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**

1/9

MODEL ANSWER
B. Sc. (Agriculture)

Semester : IInd | Academic Year : 2013-14
Course No. : STAT 121 | Course Title : Statistics
Credits : 2(1+1) | Total Marks : 40

SECTION "A"

Q. 1 Define "Statistics". Describe the uses and limitations of the subject Statistics.

Definitions of Statistics are given by various authors described as follows:

1.**HORACE SECRIST**: By Statistics we mean aggregates of facts affected to a marked extent by multiplicity of causes, numerically expressed, enumerated or estimated according to reasonable standards of accuracy, collected in a systematic manner for a pre-determined purpose and placed in relation to each other.

2. **KING**: The science of Statistics is the method of judging collective natural or social phenomenon from the results obtained from the analysis of enumeration or collection of the estimates.

3 **BUDDINGION**: Statistics is a science of estimates and probabilities.

4. **A.L. BOWELY**: Statistics may be called as a science of counting.

5. **COWDEN and CROXTON**: Statistics defined as collection, analysis and interpretation of data.

In general the science of Statistics can be defined as the body of concepts, principles and methods dealing with collection, summarization, analysis and interpretation data.

Uses of the subject Statistics:

1. Simplifies complexity of the data.
2. Computes the averages, standard deviation etc.
3. Studies the relationship between variables.
4. Useful for estimation purposes.
5. Useful for surveys.
6. Used in the subjects like Agricultural science and biological sciences, medical sciences and various research fields.

Limitations of science Statistics:

1. Statistics does not study qualitative data. i.e. It studies only numerical type of data.
2. It is not the study of individuals but it studies only aggregates.
3. Statistics is likely to be misused.
4. Statistical results some times lead to wrong conclusions unless they are properly used.
5. Statistical laws are not exact.

Q. 2 What do you mean by the measure of Central Tendency? States the characteristics of ideal average.

The meaning of measure of central tendency i.e. average is given in the following definitions.

"A measure of central tendency is a typical value around which other figures congregate."

"An average stands for the whole group of which it forms a part and it represents the whole."

"One of the most widely used set of summary figures is known as measures of location."

Characteristics for a good or an ideal average :

A good average possess for an ideal average.

1) It should be rigidly defined.
2) It should be easy to understand and compute.
3) It should be based on all items in the data.
4) Its definition shall be in the form of a mathematical formula.
5) It should be capable of further algebraic treatment.
6) It should have sampling stability.
7) It should be capable of being used in further statistical computations

Q. 3 Explain the term "measure of dispersion" with some suitable example. Define standard deviation, variance and coefficient of variation.

The estimation of variation or deviation of the different values of a variable form the average is known as measures of dispersion. For Eg., consider the yield of two crops A and B.

Variety A(yield per plant in grams): 5, 6, 7

Variety B(yield per plant in grams) 2, 6, 10

If mean alone is taken to explain the data, then it will be concluded that the means of the two variety are equal with respect to yield per plant (because the mean value of variety A and variety B are equal to 6 grams). Actually the means of two varieties are different when the variation of observations is concerned. Therefore, measure of central tendency alone is not sufficient to explain behaviour of data. In addition to means, the dispersion should also be considered simultaneously.

1) **Standard Deviation** (SD) is the positive square root of the mean of the squared deviations from mean. It is denoted by σ.
2) **Variance:** Square of standard deviation is known as variance. It is the mean of the squared deviations from mean. It is denoted by σ^2
3) **Co-efficient of Variation** CV (%) = (Standard Deviation / Mean) X 100

Q.4 Define the term correlation between two variables. Distinguish between positive correlation and negative correlation between two variables with suitable examples.

Correlation is statistical analysis which measures and analyses the degree or extent to which the two variables fluctuate with reference to each other. The word relationship is important. It indicates that there is some connection between the variables. It measures the closeness of the relationship. Correlation does not indicate cause and effect relationship. Price and supply, income and expenditure are correlated.

Definitions: 1. Correlation Analysis attempts to determine the degree of relationship between two variables- Ya-Kun-Chou. Or

2. Correlation is an analysis of the covariation between two variables.- A.M.Tuttle.

Positive Correlation: It depends upon the direction of change of the variables. If the two variables tend to move together in the same direction (ie) an increase in the value of one variable is accompanied by an increase in the value of the other, (or) a decrease in the value of one variable is accompanied by a decrease in the value of other, then the correlation is called positive or direct correlation. Price and supply, height and weight, yield and rainfall, are some examples of positive correlation.

Negative Correlation: If the two variables tend to move together in opposite directions so that increase (or) decrease in the value of one variable is accompanied by a decrease or increase in the value of the other variable, then the correlation is called negative (or) inverse correlation. Price and demand, yield of crop and price, are examples of negative correlation.

Q.5 Define the term regression. Define linear regression of Y on X and X on Y using suitable notations. Also state the inter-relation between 'r' 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 is denoted by X. The variable X is also known as predictor variable or repressor. 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 the independent variable may be expressed as a function and such functional relationship is termed as regression. When there are only two variables the functional relationship is known as simple regression and if the relation between the two variables is a straight line, it is known a simple linear regression. The regression line is of the form $Y = a + b X$ where a is a constant or intercept and b is the regression coefficient or the slope. The values of and b can be calculated by using the method of least squares.

