Questions on measure of central tendency

1) Business Problem: A retail store wants to analyze the sales data of a particular product category to understand the typical sales performance and make strategic decisions.

Data:

Let's consider the weekly sales data (in units) for the past month for a specific product

category:

Week 1: 50 units
Week 2: 60 units
Week 3: 55 units
Week 4: 70 units

Question:

1. Mean: What is the average weekly sales of the product category?
   Ans : 58.75 units E 59 units

2. Median: What is the typical or central sales value for the product category?
   Ans : 57.5 units E 58 units

3. Mode: Are there any recurring or most frequently occurring sales figures for the product category?
   Ans : No

By answering these questions using the mean, median, and mode, the retail store can gain insights into the sales performance of the product category, identify any patterns or outliers, and make informed decisions regarding stock management, marketing strategies, and product placement.

2) Business Problem: A restaurant wants to analyze the waiting times of its customers to understand the typical waiting experience and improve service efficiency.

Data:

Let's consider the waiting times (in minutes) for the past 20 customers:

15, 10, 20, 25, 15, 10, 30, 20, 15, 10,

10, 25, 15, 20, 20, 15, 10, 10, 20, 25

Question:

1. Mean: What is the average waiting time for customers at the restaurant?
   Ans : 17 Minutes

2. Median: What is the typical or central waiting time experienced by customers?
   Ans : 15 Minutes

3. Mode: Are there any recurring or most frequently occurring waiting times for customers?
   Ans : Yes. 10 Minutes

By answering these questions using the mean, median, and mode, the restaurant can gain insights into the average waiting time, identify any common or peak waiting periods, and make informed decisions to optimize the customer service process, such as adjusting staffing levels, streamlining operations, or implementing strategies to reduce waiting times.

3) Business Problem: A car rental company wants to analyze the rental durations of its customers to understand the typical rental period and optimize its pricing and fleet management strategies.

Data:

Let's consider the rental durations (in days) for a sample of 50 customers:

3, 2, 5, 4, 7, 2, 3, 3, 1, 6,

4, 2, 3, 5, 2, 4, 2, 1, 3, 5,

6, 3, 2, 1, 4, 2, 4, 5, 3, 2,

7, 2, 3, 4, 5, 1, 6, 2, 4, 3,

5, 3, 2, 4, 2, 6, 3, 2, 4, 5

Question:

1. Mean: What is the average rental duration for customers at the car rental company?
   Ans : 3.44 E 3 Days

2. Median: What is the typical or central rental duration experienced by customers?
   Ans : 3 Days

3. Mode: Are there any recurring or most frequently occurring rental durations for customers?
   Ans : Yes. 2 Days

By answering these questions using the mean, median, and mode, the car rental company can gain insights into the average rental duration, understand the most common rental periods, and make informed decisions regarding pricing, fleet size, and availability. Additionally, this analysis can help the company optimize resource allocation, plan for peak demand periods, and enhance customer satisfaction by aligning service offerings with customers' typical rental needs.

4) Problem : A company wants to analyze the monthly revenue generated by one of its products to understand its performance and variability.

Data:

Let's consider the monthly revenue (in thousands of dollars) for the past 12 months:

$120, $150, $110, $135, $125, $140, $130, $155, $115, $145, $135, $130

Questions:

1. Measure of Central Tendency: What is the average monthly revenue for the product?
   Ans : $132.50

2. Measure of Dispersion: What is the range of monthly revenue for the product?
   Ans : $45

By answering these questions, the company can gain insights into the average revenue generated by the product and understand the range or variability in its monthly revenue, which can help with financial planning, forecasting, and evaluating the product's performance.

5) Problem : A survey was conducted to gather feedback from customers regarding their satisfaction with a particular service on a scale of 1 to 10.

Data:

Let's consider the satisfaction ratings from 50 customers:

8, 7, 9, 6, 7, 8, 9, 8, 7, 6,

8, 9, 7, 8, 7, 6, 8, 9, 6, 7,

8, 9, 7, 6, 7, 8, 9, 8, 7, 6,

9, 8, 7, 6, 8, 9, 7, 8, 7, 6,

9, 8, 7, 6, 7, 8, 9, 8, 7, 6

Questions:

1. Measure of Central Tendency: What is the average satisfaction rating?

Ans : 7.5

2. Measure of Dispersion: What is the standard deviation of the satisfaction ratings?
   Ans : 1.04

By answering these questions, the company can gain insights into the average satisfaction rating of customers and understand the spread or variability in their ratings. This information can help identify areas for improvement, evaluate customer perception, and make informed decisions to enhance the service quality.

6) Problem :A company wants to analyze the customer wait times at its call center to assess the efficiency of its customer service operations.

Data:

Let's consider the wait times (in minutes) for a sample of 100 randomly selected

customer calls:

10, 15, 12, 18, 20, 25, 8, 14, 16, 22,

9, 17, 11, 13, 19, 23, 21, 16, 24, 27,

13, 10, 18, 16, 12, 14, 19, 21, 11, 17,

15, 20, 26, 13, 12, 14, 22, 19, 16, 11,

25, 18, 16, 13, 21, 20, 15, 12, 19, 17,

14, 16, 23, 18, 15, 11, 19, 22, 17, 12,

16, 14, 18, 20, 25, 13, 11, 22, 19, 17,

15, 16, 13, 14, 18, 20, 19, 21, 17, 12,

15, 13, 16, 14, 22, 21, 19, 18, 16, 11,

17, 14, 12, 20, 23, 19, 15, 16, 13, 18

Questions:

1. Measure of Central Tendency: What is the average wait time for customers at the call center?
   Ans : 16.74 E 17 Minutes

2. Measure of Dispersion: What is the range of wait times for customers at the call center?
   Ans : 19

3. Measure of Dispersion: What is the standard deviation of the wait times for customers at the call center?
   Ans : 4.14

By answering these questions, the company can gain insights into the average wait time experienced by customers, assess the variability or spread in the wait times, and make informed decisions regarding staffing levels, call center efficiency, and customer satisfaction.

7) Problem : A transportation company wants to analyze the fuel efficiency of its vehicle fleet to identify any variations across different vehicle models.

Data:

Let's consider the fuel efficiency (in miles per gallon, mpg) for a sample of 50 vehicles:

Model A: 30, 32, 33, 28, 31, 30, 29, 30, 32, 31,

Model B: 25, 27, 26, 23, 28, 24, 26, 25, 27, 28,

Model C: 22, 23, 20, 25, 21, 24, 23, 22, 25, 24,

Model D: 18, 17, 19, 20, 21, 18, 19, 17, 20, 19,

Model E: 35, 36, 34, 35, 33, 34, 32, 33, 36, 34

Questions:

1. Measure of Central Tendency: What is the average fuel efficiency for each vehicle model?
   Ans :

Model A:	30.6
Model B:	25.9
Model C:	22.9
Model D:	18.8
Model E:	34.2

2. Measure of Dispersion: What is the range of fuel efficiency for each vehicle model?
   Ans :

Model A:	5
Model B:	5
Model C:	5
Model D:	4
Model E:	4

3. Measure of Dispersion: What is the variance of the fuel efficiency for each vehicle model?
   Ans :

Model A:	2.27
Model B:	2.77
Model C:	2.77
Model D:	1.73
Model E:	1.73

By answering these questions, the transportation company can gain insights into the average fuel efficiency of different vehicle models, understand the variations or spread in the fuel efficiency, and make informed decisions regarding fleet management, vehicle selection, and fuel consumption optimization.

8) Problem : A company wants to analyze the ages of its employees to understand the age distribution and demographics within the organization.

