BIO-5023YB
2020
Spring term – week 5
Model assumptions & Checking

Dr Philip Leftwich – p.leftwich@uea.ac.uk

Four main assumptions for linear regression:

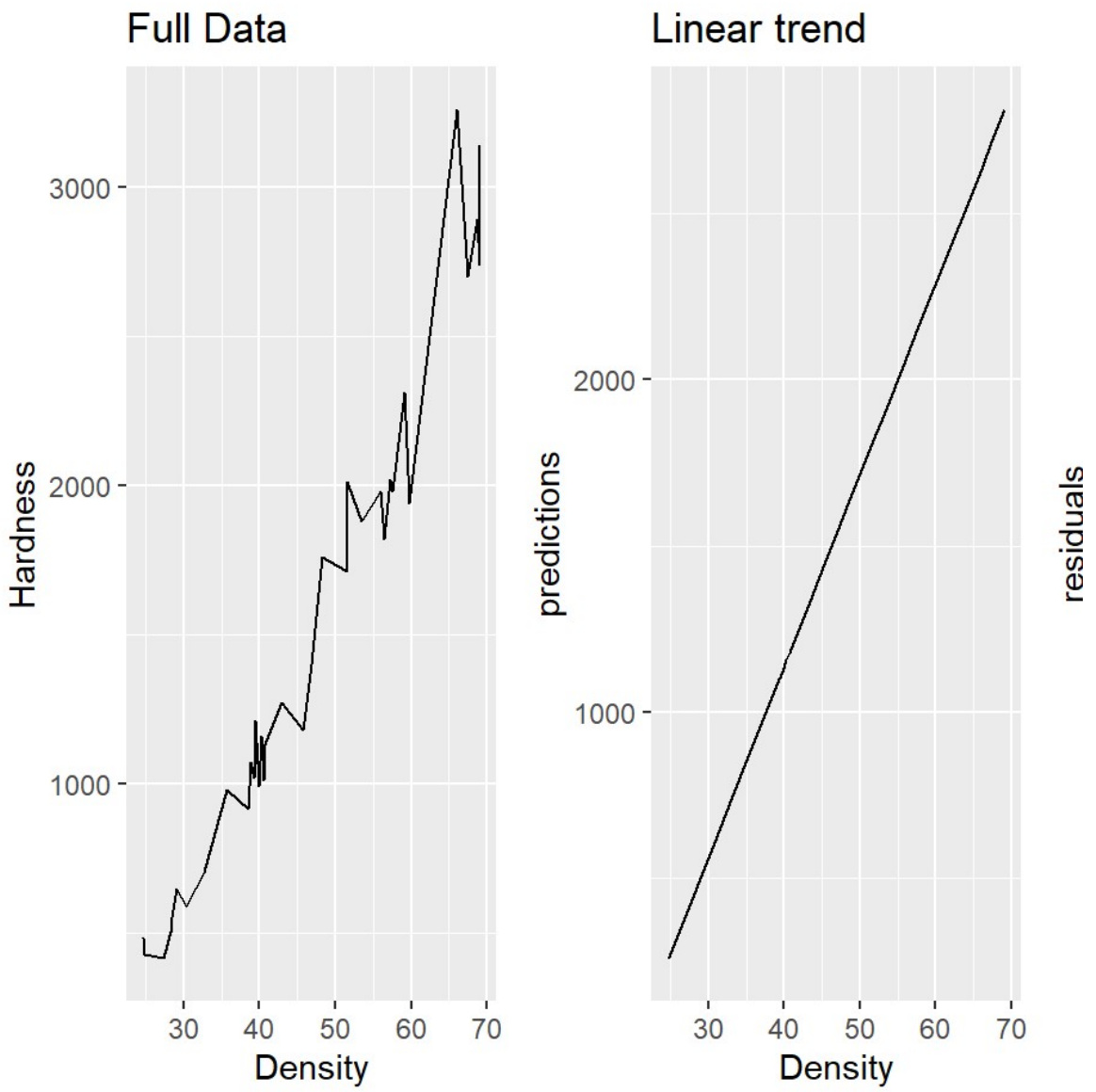
(i) linearity of the relationship between dependent and independent variables

(ii) independence of errors (no serial correlation)

(iii) homoscedasticity (constant variance) of the errors (or robust SE)

(iv) normality of the error distribution

(i)
linearity of the relationship between dependent and independent variables



Four main assumptions for linear regression:

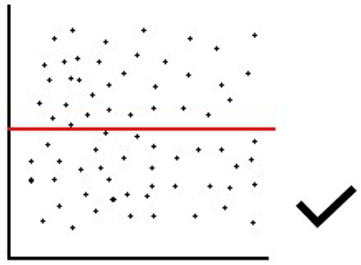
(i) linearity of the relationship between dependent and independent variables

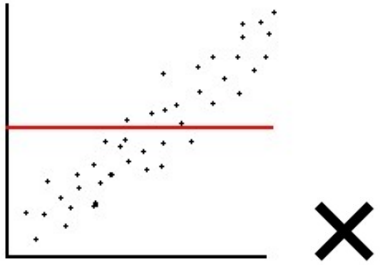
(ii) independence of errors (no serial correlation)

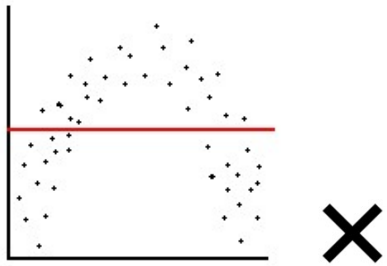
(iii) homoscedasticity (constant variance) of the errors (or robust SE)

(iv) normality of the error distribution

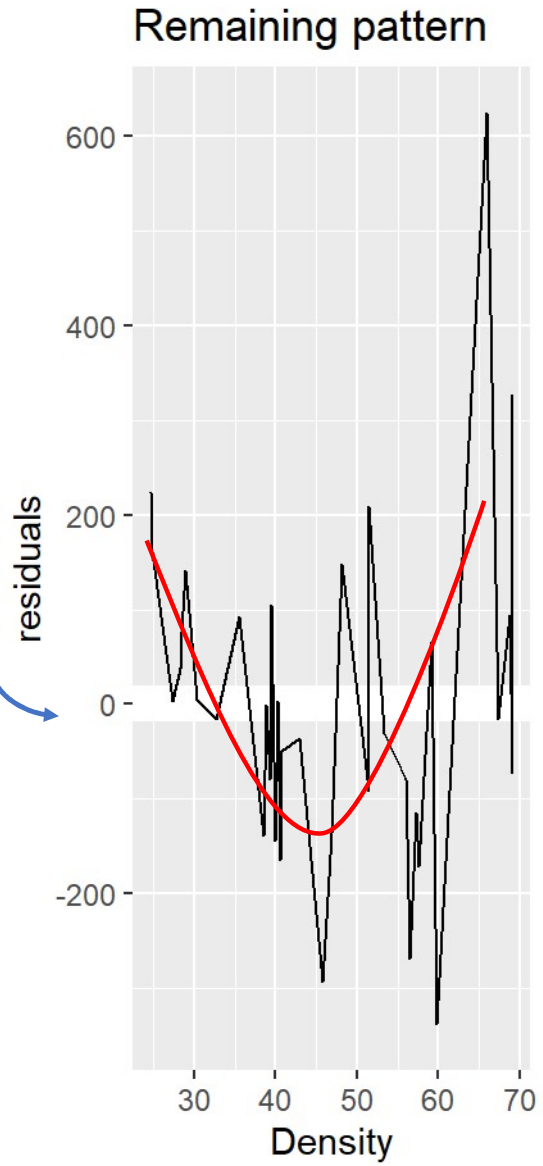
(ii) independence of errors
(no serial correlation)







Changing pattern of residuals with y = non-independence



Four main assumptions for linear regression:

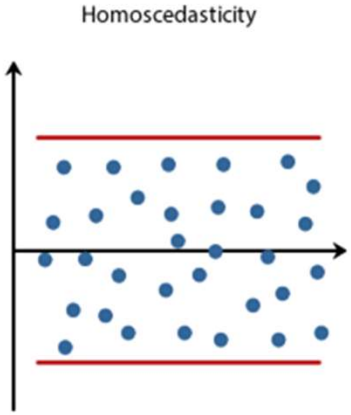
(i) linearity of the relationship between dependent and independent variables

(ii) independence of errors (no serial correlation)

