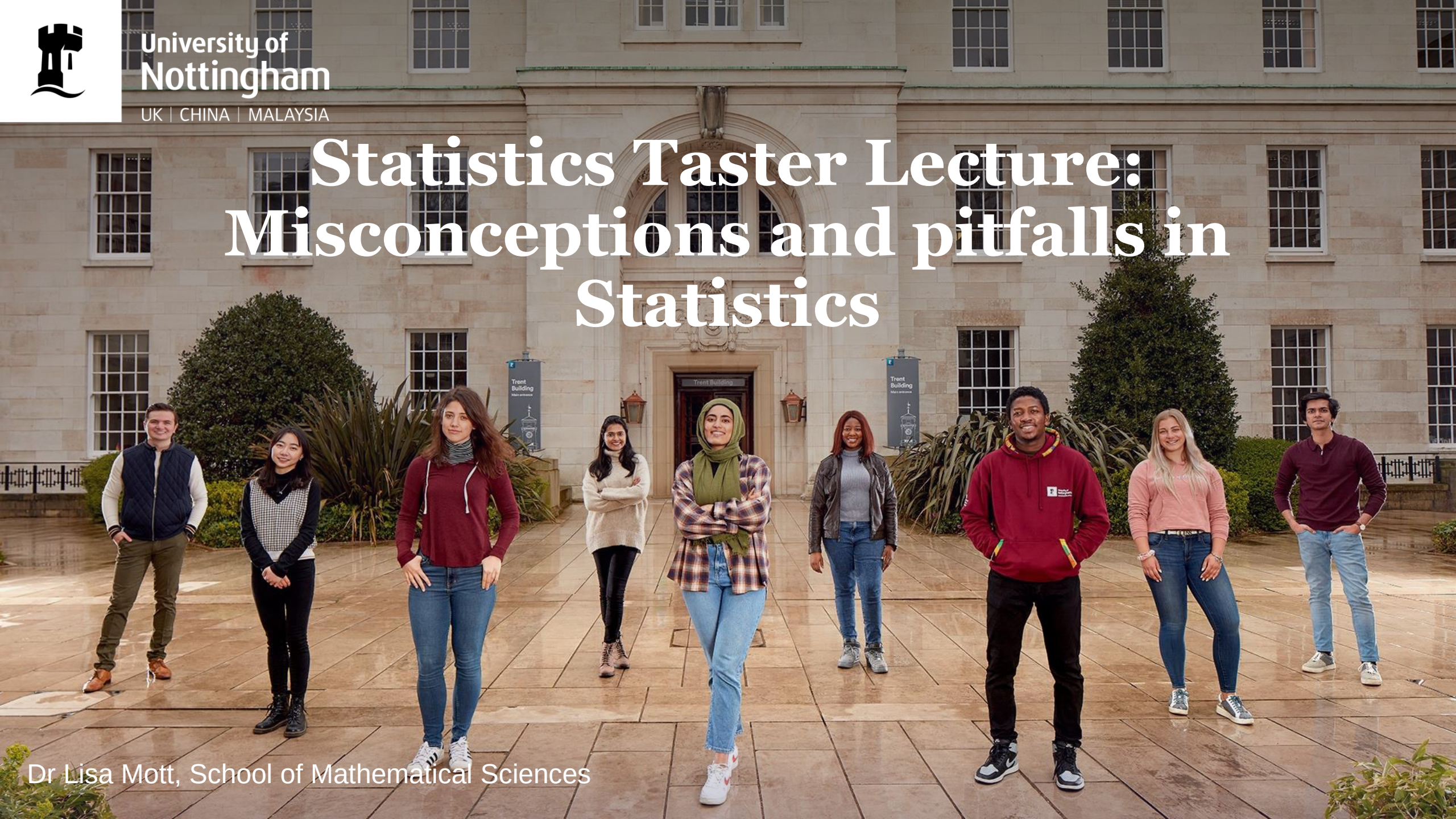
University of Nottingham
UK | CHINA | MALAYSIA

Statistics Taster Lecture: Misconceptions and pitfalls in Statistics



Dr Lisa Mott, School of Mathematical Sciences

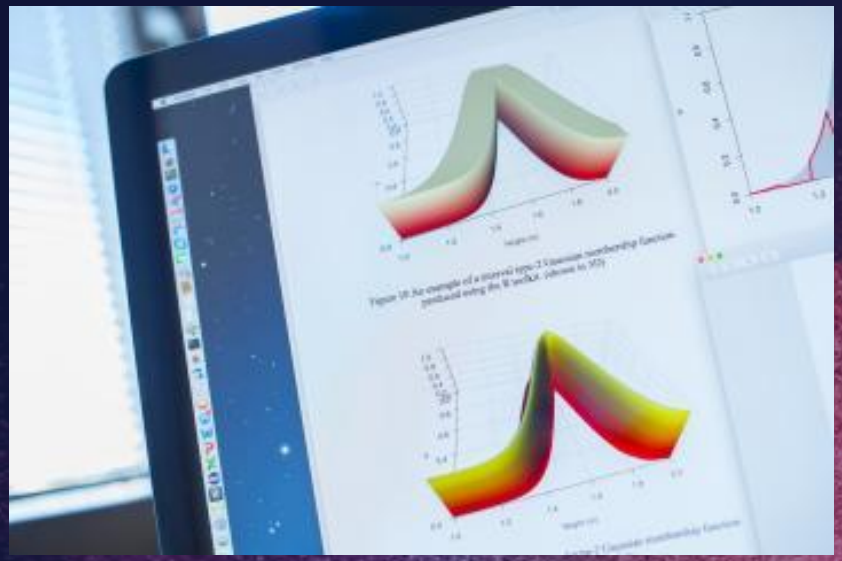
About me

- Dr Lisa Mott
- Teaching Associate
- I teach on the Degree Apprenticeship in Data Science
- First in my family to go to University.

Aims and Objectives

- Understanding the importance of visualising data.
- Considering extrapolation and its limitations.
- Introducing the normal distribution and the combination of two normal distributions.

University of Nottingham
UK | CHINA | MALAYSIA



An introduction

A motivating example

Anscombe data

- Francis Anscombe 1973

Data $x_1, x_2, x_3, \ldots, x_n$

Sample mean: $\frac{x_1 + x_2 + \ldots + x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n}$

	w_1	w_2	w_3	w_4	z_1	z_2	z_3	z_4
Mean	9	9	9	9	7.50	7.50	7.50	7.50
Variance	11	11	11	11	4.13	4.13	4.12	4.12

Sample variance: $\frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \ldots + (x_n - \bar{x})^2}{n-1}$

	w_1 and z_1	w_2 and z_2	w_3 and z_3	w_4 and z_4
Correlation	0.82	0.82	0.82	0.82

Pearson's Correlation Coefficient

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}}$$

Comment, are these data the same? What would you suggest we do to check?

