H9SDA: Statistics for Data Analytics

Module Code:	H9SDA
Long Title	Statistics for Data Analytics APPROVED
Title	Statistics for Data Analytics
Module Level:	LEVEL 9
EQF Level:	7
EHEA Level:	Second Cycle
Credits:	10
Module Coordinator:	TONY DELANEY
Module Author:	Margarete Silva
Departments:	School of Computing
Specifications of the qualifications and experience required of staff	This module requires a lecturer holding a Master's degree or higher, in a discipline with a significant statistics component. e.g. Statistics, Mathematics, Economics.

Learning Outcomes	
On successful completion of this module the learner will be able to:	
#	**Learning Outcome Description**
LO1	Apply appropriate statistical inference techniques to the analysis of data across a variety of domains.
LO2	Interpret the outputs from statistical software packages and programming languages
LO3	Report and communicate statistical results in a comprehensive, ethical and professional manner
LO4	Apply appropriate forecasting techniques to time series.
LO5	Identify patterns in data and implement dimension reduction techniques.

Dependencies	
Module Recommendations	
No recommendations listed	
Co-requisite Modules	
No Co-requisite modules listed	
Entry requirements	A cognate level 8 degree. Introductory statistics.

H9SDA: Statistics for Data Analytics

Module Content & Assessment

Indicative Content

Review of Fundamental Statistical Concepts
Fundamentals of probability Sampling Estimation & confidence intervals Hypothesis testing & t-tests ANOVA techniques Hypothesis testing & ANOVA exercises

Review of Fundamental Statistical Concepts
Non-parametric tests Effect size in research & effect size metrics Statistical power and sample size Reporting test results Ethics in the use of data Correlation/partial correlation Basic linear regression models Exercises using non-parametric tools Examples of misuse of NHST Ethics in the use of data and statistical reporting

Topics in Multiple Linear Regression I
Model accuracy Qualitative variables Transformations Collinearity & collinearity diagnostics / VIF Diagnostics for leverage and influence Heteroscedasticity in regression models Correlation of error terms Non-linearity of data Use of statistical software & R to estimate regression models

Topics in Linear Regression II
Principles of regression model building Adjusted R2, AIC, BIC, Cp Best subset selection, forward selection, backward selection Modelling interactions Use of statistical software & R to estimate regression models

Logistic Regression
Principles behind the binary logistic regression model Odds & odds ratios The logit transformation Maximum likelihood estimation Estimating logistic regression coefficients Wald statistic – contribution of predictors Prediction using logistic regression Practical estimation of logistic regression models

Multinomial Logistic Regression & Linear Discriminant Analysis
Multinomial logistic regression Introduction to linear discriminant analysis Exercises in multinomial logistic regression and linear discriminant analysis

Dimension Reduction
Applications of PCA & exploratory factor analysis Suitability of data for PCA / factor analysis Kaiser's criterion Interpretation of principal components Factor rotation Clustering methods Practical application of PCA in R / statistical software

Multivariate Analysis of Variance (MANOVA)
ANOVA vs MANOVA Applications of MANOVA SSCP matrices MANOVA test statistics Interpretation of MANOVA software output Practical application of MANOVA

Bayesian Statistics
Frequentists vs. Bayesians Bayes rule & applications Introduction to Bayesian networks Bayesian Statistics problems

Time Series I
Decomposition of Time Series Seasonality Stationarity Data Transformations Mean & Linear Trend models Random Walk Models Averaging & smoothing models Auto regressive models Applications of time series forecasting

Time Series II
Non-seasonal ARIMA models Orders of AR and MA terms Seasonal ARIMA models Model estimation ARCH Applications of time series forecasting

Revision
Revision

Assessment Breakdown	%
Coursework	35.00%
End of Module Assessment	65.00%

Assessments

Full Time

Coursework

Assessment Type:	Continuous Assessment (0200)	**% of total:**	35
Assessment Date:	n/a	**Outcome addressed:**	1,2,3,4,5
Non-Marked:	No		

Assessment Description:
Learners, individually or in a group, will be directed towards appropriate datasets and asked to produce a statistical report that incorporates the estimation of statistical models and reports findings in an appropriate manner. Estimation using some/all of multiple linear regression, logistic regression, time series analysis and dimension reduction techniques is likely to be required

Assessment Type:	Formative Assessment	**% of total:**	Non-Marked
Assessment Date:	n/a	**Outcome addressed:**	1,2,3,4,5
Non-Marked:	Yes		

Assessment Description:
Formative assessment will be undertaken utilising exercises and short answer questions during certain tutorials. In class discussions will be undertaken on contemporary topics. Feedback will be provided individually or as a group in oral format.

End of Module Assessment

Assessment Type:	Terminal Exam	**% of total:**	65
Assessment Date:	End-of-Semester	**Outcome addressed:**	1,2,3,4,5
Non-Marked:	No		

Assessment Description:
The examination will be of two hours duration and may include a mix of: theoretical, applied and interpretation questions.

No Workplace Assessment

Reassessment Requirement

Repeat examination
Reassessment of this module will consist of a repeat examination. It is possible that there will also be a requirement to be reassessed in a coursework element.

H9SDA: Statistics for Data Analytics

Module Workload

Module Target Workload Hours 0 Hours

Workload: Full Time

Workload Type	Workload Description	Hours	Frequency	Average Weekly Learner Workload
Lecture	No Description	24	Every Week	24.00
Tutorial	No Description	24	Every Week	24.00
Independent Learning	No Description	202	Every Week	202.00
			Total Weekly Contact Hours	48.00

Workload: Part Time

Workload Type	Workload Description	Hours	Frequency	Average Weekly Learner Workload
Lecture	No Description	2	Every Week	2.00
Tutorial	No Description	2	Every Week	2.00
Independent Learning	No Description	17	Every Week	17.00
			Total Weekly Contact Hours	4.00

Module Resources

Recommended Book Resources

Carlos Cortinhas,Ken Black. (2012), Statistics for Business and Economics, 1st European Edition. John Wiley & Sons, p.862, [ISBN: 1119993660].

Jeremy J Foster,Emma Barkus,Christian Yavorsky. (2006), Understanding and Using Advanced Statistics, SAGE, p.178, [ISBN: 141290014X].

Wolfgang Karl Härdle,Léopold Simar. (2012), Applied Multivariate Statistical Analysis, Springer Science & Business Media, p.516, [ISBN: 978-3-642-17229-8].

Rob J. Hyndman,George Athanasopoulos. (2013), Forecasting, Otexts, p.292, [ISBN: 978-0987507105].

Gareth James,Daniela Witten,Trevor Hastie,Robert Tibshirani. (2014), An Introduction to Statistical Learning, Springer, p.426, [ISBN: 978-1461471370].

Ben Lambert. (2018), A Student's Guide to Bayesian Statistics, SAGE Publications Limited, p.520, [ISBN: 9781473916364].

Field A.. (2018), Discovering statistics using SPSS statistics, 5th edition. SAGE, London.

Supplementary Book Resources

Chris Brooks. (2019), Introductory Econometrics for Finance, Cambridge University Press, p.750, [ISBN: 978-1108436823].

Christian Heumann,Michael Schomaker,Shalabh. (2017), Introduction to Statistics and Data Analysis, Springer, p.456, [ISBN: 978-3-319-46162-5].

Julie Pallant. SPSS Survival Manual, [ISBN: 9780335261543].

Ruey S. Tsay. (2012), An Introduction to Analysis of Financial Data with R, John Wiley & Sons, p.416, [ISBN: 9780470890813].

This module does not have any article/paper resources

This module does not have any other resources

Discussion Note: