**Learning Objective (Short Explanation)**

Is chapter me aap **Statistics aur Probability** ke 3 main topics seekhoge:

1️⃣ **Descriptive Statistics:** Data ko summarize aur analyze karna using **Mean, Median, Mode, Variance, Standard Deviation, aur Correlation**.  
✅ Example: Class ke students ke marks ka average aur variation nikalna.

2️⃣ **Inferential Statistics:** Sample data se **poori population ka prediction** karna using **Sampling, Hypothesis Testing, aur Confidence Intervals**.  
✅ Example: 1000 logon ke salary data se poori city ki average salary predict karna.

3️⃣ **Probability:** Kisi event ke hone ke **chances** calculate karna using **Probability Rules, Distributions, aur Conditional Probability**.  
✅ Example: Toss karne pe Head aane ka chance 50% hai.

**📊 Statistics: Data ko Collect, Analyze aur Interpret karna**

👉 **Kya hota hai?**  
Statistics ek technique hai jo **data ko collect, analyze aur interpret** karne me madad karti hai. Matlab, hume agar **bahut saara data diya ho**, toh Statistics ki madad se hum **usko summarize karke patterns ya trends** nikal sakte hain.

👉 **Key Points:**

* **Data Collection:** Pehle hume data ikattha karna hota hai. (Survey, Experiment ya Observation se)
* **Data Analysis:** Fir hum us data ka analysis karte hain taaki samajh sake ki **trend kya hai?**
* **Interpretation:** Finally, hum data ko interpret karke **decision-making** me use karte hain.

👉 **Real-World Example:**  
Maan lo ek **company ko apne customers ka feedback analyze karna hai**.

* Wo **1000 customers ka survey karti hai**
* Data ko analyze karke dekhti hai ki **80% log service se khush hain**
* Iska matlab hai ki company ka performance acha hai, aur bas **20% customers ke concerns pe kaam karna padega**

📌 **Conclusion:** Statistics ka kaam **data-driven decisions** lene me help karna hai.

**🎲 Probability: Kisi Event ke Hone ke Chances**

👉 **Kya hota hai?**  
Probability ka matlab hai **kisi event ke hone ke chances nikalna**. Ye uncertainty ko quantify karta hai aur predict karta hai ki **kisi random scenario me kya outcome ho sakta hai?**

👉 **Key Points:**

* Probability **0 se 1 ke beech hoti hai**, jisme:
  + **0 ka matlab** - Event kabhi nahi hoga.
  + **1 ka matlab** - Event pakka hoga.
  + **0.5 ka matlab** - Event hone aur na hone ka chance equal hai.

👉 **Real-World Example:**  
Maan lo **ek coin toss kar rahe ho**:

* **Head aane ka chance = 50% (0.5 probability)**
* **Tail aane ka chance = 50% (0.5 probability)**

Agar hum **cricket match me predict karna chahein ki ek team jeetegi ya nahi**, toh hum probability ka use karke uska **winning percentage** nikal sakte hain.

📌 **Conclusion:** Probability ka use **random events aur uncertainty ko model karne** ke liye hota hai, jaise **weather forecast, stock market predictions, aur AI models**. 🚀

Descriptive Statistics ka main kaam **data ko summarize aur analyze** karna hota hai taaki hum usme **patterns aur trends** samajh sakein.

**1️⃣ Introduction**

📌 **Definition:**  
Descriptive Statistics ka matlab hai **data ko summarize, organize aur represent** karna taaki hum usko easily samajh sakein.

📌 **Importance:**

* **Data ko simple aur meaningful banata hai**
* **Patterns aur trends identify karne me help karta hai**
* **Decision making easy ho jati hai**

✅ **Example:**  
Ek school ke students ke marks ka data agar raw format me ho, toh usko analyze karna mushkil hoga. **Descriptive Statistics ki madad se hum average marks (Mean), beech ka value (Median), aur sabse common marks (Mode) nikal sakte hain.**

**2️⃣ Measures of Central Tendency**

Yeh **data ke center ya middle value** ko represent karta hai.

📌 **Mean (Average):**  
Total values ka sum leke number of values se divide karna.  
✅ **Example:** Agar kisi class ke students ke marks hain:

* **40, 50, 60, 70, 80**
* Mean = (40+50+60+70+80) / 5 = **60**

📌 **Median (Middle Value):**  
Agar values ko ascending order me rakhein toh **jo value beech me hogi, wahi median hoga.**  
✅ **Example:** **10, 20, 30, 40, 50** ka median **30** hoga.

📌 **Mode (Most Frequent Value):**  
Jo value **sabse zyada baar repeat hoti hai**, wahi Mode hoti hai.  
✅ **Example:** **2, 3, 3, 3, 5, 6, 7** ka mode **3** hai kyunki yeh sabse zyada repeat ho raha hai.

**3️⃣ Measures of Dispersion**

Yeh **data ke spread aur variability** ko batata hai.

📌 **Range:**  
Maximum aur Minimum value ka difference.  
✅ **Example:** **Marks: 10, 20, 30, 80, 90**

* Range = **90 - 10 = 80**

📌 **Variance:**  
Data values aur mean ke beech **kitna variation hai**, yeh batata hai.

📌 **Standard Deviation:**  
Variance ka **square root** hota hai, jo batata hai ki **data mean ke around kitna spread hai**.  
✅ **Example:** Agar ek test me students ke marks **50, 55, 60, 65, 70** hain, toh **Standard Deviation low hoga** kyunki sabhi values ek doosre ke paas hain.  
Agar marks **10, 20, 30, 80, 90** hain, toh **Standard Deviation high hoga** kyunki values ka spread zyada hai.

**4️⃣ Correlation and Covariance**

Yeh **do variables ke beech relationship** ko measure karta hai.

📌 **Covariance:**  
Agar **do variables ek saath increase ya decrease ho rahe hain**, toh **positive covariance** hoti hai. Agar ek badhta hai aur dusra ghatta hai, toh **negative covariance** hoti hai.

📌 **Correlation Coefficient:**  
Yeh **-1 se +1 ke beech hota hai** aur batata hai ki do variables kitne strongly related hain.  
✅ **Example:**

* **Height aur Weight:** Agar kisi ka height zyada hai toh weight bhi zyada ho sakta hai → **Positive Correlation**
* **Temperature aur Jacket Sales:** Temperature badhne se log jackets kam kharidenge → **Negative Correlation**

📌 **Conclusion:**  
Descriptive Statistics data ko **summarize aur analyze karne** me help karta hai taaki hum **better decisions** le sakein. 🚀

Descriptive Statistics ka kaam **data ko summarize karna** hota hai taaki bada data set **easy aur samajhne layak format me convert ho sake**. Yeh sirf data ko describe karta hai, **koi prediction ya conclusion nahi nikalta**.

**1️⃣ Summarizes Data**

👉 **Kya matlab hai?**  
Descriptive Statistics ka use **large data sets ko chhoti aur simple form me present** karne ke liye hota hai. Isse hume **data me patterns aur trends** samajhne me madad milti hai.

✅ **Example:**  
Maan lo ek school me **1000 students ke marks available hain**. Agar hume har ek student ka mark individually dekhna pade, toh analysis difficult ho jayega.

