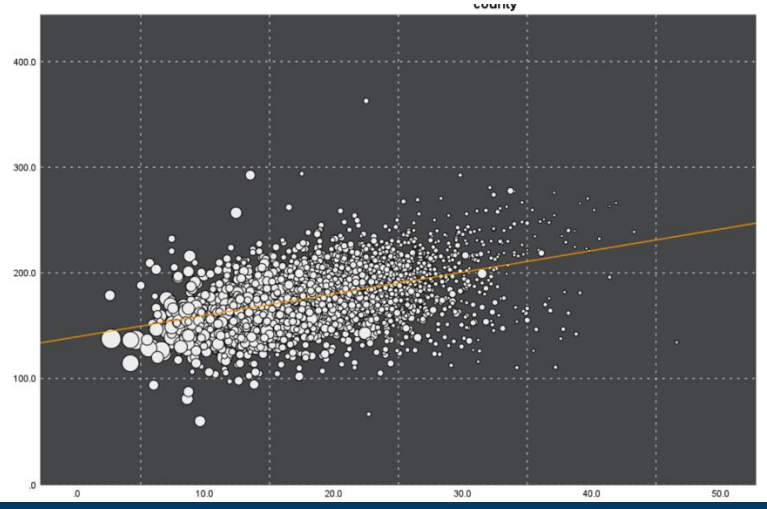
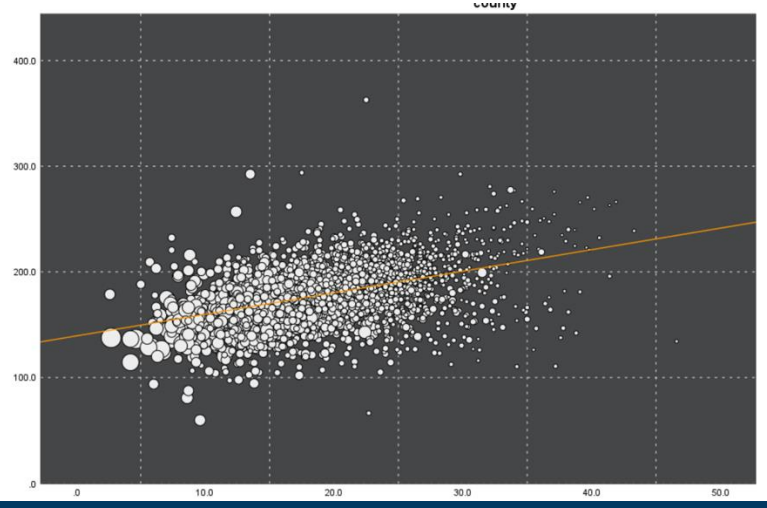


Getting started with regression techniques in SPSS

Jarlath Quinn

SMARTVISION Europe



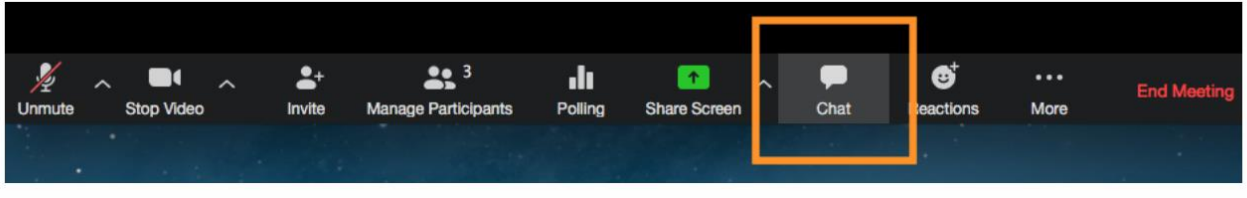
Just waiting for all attendees to join…

Getting started with regression techniques in SPSS

Jarlath Quinn

FAQ's

- Is this session being recorded? Yes
- Can I get a copy of the slides? Yes, we'll email links to download materials after the session has ended.
- Can we arrange a re-run for colleagues? Yes, just ask us.
- How can I ask questions? All lines are muted so please use the chat panel – if we run out of time we will follow up with you.



- Gold accredited partner to IBM, Predictive Solutions and DataRobot specialising in advanced analytics & big data technologies
- Work with open source technologies (R, Python, Spark etc.)
- Team each has 15 to 30 years of experience working in the advanced and predictive analytics industry

- Deep experience of applied advanced analytics applications across sectors
  - Retail
  - Gaming
  - Utilities
  - Insurance
  - Telecommunications
  - Media
  - FMCG

Agenda

- Overview of regression techniques and linear relationships
- Performing a Simple Linear Regression
- Using Multiple Linear Regression to make predictions
- Predicting response *probability* with Logistic Regression

