

# Multiple Regression in SPSS worksheet (Quiz)



## Multiple Regression practical questions

In this example you will explore the association of two further indicators of parental resources with the PISA science test score. The variable **HEDRES** measures *home educational resources* and is derived from indicators of ownership of seven items, such as a desk to study at and books to help with schoolwork. The variable **CULTPOSS** measures *cultural possessions at home* and is derived from five ownership indicators such as books of poetry and works of art (see the PISA datafile description for more details).

Use SPSS to fit a regression of **Science test score[SCISCORE]** on the single predictor **Home educational resources score[HEDRES]** and answer the following questions:

- Question: Looking at the model summary table what proportion of the variance in the response is explained by the regression?
- Question: Looking at the Coefficients table what is the predicted value of **SCISCORE** when **HEDRES** = 0?
- Question: Looking at the Coefficients table how much is a one unit increase of **HEDRES** predicted to change **SCISCORE**?
- Question: What is the standardised slope coefficient and what does it mean?
- Question: Is the slope significant?
- Question: What is the 95% CI for the intercept?
- Question: What is the 95% CI for the slope?

Use SPSS to fit a regression of **Science test score[SCISCORE]** on **Home cultural possessions score[CULTPOSS]** and answer the following questions:

- Question: Looking at the model summary table what proportion of the variance in the response is explained by the regression?
- Question: Looking at the Coefficients table what is the predicted value of **SCISCORE** when **CULTPOSS** = 0?
- Question: Looking at the Coefficients table how much is a one unit increase of **CULTPOSS** predicted to change **SCISCORE**?
- Question: What is the standardised slope coefficient and what does it mean?
- Question: Is the slope significant?
- Question: What is the 95% CI for the intercept?
- Question: What is the 95% CI for the slope?

Use SPSS to fit a regression of **Science test score[SCISCORE]** on both **Home educational resources score[HEDRES]** and **Home cultural possessions score[CULTPOSS]** and answer the following questions:

- Question: Looking at the model summary table what proportion of the variance in the response is explained by the regression?
- Question: How does the adjusted R squared compare with the two simpler regressions?
- Question: Looking at the Coefficients table what is the value of the intercept?
- Question: Looking at the Coefficients table what is the value of the slope for **HEDRES**?
- Question: Looking at the Coefficients table what is the value of the slope for **CULTPOSS**?
- Question: What is the standardised slope coefficient for **HEDRES** and what does it mean?
- Question: What is the standardised slope coefficient for **CULTPOSS** and what does it mean?
- Question: Is the slope for **HEDRES** significant?
- Question: Is the slope for **CULTPOSS** significant?
- Question: What is the 95% CI for the intercept?
- Question: What is the 95% CI for the slope for **HEDRES**?
- Question: What is the 95% CI for the slope for **CULTPOSS**?

Use SPSS to build up the regression models by fitting first a regression of **Science test score[SCISCORE]** on **Home educational resources score[HEDRES]** and then in block 2 both **Home educational resources score[HEDRES]** and **Home cultural possessions score[CULTPOSS]** and answer the following questions: Make sure you tick the box for Standardized found under Save... -> Residuals.:

- Question: What do the three output tables show?

Use Histogram from the Legacy diagnostics on the residuals from the last model to answer the following question:

- Question: Looking at the histogram do the residuals look normally distributed?

Use Scatter from the Legacy diagnostics to look at the relationship between the residuals from the last model and **HEDRES**.

- Question: Is there any specific pattern in the plot against **HEDRES**?

Use Scatter from the Legacy diagnostics to look at the relationship between the residuals from the last model and **CULTPOSS**.

- Question: Is there any specific pattern in the plot against **CULTPOSS**?

## Solutions to Multiple Regression practical questions

The SPSS instructions for the first simple regression are as follows:

- Select **Linear** from the **Regression** submenu available from the **Analyze** menu.
- Copy the **Science test score[SCISCORE]** variable into the **Dependent** box.
- Copy the **Home educational resources score[HEIRES]** variable into the **Independent(s)** box.
- Click on the **Statistics** button.
- On the screen appears add the tick for **Confidence Interval** to those for **Estimates** and **Model fit**.
- Click on the **Continue** button to return to the main window.
- Click on the **OK** button to run the command.

Question: Looking at the model summary table what proportion of the variance in the response is explained by the regression?

Solution: The output from SPSS is as follows:

**Model Summary**

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.227 <sup>a</sup>	.051	.051	99.10189

a. Predictors: (Constant), Home educational resources score

The R squared provides the proportion of the variance explained by the regression and has the value .051.

- Question: Looking at the Coefficients table what is the predicted value of **SCISCORE** when **HEIRES** = 0?
- Question: Looking at the Coefficients table how much is a one unit increase of **HEIRES** predicted to change **SCISCORE**?
- Question: What is the standardised slope coefficient and what does it mean?
- Question: Is the slope significant?
- Question: What is the 95% CI for the intercept?
- Question: What is the 95% CI for the slope?

Solutions: The output from SPSS is as follows:

**Coefficients**

Model		Unstandardized Coefficients		Standardized Coefficients		95.0% Confidence Interval for B		
		B	Std. Error	Beta	t	Sig.	Lower Bound	Upper Bound
1	(Constant)	523.882	1.403		373.505	.000	521.133	526.632
	Home educational resources score	22.733	1.375	.227	16.533	.000	20.037	25.428

The answers are as follows:

- The predicted value of **SCISCORE** when **HEIRES** = 0 is 523.882.
- A one unit increase of **HEIRES** is predicted to change **SCISCORE** by 22.733.
- The standardised slope coefficient takes value .227 which represents the predicted change in **SCISCORE** for an increase of 1 standard deviation in **HEIRES**.
- The p value (quoted under Sig.) is .000 (reported as  $p < .001$ ) which is less than 0.05. We therefore have significant evidence to reject the null hypothesis that the slope is zero.
- The intercept confidence interval is 521.133 to 526.632.
- The **HEIRES** slope confidence interval is 20.037 to 25.428.

The SPSS instructions for the second simple regression are as follows:

- Select **Linear** from the **Regression** submenu available from the **Analyze** menu.
- Remove the **Home educational resources score[HEIRES]** variable from the **Independent(s)** box.
- Copy the **Home cultural possessions score[CULTPOSS]** variable into the **Independent(s)** box.
- The other options will be remembered from last time.
- Click on the **OK** button to run the command.

Question: Looking at the model summary table what proportion of the variance in the response is explained by the regression?

Solution: The output from SPSS is as follows:

**Model Summary**

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.227 <sup>a</sup>	.051	.051	99.10189

a. Predictors: (Constant), Home educational resources score

The R squared provides the proportion of the variance explained by the regression and has the value .051.

- Question: Looking at the Coefficients table what is the predicted value of **SCISCORE** when **CULTPOSS** = 0?
- Question: Looking at the Coefficients table how much is a one unit increase of **CULTPOSS** predicted to change **SCISCORE**?
- Question: What is the standardised slope coefficient and what does it mean?

- Question: Is the slope significant?
- Question: What is the 95% CI for the intercept?
- Question: What is the 95% CI for the slope?

Solutions: The output from SPSS is as follows:

		Coefficients						
		Unstandardized Coefficients		Standardized Coefficients			95.0% Confidence Interval for B	
Model		B	Std. Error	Beta	t	Sig.	Lower Bound	Upper Bound
1	(Constant)	523.882	1.403		373.505	.000	521.133	526.632
	Home educational resources score	22.733	1.375	.227	16.533	.000	20.037	25.428

The answers are as follows:

- The predicted value of **SCISCORE** when **CULTPOSS** = 0 is 523.882.
- A one unit increase of **CULTPOSS** is predicted to change **SCISCORE** by 22.733.
- The standardised slope coefficient takes value .227 which represents the predicted change in **SCISCORE** for an increase of 1 standard deviation in **CULTPOSS**.
- The p value (quoted under Sig.) is .000 (reported as  $p < .001$ ) which is less than 0.05. We therefore have significant evidence to reject the null hypothesis that the slope is zero.
- The intercept confidence interval is 521.133 to 526.632.
- The **CULTPOSS** slope confidence interval is 20.037 to 25.428.

The SPSS instructions for the multiple regression are as follows:

- Select **Linear** from the **Regression** submenu available from the **Analyze** menu.
- Copy the **Home educational resources score[HEIRES]** variable into the **Independent(s)** box to join **Home cultural possessions score[CULTPOSS]**.
- The other options will be remembered from last time.
- Click on the **OK** button to run the command.

