in association with

Curtin College | Curtin University

Test 2 – TRIMESTER3

IPDA1005 Introduction to Probability and Data Analysis

Date and Time: 18 January 2023, (4:00 PM-5:40 PM)

NAME & ID:____.

Question:	A1	A2	A3	A4	A5	A6	B1	B2	B3	B4	B5	Total
Marks:												
marks possible	1	1	2	2	2	2	12	6	5	9	8	50

Instructions to Candidates:

Solution Key

NB give generous grades

1. This is a 100 minutes exam,
2. It is divided into two parts
   (a) Part A consists of multiple-choice/short answer/fill-in-the-gap questions, students do not need to show working;
   (b) Part B consists if short answer questions and students **MUST** show working;
3. Students are allowed to:
   (a) Use any material in this unit from Moodle. (Do not search websites or communicate with others)
   (b) Bring four pages of 2-sided A4 page formula (or cheat sheet) for this exam;
   (c) Use R and/or a scientific calculator
   (d) Use statistical tables.

PART A

A1 A professor records the values of several variables for each student in her class. These include the variables listed below.

- The number of lectures the student missed (Lectures).
- Attending or not attending an in-class test (Attendance).
- Final Grade for the topic (Fail, Pass, Credit, Distinction, or High Distinction) (Grade).
- The total number of marks earned in the class (i.e, the total of the marks on all assessments in the course. The maximum number of marks possible is 500) .

The types of variables Lectures, Attendance, Grade and Marks are respectively: [1 mark]

(A) Discrete, nominal, ordinal and continuous.

(B) Nominal, ordinal, continuous and discrete.

(C) Ordinal, nominal, ordinal and discrete.

(D) Nominal, continuous, ordinal and continuous.

A2 A television station is interested in predicting whether or not voters in its listening area are in favour of federal funding for abortions. It asks its viewers to phone in and indicate whether they are in favour of or opposed to this. Of the 2241 viewers who phoned in, 1574 (70.24%) were opposed to federal funding for abortions. The number 70.24% is a [1 mark]

(A) parameter

(B) population

(C) sample

(D) statistic

A3 What is the sample size required if we need to estimate a population proportion p such that a 90% confidence interval has margin of error no more than 0.15? [2 marks]

(A) 15.

$$n = \left(\frac{z_{\alpha/2}}{2e}\right)^2 = \left(\frac{z_{0.05}}{2(0.15)}\right)^2 = \left(\frac{1.645}{2(0.15)}\right)^2 = 30.067 \approx 31$$

(B) 30.

(C) 50.

(D) None of the above.

PART B

B1 In a study conducted by the Department of Mechanical Engineering at Virginia Polytechnic Institute, the steel rods supplied by two different companies were compared. Ten sample springs were made out of the steel rods supplied by each company and the bounciness were studied. The data are as follows.

Company A: 9:3 8:9 6:7 8:5 8:6 6:8 8:0 6:9 9:1 7:0

Company B: 11:0 9:7 9:6 10:2 10:1 10:9 9:5 10:1 9:2 9:6

Can we conclude that there is no difference between the steel rods supplied by the two companies? Use a significance level of 5%.

Hint. Copy the following into *R* console or a chunk in Rmd, to read and analyse the data.

A <-c(9.3, 8.9, 6.7, 8.5, 8.6, 6.8, 8.0, 6.9, 9.1, 7.0)

B <-c(11.0, 9.7, 9.6, 10.2, 10.1, 10.9, 9.5, 10.1, 9.2, 9.6)

(a) Write down the null and alternative hypotheses. [2 marks]

$H_0: \mu_A = \mu_B$ OR $H_0: \mu_A - \mu_B = 0$ . . .

$H_A: \mu_A \neq \mu_B$ $H_A: \mu_A - \mu_B \neq 0$

(b) Calculate the p-value. [4 marks]

mean diff : $\mu_d = -2.01$

sd diff : 1.1902

test stat : $\dfrac{\mu_d - 0}{s/\sqrt{n}} = \dfrac{-2.01 - 0}{1.1902/\sqrt{10}} = -5.3404$

p-value = pt(−5.3404, 9) = 0.00023411

(c) Based on this p-value, express your conclusion to your analysis. [2 marks]

A4 Observation of a small number of patients, randomly selected from those waiting for treatment in the casualty room of a busy hospital, showed the standard deviation of waiting time to be 16.4 min.

How many patients' progress must be monitored in order to estimate, with 95% confidence, the true average waiting time to within 5 min of the real value? [2 marks]

$e = 5 \quad s = 16.4 \quad \alpha = 5\% \Rightarrow z_{\alpha/2} = 1.96$

$n = \left(\frac{z_{\alpha/2}\,\sigma}{e}\right)^2 = 41.3295 \approx 42$

(A) 45.

(B) 41.

(C) 42.

(D) 43.

A5 A random sample of 15 eleven-year-old boys was obtained with a mean height of 142 centimetres and a standard deviation of 5 centimetres. Which of the following could be the 99% confidence interval based on that data? The following information may be useful: *t(df=14, confidence level=0.99)* = *2.977* [2 marks]

$\bar{x} \pm t_{\alpha/2}\frac{s}{\sqrt{n}}$

$142 \pm 2.977\frac{5}{\sqrt{15}}$

142 ± 3.8433

(A) 142 ± 2.27

(B) 142 ± 2.77

(C) 142 ± 2.98

(D) 142 ± 3.84

A6 A random sample of 100 Curtin students was asked how much sleep they get per night. Do Curtin students sleep on average less than 8 hours per night?

Choose the correct null and alternative hypothesis for the above statement. [2 marks]

(A) $H_0 : \mu = 8,\ H_A : \mu \neq 8$

(B) $H_0 : \mu = 8,\ H_A : \mu < 8$

(C) $H_0 : \mu = 8,\ H_A : \mu > 8$

(D) None of the above.

B2 In a random sample of $n = 1000$ families owning television sets in the city of Hamilton, Canada, it was found that $x = 560$ subscribed to HBO.

(i) Find a 95% confidence interval for the actual proportion of families in this city who subscribe to HBO. [3 marks]

$$\hat{p} = \frac{560}{1000} = 0.56 \quad (1/2)$$

$$\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \quad (1/2)$$

$$z_{\alpha/2} = z_{0.025} = 1.96 \quad (1)$$

$$0.56 \pm 1.96\sqrt{\frac{0.56(1-0.56)}{1000}} \quad (1)$$

(b) How large a sample is required if we want to be 95% confident that our estimate of p is within 0.02? [3 marks]

$$n = \left(\frac{z_{\alpha/2}}{e}\right)^2 \hat{p}(1-\hat{p}) \qquad e = 0.02$$

$$n = \left(\frac{1.96}{0.02}\right)^2 (0.56)(0.44)$$

$$n = 2366.426$$

$$n = 2367$$

(d) Construct a 99% confidence interval for the difference between the population means. [2 marks]

$t(\alpha/2, 9) = qt(0.05, 9) = -1.833113$

$\cancel{\pi}.\ \bar{x} \pm t_{\alpha/2, df} \cdot \frac{s}{\sqrt{n}}$

$-2.\cancel{0}1 \pm 1.833113 \frac{(1.1902)}{\sqrt{10}}$

-2.01 ± 0.68994 or any valid variation

B3 Sample proportions obtained from two independent samples of size 25 and 36 were 0.32 and 0.25 respectively. Construct a 95% confidence interval for the difference of proportions. [5 marks]

$\hat{p}_1 = 0.32$ $\hat{p}_2 = 0.25$

$n_1 = 25$ $n_2 = 36$

$$\hat{p}_1 - \hat{p}_2 \pm z_{\alpha/2} \sqrt{\frac{\hat{p}_1(1-\hat{p})}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

$$0.32 - 0.25 \pm 1.96 \sqrt{\frac{0.32(1-0.32)}{25} + \frac{0.25(1-0.25)}{30}}$$

$$0.07 \pm 0.2397$$

B4 Test the hypothesis that the average content of containers of a particular lubricant is 8 liters if the contents of a random sample of 10 containers are 8.21 7.95 8.13 8.16 7.92 7.87 8.02 7.85 8.09 7.89:

Use a significance level of 10%.

[9 marks]

$H_0 : \mu = 8$

$H_A = \mu \neq 8$

$\bar{x} = 8.009 \quad s = 0.1311022$

test statistic $q = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{8.009-8}{0.1311022} = 0.217086$

pvalue : $2pt(q, df) = 2pt(0.2171, 9) \ngtr 0.1$

Insufficient evidence to reject H_0.

B5 It is thought that 40% of the adult population are coffee drinkers. A random sample of size 240 contained 80 coffee drinkers. Test the hypothesis that the proportion of adult population are coffee drinkers is 40%. Use a $\alpha = 0.01$.

(a) What is the standard deviation of $\hat{p}$. [2 marks]

$$S = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{0.4(1-0.4)}{240}} = 0.031$$

(b) What is the approximate distribution of $\hat{p}$? [3 marks]

$$n\hat{p} = 96 > 10 \quad \& \quad n(1-\hat{p}) = 144 > 10$$

$$\Rightarrow \hat{p} \sim N(0.4, 0.031^2)$$

(c) What is the probability that $\hat{p}$ is between 87% and 93%? (This is the probability that $\hat{p}$ estimates p within 3%.). [3 marks]

$$P(0.87 < p < 0.93)$$

$$= P(p < 0.93) - P(p < 0.87)$$

$$= P\left(z < \frac{0.93-0.4}{0.031/\sqrt{240}}\right) - P\left(z < \frac{0.87-0.4}{0.031/\sqrt{240}}\right)$$

$$= 1 - 1$$

$$= 0$$