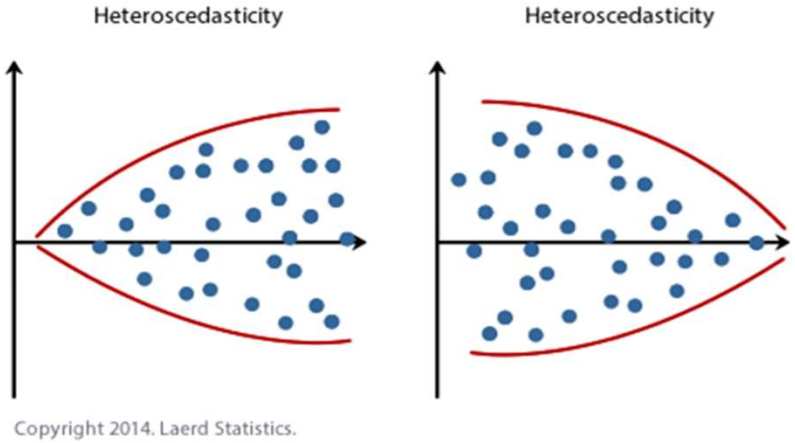
(iii) homoscedasticity (constant variance) of the errors (or robust SE)

(iv) normality of the error distribution

(iii) homoscedasticity (constant variance) of the errors (or robust SE)

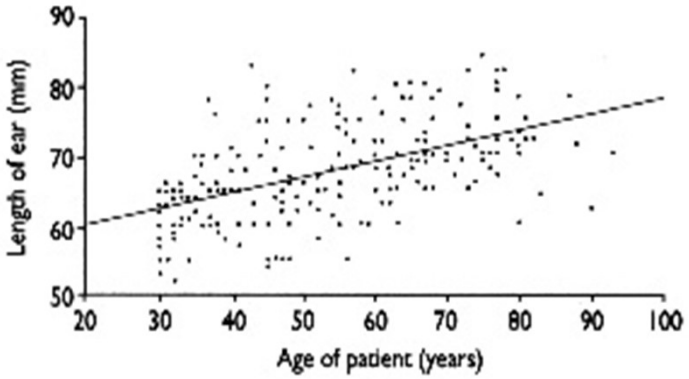


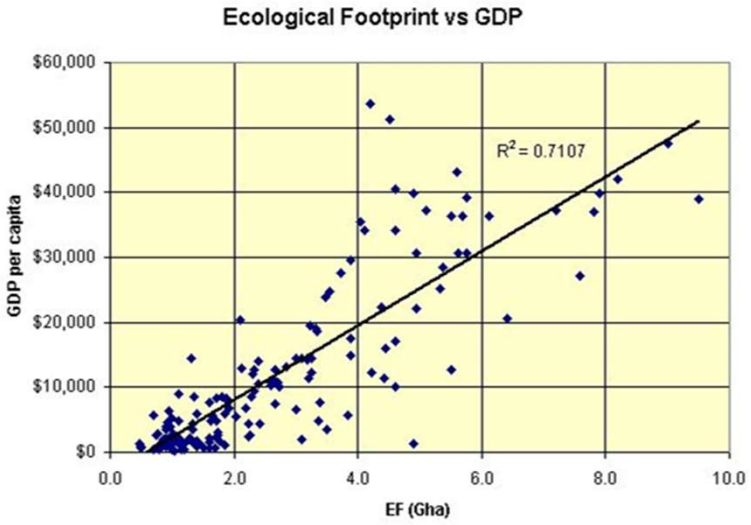
Homoscedastic: Residuals variance about the model remains relatively even over 'x'

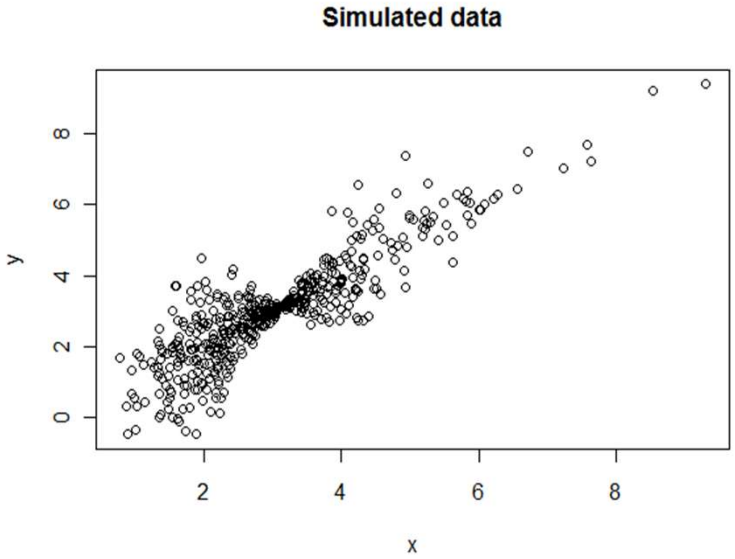


Heteroscedastic: Spread of residuals about the model changes (in a non-random way) over 'x'

Homoscedastic or Heteroscedastic?







