t_{.05}$ for $(12 + 15 - 2)$ $t_{.05}$ for 25 d.f. $= 2.06$

$|t|_{tab} = 6.09 < 2.06 \Rightarrow$ two diets not differ significantly as 5% level of significance.

Example 19 A certain stimulus administrated 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. (V.T.U. 2007)

Solution We have to test that stimulus will increase the blood pressure. To test this $t = \frac{\bar{d}}{s/\sqrt{n}}$ has a t-dist. with $n - 1$ d.f.

Here $n = 12$,

$d = x_i - y_i$	5	–2	8	–1	3	0	–2	1	5	0	4	6
d_i^2	25	4	64	1	9	0	4	1	25	0	16	36

$$\therefore \quad \bar{d} = \sum \frac{d_i}{n} = \frac{31}{12} = 2.583$$

$$s^2 = \frac{1}{n-1}\sum_{i=1}(d_i - \bar{d})^2 = 9.5382$$

$$\therefore \quad t = \frac{\bar{d}}{s/\sqrt{n}} \frac{2.58}{\sqrt{9.5382}/\sqrt{12}} = 2.89$$

t_{tab} at 5% with (12 – 1) d.f. = 2.20

$|t|_{cal} = 2.89 > 2.20 \Rightarrow$ The stimulus does not increase the blood pressure

Example 20 In a certain experiment to compare two types of animal foods A and B, the following results 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 there is no difference in Food A and B. (Use $\alpha = 0.05$)

Scanned by CamScanner

(ii) Also, examine the case when the same set of eight animals were used in both the food (use $\alpha = 0.05$)

Solution (i) $t = \dfrac{\bar{x} - \bar{y}}{s\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$

$$\bar{x} = \sum_i \frac{x_i}{n} = \frac{407}{8} = 50.875,\ \bar{y} = \sum \frac{y_i}{n} = \frac{423}{8} = 52.875$$

$$s^2 = \frac{1}{n_1 + n_2 - 2}\left[\sum_i (x_i - \bar{x})^2 + (y_i - \bar{y})^2\right] = 3.41$$

$$\therefore \quad t = \frac{50.875 - 52.875}{\sqrt{3.41\left(\dfrac{1}{8} + \dfrac{1}{8}\right)}} = -2.17$$

$|t| = 2.17$

$t_{.05}$ with 14 d.f. 2.14

$|t|_{cal} = 2.17 > 2.14 \Rightarrow$ food A and B do not differ significantly at 5% level of significance with regards to their effect on increase in weight.

(ii) $t = \dfrac{\bar{d}}{s/\sqrt{n}}$ where $\bar{d} = \dfrac{\sum d_i}{n} = \sum\left(\dfrac{x_i - y_i}{n}\right)$

x_i	49	53	51	52	47	50	52	53	Total 407
y_i	52	55	52	53	50	54	54	53	423
d_i	–3	–2	–1	–1	–3	–4	–2	0	–16
d_i^2	9	4	1	1	9	16	0	0	44

$$\bar{d} = \sum \frac{d_i}{n} = -\frac{16}{8} = -2$$

$$s^2 = \frac{1}{n-1}\sum\left[(d_i - \bar{d})^2\right] = 1.7143$$

$$\therefore \quad |t| = \left|\frac{\bar{d}}{\sqrt{s^2}/\sqrt{n}}\right| = \frac{2}{\sqrt{1.7143/8}} = 4.32$$

tab $t_{.05}$, for $(8 - 1) = 7$ dif $f = 2.36$

$|t|_{cal} > 2.36 \Rightarrow$ 't' is significant at 5% level of significance and we conclude that both foods A and B do not differ.

Scanned by CamScanner

EXERCISE 7.3

1. Nine items of a sample holds the following values: 45, 47, 50, 52, 48, 47, 49, 53, 51. Does the mean of these differ significantly from the assumed mean of 47.5? **(V.T.U. 2010)**
2. A mechanic is making engine parts with axle diameter of 0.7 inch. A random sample of 10 parts show mean diameter 0.742 inch with a standard deviation of 0.04 inch. On the basis of this sample would you say that the work is inferior. **(V.T.U. 2009)**
3. Prove that 95% confidence limits for the population mean μ are $\bar{x} \pm \frac{\sigma}{\sqrt{n}} t_{0.05}$.
4. A random sample of 10 boys had the following IQ.
70, 120, 110, 101, 88, 83, 95, 98, 107, 100
Do these data support the assumption of a population mean IQ of 100 at 5% level of significance? **(V.T.U. 2006, Coimbatore 2001)**
5. A random sample of size 25 from a normal population has the mean $\bar{x} = 47.5$ and s.d. $s = 8.4$. Does this information refute the claim that the mean of the population is $\mu = 49.1$. **(J.N.T.U. 2003)**
6. The means of two random samples of sizes 9 and 7 are 196.42 and 198.82, respectively. The sum of squares of the deviations from the means are 26.94 and 18.73, respectively. Can the sample considered to have been drawn from the same normal population? **(Mumbai 2004)**
7. For a random sample of 10 pigs, fed on a diet A, the increase in weights in a certain period were:
10, 6, 16, 17, 13, 13, 12, 8, 14, 15, 9 lbs
For another random sample of 12 pigs fed on diet B, the increase in the same period were.
7, 13, 22, 15, 12, 14, 18, 8, 21, 23, 10, 17 lbs.
Find if the two samples are significantly different regarding the effect of diet. (Use $\alpha = 0.05$)
8. Eleven students were given a test in statistics. They were given a month's further tuition and a second list of equal difficulty was held at the end of it. Do the marks give evidence that students have been benefited by extra coaching?

Boys	1	2	3	4	5	6	7	8	9	10	11
Marks I test:	23	20	19	21	18	20	18	17	23	16	19
Marks II test:	24	19	22	18	20	22	20	20	23	20	17

(V.T.U. 2011S)

9. Test runs with 6 models of an experimental engine showed that they operated for 24, 28, 21, 23, 32 and 22 minutes with a gallon of fuel. If the probability of a type I error is at numbers 0.01, is this an evidence against a hypothesis that with this average, engine will operated for at least 29 minutes per gallon of the same fuel. Assume normality. **(J.N.T.U. 2003)**
10. Two houses A and B were tested according to the time (in seconds) to run a particular race with the following results:

House A:	28	30	32	33	33	29 and 34
House B:	29	30	30	24	27 and 29	

Test whether you can discriminate between two houses? **(Rohtak 2005, Coimbatore 2001)**

11. A group of 10 rats fed on a diet A and another group of 8 rats fed on a different diet B, recorded the following increase in weight:
Diet A: 5, 6, 8, 1, 12, 4, 3, 9, 6, 10 gms
Diet B: 2, 3, 6, 8, 10, 1, 2, 8 gms

Scanned by CamScanner

(Madras 2003)

Does it show the superiority of diet A over diet B?

12. The average number of articles produced by two machines per day are 200 and 250 with standard deviations 20 and 25, respectively, on the basis of records of 25 days production. Can you regard both the machines equally efficient at 5% level of significance.

13. A company is interested in knowing if there is a difference in the average salary received by foremen in two divisions. Accordingly samples of 12 foremen in the first division and 10 foremen in the second division are selected at random. Based upon experience foremen's salaries are known to be approximately normally distributed and the standard deviations are about the same.

Sample size	First division	Second division
	12	10
Average weekly salary of foremen (Rs)	1050	980
Standard deviation of salaries (Rs)	68	74

The table value of t for 20. d.f. at 5% level of significance is 2.09.

14. A random sample of size 25 from a normal population has mean $\bar{x} = 47.5$ and s.d. $s = 8.4$. Does this information refute the claim that the mean of the population is $\mu = 42.1$ (J.N.T.U. 2003)

Answers

1. $t = 1.83$, $t_{.05}$ for 8 d.f. = 2.31, not significant.
2. $t = 3.16$.
3. $t = 0.62$, yes.
4. Refute the claim.
5. No.
6. $t = 1.51$.
7. $t = 1.48$.
8. Accept null hypothesis.
9. Yes with 75% confidence.
10. No.
11. $|t| = 7.65$.
12. $t = 2.2$.
14. Refute the claim as $t = 3.21$.

7.24 CHI-SQUARE (χ^2) TEST

Introduction: The square of a standard normal variable $Z \sim N(0, 1)$ is known as a Chi-square variate with let f.

