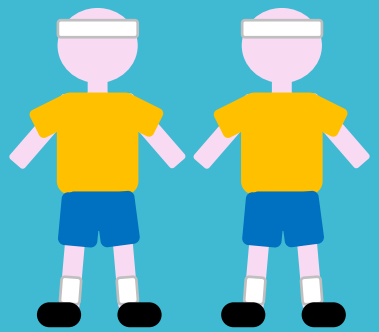
Linear regression: Part 1



Lecture Outline

What are linear models?

What is linear regression?

How do we fit linear regressions?

Fitting in R

Lecture Outline

What are linear models?

- EX1: What is the ‘best’ line?

What is linear regression?

- EX2: When to use linear regression

How do we fit linear regressions?

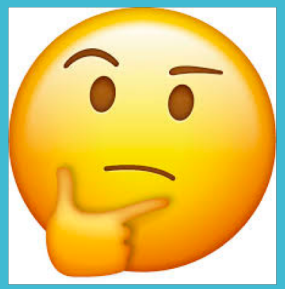
- EX3: Trying fitting a line

Fitting in R

Reading

Chapter 4 – The New Statistics with R

What are linear models?



Definition

Linear models:

Models with a **continuous response** variable as a function of one or more **explanatory** variable. Variables are connected by **linear equations**.

We want to explain variable Y with variable X.

Definition

Linear models:

Models with a **continuous response** variable as a function of one or more **explanatory** variable. Variables are connected by **linear equations**.

$$Y_i = \alpha + \beta X_i + \varepsilon_i$$

Definition

Linear models:

Models with a continuous **response** variable as a function of one or more **explanatory** variable. Variables are connected by **linear equations**.

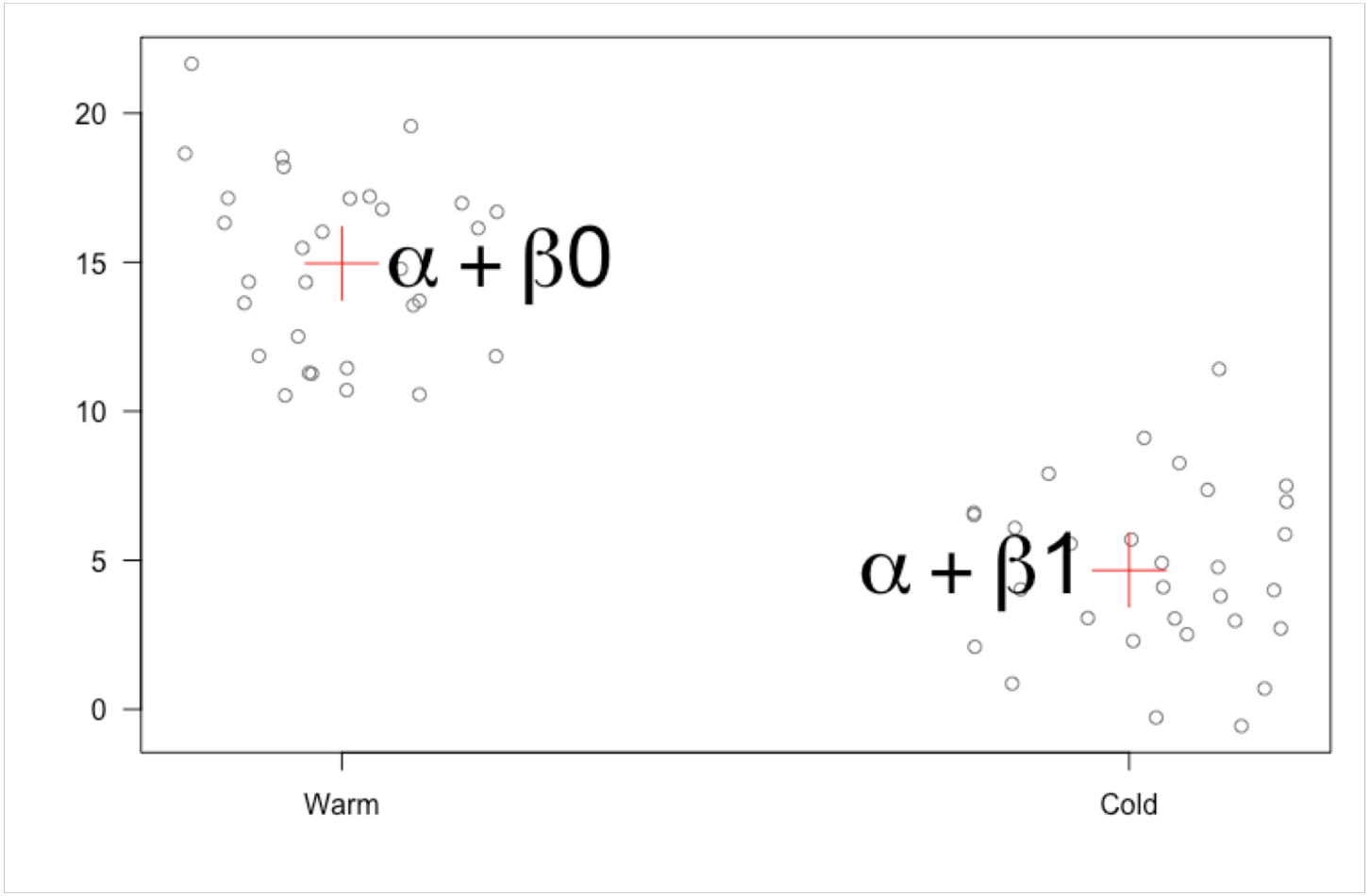
$$Y_i = \alpha + \beta X_i + \varepsilon_i$$

