Mock Exam 02 (Part 02)

Subject: Introductory Statistics

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

1) A social media platform tests two versions of a notification design. Version A is shown to 500 users; 185 click through. Version B is shown to 500 users; 215 click through.

   a) State the null and alternative hypotheses for testing whether the two versions have different click-through rates. Verify the conditions and compute the pooled sample proportion $\hat{p}_{\text{pooled}}$.

   b) Compute the two-proportion z-test statistic and p-value. State your conclusion at $\alpha = 0.05$ and interpret the result for the platform's design team.

   c) Construct a 95% confidence interval for the difference in true click-through rates $(p_B - p_A)$. Explain how this interval provides additional information beyond the hypothesis test result.

2) A die is suspected of being biased. It is rolled 120 times with the following observed frequencies: Face 1: 14, Face 2: 22, Face 3: 18, Face 4: 25, Face 5: 28, Face 6: 13.

   a) State the null and alternative hypotheses. Compute the expected frequency for each face under H_0 and verify the conditions for the goodness-of-fit test.

   b) Compute the chi-square goodness-of-fit statistic. Determine the degrees of freedom, find the p-value range from a table, and state your conclusion at $\alpha = 0.05$.

   c) If the test concludes the die is biased, identify which faces appear most discrepant from expectation and explain the distinction between a statistically significant result and a practically meaningful one in the context of a gambling game.

3) Researchers study the effect of study method (Flashcards, Practice Tests, Re-reading) on exam scores for students across three degree programmes (Science, Arts, Business), with 5 students per cell. A one-way ANOVA on study method yields $SS_{Between} = 320$, $SS_{Within} = 540$, with 3 groups and 45 total observations.

   a) Complete the ANOVA table: compute all degrees of freedom, mean squares, and the F-statistic. State the conclusion at $\alpha = 0.05$ (use F-critical ≈ 3.21).

   b) Assuming the F-test is significant, the researchers plan to apply Bonferroni post-hoc tests. How many pairwise comparisons are there? State the Bonferroni-adjusted significance level and explain the purpose of this adjustment.

   c) The researchers want to extend to a two-way ANOVA including degree programme as a second factor. Define the concept of an interaction effect and explain, with a sketch or verbal description, what it would mean if study method and degree programme showed a significant interaction for exam scores.

4) A real estate analyst models house sale price (€000s) using three predictors: floor area in m² (x_1), distance to city centre in km (x_2), and number of bedrooms (x_3). Fitted model: $\hat{y} = 85.4 + 1.92(x_1) - 3.11(x_2) + 4.80(x_3)$. Model is fitted on $n = 60$ observations. $R^2 = 0.74$, Adjusted $R^2 = 0.72$.

   a) Interpret each regression coefficient in context, paying attention to units and the meaning of holding other variables constant. Interpret the intercept and comment on whether it is meaningful in this setting.

   b) Explain the difference between R^2 and Adjusted R^2. Given that $R^2 = 0.74$ and Adjusted $R^2 = 0.72$, what does the small gap suggest about the number of predictors relative to the sample size? Compute the number of additional predictors that would cause Adjusted R^2 to fall noticeably below R^2.

   c) Define multicollinearity. Explain why it might be a concern in this model if floor area and number of bedrooms are highly correlated. Describe two ways a researcher can detect multicollinearity and one approach to address it.