

VU LEARNING HUB

vulearninghub.xyz

HIGHLIGHTED HANDOUT

STAT406

Statistical Inference Methods

STAT406 by VU Learning HUB

Prepared for VU students • Free study resources

<https://vulearninghub.xyz>

STAT406

Statistical Inference Methods

This handout covers the complete course topic by topic in simple, exam-focused language. Each chapter begins with learning objectives and ends with key points and exam pointers. Definitions are boxed in green-friendly format, examples are worked out step by step, and very important exam points are flagged separately. Study the highlighted points carefully before your papers. 

Course Information	
Course Code	STAT406
Course Title	Statistical Inference Methods
Credit Hours	3
Level	BS/MS

Table of Contents

Table of Contents	2
1. Introduction to Statistical Inference	3
2. Point Estimation	4
3. Interval Estimation	6
4. Hypothesis Testing: Foundations	8
5. Tests for Means and Proportions	9
6. Analysis of Variance (ANOVA)	11
7. Regression Inference	13
8. Nonparametric Methods	14
9. Likelihood Ratio Tests and Goodness of Fit	16
10. Bayesian Inference Basics	18
11. Resampling: Bootstrap and Permutation	19
Glossary of Key Terms	21
Practice Questions for Exams	21
MCQ Practice Test	26
Exam Strategy Guide 	30

1. Introduction to Statistical Inference

Learning Objectives

- Distinguish populations from samples
- Contrast estimation with hypothesis testing

Statistical inference is defined as drawing conclusions about a population from sample data, quantifying the uncertainty of those conclusions.

- Population: the whole group of interest; sample: the observed subset; parameter: population quantity (μ , σ); statistic: sample quantity ($\bar{x}$, s).
- Two branches: estimation (point + interval) and hypothesis testing.
- Sampling matters: random sampling justifies inference; biased sampling invalidates it.

Sampling distributions

- The sampling distribution of $\bar{x}$ centres on μ with standard error $\sigma/\sqrt{n}$.
- Central limit theorem: for large n , $\bar{x}$ is approximately normal regardless of population shape — the engine of classical inference.

Standard error of mean: $SE = \sigma / \sqrt{n}$

Key Points to Remember

- Larger $n \rightarrow$ smaller SE $\rightarrow$ sharper inference.
- CLT lets us use normal theory on non-normal data.

Exam Pointer

Very important for exams: Compute SE and explain what it measures.

Why inference beats description

- Description summarizes the sample; inference generalizes.
- Uncertainty quantification is the value-add.
- Decisions need inference, not just summaries.

Finite population correction

- Sampling without replacement from small populations.
- $FPC = \sqrt{(N-n)/(N-1)}$.
- Ignore when $n/N < 5\%$.

Sufficiency intuition

- Sufficient statistic captures all parameter info.
- Factorization theorem identifies them.
- Rao-Blackwell improves via sufficiency.

Law of large numbers

- Sample means converge to population means.
- Weak vs strong versions.
- Why statistics works at all.

Populations vs samples

- Population: everything; sample: what you observe.
- Parameters describe populations; statistics describe samples.
- Inference bridges the two.
- Random sampling justifies the bridge.
- Convenience samples break it.

Sampling distributions

- Distribution of a statistic over repeats.
- Standard error is its standard deviation.
- CLT describes the mean's shape.
- Simulation makes it concrete.
- The foundation of all inference.

2. Point Estimation

Learning Objectives

- Define desirable estimator properties
- Derive MLEs for common distributions

Properties of estimators

Property	Meaning
Unbiasedness	$E(\text{estimator}) = \text{parameter}$
Consistency	Converges to parameter as n grows
Efficiency	Smallest variance among unbiased estimators
Sufficiency	Uses all sample information about the parameter

Methods

- Method of moments: equate sample moments to population moments and solve.
- Maximum likelihood (MLE): choose the parameter maximizing the likelihood of observed data.
- MLEs are consistent, asymptotically efficient and invariant under transformation — very important for exams: MLE is the default method in modern statistics.

Example

Bernoulli(p): likelihood $L = p^{\sum(x)} (1-p)^{(n-\sum(x))}$; maximizing gives $\hat{p} = \bar{x}$ — the sample proportion.

Key Points to Remember

- Unbiased + low variance = good estimator.
- MLE: write likelihood, log it, differentiate, solve.

Exam Pointer

Very important for exams: Derive the MLE of the Poisson mean.

Bias-variance in estimation

- $MSE = \text{variance} + \text{bias}^2$.
- Slight bias can buy big variance cuts.
- Shrinkage estimators exploit this.

Cramer-Rao bound

- Lower bound on unbiased estimator variance.
- Efficient estimators attain it.
- Fisher information sets the limit.

Invariance of MLE

- MLE of $g(\theta)$ is $g(\text{MLE})$.
- Unique among methods.
- Transforms inference effortlessly.

