MONASH University

MONASH BUSINESS SCHOOL

Monash Business Analytics Major

Unit Summaries

Di Cook and Rob Hyndman

1 August 2025

Business Analytics is a major area of study in the Bachelor of Commerce, and has been in operation since 2017. Students completing this major are expected to be able to:

1. Use their data wrangling, analysis and visualisation skills to inform business decisions made in a variety of settings.
2. Employ reliable statistical models and machine learning algorithms in practice to understand uncertainty, and predict future outcomes.
3. Interpret and communicate the results of data analysis and modelling, using collaborative and reproducible practices.

A recent review led to small changes that are planned for 2025 and beyond, reflected in this document.

Students must complete 3 core units and 5 elective units. At least 2 of the elective units must be at Level 3.

The following are core units and must be taken by all students:

- ETC1000: Business and economic statistics (Level 1)
- ETC1010: Introduction to data analysis (Level 1)
- ETC2420: Statistical thinking (Level 2)
- ETC3250: Introduction to machine learning (Level 3)

The following are elective units that can contribute towards the major (first digit indicates level):

- ETC2410 Introductory econometrics
- BEX3726: Engaging with international business
- ETC3550: Applied forecasting
- ETC3555: Statistical machine learning
- ETC3580: Advanced statistical modelling
- ETF3500: High dimensional data analysis
- FIT1045: Algorithms and programming fundamentals in python
- FIT2094: Databases
- FIT3003: Business intelligence and data warehousing
- FIT3154: Advanced data analysis
- FIT3171: Databases
- FIT3179: Data visualisation

ETC1000 Business and economic statistics

Unit synopsis

Presentation and analysis of grouped and ungrouped data using tables, charts and measures of location and dispersion; standardisation techniques, including index numbers, with application to share price indices and the CPI; analysis of relationships between variables using simple and multiple regression; extensions to multiple regression, including nonlinear and categorical explanatory variables and time series data with applications to forecasting; main ideas of probability theory as a foundation for statistical inference; concept of sampling as a way of capturing uncertainty about data; estimators and their properties; constructing and interpreting confidence intervals; fundamentals of hypothesis testing, testing hypotheses about the mean, difference between means and the slope, including analysis of variance.

Learning outcomes

1. Interpret business and economic data using tables, charts and descriptive statistics techniques, applying standardisation techniques where appropriate.
2. Describe the concept of a sampling distribution, estimators and their properties
3. Make inference on single and multiple population means, difference between means and the slope for business and economic decision-making.
4. Interpret and evaluate relationships between variables for business and economic decision-making using simple and multiple linear regression.
5. Apply the main ideas of probability theory to business and economic decision-making

Topics covered include

- Descriptives for categorical data
- Descriptives for numerical data
- Correlation and simple regression
- Sampling distributions and confidence intervals
- Hypothesis testing
- Multiple regression modeling
- Time series modeling
- Spreadsheet modeling

ETC1010: Introduction to data analysis

Unit synopsis

This unit introduces principles and techniques for exploring different types of data. You will learn concepts and practice in taking "data from the wild", reading different formats, tidying and wrangling it into shape for analysis, and creating useful visualisations to achieve effective data-driven decision-making. You will learn to use data to solve problems, and develop oral and written communication skills, using hands-on learning and live data analysis with code.

Learning outcomes

1. Learn to read different data formats, learn about tidy data and wrangling techniques.
2. Apply effective visualisation, descriptive analysis and modelling to understand relationships between variables, and make decisions with data.
3. Develop computing, communication, and reproducible reporting skills.

