ASSIGNMENT No. 01

Statistics for Management (1430) BA/ B.COM

Spring, 2025

Q. 1 Describe the formation of frequency distribution and characteristics of relative frequency distribution. (20)

The *frequency* (**f**) of a particular observation is the number of times the observation occurs in the data. The *distribution* of a variable is the pattern of frequencies of the observation. Frequency distributions are portrayed as frequency tables, histograms, or polygons.

Frequency distributions can show either the actual number of observations falling in each range or the percentage of observations. In the latter instance, the distribution is called a *relative frequency distribution*.

Frequency distribution tables can be used for both categorical and numeric variables. Continuous variables should only be used with class intervals, which will be explained shortly.

Example 1 – Constructing a frequency distribution table

A survey was taken on Maple Avenue. In each of 20 homes, people were asked how many cars were registered to their households. The results were recorded as follows:

1, 2, 1, 0, 3, 4, 0, 1, 1, 1, 2, 2, 3, 2, 3, 2, 1, 4, 0, 0

Use the following steps to present this data in a frequency distribution table.

Divide the results (*x*) into intervals, and then count the number of results in each interval. In this case, the intervals would be the number of households with no car (0), one car (1), two cars (2) and so forth.

Make a table with separate columns for the interval numbers (the number of cars per household), the tallied results, and the frequency of results in each interval. Label these columns *Number of cars*, *Tally* and *Frequency*.

Read the list of data from left to right and place a tally mark in the appropriate row. For example, the first result is a 1, so place a tally mark in the row beside where 1 appears in the interval column (*Number of cars*). The next result is a 2, so place a tally mark in the row beside the 2, and so on. When you reach your fifth tally mark, draw a tally line through the preceding four marks to make your final frequency calculations easier to read.

Add up the number of tally marks in each row and record them in the final column entitled *Frequency*. Your frequency distribution table for this exercise should look like this:

Table 1. Frequency table for the number of cars registered in each household

Number of cars (x)	Tally	Frequency (f)
0		4
1		6

2		5
3		3
4		2

By looking at this frequency distribution table quickly, we can see that out of 20 households surveyed, 4 households had no cars, 6 households had 1 car, etc.

Example 2 – Constructing a cumulative frequency distribution table
A *cumulative frequency distribution table* is a more detailed table. It looks almost the same as a frequency distribution table but it has added columns that give the cumulative frequency and the cumulative percentage of the results, as well.

At a recent chess tournament, all 10 of the participants had to fill out a form that gave their names, address and age. The ages of the participants were recorded as follows:

36, 48, 54, 92, 57, 63, 66, 76, 66, 80

Use the following steps to present these data in a cumulative frequency distribution table.

Divide the results into intervals, and then count the number of results in each interval. In this case, intervals of 10 are appropriate. Since 36 is the lowest age and 92 is the highest age, start the intervals at 35 to 44 and end the intervals with 85 to 94.

Create a table similar to the frequency distribution table but with three extra columns.

In the first column or the *Lower value* column, list the lower value of the result intervals. For example, in the first row, you would put the number 35.

The next column is the *Upper value* column. Place the upper value of the result intervals. For example, you would put the number 44 in the first row.

The third column is the *Frequency* column. Record the number of times a result appears between the lower and upper values. In the first row, place the number 1.

The fourth column is the *Cumulative frequency* column. Here we add the cumulative frequency of the previous row to the frequency of the current row. Since this is the first row, the cumulative frequency is the same as the frequency. However, in the second row, the frequency for the 35–44 interval (i.e., 1) is added to the frequency for the 45–54 interval (i.e., 2). Thus, the cumulative frequency is 3, meaning we have 3 participants in the 34 to 54 age group.

1 + 2 = 3

The next column is the *Percentage* column. In this column, list the percentage of the frequency. To do this, divide the frequency by the total number of results and multiply by 100. In this case, the frequency of the first row is 1 and the total number of results is 10. The percentage would then be 10.0.

