



Mock Exam 01 (Part 02)

Subject: Introductory Statistics

Total marks: 100 marks (25 each)

Exam time: 60 min

- 1) Two independent groups of students sit a standardised mathematics test before and after different teaching interventions. Group A (traditional instruction, $n = 30$) has a mean score of 72.4 with standard deviation 8.1. Group B (active-learning instruction, $n = 28$) has a mean score of 77.9 with standard deviation 9.6.
 - a) State the null and alternative hypotheses for testing whether the two instruction methods produce different mean scores. Specify all assumptions required.
 - b) Compute the two-sample t -statistic using the unpooled (Welch) formula. Approximate the degrees of freedom and determine the p-value range using a t -table. State your conclusion at $\alpha = 0.05$.
 - c) Construct a 90% confidence interval for the difference in population means ($\mu_B - \mu_A$). Interpret the interval in plain language and explain what it implies about the practical significance of the difference.

- 2) A university surveys 200 students on their preferred study environment (Library, Café, Home) and their academic year (First Year, Upper Year). The observed frequencies are: Library: 40 First Year, 55 Upper Year; Café: 30 First Year, 20 Upper Year; Home: 20 First Year, 35 Upper Year.
 - a) State the null and alternative hypotheses. Compute all expected cell counts and verify that conditions for the chi-square test are satisfied.
 - b) Compute the chi-square test statistic χ^2 . State the degrees of freedom and find the p-value range. State your conclusion at $\alpha = 0.05$.
 - c) Identify the cell(s) contributing most to the chi-square statistic. Interpret what this tells you about the relationship between academic year and preferred study environment.



-
- 3) A botanist grows plants under three fertiliser conditions (A, B, C), with 8 plants per group. The group means are $\bar{y}_A = 14.2$ cm, $\bar{y}_B = 18.7$ cm, $\bar{y}_C = 16.4$ cm and the grand mean is $\bar{y} = 16.43$ cm. The within-group sum of squares $SS_{\text{Within}} = 168.0$.
- a) State the null and alternative hypotheses for the ANOVA. List the three conditions that must hold for ANOVA to be valid and describe how you would check each one.
 - b) Compute SS_{Between} , MS_{Between} , MS_{Within} , and the F-statistic. Using an F-table with appropriate degrees of freedom, determine whether the result is significant at $\alpha = 0.05$.
 - c) Explain why it would be inappropriate to use multiple pairwise t -tests instead of ANOVA to compare the three groups. If the ANOVA is significant, describe what a post-hoc test is designed to determine.
- 4) A dataset of 20 cities records average annual temperature ($^{\circ}\text{C}$) as the explanatory variable and annual ice cream sales (thousands of units) as the response. Summary statistics: $\bar{x} = 15.4$, $\bar{y} = 82.6$, $s_x = 4.2$, $s_y = 18.3$, $r = 0.87$.
- a) Compute the slope b_1 and intercept b_0 of the least-squares regression line. Write the equation of the line and interpret the slope in context.
 - b) The standard error of the slope is $SE(b_1) = 0.52$. Test $H_0: \beta_1 = 0$ against $H_a: \beta_1 \neq 0$ at $\alpha = 0.05$. State the test statistic, degrees of freedom, and conclusion.
 - c) State the four conditions for inference in regression. For each condition, name the diagnostic tool or plot used to check it.