Generating SPSS training materials in Stat-JR

Hi, my name is Bill Browne, and I am co-director for the centre of multilevel modelling at Bristol, and today I'm going to talk to you about the third of a series of talks about the statistical software package Stat-JR that we develop at the multilevel centre in Bristol. In this talk we will describe the SPSS training material generation features of Stat-JR.

So if you haven't seen the other talks, Stat-JR is a software package with its own estimation engine, but which can also interoperate with other software packages. The regression template in Stat-JR for example will fit a regression model but it will fit it in a variety of statistical software packages including R, Stata, MLwiN, and also SPSS, and SPSS is a software packages often used for training in the social sciences. I want to describe here how we have extended Stat-JR's functionality to allow it to create bespoke training materials automatically for an individual lecturer using their own data.

This work was done as part of a British Academy funded grant with my two colleagues Chris Charlton and Liz Washbrook. And it builds on our work on statistical e-books and statistical analysis assistants that I have talked about in the second of this series of talks. It also came about due to the perceived shortage of quantitatively skilled social science graduates, so the British Academy put aside some money to try to tackle this problem. So, what was the plan? Well firstly, the simple brief was to produce a set of static training materials guiding students through various statistical topics that are covered in an introductory statistics course using SPSS and an example dataset. Then we would create a tool that would allow quick customization of these materials for other disciplines like the list you see behind me here. In other words, we create one set of training materials and along comes another teacher and creates another set using a different data set.

So, what does the student get? Essentially, we've produced 13 sets of exercises with for each one three components. A concept practical: it takes the student through an example and gives SPSS instructions related to that example. Complementing this is a practice component which asks different questions around the same subject area and has, as a third component, a corresponding solutions component that gives answers to those practical questions. So, what are the topics that were covered? As you can see in

the list here they range from the very basic descriptive statistics, through some of the simple statistical tests that one might use, up to more advanced topics like regression and analysis of variance. Each practical uses an extract from the PISA dataset that does international comparison in education.

So, let's look at an example. Here behind me we see a correlation practical. This is one of our 13 topics. The bulk of the practical is generated by the Stat-JR package and includes instructions for SPSS. Whilst as you'll see some of the text is in blue, and that is example specific that the user has to write. Here this text describes the PISA example that we're going to look at and puts the technique into context. Scrolling lower down this PDF we can see SPSS instructions that tell the student which variables to use in the operations to calculate the correlation coefficient. In this case it will not really trip off the tongue, but they are SCIEEFF and INSMOVSCI. These are two variables from the Pisa dataset. Next the PDF shows the output as it will appear in SPSS, and crucially we see that the PDF gives some interpretation of the size and the significance of the correlation that has been found. Here the text will depend on the specific example. So, when you come on to create your own training materials, Stat-JR will create specific text for your example. Moving on, we next look at the quiz PDF document, again for the topic of correlations. This is a separate document that asks the student to try out what they have learnt using a different example with different variables. There is then, as we show here, a corresponding solutions document. This combines the commands in SPSS that are required to create the solution, the outputs that will be shown in SPSS, and some interpretation of those outputs.

So, what is so great you may ask Aren't you just creating a set of training materials? Well let's show you how the system works if you are a teacher yourself. Let us suppose I work in political sciences and I want to create learning materials for my own training course, in my own discipline. Perhaps I have a data set on voting behaviour and I want to use it with the listed variables you see here. So, to start with do I do? Well I load up the TREE interface in Stat-JR, that's one of StatJRs three interfaces, and I load up the correlations practical eBook. These will be on my machine. Then it asks me firstly to choose a data set, and as shown here behind me, I choose the elections data set. You will have loaded your own data in if you were trying to do this yourself and you choose your data set. I next choose from the pulldown lists the two variables of interest in my practical. In this case it is the conservative vote share and, the proportion of the

constituency that are over 65. So, I'm going to construct a practical that uses that as my example. What is the correlation between conservative vote share and older constituents? Doing this as we see now in a couple of seconds will generate the practical shown behind me. All the text apart from what you see in blue is automatically generated, and the user can now type in their own example specific text to replace the blue text we have from the PISA example.

Moving on, as you can see here, the various chunks of the practical are in their own self-contained boxes. When we look at correlations, we look at different types of correlations and so each one is in its own box. And if you don't want to cover all the materials that are in our correlations practical you can simply delete some of them. When the user is happy with the content they can see they can then click on the download as eBook button on the bottom of the screen and an eBook will be created. This will allow them to save the eBook as a file and then you can move to the Stat-JR Deep eBook interface, the second of Stat-JR's interfaces and we can see the browser as shown behind me. Clicking with the right mouse button will allow us to print what we see on the screen as a PDF and hey presto, we have a hard copy of our new practical. Looking at the new PDF, as we see behind me, we would replace the blue X's with our new text, but you can clearly see the SPSS instructions refer to the new variables that we've incorporated. If we move further down, here again we see the variable names and their labels are updated throughout the SPSS instructions, and further still we can see that the SPSS output is updated, as is the interpretation that we see, on the commentary given.

So, to summarize, all the text and objects update automatically in seconds. All the user has to do is to write a new introduction to the text. Once again, Stat-JR is a computer package. It can't do everything that humans do, so context that's very important the social science has to be input by a human.

Moving on, here is an example of a generated quiz PDF using the election data, and you can see here we're going to look at the Green share of the vote against the non-native proportion in each constituency. So, it might be another interesting political example. Here behind, we can see the corresponding outputs for the solutions to that quiz. And really this kind of illustrates this diagram you can see behind me, some of the limitations of the system. So Stat-JR is a computer program, it's not a statistician. And so, you give it an example, and if it isn't the best it's a case of a bit of

garbage in, and garbage out. So, what you can see behind me is an illustration of an outlier. So, this is the Green vote, and you can see Brighton on there. If you look, one X is very different from the rest, and this is really influencing our conclusions. But maybe this would be a good example to show students correlations and how they don't always work the same way for every different data set.

Okay, so that concludes my very quick description of the SPSS features in Stat-JR. To recap, we have 13 different topics, and if you have a set of data sets in a matter of seconds you can construct training materials using your data sets. What I have included behind are a series of links that have more information.

Thank you very much for listening, have a nice day.