The regression equation of y on x is given by $Y = a_1 + b_1 X$

Similarly the regression equation of X on Y is given by $X = a_2 + b_2 Y$

Correlation coefficient r is the geometric mean of the two regression coefficients b_1 & b_2

i.e. $r = (b_1 \times b_2)^{1/2}$

Q. 6 Define the terms 1) Random experiment, 2) Event, 3) Equally likely events, 4) Mutually exclusive events.

1) **Random experiment:** Random experiment is one whose results depend on chance, that is the result cannot be predicted. e.g. tossing of coin, throwing of die are some examples of random experiments.
2) **Event:** An outcome or a combination of outcomes of a random experiment is called an event. For example tossing of a coin is a random experiment which results into getting a head or tail is an event.
3) **Equally likely events:** Two or more events are said to be equally likely events if each one of them has an equal chance of occurring. For example in tossing of a coin, the event of getting a head and the event of getting a tail are equally likely events.
4) **Mutually exclusive events:** Two or more events are said to be mutually exclusive, when the 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 face up words. Therefore, the occurrence of the head completely excludes the occurrence of the tail. Thus getting head or tail in tossing of a coin is a mutually exclusive event.

Q. 7 Define normal distribution. Describe the probability curve of the normal distribution. Also state the properties of normal distribution.

A continuous random variable X is said to follow normal distribution with mean μ and standard deviation σ, and its probability density function is given as

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

In the above formula,

π = a constant equality (22 / 7) e = base of the natural logarithm = 2.7183

μ = population mean σ = population standard deviation

x = a given value of the random variable in the range $-\infty \le x \le \infty$

The mean μ and standard deviation σ are called the parameters of Normal distribution.

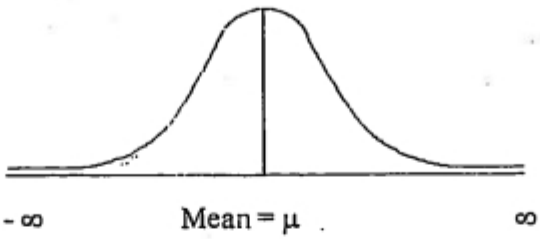
The normal distribution is expressed by $X \sim N(\mu, \sigma^2)$

Normal probability curve:

The curve representing the normal distribution is called the normal probability curve.

The curve is a) symmetrical about the mean (μ), b) bell-shaped and c) the tails on the

right and left sides of the mean extends to the infinity. The shape of the curve is shown in the following figure.



Properties of normal distribution:

1. The normal curve is bell shaped and is symmetric at $x = \mu$
2. Mean, median, and mode of the distribution are coincide i.e.,

   Mean = Median = Mode
3. It has only one mode at $x = \mu$ (i.e., distribution is unimodal)
4. Since the curve is symmetrical, Skewness = $\beta 1 = 0$ and Kurtosis = $\beta 2 = 3$.
5. The points of inflection are at $x = \mu - \sigma$ and at $x = \mu + \sigma$
6. The maximum ordinate occurs at $x = \mu$ and its value is = $1/(\sigma \sqrt{2\Pi})$
7. The x axis is an asymptote to the curve

   (i.e. the curve continues to approach the x axis, but never touches the axis)
8. The first and third quartiles are equidistant from median.
9. The mean deviation about mean is 0.8 times the standard deviation.
10. Quartile deviation = (0.6745) (σ)
11. If X and Y are independent normal variates with means $\mu 1$ and $\mu 2$, and Variance σ^2_1 and σ^2_2 respectively then their sum (X + Y) is also a normal variate with mean $(\mu 1 + \mu 2)$ and variance $(\sigma^2_1 + \sigma^2_2)$
12. Area Property $p(\mu - \sigma < x < \mu + \sigma) = 0.6826$

    $p(\mu - 2\sigma < x < \mu + 2\sigma) = 0.9544$

    $p(\mu - 3\sigma < x < \mu + 3\sigma) = 0.9973$

Q. 8 Define the terms 1) Sampling distribution 2) Standard error 3) Null hypothesis and 4) alternative hypothesis.

Sampling Distribution: The distribution of all possible values which can be assumed by some statistic measured from samples of same size ' n' which are randomly drawn from the same population of size N, is called as sampling distribution of the statistic. Consider a population with N values .Let us take a random sample of size n from this population, then there are $NCn = N! / (n!(N - n)!) = k$ (say), possible samples. From each of these k samples if we compute a statistic (e.g mean, variance, correlation coefficient, skewness etc) and then we form a frequency distribution for these k values of a statistic, then such a distribution is called sampling distribution of that statistic. For example, we can compute some statistic $t = t(x_1, x_2, \ldots x_n)$ for each of these k samples. Then $t_1, t_2 \ldots, t_k$ determine the sampling distribution of the statistic t. In other words

statistic *t* may be regarded as a random variable which can take the values $t_1, t_2 \ldots\ldots, t_k$ and we can compute various statistical constants like mean, variance, skewness, kurtosis etc., for this sampling distribution.

Standard Error: The standard deviation of the sampling distribution of a statistic is known as its standard error. It is abbreviated as S.E. For example, the standard deviation of the sampling distribution of the mean *x is* known as the standard error of the mean.

Null Hypothesis and Alternative Hypothesis: Hypothesis testing begins with an assumption called a Hypothesis, that we make about a population parameter. A hypothesis is a supposition made as a basis for reasoning. The conventional approach to hypothesis testing is not to construct a single hypothesis about the population parameter but rather to set up two different hypothesis. So that as a result of testing of hypothesis, either accepted, or rejected.

Null Hypothesis: A hypothesis of no difference is called null hypothesis and is usually denoted by H_0 "Null hypothesis is the hypothesis" which is tested for possible rejection under the assumption that it is true ". It is very useful tool in test of significance. For example: If we want to find out whether the special classes after college hours have benefited the students or not, we shall set up a null hypothesis that "H_0: special classes after college hours have not benefited the students".

Alternative Hypothesis: Any hypothesis, which is complementary to the null hypothesis, is called an alternative hypothesis and is usually denoted by H_1. For example, if we want to test the null hypothesis that the population has a specified mean μ_0 (say),

i.e., : Step 1: Null hypothesis $H_0: \mu = \mu_0$

then Step 2: Alternative hypothesis may be

i) $H_1: \mu \neq \mu_0$ (ie : $\mu < \mu_0$ or : $\mu > \mu_0$) or ii) $H_1: \mu < \mu_0$ or iii) $H_1: \mu > \mu_0$

the alternative hypothesis in (i) is known as a two – tailed alternative and the alternative in (ii) is known as right-tailed (iii) is known as left –tailed alternative respectively. The settings of alternative hypothesis is very important since it enables us to decide whether we have to use a single – tailed (right or left) or two tailed test.

