A Level Statistics
Practice Test 3: Advanced Topics

Instructions:
Answer all questions. Show your working clearly.
Calculators may be used unless stated otherwise.
Draw diagrams where appropriate to illustrate your solutions.
Time allowed: 3 hours

Section A: Fundamental Concepts [25 marks]

1. [**12 marks**] Define and explain fundamental concepts:

(a) Define regression analysis and explain its purpose.

(b) Explain what is meant by "correlation" and "causation."

(c) State the difference between simple and multiple regression.

(d) Define the coefficient of determination (R^2) and its interpretation.

(e) Distinguish between fitted values and residuals.

(f) Explain how regression relates to prediction and inference.

2. [**8 marks**] Explain the importance of these concepts:

(a) Why is the least squares method used for fitting regression lines?

(b) Explain how regression analysis helps in understanding relationships between variables.

(c) Describe the role of assumptions in regression modeling.

(d) Explain the relationship between correlation and regression slopes.

3. [**5 marks**] Practical and theoretical context:

(a) Explain why regression is fundamental to data science and machine learning.

(b) Describe the role of regression in quality control and process improvement.

(c) Explain how regression applies to economic forecasting and business analytics.

Section B: Simple Linear Regression - Theory [30 marks]

4. **[15 marks]** State and explain simple linear regression:

   (a) Write the mathematical model for simple linear regression.

   (b) Explain the meaning of each parameter in the regression equation.

   (c) Describe the least squares criterion for parameter estimation.

   (d) State the normal equations for finding and .

   (e) Explain the assumptions underlying simple linear regression.

   (f) Describe how to interpret the regression coefficients.

5. **[15 marks]** Properties and inference in regression:

   (a) Explain the properties of least squares estimators.

   (b) Describe the sampling distributions of and .

   (c) State the conditions for unbiased estimation.

   (d) Explain the Gauss-Markov theorem and its significance.

   (e) Describe how to construct confidence intervals for regression parameters.

   (f) Explain prediction intervals versus confidence intervals for mean response.

   (g) Describe the concept of leverage and influential observations.

   (h) Explain how outliers affect regression analysis.

   (i) Describe methods for detecting assumption violations.

Section C: Regression Applications [35 marks]

6. **[18 marks]** A study examines the relationship between study hours (x) and exam scores (y). Data: n = 12, x = 84, y = 876, x^2 = 674, y^2 = 65,248, xy = 6,534:

   (a) Calculate the sample means $\bar{x}$ and $\bar{y}$.

   (b) Calculate the sample correlation coefficient r.

   (c) Determine the least squares estimates $\hat{\beta}_0$ and $\hat{\beta}_1$.

   (d) Write the fitted regression equation.

   (e) Interpret the slope and intercept in context.

   (f) Calculate the coefficient of determination R^2.

   (g) Predict the exam score for a student who studies 8 hours.

   (h) Calculate the residual sum of squares (RSS).

   (i) Estimate the standard error of regression .

7. **[17 marks]** For the regression in question 6, perform statistical inference:

   (a) Calculate the standard error of the slope estimate SE($\hat{\beta}_1$).

   (b) Test H: = 0 versus H: 0 at = 0.05.

(c) Calculate the p-value for this test.

(d) Construct a 95

(e) Interpret the confidence interval in context.

(f) Test whether the slope equals 5 points per hour at $= 0.05$.

(g) Calculate a 95

(h) Calculate a 95

(i) Explain the difference between these two intervals.

Answer Space

Use this space for your working and answers.

Formulae and Key Concepts

Simple Linear Regression:

Model: $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$

Slope: $\hat{\beta}_1 = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2} = \frac{S_{xy}}{S_{xx}}$

Intercept: $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$

Correlation and R²:

Correlation: $r = \frac{S_{xy}}{\sqrt{S_{xx} S_{yy}}}$

Coefficient of determination: $R^2 = r^2 = \frac{SSR}{SST}$

Standard Errors:

Regression SE: $\hat{\sigma} = \sqrt{\frac{SSE}{n-2}}$

SE of slope: $SE(\hat{\beta}_1) = \frac{\hat{\sigma}}{\sqrt{S_{xx}}}$

SE of intercept: $SE(\hat{\beta}_0) = \hat{\sigma}\sqrt{\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}}$

Confidence and Prediction Intervals:

CI for : $\hat{\beta}_1 \pm t_{\alpha/2, n-2} \cdot SE(\hat{\beta}_1)$

CI for mean response: $\hat{y} \pm t_{\alpha/2, n-2} \cdot \hat{\sigma}\sqrt{\frac{1}{n} + \frac{(x-\bar{x})^2}{S_{xx}}}$

Prediction interval: $\hat{y} \pm t_{\alpha/2, n-2} \cdot \hat{\sigma}\sqrt{1 + \frac{1}{n} + \frac{(x-\bar{x})^2}{S_{xx}}}$

Sums of Squares:

$S_{xx} = \sum(x_i - \bar{x})^2 = \sum x_i^2 - n\bar{x}^2$

$S_{yy} = \sum(y_i - \bar{y})^2 = \sum y_i^2 - n\bar{y}^2$

$$S_{xy} = \sum(x_i - \bar{x})(y_i - \bar{y}) = \sum x_i y_i - n\bar{x}\bar{y}$$

ANOVA for Regression:

SST = SSR + SSE

$$SSR = \hat{\beta}_1^2 S_{xx}$$

$$SSE = S_{yy} - \hat{\beta}_1^2 S_{xx}$$

F-test: $F = \frac{MSR}{MSE} = \frac{SSR/1}{SSE/(n-2)}$

END OF TEST

Total marks: 95

For more resources and practice materials, visit:
stepupmaths.co.uk