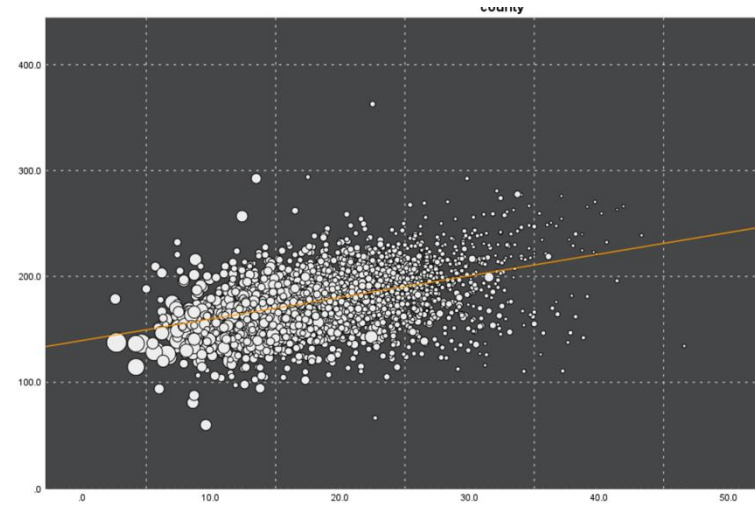




Getting started with regression techniques in SPSS

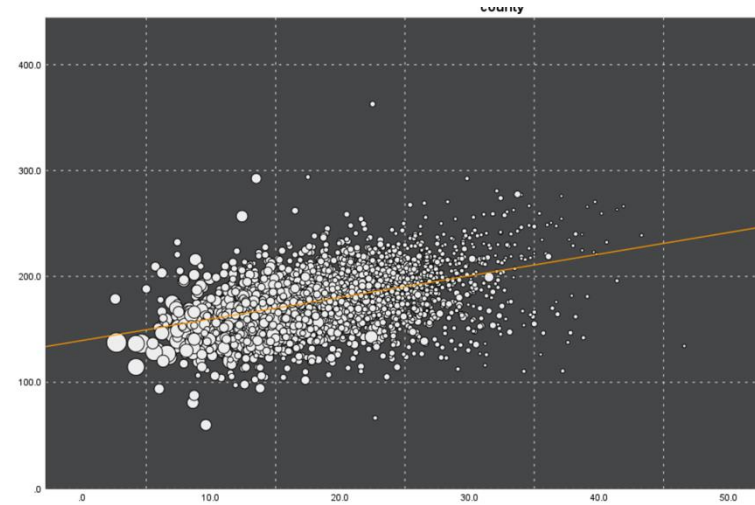
Jarlath Quinn

www.sv-europe.com

A SELECT INTERNATIONAL COMPANY



Just waiting for all attendees to join...



Getting started with regression techniques in SPSS

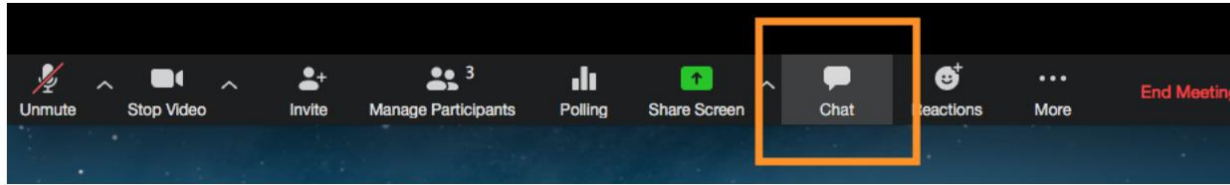
Jarlath Quinn

www.sv-europe.com

A SELECT INTERNATIONAL COMPANY

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 EDB specialising in advanced analytics & AI technologies
- Work with open-source technologies (R, Python 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/Pharma
 - Finance/Insurance
 - Media/Telecoms
 - Utilities
 - FMCG
 - Charity/Housing/Government



