Mock Exam 01 (Part 01)

Subject: Introductory Statistics

Total marks: 100 marks (25 each) **Exam time: 60 min**

1) A city council wants to estimate the proportion of residents who support a proposed cycling lane on the main road. A researcher stands outside the central train station on weekday mornings and asks every 10th person exiting the station for their opinion, recording 200 responses.

    a) Identify the sampling method used. State one likely source of bias this method introduces and explain precisely why it would make the estimate unrepresentative of all city residents.

    b) Describe how you would redesign this study using stratified random sampling. Define the strata you would use, justify your choice of strata, and explain how sample sizes within strata would be determined.

    c) Suppose the researcher instead conducts a randomised controlled experiment to test whether providing cycling-lane information changes support levels. Identify the explanatory variable, response variable, and one variable that should be controlled. Explain the role of randomisation in this design.

2) A nutritionist measures the daily caloric intake of a random sample of 25 university students. The sample yields a mean of $\bar{x}$ = 2,140 kcal and a standard deviation of sd = 310 kcal. Assume the population is approximately normally distributed.

    a) State and verify the conditions required to construct a valid t-interval for the population mean caloric intake.

    b) Construct a 95% confidence interval for the true mean daily caloric intake. Show all working including the critical value used.

    c) A student interprets the interval as "95% of students eat between [lower] and [upper] kcal per day." State whether this interpretation is correct. Provide the correct interpretation and explain the conceptual error in the student's statement.

3) A smartphone manufacturer claims that 70% of its customers renew their device within 2 years. A consumer watchdog surveys 150 randomly selected customers and finds that 96 renewed within 2 years.

   a) State appropriate null and alternative hypotheses. Justify whether a one-tailed or two-tailed test is appropriate in this context.

   b) Verify the conditions for the test and compute the test statistic and p-value. State your conclusion at the $\alpha = 0.05$ significance level.

   c) In the context of this test, describe what a Type I error and a Type II error would each mean. State which type of error is directly controlled by the choice of $\alpha = 0.05$.

4) A medical diagnostic test for a rare disease has a sensitivity of 92% (probability of testing positive given disease is present) and a specificity of 88% (probability of testing negative given no disease). The disease affects 1% of the general population.

   a) Define the events D = "has disease" and T = "tests positive." Express sensitivity and specificity formally in terms of conditional probabilities.

   b) Using the Law of Total Probability, compute P(T), the probability that a randomly selected individual from the general population tests positive.

   c) Apply Bayes' Theorem to determine $P(D \mid T)$, the probability that a person actually has the disease given a positive test result. Comment on why this value may be surprising to clinicians who are unfamiliar with base rates.