# Victoria University of Bangladesh
# Mid Term Assessment
# Summer Semester 2023

**Name : MD Sujan ali**

ID : 1119460091

Program : BBA

Batch : 46th

COURSE CODE : STA 220

COURSE TITLE : Business Statistics

**Submitted to : TANIYA NASHIN**

1

Course title : Business Statistics
Course code : STA - 220

## Ans : to the Q.N (1)

(a) Define statistics

Statistics is a branch of mathematics that involves collecting, organizing, analyzing interpreting, and presenting data. It helps us understand and draw conclusions from information obtained by using techniques such as graphs, charts, and various statistics measures.

Example : A statistics is a number that represents a property of the sample. For example, if we consider one math class to be a sample of the population of all math classes. Then the average number of points earned by students in that one math class at the end of the term is an example of a statistics

P.t.o

Ans: to the q.n (2)

(2) Shapes of histogram with proper diagram: A histogram is type of chart that allows us to visualize the distribution of values in a dataset.

The x-axis displays the values in the dataset and the y-axis shows the frequency of each value.

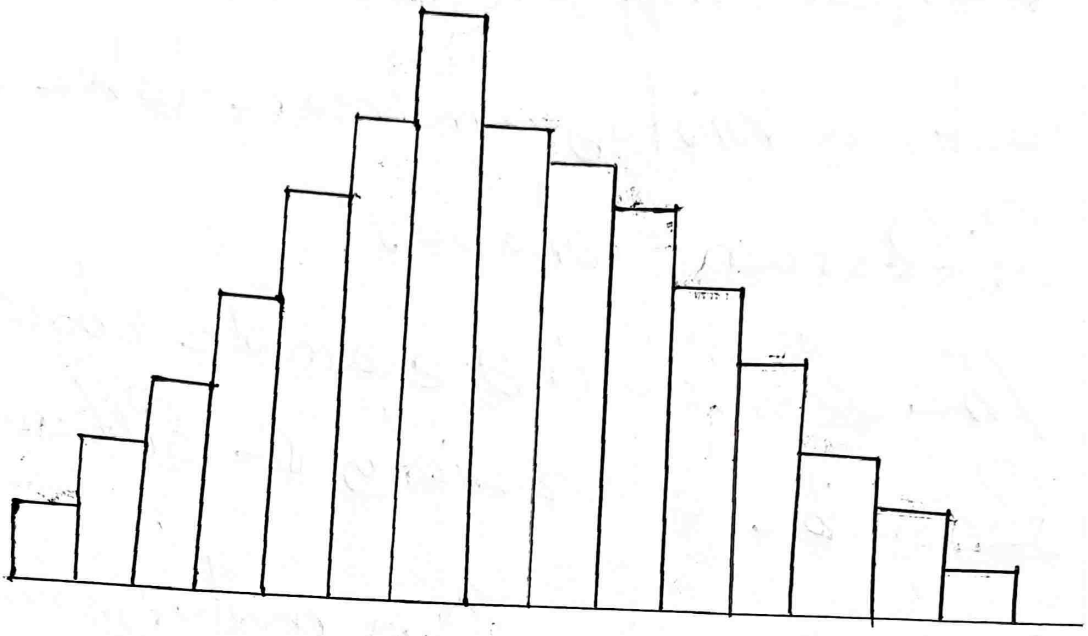
p.t.o

Depending on the values in the data set, a histogram can take on many different shapes.

The following example show how to describe a variety of different histogram

① Bell-shaped: A histogram is bell-shaped if it resembles a bell curve and has one single peak in the middle of the distribution. The most common real-life example of this type of distribution is the normal distribution curve. This is a symmetric distribution where data clusters around a central value, creating a bell-shaped curve

p.t.o



Uniform Distribution:

Also known as a rectangular distribution. all values occur with approximately the same frequency, resulting in a flat histogram.