- Question: Looking at the model summary table what proportion of the variance in the response is explained by the regression?
- Question: How does the adjusted R squared compare with the two simpler regressions?

Solutions: The output from SPSS is as follows:

Model Summary				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.301 <sup>a</sup>	.091	.090	96.77982

a. Predictors: (Constant), Home educational resources score, Home cultural possessions score

The answers are as follows:

- The R squared provides the proportion of the variance explained by the regression and has the value .091.
- The adjusted R square measure takes value .090 which we can compare with .051 for just **HEIRES** and .074 for just **CULTPOSS**.
- Question: Looking at the Coefficients table what is the value of the intercept?
- Question: Looking at the Coefficients table what is the value of the slope for **HEIRES**?
- Question: Looking at the Coefficients table what is the value of the slope for **CULTPOSS**?
- Question: What is the standardised slope coefficient for **HEIRES** and what does it mean?
- Question: What is the standardised slope coefficient for **CULTPOSS** and what does it mean?
- Question: Is the slope for **HEIRES** significant?
- Question: Is the slope for **CULTPOSS** significant?
- Question: What is the 95% CI for the intercept?
- Question: What is the 95% CI for the slope for **HEIRES**?
- Question: What is the 95% CI for the slope for **CULTPOSS**?

Solutions: The output from SPSS is as follows:

		Coefficients						
		Unstandardized Coefficients		Standardized Coefficients			95.0% Confidence Interval for B	
Model		B	Std. Error	Beta	t	Sig.	Lower Bound	Upper Bound
1	(Constant)	528.024	1.394		378.701	.000	525.291	530.758
	Home cultural possessions score	20.497	1.429	.214	14.340	.000	17.695	23.299
	Home educational resources score	14.114	1.481	.142	9.531	.000	11.211	17.017

The answers are as follows:

- The regression intercept takes value 528.024.
- The regression slope for **HEIRES** takes value 20.497.
- The regression slope for **CULTPOSS** takes value 14.114.
- The standardised slope coefficient for **HEIRES** takes value .214 which represents the predicted change in **SCISCORE** for an increase of 1 standard deviation in **HEIRES**.
- The standardised slope coefficient for **CULTPOSS** takes value .142 which represents the predicted change in **SCISCORE** for an increase of 1 standard deviation in **CULTPOSS**.
- The p value (quoted under Sig.) is .000 (reported as  $p < .001$ ) which is less than 0.05. We therefore have significant evidence to reject the null hypothesis that the slope for **HEIRES** is zero.
- The p value (quoted under Sig.) is .000 (reported as  $p < .001$ ) which is less than 0.05. We therefore have significant evidence to reject the null hypothesis that the slope for **CULTPOSS** is zero.

- The confidence interval for the intercept is 525.291 to 530.758.
- The confidence interval for the slope for **HEDRES** is 17.695 to 23.299.
- The confidence interval for the slope for **CULTPOSS** is 11.211 to 17.017.

The SPSS instructions to build the regression in blocks are as follows:

- Select **Linear** from the **Regression** submenu available from the **Analyze** menu.
- Remove the **Home cultural possessions score[CULTPOSS]** variable from the **Independent(s)** box to leave just **Home educational resources score[HEDRES]**.
- Click the **Next** button.
- Copy the **Home cultural possessions score[CULTPOSS]** variable into the now empty **Independent(s)** box.
- Click on the **Save** button.
- On the screen appears select the tick for **Standardized** found under **Residuals**.
- Click on the **Continue** button to return to the main window.
- Click on the **OK** button to run the command.

- Question: What do the three output tables show?

Solutions: The output from SPSS is as follows:

#### Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.229 <sup>a</sup>	.053	.052	98.77069
2	.301 <sup>b</sup>	.091	.090	96.77982

a. Predictors: (Constant), Home educational resources score

b. Predictors: (Constant), Home educational resources score, Home cultural possessions score

Here we see the model summaries for the first and third regression models earlier i.e. we fit a model with just **HEDRES** and then a second model where we introduce **CULTPOSS**.