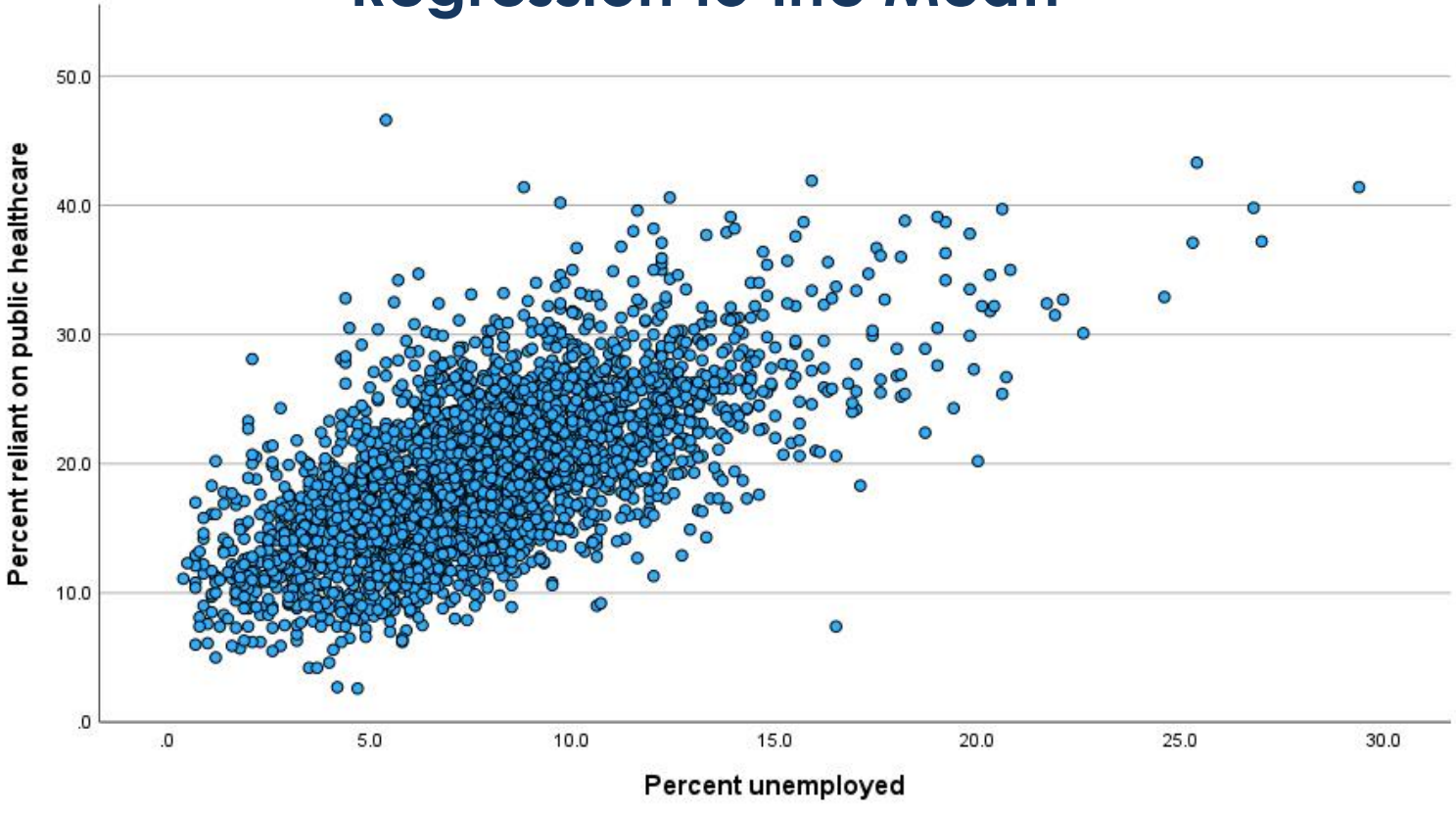
What do we mean by ‘Regression’?

- A family of statistical techniques used to predict outcomes and generate estimates for hundreds of applications
- Linear Regression is used
  - when the outcome is continuous (or scale) data
  - the relationships between the fields can be described using straight lines
- Quadratic Regression
  - Is a variant of Linear Regression when the outcome is continuous
  - the relationship with the dependent variable is curvilinear
- Logistic Regression is used
  - When the outcome consists or 2 (or more) categories
- Poisson regression
  - is commonly used when the dependent variable records counts of events

Where are Regression Techniques Used

- Modelling the relationship between promotion spend and revenue
- Estimating pollution levels following heavy rainfall
- Predicting tourism revenue based on exchange rates and air travel costs
- Predicting student test scores based on previous test results and peer-group performance
- Estimating website hits based on re-tweets and follower numbers
- Predicting sales of barbeques based on temperature forecasts