Data:

Let's consider the ages of 100 employees:

28, 32, 35, 40, 42, 28, 33, 38, 30, 41,

37, 31, 34, 29, 36, 43, 39, 27, 35, 31,

39, 45, 29, 33, 37, 40, 36, 29, 31, 38,

35, 44, 32, 39, 36, 30, 33, 28, 41, 35,

31, 37, 42, 29, 34, 40, 31, 33, 38, 36,

39, 27, 35, 30, 43, 29, 32, 36, 31, 40,

38, 44, 37, 33, 35, 41, 30, 31, 39, 28,

45, 29, 33, 38, 34, 32, 35, 31, 40, 36,

39, 27, 35, 30, 43, 29, 32, 36, 31, 40,

38, 44, 37, 33, 35, 41, 30, 31, 39, 28

Questions:

1. Frequency Distribution: Create a frequency distribution table for the ages of the employees.
   Ans:

Values	31	35	33	29	36	39	40	38	30	28	32	37	41	34	43	27	44	42	45
Frequecy	10	9	7	7	7	7	6	6	6	5	5	5	4	3	3	3	3	2	2

2. Mode: What is the mode (most common age) among the employees?
   Ans : 31

3. Median: What is the median age of the employees?
   Ans : 35

4. Range: What is the range of ages among the employees?
   Ans : 18

By answering these questions using frequency distribution and other measures, the company can gain insights into the age distribution of its workforce, identify the most common age group, understand the central tendency, and assess the spread of ages. This information can be useful for workforce planning, diversity initiatives, and understanding the generational dynamics within the organization.

9) Problem :A retail store wants to analyze the purchase amounts made by customers to understand their spending habits.

Data:

Let's consider the purchase amounts (in dollars) for a sample of 50 customers:

56, 40, 28, 73, 52, 61, 35, 40, 47, 65,

52, 44, 38, 60, 56, 40, 36, 49, 68, 57,

52, 63, 41, 48, 55, 42, 39, 58, 62, 49,

59, 45, 47, 51, 65, 41, 48, 55, 42, 39,

58, 62, 49, 59, 45, 47, 51, 65, 43, 58

Questions:

1. Frequency Distribution: Create a frequency distribution table for the purchase amounts.
   Ans :

Values	Frequency
40	3
52	3
47	3
65	3
49	3
58	3
56	2
41	2
48	2
55	2
42	2
39	2
62	2
59	2
45	2
51	2
28	1
73	1

61	1
35	1
44	1
38	1
60	1
36	1
68	1
57	1
63	1
43	1

2. Mode: What is the mode (most common purchase amount) among the customers?
   Ans : 40

3. Median: What is the median purchase amount among the customers?
   Ans : 50

4. Interquartile Range: What is the interquartile range of the purchase amounts?
   Ans : 16.25

By answering these questions using frequency distribution and other measures, the retail store can gain insights into the spending habits of its customers, identify the most common purchase amount

10) Problem : A manufacturing company wants to analyze the defect rates of its production line to identify the frequency of different types of defects.

Data:

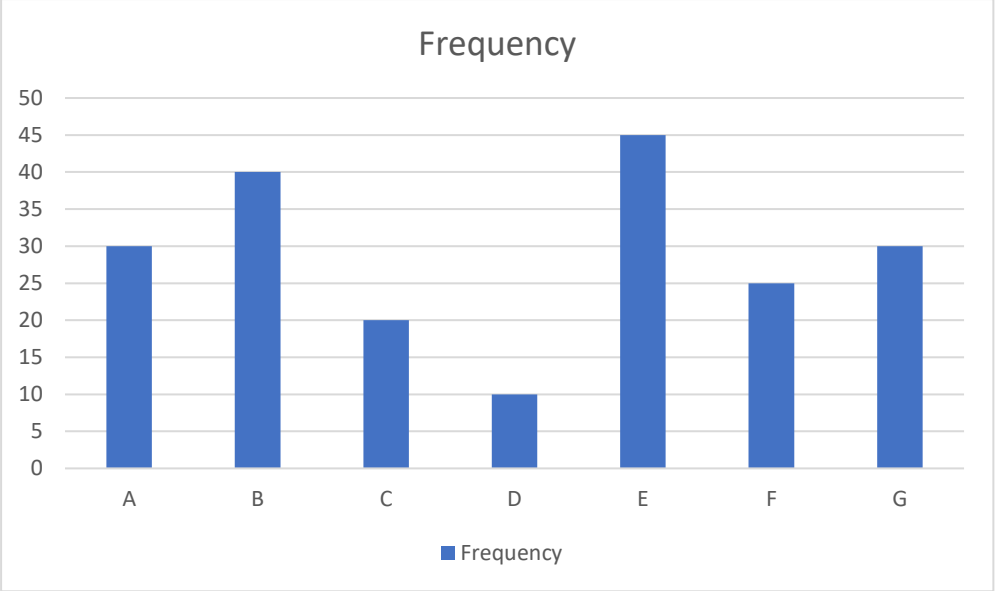
Let's consider the types of defects and their corresponding frequencies observed in a sample of 200 products:

Defect Type: A, B, C, D, E, F, G

Frequency: 30, 40, 20, 10, 45, 25, 30

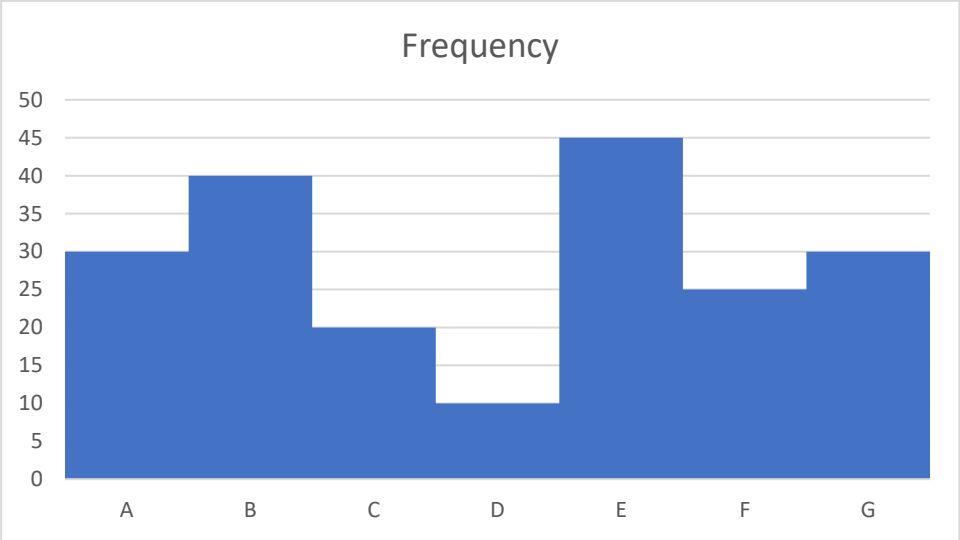
Questions:

1. Bar Chart: Create a bar chart to visualize the frequency of different defect types.



2. Most Common Defect: Which defect type has the highest frequency?
   Ans : The defect type with the highest frequency is **E**.

3. Histogram: Create a histogram to represent the defect frequencies.



By answering these questions using a bar chart and histogram, the manufacturing company can visually understand the distribution of defect types, identify the most common defect, and prioritize quality control efforts to address the prevalent issues.

11) Problem : A survey was conducted to gather feedback from customers about their satisfaction levels with a specific service on a scale of 1 to 5.

