DV QP CAT 1 Answers

1.b) Differentiate between Descriptive and Inferential Statistics.

Ans. 1.b)

S.No.	Descriptive Statistics	Inferential Statistics
1.	It gives information about raw data which describes the data in some manner.	It makes inferences about the population using data drawn from the population.
2.	It helps in organizing, analyzing, and to present data in a meaningful manner.	It allows us to compare data and make hypotheses and predictions.
3.	It is used to describe a situation.	It is used to explain the chance of occurrence of an event.
4.	It explains already known data and is limited to a sample or population having a small size.	It attempts to reach the conclusion about the population.
5.	It can be achieved with the help of charts, graphs, tables, etc.	It can be achieved by probability.

1.c) Explain any three data types of the R-object.

Ans.1.c)

There are 5 basic data types of objects in the R language:

1. **Vectors**
   Atomic vectors are one of the basic types of objects in R programming. Atomic vectors can store homogeneous data types such as character, doubles, integers, raw, logical, and complex. A single element variable is also said to be vector.
2. **Lists**
   List is another type of object in R programming. List can contain heterogeneous data types such as vectors or another lists.
3. **Matrices**
   To store values as 2-Dimensional array, matrices are used in R. Data, number of rows and columns are defined in the matrix() function.
4. **Factors**
   Factor object encodes a vector of unique elements (levels) from the given data vector.
5. **Arrays**
   array() function is used to create n-dimensional array. This function takes dim attribute as an argument and creates required length of each dimension as specified in the attribute.

2.b) Explain the statement of Hypothesis in detail.

Ans.2.b)

Statement of Hypothesis

A statistical hypothesis is defined as a statement, which may or may not be true about the population parameter or about the probability distribution of the parameter that we wish to validate on the basis of sample information. Most times, experiments are performed with random samples instead of the entire population and inferences drawn from the observed results are then generalised over to the entire population. But before drawing inferences about the population, it should be always kept in mind that the observed results might have come due to chance factor. In order to have an accurate or more precise inference, the chance factor should be ruled out.

Null Hypothesis

The probability of chance occurrence of the observed results is examined by the null hypothesis (H0). Null hypothesis is a statement of no differences. The other way to state null hypothesis is that the two samples came from the same population. Here, we assume that population is normally distributed and both the groups have equal means and standard deviations. Since the null hypothesis is a testable proposition, there is counter proposition to it known as alternative hypothesis and denoted by H1 . In contrast to null hypothesis, the alternative hypothesis (H1) proposes that

i) the two samples belong to two different populations,
ii) their means are estimates of two different parametric means of the respective population,
iii) there is a significant difference between their sample means.

2.c) Explain in short the following term.

i) Random Variables **ii) Normal Probability Distribution**

Ans.2.c)

i) Random Variables

A random variable, X, is a variable whose possible values are numerical outcomes of a random phenomenon. There are two types of random variables, discrete and continuous.

Example of Random variable

- A person's blood type
- Number of leaves on a tree
- Number of times a user visits LinkedIn in a day
- Length of a tweet.

Types of Random Variables

a. **Discrete Random Variables :**

A discrete random variable is one which may take on only a countable number of distinct values such as 0,1,2,3,4,..... Discrete random variables are usually counts. If a random variable can take only a finite number of distinct values, then it must be discrete. Examples of discrete random variables include the number of children in a family, the Friday night attendance at a cinema, the number of patients in a doctor's surgery, the number of defective light bulbs in a box of ten.

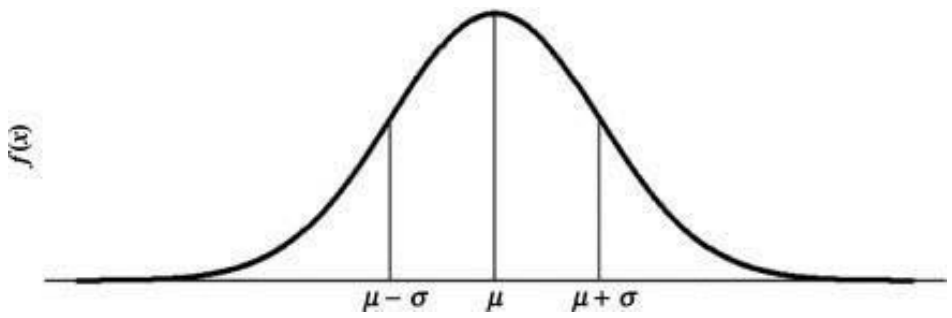
b. **Continuous Random Variables:**

A continuous random variable is one which takes an infinite number of possible values. Continuous random variables are usually measurements. Examples include height, weight, the amount of sugar in an orange, the time required to run a mile. A continuous random variable is not defined at specific values. Instead, it is defined over an interval of values, and is represented by the area under a curve (known as an integral). The probability of observing any single value is equal to 0, since the number of values which may be assumed by the random variable is infinite.