$\therefore$ If $X \sim N(\mu, \sigma^2)$, then $Z = \dfrac{X-\mu}{\sigma} \sim N(0.1)$, and

$$Z^2 = \left(\frac{X-\mu}{\sigma}\right)^2 \text{ is a Chi-square variate with 1 d.f.}$$

In general, if $X_1, X_2, \ldots, X_n$ are n independent normal variates with means $\mu_1, \mu_2, \ldots, \mu_n$ and variance $\sigma_1^2, \sigma_2^2, \cdots, \sigma_n^2$, respectively, then

$$\sum_{i=1}^{n}\left(\frac{X_i-\mu_i}{\sigma_i}\right)^2 = \chi^2, \text{ is a Chi-square variate with n.d.f.}$$

The p.d.f. of χ^2 is given by

$$f(\chi^2) = \frac{\left(\frac{1}{2}\right)^{\frac{n}{2}}}{\left|\frac{n}{2}\right.}\left[\exp\left(-\frac{1}{2}\chi^2\right)(\chi^2)^{\frac{n}{2}-1}\right]$$

$$= \frac{1}{2^{\frac{n}{2}}\sqrt{\frac{x}{2}}}\left[\exp\left(-\frac{1}{2}\chi^2\right)\right](\chi^2)^{\frac{n}{2}-1}, 0 \le \chi^2 < \infty$$

or

$$f(\chi^2) = C\, e^{-\frac{1}{2}\chi^2}\,(\chi^2)^{\frac{(v-1)}{2}} \quad (3)$$

where C is a constant and $v = n - 1$

The graph of the equation of χ^2 – curve which is given in eq. (3) is shown in Fig. 7.3.

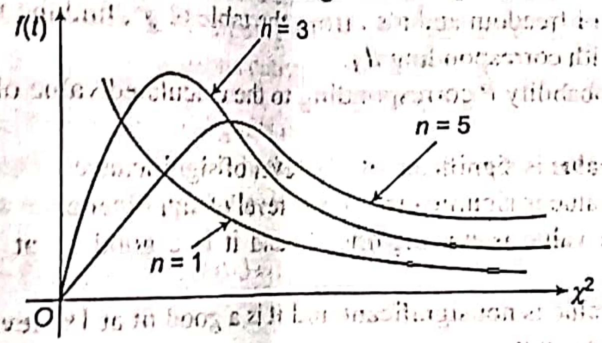


Fig. 7.3

7.24.1 Properties of χ^2-distribution

1. If $v = 1$, then Eq. (3) becomes $y = Ce^{-\frac{1}{2}\chi^2}$, which is exponential distribution.
2. The mean of a χ^2 distribution with v d.f. is v and various is $2v$.
3. Mode of χ^2 distribution with v d.f. $= n - 2$
4. If $v = 1$, the curve of χ^2-variate is tangential to x-axis at origin and positively skewed.
5. The probability P that the value of χ^2 from a random sample will exceed χ_α^2 is given by

$$P\left[\chi^2 \ge \chi_\alpha^2\right] = \int_{\chi^2}^{\infty} f(\chi^2)\, dx = P$$

and the values of χ_α^2 have been tabulated for various values of P and for values of degrees of freedom v from 1 to 30.

If $v > 30$, then χ^2-distributin fallows normal distribution and we use normal distribution tables for significant values of χ^2.

6. As the equation of χ^2-curve does not involve any of the parameters of the population, therefore χ^2-distribution does not depend on the distribution of the population and hence very useful in statistical theory.

Scanned by CamScanner

7.24.2 Chi-Square (χ^2) Test of Goodness of Fit

A very powerful test for testing the significance of the discrepancy between theory and experime[nt is] given by Prof. Karl Pearson in 1900 and is known as "Chi-Square test of goodness of fit". It e[nables] us to find if the deviation of the experiment from theory is just by chance or is it really due [to] inadequacy of the theory to fit the observed data.

Let $o_i (i = 1, 2, \ldots, n)$ be a set of observed (experimental) frequencies and $e_i (i = 1, 2, \ldots, n)$ b[e the] corresponding set of expected (theoretical or hypothetical) frequencies, then

$$\chi^2 = \sum_{i=1}^{n}\left(\frac{o_i - e_i}{e_i}\right)^2 \text{ has a } \chi^2\text{-distribution with } (n-1) \text{ d.f.}$$

Procedure to test significance and goodness of fit:

(a) Set up a 'null hypothesis'.

(b) Compute $\chi^2 = \sum_{i=1}^{n}\left(\frac{o_i - e_i}{e_i}\right)^2$

(c) Find the degrees of freedom and also from the table of χ^2, find the value of χ^2 at desired le[vel] of significance with corresponding $d.f.$

(d) Also, find the probability P corresponding to the calculated value of χ^2 for the given d.f. fr[om] χ^2 table.