Topics covered include

- Workflow and introduction to R
- Tidy data
- Wrangling verbs/operations
- Visualising data
- Missing value handling
- Working with temporal and spatial variables
- Data collection: scraping, JSON
- Basic statistical models: form and fitting and diagnosis
- Basic computational models: algorithm (regression & classification trees), stopping, diagnostics
- Text analysis: wrangling text into data, sentiment tagging
- Basic algorithmic clustering

Textbook

Wickham, Çetinkaya-Rundel & Grolemund (2023). *R for Data Science*. 2nd edition, O'Reilly. https://r4ds.hadley.nz

ETC2410: Introductory econometrics

Unit synopsis

This unit introduces you to the empirical analysis of relationships between economic variables. The approach is based on linear regression theory, and emphasises 'hands on' data analysis. Topics studied will include properties of least squares estimators, hypothesis testing, the choice of appropriate functional form, the use of dummy variables, issues around modelling survey data and the problems of serial correlation, heteroscedasticity and multicollinearity.

Learning outcomes

1. Understand and derive the properties of ordinary least squares in summation and matrix notation.
2. Interpret, evaluate and apply inferential methods to multiple linear regression.
3. Understand the use and implications of data scaling, functional form and dummy variables in regression modelling.
4. Identify the presence of heteroscedasticity, adjust OLS standard errors and perform feasible GLS in regression models.
5. Understand issues related to modelling with time-series data.

Topics covered include

- Introduction to Econometrics
- Review of Statistical Concepts
- Simple and Multiple Regression
- Multiple Regression - Inference
- Qualitative Information and Dummy Variables
- Functional Form and Model Selection
- Statistical Properties of the OLS Estimator
- Heteroskedasticity
- Serial Correlation
- Modelling Dynamics
- Large Sample Properties of the OLS Estimator

Textbook

Wooldridge (2020). *Introductory Econometrics: A Modern Approach*, 7th edition, Cengage.

ETC2420: Statistical thinking

Unit synopsis

This unit presents data analysis, statistical modelling and decision-making in the presence of uncertainty, using a computational approach. You will use different frameworks of probability that are helpful for analysing real-world problems. Topics covered will include probability distributions, statistical inference (classical & Bayesian), simulation, permutation and randomisation methods, regression models, decision theory, and model assessment and diagnosis.

Learning outcomes

1. Characterise and understand uncertainty using data.
2. Build statistical models to support decision-making, hypothesis testing and risk assessment.
3. Use randomisation methods in data collection and to assess causality and uncertainty.
4. Learn about and use concepts from probability.
5. Understand Bayesian and frequentist approaches to statistical modelling.
6. Further develop computational skills for statistical analysis.

Topics covered include

- Review of tidy data
- Estimation
- Hypothesis testing
- Resampling techniques, including the bootstrap
- Probability and probability distributions
- Randomisation
- Maximum likelihood estimation
- Bayesian inference
- Decision theory
- Regression models, including multiple regression
- Model performance and diagnostics

Recommended reading

Introductory Statistics with Randomisation and Simulation, by D. M. Diez, C. D. Barr and M. Çetinkaya-Rundel, published by OpenIntro, 2014.

Modern Dive: Statistical Inference via Data Science, by C. Ismay and A. Y. Kim. Hardcopy published by CRC Press, 2020.

R for Data Science (2nd edition), by H. Wickham, M. Çetinkaya-Rundel and Garrett Grolemund. Hardcopy published by O'Reilly Media, Inc, 2023.

Statistical Thinking for the 21st Century, by R. A. Poldrack.

Doing Bayesian Data Analysis: A tutorial with R, JAGS and Stan, by J. K. Kruschke. 2nd edition, published by Elsevier, 2015.

ETC3250: Introduction to machine learning

Unit synopsis

This unit develops your ability to model multi-dimensional data using statistical and machine learning techniques. Topics covered include: dimension reduction with linear and nonlinear methods; supervised learning such as discriminant analysis, decision trees and forests, neural networks; and unsupervised learning such as k-means, hierarchical and model-based clustering. You will learn about conceptualising problems using the bias-variance trade-off and how to balance this when fitting models. Complex model fitting techniques will be covered including bagging, boosting, cross-validation, regularisation and constructing ensembles. An important component is learning how to diagnose your model, especially utilising high-dimensional visualisation methods, and explain your model with explainable artificial intelligence (XAI). You will develop practical skills in applying techniques to different problems using a suitable software environment that involves doing reproducible analyses.

