About Myself

Hi, I'm **Yash Pandey**, a passionate tech enthusiast From **Maharaja Agrasen Institute Of Technology (MAIT)** with a strong interest in **Artificial Intelligence (AI), Machine Learning (ML), and Data Science**. I have a deep curiosity for understanding how data-driven technologies shape the world, and I constantly strive to enhance my technical skills to stay ahead in this ever-evolving field.

My journey in the tech world has been driven by a problem-solving mindset and a desire to build impactful solutions. I enjoy working with complex datasets, developing predictive models, and optimizing machine learning algorithms. Apart from AI/ML, I have experience in **software development, competitive programming, and full-stack web development**. I believe in continuous learning and love sharing my knowledge with the tech community.

My Skills

- **Machine Learning & AI**: Supervised & Unsupervised Learning, Deep Learning, Model Optimization
- **Web Development**: React.js, Node.js, Express.js, Database Management
- **Problem-Solving & DSA**: DSA In Java With 600+ LeetCode Problems .

About These Notes

This document contains **comprehensive statistics notes** curated for the AIML learning path. It covers key statistical concepts essential for data analysis, machine learning, and AI applications. These notes are structured to be beginner-friendly yet detailed enough for deeper insights. They can serve as a reference for anyone looking to strengthen their foundation for AI/ML.

Connect With Me

🖇️ **GitHub**: github.com/YashPandey1405
🖇️ **LinkedIn**: linkedin.com/in/yashpandey29
🖇️ **LeetCode**: leetcode.com/u/pandeyyash041

For **Statistics Code Implementation using Python**, check out my repository:
🖇️ GitHub - Statistics Using Python 🚀

"Statistics for AIML"

→ Statistics Is an Mathematical Science Including methods of Collecting, Organizing and Analysis of data in such an Way That Meaningful Conclusion Can be drawn from The Avoidable data.

→ Data Is Facts / Peice of data That can be stored and measured.

Uses of Statistics
- Weather Forecast.
- Sports Analysis.
- Election Campaign.
- E-Commerce Platform.
- Medical Industry.

→ Types of Statistics

① Descriptive Statistics [Describe]

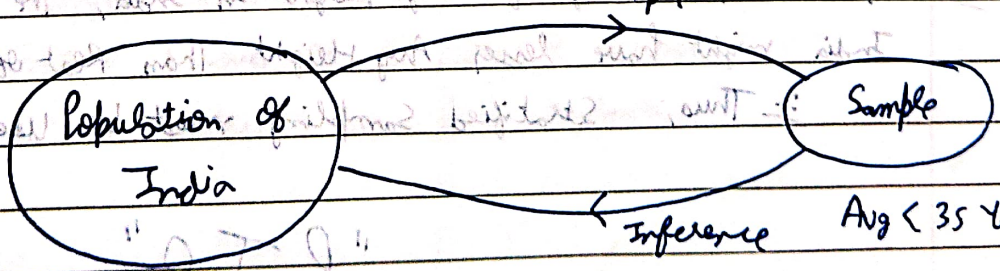
→ It Consist of Organizing and summarize The data/Population.

Eg:- Virat Kohli Average In ODI's.

② Inferential Statistics [Inference]

→ It Consist of Using The data that has been measured To form The Conclusions about The Population.

Eg:-



PAGE NO.:

Types of Descriptive Statistics
- Measure of Central Tendency (Mean, Median, Mode)
- Measure of Symmetricity (Skewness, Kurtosis)
- Measure of Dispersion: (St. Deviation, Variance)

Note → Understand Descriptive Vs Inferential by below :-

① Descriptive : What Is The Avg. Height of The Class?

② Inferential : Are The Avg. Height of Students To Save What we Expect from The Entire School.

→ Types of Sampling

1. Simple Random Sampling

→ Every Member of The Population (N) has an Equal Chance of Being Selected of your Sample.

⊙ Eg :- For Exit Poll, In an 12-Member Group, Each Person has Selection Probability of 1/12.

2. Stratified Sampling [Strata ≡ layers / Groups]

→ In This Sampling, different distinct Categories are There.

→ An Simple Sample would be chosen from Each Strata or layer.

⊙ Eg :- In Survey of Avg. Height In India, The People of North-East India might have lesser Avg Height Than Rest of India.

∴ Thus, Stratified Sampling must be Used.

PAGE NO.:

③ Clustered Sampling

→ The division of the Sample Into Groups and Clusters where some of the Clusters are been randomly selected.

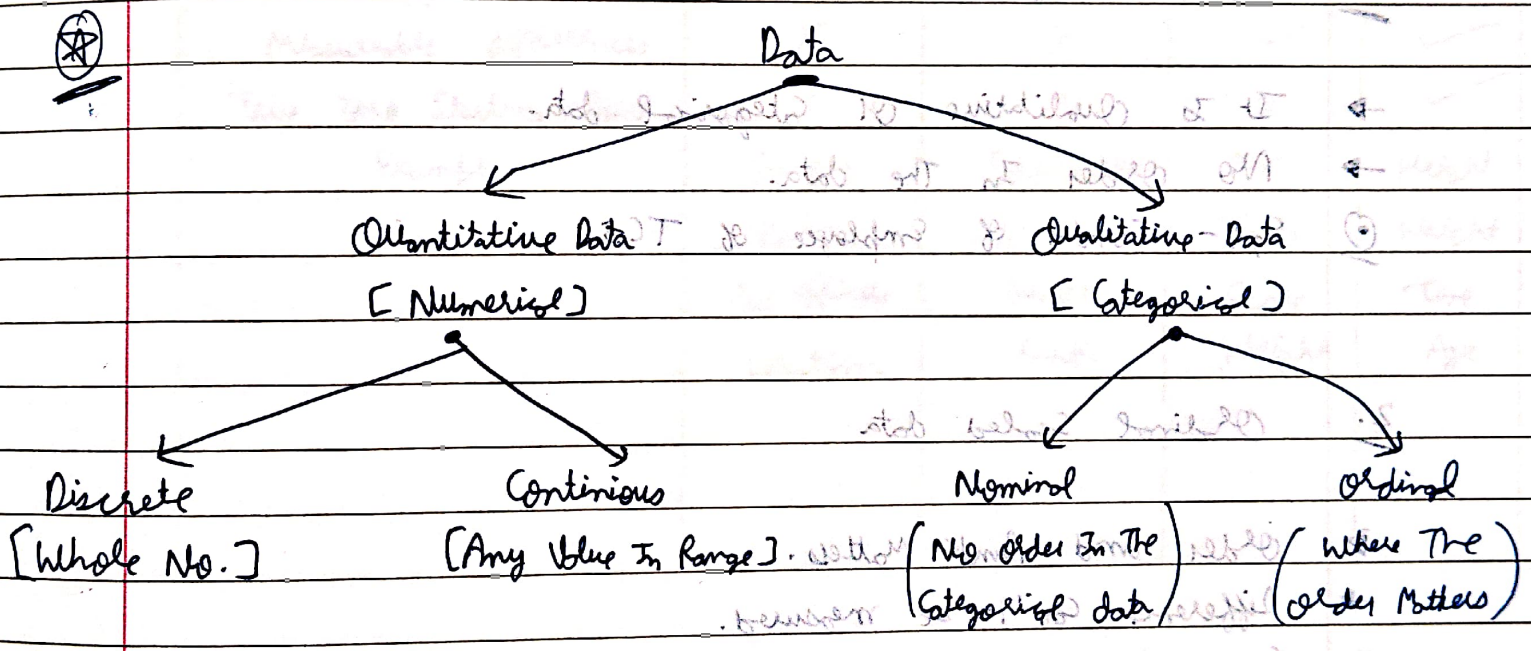
→ Thus, all Individual in the chosen Cluster are selected In the Sample.

④ Symmetric Symmetry

→ Every nTh Element will be selected.

⊙ Eg :- Odd Roll Numbers.
People Born In odd Years.

✱



① Discrete : N.o of Childrens In an House.
② Continious : Weight of Students of Class III.
③ Nominal : Blood Group of Students of School.
④ Ordinal : Grades Obtained by the Students.

Scales of Measurements

① Interval Scale data
② Ratio Scale data
} Quantitative Measures (Measurements are Numeric)

③ Nominal Scale data
④ Ordinal Scale data
} Qualitative data / Measures (It classified In Non-Numeric)

1. Nominal Scale data

→ It is Qualitative or Categorical data.
→ No order In The data.
⊙ Eg :- Gender of Employee of TCS.

2. Ordinal Scaled data

→ Order and Ranks Matters.
→ Difference can't be measured.
⊙ Eg :- Educational Qualification of Interns In HCL.

3. Interval Scaled data

→ The Rank and The order has an meaning.
→ The difference can be measured by Excluding Ratio.
→ It doesn't have 0 starting value.
⊙ Eg :- Height / Weight.

4. Ratio Scaled data.

→ The order and Rank has an meaning.
→ Differences and The Ratio are been measurable.
→ It does have an '0' starting blue.

⊙ Eg :- Person A = 48 kg
Person 2 = 24 kg
} A : B = 2 : 1

Data	Nominal	Ordinal	Interval	Ratio
Labelled	✓	✓	✓	✓
Meaningful order	×	✓	✓	✓
Measurable differences	×	×	✓	✓
True zero Starting Point	×	×	×	✓
Example	Gender	Satisfaction	IQ	Height
	Religion	Rating	Temp	Weight
	Post offices	Grade	Score	Time
	Location	Rank	Weight	Age

Descriptive Statistics

→ Summarization of data without Adding / subtracting anything at any specific Instance of the data.

→ 3 Types are
- Measure of Central Tendency.
- Measure of Dispersion.
- Measure of Symmetricity.

Measure of Central Tendency

→ Central Tendency represents the Centre Point of an dataset.

→ It's 3 Types are Mean, Median and Mode.

① Mean (Airthmatic mid Value of the data)

→ For Population, $\mu_p = \sum_{i=1}^{N} \frac{x_i}{N}$

→ For Sample, $\mu_s = \sum_{i=1}^{n} \frac{x_i}{n}$

Eg:- [1, 2, 3, 4, 5], mean $= \frac{1+2+3+4+5}{5} = \frac{15}{5} = 3$

② Median (Physical Mid-Value of data)

→ If, Odd i.e. [1, 2, 3, 4, 5], Then, Median = 3 (Mid Value)

→ If ~~Odd~~ Even i.e. [1, 2, 3, 4, 5, 6], Then, Median $= \frac{3+4}{2} = 3.5$

★ Let, List = [1, 2, 3, 4, 5, 10000] → Here, 10000 is an outlier as it is very different from the Rest of the data Given In List.

→ Outliers are the numbers which are much higher or lower than the other members. They are extreme values.

Note → Mean is affected by outliers, but not with median.

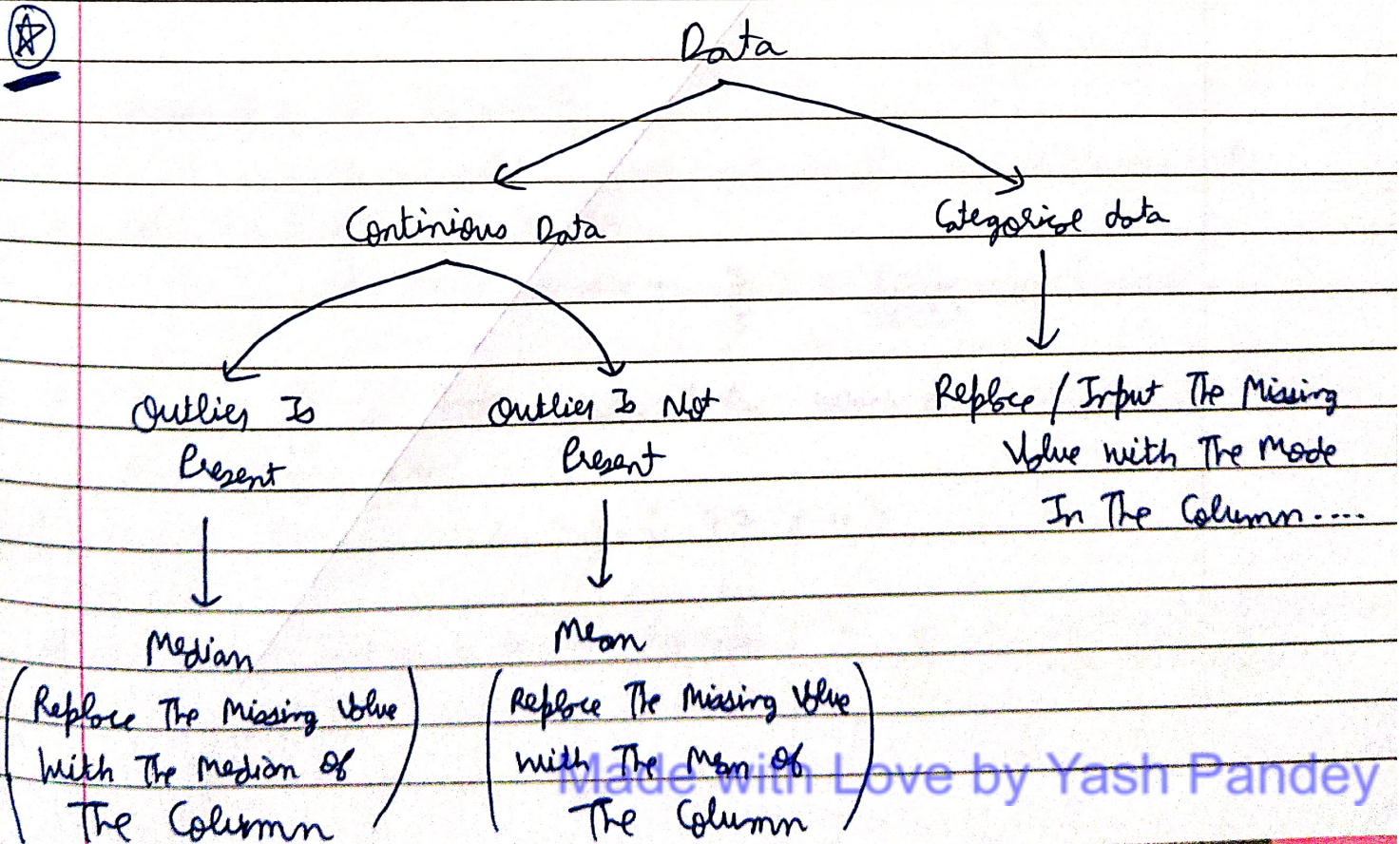
Case - I	Case - II
[1, 2, 3, 4, 5]	[1, 2, 3, 4, 99]
mean = 3	mean = 21.8
median = 3	median = 3

∴ Thus, outliers affect the mean but not median.

— JAI LOGIC

③ Mode (The Highest Frequency Element)

→ Let, [1, 2, 2, 3, 3, 3, 4] ⟶ mode = 3



PAGE NO.:

Implementation of Central Tendency Using Python

→ import "numpy as np" In The Program

(1) Mean ⟶ np.mean(list)

(2) Median ⟶ np.median(list)

(3) Mode ⟶ Mode Is Not Present In The Numpy, for Mode,
import Statistics # Inbuilt Library In Python
statistics.mode(list)

— JAI LOGIC

```python
# Measures Of Central Tendancy....
import numpy as np # type: ignore
import statistics

age=[15,75,29,86,45,58,37,41]

# Mean Calculation Using Numpy.....
print(f"Mean Of {age} Is : {np.mean(age)}")

# Median Calculation Using Numpy.....
print(f"Median Of {age} Is : {np.median(age)}")

# Mode Calculation Using Statistics.....
print(f"Mode Of {age} Is : {statistics.mode(age)}")
```

→ Measures of Dispersion

→ Let, $S_1 = 1, 2, 3, 4, 5$, $S_2 = 3, 3, 3, 3, 3$

Mean = 3 | Mean = 3

Median = 3 | Median = 3

→ Spread of data →
- Range
- Percentage and Percentile.
- Quartiles.
- Variance.
- Standard deviation.

① Range :- diff. b/w Max. Value & Min. Value

{ 1, 2, 3, 4, 5 } ⟶ Range = 4

{ 1, 2, 3, 4, 1000 } ⟶ Range = 999

* Outliers affects The Range.

② Percentage or Percentile

→ Let, S = { 1, 2, 3, 4, 5 }, Then,

% Numbers as odd = $\frac{3}{5}$ = 60%

→ Percentile Is an ~~Value~~ Value below which an certain Percentage of The observation lies.

Eg:- S = { 1, 2, 3, 4, 4, 6, 7, 7, 8, 10 }

→ Percentile Rank of 6 = $\frac{\#\text{Values below } 6}{\#\text{Values}} = \frac{5}{10} = 50^{\text{th}}$ Percentile.

PAGE NO.:

① Percentile of an Value = $\frac{\#\text{Values below the Value}}{\#\text{Values}} \times 100$

② Value = $\frac{\text{Percentile}}{100} \times (n+1)$ ← n = #Values

◉ Eg:- $S = \{1, 2, 2, 2, 3, 4, 4, 4, 5, 6\}$ (4 is the 8th Number)

Then, Value at 75th Percentile will be

Value = $\frac{\text{Percentile}}{100} \times (n+1) = \frac{75}{100} \times (10+1) = 11 \times 0.75$

$= 8.25^{Th}$ Number

∴ 75th Percentile = 8.25th Number of the Sample

= 8th Number of the Sample = 4 ①

8.25th ⟶ Take 8th Number

∴ Thus, 8.5th ⟶ Take Avg. of 8th & 9th Number

8.75th ⟶ Take 9th Number.

3. Quartiles

→ Quartiles are Values that divide a list of Numbers into Quarters.

① Put the Numbers in Order.

② Cut the Numbers into 4 Equal Parts.

③ The Quartiles will be at the Cut.

PAGE NO.:

① Eg:- {6, 8, 5, 5, 7, 3, 9}

order = {3, 5, 5, 6, 7, 8, 9}

Than, 3 | 5 | 6 | 7 | 8 | 9

Cut Quartile-1 (Q_1) | Cut Quartile-2 (Q_2) | Cut Quartile-3 (Q_3)

Imp. Trick for Quartiles

① When n = odd	② When n = Even
→ $Q_1 = \left(\frac{n+1}{4}\right)^{Th}$	→ $Q_1 = \frac{n}{4}^{Th}$
→ $Q_3 = \frac{3(n+1)}{4}^{Th}$	→ $Q_3 = \frac{3n}{4}^{Th}$
→ $Q_2 = \left(\frac{n+1}{2}\right)^{Th}$	→ $Q_2 = Avg\left(\frac{n}{2}^{Th}, \left(\frac{n}{2}+1\right)^{Th}\right)$

Q → Let, Value = {1, 2, 3, 3, 4, 4}, n = 6

$\therefore Q_1 = \frac{n}{4}^{Th} = \frac{6}{4}^{Th} = 1.5^{Th} = Avg(1^{Th}, 2^{Th}) = \frac{1+2}{2} = 1.5$

$Q_3 = \frac{3n}{4}^{Th} = \frac{3\times 6}{4}^{Th} = 4.5^{Th} = Avg(4^{Th}, 5^{Th}) = \frac{3+4}{2} = 3.5$

$Q_2 = Avg\left(\frac{6}{2}^{Th}, \frac{6}{2}+1^{Th}\right) = Avg(3^{Th}, 4^{Th}) = \frac{3+3}{2} = 3$

∴ Thus, $Q_1 = 1.5$, $Q_2 = 3$, $Q_3 = 3.5$.

PAGE NO.:

⇒ Five Point Summary

→ Q_0 (min) ⟶ 0% ⟶ 0^{Th} Percentile

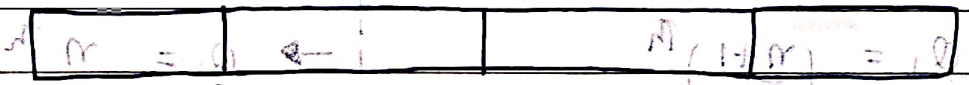
Q_1 ⟶ 25% ⟶ 25^{Th} Percentile

Q_2 ⟶ 50% ⟶ 50^{Th} Percentile.

Q_3 ⟶ 75% ⟶ 75^{Th} Percentile

Q_4 (max) ⟶ 100% ⟶ 100^{Th} Percentile

⇒ Inter-Quartile Range (IQR)



$$IQR = Q_3 - Q_1$$

Here, outliers are the Extreme values, Then,

Lower fence = $Q_1 - (1.5 * IQR)$

Upper fence = $Q_3 + (1.5 * IQR)$

∴ Thus, Values $\in$ [Lower fence, Upper fence] are valid values. and The values outside this Range are outliers.

"Box - Whisker Plot"

PAGE NO.:

Q → Let, Values = {2, 3, 3, 3, 3, 4, 4, 5, 5, 5, 6, 6, 6, 7, 8, 99}

$(n = 16)$

→ Q_1 = min Value = 0^{Th} = 2

Q_1 = 25^{Th} Percentile = $\frac{25}{100} \times 17^{Th} = 4.25 = 4^{Th} = 3$

Q_2 = 50^{Th} Percentile = $\frac{50}{100} \times 17^{Th} = 8.5^{Th} = \frac{5+5}{2} = 5$

Q_3 = 75^{Th} Percentile = $\frac{75}{100} \times 17^{Th} = 12.75^{Th} = 13^{Th} = 6$

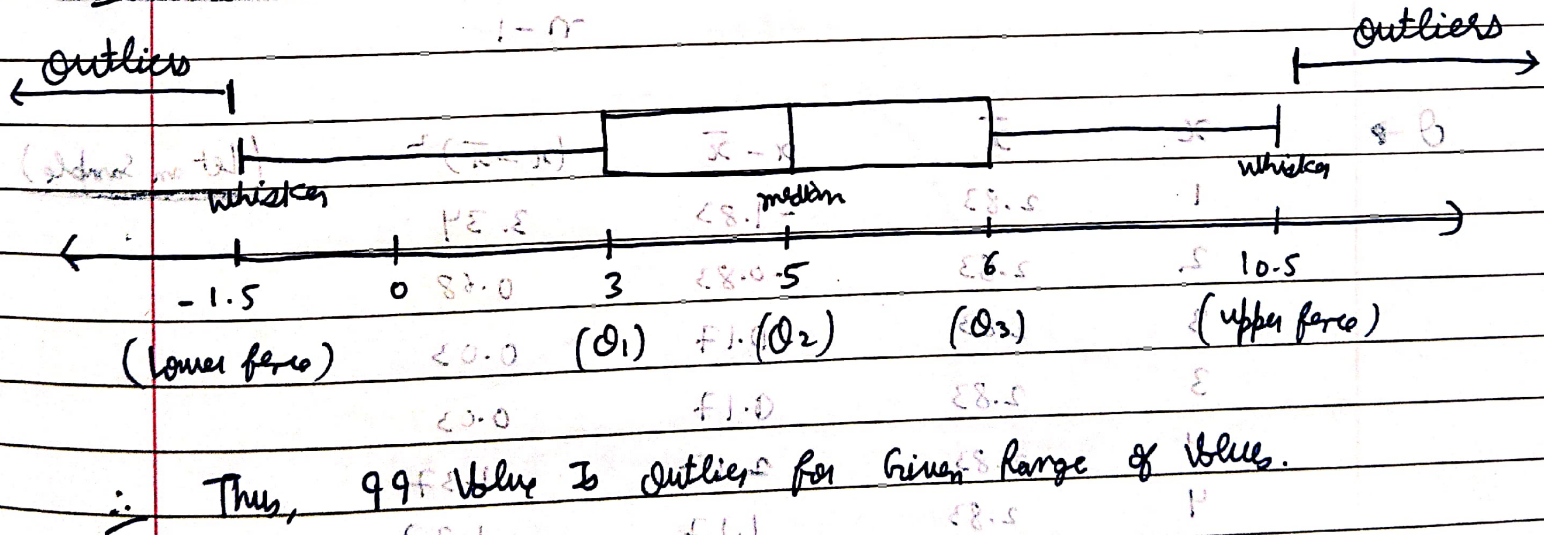
Q_4 = max Value = 100^{Th} = 99

∴ Thus, IQR = $Q_3 - Q_1 = 6 - 3 = 3$

Lower Fence = $Q_1 - (1.5 * IQR) = 3 - (1.5 * 3) = -1.5$

Upper Fence = $Q_3 + (1.5 * IQR) = 6 + (1.5 * 3) = 10.5$

The Box - Whisker Plot will be :-



∴ Thus, 99 Value is outlier for Given Range of Values.

PAGE NO.:

→ Mean - Deviation : Average of $(x - \mu)$.

Eg :- Let, Values = [1, 2, 3, 4, 5]

$$\text{Mean} = \frac{1+2+3+4+5}{5} = 3$$

$$\text{Mean Dev.} = \frac{|1-3| + |2-3| + |3-3| + |4-3| + |5-3|}{5}$$

$$\text{Mean Deviation} = \frac{2+1+0+1+2}{5} = \frac{6}{5} = 1.2$$

∴ On an Average, Each of the data Point are 1.2 Unit away from mean Value

4. Variance : The Average of Squared differences from the mean.

(i) Population Variance $= \sigma^2 = \frac{\sum_{i=1}^{N} (x_i - \mu)^2}{N}$ (Complete Variance)

(ii) Sample Variance $= s^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}$ (Sample Variance)

→

x	$\bar{x}$	$x - \bar{x}$	$(x - \bar{x})^2$
1	2.83	-1.83	3.34
2	2.83	-0.83	0.68
3	2.83	0.17	0.03
3	2.83	0.17	0.03
4	2.83	2.17	1.37
4	2.83	1.17	1.37
			6.82

(Let on Sample)

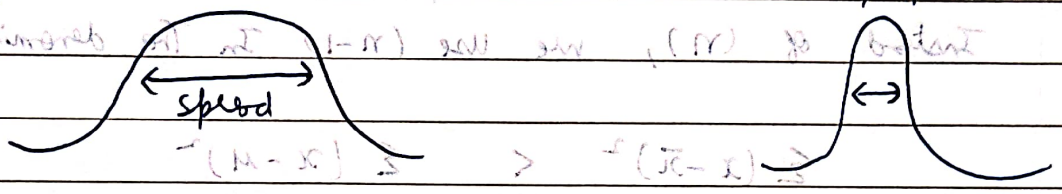
$$\therefore \sigma_5^2 = \frac{\Sigma (x - \bar{x})^2}{n-1} = \frac{6.82}{6-1} = \frac{6.82}{5} = 1.37$$

PAGE NO.:

How To Calculate Variance

1. Calculate The Mean.
2. For Each No. In data, subtract The mean from Number.
3. Square The differences.
4. Calculate The Average of Square of differences.

Note → When; $\sigma^2 \uparrow$, Spread ↑ $\sigma^2 \downarrow$, Spread ↓



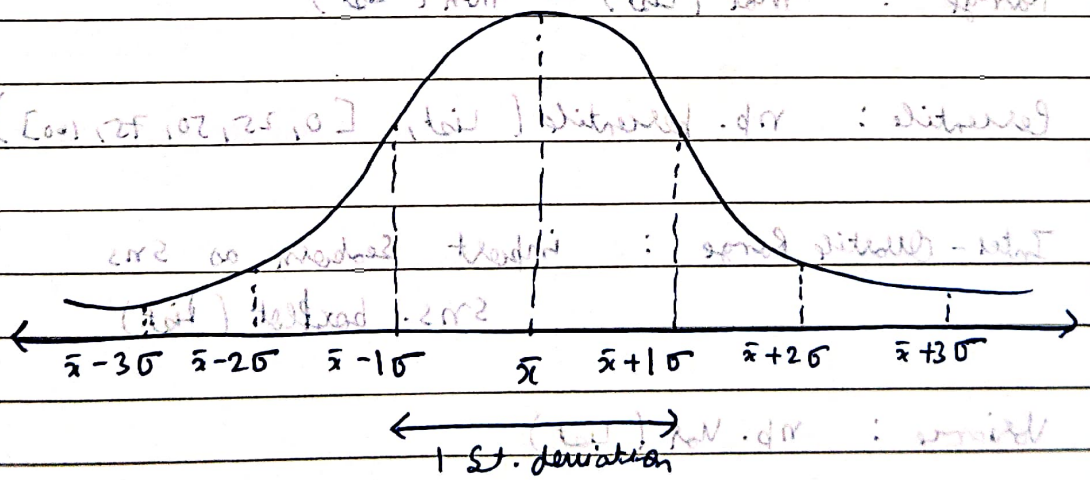
5. Standard Deviation : The Measure of How Spread Numbers are :-

$$\therefore S = \sqrt{\sigma^2} = \sqrt{\text{Variance}} = \sqrt{1.37} = 1.17$$

* $\sigma_P = \sqrt{\sigma^2_P}$ (Population)

$\sigma_S = \sqrt{\sigma^2_S}$ (Sample)

$\therefore$



⊙ Eg :- The Range for 1 Standard deviation will be :-

$(2.83 - 1.17,\ 2.83 + 1.17) = [1.16, 4]$

↳ All Points In This Range are within at 1 St. dev. from Mean.....

PAGE NO.:

Note → In Population, $\sigma^2_p = \frac{\sum_{i=1}^{N} (x-\mu)^2}{N}$

In Sample, $\sigma^2_s = \frac{\sum_{i=1}^{n} (x-\bar{x})^2}{(n-1)}$

∴ Now, $\sum (x-\mu)^2 >> \sum (x-\bar{x})^2$

∴ Thus, we use den. of Sample Smaller To make Ratio Equal. Thus, Instead of (n), we use (n-1) In The denominator.

$$\frac{\sum (x-\bar{x})^2}{n-1} < \frac{\sum (x-\mu)^2}{N}$$

∴ Thus, $\sigma^2_s \approx \sigma^2_p$: (JAI LOGIC)

Implementation of Method of Dispersion In Python :-

→ Use (Numpy as np) In The Program.

① Range : max (List) − min (List)

② Percentile : np. percentile (List, [0, 25, 50, 75, 100])

③ Inter-Quartile Range : import Seaborn as sns
sns. boxplot (List)

④ Variance : np. Var (List)

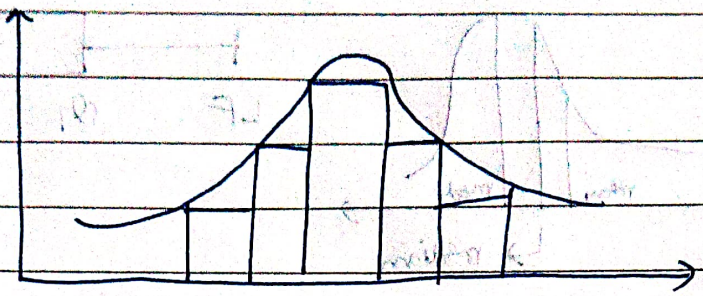
⑤ St. dev. : np. Std (List)

— JAI LOGIC

```python
import numpy as np # type: ignore
import seaborn as sns # type: ignore
import statistics

Numbers=[2,3,4,5,7,8,9]

# Range
print(f"The Range Will Be : {max(Numbers)-min(Numbers)}")

# Percentile
ans=np.percentile(Numbers,[0,25,50,75,100])
print(f"The Percentile Will Be : {ans}")

# Inter-Quartile Range(ITR) -> Q3 - Q1...
# Lower-Fence -> Q1 - (1.5*IQR)
# Upper-Fence -> Q3 + (1.5*IQR)
print(sns.boxplot(Numbers))

# Variance
print(f"The Variance By Numpy Will Be : {np.var(Numbers)}")
print(f"The Sample-Variance By Statistics Will Be : {statistics.variance(Numbers)}")
print(f"The Public-Variance By Statistics Will Be : {statistics.pvariance(Number
s)}")
# Standard-Deviation
print(f"The Standard-Deviation Will Be : {np.std(Numbers)}")
```

→ Measure of Symmetricity.

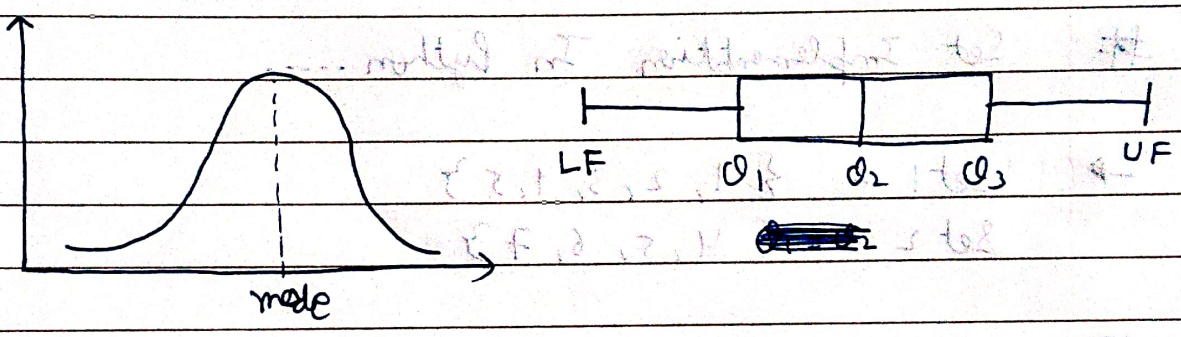
→ If Anything is Exactly Towards The left and Right.



→ Skewness : measure of data Symmetry

① No Skewness (Skewness = 0)

→

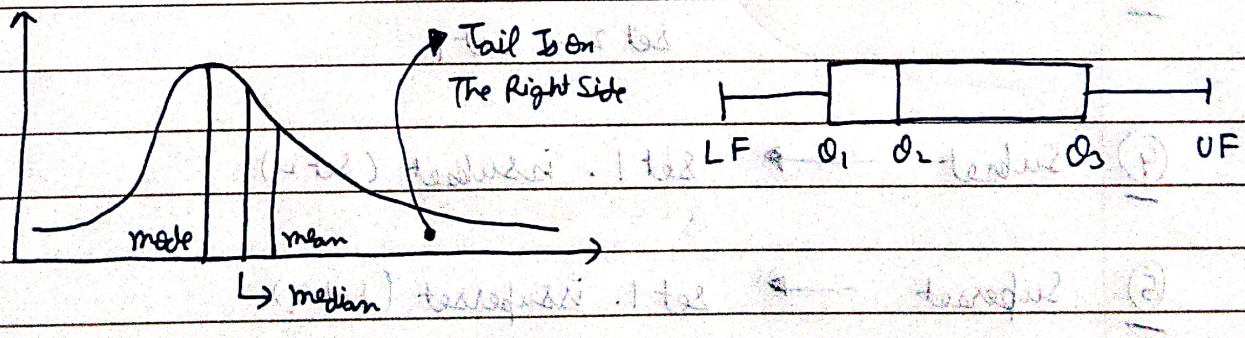


→ In No Skewness, mean = median = mode

$$Q_2 - Q_1 \approx Q_3 - Q_2$$

② +ve Skewed / Right Skewed (Skewness = ⊕)

→

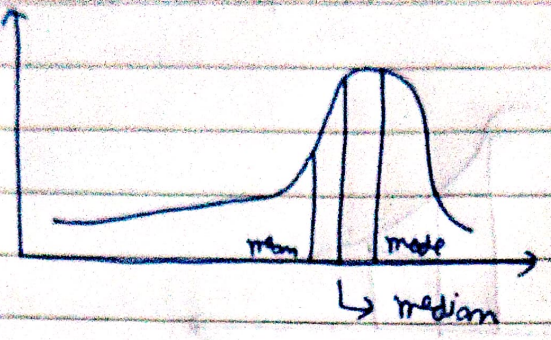


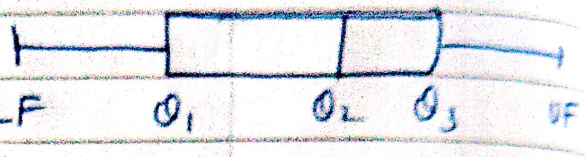
→ In +ve Skewness, mode ≤ median ≤ mean

$$Q_2 - Q_1 \leq Q_3 - Q_2$$

③ -Ve Skewed / left Skewed (Skewness = -ve)

→





→ In -Ve Skewed, mean ≤ median ≤ mode

$Q_1 - Q_2 \geq Q_3 - Q_2$

Set Implementation In Python ....

→ Set 1 = {1, 2, 3, 4, 5}

Set 2 = {4, 5, 6, 7}

① Union → Set1.Union(Set2) , (Set1 | Set2)

② Intersection → Set1.intersection(Set2), (Set1 & Set2)

③ Difference → Set1 - Set2

Set 2 - Set 1

④ Subset → Set1.issubset(Set2)

⑤ Superset → Set1.issuperset(Set2)

⑥ Symm. Diff. → Set1.symmetric_difference(Set2), (Set1 ^ Set2)

```python
set1 = {1,2,3,4,5}
set2 = {4,5,6,7,8}

# Union Of 2 Sets -> union = set1 | set2
print("Union of set1 and set2: ", set1.union(set2))

# Intersection of 2 sets -> -> intersection = set1 & set2
print("Intersection of set1 and set2: ", set1.intersection(set2))

# Difference of 2 sets -> diff = set1 - set2
print("Intersection of set1 and set2: ", set1.difference(set2))

# Subset And Superset In Python Sets...
print("Subset of set1 and set2: ", set1.issubset(set2))
print("Superset of set1 and set2: ", set1.issuperset(set2))

# Symmetric Difference Of 2 Sets -> SymmDiff = set1 ^ set2
print("Symmetric Difference of set1 and set2: ", set1.symmetric_difference(set2))
```

⇒ Covariance and Correlation

→ There are 2 Types of Relation as :-

① Direct ⟶ $X \propto Y$ (Eg :- House Area ∝ Buy Price)

② Indirect ⟶ $X \propto \frac{1}{Y}$ (Eg :- Age ∝ $\frac{1}{\text{Health}}$)

① Covariance

→ $$COV(x,y) = \frac{\sum_{i=1}^{n} (x_i - \bar{x}) \cdot (y_i - \bar{y})}{n-1}$$ ⊛ रटो !

→ Variance Was Spread of data , whereas, Covariance means you are been Trying To Understand The Relationship of an feature wrt others.

θ→ Let, $X = [2, 3, 6, 1] \longrightarrow \bar{x} = 3$

$Y = [3, 5, 6, 8] \longrightarrow \bar{y} = 5.5$

→

X	Y	$x - \bar{x}$	$Y - \bar{Y}$	$(X-\bar{x}) \cdot (Y-\bar{Y})$
2	3	-1	-2.5	2.5
3	5	0	-0.5	0
6	6	3	0.5	1.5
1	8	-2	2.5	-5
				-1

$\therefore$ $$Cov(x,y) = \frac{\Sigma (x-\bar{x})(y-\bar{y})}{n-1} = \frac{-1}{4-1} = \frac{-1}{3} = -0.34$$

$\therefore$ Thus, x and y are ⊖ve Related To Each other.

$\therefore$ Thus, $(X \propto 1/Y)$.

Q → Let, $X = [2, 4, 6] \longrightarrow \bar{x} = 4$

$Y = [3, 5, 7] \longrightarrow \bar{y} = 5$

→

X	Y	$X - \bar{x}$	$Y - \bar{y}$	$(X-\bar{x}).(Y-\bar{y})$
2	3	-2	-2	4
4	5	0	0	0
6	7	2	2	4
				8

$\therefore \quad Cov(x,y) = \frac{\Sigma(X-\bar{x})(Y-\bar{y})}{n-1} = \frac{8}{3-1} = \frac{8}{2} = 4 \equiv (+)$

$\therefore$ Thus, x and y are ⊕ ve Related i.e. $X \propto Y$.

Disadvantage of Corbriance

① No Comparison of The stergth In The Corbriance.

② No any standardization scale To Interpret The stergth.

③ Corbriance has dimensions, which makes The Comparin more difficult.

⊙ Eg :- Let, Cov (Height, Weight) = 24 m.kg —— ①

Cov (Height, Salery) = 54 m.Rs —— ②

Cov (Weight, Salery = 148 Kg.Rs —— ③

→ No Comparin of data Possible with ①, ② & ③.

— JAI LOGIC

② Person Correlation Coeff. $[-1, 1]$

→ $\rho_{x,y} = \dfrac{Cov(x,y)}{\sigma_x \cdot \sigma_y} = [-1 \text{ To } 1]$

→ $-1 \longleftarrow 0 \longrightarrow 1$

The more ⊖, The more -ve Correlated feature

The more ⊕, The more ⊕ve Correlated feature

① Eg:- $\rho_{X,Y} = 0.4$, $\rho_{A,B} = 0.8$ } Feature (A, B) are Highly Correlated As Compared To Feature (X, Y)....

⊖→ Calculate $\rho_{x,y}$, $X = [2, 3, 6, 1] \longrightarrow \bar{x} = 3$

$Y = [3, 5, 6, 8] \longrightarrow \bar{y} = 5.5$

→

X	Y	$X - \bar{X}$	$Y - \bar{Y}$	$(X - \bar{X})(Y - \bar{Y})$
2	3	-1	-2.5	2.5
3	5	0	-0.5	0
6	6	3	0.5	+1.5
1	8	-2	2.5	-5
				-1

① $\sigma^2_x = \dfrac{\Sigma (X - \bar{x})^2}{n-1} = \dfrac{(-1)^2 + 0^2 + 3^2 + (-2)^2}{4-1}$

$\sigma^2_x = \dfrac{1 + 0 + 9 + 4}{3} = \dfrac{14}{3} = 4.67$

∴ $\sigma_x = \sqrt{\sigma^2_x} = \sqrt{4.67} = 2.161$

PAGE NO.:

(2) $\sigma^2_y = \frac{\Sigma(y-\bar{y})^2}{n-1} = \frac{(-2.5)^2 + (-0.5)^2 + (0.5)^2 + (2.5)^2}{4-1}$

$= \frac{6.25 + 0.25 + 0.25 + 6.25}{3} = \frac{13}{3} = 4.33$

$\therefore \sigma_y = \sqrt{\sigma^2_y} = \sqrt{4.33} = 2.08$

(3) $Cov(x,y) = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{n-1} = \frac{-1}{4-1} = \frac{-1}{3} = -0.33$

$\therefore \rho_{x,y} = \frac{Cov(x,y)}{\sigma_x \cdot \sigma_y} = \frac{-0.33}{2.16 \times 2.08} = 0.0734$

$\therefore$ Thus, X Is Very less Correlated with Y.

Note → Pearson Correlation Coeff. always measures The Linear Relationship.

→ For Non-Linear Relationship, Use Spearman Rank Correlation

⇒ Spearman Rank Correlation (γ)

→ $\gamma_{x,y} = \frac{Cov(R(x), R(y))}{\sigma_{R(x)} \cdot \sigma_{R(y)}}$

⊙ Where, $R(x)$ = Rank of x ...

$R(y)$ = Rank of y ...

Q → Find $\gamma_{X,Y}$, $X = [5, 7, 8, 1, 2] \longrightarrow \bar{X} = 4.6$

$Y = [6, 4, 3, 1, 2] \longrightarrow \bar{Y} = 3$

→ For R(X), Sort X → $X = [\underset{⑤}{1}, \underset{④}{2}, \underset{③}{5}, \underset{②}{7}, \underset{①}{8}]$

(Numbers In Reverse Order)

→ For R(Y), Sort Y → $Y = [\underset{⑤}{1}, \underset{④}{2}, \underset{③}{3}, \underset{②}{4}, \underset{①}{6}]$

∴

X	R(X)	Y	R(Y)
5	3	6	1
7	2	4	2
8	1	3	3
1	5	1	5
2	4	2	4

∴ Thus, $R(X) = [3, 2, 1, 5, 4] \longrightarrow \overline{R(X)} = 3$

$R(Y) = [1, 2, 3, 5, 4] \longrightarrow \overline{R(Y)} = 3$

∴

R(X)	R(Y)	$R(X) - \overline{R(X)}$	$R(Y) - \overline{R(Y)}$	(Product)
3	1	0	-2	-2
2	2	-1	-1	1
1	3	-2	0	0
5	5	2	2	4
4	4	1	1	1
				4

① $\sigma^2_{R(X)} = \dfrac{\sum (R(X) - \overline{R(X)})^2}{n-1} = \dfrac{0^2 + (-1)^2 + (-2)^2 + 2^2 + 1^2}{5-1}$

$\sigma^2_{R(X)} = \dfrac{0+1+4+4+1}{4} = \dfrac{10}{4} = 2.5$

∴ Thus, $\sigma_{R(X)} = 1.581$

PAGE NO.:

(2) $\sigma^2_{R(Y)} = \frac{\Sigma (R(Y) - \overline{R(Y)})^2}{n-1} = \frac{(-2)^2 + (-1)^2 + 0^2 + 2^2 + 1^2}{5-1}$

$\sigma^2_{R(Y)} = \frac{4+1+0+4+1}{4} = \frac{10}{4} = 2.5$

$\therefore \sigma_{R(Y)} = \sqrt{\sigma^2_{R(Y)}} = \sqrt{2.5} = 1.581$

(3) $\text{Cov}(R(X), R(Y)) = \frac{\Sigma (R(X) - \overline{R(X)}) \cdot (R(Y) - \overline{R(Y)})}{n-1} = \frac{4}{5-1} = 1$

$\therefore \rho_{R(X), R(Y)} = \frac{\text{Cov}(R(X), R(Y))}{\sigma_{R(X)} \cdot \sigma_{R(Y)}} = \frac{1}{1.581 \times 1.581} = 0.4$

$\therefore$ Thus, $\rho_{R(X), R(Y)} = 0.4$

$\therefore$ Thus, Spearman Rank Correlation (X, Y) Is 0.4.

Note → If we have To Try Combo which can Result In Better Growth of The Company, we can use Correlation There.

(e) Let, Product = X , Options = {A, B, C}

→ Let, On Computation from data, $\rho_{X,A} = -0.8$ → ⊖ve Relⁿ

$\rho_{X,B} = 0.01$ → No Relⁿ

$\rho_{X,C} = 0.9$ → +ve Relⁿ

$\therefore$ Thus, we can use "X - C" Combo Here.

- JAI LOGIC

```python
import pandas as pd
import numpy as np

# Load your DataFrame (Example data, replace with actual dataset)
df = pd.DataFrame({
    'A': [1, 2, 3, 4, 5],
    'B': [2, 3, 4, 5, 6],
    'C': ['Yes', 'No', 'Yes', 'No', 'Yes']  # Non-numeric column
})

# Convert categorical values ('Yes'/'No') to numeric if present
df.replace({'Yes': 1, 'No': 0}, inplace=True)

# Select only numeric columns
df_numeric = df.select_dtypes(include=[np.number])

# Calculate the correlation matrix
corr_matrix = df_numeric.corr()

# Print the correlation matrix
print("Correlation Matrix:")
print(corr_matrix)

```

PAGE NO.:

Implementation of Cov(X,Y) and $S_{X,Y}$ In Python

① Load The data In The Variable Let "df".

② For Covariance : df.cov()
For Coorelation : df.corr()

— JAI LOGIC

→ Random Variable

→ A Set of possible values from an Random Experiment.

⊙ Eg :- Let, Sample = {HHH, HHT, HTH, THH, HTT, THT, TTH, TT
X = # Heads

∴ Then, P(X=1) = 3/8
P(X=2) = 3/8
P(X=3) = 1/8
P(X=0) = 1/8

→ Random Variable Is an set of possible values from an Random Exp.
→ An Random Variable Value Is been Unknown.
→ An function which assignes Values To Each of Exp. outcomes.

Probability Dist^n F^n (PDF)

→ Irrespective of outcome Nature, draw outcomes in form of distribut
This is known as Prob. dist^n F^n.

→ The 2 Types are → Discrete (Prob. Mass F^n).
→ Continious (Prob. density F^n).

① Probability Mass Function

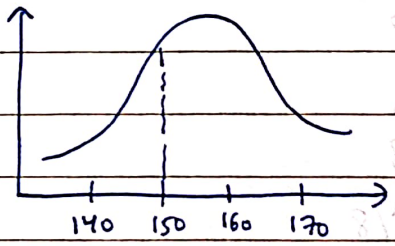
→ The Distribution of the discrete Random Variable.

⊙ Eg:- Rolling on dice.

② Probability Density Function

→ The Distribution of the Continious Data.

⊙ Eg:-



, find P(X ≤ 150)

③ Cummulative Distribution Function (CDF)

→ CDF is the Summation of all probabilities possible upto an Given point.

⊙ Eg:- Let the case for Rolling of dice :-

X	P(X)	cdf
1	1/6	1/6
2	1/6	2/6
3	1/6	3/6
4	1/6	4/6
5	1/6	5/6
6	1/6	1

Different Types of Probability Distribution Fⁿ (PDF)

→ The 2 Types of Probability distribution Fⁿ are ;-

1. Prob. density Fⁿ
   - → Normal Distⁿ / Gaussian Distⁿ.
   - → Standard Normal Distⁿ.
   - → Log - Normal Distⁿ.
   - → Chi - Square Distⁿ.
   - → F - Distribution.
2. Prob. Mass Fⁿ
   - → Bernoulli Distⁿ.
   - → Binomial Distⁿ.
   - → Poisson Distⁿ.

⍟ Uniform Distⁿ Is Both Discrete & Continious PDF.

→ Discrete Uniform Distributions → DISCRETE ....

→ An Uniform Distribution refers To The Type of an Probability Distⁿ in which The outcomes are Equally Likely.

→ Uniform Distⁿ are Both Continious and Discrete.

→ Uniform Distⁿ Fⁿ

→ Notation of Uniform Distⁿ = U(a, b) | Unif{a, b}

→ Mean of Discrete Uniform Distⁿ = $\frac{a+b}{2}$; (a = Start, b = End)

Variance of Discrete Uniform Distⁿ = $\frac{n^2-1}{12}$

PAGE NO.:

DISCRETE

→ Bernoulli Dist^n (Binary → 2 outcomes)

→ An Discrete Prob. Distribution of an Random Variable which Takes only Two Possible outcomes, Typically labelled as Success (Coded As 1), Failure (Coded as 0) with an fixed Prob. of Success & Failure.

⊙ Eg:- X = "Heads In an Toss → Success (Head) = 0.5 ≡ p
Failure (Tail) = 0.5 ≡ 1−p

∴ bmf = P(X=k) = { p if k = 1 (Success); 1−p if k = 0 (Failure) }

$$P(x=K) = p^K (1-p)^K$$

Conditions of Bernoulli Dist^n

→ No. of Trials must be finite.
→ Each Trial must be Independent.
→ There are only 2 Possible outcomes as: Success & Failure.
→ Prob. of Each output should be same In Every Trial.

✪ In Bernoulli Dist^n, Mean = p, Variance = p(1−p)

O→ Rumesh Bowls 6 Balls at wicket with The Prob. of 0.6 at hitting The Stumps with Each Ball. What Is The Prob. of Not Hitting an wicket?

→ Prob. of Hitting (Success) = p = 0.6
Prob. of Not-Hitting (Failure) = 1−p = 0.4

⊙ ///ly, Mean = p = 0.6
Variance = p(1−p) = 0.6 × 0.4 = 0.24
St. Dev = $\sqrt{\text{Var}} = \sqrt{0.24} = 0.4899$

DISCRETE

→ Binomial Dist^n (Bernoulli with n-Trials)

→ Binomial Dist^n Is The 'n' Bernoulli Trial.

$$\text{pmf} = {}^nC_k \, p^k \, (1-p)^{n-k}$$ ⭐ रटो!

Imp. Points about Binomial Dist^n

→ The Binomial dist^n Is PdB i.e. '$\sum P(x) = 1$'.

→ Moment Generating F^n (MGF) = $(P e^t + q)^n$

→ Characteristic F^n (Char. F^n) = $(P e^{it} + q)^n$

→ Prob. Generating F^n (PGF) = $(P Z + q)^n$

→ In Binomial Dist^n, Mean = np } ⭐ रटो!
Variance = npq }

Q → With 3 Tosses, What Is The Prob. Of Getting Exactly 2 Heads?

→ Here, $n = 3$, $k = 2$, $P(\text{Head}) = p = 0.5$, $q = 1-p = 0.5$

① $P(X = 2) = {}^nC_k \, p^k \, q^{n-k}$

$= {}^3C_2 \, (0.5)^2 \, (0.5)^{3-2}$

$= \frac{3!}{2!1!} \cdot \left(\frac{1}{2}\right)^2 \cdot \left(\frac{1}{2}\right)^1$

$= \frac{3 \times 2 \times 1}{2 \times 1} \times \frac{1}{4} \times \frac{1}{2} = \frac{3}{8} = 0.375$ ∴

② Mean = $np = 3 \times 0.5 = 1.5$

③ Variance = $npq = 3 \times 0.5 \times 0.5 = 0.75$

St. Dev = $\sqrt{\text{Var}} = \sqrt{0.75} = 0.8660$

DISCRETE

→ Poisson Distribution

→ The Poisson Distribution is an discrete prob. distribution that describ The # Events that occur within an fixed Interval of Time / Sp given an Known Average Rate of Occurances.

→ No. of Events Occuring in an fixed Time Interval.

$$P(x) = \frac{\lambda^x e^{-\lambda}}{x!}$$

(★ रटो)

Imp. Points about Poisson Dist^n

→ Poisson Dist^n Is an pmf i.e. $\sum P(x) = 1$.

→ Moment Generating F^n (MGF) $= e^{\lambda(e^t - 1)}$

→ Characteristic F^n (Char F^n) $= e^{\lambda(e^{it} - 1)}$

→ Prob. Generating F^n (PGF) $= e^{\lambda(z-1)}$

→ In Poisson Dist^n, Mean = Variance = λt (t = # Intervals)

Q → The Avg. Number of Customers Entering a Store In an House Is 5. What Is The Prob. That Exactly 3 Customers will Enter The next Hour?

→ Here, $\lambda = 5$

$\therefore$ $P(X = 3) = \frac{\lambda^x e^{-\lambda}}{x!} = \frac{5^3 (2.7182)^{-5}}{3!} = 0.142$

Q /// ly, Mean = $\lambda t = \lambda (t = 1)$
Variance = $\lambda t = \lambda (t = 1)$ → 5

PAGE NO.:

Continious Uniform Distribution → CONTINIOUS....

→ An Uniform Distribution refers to an Type of Prob. Distribution in which outcomes are been Equally Likely.

→ Uniform Dist^n → Discrete Uniform Dist^n (pmf). (See 2 Page Back)
→ Continious Uniform Dist^n (pdf).

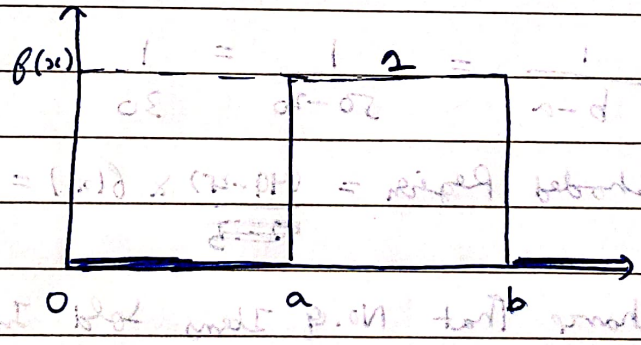
→ An Continious Distribution F^n Is an dist^n that Has an Infinite No. of Values defined In an Specified Range / Bound.

→ The Random Variable Is Continious In This.

→ Notation : $U(a, b)$

Parameters : $-\infty < a < b < \infty$, $b >> a$

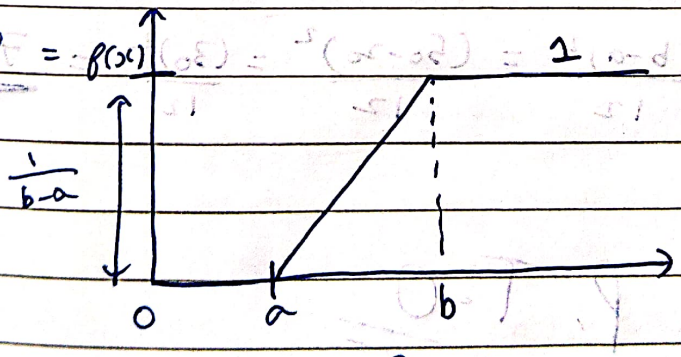
→ pdf =



→ Area of Rectangle = L X B = $(b-a) \times f(x)$

$1 = (b-a) \times f(x)$ ⇒ $\boxed{f(x) = \frac{1}{b-a}}$

→ cdf =



∴ Thus, $Cdf = \begin{cases} 0 & , x < a \\ \frac{x-a}{b-a} & , x \in [a,b] \\ 1 & , x > b \end{cases}$

PAGE NO.:

Note → For Continious Uniform Dist^n, Mean $= \frac{a+b}{2}$

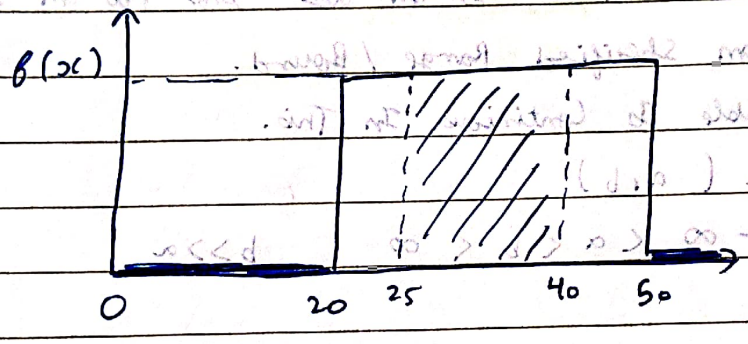
Variance $= \frac{(b-a)^2}{12}$

No Need for Learning for AIML....

Q → The No. of Items sold at shop daily is Uniformly distributed with the Max & Min Items Sold as 50 & 20.

→ Find The Prob. of Sales to fall b/w 25 & 40

→



→ As, $f(x) = \frac{1}{b-a} = \frac{1}{50-20} = \frac{1}{30}$

→ Now, Area of Shaded Region $= (40-25) \times f(x) = \frac{15}{} \times \frac{1}{30} = 0.5$

∴ Thus, 50 % Chance That No. of Items Sold In [25, 30]

⊙ For Range a = 20 and b = 50,

$$\text{Mean} = \frac{b+a}{2} = \frac{50+20}{2} = \frac{70}{2} = 35$$

$$\text{Variance} = \frac{(b-a)^2}{12} = \frac{(50-20)^2}{12} = \frac{(30)^2}{12} = 75$$

CONTINUOUS

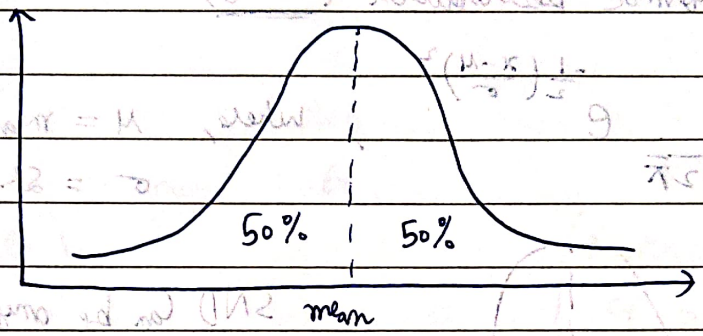
→ Normal Distribution / Gaussian Distⁿ

→ An Continious Prob. distribution.

→ It Is an Bell - Shaped distribution.

→ Most of The Real World data follows an Normal distribution.

⊙ Eg :- Height of an Population, Exam Score of an class



Characteristic of Normal Dist.

→ ND are Symmetrical about The mean.

→ Mean = Median = Mode

→ Skewness of ND Is 0.

Emperical Rule of an Normal Distribution

68% — 95% — 99.7% Rule

68% of The Total data Points will lie within 1 St. dev Range
68% In $[\bar{x}-\sigma, \bar{x}+\sigma]$

95% of The Total data Points will lie within 2σ
95% In $[\bar{x}-2\sigma, \bar{x}+2\sigma]$

99.7% of Points lie with 3σ Range of ND
$[\bar{x}-3\sigma, \bar{x}+3\sigma]$

∴ Thus, Based on The Emperical Rule :-

$P(\mu - \sigma \leq X \leq \mu + \sigma) = 68\% = 0.68$

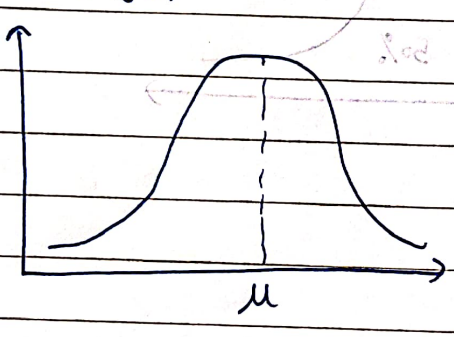
$P(\mu - 2\sigma \leq X \leq \mu + 2\sigma) = 97\% = 0.97$

$P(\mu - 3\sigma \leq X \leq \mu + 3\sigma) = 99.7\% = 0.997$

— JAI LOGIC

→ Standard Normal Distribution (SND)

→ $pdf = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$, where, μ = mean, σ = St. dev.

→ 

SND can be any value of μ & σ....

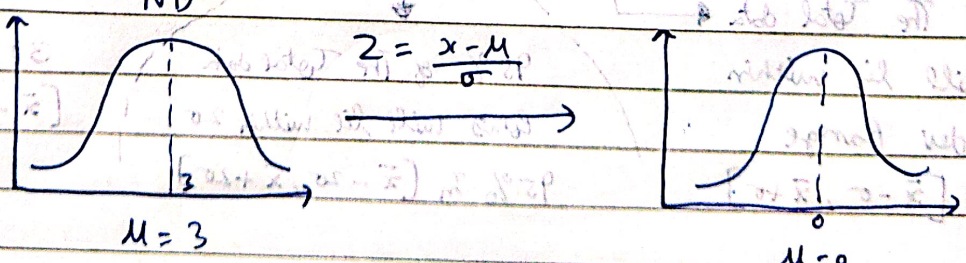
→ St. Normal Distⁿ Is an Continious Probability distribution.

→ It Is an Special Case of Normal Distⁿ where $\mu = 0$ & $\sigma = 1$.

Note → Let, Graph 1 Represent Normal Distⁿ Curve for Marks In Science Stream.
Graph 2 " " " " " " " Arts Stream.

→ As, There are Many different factors In Both Streams, Thus, We Can't Compare The 2 Graph directly, Thus, For Comparin,
We will Convert ND curve To SND curve, Then Compare Them.

⊙ Eg :-



(Replace all Points of ND with $\frac{x-\mu}{\sigma}$)

Q→ Standardize The below data of Normal Dist^n :-

→ Given That $\mu = 1010$ & $\sigma = 20$

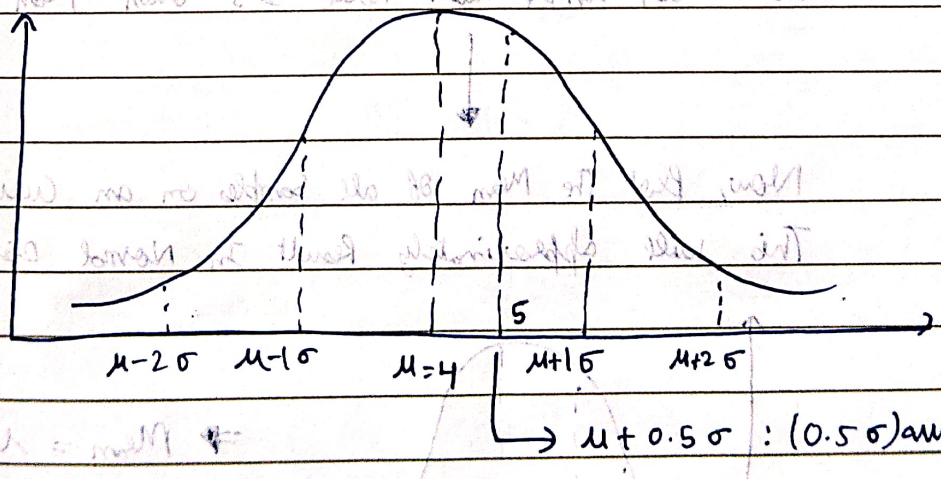
x	$x - \mu$	$Z = (x-\mu)/\sigma$
950	-60	-3
970	-40	-2
990	-20	-1
1010	0	0
1030	20	1
1050	40	2
1070	60	3
(X-Axis Scale of ND)		(X-Axis Scale of SND)

Note → $Z = \frac{x-\mu}{\sigma}$ } The Z-Value Simply Represent That How much St. dev The point Is away from mean....

⊙ Eg:- $x = 5$, $\mu = 4$, $\sigma = 2$

Then, $Z = \frac{x-\mu}{\sigma} = \frac{5-4}{2} = 0.5$ ∴ The point Is (0.5σ) Away from The mean of ND.....

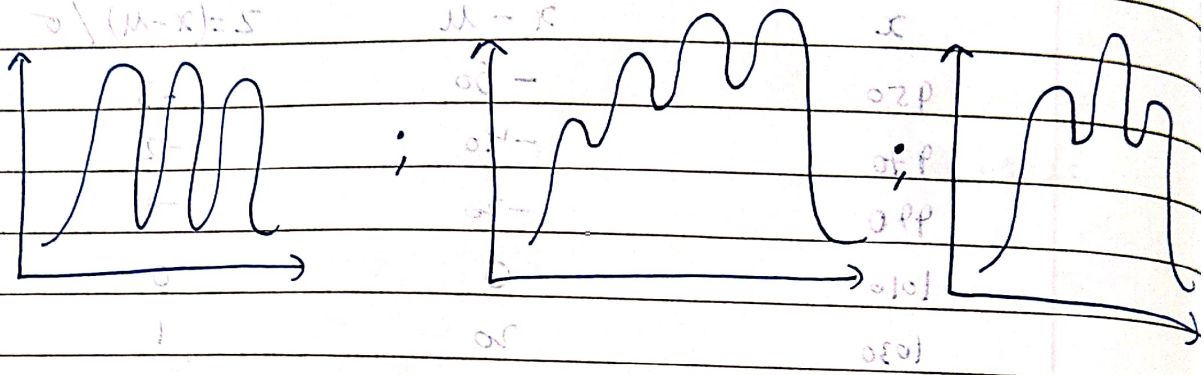
→



- JAI LOGIC

Central Limit Theorem (CLT)

→ Sometimes, The distributions can be Irregular.



→ The Central Limit Theorem states that If you have an population with mean μ and st. dev σ and Take sufficiently large samples (random) from the population with replacement, Then, the distribution of the sample will be Approximately Normal Dist^n.

→ Sampling Mean of population (μ) will be Approx. To Normal Dist^n.

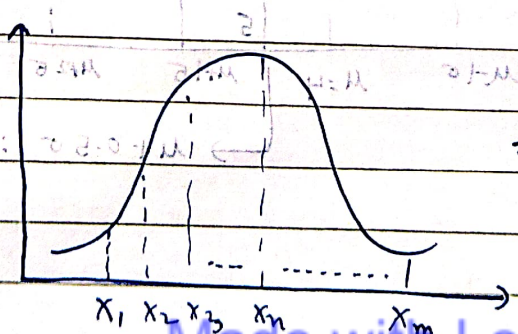
Given an population curve which is Irregular and Not Normal Dist^n having Mean = μ & St. dev = σ

↓

Divide the population into samples (large No. of samples) where Each sample will have it's own mean value

↓

Now, Plot the Mean of all samples on an curve, Then, This will approximately Result In Normal Dist^n curve



→ Mean = μ (Same as original curve)

St. Dev. = $\frac{\sigma}{\sqrt{n}}$

n → Samples Size

PAGE NO.:

The Two Conditions of CLT

→ The # Samples should be large.

→ The Sample Size should be Greater Than Equal To 30.
(Except The Population distⁿ which is already on Normal distⁿ)

Note → In Mean Curve, Mean = μ

$$\text{St. dev} = \frac{\sigma}{\sqrt{n}} \quad \text{(Standard Error)}$$

Where, Standard Error $\propto \frac{1}{n}$ ⇒ $n \uparrow$ SE $\downarrow$; $n \downarrow$ SE $\uparrow$

Q → You have an Population distⁿ with $\mu = 100$ and $\sigma = 20$. If You Take Sample size 50 from The Population, Then, what will be Prob. That mean will be less Than 105?

→ $\mu = 100$, $\sigma = 20$ and $n = 50$

→ ATQ, We have to Find Prob. Such That Sample Mean ≤ 105.

$\therefore$ $\bar{X} = 105$, $\mu_s = \mu = 100$

$$\sigma_s = \frac{\sigma}{\sqrt{n}} = \frac{20}{\sqrt{50}} = \frac{20}{5\sqrt{2}} = 2\sqrt{2} = \underline{2.83}$$

$\therefore$ Z_{score} (for Sample) $= \dfrac{\bar{X} - \mu_s}{\sigma_s} = \dfrac{105 - 100}{2\sqrt{2}} = \dfrac{5}{2\sqrt{2}} = \underline{1.7675}$

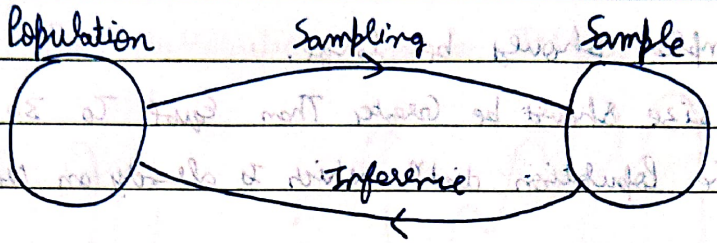
$\therefore$ $P(Z < 1.7675) = \underline{0.8944}$ (Value from Z-Table)

$\therefore$ Thus, The Probability will be $\underline{0.89}$.

PAGE NO.:

→ Estimate In Statistics

→



→ An Specified observed Value which Is been Used To Estimate an Uniform Population Parameter Using The Samples.

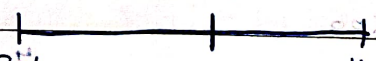
① Point Estimate → An Point Estimate is an Single Value that Is been used To Estimate The True Value of The Population Parameter.

⊙ Eg:- Avg. Salary of IT-Employee of TCS.

② Interval Estimate → Range of The Values used To Estimate The Unknown Population Parameters.

⊙ Eg:- The Avg. Salary Is 35K, whereas The Range (Confidence Range) where The Max Employee Lie Is 30K To 40K.

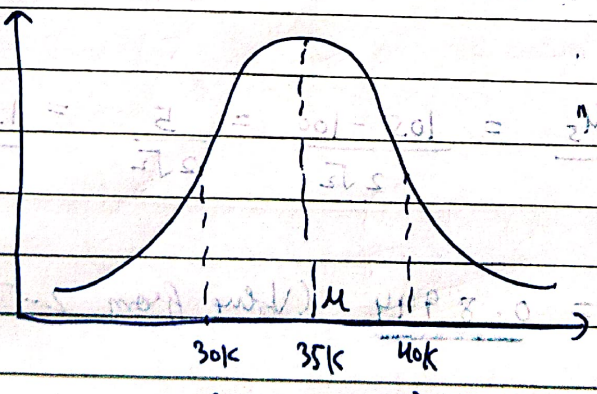
∴ Thus,



∴ 35K Is The Point Estimate

(30K-40K) Is The Interval Estimate.

∴



"35K Is Point Estimate" ∴

(30K-40K) Is The Confidence Interval ....

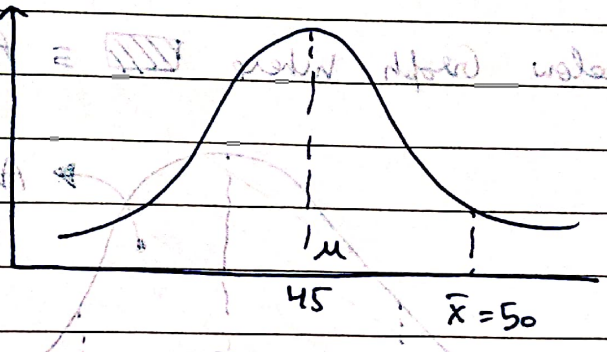
PAGE NO.:

P-Value

→ The P-Value Is The Probability Value Calculated from an Statistical Test.

→ P-Value Is Hypothesis Testing Which Is been used to decide whether To Reject The Null Hypothesis Or Not.

◉ Eg:- Mage of Employee Is 45 Years. Let, $\bar{X}$ of Sample Comes 50.



$$Z\text{ score} = \frac{\bar{X} - \mu}{(\sigma/\sqrt{n})}$$ } Corresponding To This Z-Value We will Get Probability Value....

◉ Eg:- Let, an Medicine Is 95% Effective which means, on 100 People, 95 will See Result whereas 5 willn't

∴ Thus, We are 95% Confident & 5% Margin of Error.

✱ Level of Significance (α) = Margin of Error = 5 %

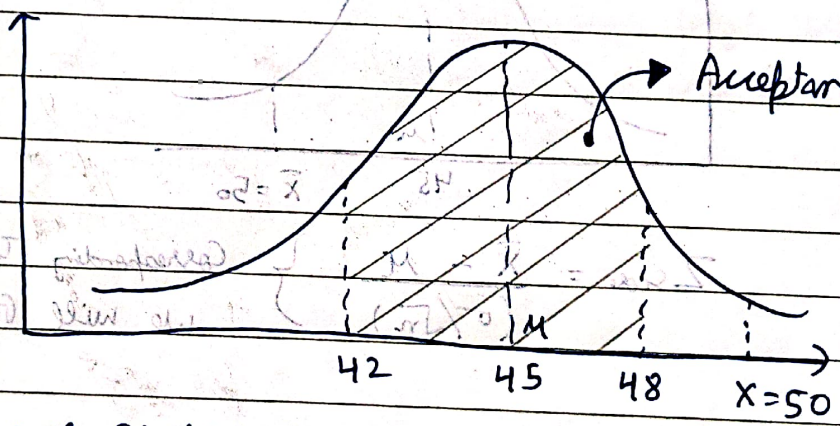
∴ Thus, Confidence Interval = 1 − (Error) = 1 − α = 0.95

Note → If, The Value Comes out of The Confidence Value Range, Then, Calculate Z Value of That Value & Corresponding Prob. To That Z-Value.

→ If, (Prob. < Level of Sig.) ⟶ Reject The Value
(Prob ≥ Level of Sig.) ⟶ Accept.

— JAZ LOGIC

Eg:- Let The below Graph where ▨ ≡ Acceptance Region.



→ Let, Level of Sig (α) = 5 % = 0.05
Then, for The Value X = 50, Calculate Z score,

$$Z\ score = \frac{X-\mu}{\sigma/\sqrt{n}} = \frac{50-45}{20/\sqrt{n}} = Something$$

Then, P(Z score) = wwww

→ If, Prob (Z score) < α ⟶ Reject The Value X = 50
Prob (Z score) ≥ α ⟶ Accept The Value X = 50

PAGE NO.:

→ Applications of Z-Score

→ Let, for an Normal Distⁿ ($\mu = 50$, $\sigma = 20$, $X = 110$)

Then, Let $Z = \frac{X_1 - \mu}{\sigma} = 3$ → $X = 110$ is 3σ Away from The Mean.....

⊙ Let, We have To find $P(Z > 0.5)$

Then, $P(Z > 0.5) = 1 - P(Z \leq 0.5)$

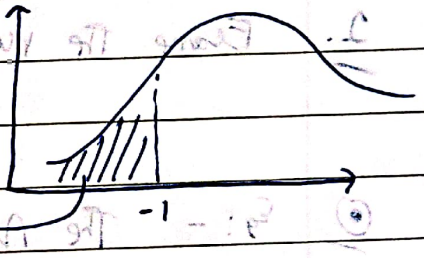
$= 1 - 0.6915$ (from Z-Table)

$= 0.3185$

θ→ What Is The P(Marks less Than 3) where $\mu = 4$ & $\sigma = 1$

→ $Z = \frac{X - \mu}{\sigma} = \frac{3-4}{1} = -1$

Then, $P(Z = -1) = 0.1587$



∴ The P(Marks lesser Than 3) Is 0.1587.

⊙ Similarly, $P(2 \leq Z \leq 3) = P(Z \leq 3) - P(Z \leq 2)$

— JAI LOGIC

⊙ Ex:- $P(2 < Z < 3) = P(Z \leq 3) - P(Z \leq 2)$

$= 0.9987 - 0.9772$

$= 0.0215 = 2.15\%$

PAGE NO.:

→ Hypothesis and Mechanism

→ Hypothesis is an claim as a Statement / Assumption about the Population Parameter That Can be Tested Using The Statistic methods.

◉ Eg:- Avg Salary of Employee of TCS is 7 LPA.

① Null Hypothesis : The Initial / default Assumption.

② Alternate Hypothesis :: The opposite of The Alternate Hypothesis.

Mechanism of Hypothesis

1. Frame The Hypothesis → Null Hypothesis (H_0).
   → Alternate Hypothesis (H_A).

◉ Eg:- The Age of PW Skills is Atleast 35 Years.

(H_0 will always have = Sign)

$\mu \geq 35$ (H_0) $\mu < 35$ (H_A)

◉ Eg:- The Avg Salary of TCS Employee is 7 LPA.

$\mu = 7$ (H_0) $\mu \neq 7$ (H_A)

◉ Eg:- The Age for M.Tech is Atmost 30 Years.

$\mu \leq 30$ (H_0) $\mu > 30$ (H_A)

PAGE NO.:

② Experiment / Statistical Analysis. (P-Value / Significance Value)

③ Reject H_0 / Fail To Reject H_0.

⇒ Hypothesis Testing & Statistical Analysis

→ Z-Test.] ⟶ Average ....
→ T-Test.
→ Chi-Square Test ⟶ Categorical Data ....
→ ANNOVA. ⟶ Variance ...

→ Z-Test

→ A/C To Central Limit Theorem,

$$\mu_s = \mu \quad \text{and} \quad \sigma_s = \frac{\sigma}{\sqrt{n}}$$

→ For Sample Z-Test, Sample size ≥ 30

σ_p must be Given

Then, $Z\ score = \dfrac{\bar{X} - \mu}{\sigma/\sqrt{n}}$

∴ Then, 5 steps are
- Frame The Hypothesis.
- Level of Significance.
- Type of Test Needed.
- Z score Evaluation.
- Probability.

Q → Suppose an Child Physologist says That The Avg. Time Working mothers Spend Talking Their Children Is upto 11 minutes per day. To Test The Hypothesis, We Conducted The Exp. on Random Sample of 100 Working mothers and found That They Spend 11.5 min per day with Their Childrens. Assume That The Pop. σ = 2.3 mins. Then, Conduct The Test with The 5% Level of Significance ($\alpha = 0.05$).

