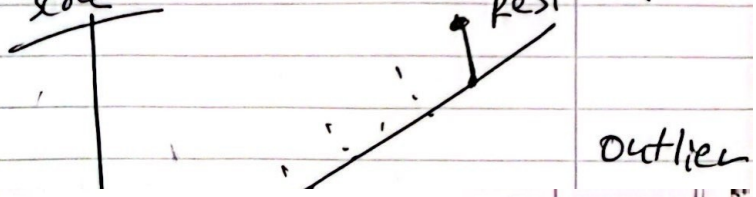
- How do I perform Bouman and Jacobsen's study on my data?
- Calculating the important and relevant values for my data (standard deviation etc) How does their regressions work?
- Have I understood the regressions correctly?
- Do my regressions work as they should?
- How do I test for autocorrelation and heteroscedasticity for the specific regressions?

Simple linear Regression

x Hour	Score y
1	8
2	12
3	18



Multiple Linear Regression

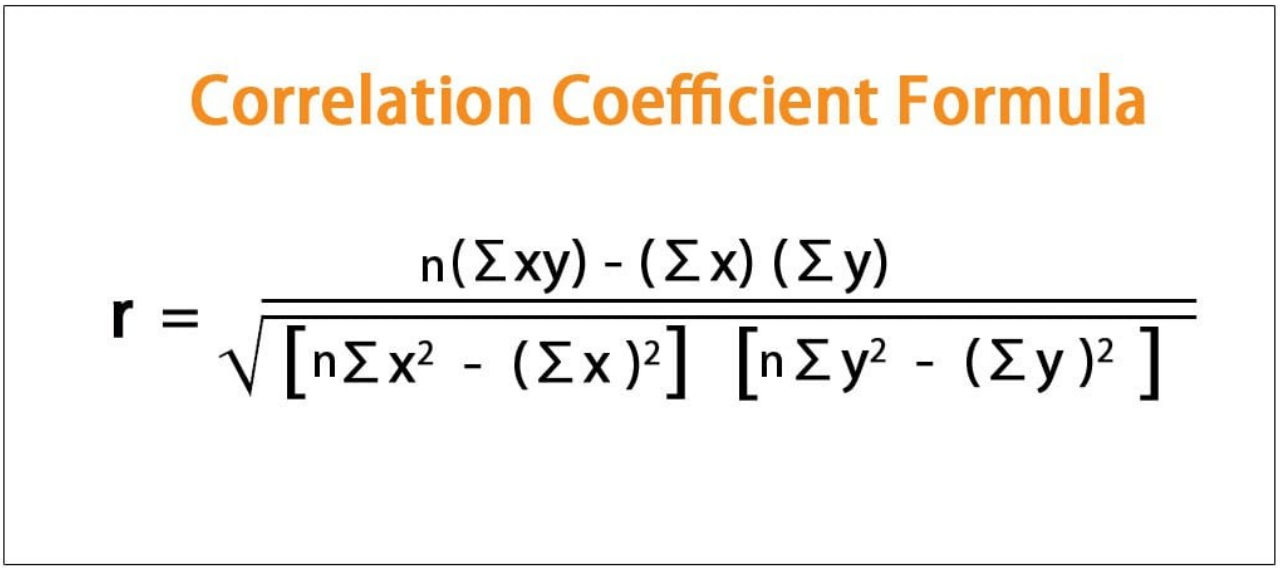
$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon$$

