

VU LEARNING HUB

vulearninghub.xyz

HIGHLIGHTED HANDOUT

STA631

Inferential Statistics

STA631 by VU Learning HUB

Prepared for VU students • Free study resources

<https://vulearninghub.xyz>

STA631

Inferential Statistics

Lecture-Wise Study Handout
Virtual University of Pakistan

22 lectures • prepared for exam study 

Table of Contents

1. From Samples to Populations
2. Point Estimation
3. Confidence Intervals
4. Hypothesis Testing Logic
5. Comparing Groups
6. ANOVA
7. Regression Inference
8. Power and Effect Sizes
9. Nonparametrics and a Bayesian Glimpse
10. Survey Sampling Design
11. Variances and Proportions in Depth
12. Multiple Comparisons and Two-Way Designs
13. Transformations, Diagnostics, and Logistic Regression
14. Bootstrap and Meta-Analysis
15. Reporting and Final Drills
16. Solved Problems: Estimation
17. Solved Problems: Hypothesis Tests
18. Solved Problems: ANOVA and Regression
19. The t, Chi-Square, and F Distributions
20. Question Bank with Answer Outlines
21. Nonparametric Methods
22. Software Output and Reporting

About This Course

STA631 – Inferential Statistics studies how to draw conclusions about populations from samples: sampling distributions, point and interval estimation, hypothesis testing, ANOVA, regression inference, and nonparametric methods. This handout emphasizes the logic (why the procedures work), the computations (worked examples), and the interpretation (what results actually mean) — with Pakistani data examples.

Each lecture gives learning objectives, core explanations with definitions, worked examples, exam key points, a summary, and review questions. 

Lecture 1: From Samples to Populations

Learning Objectives

- Define the inferential framework
- Apply the CLT

Descriptive statistics summarizes data; inferential statistics generalizes from sample to population. The bridge is probability: if we know how statistics vary across samples (sampling distributions), we can quantify uncertainty about population parameters. The population (the whole — parameters like μ , σ , fixed but unknown) vs the sample (the observed subset — statistics like $\bar{x}$, s , random before observation).

The sampling distribution of $\bar{x}$: mean μ , standard deviation $\sigma/\sqrt{n}$ (the standard error — shrinking with $\sqrt{n}$), and — by the Central Limit Theorem — approximately normal for large n regardless of the population's shape. The CLT is the most important theorem in applied statistics: it justifies normal-based inference for means from almost any data.

Example: Pakistani household incomes (skewed population) — samples of $n=100$ give $\bar{x}$ values normally distributed around μ with $SE = \sigma/10$. Exam intuition: bigger $n \rightarrow$ narrower sampling distribution $\rightarrow$ more precise inference. The standard error (not the standard deviation) measures the statistic's precision.

Key Definitions

Standard error is defined as the standard deviation of a statistic's sampling distribution — measuring its precision ($\sigma/\sqrt{n}$ for $\bar{x}$).

Examples

- SE: $\sigma=5000$, $n=100 \rightarrow SE=500$: $\bar{x}$ within ± 1000 of μ about 95% of the time.
- CLT: skewed income data, yet $\bar{x} \sim$ normal for $n=100$.

Key Points — very important for exams

Inference = generalizing via sampling distributions.

Population parameters fixed/unknown; sample statistics random/known.

$\bar{x}$: mean μ , SE $\sigma/\sqrt{n}$, CLT $\rightarrow$ normal for large n .

Standard error measures precision — not standard deviation.

Note:

Note: 'explain why the CLT matters' is a perennial short question — it's the license for normal-based inference.

Review Questions

? Why does SE shrink with $\sqrt{n}$, not n ?

? A sample of $n=4$ from skewed data — is $\bar{x}$ normal?

Section Summary

Sampling distributions turn data into evidence: the CLT makes the sample mean's behavior predictable, and precision grows with $\sqrt{n}$.

Exam Pointers

- Distinguish population, sample, parameter, statistic.
- State and explain the Central Limit Theorem.

Lecture 2: Point Estimation

Learning Objectives

- Evaluate estimators
- Derive MLEs

Point estimation: a single best guess — $\bar{x}$ for μ , s^2 for σ^2 , p for p . Desirable properties: unbiasedness ($E[\hat{\theta}] = \theta$ — right on average), efficiency (minimum variance among unbiased — $\bar{x}$ beats the median for normal data), consistency (converges to θ as n grows), and sufficiency (uses all sample information). The sample variance uses $n-1$ (Bessel's correction) to be unbiased.

Methods: method of moments (equate sample and population moments — simple, often inefficient) and maximum likelihood (choose θ maximizing the data's likelihood — the workhorse: efficient, asymptotically normal, invariant). MLE example: for Bernoulli(p), $p = \bar{x}$ — the sample proportion; for Normal(μ, σ^2), $\mu = \bar{x}$, $\sigma^2 = (n-1)/n \cdot s^2$ (MLE is biased — the n vs $n-1$ story).

Pakistani example: estimating the unemployment rate — p from the Labour Force Survey, with its SE $\sqrt{p(1-p)/n}$. Exam skill: derive a simple MLE (write likelihood, log, differentiate, solve) and check bias.

Key Definitions

Maximum likelihood estimation is defined as choosing the parameter value maximizing the probability of the observed data.

Examples

- Bessel: dividing by $n-1$ unbias s^2 — the degree-of-freedom cost of estimating μ .
- LFS: $p=0.07$ unemployment, $n=50,000 \rightarrow SE \approx 0.001$ — precise.

Key Points — very important for exams

Properties: unbiased, efficient, consistent, sufficient.

s^2 uses $n-1$ for unbiasedness — know why.

MLE: maximize likelihood — the standard method.

Derive simple MLEs; check bias mechanically.

Review Questions

- ? Is $\bar{x}$ unbiased for μ ? Prove it.
- ? Why is MLE's $\hat{\sigma}^2$ biased?

Section Summary

Point estimation picks the best single guess: judge by unbiasedness and efficiency, compute by maximum likelihood.

Exam Pointers

- What makes a good estimator?
- Derive the MLE for Bernoulli(p).

Lecture 3: Confidence Intervals

Learning Objectives

- Construct intervals
- Plan sample sizes

Interval estimation: a range with stated confidence — $\bar{x} \pm z^* \cdot \sigma / \sqrt{n}$ (σ known) or $\bar{x} \pm t^* \cdot s / \sqrt{n}$ (σ unknown — the realistic case). The t-distribution (heavier tails than normal, $df = n-1$) corrects for estimating σ ; as n grows, $t \rightarrow z$. Interpretation (the exam trap): '95% confident the interval covers μ ' — the confidence is in the procedure (95% of such intervals cover), not a 95% probability for this specific interval.

