Model Answer

MAHARASHTRA AGRICULTURAL UNIVERSITIES EXAMINATION BOARD, PUNE
SEMESTER END THEORY EXAMINATION
B.Sc. (Agri.)

Semester	**: II (New)**		
Course No.	**: STAT –121**	**Academic Year : 2014-15**	
Credits	**: 2(1+1)**	**Title**	**: Statistics**
Day & Date	**:**	**Time :**	**Total Marks : 40**

SECTION "A"

Q.1 Define statistics and State the functions and limitations of statistics.

Definition: Statistics is defined as the science of collection, organization, presentation, analysis and interpretation of numerical data.

Functions :

It presents facts in a definite form.
It simplifies mass of figures.
It facilitates comparison.
It helps in formulating and testing of hypothesis.
It helps in prediction
It helps in the formulation of suitable policies.

Limitations :

Statistics does not deal with individuals.
Statistics deals only with quantitative characteristics.
Statistical results are true only on an average.
Statistics is only one of the methods of studying a problem.

Q.2 Define mode and explain the method to compute mode in case of discrete and continuous series.

Mode is defined as item which occur maximum number of times in a data

Mode in case of Discrete series :

Mode is the item with highest frequency.

Mode in case of continuous series :

$$\text{Mode} = L + \frac{f_1 - f_o}{2f_1 - f_o - f_2} \times h$$

Where,

L = Lower limit of the modal class (Class with highest frequency)
f_1 = Frequency of the modal class
f_o = Frequency of the class preceding to modal class
f_2 = Frequency of the class succeeding to modal class
h = Class width.

Q.3 What does Dispersion indicate about the data? State the objectives and characteristics of a good dispersion.

Dispersion : The measurement of the scatterness of the mass of figures in a series about an average is called measure of variation or dispersion.

Objectives :

To determine the reliability of an average.
To serve as a basis for the control of the variability.
To compare two or more series with regard to their variability.
To facilitate the use of other statistical measures.

Characteristics :

It should be simple to understand.
It should be easy to compute.
It should be rigidly defined.
It should be based on each and every item of the distribution.
It should be amenable to further algebraic treatment.
It should have sampling stability.
It should not be unduly affected by extreme values.

Q.4. Define Correlation. What is Karl Pearson's correlation coefficient? State its use.

Correlation Coefficient is the index of degree of relationship between two variables.

Karl Pearson's correlation Coefficient (r) = cov(X,Y) /Std.Dev(X). Std.Dev(Y)

i.e. $r = cov(X,Y)/\sigma_x \sigma_y$ ----------(A)

Karl Pearson's correlation coefficient is the ratio of covariance between two variables to the product of their standard deviation. Its magnitude gives the degree of association between the two variables and sign gives the direction of correlation. It is a measure of correlation hence used to study the degree of association between two variables.

Q.5 What do you understand by a regression model and Define linear regression of Y on X and X on Y. Also give relationship between the correlation coefficient and regression coefficients.

Regression is the functional relationship between two variables and of the two variables one may represent cause and the other may represent effect. The variable representing cause is known as independent variable and it is denoted by X. The variable X is also known as predictor variable or regressor. The variable representing effect is known as dependent variable and is denoted by Y. Y is also known as predicted variable. The relationship between the dependent and independent variable may be expressed as a function and such a functional relationship is termed as regression. When there are only two variables the functional relationship is known as a simple regression and if the relation between two variables is a straight line, it is known as a simple linear regression. The regression line is of the form $Y = a + byx\ X$ where 'a' is a constant or intercept and byx is the regression coefficient or the slope. The values of a and byx can be calculated by using method of least squares.

The regression equation of Y on X is given by $Y = a_1 + byx\ X$

Similarly the regression equation of X on Y is given by $X = a_2 + bxy\ Y$

Correlation coefficient is the geometric mean between the two regression coefficient byx & bxy

Q.6 Define the following terms.

1) Probability
2) Type I Error
3) Degree of Freedom
4) Mutually exclusive events

1) Probability :

i) The ratio of favorable cases to the total number of exhaustive, mutually exclusive and equally likely cases.

ii) If a trail is repeated number of times under essential homogeneous and identical conditions than the limiting value of the ratio of number of times event happens to the number of times trail conducted is called the probability of event.

2) Type I Error

We may reject H_0 when in fact it is true. This error in decision is known as Type I Error.

3) Degree of Freedom

The Degree of Freedom is the number of observation that are free to vary after certain restriction have been placed on the data.

4) Mutually exclusive events

Two or more events are said to be mutually exclusive, when occurrence of any one event excludes the occurrence of the other event . Mutually exclusive events cannot occur simultaneously. For example when a coin is tossed, either the head or the tail will occur.

Q.7 What is Normal Distribution? Give its properties.

In Answer there should be Probability density function of Normal Distribution

Properties of Normal Distribution

1. The curve of Normal Distribution is bell shaped.
2. The height of the normal curve is maximum at mean.
3. Curve is unimodal
4. For Normal Distribution
   coefficient of Skewness = $\beta_1 = 0$
   coefficient of Kurtosis = $\beta_2 = 3$
5. The total area under the normal probability curve is unity/
6. In this distribution mean = median = mode

The area under the normal probability curve is distributed as

i. $\mu \pm \sigma$ covers 68.26 per cent area.
ii. $\mu \pm 2\sigma$ covers 95.45 per cent area.
iii. $\mu \pm 3\sigma$ covers 99.73 per cent area.

Q.8 What do you understood by Sampling? Explain simple random sampling.

Some representative items are selected from the population, so that all important characteristics of population are covered in the items of this group. Such a group is called as sample and the method of selecting such a group is called as sampling.
Simple random sampling is the method of sampling in which the each element of population is given same chance of getting selected in the sample.

Q.9 Write short notes on (Any Two)

1) Layout in Latin Square Design
2) F-test

1) Layout in Latin Square Design

Randomized Latin square for 3 x 3 to 12 x 12 are already available in standard books of on Randomization tables. One has to peak anyone them randomly for m number of treatments. In these squares care is taken that no treatment should be repeated in a row and in a column and number of treatments and the replications, rows, columns are same.

2) F-test

Objective of a F-test is to compare the variability of two populations

$H_0 : \sigma_1^2 = \sigma_2^2$

$H_1 : \sigma_1^2 \neq \sigma_2$

Method

Given samples of size n_1 with values $X_1, X_2, \ldots\ldots Xn_1$ and size n_2 with values $Y_1, Y_2 \ldots\ldots Yn_2$ from two population. The sample variance s_1^2 and s_2^2 are calculated.

Test statistic

$$F = s_1^2 / s_2^2 \qquad (s_1^2 > s_2^2)$$

Follows the F- distribution with $(n_1\text{-}1, n_2.1)$ degrees of freedom.

Compare calculated value of F with table value and write conclusion.

Q.10 Write the formulae.

1) Spearman's Rank Correlation Coefficient = $1\text{-}6 \sum di^2/(N^3\text{-}N)$
2) Coefficient of Variation. = $(\sigma/\text{mean}) \times 100$
3) Paired t-test Statistics $= t = |\bar{d}| / s/\sqrt{n}$
4) Median for continuous series = $L + [(N/2\text{-}c.f.)/f] \times h$

SECTION 'B'

Q.11 State True or False.

1) Regression technique is useful for prediction purpose. = **TRUE**
2) Probability of impossible event is Zero. = **TRUE**
3) t-test is also called as variance ratio test. = **FALSE**
4) Technique of analysis of variance is developed by R.A. Fisher. = **TRUE**

Q.12 Fill in the blanks.

1) The sum of deviations of items from their arithmetic mean is **zero.**
2) **Local Control** is basic principle not used in CRD.
3) Independence of attributes is tested by χ^2 **test .**
4) If Calculated value is less than table value result is non significant than null hypothesis is **Accepted**