Revision – Module 1

Biology **E**xperimental **D**esign and **A**nalysis (**BEDA**)

Januar Harianto

The University of Sydney

Semester 2, 2025

THE UNIVERSITY OF SYDNEY

You’ve made it to the end of Module 1!

Important

Please complete the Assessment Quiz for Module 1 by **THIS FRIDAY** – worth **10%** of your final grade.

What have we learnt?

1. Study design – IMPORTANT.
2. (Almost) everything is a **general linear model**.
3. Block what you can, randomise what you cannot – **controls**.

Experimental design

There is no *magical* statistical method that will make up for a poorly designed study.

🗑️ Garbage in ▶ Garbage out 💩

Abstraction comes before data collection

The **abstraction process** defines the problem, question, and hypothesis before collecting data, and includes:

- Formulating the model i.e. the relationship between the **response** and **predictor/explanatory** variables.
- Thinking about sources of variation and ensuring that the study design can account for them, i.e. **blocking** and **randomisation**.
- Deciding on the appropriate statistical model to analyse the data (even without the data) by considering the structure of each variable (**categorical** or **continuous**).

Modelling

What is a model?

A way to describe the **relationship** between the response variable and the explanatory variable(s).

$$y \sim x$$

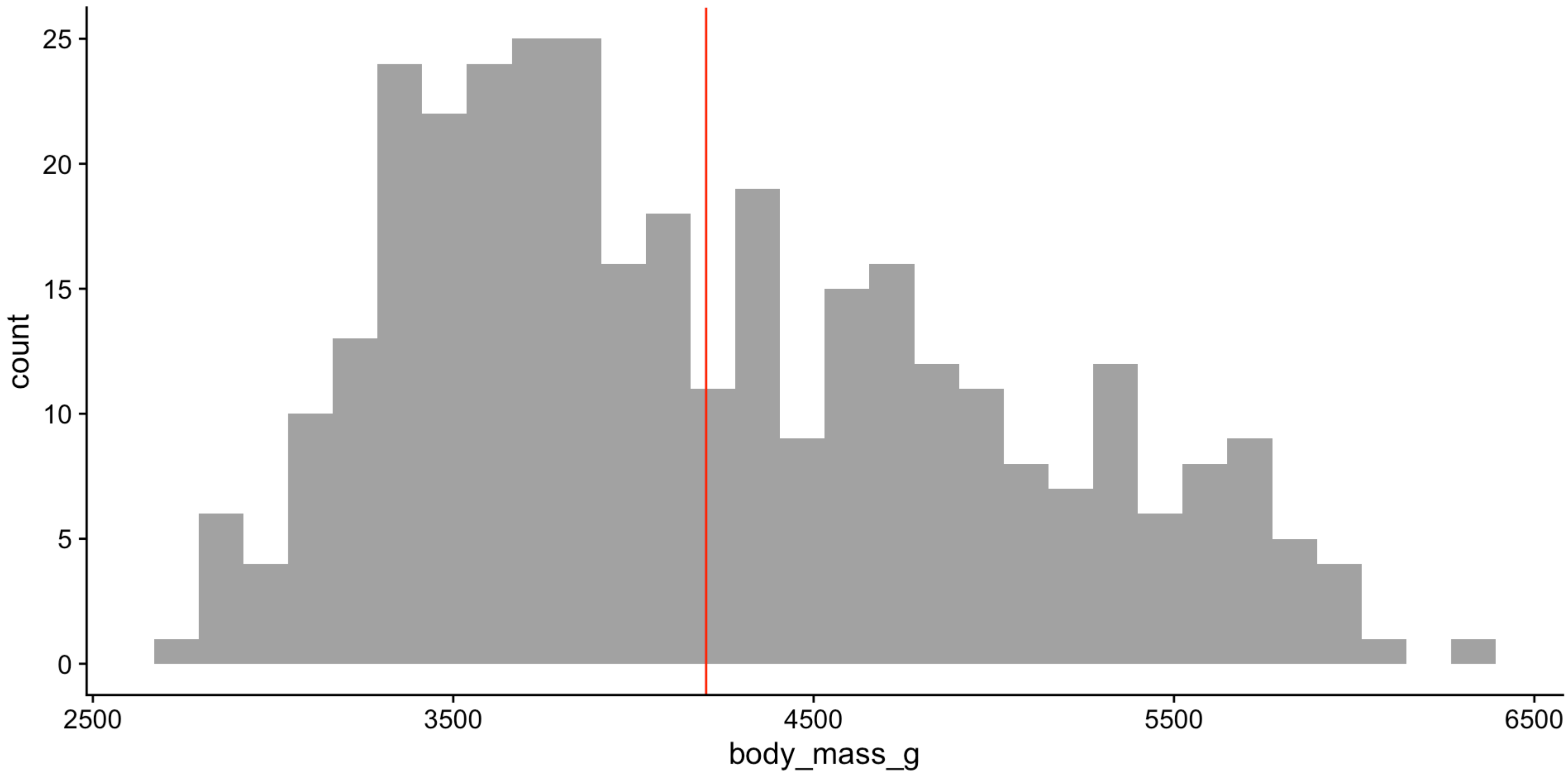
- y is influenced by x.
- A change in y is dependent on x.
- y is the response to the explanatory variable x.
- **y is a function of x.**

