⊛ Paired t-test:

In the previous test it was assumed that the two samples were independent. However, there are many situations in which this condition does not hold true. i.e we have dependent samples. Two samples are said to be dependent when the observations in one sample are related to those in the other in a meaningful manner. In fact, the two samples may consist of pairs of observations made on the same object or individual. When the samples are dependent, they have equal sample size. We may carry out certain experiment to find out the effect of training on a group of employees.

Let $(x_1, x_2, x_3, \ldots, x_n)$ and $(y_1, y_2, \ldots, y_n)$ be two samples taken on the same units. Let μ_1 and μ_2 be the population means. Suppose the variances of the two populations are equal to σ^2.

We want to test the hypothesis $H_0 : \mu_1 = \mu_2$ against $H_1 : \mu_1 \neq \mu_2$

We use the statistic $\boxed{t = \dfrac{\bar{d}}{s/\sqrt{n}}}$;

where $\boxed{\bar{d} = \Sigma d_i / n}$; $\boxed{d_i = x_i - y_i}$ and

s = standard deviation of differences.

$$\therefore \boxed{s = \sqrt{\frac{\sum d^2 - n(\bar{d})^2}{n-1}}} \quad \underline{OR} \quad \boxed{s = \sqrt{\frac{1}{n-1}\left[\sum d^2 - \frac{(\sum d)^2}{n}\right]}}$$

$$\underline{OR} \quad \boxed{s = \sqrt{\frac{1}{n-1}\sum (d_i - \bar{d})^2}}$$

$$\boxed{d.f = n-1}$$

We find $t_{n-1, 0.05}$ value from the table of t-distribution.

If $|t| \le t_{n-1, 0.05}$, we accept H_0 at 5% level of significance, otherwise we reject it.

Ex 1. The systolic blood pressure (BP) of 9 normal individuals was recorded. Then 2ml of 0.5% solution of hypotensive (BP lowering) drug was given and the BP was recorded again. The difference in the blood pressure of the individuals before and after giving the injection is given below. Did the injection of the drug lower the blood pressure? ($t_{8, 0.05} = 2.31$)

Patient No.	1	2	3	4	5	6	7	8	9
Difference in BP	2	3	5	5	4	0	4	1	3

Sol^n Null Hypothesis H_0: The drug has no significant effect on the change of the blood pressure.

Given d: 2, 3, 5, 5, 4, 0, 4, 1, 3.

Scanned with CamScanner

$$\bar{d} = \frac{\Sigma d}{n} = \frac{27}{9} = 3 \quad \text{and} \quad s = \sqrt{\frac{\Sigma d^2 - n(\bar{d})^2}{n-1}}$$

$$= \sqrt{\frac{105 - 9(3)^2}{9-1}}$$

$$= 1.3229$$

$$\therefore t = \frac{\bar{d}}{s/\sqrt{n}} = \frac{3}{1.3229/\sqrt{9}} = \frac{9}{1.3229} = 6.80$$

$d.f = n-1 = 9-1 = 8$

given that $t_{8,0.05} = 2.31$

$\therefore |t| > t_{8,0.05}$

$\therefore H_0$ reject at 5% level of significance and conclude that the drug has significant effect in lowering the BP.

EX-2 An I.Q test was administered to 5 computer engineers before and after they were trained. The results are given below.

Candidate No:	1	2	3	4	5
I.Q. before training	110	120	123	132	125
I.Q after training	120	118	125	136	127

Test whether there is any change in I.Q. after the training Programme.

sol^n : We wish to test the hypothesis.

H_0 : There is no significant change in I.Q. due to training.

$$d_i = x_i - y_i = -10, 2, -2, -4, -2.$$

$$\therefore \sum d_i = -16,$$

$$d_i^2 = 100, 4, 4, 16, 4 \Rightarrow \sum d_i^2 = 128$$

$$\therefore \bar{d} = \frac{\sum d_i}{n} = \frac{-16}{5} = -3.2$$

$$S = \sqrt{\frac{\sum d^2 - n(\bar{d})^2}{n-1}} = \sqrt{\frac{128 - 5(-3.2)^2}{5-1}} = 4.38.$$

$$|t| = \left|\frac{\bar{d}}{s/\sqrt{n}}\right| = \left|\frac{-3.2}{4.38/\sqrt{5}}\right| = 1.63.$$

$$d.f = n-1 = 5-1 = 4.$$

Find the value from the table of t-distributio

$$t_{4,\,0.05} = 2.776.$$

$$\therefore |t| < t_{4,\,0.05}$$

$\therefore$ we accept H_0 at 5% level of significance and conclude that there is no significant change in I.Q due to training.

Ex-3 12 students were given intensive coaching and two tests were conducted in a month. The scores of test 1 and 2 are given below. Do the scores from the test 1 to the test 2 show an improvement?

No of students	1	2	3	4	5	6	7	8	9	10	11	12
Marks in 1st test	50	42	51	26	35	42	60	41	70	55	62	38
marks in 2nd test	62	40	61	35	30	52	68	51	84	63	72	50

Soln: Let x_i = marks of a students in the first test.

y_i = marks of a students in the second test.

$d_i = y_i - x_i$

$\therefore d_i = 12, -2, 10, 9, -5, 10, 8, 10, 14, 8, 10, 12.$

$d_i^2 = 144, 4, 100, 81, 25, 100, 64, 100, 196, 64, 100, 144$

$$\therefore \bar{d} = \frac{\Sigma d_i}{n} = \frac{96}{12} = 8.$$

$$s = \sqrt{\frac{\Sigma d_i^2 - n(\bar{d})^2}{n-1}} = \sqrt{\frac{1122 - 12(8)^2}{12-1}} = 5.6729$$

We wish to test H_0: There is no significant difference in the marks of the two tests.

$$\text{Here } t = \frac{\bar{d}}{s/\sqrt{n}} = \frac{8}{5.6729/\sqrt{12}} = \frac{\sqrt{12}\times 8}{5.6729} = 4.89$$

$d.f = n-1 = 12-1 = 11$

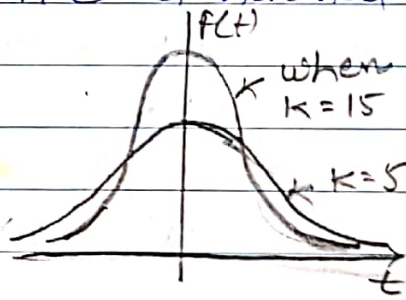
From the table of t-distribution, we find $t_{11, 0.05} = 2.201$.

$\therefore t > t_{11, 0.05}$ $\therefore$ we reject H_0

$\therefore$ conclude that the scores show improvement from test 1 to test 2.

⊛ Properties of t-distribution.

1) t-distribution is unimodal distribution, curve is symmetrical about the line $t=0$ and it is bell shaped curve just like a normal curve.



2) t-distribution has only one parameter K = degree of freedom. Its spread increase as K increases.

3) The constants are as follows:

   Mean = 0, for $K \geq 2$

   Variance $\sigma^2 = \dfrac{K}{K-2}$; for $K \geq 3$.

4) The area under t-distribution curve for $t < t_0$ is determined by the equation.

$$P(t < t_0) = \int_{-\infty}^{\infty} f(t)\, dt$$

(But need not to integrat because use t-distribution table.)

5) t-distribution has tremendous utility in testing of hypothesis when $n < 30$ and standard deviation of population is not known.

Scanned with CamScanner