* Descriptive Statistics ki help se hum **Average Marks (Mean), Beech ka Value (Median), aur Most Frequent Marks (Mode)** nikal sakte hain.
* Isse hume ek **overall idea mil jata hai ki students ka performance kaisa hai**.

**2️⃣ Measures of Central Tendency**

👉 **Kya matlab hai?**  
Yeh **data ka central value ya representative value** batata hai. Iske 3 important measures hote hain:

1. **Mean (Average)** – Sabhi values ka sum lekar total numbers se divide karte hain.
2. **Median (Middle Value)** – Data ko ascending order me rakhkar beech ki value lete hain.
3. **Mode (Most Frequent Value)** – Jo value sabse zyada repeat hoti hai, wahi Mode hoti hai.

✅ **Example:**  
Maan lo kisi class ke students ke marks hain: **40, 50, 60, 70, 80**

* **Mean** = (40+50+60+70+80) / 5 = **60**
* **Median** = **60** (middle value)
* **Mode** = **Agar ek value zyada baar repeat hoti, toh wahi mode hoti**

**3️⃣ Measures of Dispersion**

👉 **Kya matlab hai?**  
Yeh **batata hai ki data kitna spread out hai**. Agar values ek dusre ke paas hain toh spread **kam** hoga, aur agar values bahut alag hain toh spread **zyada** hoga.

🟢 **Important Measures:**

* **Range** = Maximum - Minimum value
* **Variance** = Mean se data ka deviation measure karta hai
* **Standard Deviation** = Variance ka square root, jo batata hai ki values mean ke around kitni spread hain

✅ **Example:**  
Marks = **10, 20, 30, 80, 90**

* **Range** = 90 - 10 = **80**
* **Standard Deviation high hoga**, kyunki values zyada spread hain

Agar marks **45, 48, 50, 52, 55** hote, toh spread **kam hota** aur standard deviation bhi **kam hota**.

**4️⃣ Data Visualization**

👉 **Kya matlab hai?**  
Data ko **graph aur charts** ke through represent karna taaki analysis easy ho sake.

🟢 **Common Charts:**

* **Histogram** – Data ka distribution dikhata hai
* **Pie Chart** – Percentage representation ke liye
* **Box Plot** – Outliers aur spread analysis ke liye

✅ **Example:**  
Agar ek company ke **sales data ko histogram me dikhaya jaye**, toh hume samajh aayega ki **kis month me zyada sales huyi aur kis month me kam**.

**5️⃣ No Inference (Sirf Summary, Koi Prediction Nahi)**

👉 **Kya matlab hai?**  
Descriptive Statistics sirf **data ko summarize karta hai**, lekin **koi future prediction nahi karta**.

✅ **Example:**  
Agar ek survey me pata chalta hai ki **80% log ek mobile brand ko pasand karte hain**, toh yeh sirf **past data ka analysis hai**.  
Yeh **nahi bata sakta ki future me bhi log yahi brand pasand karenge ya nahi**.

📌 **Conclusion:**  
Descriptive Statistics ka kaam **data ko summarize, visualize aur analyze karna** hota hai. Yeh **data ke baare me insights deta hai**, lekin **koi future prediction nahi karta**. 🚀

**Measures of Central Tendency** ka matlab hota hai **data ka ek central ya representative value** nikalna, jo **poore dataset ka overall trend** represent kare. Iska main purpose **data ko simplify karna** aur ek **single value se entire dataset ko summarize karna** hota hai.

**1️⃣ Central Tendency Kya Hai?**

👉 **Kya matlab hai?**  
Central Tendency ek aisa concept hai jo **data ke beech ka value dikhata hai** taaki hume uska **overall trend samajh aaye**. Iska kaam **data ko simplify karna aur ek aisi value dena hai jo poore dataset ka representative ho**.

✅ **Example:**  
Maan lo 5 students ke marks diye gaye hain:  
👉 **50, 60, 70, 80, 90**

* Agar hume poore dataset ka **ek representative value** chahiye, toh hum **Mean, Median ya Mode** ka use kar sakte hain.

**2️⃣ Central Tendency Ki Importance**

👉 **Kya matlab hai?**  
🔹 **Data ko summarize karta hai** – Ek hi value se poore dataset ka idea mil jata hai  
🔹 **Distribution aur general characteristics samajhne me madad karta hai** – Data ka pattern aur uska spread dikhata hai  
🔹 **Comparison easy ho jata hai** – Alag-alag datasets ke beech comparison karna simple ho jata hai

✅ **Example:**  
Agar kisi **school ki average attendance 85%** hai, toh yeh ek single value me school ki attendance ka overall trend batata hai.

**3️⃣ Common Measures of Central Tendency**

👉 **Kya matlab hai?**  
Central Tendency ke 3 main measures hote hain:

1. **Mean (Average)** – Sabhi values ka sum lekar total numbers se divide karte hain
2. **Median (Middle Value)** – Data ko ascending order me rakhkar beech ki value lete hain
3. **Mode (Most Frequent Value)** – Jo value sabse zyada repeat hoti hai

✅ **Example:**  
Agar students ke marks diye gaye hain: **40, 50, 60, 70, 80**

* **Mean** = (40+50+60+70+80) / 5 = **60**
* **Median** = **60** (Middle value)
* **Mode** = **Agar ek value zyada baar repeat hoti, toh wahi mode hoti**

💡 **Example for Mode:**  
Agar marks **30, 40, 40, 50, 60** hain, toh **Mode = 40** (kyunki 40 do baar aaya hai)

**4️⃣ Real-World Applications of Central Tendency**

👉 **Kahan Use Hota Hai?**  
**Central Tendency ka use kaafi saare fields me hota hai:**  
🔹 **Economics** – Average income ya GDP nikalne ke liye  
🔹 **Healthcare** – Patients ke average blood pressure ya sugar level analyze karne ke liye  
🔹 **Education** – Students ke average marks ya performance analyze karne ke liye

✅ **Example:**  
Agar ek **company ka average employee salary ₹50,000** hai, toh yeh ek single value me bata raha hai ki **mostly log ₹50,000 ke aaspaas kama rahe hain**.

📌 **Conclusion:**  
Measures of Central Tendency **ek dataset ka center ya typical value batate hain**, jo **data ko summarize karne aur samajhne me madad karta hai**. Yeh **real-world problems me decision-making ke liye useful hota hai**, jaise salary comparison, student performance analysis, aur business trends samajhna. 🚀

**📊 Mean – Detailed Explanation (Hinglish)**

**Mean** ek **sabse common measure of central tendency** hai, kyunki yeh **dataset ki sabhi values ko consider karta hai** aur ek **single representative value** nikalta hai.

**1️⃣ Mean Kya Hai?**

👉 **Kya matlab hai?**  
Mean ka matlab hota hai **sabhi values ka sum lekar total numbers se divide karna**. Isko **Arithmetic Average** bhi bolte hain.

✅ **Formula:**

Mean=∑ValuesTotal Number of Values\text{Mean} = \frac{\sum \text{Values}}{\text{Total Number of Values}}Mean=Total Number of Values∑Values​

✅ **Example:**  
Agar 5 students ke marks diye gaye hain:  
👉 **50, 60, 70, 80, 90**  
Toh **Mean** = (50 + 60 + 70 + 80 + 90) ÷ 5  
👉 **Mean = 70**

**2️⃣ Daily Life Examples of Mean**

**Mean ka use bahut saare real-world situations me hota hai.** Jab bhi hum kisi data ka **"average"** nikalte hain, toh hum mean ka use kar rahe hote hain.

✅ **Example 1: School Marks**  
Agar ek school me students ke marks ka **average** nikalna ho, toh hum mean ka use karenge.

✅ **Example 2: Salary Calculation**  
Agar hume pata karna ho ki **India me software engineers ki average salary kitni hai**, toh hum different salaries ka **Mean** calculate karenge.

✅ **Example 3: Mobile Sales**  
Agar ek shop me **har mahine kitne mobile phones bik rahe hain**, iska **average nikalne ke liye Mean ka use hota hai**.

✅ **Example 4: Cricket Batting Average**  
Ek cricketer ne **5 matches me ye scores kiye: 40, 50, 60, 30, 70**  
Toh uska batting average (Mean Score) = (40+50+60+30+70) ÷ 5 = **50 runs per match**

**3️⃣ Why is Mean Important?**

👉 **Kya matlab hai?**  
🔹 **Ek representative value deta hai jo poore dataset ka trend dikhata hai**  
🔹 **Decision making me madad karta hai** (jaise business planning, salary estimation, etc.)  
🔹 **Data ka summary provide karta hai taaki hum easily compare kar sakein**

📌 **Conclusion:**  
Mean ek **bahut useful statistical measure hai** jo **poore dataset ka ek single representative value nikalta hai**. Yeh **daily life aur business decisions** me important role play karta hai! 🚀

**Mean – Formula & Examples (Hinglish)**

**📌 Mean Ka Formula:**

Mean ko calculate karne ka **standard formula** yeh hai:

xˉ=∑xin\bar{x} = \frac{\sum x\_i}{n}xˉ=n∑xi​​

🔹 **xˉ\bar{x}xˉ (Mean):** Dataset ka **average value**  
🔹 **∑xix\_ixi​ (Summation of Values):** Sabhi values ka **total sum**  
🔹 **nnn (Number of Values):** Dataset me **total values ki sankhya**

**🧮 Example 1: Average Marks of 5 Students**

👉 **Maan lo 5 students ke mathematics exam ke marks diye gaye hain:**  
🔹 45, 50, 55, 60, 70

✅ **Step 1: Sabhi marks ka sum nikalna**

45+50+55+60+70=28045 + 50 + 55 + 60 + 70 = 28045+50+55+60+70=280

✅ **Step 2: Total students ki sankhya n=5n = 5n=5**

✅ **Step 3: Mean ka formula lagana**

xˉ=2805=56\bar{x} = \frac{280}{5} = 56xˉ=5280​=56

👉 **Toh, in students ka average marks = 56** 🎯

**💡 Mean Kyu Zaroori Hai?**

✅ **Summarization:** Large dataset ko **ek single value me summarize** karne ke liye  
✅ **Comparison:** Alag-alag groups ka **comparison** karne ke liye  
✅ **Trend Analysis:** Kisi **pattern ya trend** ko samajhne ke liye

📌 **Conclusion:**  
Mean ek **bahut important statistical tool** hai jo **kisi bhi dataset ka representative average nikalta hai**. Yeh **education, finance, sports, business** jaise har field me use hota hai! 🚀

**📊 Example 2: Effect of an Outlier on the Mean (Hinglish)**

**📌 Pehle Wali Situation**

🔹 **5 students ke marks diye gaye hain:**  
👉 **45, 50, 55, 60, 70**

✅ **Mean (Average) pehle calculate kiya tha:**

xˉ=45+50+55+60+705=2805=56\bar{x} = \frac{45 + 50 + 55 + 60 + 70}{5} = \frac{280}{5} = 56xˉ=545+50+55+60+70​=5280​=56

👉 **Toh, pehle average marks = 56**

**📌 Ab Ek Naya Student Add Ho Raha Hai**

👉 **Ek naye student ke marks = 99**

✅ **Naya Mean Calculate Karein:**

xˉ=45+50+55+60+70+996\bar{x} = \frac{45 + 50 + 55 + 60 + 70 + 99}{6}xˉ=645+50+55+60+70+99​

✅ **Total Sum:**

∑xi=45+50+55+60+70+99=379\sum x\_i = 45 + 50 + 55 + 60 + 70 + 99 = 379∑xi​=45+50+55+60+70+99=379

✅ **New Mean:**

xˉ=3796=63.17\bar{x} = \frac{379}{6} = 63.17xˉ=6379​=63.17

👉 **Ab Average Marks = 63.17** 🔺 (Jo pehle wale 56 se kaafi zyada hai!)

**💡 Key Takeaway: Outliers Mean Ko Affect Karte Hain!**

🔹 **Outlier kya hai?** - Bahut zyada **chhoti ya badi** value jo dataset se alag dikh rahi ho.  
🔹 **Effect:** Ek single **outlier** bhi mean ko **kaafi increase ya decrease** kar sakta hai.  
🔹 **Problem:** Agar **data me extreme values hain**, toh **mean misleading ho sakta hai**!  
🔹 **Better Alternative:** **Aise cases me ‘median’ use karna better hota hai**, jo outliers se **zyada impact nahi leta**.

📌 **Conclusion:**  
Mean **outliers ke wajah se biased ho sakta hai**, isliye **agar data me extreme values hain, toh median ka use karna better hota hai**! 🚀

**Mean ke Properties (Gun) – Hinglish me Samjhaye**

1️⃣ **Sum of Deviations Zero**

* Agar hum har value me se **mean minus karein** aur sabko **add karein**, toh total **hamesha zero hota hai**.
* Formula: ∑(xi−xˉ)=0\sum (x\_i - \bar{x}) = 0∑(xi​−xˉ)=0
* Matlab **mean dataset ka perfect balancing point hota hai**. ⚖️

2️⃣ **Affected by Values Koi bhi value change** hone par mean bhi **change ho jata hai**.

* Har ek data point **mean ko influence karta hai**.

3️⃣ **Unique and Defined** Har dataset ka **sirf ek hi mean hota hai**.

* Mean **hamesha ek clear value deta hai**, jo **pure data ko represent karta hai**.

4️⃣ **Sensitive to Outliers**

* **Agar data me koi bahut bada ya chhota number aa jaye** (outlier), toh mean **distort ho sakta hai**.
* Is wajah se, kabhi kabhi **median zyada reliable hota hai**.

📌 **(Conclusion):**  
Mean ek **useful measure** hai, lekin **agar outliers zyada hain toh median better ho sakta hai**! 🚀

**Python ka Use karke Mean Calculate Karna**

Python me mean **bahut easily calculate** kiya ja sakta hai.

📌 **Step-by-Step Explanation:**  
✅ **statistics library ka use karke mean nikalna**  
✅ **mean() function se calculation karna**

**Python Code:**

python

Copy code

import statistics

# Students ke marks ka list

marks = [45, 50, 55, 60, 70]

# Mean calculate karna

mean\_marks = statistics.mean(marks)

# Result print karna

print("Mean Marks:", mean\_marks)

**Output:**

css

Copy code

Mean Marks: 56

**📌 Mean ka Summary:**

✅ **Sabse common central tendency measure hai**  
✅ **Formula:** (Sabhi values ka sum) ➗ (Total values)  
✅ **Extreme values se effect hota hai (outliers)**  
✅ **Weighted mean use hota hai jab kuch values ki importance zyada ho**  
✅ **Python ka use karke bada dataset easily handle kiya ja sakta hai** 🚀

**Median – Data ka Middle Value**

🔹 **Real-Life Examples:**

* Kisi company me **middle salary** kya hai?
* Mumbai me **houses ka middle price** kya hai?
* Students ke **exam score ka beech ka value** kya hai?