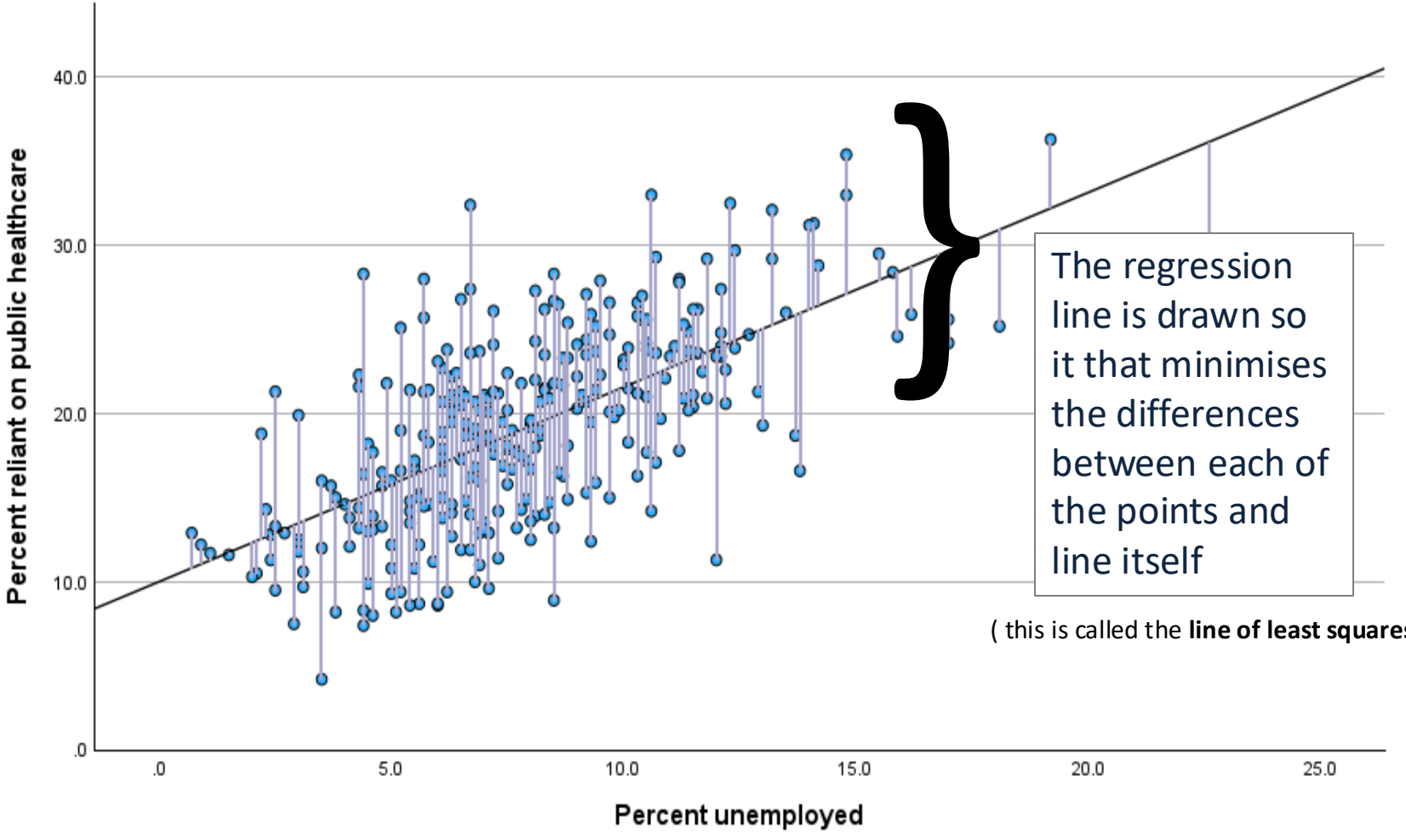
“Regression to the Mean”



SMARTVISION Europe

A SELECT INTERNATIONAL COMPANY

"Regression to the Mean"

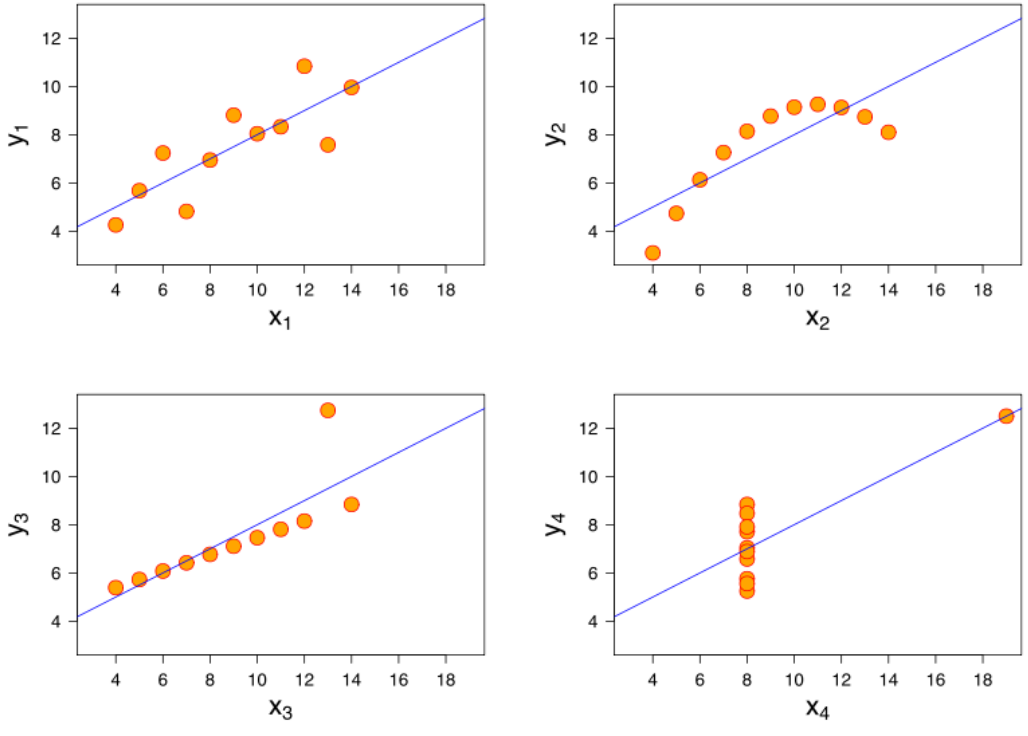


The regression line is drawn so it that minimises the differences between each of the points and line itself

(this is called the **line of least squares**)

Regression to the Mean

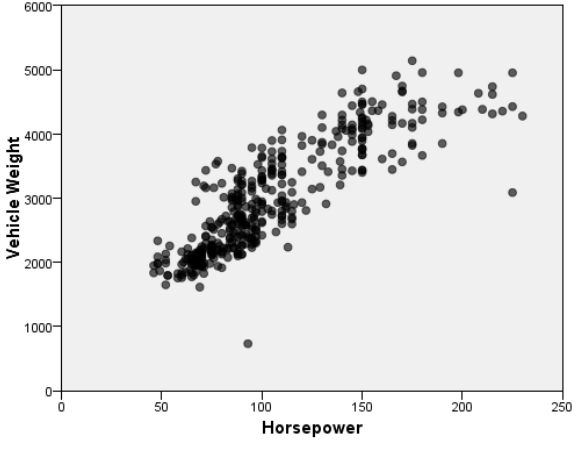
*

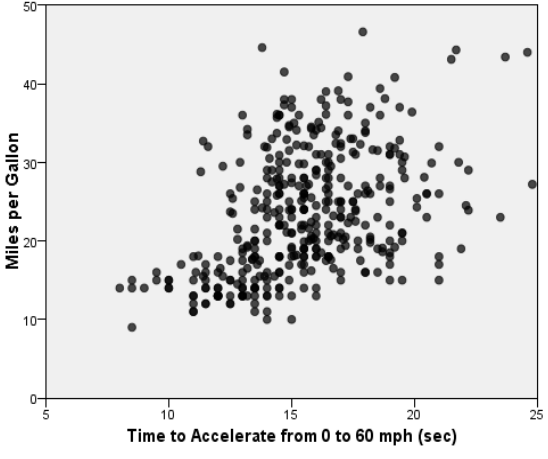


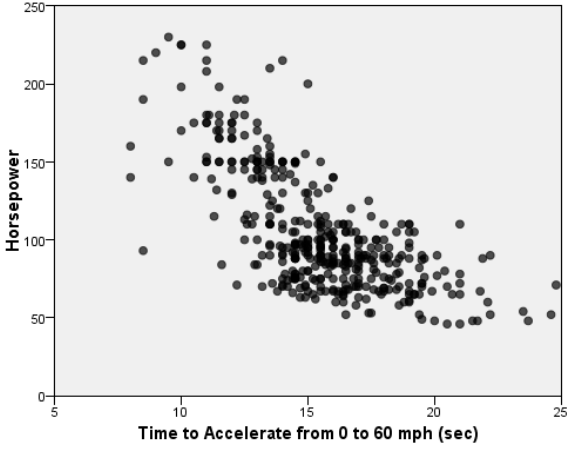
- But be careful…
- It is just an average after all…