Anscombe data

- Francis Anscombe 1973

This seems different.

x_1	x_2	x_3	x_4	y_1	y_2	y_3	y_4
10	10	10	8	8.04	9.14	7.46	6.58
8	8	8	8	6.95	8.14	6.77	5.76
13	13	13	8	7.58	8.74	12.74	7.71
9	9	9	8	8.81	8.77	7.11	8.84
11	11	11	8	8.33	9.26	7.81	8.47
14	14	14	8	9.96	8.10	8.84	7.04
6	6	6	8	7.24	6.13	6.08	5.25
4	4	4	19	4.26	3.10	5.39	12.50
12	12	12	8	10.84	9.13	8.15	5.56
7	7	7	8	4.82	7.26	6.42	7.91
5	5	5	8	5.68	4.74	5.73	6.89

Anscombe data

- Francis Anscombe 1973

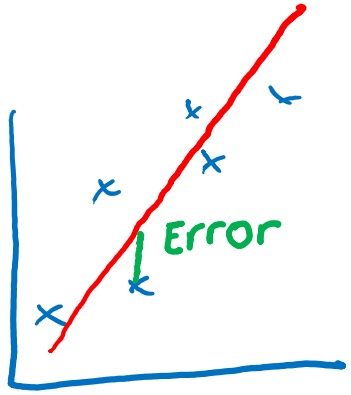
Linear regression line:

$$y_1 = 3.00 + 0.50x_1$$

$$y_2 = 3.00 + 0.50x_2$$

$$y_3 = 3.00 + 0.50x_3$$

$$y_4 = 3.00 + 0.50x_4$$

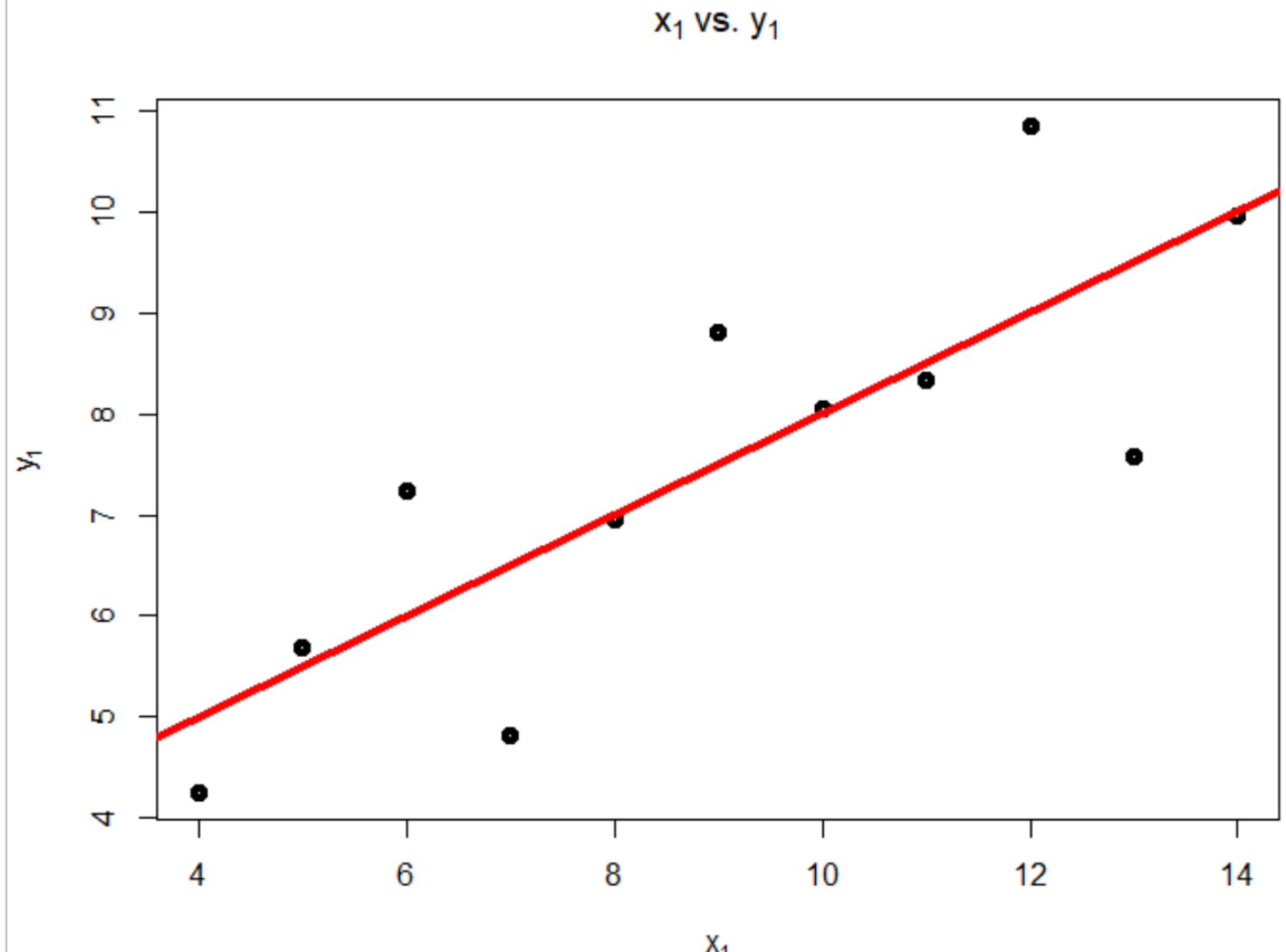


Least squares regression:

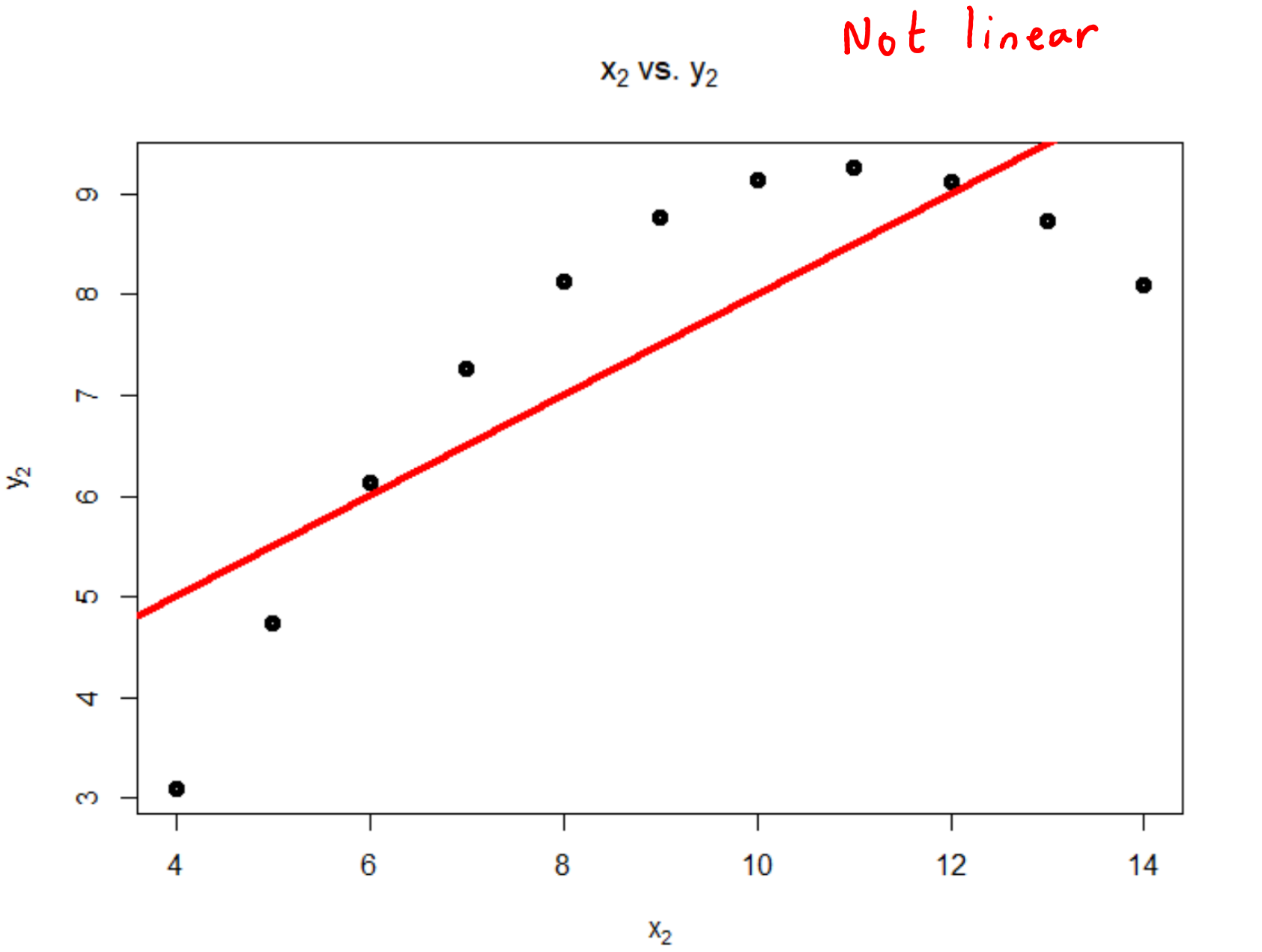
Minimise $\sum (\text{error})^2$

There are other methods.

