Department of Biotechnology

Biostatistics BT2023

Fall semester Aug-Dec 2022

Email: hjoshi@bt.iith.ac.in

Mid Term Exam

18 October 2022

Maximum Marks: 60

Maximum marks for each question is 6

Maximum Time: 1 hour 30 mintutes

1. (a) What is biostatistics, in the context of today's world describe it's importance. [Maximum 80 words]
   (b) What do you understand by descriptive and inferential statistics.

2. (a) When is *mean* not a good measure of the central tendency and *mode* or *median* is preferred to describe the data set.
   (b)What are the measures of dispersion in a data set.
   (c)What is the standard error of a distribution of the data having variance σ^2 and n number of data points.

3. A data set of five numbers has **mode, mean, median 25, 30 and 21** respectively. Write down any possible data set which will satisfy the given values of mode, mean, median. Also compute **standard deviation, skewness and kurtosis** coefficients (any coefficients you like) for the data set.

4. Sketch (hand drawn plot showing the key features of the function) the the following function with respect to x,
   (a) $y = \frac{1}{3\sqrt{2\pi}} e^{-\frac{(x-3)^2}{18}}$.
   (b) $y = \frac{1}{2\sqrt{2\pi}} e^{-\frac{(x-4)^2}{8}}$.
   (c) What is the area under these curves in the range x=0 to $x = \infty$ *i.e.* $\int_0^\infty y\, dx$

5. (a)What is *Z-score* in the context of a statical measurement.

   (b) Given a normal distribution of a random variable, what is the probability of finding a value having z score >2.

6. In the context of a probability distribution function,

   (a) what does *Full-Width-Half-Maxima* represents.

   (b) what is *68-95-99* rule in normal distribution.

   (c) what is skewness and kurtosis for a normal distribution.

7. (a) What is hypothesis testing.

   (b) Please perform "Student's *t-test* " to find out whether the null hypothesis is valid or not for the data set given below. The data contains body mass index of two group of people (measured in kg/m^2)

23	25	26	24	23	27	27	24	21	28
26	26	23	24	25	22	19	25	26	25

   Given that the *p-value* for 8 degree of freedom and 0.025 probability is 2.75.

8. (a)Describe Chebyshev's inequality for a hypothetical data set having mean μ and variance σ^2

   (b) For a given random variable x, having a mean and variance 4 and 100 respectively, use the Chebyshev's inequality to find out the appropriate bound for the random variable $P[|x-4| \geq 10]$

9. Find out the correlation coefficient,r, for the following hypothetical sample data set of blood sugar *vs* the annual income,

Blood Pressure (mm/l)	4.5	3.2	5.7	5.9	6.3	3.6	10.7	3.5	7.9
Annual Income (in lacks)	6	1	3	10	18	19	25	2	25

10. (a) What is a *scatter plot*, illustrate by plotting the data set (blood sugar *vs* the annual income) provided in previous question, the plot does't have to be accurate.

    (b) Schematically draw a *violin* plot and describe various features.