👉 **Median hamesha beech ka value hota hai**, jab data **ascending order me** arrange hota hai.  
👉 **Mean sabhi values ko consider karta hai**, lekin **Median sirf middle value dekhta hai**.  
👉 **Agar dataset me outliers (extreme values) ho**, toh **Median better choice hota hai**! 🚀

**Median Kaise Calculate Karein?**

👉 **Step 1:** Numbers ko **ascending order me** arrange karein.  
👉 **Step 2:** **Middle number find karein**:

* **Agar values ki count odd hai**, toh **middle value hi median hota hai**.
* **Agar values ki count even hai**, toh **do middle values ka average median hota hai**.

**📝 Example 1: Odd Set of Numbers**

🔹 5 students ki age: **12, 15, 14, 17, 13**  
🔹 **Step 1:** Arrange in **ascending order** → **12, 13, 14, 15, 17**  
🔹 **Step 2:** **Middle value = 14**  
✅ **Median = 14**

**📝 Example 2: Outlier Ka Effect**

🔹 **House Prices (₹ lakhs):** **50, 55, 60, 62, 500**

**📊 Mean Calculation:**

Mean=50+55+60+62+5005=145.4\text{Mean} = \frac{50 + 55 + 60 + 62 + 500}{5} = 145.4Mean=550+55+60+62+500​=145.4

✅ **Mean ₹145.4 lakh aaya, jo galat lag raha hai kyunki ₹500 lakh ek extreme outlier hai.**

**📊 Median Calculation:**  
🔹 **Sorted Data:** **50, 55, 60, 62, 500**  
🔹 **Middle value = 60**  
✅ **Median = ₹60 lakh (Jo zyada accurate representation hai!)**

📌 **Key Takeaway:**  
👉 **Median outliers se affect nahi hota** aur **actual central value ko accurately represent karta hai!** 🚀

**Median ki Properties (Gun) - Hinglish Me Samjhayein**

1️⃣ **Focus on Middle Values (Mukhya Mulya Par Dhyan)**  
🔹 Median **sirf beech wale value** ko dekhta hai, **bahar ke values ignore karta hai**.

2️⃣ **Resistance to Outliers (Extreme Values se Prabhavit Nahi Hota)**  
🔹 **Agar koi value bahut zyada chhoti ya badi ho (outlier), toh median par koi effect nahi hota**.

3️⃣ **Usefulness for Skewed Data (Tedi-Medi Data Me Upyogi)**  
🔹 **Jab data ek taraf jhukta hai (skewed data), toh mean accurate nahi hota**, lekin **median accurate hota hai**.

4️⃣ **Existence (Hamesha Maujood Hota Hai)**  
🔹 **Har dataset ka ek clear aur defined median hota hai**.

✅ **Key Takeaway:** Median **beech ki value ko dhyan me rakhta hai** aur **outliers se effect nahi hota**, isliye **zyada accurate measure hota hai!** 🚀

**Python Me Median Kaise Nikalein? (Hinglish Explanation)**

Python ka **statistics** module use karke **median** aasani se find kar sakte hain.

**📝 Example - Python Code for Median Calculation**

python

Copy code

import statistics

# Student ke scores ka dataset

scores = [12, 15, 14, 17, 13]

# Median find karna

median\_score = statistics.median(scores)

print("Median Score:", median\_score)

🔹 **Output:**

yaml

Copy code

Median Score: 14

**📌 Summary of Median (Sankshipt Saransh)**

✔ **Middle Value ka Concept:** Ordered data ka **beech wala value** median hota hai.

✔ **Outliers ka Asar Nahi Hota:** Agar dataset me **koi extreme values ho**, toh **median usse prabhavit nahi hota**.

✔ **Even Number of Values:** Agar dataset me **even values ho**, toh **dono middle values ka average median hota hai**.

✔ **Practical Use Cases:** **Economics, business, real estate, aur machine learning me median kaafi use hota hai.**

✅ **Conclusion:** Jab **data me bahut extreme values ho**, tab **mean ke bajaye median ka use zyada sahi hota hai!** 🚀

**📌 Mode Kya Hota Hai? (Hinglish Explanation)**

**Mode** ek **aisa value hota hai jo dataset me sabse zyada baar repeat hota hai**.

**Daily Life Examples:**  
📱 **Sabse popular mobile brand kaunsa hai?**  
👟 **Students ka sabse common shoe size kya hai?**  
💻 **Data Science me sabse zyada use hone wali programming language kaunsi hai?**

**📌 Mode ka Use**

✔ **Numerical aur Categorical Dono Data ke liye Useful:**

* **Numerical Data:** Sabse zyada baar aane wala number.
* **Categorical Data:** Sabse popular choice, jaise **favourite color, best-selling laptop, etc.**

✔ **Mean aur Median se Alag:**

* Mean aur median **average aur middle value** pe focus karte hain.
* Mode **sabse zyada baar repeat hone wale value** pe focus karta hai.

✅ **Conclusion:** Jab **humein sabse common value pata karni ho**, tab **mode ka use sabse best hota hai!** 🚀

**Mode Kaise Calculate Karein? (Hinglish Explanation)**

Mode find karne ke liye ye steps follow karein:

**🛠 Steps to Find Mode:**

🔹 **Step 1:** Har value ki frequency (kitni baar repeat ho raha hai) count karein.  
🔹 **Step 2:** Jo value(s) **sabse zyada baar repeat ho rahi hai**, wahi **Mode** hai.

**📌 Example 1: Simple Dataset ka Mode**

Maan lo ek test me students ke marks diye gaye hain:  
👉 **45, 50, 55, 50, 60, 50, 70, 75**

✅ Yaha **50** **3 baar** repeat ho raha hai.  
**📌 Mode = 50** (Sabse zyada baar repeat hone wala value)

**📌 Example 2: Bimodal Dataset (2 Modes)**

Agar dataset me **2 different values** barabar baar repeat ho rahi ho, toh **Bimodal** kehlata hai.

Maan lo students ki height (cm) di gayi hai:  
👉 **150, 160, 165, 150, 170, 175, 165, 180**

✅ **150** aur **165** dono **2 baar** repeat ho rahe hain.  
**📌 Modes = 150 & 165 (Bimodal)**

**📌 Example 3: Multimodal Dataset (3+ Modes)**

Agar **3 ya usse zyada values** equal frequency ke saath repeat ho rahi ho, toh **Multimodal** kehlata hai.