(e) If $P < 0.01$, the value is significant at 1% level of significance

If $P < 0.05$, the value is significance at 5% level of significance

If $P > 0.05$, the value is not significant and it is a good fit at 5% level of significan[ce]

Similarly

If $P > 0.01$, the value is not significant and it is a good fit at 1% level of significance

In general, we can say if

$P > \alpha$, the value is not significant at a given level of significance α and it is a good fit.

If $P \leq \alpha$ then value is significant.

Remark: As we don not make assumptions on the population distribution, the test of χ^2 is al[so] known as non-parametric test

Example 21 The demand for a particular spare part in a factory was found to vary from day to da[y]. In a sample study the following information was obtained:

Days	Mon	Tue	Wed	Thu	Fri	Sat
No. of parts demanded	1124	1125	1110	1120	1126	1115

Test the hypothesis that the number of parts does not depend on the day of the week.

Solution The null hypothesis Ho. The number of parts demanded does not depend on the day of week. Under the null hypothesis the expected frequencies of the spare part demanded on each of the six days would be

$$\frac{1}{6}(1124 + 1125 + 1110 + 1120 + 1126 + 1115) = \frac{6720}{6} = 1120 \text{ calculated by } \chi^2$$

Days:	Mon	Tue	Wed	Thu	Fri	Sat	Total
Observed frequencies (o_i)	1124	1125	1110	1120	1126	1115	6720
Expected Frequencies (e_i)	1120	1120	1120	1120	1120	1120	6720
$(o_i - e_i)^2$	16	25	100	0	36	25	202
$\dfrac{(O_i - e_i)^2}{e_i}$	0.014	0.022	0.089	0	0.032	0.022	0.179

Number of d.f. = (6 – 1) = 5

Tabulated value of $\chi^2_{.05}$ with 5 d.f. = 11.07

$\chi^2_{cal} = 0.179 < 11.07 \Rightarrow$ It is not significant and null hypothesis is accepted at 5% level of significance. Hence, we conclude that the number of parts demanded are same over the 6 days period.

Example 22 A sample analysis of examination results of 200 engineering students was made. It was found that 46 students failed, 68 secured a third division, 62, secured a second division and rest were placed in a first division. Are these figures commensurate with the general examination result which is in the ratio of 4 : 3 : 2 : 1 for various categories respectively?

Solution Null hypothesis: The observed figures do not differ significantly from the hypothetical frequencies which are in the ratio of 4 : 3 : 2 : 1. In three words, the given data are commensurate with the general examination result which is in the ratio of 4 : 3 : 2 : 1.

$$\therefore \quad e_1 = \frac{1}{10}(200 \times 4) = 80,\ e_2 = \frac{1}{10}(200)\cdot 3 = 60,\ e_3\ \frac{1}{10}(200)\cdot 2 = 40, \text{ and } e_4 = \frac{1}{10}(200)\cdot 1 = 20$$

Calculation for χ^2

Category	Frequency Observed (o_i)	Frequency Expected (e_i)	$(o_i - e_i)^2$	$\left(\dfrac{o_i - e_i}{e_i}\right)^2$
Failed	46	80	1156	14.450
III division	68	60	64	1.069
II division	62	40	484	12.100
I division	24	20	16	0.800
Total	200	200		28.417

d.f = 4 – 1 = 3, Tab $\chi^2_{0.05}$ for 3 d.f. = 7.815

$\chi^2_{cal} = 28.417 > 7.815$. It is significant and null hypothesis is rejected. Hence, we may conclude that data are not commensurate with the general examination.

Example 23 When the first proof of 392 pages of a book of 1200 pages were read, the distribution of printing mistakes were found to be as follows:

No. of mistakes in a page (x)	0	1	2	3	4	5	6
No. of pages (o_i)	275	72	30	7	5	2	1

Fit a Poisson distribution data and test the goodness of the given.

Solution *Null hypothesis:* The given data fit the Poisson distribution

Here $\bar{x} = \frac{1}{N} \sum_{i=0}^{6} f_i x_i = \frac{189}{392} = 0.482$

The frequency of x mistakes per page is given by the Poisson law as follows:

$$e(x): N\,p(x) = \frac{392.\,e^{0.482}\,(0.482)^x}{x!}, x = 0, 1, 2, \cdots, 6$$

$\therefore$ $e_o = 242.1, e_1 = 116.7, e_2 = 28.1, e_3 = 4.5, e_4 = 0.5, e_5 = 0.1, e_6 = 0$

Calculation of χ^2

Mistakes per page x	Frequency Observed (o_i)	Frequency Expected (e_i)	$(o_i - e_i)^2$	$\left(\frac{o_i - e_i}{e_i}\right)^2$
0	275	242.1	1082.41	4.471
1	72	116.7	1998.09	17.121
2	30	28.1	3.61	0.128
3	7 } 15	4.5 } 5.1	98.01	19.217
4	5	0.5		
5	2	0.1		
6	1	00		
Total	392	392		40.937

d.f. = 7 – 1 – 1 – 3 = 2

($\because$ One d.f. being lost because of calculating $\bar{x}$ and 3 d.f. are lost because of pooling the last four expected cell frequencies, which are less than 5).