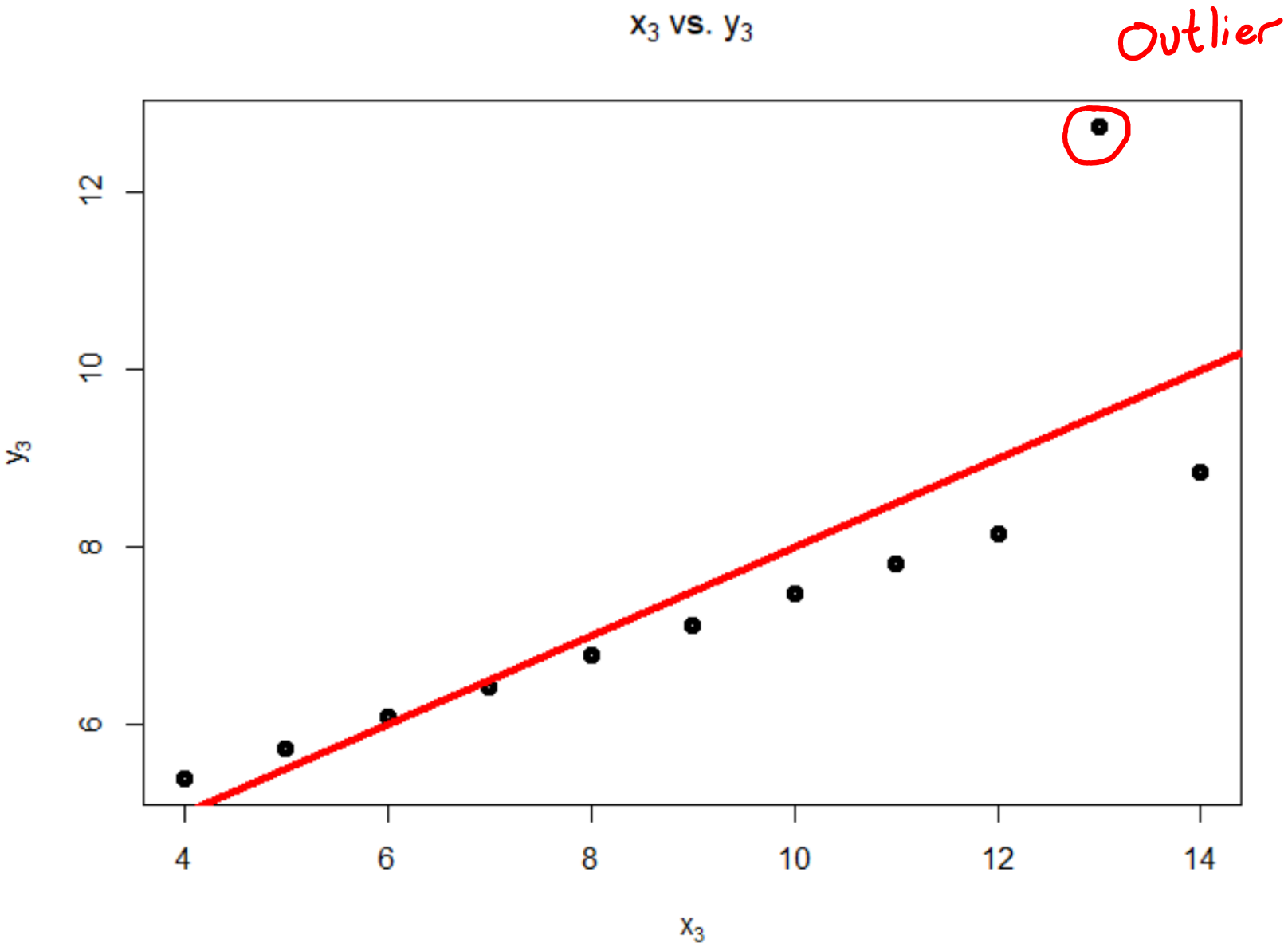
Anscombe data



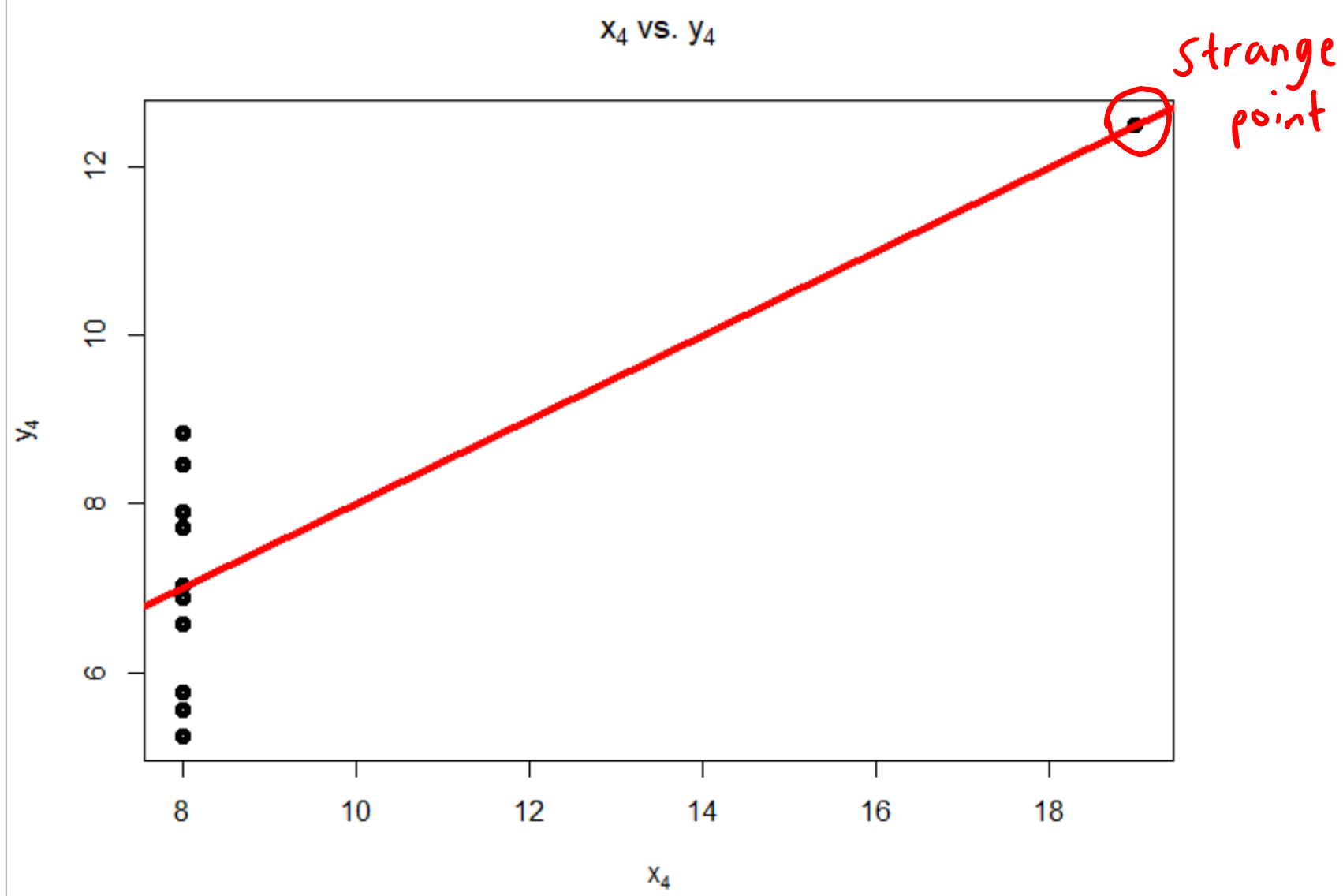
Anscombe data



Anscombe data



Anscombe data



The moral?

Always plot the data first.

Make sure all assumptions are met to use the method.

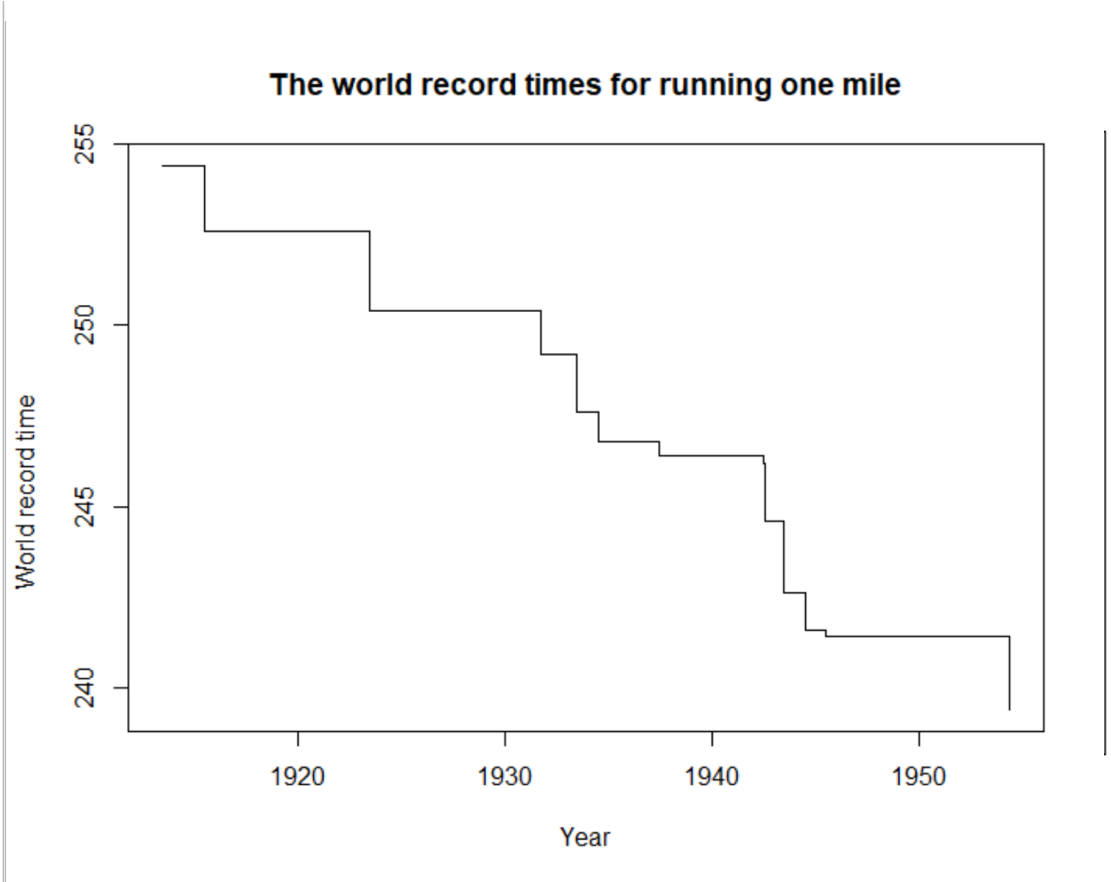
University of Nottingham
UK | CHINA | MALAYSIA