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

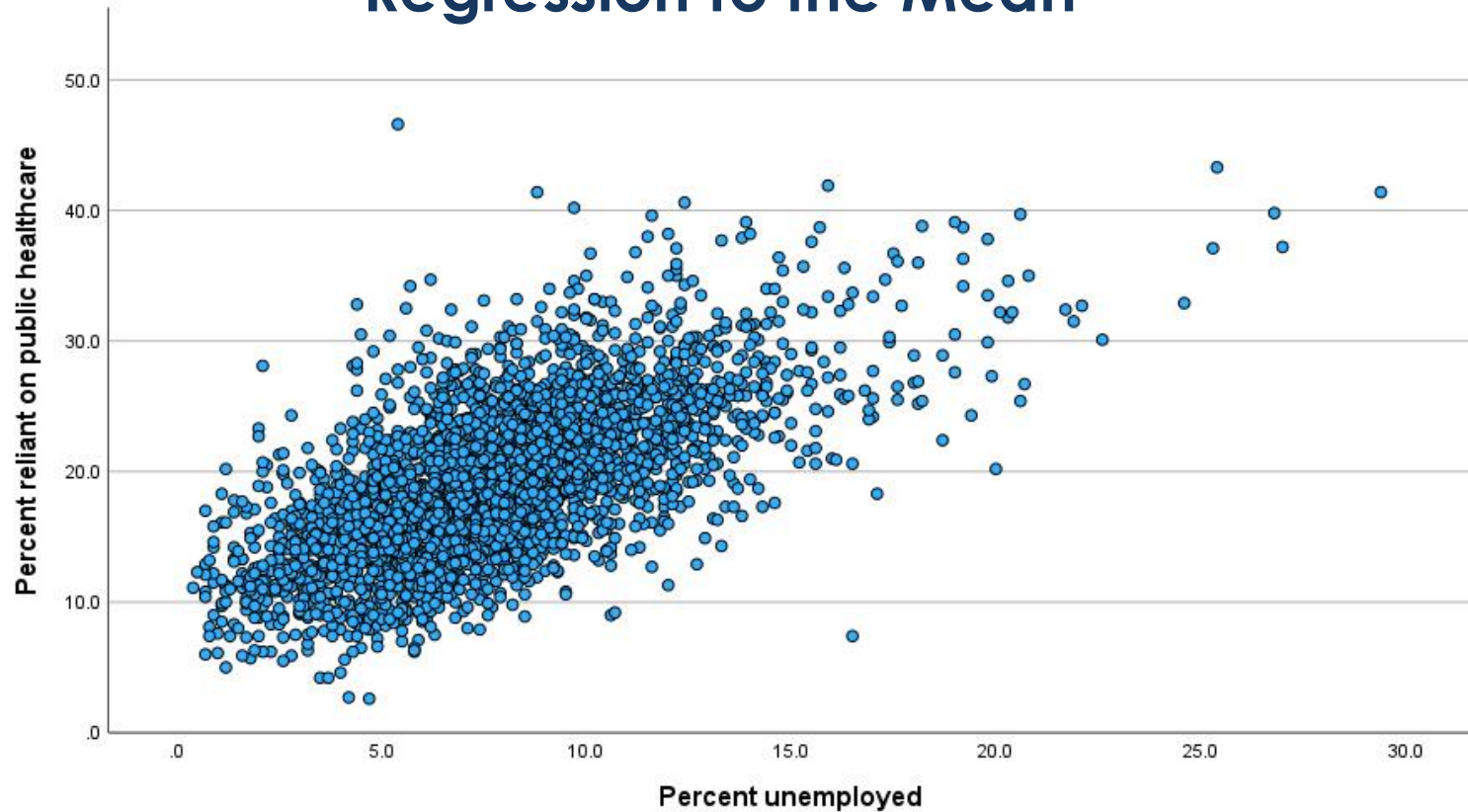
What do we mean by 'Regression'?

- [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 of 2 (or more) categories
- Poisson regression
 - is commonly used when the dependent variable records counts of events