Tabulated value of χ^2 for 2 d.f. at 5% level = 5.99

$\chi^2_{cal} = 40.937 > 75.98$. It is highly significant. Hence, Poisson distribution is not good fit.

Note:

1. If any parameter(s) is(are) calculated based on sample data, the degrees of freedom will be reduced to the no. of parameter(s) computed
2. If any cell frequency is <5 then add these cell frequencies to the proceeding or succeeding cell and number of degrees will be reduced the number of those cells for which frequencies are added to another cell.
3. In above example $\bar{x}$ is calculated from sample data and 3 frequencies are added to their proceeding cell there fore number of d.f.

$$= (n - 1) - 1 - 3 = (7 - 1) - 1 - 3 = 2.$$

Scanned by CamScanner

...CISE 7.4

1. A set of 5 similar coins is tossed 320 times and the result is

Number of heads	0	1	2	3	4	5
Frequencies	6	27	72	112	71	32

Test the hypothesis that the data follows a binomial distribution

(Kottayam, 2005: P.T.U., 2005; V.T.U. 2004)

2. Fit a Poisson distribution to the following data and test the goodness of fit at a level of significance 0.05.

x	0	1	2	3	4
f	419	352	154	56	19

(V.T.U., 2008)

3. The following table gives the number of aircrafts accidents that occurred during the various days of the week. Whether the occident are uniformly distributed over the week?

Days	Sun	Mon	Tue	Wed	Thu	Fri	Sat	Total
Number of accident	14	16	08	12	11	9	14	84

(Hissar, 2005)

4. A survey of 800 families with four children each revealed the following distribution:

Number of boys	0	1	2	3	4
Number of girls	4	3	2	1	0
Number of families	32	178	290	236	64

Is this result consistent with the hypothesis that male and female births one equally probable?

5. The following figure show the distribution of digits in numbers chosen at random from a telephone directory:

Digits	0	1	2	3	4	5	6	7	8	9	Total
Frequency	1026	1107	997	966	1075	933	1107	972	964	853	10,000

Test whether the digits may be taken to occur equally frequently in the directory.

6. Fit a normal distribution to the following data of weights of 100 students of an university and test the goodness of its

Weight (kg)	60-62	63-65	66-68	69-71	72-74
Frequency	5	18	42	27	8

7. Fit a Poisson distribution to the following data and test the goodness of fit.

x	0	1	2	3	4	5	6	7
f	305	366	210	80	28	9	2	1

8. A dice was thrown 60 times and the following frequency distribution was observed. Test whether the dice in unbiased?

Faces	1	2	3	4	5	6	Total
f	15	6	4	7	11	17	60

Answers

1. $\chi^2_{cal} = 78.68$, data do not follow binomial distribution at 5%.
2. $\chi^2_{cal} = 5.748$, data fit the Poisson distribution at 5%.
3. Yes, accidents are uniformly distributed.
4. $\chi^2_{cal} = 19.63$, male and female births are not equally probable at 5%.
5. $\chi^2_{cal} = 58.542$, digits are not uniformly distributed.
6. $\chi^2_{cal} = 0.8362$ (here d.f. 2 as mean and variance are computed.
7. $\chi^2 = 3.097$, Poisson distribution gives a good fit at 5% level.
8. Significant at 5%.

7.25 F-DISTRIBUTION

F-distribution introduced by English statistician R.A. Fisher is defined as follows:

Let X and Y are two independent Chi-square variates with ν_1 and ν_2 d.f., respectively, then *F*-statistics is given by

$$F = \frac{X/\nu_1}{Y/\nu_2}$$

In other words, F is defined as the ratio of two independent Chi-square variates divided by their respective degrees of freedom and devoted by $F(\nu_1, \nu_2)$.