Delta method

- SE of transformed estimators via Taylor.
- $\text{SE}(g) \sim |g'| \text{SE}$.
- Ratios and logs handled easily.

Method of moments

- Equate sample moments to population moments.
- Simple and often closed-form.
- Less efficient than MLE generally.
- Good starting values for MLE.
- Know when it is used.

Fisher information

- Curvature of the log-likelihood.
- Larger means more precise estimation.
- Inverse gives asymptotic variance.
- Additive over independent observations.
- Observed vs expected versions.

3. Interval Estimation

Learning Objectives

- Construct confidence intervals
- Interpret them correctly

- Confidence interval: estimate $\pm$ margin of error; margin = critical value $\times$ SE.
- One mean (sigma known): $\bar{x} \pm z_{(\alpha/2)} \sigma/\sqrt{n}$; sigma unknown: use t with $n-1$ df.
- Proportions: $\hat{p} \pm z \sqrt{\hat{p}(1-\hat{p})/n}$.
- Correct interpretation: "95% confident" means 95% of such intervals cover the parameter — NOT "95% probability the parameter is inside this one".

$$\text{CI: } \bar{x} \pm t_{(n-1, \alpha/2)} \times s/\sqrt{n}$$

- Width shrinks with larger n and lower confidence; grows with more variability.
- Sample-size formula: $n = (z \sigma / E)^2$ for a desired margin E .

Key Points to Remember

- The t-distribution replaces z when σ is estimated.
- Interpretation wording is examined precisely.

Exam Pointer

Very important for exams: Compute a 95% CI and state the interpretation in words.

t-distribution intuition

- Heavier tails than normal — extra uncertainty.
- Approaches normal as df grows.
- Gosset invented it for Guinness brewing.

Pivotal quantities

- Statistics with parameter-free distributions.
- $(\bar{x} - \mu)/(s/\sqrt{n}) \sim t$.
- The machinery behind CIs.

Wilson interval

- Better than Wald for proportions.
- Works near 0 and 1.
- Use it by default.

Exact binomial CI

- Clopper-Pearson: conservative but exact.
- Inverts the binomial test.
- Small samples need it.

Interpreting confidence levels

- 95% of such intervals cover the truth.
- Not 95% probability for this interval.
- Width trades off against confidence.
- One-sided intervals exist too.

- The classic exam trap — state it right. 

One-sample z-interval

- $\bar{x} \pm z^* \sigma / \sqrt{n}$, sigma known.
- Rare in practice, common in exams.
- $z^* = 1.96$ for 95%.
- Derive from the pivot.
- Stepping stone to t-intervals.

4. Hypothesis Testing: Foundations

Learning Objectives

- Set up null and alternative hypotheses
- Explain p-values and error types

- H_0 (null): the default/no-effect claim; H_1 (alternative): what we seek evidence for.
- Test statistic measures distance from H_0 ; p-value = probability of data this extreme under H_0 .
- Reject H_0 if $p < \alpha$ (significance level, often 0.05).

Decision \ Truth	H_0 true	H_0 false
Reject H_0	Type I error (alpha)	Correct (power)
Fail to reject	Correct	Type II error (beta)

- Power = $1 - \beta$: probability of detecting a real effect; rises with n , effect size and α .
- Practical vs statistical significance: huge n can make trivial effects "significant".

Key Points to Remember

- p-value is about the data under H_0 , not $P(H_0 \text{ is true})$.
- Never "accept" H_0 — we fail to reject it.

Exam Pointer

Very important for exams: Given a scenario, state H_0/H_1 and identify error types.

One-sided vs two-sided tests

- Two-sided: default, detects either direction.
- One-sided: needs strong prior justification.

- Halves the p-value — don't cheat with it.

Likelihood principle

- All evidence is in the likelihood.
- Bayesians embrace it; frequentists partly.
- Stopping rules shouldn't matter — but do.

p-hacking

- Trying analyses until $p < 0.05$.
- Pre-registration prevents it.
- Replication crisis roots here.

Sequential testing

- Peek with proper alpha spending.
- O'Brien-Fleming boundaries.
- Clinical trials use this.

Critical region design

- Reject H_0 for extreme test statistics.
- Size alpha set before seeing data.
- Two-sided splits alpha both ways.
- Power grows with sample size.
- Draw the regions for exams.

Power analysis

- Choose n before collecting data.
- Needs effect size, alpha, power.
- Underpowered studies waste resources.
- Software: G*Power, R pwr.
- Ethics boards demand it.

5. Tests for Means and Proportions

Learning Objectives

- Choose the right test for the design
- Run z , t and chi-square tests

Situation	Test
One mean, sigma known	z-test
One mean, sigma unknown	one-sample t-test
Two independent means	two-sample t-test
Paired measurements	paired t-test
One/two proportions	z-test for proportions
Categorical association	chi-square test

- Paired vs independent: pairing (before-after) removes subject variability — more powerful.
- Assumptions: approximate normality (or large n via CLT), independent observations.