Learning outcomes

1. Develop, select, and diagnose statistical and machine learning methods for supervised and unsupervised tasks.
2. Measure the uncertainty of a prediction or classification using resampling methods.
3. Efficiently conduct analysis tasks in a contemporary software environment.
4. Explain and interpret the analyses undertaken clearly and effectively.
5. Apply analytic tools to contemporary business problems.

Topics covered include

- Conceptual overview of ML, bias/variance, diagnostics, model choice, feature engineering, math/notation/computing
- Resampling: LOOCV, cross-validation, training/test
- Dimension reduction: PCA, PP, MDS, regularisation
- High-d vis: tours, par coords, mosaic plots
- Classification: logistic, LDA, trees, SVM, NN
- Ensembles and boosting: forest, xgboost
- Explainability and interpretability
- Unsupervised classification: k-means, hierarchical clustering, mixture models
- Model choice, diagnostics

Textbook

See the materials on the unit website https://iml.numbat.space.

- James, Witten, Hastie, & Taylor (2021). *An Introduction to Statistical Learning*. 2nd edition, Springer. https://www.statlearning.com
- Boehmke & Greenwell (2020) *Hands-On Machine Learning with R* https://bradleyboehmke.github.io/HOML/
- Kuhn and Silge (2023) *Tidy Modeling with R* https://www.tmwr.org

- Molnar (2023) *Interpretable Machine Learning* https://christophm.github.io/interpretable-ml-book/
- Hvitfeldt (2024) *Feature Engineering A-Z* https://feaz-book.com
- Cook & Laa (2024) *Interactively Exploring High-dimensional Data and Models in R* https://dicook.github.io/mulgar_book/

BEX3726: Engaging with international business

Unit synopsis

This unit requires you to examine a broad range of social, cultural, legal and political issues that can impact on the strategy and operations of businesses operating in a global business environment. Through the use of readings, case studies, and an analysis of current events you are confronted with some specific challenges of doing business in a global context. This unit focuses on developing skills, addressing various global issues and how they affect business, as well as aspects of global citizenship, cultural sensitivity and effective teamwork.

Learning outcomes

1. Recognise issues affecting businesses operating in a global context.
2. Develop and present strategic responses to specific issues affecting businesses operating in a global context with due consideration to sustainable, responsible and ethical business practices.
3. Acquire team skills including team leadership, collaboration, problem-solving, decision-making, communication and presentation skills.
4. Develop reflective practice skills in relation to individual learning.

ETC3550: Applied forecasting

Unit synopsis

Business and economic data are collected over time against a backdrop of a changing environment. This unit introduces data wrangling and exploratory data analysis for such temporal context data. You will learn how to analyse, model and forecast time series data in diverse applications. Methods to be covered include the decomposition of time series, exponential smoothing methods, ARIMA modelling, and regression with auto-correlated disturbances. You will enhance your computational skills with exercises using R.

Learning outcomes

1. Understand common statistical methods used in business and economic forecasting.
2. Develop computer skills for forecasting business and economics time series data.
3. Provide insights into the problems of implementing and operating large scale forecasting systems.

Topics covered include

- Introduction to forecasting and R
- Time series graphics
- Time series decomposition
- The forecaster's toolbox
- Exponential smoothing
- Forecasting with ARIMA models
- Multiple regression and forecasting
- Dynamic regression

Textbook

Hyndman & Athansopoulos (2021). *Forecasting: principles and practice*. 3rd edition, OTexts. https://OTexts.com/fpp3/

ETC3555: Statistical machine learning

Unit synopsis

This unit covers the methods and practice of statistical machine learning for modern data analysis problems and takes a deeper look at many of the methods introduced in ETC5250: Introduction to Machine Learning. Particular focus will be placed on how to effectively learn model parameters and methods to guard against overfitting. Topics covered will include deep neural networks with dropout, image recognition, text mining and generation, and sparse multivariate methods. All computing will be conducted using the R programming language.