7.25.1 Application of *F*-distribution

F-distribution has wide applications in the theory of statistics. One of them is F-test for equality of two population variances, which is defined as follows: Let $x_1, x_2, \ldots, x_{n_1}$, and $y_1, y_2, \ldots, y_{n_2}$ be the two independent random samples drawn from the normal populations with the same variance. Let $\bar{x}$ and $\bar{y}$ be the sample means and s_x^2 and s_y^2 be the sample variances of two samples, and

$\bar{x}_1 = \frac{1}{n}\sum_{i=1}^{n_1} x_i$, $\bar{y} = \frac{1}{n_2}\sum_{j=1}^{n_2} y_j$, $s_x^2 = \frac{1}{n_1 - 1}\sum_{i=1}^{n_1}(x_i - \bar{x})^2$, $s_y^2 = \frac{1}{n_2 - 1}\sum_{j=1}^{n_2}(y_j - \bar{y})^2$. Then $F = \frac{s_x^2}{s_y^2}$ has *F*-distribution with $n_1 - 1 = \nu_1$ and $n_2 - 1 = \nu_2$ degrees of freedom. The larger variance is always taken as numerator, so that the values of F is always positive.

7.25.2 Properties of *F*-distribution

1. The graph of *F*-distribution is shown as follows in Fig. 7.4

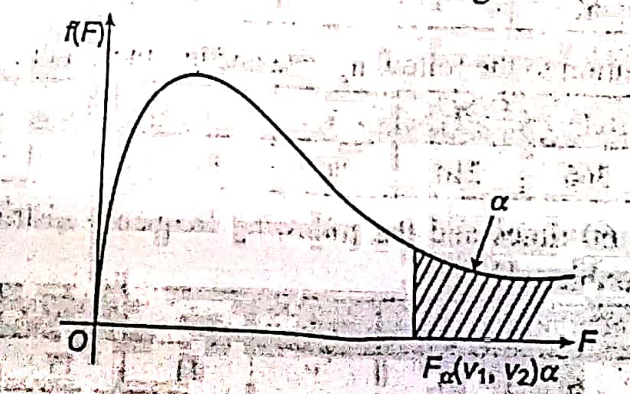


Fig. 7.4

Scanned by CamScanner

3. The curve of F-distribution always lie in first quadrant as $F > 0$ (always).
4. F-distribution is unimodel and its mode is always < 1.
4. F-distribution depends on ν_1 and ν_2 not on the population variants σ^2 i.e, it is independent of σ^2.
5. ν_1, ν_2 are known as degrees as degrees of freedom of numerator and denominator, respectively.
6. $P\left[F(\nu_1, \nu_2) \geq F_\alpha(\nu_1, \nu_2)\right] = \alpha.$
7. $F_\alpha(\nu_1, \nu_2) \dfrac{1}{F_{1-\alpha}(\nu_2, \nu_1)}.$

7.25.3 Significance Test

F tables gives 5% and 1% points of significance for F. 1% points of F mean that area under the F-curve to the right of the ordinate at a value of F, is 0.01. The value of at 1% significance level is more that at 5%. F-distribution has wide applications and has a base for analysis of variance.

Example 24 In one sample of 8 observations, the sum of squares of derivations 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% level of significance.

Solution Here $n_1 = 8, n_2 = 10, \sum_{i=1}^{8}(x_i - \bar{x})^2 = 84.4, \sum_{j=1}^{10}(y_j - \bar{y})^2 = 102.6$

$$\therefore \quad s_x^2 = \frac{1}{n_1 - 1}\sum(x_i - \bar{x})^2 = \frac{1}{7}(84.4) = 12.057$$

$$\text{and} \quad s_y^2 = \frac{1}{n_2 - 1}\sum(y_j - \bar{y})^2 = \frac{1}{9}(102.6) = 11.4$$

$$\therefore \quad F = \frac{s_x^2}{s_y^2} = \frac{12.057}{11.4} = 1.0576$$

Tabulated value

$F_{0.5}(7, 9) = 3.29$ (from F-table)

$F_{cal} = 1.0576 < 3.29$. It is not significant

7.26 FISHER'S Z-DISTRIBUTION

If we put $z = \frac{1}{2}\log_e F$ or. $F = e^{2z}$ in the F-distribution, we get Fisher's Z-distribution.

Z-distribution is more symmetrical then F-distribution. For various values of ν_1 and ν_2 degrees of freedom, a table showing the values of z that will be exceeded in simple sampling with probabilities 0.01 and 0.05.

7.26.1 Significance Test

In Z-table, we have critical values as $P[Z \geq Z_d] = \alpha$

Therefore, the 1% or 5% points of Z imply that the area to the right of ordinate $Z_{.05}$ or $Z_{.01}$ is 0.05 or 0.01. In other words, 1% and 5% points of Z-correspond to 2% and 10% level of significance respectively. As we generally use two-tailed test.

