Model Answer

Department of forestry, Wildlife and Environmental Sciences

Course: General Statistical Methods and Research Methodology

Class: M.Sc. (Forestry) I Semester MM:60

Note: Calculator is allowed. First question is compulsory which carries 20 marks.

Section-A

Q.1. **(a) Choose the correct answer**

(i) An average used to determine soil fertility is

a. Mean b. Mode c. Median d. None of these

Ans. b) Mode

(ii) Two tailed test means

a. $\mu_1 > \mu_2$ b. $\mu_1 < \mu_1$ c. $\mu_1 = \mu_1$ d. $\mu_1 \neq \mu_1$

Ans. d) $\mu_1 \neq \mu_1$

(iii) The degrees of freedom for 10 treatments in the CRD is

a. t-10 b. 10-t c. 9 d. None of these

Ans. c) 9

(iv) Replication is the repetition of the___________ in an experiment

a. Experimental unit b. Treatments c. Error d. None of these

Ans. b) Treatments

(v) Factorial design is

a. Split plot design b. CRD c. RBD d. None of these

Ans. a) Split plot design

(vi) If the sample size is large (mare than30) than the used test to determine test of significance is

a. Chi square test b. t- test c. paired t-test d. z-test

Ans. d) z-test

(vii) Calculate mean, if mode is 10 and median is 8

a. 7 b. 9 c. 10 d. None of these

Ans. a) 7

(viii) Correlation coefficient is lies between

a. +1 to 0 b. +1 to - 1 c. +1 to 0 d. None of these

Ans. b) +1 to - 1

(ix) The different possible results are known as

a. Experiment b. Event c. Outcomes d. None of these

Ans. c) Outcomes

(x) When we reject the hypothesis when in reality it is true. Probably committed

a. Type –I error b. Type-II error c. Standard error d. Probability

Ans. a) Type –I error

(b) Fill in the blanks (10)

(i) If median= 3 and mean=2.5, than Mode=___________.

Ans. 4

(ii) 20, 9, 6, 21, 30, 21, 5 are the different observation has been taken. Range coefficient for the mentioned observation is= ____________.

Ans. 0.6

(iii) From a pack of 52 cards drawn of a card at random, find the probability of getting a king is ___________.

Ans. 1/13

(iv) If the Standard deviation is 2 and sample mean is 68, than the coefficient of variance is = __________.

Ans. 2.94

(v) Formula for χ^2 test= ___________.

Ans. $\Sigma(O-E)^2/E$

Section -B

Note: Attempt any four questions. Each question carries 10 marks.

Q.2. (a). What is diagrammatic representation? Discuss?

Ans. One of the most effective and interesting alternative way in which a statistical data may be presented is through diagrams. There are several ways in which statistical data may be displayed pictorially such as different types of graphs and diagrams. The commonly used diagrams and graphs to be discussed in subsequent paragraphs are given as under:

Types of Diagrams

1. One dimensional diagrams: Line, simple bars, multiple and compound, sub-divided bars
2. Two dimensional diagrams: Square and pie diagram
3. Three dimensional diagrams: solid cubes, square
4. Pictograms: Maps and pictures

Utility of diagrams:

1. Attractiveness and effective
2. Easy to understand

3. Readily intangible
4. Comparisons possible
5. Saving of much valuable time and energy
6. Visible and clear at a glance

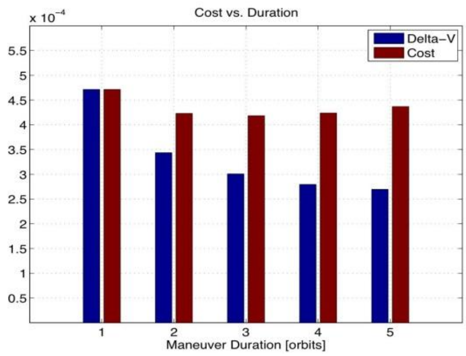
1) Simple Bar

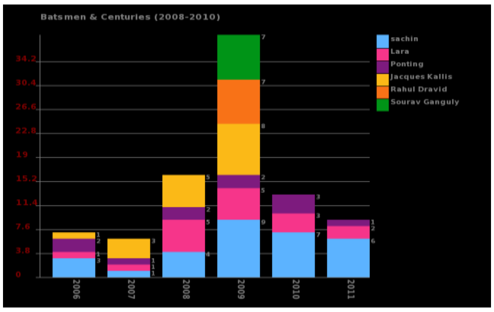
a) Label all the items, units and scales on the horizontal and vertical axes

(b) All the bars should have equal width and be evenly spaced

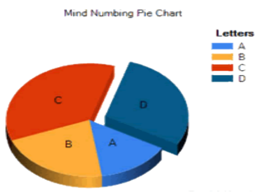
(c) Give a title to the bar chart

2) Multiple Bar **3) Staked Bar or Sub-Divided Bar or Component Bar**





4) Pie chart: Here instead of comparing the length of a bar, the areas of segments of a circle are compared. The area of each segment depends upon the percentage, which is converted to angle and drawn.