→ Time upto 11 min ⟶ $\mu \leq 11$ (H_0)

$\mu > 11$ (H_A)

① Frame The Hypothesis

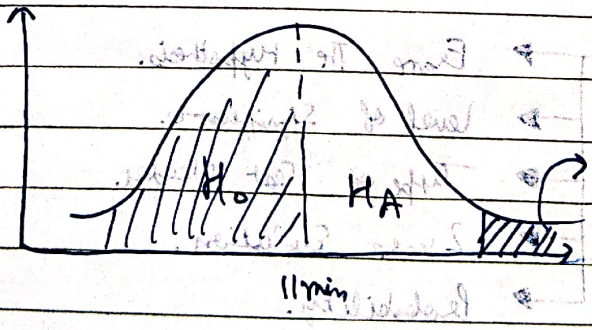
→ Then, $\mu \leq 11$ (H_0) & $\mu > 11$ (H_A)

② Level of Sig. (α) = 0.05

③ Type of Test ⟶ Z-Test [As, $n > 30$, σ_p Is Given as 2.3]

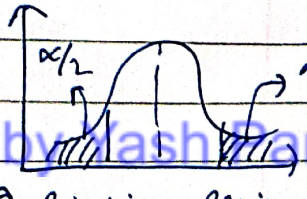
④ $Z_{score} = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} = \frac{11.5 - 11}{2.3/\sqrt{100}} = 2.17$

⑤



One-Tail Test (α)
(Rejection Region)

Note → Whereever, H_A has $\mu \neq$ mw, Then, It will be Two Tail Test



(2 Rejection Regions)

→ Now for checking of Hyp., we have 2 methods :-

PREFER

(i) Z - Critical : $\alpha = 0.05$ (from Z-Table, find the closest prob value to this, then, take Z-value of that)

$Z_{critical} = 1.64$

→ $Z \leq 1.64$ (Accept)

$Z > 1.64$ (Reject)

→ As, $Z_{score} = 2.17 > 1.64$ ⟶ Reject the H_0.....

(ii) p-Value : $P(Z_{score} = 2.17) = 0.9850$

$P(Z > 2.17) = 1 - 0.9850 = 0.015 < \alpha$

As, $0.015 < \alpha$, Reject the H_0.....

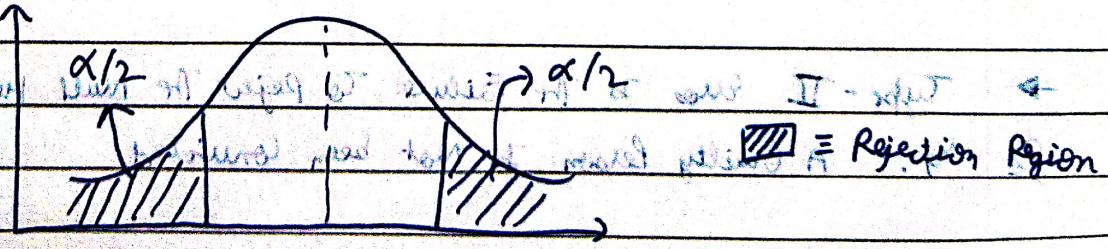
Q→ The Average Heights of all Residents in an city is 168 cm with an $\sigma = 3.9$ cm. On Research, they believe that mean to be different. He measured the height of 36 individuals and found as 169.5 cm. Test the Hypothesis with 95% Confidence Interval.

→ 95% Confidence Interval ⟶ $\alpha = 0.5$

① Frame the Hypothesis ⟶ $\mu = 168$ (H_0)

$\mu \neq 168$ (H_A)

→ As, H_A has (≠) sign, Thus, this will be an Two-Tail Test.



(2) Level of Significance : As, It Is an Two - tail Test,

$$\alpha = 5/2 = 2.5\ \% = 0.025$$

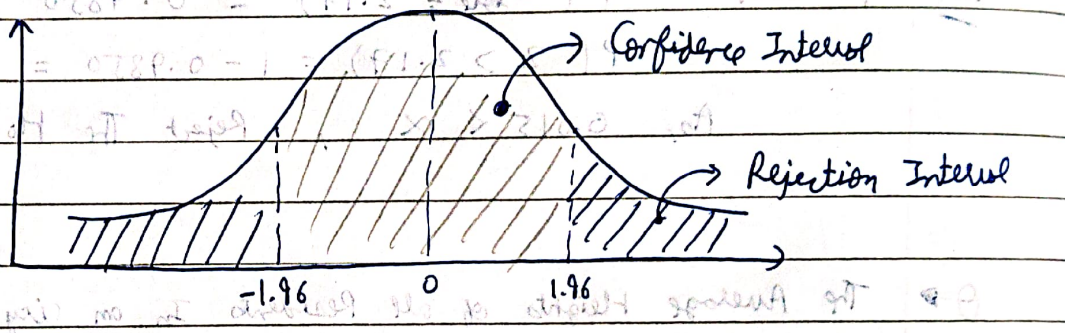
(3) Type of Test : Z - Test

(4) $$Z_{score} = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{169.5 - 168}{3.9/\sqrt{36}} = 2.31$$

(5) $Z_{critical}$: Here, $\alpha = 0.025$, Then,

$Z_{critical} = 1.96$ (Evaluated from Z - Table)

∴ Thus,



∴ Since, $Z_{score} = 2.31 > 1.96$ → Reject The H_0.

∴ Thus, we will Reject The Null Hypothesis.

EASY

→ Type - I and Type - II Error

→ Type - I Error Is The Rejection of The Null Hypothesis when It Is been actually True.

⊙ Eg :- An Innocent Person Is been Convicted.

→ Type - II Error Is The Failure To Reject The Null Hypothesis which was actualy false.

⊙ Eg :- A Guilty Person Is Not been Convicted.

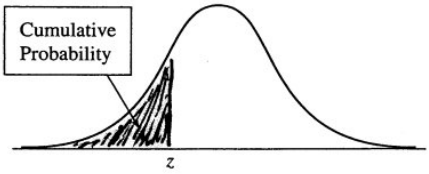
True Nature of Hypothesis (Actual)

Conclusion (Based on The Analysis)	H_0 Is True	H_A Is True
Support H_0	Correct Conclusion	Type - II Error
Support H_a	Type - I Error	Correct Conclusion

- H_0 Is True
  - Accept H_0 → Correct Conclusion
  - Reject H_0 → Type - I Error
- H_0 Is False
  - Accept H_0 → Type - II Error
  - Reject H_0 → Correct Conclusion

— JAI LOGIC

APPENDIX A

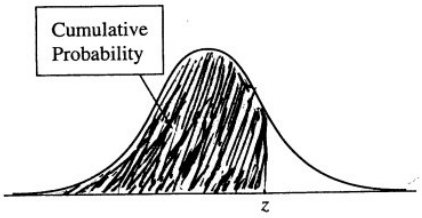


Cumulative probability for z is the area under the standard normal curve to the left of z

z	.00
−5.0	.000000287
−4.5	.00000340
−4.0	.0000317
−3.5	.000233

TABLE A Standard Normal Cumulative Probabilities

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
−3.4	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0002
−3.3	.0005	.0005	.0005	.0004	.0004	.0004	.0004	.0004	.0004	.0003
−3.2	.0007	.0007	.0006	.0006	.0006	.0006	.0006	.0005	.0005	.0005
−3.1	.0010	.0009	.0009	.0009	.0008	.0008	.0008	.0008	.0007	.0007
−3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010
−2.9	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
−2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
−2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
−2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
−2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
−2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
−2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
−2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
−2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
−2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
−1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
−1.8	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
−1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
−1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
−1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
−1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681
−1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
−1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
−1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
−1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
−0.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
−0.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
−0.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
−0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
−0.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
−0.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
−0.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
−0.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
−0.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
−0.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641



Cumulative probability for z is the area under the standard normal curve to the left of z

TABLE A Standard Normal Cumulative Probabilities *(continued)*

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998

z	.00
3.5	.999767
4.0	.9999683
4.5	.9999966
5.0	.999999713

→ Student - t - Distribution (t - distⁿ)

→ As, for Z - Test, σ_{Pop} must be Given and $n \geq 30$ Is Required

→ But, In majority of The Cases (σ_{Pop}) Wouldn't be Known.

→ Whereever, $n \leq 30$ & σ_{Pop} not Given } Use t - Test (t - distⁿ)

→ for Z - Test, $Z\ score = \dfrac{\bar{X} - \mu}{\sigma/\sqrt{n}}$

→ for T - Test, $T\ score = \dfrac{\bar{X} - \mu}{S/\sqrt{n}}$

S → Sample's St. dev....

Note → Degree of freedom of T-Test Is $n-1$.

→ Let, There are 5 Values, we have to select any 5 values such That The Avg. of These values Is 10.

Avg. (_ _ _ _ _ _) = 10

We Can Take Any Values for These 4 Spots of our Choice

We have To Take This Value Correctly and depend on The 4 Values such That Avg will be 10.

∴ Thus, for n-Values, (n-1) values are Unconstrained
2 Values are Constrained

∴ Thus, The degree of freedom will be (n-1).

— JAI LOGIC

Q → Suppose a Child Physiologist Say That The alleroge Time Working mothers Spend Talking Their children Is upto 11 min. To Test The Hypothesis, you Took an Random sample 20 Working mothers and found That The Avg. Time Spend Is 11.5. The Sample St. dev. Is 2.3. ($\alpha = 0.05$)

→ Sample Size < 30, σ_{pop} Is Not Given } Thus, Apply T-Test.....

① Ecare The Hypothesis

→ $M \leq 11$ ⟶ $H_0 = M \leq 11$
$H_A = M > 11$
(No ≠ Sign In H_A, Thus, One-Tailed Test)

PAGE NO:

(2) level of Sig. (α) = 0.05 (One-Tailed Test)

(3) Type of Test : T-Test

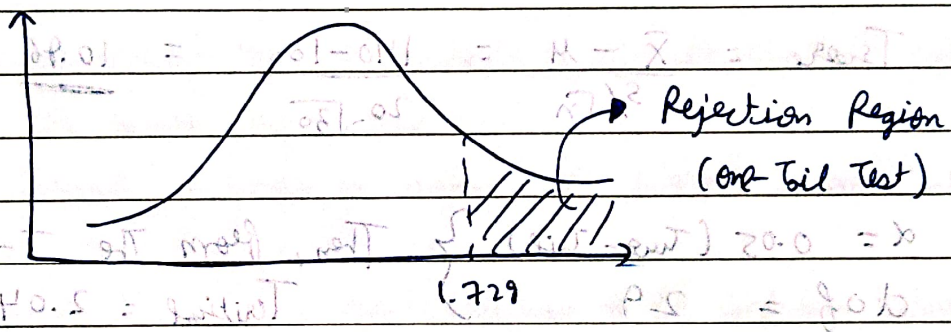
(4) $T_{score} = \frac{\bar{X} - \mu}{s/\sqrt{n}} = \frac{11.5 - 11}{2.3/\sqrt{20}} = 0.97$

(5) Test Score : As, $\alpha = 0.05$

for This Question, degree of freedom = $n - 1 = 20 - 1 = 19$

→ from T-Table with $\alpha = 0.05$ & dof = 19, $T_{score} = 1.729$ ($T_{critical}$)

∴



→ As, $T_{score} = 0.97 < 1.729$ → Accept The H_0 .....

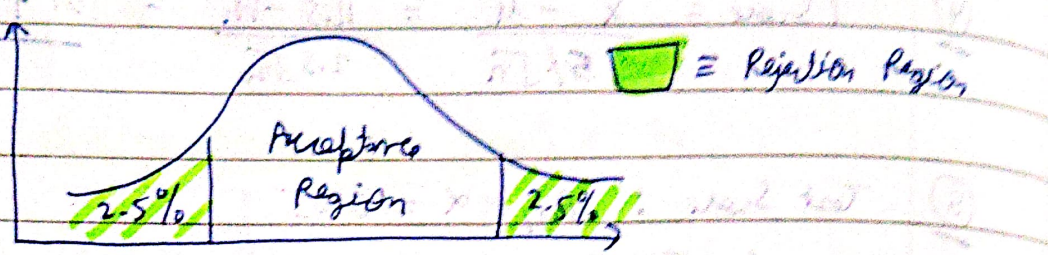
∴ Thus, we failed To Reject The Null Hypothesis.

Q→ In The Population, The Avg. IQ Is 100. A Team of Reaches Want To Test on medicine To check The ⊕ & ⊖ effect on IQ. A Sample of 30 Participate who Took medicine has mean Value of IQ as 140 with St. dev. of 20. Did The medicine effect The IQ? ($\alpha = 0.05$)

→ Sample Size ≤ 30
σ_{Pop} Is Not Given But
σ_{Sample} Is Given
} t-Test

① Frame The Hypothesis → $\mu = 100$ (H_0)
$\mu \neq 100$ (H_A) → Two-Tailed Test

② level of Sig. (α) = 0.05 (Two Tailed Test)

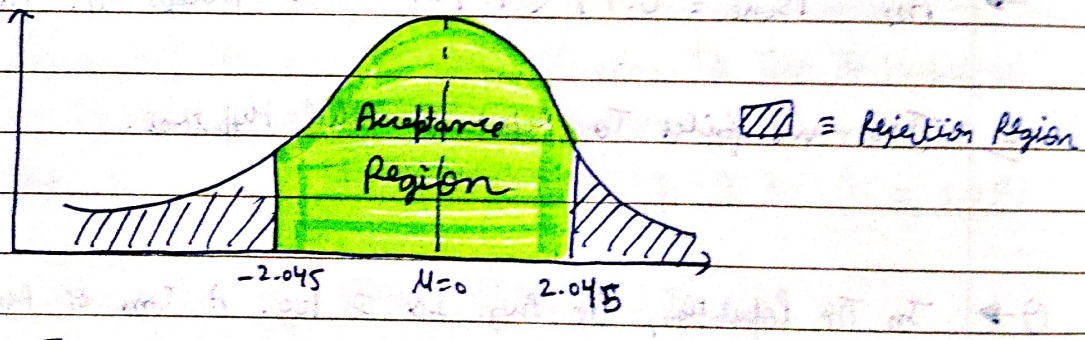


→ degree of freedom (dof) = $n - 1 = 29$

③ Type of Test : T-Test

④ $T_{score} = \frac{\bar{X} - \mu}{s/\sqrt{n}} = \frac{140 - 100}{20/\sqrt{30}} = 10.96$

⑤ $\alpha = 0.05$ (Two-Tail)
dof = 29
} Then, from The T-Table
$T_{critical} = 2.045$



→ Now, $T_{score} = 10.96 >> 2.045$ → Reject The H_0 .....

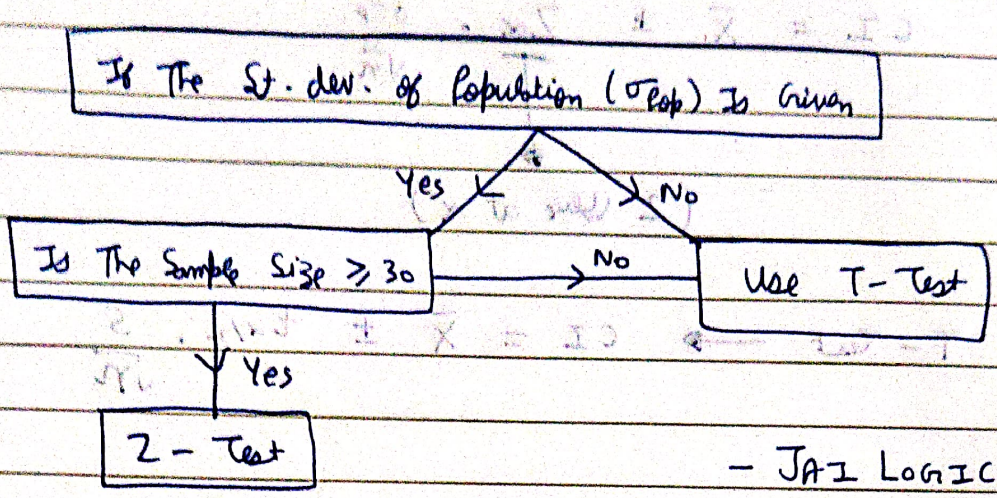
∴ Thus, We Rejected The H_0.