Maan lo ek survey me favourite colours collect kiye gaye hain:  
👉 **Red, Blue, Green, Blue, Yellow, Green, Blue, Red, Green, Red**

✅ **Green (3), Blue (3), aur Red (3)** sabse zyada baar aaye hain.  
**📌 Modes = Green, Blue & Red (Multimodal)**

✅ **Conclusion:**  
Mode **sabse zyada baar repeat hone wala value** hota hai aur agar ek se zyada modes ho, toh **Bimodal ya Multimodal** kehate hain. 🚀

Mode vs Mean vs Median

**📌 Python me Mode Kaise Calculate Karein? (Hinglish Explanation)**

Python me **statistics** module ka use karke **Mode** nikalna bohot easy hai. Chaliye example dekhte hain:

**🛠 Python Code for Mode Calculation**

python

Copy code

import statistics

numbers = [45, 50, 55, 60, 50, 70, 75]

mode\_value = statistics.mode(numbers)

print("Mode:", mode\_value)

🔹 **Output:**

makefile

Copy code

Mode: 50

✅ **Yaha 50 sabse zyada baar repeat ho raha hai, isliye Mode = 50 hai.**

**📌 Summary of the Mode**

✅ **Mode sabse zyada baar repeat hone wala value hota hai.**  
✅ Ek dataset me **1, 2, multiple, ya koi mode nahi** ho sakta.  
✅ **Categorical data** (jaise most popular brand) ke liye Mode useful hota hai.  
✅ **Business, fashion, IT, aur education** me Mode ka use hota hai.

🚀 **Mode ka calculation simple hai aur Python ise aur bhi easy bana deta hai!**

Lab

**📌 Measures of Dispersion (Hinglish Explanation)**

📌 **Measure of Dispersion** ka matlab hota hai **data ka spread** ya **variability** check karna. Yeh batata hai ki data **kitna scattered hai mean ya median ke around**.

**🔹 Kyu Zaroori Hai?**

✅ **Data consistency aur variability** samajhne ke liye use hota hai.  
✅ **Risk analysis aur future prediction** me kaam aata hai.  
✅ Yeh batata hai ki **data points mean ke pass clustered hain ya bohot dur dur hain**.

**📌 Types of Measures of Dispersion**

1️⃣ **Range** – Maximum aur minimum value ka difference.  
2️⃣ **Variance** – Data points mean se kitna deviate kar rahe hain.  
3️⃣ **Standard Deviation (SD)** – Variance ka square root, jo data spread batata hai.  
4️⃣ **Interquartile Range (IQR)** – Middle 50% data ka spread.

**Range (Hinglish Explanation)**

📌 **Range** ek **simplest measure of dispersion** hai, jo **data ka spread** check karta hai. Yeh **highest aur lowest values ka difference** batata hai.

**🔹 Range ka Formula**

Range=Maximum Value−Minimum ValueRange = Maximum\ Value - Minimum\ ValueRange=Maximum Value−Minimum Value

**🔹 Example 1: Simple Range Calculation**

Agar ek classroom me **5 students** ki ages di gayi hain:

✅ **Maximum age** = 22  
✅ **Minimum age** = 12

👉 **Range = 22 - 12 = 10**

🔹 **Iska Matlab?**  
Students ki **ages ka difference 10 years hai**, yaani **data 10 units tak spread hai**.

**📌 Range ki Limitations**

❌ **Outliers** pe sensitive hota hai (Agar ek bohot chhoti ya bohot badi value ho to range bohot badh sakti hai).  
❌ **Data distribution ka proper idea nahi deta**, sirf spread batata hai.

**Variance (Hinglish Explanation)**

📌 **Variance** ek **important measure of dispersion** hai jo **data ke spread** ko represent karta hai. Yeh batata hai ki **data points mean se kitne dur ya paas hai**.

🔹 **Range useful hoti hai, par variance data distribution ka aur accurate analysis provide karta hai**.

**🔹 Variance ka Formula**

𝝈^2 = \frac{\sum (𝒙\_𝒊 - 𝒙̅ )^2}{𝑵}

✅ **Jahaan:**

* **𝝈²** = Variance
* **𝒙\_𝒊** = Har ek data value
* **𝒙̅** = Mean (Average)
* **N** = Total number of values

**🔹 Variance Kaise Calculate Karein? (Step-by-Step)**

🔸 **Step 1:** Mean (𝒙̅) find karein.  
🔸 **Step 2:** Har ek data point ka **mean se difference nikaalein** (𝒙\_𝒊 - 𝒙̅).  
🔸 **Step 3:** In differences ka **square karein** taaki negative values avoid ho sakein.  
🔸 **Step 4:** In squared values ka **sum karein**.  
🔸 **Step 5:** Final result ko **number of values (N) se divide karein**.

**🔹 Example: Variance Calculation**

Agar ek dataset diya gaya hai:  
**5, 7, 9, 12, 15**

**Step 1:** Mean Calculate karein:

𝒙̅ = \frac{(5+7+9+12+15)}{5} = \frac{48}{5} = 9.6

**Step 2:** Mean se difference nikaalein:

| **Data (𝒙\_𝒊)** | **(𝒙\_𝒊 - 𝒙̅)** | **(𝒙\_𝒊 - 𝒙̅)²** |
| --- | --- | --- |
| 5 | (5 - 9.6) = -4.6 | 21.16 |
| 7 | (7 - 9.6) = -2.6 | 6.76 |
| 9 | (9 - 9.6) = -0.6 | 0.36 |
| 12 | (12 - 9.6) = 2.4 | 5.76 |
| 15 | (15 - 9.6) = 5.4 | 29.16 |

**Step 3:** Squared differences ka sum:

21.16+6.76+0.36+5.76+29.16=63.221.16 + 6.76 + 0.36 + 5.76 + 29.16 = 63.221.16+6.76+0.36+5.76+29.16=63.2

**Step 4:** Total values (N) se divide karein:

𝝈2=63.25=12.64𝝈^2 = \frac{63.2}{5} = 12.64𝝈2=563.2​=12.64

✅ **Variance = 12.64**

**📌 Variance ke Key Points**

✅ **Higher Variance → Data zyada spread hai**  
✅ **Lower Variance → Data mean ke aas paas concentrated hai**  
✅ **Variance ka square root le kar Standard Deviation nikaalte hain** (jo zyada readable hota hai)

👉 **Agla topic chahiye? Standard Deviation ya aur examples?** 😃

**📌 Example 1: Variance Calculation (Simple Example in Hinglish)**

Agar humare paas **marks dataset** diya gaya hai:  
**10, 20, 30, 40, 50**

**🔹 Step 1: Calculate the Mean**

𝒙̅ = \frac{10 + 20 + 30 + 40 + 50}{5} = \frac{150}{5} = 30

**🔹 Step 2: Find the Squared Differences**

Har ek data point ka mean se difference aur uska square calculate karein:

| **Marks (𝒙\_𝒊)** | **(𝒙\_𝒊 - 𝒙̅)** | **(𝒙\_𝒊 - 𝒙̅)²** |
| --- | --- | --- |
| 10 | (10 - 30) = -20 | 400 |
| 20 | (20 - 30) = -10 | 100 |
| 30 | (30 - 30) = 0 | 0 |
| 40 | (40 - 30) = 10 | 100 |
| 50 | (50 - 30) = 20 | 400 |

**🔹 Step 3: Calculate Variance**

𝝈2=(400+100+0+100+400)5𝝈^2 = \frac{(400 + 100 + 0 + 100 + 400)}{5}𝝈2=5(400+100+0+100+400)​ 𝝈2=10005=200𝝈^2 = \frac{1000}{5} = 200𝝈2=51000​=200

✅ **Variance = 200**

**📌 Key Takeaways**

✔️ **Agar variance high hai (200), toh iska matlab marks zyada spread hain.**  
✔️ **Agar variance low hota, toh iska matlab marks ek common range ke andar hote.**  
✔️ **Variance ka square root le kar Standard Deviation nikalte hain jo data spread ko aur clearly represent karta hai.**

**Variance Calculation Using Python**

Python mein variance calculate karna bahut easy hai using **NumPy** library.

