ASSIGNMENT

Statistics for Management

QUESTION 1

According to the definition of Statistics given by Croxton and Cowden, what are the four components of Statistics?

1st Component

2nd Component

3rd Component

4th Component

ANSWER 1

Definition by Croxton and Cowden: Statistics may be defined as the science of collection, presentation analysis and interpretation of numerical data from the logical analysis.

It is clear that the definition of statistics by Croxton and Cowden is the most scientific and realistic one.

According to this definition there are four stages:

1st Component

Collection of Data

- First statistical investigation
- Most important step because the entire results will depend upon the quality of this collection
- There are different methods of collection of data such as census, sampling, primary, secondary, etc.
- The investigator should make use of correct method

2nd Component

Presentation of data

- After the data have been collected they are ready for presentation

- Two type od presentation diagrammatic or graphic form.

3rd Component

Analysis of data

- The data presented should be carefully analysed.

4th Component

Interpretation of data

- The last stage in statistical investigation is interpretation
- A valid conclusion must be drawn on the basis of analysis.
- A high degree of skill and experience is necessary for the interpretation

QUESTION 2

What do you mean by Statistical Averages? List various requisites of a Good Average.

Statistical Averages.

Requisites of a Good Average.

ANSWER 2

Statistical Averages

Statistical averages are important in the measurement of quantities that are obscured by random variations.

After the data have been properly checked for its quality, the first and foremost analysis is usually for the descriptive statistics. The general aim is to summarize the data, iron out any peculiarities and perhaps get ideas for a more sophisticated analysis. The data summary may help to suggest a suitable model which in turn suggests an appropriate inferential procedure. The first phase of the analysis will be described as the initial examination of the data or initial data analysis.

It has many things in common with explanatory data analysis which includes a variety of graphical and numerical techniques for exploring data. Thus explanatory data analysis is an essential part of nearly every analysis.

It provides a reasonably systematic way of digesting and summarizing the data with its exact form naturally varies widely from problem to problem. In general, under initial and exploratory data analysis, the following are given due importance.

As an example to motivate the discussion, consider the problem of measuring a voltage level with a noisy instrument. Suppose that the unknown voltage has value a and that the instrument has an uncertainty x.

The observed value may be y = a + x. Suppose that n independent measurements are made under identical conditions, meaning that neither the unknown value of the voltage nor the statistics of the instrument noise change during the process.

Under our model of the process, it must be the case that yi = a + xi.

Requisites of a Good Average

Since an average is a single value representing a group of values, it is desired that such a value satisfies the following properties:

(i) It should be easy to understand.

Since statistical methods are designed to simplify complexity. It is desirable that an average is such that can be readily understood; otherwise, its use is bound to be very limited.

(ii) It should be simple to compute.

Not only an average should be easy to understand but also it should be simple to compute so that it can be used widely.

However, though ease of computation is desirable is should not be sought at the expense of the other advantages, i.e., if in the interest of greater accuracy, use of a more difficult average is desirable, one should prefer that.

(iii) It should be based on all the items.

The average should depend upon each and every item of the series so that if any of the items is dropped the average itself is altered.

(iv) It should not be unduly affected by extreme observations.

Although each and every item should influence the value of the average, none of the items should influence it unduly. If one or two very small or very large items unduly affect the average, i.e., either increase its value or reduce its value, the average cannot be really typical of the entire series. In other words. Extremes may distort the average and reduce its usefulness.

(v) It should be rigidly defined.

An average should be properly defined so that it has one and only one interpretations. It should preferably be defined by an algebraic formula so that if different people compute that average from the same figures they all get the same answer (barring arithmetical mistakes).

The average should not depend upon the personal prejudice and bias of the investigator; otherwise the results can be manipulated.

(vi) It should be capable of further algebraic treatment.

We should prefer to have an average that could be used for further statistical computations so that its utility is enhanced.

(vii) It should have sampling stability.

Last, but not the least, we should prefer to get a value which has what the statisticians call 'sampling stability' This means that if we pick 10 different groups of college students, and compute the average of each group, we should expect to get approximately the same value.

It does not mean, however, that there can be no difference in the values of different samples. There may be some difference but those samples in which this difference (technically called sampling fluctuation) is less are considered better than those in which this difference is more.

QUESTION 3

Explain the concept of combined mean.

Find the mean for the entire group of workers from the following data :

	Group – 1	Group – 2
Mean wages	150	200
No. of workers	500	450

Combined Mean

Mean of Entire group

ANSWER 3

Combined Mean

A combined mean is a mean of two or more separate groups, and is found by :

1. Calculating the mean of each group,
2. Combining the results.

Combined Mean Formula

More formally, a combined mean for two sets can be calculated by the formula

$$x_c = \frac{m \cdot x_a + n \cdot x_b}{m + n}$$:

Where:

- x_a = the mean of the first set,
- m = the number of items in the first set,
- x_b = the mean of the second set,
- n = the number of items in the second set,
- x_c the combined mean.

A combined mean is simply a weighted mean, where the weights are the size of each group.

Mean of Entire group

For Group 1,
Number of Worker = 500 (given)
Mean of wages = $\bar{X}$=Rs.150 (given)
Amount paid by the firm A= Number of wage earners ×× mean of monthly wages.
= 500×150
= 75,000

For Group 2,
Number of Worker = 450 (given)
Mean of wages = $\bar{X}$=Rs.200 (given)
Amount paid by the firm A= Number of wage earners ×× mean of monthly wages.
= 450×200
= 90,000

Mean of Entire group = Sum of Observation / No. of Observation
= 75000 + 90000 / 2
= 165000 / 2
= 82,500

SET - II

QUESTION 1

Write short notes on

a. Standard Deviation

b. Addition rule of probability

c. Sampling and Non Sampling errors in statistics

d. Advantages of Business forecasting

ANSWER 1

a. Standard Deviation

The standard deviation (σ) is a measure that is used to quantify the amount of variation or dispersion of a set of data values.

