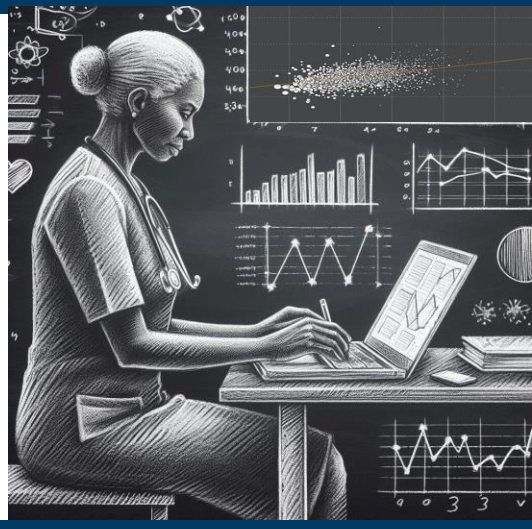
SMARTVISION Europe

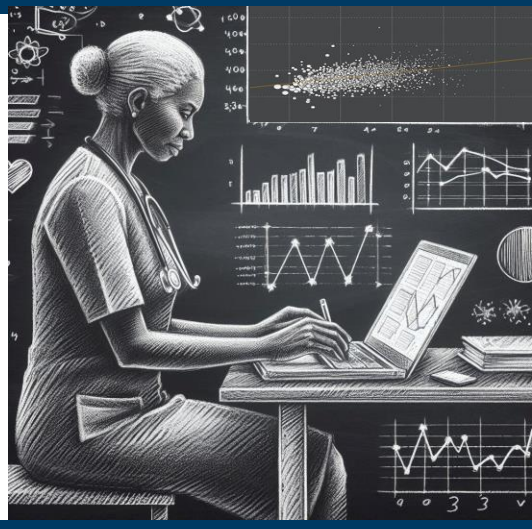


Regression Techniques for Healthcare Applications

Jarlath Quinn

SMARTVISION Europe

Just waiting for all attendees to join…

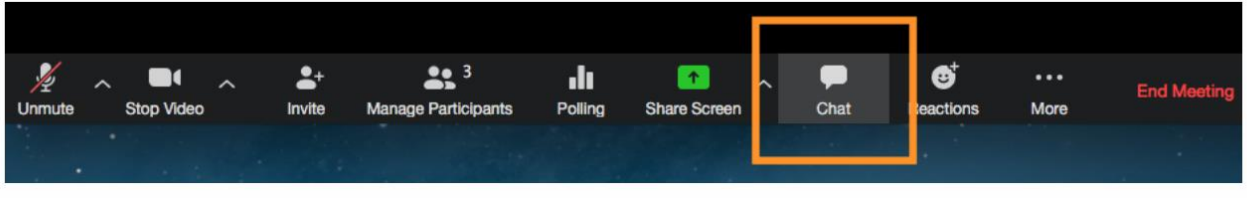


Regression Techniques for Healthcare Applications

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
  - Healthcare
  - Utilities
  - Insurance
  - Telecommunications
  - Housing
  - FMCG

Agenda

- Exploring relationships between scale variables with correlations
- Introducing Simple Linear Regression
- Using Multiple Linear Regression to model regional per-capita cancer mortality rates
- Identifying problems in Linear Regression models
- Introducing Logistic Regression
- Working with Logistic Regression to model low birthweight

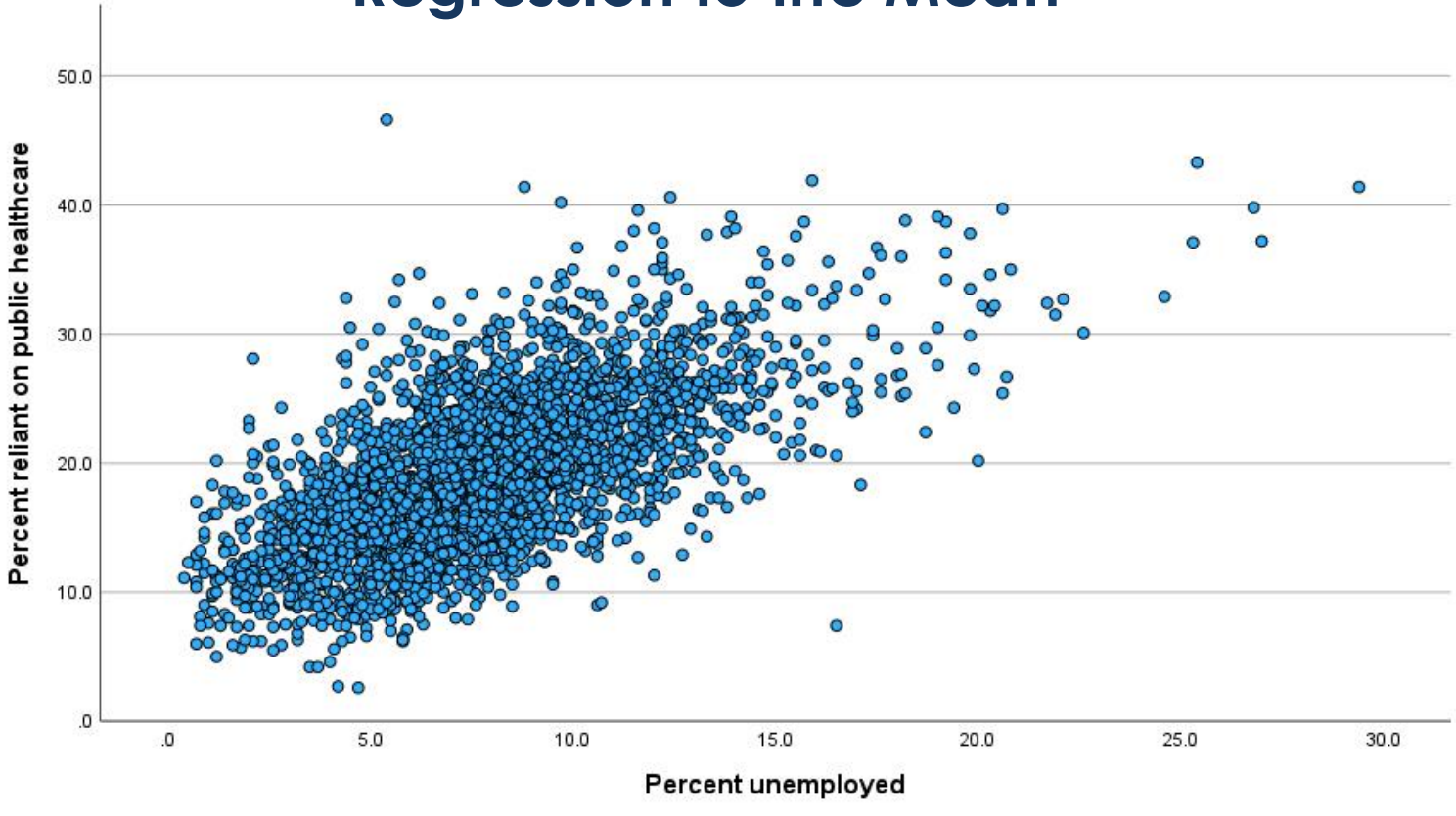
What do we mean by 'Regression'?