x_1 = Study Hour
x_2 = Course Teacher

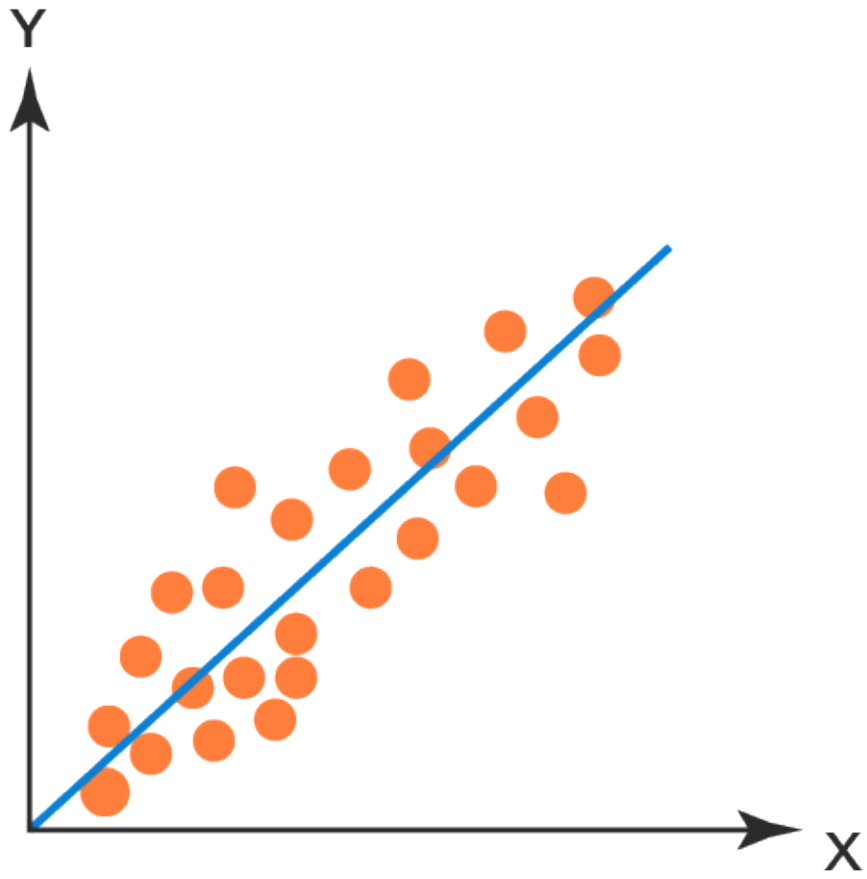
Y = 3 + 4.80x1

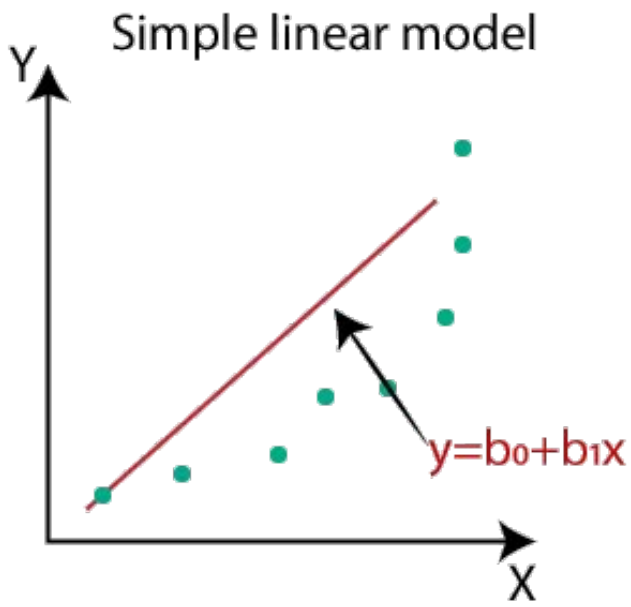
R Square = 0.7

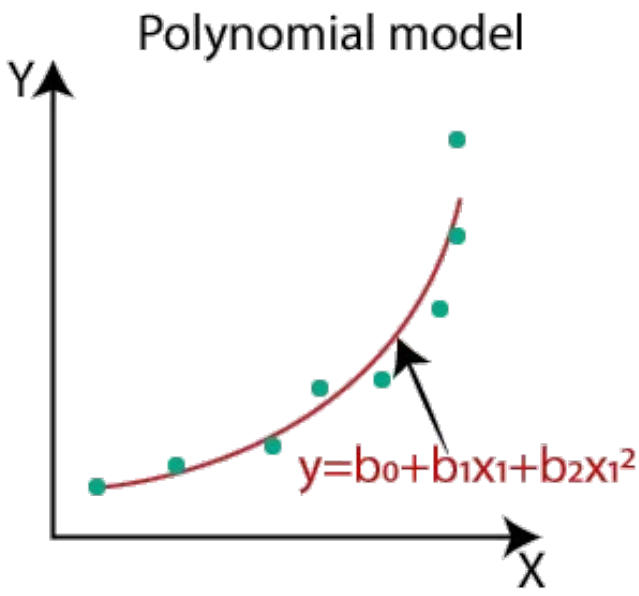
Adjusted R Square

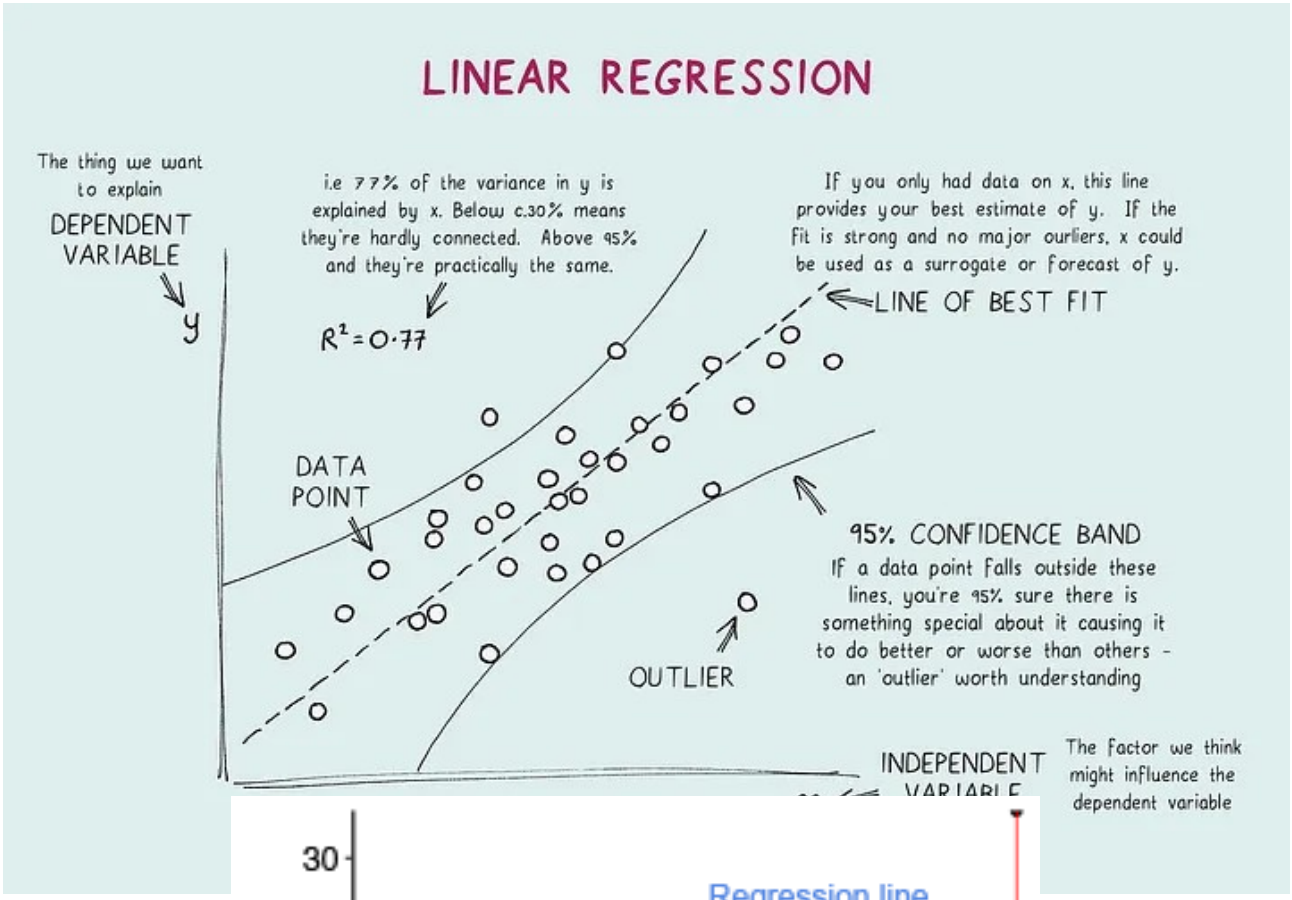


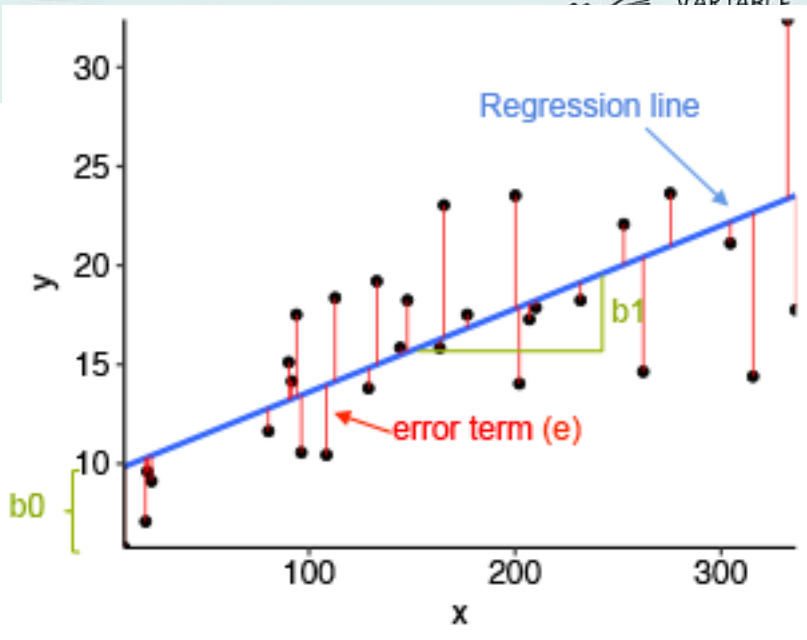
Regression Analysis Graph




Simple linear model
Y
X
y=b0+b1x


Polynomial model
Y
X
y=b0+b1x1+b2x1²


LINEAR REGRESSION
The thing we want to explain
DEPENDENT VARIABLE
y
i.e 77% of the variance in y is explained by x. Below c.30% means they're hardly connected. Above 95% and they're practically the same.
R² = 0·77
If you only had data on x, this line provides your best estimate of y. If the fit is strong and no major ourliers, x could be used as a surrogate or forecast of y.
LINE OF BEST FIT
DATA POINT
95% CONFIDENCE BAND