Sample-size planning: $n = (z^* \cdot \sigma / m)^2$ for margin m — precision costs quadratically: halving the margin quadruples n . For proportions: $\bar{p} \pm z^* \sqrt{\bar{p}(1-\bar{p})/n}$, with $n = (z^*/2m)^2$ worst case ($p=0.5$). Pakistani polling: a $\pm 3\%$ margin at 95% needs $\sim 1,067$ respondents — the arithmetic behind survey headlines.

Worked example: $n=25$ students, $\bar{x}=68$, $s=10 \rightarrow 95\%$ CI: $68 \pm 2.064 \cdot (10/5) = (63.9, 72.1)$. Report: 'we are 95% confident the mean score is between 63.9 and 72.1.' Conditions: random sample, roughly normal data (or $n \geq 30$ via CLT).

Key Definitions

Confidence interpretation is defined as the correct interpretation: 95% of intervals from repeated sampling cover the parameter — not $P(\text{parameter in this interval}) = 0.95$.

Examples

- t vs z : $n=10 \rightarrow t^*=2.262$ vs $z^*=1.96$ — small samples pay wider intervals.
- Polling: $\pm 3\%$ at 95% $\rightarrow n \approx 1067$ — the survey industry's standard.

Key Points — very important for exams

$\bar{x} \pm t^* \cdot s / \sqrt{n}$ (σ unknown); t corrects small samples.

Confidence = procedure reliability, not interval probability — exam trap.

$n = (z^* \sigma / m)^2$: precision costs quadratically.

Check conditions: randomness, approximate normality.

Review Questions

? Why is the t -interval wider than z ?

? A 90% vs 95% interval: which is wider?

Section Summary

Intervals quantify precision honestly: center at the estimate, width from SE and confidence, interpretation about the procedure.

Exam Pointers

→ Interpret a 95% confidence interval correctly.

→ How large a sample for $\pm 2\%$ margin?

Lecture 4: Hypothesis Testing Logic

Learning Objectives

- Test hypotheses
- Interpret p -values correctly

Hypothesis testing logic: state H_0 (the null — no effect, $\mu = \mu_0$) and H_a (the alternative — $\mu \neq \mu_0$, or one-sided); compute the test statistic (how many SEs the data is from H_0 : $z = (\bar{x} - \mu_0) / (\sigma / \sqrt{n})$); find the p -value (probability of data this extreme if H_0 true); reject H_0 if $p < \alpha$ (usually 0.05). The p -value is NOT $P(H_0 \text{ true})$ — the exam's most-tested misconception.

Errors: Type I (rejecting true H_0 — rate α) vs Type II (missing real effects — rate β); power = $1 - \beta$. Trade-off: lowering α raises β . Consequences decide α : drug safety (low α) vs exploratory research (tolerate higher α). Pakistani example: testing if a fertilizer raises wheat yield — H_0 : no effect; Type I wastes farmers' money, Type II misses real gains.

One-sample t -test worked: $H_0: \mu = 70$, $n = 25$, $\bar{x} = 74$, $s = 8 \rightarrow t = 4 / (8 / 5) = 2.5$, $df = 24$, $p \approx 0.02 \rightarrow$ reject at 5%: evidence the mean exceeds 70. Always report the effect size too — significance $\neq$ importance.

Key Definitions

p -value is defined as $P(\text{data this extreme} \mid H_0 \text{ true})$ — not $P(H_0 \text{ true} \mid \text{data})$.

Examples

■ $p=0.02$: reject H_0 at 5% — but check the effect size (4 points — is it meaningful?).

■ Fertilizer: Type I wastes money; Type II misses yield — asymmetric costs.

Key Points — very important for exams

H_0 vs H_a → test statistic → p-value → decide at α .

p-value $\neq$ $P(H_0 \text{ true})$: the #1 misconception.

Type I (α) vs Type II (β); power = $1-\beta$; trade-off.

Report effect sizes: significance isn't importance.

Review Questions

? $p=0.06$ at $\alpha=0.05$ — your conclusion and its limits?

? When should α be 0.01 vs 0.10?

Section Summary

Testing is structured skepticism: assume H_0 , measure surprise, and decide — while minding errors and effect sizes.

Exam Pointers

→ What is a p-value? State the common misinterpretation.

→ Distinguish Type I and Type II errors.

Lecture 5: Comparing Groups

Learning Objectives

- Choose comparison tests
- Compute and interpret

Two-sample comparisons: independent samples (two groups — treatment vs control) use the two-sample t-test (pooled variance if σ s equal, Welch's if not); paired data (before/after on same subjects) use the paired t-test on differences — far more powerful when pairing removes variability. Pakistani example: wheat yields, same farms before/after fertilizer — paired; two different districts — independent.

Categorical data: the chi-square test — goodness-of-fit (do observed counts match expected? e.g., are births uniform across seasons?) and independence (is smoking related to lung disease? — contingency tables). Test statistic $\sum (O-E)^2/E \sim \chi^2$; conditions: expected counts ≥ 5 .

Proportions: two-proportion z-test (is the pass rate higher in school A than B?). Worked: pass rates 80/100 vs 70/100 → $z \approx 1.63$, $p \approx 0.10$ — not significant at 5%: the 10-point gap could be chance.

Exam skill: choose the right test for the design — paired vs independent is the classic fork.

Key Definitions

Paired test is defined as analyzing before/after differences on the same subjects — removing between-subject variability.

Examples

- Paired: patients' BP before/after drug — each patient their own control.
- χ^2 : 60/40 vs expected 50/50 → $\chi^2=4$, $p<0.05$ — significant deviation.

Key Points — very important for exams

Paired (same subjects) vs independent (two groups) — design decides.

Paired tests are powerful: pairing removes noise.

χ^2 : goodness-of-fit and independence for counts.

Proportions via z; always check expected counts ≥ 5 .

Review Questions

- ? Crossover trial data — paired or independent?
- ? A 2×2 table with expected count 3 — what now?

Section Summary

Match the test to the design: pair when you can, chi-square for counts, and let the design — not habit — choose.

Exam Pointers

- When is a paired test better than independent?
- What does a chi-square test of independence do?

