

SBL701: Biometry Major Examination**Max. Marks: 40****Max. Time: 2 hrs**

(but since it is take-home, you may take more time!)

Submission deadline: 1900 hrs on 19th November, 2022 (includes Major Timings of 2 hrs)*Brevity is the soul of communication. Extraneous information will be penalized!***Submission details (deviations are not acceptable):**

- I. Only soft copy submissions on the email id amittal@bioschool.iitd.ac.in will be accepted (i.e. no hard copy). No late submission will be accepted (regardless of server problems etc.). To be safe, submit well in advance.
- II. Submissions should be compiled in form of one Adobe pdf file. Multiple files or files in any other software are not acceptable. One way to do this is "print" all your files as "adobe pdf" rather than sending them to a printer.
- III. Use Microsoft Equation 3.0 or Equation Editor or Math Type in MS word if you need to type any equations.
- IV. Solutions should contain each and every step followed for calculations, all raw data (properly tabulated), tables, graphs, programming codes (MATLAB), code output (in form of numbers in command window or figures).
- V. Copying is not acceptable. While you are free to consult anyone, the solutions, interpretations, answers should all be in your own language. Any material that is found to be copied from another source will automatically lead to a "zero" out of 30. Any material found to be identical in different submissions will also lead to "zero" out of 40 for all the submissions appearing identical.

1. It has been observed that amino acids occur in primary sequences of proteins in relatively fixed percentages. There could be several evolutionary reasons for this occurrence. One possibility is the number of codons corresponding to each amino acid in the genetic code. For example, it may be hypothesized that more the number of codons available for an amino acid, higher the number of ways to incorporate that amino acid in primary sequences and hence higher relative occurrence of the amino acid in primary sequences. Data pertaining to this is provided below:

Amino Acid	<u>Ala</u> (A)	Arg (R)	Asn (N)	Asp (D)	<u>Cys</u> (C)	Gln (Q)	Glu (E)	<u>Gly</u> (G)	<u>His</u> (H)	<u>Ile</u> (I)	<u>Leu</u> (L)	Lys (K)	<u>Met</u> (M)	<u>Phe</u> (F)	Pro (P)	Ser (S)	Thr (T)	<u>Trp</u> (W)	<u>Tyr</u> (Y)	<u>Val</u> (V)
Number of Codons	4	6	2	2	2	2	2	4	2	3	6	2	1	2	4	6	4	1	2	4
Avg. % Occurrence	7.8	5	4.3	5.8	1.8	3.8	7	7.2	2.3	5.8	9	6.3	2.2	3.9	4.4	6	5.5	1.3	3.4	7.1
Std % Occurrence	3.4	2.3	2.2	2	1.5	2	2.7	2.8	1.4	2.4	2.9	2.8	1.3	1.8	2	2.5	2.4	1	1.7	2.4

- (a) How will you test the hypothesis that number of codons available for an amino acid may be responsible for its relative occurrence in primary sequences?
- (b) Can you predict the average percentage occurrence of an amino acid in primary sequences based on the number of codons available for it in the genetic code?
- (c) Answer (a) and (b) for the underlined amino acids in bold (assume they represent one independent group).
- (d) What can you infer from comparing results of (a) and (b) to those obtained in (c)?

(2+3+3+2 = 10)

2. Consider the following data with X as the independent variable and Y as the dependent variable:

X	93	85	75.5	62.5	53	43	29.5	12	0
Y	3.72	4.2	4.68	5.83	5.9	6.08	6.67	8.14	8.98

Answer the following –

- (a) Obtain the residual plot after performing linear regression on the data. What can you infer from the plot?
- (b) To test the significance of the regression, what would be degrees of freedom for sum of squares of residuals and for sum of squares of regression?

