A Level Statistics
Practice Test 4: Hypothesis Testing

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: Regression Analysis and Linear Models [25 marks]

1. [**12 marks**] Define and explain regression hypothesis testing:

(a) Explain the purpose of testing regression coefficients for significance.

(b) Write the null and alternative hypotheses for testing whether a slope coefficient equals zero.

(c) Define the standard error of a regression coefficient and its role in hypothesis testing.

(d) Explain what R^2 represents and how it relates to the significance of the regression model.

(e) Describe the F-test for overall regression significance.

(f) Explain the assumptions required for valid regression hypothesis tests.

2. [**8 marks**] Explain confidence and prediction intervals in regression:

(a) Distinguish between confidence intervals for the mean response and prediction intervals.

(b) Explain why prediction intervals are wider than confidence intervals.

(c) Describe how to test whether the regression line passes through a specific point.

(d) Explain the concept of leverage and influential observations in regression.

3. [**5 marks**] Analyze regression assumptions:

(a) List the key assumptions for linear regression.

(b) Explain how to check the assumption of constant variance (homoscedasticity).

(c) Describe methods for detecting non-linearity in the relationship.

Section B: Time Series and Goodness of Fit Testing [30 marks]

4. **[15 marks]** Testing for randomness and independence:

(a) Define the runs test and explain its purpose in testing randomness.

(b) Explain autocorrelation and its implications for time series analysis.

(c) Describe the Durbin-Watson test for detecting serial correlation.

(d) Explain how to test for trend in time series data.

(e) Define stationarity and explain tests for stationarity.

(f) Describe seasonal decomposition and testing for seasonal effects.

5. **[15 marks]** A quality control manager monitors daily production defects over 20 days:
Day: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 Defects: 3, 5, 2, 4, 6, 3, 7, 4, 5, 8, 6, 9, 7, 8, 10, 9, 11, 8, 10, 12

(a) Plot the time series and comment on any apparent patterns.

(b) Test for the presence of a linear trend using correlation analysis.

(c) Calculate the correlation coefficient between defects and day number.

(d) Test whether this correlation is significantly different from zero at = 0.05.

(e) Perform a runs test to check for randomness in the sequence.

(f) Calculate the expected number of runs and the test statistic.

(g) Make conclusions about the randomness and trend in the defect data.

(h) Suggest appropriate control measures based on your findings.

(i) Calculate a prediction interval for day 21 defects assuming linear trend.

Section C: Multi-Factor and Complex Testing Scenarios [35 marks]

6. **[18 marks]** A pharmaceutical study examines the effectiveness of a new drug across different age groups and genders:

	Age Group		
Gender	Young (18-35)	Middle (36-55)	Senior (56+)
Male	85% (n=40)	78% (n=35)	72% (n=30)
Female	88% (n=45)	82% (n=38)	75% (n=32)

Success rates and sample sizes are shown in parentheses.

(a) Test whether success rates differ significantly across age groups (ignoring gender).

(b) Test whether success rates differ significantly between genders (ignoring age).

(c) Perform a two-way ANOVA-style analysis to test for age and gender effects.

(d) Test for interaction between age and gender effects.

(e) Calculate effect sizes for age and gender differences.

(f) Determine which age groups differ significantly using pairwise comparisons.

(g) Apply appropriate multiple testing corrections.

(h) Calculate confidence intervals for success rates in each subgroup.

(i) Make recommendations about drug effectiveness across different populations.

7. **[17 marks]** A marketing research study examines customer satisfaction across four service centers using multiple measures:

Quantitative Ratings (1-10 scale): Center A: 7.2, 8.1, 6.8, 7.9, 7.5, 8.3, 7.1, 7.8, 8.0, 7.4 Center B: 8.5, 8.9, 8.2, 8.7, 8.1, 8.8, 8.4, 8.6, 8.3, 8.0 Center C: 6.5, 7.2, 6.8, 7.0, 6.9, 7.1, 6.7, 6.6, 7.3, 6.8 Center D: 7.8, 8.2, 7.6, 8.1, 7.9, 8.4, 7.7, 8.0, 7.5, 8.3

Categorical Satisfaction Levels:

Center	Satisfied	Neutral	Dissatisfied
A	65	25	10
B	85	12	3
C	45	35	20
D	75	20	5

(a) Compare the mean quantitative ratings across centers using ANOVA.

(b) Test the assumptions of normality and equal variances for the ANOVA.

(c) Perform post-hoc pairwise comparisons between centers.

(d) Test for independence between center and satisfaction level using chi-squared.

(e) Calculate standardized residuals to identify which centers deviate most from expected.

(f) Compare the conclusions from quantitative versus categorical analysis.

(g) Calculate effect sizes for both types of analysis.

(h) Rank the centers by performance using both measures.

(i) Discuss any discrepancies between the two approaches and suggest explanations.

Answer Space

Use this space for your working and answers.

Formulae and Key Concepts

Regression Tests:

Slope test: $t = \frac{b_1 - \beta_1}{SE(b_1)}$ with df = n-2

Overall F-test: $F = \frac{MSR}{MSE} = \frac{R^2/(k)}{(1-R^2)/(n-k-1)}$

where k = number of predictors

Standard error: $SE(b_1) = \frac{s}{\sqrt{\sum(x_i-\bar{x})^2}}$

Confidence and Prediction Intervals:

Confidence interval for mean: $\hat{y} \pm t_{\alpha/2} \cdot SE(\hat{y})$

Prediction interval: $\hat{y} \pm t_{\alpha/2} \cdot SE(pred)$

where $SE(pred) > SE(\hat{y})$ due to additional uncertainty

Runs Test:

Expected runs: $E(R) = \frac{2n_1n_2}{n_1+n_2} + 1$

Variance: $Var(R) = \frac{2n_1n_2(2n_1n_2-n_1-n_2)}{(n_1+n_2)^2(n_1+n_2-1)}$

Test statistic: $z = \frac{R-E(R)}{\sqrt{Var(R)}}$

Trend Testing:

Correlation with time: $r = \frac{\sum(t_i-\bar{t})(y_i-\bar{y})}{\sqrt{\sum(t_i-\bar{t})^2\sum(y_i-\bar{y})^2}}$

Test statistic: $t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$

Two-Way Analysis:

Main effects and interaction testing

Cell means model: $y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijk}$

Test for interaction: F-test comparing models with/without interaction

Multiple Comparisons:

Bonferroni: $\alpha_{adj} = \frac{\alpha}{m}$

Tukey HSD for pairwise comparisons in ANOVA

False Discovery Rate (FDR) procedures

Effect Sizes:

Regression: R^2 (proportion of variance explained)

ANOVA: $\eta^2 = \frac{SS_{effect}}{SS_{total}}$

Cohen's d for mean differences

Cramer's V for categorical associations: $V = \sqrt{\frac{\chi^2}{n \cdot min(r-1,c-1)}}$

Chi-Squared for Independence:

Test statistic: $\chi^2 = \sum \frac{(O_{ij}-E_{ij})^2}{E_{ij}}$

Expected frequency: $E_{ij} = \frac{R_i \times C_j}{n}$

Standardized residual: $\frac{O_{ij}-E_{ij}}{\sqrt{E_{ij}(1-p_{i\cdot})(1-p_{\cdot j})}}$

Model Assumptions:

Regression: Linearity, independence, normality, constant variance

ANOVA: Independence, normality, equal variances

Time series: Stationarity, independence of residuals

END OF TEST

Total marks: 90

For more resources and practice materials, visit: