Agenda

1. Point Estimate ✓
2. Range of C.I ✓
3. Chi Square distribution ✓
4. F distribution. → F test ✓
5. [ANOVA] → Assignment ⇒ more than 2 groups

⇓

F Test

① Point Estimate

Dfn: The value of any statistic that estimates the value of a parameter is called Point Estimate

Point Estimate ↓

sample ← $\boxed{\bar{x}}$ → $\boxed{\mu}$ ← parameter → population

$\bar{x} = 2.95$ $\quad\mu = 3.00$

$\bar{x} = 3.5$ $\quad\mu = 3.00$ $\quad 3.5 - 0.5 = 3.0$

We rarely know if our point Estimate is correct because it is an estimation of the actual value

We construct C.I to help estimate ($\bar{x}$) what the actual value of unknown population mean is.

Point Estimate ± Margin of Error

Lower Range C.I = Point Estimate − Margin of Error

Higher Range C.I = Point Estimate + Margin of Error

T test lgega

① On the Verbal section of the CAT exam, a sample of 25 test takers has a mean of 520. with a standard deviation of 80. Construct a 95% C.I about the mean?

99%

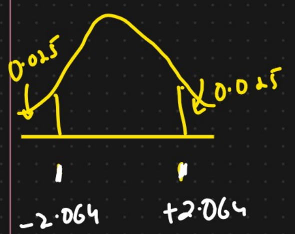
$\Downarrow$

C.I

Ans) $\bar{x} = 520$ $n = 25$ $\boxed{s = 80}$ $C.I = 0.95$ $\alpha = 0.05$

$C.I =$ Point Estimate $\pm$ Margin of Error

[425 - 564]



$$= \bar{x} \pm \boxed{t_{\alpha/2}} \left(\frac{s}{\sqrt{n}} \right) \Rightarrow \text{Standard Error}$$

$\boxed{0.05}$

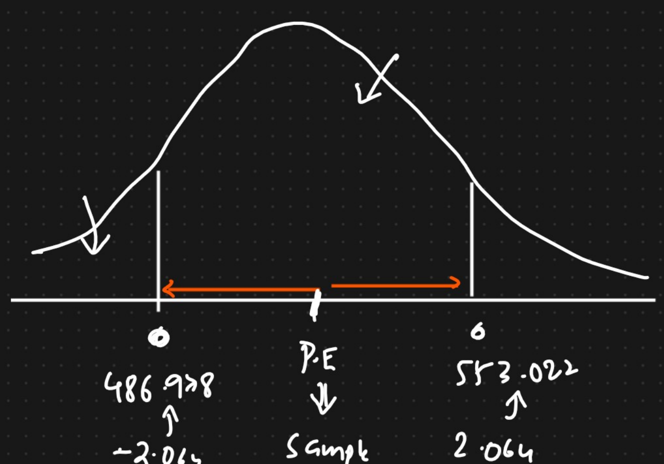
$$= 520 \pm 2.064 \left[\frac{80}{\sqrt{25}} \right]$$

degree of freedom $= n - 1 = 24$

$= 24$ //.

$$\text{Lower C.I} = 520 - 2.064 \left[\frac{80}{5} \right] = 486.978 \text{ //.}$$

$$\text{Higher C.I} = 520 + 2.064 \left[\frac{80}{5} \right] = 553.022 \text{ //}$$



mean

② CHI SQUARE TEST

whether the sample information we get is the goodness of fit for the population proportion

The Chi Square Test for Goodness of fit tests claims about population proportion.

[categorical variables]

It is a non parametric test that is performed on categorical data

[ordinal, nominal data].

Eg: There is a population of Male who likes different color of Bike

	Theory	Sample
Yellow Bike	1/3	22
Orange Bike	1/3	17
Red Bike	1/3	59

⇒ Goodness of fit.

⇒ Observed categorical distribution

⇓

Theoretical categorical distribution

①

In 2010 Census of the city, the weight of the individuals in a small city were found to be the following

<50kg	50 - 75	>75
20%	30%	50%

In 2020, weight of n=500 individuals were sampled. Below are the results

<50	50-75	>75
140	160	200

Using $\alpha=0.05$, would you conclude the population difference of weights has changed in last 10 years?

Ans) In 2010

Expected

< 50kg	50-75	> 75
$\boxed{20\%}$	30%	50%
⇓ 0.2	⇓ 0.3	⇓ 0.5

In 2020

n = 500

Observed

< 50	50-75	> 75
140	160	200

In 2010

Expected

< 50kg	50-75	> 75
500 × 0.2 = 100	500 × 0.3 = 150	500 × 0.5 = 250.

① Null Hypothesis H_0 : The data meets the expectation

Alternate hypo H_1 : The data does not meet the expectation.