Q.9 Write Short Note on (Any Two)

1) **Scatter Diagram**
2) **Paired t test**
3) **Basic principles of Design of Experiment**
4) **Completely Randomised Design C.R.D.**

1) Scatter Diagram: To investigate whether there is any relationship between the variables X and Y we use scatter diagram. Let (x_1,y_1), (x_2,y_2)....(x_n,y_n) be n pairs of observations. If the variables X and Y are plotted along the X-axis and Y-axis respectively in the x-y plane of the graph sheet, the resultant diagram of dots is known as scatter diagram. From the scatter diagram we can say whether there is any

correlation between x and y and whether it is positive or negative or the correlation is linear or curvilinear.

1) If all the plotted points form a straight line from lower left hand corner to the upper right hand corner then there is Perfect positive correlation. We denote this as r = +1
2) If all the plotted dots lie on a straight line falling from upper left hand corner to lower right hand corner, there is a perfect negative correlation between the two variables. In this case the coefficient of correlation takes the value r = -1.
3) If the plotted points in the plane form a band and they show a rising trend from the lower left hand corner to the upper right hand corner the two variables are highly positively correlated.
4) If the points fall in a narrow band from the upper left hand corner to the lower right hand corner, there will be a high degree of negative correlation.
5) If the plotted points in the plane are spread all over the diagram there is no correlation between the two variables.

2) Paired t test: When the samples are correlated or paired samples wherein observations x and y are recorded from same individuals, we apply paired t test. The test statistics t is given by $t_o=\bar{d}/SE(d)$ where, SE(d) is given by $\sqrt{[\sum d_i^2-(\sum d_i)^2/n]/(n-1)}$, where $d_i=(x_i-y_i)$; $\bar{d}=\sum d_i/n$. This t_o follows t distribution with (n-1) degrees of freedom. We can also use a paired t test to compare samples that are subjected to different conditions, provided the samples in each pair are identical otherwise. For example, one might test the effectiveness of a water additive in reducing bacterial numbers by collecting water samples from different sources and comparing bacterial counts in the treated versus untreated water sample. Each different water source would give a different pair of data points.

3) Basic Principles of Design of Experiment: The basic principles of any design of experiment are replication, randomization and local control. error.

Replication in an Experiment: The application of the treatment to different experimental units is known as replication. The advantages of replication are:

1. Replications are essential to obtain a valid estimate of the experimental error variance.
2. Replications are necessarily required to attach a probability statement with estimate treatment differences
3. The larger the number of replications reduces the standard error of the treatment mean's But one can not take more than certain number of replications because of the management problems and limitation of resources

Randomization: When all the treatments have an equal chance of being allocated to different experimental units, it is known as randomization. The procedure of allocating

treatments to experimental unit may vary according to the design of the experiment. The advantages of randomization are:

1. To get an unbiased estimate of experimental error
2. To get valid conclusions

Local control :

Local control is a device to maintain greater homogeneity of experimental units a block of an experiment (or) as a whole. For instance, Soil fertility is a factor that affects the plant growth and yield. So all the neighboring plots having the same soil fertility should constitute a block. Conducting a uniformity trial can assess the soil fertility of land. Before conducting the actual field experiment, we prepare the fertility contour map on the basis of fertility gradient. These contours help to form blocks. Local control is also known as error control. It increases the efficiency of an experiment. The advantages of local control are:

1. A reduction in Experimental error consequently helps the Local control reduces the experimental error
2. Local control is meant to make designs more efficient
3. It makes any test of significance more sensitive and powerful
4. Any small real difference between treatment means can be detected easily.

4) Completely Randomized Design (CRD)

Completely randomized design is the basic single-factor experiment. When the treatments are arranged randomly over the whole of a previously determined set of experimental units, the design in known as completely randomized design. CRD is appropriate only when the experimental material is homogeneous. However field experiments there is generally large variation among experimental plots due to soil heterogeneity. Hence CRD is not preferred in field experiments. In laboratory experiments and greenhouse studies it is easy to achieve homogeneity of experimental materials. Therefore CRD is most useful in such experiments.

Statistical Model for CRD with one observation per unit is

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

Q. 10 Write the formulae

(1) Arithmetic mean for ungrouped data = $\Sigma x i / N$

(2) Spearman's Rank correlation Coefficient = $1 - 6\Sigma di^2/(N^3 - N)$

(3) SND test statistic for single sample mean

$$z = \frac{\bar{x} - \mu_o}{s/\sqrt{n}}$$

(4) t- statistic for testing the significance of simple Correlation coefficient

$$t = r \times (\sqrt{(n-2)} / \sqrt{(1 - r^2)}$$

SECTION "B"

Q.11 State True or False

(1) Regression technique is useful for prediction purpose. **True**

(2) Probability of sure event is one. **True**

(3) Student's t test is applicable in case of small sample. **True**

(4) Technique of analysis of variance is developed by R.A. Fisher **True**

Q.12 Fill in the blanks

(1) The difference between the upper limit and lower limit of class interval is called **width of class interval**.

(2) Homogeneity of two sample variances is tested by **F-test**.

(3) Large sample test is applied when sample size is greater than **30** .

(4) Independence of attributes is tested by **Chi square** test.