5) Pictorial representation

6) Square: Two dimensional

7) Cubes: Three dimensional

(b). Define statistics? Role of statistics in forestry research?

Ans. Statistics is the study of the collection, organization, analysis, interpretation and presentation of data.

Role of statistics in forestry research as follows:

General statistics: In classical or Mendelian genetics the focus of interest centered on the inheritance of qualitative characters and the statistical methods were those applicable to the nominal scale like binomial or chi-square test. Testing the agreement of observed frequency data with those expected by rhe hypothesis of Mendelian segregation was the major task and this included problems such as detection and estimation of linkage. In contrast, the methods of the quan- titative genetics dealt with the study of continuous variation caused by action of polygenes, environmental factors and their interaction.

Numerical taxonomy : Numerical taxonomy deals with grouping of taxonomic units into taxa by numerical methods on the basis of their character states. The term includes the drawing of phylogenetic inferences from the data by statistical or other mathematical methods to the extent to which this is possible. The major advantages of using numerical methods for classification are repeatability and objectivity. After having chosen the organisms and the characters to be recorded on them, further steps include working out the resemblances between organisms and constructing taxa based upon these resemblances. Generalizations are then made about taxa such as inference about their phylogeny, choice of discriminatory characters.

Statistical ecology: Most of the mathematical and statistical applications in ecology deaf with the study of temporal and spatial patterns of populations of organisms.

Quantitative ethology : Like in the case of many other disciplines the need for quantification in ethology arises from the presence of individual differences in animal behaviour and the need for an objective criterion to decide between cases where marginal differences exist. Behavioural studies in the usual case yield time series data.

Mensuration: Mensuration is concerned with measurements in forestry. Measurements arise in terms of billets stacked or otherwise, individual trees or whole stands and the features measured will quite often be length, area or volume and weight in a few cases.

Site quality: Site quality is supposed to be a measure of the productive capacity of a designated land area as measured from the existing condition of a stand growing there. Naturally it is a post planting concept and is highly related to the planting material and the associated envlronment 'including management.

Yield table: Yield table is a tabular statement which summarises all the essential data relating to the development of a fully stocked and regularlythinned, even: aged crop at periodic intervals covering greater part of its useful life.

Volume table: Volume table is essentially the output of a regression function with tree volume as the dependent variable and diameter and/or height as independent variable (s). In practice the best suited function out of a set of polynomial or exponential models is selected based on some goodness of fit criterion.

Taper functions : Taper is the decrease in diameter of a tree stem or a log from base upwards and the rate of taper is termed form. Taper varies not only between but also within trees.

Biomass estimation: Biomass refers to the total mass of living material in a location and its estimation has assumed considerable importance in recent years. Foresters are mostly concerned with tree biomass and the common procedure of estimation is through the use of regression equations and stand tables.

Forest management: Forest management in the broad sense, integrates all of the biological, social, economic and other factors that affect management decisions about the forest. Historically, forest management has dealt primarily with silviculture and the biological management of the forests.

Econometrics: Econometrics deals with the application of statistical techniques in the field of economics.

Forest regulation: Traditional forest regulation has centered around the concepts of sustained yield and normal forest. The implicit objectives in most historical regulation schemes are to maximize the volume harvested and to maintain an increasing or eventually even wood flow.

Forest valuation: Other than the objectives and constraints themselves a prerequisite for decision making is an appraisal of the alternatives. The alternatives can be assessed by placing a value on forest production.

Q.3. (a) Calculate the mean by the step deviation method for the following frequency distribution of diameter class of Peltophorum trees:

Diameter class (mm)	0-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80	80-90	90-100
Frequency	5	6	6	4	7	12	13	1	2	4

Solve:

Class	**Frequency (f)**	**Mid value**	**d= x-A**	**fd**
0-10	5	5	-50	-250
10-20	6	15	-40	-240
20-30	6	25	-30	-180
30-40	4	35	-20	-80
40-50	7	45	-10	-70
50-60	12	55	0	0
60-70	13	65	10	130
70-80	1	75	20	20
80-90	2	85	30	60
90-100	4	95	40	160
Total	**60**			**-450**

Mean= A+ Σ fd/N

A=55

Mean= 55+ (-450/60)

Mean= 55-7.5

Mean= 47.5 Ans.