Example

Drug trial: paired t-test on before-after blood pressure; independent t-test comparing drug vs placebo groups.

Key Points to Remember

- Match the test to the design — pairing changes everything.
- Check assumptions before trusting p-values.

Exam Pointer

Very important for exams: Select the test for four scenarios and justify.

Welch vs pooled t-test

- Welch: no equal-variance assumption.
- Safer default in practice.
- Know when each applies.

Paired design power

- Removes between-subject noise.
- Needs fewer subjects for same power.
- Carryover effects are the risk.

Effect sizes

- Cohen's d for means; odds ratios for binary.
- Report with every p-value.

- What matters, quantified.

Non-inferiority trials

- Show "not worse by margin".
- Flipped hypotheses from superiority.
- Common in drug approvals.

Pooled variance

- Weighted average of two sample variances.
- Valid under equal-variance assumption.
- Test the assumption first (F-test).
- Welch avoids the assumption.
- Know both formulas.

Proportion z-test

- Normal approximation needs $np \geq 10$.
- Pooled p under H_0 .
- Continuity correction optional.
- Exact binomial for small n .
- Election polling uses it.

6. Analysis of Variance (ANOVA)

Learning Objectives

- Explain the ANOVA logic
- Run one-way ANOVA and post-hoc tests

- Problem: comparing 3+ means with t-tests inflates Type I error — ANOVA tests all at once.
- Logic: $F = \text{between-group variance} / \text{within-group variance}$; large F means real group differences.
- Assumptions: normality within groups, equal variances, independence.

$$F = MS_{\text{between}} / MS_{\text{within}}, \text{ df} = (k-1, N-k)$$

- Significant F only says "some differ" — post-hoc tests (Tukey) find which pairs.
- Two-way ANOVA adds a second factor and their interaction.

Example

Compare wheat yields of 4 fertilizer treatments across plots — one-way ANOVA, then Tukey to find the best treatment.

Key Points to Remember

- ANOVA partitions total variation into explained + unexplained.
- Post-hoc only after significant F.

Exam Pointer

Very important for exams: Complete a partial ANOVA table and interpret.

ANOVA assumptions check

- Residual plots, Levene's test.
- Transform or go nonparametric if violated.
- Robust ANOVA exists too.

Tukey HSD logic

- Controls family-wise error across pairs.
- Honest significant differences.
- Conservative but trustworthy.

Interaction in ANOVA

- Effect of A depends on B level.
- Plot it — tables hide it.
- Interpret main effects cautiously.

Randomized block designs

- Block known nuisance variation.
- Agriculture invented this (Fisher).
- Precision gains are free.

Two-way ANOVA

- Two factors plus their interaction.
- Interaction means effects depend on each other.
- Partition SS into A, B, AB, error.

- Plot interactions always.
- Randomized blocks are a special case.

Kruskal-Wallis test

- Nonparametric one-way ANOVA.
- Ranks across all groups.
- Post-hoc: Dunn's test.
- No normality needed.
- Report medians with it.

7. Regression Inference

Learning Objectives

- Test regression coefficients
- Build prediction intervals

- Model: $y = \beta_0 + \beta_1 x + \epsilon$; least squares minimizes sum of squared residuals.
- Inference on slope: $t = b_1 / SE(b_1)$; tests whether x really affects y .
- R-squared: proportion of variation explained; adjusted R-squared penalizes extra predictors.
- Confidence interval for the mean response vs wider prediction interval for a new observation.

Diagnostics

- Residual plots reveal non-linearity, heteroscedasticity and outliers.
- Influential points (leverage) can distort the line — check Cook's distance.

Key Points to Remember

- A significant slope with tiny R-squared is real but weak.
- Prediction intervals are wider than confidence intervals.

Exam Pointer

Very important for exams: Interpret a regression output table fully.

Interpreting R-squared

- Proportion of variance explained.

- High R^2 with wrong model still misleads.
- Adjusted R^2 penalizes kitchen-sink models.

Regression through origin

- Rarely justified — needs theory.
- R-squared definition changes.
- Test, don't assume, zero intercept.

Heteroscedasticity fixes

- Weighted least squares; robust SEs.
- Transform the response.
- Diagnose via residual plots.

Leverage points

- Extreme x with pull on the line.
- Cook's distance combines leverage + residual.
- Investigate, don't auto-delete.

Prediction intervals

- Wider than confidence intervals.
- Cover future observations, not the mean.
- Include residual variance.
- What managers actually want.
- Distinguish the two precisely.

Influential observations

- Leverage $\times$ discrepancy = influence.
- Cook's $D > 1$ deserves investigation.
- DFBETAS flag coefficient-specific influence.
- Deleting data needs justification.
- Robust regression resists automatically.

