1. z-test
2. t-test

Hyphothesis

$$\boxed{z\_Score = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}}$$

$$\boxed{t\_test = \frac{\bar{x} - \mu}{s/\sqrt{n}}}$$

** 1. $n \leq 30$
2. there won't be any Population Std dev.

type of t-test

1. One sample t-test
2. two sample t-test (independent pop)
3. Paired t-test (same Population)

** Practical implementation

Question → test

1. z-test (z-score)
2. t-test

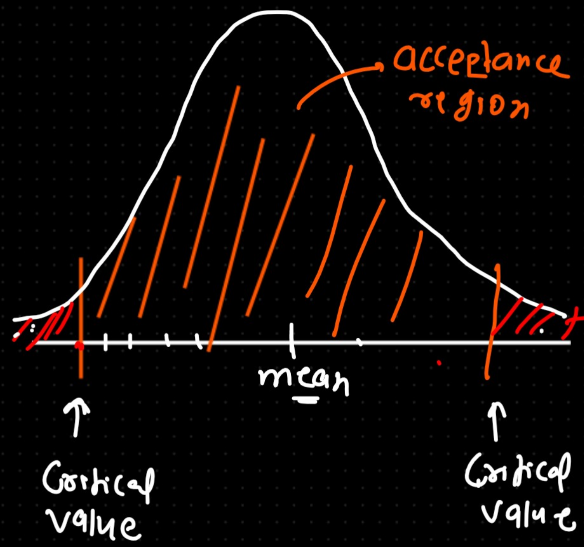
Null hypothesis / reject Null hypo

$CI = 95\%$

$\alpha = 5\%$

two-tail test

one-tail test (left hand or right hand side)



rejection region

Data
- Numeric
  - Conti
  - Discreate
- Categorical Data → Chi-square test (feature selection, Corr b/w two cat. feature)
  - Nominal
  - Ordinal

Chi-Squre ⇒ Independent | Dependent (Goodness of fit)

1. Gender ⇒ male | female
2. color ⇒ Red | blue | Green | yellow

· Expected | observed

- Expected ⇒ 11 female 20 male
- observed ⇒ 120 F 150 m

Categorical

	Cat	Dog
Men	207	282
Women	231	242

(Number must be large enough)

Independent | not independent (Goodness of fit)

hyphothesis H_0 (Null hyphothesis) ⇒ 1. Gender and Prefrence for Cat and dog are independent.

2. Gender and prefrence for Cat and dog are not independent.

Observed value

	Cat	Dog	
Men	207	282	$\boxed{489}$
Women	231	242	$\boxed{473}$
	$\boxed{438}$	524	962

Excepted value ⇒

	Cat	Dog
men	$\frac{489 \times 438}{962}$	$\frac{489 \times 524}{962}$
Women	$\frac{473 \times 438}{962}$	$\frac{473 \times 524}{962}$

	Cat	Dog	
men	222.64	266.36	
women	215.36	257.64	
			962

$X \rightleftarrows Y$

goodnes of fit wrt (x, y)

Coin

↓

$\boxed{H/T}$

↓

3 → H, H, H

$\boxed{\text{Exception}}$

$\boxed{\text{Observation}}$

H, T, H
T, T, T
H, H, T

	Cat	Dog	
Men	207	282	489
Women	231	242	473
	438	524	962

	Cat	Dog	
Men	222.64	266.36	489
Women	215.36	257.64	473
	438	524	962

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

	Cat	Dog
Men	$\frac{(207-222.64)^2}{222.64}$	$\frac{(282-266.36)^2}{266.36}$
Woman	$\frac{(231-215.36)^2}{215.36}$	$\frac{(242-257.64)^2}{257.64}$

	Cat	Dog
Men	1.099	0.918
Woman	1.136	0.949

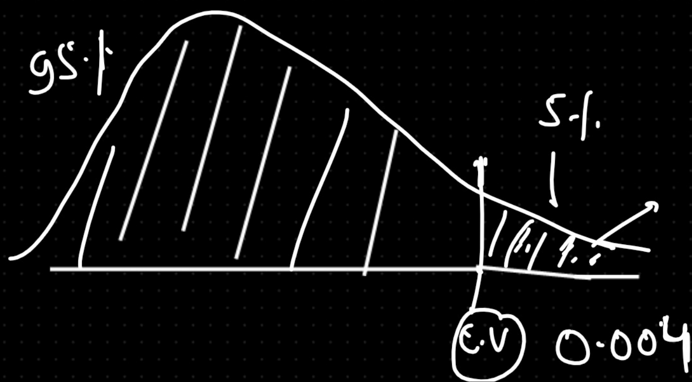
$\Downarrow$

$$\boxed{\text{chi-square value } (X^2) \Rightarrow \sum \frac{(O-E)^2}{E}}$$

$$1.099 + 0.918 + 1.136 + 0.949 = \boxed{4.102}$$

(90% | 95% | 99%)

