Inferential Statistics Cheat Sheet

BCB744 — Biostatistics Version 3 (3 April 2026)

Professor A.J. Smit

Chapters 1–5 and 26 develop foundations, workflow, and interpretation. The sheet below covers the inferential, diagnostic, and model-based procedures introduced across Chapters 6–25.

QUICK DECISION FLOW

1. **Response?** Continuous → Gaussian methods; binary or proportion → binomial GLM; count → Poisson or quasi-Poisson.
2. **Independent?** No → paired procedure or mixed/grouped model.
3. **Linear enough?** No → polynomial, GAM, nonlinear, or quantile model as appropriate.

Ch.	Cue	Method	Use When	Null / Model Target	Main R Function(s)	Notes, Assumptions, or Follow-up
TESTING DIFFERENCES AND BASIC CHECKS						
Diagnostics and assumption checks						
6	μ **G**	Shapiro–Wilk normality test	One sample, paired differences, or within-group values may need a formal normality check before a Gaussian mean-based test.	H_0: the checked values are compatible with a normal distribution.	`shapiro.test(x)`	Check values on the analysis scale. Prioritise histograms and Q–Q plots; don't use alone.
6	μ **G**	Levene's test for equal variance	Two or more groups are being compared and homogeneity of spread needs checking.	H_0: group variances are equal.	`car::leveneTest(y ~ g,` `data = df)`	Preferred over Bartlett when normality is doubtful. Most relevant before Student's t-test and classical ANOVA.
6	μ **G**	Bartlett's test	Several groups are compared and a normal-theory variance test is acceptable.	H_0: group variances are equal.	`bartlett.test(y ~ g,` `data = df)`	More sensitive to non-normality than Levene. Use only when normality is already plausible.
6	μ **G**	Transformation of the response	Assumptions fail on the raw scale but a transformed scale may stabilise spread or reduce skew.	No null; the aim is a more defensible analysis scale.	`mutate(log_y = log(y))` `mutate(sqrt_y = sqrt(y))`	Transform before the final Gaussian analysis, then re-check assumptions.
Mean comparisons: *t*-tests and rank analogues						
7	μ **G**	One-sample t-test	One continuous sample is compared with a biologically meaningful reference mean.	H_0: $\mu = \mu_0$.	`t.test(x, mu = m0)`	Directional alternatives use `alternative = "less"` or `"greater"`.
7	μ **G**	Two-sample Welch t-test	Two independent groups are compared and equal variances should not be assumed.	H_0: $\mu_1 = \mu_2$.	`t.test(y ~ g, data = df)`	Usually the default. Assumes independence and approximate within-group normality.
7	μ **G**	Two-sample Student t-test	Two independent groups are compared and equal variances are defensible.	H_0: $\mu_1 = \mu_2$.	`t.test(y ~ g, data = df,` `var.equal = TRUE)`	Use only when group spreads are sufficiently similar.
7	μ **G**	Paired t-test	Two linked measurements are compared within the same sampling units.	H_0: mean paired difference = 0.	`t.test(x, y, paired = TRUE)`	Assumptions apply to within-pair differences, not the raw before/after values.
7	◇ **Rk**	Wilcoxon signed-rank test	One-sample or paired problem when a rank-based alternative is more defensible.	H_0: the location shift is 0.	`wilcox.test(x, mu = m0)` `wilcox.test(x, y, paired = TRUE)`	Use for one-sample and paired designs. Still requires the correct design logic.
7	◇ **Rk**	Wilcoxon rank-sum / Mann–Whitney test	Two independent groups are compared without relying on Gaussian assumptions.	H_0: the two groups come from the same distributional location.	`wilcox.test(y ~ g,` `data = df)`	Rank-based alternative to the two-sample t-test.
ANOVA family and rank-based group comparisons						
8	μ **G**	One-way ANOVA	One continuous response is compared across more than two independent groups.	H_0: $\mu_1 = \mu_2 = \cdots = \mu_k$.	`aov(y ~ g, data = df)`	Check residual normality and equal variance. This is the multi-group mean-comparison model.
8	μ **G**	Factorial / two-way ANOVA	Two or more categorical predictors are fitted together without continuous covariates.	H_0: no main effects and, if included, no interaction effect.	`aov(y ~ A * B, data = df)`	Use `A + B` for additive structure and `A * B` when the interaction is part of the question.
8,14	μ **G**	ANCOVA as a linear model	One continuous response is modelled using both categorical and continuous predictors.	Model target: adjusted group differences plus covariate slope(s).	`lm(y ~ g + x, data = df)`	ANCOVA is a standard linear model with one factor and one or more covariates.
8	μ **G**	Tukey's HSD post hoc comparison	One-way ANOVA is significant and pairwise mean differences among groups are needed.	H_0: each pairwise group mean difference = 0.	`TukeyHSD(fit)`	Use after a fitted `aov()` object. Controls family-wise error for all pairwise comparisons.
8	◇ **Rk**	Kruskal–Wallis test	More than two independent groups are compared with a rank-based omnibus test.	H_0: groups have the same distributional location.	`kruskal.test(y ~ g,` `data = df)`	Rank-based alternative to one-way ANOVA.
8	◇ **Rk**	Pairwise Wilcoxon comparisons	Kruskal–Wallis or multi-group rank comparisons require pairwise follow-up.	H_0: each pair of groups has the same location.	`pairwise.wilcox.test(y, g,` `p.adjust.method = "BH")`	Used as a follow-up rather than as the primary omnibus test.
CONTINUOUS MODELING						
Association and correlation						
9	↔ **Asc**	Pearson correlation	Two continuous variables are paired and a linear association is of interest without fitting a response model.	H_0: $\rho = 0$.	`cor.test(x, y,` `method = "pearson")`	Appropriate for approximately linear continuous relationships.
9	↔ **Asc**	Spearman rank correlation	Two paired variables show a monotonic but not necessarily linear association, or ranks are more appropriate.	H_0: $\rho_s = 0$.	`cor.test(x, y,` `method = "spearman")`	Works for ordinal or non-Gaussian paired data.
9	↔ **Asc**	Kendall's τ	Two paired variables are associated, often with small samples or many ties.	H_0: $\tau = 0$.	`cor.test(x, y,` `method = "kendall")`	Often more conservative than Spearman.
Gaussian regression: mean structure						
11	μ **G**	Residual diagnostics for fitted Gaussian models	A regression model has been fitted and the adequacy of mean and variance structure must be checked.	No null; the aim is to inspect residual structure.	`plot(mod, which = 1)` `plot(mod, which = 2)`	Residual-vs-fitted checks mean structure and spread. Q–Q checks residual normality. Independence comes from design.
12	μ **G**	Simple linear regression	One continuous response is modelled as a linear function of one continuous predictor.	Model target: mean response changes linearly with the predictor.	`lm(y ~ x, data = df)`	Inference on slopes uses the fitted linear model, not a correlation test in disguise.

