
STAT5003

Computational Statistical Methods

Independent Study Review Notes

Weeks 1–13 • Key Formulas • Comparisons • R Code

Assessment context recorded in the working notes only — not an official assessment document; verify current requirements independently.

Exam Format: 12 MCQ (24 marks) + 4 Extended Answer (36 marks) | 60 marks total

Allowed: Non-programmable calculator + 1 A4 handwritten notes sheet (double-sided)

MCQ: Each question has **two** correct answers — 1 mark each, no penalty for wrong selections

Independent study notice

These are independently authored review notes prepared from the author's study of STAT5003 (Computational Statistical Methods) at the University of Sydney in Semester 1, 2026. They are not official University of Sydney materials, are not endorsed by or affiliated with the University, and do not reproduce or redistribute university lecture slides, readings, assessments, answer keys, or other course resources. Explanations, examples, and diagrams were created by the author, with external sources identified where applicable. University/course names are used only to identify the subject context.

Contents

Quick Reference: Key Formulas	3
1 Week 1: Basics of Statistical Computing and Visualization	5
1.1 Descriptive Statistics	5
1.2 Data Types and R Workflow	5
1.3 Base R vs Tidyverse	5
2 Week 2: Regression and Smoothing	6
2.1 Statistical Learning Framework	6
2.2 Simple Linear Regression	6
2.3 Inference for Coefficients	7
2.4 Multiple Linear Regression	7
2.5 Non-parametric Regression (Smoothing)	8
3 Week 3: Density Estimation	8
3.1 Probability Distribution Functions	8
3.2 Maximum Likelihood Estimation	9
3.3 Kernel Density Estimation	9
3.4 Bias-Variance Tradeoff	10
4 Week 4: Clustering and Dimension Reduction	11
4.1 K-Means Clustering	11
4.2 Hierarchical Clustering	12
4.3 Principal Components Analysis	13
4.4 t-SNE and MDS	14
5 Week 5: Classification Techniques	14
5.1 Classification Basics	14
5.2 Logistic Regression	15
5.3 Linear Discriminant Analysis	16
5.4 k-Nearest Neighbours	16
5.5 Support Vector Machines	16
5.6 Method Comparison	17
6 Week 6: Cross-Validation and Classification Evaluation	17
6.1 Training Error vs Test Error	18
6.2 Validation Set Approach	18
6.3 k-Fold Cross-Validation	18
6.4 Repeated k-Fold Cross-Validation	18
6.5 Confusion Matrix and Metrics	19
6.6 Multi-Class Classification	20
6.7 ROC Curve and AUC	20

7 Week 7: Bootstrap, Missing Data, and Class Imbalance	20
7.1 The Bootstrap	20
7.2 Missing Data	21
7.3 Class Imbalance	21
8 Week 8: Feature Selection and Regularisation	21
8.1 Subset Selection	22
8.2 Model Selection Criteria	22
8.3 Ridge Regression	22
8.4 The Lasso	23
8.5 Ridge vs Lasso vs Elastic Net	23
9 Week 9: Decision Trees and Ensemble Methods	23
9.1 Decision Trees	23
9.2 Cost-Complexity Pruning	24
9.3 Bagging and Random Forests	24
9.4 Boosting	24
10 Week 10: Monte Carlo Methods	25
10.1 Monte Carlo Integration	25
10.2 Simulating Random Variables	25
10.3 Applications	26
10.4 Exam-Style Worked-Example Templates	26
11 Week 11: Markov Chain Monte Carlo (MCMC)	26
11.1 Monte Carlo vs MCMC	26
11.2 Markov Chains	27
11.3 Metropolis-Hastings Algorithm	27
11.4 MCMC for Bayesian Inference	28
12 Cross-Topic Comparison Tables	28
12.1 All Classification Methods	29
12.2 Bagging vs Random Forest vs Boosting	29
12.3 All Resampling Methods	29
12.4 Ridge vs Lasso	29
12.5 Missing Data Mechanisms	30
13 Complete R Functions Reference	30
14 Revision Prompts by Week	31

Quick Reference: Key Formulas

Regression and Smoothing

Simple Linear Regression: $Y = \beta_0 + \beta_1 X + \varepsilon$

Least Squares: $b_1 = \frac{\text{Cov}(x, y)}{\text{Var}(x)}, \quad b_0 = \bar{y} - b_1 \bar{x}$

R^2 : $R^2 = 1 - \frac{\text{RSS}}{\text{TSS}}$

KDE: $\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{X_i - x}{h}\right)$

Symbol key: RSS (Residual Sum of Squares) = $\sum (y_i - \hat{y}_i)^2$ — total squared prediction error. TSS (Total Sum of Squares) = $\sum (y_i - \bar{y})^2$ — total sum of squares (proportional to sample variance). h = bandwidth, K = kernel function, n = sample size.

