BINDURA UNIVERSITY OF SCIENCE EDUCATION

NOV 2024

DM006: INFERENTIAL STATISTCS

Time: 3 hours

Candidates may attempts ALL questions in Section A and at most TWO questions in Section B. Each question should start on fresh page.

SECTION A (40 marks)

A1. (a) Define random sampling and give two examples. [3]

(b) Define non-random sampling and give two examples. [3]

A2. $X \sim N(100, \sigma^2)$ and $P(X < 106) = 0.8849$.

Find the value of the standard deviation. [6]

A3. Draw a sketch diagram for each of the following, indicating the position of the mean, mode and median;

(a) normal distribution [3]

(b) positively skewed-distribution [3]

(c) negatively skewed distribution [3]

A4. The birth weight of babies is normally distributed with mean 3500g and standard deviation 500g.

What is the probability that a baby is born that weighs less than 3100g? [10]

A5. . (a) Explain three factors that affect the power of a test. [6]

(b) Explain three factors that affect the interval width. [3]

SECTION B (60 Marks)

Candidates may attempt TWO questions being careful to number them B6 to B8.

B6. (a) The weight of a large loaf of bread is a normal variable with mean 420 g and standard deviation is 30 g. The weight of a small loaf of bread is a normal variable with mean 220 g and standard deviation is 10 g.

(i) Find the probability that 5 large loaves weigh more than 10 small loaves. [6]

(ii) Find the probability that the total weight of 5 large loaves and 10 small loaves lies between 4.25 kg and 4.4 kg. [4]

(b) It is known that 3 % of frozen pies delivered in a canteen are broken.

What is the probability that, on a morning when 500 pies are delivered,

5 % or more are broken? [8]

(c) Suppose that we have a 6-sided die. We assume that the die is unbiased (upon rolling the die, each outcome is equally likely). An experiment is conducted in which

the die is rolled 240 times. The outcomes are in the table below. At a significance level of $\alpha = 0.05$, is there enough evidence to support the hypothesis that the die is unbiased?

Outcome	1	2	3	4	5	6
Frequency	34	44	30	46	51	35

[12]

B7. (a) When a production machine is properly calibrated, it requires an average of 25 seconds per unit produced, with a standard deviation of 3 seconds. For a simple random sample of $n = 36$ units, the sample mean is found to be 26.2 seconds per unit. When the machine is properly calibrated, what is the probability that the mean for a simple random sample of this size will be at least 26.2 seconds? [5]

(b) For many years, the mean gas mileage on a long trip was known to be 26.5 miles per gallon for a certain type of automobile. Each automobile in a simple random sample of those with the new design is driven 100 miles and the following gas mileages are recorded:

26.2 26.5 27.2 27.3 27.1 27.6 25.9 27.5 26.8.

(i) When a newly designed engine is incorporated into the automobile, is there is any evidence that the mean gas mileage with the new design is different from 26.5 miles per gallon at 0.05 significance level. [10]

(ii) Why would a box plot be an appropriate graphical display for the data used in this hypothesis test? [3]

(iii) Construct the box plot, and comment on whether there is any reason to think the t statistic is not appropriate. [3]

(iv) Considering the results of the hypothesis test, decide which of the Type I or Type II errors is possible, and describe this error. [3]

(v) Decide whether H_0 would have been rejected or would not have been rejected with each of the following significance levels: $\alpha = 0.01$ and $\alpha = 0.10$. [6]

B8. An old film is treated with a chemical in order to improve the contrast. Preliminary test on 9 samples drawn from a segment of the film produced the following results.

Sample	A	B	C	D	E	F	G	H	I
X	1,0	1,5	2,0	2,5	3,0	3,5	4,0	4,5	5,0
Y	49	60	66	62	72	64	89	90	96

The quantity x is a measure of the amount of chemical applied, and y is the contrast index, which takes values between 0 (no contrast) and 100 (maximum contrast).

(a) Plot a scatter plot diagram to illustrate the data. [5]

(b) It was subsequently discovered that on of the samples of the film was damaged and produced an incorrect result. State which sample you think was. [2]

In all subsequent calculations this incorrect sample is ignored. The remaining data can be summarized as follows:

$$\sum x = 23{,}5 \quad \sum x^2 = 83.75 \quad \sum y = 584 \quad \sum y^2 = 44\,622 \quad \sum xy = 1883 \quad n = 8$$

(c) Calculate the product moment correlation coefficient. [8]

(d) State, with a reason, whether it is sensible to conclude from your answer to part (c) that x and y are linearly related. [2]

(e) The line of regression of y on x has equation $y = a + bx$. Calculate the value of a and b, each correct to three significant figures. [9]

(f) Use your line of regression to estimate what the contrast index corresponding to the damaged piece of film would have been if the piece had been undamaged. [2]

(g) State, with a reason; whether it would be sensible to use your regression equation to estimate the contrast index when the quantity of chemical applied to the film is zero. [2]

END OF THE PAPER