**Anscombe's Quartet*

Measuring Linear Relationships







0.859

0.434

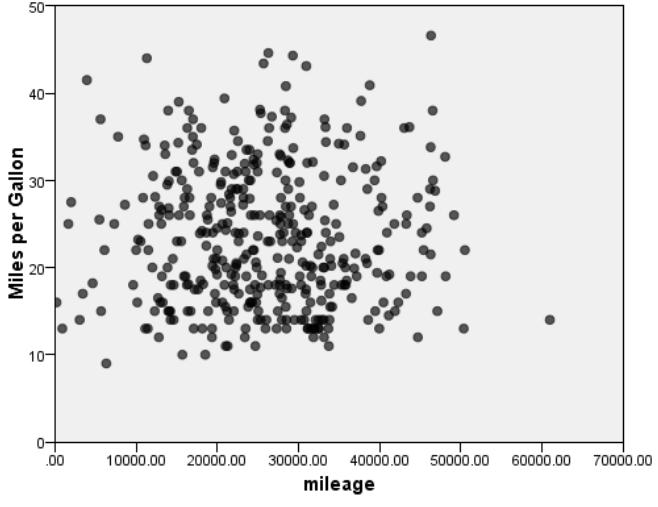
-.701

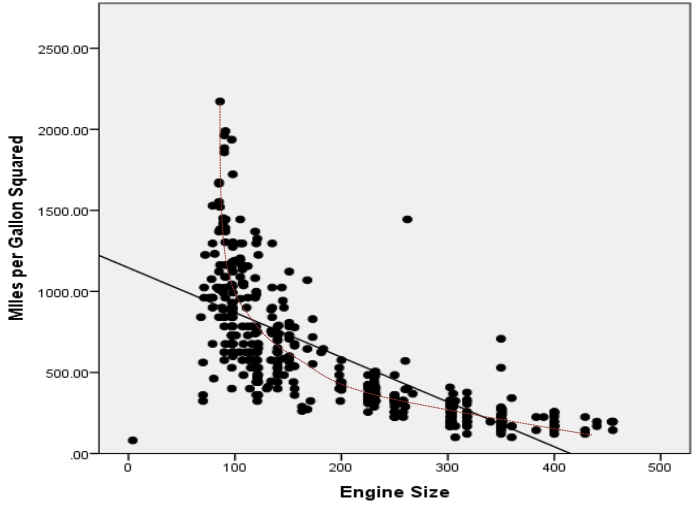
Pearson Correlation Values

SMARTVISION Europe

A SELECT INTERNATIONAL COMPANY

Non-Linear Relationships





-0.005

-.671

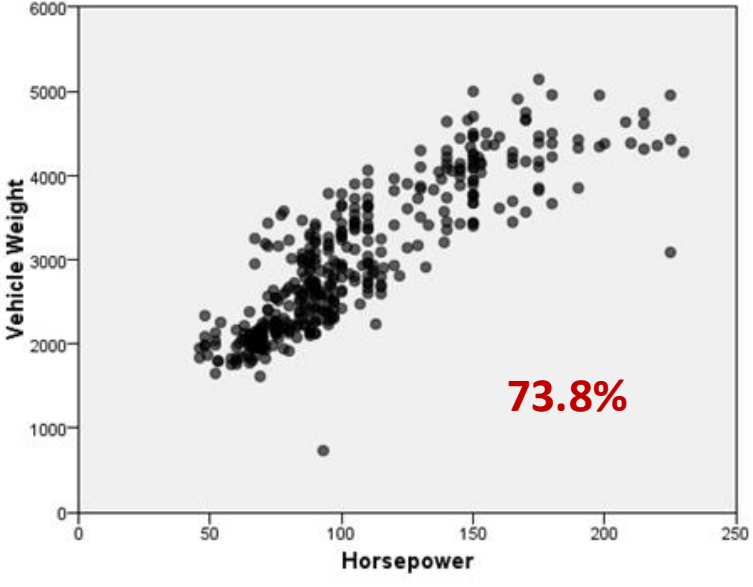
Pearson Correlation Values

SMARTVISION Europe

A SELECT INTERNATIONAL COMPANY

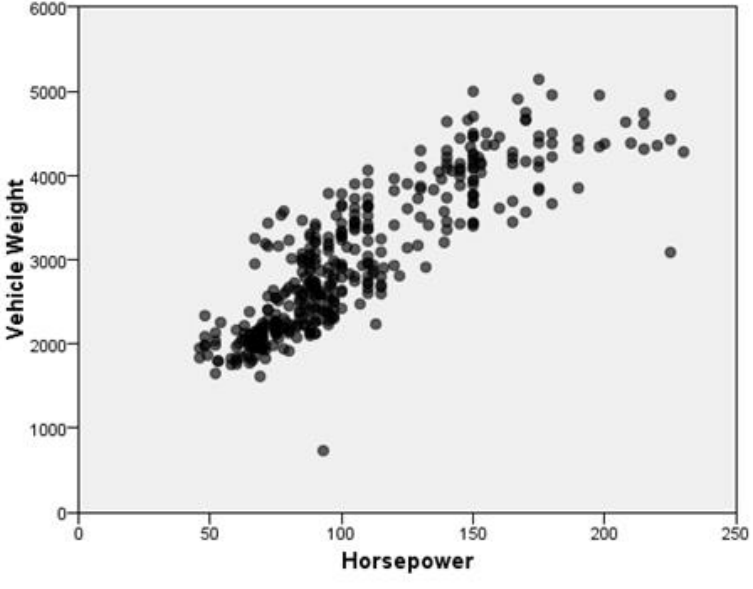
Correlations as Percentages

- Correlation = 0.859
- 0.859 x 0.859 = 0.738
- 0.738 = 73.8 %
- Correlation Squared = ‘R Square’



From Correlation to Prediction

How can we express linear relationships as predictive models?



SMARTVISION
Europe

How long does it take to cook a chicken?

How long does it take to cook a chicken?