8. Nonparametric Methods

Learning Objectives

- Know when parametric tests fail
- Apply rank-based tests

- Use when: small samples, skewed data, ordinal scales — no normality assumed.
- Sign test: median of paired differences.
- Wilcoxon signed-rank: paired samples using ranks; Mann-Whitney U: two independent samples.
- Kruskal-Wallis: nonparametric ANOVA; Spearman rank correlation for monotonic association.

Key Points to Remember

- Ranks sacrifice some power for robustness.
- Nonparametric = fewer assumptions, not assumption-free.

Exam Pointer

Very important for exams: Choose parametric vs nonparametric for three datasets.

Rank tests intuition

- Replace values by ranks, test ranks.
- Outliers lose their leverage.
- Small price in power for big robustness.

Sign test example

- Count + vs - differences; binomial test.
- No distributional assumptions at all.
- Low power but bulletproof.

Spearman vs Pearson

- Spearman: monotonic; Pearson: linear.
- Spearman robust to outliers.
- Choose by relationship shape.

Kendall's tau

- Concordant vs discordant pairs.
- Interpretable probability.

- Alternative to Spearman.

Wilcoxon signed-rank

- Paired nonparametric alternative.
- Uses ranks of differences.
- More powerful than the sign test.
- Assumes symmetric differences.
- Know when each rank test applies.

Mann-Whitney details

- U statistic from pairwise comparisons.
- Tests stochastic dominance.
- Ties need correction.
- Effect size: rank-biserial correlation.
- Report with medians and IQR.

9. Likelihood Ratio Tests and Goodness of Fit

Learning Objectives

- Construct likelihood ratio tests
- Test distributional fit

- Likelihood ratio: $\lambda = L(H_0)/L(\text{unrestricted})$; $-2 \log \lambda \sim \text{chi-square}$ — the general engine for nested hypotheses.
- Chi-square goodness of fit: compare observed vs expected counts; $df = \text{categories} - 1 - \text{estimated parameters}$.
- Test of independence: contingency tables (e.g. treatment vs recovery).

Example

Die-rolling: 60 rolls, expected 10 per face — chi-square tests fairness.

Key Points to Remember

- LRT is the master key: t, F and chi-square tests are special cases.
- Expected counts should exceed 5 for chi-square validity.

Exam Pointer

Very important for exams: Run a chi-square independence test on a 2x2 table. 

Contingency table example

- Treatment vs recovery counts.
- Expected = $\text{row} \times \text{col} / \text{total}$ under independence.
- Chi-square sums standardized squared gaps.

Degrees of freedom intuition

- Independent pieces of information.
- Estimating parameters consumes them.
- Explains $n-1$ in sample variance.

Neyman-Pearson lemma

- LRT is most powerful for simple hypotheses.
- Theoretical crown jewel.
- Justifies the LRT's centrality.

Score test

- LRT's sibling: uses gradient at null.
- No alternative fit needed.
- Rao's contribution.

Chi-square test of independence

- Contingency tables: rows vs columns.
- Expected counts assume independence.
- $df = (r-1)(c-1)$.
- Small expected counts need Fisher's exact.
- A workhorse of applied statistics.

Wald test

- $(\text{estimate} - \text{null}) / \text{SE}$, squared.
- Asymptotically chi-square.
- Simplest of the trinity.
- Can misbehave at boundaries.
- LRT vs Wald vs Score.

10. Bayesian Inference Basics

Learning Objectives

- Explain the Bayesian philosophy
- Compute simple posteriors

- Frequentist: parameters fixed, data random. Bayesian: data fixed, parameters have distributions.
- Bayes' theorem: posterior proportional to likelihood x prior.
- Conjugate priors (beta-binomial) give closed-form posteriors.
- Credible intervals: direct probability statements about parameters — "95% probability μ is in (a,b) ".

Example

Coin bias: uniform prior + 7 heads in 10 flips $\rightarrow$ posterior $\text{Beta}(8,4)$; posterior mean $8/12 = 0.67$.

Key Points to Remember

- Priors let past knowledge in; with big data, priors wash out.
- Credible $\neq$ confidence interval — different philosophy.

Exam Pointer

Very important for exams: Contrast Bayesian and frequentist interpretations of a 95% interval.

Prior sensitivity

- Try several priors; report stability.
- With big data, priors wash out.
- Sensitivity analysis is honest Bayes.

MCMC in one paragraph

- Sample from posterior via Markov chains.
- Gibbs and Metropolis-Hastings algorithms.
- Convergence diagnostics required.

Empirical Bayes

- Estimate priors from the data.
- James-Stein: shrink towards the mean.

- Beats MLE in high dimensions.

Hierarchical Bayes

- Priors with hyperpriors.
- Borrow strength across groups.
- Modern standard for complex models.

Conjugate priors

- Beta-binomial: posterior stays Beta.
- Normal-normal with known variance.
- Simplify computation elegantly.
- Modern MCMC reduced their necessity.
- Still the teaching standard.

Posterior predictive checks

- Simulate data from the fitted model.
- Compare with observed data.
- Misfit reveals model flaws.
- Bayesian goodness-of-fit.
- Visual checks beat single numbers.