If a data point falls outside these lines, you're 95% sure there is something special about it causing it to do better or worse than others - an 'outlier' worth understanding
OUTLIER
INDEPENDENT
The factor we think might influence the dependent variable


Regression line
b1
error term (e)
b0
y
x
30
25
20
15
10
100
200
300

Multiple Linear Regression

House Rent = Interest Rate + Inflation + Location + Facilities/Amenities + Space size

Yearly Rice yield = Soil + Water + Fertilizer

Ice Cream Sales = Temperature + School Holidays + Daily Rainfalls + ? (Unexplained)

R Square = 0.61 it can explain 61 percent of the variation in size.

R-square value tells you how much variation is explained by your model.

Simple Linear Regression:

Test Score = $\beta 0 + \beta 1 * \text{Study Hours} + \varepsilon$

Multiple Linear Regression:

Test Score = $\beta 0 + \beta 1 * \text{Study Hours} + \beta 2 * \text{Prep Classes} + \beta 3 * \text{GPA} + \varepsilon$

Hypothesis

Null Hypothesis

- ✔ No correlation
- ✔ No Difference
- ✔ Proposed is not true
- ✔ Nothing new

Alternative Hypothesis:

- ✔ It has correlation
- ✔ Proposal is true
- ✔ New Findings is true

If p value is less than 0.05 we reject null hypothesis.

It is believed that a candy machine makes chocolate bars that are on average 5g. A worker claims that the machine after maintenance no longer makes 5g bars.

Ho: Ice Cream = 5gm // Nothing new
H1: Ice cream Not equal to 5gm // Might be greater then 5gm or less than 5 gm

In a class 10 students weight its average weight is 160 pound. We know aver weight is 170 pound.

Ho: weight is 170
H1: weight is not 170

There is no relationship between height and shoe size.

Durbin-Watson Test

H0 (null hypothesis): There is no correlation among the residuals.

HA (alternative hypothesis): The residuals are autocorrelated.