T- Test

Start at 9:03

⇒ • Recap of Z Test
• T Test 1 Sample
• T Test 2 Sample
• Paired T Test

⇒ Hypothesis Testing

1. Define H_0, H_a
2. Distribution based on H_0
3. We choose which pair test to use
4. P value
5. Give Results → α [0.05]

A french cake shop claims that the average number of pastries they can produce in a day exceeds 500. The average number of pastries produced per day over a 70 day period was found to be 530. Assume that the population standard deviation for the pastries produced per day is 125.

Test the claim using a z-test with the critical z-value = 1.64 at the alpha (significance level) = 0.05, and state your interpretation.

1) $H_0 \Rightarrow \mu = 500$

$H_a \Rightarrow \mu > 500$

125 ?

$\frac{125}{\sqrt{70}}$?

2) Define a distribution →

Normal →

$\mu = 500$

$\sigma = \frac{125}{\sqrt{70}}$

as we are considering a 70 day sample mean

3) Right tailed

4) Pvalue

$$Z \text{ Score} = \frac{530 - 500}{125/\sqrt{70}} = 2.02$$

$$\text{p-value} = 1 - \text{norm.cdf}(z)$$
$$= 0.022$$

5) Reject H_0 as p-value $< \alpha = 0.05$

Q what is critical z value for $\alpha = 0.05$??

$$\text{norm.ppf}(1 - 0.05) = \text{ppf}(0.95)$$
$$= 1.64$$

as it is right tailed

Q Left tailed test $\longrightarrow$

$$\text{ppf}(\alpha)$$

Q Two tailed test $\longrightarrow$

$ppf(1-\frac{\alpha}{2})$ $ppf(\alpha/2)$

Z Score → either be $\geq ppf(1-\alpha/2)$
in order to reject H_0 or $< ppf(\alpha/2)$

⇒ Issues with Z Test →

Q A Pharma Company is developing a cognitive enhancement drug. ?

→ Superpill → increase people's IQ

⇒ So people A → give the medicine
So people B → give placebo

⇒ A → [110, 95, 98, 102, . . .] → 103.5 ✓
B → [98, 99, . . . 99, 93] → 100.5 ✓

⇒ 1) $H_0 \to \mu_A = \mu_B$
$H_a \to \mu_A > \mu_B$ $\mu_A \neq \mu_B$

2) Dist^n →

Test Statistics $\Rightarrow$ $\rightarrow$ $\dfrac{X - \mu}{S/\sqrt{n}}$ } $\leftarrow$ T Statistic

S $\rightarrow$ Sample Std dev

$\underbrace{\hphantom{xxxxx}}$ T Test

---

① One Sample $\rightarrow$

group A $\rightarrow$ avg = 103.5 {given medicine}

population avg IQ = 100 = μ_{pop}

1) $H_0 \Rightarrow \mu_A = 100$

$H_a \Rightarrow \mu_A > 100$

4) T-Statistic = $\dfrac{X - \mu}{S/\sqrt{n}}$

t_test_1samp $\rightarrow$ p-value = 0.17

5) fail to reject H_0.

Note $\rightarrow$ If n become large $\rightarrow$ T-Stats becomes similar to Z Score

⟹ Null hypothesis of any test. (one sample)

$H_0 \Rightarrow \mu = 500$

What is the formula for the t-statistic in a one-sample t-test?

$$t\text{-stats} = \frac{X - \mu_{pop}}{\underbrace{S/\sqrt{n}}_{\text{Sample Std Error.}}}$$

$\frac{\sigma}{\sqrt{n}}$ = population std error

Two sample T Test →

⟹ S_1 → gets the medicine

S_2 → gets the placebo

⟹ we will check if both groups are similar or not.

Q 1) $H_0 \Rightarrow \mu_1 = \mu_2$

$H_a \Rightarrow$

A) $\mu_1 > \mu_2$

$\downarrow$

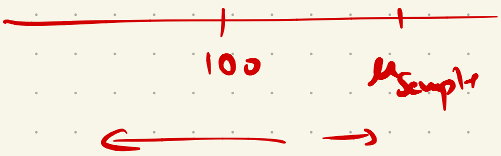
B) $\mu_1 \neq \mu_2$

$\downarrow$

Two tailed test

C) $\mu_1 < \mu_2$

$\downarrow$



$\Rightarrow \mu_2$ is right of μ_1

$\Rightarrow \mu_1$ is right of μ_2

3)

$\Rightarrow \quad \mu_1 \neq \mu_2 \Rightarrow$ test say H_a is true

$\Rightarrow \quad \mu_1 > \mu_2 \Rightarrow$ test says unable to reject or treat μ_1 same μ_2.

$\Rightarrow \quad \mu_1 \neq \mu_2$

1) ttest_ind (set1, set2)

2) $\mu_1 > \mu_2$

ttest_ind (set1, set2, alt="greater")

③ $\mu_1 < \mu_2$

t_test_ind (Set1, Set2, alt="less")

⇒ Why is this test called independent ?

Both Samples are independent of each other.

⇒ • Compare coins of BJP is North vs South India

• Compare score of Sachin in 1^{st} vs 2^{nd} inning

• Is VK better in chasing or in setting up total

⇒ Does Sachin play better in 1^{st} inning

$H_0 \rightarrow \mu_1 = \mu_2$

$H_a \rightarrow \mu_1 > \mu_2$

→ Are both innings equal

↓

$\mu_1 \neq \mu_2$

⇒ Drug recovery time

2 drugs A → 100 people → [10, 15, 9, 3, . . .]

B → 50 people → [. . . . .]