parameters

error

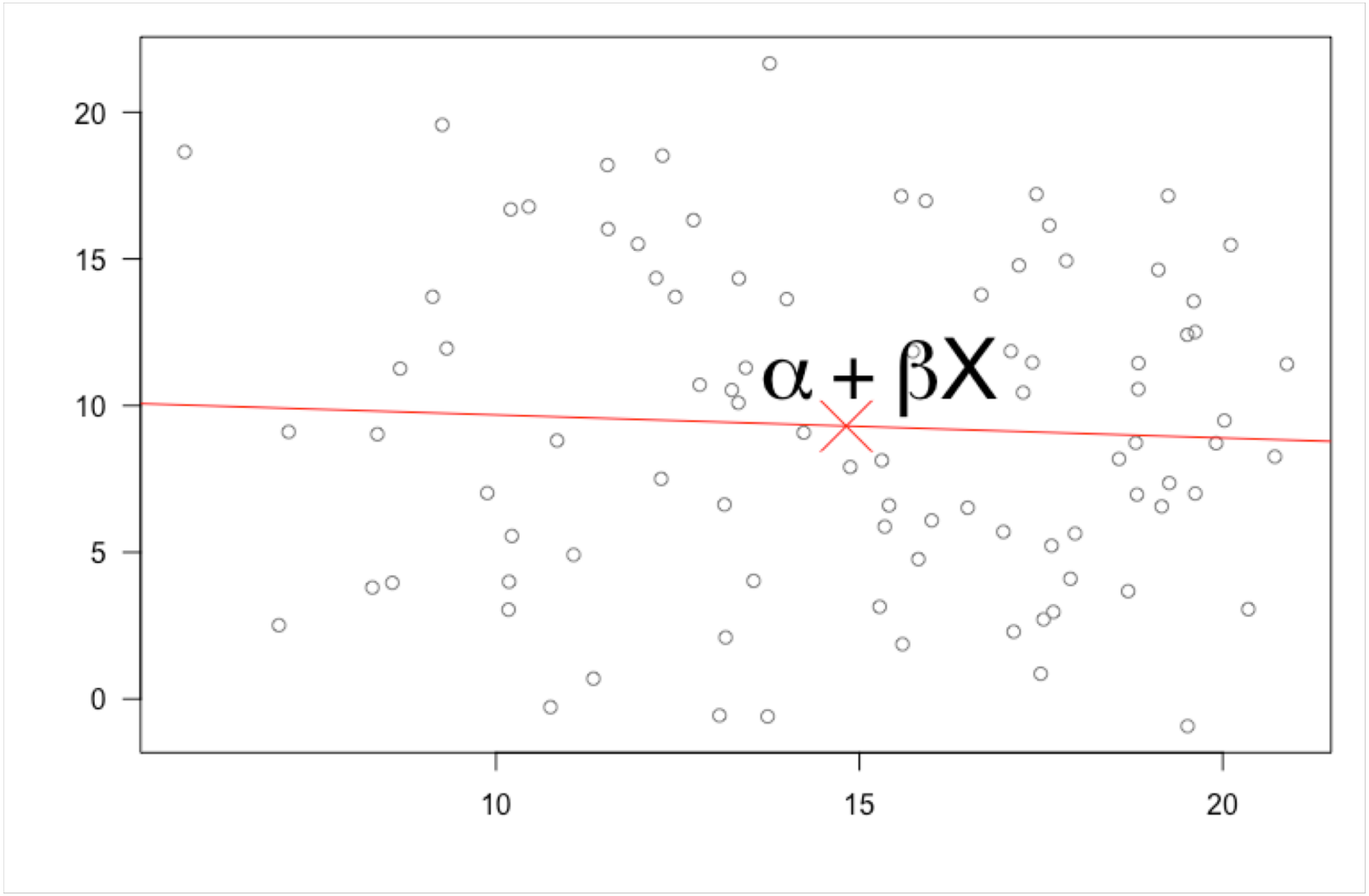
Examples of linear models

T-test (last week) $\mu_i = \alpha + \beta X_i$



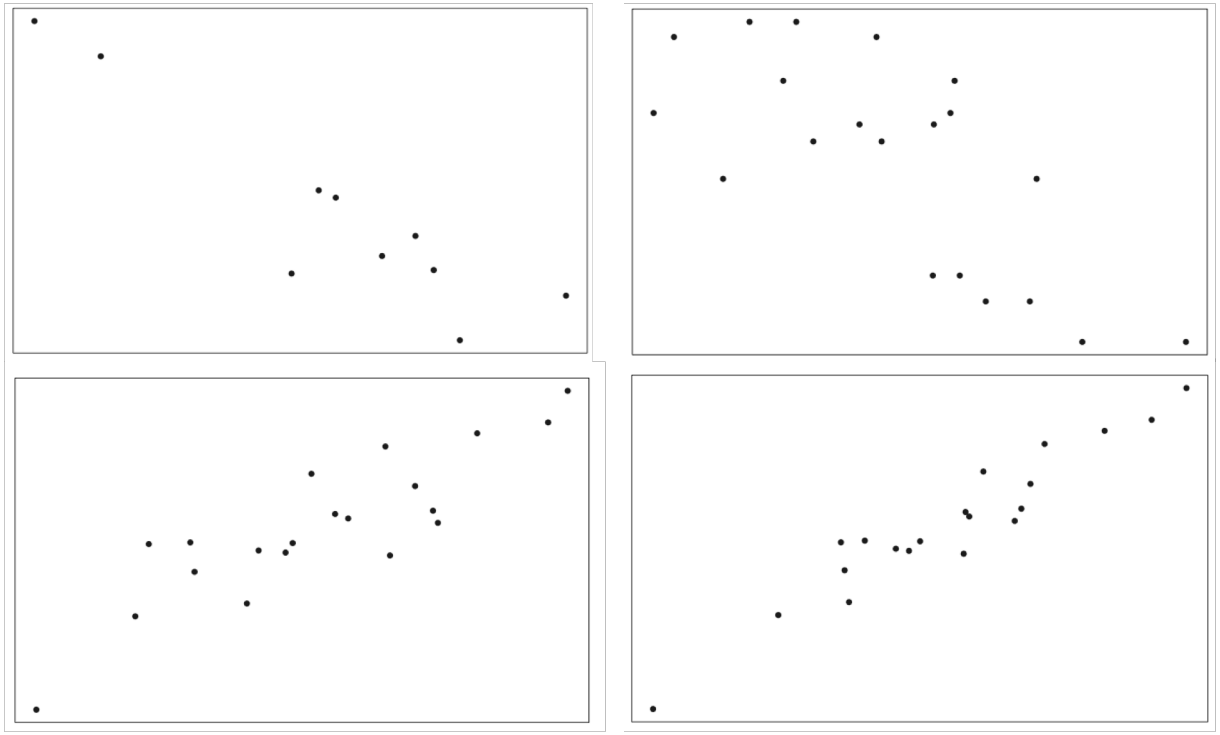
Examples of linear models

Regression (this week) $\hat{Y}_i = \alpha + \beta X_i$

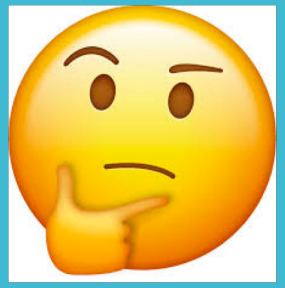


Exercise 1: What is the ‘best’ line?

- Take a look at the four datasets below.
- For each: **draw a ‘best’ line on white boards.**



What is linear regression?



Linear regression

Linear regression:

Predicts values of a response variable from values of an explanatory variable.

Linear regression

In simple terms:

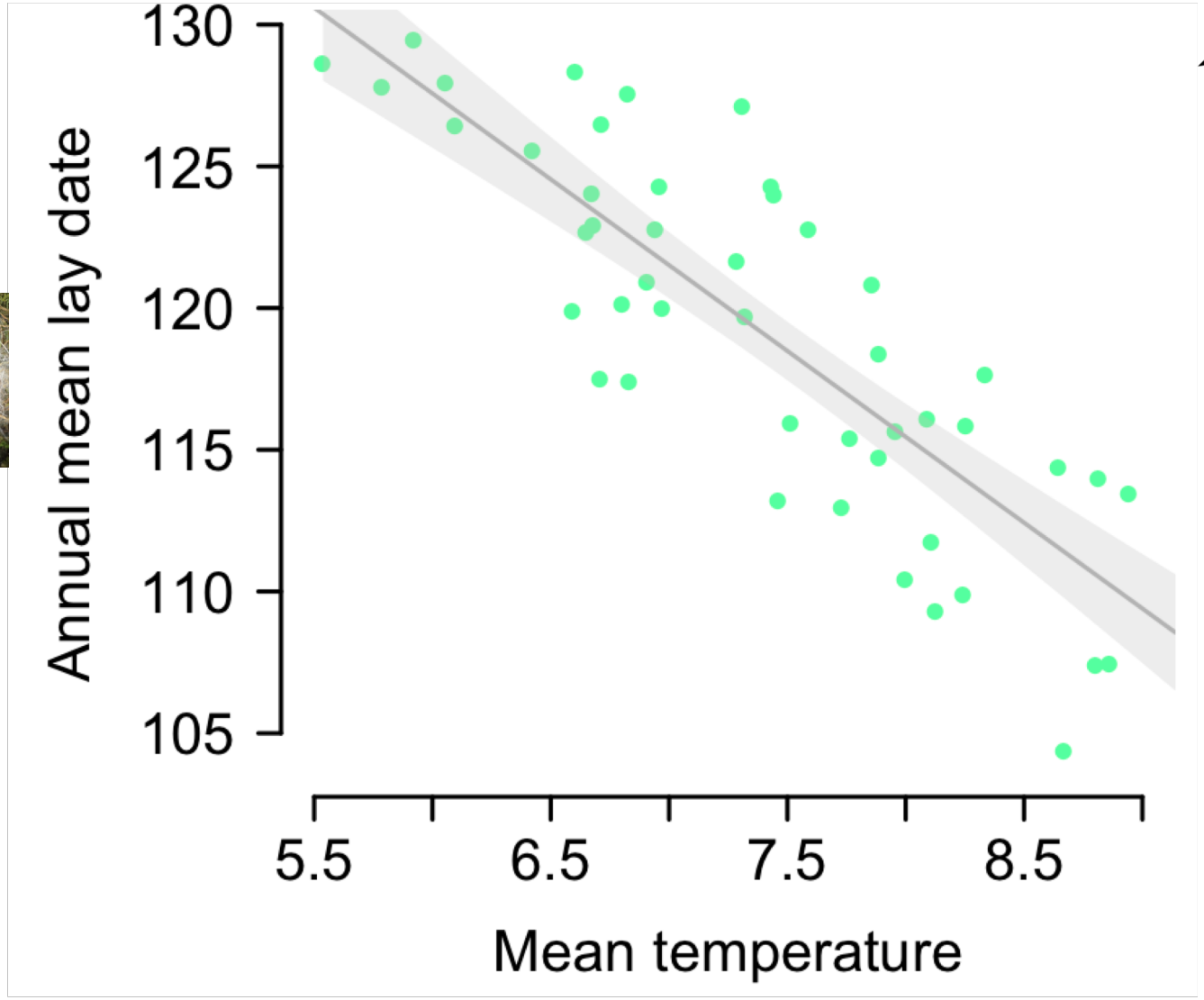
We fit a straight line to:

Estimate relationship between X and Y

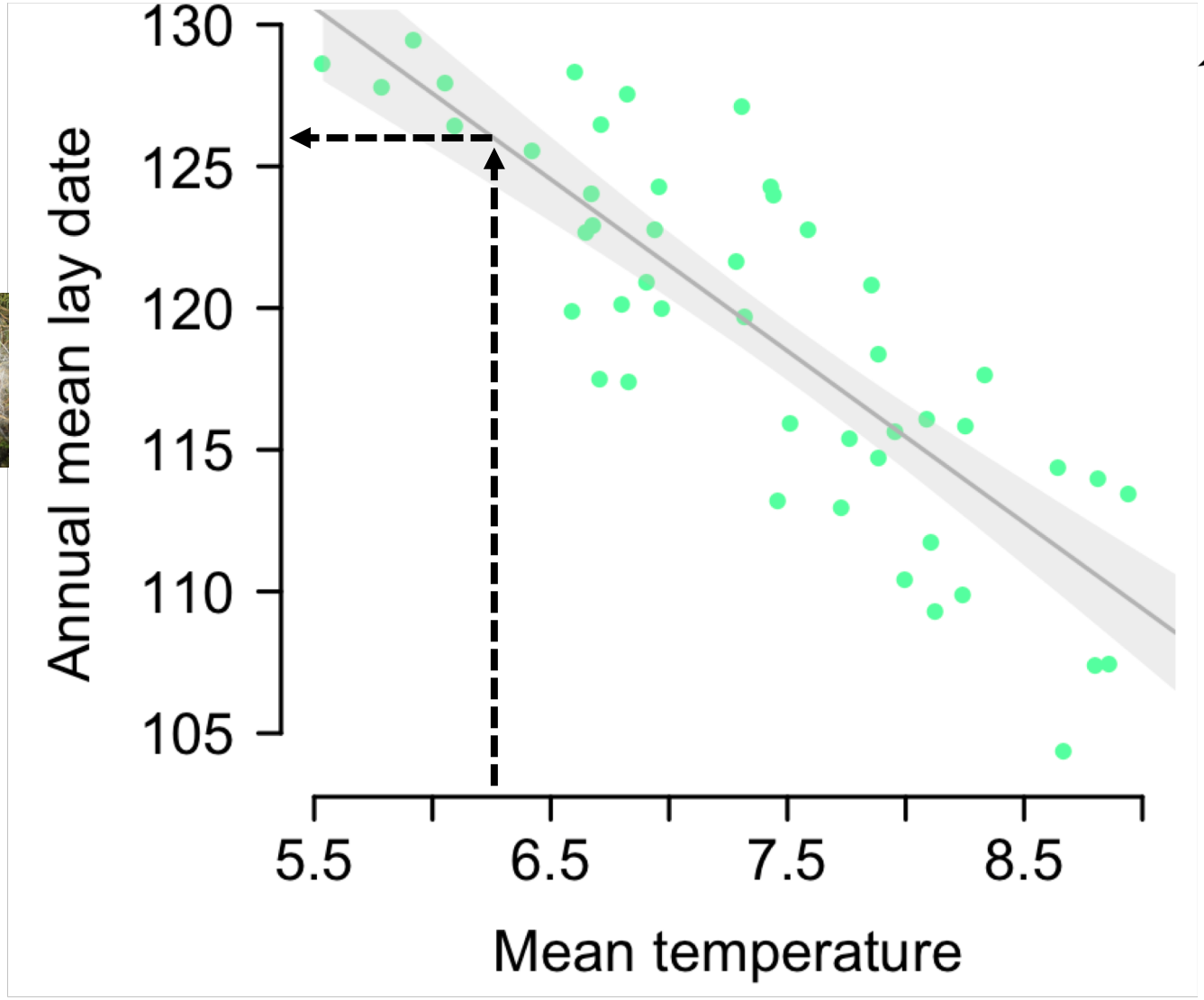
Predict change in Y from change in X

Linear regression - Example

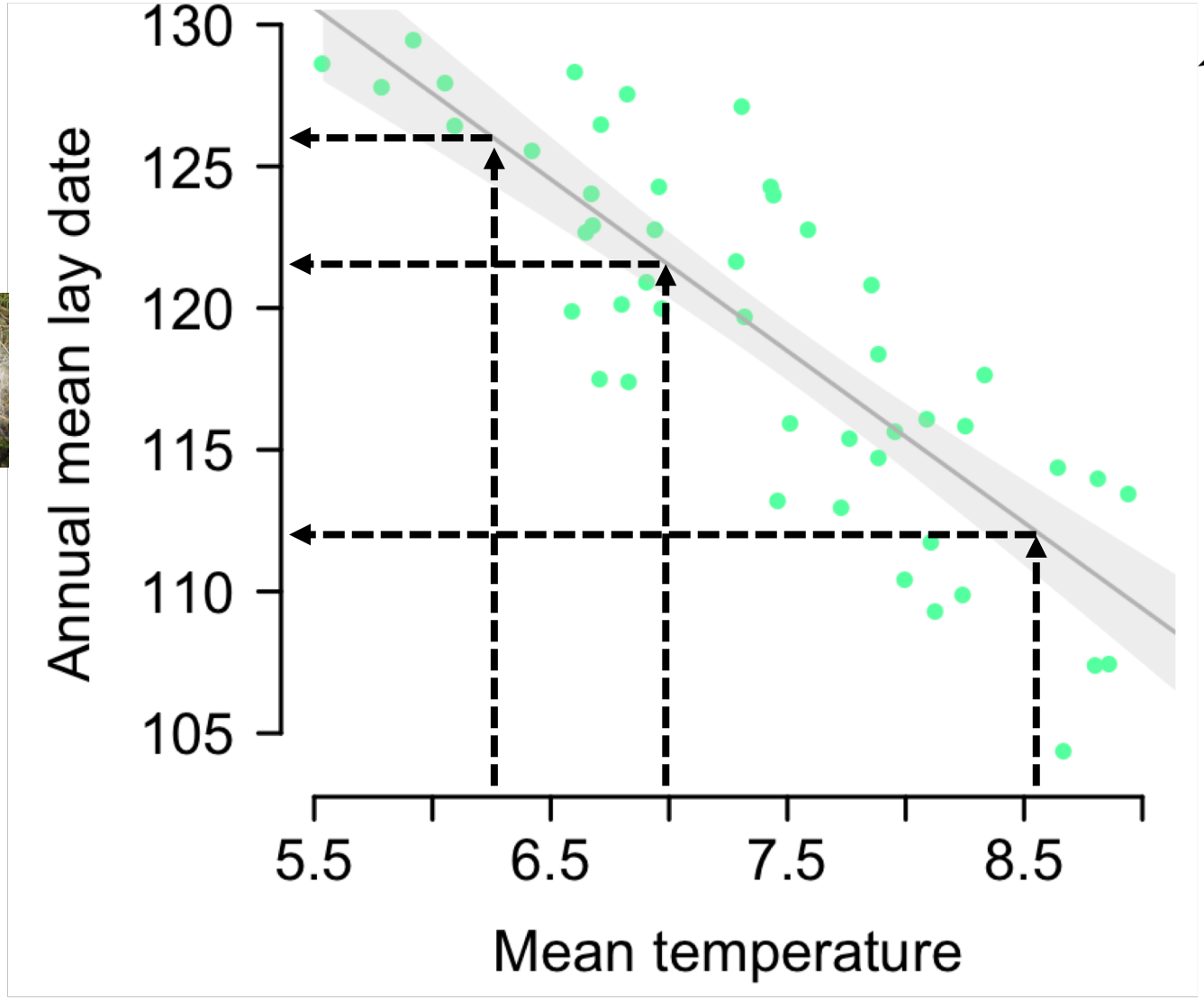
Estimate **relationship** between temperature and lay date