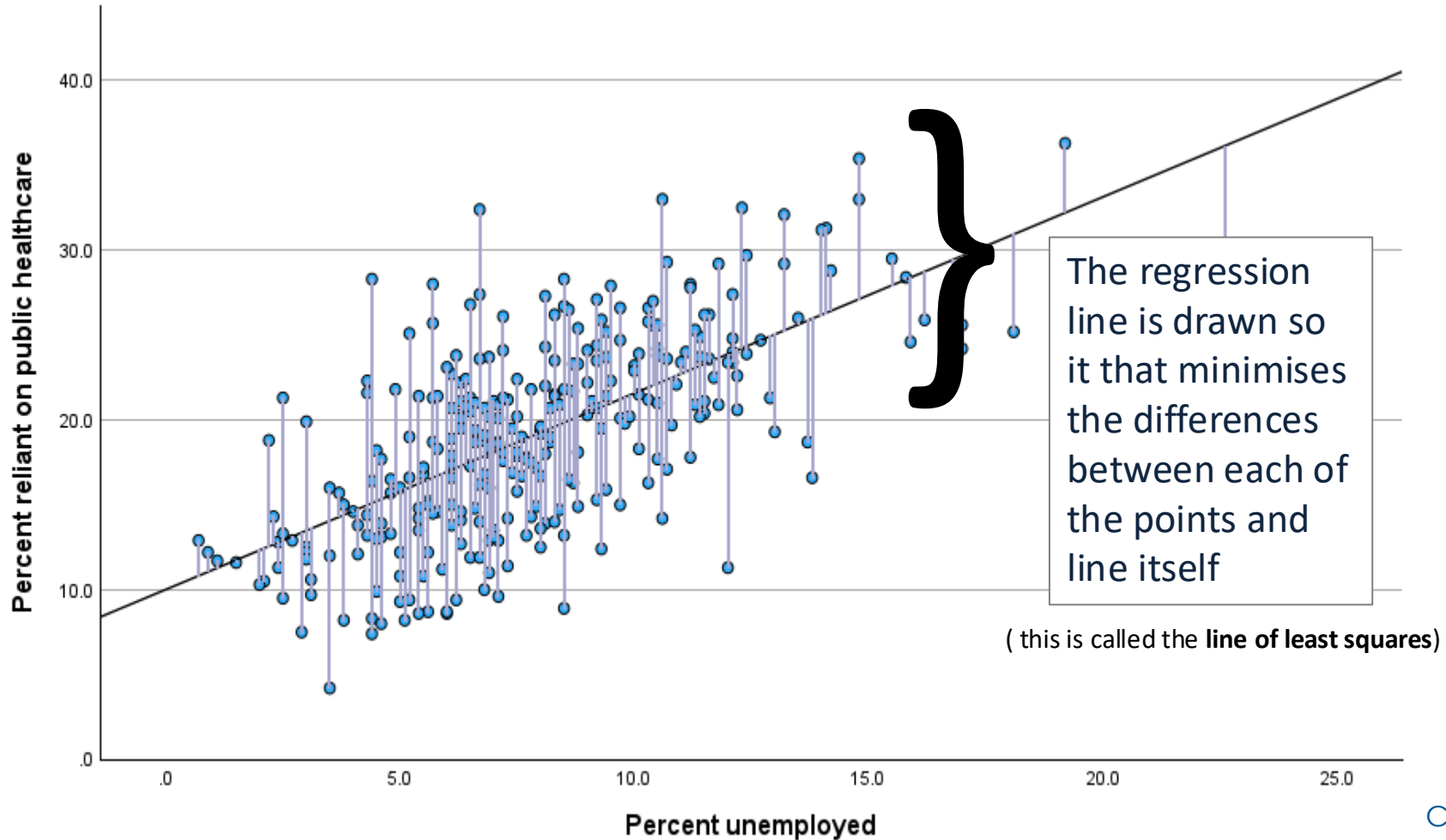
Where are Regression Techniques Used

- 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”

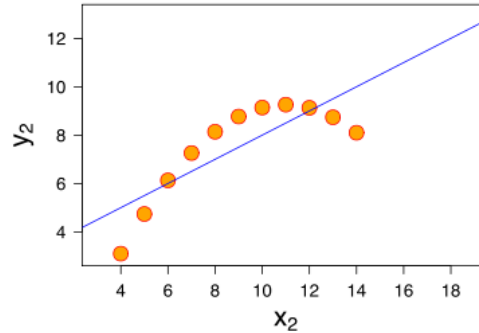
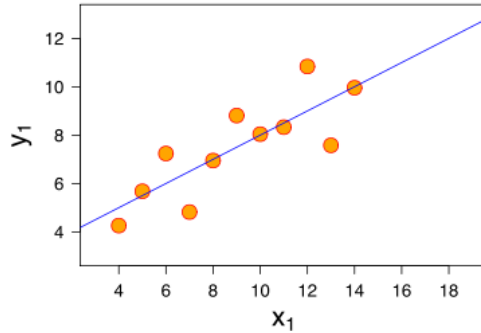


“Regression to the Mean”

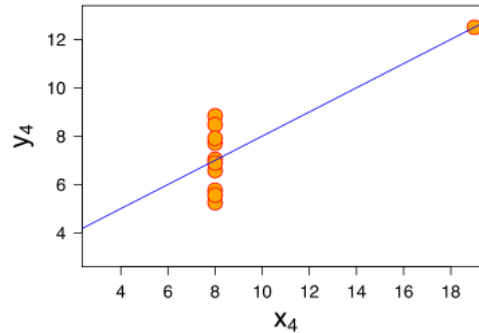
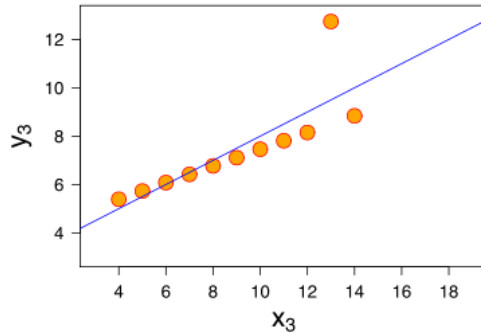


Regression to the Mean

*

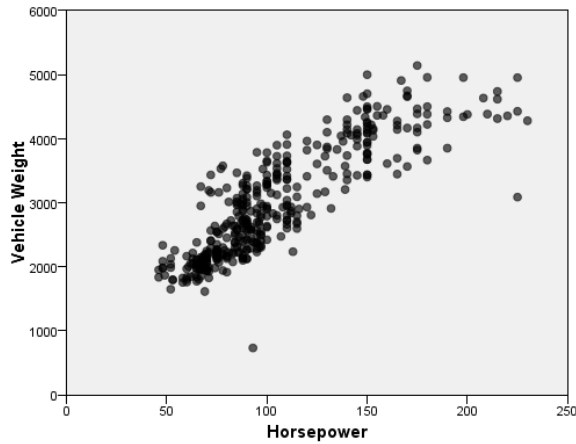


- But be careful...
- It is just an average after all...