- 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
- Logistic Regression is used
  - When the outcome consists or 2 (or more) categories
- Quadratic Regression
  - Is a variant of Linear Regression when the outcome is continuous
  - the relationship with the dependent variable is curvilinear
- Poisson Regression
  - When modelling 'count' data such rarely occurring incidents and 'never events'

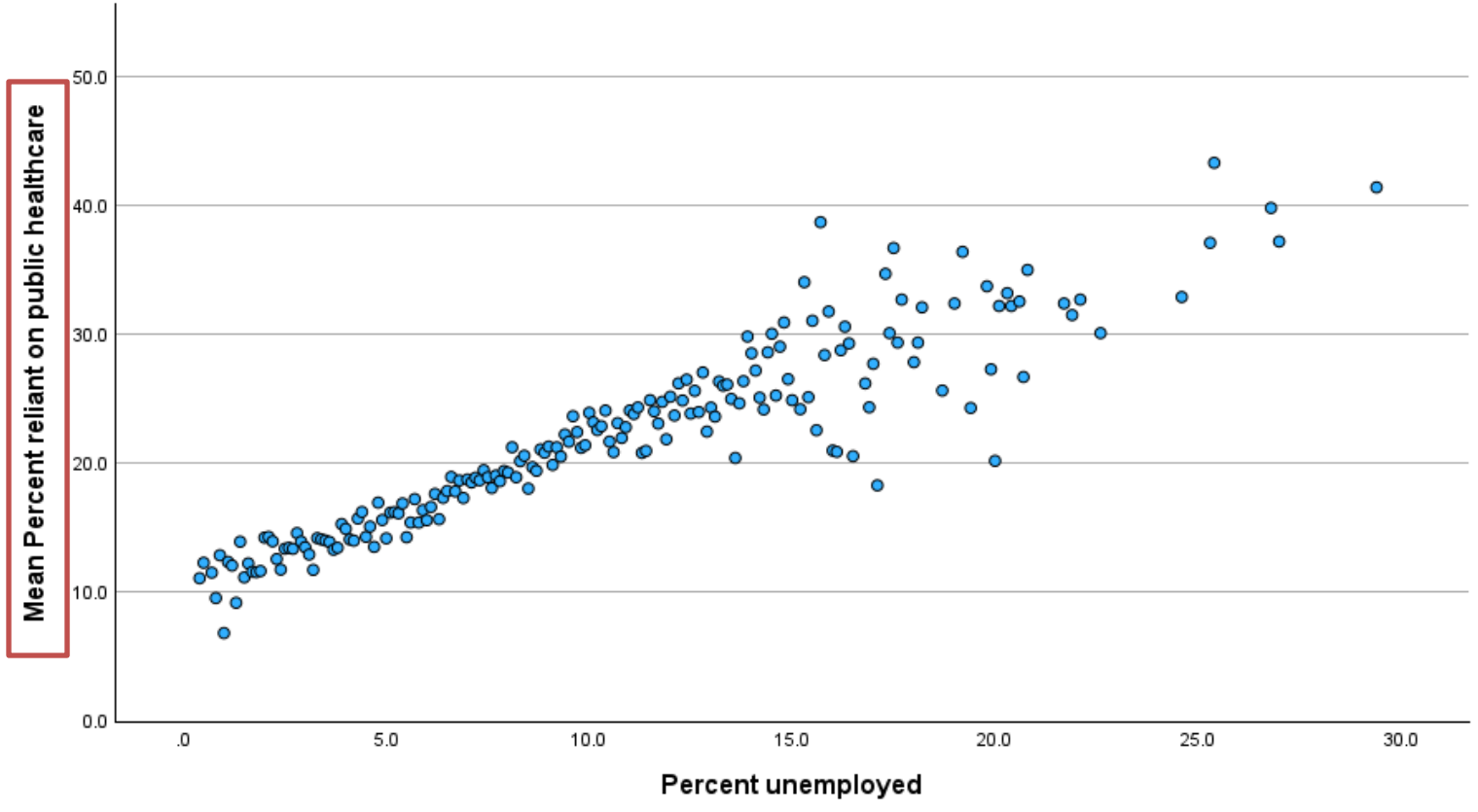
Examples of Linear Regression in Healthcare