Four main assumptions for linear regression:

(i) linearity of the relationship between dependent and independent variables

(ii) independence of errors (no serial correlation)

(iii) homoscedasticity (constant variance) of the errors (or robust SE)

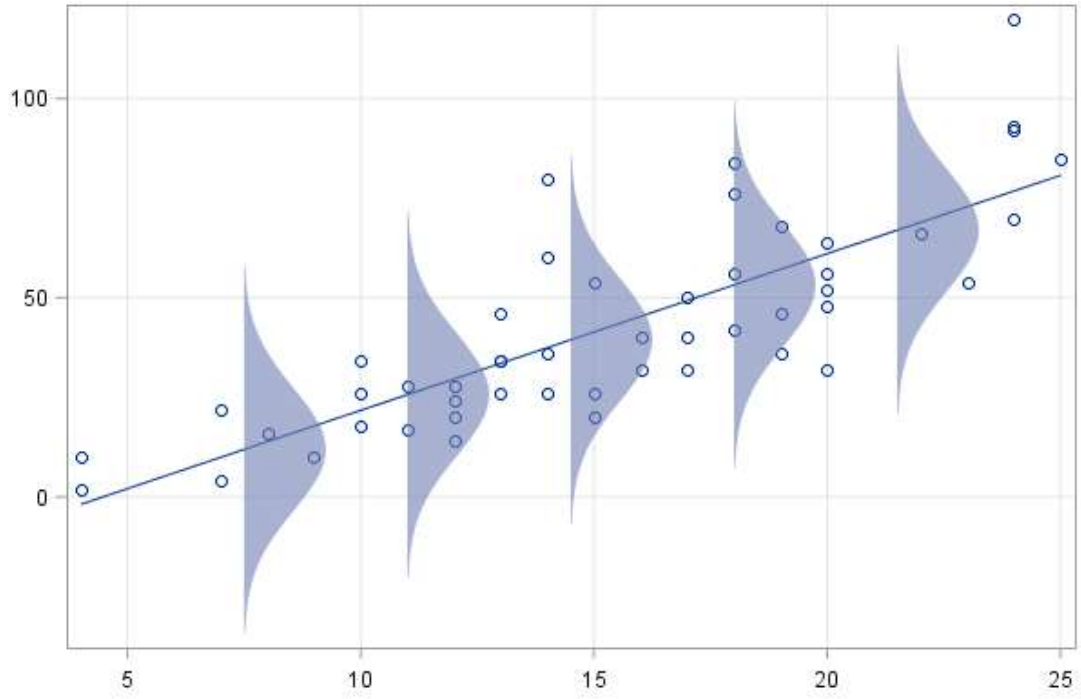
(iv) normality of the error distribution

(iv) normality of the error distribution

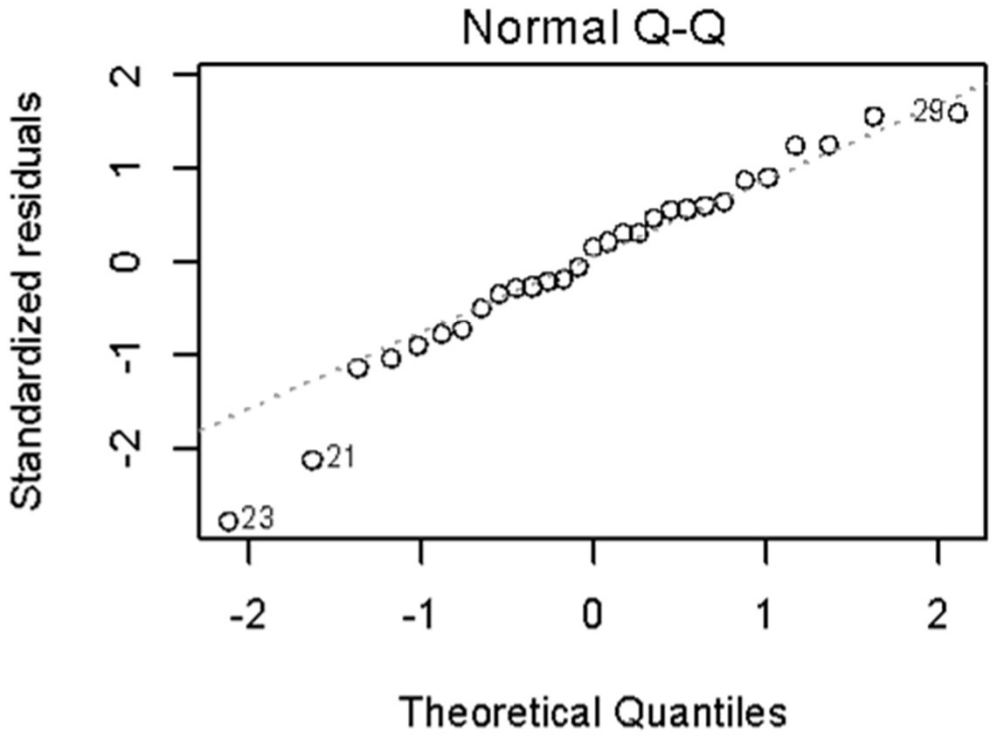
If data comes from a normal distribution then errors are likely to be normal too

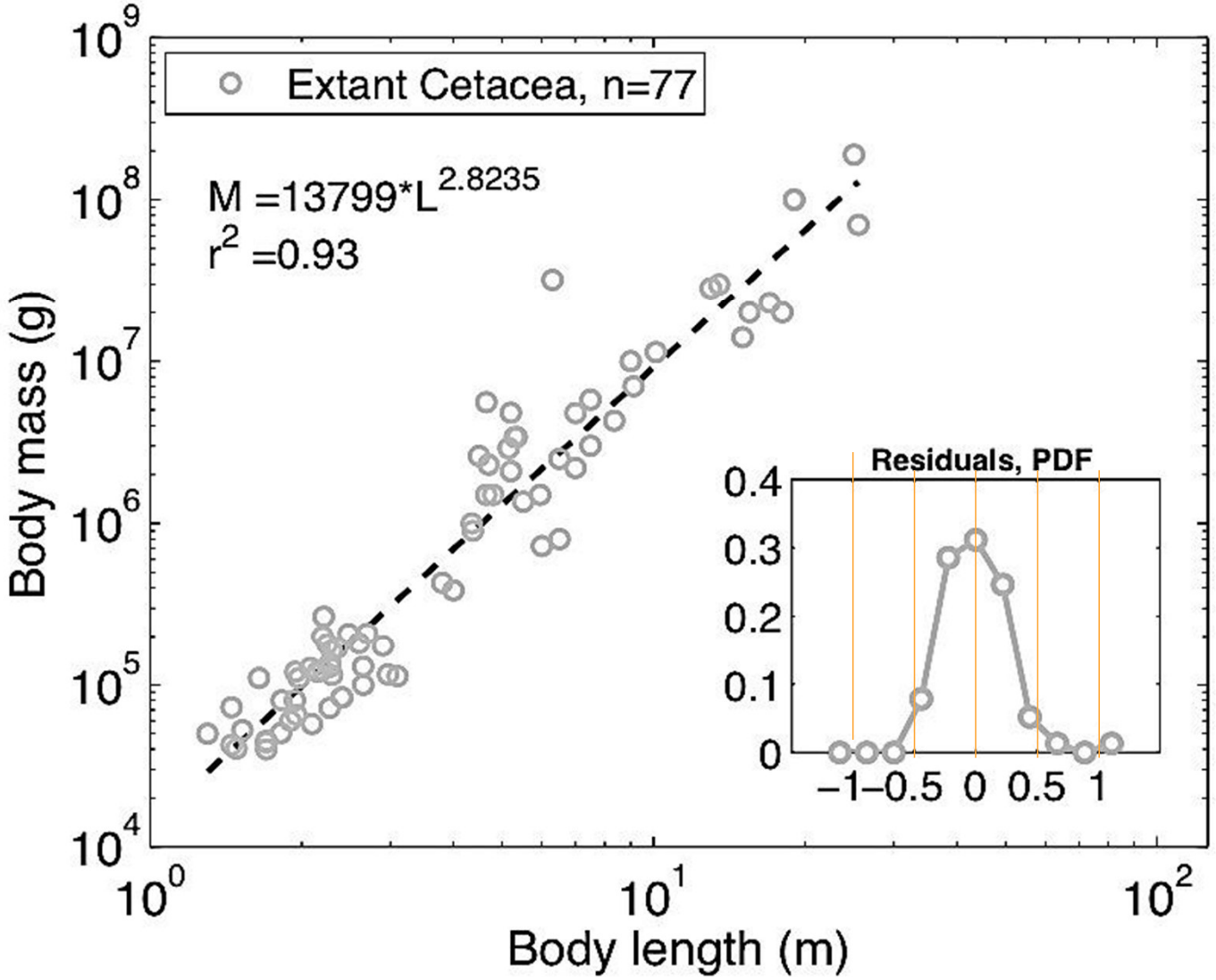
BUT it is the *errors* we care about rather than original data distribution.

