**Simulated data and Monte-Carlo simulation studies**

**Ian Hussey**

2 day workshop

Simulated data is incredibly useful. Before a study design has been finalised, you can use stimulated data to refine the design and analysis to ensure they address your research question. Before the real data has been collected, you can use simulated data to write and debug code for data processing, analysis, and visualisation. Without overfitting on your real data, you can use simulated data to learn how to use a statistical method. And using Monte Carlo simulation studies, you can better understand statistical methods, from simply how to implement and debug them to understanding their statistical power, false discovery rate, or what happens when you violate their assumptions.

This two-day workshop introduces participants to data simulation using R and the tidyverse. On the first day, participants will learn how to simulate datasets ranging from extremely simple to complex (e.g., non-normal distributions, truncated, multi-level). On the second day, participants will be introduced to simulation studies, where data is simulated and then analysed many times under different known conditions in order to estimate properties of interest (e.g., statistical power, false positive rates).

Key Takeaways:

* Simulating Unrealistic Data: In its simplest form, learning how to simulate even totally unrealistic data can be a useful starting point to help you write and debug code.
* Data Simulation for Workflow Development: Learn how to simulate realistic data to write and test your data wrangling, visualization, and analysis code, even before real data is available.
* Data Simulation for Methodological Exploration: Use realistic simulated data to practice new analysis techniques or methods in a controlled environment, helping you build confidence before applying them to real data.
* Monte Carlo Simulation Studies: Conduct simulation studies in R and purrr to improve your statistical inferences, e.g. by estimating statistical power, false positive rates, or robustness.

This workshop is particularly relevant to early-career graduate students who are working with data who want to improve their data analysis skills and confidence, but who do not yet have experience with the above tools and methods. If you have taken longer courses on these topics, this course may be too introductory for you. This course assumes some degree of experience with R and data wrangling (dplyr, tidyr) –the previous one-day course on data wrangling plus some practice will be sufficient. While this one-day workshop provides an introduction to these tools, it is important to remember that mastering them will require extensive subsequent self-study and practice.