**🔹 Code Example**

python

Copy code

import numpy as np

# Marks dataset

marks = [10, 20, 30, 40, 50]

# Variance Calculation

variance\_value = np.var(marks)

# Output

print("Variance of Marks:", variance\_value)

**🔹 Output**

css

Copy code

Variance of Marks: 200.0

**📌 Explanation**

✔️ **np.var() function** variance calculate karne ke liye use hoti hai.  
✔️ **List ke values mean se kitni deviated hain, uska square ka average nikalta hai.**  
✔️ **Variance high hone ka matlab data widely spread hai, aur low hone ka matlab data clustered hai.**

**📌 Standard Deviation (SD) Calculation**

Standard Deviation ek important measure of dispersion hai jo data ke spread ka idea deta hai.  
Agar **Standard Deviation zyada hai**, to iska matlab data **widely spread** hai, aur agar **kam hai**, to data **clustered** hai.

**🔹 Formula for Standard Deviation**

σ=σ2\mathbf{\sigma = \sqrt{\sigma^2} }σ=σ2​

**Where:**  
✔ **σ (Standard Deviation)** = Variance ka square root  
✔ **σ² (Variance)** = Variance

**📌 Example Calculation**

Agar hamare paas dataset **10, 20, 30, 40, 50** hai,  
Toh **Variance = 200** (pichle example se).

**Standard Deviation Calculation:**

σ=200≈14.14\sigma = \sqrt{200} \approx 14.14σ=200​≈14.14

**📌 Python Code for Standard Deviation**

Python mein **numpy** library ka std() function use karke SD nikal sakte hain.

**🔹 Code Example**

python

Copy code

import numpy as np

# Marks dataset

marks = [10, 20, 30, 40, 50]

# Standard Deviation Calculation

std\_value = np.std(marks)

# Output

print("Standard Deviation of Marks:", std\_value)

**🔹 Output**

css

Copy code

Standard Deviation of Marks: 14.14

**📌 Summary**

✔ **Standard Deviation variance ka square root hota hai.**  
✔ **Agar SD zyada hai, toh values zyada scattered hain.**  
✔ **Python mein np.std() function se SD easily calculate hota hai.**

Lab

**📌 Covariance in Statistics**

Covariance **do variables ke beech ka relationship** batata hai, ki agar ek variable change ho raha hai to dusra **kaise behave karega**.

**🔹 Types of Covariance**

1. **📈 Positive Covariance**
   * Dono variables **saath-saath badhte** ya **kam hote** hain.
   * Example: **Study Time 📖 & Exam Scores 📊** (Jitna zyada study, utna achha score)
2. **📉 Negative Covariance**
   * Ek variable badhta hai, dusra **kam hota hai**.
   * Example: **Speed 🚗 & Travel Time ⏳** (Jitni zyada speed, utna kam time)
3. **➖ Zero Covariance**
   * Dono variables ke beech **koi relation nahi hota**.
   * Example: **Shoe Size 👟 & IQ Level 🧠** (Ek dusre se koi lena-dena nahi)

**📌 Covariance Formula**

Cov(X,Y)=∑(Xi−Xˉ)(Yi−Yˉ)N\mathbf{Cov(X, Y) = \frac{\sum (X\_i - \bar{X}) (Y\_i - \bar{Y})}{N}}Cov(X,Y)=N∑(Xi​−Xˉ)(Yi​−Yˉ)​

**Where:**  
✔ **X,YX, YX,Y** → Do variables  
✔ **Xi,YiX\_i, Y\_iXi​,Yi​** → Individual values  
✔ **Xˉ,Yˉ\bar{X}, \bar{Y}Xˉ,Yˉ** → Mean of X and Y  
✔ **NNN** → Number of values

**📌 Covariance Calculation Using Python**

Python mein **numpy** ya **statistics** library ka use karke covariance easily find kar sakte hain.

**🔹 Example: Covariance of Study Time & Exam Scores**

python

Copy code

import numpy as np

# Study Hours & Exam Scores

study\_time = [1, 2, 3, 4, 5]

exam\_scores = [50, 55, 65, 70, 80]

# Covariance Calculation

cov\_matrix = np.cov(study\_time, exam\_scores, bias=True)

covariance = cov\_matrix[0, 1] # Extract covariance value

# Output

print("Covariance:", covariance)

**🔹 Output**

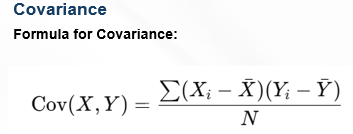
makefile

Copy code

Covariance: 12.5

**📌 Summary**

✔ **Covariance relationship batata hai ki do variables ek saath kaise change karte hain.**  
✔ **Positive covariance** → Dono variables ek saath badhte ya ghatte hain.  
✔ **Negative covariance** → Ek badhta hai, dusra ghatta hai.  
✔ **Zero covariance** → Koi relation nahi hai.  
✔ **Python mein np.cov() function se covariance easily calculate hota hai.**