11. Resampling: Bootstrap and Permutation

Learning Objectives

- Explain the bootstrap
- Use permutation tests

- Bootstrap: resample the data with replacement thousands of times; the spread of resampled statistics estimates the sampling distribution.
- Bootstrap CIs need no formula — works for medians, correlations, any statistic.
- Permutation test: shuffle group labels to build the null distribution — exact p-values with no distributional assumptions.

Key Points to Remember

- Resampling replaces maths with computation.
- Bootstrap assumes the sample represents the population.

Exam Pointer

Very important for exams: Outline bootstrap steps for the median's CI.

Jackknife vs bootstrap

- Jackknife: leave-one-out, older, simpler.
- Bootstrap: more general, needs computation.
- Both estimate standard errors.

Permutation test example

- Shuffle labels 10,000 times; compare statistic.
- Exact under exchangeability.
- No t-distribution needed.

Bootstrap failure modes

- Fails for extremes and dependent data.
- Block bootstrap for time series.
- Know the limits.

Exam revision checklist

- MLE derivation steps, CI formulas, error types table.
- Test selection guide, ANOVA table completion.
- One Bayesian vs frequentist essay.

Bootstrap confidence intervals

- Percentile method: simplest.
- BCa corrects bias and skew.
- Studentized: most accurate, needs SE.
- Thousands of resamples typical.
- Plot the bootstrap distribution.

Randomization tests

- Shuffle treatment labels, recompute.
- Exact p-values, no distributions.
- Fisher's original idea.
- Computationally intensive historically.
- Gold standard when feasible.

Glossary of Key Terms

Learning Objectives

- Revise all key terms quickly before exams

Term	Meaning
Statistical inference	drawing conclusions about a population from sample data, quantifying the uncertainty of those conclusions.

Exam Pointer

Very important for exams: Revise this glossary the night before the paper.

Practice Questions for Exams

Learning Objectives

- Test your preparation with these exam-style questions

Short questions with model answers

Example

Q: Explain: distinguish populations from samples Model answer: Larger $n \rightarrow$ smaller SE $\rightarrow$ sharper inference. CLT lets us use normal theory on non-normal data.

Example

Q: Explain: contrast estimation with hypothesis testing Model answer: Larger $n \rightarrow$ smaller SE $\rightarrow$ sharper inference. CLT lets us use normal theory on non-normal data.

Example

Q: Explain: define desirable estimator properties Model answer: Unbiased + low variance = good estimator. MLE: write likelihood, log it, differentiate, solve.

Example

Q: Explain: derive MLEs for common distributions Model answer: Unbiased + low variance = good estimator. MLE: write likelihood, log it, differentiate, solve.

Example

Q: Explain: construct confidence intervals Model answer: The t-distribution replaces z when sigma is estimated. Interpretation wording is examined precisely.

Example

Q: Explain: interpret them correctly Model answer: The t-distribution replaces z when sigma is estimated. Interpretation wording is examined precisely.

Example

Q: Explain: set up null and alternative hypotheses Model answer: p-value is about the data under H_0 , not $P(H_0 \text{ is true})$. Never "accept" H_0 — we fail to reject it.

Example

Q: Explain: explain p-values and error types Model answer: p-value is about the data under H_0 , not $P(H_0 \text{ is true})$. Never "accept" H_0 — we fail to reject it.

Example

Q: Explain: choose the right test for the design Model answer: Match the test to the design — pairing changes everything. Check assumptions before trusting p-values.

Example

Q: Explain: run z, t and chi-square tests Model answer: Match the test to the design — pairing changes everything. Check assumptions before trusting p-values.

Example

Q: Explain: explain the ANOVA logic Model answer: ANOVA partitions total variation into explained + unexplained. Post-hoc only after significant F.

Example

Q: Explain: run one-way ANOVA and post-hoc tests Model answer: ANOVA partitions total variation into explained + unexplained. Post-hoc only after significant F.

Example

Q: Explain: test regression coefficients Model answer: A significant slope with tiny R-squared is real but weak. Prediction intervals are wider than confidence intervals.

Example

Q: Explain: build prediction intervals Model answer: A significant slope with tiny R-squared is real but weak. Prediction intervals are wider than confidence intervals.

Example

Q: Explain: know when parametric tests fail Model answer: Ranks sacrifice some power for robustness. Nonparametric = fewer assumptions, not assumption-free.

Example

Q: Explain: apply rank-based tests Model answer: Ranks sacrifice some power for robustness. Nonparametric = fewer assumptions, not assumption-free.

Example

Q: Explain: construct likelihood ratio tests Model answer: LRT is the master key: t, F and chi-square tests are special cases. Expected counts should exceed 5 for chi-square validity.

Example

Q: Explain: test distributional fit Model answer: LRT is the master key: t, F and chi-square tests are special cases. Expected counts should exceed 5 for chi-square validity.

