21-04-2022

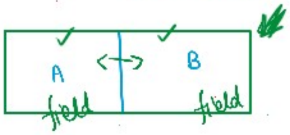
Thursday, April 21, 2022 11:59 AM

Unit-5

Student's t-test
F-test

[redacted] Unit-5 (Hypothesis Testing)

* Student's t-test — Sir William Gosset in 1908.



§ Assumptions / Conditions of t-test:

t-test
z-test

1. Parent population should follow normal distribution.
2. Sample size ~~should~~ may be small $(n < 30)$
3. The sample observations should be independent i.e the sample is random.

Z-test
Large samples

Remark :- This test can be used for large samples as well.

Variance

constant

sample variance ← s^2 $\quad S^2$ → estimated variance $\quad \sigma^2$ → population variance

$\varepsilon(s^2) = S^2$

$$\boxed{\frac{s^2}{n-1} = \frac{S^2}{n}}$$

$$\boxed{\begin{aligned} S^2 &= \frac{1}{n-1}\Sigma(x_i-\bar{x})^2 \\ s^2 &= \frac{1}{n}\Sigma(x_i-\bar{x})^2 \end{aligned}}$$

$E(s^2) \neq \sigma^2$

$E(s^2) = S^2$

$E\left(\frac{n}{n-1}s^2\right) = \sigma^2$

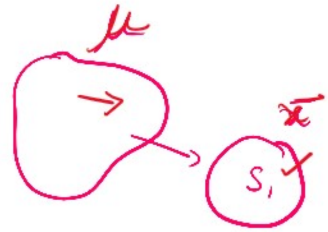
Applications for t-test:

(a) t-test for single mean — To compare the sample mean with population mean.

or

to test the significance of the sample mean

(calculated value) $$\boxed{t = \frac{\bar{x}-\mu}{\sqrt{S^2/n}} = \frac{\bar{x}-\mu}{\sqrt{s^2/n-1}}}$$



$t \leq t_{\alpha}(d.f)$ H_0 - accepted

$t > t_{\alpha}(d.f)$ H_0 - rejected

where —

- $\bar{x}$ = sample mean
- μ = population mean
- n = sample size
- s^2 = sample variance

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

$S^2 = \frac{1}{n-1}\Sigma(x_i-\bar{x})^2$

$s^2 \quad S^2$

n = sample size

s^2 = sample variance

S^2 = estimated variance

$$S^2 = \frac{1}{n-1}\Sigma(x_i - \bar{x})$$

$$\boxed{\frac{s^2}{n-1} = \frac{S^2}{n}}$$

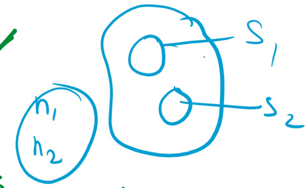
H_0 - rejected

$$s^2 = \frac{n}{n-1} S^2$$

$$\Rightarrow \frac{S^2}{n-1} = \frac{s^2}{n}$$

(2) t-test for Difference of means: To test the significance between means of two independent samples.

$$\boxed{t = \frac{\bar{x}_1 - \bar{x}_2}{S\sqrt{\frac{n_1+n_2}{n_1 n_2}}}}, \text{ where } S = \sqrt{\frac{\Sigma(x_i - \bar{x}_1)^2 + \Sigma(x_j - \bar{x}_2)^2}{n_1+n_2-2}}$$



where, S = combined estimated variance.

n_1 = size of sample 1

n_2 = size of sample 2

x_i = terms of sample 1

x_j = terms of sample 2

$\bar{x}_1$ = mean of sample 1

$\bar{x}_2$ = mean of sample 2

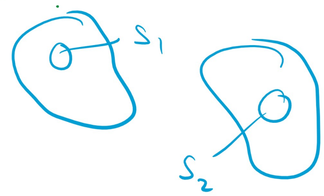
3 5

3+5-2

6

$$d.f = (n_1-1)+(n_2-1)$$

$$= \boxed{n_1+n_2-2}$$

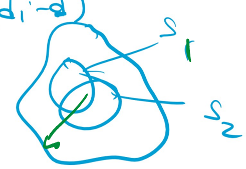


$n_1 = 6 \quad n_2 = 5 \quad \therefore t_\alpha(d.f)$

$t(9)$

3) Paired t-test: To test the difference between means of dependent samples.

$$t = \frac{\bar{d}}{\sqrt{S^2/n}}, \text{ where } S = \sqrt{\frac{(\Sigma d_i)^2 - n(\bar{d})^2}{n-1}} \Rightarrow S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(d_i - \bar{d})^2$$



$\bar{d}$ = mean of differences

S = estimated variance

$n = n_1 = n_2$

$$d_i = x_i - y_i$$

$$\bar{d} = \frac{1}{n}\Sigma x_i - y_i$$

$x_i \quad y_i$

$x_i - y_i$

Problems

Pb 1:- 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?