Linear regression - Example

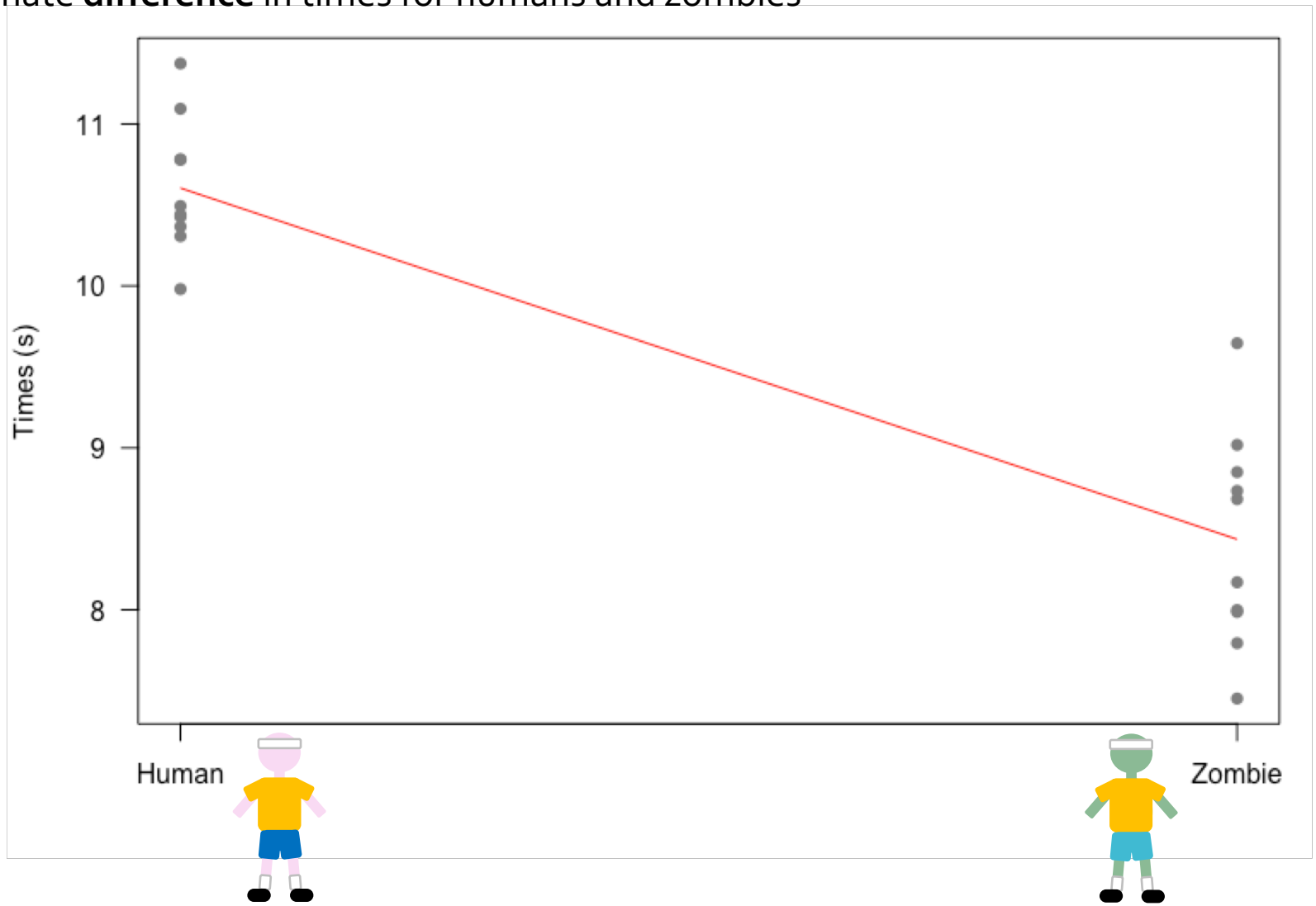


Linear regression - Example

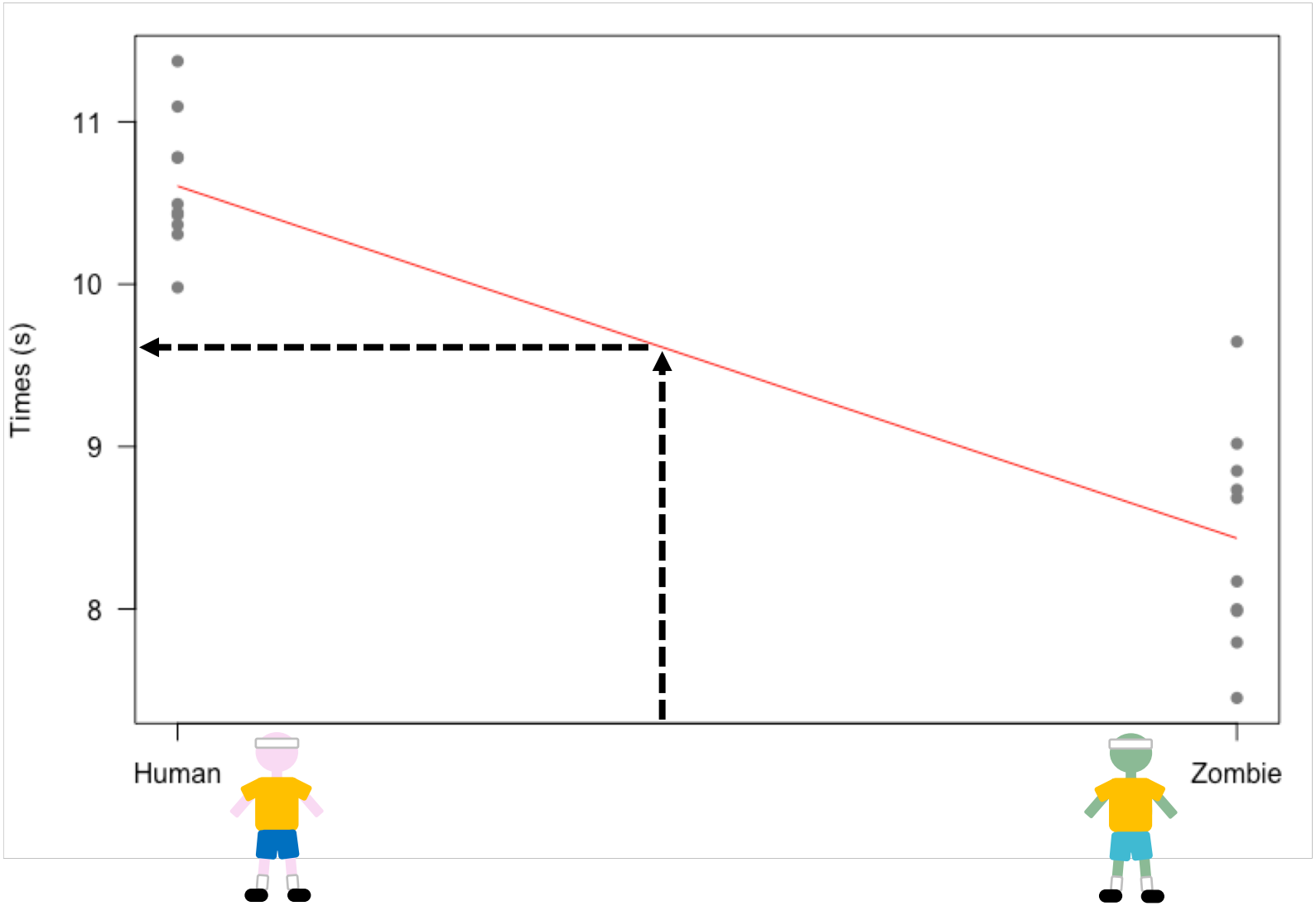


Linear regression – Example 2

Estimate **difference** in times for humans and zombies



Linear regression – Example 2

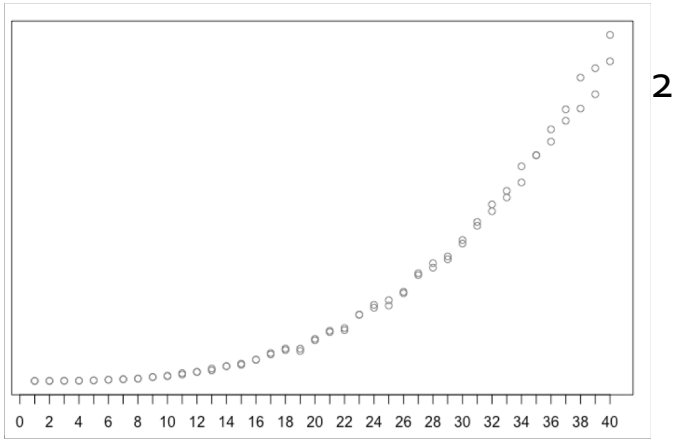


Exercise 2: When to use linear regression.

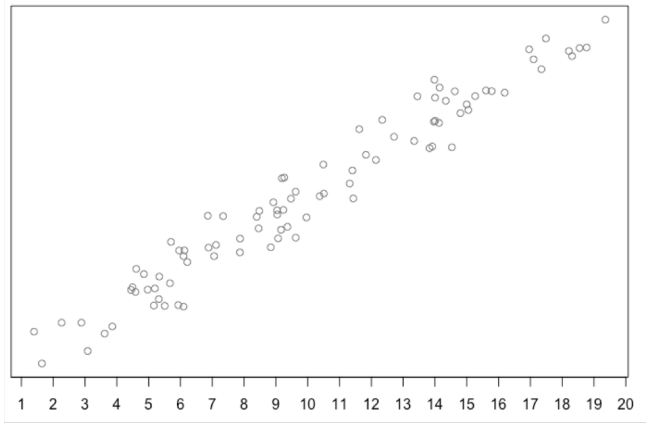
- Take a look at the four datasets below.
- For each, answer the question: **Is a linear regression a suitable model for this data?** (would a straight line work?)

1

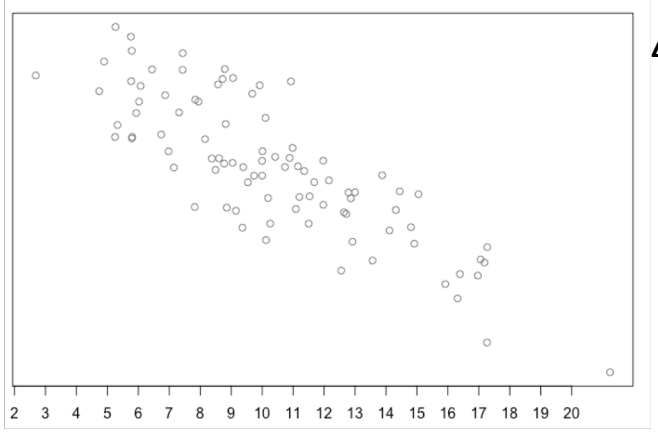
2



3



4

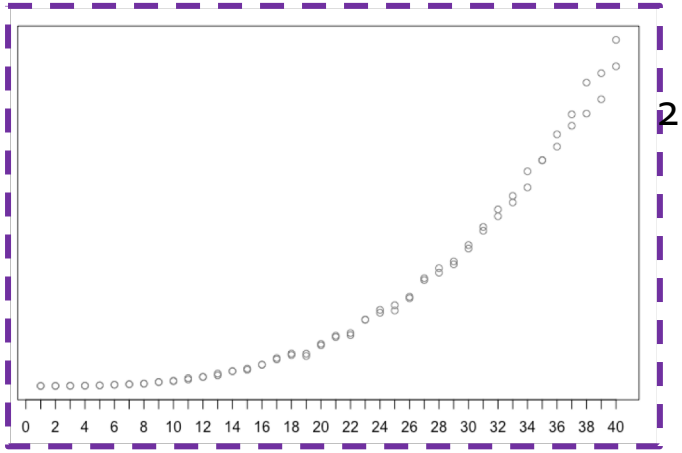


Exercise 2: When to use linear regression.

