How to Run a Simple Linear (OLS) Regression in SPSS

Introduction

There are several different types of regression analysis but the most commonly used is Ordinary Least Squares (OLS) regression which is also known as Linear Regression. The DV in OLS/Linear regression must be Interval level and the IVs must be Interval Level or Dummy versions (we will cover dummy variables in a later handout) of Nominal variables. OLS/Linear regression is usually multivariate (more than one IV involved) but to allow us to understand the 'basics' we will start by running what are known as Simple linear regressions, which have only one IV, hence the title 'simple'. In this form, they are very similar to correlations but the regression analysis, as you will see, provides us with much more information than a correlation test. You must never use simple linear regressions in your assessed work – they are strictly for practice only.

Practice Dataset

This handout uses the ***hse2002.sav*** dataset as the practice dataset. You do not need to test it for parametric assumptions, as this is just a practice – obviously, you would normally do this with a dataset prior to running a regression analysis.

Example Regression Model

When we run a simple linear regression, or indeed any regression model, we have two RQs and two NHs. Our two RQs are:

1. What is the overall fit of our model – in this case the model of systolic BP and age?

2. What is the influence of the IV(s) on the DV – in this case, age on systolic BP?

Hypothesis for RQ 1 (overall fit of the model):

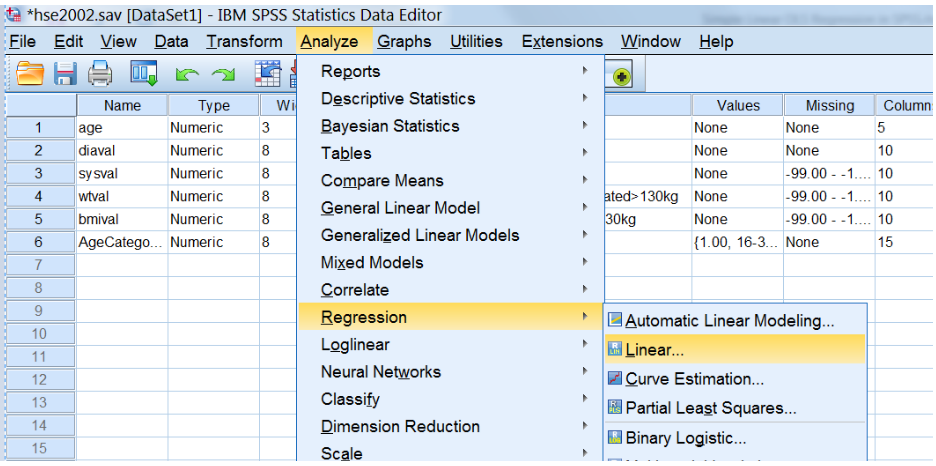
- NH: The model containing age is not significantly different from the one, which does not include the age variable (called the base model).
- RH:The model containing age is significantly different from the one, which does not include the age variable (called the base model).

RQ 2(2) (influence of IV on DV):

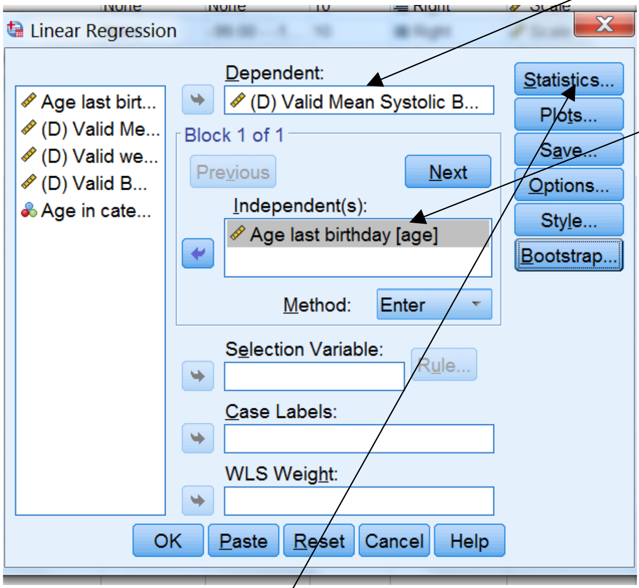
- NH: Age is not significantly associated with increased systolic blood pressure.
- RH: Age is significantly associated with increased systolic blood pressure.

How to Run a Simple Linear Regression in SPSS

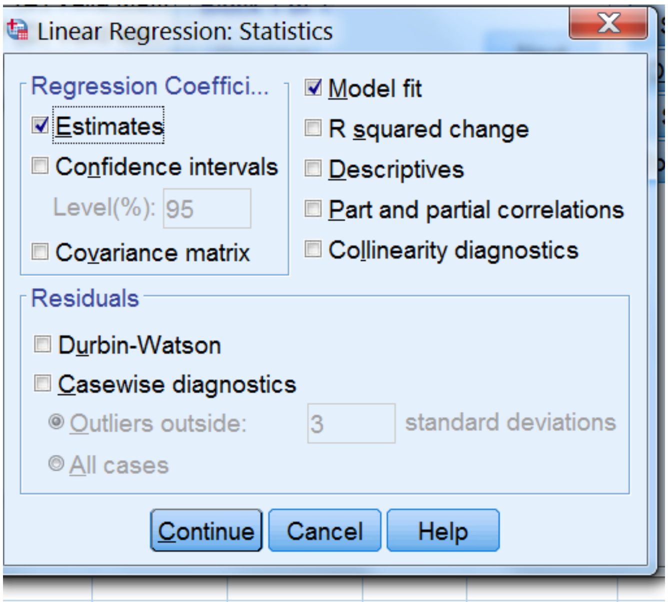
Go to the toolbar: Analyze – Regression - Linear



In the dialog box that opens move your DV to ‘Dependent’ and your IV(s) to ‘Independent(s)



Then Click ‘Statistics’.



This dialog box opens.

Everything is pre-ticked so just press ‘continue’ and then ‘OK’ to run the regression.

Exploring the SPSS Output

Variables Entered/Removed[a]

Model	Variables Entered	Variables Removed	Method
1	Age last birthday[b]	.	Enter

a. Dependent Variable: (D) Valid Mean Systolic BP

b. All requested variables entered.

This top section of the output summarises variables entered into the model. If you were building a series of models this table would be a useful summary. However, for us, we are only using one IV.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.473[a]	.224	.224	16.03956

a. Predictors: (Constant), Age last birthday

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	332133.314	1	332133.314	1291.004	.000[b]
	Residual	1151272.162	4475	257.268		
	Total	1483405.476	4476			

a. Dependent Variable: (D) Valid Mean Systolic BP

b. Predictors: (Constant), Age last birthday

This output tells us whether we reject NH1 – model fit. First we look at the Sig. which is p = .000. This should be examined in conjunction with the F-statistic, which ranges from zero upwards; the larger the figure the better, but for a threshold of 0.05 anything above 7 is good. We can see that we have significance, therefore we can reject NH1 – our model fits. The adjusted R^2 tells us how much of the variance of systolic blood pressure can be explained by age: we convert 0.224 into a percentage, i.e. 22.4%. Therefore 22.4% of the variance in systolic blood pressure can be accounted for by the variable age in this model.

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	109.782	.677		162.074	.000
	Age last birthday	.479	.013	.473	35.931	.000

a. Dependent Variable: (D) Valid Mean Systolic BP

This final bit of output refers to NH2, concerning the influence of the IV on the DV.

We can see that our standard error has reduced following the insertion of the IV into our model. In addition, the B_i is much bigger than the standard error – both good signs. We have significance (p = 0.000), to we can say that there is a relationship between the two variables; age influence blood pressure. Our B_i is 0.479, which tells us that we have a positive relationship between our variables (note absence of a minus sign; a minus sign would mean it is a negative relationship). For each increase in year of age, our systolic blood pressure increases by 0.479mmHg.

Presenting the Results of a Simple Linear Regression

You should present the findings of your regression like this:

Table 1: Simple Linear Regression analysis of rate of systolic BP by age

	B	Std. Error	Beta	t	Sig.
(constant)	109.782	.677		162.074	.000
Age (in years)	.479	.013	.473	35.931	.000

R^2 = .224; Adjusted R^2= .224; F = 1291.004; p = .000; N = 4477

The results of the linear regression (F = 1291.004; p = .000) shown in the table 1 above, suggests that the model containing age is significantly different (better) from the one which does not contain age (called the based model). Furthermore, the adjusted R^2 value indicates that 22.4% of the variance in rate of systolic blood pressure is explained by age. The regression coefficient (B = .479; p = .000) indicates a statistically positive relationship between age and rate of systolic blood pressure. An increase in one year is associated with an increase of 0.479 mmHg in the rate of systolic blood pressure.

Your Turn

Have a go at running two more simple linear regressions from this dataset, specifically weight (which we did in class) and then BMI.