- Modelling the relationship between blood pressure and diet
- Estimating treatment costs based on comorbidity factors
- Predicting visual acuity based on treatment and age-related factors
- Estimating calcium levels based on vitamin D levels, age and sex
- Predicting recovery time based on demographic, severity and treatment factors

“Regression to the Mean”



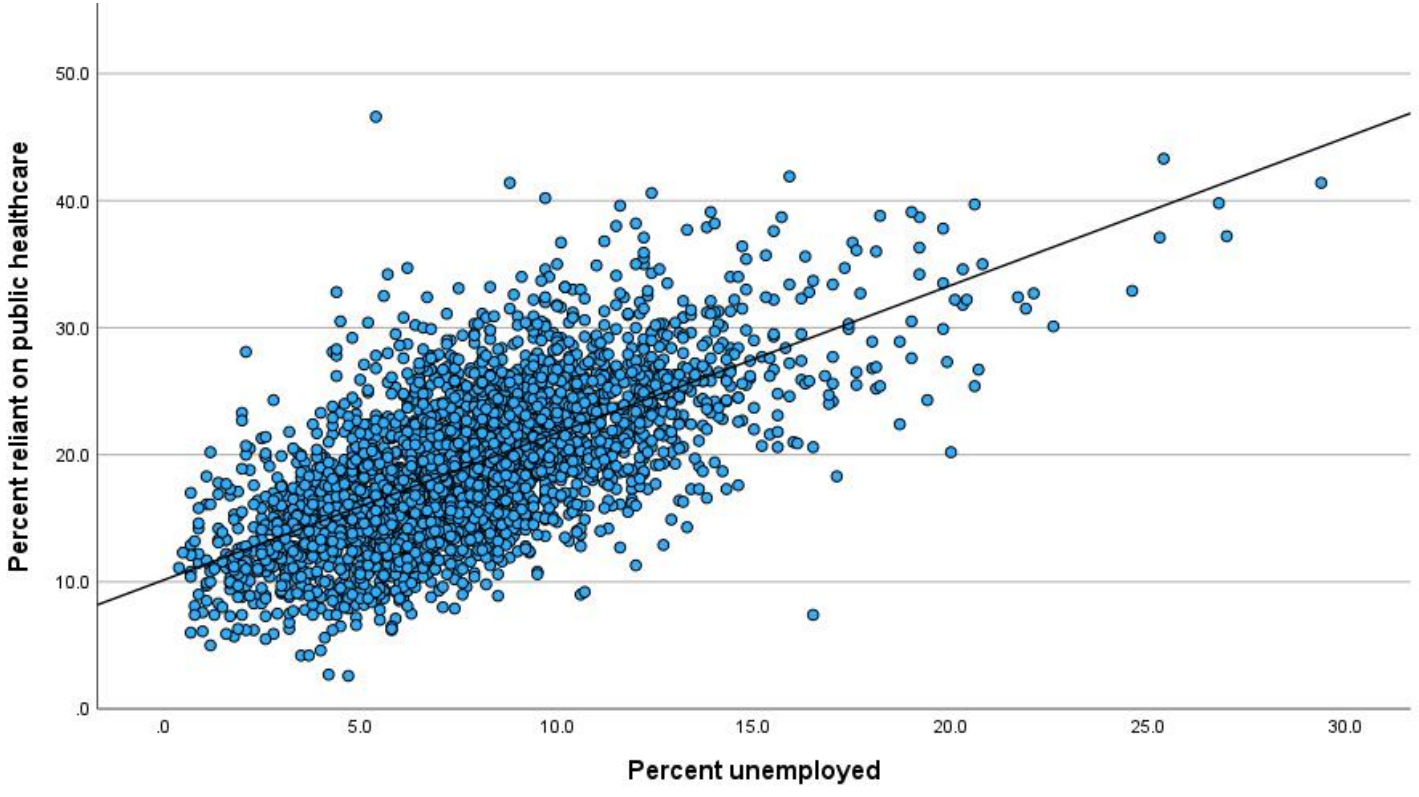
“Regression to the Mean”