#### Coefficients

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	525.710	1.413		371.942	.000	522.939	528.481
	Home educational resources score	22.799	1.379	.229	16.532	.000	20.095	25.502
2	(Constant)	528.024	1.394		378.701	.000	525.291	530.758
	Home educational resources score	14.114	1.481	.142	9.531	.000	11.211	17.017
	Home cultural possessions score	20.497	1.429	.214	14.340	.000	17.695	23.299

Similarly we have the model coefficients for the first and third models from earlier in one combined table.

#### Residuals Statistics

	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	432.7281	593.6044	527.8599	30.53889	4925
Residual	-358.69009	305.37387	.00000	96.76016	4925
Std. Predicted Value	-3.115	2.153	.000	1.000	4925
Std. Residual	-3.706	3.155	.000	1.000	4925

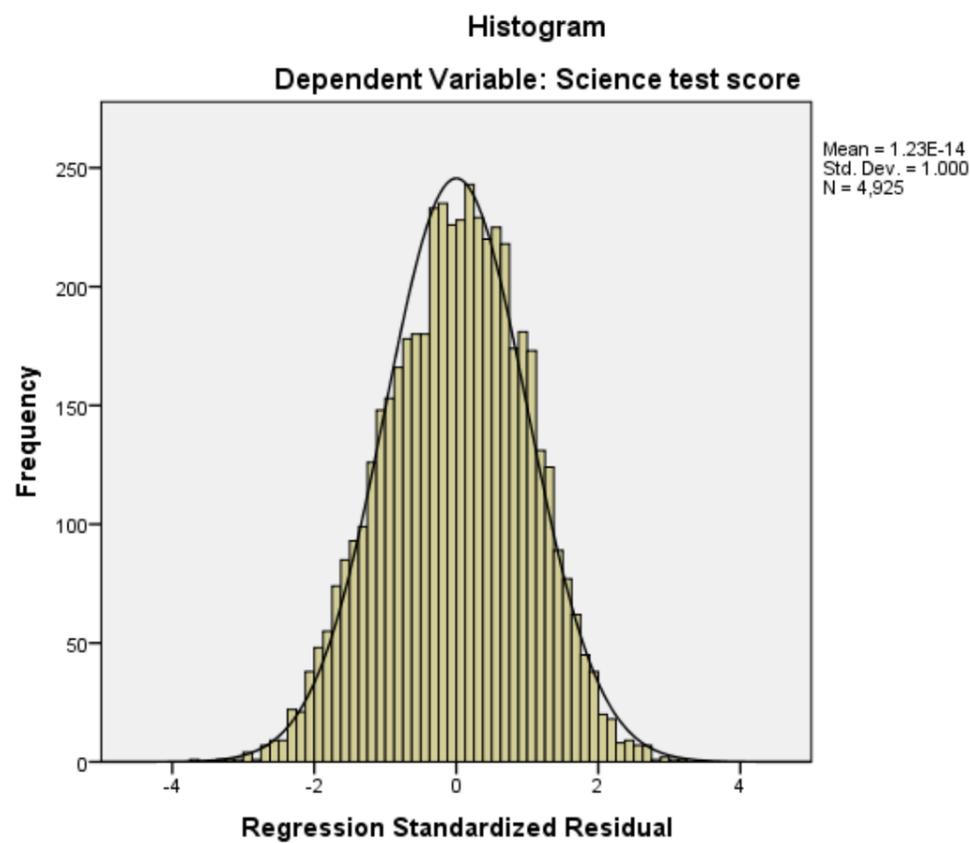
This table just summarises the predictions and residuals that come out of the final regression and it is perhaps easier to look at these via plots.

The SPSS instructions for the histogram are as follows:

- Select **Histogram** from the **Legacy diagnostics** available from the **Graphs** menu.
- Copy the **Standardized Residual [ZRE\_1]** variable into the **Variable** box.
- Click on the **Display normal curve** tick box.
- Click on the **OK** button.

- Question: Looking at the histogram do the residuals look normally distributed?

Solution: The output from SPSS is as follows:



We hope to see the histogram of residuals roughly following the shape of the normal curve that is superimposed over them. The SPSS instructions for the first scatter plot are as follows:

- 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 **Standardized Residual [ZRE\_1]** variable into the **Y Axis** box.
- Copy the **Home educational resources score[HEDRES]** variable into the **X Axis** box.
- Click on the **OK** button.