The one mile run world record

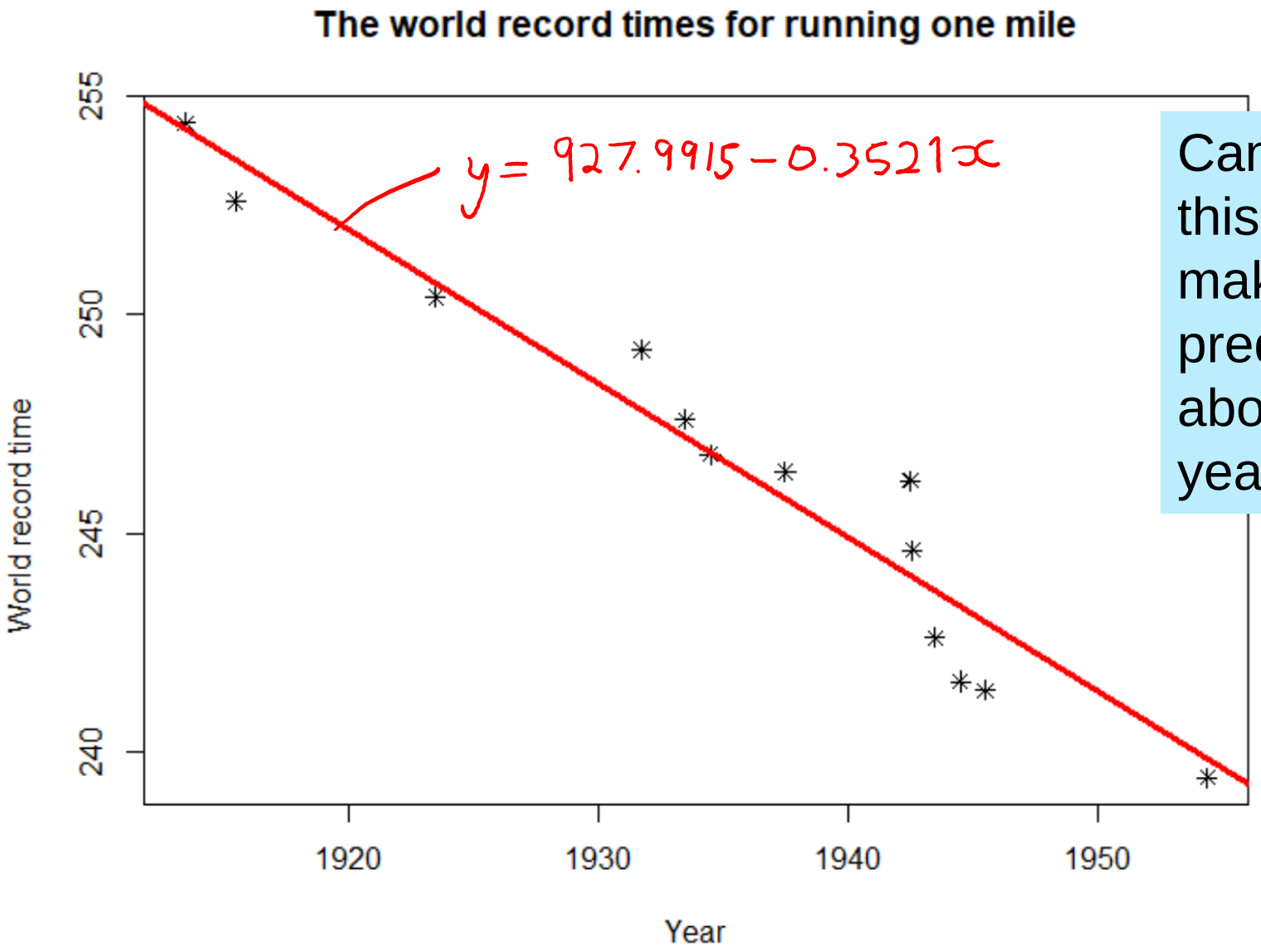
The one mile run world record

The International Association of Athletics Federations (IAAF) have been recording the world records for the time for men to complete one mile since 1913.

First let's look at the times up until 1954.