ii) Normal Probability Distribution

The Bell-Shaped Curve

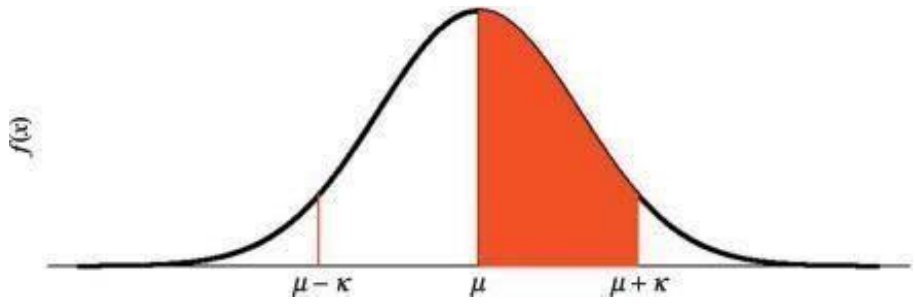
The Bell-shaped Curve is commonly called the normal curve and is mathematically referred to as the Gaussian probability distribution. Unlike Bernoulli trials which are based on discrete counts, the normal distribution is used to determine the probability of a continuous random variable.



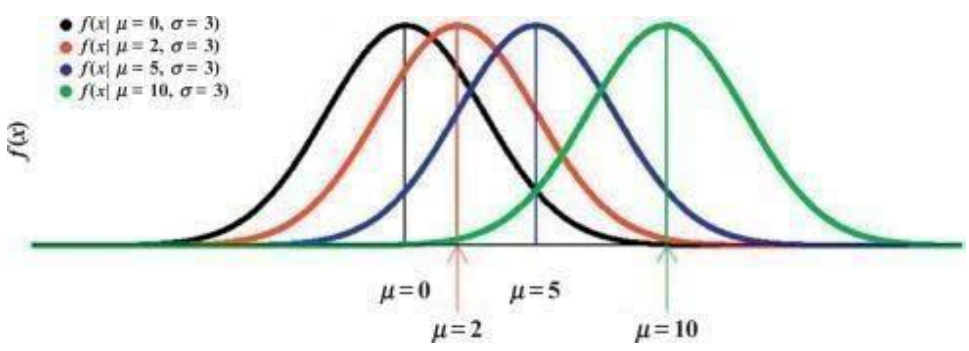
The normal or Gaussian Probability Distribution is most popular and important because of its unique mathematical properties which facilitate its application to practically any physical problem in the real world. The constants μ and σ^2 are the parameters:

- "μ" is the population true mean (or expected value) of the subject phenomenon characterized by the continuous random variable, X,
- "σ^2" is the population true variance characterized by the continuous random variable, X.
- Hence, "σ" the population standard deviation characterized by the continuous random variable X;
- The points located at μ−σ and μ+σ are the points of inflection; that is, where the graph changes from cupping up to cupping down.

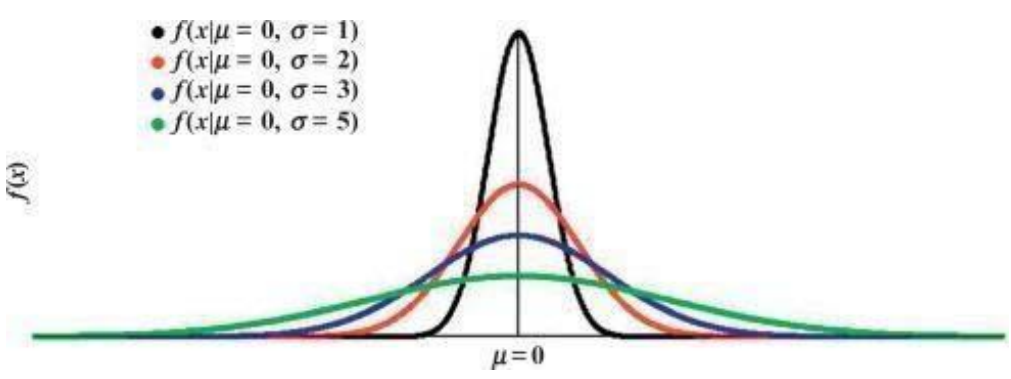
The normal curve graph of the normal probability distribution) is symmetric with respect to the mean μ as the central position. That is, the area between μ and κ units to the left of μ is equal to the area between μ and κ units to the right of μ.