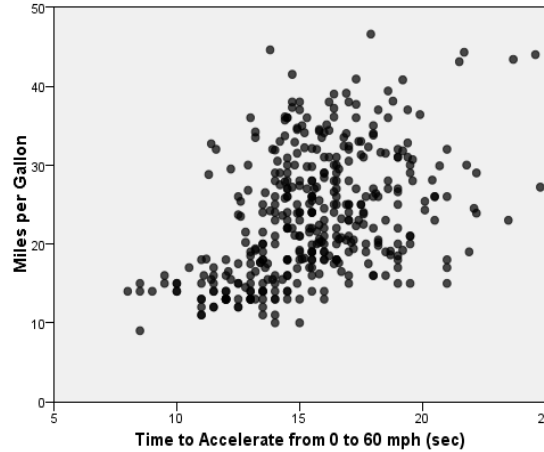


* *Anscombe's Quartet*

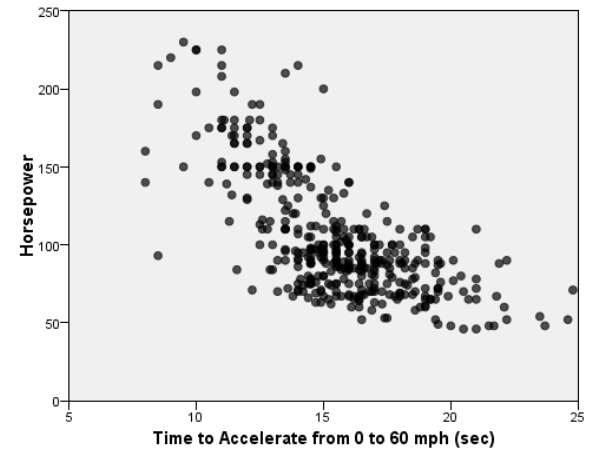
Measuring Linear Relationships



0.859



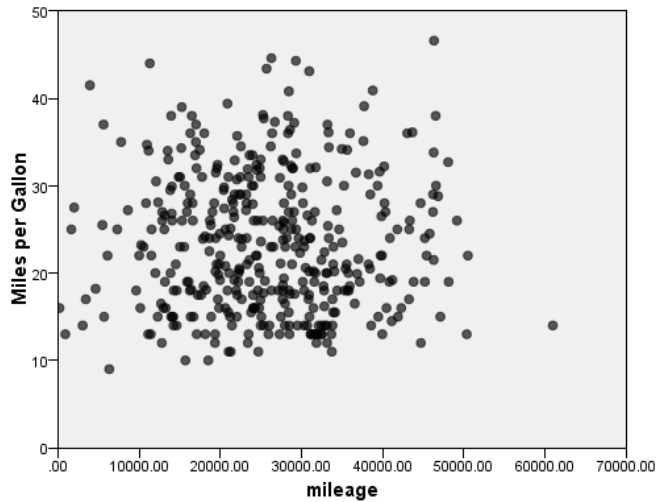
0.434



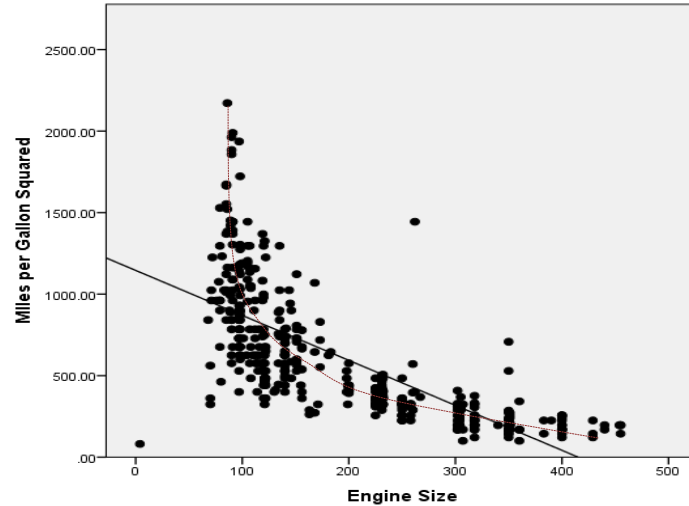
-.701

Pearson Correlation Values

Non-Linear Relationships



-0.005

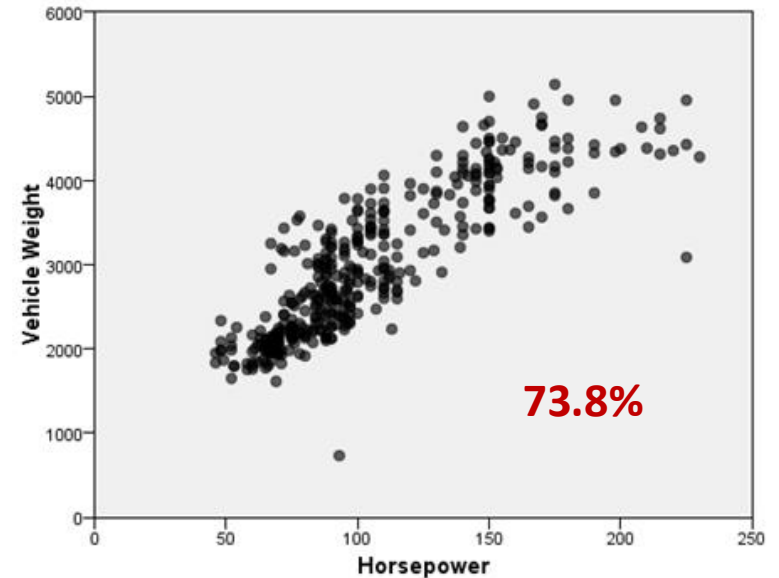


