Multiple Regression in SPSS worksheet (Quiz)

The development of this E-Book has been supported by the British Academy
This implementation is by National Centre for Research Methods and UK Data Service

Multiple Regression practical questions

The dataset we are using is an excerpt from a cut-down dataset drawn from the Living Costs and Food Survey 2013, available from the UK Data Service: http://doi.org/10.5255/UKDA-SN-7932-2. We will explore how household income can be used to predict household expenditure, and whether there is any additional effect of household size. Both income and expenditure are measured in pounds per week.

No conditions are required to use the data; however respondents are promised that their data will be kept confidential. As a result high values are grouped together to prevent households being identified by their large household expenditures or unusually high expenditure. This protects respondents, but it also affects the quality of the results produced in this workbook. Users who wish to use better quality data are encouraged to explore the full data from the Living Costs and Food Survey which is available through the UK Data Service (http://doi.org/10.5255/UKDA-SN-7702-1), for which users need to register and adhere to some conditions of use.

Use SPSS to fit a regression of **Total expenditure (top coded, formerly P550tpr)[expenditure]** on the single predictor **income** 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 **expenditure** when **income** = 0?
- Question: Looking at the Coefficients table how much is a one unit increase of **income** predicted to change **expenditure**?
- 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 **Total expenditure (top coded, formerly P550tpr)[expenditure]** on **Household size, number of people in household (recoded) formerly A049r[hhsize]** 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 **expenditure** when **hhsize** = 0?
- Question: Looking at the Coefficients table how much is a one unit increase of **hhsize** predicted to change **expenditure**?
- 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 **Total expenditure (top coded, formerly P550tpr)[expenditure]** on both **income** and **Household size, number of people in household (recoded) formerly A049r[hhsize]** 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 **income**?
- Question: Looking at the Coefficients table what is the value of the slope for **hhsize**?
- Question: What is the standardised slope coefficient for **income** and what does it mean?
- Question: What is the standardised slope coefficient for **hhsize** and what does it mean?
- Question: Is the slope for **income** significant?
- Question: Is the slope for **hhsize** significant?
- Question: What is the 95% CI for the intercept?
- Question: What is the 95% CI for the slope for **income**?
- Question: What is the 95% CI for the slope for **hhsize**?

Use SPSS to build up the regression models by fitting first a regression of **Total expenditure (top coded, formerly P550tpr) [expenditure]** on **income** and then in block 2 both **income** and **Household size, number of people in household (recoded) formerly A049r[hhsize]** 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 **income**.

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

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

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

Solutions to Multiple Regression practical questions

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

1. Select **Linear** from the **Regression** submenu available from the **Analyze** menu.
2. Copy the **Total expenditure (top coded, formerly P550tpr)[expenditure]** variable into the **Dependent** box.
3. Copy the **income** variable into the **Independent(s)** box.
4. Click on the **Statistics** button.
5. On the screen appears add the tick for **Confidence Interval** to those for **Estimates** and **Model fit**.
6. Click on the **Continue** button to return to the main window.
7. 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	.706 a	.499	.499	206.94574

a. Predictors: (Constant), income

The R squared provides the proportion of the variance explained by the regression and has the value .499.

- Question: Looking at the Coefficients table what is the predicted value of **expenditure** when **income** = 0?
- Question: Looking at the Coefficients table how much is a one unit increase of **income** predicted to change **expenditure**?
- 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 B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.	95.0% Confidence Interval for B Lower Bound	95.0% Confidence Interval for B Upper Bound
1	(Constant)	122.963	5.760		21.349	.000	111.672	134.255
	income	.575	.008	.706	71.574	.000	.559	.591

The answers are as follows:

- The predicted value of **expenditure** when **income** = 0 is 122.963.
- A one unit increase of **income** is predicted to change **expenditure** by .575. In other words, the model predicts that for each extra pound in income households have they will spend an extra 58 pence.
- The standardised slope coefficient takes value .706 which represents the predicted change in **expenditure** for an increase of 1 standard deviation in **income**.
- 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 111.672 to 134.255.
- The **income** slope confidence interval is .559 to .591.