These assumptions need to hold true in order for our estimates such as 95% CI of the mean difference to be accurate



QQ-plot: Are **residuals** normally distributed?




Extant Cetacea, n=77
M =13799*L^2.8235
r^2 =0.93
Residuals, PDF
Body mass (g)
Body length (m)
10^9
10^8
10^7
10^6
10^5
10^4
10^0
10^1
10^2
0.4
0.3
0.2
0.1
0
−1
−0.5
0
0.5
1

Outliers

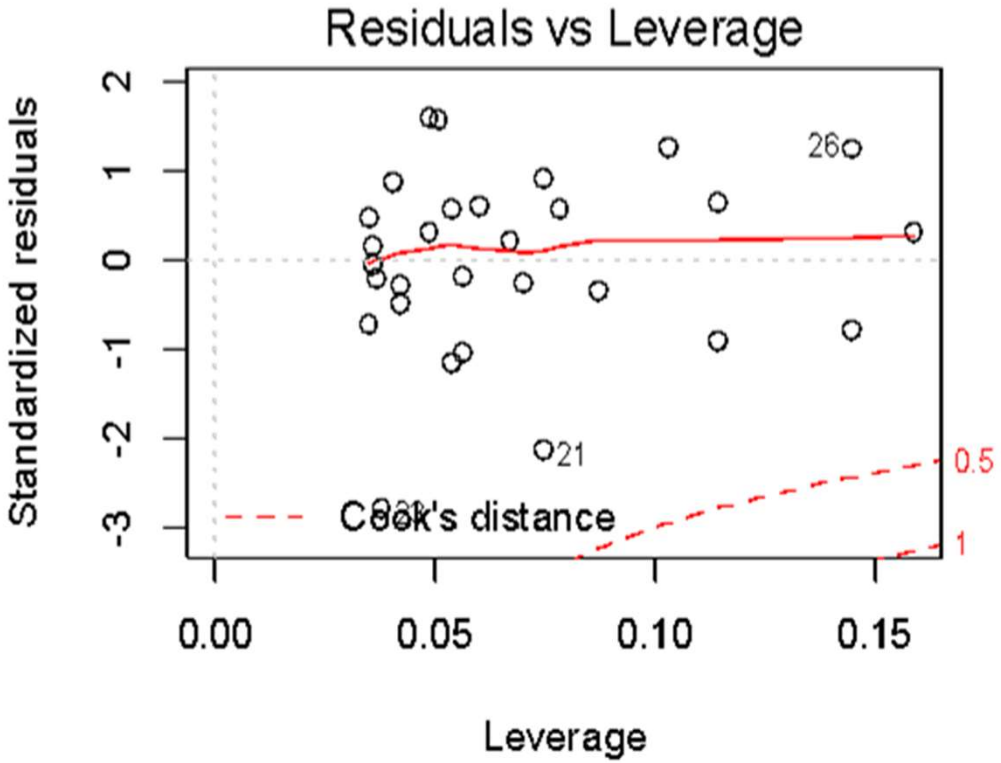
(i) linearity of the relationship between dependent and independent variables

(ii) independence of errors (no serial correlation)

(iii) homoscedasticity (constant variance) of the errors (or robust SE)