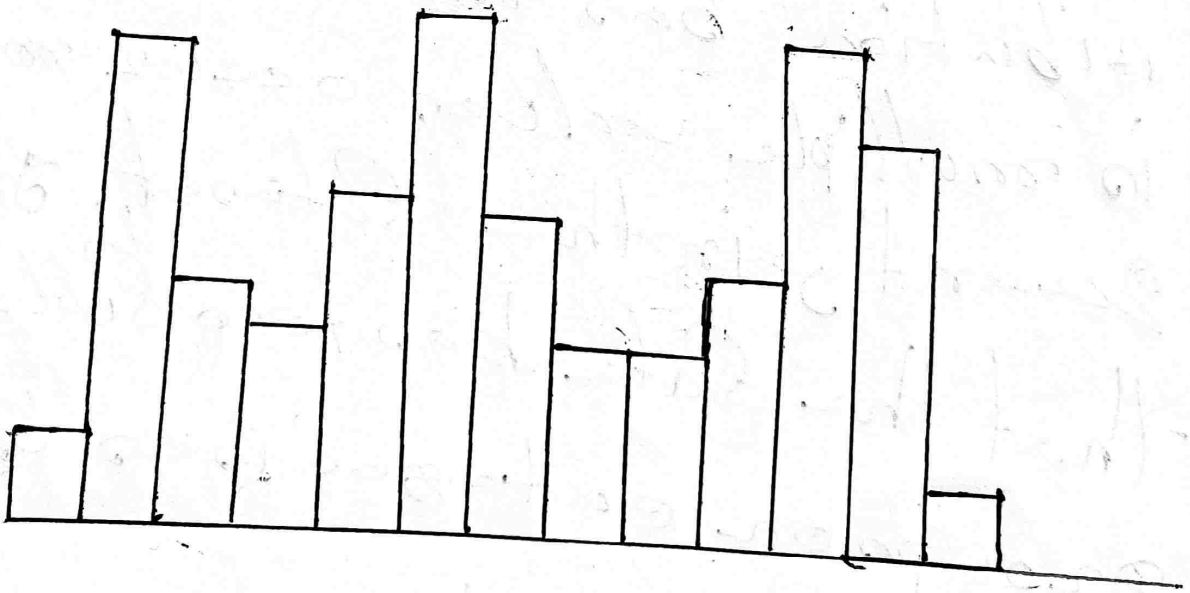
A histogram is described as uniform if every value in a dataset occurs roughly the same number of times. this type of histogram

p.t.o

Bimodal: A histogram is described as bimodal if it has two distinct peaks. We often say that this type of distribution has multiple modes-that is multiple values occur most frequently in the dataset. Indicating that the data has two different groups or categories within

p.t.o

Multimodal: A histogram is described as multimodal if it has more than two distinct peaks.



p.t.o

Exponential: This distribution shows a rapid drop off in frequency as values increase, often used to represent event that occur rarely but with a decreasing likelihood over time.

Skewed: A distribution can be positively skewed (tail on the right) or negatively skewed (tail on the left side, with a few larger values on the right, and vice versa for negatively skewed data.

P.t.o

Remember that the shape of a histogram can provide insights into the underlying data distribution and help you understand the characteristics of the dataset. If you want to visualize these shapes, you can use graphing software to create a histogram based on your data.

Ans:- to the Q.N (3) - (a)

(A) Pie Chart :-

A piechart is a circular graph used to display data as a series of slices, where each slice represents a portion or percentage of the whole. The size of each slice corresponds to the value it represents in relation to the total, making it a useful tool for visualizing relative proportions within a dataset.