12	μ
	G
		Confidence and prediction intervals	A fitted linear model is used for mean-response uncertainty or individual-response prediction.	No null; the aim is interval estimation.	`predict(mod, newdata = nd,`
					`interval = "confidence")`
					`predict(mod, newdata = nd,`
					`interval = "prediction")`
						Confidence intervals describe the mean line; prediction intervals describe future observations and are wider.

12,13	μ **G**	Breusch–Pagan test	Heteroscedasticity in a fitted linear model needs a formal supporting check.	H_0: residual variance is constant.	`lmtest::bptest(mod)`	Use as supporting evidence after reading residual plots.

Continued on next page

(Continued)

Ch.	Cue	Method	Use When	Null / Model Target	Main R Function(s)	Notes, Assumptions, or Follow-up

12,17	μ G	Influence diagnostics	Single observations may be driving the fitted model disproportionately.	No single null; the aim is to identify leverage and influence.	`cooks.distance(mod)`
					`dffits(mod)`
					`plot(mod, which = 5)`
						Inspect Cook's distance, DFFITS, leverage, and residual-vs-leverage plots before acting on a point.

13	μ G	Polynomial regression	A linear model is too rigid but the problem can still be handled within the `lm()` framework.	Model target: mean structure includes curvature.	`lm(y ~ x + I(x^2), data = df)` `lm(y ~ poly(x, k), data = df)`	Still linear in the coefficients. Compare nested degrees by `anova()` or `AIC()`.
14	μ G	Multiple regression	One continuous response depends on several predictors fitted simultaneously.	Model target: partial effects after adjusting for the other predictors.	`lm(y ~ x1 + x2 + x3, data = df)`	Interpret coefficients conditionally. Model specification matters as much as fit.
14,15	μ G	Interaction model	The effect of one predictor may depend on another predictor.	H_0: the interaction coefficient = 0.	`lm(y ~ x1 * x2, data = df)`	Use `*` to include both main effects and the interaction. Respect hierarchy.
14,15, 17	μ G	Nested model comparison for Gaussian linear models	A simpler and a more complex `lm()` model are nested and need formal comparison.	H_0: added term(s) do not reduce residual variance beyond chance.	`anova(mod_small, mod_large, test = "F")`	Appropriate when one model is a special case of the other.
MODEL SELECTION AND MODERN WORKFLOWS						
14,17, 24	△ Any	Akaike Information Criterion	Several biologically plausible models need relative comparison without a single null hypothesis.	Lower AIC indicates better trade-off between fit and complexity.	`AIC(mod1, mod2, mod3)`	Useful for non-nested and nested candidate models. Compare only models fitted to the same rows, response data, and response scale.
16	μ G	Variance Inflation Factor (VIF)	Predictors in a multiple regression may be collinear.	No null; the aim is to quantify variance inflation from shared predictor information.	`car::vif(mod)`	Large VIFs indicate unstable coefficient interpretation. Use alongside correlation matrices and scatterplots.
Dependence and grouped structure						
18	●● Grp	Design check for pseudoreplication	Measurements may have been treated as independent when the design does not support that assumption.	No formal residual null; the target is correct unit-of-analysis logic.	`No single function`	This is a design problem, not a residual issue. Mixed models handle structured dependence, but they do not rescue true pseudoreplication.
19	●● Grp	Linear mixed model	Gaussian response with grouped, repeated, or clustered observations.	Model target: fixed effects plus random-group variation.	`lme4::lmer(y ~ x + (1 | group), data = df)`	Use when dependence is structured by design, for example shared sites, individuals, years, or plots. This is not a repair for $n = 1$ pseudoreplication.
19	●● Grp	Generalised linear mixed model	Non-Gaussian response with grouped, repeated, or clustered observations.	Model target: non-Gaussian mean structure plus random effects.	`lme4::glmer(y ~ x + (1 | group), family = ..., data = df)`	Extends mixed-model logic to binomial or count responses.
