# Data Science & AI


## Probability and Statistics

### Testing of Hypothesis

**DPP Discussion Notes**

By- Rajshree mam

#Q. Ten individuals are chosen at random from a normal population of students and their marks are found to be 63, 63, 66, 67, 68 69, 70, 70, 71, 71. In the light of those data, discuss the suggestion that mean mark of the population of students is 66.

$H_0$

$$\text{Mean} = \frac{678}{10} = 67.8$$

$H_0$ :- Mean mark of population of stu is 66

$H_1$ :- — — — — — is not 66.

$\underbrace{\text{T- test}}$

$$T = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

$\mu = 66$

$\bar{X} =$ sample Mean

$n \rightarrow \underline{\underline{10}}$

$= 67.8$

$\bar{x} = \underline{\underline{67.8}}$

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

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $x_i$ | 63 | 63 | 66 | 67 | 68 | 69 | 70 | 70 | 71 | 71 | |
| $x_i - \bar{x}$ | – | – | – | – | – | – | – | – | – | – | |
| $(x_i - \bar{x})^2$ | 23.04 | 23.04 | 3.24 | 0.64 | 0.04 | 1.44 | 4.84 | 4.84 | 10.24 | 10.24 | |

$\sigma = \frac{\cancel{81.6}}{\cancel{9}} = 9.066$

$$T = \frac{67.8 - 66}{\frac{9.066}{\sqrt{10}}}$$

$\underline{\underline{0.05}}$ significance level

$$= \frac{67.8 - 66}{2.86}$$

$$= \underbrace{0.6293}$$

$\underline{\underline{T_{critical}}}$

$d.of = \underline{\underline{9}} = \underline{\underline{n-1}}$

$t_{critical} =$

$2.2\underline{\underline{6}}22$ ✓

$T < t_{\underline{\underline{critical}}}$

Acce$\underline{\underline{pted}}$

$H_0$: mean length of the cons as 46

#Q. The following values gives the lengths of 12 samples of Egyptian cotton taken from a consignment : 48, 46, 49, 46, 52, 45, 43, 47, 47, 46, 45, 50. Test if the mean length of the consignment can be taken as 46.

$n = 12$

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

$$T = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$$

$\sqrt{12} = 3.464$

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

$H_0$

$\bar{x} = 47$

$\mu = 46$

$$T = \frac{47 - 46}{\frac{6}{\sqrt{n}}}$$

$$= \frac{3.464 \times 1}{6}$$

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| $x_i$ | | | | | | | |
| $(x_i - \bar{x})^2$ | | | | | | | |
| | | | | | | | |

$$6 \rightarrow \frac{(x_i - \bar{x})^2}{n-1}$$

$H_1$: can't be taken as 46

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $x_i$ | 48 | 46 | 49 | 46 | 52 | 45 | 43 | 48 | 47 | 46 | 45 | 50 | |
| $(x_i-\bar{x})$ | 1 | 1 | 2 | 1 | 5 | 2 | 4 | 0 | 0 | 1 | 2 | 3 | |
| $(x_i-\bar{x})^2$ | 1 | 1 | 4 | 1 | 25 | 4 | 16 | 0 | 0 | 1 | 4 | 9 | |

$$\sigma = \frac{\cancel{66}}{\cancel{11}} = 6$$

$$T = \frac{\bar{X} - \mu}{\sigma} \times \sqrt{n}$$

$$= \frac{1}{6} \sqrt{n}$$

$$= \frac{1}{6} \times 3.464$$

$$= \underline{\underline{0.577}}$$

$\underline{\underline{0.05}}$ $\alpha$

$n-1 = 11$

$\underline{\underline{2.2010}}$

$\underline{\underline{T_{critical}}}$

Accepted

$\alpha$ $T < T_{critical}$

$H_0$ :- 27 units
$H_1$ : It is not 27 units

#Q. A sample of 18 items has a mean 24 units and standard deviation 3 units. Test the hypothesis that it is a random sample from a normal population with mean 27 units.

$n = 18$

$\bar{x} = 24$

$\mu = 27$

$\sigma = 3$

$$t = \frac{24-27}{\frac{3}{\sqrt{18}}}$$

$$= \frac{-\cancel{3}}{\cancel{3}} \times \sqrt{18}$$

$$= -\sqrt{18}$$

$$= -4.24$$

Modulus

$\longrightarrow$ 4.24

0.05 = $9-1 = 17$ d.f

t > t critical

Rejected

2.1098 = t critical

#Q. A random sample of 10 boys had the I.Q's 70,120,110,101,88,83,95,98, 107 and 100, Do these data support the assumption of a population mean L.Q. of 160?