> **Tip**
>
> If you can plot it, you've got a model.

Modelling in a nutshell

$response \sim \mu$

> y can be explained by a constant (e.g. mean or μ)

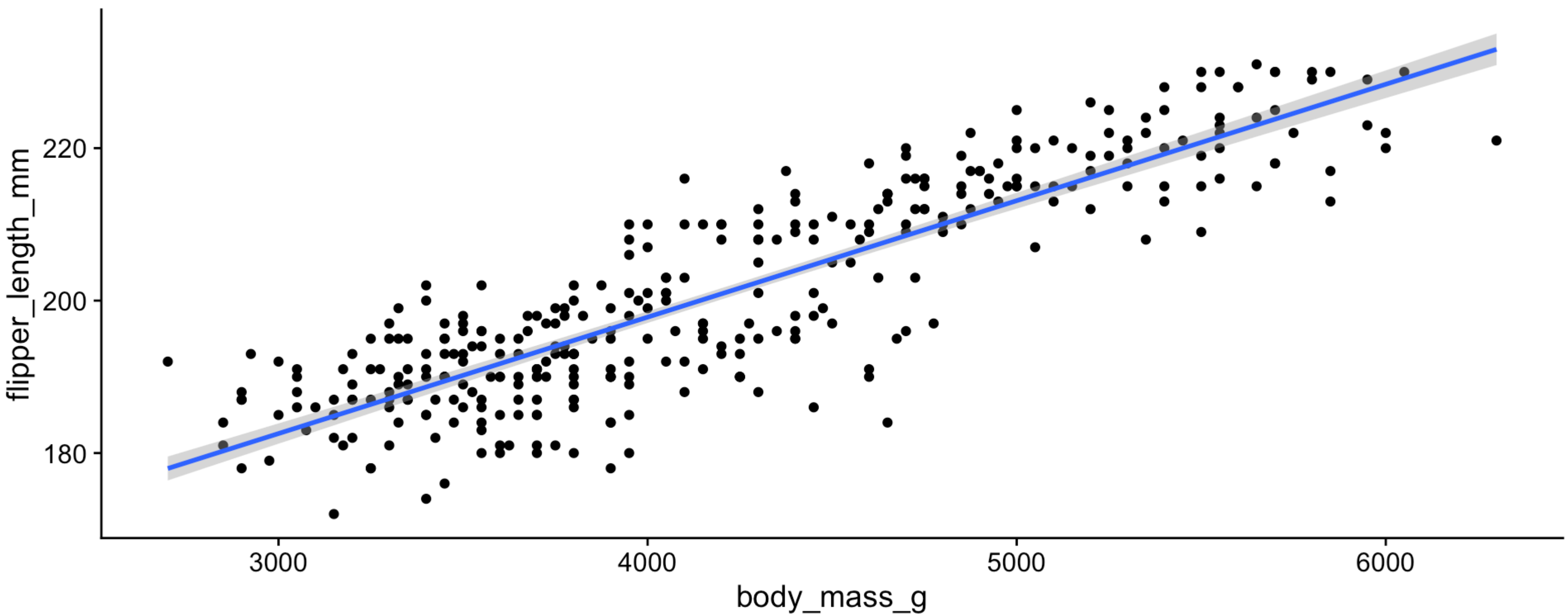


Modelling in a nutshell

$response \sim x$

> The *response* can be explained by *a predictor*.