SMARTVISION Europe

A SELECT INTERNATIONAL COMPANY

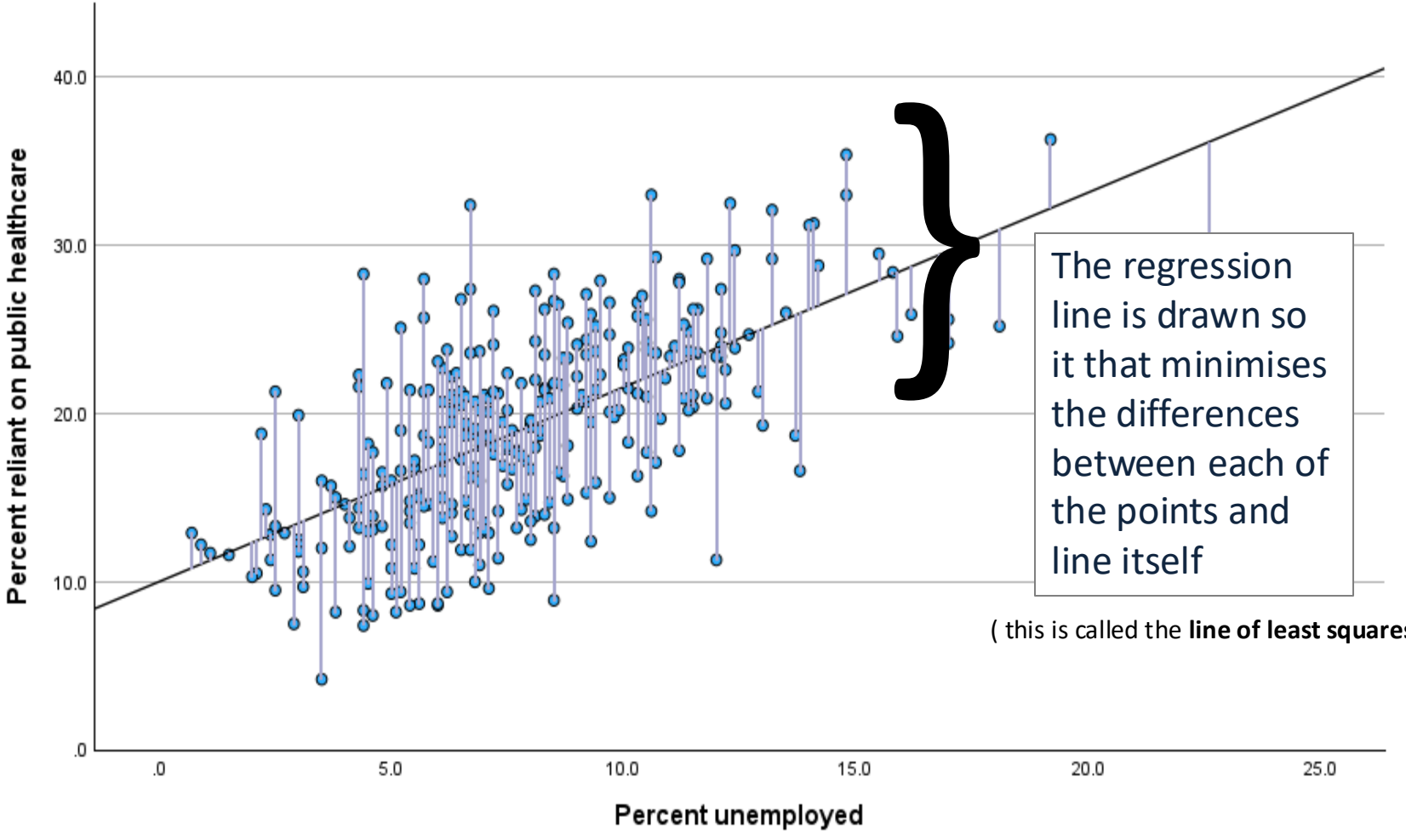
“Regression to the Mean”



SMARTVISION Europe

A SELECT INTERNATIONAL COMPANY

"Regression to the Mean"

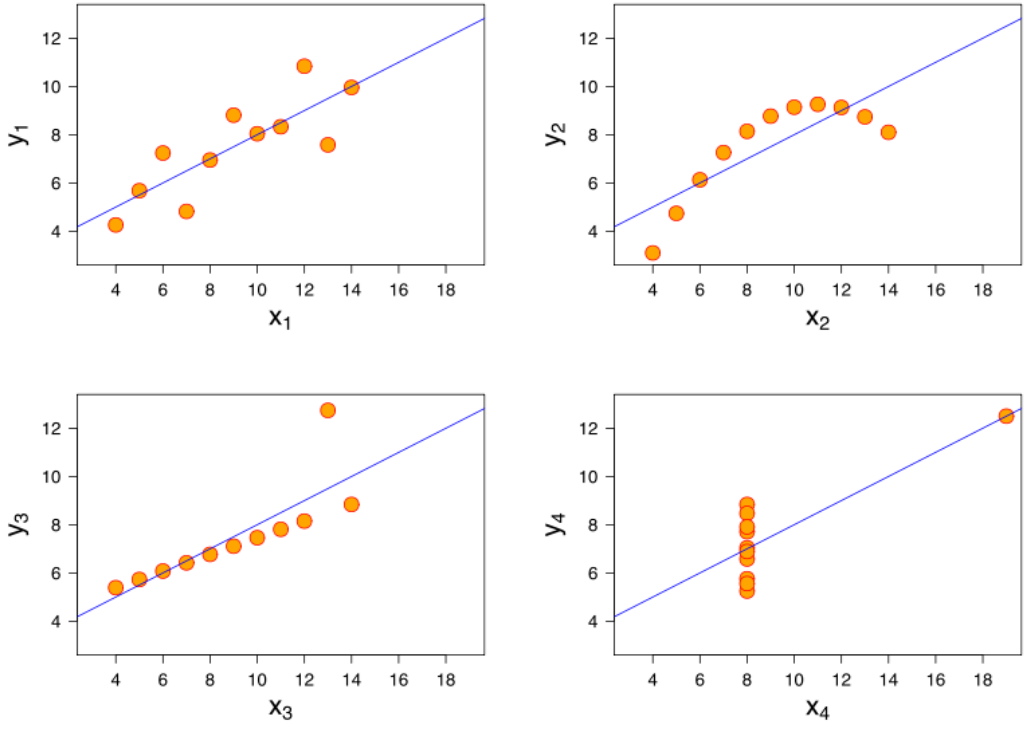


The regression line is drawn so it that minimises the differences between each of the points and line itself