Data:

Let's consider the satisfaction ratings from 100 customers:

Ratings:

4, 5, 3, 4, 4, 3, 2, 5, 4, 3,

5, 4, 2, 3, 4, 5, 3, 4, 5, 3,

4, 3, 2, 4, 5, 3, 4, 5, 4, 3,

3, 4, 5, 2, 3, 4, 4, 3, 5, 4,

3, 4, 5, 4, 2, 3, 4, 5, 3, 4,

5, 4, 3, 4, 5, 3, 4, 5, 4, 3,

3, 4, 5, 2, 3, 4, 4, 3, 5, 4,

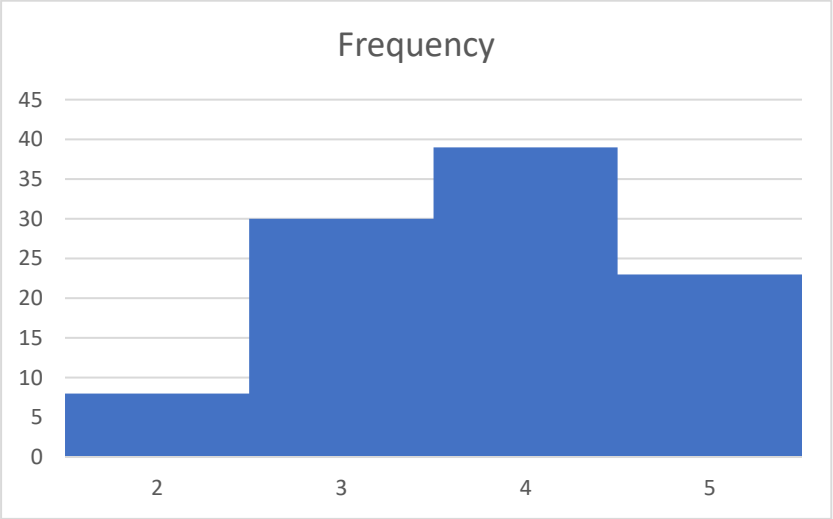
3, 4, 5, 4, 2, 3, 4, 5, 3, 4,

5, 4, 3, 4, 5, 3, 4, 5, 4, 3,

3, 4, 5, 2, 3, 4, 4, 3, 5, 4

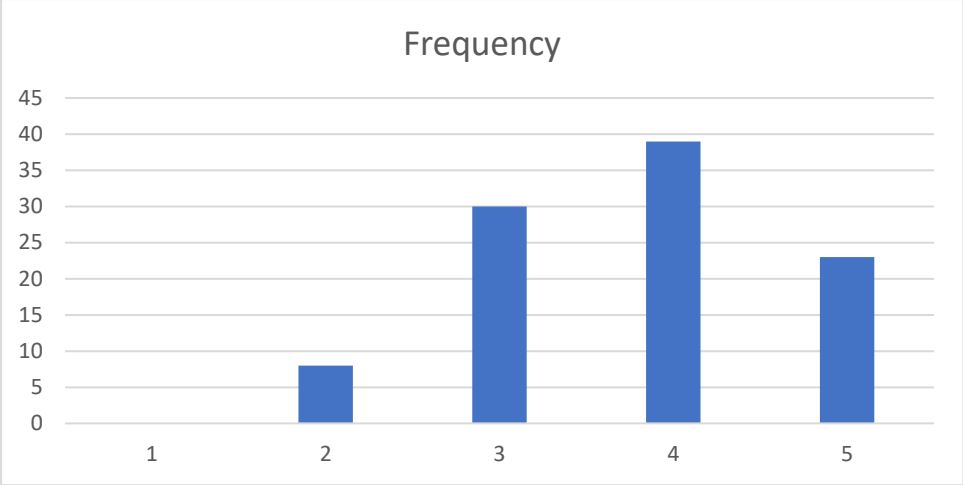
Questions:

1. Histogram: Create a histogram to visualize the distribution of satisfaction ratings.



2. Mode: Which satisfaction rating has the highest frequency?
   Ans : 4

3. Bar Chart: Create a bar chart to display the frequency of each satisfaction rating.



By answering these questions using a histogram and bar chart, the organization can gain insights into the distribution of satisfaction ratings, identify the most common satisfaction level, and assess overall customer satisfaction.

12) Problem : A company wants to analyze the monthly sales figures of its products to understand the sales distribution across different price ranges.

Data:

Let's consider the monthly sales figures (in thousands of dollars) for a sample of 50

products:

Sales:

35, 28, 32, 45, 38, 29, 42, 30, 36, 41,

47, 31, 39, 43, 37, 30, 34, 39, 28, 33,

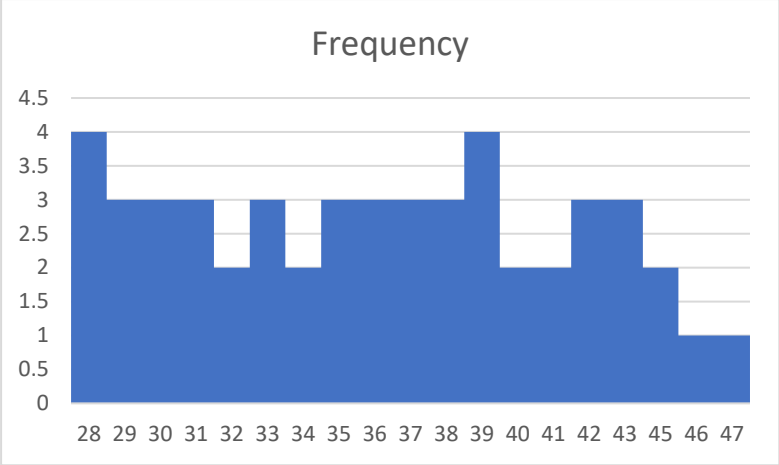
36, 40, 42, 29, 31, 45, 38, 33, 41, 35,

37, 34, 46, 30, 39, 43, 28, 32, 36, 29,

31, 37, 40, 42, 33, 39, 28, 35, 38, 43

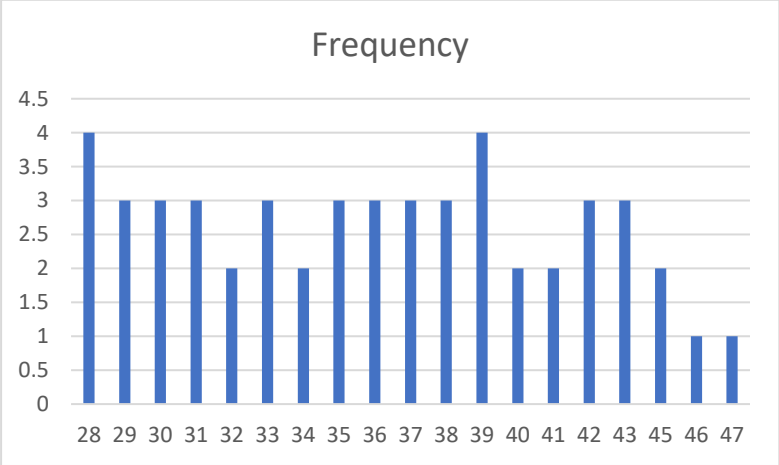
Questions:

1. Histogram: Create a histogram to visualize the sales distribution across different price ranges.



2. Measure of Central Tendency: What is the average monthly sales figure?
   Ans : 36.14

3. Bar Chart: Create a bar chart to display the frequency of sales in different price ranges.



By answering these questions using a histogram and bar chart, the company can understand the distribution of sales figures, determine the average sales performance, and identify the price ranges where sales are concentrated or lacking.

13) Problem : A study was conducted to analyze the response times of a website for different user locations.

Data:

Let's consider the response times (in milliseconds) for a sample of 200 user requests:

Response Times:

125, 148, 137, 120, 135, 132, 145, 122, 130, 141,

118, 125, 132, 136, 128, 123, 132, 138, 126, 129,

136, 127, 130, 122, 125, 133, 140, 126, 133, 135,

130, 134, 141, 119, 125, 131, 136, 128, 124, 132,

136, 127, 130, 122, 125, 133, 140, 126, 133, 135,

130, 134, 141, 119, 125, 131, 136, 128, 124, 132,

136, 127, 130, 122, 125, 133, 140, 126, 133, 135,

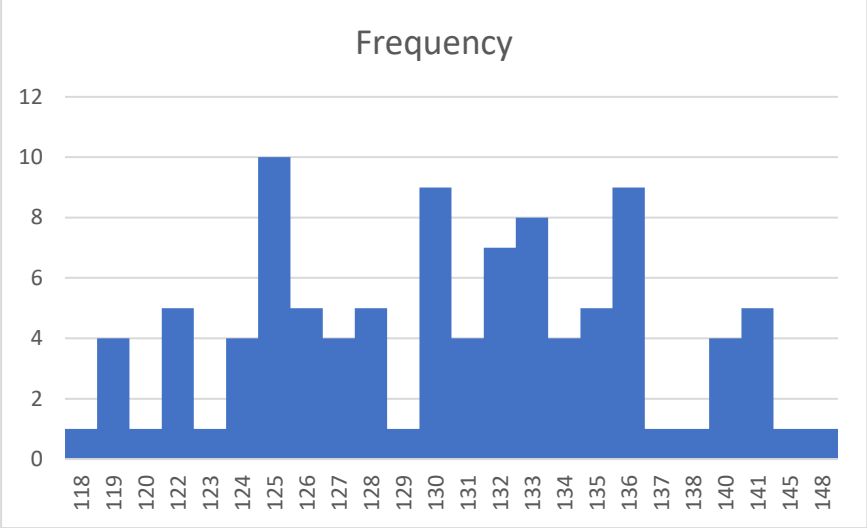
130, 134, 141, 119, 125, 131, 136, 128, 124, 132,