π π π

MAHARASHTRA AGRICULTURAL UNIVERSITIES EXAMINATION BOARD, PUNE

SEMESTER END EXAMINATION

Model Answer Paper

B.B.M. (Agri)

Semester	: IV (New)	**Academic Year**	: 2013-14
Course No.	: **STAT 241**	**Title**	: Business Statistics
Credits	: 3(2+1)		
Day & Date	:	**Time** : 14.00 to 17.00	**Total Marks: 80**

Note: 1. Solve ANY Eight questions from **SECTION "A".**
2. All questions from **SECTION "B"** are compulsory.
3. All questions carry equal marks.
4. Draw neat diagrams wherever necessary.

SECTION "A"

Q. 1 Define Statistics. State the limitations and uses of statistics.

Ans. **Statistics** may be defined as the science of collection, presentation analysis and interpretation of numerical data from the logical analysis.

Limitation of Statistics:

1. It does not study qualitative phenomena.
2. Statistical studies are true only on an average.
3. It does not study individuals.
4. It does not reveal the entire story.
5. It is liable to be misused.

Uses of statistics

In the early part of the growth period, statistics use was restricted to the states only. But today its use has spread to the study of problems which may be social, religious, economic, political, administrative, commercial, agricultural, financial, medical, relating to business management planning research education,. Psychology, forecasting and to so many other spheres.

This widely use of statistics is on account of the fact that the statistical principals have a very wide scope of application and their knowledge is very essential for any research work in any branch of study.

Statistics has pervaded almost all spheres of human activities. Statistics is useful in the administration of various states, Industry, business, economics, research workers, banking, insurance companies etc.

Q. 2 Define Correlation. Write down the properties of correlation coefficient. Give the formula for Karl Pearson's correlation coefficient with specification.

Ans. Correlation is an analysis of the co-variation between two or more variables.

Properties of Correlation:

1. Correlation coefficient lies between -1 and $+1$ (i.e) $-1 \leq r \leq +1$.
2. ' r' is independent of change of origin and scale.
3. It is a pure number independent of units of measurement.
4. Independent variables are uncorrelated but the converse is not true.
5. Correlation coefficient is the geometric mean of two regression coefficients.
6. The correlation coefficient of x and y is symmetric. $r_{xy} = r_{yx}$.

Karl pearson' s coefficient of correlation:

The formula for calculating ' r' is

$$r = \frac{Cov(xy)}{\sigma_x \sigma_y} = \frac{Cov(xy)}{S.D(x) \times S.D(y)}$$

Where, $\sigma_x \sigma_y$ are SD of X and Y respectively.

Cov (xy) is covariance between X & Y.

Q. 3. Enlist the different methods of dispersion. Give the formulae for standard error of mean and coefficient of variation. Write properties of a good measure of dispersion.

Ans. The various absolute and relative measures of dispersion are listed below.

Absolute measure	Relative measure
1. Range	1.Co-efficient of Range
2.Quartile deviation	2.Co-efficient of Quartile deviation
3.Mean deviation	3. Co-efficient of Mean deviation
4.Standard deviation	4.Co-efficient of variation

Properties of a good measure of dispersion:

An ideal measure of dispersion is expected to possess the following properties:-

1.It should be rigidly defined.

2. It should be based on all the items.

3. It should not be unduly affected by extreme items.

4. It should be capable of algebraic manipulation.

5. It should be simple to understand and easy to calculate.

Formula for CV and SE

Coefficient of variation (C.V) = $\frac{\sigma}{\overline{X}} \times 100$

Where, σ = Standard deviation and $\overline{X}$ = Arithmetic mean.

Standard Error (S.E.) of the mean = $\frac{\sigma}{\sqrt{n}}$

Where, σ = Standard deviation and n = Number of observations.

Q. 4. Discuss exhaustive events and mutually exclusive events. State law of addition and multiplication.

Ans. **Exhaustive Events:**

The total number of outcomes in any trial is known as exhaustive events.

Ex. In tossing of a coin there are two exhaustive cases i.e. head and trial.

1. **Mutually Exclusive Events:**

Events are said to be mutually exclusive if the happening of any one of them precludes the happening of all the others i.e. no two or more of them can happen simultaneously in the same trial. Ex. In tossing a coin the events head and tail are mutually exclusive.

Law of addition:

If A and B are two arbitrary (possible) events in the sample space S, the probability of the union of A and B is governed by the law,

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Law of Multiplication:

If two events A and B are independent, the probability of their product of their individual probabilities.

$$P(AB) = P(A \cap B) = P(A)\,P(B)$$

Q. 5 Define Average. State the properties of good average. Give the Empirical relationship between Mean, Median and Mode.

Ans. Average is defined as the sum of the observations divided by the number of observations.

Properties for good average

The following properties should possess for an ideal average.

1. It should be rigidly defined.
2. It should be easy to understand and compute.
3. It should be based on all items in the data.
4. Its definition shall be in the form of a mathematical formula.
5. It should be capable of further algebraic treatment.
6. It should have sampling stability.

Empirical relationship between mean, median and mode

In a symmetrical distribution the three simple averages mean = median = mode. For a moderately asymmetrical distribution, the relationship between them are brought by Prof. Karl Pearson as

mode = 3median - 2mean.

Q. 6 Define test of significance. State the assumptions and applications of t-test.

Ans. **Test of significance:**

Student's t is the deviation of estimated mean from its population mean expressed in terms of standard error. The test criteria is; reject H_0 if $t \geq t_{\alpha/2,n-1}$ or $-t \leq -t_{\alpha/2,n-1}$, otherwise accept H_0. Let Xi (i = 1,2,...n) be a random sample of size n from a normal population with mean μ and variance σ^2. Then t –test is defined as

$$t = \sqrt{n}\,(\bar{x} - \mu_0)/s, \quad -\infty \leq t \leq \infty$$

Where,

$$s = \sqrt{\Sigma(x-\bar{x})^2/(n-1)}$$

Assumption for t-test

The assumptions under which t-test is applied are given below:

1. Parent population is normal.
2. Sample has been selected randomly.
3. Sample size is small.
4. Population standard deviation is not known.