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

1. Select **Linear** from the **Regression** submenu available from the **Analyze** menu.
2. Remove the **income** variable from the **Independent(s)** box.
3. Copy the **Household size, number of people in household (recoded)formerly A049r[hhsize]** variable into the **Independent(s)** box.
4. The other options will be remembered from last time.
5. 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	.449 [a]	.202	.202	261.20871

a. Predictors: (Constant), Household size, number of people in household (recoded)formerly A049r

The R squared provides the proportion of the variance explained by the regression and has the value .202.

- Question: Looking at the Coefficients table what is the predicted value of **expenditure** when **hhsize** = 0?
- Question: Looking at the Coefficients table how much is a one unit increase of **hhsize** predicted to change **expenditure**?
- 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	t	Sig.	95.0% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	223.903	7.974		28.081	.000	208.271	239.534
	Household size, number of people in household (recoded)formerly A049r	109.878	3.046	.449	36.070	.000	103.906	115.850

The answers are as follows:

- The predicted value of **expenditure** when **hhsize** = 0 is 223.903.
- A one unit increase of **hhsize** is predicted to change **expenditure** by 109.878. Accordingly, the model predicts that households spend an extra £110 each week for every extra person in residence.
- The standardised slope coefficient takes value .449 which represents the predicted change in **expenditure** for an increase of 1 standard deviation in **hhsize**.
- 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 208.271 to 239.534. The **hhsize** slope confidence interval is 103.906 to 115.850.

The SPSS instructions for the multiple regression are as follows:

1. Select **Linear** from the **Regression** submenu available from the **Analyze** menu.
2. Copy the **income** variable into the **Independent(s)** box to join **Household size, number of people in household (recoded) formerly A049r[hhsize]**.
3. The other options will be remembered from last time.
4. 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	.721 [a]	.520	.520	202.61834

a. Predictors: (Constant), income, Household size, number of people in household (recoded)formerly A049r

The answers are as follows:

- The R squared provides the proportion of the variance explained by the regression and has the value .520.
- The adjusted R square measure takes value .520 which we can compare with .499 for just **income** and .202 for just **hhsize**.
- 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 **income**?
- Question: Looking at the Coefficients table what is the value of the slope for **hhsize**?
- Question: What is the standardised slope coefficient for **income** and what does it mean?
- Question: What is the standardised slope coefficient for **hhsize** and what does it mean?
- Question: Is the slope for **income** significant?
- Question: Is the slope for **hhsize** significant?
- Question: What is the 95% CI for the intercept?
- Question: What is the 95% CI for the slope for **income**?
- Question: What is the 95% CI for the slope for **hhsize**?

Solutions: The output from SPSS is as follows:

Coefficients

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	67.891	6.738		10.075	.000	54.681	81.101
	Household size, number of people in household (recoded)formerly A049r	39.604	2.652	.162	14.933	.000	34.405	44.803
	Income	.515	.009	.633	58.350	.000	.498	.533

The answers are as follows:

- The regression intercept takes value 67.891.
- The regression slope for **household size** takes value 39.604.
- The regression slope for **income** takes value .515.
- The standardised slope coefficient for hhsize takes value .162 which represents the predicted change in expenditure for an increase of 1 standard deviation in **income**.
- The standardised slope coefficient for income takes value .633 which represents the predicted change in expenditure for an increase of 1 standard deviation in **hhsize**.
- 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 **income** 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 **hhsize** is zero.
- The confidence interval for the intercept is 54.681 to 81.101.
- The confidence interval for the slope for **hhsize** is 34.405 to 44.803.
- The confidence interval for the slope for **income** is .498 to .533.

