H8STATS2: Statistics II

Module Code:	H8STATS2
Long Title	Statistics II APPROVED
Title	Statistics II
Module Level:	LEVEL 8
EQF Level:	6
EHEA Level:	First Cycle
Credits:	5
Module Coordinator:	Sophie Flanagan
Module Author:	ORLA LAHART
Departments:	School of Computing
Specifications of the qualifications and experience required of staff	Master's and/or PhD degree in a numerate / scientific discipline, with experience in practical applications of statistical techniques. May have industry experience also

Learning Outcomes

On successful completion of this module the learner will be able to:

#	Learning Outcome Description
LO1	Analyse and select the appropriate statistical methodology to solve data analysis problems, or make predictions
LO2	Understand the concepts of normality, independence, and homoscedasticity for the selection of statistical tests and forecast technique
LO3	Critically evaluate the outcome of statistical significant tests using advanced concepts, such as statistical power, sample size, and multiple comparisons
LO4	Conduct advanced statistical analyses to answer real life questions and demonstrate ability to solve problems.
LO5	Interpret and clearly communicate the results of statistical tests to take informed decisions using data in the appropriate contexts.

Dependencies

Module Recommendations

68040	H8STATS1	Statistics I

Co-requisite Modules

No Co-requisite modules listed

Entry requirements	N/A

H8STATS2: Statistics II

Module Content & Assessment

Indicative Content
Inferential statistics revisited Hypothesis testing. One-way ANOVA, and t-test; discussion of tools and programming languages for statistics, such as Python, R, and SPSS
Exploratory data analysis Correlations, Chi-Square test of independence, box plots, confidence intervals, simple linear regression
Regressions Multiple linear and polynomial regressions, robust and quantile regressions, stepwise regression, and model selection
Normality Tests and plots for normality, including QQ-plots, Shapiro-Wilk, and Kolmogorov-Smirnov tests, and Box Cox transformation. Reporting results
Statistical power & sample size Cohen's d and Hedges's g, and other effect size suggestions. Power calculations
Two-way ANOVA Review of main assumptions. Data preparation. Conduct, interpret, and report ANOVA results
Post-hoc tests Multiple comparisons and p-value inflation. Tukey's HSD and Bonferroni correction. Dunn's and Dunnett's tests. False discovery rate. Reporting results
Non-parametric tests on contingency tables One and two-factor analysis using Chi-squared test for count data. Reporting results
Non-parametric tests on populations Mann-Whitney, Wilcoxon, and Kruskal-Wallis tests. Reporting results
Factor Analysis and PCA Collinearity. Kaiser-Meyer-Olkin test. Screeplot. Loadings and interpretation. Reporting results
Time Series Analysis Smoothing data and seasonality. Forecasting with Holt-Winters and ARIMA. Relationship between time series forecasting and supervised learning
Survival Analysis Censored data, Kaplan-Meier estimator, and Cox model

Assessment Breakdown	%
Coursework	50.00%
End of Module Assessment	50.00%

Assessments

Full Time

Coursework

Assessment Type:	Assignment 1	**% of total:**	50
Assessment Date:	n/a	**Outcome addressed:**	1,2,3,4,5
Non-Marked:	No		

Assessment Description:
In this assignment the student will prepare data for two-way ANOVA, and any two other non-parametric tests from Week-8 and Week-9. The student may use Python, R, or SPSS, but should not rely on only one tool, variety is expected. It is not necessary to replicate any test you carry out, ie if you perform a test in R it is not necessary to repeat it in SPSS and/or Python. A data file from the Census of Ireland is suggested, though students are permitted to choose a different file if they wish (subject to approval by Lecturer). The task is to prepare a statistical report based on the data in the file

End of Module Assessment

Assessment Type:	Terminal Exam	**% of total:**	50
Assessment Date:	End-of-Semester	**Outcome addressed:**	1,2,3,5
Non-Marked:	No		

Assessment Description:
The end of semester examination paper which is two hours in duration. Usually learners are requested to answer four out of five questions. Question format will usually be of essay-style but may also include other formats (e.g., a plan for an extended business data analysis project or a technical figure). Marks will be awarded based on clarity, structure relevant examples, depth of topic knowledge and an understanding of the potential and limits of solutions

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.

Reassessment Description
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.

H8STATS2: Statistics II

Module Workload

Module Target Workload Hours 0 Hours

Workload: Full Time

Workload Type	*Workload Description*	*Hours*	*Frequency*	*Average Weekly Learner Workload*
Lecture	Weekly lecture	24	Per Semester	2.00
Tutorial	Weekly tutorial	12	Per Semester	1.00
Independent Learning Time	Research & autonomous student learning	89	Per Semester	7.42
			Total Weekly Contact Hours	3.00

Workload: Part Time

Workload Type	*Workload Description*	*Hours*	*Frequency*	*Average Weekly Learner Workload*
Lecture	Weekly lecture	24	Per Semester	2.00
Tutorial	Weekly tutorial	12	Per Semester	1.00
Independent Learning Time	Research & autonomous student learning	89	Per Semester	7.42
			Total Weekly Contact Hours	3.00

Module Resources

Recommended Book Resources

David Spiegelhalter. (2019), The Art of Statistics, Pelican, p.256, [ISBN: 0241258766].

Bernard Rosner. (2015), Fundamentals of Biostatistics, 8th. Cengage Learning, USA, [ISBN: 978-1-305-26892-0].

Peter Dalgaard. Introductory Statistics with R, Springer, p.364, [ISBN: 0387790535].

Supplementary Book Resources

Andy Field. (2018), Discovering Statistics using IBM SPSS Statistics, 5th. SAGE Publications, Incorporated, p.775, [ISBN: 9781544328225].

EMC Education Services. (2015), Data Science and Big Data Analytics, John Wiley & Sons, Incorporated, [ISBN: 111887613X].

Eugene Demidenko. (2019), Advanced Statistics with Applications in R, 10, John Wiley & Sons, p.880, [ISBN: 978-1-118-38798-6].

Recommended Article/Paper Resources

Valentin Amrhein, et al. (2019), Scientists rise up against statistical significance, Nature, 567, p.305-7, http://dx.doi.org/10.1038/d41586-019-008 57-9

Naomi Altman, Martin Krzywinski. (2017), P values and the search for significance, Nature Methods, 14, p.3–4, https://www.nature.com/articles/nmeth.41 20

Other Resources

[Website], Choosing the correct stat test. https://stats.idre.ucla.edu/other/mult-p kg/whatstat/

[Website], The Khan Academy. http://www.khanacademy.org/.

[Website], Learn with Dr Eugene O'Loughlin. http://www.youtube.com/eoloughlin.

[Website], Central Statistics office. http://www.cso.ie.

[Website], Glossary of Statistical Terms. http://bit.ly/LIRYpQ.

[Website], HyperStat Online Statistics Textbook. http://davidmlane.com/hyperstat/.

[Website], The R Project for Statistical Computing. http://www.r-project.org/.

Discussion Note:	