Applications of t-test

In small samples, students t-test is applied in the following cases.:

1. Comparison of a sample mean with the population mean (or a hypothetical mean).
2. Testing the significance of a mean difference (paired observations).
3. Comparison of two means from two independent samples.

Q. 7 Write short notes of following. (Any two)
i) 'O' give curve. ii) Normal Distribution
iii) Poisson Distribution

Ans. **i) Ogives:**

For a set of observations, we know how to construct a frequency distribution. In some cases we may require the number of observations less than a given value or more than a given value. This is obtained by an accumulating (adding) the frequencies upto (or above) the give value. This accumulated frequency is called cumulative frequency. These cumulative frequencies are then listed in a table is called cumulative frequency table. The curve table is obtained by plotting cumulative frequencies is called a cumulative frequency curve or an ogive. There are two methods of constructing ogive namely:

1. The 'less than ogive' method
2. The 'more than ogive' method.

In less than ogive method we start with the upper limits of the classes and go adding the frequencies. When these frequencies are plotted, we get a rising curve. In more than ogive method, we start with the lower limits of the classes and from the total frequencies we subtract the frequency of each class. When these frequencies are plotted we get a declining curve.

ii) Normal Distribution:

A r. v. X is said to have a normal distribution with parameters μ (called mean) and σ^2 (called variance) if its density function is given by the prob. Law:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}; \quad -\infty < x < \infty, -\infty < \mu < \infty, \sigma > 0.$$

Properties of Normal Distribution.

1. The curve is bell shape and symmetrical about the line $x = \mu$.
2. Mean, Median and mode are coincide i.e. mean = median =mode.
3. Quartiles are equidistant from median.
4. Linear combination of independent normal variates is also a normal variate.
5. The mode of the normal curve lies at the point $x = \mu$.
6. The area under the normal curve within its range $-\infty$ to ∞ is always unity.
7. The normal curve is unimodal.
8. All odd order moments of the normal distribution are zero.

iii) Poisson Distribution:

A r. v. X is said to follow a Poisson distribution if it assumes only non – negative values; and its probability mass function is given by

$$P(X = x) = \begin{cases} \frac{e^{-\lambda}\lambda^x}{x!}; x = 0,1,2,\ldots,n \\ 0; otherwise \end{cases}$$

Where λ is known as parameter of the distribution. A r. v. X is follows P.D. with parameter λ is denote as $X \sim P(\lambda)$.

Properties of Poisson distribution:

1. Poisson distribution has only one parameter.
2. Mean of Poisson distribution (variate) is λ.
3. Variance of Poisson distribution is also λ. This is the only distribution of which mean and variance are equal.
4. Moment generating function of Poisson distribution is $e^{\mu(e^{it}-1)}$.
5. If X and Y are two Poisson variates with mean μ and λ respectively, the conditional distribution of X given (X +Y) is binomial.

Q. 8 Define Geometric Mean and Harmonic Mean. Write down merits and demerits of Geometric Mean. Give the formula for Harmonic Mean of grouped data.

Ans. **Geometric Mean:** The geometric mean of a series containing n observations is the n^{th} root of the product of the values.

Harmonic mean (H.M) : Harmonic mean(H.M.) of a set of observations is defined as the reciprocal of the arithmetic average of the reciprocal of the given values.

Merits of Geometric mean :

1. It is rigidly defined
2. It is based on all items
3. It is very suitable for averaging ratios, rates and percentages
4. It is capable of further mathematical treatment.
5. Unlike AM, it is not affected much by the presence of extreme values

Demerits of Geometric mean:

1. It cannot be used when the values are negative or if any of the observations is zero
2. It is difficult to calculate particularly when the items are very large or when there is a frequency distribution.
3. It brings out the property of the ratio of the change and not the absolute difference of change as the case in arithmetic mean.
4. The GM may not be the actual value of the series.

Harmonic Mean is given as

$$H.M. = \frac{N}{\sum_{i=1}^{n} \frac{f}{x}}$$

Q. 9 Define Skewness and Kurtosis. Explain in brief the different measures of skewness.

Ans. **Skewness:** Lack of symmetry is said to be Skewness.

Kurtosis: The convexity of curve is said to be Kurtosis.

Different measures of Skewness:

i. $Sk = M - Mo$

ii. $Sk = M - Md$

iii. $Sk = (Q_3 - Md) - (Md - Q_1)$

These are the absolute measures of Skewness.

iv. Karl Pearson's Coefficient of Skewness

$$Skp = \frac{\text{Mean - Mode}}{\sigma}$$

Where, Skp is karl pearsons coefficient of skewness.

σ is standard deviation.

Bowleys coefficient of Skewness:

$$Skb = \frac{Q_3 + Q_1 - 2Md}{Q_3 - Q_1},$$

Where, Q_1 = First Quartile

Q_3 = Third Quartile

Q. 10 Explain the terms degrees of freedom, type I & type II errors, Null hypothesis, level of significance and critical region.

Ans. **Degrees of Freedom:** The number of independent observations in a set is called degrees of freedom.

Type I error: Reject H_0 when it is true. It is also known as rejection error and it is denoted by α.

Type II error: Accept H_0 when H_1 is true. It is also known as acceptance error and it is denoted by β.

Null hypothesis: The null hypothesis is defined as a particular hypothesis which is tested for rejection under the assumption that it is true, as distinct from the alternative hypotheses which are under consideration.

Level of significance: the maximum probability of rejecting the hypothesis when it is true is known as level of significance.

Critical Region: The hypothesis is rejected when an observed difference is equal to or greater thanδ, all the values, which are beyond from a region known as Critical region.