Example

Q: Explain: explain the Bayesian philosophy Model answer: Priors let past knowledge in; with big data, priors wash out. Credible != confidence interval — different philosophy.

Example

Q: Explain: compute simple posteriors Model answer: Priors let past knowledge in; with big data, priors wash out. Credible != confidence interval — different philosophy.

Example

Q: Explain: explain the bootstrap Model answer: Resampling replaces maths with computation. Bootstrap assumes the sample represents the population.

Example

Q: Explain: use permutation tests Model answer: Resampling replaces maths with computation. Bootstrap assumes the sample represents the population.

Long questions

- Write a detailed note on Introduction to Statistical Inference.
- Write a detailed note on Point Estimation.
- Write a detailed note on Interval Estimation.
- Write a detailed note on Hypothesis Testing: Foundations.
- Write a detailed note on Tests for Means and Proportions.
- Write a detailed note on Analysis of Variance (ANOVA).
- Write a detailed note on Regression Inference.
- Write a detailed note on Nonparametric Methods.
- Write a detailed note on Likelihood Ratio Tests and Goodness of Fit.
- Write a detailed note on Bayesian Inference Basics.

Long questions with model answer outlines

Example

Q: Write a detailed note on Introduction to Statistical Inference. Answer outline: • Population: the whole group of interest; sample: the observed subset; parameter: population quantity (μ , σ); statistic: sample quantity ($\bar{x}$, s). • Two branches: estimation (point + interval) and hypothesis testing. • Sampling matters: random sampling justifies inference; biased sampling invalidates it. • The sampling distribution of $\bar{x}$ centres on μ with standard error $\sigma/\sqrt{n}$. • Central limit theorem: for large n , $\bar{x}$ is approximately normal regardless of population shape — the engine of classical inference.

Example

Q: Write a detailed note on Point Estimation. Answer outline: • Method of moments: equate sample moments to population moments and solve. • Maximum likelihood (MLE): choose the parameter maximizing the likelihood of observed data. • MLEs are consistent, asymptotically efficient and invariant under transformation — very important for exams: MLE is the default method in modern statistics. • $MSE = \text{variance} + \text{bias}^2$. • Slight bias can buy big variance cuts.

Example

Q: Write a detailed note on Interval Estimation. Answer outline: • Confidence interval: estimate $\pm$ margin of error; margin = critical value $\times$ SE. • One mean (σ known): $\bar{x} \pm z_{(\alpha/2)} \sigma/\sqrt{n}$; σ unknown: use t with $n-1$ df. • Proportions: $\hat{p} \pm z \sqrt{\hat{p}(1-\hat{p})/n}$. • Correct interpretation: "95% confident" means 95% of such intervals cover the parameter — NOT "95% probability the parameter is inside this one". • Width shrinks with larger n and lower confidence; grows with more variability.

Example

Q: Write a detailed note on Hypothesis Testing: Foundations. Answer outline: • H_0 (null): the default/no-effect claim; H_1 (alternative): what we seek evidence for. • Test statistic measures distance from H_0 ; p-value = probability of data this extreme under H_0 . • Reject H_0 if $p < \alpha$ (significance level, often 0.05). • Power = $1 - \beta$: probability of detecting a real effect; rises with n , effect size and α . • Practical vs statistical significance: huge n can make trivial effects "significant".

Example

Q: Write a detailed note on Tests for Means and Proportions. Answer outline: • Paired vs independent: pairing (before-after) removes subject variability — more powerful. • Assumptions: approximate normality (or large n via CLT), independent observations. • Welch: no equal-variance assumption. • Safer default in practice. • Know when each applies.

Example

Q: Write a detailed note on Analysis of Variance (ANOVA). Answer outline: • Problem: comparing 3+ means with t-tests inflates Type I error — ANOVA tests all at once. • Logic: $F = \text{between-group variance} / \text{within-group variance}$; large F means real group differences. • Assumptions: normality within groups, equal variances, independence. • Significant F only says "some differ" — post-hoc tests (Tukey) find which pairs. • Two-way ANOVA adds a second factor and their interaction.

Example

Q: Write a detailed note on Regression Inference. Answer outline: • Model: $y = \beta_0 + \beta_1 x + \epsilon$; least squares minimizes sum of squared residuals. • Inference on slope: $t = b_1 / \text{SE}(b_1)$; tests whether x really affects y . • R-squared: proportion of variation explained; adjusted R-squared penalizes extra predictors. • Confidence interval for the mean response vs wider prediction interval for a new observation. • Residual plots reveal non-linearity, heteroscedasticity and outliers.

Example

Q: Write a detailed note on Nonparametric Methods. Answer outline: • Use when: small samples, skewed data, ordinal scales — no normality assumed. • Sign test: median of paired differences. • Wilcoxon signed-rank: paired samples using ranks; Mann-Whitney U: two independent samples. • Kruskal-Wallis: nonparametric ANOVA; Spearman rank correlation for monotonic association. • Replace values by ranks, test ranks.

