BU/EX/PG/SEM-3/304

2015

BOTANY

Paper : 304

BIOSTATISTICS AND BIOINFORMATICS

Full Marks : 80

Time : 3 hours

Part A

(Biostatistics)

The figures in the margin indicate full marks for the questions

1. Answer the following questions. 1 x 5=5

A. Founder of Biostatistics is-

i. Francis Galton ii. Karl Pearson

ii. R.A. Fisher iv. None of the above

B. Formula of mean deviation is-

i. $\sum fd/n$ ii. $\sum fd^2/(n-1)$

iii. $\sum dxdy/\sqrt{\sum dx^2.dy^2}$

iv. $\sum(O-E)^2/E$

C. Statistical method of breaking variance into its components is called-

i. Deviation ii. ANOVA

ii. Variable iv. Degree of freedom

D. Hypothesis that asserts that there is no difference between the sample and population is

P.T.O.

i. Null hypothesis ii. Alternate hypothesis

ii. Goodness of fit iv. Contingency table

E. One way F-test=

i. MSB/MSW ii. MSC/MSE

ii. MSC/MSR iv. MSR/MSE

2. Write any three of the following. 2X3=6

a. Distinguish the Skewness and Kartosis.

b. What is standard error? Mention its formula.

c. Describe the basic principles of experimental design?

d. What is research design? What are the different types of research design?

3. Answer from the following questions (any two) 2x2=4

a. Probability

b. Regression equation

c. Distinguish the population and sample.

4. Answer (any three) from the following questions. 5X3=15

a. What is goodness of fit? Define Chi-square test. According to Mendel monohybrid ration (3:1; tall: dwarf) in F_1 generation if it is carefully selfed among the 100 plants, then in F_2 generation tall plants are observed 85 while 15 are dwarf, then test the value of Chi-square. 1+2+3=5

b. What is correlation? Define the correlation coefficient. According to Karl Pearson's method calculate the co-efficient of correlation as a measure of closeness of association between the two variables as given in the following table. 1+2+3=5

Samples	X	Y
1	7	10
2	6	7
3	5	4
4	7	5
5	8	4
6	9	7
7	6	4
8	8	7

c. What is standard deviation? Calculate the standard deviation from the given population sample. 1+4=5

Class value (x)	Frequency (f)
48	8
50	32
52	75
54	52
56	28
58	5

d. Define regression. What is regression equation? Anil's weight is 45 Kg; Sunil's is 165 cm tall. Estimate the height of Anil from his weight and the weight of Sunil from his height (If the results of height and weight of the students are $\overline{Y}$ = 170 cm; X bar= 60 Kg; r = 0.6; σ_y = 6.5 cm σ_x = 5Kg). 1+4=5

4. Answer the following questions (Any one) 1x10=10

a. What is ANOVA? Arrange the steps of one way ANOVA. Test the significance of difference between the yields of the three varieties of wheat in four plots of given agricultural land by 5% Fisher's F-distribution table. 2+3+5=10

Variety	Plot yield			
	1	2	3	4
A	12	18	14	16
B	19	17	15	13
C	14	16	18	20

b. What is central tendency? Define the different types of central tendencies? Find all central values from the following frequency table of discrete series of pea seed germination. 2+3+5=10

No. of hours taken to germinate (X)	No. of seeds in frequency (f)
4	29
5	8
6	13
7	30
8	28
9	8
10	12
11	24

PART –B

(Bioinformatics)

1. Find out the correct answer of the following questions (any four) 1×4=4

A. You have a favourite gene, and you want to determine in what tissues it is expressed. Which one of the following resources is likely the most direct route to this information?

i) UniGene, ii) Entrez,

iii) PubMed, iv) PCR

B. The DBMS acts as an interface between what two components of an enterprise class database system?

i) Database application and the database

ii) Data and the database,

iii) The user and the database application

iv) Database application and SQL.

C. Orthologs are defined as

i) Homologous sequences in different species that share an ancestral gene

ii) Homologous sequences that share little amino acid identity but share great structural similarity

iii) Homologous sequences in the same species that arose through gene duplication

iv) Homologous sequences in the same species which have similar and often redundant functions.

D. How does the BLOSUM scoring matrix differ most notably from the PAM scoring matrix?

i) It is best used for aligning very closely related proteins.

ii) It is based on global multiple alignments from closely related proteins.

iii) It is based on local multiple alignments from distantly related proteins

iv) It combines local and global alignment information.

E. How can multiple sequence alignment programs improve performance?

i) By performing PSI-BLAST

ii) By incorporating data on secondary structure

iii) By incorporating data on 3D structures

iv) All of the above

F. The two main features of any phylogenetic tree are-

i) The clades and the nodes

ii) The topology and the branch lengths

iii) The clades and the root

iv) The alignment and the bootstrap

2. Match the followings in column X with appropriate items in column Y: (any four) 1×4=4

Column X		Column Y	
2.1	Microarray	A	Visualization tool
2.2	RASWIN	B	Protein modelling
2.3	KEGG	C	protein structure classification
2.4	MODELLER	D	Protein family
2.5	CATH	E	3D Structure visualization
2.6	Pfam	F	Metabolic Pathway
2.7	SPDBV	G	Gene expression

3. Define any five from the following. 2×5=10

a) Metabolomics b) Comparative genomics

c) Open Reading Frame d) Protein Data Bank

e) Ramachandran plot f) TrEMBL

4. Distinguish between (Any two) 2×5=10

a. Homology and Identity

b. Tertiary and Quaternary structure of protein

c. Pairwise sequence alignment and multiple sequence alignment

5. Answer any one from the following 1×12=12

a. What is SCOP database? What is the hierarchy of its classification? State the major Classes in this database and state their topology. Globin protein belongs to which of these classes? 2+4+6=12

b. Describe the importance of phylogenetic tree in Bioinformatics. What are the different methods of constructing phylogenetic trees? 6+6=12

—— × ——