(3+2 = 5)

3. A pharmaceutical company claims that it has developed a diagnostic test for colon cancer. It presents data that out of 1500 persons having colon cancer, the test came up positive in 1275 cases. Further, the company also shows that in another 1500 persons not having colon cancer, the test came up positive in only 45 cases. Based on these results the company asks the ministry of health to make this test mandatory for all admissions in government hospitals. This obviously means very good business for the company. You are approached by the ministry of health to advise them on the issue. Having become experts in statistical analyses because of your sincere efforts in the biometry course, you carry out your own research and present your findings that probability of having colon cancer if the test comes up positive is actually only ~ 0.22. While your results

show lower than expected reliability of the test, they also encourage the ministry of health to provide substantial new research funding to the company for improving the test. This is because your results also show that there might be a high number of colon cancer patients on our country of ~1.6 billion population. How many people of our population are expected to be having colon cancer?

(5)

4. A Ph. D. candidate, working at IIT Delhi, is asked to utilize this technique for understanding an enzyme's structure at room temperature. In January 2021, she measures the time (in seconds) for 70 wt enzyme molecules to convert one molecule of substrate into its product. She also measures the time for 7 enzyme molecules with a point mutation at position 123 of the primary sequence of the enzyme. Unfortunately, while recording the data in her notebook, she forgets to separate the data for experiments on wt and mutants. Thus, in her lab book, column Mut-1 (see on next page) is recorded as the time (in seconds) measured independently for 77 single enzyme molecules – seven (randomly mixed in the 77) of which are point mutants at position 123 in the primary sequence. A new Ph. D. student is asked to learn the technique by repeating the experiment. In June 2015, he measures the time (in seconds) for 53 wt enzyme molecules to convert one molecule of substrate into product. He also measures the time for 6 enzyme molecules with a point mutation at position 127 of the primary sequence of the enzyme. Unfortunately, working with the assistance of the “senior” Ph. D. student, he commits the same mistake in recording the data that she had committed earlier. Thus, in his lab book, column Mut-2 (see on next page) is recorded as the time (in seconds) measured independently for 59 single enzyme molecules – six (randomly mixed in the 59) of which are point mutants at position 127 in the primary sequence.
- (i) Analyze the data to test whether the mutation in position 123 affects the enzyme activity. Does it?
 - (ii) Analyze the data to test whether the mutation in position 127 affects the enzyme activity. Does it?
 - (iii) In a fresh experiment, what would be the expected time taken by a randomly chosen single wt enzyme molecule for converting one molecule of substrate into product?
 - (iv) What additional experiments should be done to rectify the mistakes of the two students without “throwing away” their results?

(5+3+8+4 = 20)

Data for above question (i.e. Q. 4) is on next page

Mut-1	Mut-2
64.0	213.4
41.3	163.7
55.2	284.2
43.9	394.2
196.6	314.0
28.8	216.4
29.6	170.9
32.6	111.3
47.3	451.3
121.3	205.0
141.0	463.0
51.2	61.2
42.4	197.5
202.4	366.2
145.5	119.4
172.6	349.1
88.2	252.2
116.0	305.0
98.9	592.2
118.8	238.6
117.5	165.5
166.1	146.7
125.7	158.7
130.9	425.6
80.2	613.7
103.6	254.0
118.5	523.2
71.5	431.5
77.5	662.1
170.0	396.8
77.1	703.7
87.9	262.9
82.6	201.5
70.8	450.1
206.4	117.3
121.7	374.8
114.4	315.2
69.7	508.1
158.1	752.5
234.7	412.2
122.0	399.6
116.7	480.8
136.1	1057.3
76.8	261.3
162.0	336.6
70.1	1277.1
156.4	753.0
87.5	299.1
252.3	316.3
75.7	493.6
12.3	363.5
119.3	288.9
159.7	672.1
84.4	403.6
33.2	343.0
27.8	172.3
113.9	815.7
58.8	488.9
97.7	686.5
54.2	
116.5	
44.5	
156.8	
93.3	
38.5	
215.0	
62.5	
179.4	
70.5	
132.2	
182.8	
238.4	
194.2	
110.0	
251.4	
73.2	
177.2	