

### Question 3 Variant 1 [25 marks]

(a) *Inference Procedures for Two Samples (8 Marks)*

*Deltatech* software has 350 programmers divided into two groups with 200 in Group A and 150 in Group B. In order to compare the efficiencies of the two groups, the programmers are observed for 1 day.

- The 200 programmers in Group A averaged 45.2 lines of code with a standard deviation of 8.4.
- The 150 programmers in Group B averaged 42.7 lines of code with a standard deviation of 5.2.

Let  $\bar{x}_A$  denote the average number of lines of code per day produced by programmers in Group A and let  $\bar{x}_B$  be the corresponding statistic for Group B. Provide an estimate of  $\mu_A\mu_B$  and calculate an approximate 95% confidence interval.

Test the claim that Group A are more efficient than Group B by

- (i) (3 Marks) Interpreting the 95% confidence interval.
- (ii) (5 Marks) Computing the appropriate test statistic.

(b) *Theory of Statistical Inference (6 Marks)*

Answer the following questions on the theory of statistical inference.

- (i) (2 Marks) What is a  $p$ -value?
- (ii) (2 Marks) Briefly describe how  $p$ -value is used in hypothesis testing.
- (iii) (1 Mark) What is meant by a Type I error?
- (iv) (1 Mark) What is meant by a Type II error?

(c) *Confidence Intervals for Proportions (4 Marks)*

A press release for a broadband provider stated that 90% percent of customers are very satisfied with the standard of service. To test this claim, the local chamber of commerce surveyed 110 randomly selected customers. Among the sampled customers, 89 stated that they are very satisfied.

- (i) (1 Mark) Compute the appropriate value for the standard error for a confidence interval.
- (ii) (2 Marks) Compute the 95% confidence interval for  $\pi$ , the true proportion.
- (iii) (1 Mark) What is your conclusion for this claim made by the press release? Justify your answer.

*Please turn over for the remaining sections of Question 3.*

(d) ***Inference Procedures with R (4 Marks)***

Consider the following inference procedure performed on data set  $X$ .

```
> shapiro.test(X)

      Shapiro-Wilk normality test

data:  X
W = 0.9619, p-value = 0.6671
```

- (i) (1 Mark) Describe what is the purpose of this statistical procedure.
- (ii) (2 Marks) What are the null and alternative hypotheses?
- (iii) (1 Mark) Write the conclusion that follows from the code output, displayed above.

(e) ***Graphical Procedures (3 Marks)***

- (i) (2 Marks) The graph below depicts a normal probability plot. Describe what this plot is used for and how to interpret one.
- (ii) (1 Mark) What is your conclusion for the data used to construct the normal probability plot below.

