t - distribution

Sampling distributions on the assumption that they are normal or approximately normal are valid when the sample size N is large. For small samples ($N \leq 30$), we consider the t - distribution

Let N be the sample size, $\bar{x}$ and μ be respectively the sample mean and the population mean, and s be the sample standard deviation. Consider the statistic t defined by

$$t = \frac{(\bar{x} - \mu)}{s}\sqrt{N-1}$$

Suppose we obtain a frequency distribution of t by computing the value of t for each of a set of samples of size N drawn from a normal or a nearly normal population. The sampling distribution so obtained is called the Student's t-distribution.

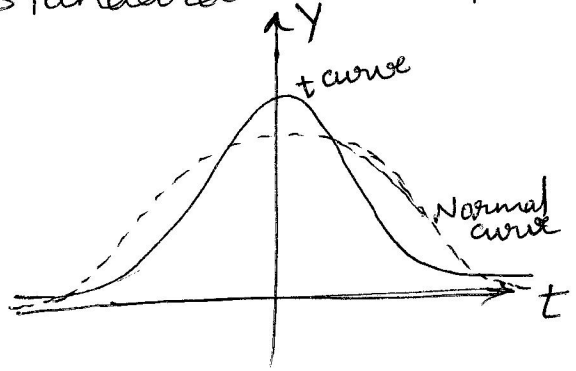
The curve represented by the function

$$Y(t) = Y_0\left(1 + \frac{t^2}{N-1}\right)^{-N/2}$$, where Y_0 is a constant

is called the t-curve.

The constant Y_0 is generally chosen in such a way that the total area under the curve is equal to unity.

For large values of N the function $Y(t)$ reduces to the standard normal distribution density function $\Phi(t) = (1/\sqrt{2\pi})\, e^{-t^2/2}$ and the t-curve becomes the standard normal curve.



Setting $\gamma = N-1$, we have

$$t = \frac{\bar{X}-\mu}{s}\sqrt{\gamma}\,, \quad Y(t) = Y_0\left(1+\frac{t^2}{\gamma}\right)^{-(\gamma+1)/2}$$

The quantity $\gamma = N-1$ is called the number of degrees of freedom for the statistic t.

For population means, the confidence limits are given by $\bar{X} \pm t_c(s/\sqrt{\gamma})$, where $\pm t_c$ are the critical values or confidence coefficients whose values depend on the level of significance desired and the sample size.

For a specified γ, $t_p(\gamma)$ is the value of t_c at $p = \alpha\%$ level of significance.

At the p level of significance, the confidence interval for the t-score is $(-t_p(\gamma), t_p(\gamma))$.

1. For a random sample of 16 values with mean 41.5 and the sum of the squares of the deviations from the mean equal to 135 and drawn from a normal population, find the 95% confidence limits and the confidence interval, for the mean of the population.

Soln $N = 16$, $\gamma = N - 1 = 15$ $\quad \Sigma s^2 = 135$

$\bar{x} = 41.5$ $\quad \therefore s^2 = \frac{135}{N} = \frac{135}{16} = 8.437$

For 95% confidence level, the significance level is 0.05

Hence the required confidence limits are

$$\bar{x} \pm t_c (s/\sqrt{\gamma}) = 41.5 \pm t_{0.05}(\gamma) \times \frac{\sqrt{8.4375}}{\sqrt{15}}$$

$$= 41.5 \pm 2.13 \times \frac{\sqrt{8.4375}}{\sqrt{15}}$$

$$= 43.1, 39.9$$

$\therefore$ the confidence interval is (39.9, 43.1)

2. A mechanist is making engine parts with axle diameter of 0.7 inch. A random sample of 10 parts showed a mean length of 0.742 inch, with a standard deviation of 0.04 inch. On the basis of this sample, can it be concluded that the work is inferior at 5% level of significance?

Soln $N = 10$ $(\because \gamma = 9)$, $\bar{x} = 0.742$, $s = 0.04$

$\mu = 0.7$, hypothesis: $\mu = 0.7$, work is not inferior.

$$t = \frac{\bar{x} - \mu}{s}\sqrt{\gamma} = \frac{0.742 - 0.7}{0.04}\sqrt{9} = 3.16$$

For $\gamma = 9$, $t_{0.05}(\gamma) = 2.26$

and $3.16 \notin (-2.26, 2.26)$.

$\therefore$ the hypothesis is rejected.

$\therefore$ at 5% level of significance, the work is inferior

3. Find the student's 't' for the following values in a sample of eight: $-4, -2, -2, 0, 2, 2, 3, 3$, taking the mean of the population to be zero.

