

~~Amna
Done~~

Q1 what is Statistics & give its limitations & functions. Importance of Statistics.



Statistics :-

As the science of collection, classification, presentation, analysis & interpretation of numerical data under study.

limitations of Statistics :-

- ① It does not deal with individual
- ② It deals with quantitative characteristics.
- ③ Statistics result are true on an average.
- ④ Statistics can be misused.

functions of Statistics.

- ① It simplifies mass figures.
- ② It presents facts in a definite form
- ③ It facilities Comprision

Importance of statistics :-

- ① Statistical methods are universal applicable.
- ② enable to condense the data & give tools of comparison.
- ③ Government & Economist use extensively statistical data for national planning.
- ④ Inter relation between two or more variable can be measured using statistical technique.
- ⑤ Statistical method is important because considerable amount of time, money manpower can be saved.
- ⑥ Statistics of Consumption tell us the relative strength of desire.
- ⑦ A businessman succeeds or fails according to forecasting.
- ⑧ The economic barometers are gift of statistical method.
- ⑨ Statistical method have got great important in weather forecasting.
- ⑩ Statistical data & statistical method are being found useful in research in different field such as agriculture, industry, business, medicine, education, Sociology etc.

Q2

What is meant by classification of data? State & explain the basis of type of classification & functions.

→

Classification.

Classification is the process of arranging data in groups or classes according to their similarities & resemblance.

Ex:-

Students in a class are classified on sex, Age, Religion, Height etc

Functions / objectives.

- ① To condense mass data eliminating unnecessary details.
- ② To facilitate comparison
- ③ To study relationship of similarities & dissimilarities
- ④ To study statistical significance of the data.

Types of classification.

- ① Geographical :-
 - ↳ on the basis of area or region
- ② Chronological :-
 - ↳ on the basis of time
- ③ Qualitative :- on the basis of characteristics.
- ④ Quantitative :- In terms of magnitude.

Q.3 Measures of Central Tendency :-
 Averages are the representation of the series or distribution as a whole.
 Measurements → Any yields → weight or volume
 Counts → No. plants → Counting.

Characteristics / properties of an ideal average

- ① It should be rigidly defined.
- ② It should be a representative of the entire data
- ③ Capable of further mathematical treatment
- ④ easy to calculate & simple to understand
- ⑤ It should be not affected by extreme values.
- ⑥ It should possess sampling stability

Commonly used measures of Central Tendency.

A] Mathematical Averages.

- ① Arithmetic mean ⑪ Geometric mean
- ⑩ Harmonic mean

B] Averages of Position.

- ① median
- ② mode.

Q.3 Measures of Central Tendency :-
Averages are the representation of the series or distribution, as a whole.
Measurements → Any yields → weight or volume
Counts → No. plants → Counting.

Characteristics / Properties of an ideal average.

- ① It should be rigidly defined.
- ② It should be a representative of the entire data
- ③ Capable of further mathematical treatment
- ④ Easy to calculate & simple to understand
- ⑤ It should be not affected by extreme values.
- ⑥ It should possess sampling stability.

Commonly used measures of Central Tendency.

A] Mathematical Averages.

- ① Arithmetic mean ⑪ Geometric mean
- ⑩ Harmonic mean

B] Averages of Position.

- ① median ② mode.

1] Arithmetic Mean :-

Date _____

Sum of all observation
divided by number of observation.

$$\bar{x} = \frac{x_1 + x_2 + x_3 + x_4 + \dots + x_n}{n}$$

* Merits :-

- ① It is rigidly defined.
- ② It is based on all the observations.
- ③ It is easy calculate & simple to understand.
- ④ It is capable of further mathematical treatment.
- ⑤ It is least affected due to fluctuation of Sampling.

* Demerits :-

- It is very much affected by extreme values & hence abnormal results are obtained.