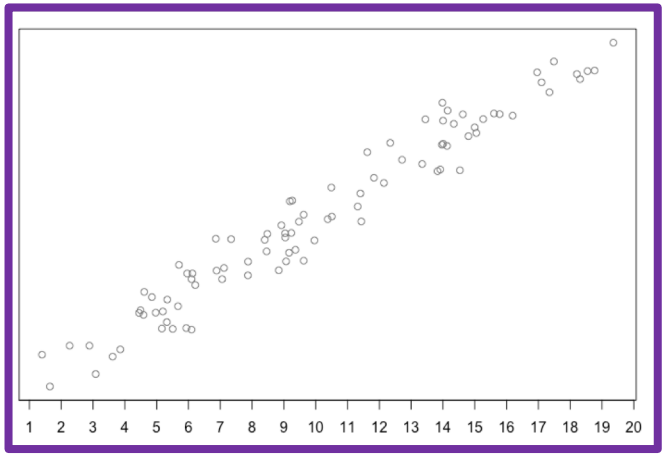
- Take a look at the four datasets below.
- For each, answer the question: **Is a linear regression a suitable model for this data?**

1

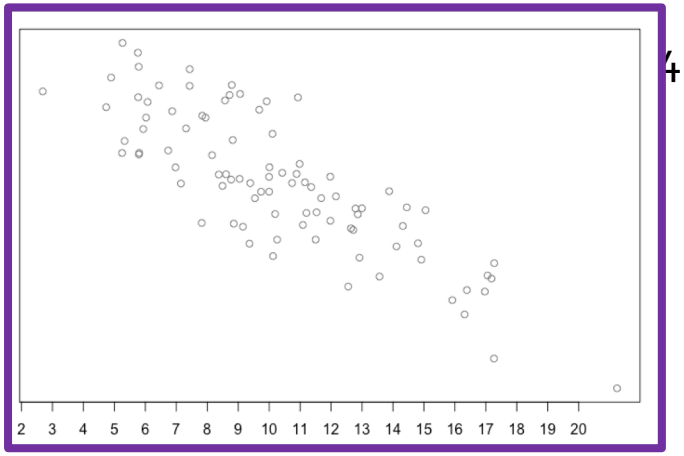
2



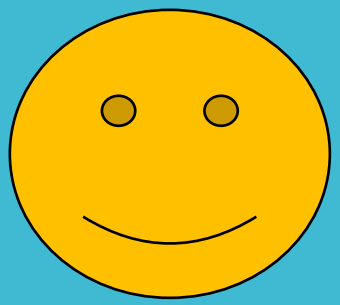
3

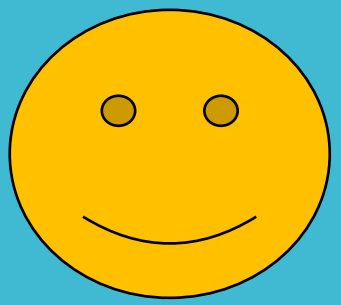


4

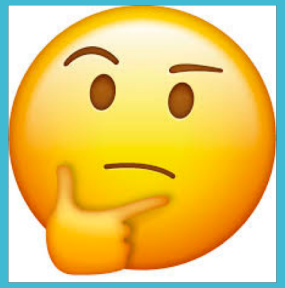


BREAK!





How do we fit a linear regression?



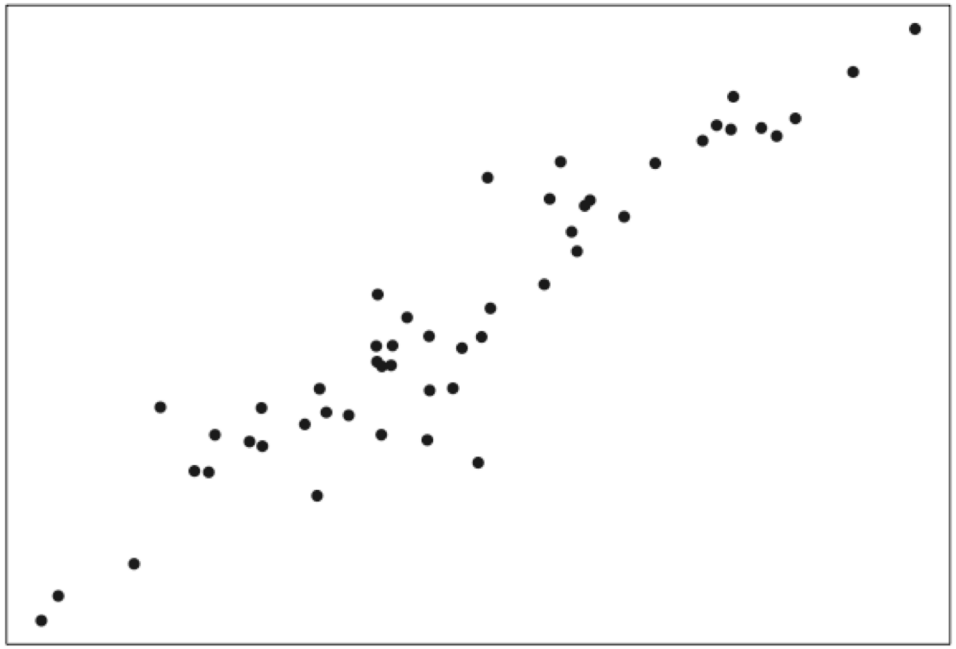
What is a ’best’ line?

Many different lines could be fitted to the same data

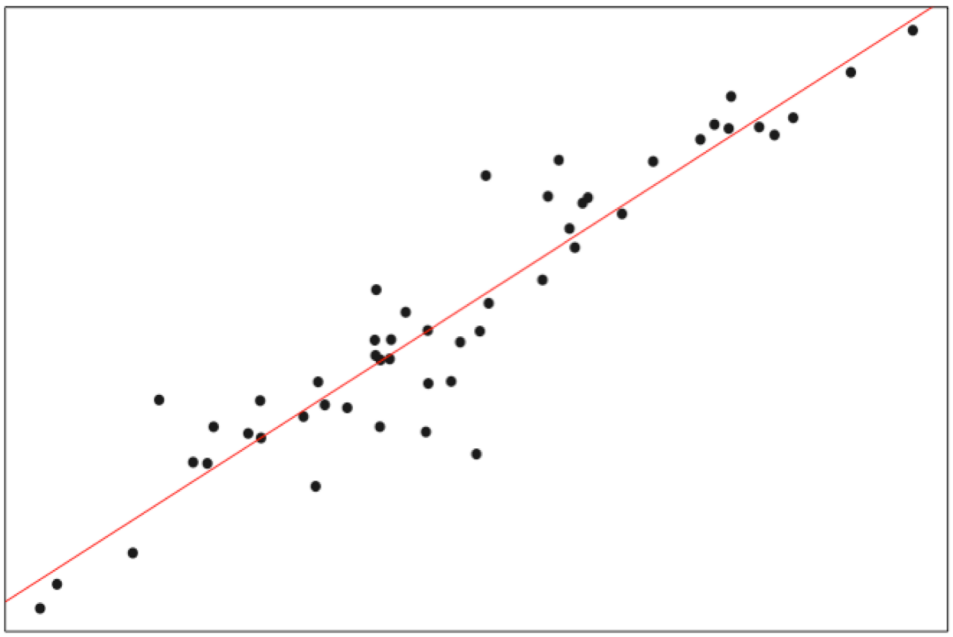
Can try to do it by eye

But also a mathematical way

Fitting the line

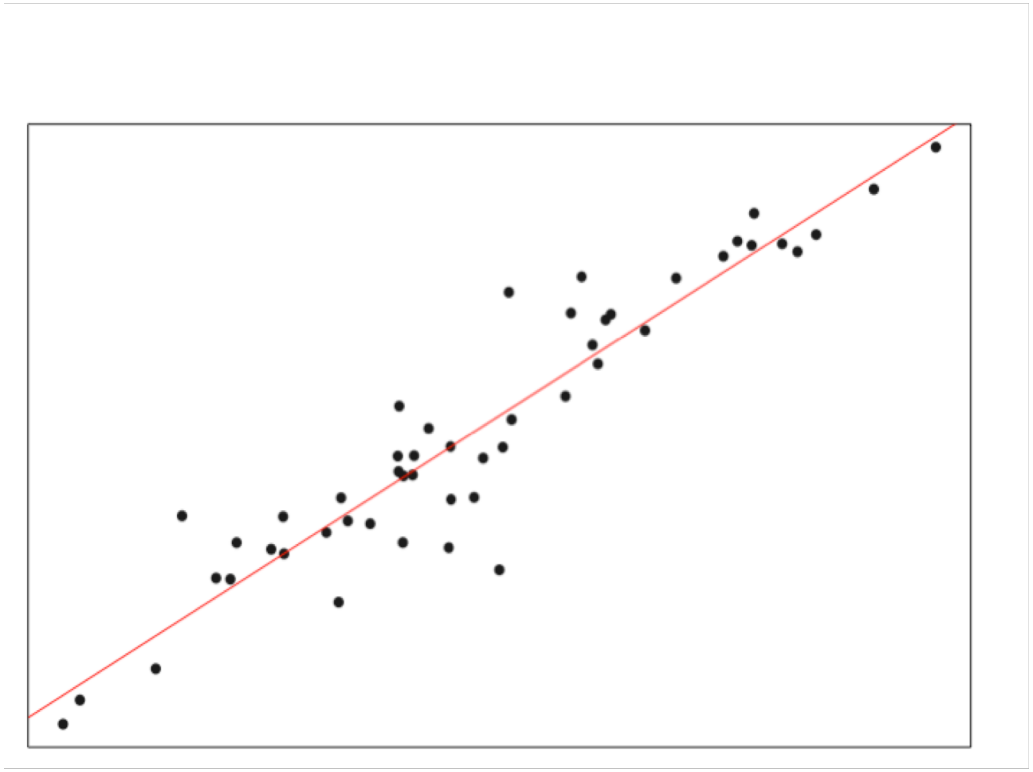


Fitting the line



Need estimates of parameters:
α and β (intercept and slope)