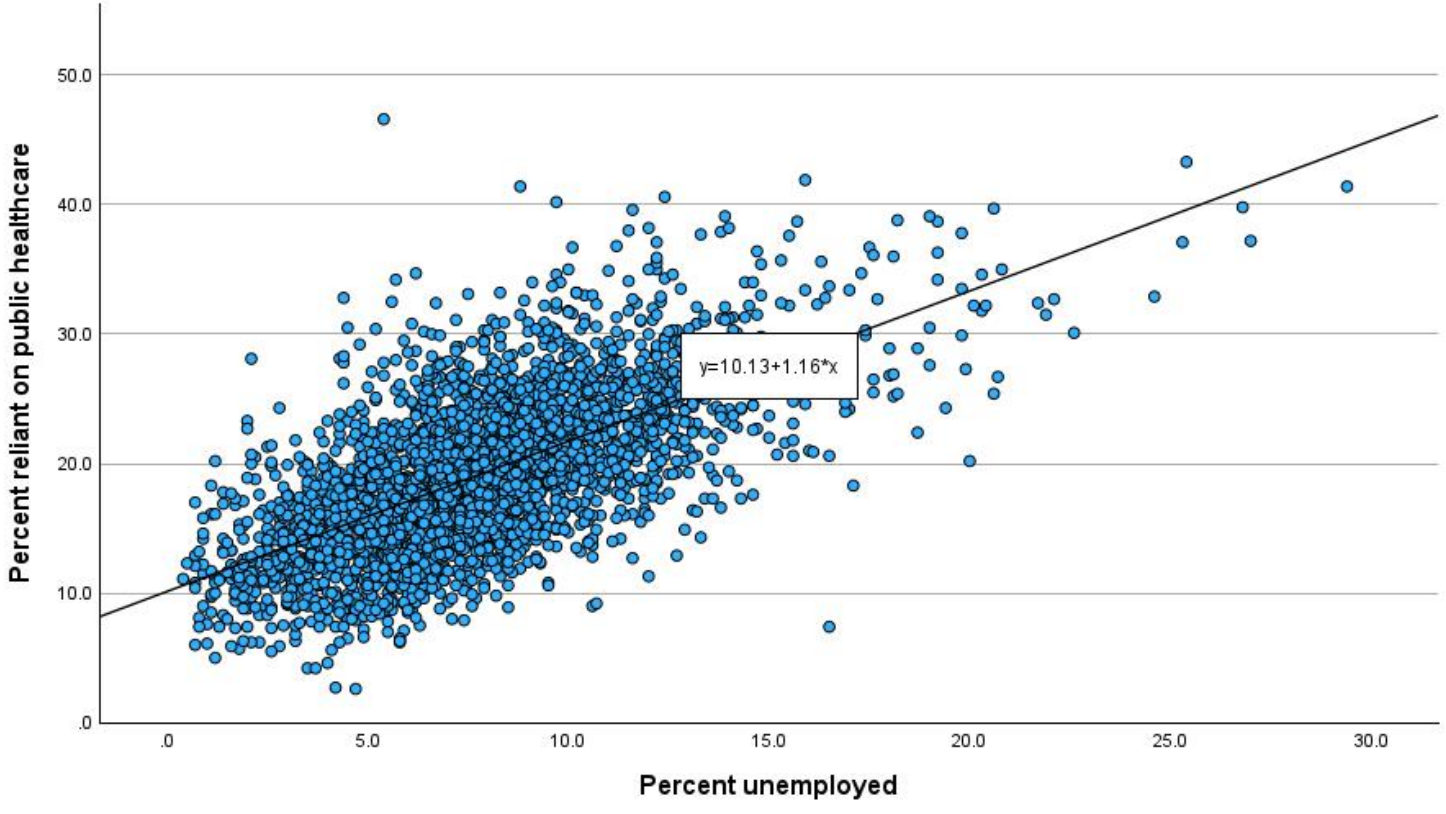
- 7 minutes per pound plus 45 minutes

$$y = mx + c$$

or

- 20 minutes per pound plus 20 minutes

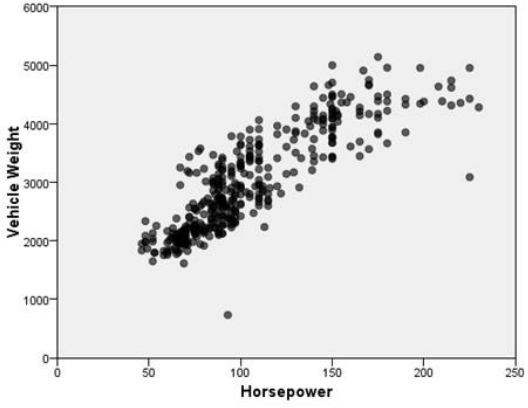
$$y = a + bx$$



SMARTVISION Europe

A SELECT INTERNATIONAL COMPANY

SMARTVISION Europe



Lets look at a demo of Linear Regression in IBM SPSS Statistics

SMARTVISION
Europe

How can we predict category outcomes?

