EXERCISE 21

Multiple regression

Before you start

The reader should study Section 12.4 before proceeding with this Exercise.

A problem in reading research

Reading comprises many different component skills. A reading researcher hypothesises that certain specific kinds of pre-reading abilities and behaviour can predict later progress in reading, as measured by performance on reading tests taken some years after the child's first formal lessons. Let us, therefore, label the dependent variable (DV) in this study *Progress*. While they are still very young indeed, many children show a considerable grasp of English syntax in their speech. Our researcher devises a measure of their syntactic knowledge, *Syntax*, based upon the average length of their uttered sentences. Some researchers, however, argue that an infant's prelinguistic babbling (which we shall label *Vocal*) also plays a key role in their later reading performance. At the pre-reading stage, some very young children can acquire a sight vocabulary of several hundreds of words. The ability to pronounce these words on seeing them written down is known as logographic reading; but many authorities do not accept that this is true reading. Our researcher, who views the logographic strategy as important, includes a measure of this skill, *Logo*, in the study.

Preparing the data set

Fifty children are studied over a period beginning in infancy and extending through their school years. Their scores on the four measures, the DV *Progress* (*P*), and the three IVs *Logo* (*L*), *Vocal* (*V*) and *Syntax* (*S*), are listed in the appendix to this Exercise. Since it would be very laborious for you to type in all the data during the exercise, we must hope that you already have them available in an accessible file, with a name such as ***Reading***. The data are also available on the Internet as ***Ex21 Reading data for multiple regression*** at: http://www.psypress.com/spss-made-simple/datasets.asp

Exploring the data

The distributions of the variables are most easily explored by using the boxplots option in **Explore**. Select **Analyze➔Descriptive Statistics➔Explore…** to open the **Explore** dialog box. Transfer the four variable names to the **Dependent List** panel and click the **Plots** radio button**.** Then click the **Plots…** button and click the **Dependents together** radio button in the **Boxplots** panel. Click off the **Stem-and-leaf** check box in the **Descriptive** panel. Click **Continue** and then **OK**. This will plot four boxplots side-by-side for easy comparison.

Regression is most effective when each IV is strongly correlated with the DV but uncorrelated with the other IVs. Although the correlation matrix can be listed from within the regression procedure, it is often more useful to scrutinise the matrix before proceeding with a regression analysis in order to make judgements about which variables might be retained and which dropped from the analysis. For example, it might be advisable to make a choice between two variables that are highly correlated with one another.

Use the **Bivariate Correlations** procedure to compute the correlation matrix. The same procedure is also useful for tabulating the means and standard deviations, which are available as an option. After transferring the variable names to the **Variables** box, click **Options** and (within the **Statistics** choice box) select **Means and standard deviations**. Click **Continue** and **OK**. Notice that the DV *Progress* shows substantial correlations with both *Logo* and *Syntax*. On the other hand, there is no appreciable correlation between *Logo* and *Syntax*. The remaining variable (*Vocal*) shows little association with any of the other variables, although there is a hint of a negative correlation with *logo*.

Running the multiple regression analysis

Run the multiple regression of *Progress* upon the three predictors, by following the procedure in Section 12.4. Remember that the **Dependent** variable is what you are predicting (*Progress*) and the **Independent** variables are the predictors (*Logo, Vocal, Syntax*). On the first run, use the **Method *Enter*** (this enters all the variables simultaneously) and on the second run the **Method *Stepwise*** (this is a forward stepwise selection procedure). One outlier is identified in **Casewise Diagnostics** – run both *Enter* and *Stepwise* again with that case deselected to see whether the results differ very much.

Output for the multiple regression

The main features of a multiple regression output, both for the simultaneous and stepwise methods, are explained in Section 12.4.

- **Do the decisions of the multiple regression procedure about which variables are important agree with your informal observations during the exploratory phase of the data analysis?**
- **Compare the values of R Square for the analyses with and without the outlier.**
- **Write out the regression equation that you would use to predict progress from a participant's scores on Logo, Vocal and Syntax for the analysis without the outlier and using the Stepwise method. Use this equation to predict a Progress score for a child scoring 50 on each of the predictor variables.**
- **You can check your calculated result by requesting the computer to do the calculation for you. Within Data Editor, enter 51 into the empty 51st row Case cell and 50 into each of the Logo, Vocal and Syntax cells. Leave the Progress cell empty. Within the Linear Regression dialog box, click Save… and click the checkbox for Unstandardized followed by Continue. Run the analysis and then inspect the data file where a new variable PRE_1 will have appeared with predicted values for Progress. Does the value for the 51st row match your calculated value?**

Finishing the session

Close down SPSS and any other windows before logging out of the computer.

Appendix to Exercise 21 - Reading data

P	L	V	S	P	L	V	S	P	L	V	S	P	L	V	S
65	75	34	48	46	55	75	32	65	50	75	68	34	32	42	27
58	29	18	67	51	31	50	66	71	65	23	64	54	64	55	32
42	40	43	38	61	69	59	46	60	56	52	44	81	82	60	69
55	55	9	48	45	19	71	59	17	10	64	20	77	66	50	79
68	81	41	54	53	48	44	45	55	41	41	55	57	30	20	54
59	28	72	68	46	45	29	45	69	51	14	62	80	82	65	58
50	39	31	42	25	28	58	28	47	49	46	59	89	51	52	48
50	26	78	56	71	70	51	54	53	14	53	77	50	34	45	60
71	84	46	50	30	55	42	25	50	40	51	31	69	49	72	72
65	71	30	52	62	53	52	57	80	45	59	90	71	69	57	60
34	30	30	20	47	20	78	69	51	18	22	61	39	25	81	49
44	71	79	22	60	46	80	67	79	58	13	82				
47	62	26	30	70	66	40	61	51	43	31	50				