When the predictor is CONTINUOUS



Modelling in a nutshell

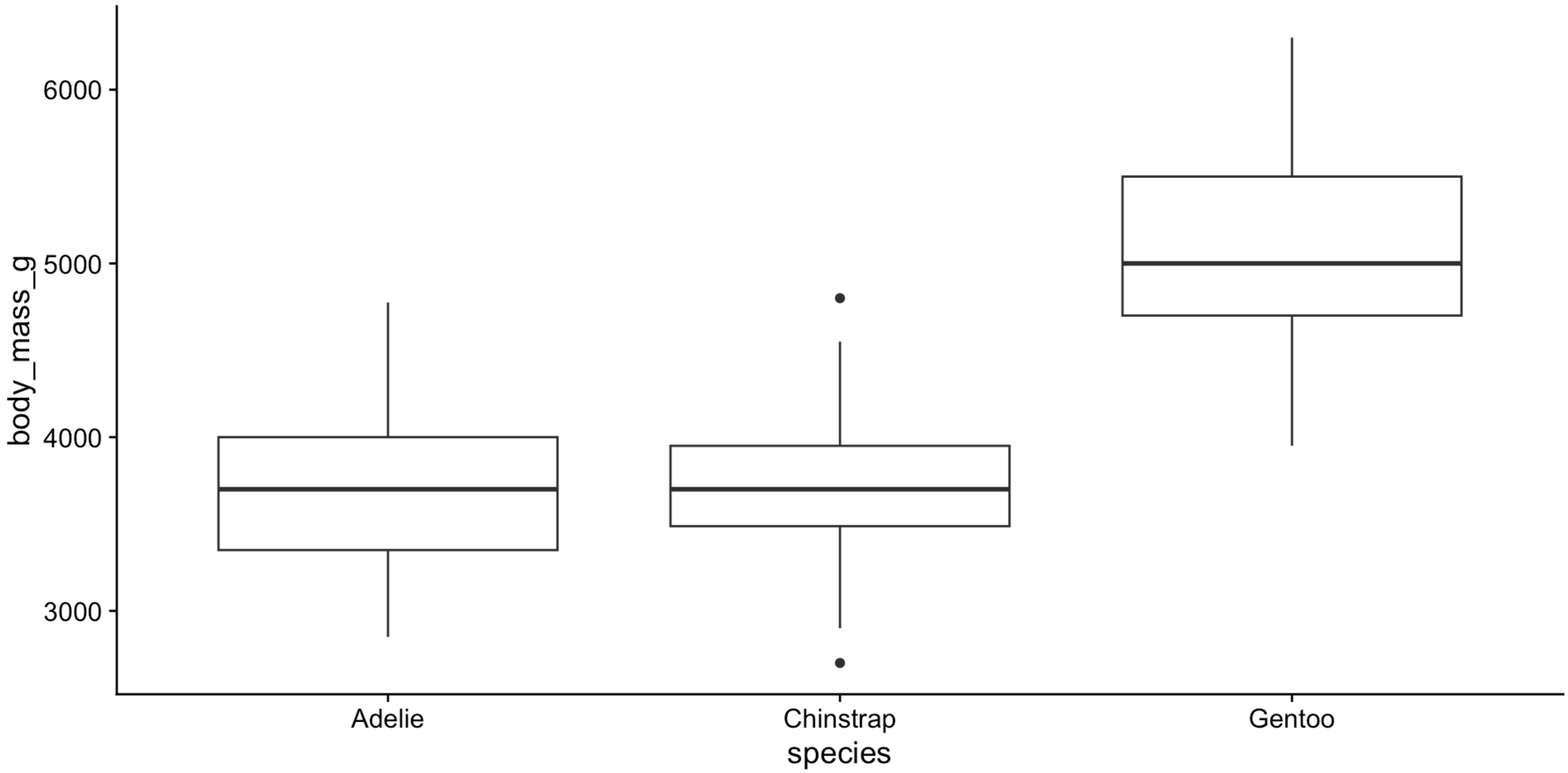
$$response \sim x$$

> The *response* can be explained by *a predictor*.