⇒ Aerofit
· Sachin
· School

} → always compared just 2 groups :

T Test → works only for 2 groups.

⇒ · used 2 groups only
· used independent groups

⇒ Drug Example →

we have data for before and after.

		Before medi:	After med
⇒	S_1	100	102
	S_2	101	103
	S_3	⋮	⋮
	⋮	⋮	⋮
	S_{100}		

as two groups are not independent → Can't use ttest_ind.

$\Rightarrow$ 1000 | 1002
1001 | 1003
1002 | 1004
1002 | 1005

100 1000 - 1010 $\rightarrow$

100 1002 - 1012 $\rightarrow$

Ind $\rightarrow$ Compare the two groups

Rel $\rightarrow$ $(\mu_1 - \mu_2)$ $\Rightarrow$ dist^n of this $\rightarrow$ $\begin{pmatrix} 2 \\ 2 \\ 2 \\ 2 \\ 2 \\ 2 \\ 2 \\ 2 \end{pmatrix}$ ✓

$\bar{T}$

✓ dist^n against 0

$\Rightarrow$ $\begin{pmatrix} 2 \\ -2 \\ -3 \\ 2 \\ 7 \\ -5 \end{pmatrix}$

$\bar{T}$

Q Does problem-Solving session helps student?

⇒ Before/After ⟶ Rel (Paired T Test)

Independent ⟶ (ind 2 sample T test)

⇒ If σ_{pop} is known ⟶ Z Test

σ_{pop} is unknown ⟶

⇒ One Sample (population mean is known & you have group) ⟶ 1 Samp T Test

⇒ 2 Samples
- Ind ⟶ ttest_ind
- Rel ⟶ ~~tt~~ttest_rel
  - Before/After

⇒ num vs Cat → 2 → T test
→ >2 → Anova

Chi-Square

Correlation

Doubts

⇒ Basic Hypothesis testing →

H_0 is true → what are chances of extreme data.

↓

p_value

p_value < 0.05 ← reject our H_0

⇒ Court room → Everyone is innocent → H_0

→ will have a knife → 60 %

→ knife will have blood → 30 %

→ Murder place have ~~your~~ fingerprint → 10 %

→ Victim blood on knife → 5 %

→ CCTV footage → 1 %

⇒ we are only looking at data from (H_0 being true perspective)

Jud Test Result \ Reality	Innoc	Crim
Innoc	True -ve	False -ve
Criminal	False +ve	True +ve

100 ⇒ less than 5% chance of being innocent

⇒ $\underline{\alpha}$ is represent of max. false +ve allowed.

⇒ β → max. allowed False -ve

$1-\beta$ → power → Representation of lesser error → False -ve

	Inno	Criminal
Inno	97	0
Crimi	0	3

95 → Inn

5 → Crimi

90	5	95
5	0	5

⇒

$\alpha \rightarrow 5\%$ p-value = 3%

$\beta \rightarrow 30\%$ power = 70%

Do not use or rely on this test.

Yes → → reject H_0 →

accept H_0 if p-value $> \alpha$

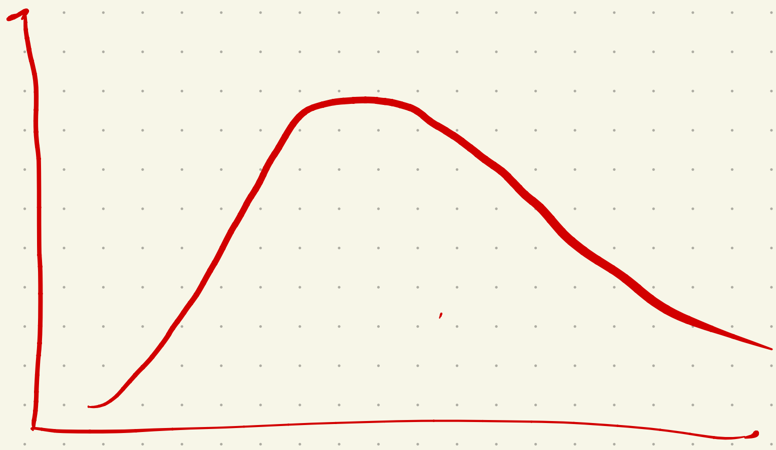
⇒ There is 5% chance that this person is not innocent.

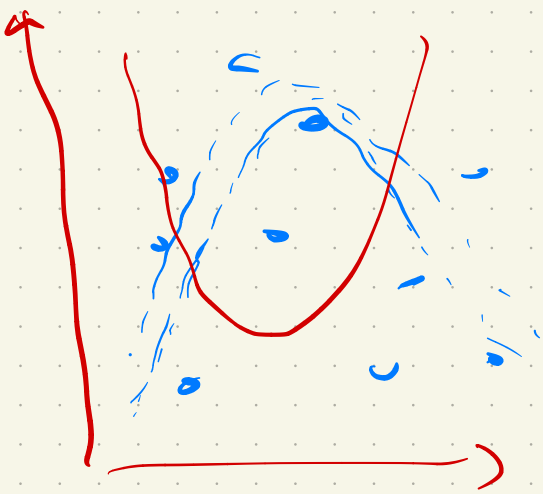
⇒ Even if I claim them to be innocent

30% of these people criminal in reality.

⇒ α is low → power is high | (α is high / power is low) $\times\alpha$

⇒ Data follow → T Distribution





T Test →

Z Test

Collab Link : https://colab.research.google.com/drive/1ExXwjUfqpDRUh6Ed3MBpA7DssrHydVlW?usp=sharing