SECTION "B"

Q. 11 Fill up the blanks.

1. Standard deviation of the series 5,5,5,5,5 and 5 is **Zero.**
2. The sum of the deviations taken from the arithmetic mean is equal to Zero.
3. For a mesokurtic curve, β_2 equal to 3.
4. The number of seeds per pod is **Discrete** variable.
5. The middle value of an ordered series is called **Second** quartile.
6. Homogeneity of two sample variances is tested by **F test.**
7. Large sample test is applied when sample size is greater than 50.
8. In dependence of attributes is tested by **Chi-square** test.

Q. 12. Match the pairs.

'A'	'B'
1. t-test	a) The real number lies between 0 & 1
2. f-test	b) random experiment
3. χ^2 test	c) The value lies between -1 & +1
4. Normal Distribution	d) Discrete distribution
5. Poisson Distribution	e) Continuous distribution
6. Binomial Distribution	f) Mean & Variance are same
7. Probability	g) Testing equality of Variances
8. Correlation	h) Testing goodness of fit
	i) Used for testing equality of two small sample mean

Ans. [1] i [2] g [3] h [4] e
[5] f [6] d [7] a [8] c

MAHARASHTRA AGRICULTURAL UNIVERSITIES EXAMINATION BOARD, PUNE
SEMESTER END THEORY EXAMINATION
B.Sc. Agricultural Biotechnology

Semester : IV
Course No : CS-248 (New)
Credits : 2(1+1)
Day & Date :

Academic Year: 2013-14
Course Title: Biostatistics
Total Marks: 40
Time :

Model Answer

SECTION A

Q.1 Measures of central tendency: It is a single value within the range of data which represents a group of individual values in a simple and concise manner.

Following are the properties of good measures of central tendency:

1. It should be based on all observations.
2. It should be rigorously defined.
3. It should be easily computable.
4. It should be least affected by extreme values.
5. It should fluctuate least from sample to sample drawn from the same population.

Q.2 State and give the properties of normal distribution.

Ans- **Definition:** (1 mark)

It is a distribution of random variable which varies continuously.

Properties of normal distribution: (3 mark)

1) . The curve of the normal distribution is bell shaped and symmetrical about mean.
2) The shape of the curve is depends on μ, σ and n.
3) The majority of observations have tendency to cluster around them.
4) The variables vary continuously.

Q.3. Classification is the process of arranging things or items in groups or classes according to their resemblance and characteristics. It give expression to the units of attributes that may subsist amongst the diversity of individuals.

Following are the different modes of classification:

1. Geographical classification: classification is according to place, area or region.
2. Chronological classification: It is according to the lapse of time, eg. Monthly, yearly.
3. Qualitative classification: Data are classified according to the attributes of subjects or items. Eg . qualification, colour.
4. Quantitative classification: Data are classified according to the magnitude of the numerical values. Eg. Age, income, height.

Q.4 Following are the different measures of dispersion.

1. Range
2. Inquartile and quartile deviation.
3. Mean deviation
4. Variance
5. Standard deviation
6. Coefficient of variation.

Following are the uses of measures of dispersion:

1. It tells about the reliability of measures of central value.
2. It makes possible to compare two series of data in respect of their variability.
3. Measures of dispersion provide the basis for the control of variability.
4. It has a wide application in almost all fields of statistics.

Q.5. Event : An event is the collection of possible outcomes which are favorable to an happening out of total outcomes which may be enumerable or denumerable.

Consider a statistical experiment which may consist of finite or infinite number of trials whereas each trials results into an outcome such as tossing three coins at a time or tossing a coin three times. We shall have in all eight outcomes.

If we consider the birth weight of children, it will consist of an infinite number of outcomes.

The totality of all possible outcomes of an experiment is called sample space and is denoted by Ω.

Q.6 Regression coefficient: The regression coefficient β is a measure of change in dependent variable Y corresponding to an unit change in independent variable X.

Examples of use of regression analysis:

1. Height of a person at a given age can be estimated by finding the regression of height Y on age X.
2. The yields of a crop can be predicted for different doses of a fertilizer, of course within the nontoxic dose of fertilizer.
3. A scientist may be able to estimate the brain size on the basis of certain related external body measurements.
4. Future demand of food may be predicted with the help of regression models.
5. Certain treatments can be compared accurately by eliminating the effect of certain variables like percentage germination per unit area, number of weeds per unit area etc.

Q.7 Properties of correlation coefficient:

1. Correlation coefficient has no unit.
2. The correlation coefficient ranges from -1 to 1
3. The correlation between two variables is known as simple correlation.
4. It is not affected by change in origin and scale.
5. Correlation coefficient are symmetric.

Q.8. Define Experiment of design. Elaborate basic principles of design of experiment.

Ans- **Definition:** (1 mark)

The choice of treatment, the method of assigning treatments to experimental units and arrangement of experimental units in various patterns to suit the requirement of particular problems are combinely known as design of experiment.

Basic principles of design of experiment: (3 marks)

1) Replication
2) Randomization
3) Local control

Q.9. Following are the differences between exclusive and inclusive class intervals:

1. In exclusive class intervals the upper limit of a class is the lower limit of the next class. Also the upper limit of a class is not included in that class.
2. In inclusive class intervals the upper limit of a class instead is not the lower limit of the next class.
3. In inclusive class interval method both the limits of a class are included.
4. To simplify the calculation procedure, inclusive classes are converted into exclusive classes.
5. Inclusive classes approach is suitable in case of data given in whole numbers. In rest of the cases exclusive class approach is suitable.

Q. 10 What is hypothesis in statistics? Give the steps involved in the test of hypothesis.

Ans- **Definition:** (1 mark)

Hypothesis in statistics is simply a quantitative statement about a population. OR Statistical hypothesis means some assumptions of statement about a population.

Steps involved in the test of hypothesis. (3 marks)

1. The null and alternative hypothesis will be formulated
2. Test statistic will be constructed
3. Level of significance will be fixed
4. The table (critical) values will be found out from the tables for a given level of significance
5. The null hypothesis will be rejected at the given level of significance if the value of test statistic is greater than or equal to the critical value. Otherwise null hypothesis will be accepted.
6. In the case of rejection the variation in the estimates will be called significant variation. In the case of acceptance the variation in the estimates will be called not-significant.

