PAPER COM 202.2

ADVANCED BUSINESS STATISTICS

Chapter 4: Parametric Test

Independent Sample t- test:

Independent sample t- test is used when we want to test whether the means of two different populations are equal or not. For testing the equality of two population means we draw two samples from two populations independently. For this reason it is also known as independent sample t test. In this case two independently drawn samples from two populations may be equal or unequal in size. For the independent samples t-test it is assumed that both samples come from normally distributed populations with homogeneous variances.

Examples of typical questions that the independent samples t-test answers are as follows:

- **Medicine** – Has the quality of life improved for patients who took drug A as opposed to patients who took drug B?
- **Sociology –** Are men more satisfied with their jobs than women?
- **Biology –** Are foxes in one specific habitat larger than in another?
- **Economics** – Is the economic growth of developing nations larger than the economic growth of the first world?
- **Marketing:** Does customer segment A spend more on groceries than customer segment B?

Calculation of t statistic:

$$t = \frac{\text{Observed value} - \text{Expected value}}{\text{Standard error}}$$

$$t = \frac{(\overline{X1} - \overline{X2}) - (\mu 1 - \mu 1)}{S\,E_{(X1 - X2)}}$$

Where, $S\,E_{(\overline{X1} - \overline{X2})} = S^{/}\sqrt{\left(\frac{1}{n1} + \frac{1}{n2}\right)}$

Where, $S^{/} = \sqrt{\frac{n1*s1*s1 + n2*s2*s2}{n1 + n2 - 2}}$

Steps of independent sample t test:

1) Establish null and alternative hypotheses
2) Fix up the level of significance of the test
3) Calculate the value of t statistic
4) Calculate the degree of freedom (df = n1 + n2 - 2)
5) Pick up the table value of t from the t table for corresponding level of significance and corresponding degrees of freedom.
6) Compare the calculated value of t with the table value (critical value)
7) If your computed value of t lies in the critical region reject the null hypothesis. Else, accept it.

Numerical problems on independent sample t test:

Problem- 1

A study was conducted to compare the efficiency of workers of two mines, one with private ownership and the other with government ownership. The researcher was interested to test whether there is any significant difference in their efficiency level. 20 workers from private sector mine and 24 from government sector mine were selected and their average output per shift was recorded.

	Mean output per shift	S. D. of output
Private mine worker	42.50	5.50
government mine worker	39.25	2.72

Test at 5% level whether there is any significant difference in efficiency of private mine worker and government mine workers.

Solution 1:

Let us take the hypothesis that there is no significant difference between the efficiency of private mine workers and government mine workers.

Therefore, our null hypothesis H0: ($\mu 1 = \mu 1$)

And the alternative hypothesis H1: ($\mu 1 \neq \mu 1$)

Here, the level of significance of the test is 5% or $\alpha = 0.05$

We know that,

$$t = \frac{\text{Observed value} - \text{Expected value}}{\text{Standard error}}$$

$$t = \frac{(\overline{X1} - \overline{X2}) - (\mu 1 - \mu 1)}{S\,E_{(X1-X2)}}$$

Where, $S\,E_{(\overline{X1}-\overline{X2})} = S^{/}\sqrt{\left(\frac{1}{n1} + \frac{1}{n2}\right)}$

Where, $S^{/} = \sqrt{\frac{n1*s1*s1+n2*s2*s2}{n1+n2-2}}$

$$= \sqrt{\frac{20*5.5*5.5+24*2.72*2.72}{20+24-2}} = 4.32$$

Computed value of $S^{/} = 4.32$

Therefore, S E = 4.32 √ (1/20 + 1/24) = 1.308

Therefore, t = [(42.5- 39.25) – 0] / 1.308

= 2.485

Here, the degree of freedom = n1 + n2 -2 = 20+24-2 = 42

The critical value of t for 5% level of significance and for 42 degrees of freedom is 1.96 [$t_{0.025,42} = 1.96$].

Since, our computed value of $|t|$ (2.485) is more than the critical value; there is no reason to accept our null hypothesis at 5% level of significance.

We may therefore conclude that there is a significant difference in the efficiency level of the workers of private mines and the government mines. Actually the private miners are more efficient than the government miners.

Problem 2:

An investigator wants to know if males spend more or fewer minutes on the phone each day. For this purpose he takes samples 10 males and 10 females and asks them about their phone calling duration. Part of the data is shown below.

	N	Average time (minute)	Standard deviation (Minute)
Female	10	22	9.42
Male	10	27	5.66

Do the data supports that the average phone time of male differ significantly from female? Use α =0.01

Solution 2:

Let us take the hypothesis that there is no significant difference between the male and female so far as the average times spend on phone per day is concerned.

Therefore, our null hypothesis H0: ($\mu 1 = \mu 1$)

And the alternative hypothesis H1: ($\mu 1 \neq \mu 1$)

Here, the level of significance of the test is 1% or $\alpha = 0.01$

We know that,

$$t = \frac{\text{Observed value} - \text{Expected value}}{\text{Standard error}}$$

$$t = \frac{(\overline{X1} - \overline{X2}) - (\mu 1 - \mu 1)}{S\,E_{(X1 - X2)}}$$

$$\text{Where, } S\,E_{(\overline{X1} - \overline{X2})} = S^{/}\sqrt{\left(\frac{1}{n1} + \frac{1}{n2}\right)}$$

$$\text{Where, } S^{/} = \sqrt{\frac{n1*s1*s1+n2*s2*s2}{n1+n2-2}}$$

$$= \sqrt{\frac{10*9.42*9.42+10*5.66*5.66}{10+10-2}} = 8.19$$

Computed value of $S^{/} = 8.19$

Therefore, S E = 8.19 √ (1/10 + 1/10) = 3.66

Therefore, t = [(22- 27) – 0] / 3.66

= 1.36

Here, the degree of freedom = n1 + n2 -2 = 10+10-2 = 18

The critical value of t for 1% level of significance and for 18 degrees of freedom is 2.878 [$t_{0.005,18}$ = 2.878].