2] Median :-

The median is the most items that divide the distribution into two equal parts when the items are arranged in ascending order.

$$M = L + \frac{\frac{n}{2} - C}{f} \times b$$

Where,

L = lower limit of class

n = total number of frequency i.e. $n = \sum f$

f = frequency of the class in which median lies.

c = class preceding the median class.

b = width of the class interval of the class in which median lies.

Merits *

- ① Easily to understand.
- ② Not affected by extreme values.
- ③ Located graphically.
- ④ Best measure for qualitative data.
- ⑤ the class-intervals in the series are unique.

Demerits.

- ① It is not suitable to algebraic treatments.
- ② It can not represent the irregular distribution series.

3] Mode :-

defined as item which occurs most frequently.

$$\text{mode} = l + b \frac{f_0 - f_1}{2f_0 - f_1 - f_2}$$

l = lower limit of modal class

f_0 = frequency or median class

f_1 = class preceding the modal class

f_2 = the class succeeding the modal class.

b = class interval.

Merits :-

- ① Easily understood.
- ② Can be located in some cases by inspection.
- ③ Capable of being ascertained graphically.
- ④ Not affected extreme values.
- ⑤ Represents the most frequent values.

Demerits.

- ① Mode is determinate.
- ② It is not rigidly defined.
- ③ It is not based on all observation of series.

Q4 Define the term measure of Dispersion with objective & characteristics of good dispersion?



Dispersion :-

is degree of variation or diversity.

* Objectives :-

- i] To judge reliability of average i.e. measure of central tendency.
- ii] To study the variation between two or more series.
- iii] To observe variation in view to control the same.
- iv] To serve as the basis for further statistical analysis.

* Characteristics :-

- ① It should be rigidly defined
- ② It should be easy to calculate & easy to understand.
- ③ It should be based on all observation
- ④ It should be capable further mathematical treatment
- ⑤ It should be not affected by extreme value.

1] Standard deviation :-

IS defined as square root of the mean of the squared deviations of individual value from the arithmetic mean.

$$S.D = \sqrt{\frac{\sum (x - \bar{x})^2}{n}}$$

Ungrouped = $\sqrt{\frac{\sum (x - \bar{x})^2}{n-1}}$ grouped $S = \sqrt{\frac{\sum f x^2 - (\sum f x)^2 / n}{n-1}}$

2) Variance :-

The square of standard deviation is known as Variance.

$$\text{i.e. Variance} = (S.D)^2$$

3] Coefficient of Variance :-

Coefficient of variance is defined as standard error per unit values of arithmetic mean.

expressed - Percentage (%)

$$C.V = \frac{\text{Standard deviation}}{\text{Mean}} \times 100$$

Q5 What do you understand by Sampling? Explain Simple random Sampling?



Sampling:-

The method of selecting sample from a population is known as Sampling.

Types of Sampling:

- ① Purposive Sampling
- ② Simple Random Sampling
- ③ Systematic Sampling
- ④ Stratified Sampling.
- ⑤ Multistage Sampling.
- ⑥ Sampling Distributions.

* Simple random Sampling *

When the unit of sample are drawn independently with equal probability of sampling method, is known as Simple Random Sampling.

- The basic probability sampling method is the simple random sampling.

- It is the simplest of all the probability Sampling method.
- It is used when the population is homogeneous
- If the population consists of N unit . the Probability of selecting any unit is $1/N$
- Suppose drawn sample size n from a populations of size N . there are (N^n) possible sample of size n .
- If all possible samples have an equal probability $1/(N^n)$ of being drawn.

Q6

Define normal distribution & give the properties of normal distribution?

→

Normal distribution :-

Normal distribution is a distribution of random variable, which varies Continuously.

$$f(x) = \frac{1}{6\sqrt{2\pi}} e^{-\frac{1}{2} \left(\frac{x-\mu}{6}\right)^2}$$

Where

x = Measurement made on biological variate

μ = Mean of normal distribution

σ^2 = Variance of normal distribution

e = Mathematical constant

Properties of normal distribution.