Lecture 6: ANOVA

Learning Objectives

- Run ANOVA
- Interpret tables

ANOVA compares ≥ 3 means: H_0 : all equal vs H_A : some differ. Logic: $F = (\text{between-group variance})/(\text{within-group variance})$ — if group means spread more than chance variation allows, reject. One-way ANOVA example: wheat yields under 3 fertilizers — F large → some fertilizer differs; then post-hoc tests (Tukey) find which pairs differ (ANOVA alone doesn't say which).

Assumptions: normality within groups, equal variances (homoscedasticity), independence. Violations: Welch's ANOVA (unequal variances), Kruskal-Wallis (non-normal — the nonparametric alternative). Two-way ANOVA adds a second factor and their interaction — e.g., fertilizer × irrigation: does

fertilizer's effect depend on water?

Worked skeleton: $SST = SSB + SSW$; $df: (k-1), (n-k)$; $F = MSB/MSW$. Exam skill: complete a partial ANOVA table and interpret — the standard 10-mark question. Pakistani example: comparing mean incomes across four provinces — one-way ANOVA, then Tukey.

Key Definitions

ANOVA is defined as partitioning total variation into between-group and within-group — the F-test for ≥ 3 means.

Examples

- Fertilizers: $F=8.2, p=0.002 \rightarrow$ means differ; Tukey: $A>B=C$.
- Interaction: fertilizer helps only with irrigation — the two-way insight.

Key Points — very important for exams

F = between/within variance; large $F \rightarrow$ means differ.

ANOVA says 'some differ'; Tukey says which.

Assumptions: normality, equal variance, independence.

Two-way: second factor + interaction — the richer design.

Review Questions

- ? Why not just do three t-tests? (Hint: error inflation.)
- ? When is Kruskal-Wallis needed?

Section Summary

ANOVA tests many means at once without inflating error: partition variance, F-test, then locate differences.

Exam Pointers

- $\rightarrow$ What is ANOVA's logic?
- $\rightarrow$ Complete an ANOVA table and interpret.

Lecture 7: Regression Inference

Learning Objectives

- Infer about regression
- Diagnose models

Simple linear regression: $y = \beta_0 + \beta_1 x + \varepsilon$ — least squares minimizes $\sum \text{residuals}^2$; $\beta_1 = r(s_y/s_x)$, $\beta_0 = \bar{y} - \beta_1 \bar{x}$. Inference: t-test for $H_0: \beta_1 = 0$ (is there a linear relationship?), CI for the slope, and

R^2 (fraction of variance explained — descriptive, not inferential). Residual analysis checks assumptions: linearity, constant variance, normality, independence — plot residuals vs fitted.

Multiple regression: $y = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k + \varepsilon$ — each coefficient is the effect holding others constant (ceteris paribus). Pitfalls: multicollinearity (correlated predictors — inflated SEs), omitted variables (bias), and overfitting (adjusted R^2 penalizes complexity). Pakistani example: house prices on area, bedrooms, location — interpreting 'holding others constant'.

Prediction vs causation: regression predicts; causal claims need design (experiments, instruments).

The exam-ready warning: a significant slope is association — confounding is the eternal rival explanation.

Key Definitions

Residual analysis is defined as the regression assumption-violations check: residuals vs fitted should show no pattern.

Examples

■ Slope t-test: study hours $\rightarrow$ marks, $\beta_1 = 3.2$, $p < 0.001$ — significant linear link.

■ Multicollinearity: area and bedrooms correlated — wide CIs for each.

Key Points — very important for exams

$\beta_1 = r(s_y/s_x)$; t-test for $H_0: \beta_1 = 0$; R^2 = variance explained.

Residual plots check linearity, variance, normality.

Multiple: ceteris-paribus coefficients; beware multicollinearity.

Association $\neq$ causation — confounding is the rival.

Review Questions

? $R^2 = 0.9$ — impressive or suspicious?

? Residuals fan out — what assumption fails?

Section Summary

Regression inference tests relationships and quantifies uncertainty — but diagnostics and design decide what it means.

Exam Pointers

$\rightarrow$ How is the slope tested?

$\rightarrow$ What does 'holding others constant' mean?

Lecture 8: Power and Effect Sizes

Learning Objectives

- Plan powered studies
- Interpret effects

Power analysis: power = $P(\text{reject } H_0 \mid \text{real effect } \delta)$ — the chance of detecting truth. It rises with n , effect size, and α , and falls with σ . Planning: choose n for 80% power at the minimum meaningful effect — the ethical and economic calculation (underpowered studies waste resources; overpowered ones waste money). Formula sketch: $n \approx 2\sigma^2(z_{\alpha} + z_{\beta})^2/\delta^2$ per group.

Effect sizes: Cohen's d (mean difference in SDs — 0.2 small, 0.5 medium, 0.8 large), correlation r , odds ratios. Report them always: p -values depend on n ; effect sizes don't. A drug trial with $n=10,000$ finds $p<0.001$ for a trivial effect — the effect size ($d=0.05$) tells the real story.

Common abuses: p -hacking (trying analyses until $p<0.05$), HARKing (hypothesizing after results known), publication bias (null results buried). Reforms: pre-registration, reporting all results, replication.

Exam point: power is a design duty — compute before collecting data, not after. 

Key Definitions

Power is defined as $1-\beta$: the probability of detecting a real effect — the design target (usually 80%).

Examples

- Planning: detect $\delta=5$ with $\sigma=15$, 80% power $\rightarrow n \approx 143/\text{group}$.
- p -hacking: 20 tests at $\alpha=0.05 \rightarrow \sim 1$ false 'discovery' expected.

Key Points — very important for exams

Power rises with n , effect, α ; falls with σ — plan for 80%.

$n \approx 2\sigma^2(z_{\alpha}+z_{\beta})^2/\delta^2$: the planning formula.

Effect sizes (Cohen's d): n -free importance measures.

Abuses: p -hacking, HARKing, publication bias — pre-register.

Review Questions

- ? Is an underpowered 'significant' result trustworthy?
- ? Design a trial for $d=0.4$, 90% power.

Section Summary

Power is planned, not hoped: size studies to detect meaningful effects, report effect sizes, and resist p -hacking.

Exam Pointers

- What determines power?
- Why report effect sizes alongside p -values?