-.671

Pearson Correlation Values

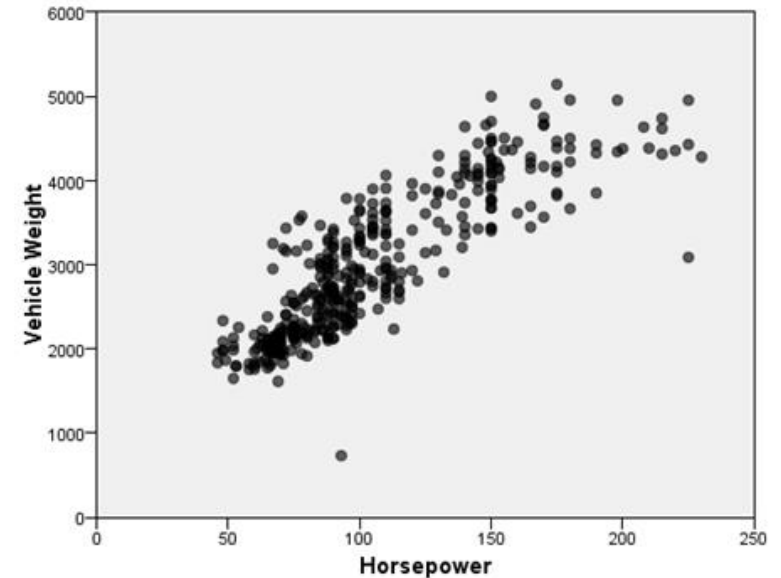
Correlations as Percentages

- Correlation = 0.859
- $0.859 \times 0.859 = 0.738$
- $0.738 = 73.8 \%$
- Correlation Squared = 'R Square'



From Correlation to Prediction

How can we express linear relationships as predictive models?