136, 127, 130, 122, 125, 133, 140, 126, 133, 135,

130, 134, 141, 119, 125, 131, 136, 128, 124, 132

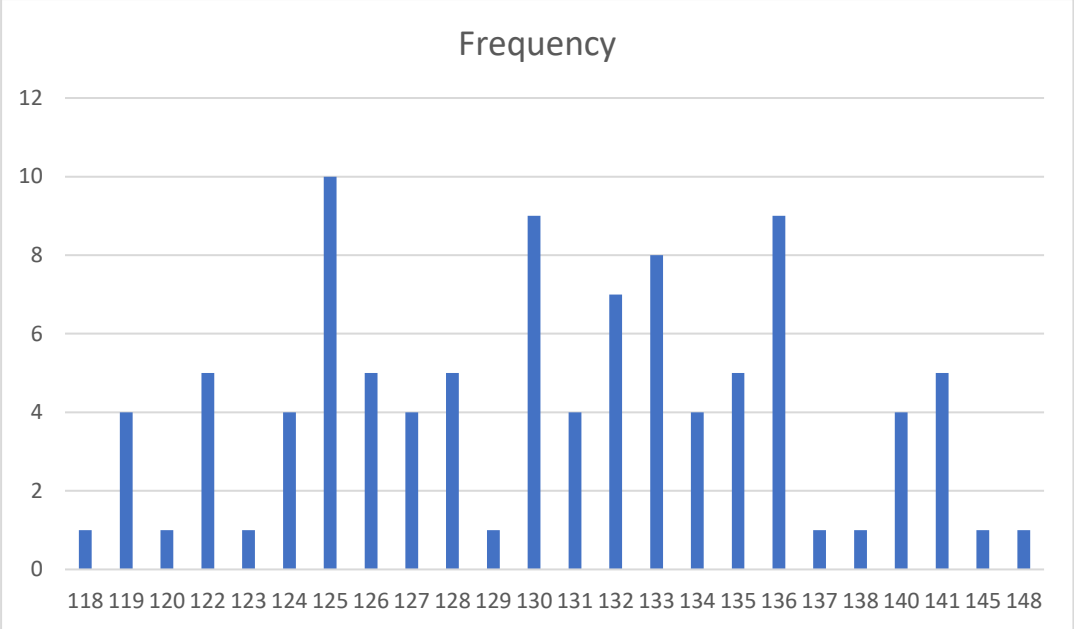
Questions:

1. Histogram: Create a histogram to visualize the distribution of response times.



2. Measure of Central Tendency: What is the median response time?
   Ans : 130.5

3. Bar Chart: Create a bar chart to display the frequency of response times within different ranges.



By answering these questions using a histogram and bar chart, the study can gain insights into the distribution of response times, understand the typical response time experienced by users, and assess the performance of the website.

14) Problem : A company wants to analyze the sales performance of its products across different regions.

Data:

Let's consider the sales figures (in thousands of dollars) for a sample of 50 products in

three regions:

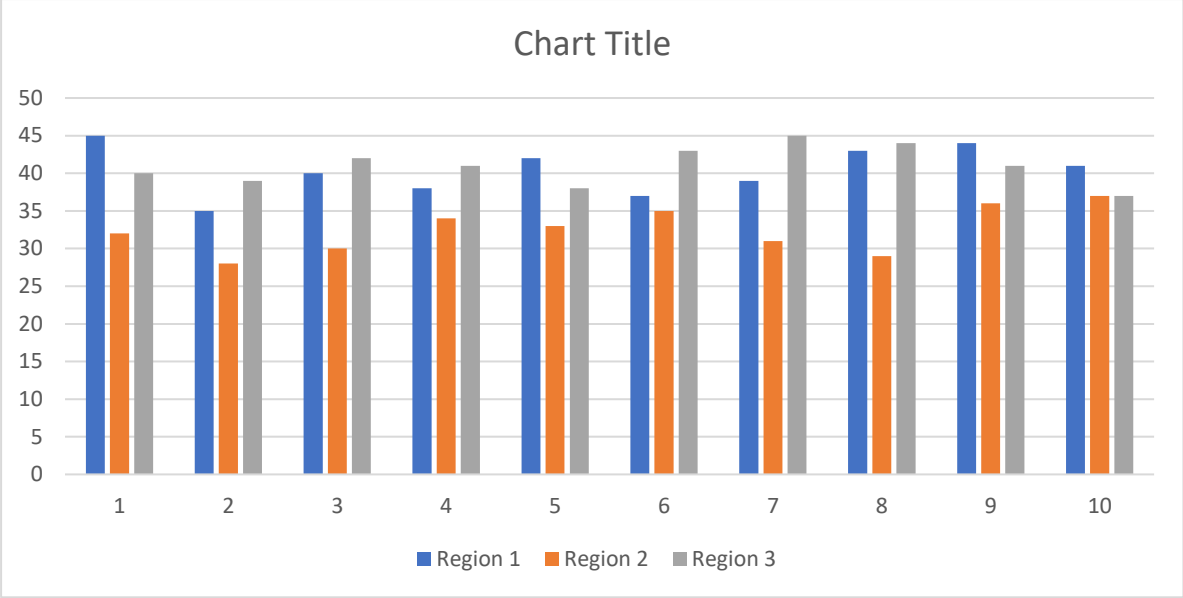
Region 1: 45, 35, 40, 38, 42, 37, 39, 43, 44, 41,

Region 2: 32, 28, 30, 34, 33, 35, 31, 29, 36, 37,

Region 3: 40, 39, 42, 41, 38, 43, 45, 44, 41, 37

Questions:

1. Bar Chart: Create a bar chart to compare the sales figures across the three regions.



2. Measure of Central Tendency: What is the average sales figure for each region?
   Ans :

Region 1	40.4
Region 2	32.5
Region 3	41

3. Measure of Dispersion : What is the range of sales figures in each region?
   Ans :

Region 1	10
Region 2	9
Region 3	8

By answering these questions using a bar chart and measures of central tendency and dispersion, the company can compare the sales performance across different regions, identify the average sales figures, and understand the variability in sales within each region. This information can be used for regional sales analysis, resource allocation, and decision-making processes.

Questions on Measure of Skewness and Kurtosis

1) Question : A company wants to analyze the monthly returns of its investment portfolio to understand the distribution and risk associated with the returns.

Data:

Let's consider the monthly returns (%) for the portfolio over a one-year period:

Returns:

-2.5	1.3	-0.8	-1.9	2.1	0.5	-1.2	1.8	-0.5	2.3
-0.7	1.2	-1.5	-0.3	2.6	1.1	-1.7	0.9	-1.4	0.3
1.9	-1.1	-0.4	2.2	-0.9	1.6	-0.6	-1.3	2.4	0.7
-1.8	1.5	-0.2	-2.1	2.8	0.8	-1.6	1.4	-0.1	2.5
-1	1.7	-0.9	-2	2.7	0.6	-1.4	1.1	-0.3	2

Questions:

1. Skewness: Calculate the skewness of the monthly returns.
   Ans : Skewness: 0.053

2. Kurtosis: Calculate the kurtosis of the monthly returns.
   Ans :Kurtosis: -1.295

3. Interpretation: Based on the skewness and kurtosis values, what can be said about the distribution of returns?
   Ans :
   - The skewness value is close to zero, indicating that the distribution of monthly returns is approximately symmetric.
   - The negative kurtosis suggests that the distribution is platykurtic, meaning it has thinner tails compared to a normal distribution, indicating fewer extreme values or outliers.