10.0. (1 ÷ 10) X 100 = 10.0

The final column is *Cumulative percentage*. In this column, divide the cumulative frequency by the total number of results and then to make a percentage, multiply by 100. Note that the last number in this column should always equal 100.0. In this example, the cumulative frequency is 1 and the total number of results is 10, therefore the cumulative percentage of the first row is 10.0.
10.0. (1 ÷ 10) X 100 = 10.0

The cumulative frequency distribution table should look like this:

Table 2. Ages of participants at a chess tournament

Lower Value	Upper Value	Frequency (f)	Cumulative frequency	Percentage	Cumulative percentage
35	44	1	1	10.0	10.0
45	54	2	3	20.0	30.0
55	64	2	5	20.0	50.0
65	74	2	7	20.0	70.0
75	84	2	9	20.0	90.0
85	94	1	10	10.0	100.0

For more information on how to make cumulative frequency tables, see the section on Cumulative frequency and Cumulative percentage.

Class intervals

If a variable takes a large number of values, then it is easier to present and handle the data by grouping the values into class intervals. Continuous variables are more likely to be presented in class intervals, while discrete variables can be grouped into class intervals or not. To illustrate, suppose we set out age ranges for a study of young people, while allowing for the possibility that some older people may also fall into the scope of our study.

The *frequency* of a class interval is the number of observations that occur in a particular predefined interval. So, for example, if 20 people aged 5 to 9 appear in our study's data, the frequency for the 5–9 interval is 20. The *endpoints* of a class interval are the lowest and highest values that a variable can take. So, the intervals in our study are 0 to 4 years, 5 to 9 years, 10 to 14 years, 15 to 19 years, 20 to 24 years, and 25 years and over. The endpoints of the first interval are 0 and 4 if the variable is discrete, and 0 and 4.999 if the variable is continuous. The endpoints of the other class intervals would be determined in the same way.

Class interval width is the difference between the lower endpoint of an interval and the lower endpoint of the next interval. Thus, if our study's continuous intervals are 0 to 4, 5 to 9, etc., the width of the first five intervals is 5, and the last interval is open, since no higher endpoint is assigned to it. The intervals could also be written as 0 to less than 5, 5 to less than 10, 10 to less than 15, 15 to less than 20, 20 to less than 25, and 25 and over.

Rules for data sets that contain a large number of observations
In summary, follow these basic rules when constructing a frequency distribution table for a data set that contains a large number of observations:

- find the lowest and highest values of the variables
- decide on the width of the class intervals
- include all possible values of the variable.

In deciding on the width of the class intervals, you will have to find a compromise between having intervals short enough so that not all of the observations fall in the same interval, but long enough so that you do not end up with only one observation per interval.

Q. 2 a) Define descriptive and inferential statistics in detail.(10+10)

Descriptive and Inferential Statistics:
Statistics is the study of numerical data. It deals with the gathering, presentation, management, organization, calculation and analysis of usually vast numerical data.
There are majorly two kinds of divisions of statistics:

1. **Descriptive Statistics**
2. **Inferential Statistics.**

The statistical analysis enables us to draw conclusions about several different statistical situations, both in descriptive and inferential statistics. Both the segments are equally important. Both have different objectives. Basically, the descriptive statistics describes the features of the sample data quantitatively. On the other hand, the inferential statistics does inferences for the population data from which the given samples were taken.Let us first discuss about the two most basic concepts: population and sample. The population is defined as the whole set of data, individuals, events or objects etc on which the researcher is performing research. The whole area of study is included in a population. While, the sample is relatively smaller. It is a subset of the population. Since it is difficult to handle and analyze each and every member in the population, a smaller and representative portion from the population is picked up. This is called sample. Usually, both descriptive and inferential statistics are used in most of the statistical researches in order to properly analyze the data and reach conclusions. Both types provide different insights about the nature of given data. They together make a powerful tool for the better analysis of the data. But what is the difference between these two types of statistics? Let us go ahead and learn about both of them in detail.

Descriptive Statistics Definition

The descriptive statistics is the type of statistical analysis which helps to describes about the data in some meaningful way. This statistics is used to describe quantitatively about the important features of the data or information. The descriptive statistics gives the summaries of the given sample as well as the observations done. These summaries or descriptions can either be graphical or quantitative.
For Example: In soccer, the individual performance of each player is said to be the a descriptive statistics.
However, descriptive statistics does not reach at conclusions beyond the given data or hypothesis made by the researcher. It is just a simple way of describing the data.
Generally, the kinds of measure that are used with descriptive statistics are:
1) Measures of Central Tendency:
The measure of central tendency describes the data which lies in the center of a given frequency distribution. The main measures of central tendency are mean and median and mode.
2) Measures of Spread:
The measure of spread describes the how the scores are spread over the whole distribution. Standard deviation, variance, quartiles, range, absolute deviation are included in the measures of spread.
3) Graphical Representation:
There are several different types of graphs that are used to describe about the statistical data. These graphs are histogram, bar graph, box and whisker plot, line graph, scatter plot, ogive, pie chart and many more.
Inferential Statistics Definition
Inferential statistics is the type of statistics which deals with making conclusions. It inferences about the predictions for the population. It also analyses the sample. Basically, the inferential statistics is the procedure of drawing predictions and conclusions about the given data which is subjected to the random variations. Inferential statistics includes detection and prediction of observational and sampling errors. This type of statistics is being utilized in order to make estimates and test the hypotheses using given data. The inferential statistics may be defined as the answer of the question "what is needed to be done next". This provides an information about the further surveys and experiments. Inferential statistics enables the researcher to draw conclusions before the implementation of some particular organizational policy.

There are two major divisions of inferential statistics:

1) Confidence Interval:

The confidence interval is represented in the form of an interval that provides a range for the parameter of given population.

2) Hypothesis Test:

Hypothesis tests are also known as tests of significance which tests some claim for the population by analyzing sample.

Difference between Descriptive and Inferential Statistics

Although descriptive and inferential statistics both are used for purpose of analysis of the data, still both of them are different in various ways. Let us learn about this difference below:

1) The descriptive statistics gives a description about a sample, while the inferential statistics predicts and infers about a much larger data or population.
3) Descriptive statistics just describes the certain characteristics about a data. Whereas, inferential statistics deeply analyzes the statistical data and observations.
3) Descriptive statistics deals with central tendency and spread of the frequency distribution. While in inferential statistics, more details such as hypothesis tests and confidence interval are studied.
4) The measures of descriptive statistics (mean, median, mode) are numbers. On the other hand, the measures in inferential statistics are not always exact numbers.

Difference between Qualitative Data and Quantitative Data

In the study of statistics, the main focus is on collecting data or information. There are different methods of collecting data, and there are different types of data collected. The different types of data are primary, secondary, qualitative, or quantitative. In this article we will focus on qualitative and quantitative data and their differences.

Statistics

Statistics is basically the study of data. Statistics is either descriptive or inferential. Descriptive data is the study of methods used for the collection of data and mathematical models in order to interpret data. Inferential statistics is the study in which different techniques and systems are used to make probability–based predictions and decisions depending on incomplete data.
Statistics uses a lot of mathematics and many major concepts like probability, populations, samples, and distribution, etc. have been made possible by statistics. To study statistics, we need to collect data, quantitative as well as qualitative.

b) Explain in detail the application of statistics in business and commerce.

Introduction to Statistics in Business and Commerce

Statistics plays an essential role in business and commerce, serving as a powerful tool for decision-making, forecasting, and optimization. The application of statistical methods enables organizations to analyze data effectively, draw meaningful conclusions, and implement strategies based on empirical evidence rather than intuition. This systematic approach leads to improved efficiencies, enhanced customer satisfaction, and increased profitability. The significance of statistics in the business environment encompasses various areas, including market research, quality control, financial analysis, decision-making processes, and strategic planning.