Learning outcomes

1. Identify and understand the statistical and computational trade-offs in modern data analysis problems.
2. Understand and apply machine learning algorithms to solve modern data analysis problems.

Topics covered include

- The learning model, hypothesis sets,
- Perceptron, Perceptron Learning Algorithm (PLA)
- Feasibility of learning, Hoefddings's inequality
- Linear models, Logistic regression, Cross-entropy
- (Stochastic) gradient descent
- Neural Networks, activation functions, backpropagation algorithm
- Deep neural networks (Keras to R interface)
- CNNs (Convolutional neural network)
- SVM, margin classifier, kernels
- Text mining, Vector space model, Bag of Words, N-grams
- Text and sequence learning, word embeddings, word2vec, RNN (Recurrent neural networks), LSTMs (Long Short Term Memory)
- Implementation of most algorithms used in class from scratch using matrix operation (Linear models, gradient descent, Neural networks with matrices by hand and with code)
- Applied Machine Learning project as group work

ETC3580: Advanced statistical modeling

Unit synopsis

This unit introduces extensions of linear regression models for handling a wide variety of data analysis problems. Three extensions will be considered: generalised linear models for handling counts and binary data; mixed-effect models for handling data with a grouped or hierarchical structure; and non-parametric regression for handling non-linear relationships. All computing will be conducted using R.

Learning outcomes

1. Understand statistical models for handling common data analysis problems.
2. Develop skills for fitting, interpreting and assessing statistical models.
3. Develop computer skills for exploring and modelling different kinds of data.

Topics covered include

- Graphics for linear models and interpreting interactions
- Leverage, influence, and residual analysis for linear models
- Regression with binary responses
- Regression with binomial and proportional responses
- Regression with count responses
- Generalized linear models
- Random effects
- Mixed-effects models
- Nonparametric regression
- Nonparametric inference
- Additive models

Textbook

Faraway (2016) *Extending the Linear Model with R: Generalized Linear, Mixed Effects and Nonparametric Regression Models*, 2nd edition, CRC Press. https://julianfaraway.github.io/faraway/ELM/

ETF3500: High-dimensional data analysis

Unit synopsis

In many fields of business, analysts must deal with data on many variables, for example, surveys with a large number of questions. In such cases, statistical tools known as multivariate methods must be used to analyse the data and drive business decisions.

This unit covers such methods in three sections: Cluster Analysis can be used to identify and predict differences between groups such as between distinct classes of customers or products; Principal Components Analysis, Correspondence Analysis and Multidimensional Scaling are dimension reduction methods that help analysts to visualise complicated datasets; and finally, Factor Analysis is used to explain and predict business outcomes.

Learning outcomes

1. Demonstrate an understanding of the role that multivariate statistical techniques such as factor analysis, categorical data analysis, cluster analysis, multidimensional scaling and correspondence analysis play in uncovering relationships and patterns in high dimensional data.
2. Appraise the strengths and limitations of these techniques.
3. Apply tools in R to analyse high dimensional data and generate solutions using the appropriate statistical techniques.
4. Demonstrate skills in using the appropriate statistical techniques from a user and provider perspective.
5. Demonstrate skills in communicating the results, including the visualization and reporting of the analysis, to aid decision making.

Topics covered include

- Distance and data
- Matrix algebra for multivariate analysis
- Cluster analysis
- Multidimensional scaling
- Principal components analysis
- Exploratory factor analysis
- Correspondence analysis

Main readings

Mardia, K. V., J.T. Kent, J.M. Bibby, (1979). *Multivariate Analysis.*

Izenman, A. J. (2008). *Modern multivariate statistical techniques. Regression, classification and manifold learning*

FIT1045: Introduction to programming

Unit synopsis

This unit introduces programming fundamentals using the Python language. It will present fundamental programming control structures, built-in and complex datatypes, mechanisms for modularity, and the use of basic libraries. Students will also be introduced to good programming practices and programming in teams.