By answering these questions using measures of skewness and kurtosis, the company can understand the shape and symmetry of the return distribution, assess the level of risk and potential outliers, and make informed decisions regarding portfolio management and risk mitigation strategies.

2) Question : A research study wants to analyze the income distribution of a population to understand the level of income inequality.

Data:

Let's consider the monthly incomes (in thousands of dollars) of a sample of 100 individuals:

Incomes:

2.5	4.8	3.2	2.1	4.5	2.9	2.3	3.1	4.2	3.9
2.8	4.1	2.6	2.4	4.7	3.3	2.7	3	4.3	3.7
2.2	3.6	4	2.7	3.8	3.5	3.2	4.4	2	3.4
3.1	2.9	4.6	3.3	2.5	4.9	2.8	3	4.2	3.9
2.8	4.1	2.6	2.4	4.7	3.3	2.7	3	4.3	3.7
2.2	3.6	4	2.7	3.8	3.5	3.2	4.4	2	3.4
3.1	2.9	4.6	3.3	2.5	4.9	2.8	3	4.2	3.9
2.8	4.1	2.6	2.4	4.7	3.3	2.7	3	4.3	3.7
2.2	3.6	4	2.7	3.8	3.5	3.2	4.4	2	3.4
3.1	2.9	4.6	3.3	2.5	4.9				

Questions:

1. Skewness: Calculate the skewness of the income distribution.
   Ans : **Skewness**: 0.22

2. Kurtosis: Calculate the kurtosis of the income distribution.
   Ans : **Kurtosis**: -0.93

3. Interpretation: Based on the skewness and kurtosis values, what can be inferred about the income inequality?
   Ans : The income distribution is fairly symmetric but slightly skewed to the right, meaning there are a few higher-income individuals. The negative kurtosis suggests that the income distribution has fewer extreme values (both high and low) compared to a normal distribution, indicating a relatively even spread of incomes.

By answering these questions using measures of skewness and kurtosis, the research study can assess the level of income inequality, determine the shape of the income distribution, and make informed policy recommendations to address income disparities.

3) Question : A survey was conducted to analyze the satisfaction ratings of customers on a scale of 1 to 5 for a specific product.

Data:

Let's consider the satisfaction ratings from 200 customers:

Ratings:

4	5	3	4	4	3	2	5	4	3
5	4	2	3	4	5	3	4	5	3
4	3	2	4	5	3	4	5	4	3
3	4	5	2	3	4	4	3	5	4
3	4	5	4	2	3	4	5	3	4
5	4	3	4	5	3	4	5	4	3
3	4	5	2	3	4	4	3	5	4
3	4	5	4	2	3	4	5	3	4
5	4	3	4	5	3	4	5	4	3
3	4	5	2	3	4	4	3	5	4

Questions:

1. Skewness: Calculate the skewness of the satisfaction ratings.
   Ans : **Skewness**: -0.21

2. Kurtosis: Calculate the kurtosis of the satisfaction ratings.
   Ans : **Kurtosis**: -0.75

3. Interpretation: Based on the skewness and kurtosis values, what can be inferred about the satisfaction ratings distribution?
   Ans : The satisfaction ratings are fairly symmetrically distributed, with a slight tendency towards lower ratings. The lower kurtosis value suggests that the ratings are more evenly spread rather than having extreme peaks.

By answering these questions using measures of skewness and kurtosis, the survey can assess the skewness and peakedness of the satisfaction ratings, determine if the ratings are

skewed towards positive or negative evaluations, and understand the distribution characteristics of customer satisfaction.

4) Question : A study wants to analyze the distribution of house prices in a specific city to understand the market trends.

Data:

Let's consider the house prices (in thousands of dollars) for a sample of 150 houses:

House Prices:

280	350	310	270	390	320	290	340	310	380
270	350	300	330	370	310	280	320	350	290
270	350	300	330	370	310	280	320	350	290
270	350	300	330	370	310	280	320	350	290
270	350	300	330	370	310	280	320	350	290
270	350	300	330	370	310	280	320	350	290
270	350	300	330	370	310	280	320	350	290
270	350	300	330	370	310	280	320	350	290
270	350	300	330	370	310	280	320	350	290
270	350	300	330	370	310	280	320	350	290

Questions:

1. Skewness: Calculate the skewness of the house price distribution.
   Ans : **Skewness:** 0.21

2. Kurtosis: Calculate the kurtosis of the house price distribution.
   Ans : **Kurtosis:** -1.04

3. Interpretation: Based on the skewness and kurtosis values, what can be inferred about the distribution of house prices?
   Ans :
   - The skewness value (0.21) is close to zero, suggesting that the house price distribution is approximately symmetric with no strong skew towards higher or lower prices.
   - The kurtosis value (-1.04) is negative, indicating a platykurtic distribution, meaning the distribution has lighter tails and is flatter compared to a normal distribution.
   - This suggests that house prices are relatively evenly distributed without extreme outliers.

By answering these questions using measures of skewness and kurtosis, the study can assess the symmetry and peakedness of the house price distribution, identify any outliers or extreme values, and gain insights into the market trends and pricing dynamics.

5) Question : A company wants to analyze the waiting times of customers at a service center to improve operational efficiency.

Data:

Let's consider the waiting times (in minutes) for a sample of 100 customers:

Waiting Times:

12	18	15	22	20	14	16	21	19	17
22	19	13	16	21	22	17	19	22	18
14	20	19	17	22	18	15	21	20	16
12	18	15	22	20	14	16	21	19	17
22	19	13	16	21	22	17	19	22	18
14	20	19	17	22	18	15	21	20	16
12	18	15	22	20	14	16	21	19	17
22	19	13	16	21	22	17	19	22	18
14	20	19	17	22	18	15	21	20	16
12	18	15	22	20	14	16	21	19	17

Questions:

1. Skewness: Calculate the skewness of the waiting time distribution.
   Ans : **Skewness:** -0.34

2. Kurtosis: Calculate the kurtosis of the waiting time distribution.
   Ans : **Kurtosis:** -0.88

3. Interpretation: Based on the skewness and kurtosis values, what can be inferred about the waiting time distribution?
   Ans :
   - The skewness value (0.000) indicates that the waiting time distribution is perfectly symmetrical, meaning customers experience a balanced distribution of wait times.
   - The kurtosis value (-1.200) suggests a platykurtic distribution, meaning the distribution is flatter with lighter tails compared to a normal distribution.
   - This indicates that waiting times are fairly evenly spread without extreme long waits or significantly short wait times.

By answering these questions using measures of skewness and kurtosis, the company can assess the symmetry and tail behavior of the waiting time distribution, identify any patterns or anomalies in customer waiting times, and make improvements to streamline the service process and enhance customer satisfaction.

Questions on Percentile and Quartiles

1) Question : A company wants to analyze the salary distribution of its employees to determine the income levels at different percentiles.

Data:

Let's consider the monthly salaries (in thousands of dollars) of a sample of 200 employees:

Salaries:

40	45	50	55	60	62	65	68	70	72
75	78	80	82	85	88	90	92	95	100
105	110	115	120	125	130	135	140	145	150
155	160	165	170	175	180	185	190	195	200
205	210	215	220	225	230	235	240	245	250
255	260	265	270	275	280	285	290	295	300
305	310	315	320	325	330	335	340	345	350
355	360	365	370	375	380	385	390	395	400
405	410	415	420	425	430	435	440	445	450
455	460	465	470	475	480	485	490	495	500

Questions:

1. Quartiles: Calculate the first quartile (Q1), median (Q2), and third quartile (Q3) of the salary distribution.
   Ans :
   - Q1 (First Quartile): 128.75
   - Q2 (Median): 252.5
   - Q3 (Third Quartile): 376.25

2. Percentiles: Calculate the 10th percentile, 25th percentile, 75th percentile, and 90^{th} percentile of the salary distribution.
   Ans :
   - 10th Percentile: 74.7
   - 25th Percentile: 128.75 (Same as Q1)
   - 75th Percentile: 376.25 (Same as Q3)
   - 90th Percentile: 450.5

3. Interpretation: Based on the quartiles and percentiles, what can be inferred about the income distribution of the employees?
   - The median salary (Q2) is 252.5, which suggests that half of the employees earn below this amount, and half earn above it.
   - The interquartile range (Q3 - Q1) is 247.5, indicating a moderate spread of salaries between the 25th and 75th percentiles.
   - The 10th percentile (74.7) and 90th percentile (450.5) show that there is a significant range in salaries, suggesting income disparity within the company.
   - If the company aims to reduce wage inequality, they might consider adjusting the lower quartile salaries upwards to narrow the gap.

