

# Course Synthesis Companion

BCB744 — Biostatistics Version 3 (3 April 2026)

Professor A.J. Smit

Map of the course chapters to complement the statistics methods cheat sheet. It is meant for revision, orientation, and linking ideas rather than rapid method lookup.

## HOW TO USE THIS PAGE

Read down the chapter map to recover the course narrative from foundations to model-based workflow.

Use the statistics methods cheat sheet when you need concrete test choice, assumptions, reporting details, or R functions.

Chapter / Focus	Key Learning
<b>Chapter 1</b>	
◦ Core	Statistics links biological questions to data, uncertainty, and decisions. Population, sample, parameter, and estimate are not interchangeable.
<b>Chapter 2</b>	
◦ Core	Descriptive statistics summarise centre, spread, and sample size, but they do not by themselves justify an inferential claim.
<b>Chapter 3</b>	
◦ Core	Graphs are part of the analysis. Choose plots that reveal structure, spread, outliers, and grouping before fitting models.
<b>Chapter 4</b>	
◦ Core	Probability distributions and sampling variation explain why estimates fluctuate. Standard errors and confidence intervals arise from repeated sampling.
<b>Chapter 5</b>	
◦ Core	Inference asks whether an observed pattern is large relative to sampling uncertainty. Hypotheses, test statistics, $p$ -values, and intervals answer different parts of that task.
<b>Chapter 6</b>	
μ G	Assumptions attach to the scale and design of the analysis, so check normality, equal variance, and transformation needs before committing to Gaussian procedures.
<b>Chapter 7</b>	
μ G	$t$ -tests compare means for one-sample, paired, and two-group designs. Pairing changes the analysis, and Welch is usually safer than pooled-variance Student's $t$ .
<b>Chapter 8</b>	
μ G	ANOVA is the multi-group extension of mean comparison. The global test asks whether any group means differ, and post hoc procedures locate the differences.
<b>Chapter 9</b>	
↔ Asc	Correlation quantifies association, not causation. Pearson targets linear association, whereas Spearman and Kendall target monotonic rank association.
<b>Chapter 10</b>	
Δ Work	Method selection follows data structure: response type, number of groups, predictor type, pairing, and independence determine the inferential methods.
<b>Chapter 11</b>	
μ G	Regression assumptions are checked on residuals because fitting removes the systematic mean structure. Therefore, use residual plots to guide model revision.
<b>Chapter 12</b>	
μ G	Simple linear regression models the conditional mean as a line. The slope inference, fitted values, and confidence versus prediction intervals answer different questions.
<b>Chapter 13</b>	
μ G	Polynomial terms allow curvature in the mean function, but coefficients should be interpreted through the fitted curve rather than term by term.
<b>Chapter 14</b>	
μ G	Multiple regression estimates partial effects while adjusting for other predictors. Model specification determines which biological questions the coefficients answer.
<b>Chapter 15</b>	
μ G	Interactions test whether the effect of one predictor changes with another. Once an interaction is present, main effects cannot be interpreted as universal averages.
<b>Chapter 16</b>	

Chapter / Focus	Key Learning
$\mu$ G	Collinearity, confounding, and measurement error destabilise coefficient interpretation. Unstable estimates are often a design or variable-definition problem.
<b>Chapter 17</b> $\Delta$ Work	Model checking separates explanation from prediction. Compare models with diagnostics, nested tests, information criteria, and out-of-sample error according to the objective.
<b>Chapter 18</b> ● Grp	Pseudoreplication is a design error in which non-independent observations are analysed as if they were independent. No later statistical fix rescues bad sampling design.
<b>Chapter 19</b> ● Grp	Mixed models represent grouped or repeated-measures data by combining fixed effects with random effects. Dependence must be modelled and not simply ignored.
<b>Chapter 20</b> $\emptyset/1$ Bin	GLMs extend regression beyond Gaussian responses by matching the response distribution and link to counts, proportions, and binary outcomes.
<b>Chapter 21</b> $\sim$ Sm	GAMs replace rigid functional forms with smooth terms when the response pattern is nonlinear but not well captured by a specific polynomial or parametric curve.
<b>Chapter 22</b> $f(x)$ NL	Nonlinear regression is appropriate when biology implies a specific nonlinear equation. Now, parameter meaning comes from the model form and biological theory. Starting values often matter.
<b>Chapter 23</b> Q Q	Quantile regression models conditional quantiles rather than only the conditional mean. It reveals how effects differ across the response distribution.
<b>Chapter 24</b> $\hat{y}$ Pred	Explanation and prediction are related but distinct goals. The best explanatory model need not be the best predictive model, and vice versa.
<b>Chapter 25</b> $\hat{y}$ Pred	Regularisation trades some bias for lower variance and better prediction in high-dimensional or collinear settings. Ridge shrinks, lasso selects, and elastic net blends both.
<b>Chapter 26</b> $\Delta$ Work	Reproducible workflow is part of statistical practice, so script the analysis, record decisions, manage outputs systematically, and keep the inferential path inspectable.