Lecture 9: Nonparametrics and a Bayesian Glimpse

Learning Objectives

- Choose nonparametric tests
- Contrast paradigms

Nonparametrics: when normality fails or data are ordinal — the sign test and Wilcoxon signed-rank (paired alternatives to t), Mann-Whitney U (independent two-sample), Kruskal-Wallis (ANOVA alternative), Spearman rank correlation (monotonic association). They trade some power for robustness — the right trade with skewed Pakistani income data or Likert scales.

Rank logic: replace values by ranks, then test — outliers lose their leverage. Example: Mann-Whitney comparing satisfaction scores (1–5 Likert) between two branches — no normality assumed. Cost: if data are truly normal, nonparametric tests are ~5% less efficient — a small insurance premium.

Bayesian glimpse: combine prior beliefs with data via Bayes' theorem → posterior distributions; credible intervals (95% probability the parameter is inside — the interpretation frequentist CIs can't give); as n grows, priors wash out. Exam note: know the philosophical difference (parameters fixed vs random) and one practical contrast (credible vs confidence intervals).

Key Definitions

Nonparametric tests is defined as tests using ranks, not raw values — robust to non-normality and outliers.

Examples

- Wilcoxon: paired recovery times, skewed — ranks tame the skew.
- Credible interval: '95% probability $\mu \in (63, 72)$ ' — the natural reading.

Key Points — very important for exams

Ranks replace values: sign, Wilcoxon, Mann-Whitney, Kruskal-Wallis, Spearman.

Robust to skew/outliers; small efficiency cost when normal.

Bayes: prior + data → posterior; parameters as random.

Credible intervals give the interpretation CIs can't.

Review Questions

- ? Likert data, two groups — which test?
- ? Why do priors wash out with big n ?

Section Summary

When assumptions fail, rank; when philosophy shifts, update beliefs — inference has more than one engine.

Exam Pointers

- When are nonparametric tests preferred?
- Contrast Bayesian and frequentist inference.

Lecture 10: Survey Sampling Design

Learning Objectives

- Design samples
- Adjust inference to design

Sampling design decides what inference is valid: simple random sampling (every subset equally likely — the gold standard), stratified (divide into strata, sample within — guarantees representation, e.g., by province), cluster (sample whole clusters — cheaper, e.g., villages), systematic (every k th — convenient, risky with periodic patterns), and multistage (clusters within clusters — the Labour Force Survey's design).

Bias sources: selection bias (non-random selection), nonresponse (the busy/refusers differ — Pakistan's surveys struggle with elite nonresponse), and measurement bias (leading questions, faulty instruments). Randomization defends against the first; callbacks and weighting adjust the second; piloting fixes the third.

Sampling distributions depend on design: stratification reduces SE (strata homogeneous), clustering raises it (design effect — intra-cluster similarity). The exam point: inference formulas assume the design — SRS formulas on cluster samples understate uncertainty.

Key Definitions

Design effect is defined as the inflation of variance from cluster sampling vs SRS — intra-cluster similarity's cost.

Examples

- LFS: stratified multistage — provinces strata, villages clusters.
- Nonresponse: wealthy households refusing — income estimates biased down.

Key Points — very important for exams

SRS, stratified, cluster, systematic, multistage — match to goal and budget.

Biases: selection, nonresponse, measurement — defend each.

Stratification cuts SE; clustering raises it (design effect).

Inference formulas assume the design — very important.

Review Questions

- ? Design a national literacy survey for Pakistan.

? A systematic sample hits a weekly pattern — what breaks?

Section Summary

Good inference starts before data: design the sample, defend against bias, and let the design set the formulas.

Exam Pointers

- Compare sampling methods.
- Why do survey designs change standard errors?

Lecture 11: Variances and Proportions in Depth

Learning Objectives

- Infer about variances
- Master proportion methods

Estimating variances: s^2 for σ^2 (unbiased via $n-1$); confidence intervals via chi-square — $((n-1)s^2/\chi^2_{upper}, (n-1)s^2/\chi^2_{lower})$ — wide and skewed (variance CIs are asymmetric — the exam's visual). Comparing two variances: the F-test (s^2/s^2) — sensitive to non-normality, so use cautiously.

Proportions in depth: Wilson score intervals (better than Wald for small n or extreme p — use Wilson); difference of proportions; relative risk and odds ratios for 2×2 tables (medical/epidemiological staples). Pakistani example: vaccine efficacy — relative risk of disease in vaccinated vs unvaccinated.

Sample-size for proportions: $n = (z^*/m)^2 \cdot p(1-p)$, worst-case $p=0.5$. Worked: estimating poverty headcount $\pm 2\%$ at 95% → $n \approx 2,401$. The exam skill: choose Wald vs Wilson, compute odds ratios, and size proportion studies.

Key Definitions

Odds ratio is defined as the ratio of odds of an outcome in two groups — the case-control study's measure.

Examples

- χ^2 interval: $n=20, s=5 \rightarrow 95\%$ CI for σ is (3.8, 7.2) — asymmetric.
- Vaccine: $RR=0.2 \rightarrow 80\%$ efficacy — the headline statistic.

Key Points — very important for exams

Variance CIs via χ^2 — asymmetric; F-test compares two.

Wilson beats Wald for proportions (small n , extreme p).

Odds ratios and relative risk for 2×2 tables.

Size proportion studies with $p(1-p)$ — worst-case 0.25.

Review Questions

- ? When is the F-test for variances unsafe?
- ? Compute an odds ratio from a 2×2 table.

Section Summary

Beyond means: variances need χ^2 , proportions need Wilson, and 2×2 tables speak in odds ratios.

Exam Pointers

- How is a variance interval constructed?
- Compare Wald and Wilson intervals.

Lecture 12: Multiple Comparisons and Two-Way Designs

Learning Objectives

- Control multiplicity
- Analyze factorial designs

Multiple comparisons: doing many t-tests inflates the family-wise error — 10 tests at 5% → ~40% chance of a false discovery. Solutions: Bonferroni (α/k — simple, conservative), Tukey HSD (all pairwise, exact for equal n), Dunnett (vs control), and False Discovery Rate control (Benjamini-Hochberg — for genomics-scale testing).

Two-way ANOVA detail: factors A and B, interaction A×B — the interaction plot (non-parallel lines) is the diagnostic visual. Example: teaching method × class size on scores — method A wins only in small classes (interaction). Interpret main effects cautiously when interaction is significant.

Randomized block designs: blocking (grouping similar units — farms by soil type) removes nuisance variation — the agricultural field-trial staple. ANCOVA blends ANOVA with regression covariates. Exam skill: read an interaction plot and choose the comparison procedure.