Scanned by CamScanner

Example 25 Two gauge operations are tested for precision in making measurements. One operator completes a set of 26 readings with a standard deviation of 1.34 and the other does 34 readings with a standard deviation 0.98. What is the level of significance of this difference. (We can use $Z_{.05} = 0.305$, $Z_{.01} = 0.432$) for $\nu_1 = 25$ and $\nu_2 = 33$)

Solution Here $n_1 = 26, n_2 = 34, s_x = 1.34, s_y = 0.98$

$\therefore$ $s_x^2 = (1.34)^2$

and $s_y^2 = (0.98)^2$

$$\therefore \quad F = \frac{s_x^2}{s_y^2} = \frac{(1.34)^2}{(0.98)^2} = 1.8696$$

$$\therefore \quad Z = \frac{1}{2}\log_e F = 1.1513 \log_{10}(1.8696) = 0.31286$$

Calculated value of $Z > .305$ but less than 0.432

$\therefore$ level of significance lies between 1% and 5% which is closer to 5%

EXERCISE 7.5

1. Two samples of sizes 9 and 8 give the sum of squares of deviation from their means equal to 160 inches^2 and 91 inches^2, respectively. Can these be regarded as drawn from the same normal population. **[V.T.U. 2002]**
2. 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, the hypothesis that two variances are equal against the alternative that they are not at 5% level.
3. Two random samples gave the following results.

Sample	Size	Sample mean	Sum of squares of deviation 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.

4. The following are the values in thousands of an inch obtained by two engineers in 10 and 9 successive measurements with the same micrometer. Are both engineers significant of each other.

Engineer A	503	505	497	505	495	502	499	493	510	501
Engineer B	502	497	492	498	499	495	497	496	498	

5. Two random samples of sizes 8 and 11, drawn from the two normal populations are characterized as follows:

Sample	Size	Sum of observation	Sum of squares of observations
1	8	9.6	61.52
2	11	16.5	73.26

Test whether two populations can be taken to have the same variance.

6. Two samples of sizes 9 and 8 have variances 1101.1 and 319.7, respectively. Is the variance 1101.1 significantly greater than variance 319.7.

Scanned by CamScanner

The IQ's of 25 students from one institute showed a variance of 20 and those of an equal number from the other institute had a variance of 10. Discuss whether there is any significant difference in variability of intelligence.

8. Show how you would use Fisher's Z test to decide whether the two sets of observations 15, 25, 16, 23, 25, 27, 25, 21, 15 and 14, 14, 18, 14, 18, 15, 13 and 19 indicate samples from the same universe.

Answers

1. $F \cong 1.54$, $F_{.05}(8.7) = 3.73$, not significant.
2. $F = 2.5$, $F_{.05}(10, 8) \cong 3.35$, not significant.
3. $F = 1.018$, $F_{0.05}(11, 9) = 3.10$, not significant.
4. $F = 2.4$.
5. $F = 1.47$.
6. $F = 3.44$, $F_{.05}(8, 7) = 3.73$ and $F_{.01}(8, 7) = 6.84$.
   It is significant at both the level of significance.
7. $F = 2$.
8. $F = 4.25$, $z = 0.72346$.

SUMMARY

In this chapter, the following topics have been discussed:

1. Population and sample.
2. Population parameters: μ and σ^2 etc.
3. Sample statistic: $\bar{x}, s^2$ etc, $\bar{x} = \sum \frac{x_i}{n}$, $s^2 = \frac{1}{n-1}\sum (x_i - \bar{x})^2$
4. Standard error: Standard deviation of sampling variance s^2
5. Null and alternate hypothesis $= H_0$ and H_1
6. P(Type I error) $= P$(rejecting null hypothesis when null hypothesis is true)
   P(Type II error) $= P$(accepting null hypothesis when it is wrong)
7. P(Type I error) and level of significance are generally same.
8. Confidence limits for μ

$$\bar{x} \pm z_\alpha \frac{\sigma}{\sqrt{n}} \ (\sigma \text{ is known})$$

$$\bar{x} \pm t_\alpha \frac{\sigma}{\sqrt{n}} \ (\sigma \text{ is unknown and } \sigma \text{ is estimated based on sample})$$

9. Central limit theorem: $\frac{X - E(x)}{\sqrt{\text{var}(x)}} = z \sim N(0, 1)$

   For large $n \geq 30$, any distribution of X will follow standard normal distribution.