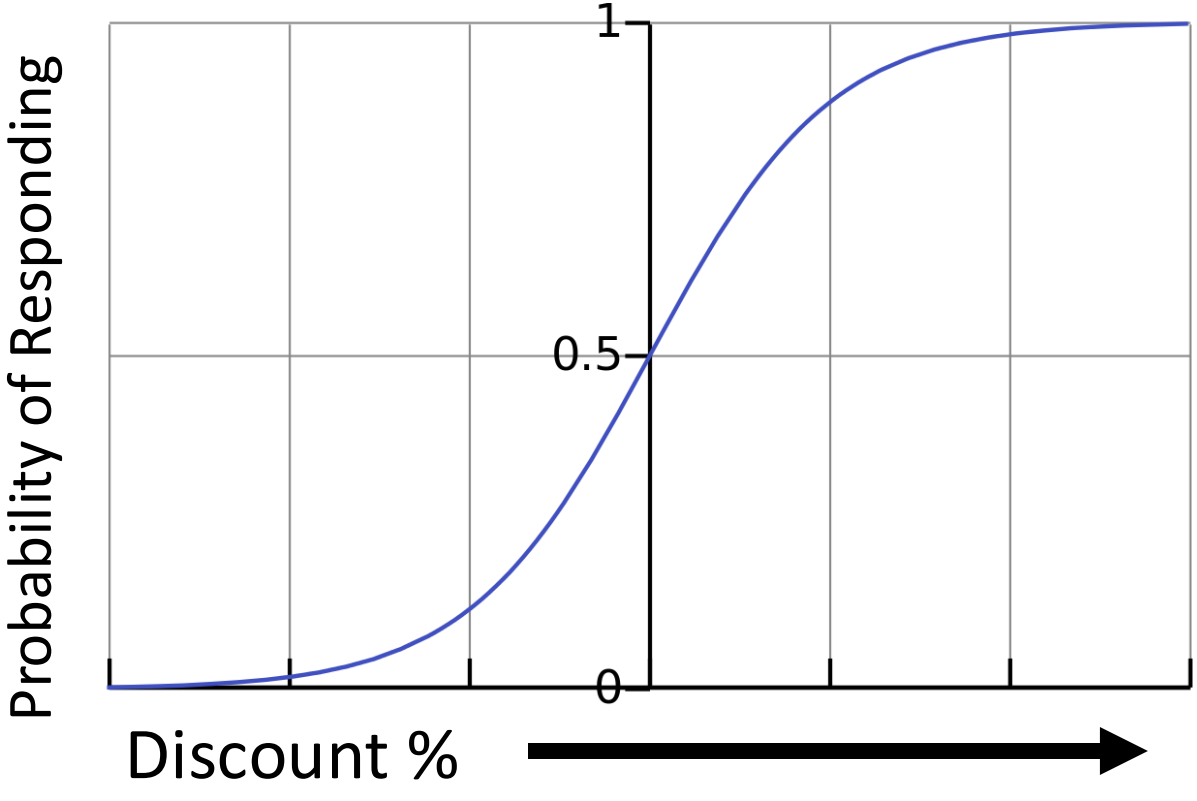
Logistic Regression

- Allows us to predict things that linear regression can't
- Such as…
  - Response to a marketing campaign
  - Credit risk
  - Whether a subscriber is likely to renew a service
  - Risk of equipment failure
  - How likely is it that a particular patient will be readmitted to hospital
  - Whether a charity donor will switch to Direct Debit

Logistic Regression

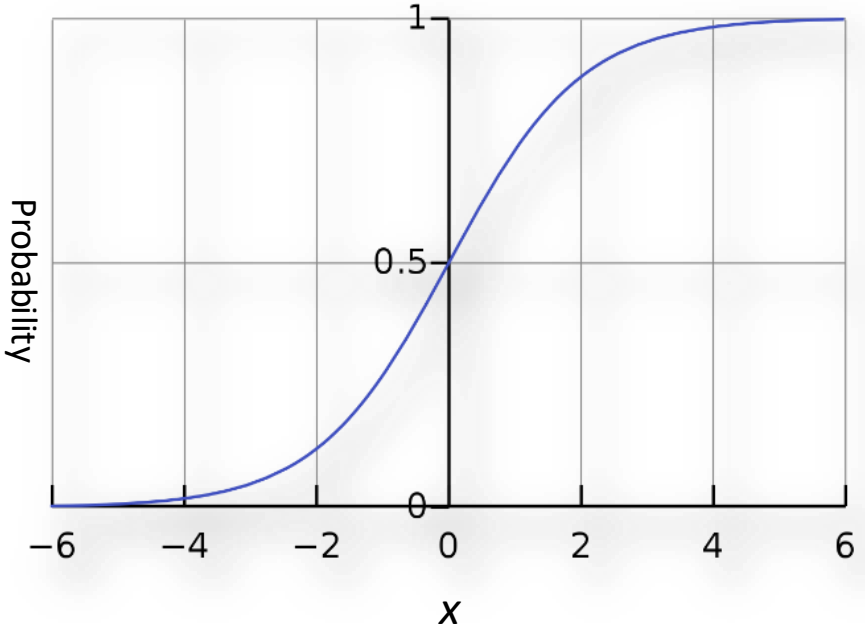
- But….
- These outcomes are not *continuous numbers* so standard linear regression won't work
- When the outcome consists of two categories we use *Binary* Logistic Regression
- When the outcome has three or more categories we use *Multinomial* Logistic Regression
- Logistic gets around the limitations of describing relationships with straight lines by using a special *sigmoid* curve