Key Definitions

FDR is defined as controlling the expected proportion of false discoveries among rejections — for many simultaneous tests.

Examples

- Bonferroni: 5 tests → $\alpha=0.01$ each.
- Interaction plot: crossing lines — method effects reverse by class size.

Key Points — very important for exams

Many tests inflate error — Bonferroni/Tukey/FDR control it.

Two-way ANOVA: factors + interaction; plots diagnose.

Interaction significant → main effects need care.

Blocking removes nuisance variation — design smarter.

Review Questions

- ? 10 comparisons at $\alpha=0.05$ — expected false positives?
- ? Design a blocked fertilizer trial.

Section Summary

More comparisons need more discipline: control the error rate, plot the interactions, and block the noise.

Exam Pointers

- Why not run many t-tests?
- How do you interpret an interaction?

Lecture 13: Transformations, Diagnostics, and Logistic Regression

Learning Objectives

- Fix violations
- Model binary outcomes

Transformations fix assumption violations: log (right-skew, multiplicative effects — incomes), square root (counts), arcsine (proportions), Box-Cox (data-driven power). Always interpret on the original scale (back-transform with care — the mean of logs $\neq$ log of the mean). Diagnostics: Q-Q plots (normality), residuals vs fitted (variance), leverage and Cook's distance (influential points).

Outliers vs influence: an outlier is unusual y ; influential points move the fit (high leverage + discrepancy). Don't delete reflexively — investigate (data error? genuine extreme?), report sensitivity (with/without). Robust regression (downweighting outliers) is the principled alternative.

Logistic regression: for binary outcomes — models log-odds as linear: $\log(p/(1-p)) = \beta_0 + \beta_1 x$. Coefficients as odds ratios (e^{β_1}); Pakistani example: loan default on income and credit history — the banking staple. Exam point: logistic for yes/no, linear for continuous — match model to outcome.

Key Definitions

Logistic regression is defined as modeling log-odds of a binary outcome as linear in predictors — coefficients as odds ratios.

Examples

- Log income: multiplicative effects become additive — and normal-ish.
- Loan default: e^{β} for credit score — odds ratio per point.

Key Points — very important for exams

Transform skew (log), counts ($\sqrt{\cdot}$), proportions (arcsine).

Diagnose: Q-Q, residuals-vs-fitted, Cook's distance.

Outliers: investigate, sensitivity-check, don't auto-delete.

Logistic: binary outcomes via log-odds — odds-ratio interpretation.

Review Questions

? Residuals fan out — transform how?

? Interpret $e^{\beta} = 1.5$ in a default model.

Section Summary

Models meet reality through diagnostics: transform, check influence, and switch to logistic when outcomes are yes/no.

Exam Pointers

→ When and how do you transform data?

→ Distinguish outliers from influential points.

Lecture 14: Bootstrap and Meta-Analysis

Learning Objectives

- Resample for inference
- Combine studies

Resampling: the bootstrap — resample the data with replacement, recompute the statistic thousands of times, and read the sampling distribution empirically. No normality needed: bootstrap CIs for medians, correlations, even complex estimators. The permutation test: shuffle group labels to test 'no difference' — exact p-values without distributional assumptions.

Why it matters: modern computing frees inference from textbook distributions — the bootstrap handles what formulas can't (e.g., CI for a ratio of means). Conditions: the sample must represent the population (garbage in, garbage out — resampling can't fix bias).

Meta-analysis: combining studies — weighted averages (inverse-variance weights), forest plots, heterogeneity (I^2 — do studies disagree?), and publication-bias checks (funnel plots). Pakistani relevance: combining small agricultural trials or education interventions. Exam synthesis: resampling = empirical sampling distributions; meta-analysis = inference across studies.

Key Definitions

Bootstrap is defined as resampling data with replacement to empirically estimate a statistic's sampling distribution.

Examples

- Bootstrap CI for the median income — no formula needed.
- Forest plot: 12 fertilizer trials combined — overall effect + heterogeneity.

Key Points — very important for exams

Bootstrap: resample → recompute → empirical distribution — no normality.

Permutation tests: shuffle labels for exact p-values.

Meta-analysis: weighted combination, forest plots, I^2 heterogeneity.

Resampling can't fix biased samples — representativeness first.

Review Questions

- ? Bootstrap a CI for a correlation.
- ? Funnel plot asymmetry suggests what?

Section Summary

Computing freed inference: bootstrap what formulas can't, and combine studies honestly — heterogeneity and all.

Exam Pointers

- How does the bootstrap work?
- What is meta-analysis?

Lecture 15: Reporting and Final Drills

Learning Objectives

- Report analyses
- Drill the paper

Reporting standards: every analysis reports the question, design, sample, methods (with assumptions checked), results (estimates + CIs + effect sizes + p-values), and limitations. The ASA's p-value statement: p-values indicate incompatibility with H_0 — not effect size, not truth probability, not importance. Write for the decision-maker: 'yield rose 8% (95% CI: 5–11%)' beats ' $p < 0.001$ '.

Common exam traps reviewed: CI misinterpretation, p-value as $P(H_0)$, confusing SE with SD, correlation → causation, multiple testing without correction, and 'non-significant = no effect' (absence of evidence $\neq$ evidence of absence — check power).

Final drill set: (1) $n=36$, $\bar{x}=50$, $s=12$ — 95% CI for μ ; (2) test $H_0: \mu=45$ at 5%; (3) two groups 30/30, means 52/48, $s=10$ — pooled t-test; (4) χ^2 independence on a given 2×2 ; (5) complete an ANOVA table; (6) interpret a regression slope's t-test. Work all six closed-book — this is the paper in miniature.

Key Definitions

ASA p-value principles is defined as the ASA's core message: p-values measure incompatibility with H_0 — nothing about size, truth, or importance.

Examples

- Report: '8% rise (CI 5–11%)' — decision-ready.
- Fallacy: 'p=0.2, so no effect' — with n=10, power was nil.

Key Points — very important for exams

Report: question, design, methods, estimates+CI+effects, limitations.

p-values: incompatibility with H_0 only — ASA.

Fallacies: CI/p misreads, SE/SD confusion, causation leaps, multiplicity.

Six-drill set: the paper in miniature — master it.

Note:

Note: do the six drills closed-book under time — exam fluency comes from timed practice, not re-reading.

Review Questions

- ? Write a results paragraph for drill (1).
- ? Which fallacy is commonest in media?