Learning outcomes

1. Effectively use variables, conditionals and loops in computer programs.
2. Design, construct and test simple programs that include user-defined functions and data structures.
3. Write programs efficiently by discovering and leveraging libraries.
4. Implement good programming practices in a team, including unit testing, basic documentation and readability.

FIT2094: Databases

Unit synopsis

This unit will provide an introduction to the concepts of database design and usage and the related issues of data management. Students will develop skills in planning, designing, and implementing a data model using an enterprise-scale relational database system (Oracle). Methods and techniques will also be presented to populate, retrieve, update and implement integrity features on data in the implemented database system.

Learning outcomes

1. Apply the theories of the relational database model.
2. Develop a sound relational database design.
3. Implement a relational database based on a sound database design.
4. Manage data that meets user requirements, including queries and transactions.
5. Contrast the differences between non-relational database models and the relational database model.

FIT3003: Business intelligence and data warehousing

Unit synopsis

Automation and the use of technological tools have resulted in the accumulation of vast volumes of data by modern business organisations. Data warehouses have been set up as repositories to store this data and improved techniques now result in the speedy collection and integration of such data. OLAP technology has resulted in the faster generation of reports and more flexible analysis based on the data repositories. This unit will explore the concepts of data warehousing and OLAP, covering the data processing technological requirements for data warehousing and OLAP and will provide hands on experience on designing data warehousing and OLAP systems

Learning outcomes

1. Design multi-dimensional databases and data warehouses.
2. Use fact and dimensional modelling.
3. Implement online analytical processing (OLAP) queries.
4. Explain the roles of data warehousing architecture and the concepts of granularity in data warehousing.
5. Create business intelligence reports using data warehouses and OLAP.

FIT3154 - Advanced data analysis

Unit synopsis

This unit introduces the problem of machine learning and the major kinds of statistical learning used in data analysis. Learning and the different kinds of learning will be covered and their usage discussed. Evaluation techniques and typical application contexts will presented. A series of different models and algorithms will be presented in an exploratory way: looking at typical data, the basic models and algorithms and their use: linear and logistic regression, support vector machines, Bayesian networks, decision trees, random forests, k-means and clustering, neural-networks, deep learning, and others. Finally, two specialist topics will be covered briefly, statistical learning theory and working with big data.

Learning outcomes

1. Describe what machine learning is.
2. Differentiate kinds of statistical learning models and algorithms.
3. Evaluate a machine learning algorithm in typical contexts.
4. Describe and apply the major models and algorithms for statistical learning.
5. Identify the most competitive algorithms for typical contexts.
6. Compare and contrast the differences between big data applications and regular applications of algorithms.
7. Describe the theoretical limits of learning.

FIT3171: Databases

Unit synopsis

This unit will provide an introduction to the concepts of database design and usage and the related issues of data management. You will develop skills in planning, designing, and implementing a data model using an enterprise-scale relational database system (Oracle). Methods and techniques will also be presented to populate, retrieve, update and implement integrity features on data in the implemented database system.

Learning outcomes

1. Apply the theories of the relational database model.
2. Develop a sound relational database design.
3. Implement a relational database based on a sound database design.
4. Manage data that meets user requirements, including queries and transactions.
5. Differentiate non-relational database model and the relational database model.
6. Develop programming structures within a database backend.

FIT3179: Data visualisation

Unit synopsis

Data visualisation is a powerful technique that allows us to use our visual system to understand data. Interactive data visualisation is now common in business, engineering and design and the social and physical sciences. This unit introduces the main kinds of information graphics and interactive visualisation systems and their areas of application. It investigates the reasons why visualisation can be effective and based on this students will gain experience in critically assessing data visualisations and in designing their own visualisations. You will learn how to create visualisations with representative computer tools and gain experience in creating a data visualisation for an application domain of their choice.

Learning outcomes

1. Critically analyse data visualisations.
2. Create effective data visualisations;
3. Describe the main applications of data visualisation in business, engineering and design, and the social and physical sciences.
4. Describe the advantages, drawbacks and pitfalls of the visual presentation of data as compared to its presentation using other media.