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

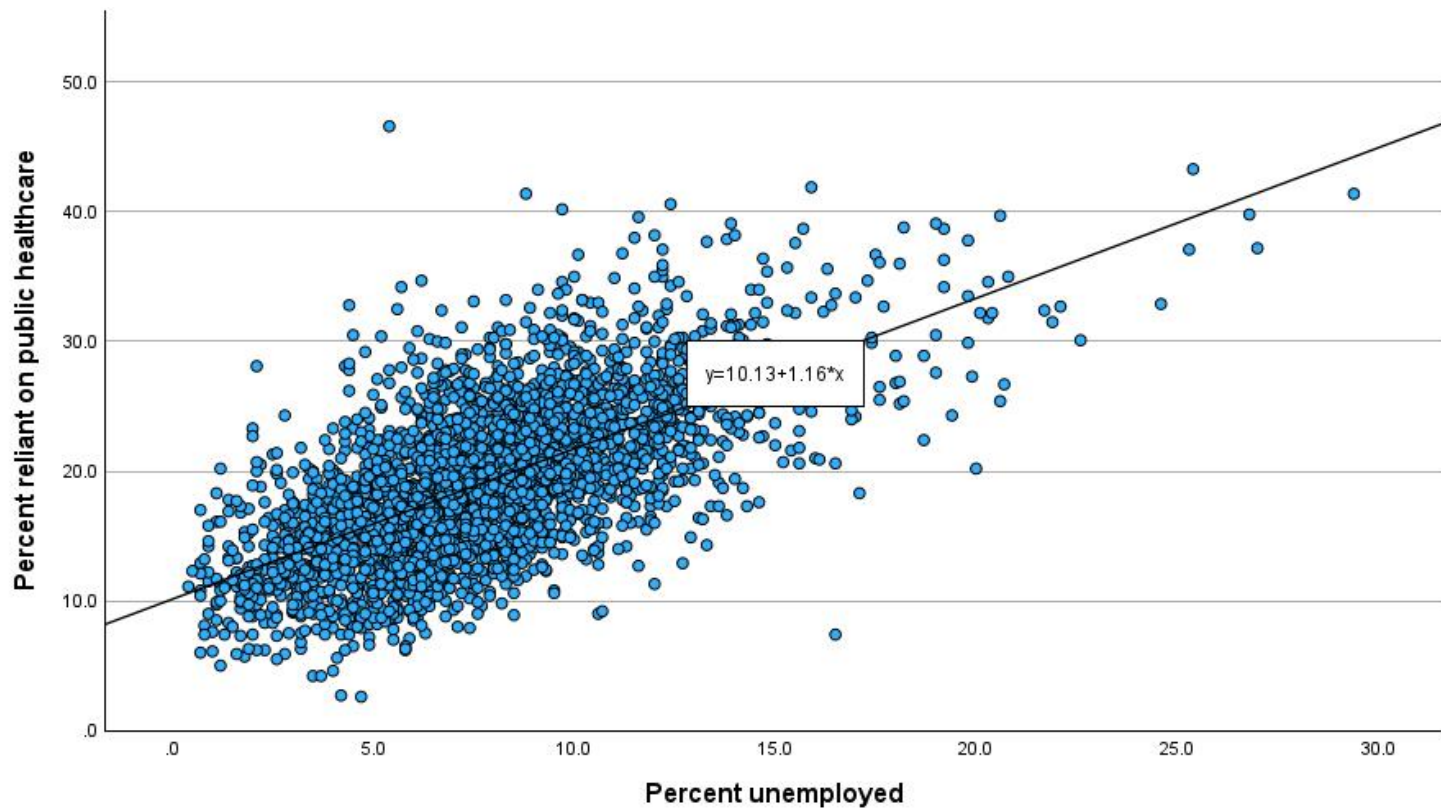
$$y = mx + c$$

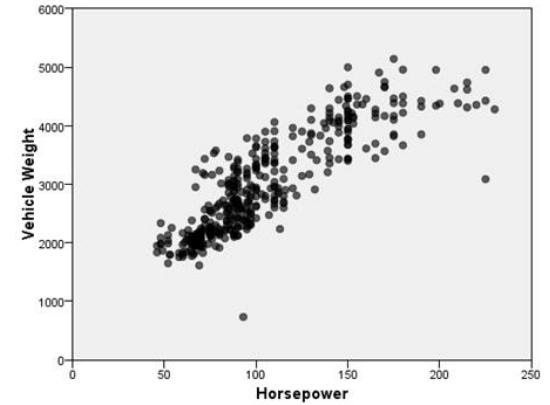
or

- 20 minutes per pound plus 20 minutes

$$y = a + bx$$







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



How can we predict category outcomes?