Sol^n $\bar{x} = \frac{1}{8}(-4 - 2 - 2 + 0 + 2 + 2 + 3 + 3) = 0.25$

$$s^2 = \frac{1}{8}\{(-4 - 0.25)^2 + (-2 - 0.25)^2 + (-2 - 0.25)^2 + (0 - 0.25)^2 + (2 - 0.25)^2 + (2 - 0.25)^2 + (3 - 0.25)^2 + (3 - 0.25)^2\}$$

$$s^2 = 6.1875$$

$\mu = 0$,

Then $t = \frac{\bar{x} - \mu}{s}\sqrt{8}$

$$= \frac{0.25 - 0}{\sqrt{6.1875}}\sqrt{(8-1)}$$

$$= 0.2659$$

4. The following are the I.Q's of a randomly chosen sample of 10 boys:

70, 120, 110, 101, 88, 83, 95, 98, 107, 100.

Does this data support the hypothesis that the population mean of I.Q's is 100 at 5% level of significance?

Soln $N = 10$

$$\bar{x} = \frac{1}{10}(70+120+110+101+88+83+95+98+107+100)$$

$$= 97.2$$

$$s^2 = \frac{1}{10}[(70-97.2)^2 + (120-97.2)^2 + (110-97.2)^2 + (101-97.2)^2 + (88-97.2)^2 + (83-97.2)^2 + (95-97.2)^2 + (98-97.2)^2 + (107-97.2)^2 + (100-97.2)^2]$$

$$s^2 = 183.36$$

$$\therefore s = \sqrt{183.36} = 13.54.$$

$\mu = 100$

$$t = \frac{\bar{x} - \mu}{s}\sqrt{n} = \frac{97.2 - 100}{13.54} \times \sqrt{9} = -0.62$$

and $t_{0.05}(n) = t_{0.05}(9) = 2.26$

$t = -0.62$, $|t| = 0.62 < t_{0.05}(9) = 2.26$

$\therefore$ the hypothesis is accepted.

5. A certain stimulus adminstered to each of 12 patients resulted in the following change in blood pressure: 5, 2, 8, −1, 3, 0, 6, −2, 1, 5, 0, 4 (in appropriate units). Can it be concluded that, on the whole, the stimulus will change the blood pressure. Use $t_{0.05}^{(11)} = 2.201$.

Soln $N = 12$, $\gamma = 11$

$$\bar{x} = \frac{1}{12}(5+2+8+(-1)+3+0+6+(-2)+1+5+0+4)$$

$$= 2.58.$$

$$s^2 = \frac{1}{12}\{(5-2.58)^2 + (2-2.58)^2 + (8-2.58)^2 + (-1-2.58)^2 + (3-2.58)^2 + (0-2.58)^2 + (6-2.58)^2 + (-2-2.58)^2 + (1-2.58)^2 + (5-2.58)^2 + (0-2.58)^2 + (4-2.58)^2\}$$

$$s^2 = 8.743$$

$$s = \sqrt{8.743} = 2.96.$$

Let the hypothesis be: the stimulus does not change the blood pressure.

then $\mu = 0$

then $t = \frac{\bar{x} - \mu}{s}\sqrt{\gamma} = \frac{(2.58-0)}{2.96}\sqrt{11} = 2.89$

$2.89 > t_{0.05}^{(11)} = 2.201$

∴ we reject the hypothesis

ie the stimulus is likely to change the blood pressure.

6. Eleven school boys were given a test in mathematics carrying a maximum of 25 marks. They were given a month's extra coaching and a second test of equal difficulty was held thereafter. The following table gives the marks in the two tests:

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

Do the marks give evidence that the students have benefited by extra coaching? Use 0.05 level of significance.

Soln: The difference in the marks is given by

$$1, -1, 3, -3, 2, 2, 2, 3, 0, 4, -2$$

$$\bar{x} = \frac{1}{11}(1-1+3-3+2+2+2+3+0+4-2) = 1$$

$$s^2 = \frac{1}{11}\{(1-1)^2+(-1-1)^2+(3-1)^2+(-3-1)^2+(2-1)^2+(2-1)^2+(2-1)^2+(3-1)^2 + (0-1)^2+(4-1)^2+(-2-1)^2\}$$

$$s^2 = 4.545$$

Let the hypothesis be: the students have not been benefited by extra coaching.

i.e., $\mu = 0$

Then $t = \frac{\bar{x}-\mu}{s}\sqrt{n} = \frac{1-0}{\sqrt{4.545}}\sqrt{10} = 1.485$

$t_{0.05}(10) = 2.23$, and $1.485 < 2.23$.

$\therefore$ we accept the hypothesis.

i.e. the students have not been benefitted by extra coaching.

Chi-square Distribution

In all random trails there exists some discrepency between the expected (theoretical) frequencies and the observed frequencies, in general. The discrepancy is analysed through a test statistic called the chi-square, denoted by χ^2.