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

“Regression to the Mean”

*

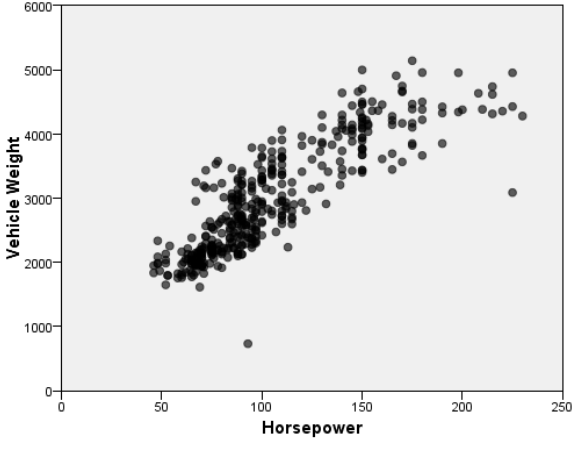


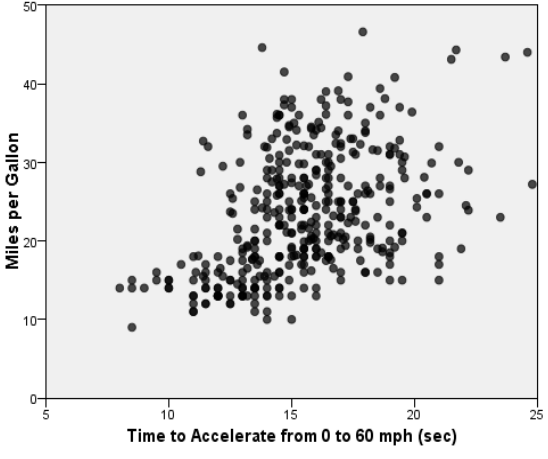
- But be careful…
- This line is just an average after all