→ $\mu = 146.3, \; n = 22, \; \bar{x} = 153.7, \; s = 17.2.$

Null-hypothesis: The adv. campaign is not successful $H_0: \mu = 146.3$.

Alternative " : $H_1: \mu > 146.3$ (Right-tailed).

Under H_0: test statistic is $t = \dfrac{\bar{x} - \mu}{\sqrt{S^2/n}} \sim t_{22-1} = t_{21}$

Under H_0. test statistic is $t = \frac{\bar{x}-\mu}{\sqrt{S^2/n-1}} \sim t_{22-1} = t_{21}$

Now,

$$t = \frac{153.7 - 143.6}{\sqrt{(17.2)^2/21}} = \frac{7.4 \times \sqrt{21}}{17.2} = 9.03$$

Tabulated value of $t_{0.05}(21)$ is 1.72.

Since $t > t_{0.05}(21)$, So H_0 is rejected $\Rightarrow H_1$ - accepted i.e compaign is successful.

Pb2: Samples of two types of electric light bulbs were tested for length of life & following data were obtained:

S_1 8 S_2 7

$\bar{x}_1 = 1234$ $\bar{x}_2 = 1036$

$S_1 = 36$ $S_2 = 40$

	Type I	Type II
Sample no:	$n_1 = 8$	$n_2 = 7$
sample means	$\bar{x}_1 = 1234$ 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 warranty that type I is superior to type II regarding length of life?

Solⁿ: Null hypothesis, $H_0: \mu_x = \mu_y$ i.e the two types I & II are identical.

Alt. " $H_1: \mu_x > \mu_y$ i.e type I is superior to type II.

Test statistic, $t = \frac{\bar{x} - \bar{y}}{S\sqrt{\frac{n_1+n_2}{n_1 n_2}}} \sim t_{n_1+n_2-2} = t_{13}$.

where $S^2 = \frac{1}{n_1+n_2-2}\left[\Sigma(x_i-\bar{x})^2 + \Sigma(y_i-\bar{y})^2\right] = \frac{1}{13}\left[n_1 s_1^2 + n_2 s_2^2\right]$

$s_1^2 = \frac{1}{n}\Sigma(x_i-\bar{x})^2$

$n\, s_1^2 = \Sigma(x_i-\bar{x})^2$

$= \frac{1}{13}(8 \cdot 36^2 + 7 \cdot 40^2)$

$\Rightarrow S^2 = 1659.08$.

$$\therefore t = \frac{1234 - 1036}{\sqrt{1659.8}\sqrt{\frac{15}{56}}} = 9.39$$

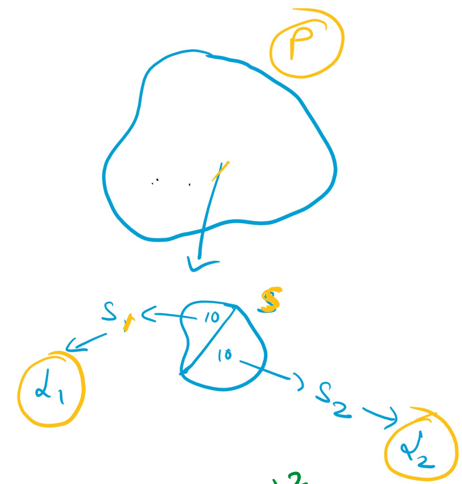
Since $t > t_{0.05}(13)$, H_0 is rejected

$\sqrt{1659.8}\ \sqrt{\frac{15}{56}}$

But $t_{0.05}(13) = 1.77$. Since $t > t_{0.05}(13)$, H_0 is rejected

$\Rightarrow$ type I is superior to type II.