Market Research and Consumer Behavior Analysis

One of the primary applications of statistics in business is market research, which involves gathering and analyzing consumer data to understand preferences, behavior, and trends. Businesses leverage statistical methods such as surveys, polls, and focus groups to collect quantitative and qualitative data. By applying statistical techniques such as regression analysis, correlation studies, and hypothesis testing, companies can gain insights into consumer needs and market dynamics. This information aids in segmenting markets, identifying target audiences, and developing tailored marketing campaigns, ultimately enhancing customer engagement and brand loyalty.

Quality Control and Process Improvement

In manufacturing and service industries, statistics is pivotal for quality control and process improvement. Statistical quality control (SQC) employs various methods, including Control Charts, Six Sigma, and Descriptive Statistics, to monitor production processes and minimize variability. By analyzing defect rates and performance metrics, businesses can identify inefficiencies and implement corrective measures. Statistical techniques help organizations maintain consistent quality in products and services, resulting in high customer satisfaction and reduced operational costs. Moreover, statistical analysis assists in forecasting demand and optimizing inventory levels, ensuring that businesses meet customer needs while minimizing excess stock.

Financial Analysis and Risk Management

Statistics is crucial in the realm of financial analysis and risk management. Organizations use statistical models to assess financial performance, evaluate investment opportunities, and manage risks. Techniques such as time series analysis, variance analysis, and financial ratios provide insights into profitability, liquidity, and solvency. Additionally, businesses employ statistical methods to forecast revenue, analyze market trends, and evaluate the potential impact of economic changes on their operations. In risk management, statistical models and simulations, such as Monte Carlo analysis, help firms quantify risks and make informed decisions regarding capital allocation, insurance coverage, and disaster recovery planning.

Decision-Making and Strategic Planning

Effective decision-making is at the heart of successful business operations, and statistics provides the tools necessary to make informed choices. By employing data-driven approaches, organizations can analyze historical data, assess current performance, and model future scenarios. Techniques such as predictive analytics and multivariate analysis enable businesses to evaluate various alternatives and choose strategies that maximize outcomes. Moreover, data visualization tools, such as dashboards and infographics, help communicate complex statistical findings in accessible formats, facilitating discussions among stakeholders and guiding strategic planning.

Human Resources and Performance Management

The application of statistics in human resources (HR) is also significant, particularly in recruitment, performance management, and employee satisfaction. Statistical analysis helps HR professionals evaluate the effectiveness of hiring processes, assess employee performance metrics, and identify trends in employee turnover. By using tools like employee surveys and performance reviews, organizations can gather data on job satisfaction and engagement levels. These insights allow businesses to implement targeted initiatives aimed at improving workplace culture, enhancing employee retention, and fostering talent development.

Marketing Analytics and Campaign Performance

In today's digital age, statistics is indispensable in evaluating marketing effectiveness and optimizing campaigns. Businesses analyze data from various channels, including social media, email marketing, and online advertising, using statistical methods to measure key performance indicators (KPIs) such as conversion rates, click-through rates, and return on investment (ROI). A/B testing, a statistical method used to compare two versions of a marketing campaign, helps organizations determine which strategies resonate more with their target audience. This data-driven approach not only improves marketing effectiveness but also enables businesses to allocate resources more efficiently.

Supply Chain and Operations Management

Statistics is vital in supply chain management (SCM) and operations management, where data-driven decisions can significantly enhance efficiency. Statistical methods enable organizations to analyze operational data, forecast demand, and optimize resource allocation. Techniques such as linear programming and inventory models help businesses minimize costs while ensuring that products are available to meet customer demand. By employing statistical analysis to streamline logistics, improve supplier relationships, and enhance production planning, organizations can reduce waste, lower lead times, and improve overall operational performance.

Conclusion

In conclusion, the application of statistics in business and commerce is expansive and transformative. From market research and quality control to financial analysis and strategic planning, statistical methods equip organizations with the necessary tools to analyze

Q. 3 a) Describe different measures of central tendency. (10+10)

In statistics, a **central tendency** (or **measure of central tendency**) is a central or typical value for a probability distribution. It may also be called a **center** or **location** of the distribution. Colloquially, measures of central tendency are often called averages. The term central tendency dates from the late 1920s. The most common measures of central tendency are the arithmetic mean, the median and the mode. A central tendency can be calculated for either a finite set of values or for a theoretical distribution, such as the normal distribution. Occasionally authors use central tendency to denote "the tendency of quantitative data to cluster around some central value.