$\Downarrow$



4.102

Reject the Null hypothesis

D.F. $\Rightarrow$ $N \Rightarrow$ $N-1$

$= (2-1) \times (2-1)$

$= 1 \times 1$

$= \boxed{DF = 1}$

$\boxed{0.004}$

Question there is a Population of male who like different color of bikes.

	theory	Sample
Yellow bike	1/3	22
orange bike	1/3	17
Red bike	1/3	50

⇒ Goodness of fit

In 2010 census of the city, the weight of the individual in a small city were found to be following

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

in 2020, weight of $n=500$ individuals were sampled. Below are the result.

<50	50-75	>75
140	160	200

Using $\alpha = 0.05$, CI = 95%, would you conclude the population difference of weight has changed in 10 years?

2010

Census (Population) Excepted

<50	50-75	>75
20%	30%	50%
0.2	0.3	0.5

2020

$\boxed{n = 500}$

<50	50-75	>75
140	160	200

In 2010

Excepted

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

(1)

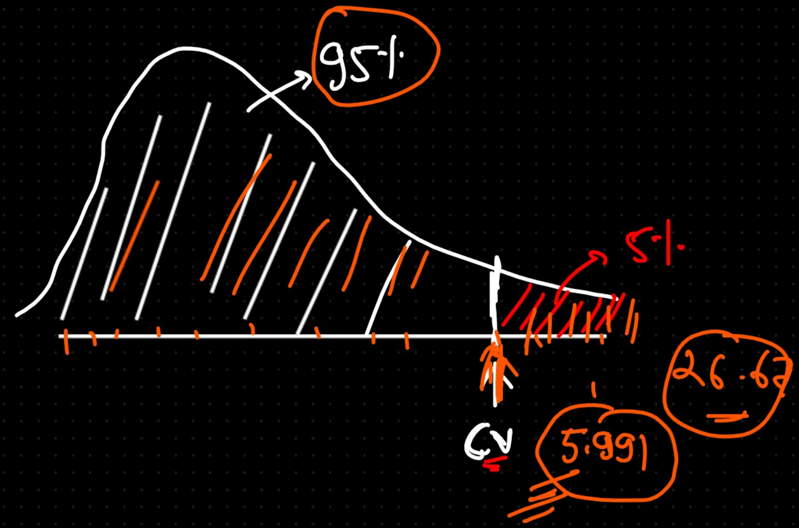
Null hypothesis H_0 ⇒ the data meets the exceptation

Alternate Hypho h_a ⇒ the data does not meet the exception

(2) $\alpha = 0.05$ (5%) $CI = 95\%$

(3) DOF $= K-1 = 3-1 = 2$

(4) Decision boundry ⇒



5. Chi-square-stats

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

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

$$= \frac{(40)^2}{100} + \frac{(10)^2}{150} + \frac{(-50)^2}{250}$$

$$\chi^2 = 16 + 0.667 + 10$$

$$\boxed{\chi^2 = 26.67}$$

$$\boxed{26.67 > 5.991}$$ ⟹ Reject the Null hypothesis

Anova test ⇒ (1) one way anova
(2) two way anova
(3) n-way anova

(Analysis of Variance)

Aim of anova

Two or more than two (3, 4, 5 ... n) mean we are going to compare

ex. 3 sample μ_1
μ_2
μ_3

$$\boxed{\begin{array}{l} H_0 \Rightarrow \mu_1 = \mu_2 = \mu_3 \\ H_a \Rightarrow \mu_1 \neq \mu_2 \neq \mu_3 \end{array}}$$

Population → Sample

1. sample must be independent
2. normally Dist.
3. equal variance
4. Random sample

Group1	Group2	Group 3
19	14	20
25	27	22
32	39	33
58	51	50
59	66	52
94	70	55

① With in a group
② B/w a group

Source of variation	Df
Between group	K-1
With in group	N-K

MSS (mean of sum of squre)

$$mssB = \frac{SSB}{k-1}$$

$$mssw = \frac{SSW}{N-k}$$

$$\boxed{F\text{-test}} = \frac{mssB}{mssw} \Rightarrow \text{value}$$

F_table

5%, 1%, 10%

Critical value

hyphothesis

One-way Anova

One-way anova

Lecture ⟷	Question-Answer ⟷	Libary
X_1	X_2	X_3
4	9	2
5	10	4
1	9	2
2	6	2

independent analysis

two way anova

[Teching method]

(Live lecture mode) (recorded lecture mode)

- Anxiety level

- low - Average - high

2way

		(teching method)	
		Live lecture method	recorded lecture
Anxiety level	(low)	-	-
	(Avg)	-	
	(high)		

Dependent Aanlysis

Assignment ⇒ ① one way anova and two way anova and implement it in the python.

