A Level Statistics
Practice Test 5: 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: Bayesian Inference and Modern Approaches [25 marks]

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

(a) Define prior probability, likelihood, and posterior probability in Bayesian inference.

(b) Explain Bayes' theorem and its role in updating beliefs about hypotheses.

(c) Compare frequentist and Bayesian approaches to hypothesis testing.

(d) Define Bayes factors and explain their interpretation.

(e) Explain the concept of credible intervals versus confidence intervals.

(f) Describe how to incorporate prior knowledge into hypothesis testing.

2. [**8 marks**] Explain modern hypothesis testing considerations:

(a) Define the replication crisis and its implications for hypothesis testing.

(b) Explain pre-registration of studies and its importance.

(c) Describe publication bias and how it affects scientific literature.

(d) Explain the difference between exploratory and confirmatory data analysis.

3. [**5 marks**] Analyze alternative approaches to traditional hypothesis testing:

(a) Explain estimation-based approaches versus hypothesis testing.

(b) Describe the American Statistical Association's statement on p-values.

(c) Explain the concept of statistical equivalence testing.

Section B: Bootstrap and Resampling Methods [30 marks]

4. [**15 marks**] Define and explain resampling techniques:

(a) Define the bootstrap method and explain its purpose in statistical inference.

(b) Explain how bootstrap confidence intervals are constructed.

(c) Describe permutation tests and when they are appropriate.

(d) Explain the difference between parametric and non-parametric bootstrap.

(e) Define jackknife resampling and its applications.

(f) Explain how resampling methods handle violations of traditional assumptions.

5. [**15 marks**] A research team studies the effectiveness of two learning methods using small samples:

Method A (n=8): 76, 82, 78, 85, 79, 83, 77, 81 **Method B (n=7):** 88, 84, 90, 86, 92, 87, 89
Traditional analysis shows: $\bar{x}_A = 80.1$, $s_A = 3.18$, $\bar{x}_B = 88.0$, $s_B = 2.71$

(a) Perform a traditional two-sample t-test to compare the methods.

(b) Explain why the small sample sizes might make traditional inference unreliable.

(c) Describe how to perform a bootstrap test for the difference in means.

(d) Outline the steps for a permutation test comparing the two groups.

(e) Calculate a bootstrap confidence interval for the difference in means (conceptual steps).

(f) Explain how bootstrap methods address the small sample size problem.

(g) Compare the assumptions required for traditional t-tests versus resampling methods.

(h) Discuss the advantages and limitations of resampling approaches.

(i) Determine which method provides more reliable inference for this scenario.

Section C: Meta-Analysis and Evidence Synthesis [35 marks]

6. [**18 marks**] Define and explain meta-analysis concepts:

(a) Define meta-analysis and explain its purpose in synthesizing research evidence.

(b) Explain fixed-effects versus random-effects models in meta-analysis.

(c) Define heterogeneity and methods for assessing it (I^2 statistic).

(d) Explain publication bias and funnel plot analysis.

(e) Describe forest plots and their interpretation.

(f) Explain how to combine effect sizes across studies.

(g) Define fail-safe N and its interpretation.

(h) Describe sensitivity analysis in meta-analysis.

(i) Explain the importance of study quality assessment in meta-analysis.

7. [**17 marks**] A meta-analysis examines the effectiveness of a new educational intervention across multiple studies:

Study	Sample Size	Effect Size (d)	Standard Error	Weight	95% CI
Study 1	120	0.45	0.18	15%	[0.10, 0.80]
Study 2	200	0.32	0.14	25%	[0.04, 0.60]
Study 3	80	0.68	0.23	10%	[0.23, 1.13]
Study 4	150	0.28	0.16	20%	[-0.04, 0.60]
Study 5	180	0.51	0.15	22%	[0.22, 0.80]
Study 6	90	0.15	0.21	8%	[-0.26, 0.56]

Additional information: Overall effect size = 0.39, Q-statistic = 12.8, I^2 = 61%

(a) Calculate the weighted average effect size and verify the given result.

(b) Test for heterogeneity using the Q-statistic (df = 5, critical value = 11.07 at = 0.05).

(c) Interpret the I^2 statistic and assess the level of heterogeneity.

(d) Identify which studies have confidence intervals that exclude zero.

(e) Assess whether there's evidence of publication bias from the pattern of results.

(f) Calculate the overall 95% confidence interval for the meta-analytic effect.

(g) Perform a sensitivity analysis by removing the most extreme study.

(h) Interpret the practical significance of the overall effect size.

(i) Make recommendations for future research based on these findings.

(j) Discuss limitations of this meta-analysis and potential sources of bias.

Answer Space

Use this space for your working and answers.

Formulae and Key Concepts

Bayesian Inference:

Bayes' Theorem: $P(H|D) = \frac{P(D|H) \cdot P(H)}{P(D)}$

Bayes Factor: $BF_{10} = \frac{P(D|H_1)}{P(D|H_0)}$ (evidence for H vs H)

Interpretation: BF ¿ 10 (strong), 3-10 (moderate), 1-3 (weak evidence)

Bootstrap Methods:

Bootstrap sample: Resample with replacement from original data

Bootstrap confidence interval: Use percentiles of bootstrap distribution

Bias-corrected bootstrap: Adjust for bias in bootstrap estimates

Number of bootstrap samples: Typically 1000-10000

Permutation Tests:
Null hypothesis: No difference between groups
Test statistic: Difference in means (or other measure)
P-value: Proportion of permutations with test statistic observed
Number of permutations: All possible or large random sample

Meta-Analysis:
Fixed effects weight: $w_i = \frac{1}{SE_i^2}$
Weighted mean: $\bar{d} = \frac{\sum w_i d_i}{\sum w_i}$
Standard error: $SE = \frac{1}{\sqrt{\sum w_i}}$

Heterogeneity:
Q-statistic: $Q = \sum w_i(d_i - \bar{d})^2$
I² statistic: $I^2 = \frac{Q-(k-1)}{Q} \times 100\%$
Interpretation: I² ¡ 25% (low), 25-75% (moderate), ¿ 75% (high)

Effect Size Interpretation:
Cohen's d: 0.2 (small), 0.5 (medium), 0.8 (large)
Practical significance depends on context and cost-benefit analysis
Statistical significance practical significance

Publication Bias:
Funnel plot: Effect size vs. standard error
Symmetric plot suggests no bias
Egger's test: Linear regression of effect size on standard error
Fail-safe N: Number of null studies needed to make effect non-significant

Modern Considerations:
Pre-registration: Specify hypotheses and analysis plan before data collection
Multiple testing: Adjust for family-wise error rate
Effect size reporting: Always report alongside p-values
Confidence intervals: Provide range of plausible values

Alternative Approaches:
Equivalence testing: Test if difference is smaller than threshold
Estimation focus: Emphasize effect sizes and confidence intervals
Bayesian model averaging: Account for model uncertainty
Robust methods: Less sensitive to outliers and violations

END OF TEST

Total marks: 90

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