BIRLA INSTITUTE OF TECHNOLOGY AND SCIENCE, PILANI
Application of Computer and Statistics in Biology/ Pharmacy BIO/ PHAG510
FIRST SEMESTER 2017 – 2018
COMPREHENSIVE EXAM
**Weightage 35%   Date: 13.12.2017   DURATION: 3Hrs.   Total Marks 70**

NOTE:
- **Answer Part A and Part B in separate answer sheets.**
- **Irrelevant answer may attract penalty.**

PART – A (CLOSED BOOK) (Max. duration: 2 Hrs., Max. Marks 46)

Q1. a) Write short note on following topic   **[1X3=3]**
   i) Noise filtering in DOT plot **ii)** Cladogram **iii)** Bootstrap value in phylogenetic tree
   b) Discuss at least three application of multiple sequence alignment.   **[2]**

Q2. a) What are the major challenges to maintain a biological sequence database?   **[1]**
   b) What is the objective of sequence alignment? ii) What is the role of scoring scheme in sequence alignment?   **[2]**
   c) Compare pairwise and multiple sequence alignment.   **[2]**

Q3. A researcher generates following unrooted and rooted tree by maximum parsimony and UPGMA method with glutamate synthatase gene of Spirulina, Synechocys, Yeast, Arabidopsi and Rice.
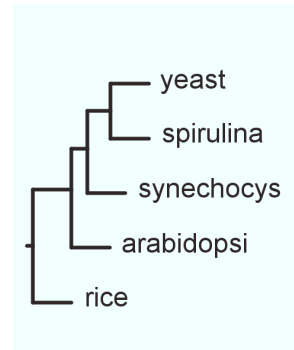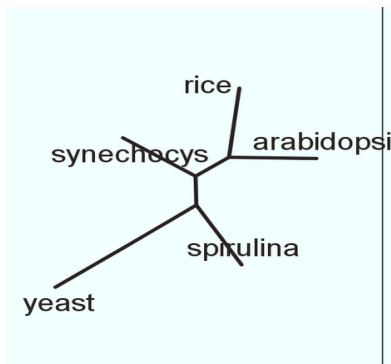
a) How many possible rooted and unrooted trees one can generate in above case? **[1]**
b) Draw two different (other than the above trees) rooted and unrooted trees with these five taxas.   **[1]**
c) Is it possible to generate the given rooted tree from the given unrooted tree? If so, mention possible root within the given unrooted tree.   **[1]**
d) Compare maximum parsimony and UPGMA method.   **[2]**

Q4. The mean age of 40 students is 16 years and the mean age of another group of 60 students is 20years. Find the mean age of 100 students together.   **[2]**

Q5. If the probability of horse 'A' winning the race is 1/5 and probability of Horse 'B' winning the race is 1/6, what would be the probability that one of the horses will win the race?   **[2]**

Q6. A genetics engineer was attempting to cross a tiger and cheetah. She predicted a phenotypic outcome of traits she was observing to follow a ratio of 4 stripes only: 3

spots only: 9 both stripes and spots. After the cross, she found 50 with stripes only, 41 with spots only and 85 with both. Did she got the predicted outcome? [Values for degree of freedom (df:1 = 3.841; df:2 = 5.991; df:3 = 7.851)] **[4]**

Q7. Write short notes on:
a) Normal Distribution curve   b) Stratified Random sampling   **[4]**

Q8. The following are the height (cm) and the weight (Kg) of 10 men. Calculate correlation coefficient between height and weight.   **[4]**

Height (cm)	162	168	174	176	180	180	182	184	186	186
Weight (in Kg)	65	65	84	63	75	76	82	65	80	81

Q9.a) Explain the difference between test of independence and test of homogeneity.   **[2]**
b) For the following, indicate whether a null hypothesis of homogeneity or a null hypothesis of independence is appropriate with reasons: **[3]**
i)   A researcher wishes to compare the status of three communities with respect to immunity against polio in preschool children. A sample of preschool children was drawn from each of the three community
ii) In a study of the relationship between smoking and respiratory illness, a random sample of adults were classified according to consumption of tobacco and extent of respiratory symptoms.
iii) A health research team believes that the incidence of depression is higher among people with hypoglycemia than among people who do not suffer from this condition.

Q10. For what two purposes, can a regression equation be used?   **[2]**

Q11. Following area under curves were observed after clinical trials of different formulations of same drug. Calculate slope and intercept.   **[4]**

Dosage	100	300	600	900	1200
AUC	1.07	5.82	15.85	25.18	33.12

Q12. For the following, indicate whether to use binomial distribution or Poisson distribution with reason (**No need to calculate)**   **[4]**
a)   A typist makes on average 2 mistakes per page. What is the probability of a particular page having no errors on it?
b)   Components are packed in boxes of 20. The probability of a component being defective is 0.1. What is the probability of a box containing 2 defective components?
c)   ICs are packaged in boxes of 10. The probability of an ic being faulty is 2%. What is the probability of a box containing 2 faulty ics?
d)   A box contains a large number of washers; there are twice as many steel washers as brass ones. Four washers are selected at random from the box. What is the probability that 0, 1, 2, 3, 4 are brass?

PART – B (OPEN BOOK) (Max. duration: 1 Hr., Max. Marks 24)

Q1.

a) While generating scoring scheme from same group of data, two researchers produce two different values (which is shown below). Which scoring scheme is probably correct and why? [A= Ala; W= Trp; E=Glu; D=Asp]   **[2]**

I

	A	W	E	D
A	4	-3	-1	-2
W		11	-3	2
E			5	2
D				6

II

	A	W	E	D
A	4	-3	-4	2
W		2	-3	2
E			2	-2
D				4

b) In the two alignments below, one is an alignment of two DNA sequences with an identity of 36%. The other alignment is of two amino acid sequences with an identity of 28%. Which of the two alignments represents greater biological similarity between sequences? Why?   **[2]**

DNA Alignment
```
Seq1    AGGCTGCCAAAACGCACTGTTTAAT
        : ::  :   :  ::        ::
Seq2    ACGCA-CGTTATGGCTAAAGCCTAT
```

Amino Acid Alignment

```
Seq3    PVALGLKEKNLYLSCVLKDKGQDIT
        :  :::           :    ::
Seq4    PADLGLMNNYNMIQLRCADELHYIT
```

c) Construct UPGMA tree for the given distance table of five taxa.   **[4]**

	A	B	C	D	E
A		5	4	7	6
B			7	10	9
C				7	6
D					5

Q2. Three chemicals A, B and C show the cleaning efficiency as given below. Find whether the differences among them are significant at 5% significance level. (Critical Value: 3.89)   **[5]**

Chemical A	80	77	76	81	71
Chemical B	70	58	72	66	74
Chemical C	77	80	82	85	76

Q3.Systolic blood pressure of 9 normal individuals was taken. Then a known hypotensive drug was given, and blood pressure again recorded. Given below is the blood pressure of nine healthy volunteers before and after injection of hypotensive drug. Did the hypotensive drug lowers the systolic blood pressure? **[4]**

BP Before	122	121	120	115	126	130	120	125	128
BP After	120	118	115	110	122	130	116	124	125

Q4.Comment on the given statement: Only a very small (5 or 10) percentage of measurements can be more than two standard deviations from the mean. Justify **[2]**

Q5.If three persons, on an average, come to ABC company for job interview, then find the probability that less than three people have come for interview on a given day. **[3]**

Q6.Of the following dotplots, which represents the set of data that has the greatest standard deviation. Justify your answer. **[2]**