Generating SPSS training materials in StatJR

Professor William Browne

Video 1 Interoperating with SPSS

- StatJR is a software system that can interoperate with other statistical software.

- For example there is a StatJR template to fit a regression in many packages including SPSS.

- SPSS is often used for training in the social sciences.

- We have extended StatJR's functionality so that it can automatically create 'bespoke' SPSS training materials.

Video 1 Interoperating with SPSS

- StatJR is a software system that can interoperate with other statistical software.

- For example there is a StatJR template to fit a regression in many packages including SPSS.

- SPSS is often used for training in the social sciences.

- We have extended StatJR's functionality so that it can automatically create 'bespoke' SPSS training materials.

British Academy training eBooks grant

Using Statistical E-books to teach undergraduate students quantitative methods and statistical software (Browne, Charlton and Washbrook)

Inspired by:

- Previous work developing Stat-JR, e-books and SAAs to help *researchers* analyse their own statistical data

- Concerns about the shortage of quantitatively-skilled UK social science graduates that prompted the Q-Step programme

The basic idea

- To produce a set of (static) training materials that guide a student through the implementation of basic statistical techniques on an example dataset in the SPSS software

- To provide teachers with a tool that allows them to QUICKLY customise the materials to their own discipline-specific data example(s)
  - Sociology, politics, geography, childhood studies, social policy, social anthropology, area studies…

The end product – what the *student* gets

13 sets of practical exercises (pdfs) with 3 components

1. Takes student through a particular statistical concept in detail, and how to implement it in SPSS, using a specific data example (*learning component)*

2. A worksheet that asks the student to try out their knowledge by applying the techniques to a second dataset or set of variables *(practice component)*

3. Solutions to the worksheet *(self-evaluation component)*

Topics and "starter" data

Data extract from the 2015 Programme for International Student Assessment (PISA) – international survey of science understanding of 15-year olds

England sub-sample (N=5194)

Topics

Over the course of the project the following twelve topics were identified. For each of these a generic eBook, which the user can use with their own data, and an example using the PISA data was created. Outputs from the PISA PDFs are linked below:

1. Descriptive statistics for categorical variables [Practical] [Quiz]
2. Descriptive statistics for continuous variables [Practical] [Quiz]
3. Tabulation [Practical] [Quiz]
4. Tests for normality [Practical] [Quiz]
5. Independent samples t-tests [Practical] [Quiz]
6. Paired t-tests [Practical] [Quiz]
7. Non-parametric tests for unpaired [Practical] [Quiz] and paired data [Practical] [Quiz]
8. Chi-squared tests [Practical] [Quiz]
9. Correlation [Practical] [Quiz]
10. Linear regression [Practical] [Quiz]
11. ANOVA [Practical] [Quiz]
12. Multiple regression [Practical] [Quiz]

Example pdf

Correlations in SPSS (Practical)

Correlation practical

In this practical we will investigate whether there is a relationship between two variables by looking how correlated they are.

Two of the key predictors of academic achievement proposed by theories of student learning are self-efficacy and instrumental motivation. Self-efficacy refers to an individual's confidence in their ability to perform a task well, and instrumental motivation to the belief that learning will be useful for one's later career. In this practical we address the question of whether these two constructs are correlated, or specifically whether students who are more confident in their abilities in science are systematically more (or less) likely to view science learning as important for their future prospects. The PISA measure of science self-efficacy (SCIEEFF) was derived from students' responses to questions on how easy they would find it to perform eight science tasks on their own, such as "Identify the better of two explanations for the formation of acid rain". The measure of instrumental motivation (INTMOVSCI) was derived from four items in which students rated their agreement with statements like "Many things I learn in my school science subjects will help me to get a job" (see PISA datafile description for further details).

To do this we will begin by simply plotting the two variables in SPSS:

- Select **Scatter/Dot** from the **Legacy diagnostics** available from the **Graphs** menu.
- Select Simple Scatter and click on Define to bring up the Simple Scatterplot window.
- Copy the **Science self-efficacy score[SCIEEFF]** variable into the **Y Axis** box.
- Copy the **Science instrumental motivation score[INSMOVSCI]** variable into the **X Axis** box.
- Click on the **OK** button.

SPSS will then draw a scatterplot of the two variables which can be seen below:

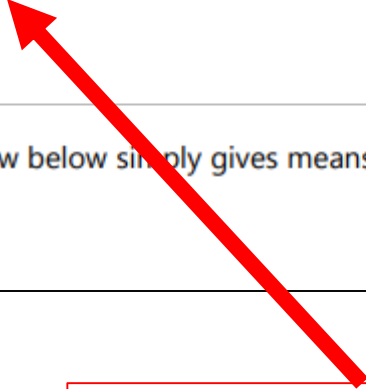Instructions for the student to carry out in SPSS

Example pdf

We will now finally turn our attention to the main topic of this practical which is the calculation of the correlation between our two variables. SPSS offers several correlation coefficients and we will consider these here in turn. All three are available through the Analyse->Correlate->Bivariate option in SPSS.

- Select **Bivariate...** from the **Correlate** option available from the **Analyse** menu.
- Copy the **Science self-efficacy score[SCIEEFF]** and the **Science instrumental motivation score[INSMOVSCI]** variables into the **Variables** box.
- Click on the **Options** button and Select the **Means and Standard deviations** tick box.
- Click on the **Continue** button to return to main window.
- Click on the **OK** button.

The correlation command will produce two output tables. The first table which we show below simply gives means and standard deviations for the two variables we are comparing.

Descriptive Statistics

SPSS instructions tell the student which variables to use in the operations (here SCIEEFF and INSMOVSCI)

Example pdf

In the next table we see the correlation matrix for the variables we are considering:

Correlations

		Science self-efficacy score	Science instrumental motivation score
Science self-efficacy score	Pearson Correlation	1	.327**
	Sig. (2-tailed)		.000
	N	4726	4694
Science instrumental motivation score	Pearson Correlation	.327**	1
	Sig. (2-tailed)	.000	
	N	4694	4791

**. Correlation is significant at the 0.01 level (2-tailed).

The Correlate option can be used for more than two variables simultaneously and will then give all correlatio

The pdf shows the output the student will generate and see in the SPSS viewer

And crucially gives some interpretation and commentary that depend on the statistical results

We would report the result as follows: The variables **SCIEEFF** and **INSMOVSCI** were significantly and moderately positively correlated $r = .327$, N = 4694, $p < .001$. Note there is no need for a table when reporting a single correlation.

Example quiz

Correlations in SPSS (Quiz)

Correlation practical questions

In this practical we will investigate whether there is a relationship between two variables by looking how correlated they are.

Family background, as well as psychological traits, has also been shown to be an important predictor of educational achievement. Family background encompasses many different dimensions of parental resources which do not necessarily vary perfectly together. In this example you will explore the association between an indicator of a family's wealth and the degree of emotional support provided by parents for a child's learning. PISA's wealth variable, WEALTH, is derived from reports of whether the family owns eight items, such as a car, a computer and a room of the child's own. The parental emotional support variable, EMOSUPS, is derived from four items with which students rated their strength of agreement, e.g. "My parents support my educational efforts and achievements" (see PISA datafile description for further details).

Firstly use SPSS to create a scatterplot of **WEALTH** and **EMOSUPS** and answer the following:

- Question: What does the scatterplot say about the relationship between **WEALTH** and **EMOSUPS**?

Next use SPSS and the Explore screen to create histograms, normality tests and QQ plots of **WEALTH** and **EMOSUPS** and answer the following:

- Question: What do the plots and tests tell us about the normality of **WEALTH**?
- Question: What do the plots and tests tell us about the normality of **EMOSUPS**?

Next use SPSS and the Correlate screen to answer the following:

- Question: What is the Pearson correlation coefficient between **WEALTH** and **EMOSUPS** and is it significant?

A separate document asks the student to try out what they have learned using different variables, and also provides a solutions sheet

Example solutions
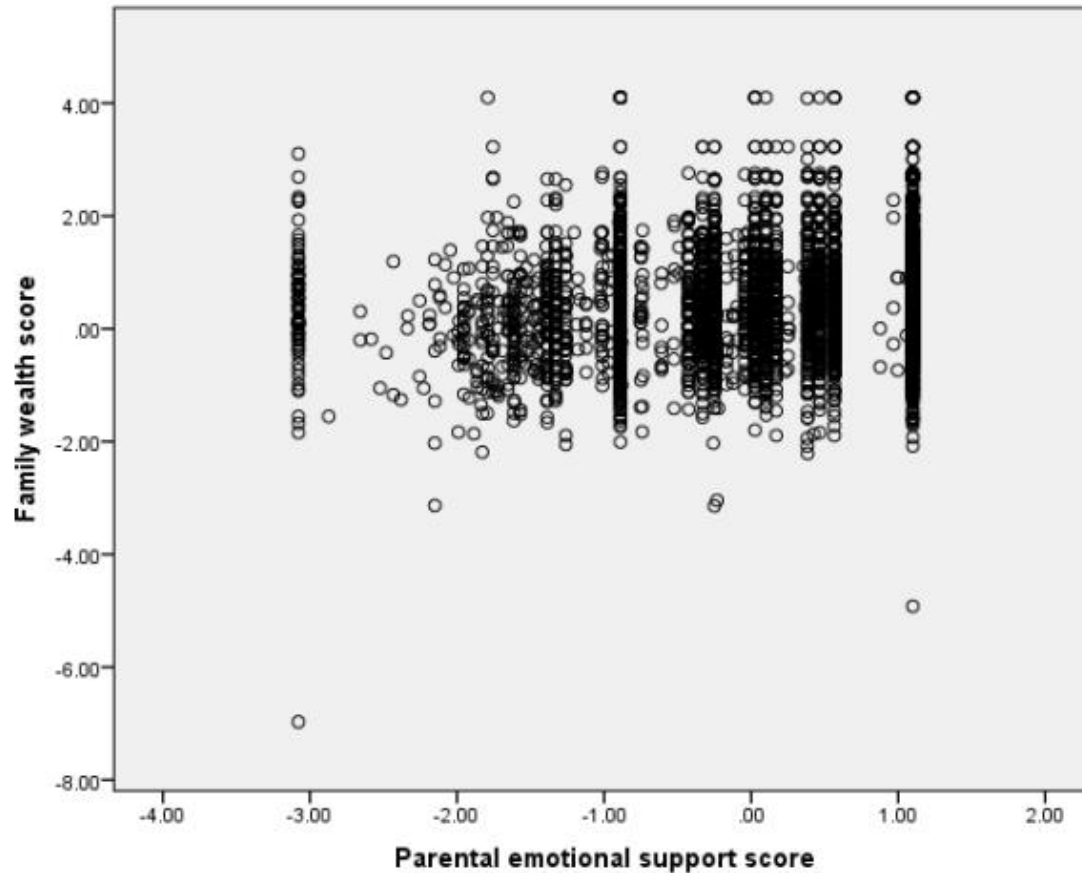
Looking at the scatterplot there appears to be a positive correlation between the variables with larger values of **WEALTH** associated with larger values of **EMOSUPS** (an upward sloping relationship) but this relationship is not that strong with possibly a few more points in the

So far so 20th century…

Now suppose I am a lecturer in political science and I would like to create learning materials using a very different dataset

- Unit of observation is parliamentary constituency rather than individual student
- Data relate to vote shares of political parties and characteristics of the population rather than individual characteristics

Perhaps I have a dataset from my own research I'd like to share with students!

