# Class 1: Basics and experimental design

Andrew Parnell, School of Mathematics and Statistics,
University College Dublin

# Learning outcomes

- Know how to create and interpret a two-sample t-test
- Understand what a p-value means
- Be able to perform a simple sample size calculation
- Understand the basics of experimental design

General goal for the course: be able to create a statistical model for a medical test and check that it is robust

# Course details

- Lectures in the morning (9:30 - 1pm), practical in the afternoon (2pm - 4:30pm). More details in the timetable.
- All course notes, code and data sets available on Github page
- All Slides available in pdf or RMarkdown (Rmd) fomat which can be opened in Rstudio

# Some basic concepts:

- One way data can be separated is via *continuous* (e.g. age, weight), or *discrete* (disease state, Gleason grade, etc)
- You can divide continuous into *interval* (temperature) or *ratio* (age, weight)
- You can divide discrete into *ordinal* (e.g. Gleason grade) or *nominal* (disease state, eye colour)

The type of statistical model we fit is almost entirely dependent on the type of data we have

# Two examples: prostate cancer (regression)

Prostate cancer data set:

- `lcavol`: log(cancer volume)
- `lweight`: log(weight)
- `age`: age
- `lbph`: log(benign prostatic hyperplasia amount)
- `svi`: seminal vesicle invasion
- `lcp`: log(capsular penetration)
- `gleason`: grade of cancer
- `pgg45`: percentage Gleason scores 4 or 5
- `lpsa`: outcome variable - log prostate specific antigen
- `train`: whether the observation should be included in the training or test set

**Task: predict `lpsa` based on other variables for the training set, and check performance on the test set**

# South African Heart Rate data (classification)

# Testing differences between groups; the two-sample t-test

# Sampling distributions of data

# Null and alternative hypotheses

# Drawing pictures

# Getting and understanding the p-value

# What the p-value is not

# Introduction to sample size calculations

# Type 1 and Type 2 error

# Drawing pictures

# The magic formula

# Getting the values to put in to the formula

# Possible extensions

# Design of Experiments

# The golden rule of designing an experiment

# Blocking

# Randomisation

# Replication

# More complicated experiments