The central tendency of a distribution is typically contrasted with its dispersion or variability; dispersion and central tendency are the often characterized properties of distributions. Analysts may judge whether data has a strong or a weak central tendency based on its dispersion.

Measures

The following may be applied to one-dimensional data. Depending on the circumstances, it may be appropriate to transform the data before calculating a central tendency. Examples are squaring the values or taking logarithms. Whether a transformation is appropriate and what it should be, depend heavily on the data being analyzed.

Arithmetic mean or simply, **mean**

the sum of all measurements divided by the number of observations in the data set.

Median

the middle value that separates the higher half from the lower half of the data set. The median and the mode are the only measures of central tendency that can be used for ordinal data, in which values are ranked relative to each other but are not measured absolutely.

Mode

the most frequent value in the data set. This is the only central tendency measure that can be used with nominal data, which have purely qualitative category assignments.

Geometric mean

the nth root of the product of the data values, where there are n of these. This measure is valid only for data that are measured absolutely on a strictly positive scale.

Harmonic mean

the reciprocal of the arithmetic mean of the reciprocals of the data values. This measure too is valid only for data that are measured absolutely on a strictly positive scale.

Weighted arithmetic mean

an arithmetic mean that incorporates weighting to certain data elements.

Purpose of measures of central tendency

Measures of Central Tendency provide a summary measure that attempts to describe a whole set of data with a single value that represents the middle or centre of its distribution. There are three main measures of central tendency: the mean, the median and the mode. When data is normally distributed, the mean, median and mode should be identical, and are all effective in showing the most typical value of a data set.

Desirable qualities of measure of central tendency

Desirable qualities of a good measure of central tendency are:-

1. It should be rigidly defined.
2. It should include all observations.
3. it should be simple to understand and easy to calculate.
4. it should be capable of further mathematical treatment.
5. It should be least affected by extreme observations.
6. it should possess sampling stability.

b) Here are the ages of forty-eight members of a country service program:

		83	**51**	**66**	**61**	**82**	**65**	**54**	**56**	**92**	**60**	**65**	**87**	**68**
64	**51**	**70**	**75**	**66**	**74**	**68**	**44**	**55**	**78**	**69**	**98**	**67**	**82**	**77**
79	**62**	**38**	**88**	**76**	**99**	**84**	**47**	**60**	**42**	**66**	**74**	**91**	**71**	**83**
80	**68**	**65**	**51**	**56**										

Obtain (i) median, (ii) mode (iii) range and coefficient of range.

To analyze the ages of the forty-eight members of the country service program, we will calculate the median, mode, range, and coefficient of range using the following data:

Ages:
83, 51, 66, 61, 82, 65, 54, 56, 92, 60, 65, 87, 68, 64, 51, 70, 75, 66, 74, 68, 44, 55, 78, 69, 98, 67, 82, 77, 79, 62, 38, 88, 76, 99, 84, 47, 60, 42, 66, 74, 91, 71, 83, 80, 68, 65, 51, 56

Step 1: Organize the Data
First, let's arrange the ages in ascending order:

38, 42, 44, 47, 51, 51, 51, 51, 54, 55, 56, 56, 60, 60, 61, 62, 65, 65, 65, 65, 66, 66, 66, 66, 67, 68, 68, 68, 68, 69, 70, 74, 74, 75, 76, 77, 78, 79, 80, 82, 82, 83, 83, 84, 87, 88, 91, 92, 98, 99

Step 2: Calculate the Median
To find the median, we first identify the middle value(s). Since there are 48 data points (an even number), the median will be the average of the 24th and 25th values in the ordered list.

- 24th value: 66
- 25th value: 67

Now, calculate the median:

$$\text{Median} = \frac{66+67}{2} = \frac{133}{2} = 66.5$$

Step 3: Calculate the Mode
The mode is the value that appears most frequently in the dataset.