When the predictor is CATEGORICAL

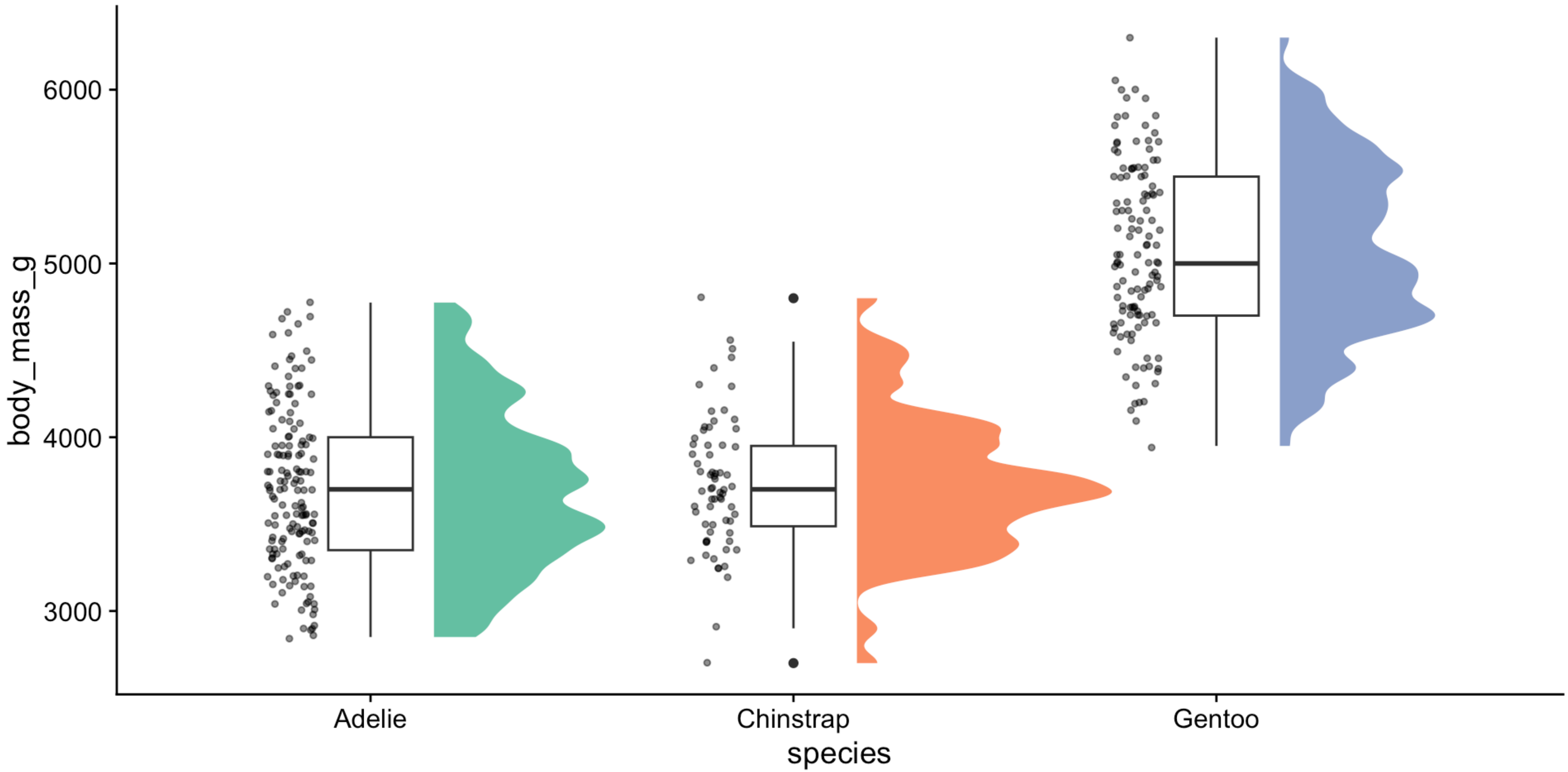
Modelling in a nutshell

When the predictor is CATEGORICAL



Modelling in a nutshell

When the predictor is CATEGORICAL