There is not a unique normal probability distribution. The figure below is a graphical representation of the normal distribution for a fixed value of σ2 with μ varying.



The figure below is a graphical representation of the normal distribution for a fixed value of μ with varying σ2 .



3.b) Explain Scatter Plot ? Also explain the advantage and limitation of scatter plot.

Ans. 3.b)

Scatter Plot

A Scatter Plot is a graph in which the values of two variables are plotted along two axes, the pattern of the resulting points revealing any correlation present. With scatter plots we can explain how the variables relate to each other. Which is defined as correlation. Positive, Negative, and None (no correlation) are the three types of correlation.

Advantages of a Scatter Plot

- Relationship between two variables can be viewed.
- For non-linear pattern, this is the best method.
- Maximum and minimum value can be easily determined.
- Observation and reading are easy to understand
- Plotting the diagram is very simple

Limitations of a Scatter Plot

- With Scatter diagrams we cannot get the exact extent of correlation.
- Quantitative measure of the relationship between the variable cannot be viewed.
- Only shows the quantitative expression.
- The relationship can only show for two variables.

3.c) What is dataframe and how it is created in R ?

Ans.3.c)

- A DataFrame is a way to represent and work with tabular data.
- Tabular data has rows and columns, just like our csv file.
- In order to read in the data, we'll need to use the pandas.read_csv function.
- This function will take in a csv file and return a DataFrame.
- Data Frames are data displayed in a format as a table.
- Data Frames can have different types of data inside it.
- While the first column can be character, the second and third can be numeric or logical.
- However, each column should have the same type of data.

Create DataFrame in R :

Use the data.frame() function to create a data frame:

```
# Create a data frame
Data_Frame <- data.frame (
  Training = c("Strength", "Stamina", "Other"),
  Pulse = c(100, 150, 120),
  Duration = c(60, 30, 45)
)

# Print the data frame
Data_Frame
```

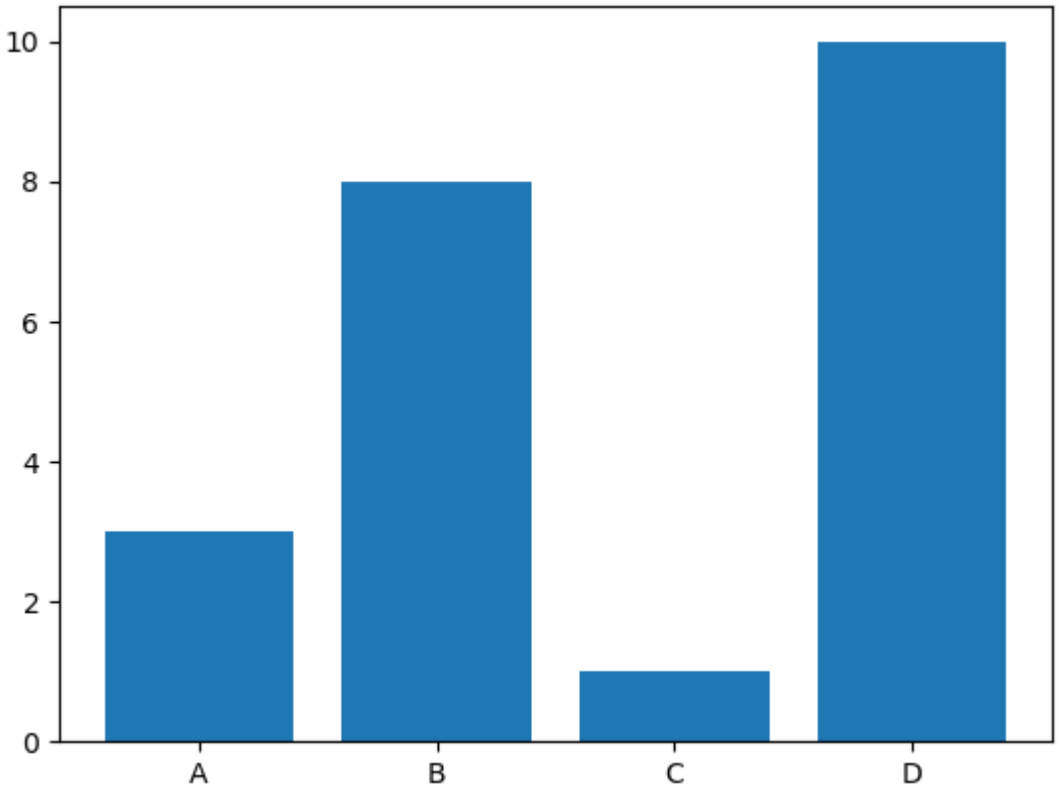
4.b) Explain Bar plot in detail.