Classification

Logistic Regression: $\log\left(\frac{p}{1-p}\right) = \mathbf{X}\boldsymbol{\beta}, \quad p = \frac{1}{1 + \exp(-\mathbf{X}\boldsymbol{\beta})}$

Log-Loss: $\ell(\boldsymbol{\beta}) = \sum_{i=1}^n [-y_i \log p(\mathbf{x}_i) - (1 - y_i) \log(1 - p(\mathbf{x}_i))]$

LDA Posterior: $p_k(\mathbf{x}) = \frac{\pi_k f_k(\mathbf{x})}{\sum_{\ell=1}^K \pi_\ell f_\ell(\mathbf{x})}$

kNN: $P(Y = \ell \mid \mathbf{x}) = \frac{1}{k} \sum_{i \in N_{\mathbf{x}}^k} \mathbf{1}_{\{y_i = \ell\}}$

SVM Margin: $\max M \text{ s.t. } y_i(\beta_0 + \mathbf{x}_i \boldsymbol{\beta}) \geq M(1 - \epsilon_i), \sum_{j=1}^p \beta_j^2 = 1, \sum \epsilon_i \leq C$

Model Selection and Regularisation

Ridge: $\text{RSS} + \lambda \sum_{j=1}^p \beta_j^2$

Lasso: $\text{RSS} + \lambda \sum_{j=1}^p |\beta_j|$

Mallow's C_p : $C_p = \frac{1}{n} (\text{RSS} + 2d\hat{\sigma}^2)$

BIC: $\text{BIC} = \frac{1}{n} (\text{RSS} + \log(n) \cdot d \cdot \hat{\sigma}^2)$

AIC: $\text{AIC} = -2\ell(\hat{\boldsymbol{\theta}}) + 2k$, where k is the number of estimated parameters.

Adjusted R^2 : $1 - \frac{\text{RSS}/(n-d-1)}{\text{TSS}/(n-1)}$

Bias-Variance and Evaluation

Bias-Variance Decomposition: $\mathbb{E}[(Y - \hat{f}(X))^2] = \text{Var}(\hat{f}(X)) + [\text{Bias}(\hat{f}(X))]^2 + \text{Var}(\varepsilon)$

k-Fold CV: $\text{CV}_k = \frac{1}{k} \sum_{i=1}^k \text{MSE}_i$

Accuracy = $\frac{\text{TP} + \text{TN}}{\text{Total}}$ **Sensitivity** = $\frac{\text{TP}}{\text{TP} + \text{FN}}$ **Specificity** = $\frac{\text{TN}}{\text{TN} + \text{FP}}$

Precision = $\frac{\text{TP}}{\text{TP} + \text{FP}}$ $F_1 = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$ $\kappa = \frac{P_o - P_e}{1 - P_e}$

Trees, Ensembles and MCMC

Bagging: $\hat{f}_{\text{bag}}(x) = \frac{1}{B} \sum_{b=1}^B \hat{f}^{*b}(x)$

Gini: $G = \sum_{k=1}^K \hat{p}_{mk}(1 - \hat{p}_{mk})$

Boosting: $\hat{f}(x) \leftarrow \hat{f}(x) + \lambda \hat{f}^b(x)$, residuals $r_i \leftarrow r_i - \lambda \hat{f}^b(x_i)$

M-H Acceptance: $\alpha(X_t, Y) = \min\left(\frac{f(Y) \cdot q(X_t | Y)}{f(X_t) \cdot q(Y | X_t)}, 1\right)$

Symmetric proposal: $\alpha = \min\left(\frac{f(Y)}{f(X_t)}, 1\right)$

Bayes: $p(\theta | D) = \frac{p(D | \theta) \cdot p(\theta)}{p(D)}$