

Class 1: An introduction to Bayesian Hierarchical Modelling

Andrew Parnell
andrew.parnell@mu.ie



https://andrewcparnell.github.io/bhm_course

Let's get started

- ▶ Tell me who you are, what you are working on, and what you hope to get out of the week
- ▶ Timetable for the week
- ▶ Pre-requisites

How this course works

- ▶ This course lives on GitHub, which means anyone can see the slides, code, etc, and make comments on it
- ▶ The timetable html document provides links to all the pdf slides and practicals
- ▶ The slides and the practicals are all written in `Rmarkdown` format, which means you can load them up in Rstudio and see how everything was created
- ▶ Let me know if you spot mistakes, as these can be easily updated on the GitHub page
- ▶ There is a `bhm_course.Rproj` R project file from which you should be able to run all the code

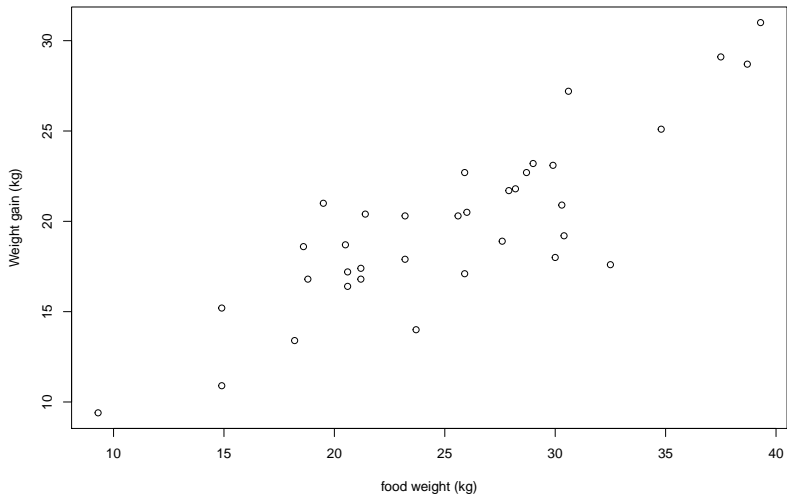
Course format and other details

- ▶ Lectures will take place in the morning, practical classes in the afternoon
- ▶ We will finish earlier on Wednesday/Thursday for a mini-trip
- ▶ Please ask lots of questions
- ▶ Some good books:
 - ▶ *Data Analysis using Regression and Hierarchical Models* by Gelman and Hill
 - ▶ *Bayesian Data Analysis* by Gelman et al

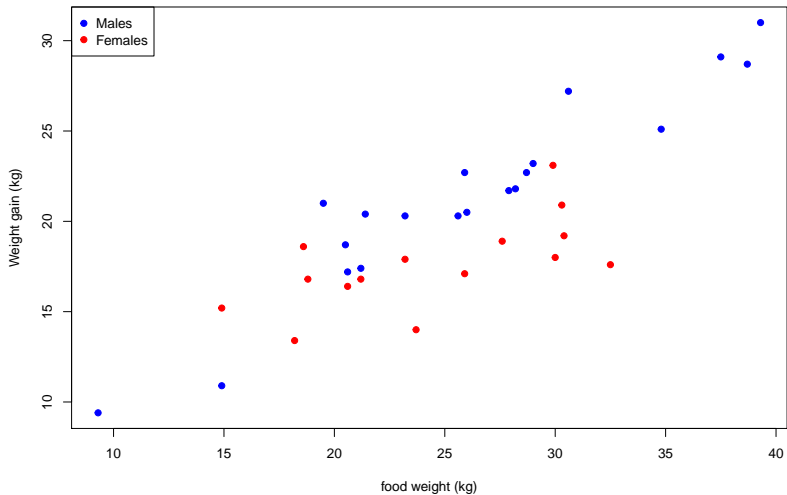
What is a Bayesian hierarchical model?

- ▶ **A model** is just a representation/approximation of the real world, here expressed in equations
- ▶ **Hierarchical** means that the model is built up in *ordered layers* which makes it easier to fit very complex models
- ▶ **Bayesian** means the model involves both a *likelihood* and a *prior* probability distribution (more on this tomorrow)

Thinking hierarchically: example 1



More information:



Example 2: 8 Schools

We have 8 schools in a region, with a relative performance score (column score) compared to the national average and a standard deviation (σ) based on 3 repeated visits

##	school	score	sigma
## 1	1	28	15
## 2	2	8	10
## 3	3	-3	16
## 4	4	7	11
## 5	5	-1	9
## 6	6	1	11
## 7	7	18	10
## 8	8	12	18

- ▶ If you had to pick an overall score for this region how would you calculate it?
- ▶ If you had to guess the score of a new measurement for school 1 what value would you use?

Example 3: Earnings data

1192 observations on earnings (in USD) and various measurements about ethnicity, age, height, etc

##	earn	age	eth	height	height_cm	y	x	x_centered
## 1	50000	2	3	74	187.96	10.81978	74	6.932011
## 2	60000	3	3	66	167.64	11.00210	66	-1.067989
## 3	30000	1	3	64	162.56	10.30895	64	-3.067989
## 4	51000	2	3	63	160.02	10.83958	63	-4.067989
## 5	9000	1	3	64	162.56	9.10498	64	-3.067989
## 6	29000	2	3	62	157.48	10.27505	62	-5.067989

- ▶ Does height affect earnings?
- ▶ Are there different rates of change for different groups (e.g. age/ethnic groups)?

Example 4: Swiss Willow Tit data

3 replicate measurements on whether Swiss Willow Tits were found with covariates on forest cover and elevation

##	rep.1	rep.2	rep.3	c.2	c.3	elev	forest	dur.1	day.2	day.3	length	alt
## 1	0	0	0	0	0	420	3	240	58	73	6.2	Low
## 2	0	0	0	0	0	450	21	160	39	62	5.1	Low
## 3	0	0	0	0	0	1050	32	120	47	74	4.3	Med
## 4	0	0	0	0	0	1110	35	180	44	71	5.4	Med
## 5	0	0	0	0	0	510	2	210	56	73	3.6	Low
## 6	0	0	0	0	0	630	60	150	56	73	6.1	Low

- ▶ How do the covariates affect the chance of finding the birds?
- ▶ Are these effects linear?
- ▶ What do we do with the missing data?

More data sets in the data directory

- ▶ The data directory contains a few more data sets which we will play with throughout the week
- ▶ The `data_descriptions.txt` file shows what they contain
- ▶ If you have some spare time it's worth loading them in, exploring relationships, and fitting some simple models

Summary

- ▶ In hierarchical models we avoid fitting models separately as much as possible
- ▶ By fitting models together we **borrow strength** from the different groups in the data and reduce uncertainty
- ▶ Bayesian models allow us to incorporate all the available data into providing information on the question we want to answer