

P.G. Diploma in Bio - Informatics
Annual Examinations – 2006

Paper PBID – 102
BIOSTATISTICS

Time allowed: Three hours

Maximum Marks: 80

1. Attempt all questions from Section I.
- 2 Attempt any six questions from Section II.
- 3 Attempt any three questions from Section III

SECTION – I

Marks
(1 x 20)

Q1. Fill in the blanks

- (i) When the principles of statistics are applied to a study of organisms or living systems, the study is called _____.
- (ii) A continuous variable can assume _____ in a given interval.
- (iii) The range is the _____ the largest value and the smallest value.
- (iv) If A and B are mutually exclusive events then $P(A \text{ and } B) = \underline{\hspace{2cm}}$.
- (v) In binomial distribution probability of r success $p(r) = \underline{\hspace{2cm}}$.
- (vi) Chi square may be defined as the _____ between observed and expected frequencies.
- (vii) The 99% confidence limit of population means μ is _____.
- (viii) Null hypothesis H_0 is rejected if table value of t is _____ the calculated value of t.
- (ix) Chi square is a _____ test.
- (x) Vital statistics signifies the _____ occurring in communities.

Write 'true' or 'false' for the following statements.

- (xi) The data being collected first time by the investigators with some objectives are called 'primary data'.
- (xii) A frequency polygon can also be drawn for continues frequency distribution without histogram.
- (xiii) The algebraic sum of the derivatives of a set of n values from their arithmetic mean is zero.
- (xiv) Whenever we need to compare the variability of two series coefficient of variation is calculated for each series.
- (xv) Skewness means the deviation from symmetry.
- (xvi) Head and tail of a fair coin are two independent events.
- (xvii) The mean of the binomial distribution is \sqrt{np} .
- (xviii) t-test is based on large sample size of data.
- (xix) In the analysis of variance, all population from which samples have been drawn is not normal.
- (xx) The two rates most frequently used in study of diseases in a community are the incidence rate and prevalence rate.

SECTION – II

Q2. (i) What are the different types of classifications? (5)

(ii) Prepare the frequency distribution for the following figures giving haemoglobin values. One class should be 9.5-10.5(exclusive method) (5)

11.6, 10.3, 9.7, 12.4, 11.4, 10.7, 11.1, 9.5
13.0, 10.3, 9.9, 9.0, 11.4, 9.6, 11.0, 11.7
8.9, 10.8, 10.0, 11.5, 7.5, 9.8, 12.3, 11.4
11.9, 10.3, 8.5, 9.8, 11.8, 10.7, 8.9, 9.5
11.8, 10.0, 7.9, 10.3, 11.4, 11.8, 11.8, 8.8
10.6, 11.0, 11.9, 12.9, 9.2, 8.0, 8.9, 12.3

(iii) Calculate Arithmetic mean and Mode of the following data (5)

Protein intake:	15-25	25-35	35-45	45-55	55-65
No. of families:	30	40	100	110	80

(iv) For a group of 200 candidates the mean and standard deviation were found to be 40 and 15. Later it was discovered that the score 43 was misread as 53. Find the correct mean and standard deviation corresponding to the correct figure. (5)

(v) Define the following terms:

(i) Probability (ii) Conditional Probability (iii) Mutually exclusive events (iv) Equally likely events (v) Multiplication theorem. (5)

(vi) Define Method of Maximum Likelihood. Write some of its properties. (5)

(vii) What do you mean by measures of variation? What are different commonly used measures of variation? (5)

(viii) Define the following

(i) Crude Birth Rate (ii) Age specific fertility rate
(iii) Standardized Mortality Rate (iv) Morbidity Indication
(v) General Fertility Rate.

SECTION – III

Q3. From the following data calculate coefficient of variation. 10

Heart Diam.	90-100	100-110	110-120	120-130	130-140	140-150	150-160
No of cases	8	18	25	22	22	14	6

Q4. A problem in statistics is given to three students A, B and C whose probability of solving it are $\frac{1}{2}$, $\frac{3}{4}$, $\frac{1}{4}$ respectively. What is the probability that the problem will be solved if all of them try independently?

Q5. Two laboratories A and B carry out independent estimates of fat content in ice cream made by a firm. A sample is taken from each batch, halved and the separate halves sent to the two laboratories. The fat content obtained by the laboratories is recorded below.

Batch No.	1	2	3	4	5	6	7	8	9	10
Lab A :	7	8	7	3	8	6	9	4	7	8
Lab B :	9	8	8	4	7	7	9	6	6	6

Is there a significance difference between the mean fat content obtained by the two labs A and B.

You may use the following extracts from t-table in answering the question

Degree of freedom :	9	10	16	18	20
t at 5% level:	2.26	2.23	2.12	2.10	2.09

Q6. In a feeding experiment on swines, three rations R_1 , R_2 , R_3 were tried. The animals were put into three classes of three each according to Litter and initial body weight. The following table gives the gain in body weight in lbs in a certain period.

Rations	CLASSES		
	I	II	III
R_1	4	16	10
R_2	14	18	19
R_3	3	14	7

Analyse the data and state your conclusion. Has division into classes prove effective
[Given $F_{2,4}(.05) = 6.94$]