

**M.S. IN BIOTECHNOLOGY (Bioinformatics)
(MSBOBI)**

Term-End Examination

December, 2014

MBOI-016 : BIOSTATISTICS

Time : 3 hours

Maximum Marks : 100

Note :

- (i) *Section I is compulsory.*
- (ii) *Answer any five questions from Section II.*
- (iii) *Assume suitable data wherever required.*
- (iv) *Scientific calculators are permitted.*
- (v) *Statistical tables are provided.*
- (vi) *Italicized figures to the right indicate maximum marks.*

SECTION I

1. Explain/Define the following : 5×2
- (a) Descriptive Statistics
- Differentiate between the following :
- (b) Statistic and Parameter
 - (c) Cross-sectional data and Time series data
 - (d) Continuous variable and Discrete variable with an example
 - (e) When is it appropriate to use the arithmetic mean as opposed to the median ?

2. The following table shows the frequency distribution of ages of 169 subjects. Compute the mean, median, variance and standard deviation.

10

Class interval	Frequency
10 – 19	4
20 – 29	66
30 – 39	47
40 – 49	36
50 – 59	12
60 – 69	4

3. Consider an experiment of tossing two coins. Assume the coins are not balanced (not fair). The design of the coins is to produce the following probabilities shown in the table :

Sample Space	Probability
HH	$\frac{4}{9}$
HT	$\frac{2}{9}$
TH	$\frac{2}{9}$
TT	$\frac{1}{9}$

What is the probability of observing

- (a) *exactly* one Head and
(b) *at least* one Head ?

10

SECTION II

4. (a) Explain the following terms : 2+2
- (i) Subjective probability
 - (ii) Conditional probability
- (b) Determine the coefficient of correlation from the following table : 10

X	5	8	10	14	17	19	21	22	25	28
Y	16	14	10	9	7	6	5	3	3	1

5. (a) What is a sampling distribution ? Why is the sample mean \bar{X} used to estimate the population mean μ ? What is the difference between a standard deviation and a standard error ? 2+2+1
- (b) Suppose it is known that the response time of healthy subjects to a particular stimulus is normally distributed with a mean of 15 sec and variance of 6. What is the probability that a random sample of 16 subjects will have a mean response time of 12 sec or more ? 5

(c) Suppose we have a sample of five values of hemoglobin A1c (HgbA1c) obtained from a single diabetic patient. HgbA1c is a serum measure often used to monitor compliance among diabetic patients. The values are 8.5%, 9.3%, 7.9%, 9.2%, and 10.3%.

(i) What is the standard deviation for this sample ?

(ii) What is the standard error for this sample ?

2+2

6. (a) Explain the characteristics of Poisson distribution in detail with possible interpretations about the data under study.

4

(b) Suppose the number of motor-vehicle fatalities in a city during a week is Poisson distributed, with an average of 8 fatalities per week.

(i) What is the probability that 12 fatalities occur in a specific week ?

(ii) What is the probability that at least 12 fatalities occur during a specific week ?

(iii) How many motor-vehicle fatalities would have to occur during a given week to conclude there are an unusually high number of events in that week ?

10

7. (a) Define Probability Density Function, Probability Mass Function, Regression, Correlation Coefficient. 4

(b) The presence of bacteria in a urine sample (bacteriuria) is sometimes associated with symptoms of kidney disease in women. Suppose a determination of bacteriuria has been made over a large population of women at one point of time and 5% of those sampled are positive for bacteriuria.

(i) If a sample size of 5 is selected from this population, what is the probability that 1 or more women are positive for bacteriuria ?

(ii) Suppose 100 women from this population are sampled. What is the probability that 3 or more of them are positive for bacteriuria ? 4

(c) One interesting phenomenon of bacteriuria is that there is a "turnover"; that is, if bacteriuria is measured on the same woman at two different points in time, the results are not necessarily the same. Assume that 20% of all women who are bacteriuric at time 0 are again bacteriuric at time 1 (1 year later), whereas only 4.2% of women who were not bacteriuric at time 0 are bacteriuric at time 1. Let X be the random variable representing the number of bacteriuric

events over the two time periods for 1 woman and still assume that the probability that a woman will be positive for bacteriuria at any one exam is 5%.

(i) What is the probability distribution of X ?

(ii) What is the mean of X ?

(iii) What is the variance of X ?

6

8. (a) Explain the method of least square in regression analysis.

4

(b) Taking x as independent variable and y as dependent variable determine the line of regression.

10

x	12	21	28	25	32	42	43	39	55
y	14	22	12	28	35	37	32	44	49

9. Explain the following in detail :

(a) Type-1 error and Type-2 error in hypothesis testing

(b) Student's t-test

(c) Define the step-by-step procedure of ANOVA and explain its application.

4+4+6

10. (a) Suppose that total carbohydrate intake in 12- to 14-year-old boys is Normally distributed, with mean = 124 g/1000 cal and standard deviation = 20 g/1000 cal.
- (i) What percentage of boys in this age range have carbohydrate intake above 140 g/1000 cal ?
 - (ii) What percentage of boys in this age range have carbohydrate intake below 90 g/1000 cal ? 2×2
- (b) In a certain developing country, 30% of the children are undernourished. In a random sample of 25 children from this area, find the probability that the number of undernourished will be $4 \times 2 \frac{1}{2}$
- (i) Exactly 10
 - (ii) Less than 5
 - (iii) Five or more
 - (iv) Between 3 and 5
-