Section Summary

Inference ends in communication: report honestly, dodge the fallacies, and let drills build exam fluency.

Exam Pointers

- What must a statistical report include?
- List five common inferential fallacies.

Lecture 16: Solved Problems: Estimation

Learning Objectives

- Solve estimation problems
- Avoid common slips

Problem 1 — CI for mean: A factory's 36 bulbs have mean life 1,200 hrs, $s = 90$. 95% CI: $1200 \pm 1.96(90/6) = 1200 \pm 29.4 = (1170.6, 1229.4)$. Interpretation: 95% of such intervals cover the true mean life. Problem 2 — sample size: want μ within ± 10 hrs, $\sigma \approx 90$: $n = (1.96 \times 90/10)^2 \approx 312$ bulbs.

Problem 3 — proportion CI: 480 of 800 voters support a policy: $p = 0.6$, 95% Wilson $\approx (0.566, 0.633)$.
Problem 4 — variance CI: $n=16$, $s^2=25$: 95% CI for $\sigma^2 = (15 \times 25 / 27.49, 15 \times 25 / 6.26) = (13.6, 59.9)$ — note the asymmetry. Problem 5 — paired CI: 10 patients' BP drops (mean 8, $s_d=4$): 95% CI = $8 \pm 2.262(4/\sqrt{10}) = (5.14, 10.86)$.

Exam technique: always (a) state the parameter, (b) check conditions, (c) compute with the right distribution ($z/t/\chi^2$), (d) interpret in context with correct confidence language. Common slip: using z when σ is unknown and n small — the t -correction exists for exactly this.

Key Definitions

z vs t choice is defined as using z^* when σ is known/large- n and t^* when σ is estimated from small samples.

Examples

- Bulb life: (1170.6, 1229.4) — the complete four-step answer.
- Poll: Wilson (0.566, 0.633) vs Wald (0.566, 0.634) — similar here, safer generally.

Key Points — very important for exams

Four-step method: parameter $\rightarrow$ conditions $\rightarrow$ compute $\rightarrow$ interpret.

t for small- n unknown σ ; χ^2 for variances (asymmetric).

Wilson for proportions; pair differences for paired data.

Sample-size formula: precision costs quadratically.

Review Questions

? $n=16$, $\bar{x}=20$, $s=6$: 99% CI for μ ?

? Why is $n=312$ for ± 10 hrs?

Section Summary

Estimation problems reward method: right distribution, checked conditions, honest interpretation.

Exam Pointers

- $\rightarrow$ When do you use z vs t for a mean interval?
- $\rightarrow$ Why are variance intervals asymmetric?

Lecture 17: Solved Problems: Hypothesis Tests

Learning Objectives

- Solve test problems
- Report completely

Problem 1 — one-sample t: $H_0: \mu=70$ vs $H_1: \mu>70$; $n=25$, $\bar{x}=74$, $s=8$: $t = (74-70)/(8/\sqrt{25}) = 2.5$, $df=24$, one-sided $p \approx 0.01 \rightarrow$ reject at 5%. Conclusion: significant evidence mean exceeds 70. Effect: 4 points — report it.

Problem 2 — two-sample: teaching methods, $n_1=n_2=30$, $\bar{x}_1=78$, $\bar{x}_2=72$, $s_p=10$: $t = 6/(10\sqrt{(2/30)}) \approx 2.32$, $df=58$, $p \approx 0.024 \rightarrow$ significant. 95% CI for difference: $6 \pm 2.0(2.58) = (0.8, 11.2)$. Problem 3 — paired: 12 runners' times before/after training, mean drop 4.2s, $s_d=5$: $t = 4.2/(5/\sqrt{12}) \approx 2.91$, $df=11$, $p \approx 0.014 \rightarrow$ training works.

Problem 4 — chi-square independence: 200 patients, treatment vs recovery table; $\chi^2=6.8$, $df=1$, $p \approx 0.009 \rightarrow$ associated. Problem 5 — proportion: 65/100 vs 52/100 pass: $z = 0.13/\sqrt{(0.585 \times 0.415 \times 0.02)} \approx 1.87$, $p \approx 0.06 \rightarrow$ not significant at 5% (but note: close — report, don't hide). Exam technique: 

hypotheses $\rightarrow$ statistic $\rightarrow$ p $\rightarrow$ decision $\rightarrow$ effect size $\rightarrow$ context sentence.

Key Definitions

Test reporting chain is defined as the complete six-step reporting chain for any hypothesis test.

Examples

■ $p=0.06$: 'not significant at 5%, but suggestive — larger study warranted.'

■ Paired $t=2.91$: each runner their own control — the power of pairing.

Key Points — very important for exams

Six steps: hypotheses $\rightarrow$ statistic $\rightarrow$ p $\rightarrow$ decision $\rightarrow$ effect $\rightarrow$ context.

Paired designs are powerful; χ^2 handles counts.

p near α : report honestly, don't dichotomize thought.

Effect sizes always — significance is not importance.

Review Questions

? When is a paired test invalid?

? $\chi^2=6.8$, $df=1$: decide at 1%.

Section Summary

Tests are arguments: state claims, quantify surprise, decide — then say how big the effect is.

Exam Pointers

$\rightarrow$ What is the six-step test procedure?

$\rightarrow$ Why report effect sizes with p-values?

Lecture 18: Solved Problems: ANOVA and Regression

Learning Objectives

- Solve ANOVA/regression problems
- Complete tables

Problem 1 — one-way ANOVA: 3 diets, $n=10$ each, $SSB=240$, $SSW=540$: $MSB=120$, $MSW=20$, $F=6.0$, $df=(2,27)$, $p\approx 0.007 \rightarrow$ diets differ. Tukey: which pairs? (Needs $MSE=20$ and means.) Problem 2 — complete the table: given $SST=1000$, $SSB=300$, $k=4$, $n=40$: $SSW=700$, $df=(3,36)$, $F=(100/19.44)=5.14 \rightarrow$ significant.

Problem 3 — regression: $n=20$, $\Sigma(x-\bar{x})(y-\bar{y})=800$, $\Sigma(x-\bar{x})^2=200$, $s=4$: $\beta=4$, $SE=4/\sqrt{200}\approx 0.283$, $t\approx 14.1 \rightarrow$ highly significant slope. 95% CI: $4 \pm 2.10(0.283) = (3.41, 4.59)$. $R^2=0.91$: 91% of variation explained — strong linear fit (check residuals!). Problem 4 — χ^2 goodness-of-fit: 120 dice rolls, face counts vs 20 each: $\chi^2=18.2$, $df=5$, $p\approx 0.003 \rightarrow$ die biased.