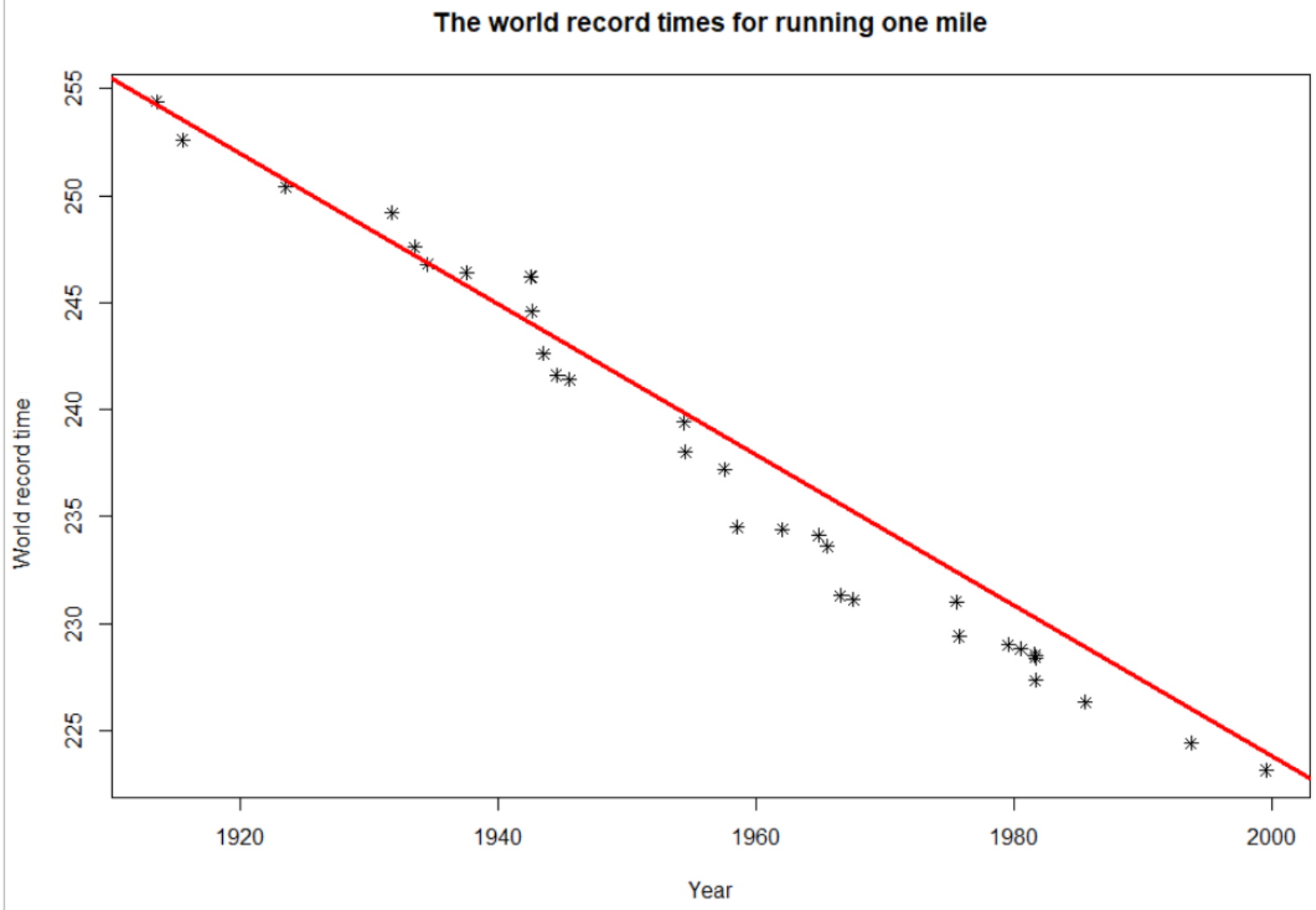


The one mile run world record

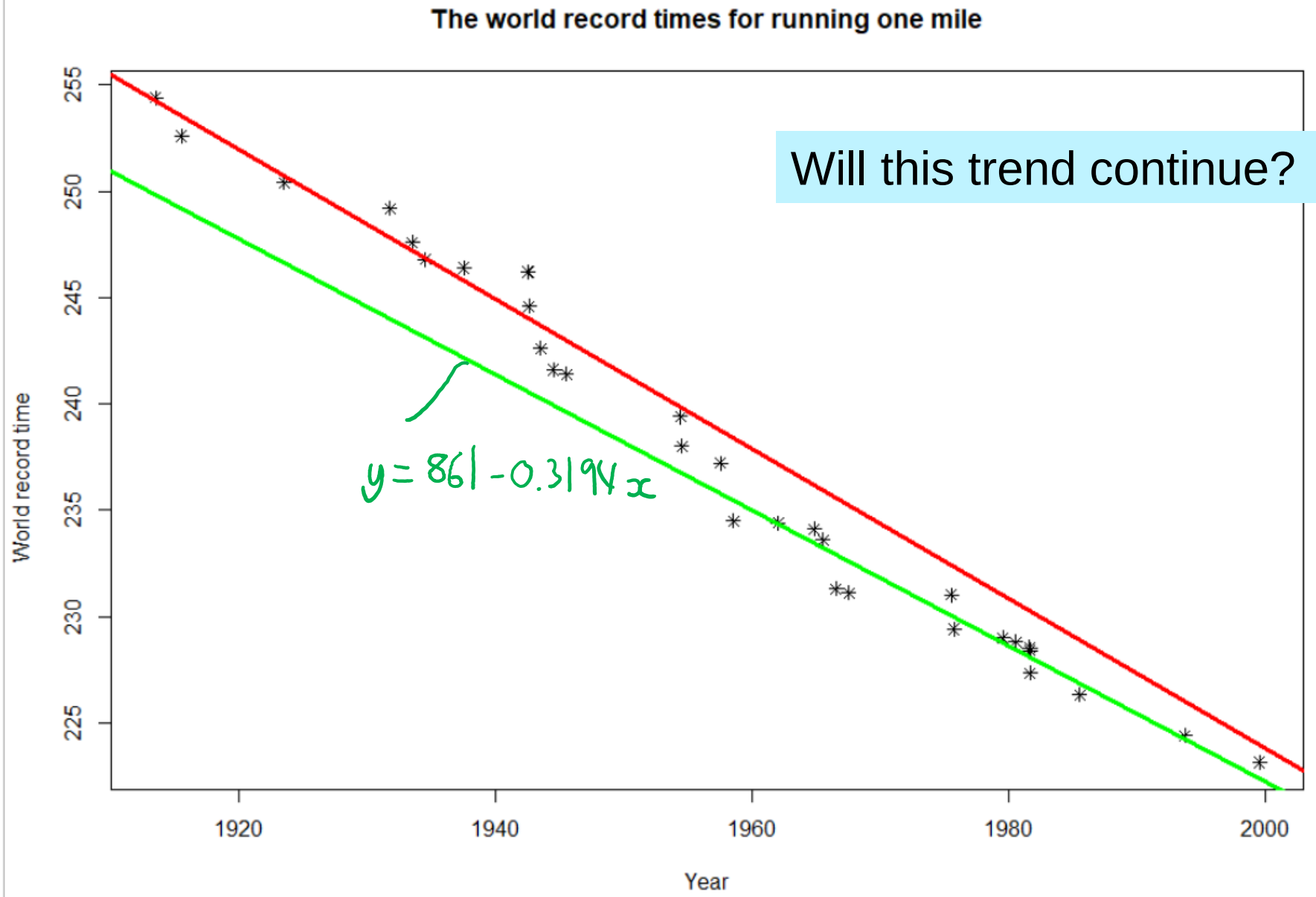


Can we use this line to make predictions about future years?

The one mile run world record



The one mile run world record



Will this trend continue?

The one mile run world record

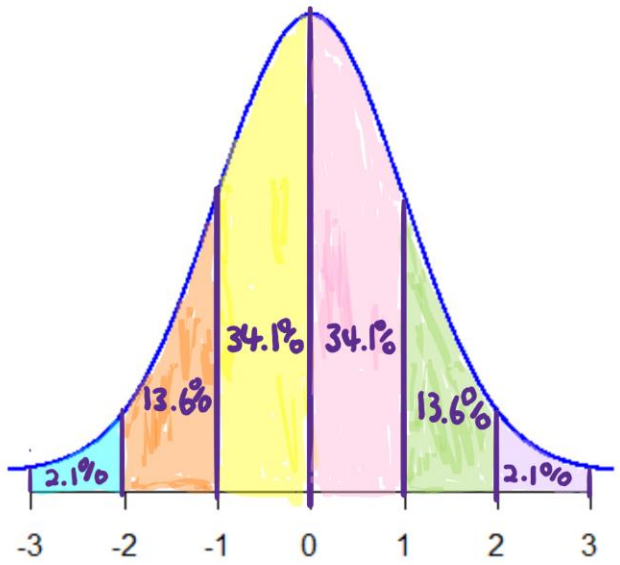
- Moral: Unexpected things may happen.
- There is probably a limit at some point, so be careful about extrapolating data.

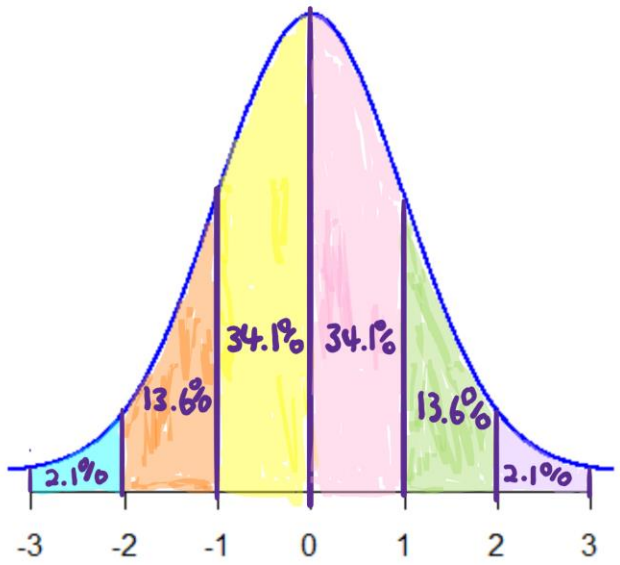
University of Nottingham
UK | CHINA | MALAYSIA

Normal distribution

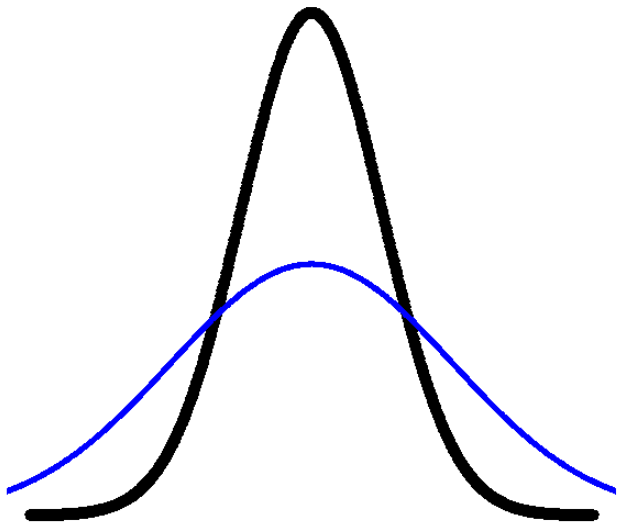
A brief introduction

- There are many applications of the normal distribution in all areas of Science.

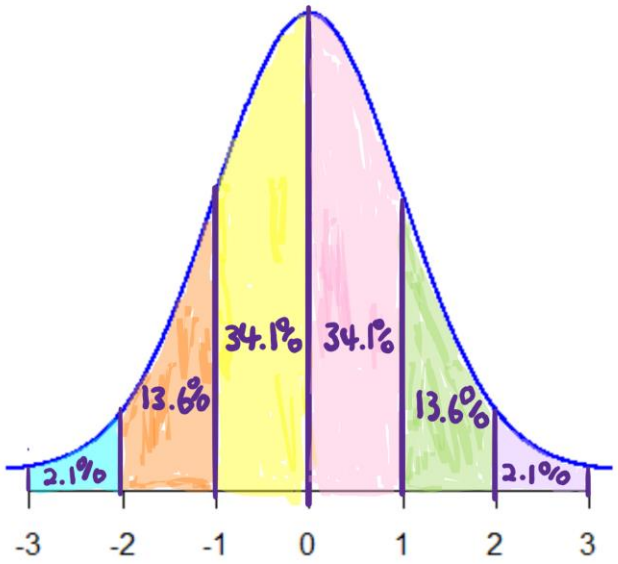




- The normal distribution has a bell-curve shape.
- The normal distribution is described by two parameters the population mean and population standard deviation.
- The standard deviation lets us know how spread the data is.



- For example the two normal distributions shown here have the same mean, but a different standard deviation.