(b) Calculate the mode for given classes:

Diameter class (mm)	5-10	10-15	15-20	20-25	25-30	30-35	35-40	40-45	45-50
Frequency	2	8	10	15	18	14	7	8	3

Solve:

Diameter class (mm)	**Frequency (f)**	
5-10	2	Mode= L+D1/D1+D2*I
10-15	8	D1= 18-15=3
15-20	10	D2= 18-14=4
20-25	15	Mode= 25+3/3+4*5
25-30	18	Mode= 27.1 Ans.
30-35	14	
35-40	7	
40-45	8	
45-50	3	

Q.4 Calculate Mean Deviation and Quartile Deviation for the given value:

Trees	1	2	3	4	5	6	7	8	9	10
Height (cm)	30	30	35	35	40	40	45	45	50	50

Solve:

S. No.	Height	$X-\overline{X}$	$\lvert X-\overline{X} \rvert$	
1	30	-10	10	$\overline{X}$=400/10
2	30	-10	10	=40
3	35	-5	5	
4	35	-5	5	Mean Deviation= 1/N Σ$\lvert X-\overline{X} \rvert$
5	40	0	0	
6	40	0	0	MD= 1/10 (60)
7	45	5	5	
8	45	5	5	**MD=6**
9	50	10	10	
10	50	10	10	
Total	400		60	

Quartile Deviation= Q3-Q1/2

Q1=N/4+1

Q1= 10+1/4

Q1= 2.75

Q1= 30+ 0.75(35-30)

Q1= 30+ 0.75(5)

Q1= 30+ 3.75

Q1= 33.75

Q3= 3(N+1/4)

Q3= 46.25

QD= 46.25-33.75/2

QD= 12.5/2

QD= 6.25

Q.5. Calculate Regression line X on Y and Y on X and Regression coefficient?

Fertilizer	8	10	12	15	18	20	22	25	27	13
Height (cm)	5	7	10	11	16	17	18	20	22	9

Solve:

X	**$X-\overline{X}$**	**$(X-\overline{X})^2$**	**Y**	**$Y-\overline{Y}$**	**$(Y-\overline{Y})^2$**	**$(X-\overline{X})(Y-\overline{Y})$**
8	-9	81	5	-8.5	72.25	76.5
10	-7	49	7	-6.5	42.25	45.5
12	-5	25	10	-3.5	12.25	17.5
15	-2	4	11	-2.5	6.25	5
18	1	1	16	2.5	6.25	2.5
20	3	9	17	3.5	12.25	10.5
22	5	25	18	4.5	20.25	22.5
25	8	64	20	6.5	42.25	52
27	10	100	22	8.5	72.25	85
13	-4	16	9	-4.5	20.25	18
Total=170		**374**	**135**		**306.5**	**335**

Solve: $\overline{X}$=170/10=17

$\overline{Y}$=135/10=13.5

X on Y

$(X-\overline{X}) = \Sigma(X-\overline{X})(Y-\overline{Y}) / \Sigma (Y-\overline{Y})^2 * (Y-\overline{Y})$

X-17=335/306.5*(Y-13.5)

X= 1.08Y+2.24

Y on X

Y-13.5=335/374* (X-17)

Y= 0.84X-1.72

Regression Coefficient (r) $=\sqrt{\text{bxy.byx}}$

r $=\sqrt{\mathbf{(0.89)(1.09)}}$

r=0.98

Q.6. Write short notes on the following

(i) Measures of central tendency

It is occasionally called an average or just the center of the distribution. The most common measures of central tendency are the arithmetic mean, the median and the mode. A central tendency can be calculated for either a finite set of values or for a theoretical distribution, such as the normal distribution. Occasionally authors use central tendency (or centrality), to mean "the tendency of quantitative data to cluster around some central value.The following may be applied to one-dimensional data. Depending on the circumstances, it may be appropriate to transform the data before calculating a central tendency.

Arithmetic mean– the sum of all measurements divided by the number of observations in the data set

Median – the middle value that separates the higher half from the lower half of the data set. The median and the mode are the only measures of central tendency that can be used for ordinal data, in which values are ranked relative to each other but are not measured absolutely.