From the ordered list:

- The frequency of 51: 4 times
- The frequency of 65: 4 times

- The frequency of 66: 4 times

Since there are multiple values (51, 65, 66) that each occur 4 times, we can say the data set is **multimodal** with modes 51, 65, and 66.

Step 4: Calculate the Range
The range is calculated by subtracting the smallest value from the largest value.

- Smallest value: 38
- Largest value: 99

Now, calculate the range:
Range=99−38=61\text{ Range} = 99 - 38 = 61Range=99−38=61

Step 5: Calculate the Coefficient of Range
The coefficient of range is calculated using the following formula:

$$\text{Coefficient of Range} = \frac{\text{Range}}{\text{Maximum value} + \text{Minimum value}}$$

Here, the maximum value is 99, and the minimum value is 38.

Now calculate:

$$\text{Coefficient of Range} = \frac{61}{99+38} = \frac{61}{137} \approx 0.445$$

Summary of Results

- Median: 66.5
- Mode: 51, 65, 66 (multimodal)
- Range: 61
- Coefficient of Range: approximately 0.445

Q. 4 Find mean deviation from mean and quartile deviation from the following data: (20)

Classes	1 – 7	8 – 14	15 – 21	22 – 28	29 – 35	36 – 42	43 – 49
Frequency	45	32	34	22	20	12	9

To calculate the mean deviation from the mean and the quartile deviation for the given frequency distribution, we will follow these steps:

Step 1: Create a Frequency Table with Midpoints

We will first find the midpoint for each class interval, then multiply the midpoint by the corresponding frequency to calculate the total.

Class Interval	**Frequency (f)**	**Midpoint (x)**	**f * x**
1 – 7	45	4	180
8 – 14	32	11	352
15 – 21	34	18	612
22 – 28	22	25	550
29 – 35	20	32	640
36 – 42	12	39	468
43 – 49	9	46	414
Total	**Number of Frequencies (N): 274**		**2826**

Step 2: Calculate the Mean

The mean (x¯\bar{x} x¯) can be calculated using the formula:

The mean ($\bar{x}$) can be calculated using the formula:

$$\bar{x} = \frac{\sum(f \cdot x)}{N}$$

Substituting the values:

$$\bar{x} = \frac{2826}{274} \approx 10.32$$

Step 3: Calculate Mean Deviation from the Mean

Now we will calculate the mean deviation from the mean, using the formula for the mean deviation (MD):

$$MD = \frac{\sum f|x - \bar{x}|}{N}$$

a) Calculate $|x-\bar{x}|$ and $f|x-\bar{x}|$

We will add a new column to our table for $|x-\bar{x}|$ and $f|x-\bar{x}|$:

Class Interval	Frequency (f)	Midpoint (x)	f * x	|x - \bar{x}|	f * |x - \bar{x}|
1 - 7	45	4	180	|4 - 10.32| = 6.32	284.4
8 - 14	32	11	352	|11 - 10.32| = 0.68	21.76
15 - 21	34	18	612	|18 - 10.32| = 7.68	261.12
22 - 28	22	25	550	|25 - 10.32| = 14.68	323.36
29 - 35	20	32	640	|32 - 10.32| = 21.68	433.6
36 - 42	12	39	468	|39 - 10.32| = 28.68	344.16
43 - 49	9	46	414	|46 - 10.32| = 35.68	321.12
Total	**274**		**2826**	**N/A**	**NDA**

Now sum up the last column:

$$\sum f|x-\bar{x}| \approx 284.4 + 21.76 + 261.12 + 323.36 + 433.6 + 344.16 + 321.12 = 1955.32$$

b) Now calculate the Mean Deviation (MD)

$$MD = \frac{1955.32}{274} \approx 7.13$$

Step 4: Calculate Quartile Deviation

The quartile deviation is calculated using the following formula:

The quartile deviation is calculated using the following formula:

$$QD = \frac{Q3 - Q1}{2}$$

a) Find $Q1$ and $Q3$

1. **Calculate $Q1$:** $N/4 = 274/4 = 68.5$
   The cumulative frequency before $Q1$ will determine the class containing $Q1$. From our frequency distribution:

   - For the class $1 - 7$: Cumulative frequency = 45
   - For the class $8 - 14$: Cumulative frequency = 77

   $Q1$ is in the class $8 - 14$.