∴ Thus, The medicine has affected The [redacted] IQ.

t Table

cum. prob	$t_{.50}$	$t_{.75}$	$t_{.80}$	$t_{.85}$	$t_{.90}$	$t_{.95}$	$t_{.975}$	$t_{.99}$	$t_{.995}$	$t_{.999}$	$t_{.9995}$
one-tail	0.50	0.25	0.20	0.15	0.10	0.05	0.025	0.01	0.005	0.001	0.0005
two-tails	1.00	0.50	0.40	0.30	0.20	0.10	0.05	0.02	0.01	0.002	0.001
df											
1	0.000	1.000	1.376	1.963	3.078	6.314	12.71	31.82	63.66	318.31	636.62
2	0.000	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	22.327	31.599
3	0.000	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	10.215	12.924
4	0.000	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	0.000	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	0.000	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	0.000	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	0.000	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	0.000	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	0.000	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	0.000	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	0.000	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055	3.930	4.318
13	0.000	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012	3.852	4.221
14	0.000	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	3.787	4.140
15	0.000	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	0.000	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921	3.686	4.015
17	0.000	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898	3.646	3.965
18	0.000	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.610	3.922
19	0.000	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861	3.579	3.883
20	0.000	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845	3.552	3.850
21	0.000	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831	3.527	3.819
22	0.000	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819	3.505	3.792
23	0.000	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807	3.485	3.768
24	0.000	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797	3.467	3.745
25	0.000	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787	3.450	3.725
26	0.000	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779	3.435	3.707
27	0.000	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771	3.421	3.690
28	0.000	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763	3.408	3.674
29	0.000	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756	3.396	3.659
30	0.000	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750	3.385	3.646
40	0.000	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704	3.307	3.551
60	0.000	0.679	0.848	1.045	1.296	1.671	2.000	2.390	2.660	3.232	3.460
80	0.000	0.678	0.846	1.043	1.292	1.664	1.990	2.374	2.639	3.195	3.416
100	0.000	0.677	0.845	1.042	1.290	1.660	1.984	2.364	2.626	3.174	3.390
1000	0.000	0.675	0.842	1.037	1.282	1.646	1.962	2.330	2.581	3.098	3.300
z	0.000	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576	3.090	3.291
	0%	50%	60%	70%	80%	90%	95%	98%	99%	99.8%	99.9%
	Confidence Level										

Z - Test Vs T - Test



- JAI LOGIC

Confidence Interval and Margin of Error

- An Estimate of an Population Parameter is an Approximation depending solely on The Sample Information.
- A Point Interval Is Single No. wheras Conf. Interval Is an Interval.

* Point Estimate Is The Exact Middle location of The Confidence Interval.

Confidence Interval ⟶ 30 - 40

Point Estimate ⟶ 35

∴ Confidence Interval (CI) = Point Estimate ± Margin of Error

① Z - Test Conf. Interval for Two - Tail Test

$$CI = \bar{x} \pm Z_{\alpha/2} \cdot \frac{\sigma_p}{\sqrt{n}}$$

σ_p ⟶ Population St. Dev...

(Z Value at $\alpha/2$)

② Z-Test Confidence Interval for One-Tailed Test

→ $CI = \bar{X} \pm Z\alpha \cdot \frac{\sigma_p}{\sqrt{n}}$

(Z value at α)

③ T-Test → $CI = \bar{X} \pm t_{\alpha/2} \cdot \frac{S}{\sqrt{n}}$

EASY

Q → In An Exam, The St. dev. of marks Is 100. An Sample of 36 students h an mean of 500. Construct The 95% of Confidence Interval about mean

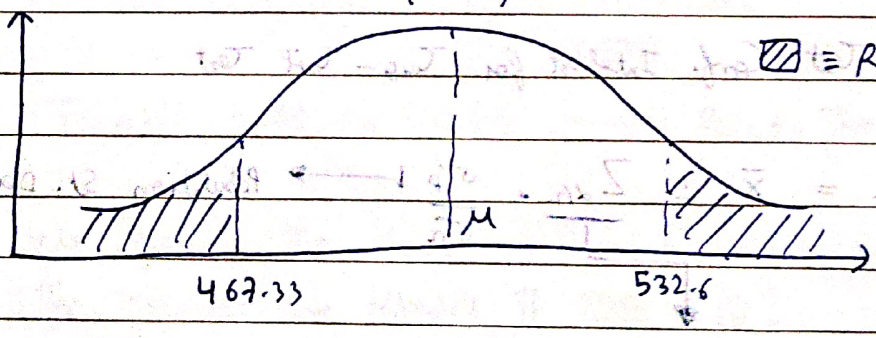
→ Level of Sig (α) = $1 - (CI) = 1 - 0.95 = \underline{0.05}$

∴ Then, $Z_{\alpha/2} = Z\left(\text{value} = \frac{0.05}{2}\right) = \underline{1.96}$

∴ $CI = \bar{X} \pm Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$

① Upper $CI = 500 + 1.96 \cdot \left(\frac{100}{\sqrt{36}}\right) = 532.6$

② Lower $CI = 500 - 1.96 \left(\frac{100}{\sqrt{36}}\right) = 467.33$



∴ I am 95% Confident That Score In The Exam lies b/w [467.33 - 532.6].

Chi - Square dist^n and Chi - Square Test

→ The Chi - Square dist^n of an Probability dist^n that describes of a sum of Squares of K - Random Variables.

→ No Two - Tail Test Possible In Chi - Square Test.

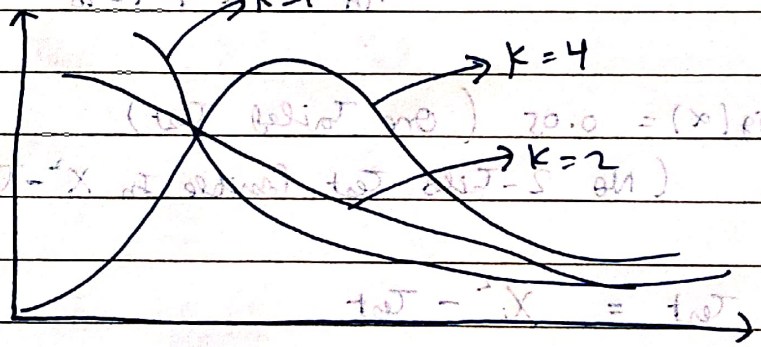
The Properties of Chi - Square Dist^n :-

→ The degree of freedom (dof) Is n-1.

→ Chi - Square dist^n Shape Is determined by 'K'

→ It Is an Non - Negative Dist^n.

→ It Is an Right Skewed Dist^n.



Chi - Square Test (X^2)

→ The Chi - Square Test Is Based on Chi - Square dist^n.

→ Goodness of fit Test Is Used To Perform This.

$$X^2 \text{ Stat.} = \sum \frac{(\text{Observed} - \text{Expected})^2}{\text{Expected}}$$

Q → 12% of People are left Handed. To Verify The Theory, We Test on Sample of 75 students where 11 were left Handed. ($\alpha = 0.05$)

→ Expected → Left-Handed = 12% of 75 = 9
Right-Handed = 75 − 9 = 66

Observed → Left-Handed = 11
Right Handed = 75 − 11 = 64

∴

	Expected (E)	Observed (O)
Left	9	11
Right	66	64

① Frame The Hypothesis → H_0 : $\mu = 12\%$
H_A : $\mu \neq 12\%$

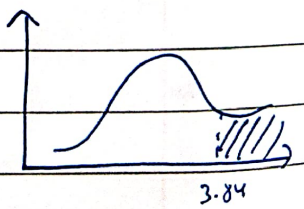
② Level of Sig (α) = 0.05 (One Tailed Test)
(No 2-Tailed Test Possible In χ^2-Test)

③ Type of Test = χ^2 - Test

④ $\chi^2_{Std.} = \sum\left(\frac{(O-E)^2}{E}\right) = \frac{(11-9)^2}{9} + \frac{(64-66)^2}{66} = 0.505$

⑤ Now, $\alpha = 0.05$, dof = $n - 1 = 2 - 1 = 1$

$\chi^2_{Critical}$ ($\alpha = 0.05$, dof = 1) = 3.84



∴ Now, As, $\chi^2_{Std.} = 0.505 < 3.84$

∴ Thus, We fail To Reject The Null Hypothesis.

∴ 12% of People are left Handed with 95% Confidence.

PAGE NO.:

Q → For an Sample of 500 people, observed & Expected Values were :-

	Expected (E)	Observed (O)
$W \leq 50$ kg	20% ≡ 100	140
$50 < W < 75$	30% ≡ 150	160
$W \geq 75$ kg	50% ≡ 250	200

→ Using $\alpha = 0.05$, Check The Expected Values are Correct ?

① → Frame The Hypothesis → H_0 : The data Is as per Expectation
H_A : " " " Not " " .

② Level of Sig (α) = 0.05

③ Type of Test : χ^2 - Test

④ $\chi^2_{Stat.} = \sum \frac{(O-E)^2}{E} = \frac{(140-100)^2}{100} + \frac{(160-150)^2}{150} + \frac{(200-250)^2}{250}$

$= 26.66$

⑤ Now, $\alpha = 0.05$ and $dof = n-1 = 3-1 = 2$

Thus, $\chi^2_{critical}\ (\alpha = 0.05, dof = 2) = 5.99$

∴ Now, $\chi^2_{Stat.} = 26.66 >> \chi^2_{critical}$

∴ Thus, Reject The Null Hypothesis.

∴ The observed Values of Weight are ~~not~~ Changed as were Expected.

Chi-square Distribution Table

d.f.	.995	.99	.975	.95	.9	.1	.05	.025	.01
1	0.00	0.00	0.00	0.00	0.02	2.71	3.84	5.02	6.63
2	0.01	0.02	0.05	0.10	0.21	4.61	5.99	7.38	9.21
3	0.07	0.11	0.22	0.35	0.58	6.25	7.81	9.35	11.34
4	0.21	0.30	0.48	0.71	1.06	7.78	9.49	11.14	13.28
5	0.41	0.55	0.83	1.15	1.61	9.24	11.07	12.83	15.09
6	0.68	0.87	1.24	1.64	2.20	10.64	12.59	14.45	16.81
7	0.99	1.24	1.69	2.17	2.83	12.02	14.07	16.01	18.48
8	1.34	1.65	2.18	2.73	3.49	13.36	15.51	17.53	20.09
9	1.73	2.09	2.70	3.33	4.17	14.68	16.92	19.02	21.67
10	2.16	2.56	3.25	3.94	4.87	15.99	18.31	20.48	23.21
11	2.60	3.05	3.82	4.57	5.58	17.28	19.68	21.92	24.72
12	3.07	3.57	4.40	5.23	6.30	18.55	21.03	23.34	26.22
13	3.57	4.11	5.01	5.89	7.04	19.81	22.36	24.74	27.69
14	4.07	4.66	5.63	6.57	7.79	21.06	23.68	26.12	29.14
15	4.60	5.23	6.26	7.26	8.55	22.31	25.00	27.49	30.58
16	5.14	5.81	6.91	7.96	9.31	23.54	26.30	28.85	32.00
17	5.70	6.41	7.56	8.67	10.09	24.77	27.59	30.19	33.41
18	6.26	7.01	8.23	9.39	10.86	25.99	28.87	31.53	34.81
19	6.84	7.63	8.91	10.12	11.65	27.20	30.14	32.85	36.19
20	7.43	8.26	9.59	10.85	12.44	28.41	31.41	34.17	37.57
22	8.64	9.54	10.98	12.34	14.04	30.81	33.92	36.78	40.29
24	9.89	10.86	12.40	13.85	15.66	33.20	36.42	39.36	42.98
26	11.16	12.20	13.84	15.38	17.29	35.56	38.89	41.92	45.64
28	12.46	13.56	15.31	16.93	18.94	37.92	41.34	44.46	48.28
30	13.79	14.95	16.79	18.49	20.60	40.26	43.77	46.98	50.89
32	15.13	16.36	18.29	20.07	22.27	42.58	46.19	49.48	53.49
34	16.50	17.79	19.81	21.66	23.95	44.90	48.60	51.97	56.06
38	19.29	20.69	22.88	24.88	27.34	49.51	53.38	56.90	61.16
42	22.14	23.65	26.00	28.14	30.77	54.09	58.12	61.78	66.21
46	25.04	26.66	29.16	31.44	34.22	58.64	62.83	66.62	71.20
50	27.99	29.71	32.36	34.76	37.69	63.17	67.50	71.42	76.15
55	31.73	33.57	36.40	38.96	42.06	68.80	73.31	77.38	82.29
60	35.53	37.48	40.48	43.19	46.46	74.40	79.08	83.30	88.38
65	39.38	41.44	44.60	47.45	50.88	79.97	84.82	89.18	94.42
70	43.28	45.44	48.76	51.74	55.33	85.53	90.53	95.02	100.43
75	47.21	49.48	52.94	56.05	59.79	91.06	96.22	100.84	106.39
80	51.17	53.54	57.15	60.39	64.28	96.58	101.88	106.63	112.33
85	55.17	57.63	61.39	64.75	68.78	102.08	107.52	112.39	118.24
90	59.20	61.75	65.65	69.13	73.29	107.57	113.15	118.14	124.12
95	63.25	65.90	69.92	73.52	77.82	113.04	118.75	123.86	129.97
100	67.33	70.06	74.22	77.93	82.36	118.50	124.34	129.56	135.81

Baye's Theorem

→ Probability Is The share of success / # Possible outcomes.

The Rules of Probability are :-

→ For any Event, $0 \leq P(A) \leq 1$

→ Σ Prob. = 1.

→ $P(\bar{A}) = 1 - P(A)$.

→ $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

→ The Multiplication Rule for Prob. Is :-

$$P(A \cap B) = P(A) \times P(B/A) \quad \text{—— (1)}$$

$$P(B \cap A) = P(B) \times P(A/B) \quad \text{—— (2)}$$

Then, $P(A) \times P(B/A) = P(B) \times P(A/B)$

$$\boxed{P(A/B) = \frac{P(A) \times P(B/A)}{P(B)}}$$ मन रहो;

Q→ 10% of Patients in an Clinic have Liver disease. 5% of The Clinic Patient are Alcoholic. Among The Healed Patients, 7% are Alcoholic.

→ What Is The Prob. of Patients having liver disease Given That He Is Alcoholic?

→ Let, A = Liver disease & B = Patient Is Alcoholic

$P(A) = 0.10$, $P(B) = 0.05$, $P(B/A) = 0.07$

$$\therefore P(A/B) = \frac{P(A) * P(B/A)}{P(B)} = \frac{0.1 \times 0.\bar{0}7}{0.05} = 0.14 = 14\%$$

∴ Thus, 14% Chance That Patient detected with Liver disease Was Alcoholic.

PAGE NO.:

F - Distribution : (Fisher - Snedecor dist^n)

- The F - Dist^n is an Probability dist^n that is been Useful in Context of Comparing Variances of Two or more Samples.
- It Is Right Skewed and Takes Only Non - Negative Values.
- The F - dist^n with d_1 and d_2 degree of freedom Is The dist^n :-

$$X = \frac{S_1 / dof_1}{S_2 / dof_2}$$

- F statistics $F = \frac{S_1^2}{S_2^2}$ (Variance Ratio Test) (No Two Tailed Test)

⊙ Where, S_1 and S_2 are Sample Std. deviation.

θ→ The Following data Is about The # Bulbs Produced daily by 2 Workers :- ($\alpha = 0.05$)

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

⊙ Can We Conclude Based on 'B' That Worker B Is more Stable & Effective ?

⊛ Mean Can't be used for This Test Because mean Is Almost Same for Both Samples. Thus, we will Compare Variance.

