

STEP 1 · OUTCOME TYPE

Continuous (measured), ordinal (scores, grades), categorical (yes/no, groups) or time-to-event?

STEP 2 · GROUPS & PAIRING

How many groups? Are they independent people, or the same people measured more than once (paired/matched)?

STEP 3 · DISTRIBUTION

Continuous data roughly normal? Use the first test; skewed, small samples or ordinal data, use the alternative.

Comparing groups: continuous or ordinal outcome

YOUR COMPARISON	NORMAL DATA (PARAMETRIC)	SKEWED / ORDINAL (NON-PARAMETRIC)
2 independent groups e.g. Hb in cases vs controls	Independent (unpaired) t-test Welch's version if variances differ	Mann-Whitney U test
2 paired measurements e.g. BP before vs after treatment	Paired t-test	Wilcoxon signed-rank test
3 or more independent groups	One-way ANOVA + post hoc test (e.g. Tukey)	Kruskal-Wallis test + post hoc (e.g. Dunn's)
3 or more repeated measurements same patients over time	Repeated-measures ANOVA	Friedman test

Comparing groups: categorical outcome

YOUR COMPARISON	TEST	WHEN TO SWITCH
Proportions in independent groups e.g. complication rate, drug A vs B	Chi-square test	Fisher's exact test if expected counts are small (2×2 table: any expected count < 5)
Paired yes/no e.g. same patients, before vs after	McNemar's test	Uses only the patients whose answer changed

Relationships, prediction and agreement

QUESTION	TEST	NOTES
Are two continuous variables related?	Pearson correlation (r)	Spearman rank correlation if skewed, ordinal or non-linear but consistent
Predict a continuous outcome / adjust for confounders	Linear regression	Reports coefficients with 95% CI
Predict a yes/no outcome / adjust for confounders	Logistic regression	Reports odds ratios with 95% CI
Do two methods or observers agree?	Bland-Altman plot (continuous)	Cohen's kappa for categorical ratings
How good is a diagnostic test?	Sensitivity, specificity, PPV, NPV	ROC curve & AUC for a continuous test value
Time until an event e.g. survival, recurrence	Kaplan-Meier + log-rank test	Cox regression to adjust (hazard ratios)

Checking normality

- Look at a **histogram** or **Q-Q plot**, then use **Shapiro-Wilk** as support, not as the only check.
- In small samples normality tests often miss real skew; in large samples they flag trivial deviations. Let the plot and your clinical sense decide.
- Ordinal scores (pain scales, grades): use non-parametric tests.

How to report

- Normal data: **mean ± SD**. Skewed data: **median (IQR)**. Categorical: **n (%)**.
- Give **exact p-values** (p = 0.032, not "p < 0.05") and **95% confidence intervals**.
- Name the test and software used in your methods section.

Common mistakes that examiners catch

- Several t-tests instead of one ANOVA** for 3+ groups. In our simulation, three t-tests gave 13% false positives versus 5% for ANOVA.
- Unpaired test on paired data**, which wastes the pairing and can miss real effects.
- Chi-square with small expected counts**: use Fisher's exact test.
- Mean ± SD for skewed data** such as length of stay or lab values with outliers.
- "p > 0.05, so no difference"**: a non-significant result can mean the study was too small. Report the CI.
- Choosing the test after seeing the results**. Plan tests in your protocol.