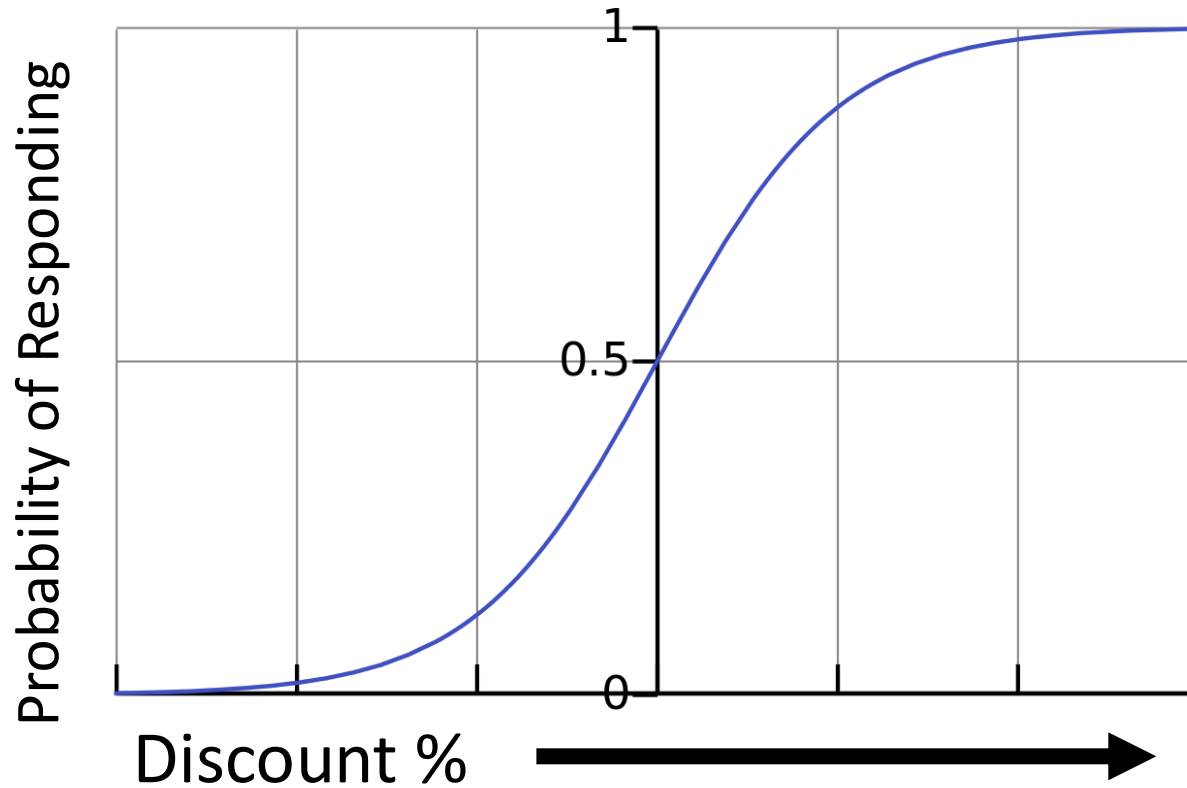
Logistic Regression

- 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

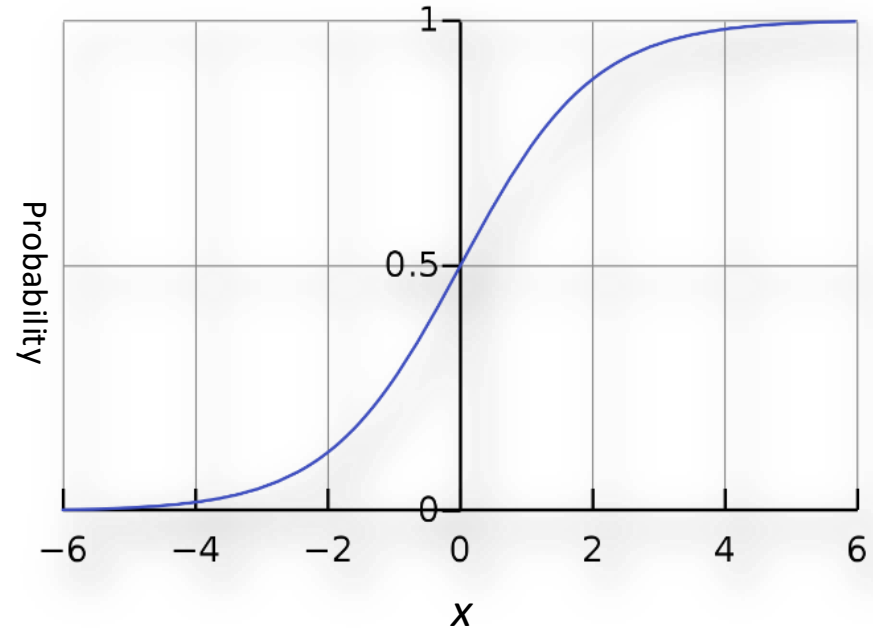
- But....
- These outcomes are not *continuous numbers* so standard linear regression won't work
- When the outcome consists of two categories we use *Binary* Logistic Regression
- When the outcome has three or more categories we use *Multinomial* Logistic Regression
- Logistic gets around the limitations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

Logistic Regression



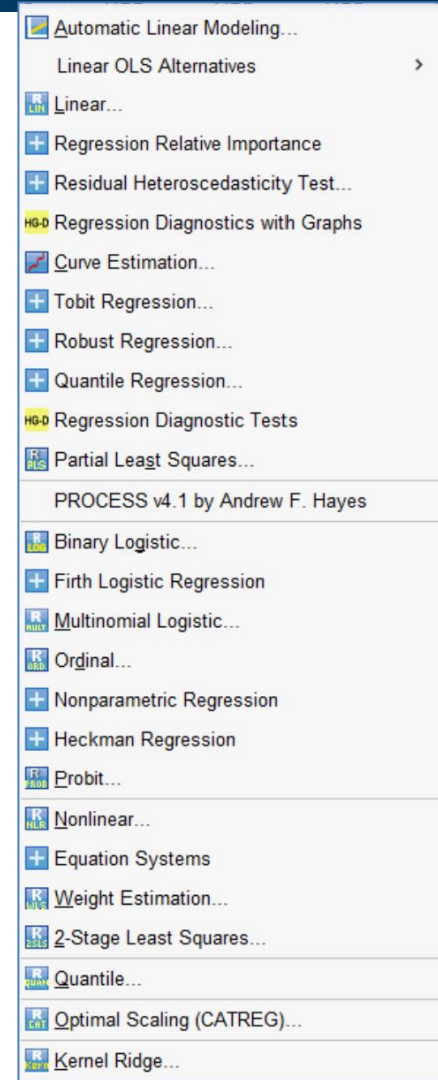
Logistic Regression

- 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**



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

Online training materials
free to Smart Vision
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