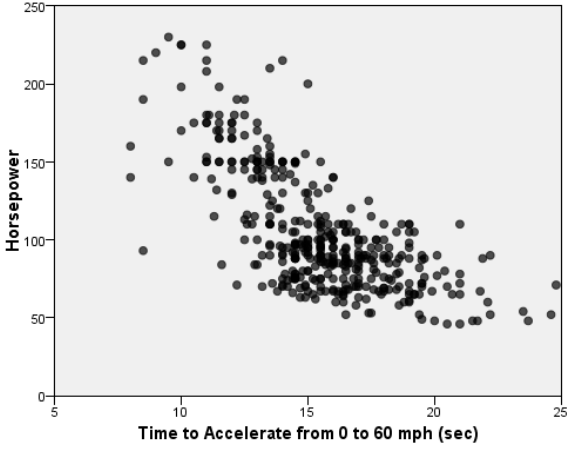
**Anscombe’s Quartet*

Measuring linear relationships with correlations

Measuring Linear Relationships







0.859

0.434

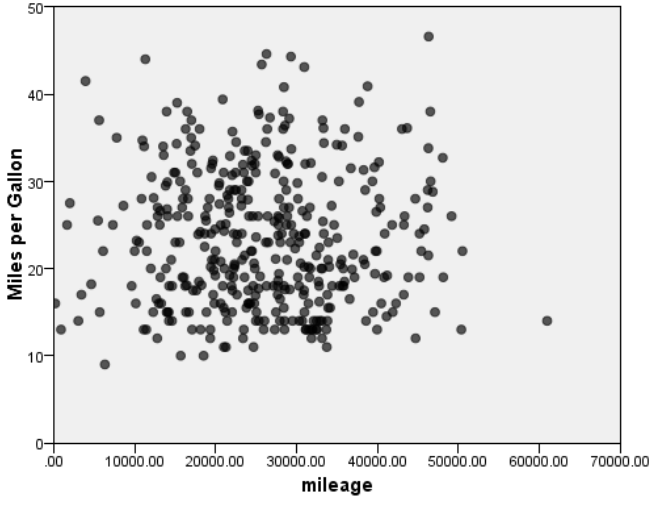
-.701

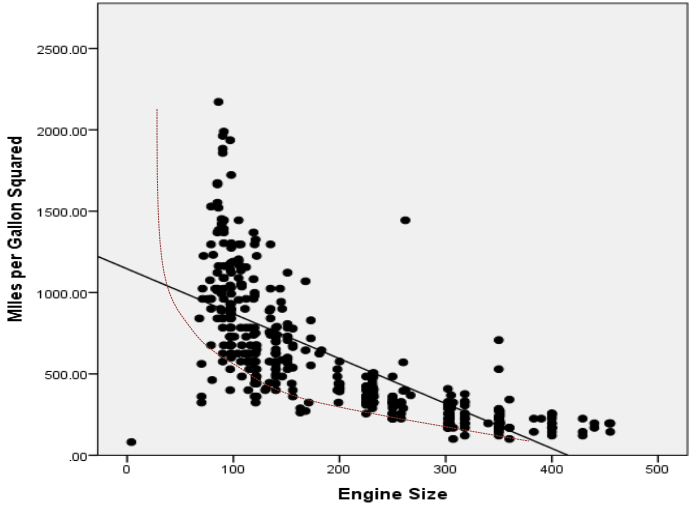
Pearson Correlation Values

SMARTVISION Europe

A SELECT INTERNATIONAL COMPANY

Non-Linear Relationships





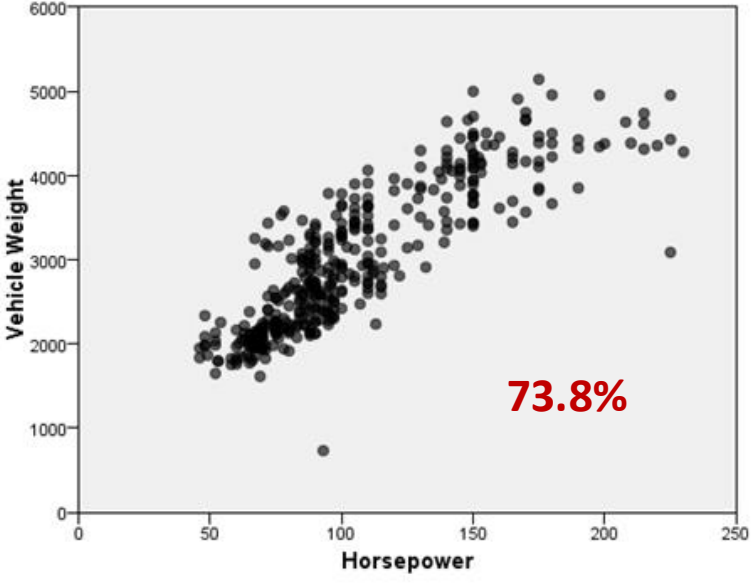
-0.005

-.671

Pearson Correlation Values

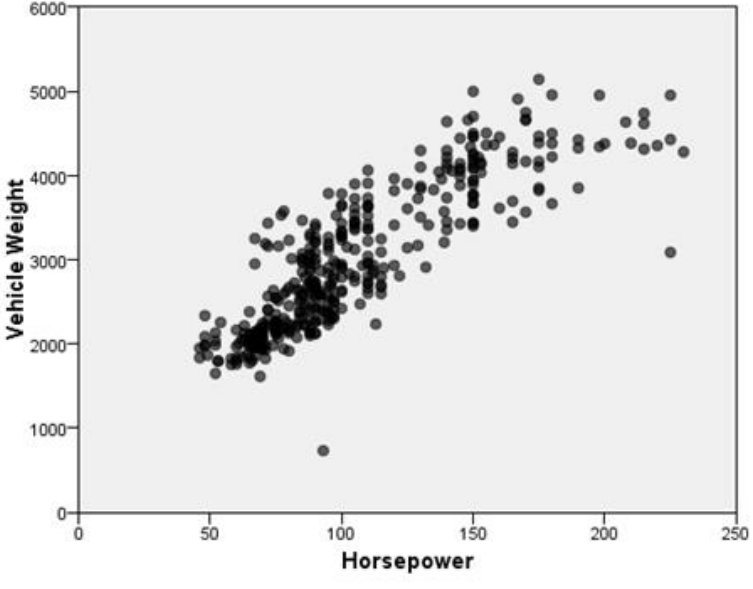
Correlations as Percentages

- 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?