(pen and paper)

② Detailed D/f b/w t-test and anova test.

t-test:
- one sample
- two sample
- Paired test

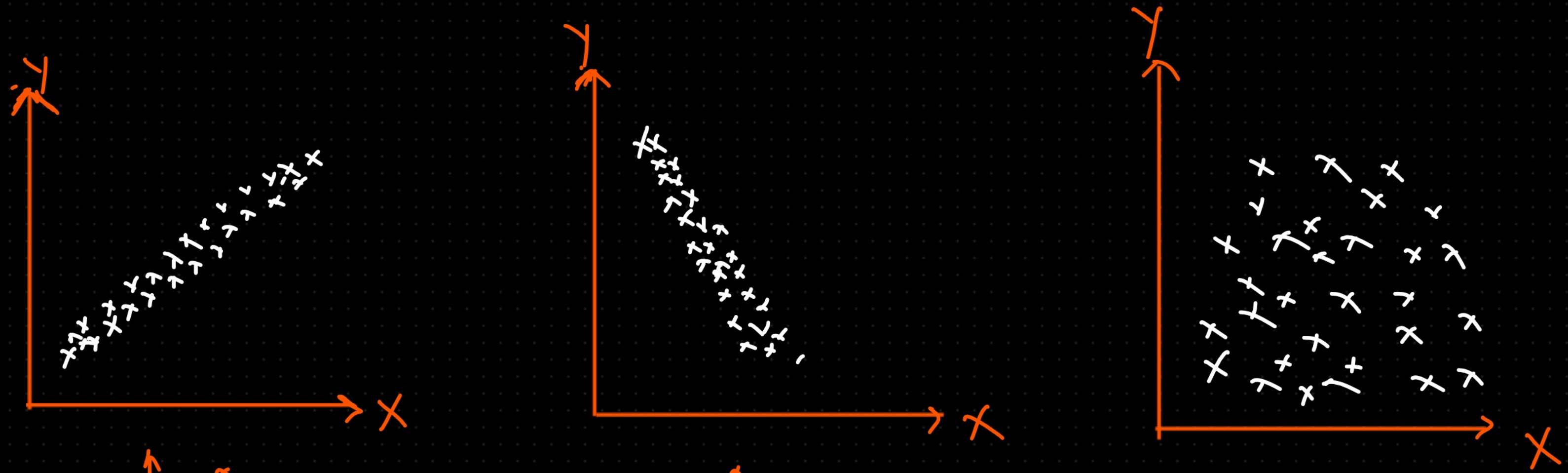
anova:
- One way
- two way
- n-way

Practical of the stats

AB testing | Possion Dist | Corr | Cov

Correlation ⇒ Relationship b/w two or more than two variable.

1. Positive
2. Negative
3. No corr b/w a variable



$X\uparrow \; Y\uparrow$ $\qquad$ $X\uparrow \; Y\downarrow$ $\qquad$ $X \neq Y$

1. Pearson corr
2. Spearman Rank corr
3. Kendall's tau

$$\text{Person corr's} = \frac{\mathrm{cov}(x, y)}{\sigma_x \, \sigma_y}$$

$$\text{Spearman Rank corr's} = \frac{\mathrm{cov}(R(x), R(y))}{\sigma(R(x)) \times \sigma(R(y))}$$

$$[-1, +1]$$

- if +ve corr than it will be towards +1
- if -ve corr than it will be towards -1
- if no corr than the value will be zero.

Covariance ⇒ it also gives relation B/w two var. or more than two var.

$$Cov(X,Y) = \sum_{i=1}^{n} \frac{(x_i - \bar{x})(y - \bar{y})}{n-1}$$

$$Cov(X,X) = \sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{n-1} \Rightarrow \sum_{i=1}^{n} \frac{(x_i - \bar{x})(x_i - \bar{x})}{n-1}$$

$$[-\infty, +\infty]$$

Question Write Down 5 difference between covariance | correlation.

10 to 1 pm IST sunday

Saturday | ⇒ Python Implementation of stats , EDA ⇒ A/B testing | Possion | corr | cov

tuesday → Pandas , tuesday - Numpy , $\boxed{\text{tuesday}}$ - matplot | seaborn

10 pm ~~to~~ 1 pm

10 am to 1 pm (mor)   1st | 2st April