- Question: Is there any specific pattern in the plot against **HEDRES**?

The output from SPSS is as follows:



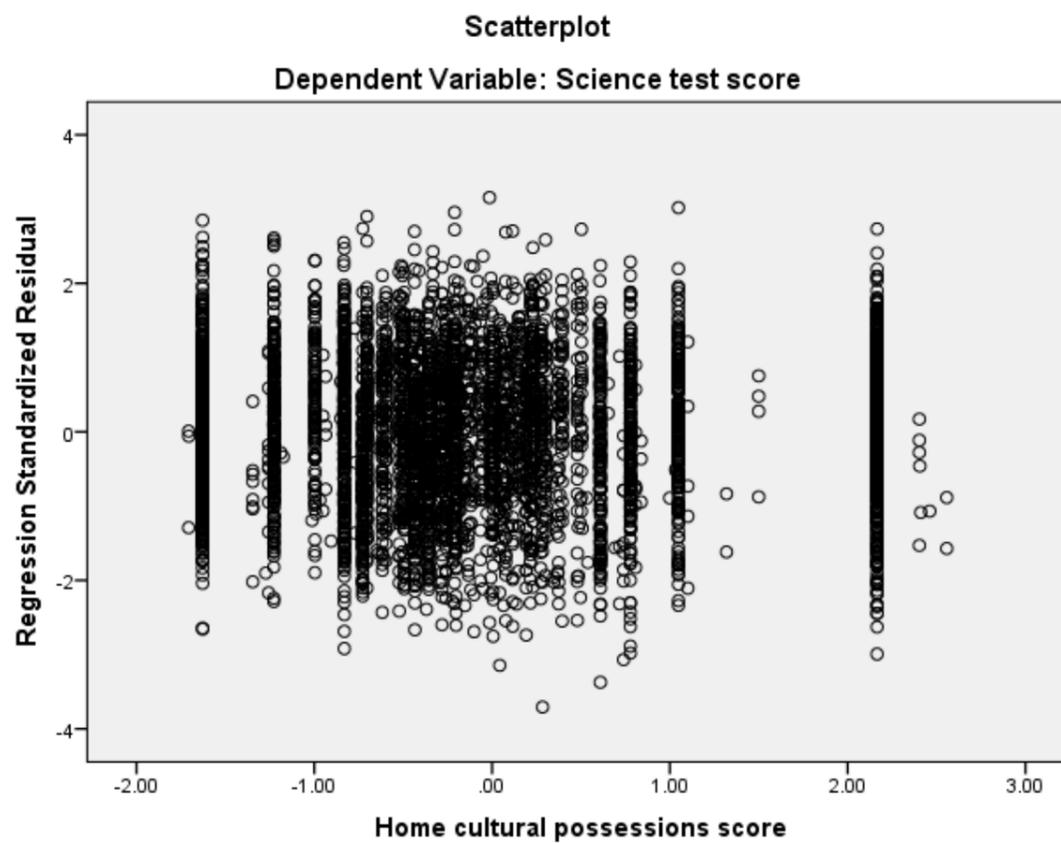
We hope that the residuals show a random scatter when plotted against the predictor variable and also that their variability is constant across different values of the predictor variable.

The SPSS instructions for the second scatter plot are as follows:

- 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
- Remove the **Home educational resources score[HEDRES]** variable from the **X Axis** box.
- Copy the **Home cultural possessions score[CULTPOSS]** variable into the **X Axis** box.
- Click on the **OK** button.

- Question: Is there any specific pattern in the plot against **CULTPOSS**?

The output from SPSS is as follows:



We hope that the residuals show a random scatter when plotted against the predictor variable and also that their variability is constant across different values of the predictor variable.

Here we have seen that home educational resources and ownership of cultural possessions are both significantly associated with a student's science achievement, and that each variable can independently predict the science test score when the other variable is held constant. The conditional effect of cultural possessions is slightly stronger than that of home educational resources (a conclusion that comes from comparing the standardized betas from the multiple regression) and, in total, they can account for around 9 percent of the overall variation in test scores, considerably more than the 2 percent accounted for by general wealth and parental emotional support that we saw in the practical example.