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

1. Select **Linear** from the **Regression** submenu available from the **Analyze** menu.
2. Remove the **Household size, number of people in household (recoded) formerly A049r[hhsize]** variable from the **Independent(s)** box to leave just **income**.
3. Click the **Next** button.
4. Copy the **Household size, number of people in household (recoded) formerly A049r[hhsize]** variable into the now empty **Independent(s)** box.
5. Click on the **Save** button.
6. On the screen appears select the tick for **Standardized** found under **Residuals**.
7. Click on the **Continue** button to return to the main window.
8. 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	.706 [a]	.499	.499	206.94574
2	.721 [b]	.520	.520	202.61834

a. Predictors: (Constant), income
b. Predictors: (Constant), income, Household size, number of people in household (recoded)formerly A049r

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

Coefficients

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	122.963	5.760		21.349	.000	111.672	134.255
	income	.575	.008	.706	71.574	.000	.559	.591
2	(Constant)	67.891	6.738		10.075	.000	54.681	81.101
	income	.515	.009	.633	58.350	.000	.498	.533
	Household size, number of people in household (recoded)formerly A049r	39.604	2.652	.162	14.933	.000	34.405	44.803

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	107.4948	876.4191	479.7584	210.80607	5144
Residual	-575.06085	967.04437	.00000	202.57893	5144
Std. Predicted Value	-1.766	1.882	.000	1.000	5144
Std. Residual	-2.838	4.773	.000	1.000	5144

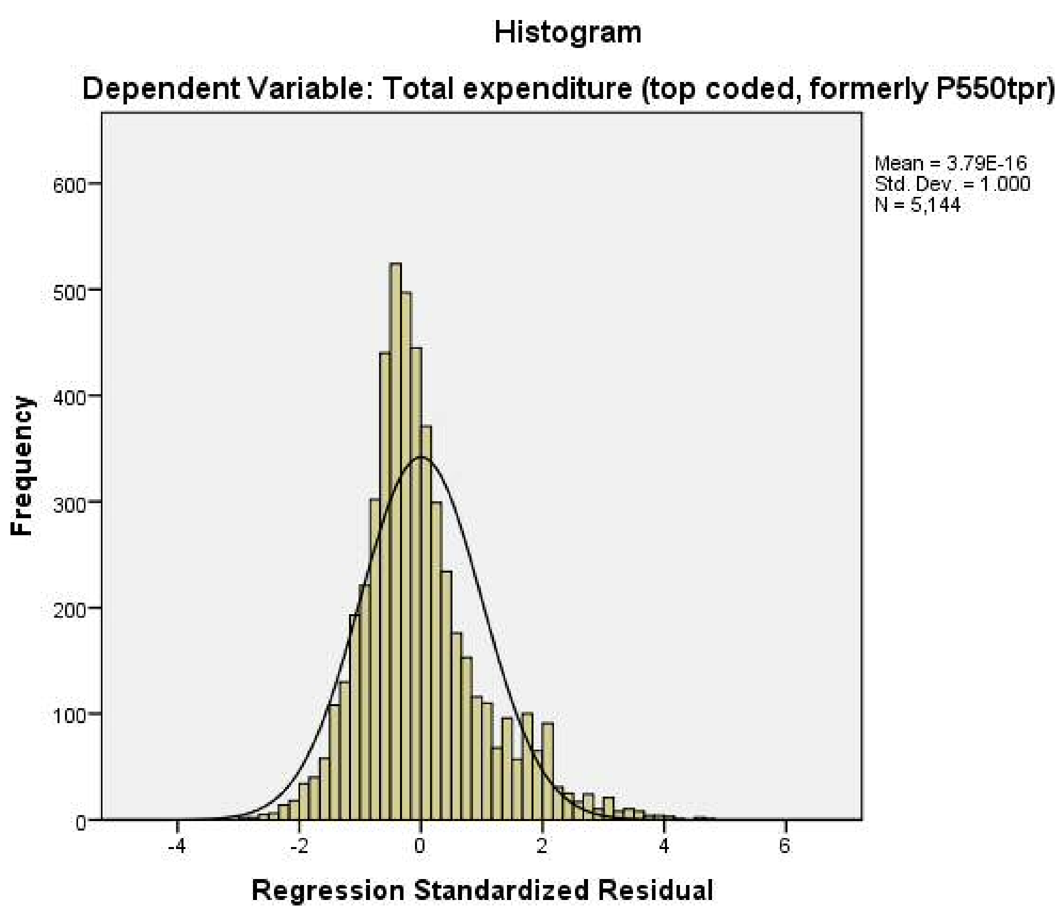
This table just summarises the predictions and residuals that come out of the final regression and it is perhaps easier to look at these via plots.

