

# Lecture 8:

## An introduction to logistic regression

PSYC234: Statistics: from association to modelling causality

Dr Amy Atkinson

Lecturer in Developmental Psychology

[amy.atkinson@lancaster.ac.uk](mailto:amy.atkinson@lancaster.ac.uk)

# The plan

---

**My aim:** to add a few final statistical tests to your toolbox for when the statistical test you've learned about might not be appropriate



# Learning objectives

---

- To develop an understanding of what binary logistic regression is and when it is appropriate to run
- To develop an understanding of the theory behind binary logistic regression
- To understanding how to conduct binary logistic regression in R when you have one categorical predictor
- To understand how to interpret binary logistic regression output

## Part 1

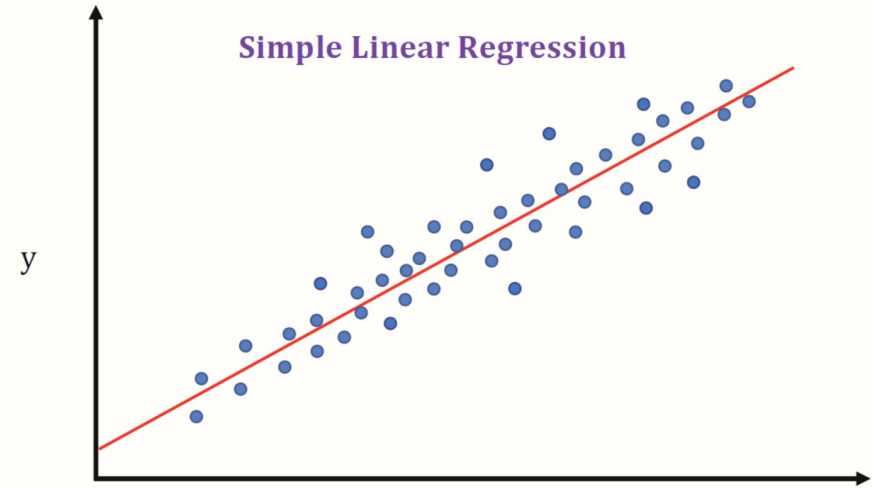
Why do I need to worry about logistic regression?

# A recap on linear regression

$$Y = b_0 + b_1 X_{1i} + \epsilon_i$$

Diagram illustrating the components of the linear regression equation:

- $Y$ : Outcome (indicated by an upward arrow from "Outcome")
- $b_0$ : Intercept (indicated by a downward arrow from "Intercept")
- $b_1$ : Slope (indicated by an upward arrow from "Slope")
- $X_{1i}$ : Value of predictor variable (indicated by a downward arrow from "Value of predictor variable")
- $\epsilon_i$ : Random error (indicated by an upward arrow from "Random error")



In linear regression, we assume a linear relationship between the predictor and the outcome

# Binary outcome variable

---

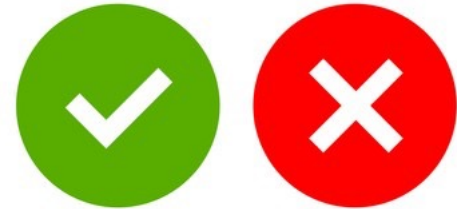
- Linear regression can work very well when you have a continuous outcome
- But what about when the outcome is binary (there are two possible outcomes)?



Happy/not happy



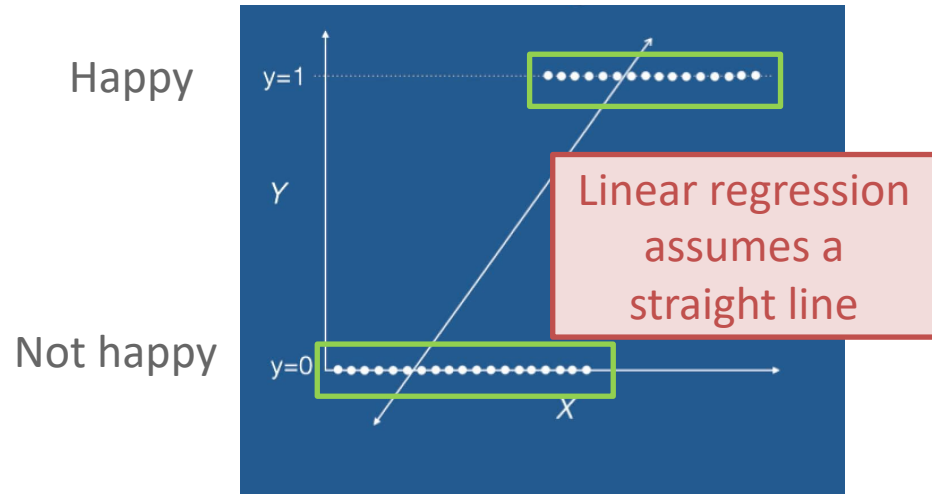
Pass/fail



Correct/Incorrect

# Two possible outcomes

- When we have two possible outcomes, participants can either be Outcome A or Outcome B
- People can either be happy ( $Y = 1$ ) or unhappy ( $Y = 0$ ) - there is no in between
- This violates the assumption of linearity



When we have two outcomes, we sometimes refer to them using the numbers 0 and 1. In our example,  $Y = 1$  could be “happy” and  $Y = 0$  is “not happy”

# How can we overcome this violation?

---


- When the outcome is binary (two possible outcomes), the assumption of linearity is **ALWAYS** violated
- We can apply a transform to the data to express the non-linear relationship in a linear way
- Binary logistic regression does this by expressing the linear regression equation in **logarithmic terms**
- This overcomes the issue of violating this assumption




# Linear and logistic regression formulae:

Binary logistic regression:

$$P(Y = 1) = \frac{1}{1 + e^{-(b_0 + b_1 X_{1i})}}$$


 Probability of a given outcome (e.g. "happy")


 Linear regression formula

Don't need to worry about this too much. Things to remember:

- We predict the probability of a given outcome occurring
- We express the linear relationship equation in logarithmic terms

# What is binary logistic regression?

---

- Logistic regression is a **generalized linear model** – flexible generalisation of linear regression
- Predicting an outcome that has only two possible outcomes  
→ Which of two outcomes is an individual likely to have (e.g. happy/not happy, pass/fail)?
- Predictors can be continuous, categorical, or a combination

# This week... let's keep it (relatively...) simple

---

- This week we will focus on binary logistic regression with **one categorical predictor**
- We aren't going to worry about different contrast coding methods – we'll just stick with R's default
  - Compares factor levels to a reference category
- We aren't going to worry about checking residuals (although this is an important step!)

## Example: Hamster and happiness

Does whether the participant have a hamster (yes/no) predict response to the following survey question:

Are you happy?

- Yes
- No



- Predictor: Hamster (Yes/No)
- Outcome: Happiness (Yes/No)