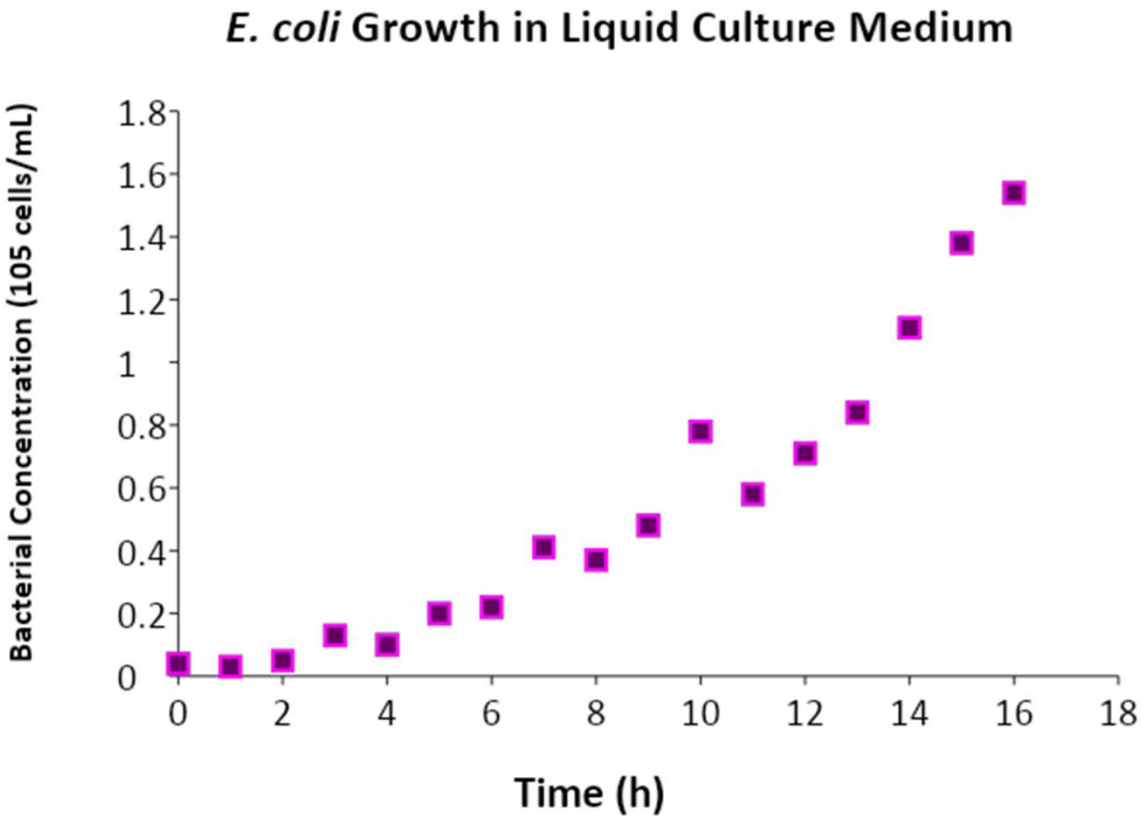
(iv) normality of the error distribution

Cook's Distance: Are Outliers Influencing Your Model?