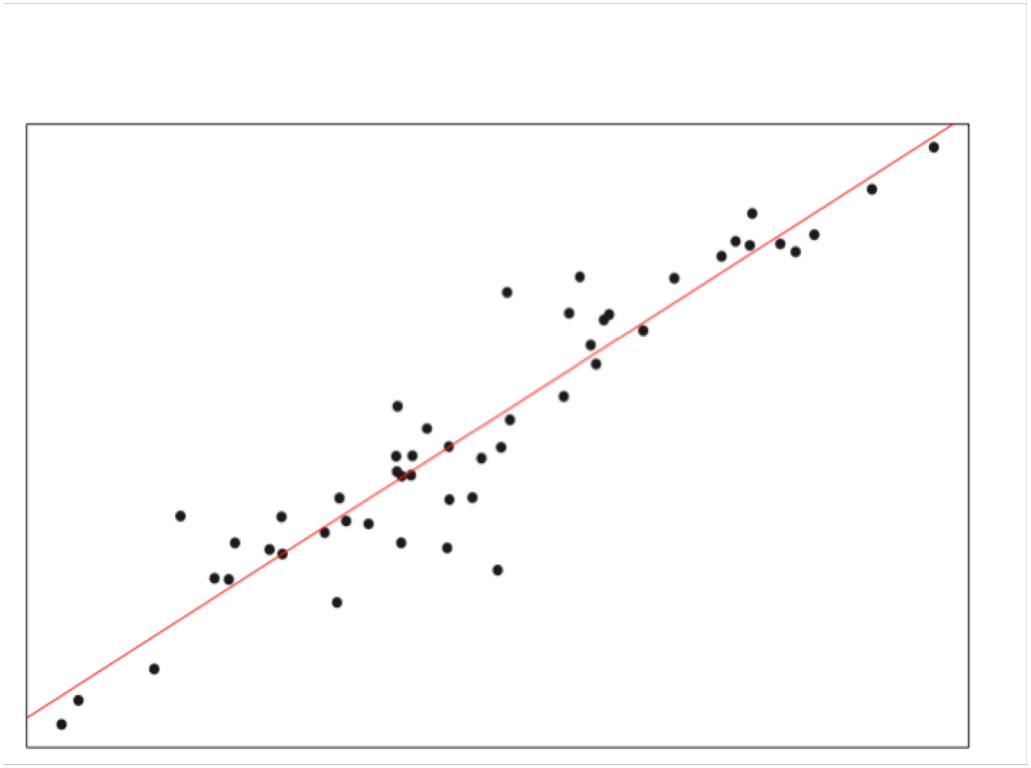
Fitting the line



Need estimates of parameters:
α and β (intercept and slope)

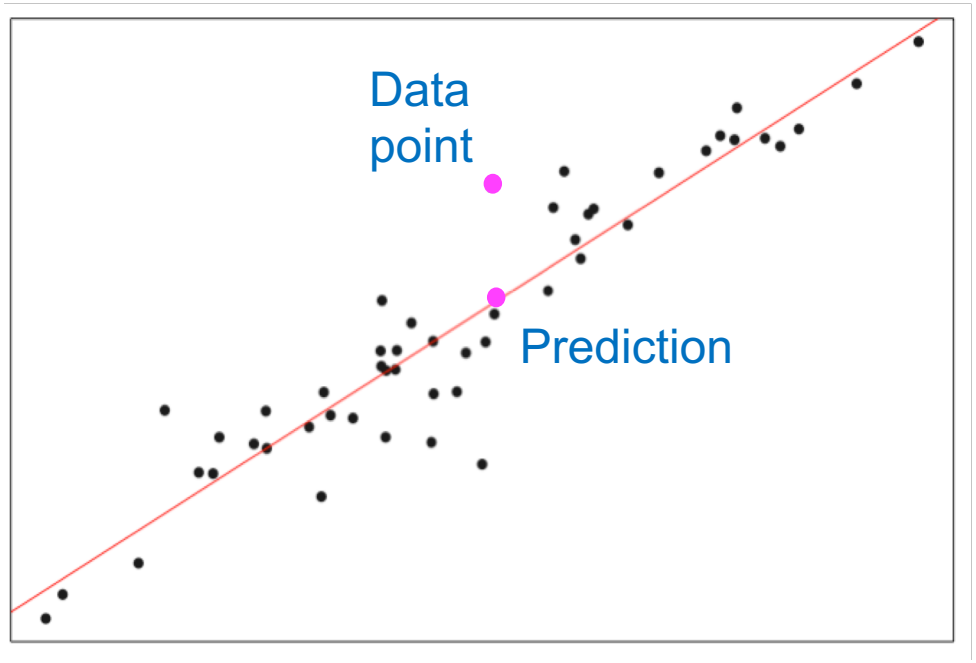
Do this using the maximum likelihood

Fitting the line



Find the parameter values that have the highest likelihood given the data (**maximum likelihood**)

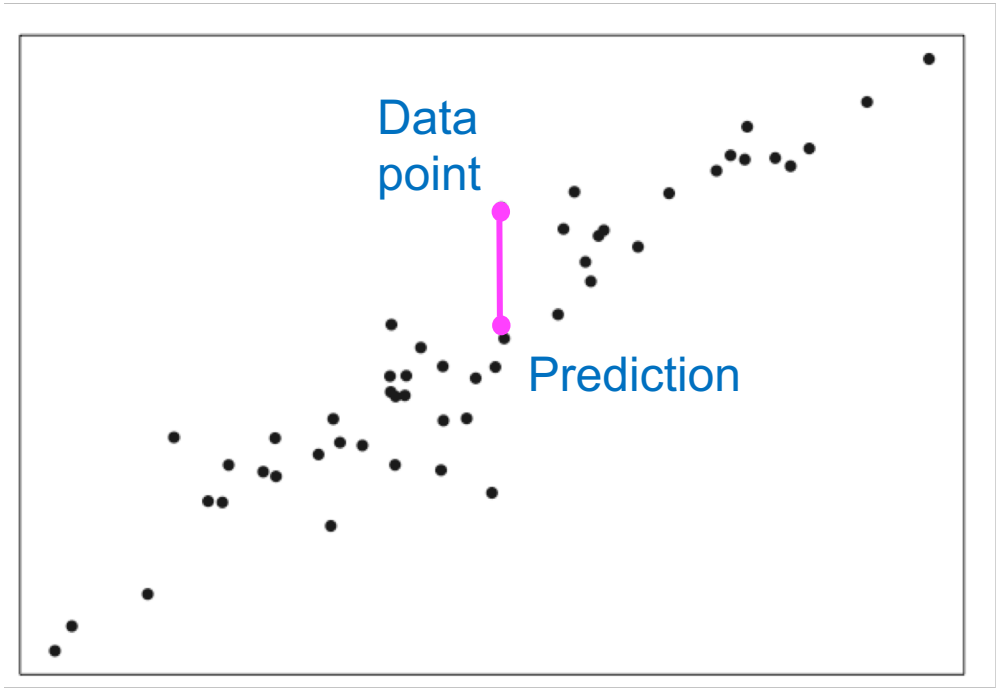
Fitting the line to data



Fitting the line

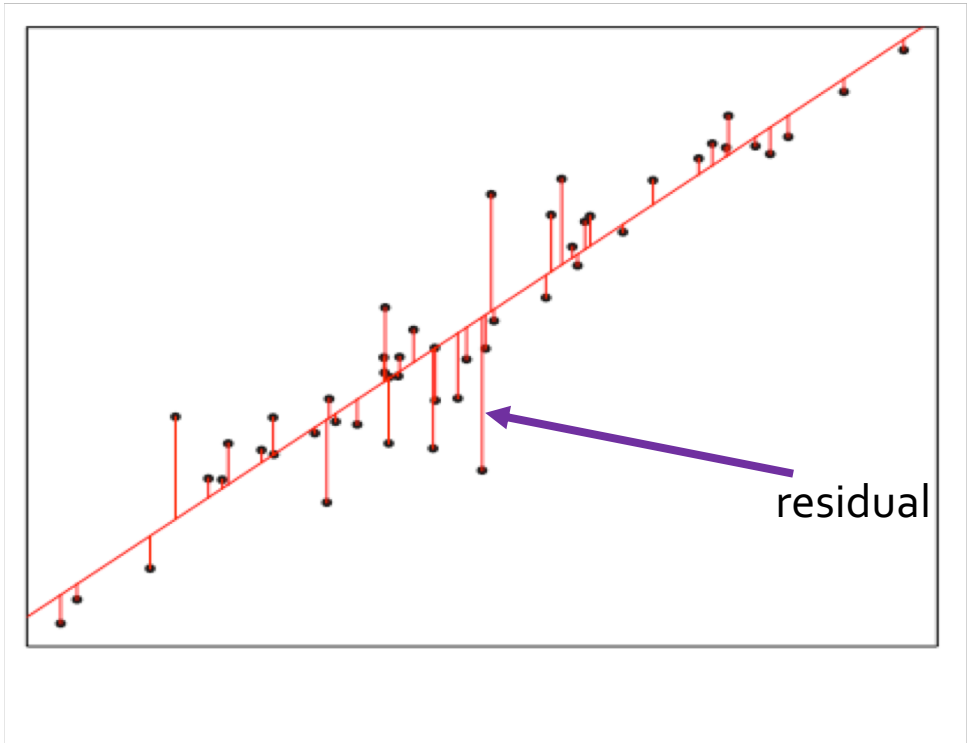
Distance between them = error (residual)

