**Running a simple linear regression analysis in SPSS**

First, it is always a good idea to plot the data before you run the regression analysis.

1. Select Graphs | Chart Builder from the SPSS menu,
2. Click on Scatter/Dot,
3. Drag the simple scatterplot icon (the first on the left) onto the chart preview window.
4. Make sure both variables are scale (yellow ruler) and not nominal or ordinal
5. Drag the independent variable to the X-axis
6. Drag the dependent variable to the Y-axis

A screenshot of a computer

Description automatically generatedIn the example shown here, there is a clear and obvious outlier. Remove this point and redraw the graph. After redrawing, you should make some adjustments. These might include

1. Placing your name and date in the title
2. Using a comma separator for the labels on the Y-axis
3. Adding tick marks on the X and Y axes.
4. Changing the scale margins to a smaller values (2% or 0%).
5. Converting the blue marker interiors to invisible
6. Changing the axes (minimum, maximum, and/or major increment)
7. Adding a trendline

Although the minimum and maximum values and major increment are fine in this graph, you should always consider alternatives.

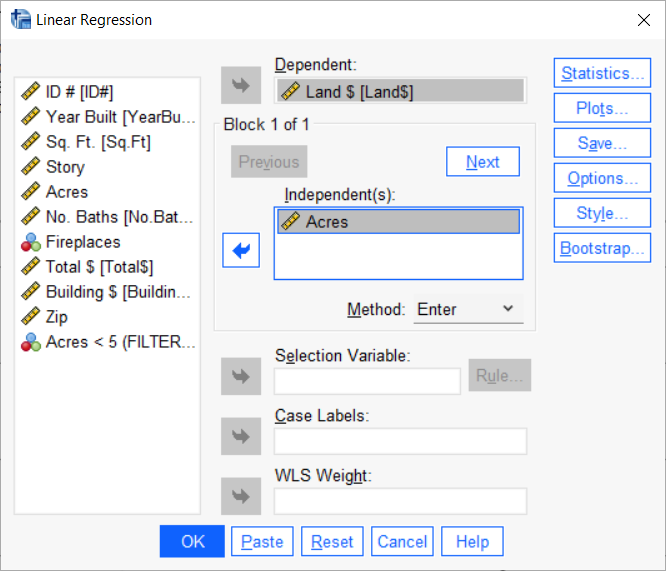
A graph of a plot

Description automatically generated with medium confidence

To run a regression analysis, you have two options:

1. Select Analyze | Regression | Linear, or
2. Select Analyze | General Linear Model | Univariate

Both have advantages and disadvantages. For simple problems, the former is best and that is the choice illustrated here.



1. Add the dependent variable to the dependent field.
2. Add the independent variable to the independent field.

Click on the Statistics button and

A screenshot of a computer

Description automatically generated

The options for Estimates and Model fit are selected by default. Add the option for Confidence intervals.

Then click on the Save button.

A screenshot of a computer

Description automatically generated

Click the boxes for unstandardized residuals and unstandardized predicted values. This produces five numeric tables.

A screen shot of a screen

Description automatically generated

The first table is only relevant in regression models with multiple independent variables, where you might consider adding or removing independent variables.

A screenshot of a graph

Description automatically generated

The second table gives the R Square value. In this model, the size of the lot explains 3% of the variation in lot price.

A table with numbers and symbols

Description automatically generated

This is the analysis of variance table. You have 1 degree of freedom for Regression and 97 degrees of freedom for Residual (or Error). The sum of squares and mean squares are rather large because the lot prices are in the tens and hundreds of thousands. The F-ratio is 3.147. This is close to 1. The p-value, 0.079, is larger than 0.05, so you would accept the null hypothesis and conclude that there is no relationship between the size of the lot in acres and its price in dollars. Some researchers would describe this as a “borderline” finding because the p-value is only slightly larger than 0.05.

A close-up of a graph

Description automatically generated

This is the table of regression coefficients. The intercept is $62,000 and the slope is $40,000. The estimated average lot price increases by 40 thousand dollars for each increase of one acre in the size of the lot. The interpretation of the intercept is a bit difficult--the estimated average price of a lot is 62 thousand dollars for a lot of zero acres.

Notice that the t-statistic for acres is close to zero and the p-value is larger than 0.05. This is testing the hypothesis that the slope is zero. You would accept the null hypothesis and conclude that there is no relationship between the size of the lot in acres and price in dollars. This is consistent with the interpretation provided earlier.

The 95% confidence interval for the slope ranges from -5 thousand dollars to 85 thousand dollars. This confidence interval includes the value of zero, leading you to draw the same conclusion as earlier.

Normally, you do not test any hypotheses or compute any confidence intervals for the intercept.

To test normality, draw a QQ plot (a histogram would also be fine). Select Analyze | Descriptive Statistics | Q-Q plots from the menu.

A screenshot of a computer

Description automatically generated

The default options are all fine.

The Q-Q plot produces three tables and two graphs. The first table provides some technical details about the plot, but none are important for the most data analysts.

A screenshot of a model description

Description automatically generated

The second table gives an accounting of missing values. This is important and you should always track the number of missing values in any data analysis.

A screenshot of a computer

Description automatically generated

The third table gives the estimated mean and standard deviation. The mean of the residuals is always zero. The scale (standard deviation) is 71,000. It can vary from data set to data set, but is usually not of direct interest.

A close-up of a box

Description automatically generated

The first plot is the important one.

A graph with a line

Description automatically generated

For this data set, the QQ plot is not anywhere close to a straight line. The assumption of normality is not met.

The final plot is a detrended plot. It has no value and should be ignored.

A graph with blue dots

Description automatically generated

You should also plot the