Suppose that, in a random experiment, a set of events $E_1, E_2, \ldots E_n$ are observed to occur with frequencies $f_1, f_2, \ldots f_n$. According to a theory based on probability rules, suppose the same events are expected to occur with frequencies $e_1, e_2, \ldots, e_n$. Then, $f_1, f_2, f_3 \ldots, f_n$ are called observed frequencies and $e_1, e_2, e_3, \ldots, e_n$ are called expected or theoretical frequencies.

Let us define the statistic χ^2 through the following formula:

$$\chi^2 = \frac{(f_1-e_1)^2}{e_1} + \frac{(f_2-e_2)^2}{e_2} + \cdots + \frac{(f_n-e_n)^2}{e_n} = \sum_{k=1}^{n} \frac{(f_k-e_k)^2}{e_k} \quad \text{--- (1)}$$

If N is the total frequency, then $N = \sum_{k=1}^{n} f_k = \sum_{k=1}^{n} e_k$

If the expected frequencies are atleast equal to 5, it can be proved that the sampling distribution of the statistic χ^2 is approximately identical with the probability distribution of the variate χ^2

whose density function is given by

$$p(\chi^2) = P_0 \chi^{\nu-2} e^{(-\chi^2/2)}$$

where ν is a positive constant, called the number of degrees of freedom, and P_0 is a constant depending on ν such that the total area under the corresponding probability curve is one. The probability distribution for which $p(\chi^2)$ is given by $P_0 \chi^{\nu-2} e^{(-\chi^2/2)}$ is called the chi-square distribution with ν degrees of freedom.

In practice, expected frequencies are computed on the basis of a hypothesis H_0. If under this hypothesis the value of χ^2 computed with the use of the formula (1) is greater than some critical value χ_c^2, we would conclude that the observed frequencies differ significantly from the expected frequencies and would reject H_0 at the corresponding level of significance, c. Otherwise, we would accept it or atleast not reject it. This procedure is called the chi-square(χ^2) Test of hypothesis or significance.

Generally the chi-square test is employed by taking $c = 0.05$ or 0.01.

The number of degrees of freedom ν is determined by the formula $\nu = n - m$, where n is the number of frequency-pairs (f_i, e_i), m is the number of quantities that are needed in the calculation of e_i.

Goodness of Fit

When a hypothesis H_0 is accepted on the basis of the chi-square test, we say that the expected frequencies calculated on the basis of H_0 form a good fit for the given frequencies. When H_0 is rejected, we say that the corresponding expected frequencies do not form a good fit.

1. In 200 tosses of a coin, 118 heads and 82 tails were observed. Test the hypothesis that the coin is fair at 0.05 and 0.01 levels of significance.

Sol^n observed frequencies $f_1 = 118$, $f_2 = 82$

expected frequencies $e_1 = 200 \times \frac{1}{2} = 100$

$e_2 = 200 \times \frac{1}{2} = 100$

$$\chi^2 = \frac{(f_1 - e_1)^2}{e_1} + \frac{(f_2 - e_2)^2}{e_2} = \frac{(118-100)^2}{100} + \frac{(82-100)^2}{100} = 6.48$$

$n = 2$, $N = \sum f_i = 200$, $m = 1$

$\therefore \nu = n - m = 2 - 1 = 1$.

$\chi^2_{0.05}(1) = 3.84$, $\chi^2_{0.01}(1) = 6.64$

$\therefore \chi^2 < \chi^2_{0.01}$ and $\chi^2 > \chi^2_{0.05}$

$\therefore$ the hypothesis is accepted at 0.01 l of s

& rejected at 0.05 l of s

2. The following table gives the number of road accidents that occured in a large city during the various days of a week. Test the hypothesis that the accidents are uniformly distributed over all the days of the week.

Day	Sun	Mon	Tue	Wed	Thu	Fri	Sat	Total
No of accidents	14	16	08	12	11	9	14	84

Sol:) $N = 84$, no of days $= 7$

$\therefore e_i = \frac{84}{7} = 12$

$$\chi^2 = \left\{ \frac{(14-12)^2}{12} + \frac{(16-12)^2}{12} + \frac{(8-12)^2}{12} + \frac{(12-12)^2}{12} + \frac{(11-12)^2}{12} + \frac{(9-12)^2}{12} + \frac{(14-12)^2}{12} \right\}$$

$$= 4.17$$

$n = 7$, $m = 1$ [$\because N = \sum f_i = 84$ is the only quantity used]

$\therefore \nu = 7 - 1 = 6$

$\chi^2_{0.05}(6) = 12.59$, $\chi^2_{0.01}(6) = 16.81$)

$\chi^2 < \chi^2_{0.05}$, $\chi^2 = \chi^2_{0.01}$

$\therefore$ the hypothesis is accepted.

ie, the accidents are distributed uniformly over the week.