10. $z = \frac{\bar{x} - \mu}{\sigma\sqrt{n}}$ and $t = \frac{\bar{x} - \mu}{s\sqrt{n}}$

Scanned by CamScanner

11. $F = \frac{s_x^2}{s_y^2}$

12. $\chi^2 = \sum_{i=1}^{n}\left(\frac{x_i - \mu}{\sigma}\right)^2 \sim \chi_{n-1}^2$

13. $z = \frac{1}{2}\log_e F$

OBJECTIVE TYPE QUESTIONS

1. If $s_1^2 = 10$ and $s_2^2 = 5$, then F is equal to
   (a) 2 (b) 3
   (c) 1 (d) 2.5
2. If S^2 be sample variance of n observations, then it is equal to
   (a) $\frac{1}{n}\sum_i (x_i - \bar{x})^2$
   (b) $\frac{1}{n-1}\sum_i (x_i - \bar{x})^2$
   (c) $\frac{1}{n}\sum_i (x_i + \bar{x})^2$
   (d) $\frac{1}{n-1}\sum_i (x_i + \bar{x})^2$
3. If a random variable X has χ^2 distribution, with 4 d.f. then its mean is equal to
   (a) 4 (b) 3
   (c) 2 (d) 1
4. If a random variable X has χ^2-distribution with 2 d.f., then its standard deviation is
   (a) 4 (b) 3
   (c) 2 (d) 1
5. If $p = 0.13$, $n = 500$, then standard error is equal to
   (a) 0.011 (b) 0.012
   (c) 0.013 (d) 0.015
6. X is normally distributed with mean 5 and variance 25. If a random sample of size 25 is drawn from X then variance of $\bar{X}$ is equal to
   (a) 1 (b) 2
   (c) 5 (d) none of these
7. If standard error of $\bar{X} = 0.2$, then its precision is equal to
   (a) 0.5 (b) 1
   (c) 2 (d) 5
8. If $n = 625$, $\bar{X} = 3.50$, $\mu = 3.20$ and $\sigma^2 = 2.25$, then z is equal to
   (a) 5 (b) 2
   (c) 1 (d) 0.5
9. If t distribution has mean zero, then its variance is
   (a) > 12 (b) < 12
   (c) = 12 (d) none of these
10. Variance of t distribution is always
    (a) Greater than 1 (b) Less than 1
    (c) Equal to 1 (d) Greater than 0.5
11. If we reject null hypothesis when it is correct we get ______ error
12. To use t-statistic, the population from which sample is drawn should have ______ distribution.
13. If the standard deviation of χ^2-distribution is 10, then its degree of freedom is
    (a) 25 (b) 50
    (c) 75 (d) 100
14. The test statistic, $F = \frac{s_x^2}{s_y^2}$ is used when
    (a) $s_x^2 > s_y^2$ (b) $s_x^2 < s_y^2$
    (c) $s_x^2 = s_y^2$ (d) none of these
15. In a t-distribution of sample size 18, the degree of freedom are
    (a) 18 (b) 17
    (c) 16 (d) 15

16. Fisher's z transformation is equal to

(a) $z=\frac{1}{2}\log_e F$ (b) $z=\frac{1}{2}\log_{10} F$

(c) $z=\log_e F$ (d) $\log_{10} F$

17. The range of t-statistic is

(a) $(0, \infty)$ (b) $(-\infty,\infty)$

(c) $(-\infty, 0)$ (d) $(2, \infty)$

18. If mean of χ^2-distribution is 4, then its mode is

(a) 4 (b) 3

(c) 2 (d) 1

19. If 6, 27, 72, 112, 71 and 32 are the observed frequencies; and 10, 50, 100, 100, 50 and 10 are respectively the expected frequencies of an experiments, respectively, then the value of χ^2 is

(a) 78.68 (b) 76.88

(c) 68.78 (d) 86.78

20. If $\bar{x}_1 = 67.5, \bar{x}_2 = 68.0, n_1 = 100, n_2 = 2000$ and $r = 2$, then z is

(a) 4.9 (b) 5.0

(c) 5.1 (d) 5.2

ANSWERS

1. (a)	2. (b)	3. (a)	4. (c)	5. (d)	6. (a)	7. (d)	8. (a)	9. (d)	10. (a)
11. Type I	12. Normal		13. (b)	14. (a)	15. (b)	16. (a)	17. (b)	18. (c)	19. (a)
20. (c)									