   Using the formula for $Q1$:

$$Q1 = L + \left(\frac{\frac{N}{4} - CF}{f}\right) \cdot h$$

Where:

- L = lower boundary of $Q1$ class = 8
- CF = cumulative frequency before $Q1$ class = 45
- f = frequency of $Q1$ class = 32
- h = width of the class = 7 (since $8 - 14$)

Substituting into the formula:

$$Q1 = 8 + (\frac{68.5 - 45}{32}) \cdot 7$$

$$Q1 = 8 + (\frac{23.5}{32}) \cdot 7 \approx 8 + 5.15625 \approx 13.15625$$

2. **Calculate** $Q3$: $3N/4 = 3 \cdot 274/4 = 205.5$

   Cumulative frequency before $Q3$:

   - For the class $22 - 28$: 77
   - For the class $29 - 35$: 97
   - For the class $36 - 42$: 109
   - For the class $43 - 49$: 118
   - For the class $43 - 49$: Cumulative frequency = 205

   $Q3$ is in class $36 - 42$.

   Using the formula for $Q3$:

$$Q3 = L + \left(\frac{\frac{3N}{4} - CF}{f}\right) \cdot h$$

   Where:

   - $L = 36$
   - $CF = 109$
   - $f = 12$
   - $h = 7$

Substituting values:

$$Q3 = 36 + (\frac{205.5 - 109}{12}) \cdot 7$$

$$Q3 = 36 + (\frac{96.5}{12}) \cdot 7 \approx 36 + 56.04167 \approx 92.04167$$

b) Calculate the Quartile Deviation

$$QD = \frac{Q3 - Q1}{2} = \frac{13.15625 - 92.04167}{2} \approx \frac{78.88542}{2} \approx 39.44271$$

Summary of Results

- Mean Deviation from Mean: **7.13**
- Quartile Deviation: **39.44**

Q. 5 a) Discuss the basic concepts in the hypothesis-testing procedure. (10+10)

The Hypothesis is an assumption which is tested to check whether the inference drawn from the sample of data stand true for the entire population or not.

Hypothesis Testing Procedure

The following steps are followed in hypothesis testing:

Set up a Hypothesis: The first step is to establish the hypothesis to be tested. The statistical hypothesis is an assumption about the value of some unknown parameter, and the hypothesis provides some numerical value or range of values for the parameter. Here two hypotheses about the population are constructed **Null Hypothesis** and **Alternative Hypothesis**.

The Null Hypothesis denoted by H_0 asserts that there is no true difference between the sample of data and the population parameter and that the difference is accidental which is caused due to the fluctuations in sampling. Thus, a null hypothesis states that there is no difference between the assumed and actual value of the parameter.

The alternative hypothesis denoted by H_1 is the other hypothesis about the population, which stands true if the null hypothesis is rejected. Thus, if we reject H_0 then the alternative hypothesis H_1 gets accepted.

Set up a Suitable Significance Level: Once the hypothesis about the population is constructed the researcher has to decide the level of significance, i.e. a confidence level with which the null hypothesis is accepted or rejected. The significance level is denoted by **'α'** and is usually defined before the samples are drawn such that results obtained do not influence the choice. In practice, we either take 5% or 1% level of significance.

If the 5% level of significance is taken, it means that there are five chances out of 100 that we will reject the null hypothesis when it should have been accepted, i.e. we are about 95% confident that we have made the right decision. Similarly, if the 1% level of significance is taken, it means that there is only one chance out of 100 that we reject the hypothesis when it should have been accepted, and we are about 99% confident that the decision made is correct.