- The shape of curve of normal distribution is bell shaped.
- the curve of normal distribution is bell shaped & symmetrical about mean.
- The number of cases below the mean is equal to number of cases above the mean in normal distribution.
- the variables vary continuously.
- the shape of the curve depends upon μ & number of observation (n)
- $-\infty \leq x \leq \infty$

Q7 Define Correlation. what is Karl Person's Correlation coefficient & state its use?



Correlation:-

Correlation is the study of relationship between two or more variable.

* Karl Person's Correlation coefficient *

* Karl Person was suggested a mathematical method for measuring Linear Relation between two variables.

* Suppose x & y are two variable which take values $(x_1, y_1), (x_2, y_2) \dots (x_n, y_n)$

* The Correlation coefficient r is given as the ratio of Covariance of Variable x & y to the product of standard deviation of x & y . Symbolically.

$$r = \frac{\text{Covariance}(xy)}{\sqrt{\text{Variance}(x) - \text{Variance}(y)}} = \frac{\text{Covariance}(xy)}{\text{SD}(x) \cdot \text{SD}(y)}$$

$$r = \frac{1}{n-1} \left[\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) \right]$$

This can be Simplified as:-

$$r = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sqrt{\sum x^2 - \frac{(\sum x)^2}{n}} \sqrt{\sum y^2 - \frac{(\sum y)^2}{n}}}$$

$$r = \frac{\sum x_i y_i - n \bar{x} \bar{y}}{\sqrt{[\sum (x_i - \bar{x})^2] [\sum (y_i - \bar{y})^2]}}$$

* Properties of Correlation Coefficient.

- ① The correlation coefficient value range between -1 & +1.
- ② The correlation coefficient is not affected change of origin or scale or both.
- ③ If $r > 0$ it denotes positive correlation & if $r < 0$ it denotes negative correlation the two variables x & y .
- ④ $r_{xy} = r_{yx}$.

* Defⁿ : Coefficient of Correlation.

Coefficient of correlation is a statistical measure, which indicate both nature & degree of relationship between two variables.

Q8

Define Correlation & explain Spearman's Rank Correlation Coefficient?



Correlation :-

Correlation is study of relationship between two or more variables.

* Spearman's Rank Correlation Coefficient *

- When both the variables are not normal, the linear correlation coefficient procedure is not applicable & we have to use rank correlation.
- The two method of computing rank correlation are one proposed by.
 - (a) Spearman
 - (b) Kendall
- Spearman's rank Correlation procedure starts with ranking of the measurement of the value of x & y separately.
- rank correlation computed by formula.

$$R_s = 1 - \frac{6 \sum d^2}{n(n^2-1)} \quad \text{for different rank}$$

$$R_s = \frac{6 \left(\sum d^2 \left[\frac{1}{n} \sum (m-1) \right] \right)}{n(n^2-1)} \quad \text{for identical rank}$$

89 What do you understand by a regression model & define linear regression of y on x & x on y also give relationship between the Correlation Coefficient & regression coefficient



Regression model :-

Measuring of the average relationship between two or more variables in terms of the original units of the data.

Linear regression of y on x :-

It is the line which give the best estimate for the value of y for any specified value of x .

$$y = a + bx$$

Linear regression of x on y :-

This line which give the best estimate for the value of x for any specified value of y .

$$x = a + by$$

* Relation between Coefficient of Correlation & Coefficient of regression (b_{xy})

- ① Correlation coefficient is geometric mean between two regression coefficients.

$$\rho = \sqrt{r_{xy} \times r_{yx}}$$

② Arithmetic mean of regression Coefficient is greater than correlation Coefficient

$$\left(\frac{b_{yx} + b_{xy}}{2} \right) > r$$

③ If one of the regression Coefficient is greater than unity. the other must be less than unity.

④ Regression Coefficient are independent of change of origin & scale.