3. A survey of 240 families with 3 children each revealed the distribution shown in the following table. Is the data consistent with the hypothesis that male and female births are equally probable? Use χ^2 test at 0.01 & 0.05 levels.

No of children	3B 0G	2B 1G	1B 2G	0B 3G
No of families	37	101	84	18

Sol'n Let p = probability of male births
q = probability of female births
Under the hypothesis male & female births are equally probable $p=\frac{1}{2}$, $q=\frac{1}{2}$.
Among 3 children, the probability that x children are boys is given by ${}^3C_x\left(\frac{1}{2}\right)^x\left(\frac{1}{2}\right)^{3-x}$

$\therefore p(3B) = {}^3C_3\left(\frac{1}{2}\right)^3\left(\frac{1}{2}\right)^{3-3} = \frac{1}{8}$; $p(2B) = {}^3C_2\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)^1 = \frac{3}{8}$

$p(1B) = {}^3C_1\left(\frac{1}{2}\right)^1\left(\frac{1}{2}\right)^{3-1} = \frac{3}{8}$; $p(0B) = {}^3C_3\left(\frac{1}{2}\right)^3\left(\frac{1}{2}\right)^{3-3} = \frac{1}{8}$

$\therefore$ Among 240 families, the expected number of families with 3B is $e_1 = 240 \times p(3B) = 240 \times \frac{1}{8} = 30$

2B is $e_2 = 240 \times \frac{3}{8} = 90$

1B is $e_3 = 240 \times \frac{3}{8} = 90$

0B is $e_4 = 240 \times \frac{1}{8} = 30$

From the table $f_1 = 37$, $f_2 = 101$, $f_3 = 84$, $f_4 = 18$

$\chi^2 = \frac{(37-30)^2}{30} + \frac{(101-90)^2}{90} + \frac{(84-90)^2}{90} + \frac{(18-30)^2}{30} = 8.1773$

$v = 4-1 = 3$ $\quad \chi^2_{0.05}(3) = 7.82$, $\chi^2_{0.01}(3) = 11.34$

$\chi^2 < \chi^2_{0.05}$ $\therefore$ consistent $\qquad \chi^2_2 > \chi^2_{0.01}$ $\therefore$ inconsistent

4. A set of five identical coins is tossed 320 times and the result is shown in the following table:

no of heads	0	1	2	3	4	5
frequency	6	27	72	112	71	32

Test the hypothesis that the data follows a binomial distribution associated with a fair coin.

Soln: $p = \frac{1}{2}$, $q = \frac{1}{2}$

Binomial distribution function is ${}^5C_x \left(\frac{1}{2}\right)^x \left(\frac{1}{2}\right)^{5-x}$

for toss of 320 times

$e_1 = 320 \times {}^5C_0 \left(\frac{1}{2}\right)^0 \left(\frac{1}{2}\right)^5 = 10$; $e_4 = 320 \times {}^5C_3 \left(\frac{1}{2}\right)^3 \left(\frac{1}{2}\right)^2 = 100$

$e_2 = 320 \times {}^5C_1 \left(\frac{1}{2}\right)^1 \left(\frac{1}{2}\right)^4 = 50$; $e_5 = 320 \times {}^5C_4 \left(\frac{1}{2}\right)^4 \left(\frac{1}{2}\right)^1 = 50$

$e_3 = 320 \times {}^5C_2 \left(\frac{1}{2}\right)^2 \left(\frac{1}{2}\right)^3 = 100$; $e_6 = 320 \times {}^5C_5 \left(\frac{1}{2}\right)^5 \left(\frac{1}{2}\right)^0 = 10$

$f_1 = 6$, $f_2 = 27$, $f_3 = 72$, $f_4 = 112$, $f_5 = 71$, $f_6 = 32$

$$\chi^2 = \frac{(6-10)^2}{10} + \frac{(27-50)^2}{50} + \frac{(72-100)^2}{100} + \frac{(112-100)^2}{100} + \frac{(71-50)^2}{50} + \frac{(32-10)^2}{10} = 78.68$$

$n = 6$, $m = 1$ $\therefore$ $\nu = 6 - 1 = 5$

$\chi^2_{0.05}(5) = 11.07$, $\chi^2_{0.01}(5) = 15.09$

$\chi^2 > \chi^2_{0.05}(5)$; $\chi^2 > \chi^2_{0.01}(5)$

$\therefore$ the hypothesis is rejected.

ie, the observed data doesnot follow the binomial distribution associated with a fair coin.