Ph.D. IN BIOCHEMISTRY (PHDBC)

Term-End Examination

00335   **June, 2018**

RBC-002 : BIOSTATISTICS AND BIOINFORMATICS

Time : 3 hours   *Maximum Marks : 100*

Note : *This question paper consists of three sections, A, B and C. Answer **all** the sections. Simple calculator is allowed.*

SECTION A

*Answer **all** the questions.*

1. Mark *True* (T) or *False* (F) against each statement given below and give reasons :   *5×2=10*

    (a) A researcher wants to compare the proportions of high triglyceride levels in smokers and non-smokers of a city then he applies the t-test.

    (b) The value of correlation coefficient always lies between "0" to "1".

    (c) Median is the value corresponding to the maximum frequency.

    (d) If the data are calculated in qualitative scale, then we apply parametric test.

    (e) The probability of a selected TB patient among 100 TB patients is 0·01.

2. Match the following : $5 \times 1 = 5$

	Group A		Group B
(a)	NCBI	(i)	Utility tool
(b)	Linux	(ii)	RCSB
(c)	Adobe Photoshop	(iii)	E-resource
(d)	Protein structure	(iv)	2° structure
(e)	Chou-Fasman	(v)	Operating system

3. Define the following terms with suitable examples : $2 \times 2\frac{1}{2} = 5$

(a) Byte in the context of computer

(b) Node in the context of phylogenetic tree

SECTION B

Answer any five questions.   *5×6=30*

4. The following data shows triglyceride levels
    (mg/dl) of 20 hypertensive males :   *6*

    153, 157, 141, 163, 168, 177, 142, 158, 149, 160,
    162, 158, 145, 172, 140, 156, 152, 159, 161, 135.

    (i)   Construct the frequency distribution by
          taking categories as 130 – 140, 140 – 150,
          etc.

    (ii) Draw a histogram of frequency distribution
          obtained in part (i).

5. The following values represent incubation
    periods (in days) of a new disease "Severe Acute
    Respiratory Syndrome" (SARS) of 15 patients :   *6*

    9, 7, 10, 6, 12, 13, 12, 11, 12, 5, 8, 8, 9, 11, 7.

    (i)   Calculate mean and mode.

    (ii) Calculate standard deviation.

6. What is sampling ? Write the names of three
    widely used sampling schemes in Biochemistry ?
    Describe any one of them.   *6*

7. What is a biological database ? Write any five
    applications of biological databases.   *6*

8. What is the significance of sequence alignment ? Explain the role of substitution scores and gap penalties in the sequence alignment process. ' 6

9. A researcher wants to know the genetic relatedness of "Peptidase-A" among the following species : (i) Homo sapiens (ii) Sus scrofa (iii) Bos taurus (iv) Mus caroli. Explain the steps involved in establishing genetic relatedness using "clustal omega" as a research tool. 6

10. What is an operating system ? Explain with suitable examples how it is important in a computer. 6

SECTION C

Answer any five questions.   5×10=50

11. Calculate the expected frequencies for the following data presuming the two attributes and test whether the condition of the home and condition of the child are independent at 5% level of significance :   10

Condition of Child	Condition of Home	
	Clean	Dirty
Clean	70	50
Fairly Clean	80	20
Dirty	35	45

Given that $f_{5\%}(2) = 5.99$.

12. An investigator is interested to know the average level of knowledge of class 12 students about HIV in three different schools of a city. A test was conducted and the scores out of 10 are given below :   10

School I (S_1)	8	6	7	5	9		
School II (S_2)	6	4	6	5	6	7	
School III (S_3)	6	5	5	6	7	8	5

Set up a table of analysis of variance and find out whether there is significant difference between average scores of the schools about the knowledge of HIV. Given that $F(2, 15) = 3.68$ and $F(2, 17) = 3.59$.

13. A dietician wishes to see whether a person's cholesterol (mg/dl) level will reduce if the diet is supplemented with Vit-E. Eight subjects were pretested and then they were given Vit-E supplement for a period of 6 weeks. The results obtained after and before the experimental period were given in the following table : (Assume the variable is approximately normally distributed).

Subject	Before experiment	After experiment
1	210	190
2	235	180
3	208	210
4	190	188
5	172	173
6	244	220
7	195	195
8	200	204

(i) Are the two samples of data paired or independent ?

(ii) State the null and alternative hypothesis.

(iii) Which test is used for testing the null hypothesis in this case and why ?

(iv) Can it be concluded that the cholesterol level has been altered at 5% level of significance ?

10

14. What is phylogeny ? Distinguish between cladograms and phylograms. Write a brief note on steps involved in constructing a phylogenic tree ? 10

15. What is propensity value ? Explain how Chou-Fasman rules are helpful in protein 2° structure prediction ? 10

16. A researcher isolated a new protein with 142 amino acid residues. Describe the steps involved in developing a 3-D model for this protein with the help of a flow chart. 10

17. Identify and explain the tools useful in performing the following tasks in protein research : 10

    (a) Plotting "Ramachandran Plot"

    (b) Protein tertiary structure prediction

    (c) Visualization of protein – ligand interactions

———————