

### MSc AREA – Linear Regression

In this workshop, you will practice conducting Linear Regression analyses using SPSS.

Imagine you want to work out how well mock exam marks can predict final exam marks. You collect data from 15 participants – as shown in the Table below.

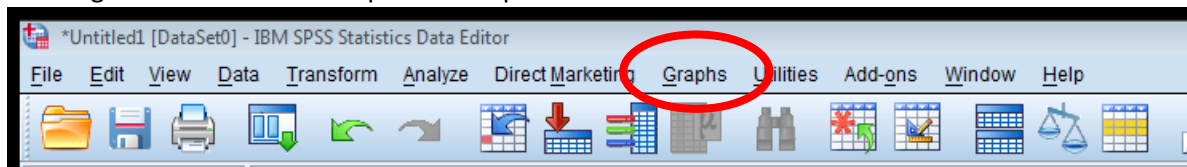
Table: Mock exam marks and Final exam marks

| Mock Exam Mark (%) – Predictor | Final Exam Mark (%) – Criterion |
|--------------------------------|---------------------------------|
| 52                             | 53                              |
| 67                             | 66                              |
| 78                             | 75                              |
| 90                             | 88                              |
| 88                             | 89                              |
| 76                             | 79                              |
| 56                             | 56                              |
| 74                             | 80                              |
| 50                             | 55                              |
| 92                             | 87                              |
| 83                             | 80                              |
| 88                             | 85                              |
| 76                             | 76                              |
| 89                             | 91                              |
| 99                             | 96                              |

Input the data in the table in to SPSS and create a scatterplot.

#### Constructing a Scatterplot in SPSS

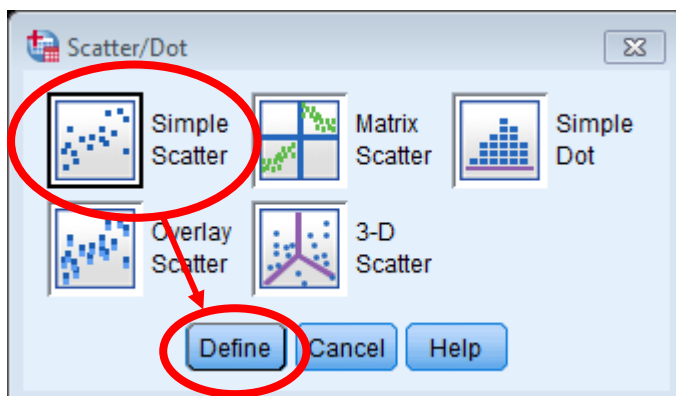
1. Along the tool bar on the top click “Graphs”.



2. Then hover the arrow over “Legacy Dialogs” and then scroll down to “Scatter/Dot”.

NOTE: You can also create a Scatterplot by clicking on Chart Builder from the Graphs menu.

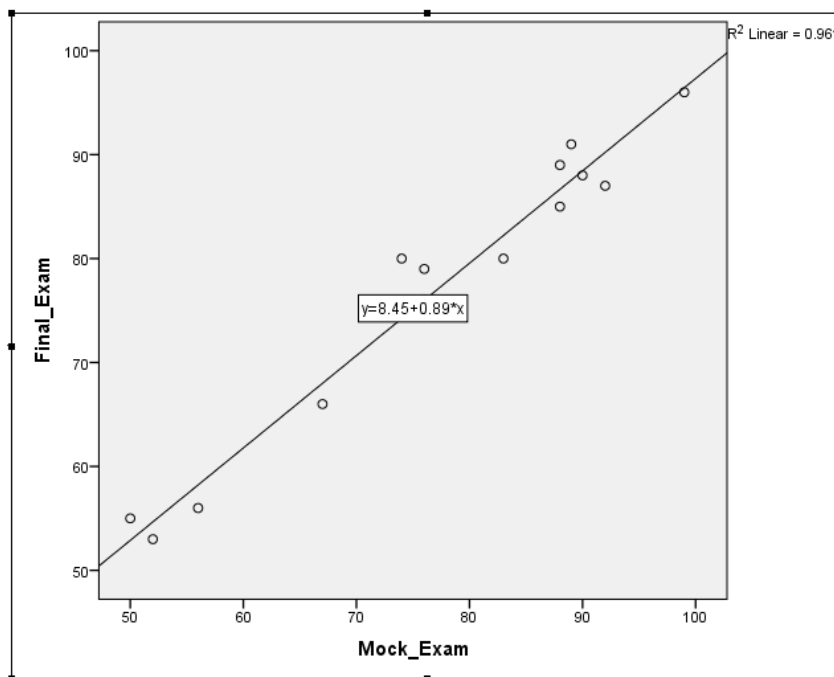
3. Then choose “Simple Scatter” and click Define



4. Then place your **Criterion in the “Y axis”** box and the **Predictor in the “X axis”** box.
5. You can then click on titles if you wish to provide fuller descriptions of what is displayed in the Scatterplot.
6. Once everything is inputted correctly, you can click ok.