PAGE NO.:

① Frame The Hypothesis → $H_0 : S_1^2 = S_2^2$

$H_A : S_1^2 \neq S_2^2$

② Level of Sig (α) = 0.05

③ Type of Test : F - Test

④ Firstly Compute S_1^2 and S_2^2

① For S_1 →

X_1	$\bar{X}_1$	$(X_1 - \bar{X}_1)^2$
40	37	9
30	37	49
38	37	1
41	37	16
38	37	1
35	37	4
		80

$\therefore S_1^2 = \frac{\Sigma (X_1 - \bar{X}_1)^2}{n-1} = \frac{80}{6-1} = 16$

② For S_2 →

X_2	$\bar{X}_2$	$(X_2 - \bar{X}_2)^2$
39	37	4
38	37	1
41	37	16
33	37	16
32	37	25
39	37	4
40	37	9
34	37	9
		84

$\therefore S_2^2 = \frac{\Sigma (X_2 - \bar{X}_2)^2}{n-1} = \frac{84}{8-1} = 12$

$$\therefore \quad F_{stat.} = \frac{S_1^2}{S_2^2} = \frac{16}{12} = \underline{1.33}$$

(5) $F_{critical}$, $\alpha = 0.05$, $dof_1 = n_A - 1 = 6 - 1 = 5$

$dof_2 = n_B - 1 = 8 - 1 = 7$

$\therefore$ Thus, By F - Table,

$F_{critical}\ (\alpha = 0.05, dof = (5, 7)) = \underline{3.97}$

$\therefore$ Thus, We Fail To Reject The Null Hypothesis.

$\therefore$ Worker B Is Not more Effective Than Worker A.

F distribution critical value landmarks

Table entries are critical values for F^* with probably p in right tail of the distribution.

Figure of *F* distribution (like in Moore, 2004, p. 656) here.

Degrees of freedom in denominator (df2)		Degrees of freedom in numerator (df1)										
	p	1	2	3	4	5	6	7	8	12	24	1000
1	0.100	39.86	49.50	53.59	55.83	57.24	58.20	58.91	59.44	60.71	62.00	63.30
	0.050	161.4	199.5	215.7	224.6	230.2	234.0	236.8	238.9	243.9	249.1	254.2
	0.025	647.8	799.5	864.2	899.6	921.8	937.1	948.2	956.6	976.7	997.3	1017.8
	0.010	4052	4999	5404	5624	5764	5859	5928	5981	6107	6234	6363
	0.001	405312	499725	540257	562668	576496	586033	593185	597954	610352	623703	636101
2	0.100	8.53	9.00	9.16	9.24	9.29	9.33	9.35	9.37	9.41	9.45	9.49
	0.050	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.41	19.45	19.49
	0.025	38.51	39.00	39.17	39.25	39.30	39.33	39.36	39.37	39.41	39.46	39.50
	0.010	98.50	99.00	99.16	99.25	99.30	99.33	99.36	99.38	99.42	99.46	99.50
	0.001	998.38	998.84	999.31	999.31	999.31	999.31	999.31	999.31	999.31	999.31	999.31
3	0.100	5.54	5.46	5.39	5.34	5.31	5.28	5.27	5.25	5.22	5.18	5.13
	0.050	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.74	8.64	8.53
	0.025	17.44	16.04	15.44	15.10	14.88	14.73	14.62	14.54	14.34	14.12	13.91
	0.010	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.05	26.60	26.14
	0.001	167.06	148.49	141.10	137.08	134.58	132.83	131.61	130.62	128.32	125.93	123.52
4	0.100	4.54	4.32	4.19	4.11	4.05	4.01	3.98	3.95	3.90	3.83	3.76
	0.050	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	5.91	5.77	5.63
	0.025	12.22	10.65	9.98	9.60	9.36	9.20	9.07	8.98	8.75	8.51	8.26
	0.010	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.37	13.93	13.47
	0.001	74.13	61.25	56.17	53.43	51.72	50.52	49.65	49.00	47.41	45.77	44.09
5	0.100	4.06	3.78	3.62	3.52	3.45	3.40	3.37	3.34	3.27	3.19	3.11
	0.050	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.68	4.53	4.37
	0.025	10.01	8.43	7.76	7.39	7.15	6.98	6.85	6.76	6.52	6.28	6.02
	0.010	16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29	9.89	9.47	9.03
	0.001	47.18	37.12	33.20	31.08	29.75	28.83	28.17	27.65	26.42	25.13	23.82
6	0.100	3.78	3.46	3.29	3.18	3.11	3.05	3.01	2.98	2.90	2.82	2.72
	0.050	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.00	3.84	3.67
	0.025	8.81	7.26	6.60	6.23	5.99	5.82	5.70	5.60	5.37	5.12	4.86
	0.010	13.75	10.92	9.78	9.15	8.75	8.47	8.26	8.10	7.72	7.31	6.89
	0.001	35.51	27.00	23.71	21.92	20.80	20.03	19.46	19.03	17.99	16.90	15.77
7	0.100	3.59	3.26	3.07	2.96	2.88	2.83	2.78	2.75	2.67	2.58	2.47
	0.050	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.57	3.41	3.23
	0.025	8.07	6.54	5.89	5.52	5.29	5.12	4.99	4.90	4.67	4.41	4.15
	0.010	12.25	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.47	6.07	5.66
	0.001	29.25	21.69	18.77	17.20	16.21	15.52	15.02	14.63	13.71	12.73	11.72
8	0.100	3.46	3.11	2.92	2.81	2.73	2.67	2.62	2.59	2.50	2.40	2.30
	0.050	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.28	3.12	2.93
	0.025	7.57	6.06	5.42	5.05	4.82	4.65	4.53	4.43	4.20	3.95	3.68
	0.010	11.26	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.67	5.28	4.87
	0.001	25.41	18.49	15.83	14.39	13.48	12.86	12.40	12.05	11.19	10.30	9.36
9	0.100	3.36	3.01	2.81	2.69	2.61	2.55	2.51	2.47	2.38	2.28	2.16
	0.050	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.07	2.90	2.71
	0.025	7.21	5.71	5.08	4.72	4.48	4.32	4.20	4.10	3.87	3.61	3.34
	0.010	10.56	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.11	4.73	4.32
	0.001	22.86	16.39	13.90	12.56	11.71	11.13	10.70	10.37	9.57	8.72	7.84

Critical values computed with Excel 9.0

Degrees of freedom in denominator (df2)		Degrees of freedom in numerator (df1)										
	p	1	2	3	4	5	6	7	8	12	24	1000
10	0.100	3.29	2.92	2.73	2.61	2.52	2.46	2.41	2.38	2.28	2.18	2.06
	0.050	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	2.91	2.74	2.54
	0.025	6.94	5.46	4.83	4.47	4.24	4.07	3.95	3.85	3.62	3.37	3.09
	0.010	10.04	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.71	4.33	3.92
	0.001	21.04	14.90	12.55	11.28	10.48	9.93	9.52	9.20	8.45	7.64	6.78
12	0.100	3.18	2.81	2.61	2.48	2.39	2.33	2.28	2.24	2.15	2.04	1.91
	0.050	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.69	2.51	2.30
	0.025	6.55	5.10	4.47	4.12	3.89	3.73	3.61	3.51	3.28	3.02	2.73
	0.010	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.16	3.78	3.37
	0.001	18.64	12.97	10.80	9.63	8.89	8.38	8.00	7.71	7.00	6.25	5.44
14	0.100	3.10	2.73	2.52	2.39	2.31	2.24	2.19	2.15	2.05	1.94	1.80
	0.050	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.53	2.35	2.14
	0.025	6.30	4.86	4.24	3.89	3.66	3.50	3.38	3.29	3.05	2.79	2.50
	0.010	8.86	6.51	5.56	5.04	4.69	4.46	4.28	4.14	3.80	3.43	3.02
	0.001	17.14	11.78	9.73	8.62	7.92	7.44	7.08	6.80	6.13	5.41	4.62
16	0.100	3.05	2.67	2.46	2.33	2.24	2.18	2.13	2.09	1.99	1.87	1.72
	0.050	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.42	2.24	2.02
	0.025	6.12	4.69	4.08	3.73	3.50	3.34	3.22	3.12	2.89	2.63	2.32
	0.010	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.55	3.18	2.76
	0.001	16.12	10.97	9.01	7.94	7.27	6.80	6.46	6.20	5.55	4.85	4.08
18	0.100	3.01	2.62	2.42	2.29	2.20	2.13	2.08	2.04	1.93	1.81	1.66
	0.050	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.34	2.15	1.92
	0.025	5.98	4.56	3.95	3.61	3.38	3.22	3.10	3.01	2.77	2.50	2.20
	0.010	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.37	3.00	2.58
	0.001	15.38	10.39	8.49	7.46	6.81	6.35	6.02	5.76	5.13	4.45	3.69
20	0.100	2.97	2.59	2.38	2.25	2.16	2.09	2.04	2.00	1.89	1.77	1.61
	0.050	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.28	2.08	1.85
	0.025	5.87	4.46	3.86	3.51	3.29	3.13	3.01	2.91	2.68	2.41	2.09
	0.010	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.23	2.86	2.43
	0.001	14.82	9.95	8.10	7.10	6.46	6.02	5.69	5.44	4.82	4.15	3.40
30	0.100	2.88	2.49	2.28	2.14	2.05	1.98	1.93	1.88	1.77	1.64	1.46
	0.050	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.09	1.89	1.63
	0.025	5.57	4.18	3.59	3.25	3.03	2.87	2.75	2.65	2.41	2.14	1.80
	0.010	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	2.84	2.47	2.02
	0.001	13.29	8.77	7.05	6.12	5.53	5.12	4.82	4.58	4.00	3.36	2.61
50	0.100	2.81	2.41	2.20	2.06	1.97	1.90	1.84	1.80	1.68	1.54	1.33
	0.050	4.03	3.18	2.79	2.56	2.40	2.29	2.20	2.13	1.95	1.74	1.45
	0.025	5.34	3.97	3.39	3.05	2.83	2.67	2.55	2.46	2.22	1.93	1.56
	0.010	7.17	5.06	4.20	3.72	3.41	3.19	3.02	2.89	2.56	2.18	1.70
	0.001	12.22	7.96	6.34	5.46	4.90	4.51	4.22	4.00	3.44	2.82	2.05
100	0.100	2.76	2.36	2.14	2.00	1.91	1.83	1.78	1.73	1.61	1.46	1.22
	0.050	3.94	3.09	2.70	2.46	2.31	2.19	2.10	2.03	1.85	1.63	1.30
	0.025	5.18	3.83	3.25	2.92	2.70	2.54	2.42	2.32	2.08	1.78	1.36
	0.010	6.90	4.82	3.98	3.51	3.21	2.99	2.82	2.69	2.37	1.98	1.45
	0.001	11.50	7.41	5.86	5.02	4.48	4.11	3.83	3.61	3.07	2.46	1.64
1000	0.100	2.71	2.31	2.09	1.95	1.85	1.78	1.72	1.68	1.55	1.39	1.08
	0.050	3.85	3.00	2.61	2.38	2.22	2.11	2.02	1.95	1.76	1.53	1.11
	0.025	5.04	3.70	3.13	2.80	2.58	2.42	2.30	2.20	1.96	1.65	1.13
	0.010	6.66	4.63	3.80	3.34	3.04	2.82	2.66	2.53	2.20	1.81	1.16
	0.001	10.89	6.96	5.46	4.65	4.14	3.78	3.51	3.30	2.77	2.16	1.22

Use StaTable, WinPepi > WhatIs, or other reliable software to determine specific *p* values

⇒ ANOVA (Analysis of Variance) & It's Assumptions

→ ANOVA Is an Statistical method used To Compare The means of 2+ Groups.

⊙ Eg:- Effect of 0mg, 10mg & 100mg of medicine.

Assumptions of ANOVA

→ The Population from which Samples are drawn should be normally distributed.

→ The Sample should be Independent from Each other.

→ There Should be The Absense of The outlies.

→ Homogeneity of Variance : Homogeneity means That The Variance among The Groups should be Approximately Equal.

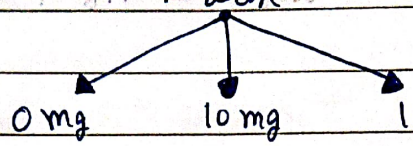
Eg $S_1, S_2, S_3 \Rightarrow S_1^2 \equiv S_2^2 \equiv S_3^2$

PAGE NO.:

Types of ANOVA

1. One Way ANOVA → One factor with Atleast Two levels and The levels are been Independent.

⊙ Eg:-



(dosage)

2. Repeated Measures ANOVA → One Factor with Atleast Two Levels but The Levels are different.

⊙ Eg:-

	day 1	day 2	day 3	day 4
GYM	2 hrs	1 hrs	½ hrs	1½ hrs

(my day workout depend on prev. day)

3. Factorial ANOVA → Two or more factors (with 2+ Levels). Levels Can be Either dependent, Independent of Both.

→ It Is Also Known as 2 Way Anova as It deals with 2 factors.

ANOVA Test

→ Hypothesis Testing In ANOVA (Partitioning of Variance In ANOVA)

$H_0 : \mu_1 = \mu_2 = \cdots = \mu_k$

H_A : Atleast One of The Sample mean Is Not Equal

→ Test Statistics = $\frac{\text{Variance b/w samples}}{\text{Variance within samples}}$

Q → There are 3 dosage of an medicine Given To Three Samples of The Patients. Rating Is done If medicine Works (1 - 10), Are There any differences In The 3 Cond^n ($\alpha = 0.05$)

(S_1)	(S_2)	(S_3)
0 mg	10 mg	100 mg
9	7	4
8	6	3
7	6	2
8	7	3
8	8	4
9	7	3
8	6	2
57	47	21

① → Exame The Hypothesis : $H_0 \longrightarrow \mu_{S_1} = \mu_{S_2} = \mu_{S_3}$

$H_A \longrightarrow$ Any of The mean Is Not Equal

② Level of Sig. (α) = 0.05 (One Tailed Test)

(No Two - Tailed Present In ANOVA - Test)

③ Type of Test : ANOVA - Test

④ F statistic = $\frac{\text{Variance b/w Samples}}{\text{Variance within Samples}}$

∴

	Sum of Squares (SS)	dof	SS/dof	F
B/W Samples	~~~	~~~	~~~	—
within Samples	~~~	~~~	~~~	—
Total	~~~	~~~	~~~	—

(i) Sum of Squares within Groups / Samples

$$SS_{within} = \sum y^2 - \sum \frac{(\sum a_i)^2}{n}$$

$$SS_{within} = (9^2+8^2+7^2+\ldots+7^2+6^2+6^2+\ldots+4^2+3^2+2^2+\ldots+3^2+2^2) - \left(\frac{57^2+47^2+21^2}{7}\right)$$

$$\boxed{SS_{within} = 10.29} \quad \text{——} (1)$$

(ii) Sum of Squares B/W Samples

$$SS_{B/W} = \sum \frac{(\sum a_i)^2}{n} - \frac{T^2}{N} \qquad (T \equiv \text{Total Sum})$$

$$SS_{B/W} = \frac{(57^2+47^2+21^2)}{7} - \frac{(57+47+21)^2}{21}$$

$$\boxed{SS_{B/W} = 98.67} \quad \text{——} (2)$$

(iii) dof B/W Samples = (# Samples) − 1 = 3 − 1 = 2 —— (3)

(iv) dof within samples = (# Total Elements) − (# Samples) = 21 − 3 = 18 —— (4)

∴

	Sum of Sq (SS)	dof	MS (SS/dof)
B/W The Samples	98.67	2	49.34
within Samples	10.29	18	0.54
	108.96	20	

∴ $F_{stat} = \frac{MS_{Between}}{MS_{within}} = \frac{49.34}{0.54} = 86.56$

(5) $F_{critical}$ for $\alpha = 0.05$, $dof_{B/W} = (\#Samples) - 1 = 3 - 1 = 2$

$dof_{within} = N - (\#Samples) = 21 - 3 = 18$

∴ By F-Table, $F_{critical}(\alpha = 0.05, dof = (2, 18)) = 3.55$

∴ As, $F_{stat.} = 86.56 >> 3.55$

∴ Thus, Reject The Null Hypothesis.

∴ Thus, Any of The mean Is Not Equal.

— JAI LOGIC