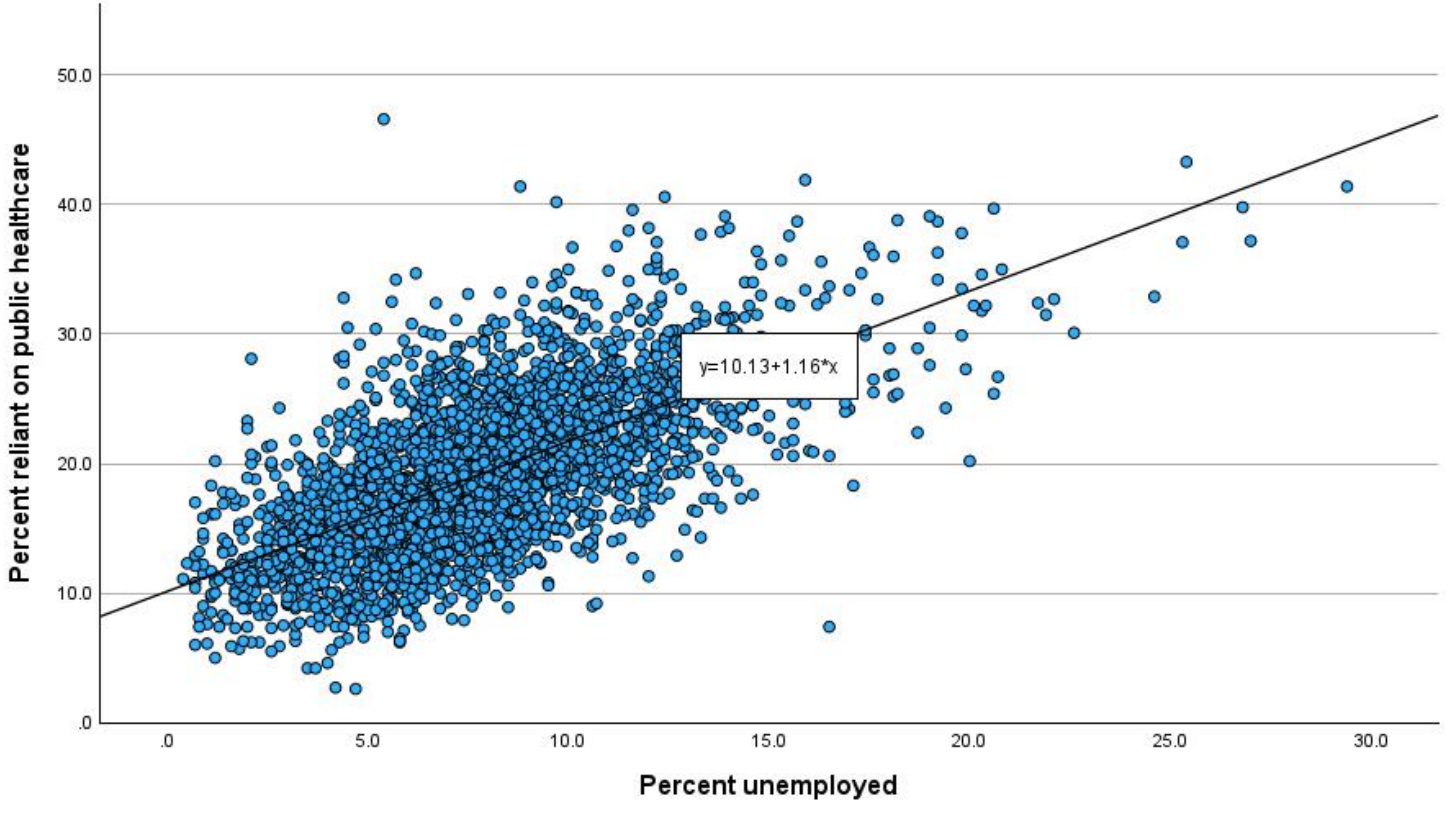
- 7 minutes per pound plus 45 minutes

$$y = mx + c$$

or

- 20 minutes per pound plus 20 minutes

$$y = a + bx$$



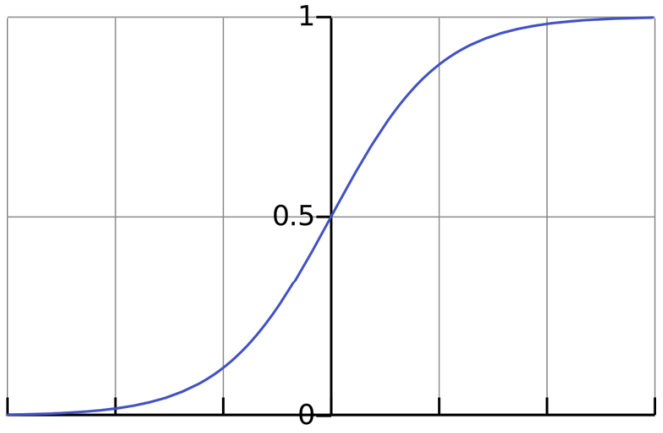
SMARTVISION Europe

A SELECT INTERNATIONAL COMPANY

SMARTVISION
Europe

Using Linear Regression to model regional per-capita cancer mortality rates

SMARTVISION
Europe



Predicting category outcomes with Logistic Regression

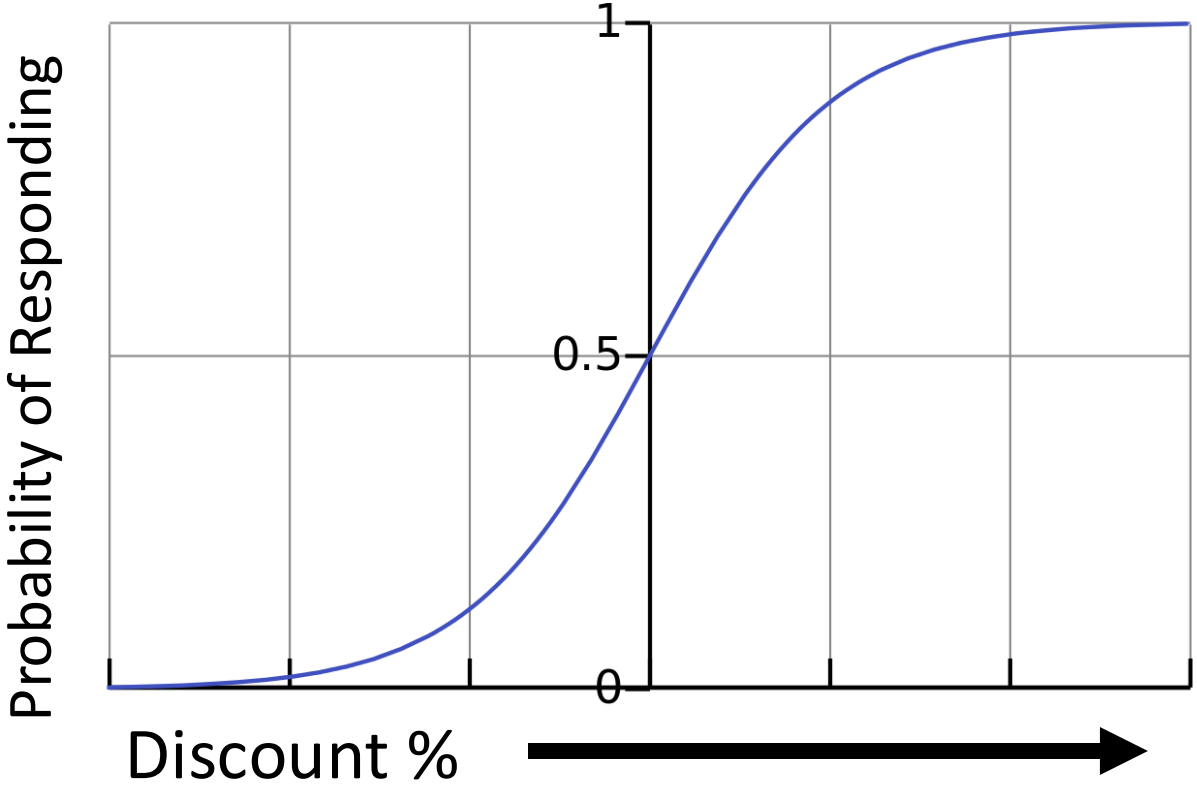
Logistic Regression

- Logistic Regression allows us to predict things that Linear Regression can't
- Such as…
  - Chance of readmission within 30 days
  - Risk of sepsis
  - Likelihood of 'no shows'
  - Probability of relapse
  - Staff retention likelihood
  - Odds Ratios in Case control studies
  - Effects of comorbidities on procedure outcomes

Logistic Regression

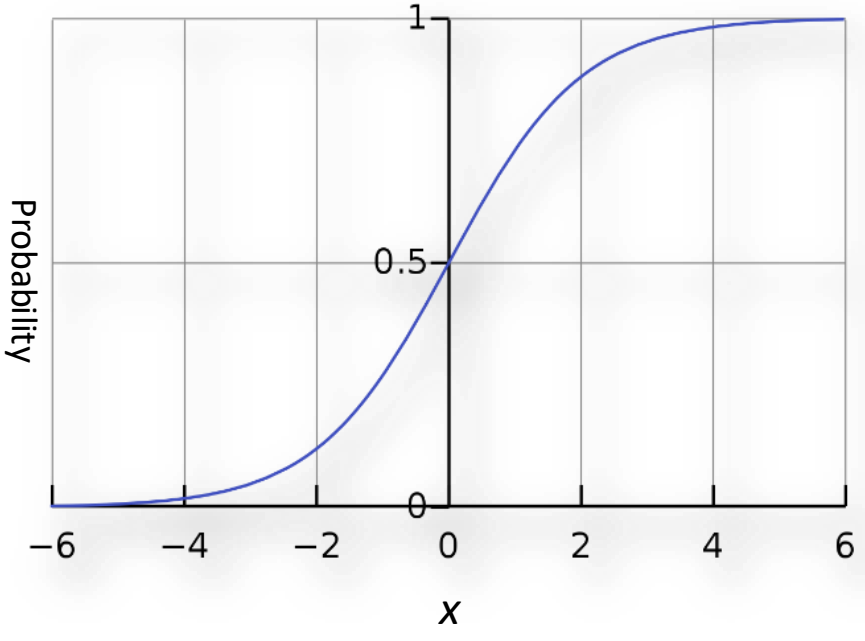
- Because outcomes are not *continuous values* standard Linear Regression won't work
- When the outcome consists of two categories, we use *Binary* Logistic Regression