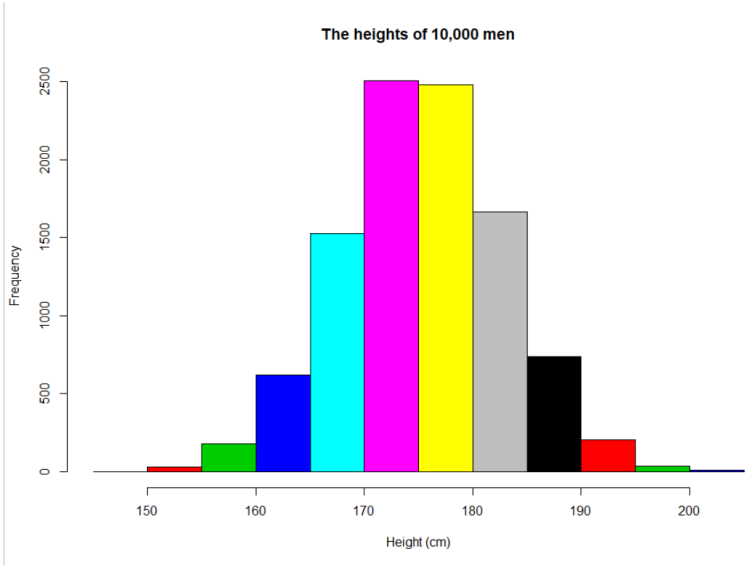


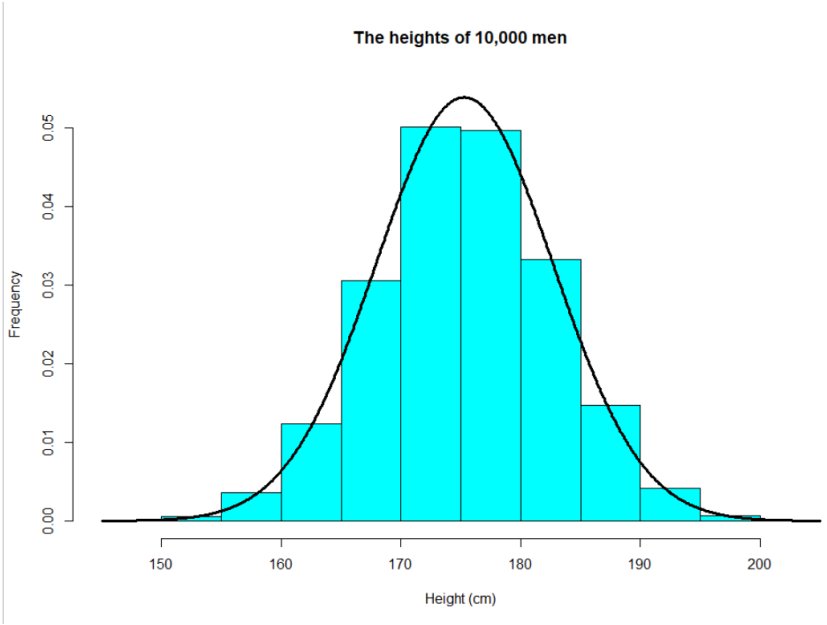
- The normal distribution has a bell-curve shape. The normal distribution is described by two parameters the population mean and population standard deviation.
- The standard deviation lets us know how spread the data is.

10,000 heights

In England the average height of men is 175.3 cm with standard deviation of 7.4 cm

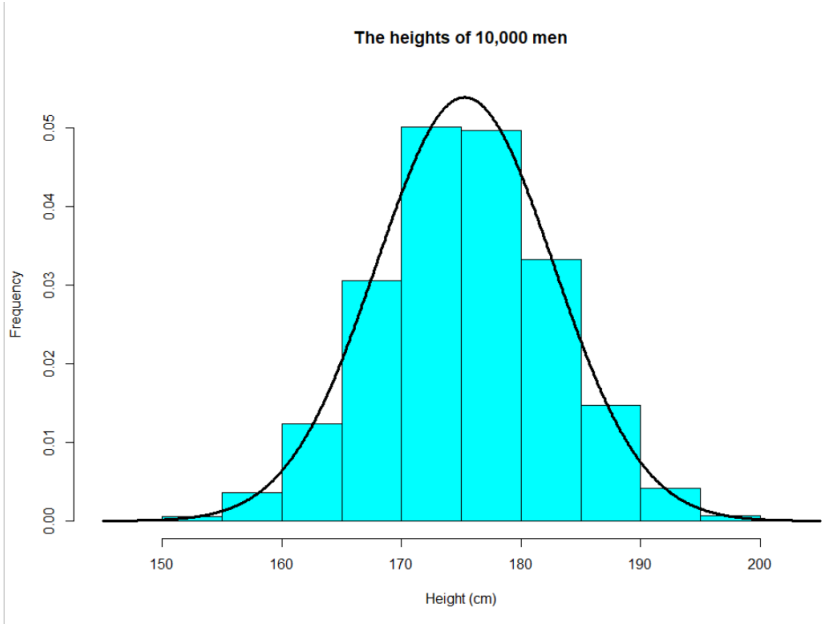
The average height of women is 161.9 cm with standard deviation of 6.9 cm





Moody A 10: Adult anthropometric measures, overweight and obesity. In Craig R, Mindell J (eds.) Health Survey for England - 2012

Answer the following questions



- Approximately how many of the 10,000 men are between 175.3 cm and 182.7 cm?
- Approximately how many of the 10,000 men are less than 160.5 cm ?
- Approximately how many of the 10,000 men are between 160.5 cm and 190.1 cm ?

Answer these using the quiz

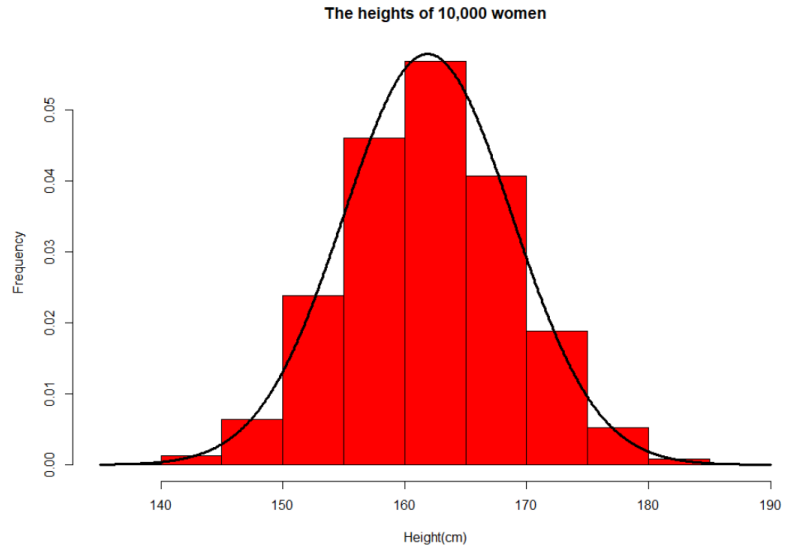
- Extra: How can we estimate the area under a curve?

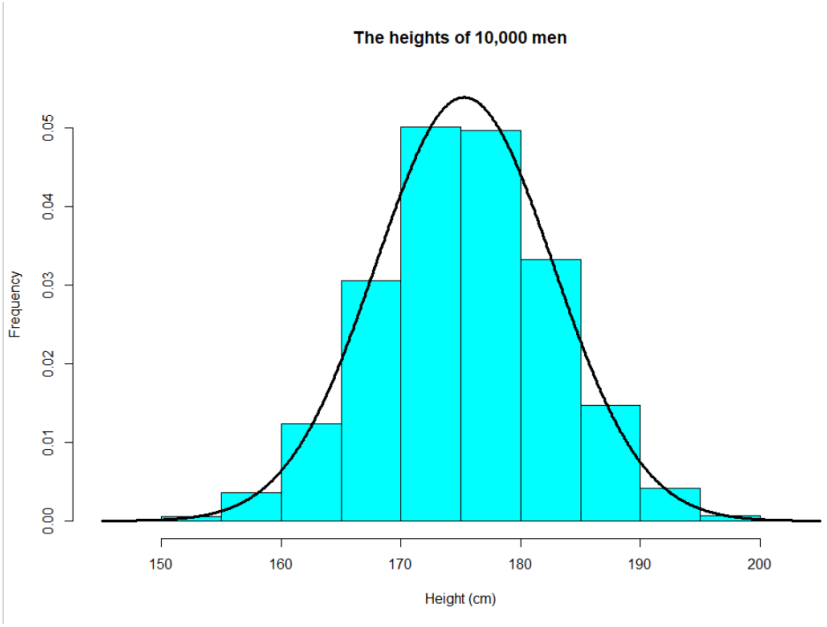
Combining normal distributions

10,000 heights

In England the average height of men is 175.3 cm with standard deviation of 7.4 cm

The average height of women is 161.9 cm with standard deviation of 6.9 cm



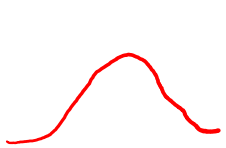


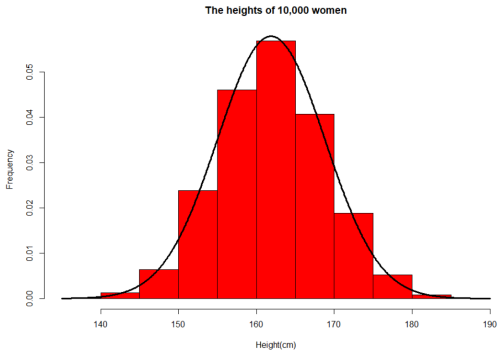
Moody A 10: Adult anthropometric measures, overweight and obesity. In Craig R, Mindell J (eds.) Health Survey for England - 2012

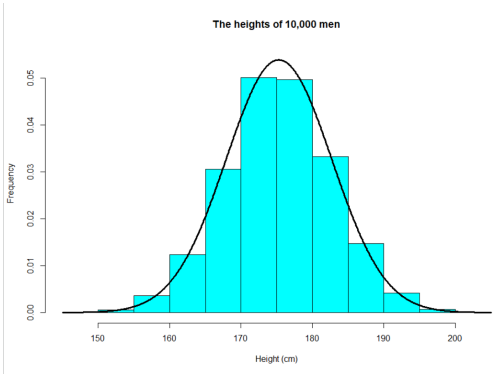
10,000 heights

If we combine the heights of 10,000 men and 10,000 women, the overall distribution will be

a) Unimodal b) Bimodal c) None of these







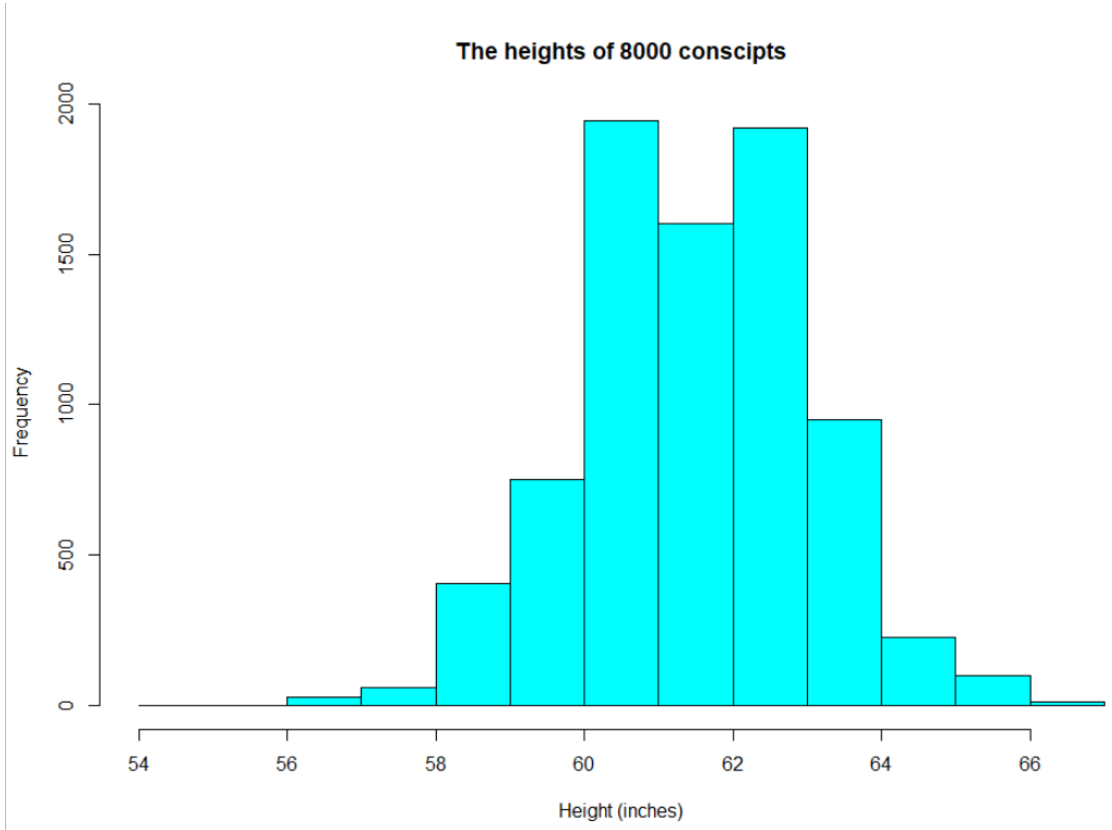
Moody A 10: Adult anthropometric measures, overweight and obesity. In Craig R, Mindell J (eds.) Health Survey for England - 2012

Conscripts in France

In *The History of Statistics: The Measurement of Uncertainty Before 1900, Stigler (1986)* looks at the heights of conscripts in France from the 1850s.

There are two groups of men Celts and Burgundians.

Conscripts in France



What does this graph appear to show?

Can you trust this hypothesis?

What might be a different reason for the 2 modes?

Conscripts in France

Moral: Don’t trust data unless you know how it has been collected and measured.

Conclusion

- We have only just touched on a few problems in Statistics
- Don’t trust all results that you see reported as fact.
- Any questions welcome!

Related Maths BSc/MMath modules

Study year

- Year 1
- Year 2
- Year 3
- Year 4 (MMath)

Related modules

- Fixed programme (includes Statistics)
- Statistical Models and Methods
- Stochastic Models, Statistical Inference, Mathematical Finance, Time Series Analysis
- Computational Statistics, Statistical Machine Learning

Before you apply

- Study with us: maths course pages

https://tinyurl.com/mathscourseuon

- Open day maths talk videos

https://tinyurl.com/uonmathsvod

Before you apply

- What do our maths graduates do?

https://tinyurl.com/uonmathscareers

- News about maths at Nottingham, interesting facts, puzzles and links to useful resources

https://tinyurl.com/uonmathsnewsform

Taster session links

- Discover some of the taster lectures and popular maths talks we've recently recorded:

- Topics have included
  - Calculus
  - Using maths in the fight against Covid-19
  - Problem solving
  - Pure mathematics

https://tinyurl.com/uonmathstaster