The Pie chart is also known as a circle chart dividing the circular statistics graphic into sectors or sections to illustrate the numerical problems

P.t.o

## Formula

The Pie chart is an important type of data representation. It contains different segments and sectors in which each segment and sector of a pie chart forms a percentage. The sum of all the data is equal to work out with the percentage for a pie chart follow the steps:

* categorize the data
* calculate the total
* Divide the categories
* convert into percentages
* Final, calculate the degrees

| Football | Hockey | cricket | Basketball | Badminton |
|---|---|---|---|---|
| 10 | 5 | 5 | 10 | 10 |

(B) Histograms

A histogram is a graphical representation of data distribution. It displays the frequency of data points falling within certain intervals, known as bins. Each bar in the histogram represents the number of data points that fall within a specific range or bin. It is commonly used to visualize the shape and spread of data, helping to identify patterns, outliers, and trends. The height of a rectangle in the vertical axis represents the distribution frequency of a variable the amount or how often that

P.T.O

## histogram used:

The histogram is a popular graphing tool. It is used to summarize discrete or continuous data that are measured on an interval scale. It is often used to illustrate the major features of the distribution of the data in a convenient form

## histogram types:

* Histogram of equal class interval
* Histogram of unequal class interval.

## (c) Mean:

In the context of business statistics, the term "Mean" refers to the average of a set of numerical values. It is calculated by adding up all the values and then dividing by the total number of values in the set. The mean is commonly used to represent the central tendency of a data set and provides insight into its overall value. Mean is the average of the given numbers and is calculated by dividing the sum of given numbers by the total number of numbers.

P.t.o

Mean = (sum of all the observations / total number of observations)

Example:
what is the mean of 2, 4, 6, 8 and 10?

Solution
first add all the numbers.
2+4+6+8+10 = 30
now divide by 5 (total number of observations)
Mean = 30/5 = 6

Mean Symbol (x bar)
The symbol mean is usually given by the symbol $\bar{x}$ the bar above the letter x, represent the mean of a number of values,

$\bar{x}$ = (sum of values ÷ number of values)

$\bar{x} = (x_1 + x_2 + x_3 + \ldots + x_n)/n$

(D) Median:-

In statistics and Probability theory, the median is the value separating the hi- gher half from the lower half of a data sample, a population, or a Probability distribution. For a data set it may be thought of as the middle value. The basic feature of the median in describing data compared to the mean often simply described as the average is that it is not skewed by a small propo- rtion of extremely large or me- small values, and therefore pro- vides a better representation of

P.t.o

the center. Median income, for example, may be a better way to describe center of the income distribution because increases in the largest incomes alone have no effect on median. For this reason the median is of central importance in robust statistics.

1, 3, 3, 6, 7, 8, 9

Median = 6

1, 2, 3, 4, 5, 6, 8, 9

Median = (4+5) ÷ 2

= 4.5

Finding the median in sets of data with an odd and even number of values.

Ans: to the Q,N (4)

(4) Different types of variables

A variable is a characteristic that can be measured and that can assume different values. Height, age income, province or country of birth, grades obtained at school and type of housing are all examples of variables. variables may be classified into two main categories, categorical and numerical. Each category is then classified into two subcategories. nominal or ordinal for categorical variables, discrete or continuous for

p.t.o

numerie variable. These types are briefly outlined in this section.

Nominal variables

A nominal variable is one that describes a name, label or category without natural order. Types of dwelling are a nominal variables.

Ordinal variables

An ordinal variables is a variables whose values are defined by an order relation between the different. The variable behaviour is ordinal because the category excellent is better than the category

p.t.o

There are several types of variables in programm

Integers (int): stores whole numbers, both Positive and negative, without decimal Points.

Floating-Point (float) and Double-Precision store numbers with decimal Points. Dou bles have higher precision compared to floats.

Character (char): stores a single character such as words or special symbol.

String (str): stores either true or fa be values, representing binary conditi ons.

P.t.o

Boolean (bool): Stores sequences of characters, such as words or sentences

Array: Stores a collection of values of the same type in a single variables

object: Used in object oriented programming to store both data and methods that operate on the data.

Pointer: stores the memory address of another variable allowing direct access to its value.

Enum: Defines a set of named constant values, often representing different options or states.

p.t.o

Structure Record: Groups multiple variables of different types under a single name.

These variable types help programmers work with different kinds of data efficiently and accurately.