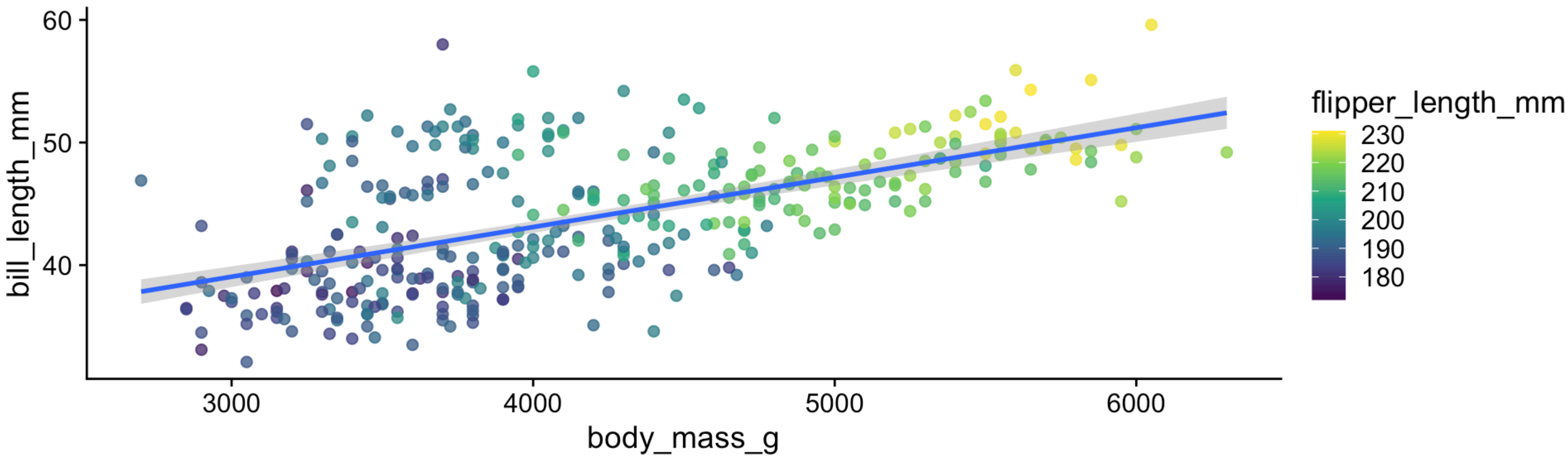


Modelling in a nutshell

$$response \sim x_1 + x_2$$

> The *response* can be explained by a linear combination of two predictors, x_1 and x_2.