Load the Correlations practical ebook in TREE

And select the dataset that will be the new example

Load the Correlations practical ebook in TREE

Then select the variables that will be used in the practical
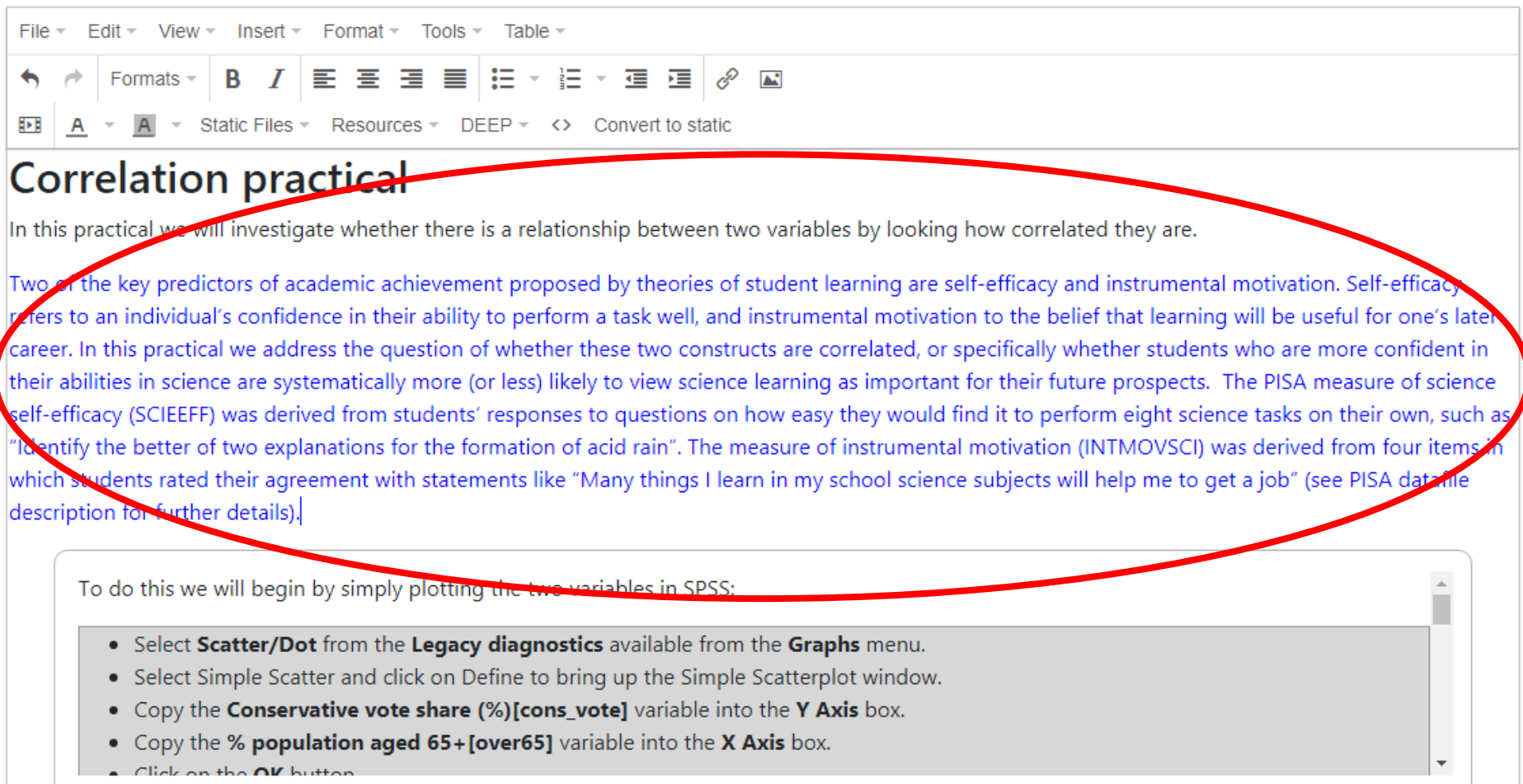
All outputs and interactive contextual text update automatically using the new variables and associated SPSS output.
The user has the opportunity to edit the example-specific blue introductory text (and any other text) by hand

File ▾   Edit ▾   View ▾   Insert ▾   Format ▾   Tools ▾   Table ▾

↶ ↷   Formats ▾   **B** *I*   ≡ ≡ ≡ ≡   ☷ ▾   ☷ ▾   ⇥ ⇥   🔗 🖼

▣   A ▾   A ▾   Static Files ▾   Resources ▾   DEEP ▾   <>   Convert to static

Correlation practical

In this practical we will investigate whether there is a relationship between two variables by looking how correlated they are.

Two of the key predictors of academic achievement proposed by theories of student learning are self-efficacy and instrumental motivation. Self-efficacy refers to an individual's confidence in their ability to perform a task well, and instrumental motivation to the belief that learning will be useful for one's later career. In this practical we address the question of whether these two constructs are correlated, or specifically whether students who are more confident in their abilities in science are systematically more (or less) likely to view science learning as important for their future prospects. The PISA measure of science self-efficacy (SCIEEFF) was derived from students' responses to questions on how easy they would find it to perform eight science tasks on their own, such as "Identify the better of two explanations for the formation of acid rain". The measure of instrumental motivation (INTMOVSCI) was derived from four items in which students rated their agreement with statements like "Many things I learn in my school science subjects will help me to get a job" (see PISA datafile description for further details).

To do this we will begin by simply plotting the two variables in SPSS:

- Select **Scatter/Dot** from the **Legacy diagnostics** available from the **Graphs** menu.
- Select Simple Scatter and click on Define to bring up the Simple Scatterplot window.
- Copy the **Conservative vote share (%)[cons_vote]** variable into the **Y Axis** box.
- Copy the **% population aged 65+[over65]** variable into the **X Axis** box.
- Click on the **OK** button.

Review Menu

EE

- Select **Bivariate...** from the **Correlate** option available from the **Analyse** menu.
- Check that the **Conservative vote share (%)[cons_vote]** and the **% population aged 65+[over65]** variables are still in the **Variables** box.

The third possible correlation is known as Kendalls Tau-b and has desirable properties when the variables take values that are ordered categories (i.e. ordinal variables). To calculate the Kendalls Tau-b we need to return to the Bivariate screen and do the following:

- Select **Bivariate...** from the **Correlate** option available from the **Analyse** menu.
- Check that the **Conservative vote share (%)[cons_vote]** and the **% population aged 65+[over65]** variables are still in the **Variables** box.
- Deselect the **Spearman** tick box

This ends our practical on correlations.

P » SPAN   29 WORDS POWERED BY TINYMCE

Download as ebook   Return to template running environment

When the user is happy with the new content, she can click on
Download as ebook …

… open it in the DEEP browser window and print off the new bespoke pdf

Example pdf v2

Introductory text replaced

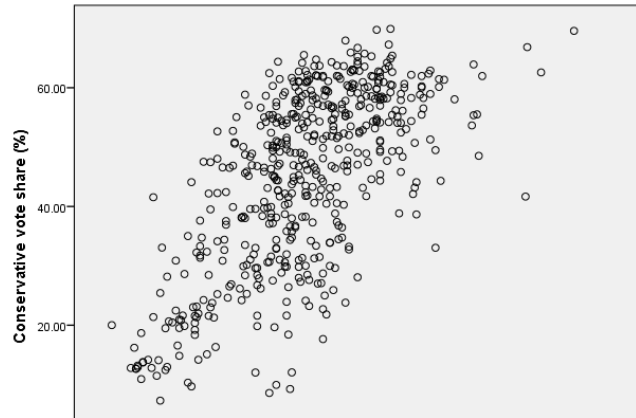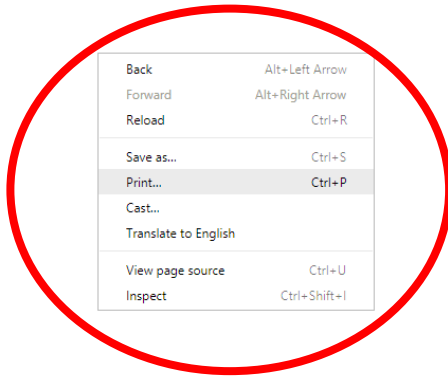## Correlations in SPSS (Practical)

Correlation practical

In this practical we will investigate whether there is a relationship between two variables by looking how correlated they are.

XXXXXXXXXXXX

To do this we will begin by simply plotting the two variables in SPSS:

- Select **Scatter/Dot** from the **Legacy diagnostics** available from the **Graphs** menu.
- Select Simple Scatter and click on Define to bring up the Simple Scatterplot window.
- Copy the **Conservative vote share (%)[cons_vote]** variable into the **Y Axis** box.
- Copy the **% population aged 65+[over65]** variable into the **X Axis** box.
- Click on the **OK** button.

Instructions incorporate new variable names and labels

SPSS will then draw a scatterplot of the two variables which can be seen below:
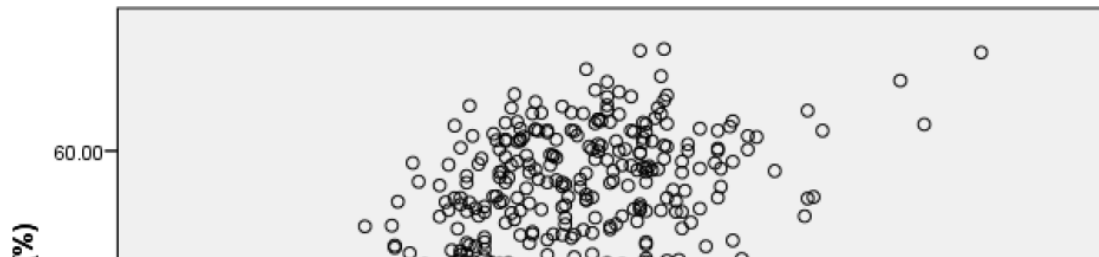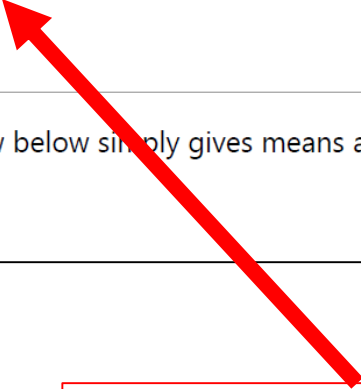
Example pdf v2

We will now finally turn our attention to the main topic of this practical which is the calculation of the correlation between our two variables. SPSS offers several correlation coefficients and we will consider these here in turn. All three are available through the Analyse->Correlate->Bivariate option in SPSS.

- Select **Bivariate...** from the **Correlate** option available from the **Analyse** menu.
- Copy the **Conservative vote share (%)[cons_vote]** and the **% population aged 65+[over65]** variables into the **Variables** box.
- Click on the **Options** button and Select the **Means and Standard deviations** tick box.
- Click on the **Continue** button to return to main window.
- Click on the **OK** button.

The correlation command will produce two output tables. The first table which we show below simply gives means and standard deviations for the two variables we are comparing.

Descriptive Statistics

Variable names and labels updated in SPSS instructions

In the next table we see the correlation matrix for the variables we are considering:

Correlations

		Conservative vote share (%)	% population aged 65+
Conservative vote share (%)	Pearson Correlation	1	.649**
	Sig. (2-tailed)		.000
	N	533	533
% population aged 65+	Pearson Correlation	.649**	1
	Sig. (2-tailed)	.000	
	N	533	533

**. Correlation is significant at the 0.01 level (2-tailed).

The Correlate option can be used for more than two variables simultaneously and will then give all correlatio

Output updated using new data

Interpretation and commentary updated on basis of statistical results

We would report the result as follows: The variables **cons_vote** and **over65** were significantly and positively correlated. r = .649, N = 533, p < .001. Note there is no need for a table when reporting a single correlation.

ALL text (in black) and SPSS objects in the body of the practical are updated automatically in seconds. The user needs only to edit the introductory text (in blue) at the beginning and end of the document

Example quiz v2

Correlations in SPSS (Quiz)

Correlation practical questions

In this practical we will investigate whether there is a relationship between two variables by looking how correlated they are.

XXXXXXXXXXXX

Firstly use SPSS to create a scatterplot of **green_vote** and **non_native** and answer the following:

- Question: What does the scatterplot say about the relationship between **green_vote** and **non_native**?

Next use SPSS and the Explore screen to create histograms, normality tests and QQ plots of **green_vote** and **non_native** and answer the following:

- Question: What do the plots and tests tell us about the normality of **green_vote**?
- Question: What do the plots and tests tell us about the normality of **non_native**?

Next use SPSS and the Correlate screen to answer the following:

- Question: What is the Pearson correlation coefficient between **green_vote** and **non_native** and is it significant?

Next use SPSS and the Correlate screen to answer the following:

Example solutions v2
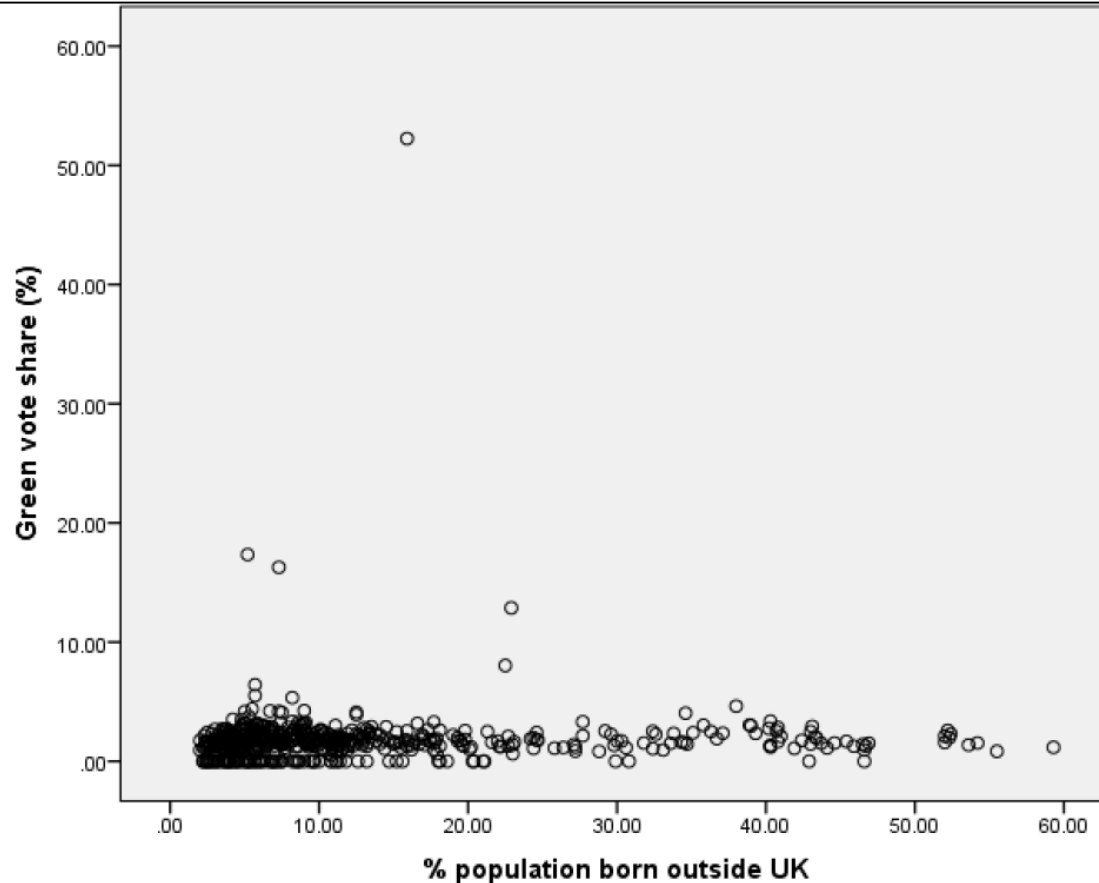
Looking at the scatterplot there appears to be a positive correlation between the variables with larger values of **green_vote** associated with larger values of **non_native** (an upward sloping relationship) but this relationship is not that strong with possibly a few more points in the bottom-left and top-right quarters of the plot.

Finding out more

- The hardcopy materials mentioned in this talk can be found at http://www.bristol.ac.uk/cmm/research/ba-teaching-ebooks/

- The teaching eBooks that can create examples with your own datasets are available at http://www.bristol.ac.uk/cmm/software/statjr/downloads/

- More general StatJR information is available at http://www.bristol.ac.uk/cmm/software/statjr/

For more information visit
www.ncrm.ac.uk