$$\hat{y}_i - y_i$$



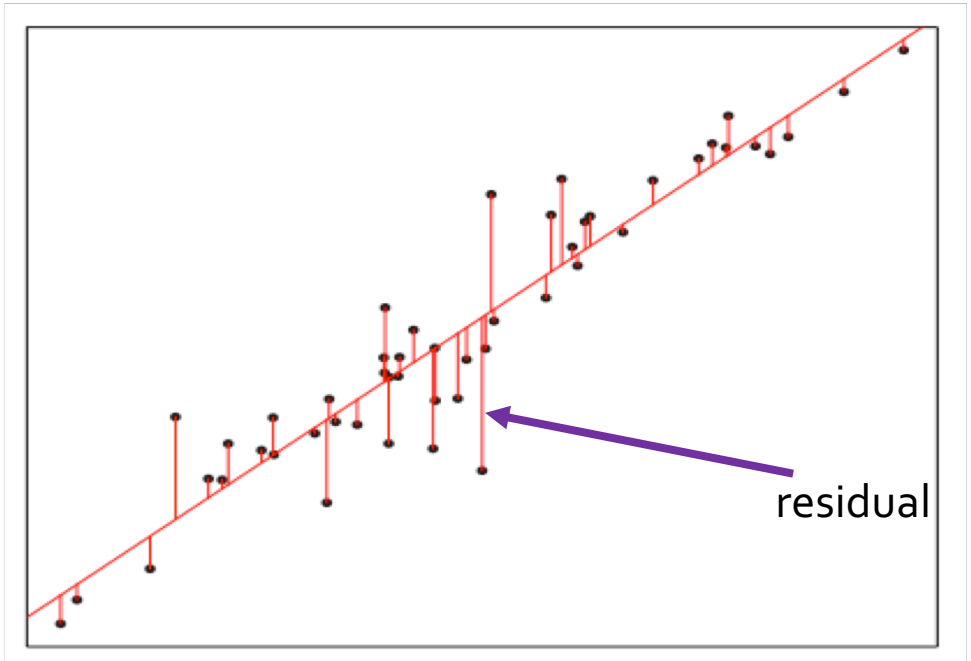
Fitting the line

We assume these residuals are normally distributed at each X value



Fitting the line

We assume these residuals are normally distributed at each X value



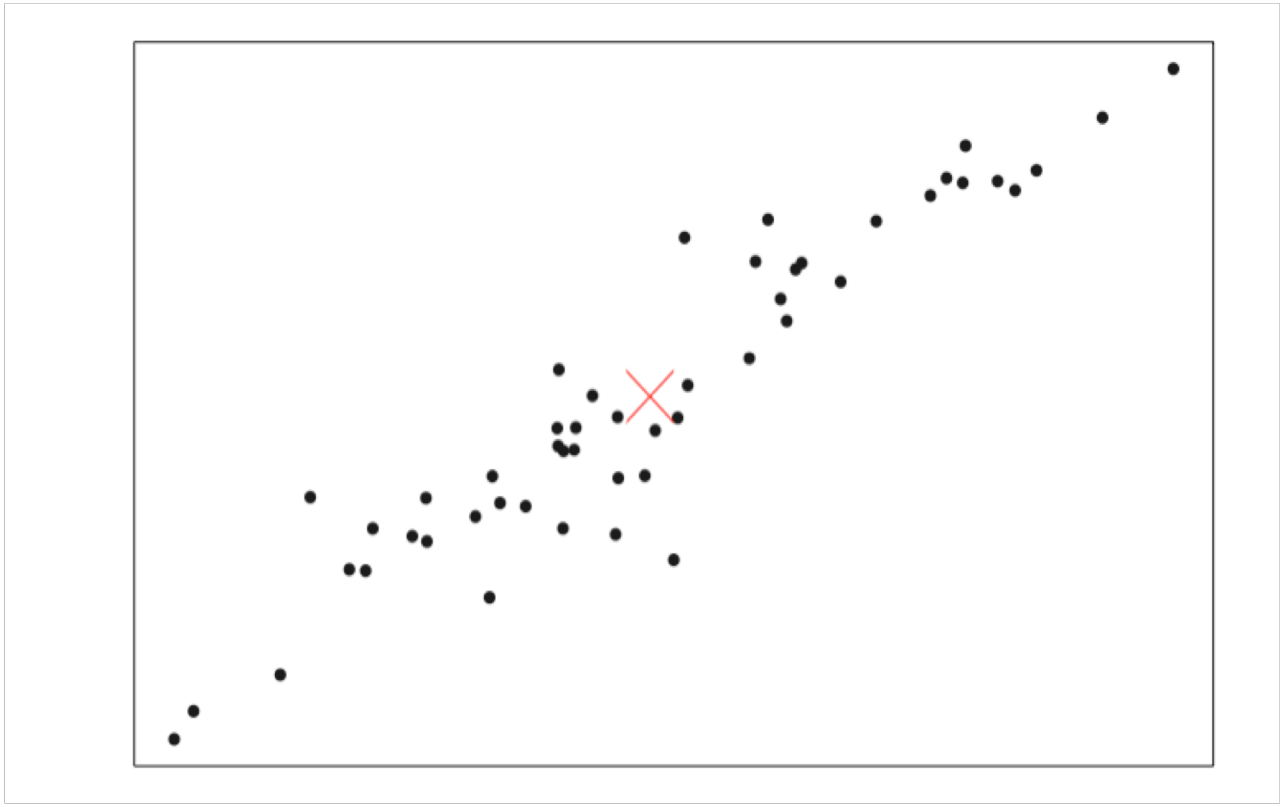
Then the maximum likelihood estimate is the same as minimizing sum of squared residuals

$$\sum_{i=0}^{n} (\hat{y}_i - y_i)^2$$

Fitting the line

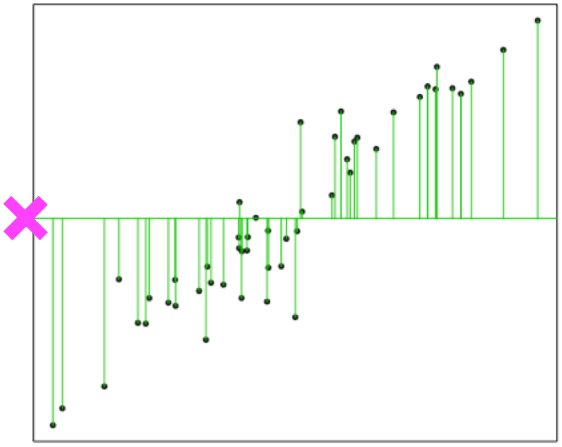
Line must go through point that = mean of X and mean of Y