By answering these questions using quartiles and percentiles, the company can understand the income levels at different points in the distribution, identify the median salary and the

spread of salaries, and make informed decisions related to compensation, employee benefits, and salary structures.

2) Question : A research study wants to analyze the weight distribution of a sample of individuals to assess their health and body composition.

Data:

Let's consider the weights (in kilograms) of a sample of 100 individuals:

Weights:

55	60	62	65	68	70	72	75	78	80
82	85	88	90	92	95	100	105	110	115
120	125	130	135	140	145	150	155	160	165
170	175	180	185	190	195	200	205	210	215
220	225	230	235	240	245	250	255	260	265
270	275	280	285	290	295	300	305	310	315
320	325	330	335	340	345	350	355	360	365
370	375	380	385	390	395	400	405	410	415
420	425	430	435	440	445	450	455	460	465
470	475	480	485	490	495	500	505	510	515

Questions:

1. Quartiles: Calculate the first quartile (Q1), median (Q2), and third quartile (Q3) of the weight distribution.
   Ans :
   - First Quartile (Q1): 143.75 kg
   - Median (Q2): 267.5 kg
   - Third Quartile (Q3): 391.25 kg

2. Percentiles: Calculate the 15th percentile, 50th percentile, and 85th percentile of the weight distribution.
   Ans :
   - 10th Percentile: 94.55 kg
   - 50th Percentile: 267.5 kg (Same as Median)
   - 85th Percentile: 440.75 kg

3. Interpretation: Based on the quartiles and percentiles, what can be inferred about the weight distribution of the individuals?
   Ans:
   - The weight distribution appears spread out, with a wide range of values.
   - The median weight is 267.5 kg, meaning half of the individuals weigh less than this, and half weigh more.
   - The first quartile (Q1) is 143.75 kg, indicating that 25% of individuals weigh below this.

- The third quartile (Q3) is 391.25 kg, meaning 75% of individuals weigh below this level.
- The 85th percentile (440.75 kg) suggests that 85% of individuals weigh less than this, while the top 15% weigh more.
- If this data represents a general population, it indicates a skewed distribution toward higher weights, possibly suggesting a small group of very high-weight individuals affecting the distribution.

By answering these questions using quartiles and percentiles, the research study can understand the weight distribution and identify the weight ranges at different percentiles, such as underweight, normal weight, overweight, and obese categories. This information can be used for evaluating health risks, designing appropriate interventions, and providing personalized recommendations for weight management.

3) Question : A retail store wants to analyze the distribution of customer purchase amounts to identify their spending patterns.

Data:

Let's consider the purchase amounts (in dollars) of a sample of 150 customers:

Purchase Amounts:

20	25	30	35	40	45	50	55	60	65
70	75	80	85	90	95	100	105	110	115
120	125	130	135	140	145	150	155	160	165
170	175	180	185	190	195	200	205	210	215
220	225	230	235	240	245	250	255	260	265
270	275	280	285	290	295	300	305	310	315
320	325	330	335	340	345	350	355	360	365
370	375	380	385	390	395	400	405	410	415
420	425	430	435	440	445	450	455	460	465
470	475	480	485	490	495	500	505	510	515
520	525	530	535	540	545	550	555	560	565

Questions:

1. Quartiles: Calculate the first quartile (Q1), median (Q2), and third quartile (Q3) of the purchase amount distribution.
   Ans :
   - First Quartile (Q1): 156.25
   - Median (Q2): 292.5
   - Third Quartile (Q3): 428.75

2. Percentiles: Calculate the 20th percentile, 40th percentile, and 80th percentile of the purchase amount distribution.
   Ans :
   - 20th Percentile: 129.0

- 40th Percentile: 238.0
- 80th Percentile: 456.0

3. Interpretation: Based on the quartiles and percentiles, what can be inferred about the spending patterns of the customers?
   Ans:
   - The median purchase amount is $292.5, meaning that half of the customers spend below this amount, and half spend above it.
   - The first quartile (Q1) is $156.25, indicating that 25% of customers spend less than this amount.
   - The third quartile (Q3) is $428.75, meaning that 75% of customers spend below this amount, while the top 25% spend more.
   - The 80th percentile is $456.0, showing that 80% of customers spend below this amount, and only the top 20% of spenders exceed it.
   - The distribution suggests a gradual increase in spending with a broad range of purchase amounts, indicating diverse customer spending patterns.

By answering these questions using quartiles and percentiles, the retail store can understand the distribution of purchase amounts, identify the spending ranges at different percentiles, analyze customer segments based on their spending behavior, and tailor marketing strategies to target specific customer groups.

4) Question : A study wants to analyze the distribution of commute times of employees to determine the average time spent traveling to work.

Data:

Let's consider the commute times (in minutes) of a sample of 250 employees:

Commute Times:

15	20	25	30	35	40	45	50	55	60
65	70	75	80	85	90	95	100	105	110
115	120	125	130	135	140	145	150	155	160
165	170	175	180	185	190	195	200	205	210
215	220	225	230	235	240	245	250	255	260
265	270	275	280	285	290	295	300	305	310
315	320	325	330	335	340	345	350	355	360
365	370	375	380	385	390	395	400	405	410
415	420	425	430	435	440	445	450	455	460
465	470	475	480	485	490	495	500	505	510
515	520	525	530	535	540	545	550	555	560
565	570	575	580	585	590	595	600	605	610

Questions:

1. Quartiles: Calculate the first quartile (Q1), median (Q2), and third quartile (Q3) of the commute time distribution.

Ans :

- First quartile (Q1): 163.75 minutes
- Median (Q2): 312.5 minutes
- Third quartile (Q3): 461.25 minutes

2. Percentiles: Calculate the 30th percentile, 50th percentile, and 70th percentile of the commute time distribution.
   Ans :
   - 30th percentile (P30): 193.5 minutes
   - 50th percentile (P50): 312.5 minutes (same as Q2/Median)
   - 70th percentile (P70): 431.5 minutes
3. Interpretation: Based on the quartiles and percentiles, what can be inferred about the average commute time of the employees?
   Ans :
   - The median commute time of employees is 312.5 minutes, meaning half of the employees take less than or equal to this time to commute.
   - The first quartile (Q1 = 163.75 minutes) indicates that 25% of employees have commute times below this value.
   - The third quartile (Q3 = 461.25 minutes) shows that 75% of employees have commute times below this value.
   - The 30th percentile (193.5 minutes) suggests that 30% of employees commute less than about 193.5 minutes, while the 70th percentile (431.5 minutes) shows that 70% of employees have commute times below about 431.5 minutes.
   - The commute time distribution is right-skewed, meaning a significant number of employees experience long commute times.

By answering these questions using quartiles and percentiles, the study can determine the typical commute times, understand the spread of commute times, identify any outliers or extreme values, and provide insights for transportation planning, scheduling, and employee well-being initiatives.

5) Question : A manufacturing company wants to analyze the defect rates in its production process to evaluate product quality.