19	●● Grp	Intra-class correlation (ICC)	Variance partitioning is needed after fitting a mixed model.	No null; the target is the proportion of variance attributable to groups.	`VarCorr(mod)`, *then compute ICC*	Used to summarise the strength of dependence at the grouping level.
Generalised linear models and proportion/count responses						
20	0/1 Bin	One-sample proportion test	A single observed proportion is compared with a null expectation.	H_0: $p = p_0$.	`prop.test(x = s, n = n, p = p0)`	Presented as a bridge to binomial GLMs when predictors are absent.
20	0/1 Bin	Two-sample proportion test	Two binomial proportions are compared without additional predictors.	H_0: $p_1 = p_2$.	`prop.test(x = c(s1, s2), n = c(n1, n2))` `prop.test(matrix_counts)`	Useful for simple comparisons; move to a binomial GLM when predictors enter.
20	0/1 Bin	Logistic regression (Bernoulli response)	Binary presence/absence or success/failure response with predictors.	Model target: $\operatorname{logit}\{P(y = 1)\} = \beta_0 + \beta_1 x + \cdots$.	`glm(y ~ x1 + x2, family = binomial, data = df)`	Use `type = "response"` in `predict()` to recover fitted probabilities.
20	0/1 Bin	Binomial GLM for grouped counts	Successes and failures are aggregated for each row.	Model target: $\operatorname{logit}(p) = \beta_0 + \beta_1 x + \cdots$.	`glm(cbind(success, failure) ~ x, family = binomial, data = df)`	Same link and interpretation as logistic regression, but with grouped trials.
20	# Cnt	Poisson regression	Count response with mean approximately equal to the variance.	Model target: $\log\{E[y]\} = \beta_0 + \beta_1 x + \cdots$.	`glm(count ~ x, family = poisson, data = df)`	Check for overdispersion before trusting Poisson standard errors.
20	# Cnt	Quasi-Poisson regression	Count response shows overdispersion relative to Poisson.	Model target: $\log\{E[y]\} = \beta_0 + \beta_1 x + \cdots$, with overdispersed variance.	`glm(count ~ x, family = quasipoisson, data = df)`	Retains the log-link mean model but adjusts the variance.
Flexible and nonlinear regression						
21	∼ Sm	Gaussian GAM	Relationship is smooth and nonlinear, but a fixed polynomial form is not appropriate.	Model target: smooth mean function estimated from the data.	`mgcv::gam(y ~ s(x), data = df, method = "REML")`	Use `mgcv::gam()` or load `library(mgcv)`. Use cyclic bases such as `bs = "cc"` for seasonal smooths.
22	$f(x)$ NL	Nonlinear least-squares regression	Mean structure follows a named nonlinear function with interpretable parameters.	Model target: explicit nonlinear response curve, for example Michaelis–Menten.	`nls(y ~ f(x, pars), start = list(...), data = df)`	Requires sensible starting values and convergence checking.
22	$f(x)$ NL	Linear-vs-nonlinear model comparison	A nonlinear model is compared with a simpler linear alternative.	Relative support for alternative mean structures.	`anova(lm_mod, nls_mod)` `AIC(lm_mod, nls_mod)`	Interpret with care because `nls()` and `lm()` may reflect different biological questions.
23	Q Q	Quantile regression	Predictor effects may differ across the response distribution rather than only at the mean.	Model target: conditional quantile(s), not conditional mean alone.	`quantreg::rq(y ~ x, tau = 0.5, data = df)`	Typical choices are `tau = 0.1`, `0.5`, and `0.9`.
23	Q Q	Quantile-slope comparison	Quantile regressions have been fitted at several τ values and slope differences need testing.	H_0: slopes do not differ across the fitted quantiles.	`anova(rq_fit_multi_tau)`	Useful when upper, median, and lower response boundaries behave differently.
Prediction, evaluation, and regularisation						
17,24	$\hat{y}$ Pred	Hold-out RMSE	Prediction quality is assessed on data not used for fitting.	No null; lower RMSE indicates better predictive performance.	`pred <- predict(mod, newdata = test)` `sqrt(mean((obs - pred)^2))`	Prediction and explanation are different goals; a more interpretable model need not minimise prediction error.