$\mu = \underline{\underline{160}}$

$n = 10$

d.o.f $\rightarrow$ $n-1 = \underline{\underline{9}}$

t critical for $\underline{\underline{0.05}} \rightarrow$ 2.2622

$H_0$:- population mean of IQ is $\underline{\underline{160}}$.

| $x_i$ | 70 | 120 | 110 | 104 | 89 | 83 | 95 | 98 | 107 | 100 |
|---|---|---|---|---|---|---|---|---|---|---|
| $x_i - \bar{x}$ | 739.84 | | | | | | | | | |

$$\bar{x} = 190 + 211 + 171 + 193 + 207$$

$$= 97.2$$

$$\Sigma = 739.84 + 519.84 +$$

$$t = \frac{\bar{x} - \mu}{\sigma} \times \sqrt{n}$$

$n = 10$

$= \sqrt{10}$

$\mu = 160$

$\bar{x} = 972$

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

$\sigma = 203.837x$

$T = 0.97$

Value of $t$

$t_{critical} = 2.2622$

$t < t_{critical}$ Accept

$t > t_{critical}$ Reject

#Q. The mean life of 10 electric motors was found to be1450 hrs with S.D. of 423 hrs. A second sample of 17motors chosen from a different batch showed a mean life of 1280 hrs with a S.D. of 398 hrs. Is there a significant difference between means of the two samples?

$n_1 = 10$

$1450 = \bar{x}$

$SD = 423$

$n_2 = 17$

$\bar{x} = 1280$

$SD = 398$

$H_1$ :- Mean is the same.

$H_0$ :- Mean has a significant diffn.

One tail test

f-test

$n_1 = 10$

$\bar{x}_1 = 1450$

$\sigma_1 = 423$

SD

$n_2 = 17$

$\bar{x}_2 = 1280$

$\sigma_2 = 398$

diffⁿ variance & diffⁿ SD

$$T_{value} = \frac{\mu_1 - \mu_2}{\sqrt{\frac{V_1}{n_1} + \frac{V_2}{n_2}}}$$

$$\boxed{V_1 = \sqrt{6}}$$

$$\boxed{V_1^2 = 6}$$

$$DOF = \frac{\left\{\left(\frac{V_1}{n_1}\right)^2 + \left(\frac{V_2}{n_2}\right)^2\right\}^2}{\frac{\left(\frac{V_1^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{V_2^2}{n_2}\right)^2}{n_2 - 1}}$$

$n = 10$ $\quad$ $d.o.f = n-1 = 9$

$SD = V^2$

$V = \sqrt{SD}$

$$T\,value = \frac{1450 - 1280}{\sqrt{\sqrt{\frac{423}{10}} + \sqrt{\frac{398}{17}}}}$$

$$= 1.794$$

$$dof = \frac{\left(\frac{423}{10}\right)^2 + \left(\frac{398}{17}\right)^2}{\frac{\left(\frac{423}{10}\right)^2}{9} + \left(\frac{398}{17}\right)^2/16} = 18.63 \approx 18$$

dof = 15 significance level

0.05

2.1009 → $t_{critical}$

Tvalue 1.794

Tvalue < $t_{critical}$

$H_0$ is accepted

$H_1$ :→ mean has no diff^n

$H_0$ :→ It has a significant diff^n

#Q. The marks obtained by a group of 9 regular course students and another group of 11 part time course students in a test are given below:

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Regular | 56 | 62 | 63 | 54 | 60 | 51 | 67 | 69 | 58 | | |
| Part time | 62 | 70 | 71 | 64 | 60 | 56 | 75 | 64 | 72 | 68 | 66 |

Examine whether the marks obtained. by regular students and part time students differ significantly at 5% and 1% level of significance.

0.05 5%

0.01 1%

$n_1 = 9$

$$\bar{x}_1 = \frac{56 + 62 + 63 + 54 + 60 + 57 + 62 + 69 + 58}{9}$$

$= 60$

$n_1 = 11$

$$\bar{x}_2 = \frac{62 + 70 + 71 + 62 + 60 + 58 + 65 + 64 + 72 + 68 + 66}{11}$$

$= 66$

$\sigma_1 = \text{diff}^n$ & $\sigma_2$ $\text{diff}^n$

$$\sigma_1 = \frac{(x_i - \bar{x})^2}{n-1}$$

$$\boxed{\sigma_1 = \frac{(x_i - \bar{x})^2}{8}}$$

$\bar{x} = 60$

$x_i \to$ data

$$\sigma_2 = \frac{(x_i - \bar{x})^2}{10}$$

$\bar{x} = 66$

Second

$\sigma_1$ & $\sigma_2$

$$\boxed{r^2 = \sigma_1}$$

5% & 1% level of significance

$\frac{5}{100}$ $\frac{1}{100}$

0.05 0.01

tcritical d.of → same formula

above question

$$d.o.f = \frac{\left(\frac{(V_1)^2}{n_1} + \frac{(V_2)^2}{n_2}\right)^2}{\frac{\left(\frac{V_1^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{V_2^2}{n_2}\right)^2}{n_2 - 1}} \simeq \text{approx}$$

table critical 0.05 0.01

0.05

t critical = A

t value = B

B < A

Accept

B > A

Reject

0.01

t critical = C

t value = D

D < C Accept

D > C Reject

#Q. Two independent samples of sizes 7 and 9 have the following values:

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Sample A: | 10 | 12 | 10 | 13 | 14 | 11 | 10 | | |
| Sample B: | 10 | 13 | 15 | 12 | 10 | 14 | 11 | 12 | 11 |

Test whether the difference between the mean is significant..