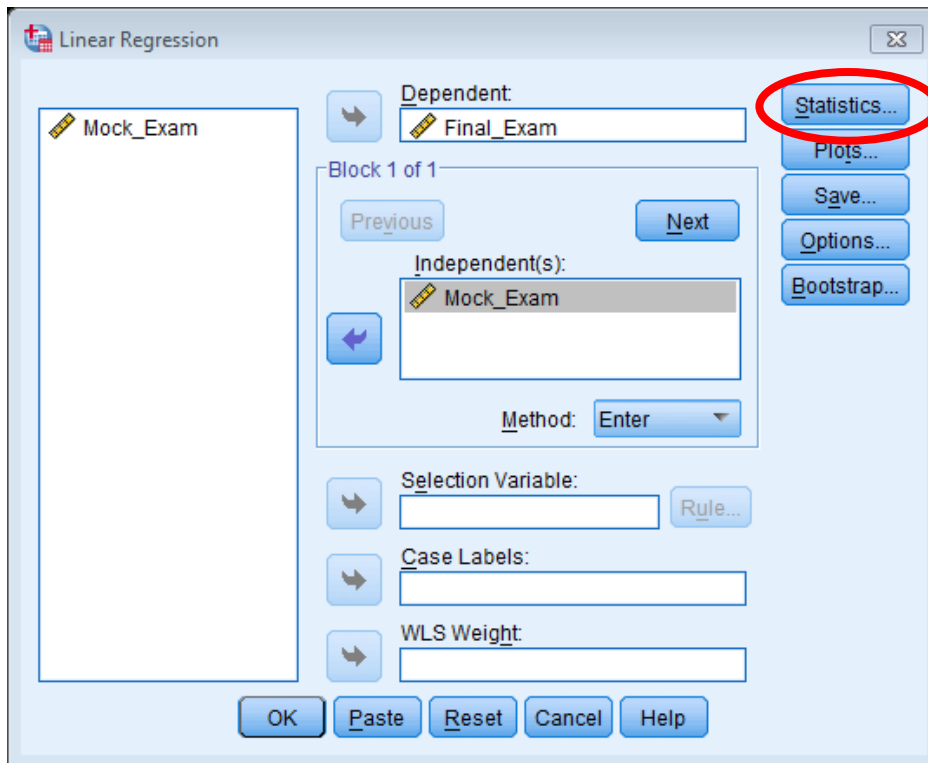
SPSS will then provide you with a Scatterplot of your data: You should end up with something that looks like the chart below.

You will need to input a “Line of best fit”. To do this, double click on the chart area. A “Chart Editor” window will then appear. Click on “Add Fit at Line Total”.

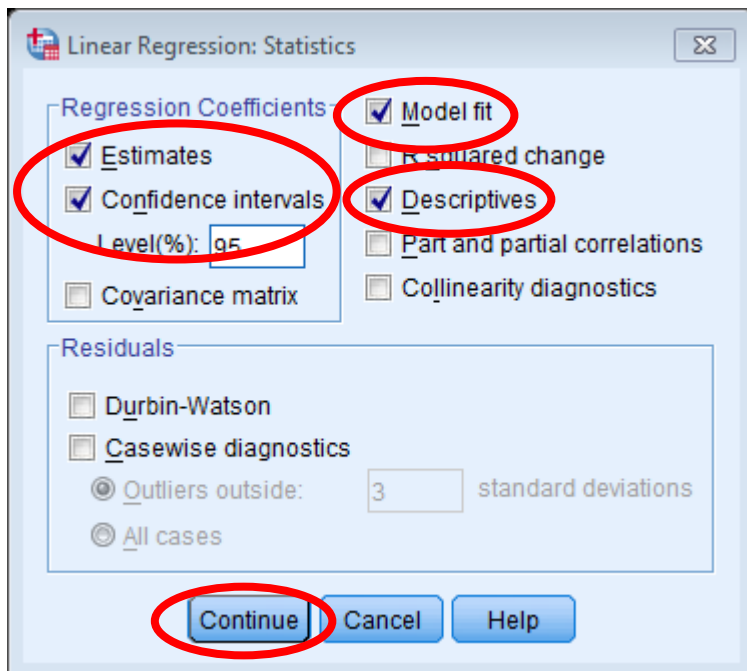


**Conducting a Linear Regression:** There are two different ways in SPSS in which you can conduct a Linear Regression - both procedures produce the same outcome. In your SPSS for Psychologists textbook the method used is to go to “Analyze, Regression, and then Curve Estimation”. But I’d like for us to run through the “Analyze, Regression, Linear” method as this will be more useful for you to understand in preparation for your multiple regression workshops next semester.