- When the outcome has three or more categories, we use *Multinomial* Logistic Regression
- Logistic gets around the limitations of describing 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?

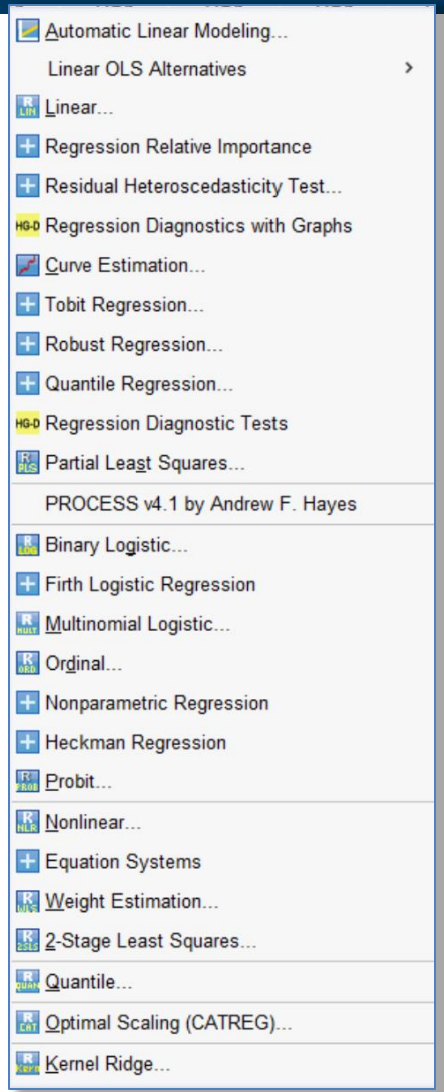


SMARTVISION
Europe

Using Logistic Regression to model low birthweight

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

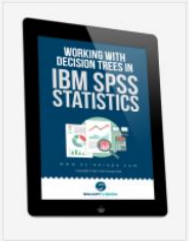
£75.00
Jarlath Quinn

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

- **Consulting Services**

Project Support

Purchase 1-2 days of consultancy time to have an expert work alongside you on your own project

Analytics Advice

Give us 3-5 days to investigate your data & analytical strategy and we'll present our recommendations re: improvements & alternatives

Analytical Deep-Dive

Let us explore your data landscape to test hypotheses, identify problem areas, find key outcome drivers or develop new applications

SMARTVISION Europe

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