Determining a Suitable Test Statistic: After the hypothesis are constructed, and the significance level is decided upon, the next step is to determine a suitable test statistic and its distribution. Most of the statistic tests assume the following form:

$$\text{Test Statistic} = \frac{\text{Sample Statistic} - \text{Hypothesized Parameter}}{\text{Standard Error of the Statistic}}$$

Determining the Critical Region: Before the samples are drawn it must be decided that which **values to the test statistic** will lead to the acceptance of H_0 and which will lead to its rejection. The values that lead to rejection of H_0 is called the critical region.

Performing Computations: Once the critical region is identified, we compute several values for the random sample of size 'n.' Then we will apply the formula of the test statistic as shown in step (3) to check whether the sample results falls in the acceptance region or the rejection region.

Decision-making: Once all the steps are performed, the statistical conclusions can be drawn, and the management can take decisions. The decision involves either accepting the null hypothesis or rejecting it. The decision that the null hypothesis is accepted or rejected depends on whether the computed value falls in the acceptance region or the rejection region.

Hypothesis testing of the equality between two population means

This is a two sample z test which is used to determine if two population means are equal
Procedure

- Sampling is from normally distributed populations with known variances

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)_0}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}}$$

- Sampling from normally distributed populations where population variances are unknown
  - population variances equal

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)_0}{\sqrt{s_p^2/n_1 + s_p^2/n_2}}$$

This is with t distributed as Student's t distribution with $(n_1 + n_2-2)$ degrees of freedom and a pooled variance.

  - population variances unequal

When population variances are unequal, a distribution of t' is used in a manner similar to calculations of confidence intervals in similar circumstances.

- Sampling from populations that are not normally distributed

If both sample sizes are 30 or larger the central limit theorem is in effect. The test statistic is

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)_0}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}}$$

Thus, to test the hypothesis, it is necessary to follow these steps systematically so that the results obtained are accurate and do not suffer from either of the statistical error Viz. Type-I error and Type-II error.

b) The average commission charged by full-service brokerage firms on a sale of common stock is $144 and the standard deviation is $52. Joel Freelander has taken a random sample of 121 trades by his clients and determined that they paid an average commission of $151. At a 0.10 significance level, can Joel conclude that his clients ‘commissions are higher than the industry average?

To determine if Joel Freelander's clients' commissions are higher than the industry average, we can perform a hypothesis test for the mean using the following steps:

Step 1: Set up the Hypotheses

- Null Hypothesis (H0H_0*H*0): The mean commission paid by Joel's clients is equal to the industry average.

$$H_0 : \mu = 144$$

- Alternative Hypothesis (H_a): The mean commission paid by Joel's clients is higher than the industry average.

$$H_a : \mu > 144$$

This is a right-tailed test.

Step 2: Given Information

- Industry average commission (μ0\mu_0*μ*0) = 144
- Standard deviation (σ\sigma*σ*) = 52
- Sample size (nn*n*) = 121
- Sample mean (x¯\bar{x}*x*¯) = 151
- Significance level (α\alpha*α*) = 0.10

Step 3: Calculate the Test Statistic

We will use the z-test for the sample mean since the population standard deviation is known. The formula for calculating the z-test statistic is:

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$$

Substituting in the provided values:

$$z = \frac{151 - 144}{52/\sqrt{121}}$$

Calculating the denominator:

$$\sigma/\sqrt{n} = \frac{52}{11} \approx 4.7273$$

Now substituting this into the z test formula:

$$z = \frac{7}{4.7273} \approx 1.480$$

Step 4: Determine the Rejection Region

For a significance level of $\alpha = 0.10$ in a right-tailed test, we can find the critical z-value from the z-table or standard normal distribution table.

- For $\alpha = 0.10$, the critical z-value (cut-off point) is approximately $z_{0.10} \approx 1.2816$.

Step 5: Make the Decision

Step 5: Make the Decision

- If the calculated z-value is greater than the critical z-value, we reject the null hypothesis.

Calculated $z = 1.480$

Since $1.480 > 1.2816$, we reject the null hypothesis.

Step 6: Conclusion

At the 0.10 significance level, there is sufficient evidence to conclude that Joel Freelander's clients' commissions are higher than the industry average of $144.