Example

Q: Write a detailed note on Likelihood Ratio Tests and Goodness of Fit. Answer outline: • Likelihood ratio: $\lambda = L(H_0)/L(\text{unrestricted})$; $-2 \log \lambda \sim \text{chi-square}$ — the general engine for nested hypotheses. • Chi-square goodness of fit: compare observed vs expected counts; $df = \text{categories} - 1 - \text{estimated parameters}$. • Test of independence: contingency tables (e.g. treatment vs recovery). • Treatment vs recovery counts. • Expected = $\text{row} \times \text{col} / \text{total}$ under independence.

Example

Q: Write a detailed note on Bayesian Inference Basics. Answer outline: • Frequentist: parameters fixed, data random. Bayesian: data fixed, parameters have distributions. • Bayes' theorem: posterior proportional to likelihood $\times$ prior. • Conjugate priors (beta-binomial) give closed-form posteriors. • Credible intervals: direct probability statements about parameters — "95% probability μ is in (a,b) ". • Try several priors; report stability.

Exam Pointer

Very important for exams: Practice writing answers within the exam time limit.

MCQ Practice Test

Learning Objectives

- Test your recall of the key points from every chapter

Q1. Which of the following was listed as a key point of 'Introduction to Statistical Inference' in this handout?

- A. Unbiased + low variance = good estimator.
- B. The t-distribution replaces z when σ is estimated.
- C. MLE: write likelihood, log it, differentiate, solve.
- D. Larger $n \rightarrow$ smaller SE $\rightarrow$ sharper inference.

Q2. Which of the following was listed as a key point of 'Introduction to Statistical Inference' in this handout?

- A. CLT lets us use normal theory on non-normal data.
- B. Unbiased + low variance = good estimator.
- C. MLE: write likelihood, log it, differentiate, solve.
- D. The t-distribution replaces z when σ is estimated.

Q3. Which of the following was listed as a key point of 'Point Estimation' in this handout?

- A. The t-distribution replaces z when σ is estimated.

B. Larger n $\rightarrow$ smaller SE $\rightarrow$ sharper inference.

C. Unbiased + low variance = good estimator.

D. CLT lets us use normal theory on non-normal data.

Q4. Which of the following was listed as a key point of 'Point Estimation' in this handout? 

A. CLT lets us use normal theory on non-normal data.

B. The t-distribution replaces z when σ is estimated.

C. Larger n $\rightarrow$ smaller SE $\rightarrow$ sharper inference.

D. MLE: write likelihood, log it, differentiate, solve.

Q5. Which of the following was listed as a key point of 'Interval Estimation' in this handout?

A. Unbiased + low variance = good estimator.

B. The t-distribution replaces z when σ is estimated.

C. Larger n $\rightarrow$ smaller SE $\rightarrow$ sharper inference.

D. CLT lets us use normal theory on non-normal data.

Q6. Which of the following was listed as a key point of 'Interval Estimation' in this handout?

A. Interpretation wording is examined precisely.

B. CLT lets us use normal theory on non-normal data.

C. Unbiased + low variance = good estimator.

D. Larger n $\rightarrow$ smaller SE $\rightarrow$ sharper inference.

Q7. Which of the following was listed as a key point of 'Hypothesis Testing: Foundations' in this handout?

A. Unbiased + low variance = good estimator.

B. Larger n $\rightarrow$ smaller SE $\rightarrow$ sharper inference.

C. p-value is about the data under H_0 , not $P(H_0 \text{ is true})$.

D. CLT lets us use normal theory on non-normal data.

Q8. Which of the following was listed as a key point of 'Hypothesis Testing: Foundations' in this handout?

A. Unbiased + low variance = good estimator.

B. Larger n $\rightarrow$ smaller SE $\rightarrow$ sharper inference.

C. Never "accept" H_0 — we fail to reject it.

D. CLT lets us use normal theory on non-normal data.

Q9. Which of the following was listed as a key point of 'Tests for Means and Proportions' in this handout?

- A. Unbiased + low variance = good estimator.
- B. CLT lets us use normal theory on non-normal data.
- C. Larger n -> smaller SE -> sharper inference.
- D. Match the test to the design — pairing changes everything.

Q10. Which of the following was listed as a key point of 'Tests for Means and Proportions' in this  handout?

- A. Unbiased + low variance = good estimator.
- B. Larger n -> smaller SE -> sharper inference.
- C. CLT lets us use normal theory on non-normal data.
- D. Check assumptions before trusting p-values.

Q11. Which of the following was listed as a key point of 'Analysis of Variance (ANOVA)' in this handout?

- A. CLT lets us use normal theory on non-normal data.
- B. Larger n -> smaller SE -> sharper inference.
- C. ANOVA partitions total variation into explained + unexplained.
- D. Unbiased + low variance = good estimator.