Ans.4.b)

Bar Plot

A barplot (or barchart) is one of the most common type of graphic. It shows the relationship between a numeric variable and a categoric variable.

Bar Plot are classified into four types of graphs - bar graph or bar chart, line graph, pie chart, and diagram.



Limitations of Bar Plot:

- When we try to display changes in speeds such as acceleration, Bar graphs won't help us.

Advantages of Bar plot:

- Bar charts are easy to understand and interpret.
- Relationship between size and value helps for in easy comparison.
- They're simple to create.
- They can help in presenting very large or very small values easily.

4.c) Explain Box and Histogram plot with advantage and limitation.

Ans.4.c)

Histogram

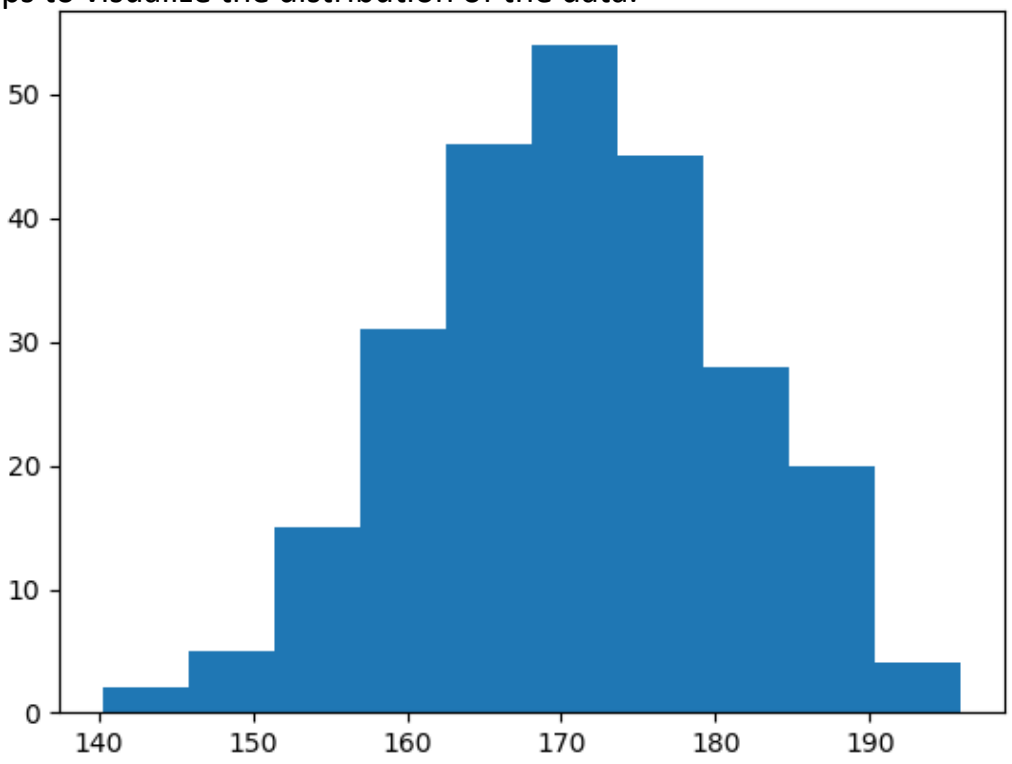
A histogram represents the frequency distribution of continuous variables. While, a bar graph is a diagrammatic comparison of discrete variables. Histogram presents numerical data whereas bar graph shows categorical data. The histogram is drawn in such a way that there is no gap between the bars.

Limitations of Histogram:

- A histogram can present data that is misleading as it has many bars.
- Only two sets of data are used, but to analyze certain types of statistical data, more than
- two sets of data are necessary

Advantages of Histogram:

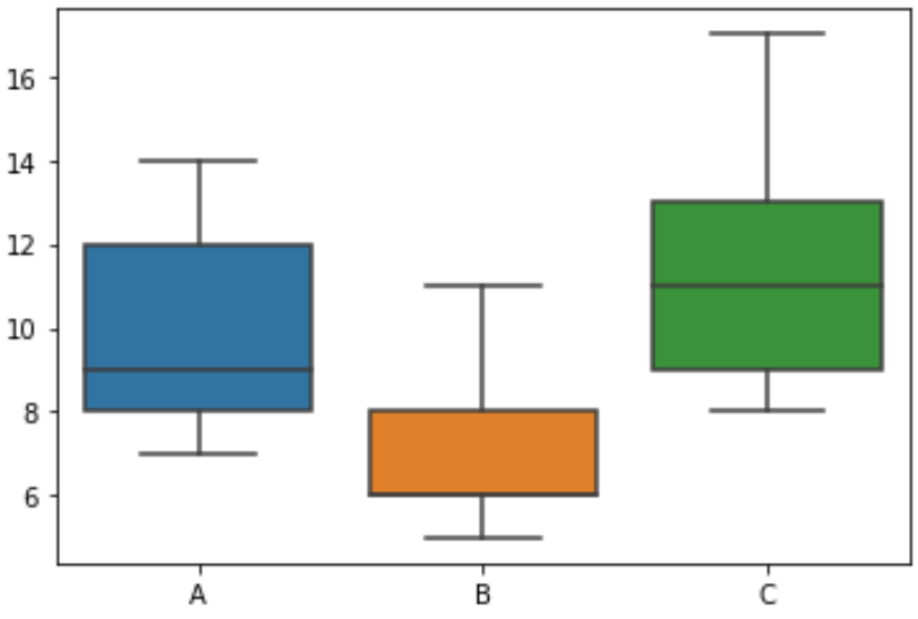
- Histogram helps to identify different data, the frequency of the data occurring in the dataset and categories which are difficult to interpret in a tabular form.
- It helps to visualize the distribution of the data.



Box Plot

A box plot is a way of statistically representing the distribution of the data through five main dimensions :

- Minimum: Smallest number in the dataset.
- First quartile: Middle number between the minimum and the median.
- Second quartile (Median): Middle number of the (sorted) dataset.
- Third quartile: Middle number between median and maximum.
- Maximum: Highest number in the dataset.



Advantages:

- The box plot organizes large amounts of data and visualizes outlier values.

Disadvantages:

- The box plot is not relevant for detailed analysis of the data as it deals with a summary of the data distribution.

5.b) What are the features and applications of Python?

Ans.5.b)

Features of Python :

1. Free and Open Source
2. Easy to code
3. Easy to Read
4. Object-Oriented Language
5. GUI Programming Support
6. High-Level Language
7. Large Community Support
8. Easy to Debug
9. Python is a Portable language
10. Python is an Integrated language
11. Interpreted Language:
12. Large Standard Library
13. Dynamically Typed Language
14. Frontend and backend development
15. Allocating Memory Dynamically

Applications of Python :

1. Web Development
2. Machine Learning and Artificial Intelligence
3. Data Science
4. Game Development
5. Audio and Visual Applications
6. Software Development
7. CAD Applications
8. Business Applications
9. Desktop GUI
10. Web Scraping Application

Some other real-world applications of Python are:

- *Robotics and automation by the use of inbuilt libraries and tools like PyDy, Dart, PyRobot, and pyro.*
- *Image processing: some of the amazing libraries and tools for image processing are Blender, OpenCV, Houdini, and PIL.*
- *Scientific applications are facilitated by popular libraries like Pandas, Matplotlib, SciPy, and many more*

6.b) What is dictionary? Explain the methods available in dictionary.

Ans.6.b)

- Dictionaries are used to store data values in key:value pairs.
- A dictionary is a collection which is ordered*, changeable and do not allow duplicates.
- Dictionaries are written with curly brackets and have keys and values.

The methods available in dictionary are

Functions Name	Descriptions
clear()	Removes all items from the dictionary
copy()	Returns a shallow copy of the dictionary
fromkeys()	Creates a dictionary from the given sequence
get()	Returns the value for the given key
items()	Return the list with all dictionary keys with values
keys()	Returns a view object that displays a list of all the keys in the dictionary in order of insertion
	Returns and removes the element with