When both predictors are CONTINUOUS

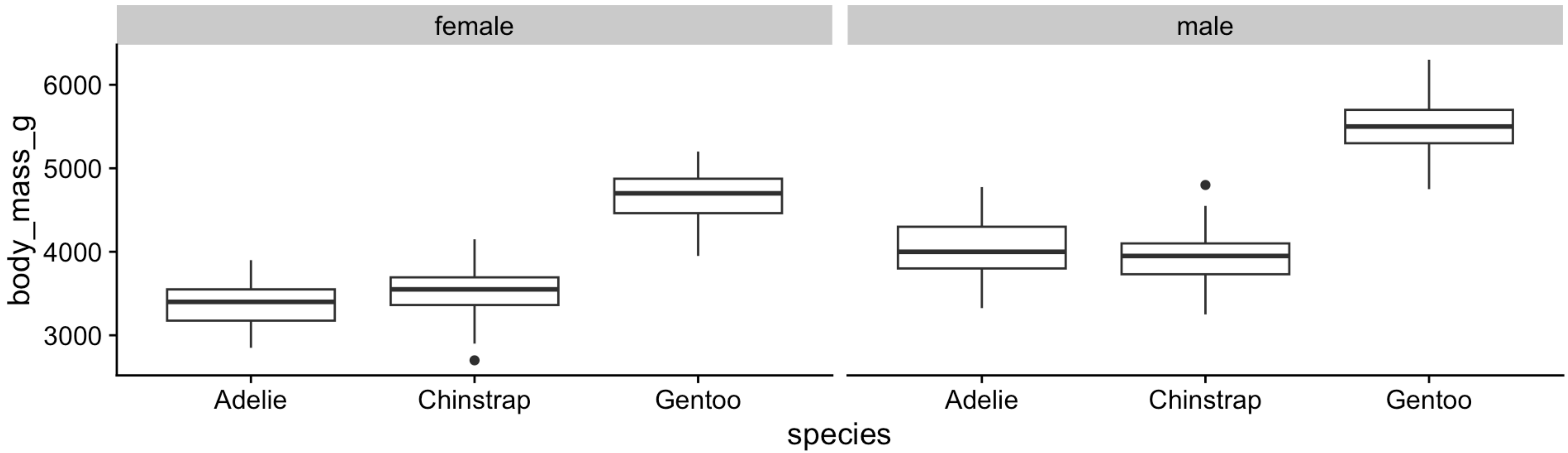


Modelling in a nutshell

$response \sim x_1 + x_2$

> The *response* can be explained by a linear combination of two predictors, x_1 and x_2.

