Course Title: Basic Statistics and Probability; Course Code: STA 215
Summer 2024 Assignment 4; Batch: 57(A+B+C+D)

Box and Whisker Plot

1. Cotinine is a chemical that is made by the body of nicotine which is found in cigarette smoke. A doctor tested the blood of 12 patients, who claimed to smoke a packet of cigarettes a day, for continue. The results, in appropriate units, are shown below.

Patient	A	B	C	D	E	F	G	H	I	J	K	L
Cotinine level, x	160	390	169	175	125	420	171	250	210	258	186	243

[You may use $\sum x^2 = 724961$]

a) Find the mean and standard deviation of the level of cotinine in a patient's blood. [4]

b) Find the median, upper and lower quartiles of these data. [3]

A doctor suspects that some of his patients have been smoking more than a packet of cigarettes per day. He decides to use $Q_3 + 1.5(Q_3 - Q_1)$ to determine if any of the cotinine results are far enough away from the upper quartile to be outliers.

c) Identify which patient(s) may have been smoking more than a packet of cigarettes a day. Show your working clearly. [4]

Research suggests that cotinine levels in the blood form a skewed distribution. One measure of skewness is found using $\frac{(Q1-2Q2+Q3)}{(Q3-Q1)}$.

d) Evaluate this measure and describe the skewness of these data. [3]

2. Rifat and Mahi play together in a basketball team. The list below shows the number of points that Rifat scored in each of 30 games.

39	19	28	30	18	21	23	15	34	24
29	17	43	12	24	25	41	19	26	40
45	23	21	32	37	24	18	15	24	36

a) Find the median and quartiles for these data. [3]

b) Represent these data with a boxplot [3]

Mahi played in the same 30 games and his lowest and highest points total in a game were 19 and 41 respectively. The quartiles for Mahi were 27,31 and 35 respectively.

c) Using the same scale draw a boxplot for Mahi's points total. [2]

d) Compare and contrast the number of points scored per game by Rifat and Mahi. [3]

Regression Analysis

3. Martin is investigating the relationship between a person's daily caffeine consumption, c milligrams, and the amount of sleep they get, h hours, per night. He collected this information from 20 people and the results are summarized below.

$\Sigma c = 3660$; $\Sigma h = 126$; $\Sigma c^2 = 973228$; $S_{hh} = 43.7349$; $\Sigma ch = 20023.4$; $S_{cc} = 303448$; $S_{ch} = -3034.6$.

a) Calculate the product moment correlation coefficient between c and h. (2)

b) Give a reason whether or not the answer in part (a) supports a linear relationship between c and h. (1)

c) The amount of sleep per night is the response variable. Explain what you understand by the term 'response variable'. (1)

d) Martin says that for each additional 100 mg of caffeine consumed, the expected number of hours of sleep decreases by 1. Determine, by calculation, whether or not the data support this statement. (3)

e) Use the data to calculate an estimate for the expected number of hours of sleep per night when no caffeine is consumed. (3)

4. Stuart is investigating the relationship between Gross Domestic Product (GDP) and the size of the population for a particular country. He takes a random sample of 9 years and records the size of the population, t millions, and the GDP, g billion dollars for each of these years.

The data are summarized as $n=9$; $\sum t = 7.87$; $\sum g = 144.84$; $\sum g^2 = 3624.41$; $S_{tt}=1.29$; $S_{tg}= 40.25$.

a) Stuart calculates the product moment correlation coefficient for these data and obtains 0.985. Give an interpretation of the product moment correlation coefficient. (1)

b) Find the equation of the least squares regression line of g on t in the form g = a + bt. (4)

c) Give an interpretation of the value of b in your regression line. (1)

d) Use the regression line from part (c) to estimate the GDP, in billions of dollars, for a population of 7,000,000. (2)

e) Comment on the reliability of your answer in part (d). Give a reason, in context, for your answer. (2)

Correlation Analysis

5. The chairman of a marketing department at a large private university undertakes a study to relate staring salary (y) in thousands after graduation for marketing majors to grade point average (GPA) in major courses. To do this, records of 10 recent marketing graduates are randomly selected. The GPA (x) and the corresponding starting salary were as follows:

GPA (x)	3.26	2.60	3.35	2.86	3.82	2.21	3.47	3.28	2.54	3.25
Salary (y) in thousands	33.8	29.8	33.5	30.4	36.4	27.6	35.3	35.0	26.5	33.8

a) Find the equation of the least squares regression line of y on x. (6)
b) Give an interpretation of the value of b in your regression line. (1)
c) Find the point prediction of starting salary corresponding to the GPA 3.75.(1)
d) Comment on the reliability of your answer in part (b). Give a reason, in context, for your answer. (2)

6. In a training scheme for young people, the average time taken for each group to reach a certain level of proficiency was measured. The data are shown below in the table:

Age x (years)	16	17	18	19	20	21	22	23	24	25
Average time y (hours)	12	11	10	9	11	8	9	7	6	8

a) Find S_{xx}, S_{yy} and S_{xy}. (6)
b) Use your answers in part (a) to calculate the Pearson's Correlation Coefficient (r). (1)
c) Describe and interpret the relationship between average time and age. (2)
d) Give a reason whether or not the answer in part (b) supports a linear relationship between x and y. (1)

Binomial Distribution

6. a) State the important properties of Binomial distribution. Give a practical example where binomial distribution would be appropriate to describe the situation.
b) In a community, the probability that a newly born child will be a boy is 2/5. Among the 4 newly born children in that community, what is the probability that:
i) all four boys.
ii) at least two boys.
iii) no boys.
iv) at most two boys.

7. The probability that a patient recovers from a rare blood disease is 0.4. If 10 people are known to have contacted this disease, what is the probability that:

i) Exactly 3 survive
ii) At most two survive
iii) At least three survive
iv) None survive.

Normal Distribution

8. a) State the important properties of Normal distribution. Name a random variable, which has a normal distribution.
b) The random variable $X \sim N(\mu, \sigma^2)$. Given that $P(X<17) = 0.8159$ and $P(X<25) = 0.9970$, find the value of μ and the value of σ.
c) The random variable $X \sim N(50,42)$. Find
i) $P(X<53)$ ii) $P(X>58)$

9. A psychologist gives a student two different tests. The first test has a mean of 80 and a standard deviation of 10 and the student scored 85.
a) Find the probability of scoring 85 or more on the first test.
The second test has a mean of 100 and a standard deviation of 15. The student scored 105 on the second test.
b) Find the probability of a score of 105 or more on the second test.
c) State, giving a reason, which of the students two test scores was better.

10. The weight of steel sheets produced by a factory are known to be normally distributed with mean 32.5kg and standard deviation 2.2kg.
a) Find the percentage of sheets that weigh less than 30kg.
Bob requires sheets that weigh between 31.6kg and 34.8kg.
b) Find the percentage of sheets produced that satisfy Bob’s requirements.