A low standard deviation indicates that the data points tend to be close to the mean (also called the expected value) of the set, while a high standard deviation indicates that the data points are spread out over a wider range of values.

The standard deviation of a random variable, statistical population, data set, or probability distribution is the square root of its variance. It is algebraically simpler, though in practice less robust, than the average absolute deviation. A useful property of the standard deviation is that, unlike the variance, it is expressed in the same units as the data.

b. Addition rule of probability

A statistical measurement which states that the probability of two events happening at the same time is equal to the probability of one event occurring plus the probability of the second event occurring, minus the probability of both events occurring simultaneously.
Possibilities:

1. The number rolled can be a 2.

2. The number rolled can be a 5.

Events: These events are mutually exclusive since they cannot occur at the same time.

c. Sampling and Non Sampling errors in statistics

Sampling Error

Sampling Error denotes a statistical error arising out of a certain sample selected being unrepresentative of the population of interest. In simple terms, it is an error which occurs when the sample selected does not contain the true characteristics, qualities or figures of the whole population.

Non-Sampling Error

Non-Sampling Error is an umbrella term which comprises of all the errors, other than the sampling error. They arise due to a number of reasons, that means

error in problem definition, questionnaire design, approach, coverage, information provided by respondents, data preparation, collection, tabulation, and analysis.

d. Advantages of Business forecasting

1. Establishing a New Business
2. Formulating Plans
3. Estimating Financial Needs
4. Facilitating Managerial Decisions
5. Quality of Management
6. Encourages Co-operation and co-ordination
7. Better Utilisation of Resources and
8. Success in Business.

QUESTION 2

Marks obtained by 50 students are given below.

Marks	10-20	20-30	30-40	40-50	50-60	60-70	70-80
No of Students	2	8	6	14	3	10	7

Calculate

Median & Mode.

ANSWER 2

COMPUTE MEDIAN

Marks	No of Students (f_i)	Cumulative Frequency
10-20	2	2
20-30	8	10
30-40	6	16
40-50	14	30
50-60	3	33
60-70	10	43
70-80	7	50
Total	**N = 50**	

Here N = 50

Median = N / 2

= 50 / 2

= 25

The Cumulative frequency just greater than 25 is 30 and the corresponding class is 40 – 50.

So, median class is 40 – 50.

L = 40

N = 50

C = 16

F = 14

H = 10

Median = L + [N/2 – C ÷ F] * H

= 40 + [50/2 – 16 ÷ 14] * 10

= 40 + [25 – 16 ÷ 14] * 10

= 40 + [9 ÷ 14] * 10

= 40 + [0.642] * 10

= 40 + 6.42

= 46.42

COMPUTE MODE

L = 40

F_1 = 14

F_0 = 10

F_2 = 8

H = 10

Mode = L + [$F_1 – F_0$ ÷ $2F_1 – F_0 - F_2$] * H

= 40 + [14 – 10 ÷ 2(14) – 10 – 8] * 10

= 40 + [4 ÷ 28 – 10 – 8] * 10

= 40 + [4 ÷ 10] * 10

= 40 + [0.4] * 10

= 40 + 4

=44

QUESTION 3

Sales for last 7 years of ABC ltd is given in following table

YEARS	2011	2012	2013	2014	2015	2016	2017
Sales(in 1000 Rs)	120	130	135	125	140	160	165

Based on above data forecast the Sales for year 2018 & 2019.

A Fitting of straight line

Sales for year 2018

Sales for year 2019

ANSWER 3

A Fitting of straight line

A mathematical procedure for finding the best-fitting curve to a given set of points by minimizing the sum of the squares of the offsets ("the residuals") of the points from the curve. The sum of the *squares* of the offsets is used instead of the offset absolute values because this allows the residuals to be treated as a continuous differentiable quantity. However, because squares of the offsets are used, outlying points can have a disproportionate effect on the fit, a property which may or may not be desirable depending on the problem at hand.

Furthermore, let's assume that the relationship between x and y is a linear one (if it's not, fitting a line to the points is worthless).

Let y = ax + b

Least squares fitting of a straight line

y=a + bx to a set of data points, when there are errors in the values of both coordinates is reviewed. It is shown that if the errors are equal or unknown, then it is possible to solve the problem by a direct approach, using a quadratic equation, and avoiding iteration.

If the errors in both coordinates are unknown, the 'best' line is not invariant under a change of scale: a possible criterion for uniquely determining the best line is suggested. If there are errors in both coordinates, and if each point has its own weighting factors for x and y, a new algorithm is given which involves the iterative solution of a quadratic equation: some remits for this algorithm are presented.

Sales for year 2018

YEARS (X)	**SALES (Y)**	**X = x – 2014**	**X^2**	**X.Y**
2011	120	-3	9	-360
2012	130	-2	4	-260
2013	135	-1	1	-135
2014	125	0	0	0
2015	140	1	1	140
2016	160	2	4	320
2017	165	3	9	495
N = 7	**∑ Y = 975**	**∑ X = 0**	**∑ X^2 = 28**	**∑ X.Y = 200**

a = ∑ Y / N

= 975 / 7

= 139.28

b = ∑ X.Y / ∑ X^2

= 200 / 28

= 7.14

Sales for the year of 2018

Y = a + bx

= 139.28 + 7.14 (4)

=139.28 + 28.56

= 167.84

x = x -2014

= 2018 – 2014

= 4

Sales for year of 2019

x = x -2014

= 2019 – 2014

= 5

Y = a + bx

= 139.28 + 7.14 (5)

=139.28 + 35.7

= 174.98

******