When both predictors are CATEGORICAL

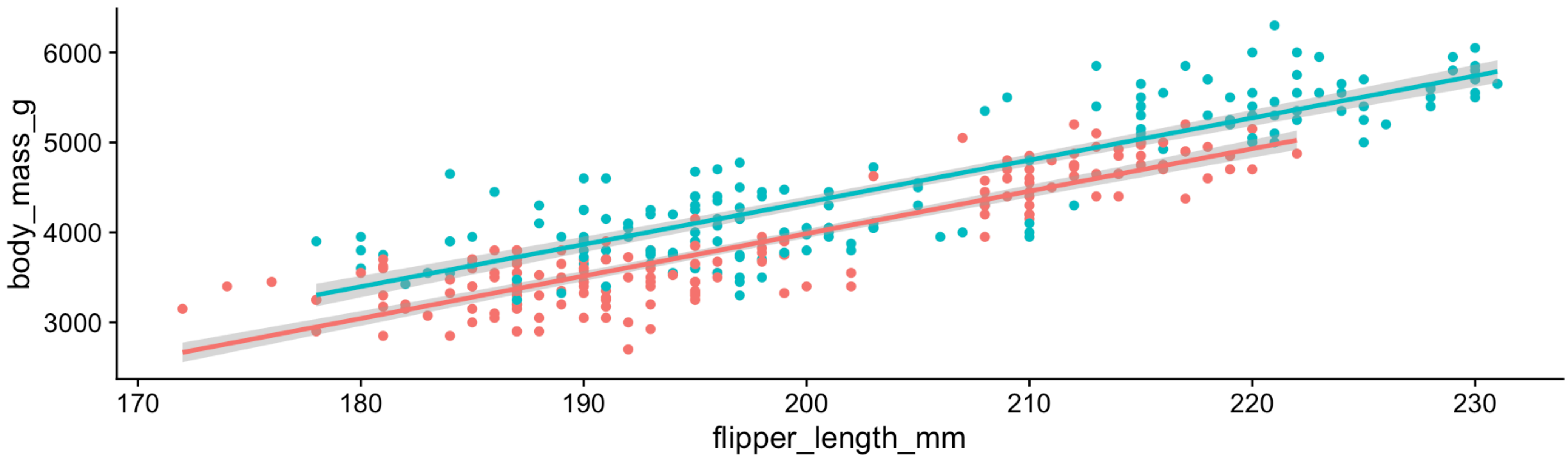


Modelling in a nutshell

$$response \sim x_1 + x_2$$

> The *response* can be explained by a linear combination of two predictors, x_1 and x_2.

When both predictors are CONTINUOUS and CATEGORICAL



Modelling in a nutshell

$$response \sim x_1 + x_2 + \cdots + x_n$$

> y can be explained by a linear combination of n predictors.

In this case, plotting is not quite useful, and it might be better to explore the summary tables (ANOVA or regression) to understand the relationship between the response and predictors.

Assumptions of a model

LINE

A model should meet assumptions like how a job applicant should meet the job requirements. Are they fit for the task and outcome?

Always check the assumptions ***before*** interpreting the results.

- **L**inearity - use residual vs fitted plot
- **I**ndependence - not checked
- **N**ormality - use QQ-plot
- **E**qual variance - use residual vs fitted plot and/or