Q12. Which of the following was listed as a key point of 'Analysis of Variance (ANOVA)' in this handout?

- A. CLT lets us use normal theory on non-normal data.
- B. Larger n -> smaller SE -> sharper inference.
- C. Post-hoc only after significant F.
- D. Unbiased + low variance = good estimator.

Q13. Which of the following was listed as a key point of 'Regression Inference' in this handout?

- A. A significant slope with tiny R-squared is real but weak.
- B. Unbiased + low variance = good estimator.
- C. Larger n -> smaller SE -> sharper inference.
- D. CLT lets us use normal theory on non-normal data.

Q14. Which of the following was listed as a key point of 'Regression Inference' in this handout?

- A. CLT lets us use normal theory on non-normal data.
- B. Unbiased + low variance = good estimator.
- C. Prediction intervals are wider than confidence intervals.
- D. Larger n -> smaller SE -> sharper inference.

Q15. Which of the following was listed as a key point of 'Nonparametric Methods' in this handout? 

- A. Unbiased + low variance = good estimator.
- B. Ranks sacrifice some power for robustness.
- C. CLT lets us use normal theory on non-normal data.
- D. Larger $n \rightarrow$ smaller SE $\rightarrow$ sharper inference.

Q16. Which of the following was listed as a key point of 'Nonparametric Methods' in this handout?

- A. Larger $n \rightarrow$ smaller SE $\rightarrow$ sharper inference.
- B. CLT lets us use normal theory on non-normal data.
- C. Nonparametric = fewer assumptions, not assumption-free.
- D. Unbiased + low variance = good estimator.

Q17. Which of the following was listed as a key point of 'Likelihood Ratio Tests and Goodness of Fit' in this handout?

- A. Larger $n \rightarrow$ smaller SE $\rightarrow$ sharper inference.
- B. CLT lets us use normal theory on non-normal data.
- C. LRT is the master key: t , F and chi-square tests are special cases.
- D. Unbiased + low variance = good estimator.

Q18. Which of the following was listed as a key point of 'Likelihood Ratio Tests and Goodness of Fit' in this handout?

- A. Larger $n \rightarrow$ smaller SE $\rightarrow$ sharper inference.
- B. CLT lets us use normal theory on non-normal data.
- C. Expected counts should exceed 5 for chi-square validity.
- D. Unbiased + low variance = good estimator.

Q19. Which of the following was listed as a key point of 'Bayesian Inference Basics' in this handout?

- A. Unbiased + low variance = good estimator.
- B. CLT lets us use normal theory on non-normal data.
- C. Larger $n \rightarrow$ smaller SE $\rightarrow$ sharper inference.
- D. Priors let past knowledge in; with big data, priors wash out.

Q20. Which of the following was listed as a key point of 'Bayesian Inference Basics' in this handout?

- A. Unbiased + low variance = good estimator.

- B. CLT lets us use normal theory on non-normal data.
- C. Larger n $\rightarrow$ smaller SE $\rightarrow$ sharper inference.
- D. Credible $\neq$ confidence interval — different philosophy.

Q21. Which of the following was listed as a key point of 'Resampling: Bootstrap and Permutation' in this handout? 

- A. Larger n $\rightarrow$ smaller SE $\rightarrow$ sharper inference.
- B. Resampling replaces maths with computation.
- C. Unbiased + low variance = good estimator.
- D. CLT lets us use normal theory on non-normal data.

Q22. Which of the following was listed as a key point of 'Resampling: Bootstrap and Permutation' in this handout?

- A. Larger n $\rightarrow$ smaller SE $\rightarrow$ sharper inference.
- B. Unbiased + low variance = good estimator.
- C. Bootstrap assumes the sample represents the population.
- D. CLT lets us use normal theory on non-normal data.

Answer Key

1. D 2. A 3. C 4. D 5. B 6. A 7. C 8. C 9. D 10. D 11. C 12. C 13. A 14. C 15. B 16. C 17. C 18. C 19. D 20. D 21. B 22. C

Exam Pointer

Very important for exams: Attempt all MCQs without looking back, then check the answer key.

Exam Strategy Guide

Learning Objectives

- Attempt papers with a winning strategy

Before the exam

- Revise the Key Points boxes and glossary first — they carry the highest marks per minute.
- Practice one past paper under timed conditions; examiners repeat question styles.
- Sleep well the night before — an exhausted brain recalls far less.

Attempting the paper

- Read the whole paper first; start with your strongest question to build momentum.
- Allocate time by marks: a 10-mark question deserves roughly double a 5-mark one.
- Structure long answers: definition, explanation, example/diagram, conclusion.

Presentation matters

- Use headings, underline key terms, and draw labelled diagrams where possible.
- Attempt all questions — partial credit beats a blank page.
- Review your paper in the last 10 minutes for missed parts.

Exam Pointer

Very important for exams: Very important for exams: neat, structured answers consistently score higher than longer messy ones.