$$(\bar{X}, \bar{Y})$$

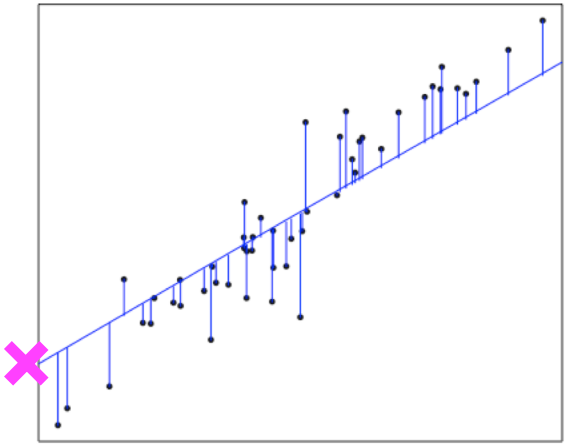


Fitting the line

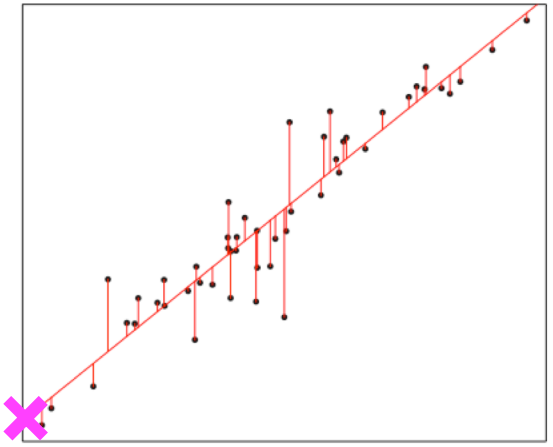
$\alpha = 10.19$
$\beta = 0$



$\alpha = 4$
$\beta = 0.6$

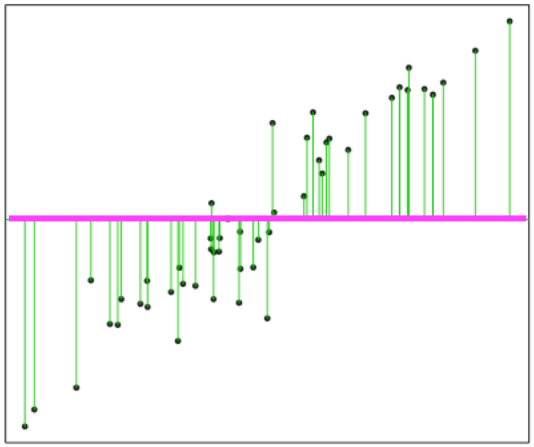


$\alpha = 1.57$
$\beta = 0.85$

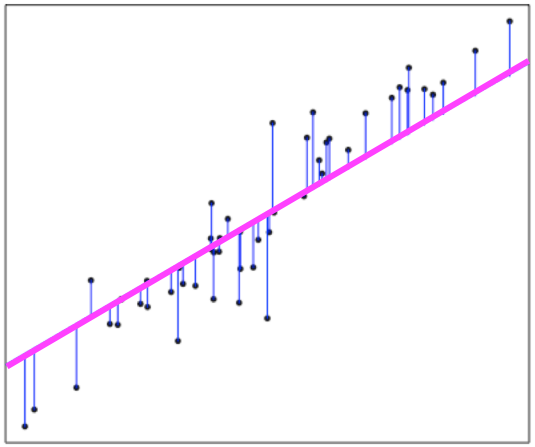


Fitting the line

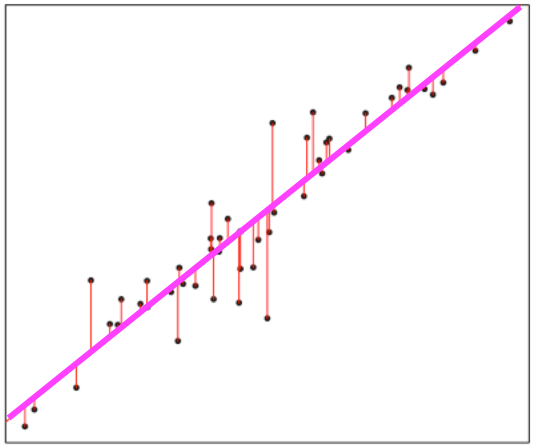
$\alpha = 10.19$
$\beta = 0$



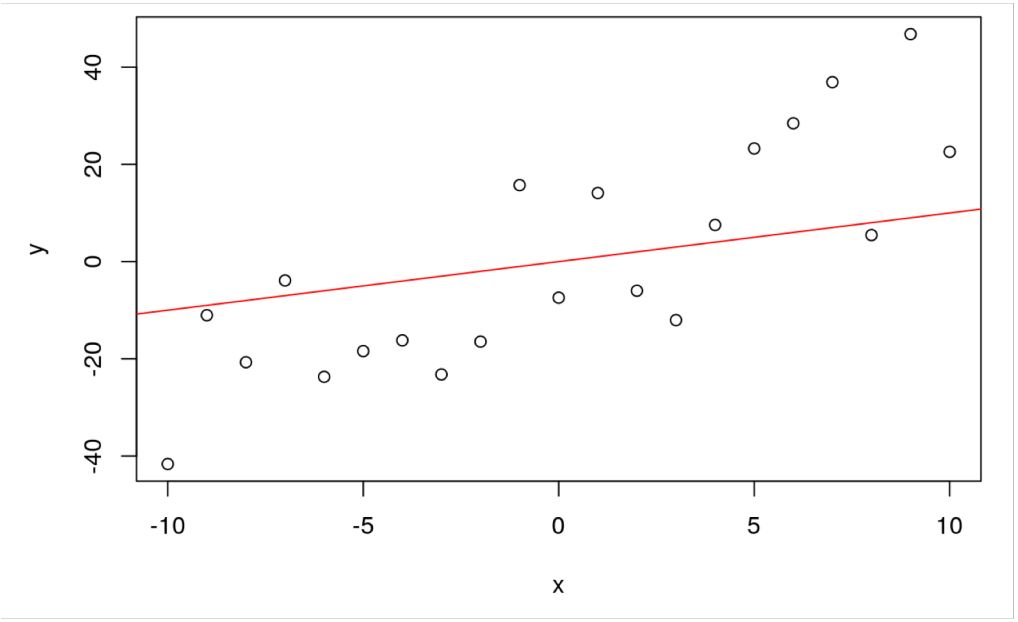
$\alpha = 4$
$\beta = 0.6$



$\alpha = 1.57$
$\beta = 0.85$

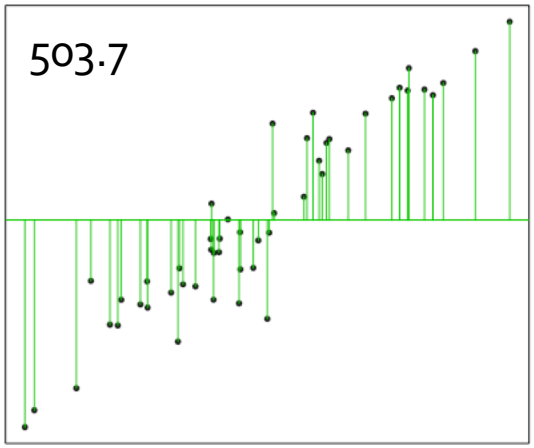


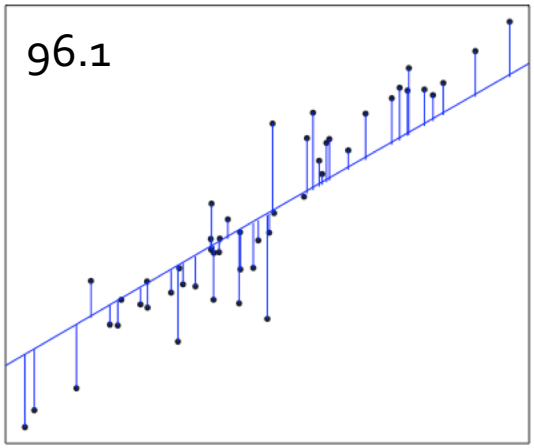
Exercise 3: Try fitting a line

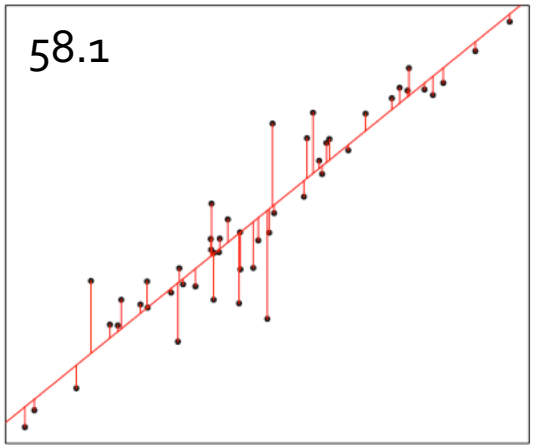


- Here is some simple data
- Both X and Y have been mean-centered (so both have a mean of 0)
- Your task is to try to find the best line (without cheating!)
- Here we have plotted a first guess at the line
- A script to help with the coding: https://www.math.ntnu.no/emner/ST2304/2019v/Week5/Script.R

Example







Fitting a linear regression in R

Fitting a linear regression in R

Use function lm() – *linear model* $Y_i = \alpha + \beta X_i + \varepsilon_i$

Fitting a linear regression in R

Arguments of lm():

lm(formula, data)

formula = Y ~ X
data = your data

Y is the response variable
X is the explanatory variable

Exercise 3: Try fitting a line…ctnd

- Use lm() to fit the actual regression line to the data from EX3 (remember to save as an object)
- Use coef(YourModelObject) to look at the estimates of the intercept and slope
- How do these compare to your estimates?
- What method does R use to estimate these parameters?
- Think about what these estimates **mean**

Lecture Summary

What are linear models?

What is linear regression?

How do we fit linear regressions?

Fitting in R

Lecture Summary

What are linear models?

Broad set of models that link a response variable to an explanatory variable with linear equations.

What is linear regression?

How do we fit linear regressions?

Fitting in R

Lecture Summary

What are linear models?

Broad set of models that link a response variable to an explanatory variable with linear equations.

What is linear regression?

A model that predicts values of a response variable from values of an explanatory variable. (lines)

How do we fit linear regressions?

Fitting in R

Lecture Summary

What are linear models?

Broad set of models that link a response variable to an explanatory variable with linear equations.

What is linear regression?

A model that predicts values of a response variable from values of an explanatory variable. (lines)

How do we fit linear regressions?

Maximise the likelihood = minimize the sum of squares

Fitting in R

Lecture Summary

What are linear models?

Broad set of models that link a response variable to an explanatory variable with linear equations.

What is linear regression?

A model that predicts values of a response variable from values of an explanatory variable. (lines)

How do we fit linear regressions?

Maximise the likelihood = minimize the sum of squares

Fitting in R

Using lm() and maximum likelihood estimation