

University of Galway

Course Module Information

Course Modules

ST2001: Statistics for Data Science 1

Semester 1 | Credits: 5

The course provides an introduction to probabilistic and statistical methods needed to make reasonable and useful conclusions from data. Topics include probabilistic reasoning, data generation mechanisms, modern techniques for data visualisation, inferential reasoning and prediction using real data and the principles of reproducible research. The course will rely heavily on R (a free open source language) and will include examples of datasets collected in a variety of domains.

(Language of instruction: English)

Learning Outcomes

1. Calculate conditional probabilities and probabilities for random variables from standard distributions (Binomial, Poisson, Normal).
2. Summarise data numerically (centre and spread) and graphically (e.g. bar charts, line, area, boxplots, histograms, density plots, scatterplots) with an emphasis on best practice for communication.
3. Summarise the importance of probabilistic based sampling schemes (e.g. simple random sampling, stratified sampling, cluster sampling).
4. Summarise the difference between observational and experimental studies and the principles of experimental design.
5. Perform probability calculations about the sample mean and use them to make inferential statements using the Central Limit Theorem.
6. Calculate interval estimates for parameter estimation in one sample problems using classical and computational (i.e. bootstrap) approaches.
7. Perform hypothesis testing (null and alternative hypotheses, type I and II errors and p-values) in a variety of scenarios.
8. Fit and interpret a simple linear regression model.
9. Compile a statistical report, i.e. prepare a typed document which introduces the statistical research question being explored, describes the data collection mechanism, provides subjective impressions on relevant numerical and graphical summaries, and outlines conclusions from all formal statistical analyses undertaken.

Assessments

- Continuous Assessment (30%)
- Written Exam (70%)

Teachers

- NICOLA FITZ-SIMON: [Research Profile](#) | [Email](#)

Reading List

1. "Open Intro Stats" by David M Diez, Christopher D Barr, Mine Cetinkaya-Rundel
Publisher: Open Intro
2. "R for Data Science" by Garrett Grolemund, Hadley Wickham
Publisher: O'Reilly
3. "Hitchhikers Guide to GGplot2" by Mauricio Vargas Sepúlveda and Jodie Burchell
Publisher: Leanpub
4. "An Introduction to Statistical and Data Sciences via R" by Chester Ismay and Albert Y. Kim