Q.1 1. Fill in the blanks

1. coefficient of determination

2. homogeneity within block

3. Chi square test

4. Zero

Q.1 2. Multiple choice questions

1. (C) equality of variance

2. (A) t-test

3. (B) Human biases.

4. (B) Poisson distribution

SECTION "B"

Ans: 11.a) Match the pairs: 4Marks

"A"	
1) Bloat	e) Accumulation of gas.
2) Pica	a) P- deficiency
3) Tamarind seed powder	b) Non – conventional feedstuff
4) Succulent feed	c) high moisture content

b) Fill in the blanks. 4Marks

1. Aqueous NaOH
2. B
3. 10-15
4. 99

Ans: 12.a) Give abbreviations: 4Marks

1. Total digestible nutrients.
2. Acid detergent fibre.
3. Non-protein nitrogen.
4. Nitrogen free extract.

b) State true or false. 4Marks

1. True
2. True
3. False
4. False

MODEL ANSWER

MAHARASHTRA AGRICULTURAL UNIVERSITIES EXAMINATION BOARD,PUNE

SEMESTER END EXAMINATION

B.Sc.(Agri.)

Semester : II (New)
Course No.: STAT–121
Credits : 2 (1+1)
Day & Date :

Academic Year : 2007-2008
Title : Statistics
Total Marks : 40
Time :

MODEL ANSWERS

SECTION 'A'

Q. 1. What is statistics? What are the uses of statistics and Give limitations of the same.

Ans. Statistics is a science of counting. Statistics is science of estimates and probabilities statistics is defined as the science of collection, presentation, analysis and interpretation of numerical data.

Use of statistics:

In the early part of the growth period, statistics use was restricted to the states only. But today its scope has spread to the study of problems which may be social, religious, economic, political, administrative, commercial, agricultural, financial, medical, relating to business management, planning research education,. Psychology, forecasting and to so many other spheres.

This widely use of statistics is on account of the fact that the statistical principles have a very wide scope of application and their knowledge is very essential for any research work in any branch of study.

Limitation of statistics:

1) It does not study qualitative phenomena
2) Statistical studies are true only on an average
3) It does not study individuals
4) It does not reveal the entire story
5) It is liable to be misused

Q. 2. State properties of good average.Give the empirical relationship between mean, median and mode.

Ans. Properties of good average:

a) It should be rigidly defined
b) It should be rigidly understopd & easy to calculate
c) It should be based upon all the observations
d) It should suitable for further mathematical calculations
e) It should not be affected by the fluctuations of sampling.

Empirical Relation between mean, median and mode.

Mode = 3 median – 2 mean. Or 2 mean + 1 mode = 3 median

Q. 3. Illustrate the meaning of the terms a) simple event, b) compound event, c) Independent event, d) Mutually exclusive event, e) complementary event.

a) Simple event:

The occurrence of a single event is known as the simple event.

b) Compound event:

When two or more of simple events occur in connection with each other the joint occurrence is called the compound event

c) Independent event:

Events are said to be dependent or independent accordingly as the occurrence of one does or does not affect the occurrence of the others.

d) Mutually Exclusive event:

When the events are such that occurrence of any one is incompatible with the occurrence of any of the others, they are known as mutually exclusive events mutually exclusive events are such that only one will occur at a time.

e) Complementary event:

If $E_1, E_2, \text{----------}, E_n$ be n events, then $P(E_i)$ denotes the probability of happening E_i and $P(\bar{E}_i)$ denotes the probability that the event Ei does not happen. The event $\bar{E}_i$ is known as complementary event of Ei .(P)$E_i + P(\bar{E}_i) = 1$

Q. 4. Define the term sample. Explain the concept of Random sampling and standard error.

Ans. Sample:

A sample is the number of individuals each of which is a member of the population to be studied . It is expected to be the representative of the whole population.

Random sampling:

A simple random sampling is a method of selecting a sample of n units such that every one of all possible samples of size n, which are ${}^{N}C_{n}$ in number has an equal chance of being selected or it in that the chance of selecting every unit is the same.

Standard Error:

The term standard error of any estimate is used for a measure of the average magnitude of the difference between the sample estimate and the population parameter taken over all possible samples of the same size from the population.

The standard deviation of the sampling distribution of a statistic (estimate) is known as the standard error of that statistic (estimate)

It we take all possible samples from the population of the same size and get a sampling distribution of means, it can be proved that the mean of this sampling distribution of means is the population mean and its standard deviation the standard error of mean.

It in not possible to draw all possible samples, we get the estimate of the standard error from a single sample. If S be the standard deviation of sample of size N, the estimate of standard error of mean is given by $\frac{S}{\sqrt{n}}$

Q. 5. Define correlation coefficient between two variables. State its properties. How do you test its significance. Explain

Ans. The two variables are said to be correlated, if the change in one variable affect a change in other variable.

Properties of correlation coefficient (r):

1) Correlation coefficient remains unaffected by change of origin and scale.
2) Correlation coefficient in the geometric mean between the two regression coefficients. i,e, $r = \sqrt{b_{xy} * byx}$
3) When the change in one variable (X) is proportional to the change in ;the other variable (Y), the two variables are said to be perfectly correlated with each other. In such a case $r = +1$ or $r = -1$
4) Two independent variables must be uncorrelated (i,e $r = 0$) , but the converse is not true. Two uncorrelated variables may not be necessarily independent.
5) The correlation coefficients (r) lies between –1 and +1.

The standard error of r is of very little in the test of significance for testing the significance of r in case of small samples, we calculate t by using the following formula.

$$t = \frac{r}{\sqrt{1-r^2}}\sqrt{n-2}$$

This value of t is based on (n-2) d. f.