② $\alpha = 0.05$ C.I = 95%

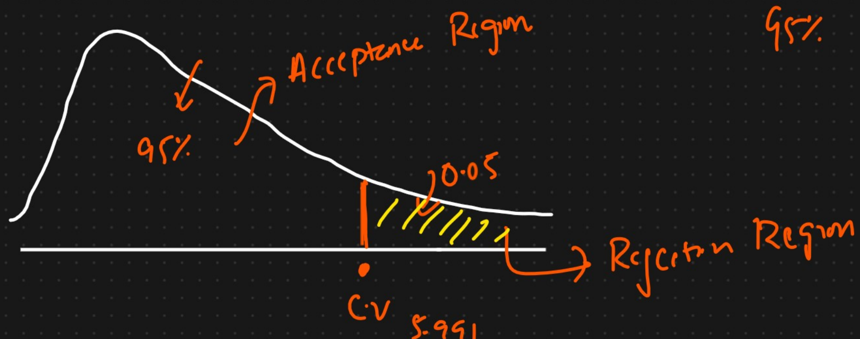
③ Degree of freedom

No. of categories

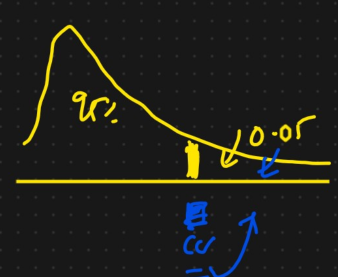
$$dof = K - 1 = 3 - 1 = 2$$

④ Decision Boundary

Now we need to use chi square table, dof and alpha value to find the c.v (critical value)



95%



Chi square Test $\chi^2 > 5.991$ { Reject the Null Hypothesis }.

⑤ Calculate Chi Square Test Statistics

observed

$$\chi^2 = \sum \frac{(O-E)^2}{E}$$

expected

$$= \frac{(140-100)^2}{100} + \frac{(160-150)^2}{150} + \frac{(200-250)^2}{250}.$$

$$\chi^2 = 26.67$$

$\chi^2 > 5.991$ { Reject the Null Hypothesis }.

③ F-distribution

The F-distribution with d_1 and d_2 degree of freedom is the distribution of

$$X = \frac{S_1/d_1}{S_2/d_2}$$

$S_1 \rightarrow$ Independent random variables
$S_2 \rightarrow$ Independent random variable
} Chi square distribution

$d_1 \rightarrow$ Degree of freedom (S_1)
$d_2 \rightarrow$ Degree of freedom (S_2)

F test $\rightarrow$ Variance Ratio Test { Comparing the variance between 2 groups }.

F Test [Variance Ratio Test].

① The following data shows the no. of bulbs produced daily for some days by 2 workers A and B

A	B
40	39
30	38
38	41
41	33
38	32
35	39
	40
	34

Can we consider based on the data Worker B is more stable or not and efficient

$\alpha = 0.05$ 95% C.I.

Ans) Null Hypothesis $H_0 : \sigma_1^2 = \sigma_2^2$

Alternate Hypothesis $H_1 : \sigma_1^2 \neq \sigma_2^2$

A

X_1	$\bar{x}$	$(x_i - \bar{x})^2$
40	37	9
30	37	49
38	37	1
41	37	16
38	37	1
35	37	4

$\bar{X_1} = 37$ $\Sigma(x_i - \bar{x})^2 = 80$

B

X_2	$\bar{X_2}$	$(X_2 - \bar{x})^2$
39	37	4
38	37	1
41	37	16
33	37	16
32	37	25
39	37	4
40	37	9
34	37	9

$\bar{X_2} = 37$ 84.

$$S_1^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1} = \frac{80}{5} = 16.$$

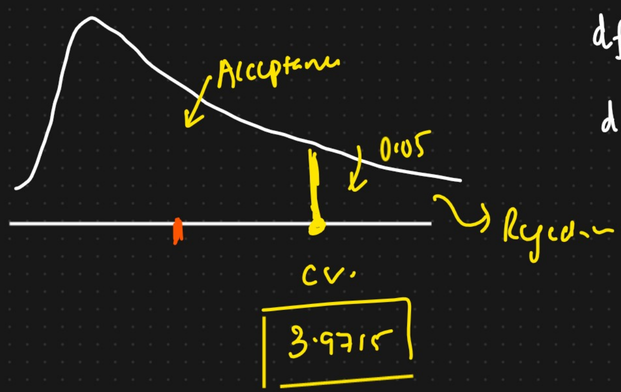
$$S_2^2 = \frac{84}{7} = 12$$

* Variance Ratio. [F-test]

$$\boxed{F = \frac{S_1^2}{S_2^2} = \frac{16}{12} = 1.33}$$

* Decision Rule [F distribution]

to find c.v (critical value), we will use f-table, dof and alpha



$df_1 = 6-1 = 5$

$df_2 = 8-1 = 7$

$\alpha = 0.05$

F-test > 3.9715 {Reject H_0}.

Is $1.33 > 3.9715 \Rightarrow$ False

We fail to Reject the Null Hypothesis

Worker A $\approx$ Worker B