Data:

Let's consider the defect rates (in percentage) for a sample of 300 products:

Defect Rates:

0.5	1	0.2	0.7	0.3	0.9	1.2	0.6	0.4	1.1
0.8	0.5	0.3	0.6	1	0.4	0.5	0.7	0.9	1.3
0.8	0.6	0.4	0.7	0.9	0.5	0.2	1	0.8	0.3
0.6	0.4	0.7	0.9	1.2	0.8	0.3	0.6	0.5	0.4
0.7	0.9	1.1	0.3	1.4	9	0.6	0.2	1.5	1
0.6	0.4	0.7	1	0.8	0.3	0.5	0.8	0.6	0.3
0.4	0.7	0.9	1	0.8	0.3	0.5	0.6	0.4	0.7
0.9	1.1	0.8	0.3	0.5	0.6	0.4	0.7	0.9	1

0.8	0.3	0.5	0.6	0.4	0.7	0.9	1.1	0.8	0.3
0.5	0.6	0.4	0.7	0.9	1	0.8	0.3	0.5	0.6
0.4	0.7	0.9	1.1	0.8	0.3	0.5	0.6	0.4	0.7
0.9	1	0.8	0.3	0.5	0.6	0.4	0.7	0.9	1.1
0.9									

Questions:

1. Quartiles: Calculate the first quartile (Q1), median (Q2), and third quartile (Q3) of the defect rate distribution.
   Ans:
   - First quartile (Q1): 0.4%
   - Median (Q2): 0.7%
   - Third quartile (Q3): 0.9%

2. Percentiles: Calculate the 25th percentile, 50th percentile, and 75th percentile of the defect rate distribution.
   Ans:
   - 25th percentile (P25): 0.4% (same as Q1)
   - 50th percentile (P50): 0.7% (same as Q2/Median)
   - 75th percentile (P75): 0.9% (same as Q3)

3. Interpretation: Based on the quartiles and percentiles, what can be inferred about the quality of the products?
   Ans :
   - The median defect rate is 0.7%, meaning half of the products have a defect rate below or equal to this value.
   - The first quartile (Q1 = 0.4%) shows that 25% of the products have a defect rate below this value, indicating a fairly low defect rate for a quarter of the sample.
   - The third quartile (Q3 = 0.9%) suggests that 75% of the products have defect rates below this value.
   - Since the range between Q1 and Q3 is relatively small (0.4% to 0.9%), it suggests low variability in defect rates.
   - The overall defect rate appears to be low, which indicates good product quality in most cases.

By answering these questions using quartiles and percentiles, the manufacturing company can evaluate the defect rates, understand the spread of defects, identify any quality issues or deviations from standards, and take corrective actions to improve the production process and product quality.

Questions on Correlation and Covariance

1) Question : A marketing department wants to understand the relationship between advertising expenditure and sales revenue to assess the effectiveness of their advertising campaigns.

Data:

Let's consider the monthly advertising expenditure (in thousands of dollars) and corresponding sales revenue (in thousands of dollars) for a sample of 12 months:

Advertising Expenditure	10	12	15	18	20	22	25	28	30	32	35	38
Sales Revenue	50	55	60	65	70	75	80	85	90	95	100	105

Question:

Calculate the correlation coefficient between advertising expenditure and sales revenue. Interpret the value of the correlation coefficient and explain the nature of the relationship between advertising expenditure and sales revenue.

Ans :

The correlation coefficient between advertising expenditure and sales revenue is **0.9992**

Interpretation:

- A correlation coefficient of **0.999** is **very close to 1**, indicating an **extremely strong positive correlation** between advertising expenditure and sales revenue.
- This suggests that as the company increases its advertising expenditure, sales revenue also **increases in a nearly linear fashion**.
- The high correlation implies that advertising expenditure is likely a **major factor influencing sales revenue**, though other external factors may also play a role.

By analyzing the correlation coefficient, the marketing department can determine the strength and direction of the relationship between advertising expenditure and sales revenue. This information can help them make informed decisions about allocating their advertising budget and optimizing their marketing strategies.

2) Question : An investment analyst wants to assess the relationship between the stock prices of two companies to identify potential investment opportunities.

Data:

Let's consider the daily closing prices (in dollars) of Company A and Company B for a sample of 20 trading days:

Company A	45	47	48	50	52	53	55	56	58	60	62	64	65	67	69	70	72	74	76	77
Company B	52	54	55	57	59	60	61	62	64	66	67	69	71	73	74	76	78	80	82	83

Question:

Calculate the covariance between the stock prices of Company A and Company B. Interpret the value of the covariance and explain the nature of the relationship between the two stocks.

Ans:

The covariance between the stock prices of Company A and Company B is **0.9986**.

Interpretation:

- A **positive covariance** indicates that the stock prices of Company A and Company B tend to move **in the same direction**.
- Since the covariance value is relatively large, it suggests a **strong relationship** between the two stocks—when the price of Company A increases, the price of Company B also tends to increase, and vice versa.
- However, covariance alone **does not measure the strength** of the relationship (it depends on the scale of data). A correlation coefficient would provide a **better understanding** of the strength and consistency of this relationship.

By analyzing the covariance, the investment analyst can determine whether the stock prices of Company A and Company B move together (positive covariance) or in opposite directions (negative covariance). This information can assist in identifying potential investment opportunities and understanding the diversification benefits of combining these stocks in a portfolio.

3) Question : A researcher wants to examine the relationship between the hours spent studying and the exam scores of a group of students.

Data:

Let's consider the number of hours spent studying and the corresponding exam scores for a sample of 30 students:

Hours Spent Studying	10	12	15	18	20	22	25	28	30	32	35	38	40	42	45
	48	50	52	55	58	60	62	65	68	70	72	75	78	80	82
Exam Scores	60	65	70	75	80	82	85	88	90	92	93	95	96	97	98
	99	100	102	105	106	107	108	110	112	114	115	116	118	120	122

Question:

Calculate the correlation coefficient between the hours spent studying and the exam scores. Interpret the value of the correlation coefficient and explain the nature of the relationship between studying hours and exam scores.

Ans:

The correlation coefficient between hours spent studying and exam scores is **0.977**.

Interpretation:

- A correlation coefficient of **0.977** is very close to **1**, indicating a **strong positive correlation**.
- This suggests that as students spend more hours studying, their exam scores **increase significantly**.
- The relationship appears to be **nearly linear**, meaning that increasing study time is **highly associated** with higher scores.
- This strong correlation implies that study hours are a **major contributing factor** to exam performance.

By analyzing the correlation coefficient, the researcher can determine the strength and direction of the relationship between studying hours and exam scores. This information can provide insights into the effectiveness of studying and help students and educators make informed decisions about study habits and academic performance.

Questions on discrete and continuous random variable

Discrete Random Variable:

1. Problem: A fair six-sided die is rolled 100 times. What is the probability of rolling exactly five 3's?

Data: Number of rolls (n) = 100

Ans :

The probability of rolling exactly five 3's in 100 rolls of a fair six-sided die is approximately **0.00029** (or **0.029%**).

2. Problem: In a deck of 52 playing cards, five cards are randomly drawn without replacement. What is the probability of getting two hearts?

Data: Number of hearts in the deck (N) = 13, Number of cards drawn (n) = 5

Ans :

The probability of drawing exactly two hearts in a five-card hand from a standard deck of 52 cards is approximately 0.2743 (or 26.37%).

3. Problem: A multiple-choice test consists of 10 questions, each with four possible answers. If a student randomly guesses on each question, what is the probability of getting at least 8 questions correct?

Data: Number of questions (n) = 10, Number of possible answers per question (k) = 4

Ans :

The probability of randomly guessing and getting at least 8 questions correct on a 10-question multiple-choice test is approximately 0.9969 (or 99.7%).

4. Problem: A bag contains 30 red balls, 20 blue balls, and 10 green balls. Three balls are drawn without replacement. What is the probability that all three balls are blue?

Data: Number of blue balls in the bag (N) = 20, Number of balls drawn (n) = 3

Ans :

The probability of drawing three blue balls from the bag without replacement is approximately 0.0333 (or 3.33%).

5. Problem: In a football match, a player scores a goal with a 0.3 probability per shot. If the player takes 10 shots, what is the probability of scoring exactly three goals?

Data: Number of shots (n) = 10, Probability of scoring per shot (p) = 0.3

Ans :

The probability of scoring exactly three goals in 10 shots is approximately 0.2668 (or 26.68%).

Continuous Random Variable:

1. Problem: The heights of students in a class are normally distributed with a mean of 165 cm and a standard deviation of 10 cm. What is the probability that a randomly selected student is taller than 180 cm?