Q. 6. Give merits & demerits of Completely Randomized Design.

Ans.: Merits & Demerits of CRBD (CRD):

Merits:

i) Any number of treatments and replications can be used.
ii) The analysis is very easy and it remains easy if the no. of replications for each treat is different
iii) The method of analysis remains simple even if some data is missing
iv) The relative loss of information due to missing data is smaller than that with any other design.
v) The design in specially useful in small experiments where the supply of material is scarce and homogeneous, as the whole of the material is used in the Experiment.
vi) This design provides maximum number of degrees of freedom for the estimation of error as compared with other designs, for given number of treatments and for given number of number of replications.

Demerits:

There is one and the main objection in this design is on the ground of accuracy. Since there is no restriction on the randomization of the treatments , we cannot be sure about the fact that ;the units receiving one treatment are similar to those receiving the other treatment and therefore, the whole of the variation among the units enters into the experimental error.

Q. 7. Write short notes on (ANY TWO)

i) Normal Distribution:

The normal distribution was first discovered in 1733 by English mathematician De – Moivre, who obtained this continuous distribution as a limiting case of the binomial distribution

A random variable X is said to follow normal distribution with parameter μ (called mean) and σ^2 (Called variance) if it's density function is given by the probability law,

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}} \qquad -\infty < X < +\infty,\ -\infty < \mu < +\infty,\ \sigma > 0$$

The parameters of normal distribution is μ (Mean) and σ^2 (Variance).

Characteristics of Normal Distribution:

1) Normal distribution is a the oretical continuous distribution
2) Limit of this distribution is $-\infty$ to $+\infty$,
3) Normal distribution is a limiting form of binomial distribution Mean, mode and median of normal distribution are equal
4) Normal probability curve

The normal probability, curve with mean (μ) and variance (σ^2) is given by equation

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}} \qquad -\infty < X < +\infty,\ -\infty < \mu < +\infty,\ \sigma > 0$$

and has the following properties.

i) The curve is a bell shaped and symmetrical about the line $X = \mu$
ii) Mean, mode and median of distribution coincides.
iii) X axis is an asymptote to the curve.

Draw the Normal probability curve

ii) Ogive curve:

In drawing an ogive curve, variate values are represented on the x – axis and the cumulative frequencies on the Y – axis . It can be drawn in two ways.

1) 'Less than' cumulative frequency curve.
In this case the cumulation of frequencies is done from the end of lowest variate value.

The co-ordinates of the points, which are plotted, are as.

X - Co-ordinates = upper limit of the class interval

Y – co-ordinate = corresponding cumulative frequency.

Now a free hand smooth curve is drawn through these plotted points.

2) 'More than' cumulative frequency curve.

In this case the cumulation of frequencies is done from the end of highest variate value

The co-ordinates of the points, which are plotted, are as

X - Co-ordinates = Lower, limit of the class interval

Y – co-ordinate = corresponding cumulative frequency.

Here also, a free hand smooth curve is drawn through plotted points

Draw the curve.

The point of intersection of the 'less than' and "equal to or greater than" graphs represents the point of median.

iii) Lorenz curve:

Dispersion can be studied graphically with the help of Lorenz curve. In this method size of the items and frequencies are cumulated and taking total as 100, percentages are calculated for various cumulated values, these percentages are plotted on graph. Straight line is called as line of equal distribution ,curve away from the straight line indicates variability.

Draw the curve.

iv) Yates correction for 2 x 2 contingency table

In order to avoid irregularities caused by smaller frequencies in a 2 x 2 contingency table, a correction for continuity, known as Yate's correction, is applied.

This correction amounts to subtracting ½ from each of a and d and adding ½ to each of b and c and then calculating the corrected value of χ^2 from these adjusted frequencies. This applies when ad > bc. But if ad < bc the procedure is reversed.

The expression for calculating χ_c^2, the corrected value of

$$\chi_c^2 = \frac{\left[(ad - bc) - N/2\right]^2}{R_1 R_2 C_1 C_2} * N$$

Here, the difference between ad & bc is always positive.

Which shows that the correction N/2 always reduces the value of χ^2

SECTION'B'

Q. 8. State TRUE or FALSE.

i) Sum of deviations of a variable from arithmetic mean is equal to zero.

Ans.: True

ii) The variance of standard Normal variate is zero.

Ans.: False

iii) Probability of Null event is one

Ans.: False

iv) All the three principles of field Experimentation are used in Randomized Block Design.

Ans.: True

v) A series is said to be consistent when its coefficient of variation is least.

Ans.: True

Q. 9. Write down the formulae with terms and specifications used.

i) Median for continuous series

Ans.: $Median = l_1 + (\frac{m-c}{f}) * (l_2 - l_1)$

Where, l_1 is the lower limit of the median class

l_2 is the upper limit of the median class

m= (n+1)/2

c is the cumulative frequency preceding to the median class.

f = frequency of the median class

ii) Standard Deviation for Discrete series.

$$S\tan dardDeviation = \sqrt{\frac{\sum f\left(X - \bar{X}\right)^2}{n}}$$

iii) The line of regression of X on Y.

Y= c + d Y Where, c= constant and d is the regression coefficient

iv) Treatment sum of square in completely Randomised Design with unequal number of replications.

$$Treatment Sum of square = \sum \frac{T_i^2}{r_i} - \frac{G^2}{N}$$

v) Karl Personas coefficient of correlation

$$r = \frac{\sum XY - \frac{\sum X \sum Y}{n}}{\sqrt{\sum X^2 - \frac{(\sum X)^2}{n}} * \sqrt{\sum Y^2 - \frac{(\sum Y)^2}{n}}}$$

Q. 10. Match the pairs.

1) Significance of attributes	e) Chi square test
2) Grand Total	d) Σ Xi
3) Latin square	b) Two way variation
4) C R D	a) Laboratory Experiment
5) Test for homogeneity of variance	c) 'F' test