Pb 3 :- 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 two laboratories. The foll^n data is obtained:



No. of samples	, n	10
Mean value of the difference of estimates	$\bar{d}$ =	0.6
Sum of the squares of the differences		20

Is the difference significant?

$\Sigma(d-\bar{d})^2$

Sol^n:

H_0: different is not significant; $\mu_1 = \mu_2$.

H_1: $\mu_1 \neq \mu_2$ ($\mu_1 > \mu_2$, $\mu_1 < \mu_2$).

$\bar{d} = 0.6$, $n = 10$, $\Sigma(d-\bar{d})^2 = 20$.

Under null hypothesis, $t = \dfrac{\bar{d}}{\sqrt{S^2/n}} \sim t_{10-1} = t_9$

$\left\{ \begin{array}{l} var(x) = \\ \varepsilon(x-\varepsilon(x))^2 \\ = \varepsilon(x^2) - (\varepsilon(x))^2 \end{array} \right.$

$S^2 = \frac{1}{n-1}\Sigma(d-\bar{d})^2 = \frac{1}{9}20 = 2.22$

$\therefore t = \dfrac{0.6}{\sqrt{2.22/10}} = 1.274$. $\quad t_{0.05}(9) = 2.262$ (given)

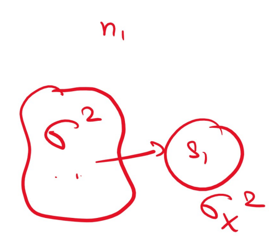
Since $t < t_{0.05}(9)$, so H_0 is accepted. The diff is insignificant.

---

* **F-test** :- To test the equality of two population Variances –

Suppose we want to test whether two independent samples x_i ($i = 1$ to n_1) and y_i ($i = 1$ to n_2) have been drawn from the normal populations with the same variance σ^2 (say).

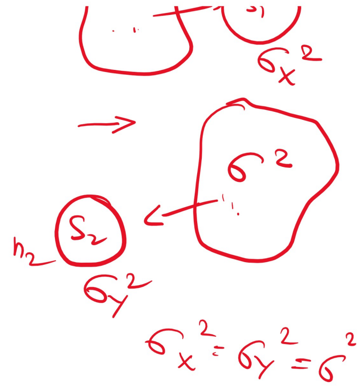
Under the null hypothesis, $H_0: \sigma_x^2 = \sigma_y^2 = \sigma^2$, the statistic F is given by –



Under the null hypothesis, $H_0: \sigma_x^2 = \sigma_y^2 = \sigma^2$, the statistic F is given by –

$$\rightarrow F = \frac{S_x^2}{S_y^2}, \text{ where –}$$

$$\underbrace{S_x^2 = \frac{1}{n_1 - 1}\sum_{i=1}^{n_1}(x_i - \bar{x})^2} \quad \& \quad \underline{S_y^2 = \frac{1}{n_2 - 1}\sum_{i=1}^{n_2}(y_i - \bar{y})^2}.$$



Pb4:- 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 the difference is significant at 5% level, given that the 5% of F for $n_1 = 7$ & $n_2 = 9$ degrees of freedom is 3.29.

Solⁿ:

$$\sum_{i=1}^{8}(x - \bar{x})^2 = 84.4 \qquad \sum_{i=1}^{10}(y - \bar{y})^2 = 102.6$$

$$\therefore S_x^2 = \frac{1}{n_1 - 1}\Sigma(x - \bar{x})^2 = \frac{1}{7} \times 84.4 = 12.057$$

$$S_y^2 = \frac{1}{n_2 - 1}\Sigma(y - \bar{y})^2 = \frac{1}{9} \times 102.6 = 11.4$$

Under null hypothesis, $H_0: \sigma_x^2 = \sigma_y^2 = \sigma^2$, the test statistic is,

$$F = \frac{S_x^2}{S_y^2} = \frac{12.057}{11.4} = 1.057$$

Tab. value $F_{0.05}(7, 9)$ is 3.29. Since $F < F_{0.05}(7, 9)$, so H_0 is accepted $\Rightarrow$ diffⁿ is insignificant.

Pb Which of the follⁿ test can be used for small sample size?

(A) χ^2-test → frequencies

(B) Z-test

(C) t-test

(D) F-test

} parametric

Syllabus is done