

U.G. 3rd Semester Examination - 2025**ZOOLOGY****[Skill Enhancement Course (SEC)]****Course Code : ZOO-SEC-T-3****(Statistical and Computational Biology)****[NEP-2020]****Full Marks : 35****Time : 1½ Hours***The figures in the right-hand margin indicate marks.**Candidates are required to give their answers in their own words as far as practicable.*

1. Answer any **five** of the following: 1×5=5
- a) What does $P < 0.05$ signify?
 - b) Write down one common file format in which nucleotide sequence data is retrieved.
 - c) Define central tendency. How can it be measured?
 - d) What is meant by frequency?
 - e) How does a measurement variable differ from a categorical variable?
 - f) Define population in statistics.
 - g) Name one database each for storage and retrieval of protein and nucleic acid sequence data.
 - h) What is a null hypothesis (H_0)?

[Turn Over]

2. Answer any two of the following: $5 \times 2 = 10$

- a) i) Define standard deviation.
 ii) The following data represent the body weight (in kg) of five fish:
 20, 22, 24, 26, 28

Calculate the standard deviation. $2+3$

- b) Calculate arithmetic mean and standard deviation from the given values of rainfall over five consecutive days:

Date	21st. Nov	22nd. Nov	23rd. Nov	24th. Nov	25th. Nov
Rainfall (mm)	27	60	40	30	32

$2\frac{1}{2} + 2\frac{1}{2}$

- c) i) What are sequence submission tools? Give one example each for submitting simple sequences and protein sequences determined by Edman degradation.
 ii) Differentiate between global, local and multiple-sequence alignment. $2+3$
 d) i) State two aims of bioinformatics.
 ii) Match the following databases with their contents:

Database	Content
GenBank	?
PDB	?
UniProt	?

$2+3$

3. Answer any two of the following: $10 \times 2 = 20$

- a) In *Drosophila*, grey body colour (G) is dominant over black (g). A heterozygous grey fly was crossed with another heterozygous grey fly. Out of 200 offspring, 152 were grey and 48 were black. Test the goodness of fit to the expected Mendelian ratio using the chi-square test ($df = 1$, $\chi^2_{0.05} = 3.84$). 10

- b) The hemoglobin levels (g/dL) of patients receiving two different treatments were recorded to compare their effectiveness.

Treatment A: 13.2, 13.5, 13.1, 13.6, 13.4

Treatment B: 12.6, 12.8, 12.5, 12.7, 12.9

Using a t-test, whether there is a significant difference between the two treatments at 5% level of significance ($df = 8$, $t_{0.05} = 2.31$). 10

- c) i) The table below presents the results of a study showing individual marks scored by 7 candidates and their hours of study before the exam. Calculate r and state your conclusions as to whether hours of study affect exam scores.

Student #	1	2	3	4	5	6	7
Study (hr)	3	15	6	8	4	2	10
Score	75	95	65	70	85	80	65

- ii) Find the regression coefficient and regression equation of X on Y from the following data:

X	1	2	3	4	5	6	7
Y	9	8	10	12	11	13	14

5+5

- d) Write short notes on (any four): $2\frac{1}{2} \times 4 = 10$

- i) Types and uses of the various types of BLAST;
 - ii) Entrez databases and retrieval systems;
 - iii) Applications of sequence alignment;
 - iv) Major types of protein databases;
 - v) Pearson's Correlation Coefficient.
-