The SPSS instructions for the histogram are as follows:

1. Select **Histogram** from the **Legacy Dialogs** available from the **Graphs** menu.
2. Copy the **Standardized Residual [ZRE_1]** variable into the **Variable** box.
3. Click on the **Display normal curve** tick box.
4. Click on the **OK** button.

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

Solution: The output from SPSS is as follows:

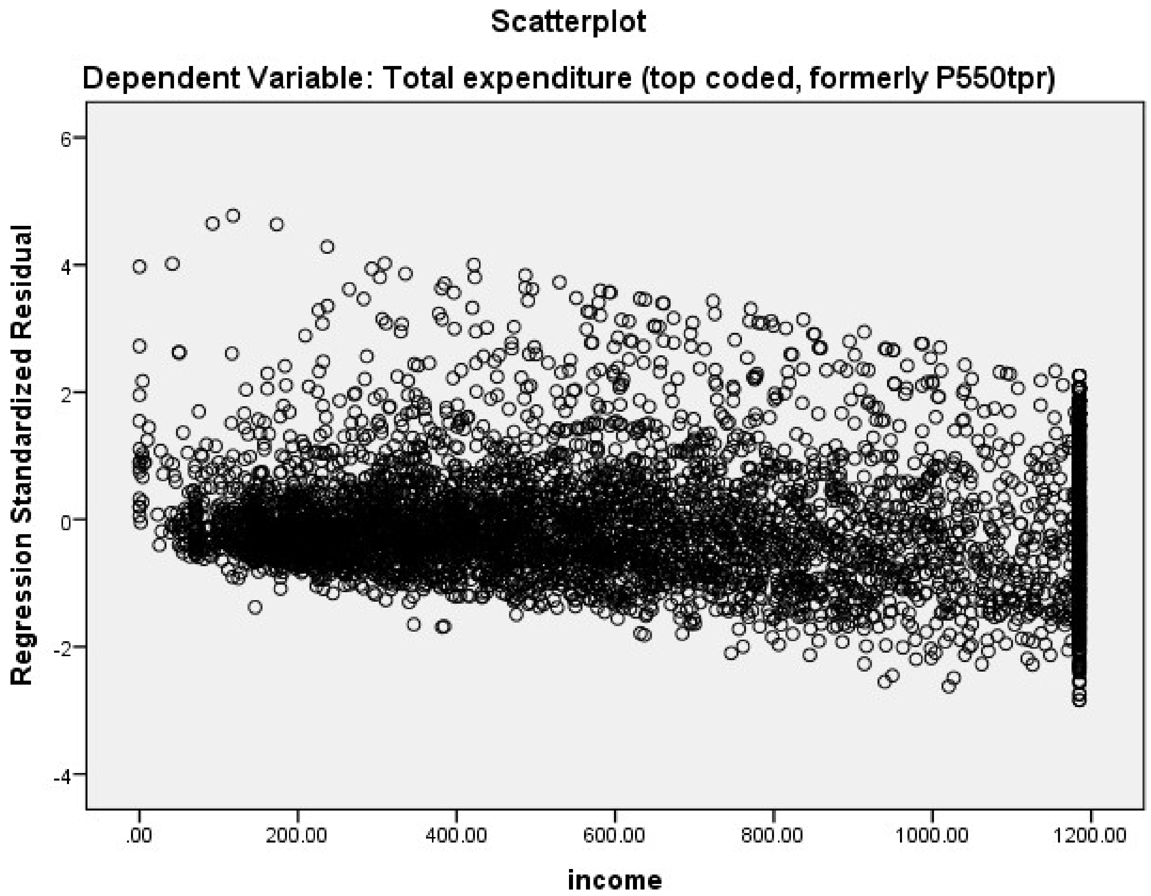


We hope to see the histogram of residuals roughly following the shape of the normal curve that is superimposed over them. The SPSS instructions for the first scatter plot are as follows:

1. Select **Scatter/Dot** from the **Legacy Dialogs** available from the **Graphs** menu.
2. Select Simple Scatter and click on Define to bring up the Simple Scatterplot window.
3. Copy the **Standardized Residual [ZRE_1]** variable into the **Y Axis** box.
4. Copy the **income** variable into the **X Axis** box.
5. Click on the **OK** button.

