SS2: Foundations in Statistics

SS2.3: The design and implementation of experiments and the management of data

Professor Caroline Brophy

School of Computer Science and Statistics
Trinity College Dublin

LegumeLegacy Event 2, Dublin, Ireland

November 2023

LegumeLegacy funding

Summary of SS2: Foundations in Statistics II

- ▶ SS2.1: Regression Analysis

  - ▶ Simple linear regression and One-way ANOVA models

  - ▶ Two-way ANOVA models

  - ▶ Multiple regression models

- ▶ SS2.2: Challenges with data and regression analysis

- ▶ SS2.3: The design and implementation of experiments and the management of data (this session)

SS2.2: The design and implementation of experiments and the management of data

SS2.2.1: Statistical considerations when implementing experiments

Choice of experiment design

Fisher: *"To consult the statistician after an experiment is finished is often merely to ask him to conduct a post mortem examination. He can perhaps say what the experiment died of."*

Types of studies

- ▶ Observational studies: explanatory variables are not controlled

- ▶ Experiments: explanatory variable values on experimental units are controlled / manipulated.

  - ▶ Comparison of treatments

  - ▶ Replication

  - ▶ Randomisation

Replication and randomisation are particularly crucial for any experiment.

SS2.2.2: Experimental designs

Various designs

- Completely randomised design

- Randomised block design

- Factorial design

- Split plot design.

Plan your design

- ▶ Hypotheses and questions to be addressed.

- ▶ Sample size calculations.

- ▶ Logistical constraints.

- ▶ Ethical issues.

- ▶ Process for randomisation.

- ▶ Measurements to be recorded. Sufficient for the hypotheses / questions?

Having a good design is important because it can waste time and money if an investigation is badly designed and uses too many or too few resources.

SS2.2.3: Data management

Plan, plan, plan ahead

- How you will record your data in the field or lab?

- How will you record your data after the field or lab?

- Make sure to plan and prepare templates for data recording **before** you arrive in the field or lab.

- How will you check your data and ensure its quality?

- How will you manage version control?

SS2.2.4: Recommended reading

Some suggested reading material

Books

OpenIntro Statistics, by Diez, Çetinkaya-Rundel and Barr.

Applied Linear Regression, by Sandford and Weisberg.

Regression Analysis by Example, by Chatterjee and Hadi.

The Statistical Sleuth, by Ramsey and Schafer.

Statistics for experimenters, by Box, Hunter and Hunter

Papers

Hector, von Felten and Schmid (2010) Analysis of variance with unbalanced data: an update for ecology & evolution. Journal of Animal Ecology, 79, 308–316.

Zuur, Ieno and Elphick (2009) A protocol for data exploration to avoid common statistical problems. Methods in Ecology & Evolution, 1, 3–14.

Box and Cox (1964) An Analysis of Transformations. Journal of the Royal Statistical Society. Series B (Methodological), 26, 211–52.