Numeric variables: A numeric variable (also called quantitative variable) is a quantifiable characteristic whose values are numbers (except numbers which are codes standing when categorized). numeric variables may be either continuous or discrete.

p.t.o

## Continuous variables:

A variable is said to be continuous if it can assume an infinite number of real values within a given interval. For instance, consider the height of a student. The height can't take any values. It can't be negative and it can't be higher than three metres. But between 0 and 3, the number of possible values is theoretically infinite. In practice, the method used and the accuracy of the measurement instrument will

P.T.O

restrict the precision of the variable. The reported height would be rounded to the nearest centimetres, so it would be 1.63 centimetres. The age is another example of a continuous variable that is typically rounded down.

Discrete variables:

As opposed to a continuous variable a discrete variable can assume only a finite number of real values within a given interval. An example of a discrete variable would be the score given

p.t.o

by a judge to a gymnast in competition. The range is 0 to 10 and the score is always given to one decimal a score of 8.3. You can enumerate all possible values (0, 0.1, 0.2...) and so the number of possible values is finite: It is not another example of a discrete variable is the number of people in a household or even a household of size 20 or less. The number of possible values is 20, because it's not possible to have a household of

## Ans: to the Q.N (5)

### (5) chebyshev's theorem:

chebyshev's theorem, also known as the chebyshev inequality, is a mathematical theorem that provides bounds on the proportion of data that falls within a certain number of standard deviations from the mean in a probability distribution. Specifically, for any probability distribution with a finite mean and variance, it states that.

p.t.o

At least $(1-1/k^2)$ of the data falls within k standard deviations from the mean, where k is any positive real number greater than 1.

At least $(1-1/k^2)$ of the data falls outside of k standard deviation on the mean.

In essence, Chebyshev's theorem is a general principle that applies to any distribution, whether it is normal or not, and it gives an idea of how data spread out from the mean. It is important

p.t.o

to note that while chebyshevs theorem provides a guarantee about the proportion of data within a range. it does not provide specific information about the actual distribution shape.

Using chebyshevs theorem:

By entering values for k into the equation, I have created the table below that displays proportions for various standard deviations. If you have a mean and standard deviation you might need to know the proportion

P.T.O

| Standard Deviations | Minimum % within | Max % outside |
|---|---|---|
| $\sqrt{2} = 1.41$ | 0.50 | 0.50 |
| 1.5 | 0.56 | 0.44 |
| 2 | 0.75 | 0.25 |
| 3 | 0.89 | 0.11 |
| 4 | 0.94 | 0.06 |
| 5 | 0.96 | 0.04 |

For example, if you're interested in a range of three standard deviations around the mean, Chebyshev's theorem states that at least 89% of the observations fall inside that range, and no more than 11% fall outside that range

P.t.o

A crucial point to notice is that Chebyshev theorem states that at Produces minimum and maximum Proportions. For example, at least 56% of the observations fall inside 1.5 standard deviations, and a mini- mum of 44% fall outside

The theorem does not provide exact answers but it places limited on the possible Proportions. For the example above, more than 56% of the observations can lie within 1.5 standard deviations of the means there then

P.t.o

the minimum and maximum proportions arise due to uncertainties about the data's distribution while the theorem is valuable because it applies to all distribution. it also limits the precision of the results.

## chebyshev's theorem compared to the empirical rule:

the Empirical Rule also describes the proportion of data that fall within a specified number of standard deviations from the mean. however, there

p.t.o

are several crucial differences between chebyshevs theorem and the Empirical Rule.

chebyshevs theorem applies to all probability distribution where you can calculate the mean and standard deviation. on the other hand, the Empirical Rule applies only to the normal distribution.

As you saw above. chebyshevs theorem provides approximations. conversely, the Empirical Rule provides exact answers for the proportions because

P.t.o

the data are known to follow the normal distribution.

| Standard Deviations | Empirical Rule | Chebyshev's theorem |
|---|---|---|
| 1 | 68% | NA |
| 2 | 95% | $\geq 75\%$ |
| 3 | 99.7% | $\geq 88.9\%$ |

Again, notice that the empirical Rule provides exact answers while chebyshevs theorem gives approximations.