Mode – the most frequent value in the data set. This is the only central tendency measure that can be used with nominal data, which have purely qualitative category assignments.

Geometric mean – the nth root of the product of the data values, where there are n of these. This measure is valid only for data that are measured absolutely on a strictly positive scale.

Harmonic mean – the reciprocal of the arithmetic mean of the reciprocals of the data values. This measure too is valid only for data that are measured absolutely on a strictly positive scale.

Weighted mean – an arithmetic mean that incorporates weighting to certain data elements.

(ii) Randomization

The first principle of an experimental design is randomization, which is a random process of assigning treatments to the experimental units. The random process implies that every possible allotment of treatments has the same probability. An experimental unit is the smallest division of the experimental material and a treatment means an experimental condition whose effect is to be measured and compared. The purpose of randomization is to remove bias and other sources of extraneous variation, which are not controllable. Another advantage of randomization (accompanied by replication) is that it forms the basis of any valid statistical test. Hence the treatments must be assigned at random to the experimental units. Randomization is usually done by drawing numbered cards from a well-shuffled pack of cards, or by drawing numbered balls from a well-shaken container or by using tables of random numbers.

(iii)Probability

Ans.: Most experimental searches for paranormal phenomena are statistical in nature. A subject repeatedly attempts a task with a known probability of success due to chance, then the number of actual successes is compared to the chance expectation. If a subject scores consistently higher or lower than the chance expectation after a large number of attempts, one can calculate the probability of such a score due purely to chance, and then argue, if the chance probability is sufficiently small, that the results are evidence for the existence of some mechanism (precognition, telepathy, psychokinesis, cheating, etc.) which allowed the subject to perform better than chance would seem to permit. Some terms used in probability as follows:

Experiment and outcomes: Any operation on certain objects which gives different results known as experiment and the different possible results are known as its outcomes.

Events: One outcome or a group of outcomes associated with certain conditions is known as an event.

Equally likely cases or event: the cases or outcomes are said to be equally likely when we have no reason to except any one more rather than the other.

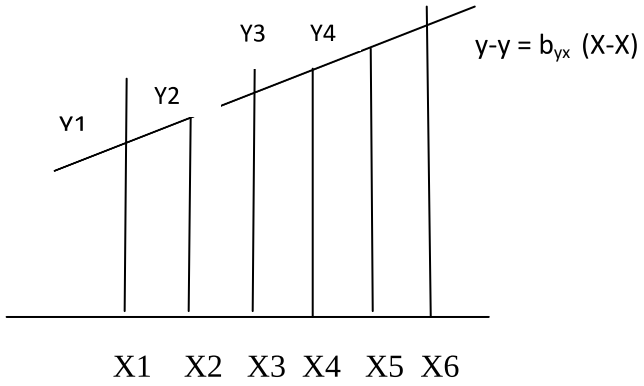
Favourable cases: Out of all equally likely cases those cases which fulfill the conditions of the event or the cases which entail the happening of the event are known as favourable cases for that event.

(iv) Scatter diagram

Ans.: This is a graphical representation of a bivariate ungrouped data which is also known as dot diagram. In this representation, independent variable variable is represented on the axis of x and the axis of X and the dependent one on the axis of Y. To each member of the sample there is X and whose ordinate (y-coordinate) is Y. with these coordinates points are plotted on a graph paper and the figure, which is thus obtained is called a scatter diagram.

Variable: X	X1	X2	X3	X4	X5	X6
Variable: Y	Y1	Y2	Y3	Y4	Y5	Y6

The scatter diagram for these values will be as shown in the following diagram:



With the help of this diagram we can interpret the ways in which the points cluster or scatter, as the properties of the relationship between the two variables.

If data is given in pairs then the scatter diagram of the data is just the points plotted on the xy-plane. The scatter plot is used to visually identify relationships between the first and the second entries of paired data.

The scatter plot above represents the age vs. size of a plant. It is clear from the scatter plot that as the plant ages, its size tends to increase. If it seems to be the case that the points follow a linear pattern well, then we say that there is a high linear correlation, while if it seems that the data do not follow a linear pattern, we say that there is no linear correlation. If the data somewhat follow a linear path, then we say that there is a moderate linear correlation.

Q**.7. (a) Discuss about the principles of experimental design?**

Ans. **Principles of design-** An experimental study is a scientific test that investigates the relationship between an outcome and one or more conditions manipulated by the researcher.