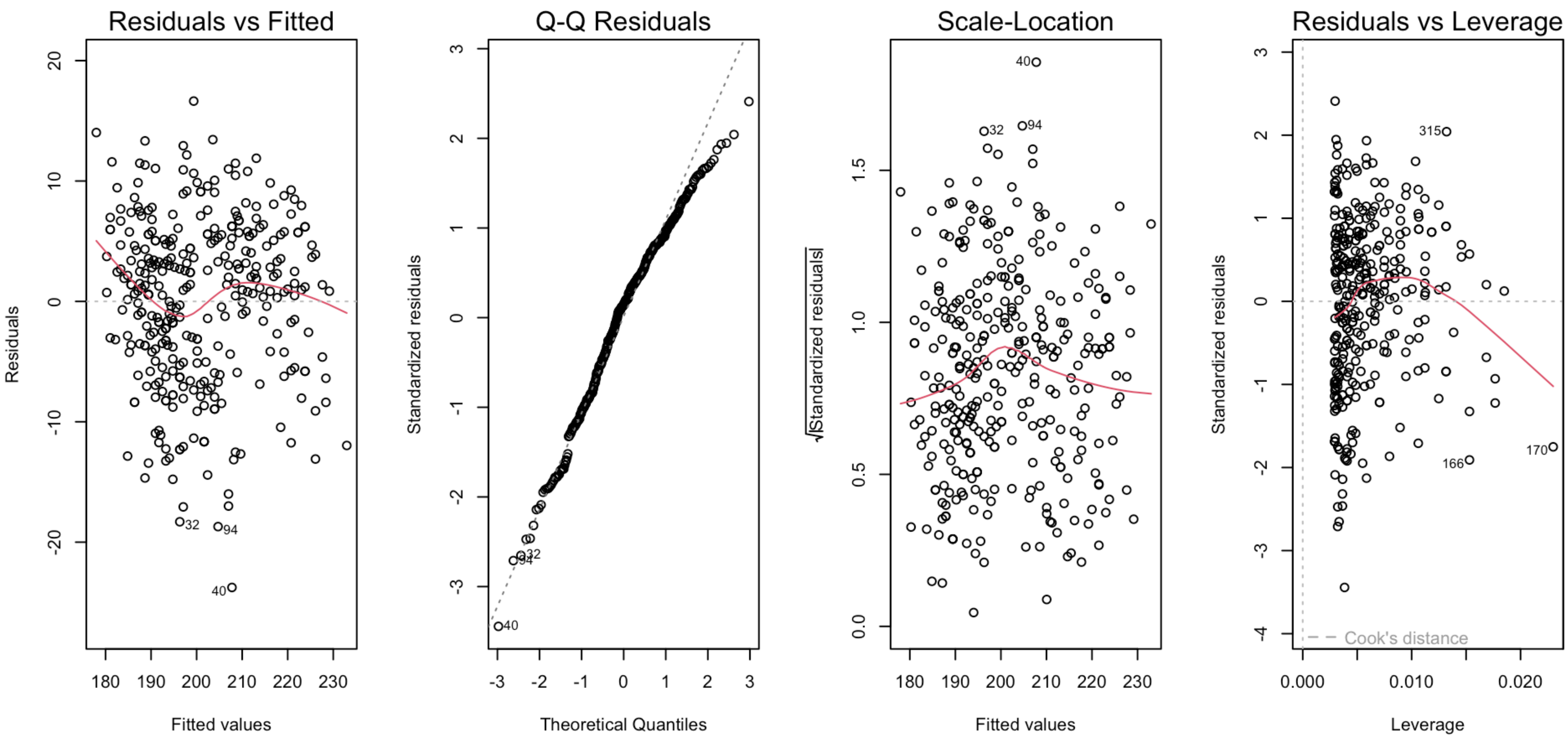
Notes on assumptions

Normality

- **Avoid using formal tests** (e.g. Shapiro-Wilk, Levene's) to check assumptions, unless sample size is so small that visual inspection is difficult.
- If sampling is random and representative, the **Central Limit Theorem** will ensure that the distribution of the sample means will be approximately normal.
- **Transformations** can be used to meet assumptions.

Equal variance

- **Homoscedasticity** – the variance of the residuals is constant across all levels of the predictor.
- Draw a perimeter around the points in a residual plot! if what you see is not a circle or rectangle, then you may have a problem.
- **Transformations** can be used to meet assumptions.



Useful resource: How to interpret a QQ-plot?

Exam

What to expect (Module 1)

- **Long answer questions** (but this applies to all modules).
- You will be asked to **interpret** the results of a statistical analysis.
- You will be asked to explain one or more study design concepts, e.g. pseudoreplication, with **example(s)**.
- You *may* be asked to *re-design* a study to address a specific issue.
- **No coding**
- **No calculations**

Explain....

Exam questions will focus on a case study to assess your understanding of key concepts. Some critical thinking will be required. If you can demonstrate a deep understanding of concepts like below, you will be well prepared:

- **What is confounding?** How does it influence the interpretation of study results? Explain with **examples**.
- Explain the purpose of the four **assumptions** of linear models (LINE).
- For each assumption, describe a scenario where **violating** it would lead to misleading conclusions about the relationship between a predictor and a response variable.

Interpret…

- A model is presented to you in empirical form. Can you describe its key features and the **relationships** it suggests between variables?
- An experiment was done, and its results analysed. You see ANOVA, regression summary and post-hoc tables. Are you able to pick out the **main findings** and their **implications**?
    - How would you interpret **coefficients** and **95% confidence intervals**?
    - What are **pairwise comparisons** and how are they interpreted in the context of **post-hoc** results?
    - When there are significant **interactions** in a model, how would you approach interpreting the **main effects**?
    - Draw a figure to illustrate the relationship between the predictor/explanatory and response variables (**boxplot, scatterplot**)

Re-design…

- We may ask you to **critique** a study design and suggest improvements, perhaps to address potential biases or limitations in the original design.
- You may also be asked to propose a completely **new** study design to investigate a specific research question.
- A flawed study design was identified and you are asked to re-design the study to **address** the identified **issues**.

Week 13: revision

- Lectures will be dedicated to revision
- I will run a mock exam during the practicals
- **We will NOT provide past exam papers**

Thanks!