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

The output from SPSS is as follows:



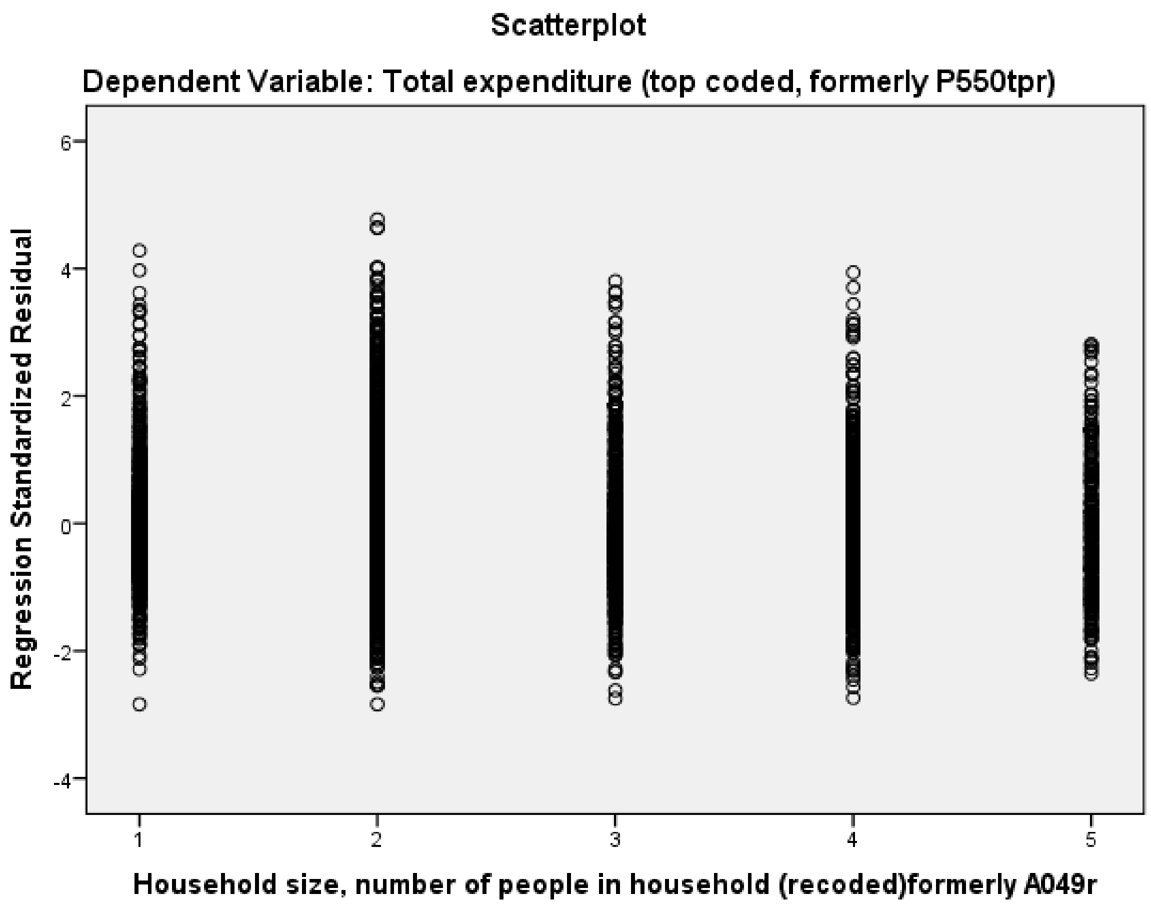
We hope that the residuals show a random scatter when plotted against the predictor variable and also that their variability is constant across different values of the predictor variable.

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

1. Select **Scatter/Dot** from the **Legacy Dialogs** available from the **Graphs** menu.
2. Select Simple Scatter and click on Define to bring up the Simple Scatterplot window
3. Remove the **income** variable from the **X Axis** box.
4. Copy the **Household size, number of people in household (recoded) formerly A049r[hhsize]** variable into the **X Axis** box.
5. Click on the **OK** button.

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

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.

Copyright and citation