Logistic Regression



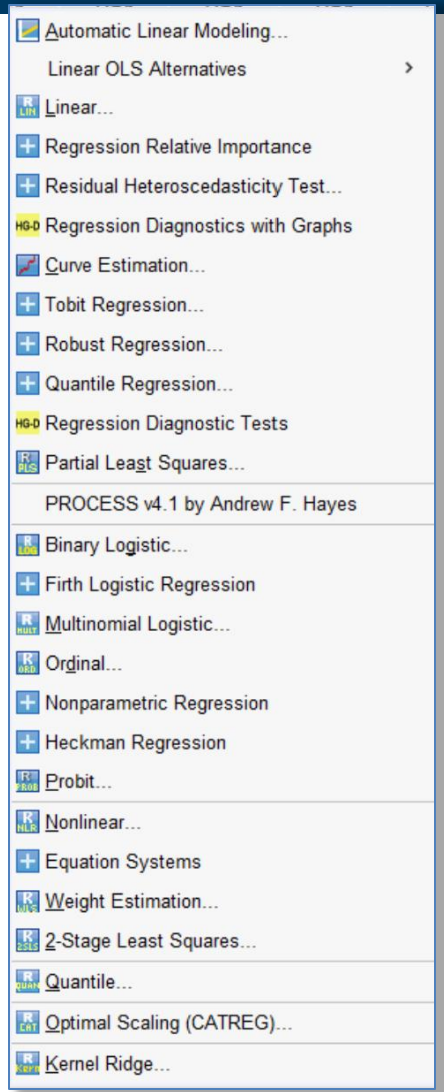
Logistic Regression

- There is a special formula that converts the values of the predictor coefficients on the x axis to the values on predicted probabilities on the y-axis
- But what *are* these numbers on the x-axis?



IBM SPSS Regression Methods

- SPSS Statistics has a **lot** of regression methods
- The **Regression Models** module adds several key methods like **Logistic Regression**
- Many additional methods are made available through SPSS' integration with **R** and **Python**



SMARTVISION Europe

Additional Resources

- How to model non-linear relationships
- Introduction to Moderation Analysis
- Introduction to Mediation Analysis

- Check what version / modules of SPSS you have installed
- See exactly what is included in the Regression Module

- Choosing the correct statistical test
- How to interpret significance tests
- Eat your greens blog series on statistical testing and procedures

Smart Vision provides a portfolio of online training materials **free** to existing customers or available for purchase

Factor and Cluster Analysis with IBM SPSS Statistics

£75.00
Jarlath Quinn

Introduction to Time Series Forecasting with IBM SPSS Statistics

£75.00
Jarlath Quinn

Understanding and applying logistic regression techniques in SPSS Statistics

£75.00
Jarlath Quinn

Understanding and Applying Linear Regression Techniques in SPSS Statistics

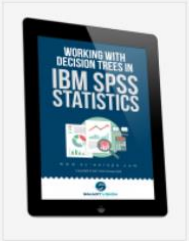
£75.00
Jarlath Quinn

Building predictive models in SPSS Modeler

£75.00
Jarlath Quinn

Statistical and significance testing in SPSS Statistics

£75.00
Jarlath Quinn



Working with decision trees in SPSS Statistics

Introduction to SPSS Modeler course

Introduction to IBM SPSS Statistics course

SMARTVISION
Europe

Working with Smart Vision Europe

We can help with processing and analysing your data.

- Self-paced, virtual and in-person training courses in how to use SPSS products and appropriate statistical techniques
- A mix of consultancy and training whereby we do the initial work and then teach you how to replicate it

Working with Smart Vision Europe Ltd.

- **Sourcing Software**
  - You can buy your analytical software from us often with discounts
  - Assist with selection, pilot, implementation & support of analytical tools
  - http://www.sv-europe.com/buy-spss-online/
- **Training and Consulting Services**
  - Guided consulting & training to develop in house skills
  - Delivery of classroom training courses / side by side training support
  - Identification & recruitment of analytical skills into your organisation
- **Advice and Support**
  - offer 'no strings attached' technical and business advice relating to analytical activities
  - Technical support services

SMARTVISION Europe

Contact us:

+44 (0)207 786 3568
info@sv-europe.com
Twitter: @sveurope
Follow us on Linked In
Sign up for our Newsletter

Thank you