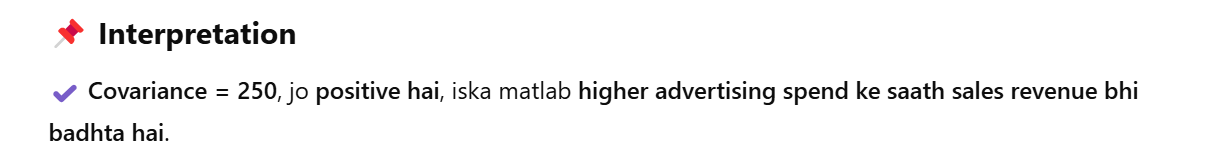
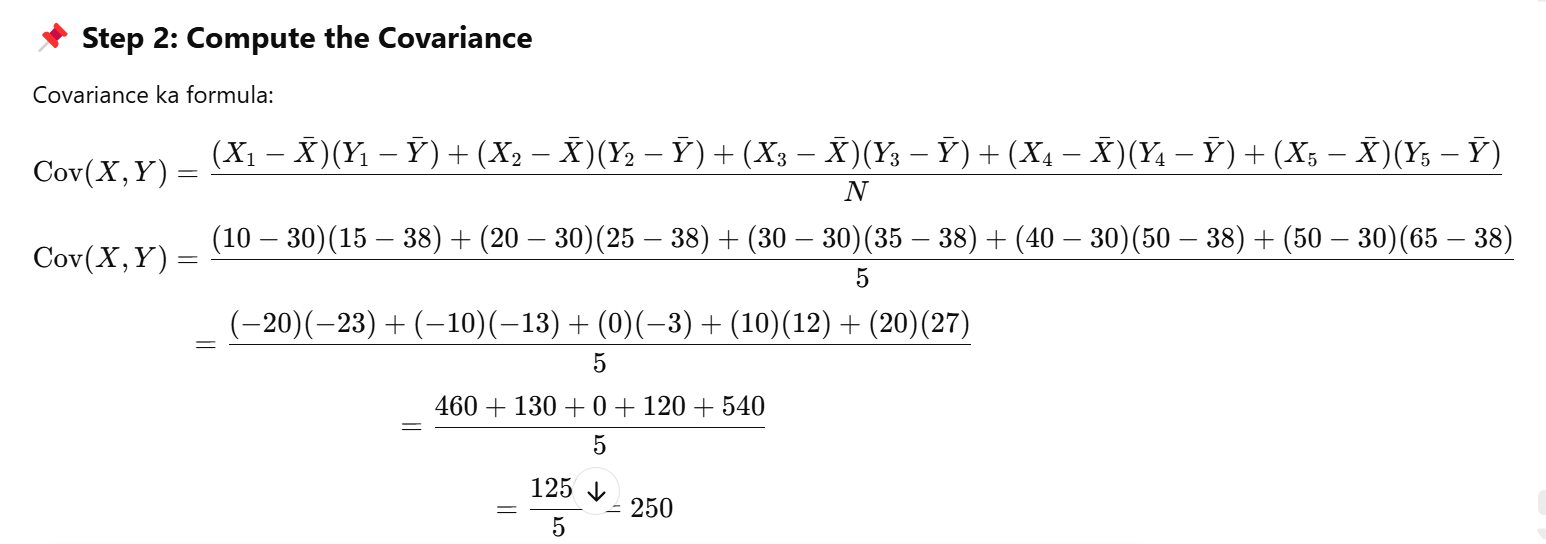
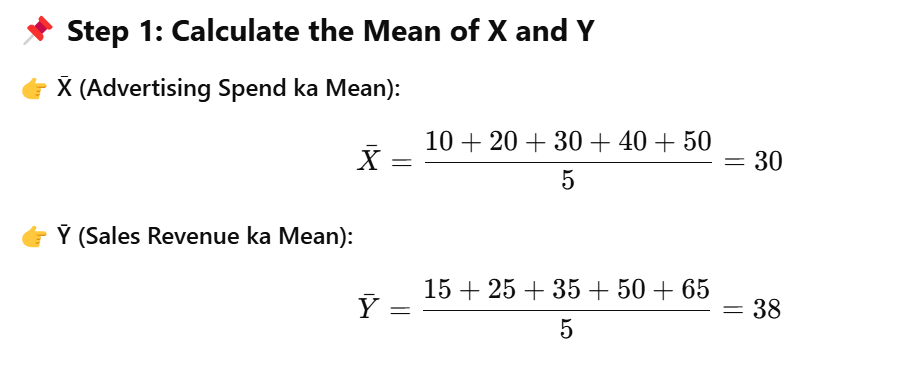
Covariance ka **use business analysis mein hota hai**, jaise ki **advertising spend** aur **sales revenue** ke beech relation samajhne ke liye.

*Where:*

*Xi, Yi*= Data values for two variables

= Means of the two datasets

*N* = Number of data points



**Correlation Overview**

👉 **Correlation** ek **statistical measure** hai jo **do variables ke beech relation ka direction aur strength** batata hai.

**📌 Correlation Ki Range**

−1≤r≤1-1 \leq r \leq 1−1≤r≤1

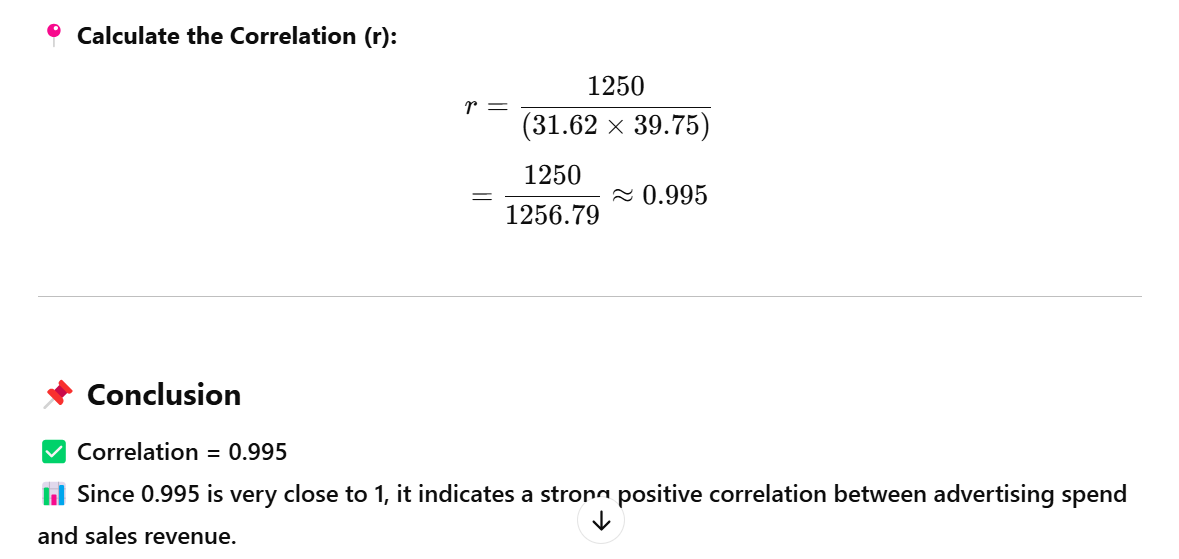
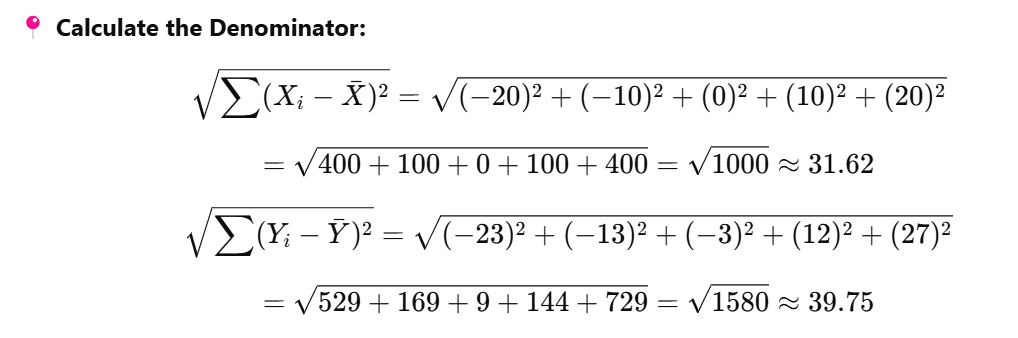
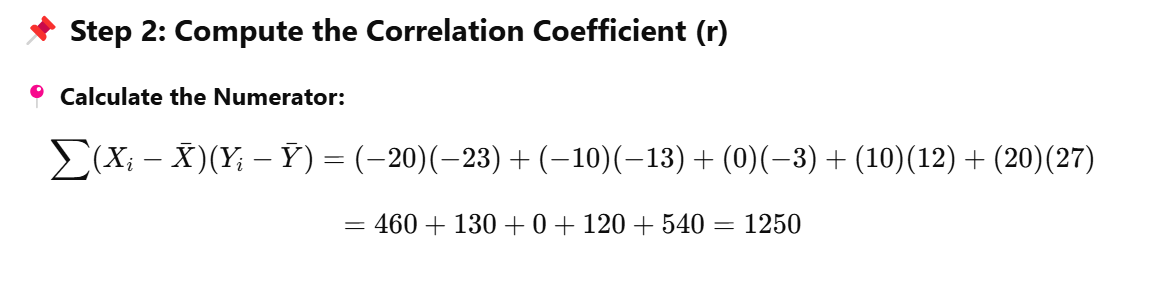
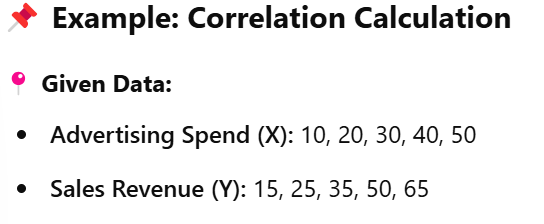
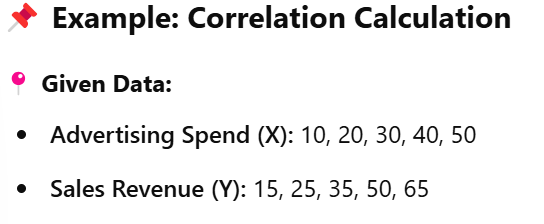
1️⃣ **Positive Correlation ( r > 0 )** – Jab **dono variables ek saath badhte hain**.  
2️⃣ **Negative Correlation ( r < 0 )** – Jab **ek variable badhta hai aur dusra girta hai**.  
3️⃣ **Zero Correlation ( r = 0 )** – Jab **koi relation nahi hota**.