Once you have gone to “Analyze, Regression, Linear”, a linear regression window will appear. Here you will need to move your DV/Criterion variable (Final Exam mark) in to the “Dependant” box, and move your IV/Predictor variable (Mock Exam mark) in to the “Independent(s)” box. Then click on “Statistics”.



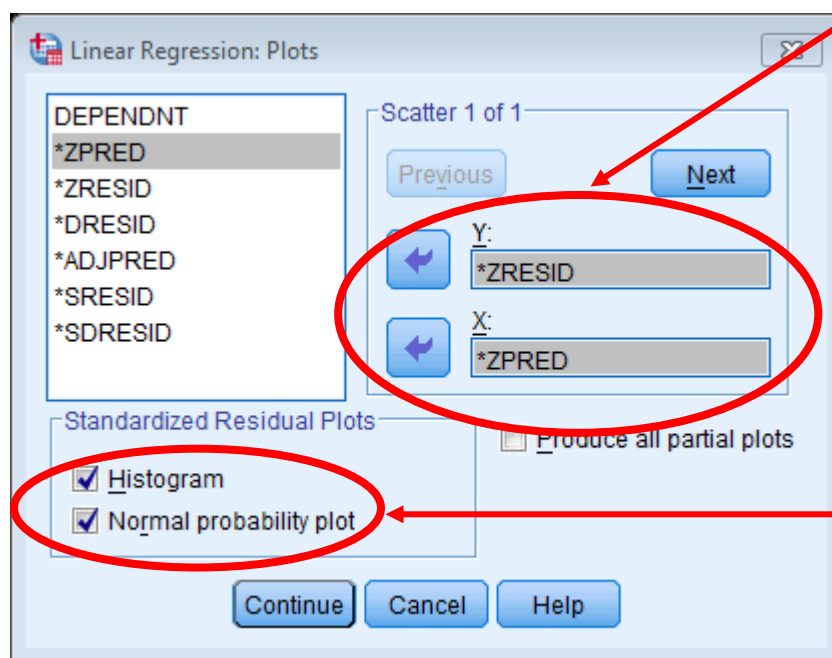
A "Linear Regression: Statistics" window will appear. Make sure "Estimates" and "Confidence Intervals" are ticked in the "Regression Coefficients" section, and that "Model fit" and "Descriptives" are also ticked. Then click "Continue" and then click OK



## Data Checking and Screening

There are 2 assumptions regarding the residuals that we need to consider. 1) They should be normally distributed. 2) There should not be heteroscedasticity (i.e., for each value of  $y$  the variance of the residuals should be similar).

You can click on Plots to ensure these assumptions are met. Move “\*ZRESID” in to the Y box and move “\*ZPRED” in to the X box. Hopefully, the chart plotted from this will **not** show any pattern – so you can be sure that the amount of error is similar for both variables (i.e., no heteroscedasticity).



You can also tick ‘Histogram’ and ‘Normal probability plot’ under the “Standardised Residuals Plots” section. From this you should obtain a histogram that will hopefully show a normal distribution for your Criterion variable (DV). You should also obtain a Normal P-P plot of regression standardised Residual, where hopefully you will see all the data points reasonably close to the line. If this is the case then you can assume normality.

## Linear Regression Output

### Regression

[DataSet0]

#### Descriptive Statistics

|            | Mean  | Std. Deviation | N  |
|------------|-------|----------------|----|
| Final_Exam | 77.07 | 13.719         | 15 |
| Mock_Exam  | 77.20 | 15.134         | 15 |

#### Correlations

|                     |            | Final_Exam | Mock_Exam |
|---------------------|------------|------------|-----------|
| Pearson Correlation | Final_Exam | 1.000      | .980      |
|                     | Mock_Exam  | .980       | 1.000     |
| Sig. (1-tailed)     | Final_Exam | .          | .000      |
|                     | Mock_Exam  | .000       | .         |
| N                   | Final_Exam | 15         | 15        |
|                     | Mock_Exam  | 15         | 15        |

Correlation between the predictor and criterion variables and the associated level of significance

Variables Entered/Removed<sup>a</sup>

| Model | Variables Entered      | Variables Removed | Method |
|-------|------------------------|-------------------|--------|
| 1     | Mock_Exam <sup>b</sup> | .                 | Enter  |

a. Dependent Variable: Final\_Exam

b. All requested variables entered.

**IMPORTANT!**

| Model Summary <sup>b</sup> |                   |          |                   |                            |
|----------------------------|-------------------|----------|-------------------|----------------------------|
| Model                      | R                 | R Square | Adjusted R Square | Std. Error of the Estimate |
| 1                          | .980 <sup>a</sup> | .961     | .958              | 2.802                      |

a. Predictors: (Constant), Mock\_Exam  
b. Dependent Variable: Final\_Exam

The correlation (same value as in the correlations box above).