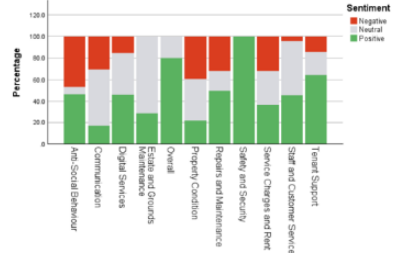
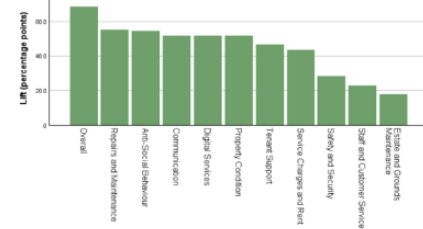
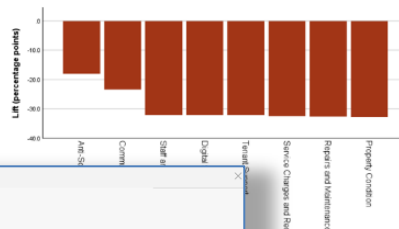
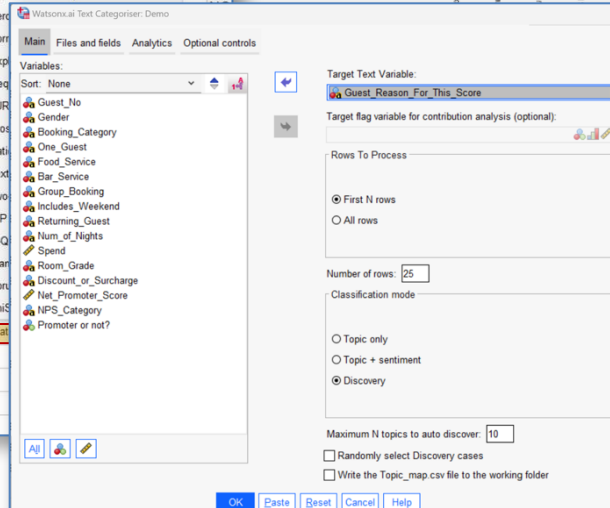
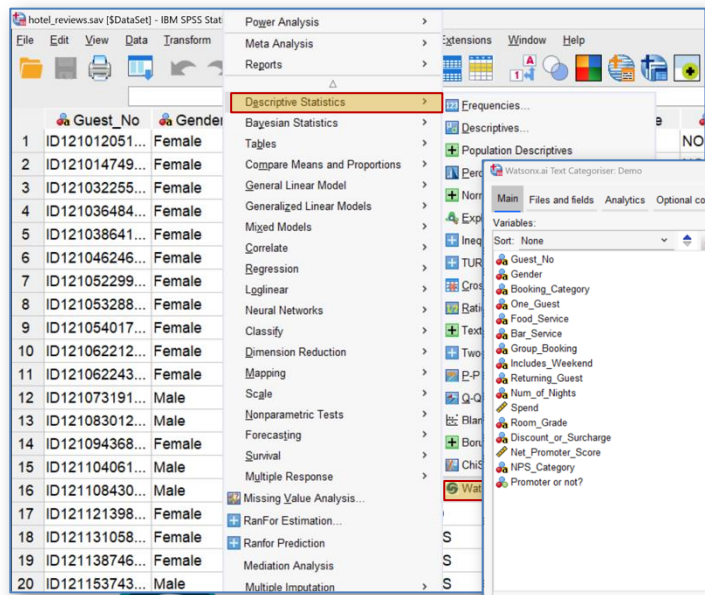
Introduction to IBM SPSS
Statistics course



ChatSPSS is a free generative AI chatbot developed by Smart Vision Europe Ltd based on official IBM documentation of SPSS procedures

A screenshot of the ChatSPSS web interface. The browser address bar shows "https://chatspss.com". The page has a dark blue header with "NEW CONVERSATION", "EXPORT PDF", "LOG OUT", "PRIVACY", and "SMARTVISION" links. The main content area is divided into two columns. The left column, titled "CONVERSATIONS", shows a list of questions: "How can I recode a string variable called region to create a numeric equivalent?" and "How can I recode a string variable called region to create a numeric equivalent?". The right column shows the answer to the selected question. At the top of the right column is a blue box with the question: "How can I recode a string variable called region to create a numeric equivalent?". Below this is the answer: "To recode a string variable called region to create a numeric equivalent, you can use the RECODE command with the INTO keyword." followed by the section "Steps to Recode the Variable" with three bullet points: "Use the RECODE command to specify the original values and the new values.", "Use the INTO keyword to specify the target numeric variable.", and "Ensure that the target variable is declared before using it." Below this is the section "Example Syntax" with the code: "RECODE region ('North'=1) ('South'=2) ('East'=3) ('West'=4) INTO region_num." and two bullet points: "In this example, the string variable region is recoded into the numeric variable region_num." and "The original values 'North', 'South', 'East', and 'West' are replaced with the numeric values 1, 2, 3, and 4, respectively." Below this is the section "Alternative Method" with the text "Alternatively, you" followed by a text input field containing "region_num" and the text "riable." followed by another text input field containing "region_num". Below these fields is the code: "AUTORECODE VAI", "/INTO region_num", and "/PRINT.".

- **NEW: Text Analytics with AI**
- A new downloadable tool from Smart Vision Europe
- Powered by IBM Watsonx.ai
- Works seamlessly as part of IBM SPSS Statistics



Smart Vision Europe: Services and Expertise



We have decades of experience providing guidance, training and consultancy in the delivery of effective data science initiatives.



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