Exam technique for ANOVA tables: fill df first (they're mechanical), then MS , then F — and always state the conclusion about means in words, plus the follow-up (Tukey) when H_0 is rejected.

Key Definitions

ANOVA table completion is defined as filling ANOVA tables mechanically: $df \rightarrow SS \rightarrow MS \rightarrow F$ — the standard exam format.

Examples

■ $F=6.0$, $p=0.007$: 'diets differ; Tukey locates the pairs.'

■ $t=14.1$ for slope: overwhelming linear evidence — but plot residuals.

Key Points — very important for exams

ANOVA: $df \rightarrow MS \rightarrow F$; reject $\rightarrow$ Tukey.

Slope t-test: β/SE ; R^2 describes fit.

χ^2 goodness-of-fit: observed vs expected counts.

Conclusions in words + follow-ups — never bare numbers.

Review Questions

? $SST=500$, $SSB=150$, $k=5$, $n=50$: compute F .

? $R^2=0.91$ but residuals curve — your verdict?

Section Summary

ANOVA and regression problems are table-filling plus judgment: compute mechanically, conclude thoughtfully.

Exam Pointers

$\rightarrow$ Complete a partial ANOVA table.

$\rightarrow$ How do you test a regression slope?

Lecture 19: The t, Chi-Square, and F Distributions

Learning Objectives

- Understand the three distributions
- Use tables fluently

The t-distribution: arising when x is standardized with s instead of σ — heavier tails reflecting the extra uncertainty; $df=n-1$; approaches normal as $n \rightarrow \infty$. The chi-square: distribution of $\sum z^2$ — sum of squared standard normals; $df=k$; used for variances and counts; skewed right, $\text{mean}=df$. The F: ratio of two scaled chi-squares — $(\chi^2/d)/(\chi^2/d)$; used in ANOVA and variance comparison; skewed right.

Derivation sketches: $(x-\mu)/(s/\sqrt{n}) \sim t_{n-1}$ because numerator $\sim N(0,1)$ independent of denominator's $\sqrt{(\chi^2/df)}$; $(n-1)s^2/\sigma^2 \sim \chi^2_{n-1}$; $F = (SSB/(k-1))/(SSW/(n-k))$ under H_0 . You need the shapes, df rules, and uses — not full proofs — but understanding the constructions prevents df errors.

Relationships: $t^2_{df} = F_{1,df}$; as $df \rightarrow \infty$, $t \rightarrow z$, $\chi^2/df \rightarrow 1$, $F \rightarrow 1$. Reading tables: t^* for CIs (two-sided), t critical for tests (one/two-sided — the halving trap), χ^2 upper/lower for variance intervals, F upper for ANOVA. Exam drill: given α and df , find each critical value — table fluency is free marks. 

Key Definitions

t, χ^2 , F families is defined as the three sampling distributions beyond normal: t (means, σ unknown), χ^2 (variances, counts), F (variance ratios).

Examples

■ $t^2_{df} = F_{1,df}$: the two-sample t-test as ANOVA with $k=2$.

■ Halving trap: one-sided 5% test uses $t_{0.05}$, not $t_{0.025}$.

Key Points — very important for exams

t : $(x-\mu)/(s/\sqrt{n})$; χ^2 : $\sum z^2$ and $(n-1)s^2/\sigma^2$; F : variance ratios.

df rules: $n-1$, $n-1$, (d, d) — mechanical once understood.

$t^2=F$; all $\rightarrow$ normal-like limits as $df \rightarrow \infty$.

Table fluency: one- vs two-sided critical values — free marks.

Review Questions

? Why does t have heavier tails than z ?

? Find $F_{0.05}(3,20)$.

Section Summary

Three distributions, three jobs: t for means, χ^2 for variances and counts, F for comparing variances.

Exam Pointers

→ How do t , χ^2 , and F arise?

→ What are their df rules?

Lecture 20: Question Bank with Answer Outlines

Learning Objectives

- Drill exam archetypes 
- Practice complete answers

Question bank (with answer outlines): Q1. Prove $\bar{x}$ is unbiased for μ . ($E[\bar{x}] = \Sigma E[x_i]/n = \mu$.) Q2. Derive the MLE of λ for Poisson data. ($L = \prod e^{-\lambda} \lambda^{x_i} / x_i!$; $\bar{x} = n^{-1} \Sigma x_i$; $\ln L = -n\lambda + \lambda \bar{x}$; $\lambda = \bar{x}$.) Q3. A 95% CI is (10,20): interpret correctly and list two misinterpretations. Q4. $p=0.04$ at $\alpha=0.01$: decide and explain. Q5. Paired vs independent: a diet study weighs subjects before/after — which test and why?

Q6. Complete: $SSB=120$, $SSW=480$, $k=4$, $n=36 \rightarrow F?$ ($MSB=40$, $MSW=15$, $F=2.67$, $df=(3,32)$, $p \approx 0.06$ — not significant.) Q7. $R^2=0.64$: interpret; give one caution. (64% explained; causation not shown.) Q8. Power 0.5: what does it mean? (Coin flip of detecting a real effect — underpowered.) Q9. Expected cell count 3 in χ^2 : problem and fix? (Combine categories or Fisher's exact.) Q10. Bootstrap in one paragraph.

Q11. Bayesian vs frequentist: parameters random or fixed? Q12. n for $\pm 1\%$ margin on a proportion at 95%? ($(1.96/0.02)^2 \times 0.25 = 2401$.) Long Q: Design a study comparing two fertilizers: sampling, tests, power, reporting — full plan. Work each with the six-step reporting chain; the paper rewards complete arguments, not bare computations.

Key Definitions

Question archetypes is defined as the exam's favorite numerical formats: complete-the-table, derive-the-MLE, interpret-the-output.

Examples

■ Q6: $F=2.67$, $p \approx 0.06$ — 'no significant difference at 5%.'

■ Q12: 2,401 respondents — the $\pm 1\%$ price tag.

Key Points — very important for exams

Twelve questions covering the course's core computations.

MLE derivations: write likelihood, log, differentiate.

Design question: sampling $\rightarrow$ tests $\rightarrow$ power $\rightarrow$ reporting.

Six-step chain on every answer — completeness earns marks.