Continued on next page

(Continued)

Ch.	Cue	Method	Use When	Null / Model Target	Main R Function(s)	Notes, Assumptions, or Follow-up
17	$\hat{y}$ **Pred**	K-fold cross-validation for Gaussian models	Generalisation error is estimated by repeated train/test splits across folds.	No null; compare average out-of-fold RMSE across candidate models.	fit with `lm()` inside each fold, use `predict()`, then summarise RMSE	Implemented manually in the chapter rather than through a single package wrapper.
25	$\hat{y}$ **Pred**	Ridge regression	Many correlated predictors are present and coefficient shrinkage is acceptable.	Model target: stable prediction with L_2 penalty.	`glmnet::cv.glmnet(X, y, alpha = 0)`	Shrinks coefficients toward zero but typically keeps all of them non-zero.
25	$\hat{y}$ **Pred**	Lasso regression	Prediction and variable selection are both of interest.	Model target: sparse regression with L_1 penalty.	`glmnet::cv.glmnet(X, y, alpha = 1)`	Can set some coefficients exactly to zero.
25	$\hat{y}$ **Pred**	Elastic net	Predictors are correlated and a compromise between ridge and lasso is needed.	Model target: mixed L_1/L_2 penalty.	`glmnet::cv.glmnet(X, y, alpha = 0.5)`	Choose `alpha` between 0 and 1. Particularly useful when predictors cluster strongly.

Design first. Choose the method from the data structure and not from the software options. The same response variable may require a *t*-test, ANOVA, regression, mixed model, or GLM depending on independence, grouping, pairing, and whether predictors are categorical or continuous.

Response scale. Gaussian methods such as `t.test()`, `aov()`, and `lm()` assume that the analysis is defensible on the chosen response scale. For `aov()` and `lm()`, this is checked mainly through residuals. For `t.test()`, checks apply to the relevant raw values or paired differences. If the assumptions fail, transform the response or move to a method better matched to the data.

Model hierarchy. When an interaction is included, keep its main effects. Use `y ~ x1 * x2` so the interaction and main effects are retained together. When grouped structure is present, fit the grouping rather than pretending the rows are independent.

Model comparison. Use nested `anova(..., test = "F")` when one model is a special case of another. Use `AIC()` for broader candidate-model comparison, but not across different response scales such as y versus $\log(y)$ or across models fitted to different numbers of observations after NA-driven row loss. Use hold-out or cross-validated RMSE when prediction is the objective.