Since, our computed value of $|t|$ (1.36) is less than the critical value; there is no reason to reject our null hypothesis at 1% level of significance.

We may therefore conclude that there is no significant difference in the average telephone time spent by male and female.

Paired t test:

The paired sample *t*-test, sometimes called the dependent sample *t*-test, is a statistical procedure used to determine whether there is any significant difference in the mean values of two sets of observations. In a paired sample *t*-test, each subject or entity is measured twice; resulting in *pairs* of observations, where the observations are not independent rather those are pair-wise dependent. Suppose we are interested in evaluating the effectiveness of a company training program. One approach you might consider would be to measure the performance of a sample of employees before and after completing the training program, and analyze the differences using a paired sample *t*-test. In case of paired sample t test n_1 must be equal to n_2 ($n_1 = n_2$).

Calculation of paired-t statistic:

$$t = \bar{d} / [S_d / \sqrt{(n-1)}]$$

($\bar{d}$ Divided by Standard deviation of d (S_d) and that standard deviation of d is divided by root over n-1)

Here, d = difference in observed values of each pair.

Problem 3: IQ test was administered on 5 persons before and after they were trained. The results are given below:

Candidate	I	II	III	IV	V
IQ before training	110	120	123	132	125
IQ after training	120	118	125	136	121

Test whether there is any significant change (gain) in IQ level after the training program.

Solution 3:

Let us take the hypothesis that there is no significant change (improvement) in the average IQ level after the training program.

Therefore, our null hypothesis H0: ($\mu x = \mu y$)

And the alternative hypothesis H1: ($\mu x > \mu y$)

Here, μy is the average IQ level after the training program and μx is the average IQ level before the training program.

Let us fix the level of significance of the test at 5%.

We know that, $t = \bar{d} / [S_d / \sqrt{(n-1)}]$

Table showing for necessary calculations:

Candidate	IQ after training (x)	IQ before training (y)	D = x-y	D^2
I	120	110	10	100
II	118	120	-2	4
III	125	123	2	4
IV	136	132	4	16
V	121	125	4	16
		SUM	**$\sum D = 10$**	**$\sum D^2 = 140$**

Here, $\bar{d} = 10/5 = 2$

$S_d = \sqrt{[\sum D^2 / n - (\sum D/n)^2]} = \sqrt{[140/5 - (2)^2]} = \sqrt{24} = 4.9$

$$t = \bar{d} / [S_d / \sqrt{(n-1)}]$$

$$= 2 / [4.9 / / \sqrt{(5-1)}]$$

$$= 2 / 2.45$$

$$\mathbf{= 0.816}$$

Here, degree of freedom = n-1 = 5-1 = 4

The critical value of t at 5% level of significance and for 4 degree of freedom is 2.132 [$t_{0.05,4} = 2.132$].

Since, our computed value of t (0.816) is less than the critical value there is no reason to reject our null hypothesis at 5% level of significance.

We may therefore conclude that there is no significant improvement in average IQ level after the training program.

Problem 4:

Suppose a sample of 15 students were given a diagnostic test before studying a particular module and then again after completing the module. We want to find out if, in general, whether the module helped to improvements in students' knowledge/skills (i.e. test scores).

Students	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Pre- module score	18	21	16	22	19	24	17	21
Post module score	22	25	17	24	16	29	20	23
Students	**9**	**10**	**11**	**12**	**13**	**14**	**15**	--
Pre- module score	23	18	14	16	16	19	18	--
Post module score	19	20	15	15	18	26	18	--

Apply paired t test to find whether there is any significant improvement in the score after studying the module. Use $\alpha = 0.05$.

Solution 4:

Let us take the hypothesis that there is no significant change (improvement) in the average score after studying the module.

Therefore, our null hypothesis H0: (μx = μy)

And the alternative hypothesis H1: (μx > μy)

Here, μy is the average score of the students after studying the module and μx is the average score of the students before studying the module.

Here, level of significance of the test at 5%.

We know that, $t = \bar{d} / [S_d / \sqrt{(n-1)}]$

Table showing for necessary calculations:

Students	**IQ after training (x)**	**IQ before training (y)**	**D = x-y**	**D^2**
1	22	18	4	16
2	25	21	4	16
3	17	16	1	1
4	24	22	2	4
5	16	19	-3	9
6	29	24	5	25
7	20	17	3	9
8	23	21	2	4
9	19	23	-4	16
10	20	18	2	4
11	15	14	1	1
12	15	16	-1	1
13	18	16	2	4
14	26	19	7	49
15	18	18	0	0
		SUM	**$\sum D = 25$**	**$\sum D^2 = 159$**

Here, $\bar{d}$ = 25/ 15 = 1.67

$S_d = \sqrt{[\sum D^2 /n - (\sum D/n)^2]} = \sqrt{[159/15 - (1.67)^2]}$ = 2.8

$t = \bar{d} / [S_d /\sqrt{(n-1)}]$

$= 1.67 / [2.8 /\sqrt{(15-1)}]$

= 1.67 / 0.75

= 2.23

Here, degree of freedom = n-1 = 5-1 = 14

The critical value of t at 5% level of significance and for 14 degree of freedom is 1.761 [$t_{0.05,4}$ = 1.761].

Since, our computed value of t (2.23) is more than the critical value there is no reason to accept our null hypothesis at 5% level of significance.

We may therefore conclude that there is a significant improvement in average score of the students after studying the module.