Note:

Review Questions

- ? Which archetype is your weakest?
- ? Time yourself: 15 minutes per long question.

Section Summary

The paper in twelve questions: derive, compute, interpret, design — each answered as a complete argument.

Exam Pointers

- Derive the Poisson MLE.
- Design a two-fertilizer study fully.

Lecture 21: Nonparametric Methods

Learning Objectives

- Choose rank tests
- Compute Wilcoxon

Nonparametric tests: when normality fails or data is ordinal — sign test (median of paired differences), Wilcoxon signed-rank (paired, uses magnitudes), Wilcoxon rank-sum / Mann-Whitney (two independent samples), Kruskal-Wallis (ANOVA's rank analogue), and Spearman rank correlation (monotonic association). The logic: replace values by ranks, then test — distribution-free.

Worked: 8 patients' pain scores (ordinal 1–10) before/after treatment: Wilcoxon signed-rank — rank the absolute differences, sum positive ranks, compare to critical value. Exam technique: justify the choice ('ordinal data, n small, normality doubtful') before computing — the justification earns marks, the ranks earn the rest.

Power trade-off: nonparametric tests are safer but weaker — if data is truly normal, the t-test wins. The decision tree: normal + interval → parametric; else → ranks. Very important: never apply a t-test to ordinal Likert data without comment — examiners notice.

Key Definitions

Kruskal-Wallis is defined as Kruskal-Wallis: the rank-based analogue of one-way ANOVA for 3+ groups.

Examples

- Pain scores: ranks replace raw values — the distribution-free move.
- Likert data: ordinal — justify, don't just t-test.

Key Points — very important for exams

Sign, Wilcoxon, Mann-Whitney, Kruskal-Wallis, Spearman — the rank toolkit.

Replace values by ranks; test distribution-free.

Justify the choice — then compute.

Safer but weaker: parametric wins when assumptions hold.

Review Questions

? Kruskal-Wallis vs ANOVA: decide for skewed data.

? Rank: 3, 1, 2, 2, 5 — handle the tie.

Section Summary

When normality fails, rank: the nonparametric toolkit keeps inference honest on ordinal and skewed data.

Exam Pointers

→ When do you choose nonparametric tests?

→ Do a Wilcoxon signed-rank by hand.

Lecture 22: Software Output and Reporting

Learning Objectives

- Interpret output
- Report properly

Statistical software and reporting: R (`t.test`, `aov`, `lm`, `chisq.test` — the exam may ask for output interpretation), SPSS (menus + syntax), and Python (`scipy.stats`). Reading R output: coefficients table (Estimate, Std. Error, t, p), ANOVA table, and the diagnostic plots — interpret, don't just quote p-values.

The reproducible report: code + data + narrative (R Markdown / Jupyter) — the modern standard replacing copy-paste tables. APA-style reporting: ' $t(24) = 2.50, p = .020, d = 0.50$ ' — statistic, df, p, effect size, in one line. Figures: every plot labeled, every table captioned, no chartjunk.

Exam skill: given computer output, write the conclusion — identify the test, read the p-value, state the decision, add the effect size, and note assumption checks. The most tested output: regression summaries and ANOVA tables. Very important: software computes; the statistician concludes — practice the concluding sentence.

Key Definitions

APA reporting is defined as the one-line statistical report: statistic, df, p-value, effect size — e.g. $t(24)=2.50, p=.020, d=0.50$.

Examples

- lm output: 'each extra hour predicts +2.1 marks (SE 0.4), $p < .001$.'
- APA: $t(24) = 2.50$, $p = .020$, $d = 0.50$ — the complete line.

Key Points — very important for exams

R/SPSS/Python: know the key functions and menus.

Read output: coefficients, ANOVA tables, diagnostics.

APA: statistic(df) = value, p, effect size.

Reproducible reports: code + data + narrative.

Review Questions

- ? Write the APA line for $F(2,27)=6.0$, $p=.007$.
- ? What diagnostic plots for lm?

Section Summary

Software computes, you conclude: read outputs completely and report in the one-line APA style.

Exam Pointers

- Interpret a regression summary.
- What goes in a reproducible report?

Glossary of Key Terms

Standard error — the standard deviation of a statistic's sampling distribution — measuring its precision ($\sigma/\sqrt{n}$ for $\bar{x}$).

Maximum likelihood estimation — choosing the parameter value maximizing the probability of the observed data.

Confidence interpretation — the correct interpretation: 95% of intervals from repeated sampling cover the parameter — not $P(\text{parameter in this interval}) = 0.95$.

p-value — $P(\text{data this extreme} \mid H_0 \text{ true})$ — not $P(H_0 \text{ true} \mid \text{data})$.

Paired test — analyzing before/after differences on the same subjects — removing between-subject variability.

ANOVA — partitioning total variation into between-group and within-group — the F-test for ≥ 3 means.

Residual analysis — the regression assumption-violations check: residuals vs fitted should show no pattern.

Power — $1 - \beta$: the probability of detecting a real effect — the design target (usually 80%).

Nonparametric tests — tests using ranks, not raw values — robust to non-normality and outliers.

Design effect — the inflation of variance from cluster sampling vs SRS — intra-cluster similarity's cost.

Odds ratio — the ratio of odds of an outcome in two groups — the case-control study's measure.

FDR — controlling the expected proportion of false discoveries among rejections — for many simultaneous tests.

Logistic regression — modeling log-odds of a binary outcome as linear in predictors — coefficients as odds ratios.

Bootstrap — resampling data with replacement to empirically estimate a statistic's sampling distribution.

ASA p-value principles — the ASA's core message: p-values measure incompatibility with H_0 — nothing about size, truth, or importance.

z vs t choice — using z^* when σ is known/large- n and t^* when σ is estimated from small samples.

Test reporting chain — the complete six-step reporting chain for any hypothesis test.

ANOVA table completion — filling ANOVA tables mechanically: $df \rightarrow SS \rightarrow MS \rightarrow F$ — the standard exam format. 

t, χ^2 , F families — the three sampling distributions beyond normal: t (means, σ unknown), χ^2 (variances, counts), F (variance ratios).

Question archetypes — the exam's favorite numerical formats: complete-the-table, derive-the-MLE, interpret-the-output.

Kruskal-Wallis — Kruskal-Wallis: the rank-based analogue of one-way ANOVA for 3+ groups.

APA reporting — the one-line statistical report: statistic, df, p-value, effect size — e.g. $t(24)=2.50$, $p=.020$, $d=0.50$.