Cue legend. μ **G** Gaussian continuous; ◇ **Rk** rank-based; ↔ **Asc** association; 0/1 **Bin** binary or proportion; # **Cnt** count; ●● **Grp** grouped or repeated; ∼ **Sm** smooth model; $f(x)$ **NL** nonlinear; Q **Q** quantile; $\hat{y}$ **Pred** predictive workflow; Δ **Any** model-comparison workflow.

Icons. Icons are mnemonics: μ mean-based Gaussian, ◇ rank order, ↔ paired association, 0/1 binary response, # counts, ●● clustered observations, ∼ smooth function, $f(x)$ explicit nonlinear form, Q quantiles, $\hat{y}$ prediction, Δ model comparison.

Row tinting. Cream rows mark diagnostics or assumption checks. Green rows mark workflow, comparison, or prediction entries rather than primary fitted procedures.

Abbreviations. GLM = generalised linear model; GAM = generalised additive model; ICC = intra-class correlation; RMSE = root mean squared error; VIF = variance inflation factor.

Packages beyond base R used in the course.

- `car` for `leveneTest()` and `vif()`
- `lmtest` for `bptest()`
- `lme4` for `lmer()` and `glmer()`
- `mgcv` for `gam()`
- `quantreg` for `rq()`
- `glmnet` for ridge, lasso, and elastic net

Not every chapter adds a new test. Chapters 1–5, 10, 18, 24, and 26 mainly refine framing, design logic, model choice, interpretation, prediction, and workflow rather than introducing a brand-new inferential procedure.

Map-to-sheet connection. Use the course synthesis companion to decide whether a model is needed and which inferential frame applies (Chapters 1–5). Use this sheet to decide how to fit, check, and report it (Chapters 6–25).

Safer Use Checks

Replication Checks

- The sampling unit is the smallest independently sampled biological or experimental replicate to which inference is intended to apply.
- One-sample methods: one value per independent sampling unit.
- Two-sample methods: one row per independent sampling unit, and no sampling unit appears in both groups.
- Paired methods: the sampling unit is the matched pair; analyse within-pair differences, not raw columns as independent samples.
- ANOVA and ANCOVA: groups must contain independent sampling units unless grouping is modelled explicitly.
- Correlation: both variables are measured on the same sampling units; neither variable is a replicated treatment label in disguise.
- Regression: repeated rows from the same site, tank, plot, year, or individual are not independent sampling units.
- Visual red flag: if the same ‘ID’, ‘Site’, or equivalent grouping label appears in multiple rows without being modelled, the analysis is likely pseudoreplicated.
- Interaction hierarchy: once an interaction is retained, keep the main effects with it. Use `y ~ x1 * x2`, not an interaction-only form.
- Mixed and grouped models: use when rows are clustered within higher-level sampling units. They partition variance and dependence; they do not create new independent replicates.

Do Nots

- Do not choose a test from the software menu before identifying the sampling unit and dependence structure.
- Do not treat technical replicates, subsamples, or repeated measures as biological replicates.
- Do not use Student’s two-sample *t*-test by default; Welch is usually the safer default.
- Do not read a significant omnibus ANOVA as evidence that every pair of groups differs.
- Do not read a correlation coefficient as a causal effect or as a substitute for a regression with a defined response.
- Do not use residual normality tests to rescue a misspecified mean structure; fix the model first.
- Do not drop main effects when an interaction is retained.
- Do not interpret collinear coefficients as stable biological effects without checking VIFs, plots, and study design.
- Do not use AIC across different response datasets or silently different missing-data subsets.
- Do not treat lower predictive error as proof of better biological explanation.

Reporting Checklist

- *t*-tests and ANOVA: group summaries, effect direction, effect size, test statistic, degrees of freedom, *p*-value, and interval where relevant.
- Rank tests: sample summaries, test statistic, *p*-value, and a clear statement that inference is rank-based.
- Correlation: coefficient type, coefficient estimate, sample size, confidence interval if available, and *p*-value.
- Linear models: formula, coefficients, effect sizes (β estimates), standard errors, test statistics, *p*-values, uncertainty intervals, and residual-diagnostic conclusion.
- Mixed models: fixed-effect estimates, random-effect structure, variance components or ICC, and the grouping variable.
- GLMs: family and link, coefficients on the model scale, transformed interpretation where needed, and any dispersion check.
- GAMs and nonlinear models: fitted form, smoothing or parameter details, uncertainty summary, and graphical diagnostic conclusion.
- Predictive comparisons: data split or resampling scheme, performance metric, and whether the goal is explanation or prediction.