• Before considering the appropriate experimental design, it is important to be clear about the aims of any experiment, which are usually associated with one or more scientific questions or hypotheses to be tested.

• A thorough definition of the objectives of the design is required to make easy to assess whether the chosen treatments are sufficient the assess these

• The term treatment will be understood as the set of different experimental conditions to be tested.

• Asides from identifying the experimental treatments, experimental units must be chosen. Experimental Units (EU) are “the smallest division of the experimental material such that any two units may receive different treatments in the actual experiment”.

The basic principles of experimental designs are randomization, replication and local control. These principles make a valid test of significance possible. Each of them is described briefly in the following subsections.

(1) Randomization: The first principle of an experimental design is randomization, which is a random process of assigning treatments to the experimental units. The random process implies that every possible allotment of treatments has the same probability. An experimental unit is the smallest division of the experimental material and a treatment means an experimental condition whose effect is to be measured and compared. The purpose of randomization is to remove bias and other sources of extraneous variation, which are not controllable. Another advantage of randomization (accompanied by replication) is that it forms the basis of any valid statistical test. Hence the treatments must be assigned at random to the experimental units. Randomization is usually done by drawing numbered cards from a well-shuffled pack of cards, or by drawing numbered balls from a well-shaken container or by using tables of random numbers.

(2) Replication: The second principle of an experimental design is replication; which is a repetition of the basic experiment. In other words, it is a complete run for all the treatments to be tested in the experiment. In all experiments, some variation is introduced because of the fact that the experimental units such as individuals or plots of land in agricultural experiments cannot be physically identical. This type of variation can be removed by using a number of experimental units. We therefore perform the experiment more than once, i.e., we repeat the basic experiment. An individual repetition is called a replicate. The number, the shape and the size of replicates depend upon the nature of the experimental material. A replication is used

(i) to secure more accurate estimate of the experimental error, a term which represents the differences that would be observed if the same treatments were applied several times to the same experimental units;

(ii) to decrease the experimental error and thereby to increase precision, which is a measure of the variability of the experimental error; and

(iii) to obtain more precise estimate of the mean effect of a treatment, since , where denotes the number of replications.

(3) Local Control : It has been observed that all extraneous sources of variation are not removed by randomization and replication. This necessitates a refinement in the experimental technique. In other words, we need to choose a design in such a manner that all extraneous sources of variation are brought under control. For this purpose, we make use of local control, a term referring to the amount of balancing, blocking and grouping of the experimental units. Balancing means that the treatments should he assigned to the experimental units in such a way that the result is a balanced arrangement of the treatments. Blocking means that like experimental units should be collected together to form a relatively homogeneous group. A block is also a replicate. The main purpose of the principle of local control is to increase the efficiency of an experimental design by decreasing the experimental error. The point to remember here is that the term local control should not be confused with the word control. The word control in experimental design is used for a treatment. Which does not receive any treatment but we need to find out the effectiveness of other treatments through comparison.

(**b) Write the structure of analysis of variance for CRD?**

Computing total SS: The SStotal is the SS based on the entire set of scores in the study. So computing this SS is the same as if you just stacked your different treatment samples together to form a single sample and then computed the Sum of Squares on that one larger sample. In terms of our new notation:

	V1	**V2**	**V3**	**V4**	**TOTAL**
	1	2	3	4	**10**
	1	2	3	3	**9**
	2	2	3	4	**11**
	1	1	2	4	**8**
Total	**5**	**7**	**11**	**15**	**38**

Grand total = 38

Correction factor= GT/N

CF= 38/16

CF=2.38

Total S S=$1^2+1^2+2^2+1^2+$-----------------------$+4^2$-CF

Total S S= 1+1+4+1+4+4+4+1+9+9+9+4+16+9+16+16-2.38

Total S S= 105.62

2) Treatment S S:

To compute treatment S S = $5^2+7^2+11^2+15^2/4-2.38$
Treat. S S= 25+ 49+121+225/4-2.38
Treat. S S= 25+ 49+121+225/4-2.38
Treat. S S= 105-2.38
Treat. S S= 102.62
3) Error SS (within treatment):

Error SS = Total SS- Treatment S S
Error SS =105.62-102.62

Error SS (within treatment) = 3

In computing the degrees of freedom, you should keep in mind that:
Finally, you should always present what is called an Anova Summary Table that contains the results of all of these calculations. It should look like:

Source	df	SS	MS	F
Treatment	3	102.62	34.20	136.8
Error	12	3	0.25	