⑤ $r = \frac{b_{yx} \frac{\sigma_y}{\sigma_x}}{\sqrt{b_{yx} b_{xy}}} \text{ if } b_{xy} \neq 0$

Q10 Define chi-square (χ^2) test & Give Contingency 2×2 table ?

→

chi-square (χ^2) test :-

The chi-square (χ^2) test provide a measurement of agreement between such observed & expected frequencies.

- : 2×2 Contingency table :-

- If each attributes divided into two classes it is known as 2×2 Contingency table
- When one attribute divided into two classes & another into r or c the resultant Contingency table is known as 2×2 Contingency table.

- In similar way get $r \times c$ Contingency table where, r denotes no. of rows & c the no. of columns.

- Statistical hypothesis under test is that the two attribute are independent of one another.

$$\chi^2 = \sum_{E} (O - E)^2$$

- This χ^2 follows χ^2 distribution with $(r-1)$ $(c-1)$ degrees of freedom.

- The degree of freedom in 2×2 Contingency table is
- Consider the 2×2 Contingency table with cell frequencies a, b, c & d .

Group	Category		Total
	I	II	
1	a	b	$a+b = r_1$
*	c	d	$c+d = r_2$
3	$a+c = c_1$	$b+d = c_2$	$a+b+c+d = N$

Q 11

Define Completely Randomized Design & give merit & Demerit.



CRD :-

When the treatment are arranged randomly over the whole of previously determined set of experimental unit the design is known as CRD.

$$X_{ij} = \mu + t_i + e_{ij}$$

Where:

μ = overall mean effect

t_i = true effect of the i^{th} treatment

e_{ij} = error effect of the j^{th} unit receiving i^{th} treatment.

Merits of CRD : [LFWys]

- Layout is very easy
- There is complete flexibility in this design.
- Whole experimental material can be utilized in this design.
- This design yield maximum degree of freedom for experimental error.
- The analysis of data is simplest and compared to any other design.

Demerits :-

- The design is suitable for a small no. of treatment.
- It is difficult to find homogeneous experimental unit in all aspects.

Explain local Controls :-

The principle of making greater Homogeneity in experimental units for reducing the experimental error is known as local controls.

Advantages of local control.

- i] Local control reduces the experimental error.
- ii] Local control is meant to make design more efficient.
- iii] It makes any test of significance more sensitive & powerful.
- iv] A reduction in experimental error consequently helps in investigator to detect the small real difference between treatment means.

Q12 Define term Randomized Block Design. Explain the ANOVA TABLE & advantages & disadvantages of RBD.



RBD.

An arrangement of grouping the heterogeneous units into homogenous blocks is known as randomized block design.

Sources of Variation	D.f	SS	T _{MSS}	F value
Treatment	t-1	T _{rSS}	T _{rMSS}	T _{rMSS} /E _{MSS}
Block / Replication	r-1	R _{SS}	B _{MSS} /R _{MSS}	R _{MSS} /E _{MSS}
Error	(r-1)(t-1)	E _{rSS}	E _{rMSS}	R _{MSS} /E _{mss}
Total	T _{r-1}	total SS		

Where

$$S.E.(d) = \sqrt{\frac{2 E_r.MSS}{r}}$$

- t+ table value of t for a specified level of significance (α) & error degree of freedom.
- Based on the cd value the bar chart can be drawn using the bar chart conclusion can be written.

* Advantages of RBD.

- $T +$ increases the precision of the experiments.
- This is due to reduction of experiment error by adoption of local control.
- The amount of information got in RBD is more as compared to CRD.
- Hence,
- The RBD is more efficient than PBD.
- Flexibility is advantages of RBD Any no. of replication can be included in RBD.
- If large no. of homogeneous unit are available.
- Large no. of treatment can be included in this design.

* Disadvantages:-

- No. of treatments increased the block size will increases.
- If the block size is large may be difficult to maintain homogeneity within blocks.
- RBD is a versatile design.
- It is the most frequently used design in agricultural experiments.

