STQD6214 Assignment 1

Answer all questions. Submit your answer via UKMFolio by 11.59pm on 9th December 2024 or 15th December 2024 for the executive program. You are allowed to discuss with other students to answer the questions. However, do not copy their answer directly and make sure you understand your own answer.

---

1. A survey was conducted on the favorite flavor of a coffee drink a person prefers. The responses were V = Vanilla, C = Caramel, M = Mocha, H = Hazelnut, and P = Plain.

V	C	P	P	M	M	P	P	M	C
M	M	V	M	M	M	V	M	M	M
P	V	C	M	V	M	C	P	M	P
M	M	M	P	M	M	C	V	M	C
C	P	M	P	M	H	H	P	H	P

   a) Construct a frequency distribution table. Include the relative frequency and the percentage in the table.

   b) What is the mode for this data? Explain your answer.

   c) How many percentages of the respondents like Caramel or Hazelnut flavor?

   d) Draw a bar graph for the relative frequency distribution.

[10 Marks]

2. The data represents the number of minutes a sample of 25 people spent on exercise each week.

108	139	120	123	120	132	123	131	131
157	150	124	111	106	135	119	116	117
	127	128	139	119	118	114	127	

   a) Calculate the mean, median, and variance for the given data.

   b) Find the three quartiles of the data set. Then, draw a box-and-whisker plot by hand. From the box-and-whisker plot, is there any outlier in the data?

   c) Is the data skewed? If yes, in which direction does the data skewed to? Explain your answer.

[11 marks]

3. Six hundred adults in a city were asked whether they watch for calories and fat content when they buy groceries. The following table gives the two-way classification of their responses, where yes means that an adult watches for calories and fat content and no means he/she does not watch.

	Yes	No	No Opinion
Men	74	168	58
Women	106	124	70

Suppose an adult in the city is selected at random.

a) Find the probability that the selected adult is a man and has no opinion.

b) Find the probability that the selected adult is a woman, given that she watches calories and fat content.

c) Find the probability that the selected adult is a man, or watches for calories and fat content.

d) Are the events "Men" and "No" independent? Explain your answer.

e) Are the events "Yes" and "No Opinion" mutually exclusive? Explain your answer.

[10 marks]

4. A drawer contains 11 identical red socks and 8 identical black socks. Suppose that you choose 2 socks at random in the dark.

a) What is the probability that you get a pair of red socks?

b) What is the probability that you get a pair of black socks?

c) What is the probability that you get 2 unmatched socks?

[6 marks]

5. Let Y be a discrete random variable with probability mass function

$$P(Y = y) = \frac{10 - y}{2c}$$

where c is an integer and Y can take the values 0, 1, 2, 3, or 4.

a) Show that $c = 20$.

b) Construct a probability distribution table of Y.

c) Find the probability that Y is less than 2.

d) Find the probability that Y is at least 3.

e) Calculate the mean and variance of the distribution.

[10 marks]

6. An office supply company conducted a survey before marketing a new paper shredder designed for home use. In the survey, 90% of the people who used the shredder were satisfied with it. Because of this high acceptance rate, the company decided to market the new shredder. Assume that 90% of all people who will use the new shredder will be satisfied. On a particular day, 10 customers bought this shredder.

a) State a suitable probability distribution to model the number of satisfied customers on that day. Hence, find the mean and variance of the distribution.

b) Find the probability that

i) none of the customers are satisfied with the shredder.

ii) at least 8 customers are satisfied with the shredder.

iii) at most 2 customers are satisfied with the shredder.

[10 marks]

7. In the production factory of the new shredder, the company found that an average of 1 accident occurs every four months. Given this information, the company wishes to model the number of accidents occurring in the factory for a year using a probability distribution.

a) State a suitable probability distribution to model the number of accidents occurring in the factory within a year. Hence, find the mean and variance of the distribution.

b) Find the probability that

i) no accident occurs within a year.

ii) less than 2 accidents occur within a year.

iii) at least 4 accidents occur within a year.

[8 marks]

8. The total lifetime (in years) of five-year-old dogs of a certain breed is a random variable whose distribution function is given by

$$F(x) = \begin{cases} 0, & \text{for } x \leq 5 \\ 1 - \dfrac{25}{x^2}, & \text{for } x > 5 \end{cases}$$

a) Find the probabilities that such a five-year-old dog will live

i) beyond 10 years.

ii) less than eight years.

iii) anywhere from 12 to 15 years.

b) Find the probability density function for the random variable.

c) Calculate the mean and variance of the distribution.

[12 marks]

9. A study has found that the systolic blood pressure of an adult can be approximated by a normal distribution with mean of 120 mmHg and standard deviation of 4 mmHg. Let X be the blood pressure of a randomly selected adult and assume X is a normally distributed random variable with the given mean and standard deviation.

a) Find the probability $P(112 \leq X < 125)$.

b) To qualify for a medical study, an applicant must have a systolic pressure within the 80% of the middle range. Determine the lower and upper limits of blood pressure a person must have to qualify for the study.

c) In another study, 5 adults are randomly selected, and the mean of their blood pressure is recorded.

i) What is the distribution of $\bar{X}$?, the mean blood pressure of the 5 randomly selected adults? State the mean and variance for $\bar{X}$?.

ii) If X is not normally distributed, can the Central Limit Theorem be used to approximate the distribution of $\bar{X}$? Why or why not?

iii) Find the probability $P(116.5 < \bar{X} < 120)$.

[12 marks]

10. A student from University A believes that University A is harder than University B and typically, students in University A receive lower GPAs than students in University B. The student collected a sample of data given below.

GPAs of students from University A:

1.33	1.52	2.02	2.05	2.14	2.29	2.36	2.42	2.48	2.62	2.71	2.81	2.81
2.82	2.96	2.99	3.11	3.12	3.12	3.18	3.21	3.25	3.37	3.58	3.70	

GPAs of students from University B:

1.57	1.97	2.00	2.06	2.07	2.11	2.48	2.51	2.64	2.75	2.77	2.82	3.04
3.08	3.11	3.15	3.25	3.27	3.27	3.36	3.45	3.45	3.49	3.74	3.80	

a) Based on the given situation, suggest a suitable hypothesis test that the student can perform. Then, write down a suitable null and alternative hypothesis that can be used for the test.

b) Using the data, conduct an exploratory data analysis (EDA). Calculate useful numerical measures, produce relevant plots or figures, and explain what you can observe.

c) When a hypothesis testing is conducted to the data, it is found that the *p*-value is 0.21. (You do not need to find this *p*-value.) How would you conclude the hypothesis testing using this *p*-value and the hypotheses given in (a).

Note: This question is more open-ended. You can get marks as long as your answers are valid. For part (a) of this question, you do not need to conduct the hypothesis test. For part (b) of this question, you are allowed to use computers to perform the analysis, and you are not required to show the steps to calculate the measures and plot the graph. I am more interested in measures and graphs you choose to include, as well as the discussions.

[11 marks]

[Total: 100 marks]

'GOOD LUCK'