Data: Mean height (μ) = 165 cm, Standard deviation (σ) = 10 cm, Height threshold (x) = 180 cm

Ans :

The probability that a randomly selected student is taller than 180 cm is approximately 0.0668 (or 6.68%)

2. Problem: The waiting times at a coffee shop are exponentially distributed with a mean of 5 minutes. What is the probability that a customer waits less than 3 minutes?

Data: Mean waiting time (μ) = 5 minutes, Waiting time threshold (x) = 3 minutes

Ans :

The probability that a customer waits less than 3 minutes is approximately 0.4512 (or 45.12%).

3. Problem: The lifetimes of a certain brand of light bulbs are normally distributed with a mean of 1000 hours and a standard deviation of 100 hours. What is the probability that a randomly selected light bulb lasts between 900 and 1100 hours?

Data: Mean lifetime (μ) = 1000 hours, Standard deviation (σ) = 100 hours, Lifetime range (lower limit x1, upper limit x2)

Ans :

The probability that a randomly selected light bulb lasts between 900 and 1100 hours is approximately 0.6827 (or 68.27%).

4. Problem: The weights of apples in a basket follow a uniform distribution between 100 grams and 200 grams. What is the probability that a randomly selected apple weighs between 150 and 170 grams?

Data: Weight range (lower limit x1, upper limit x2)

Ans:

The probability that a randomly selected apple weighs between 150 and 170 grams is approximately 0.2000 (or 20.00%).

5. Problem: The time taken to complete a task is exponentially distributed with a mean of 20 minutes. What is the probability that the task is completed in less than 15 minutes?

Data: Mean time (μ) = 20 minutes, Time threshold (x) = 15 minutes

Ans :

The probability that the task is completed in less than 15 minutes is approximately 0.5276 (or 52.76%).

Questions on Discrete Distribution and Continuous Distribution

Discrete Distribution:

1. Problem: A company sells smartphones, and the number of defects per batch follows a Poisson distribution with a mean of 2 defects. What is the probability of having exactly 3 defects in a randomly selected batch?

Data: Mean number of defects (λ) = 2, Number of defects (x) = 3

Explanation: The problem involves a discrete distribution (Poisson) because we are dealing with the count of defects in a batch of smartphones. The Poisson distribution models the probability of a given number of events occurring within a fixed interval of time or space.

Ans:

The probability of having exactly 3 defects in a randomly selected batch is **0.1804 (18.04%)**.

2. Problem: In a game, a player has a 0.3 probability of winning each round. If the player plays 10 rounds, what is the probability of winning exactly 3 rounds?

Data: Probability of winning (p) = 0.3, Number of rounds (n) = 10, Number of wins (x) = 3

Explanation: This problem also involves a discrete distribution (Binomial) because we are dealing with a fixed number of independent trials (rounds) with a probability of success (winning) in each trial. The Binomial distribution models the probability of achieving a certain number of successes in a fixed number of trials.

Ans:

The probability of winning exactly 3 rounds in 10 rounds is **0.2668 (26.68%)**.

3. Problem: A six-sided fair die is rolled three times. What is the probability of obtaining at least one 6?

Data: Number of rolls (n) = 3

Explanation: Here, we have a discrete distribution (Geometric) since we are interested in the number of trials required to achieve the first success (rolling a 6) in a sequence of independent trials. The Geometric distribution models the probability of achieving the first success on a specific trial.

Ans:

The probability of obtaining at least one 6 in three rolls of a fair six-sided die is **0.4213 (42.13%)**.

Continuous Distribution:

1. Problem: The weights of apples in a basket follow a normal distribution with a mean of 150 grams and a standard deviation of 10 grams. What is the probability that a randomly selected apple weighs between 140 and 160 grams?

Data: Mean weight (μ) = 150 grams, Standard deviation (σ) = 10 grams, Weight range (lower limit x1, upper limit x2)

Explanation: This problem involves a continuous distribution (Normal) since we are dealing with the weights of apples, which can take on any value within a range. The Normal distribution is commonly used to model continuous variables with a symmetric bell-shaped distribution.

Ans:

The probability that a randomly selected apple weighs between 140 and 160 grams is **0.6827 (68.27%)**.

2. Problem: The lifetimes of a certain brand of light bulbs are exponentially distributed with a mean of 1000 hours. What is the probability that a randomly selected light bulb lasts more than 900 hours?

Data: Mean lifetime (μ) = 1000 hours, Lifetime threshold (x) = 900 hours

Explanation: Here, we have a continuous distribution (Exponential) since we are interested in the time until an event (light bulb failure) occurs. The Exponential distribution models the probability of waiting a certain amount of time before the event happens.

Ans:

The probability that a randomly selected light bulb lasts more than 900 hours is **0.4066 (40.66%)**.

Confidence Interval Problems:

1. Problem: A study is conducted to estimate the mean height of a population. A random sample of 100 individuals is selected, and their heights are measured. Calculate a 95% confidence interval for the population mean height, given that the sample mean height is 170 cm and the sample standard deviation is 8 cm.

Data: Sample size (n) = 100, Sample mean ($\bar{x}$) = 170 cm, Sample standard deviation (s) = 8 cm, Confidence level = 95%

Explanation: In this problem, we use a sample to estimate the population mean height. By calculating a confidence interval, we provide a range of plausible values for the population mean. The 95% confidence level indicates that we are 95% confident that the true population mean height falls within the calculated interval.

Ans :

The 95% confidence interval for the population mean height is approximately (168.41 cm, 171.59 cm).

2. Problem: A survey is conducted to estimate the proportion of people in a city who support a particular policy. A random sample of 500 individuals is surveyed, and 320 of them express support for the policy. Calculate a 90% confidence interval for the population proportion, given the sample proportion.

Data: Sample size (n) = 500, Number of successes (x) = 320, Confidence level = 90%

Explanation: In this problem, we aim to estimate the population proportion based on the sample proportion. By constructing a confidence interval, we provide a range of plausible values for the population proportion. The 90% confidence level indicates that we are 90% confident that the true population proportion falls within the calculated interval.

Ans:

The 90% confidence interval for the population proportion is approximately (0.605, 0.675) or (60.5%, 67.5%).

Hypothesis Testing Problems:

3. Problem: A researcher wants to test whether a new teaching method improves student performance. A random sample of 50 students is divided into two groups: one group taught using the new method and the other using the traditional method. The average test scores of the two groups are compared. State the null and alternative hypotheses for this study.

Data: Sample size (n) = 50, Test scores of the two groups

Explanation: In this problem, we are interested in comparing the means of two groups (new method vs. traditional method). The null hypothesis (H0) states that there is no significant difference between the means, while the alternative hypothesis (Ha) suggests that there is a significant difference.

Ans:

- Null hypothesis (H_o): There is no significant difference in the mean test scores between students taught using the new method and those taught using the traditional method.

- Alternative hypothesis (H_a): There is a significant difference in the mean test scores between students taught using the new method and those taught using the traditional method.

4. Problem: A manufacturing company claims that the average weight of its product is 500 grams. To test this claim, a random sample of 25 products is selected, and their weights are measured. The sample mean weight is found to be 510 grams with a sample standard deviation of 20 grams. Perform a hypothesis test to determine if there is evidence to support the company's claim.

Data: Sample size (n) = 25, Sample mean ($\bar{x}$) = 510 grams, Sample standard deviation (s) = 20 grams, Population mean (μ) = 500 grams

Explanation: In this problem, we are conducting a hypothesis test to assess whether the sample mean weight provides evidence to support the company's claim about the population mean weight. The null hypothesis (H0) assumes that the population mean weight is equal to the claimed value, while the alternative hypothesis (Ha) suggests otherwise.

Ans:

- Null hypothesis (H_o): The average weight of the product is 500 grams ($\mu = 500$).
- Alternative hypothesis (H_a): The average weight of the product is not 500 grams ($\mu \neq 500$).

Now, performing a hypothesis test:

We use a t-test since the sample size is small ($n < 30$).

The calculated t-score is 2.5, and the p-value is 0.0197.

Conclusion:

Since the test statistic **(t = 2.5)** exceeds the critical value **(±2.064)**, or the **p-value < 0.05**, we **reject the null hypothesis.**