The amount of variance explained: transfer to %

Takes account of the No. of Ps & variables in the model. Gives a more realistic estimate. Again, you need to transfer this to a %.

ANOVA<sup>a</sup>

| Model |            | Sum of Squares | df | Mean Square | F       | Sig.              |
|-------|------------|----------------|----|-------------|---------|-------------------|
| 1     | Regression | 2532.859       | 1  | 2532.859    | 322.581 | .000 <sup>b</sup> |
|       | Residual   | 102.074        | 13 | 7.852       |         |                   |
|       | Total      | 2634.933       | 14 |             |         |                   |

a. Dependent Variable: Final\_Exam  
b. Predictors: (Constant), Mock\_Exam

Associated degrees of freedom values

Shows whether the regression line is significantly different from 0. (For a Bivariate Regression we could just use the correlation significance – because there's only 1 predictor variable. The ANOVA F significance relates more toward multiple regression because it includes more than 1 predictor variable).

| Coefficients <sup>a</sup> |                             |            |                           |        |      |                                 |             |
|---------------------------|-----------------------------|------------|---------------------------|--------|------|---------------------------------|-------------|
| Model                     | Unstandardized Coefficients |            | Standardized Coefficients | t      | Sig. | 95.0% Confidence Interval for B |             |
|                           | B                           | Std. Error | Beta                      |        |      | Lower Bound                     | Upper Bound |
| 1 (Constant)              | 8.452                       | 3.888      |                           | 2.174  | .049 | .053                            | 16.852      |
| Mock_Exam                 | .889                        | .049       | .980                      | 17.961 | .000 | .782                            | .996        |

a. Dependent Variable: Final\_Exam

This is a  
(The intercept)

This is b  
(The Slope)

This has converted the slope (b) in to a standardised regression coefficient, which is measured in units of standard deviation.

Lower and Upper Bound Confidence Interval (CI). Tells us the range in which the value of the population slope can be found. The narrower the interval the better, as a narrow range gives us good confidence in our findings.

| Residuals Statistics <sup>a</sup> |         |         |       |                |    |
|-----------------------------------|---------|---------|-------|----------------|----|
|                                   | Minimum | Maximum | Mean  | Std. Deviation | N  |
| Predicted Value                   | 52.89   | 96.44   | 77.07 | 13.451         | 15 |
| Residual                          | -3.221  | 5.777   | .000  | 2.700          | 15 |
| Std. Predicted Value              | -1.797  | 1.440   | .000  | 1.000          | 15 |
| Std. Residual                     | -1.149  | 2.062   | .000  | .964           | 15 |

a. Dependent Variable: Final\_Exam

The Residual Statistics box will be displayed if we requested any plots.

## Writing up the results of a Linear Regression

"A simple Linear Regression was carried out to determine whether marks on a final exam could be predicted by marks on a mock exam. A significant regression equation was found ( $F(1, 13) = 322.581, p < 0.001$ ). The adjusted  $R^2$  was 0.958, suggesting that 96% of the variation in final exam marks can be explained by mock exam marks. For every one mark increase in the mock exam, final exam marks increased by 0.889. Confidence limits were narrow, showing that we are 95% confident that the population slope is between 0.782 and 0.996."

## Predicting new cases

We now have all the information we need if we wanted to predict y for new cases in the regression equation. So:

Y = Final Exam Mark, X = Mock Exam Mark, a = 8.45, b = 0.89

Final Exam mark =  $(0.89 \times X) + 8.45$ . Try putting a pretend mock exam mark in x to calculate the equation and see what result you get. This would be the persons predicted Final Exam mark.

\*\*\*Now complete the question below\*\*\*

A Psychologist was interested in whether anxiety about exams related to the number of hours students spent studying. Data from the study are presented below.

Table: Level of Anxiety and Number of hours spent studying

| Level of Anxiety | No. of hours spent studying |
|------------------|-----------------------------|
| 33               | 45                          |
| 64               | 68                          |
| 33               | 100                         |
| 22               | 44                          |
| 70               | 62                          |
| 66               | 61                          |
| 59               | 52                          |
| 84               | 66                          |
| 56               | 79                          |
| 44               | 44                          |
| 22               | 16                          |
| 44               | 61                          |
| 80               | 60                          |
| 66               | 61                          |
| 79               | 60                          |

- a.) Perform a Linear Regression on the data.
- b.) Looking at your output – Does anxiety about exams relate to number of hours studied? – **Write your answer in the space below**
  
- c.) How well does the regression equation predict? – **Write your answer in the space below**
  
- d.) Write-up your results in the space below as you would for a report.