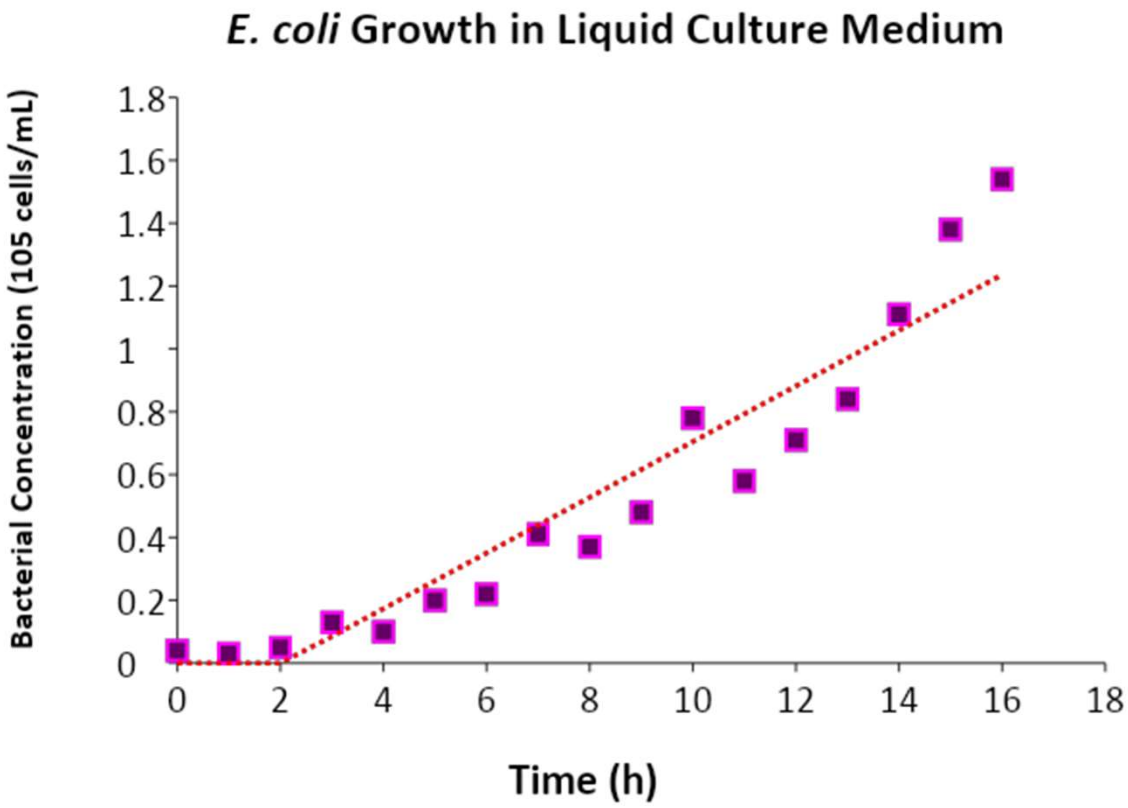


Cook's distance is a measure of how much influence a single observation has on your model. If points are ***labeled OR outside of the dashed red lines, they may be outliers that influence your model.***

What happens if we fit a linear model to some data that is clearly *not* linear…



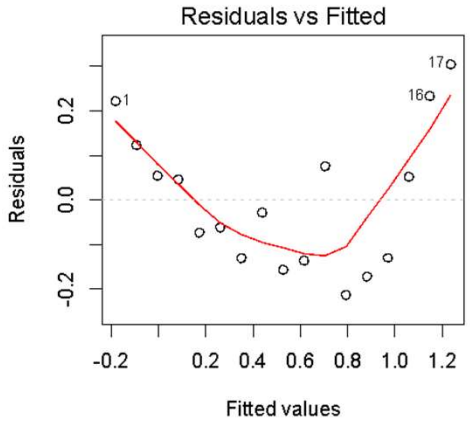
So we perform linear regression (bad idea)…

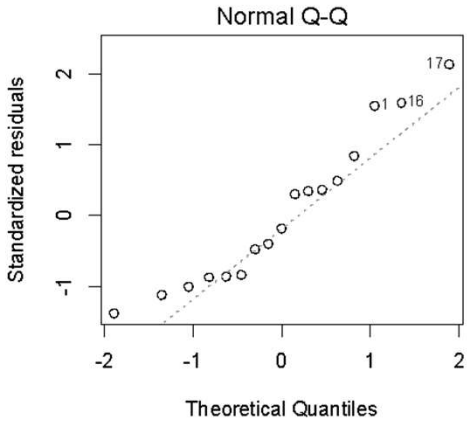


But how do we **know** it's a bad idea?

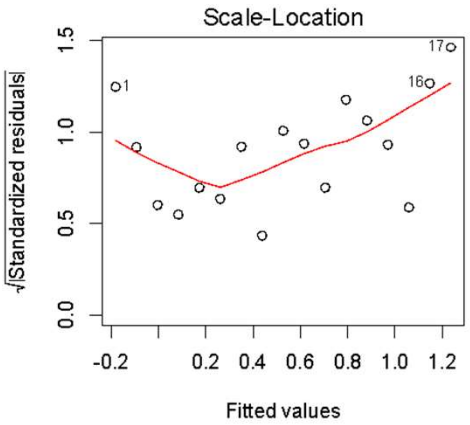
A. Because it's bacterial growth (exponential), and the data doesn't look linear.

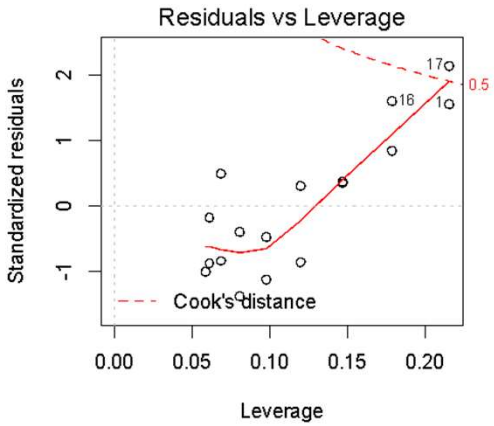
B. Because there is an obvious pattern to the residuals! Not distributed evenly or randomly around horizontal dashed line…





C. Because we might be concerned that the residuals are for the most part all above this gray dashed line…not normally distributed?





D. Because there may be **outliers** that are strongly influencing the linear model (which would make sense…because it's the wrong model to use!)

Next Time

- Testing Interactions in Linear Models
- Model Selection

This week's assignments

1) Finish any outstanding worksheets

2) **Fork** & Clone this week's GitHub Repo Philip-Leftwich/5023Y-Week5-Statistics

3) Open the Classroom Assignment for your Second Summative