$H_0$:- Mean diffⁿ is significant

$H_1$:- Mean diffⁿ is zero.

$n_1 = 7$

$$\bar{X}_1 = \frac{10+12+10+13+14+11+10}{7} = 11.42$$

$n_2 = 9$

$$\bar{X}_2 = \frac{10+13+15+12+10+14+11+13+11}{9} = 12$$

$$\text{T value} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{V_1}{n_1} + \frac{V_2}{n_2}}}$$

$$\boxed{\Sigma(x_i - \bar{x})^2}$$

$V_1 \to S.D$

$$\boxed{S.D_1 = \frac{\Sigma(\bar{x}_1 - x_i)^2}{n_1 - 1}}$$

$$SD_2 = \frac{\Sigma(\bar{x}_2 - x_i)^2}{n_2 - 1}$$

$$\boxed{S \cdot D_1^2 = V_1}$$

$$\boxed{S \cdot D_2^2 = V_2}$$

P value

$$d.o.f = \frac{\left(\frac{(V_1)^2}{n_1} + \frac{(V_2)^2}{n_2}\right)^2}{\frac{\left(\frac{V_1^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{V_2^2}{n_2}\right)^2}{n_2 - 1}}$$

d.o.f

0.05

$t > t_{critical}$

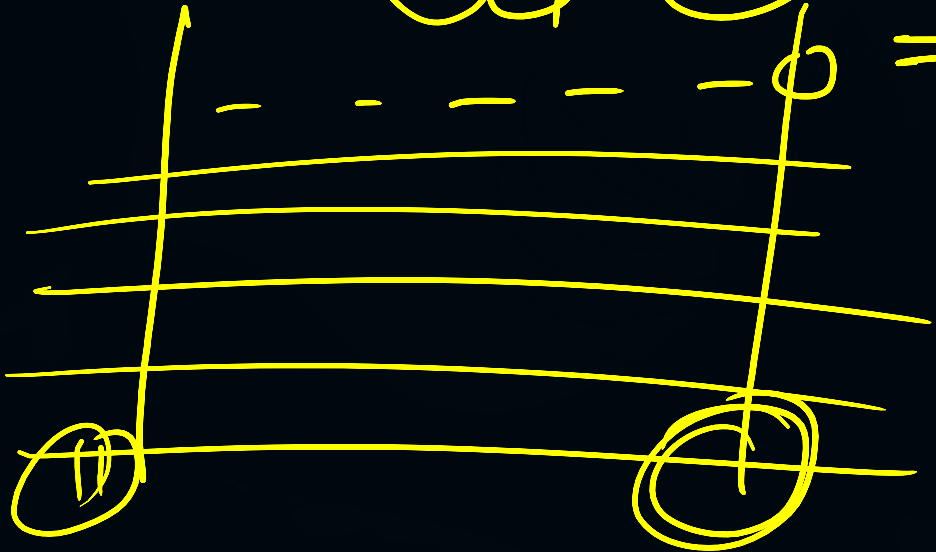
T-test

$t < t_{critical}$ ✗

$t_{critical}$

$\Rightarrow$ d.f (n) Integer

table



## Topic : Probability and Statistics

$H_0$ :- equally effi
$H_1$ :- Not equally effi

#Q. The average no. of articles produced by two machines per day are 200 and 250 with standard deviations 20and 25 respectively on the basis of records of 25 days production. Can you regard both the machines equally efficient at 5% level of significance?

$\bar{x}_1 = 200$ $\bar{x}_2 = 250$

$\sigma_1 = 20$ $\sigma_2 = 25$

$v_1^2 = 20$ $v_2^2 = 25$

$n_1 = 25$ $n_2 = 25$

$\underline{\underline{0.05}}$ ✓

$\underline{\underline{d.o.f = 48}}$

$$T\text{value} = \frac{200 - 250}{\dfrac{24 \times 20 + 24 \times 25}{25 + 25 - 2} \times \sqrt{\dfrac{1}{25} + \dfrac{1}{25}}}$$

$$= \frac{50 \times 48 \times 5}{1080 \times \sqrt{2}}$$

$$= 7.85 =$$

at a significance level

$t_{critical}$ 0.05

for dof = 48

$t_{critical}$

$t < t_{critical}$

Accept

$t > t_{critical}$

Reject

T - test

0.05

0.01

$t = \dfrac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$

$n_1$ & $n_2$

$s_1$ & $s_2$

formula

$n_1 = n_2$

$s_1$ & $s_2$ diff

formula

t critical

# THANK - YOU

Topics to be Covered