**📌 Formula for Correlation (Pearson’s Correlation Coefficient)**

r=Cov(X,Y)σX⋅σYr = \frac{\text{Cov}(X,Y)}{\sigma\_X \cdot \sigma\_Y}r=σX​⋅σY​Cov(X,Y)​

📍 **Where:**  
✔ **Cov(X, Y)** = **Covariance** of X and Y  
✔ **σX, σY** = **Standard Deviations** of X and Y  
✔ **X̄, Ȳ** = **Means** of X and Y

✅ **Ye formula covariance ko scale karta hai** by dividing with the product of standard deviations, taaki **different datasets ke relationships compare kiye ja sakein**.



**Python Code to Calculate Correlation**

Python makes it easy to compute correlation using NumPy and Pandas.

**Python code**:

python

Copy code

import numpy as np

x = [10, 20, 30, 40, 50] # Advertising spend

y = [15, 25, 35, 50, 65] # Sales Revenue

# Compute correlation matrix

correlation\_matrix = np.corrcoef(x, y) # This will create a 2×2 correlation matrix

# Extract the correlation coefficient

correlation\_value = correlation\_matrix[0, 1] # Correlation is located at [0,1] in the matrix

print("Correlation Coefficient:", correlation\_value)

**🛠 Output:**

mathematica

Copy code

Correlation Coefficient: 0.995

✅ This code correctly calculates the correlation coefficient between advertising spend and sales revenue.

Lab

**Descriptive Statistics vs. Inferential Statistics**

| **Feature** | **Descriptive Statistics** | **Inferential Statistics** |
| --- | --- | --- |
| **Definition** | Summarizes and describes data. | Makes predictions or generalizations about a population using sample data. |
| **Purpose** | Provides insights into data distribution, central tendency, and variability. | Helps infer conclusions about a larger population. |
| **Techniques Used** | Mean, median, mode, standard deviation, variance, frequency distribution, graphs, etc. | Hypothesis testing, confidence intervals, regression analysis, etc. |
| **Scope** | Focuses only on the given dataset. | Extends conclusions beyond the observed dataset. |
| **Example** | "The average age of students in a class is 20 years." | "Based on a survey of 100 students, the average age of all university students is estimated to be 21 years." |

**Key Difference**

* **Descriptive Statistics** → Summarizes **what the data shows**.
* **Inferential Statistics** → **Makes predictions** based on data samples.

**Inferential Statistics**

🔹 **Definition:** Inferential statistics allows us to **make predictions or draw conclusions** about a population based on a sample.

🔹 **Purpose:** It helps in **generalizing findings** from a sample to a larger population.

**Key Concepts**

📌 **Population:** The **entire group** we want to study or draw conclusions about.  
📌 **Sample:** A **smaller subset** selected from the population for analysis.  
📌 **Why Use Samples?**  
✔ Collecting data from an entire population is often **impractical** or **impossible** due to time and cost constraints.

📌 **Statistical Inference:**  
✅ Uses sample data to **estimate population parameters**, such as:

* **Mean (Average)**
* **Proportions**
* **Trends and Patterns**

📌 **Key Goal:**  
🎯 To **draw conclusions** that can be applied to the whole population **based on sample data**.

**Inferential Statistics (अनुमानात्मक Statistics)**

🔹 **Definition:** Inferential statistics हमें **sample (chhoti dataset)** ke basis par **poori population** ke baare me conclusion nikalne ki technique hai.

🔹 **Purpose:** Ye help karta hai **sample ke findings ko generalize** karke **poori population** pe apply karne me.

**Key Concepts (Main Cheezein)**

📌 **Population:** Wo **poora group** jisko hum study karna chahte hain.  
📌 **Sample:** Population ka **chhota part**, jo analysis ke liye liya jata hai.  
📌 **Samples kyu zaroori hain?**  
✔ Pura data collect karna **impossible ya costly** ho sakta hai, isliye sample ka use hota hai.

📌 **Statistical Inference:**  
✅ Sample data ka use karke **Population parameters** ka estimation kiya jata hai, jaise:

* **Mean (average nikalna)**
* **Proportions (kitna % kis category me hai)**
* **Trends aur Patterns (kya pattern follow ho raha hai)**

📌 **Key Goal:**  
🎯 Sample data ko analyze karke **poori population ke liye valid conclusions nikalna**.

Sampling ka matlab hai **poori population me se chhota set (sample) choose karna** taaki uska analysis karke **poori population ke baare me conclusion nikal sakein**.

Isme **random selection** hota hai, yani har member ko **equal chance** milta hai sample me aane ka.

🔹 **(a) Simple Random Sampling (SRS)**

* **Random tarike se** sample select hota hai.
* Jaise ek **list me se randomly naam pick karna**.
* **Example:** Lottery system, computer-generated random numbers.

🔹 **(b) Stratified Sampling**

* Population ko **groups (strata) me divide** karke har group me se **random sample** liya jata hai.
* **Example:** Ek school me students ko **class-wise** divide karke har class me se **randomly kuch students** lena.

🔹 **(c) Systematic Sampling**

* **Fixed interval pe selection hota hai**.
* **Example:** Agar 100 logon ka survey karna ho, toh **har 10th person** ko select karna.

🔹 **(d) Cluster Sampling**

* **Puri population ko groups (clusters) me divide karna** aur **random clusters select karna**.
* **Example:** Ek country ko **cities me divide karke kuch cities ko randomly select karna** aur sirf unhi pe study karna.

Lab

**Summary (Ek Nazar Me Pura Chapter)**

✅ **Descriptive Statistics**: Data ko summarize aur analyze karne ka tareeka hai, jisme **mean, median, standard deviation** jaise measures use hote hain.

✅ **Inferential Statistics**: Sample data ke basis par **poori population ke baare me prediction** karne ka method hai. Isme **hypothesis testing aur confidence intervals** jaise techniques use hoti hain.

✅ **Correlation & Covariance**: **Variables ke beech relation** measure karne ke liye use hote hain, taaki **data patterns** samjhe ja sakein.