

Lab Activity: Linear Regression and Correlation

In this lab activity, you will collect sample data of two variables, determine if a linear correlation exists between the two variables, and perform linear regression.

Student Learning Outcomes

By the end of this chapter, you should be able to do the following:

- Perform linear regression and correlation using Statcato
- Interpret the results of linear regression and correlation

Preliminary

Read Chapter 12 Linear Regression and Correlation in:

Illowsky, Barbara, and Susan Dean. Collaborative Statistics. Connexions. 2 Mar. 2010
<http://cnx.org/content/col10522/1.37/>.

Make sure you understand the following **key terms** (LR:Key Terms):

linear regression, linear equation, independent variable, dependent variable, scatterplot, line of best fit, least-squares regression line, residual, sum of squared errors (SSE), correlation coefficient, outlier

Background

Think about two variables between which you suspect a relationship exists. For example, one might suspect that the number of hours a student studies per day is positively related to the GPA of a student (i.e. the more a student studies, the higher his/her GPA). In this case, number of hours studies per day should be the independent variable (x), and GPA should be the dependent variable (y). Choose two variables that you are curious about and for which you can collect sample data. Record the independent and dependent variables below and in LR: Variables:

independent variable (x): _____
dependent variable (y): _____

Collecting Data

Collect 10 paired sample values. You may sample values from your environment, survey human subjects, or use published data sets from a respectable source, as appropriate for the two variables you have chosen. Discuss the details and results of the data collection process in LR: Data.



Here you will input the samples into Statcato for further processing later in the lab.

- Go to **File > Save Project** in order to save the project.
- Enter your sample x values in column **C1**.
- Enter your sample y values in column **C2**.

Performing Linear Regression and Correlation

Using Statcato, you will perform a test of linear correlation and determine the linear regression equation. Use a significance level of 5%.



Performing Linear Regression and Correlation

Go to **Statistics > Correlation and Regression > Linear (Two-Variables)**.

- In the **Inputs** panel, select the columns containing the x and y variables in the corresponding drop-down menus. Click the **Add Series** button. You should see the pair of column names in the **Independent/dependent variable series** list.
- For **Significance Level**, enter **0.05**.
- Check the box **Show a scatterplot for all pairs of data values**.
 - Enter appropriate labels for the x - and y -axes.
 - Change the plot title as desired.
 - Check **Show legend** and **Show regression line**.
- Click **OK**.

You should see a scatterplot of the data pairs. The results of the linear regression and test of correlation are shown in the Log window. Copy the generated scatterplot to **LR: Scatterplot**. Copy the computation results from the Log window to **LR: Linear Correlation and Regression**.

Interpreting the Results

Based on the computer-generated results, you will make interpretations on the relationship between the two variables of your study. Record your answers in **LR: Interpretation**.

Linear Correlation

The null hypothesis of the test of linear correlation is that there is no linear correlation, whereas the alternative hypothesis is that there is. There are two ways to decide if we should reject the null hypothesis and conclude that there is sufficient evidence for linear correlation:

- if the linear correlation coefficient r is greater than the positive critical value or less than the negative critical value
- if the p-value is less than the significance level α

Moreover, the value of r indicates the type of correlation:

- $r > 0$: positive correlation
- $r < 0$: negative correlation
- $r = 0$: no correlation

In the results generated by Statcato, you should see the p-value, the test statistic r and the critical value (for $\alpha = 0.05$). Based on the results, explain whether there is a linear correlation between the two variables. If there is a correlation, state what type of correlation it is (positive or negative). Does the type of correlation match how the data appears in the scatterplot?

Linear Regression

In the results computed by Statcato, you should also see the regression equation ($Y = b_0 + b_1x$), for which the y-intercept (b_0) and slope (b_1) are given. What is the regression equation for your sample data? Does the slope indicate that the relationship between the two variables is positive or negative?

The regression equation can be used to predict values within the domain of x . Choose an x -value in the domain of your independent variable that is not one of your sample x -values. Substitute the value into your regression equation to obtain a prediction.

Discussion

Answer the following questions in **LR: Discussion**.

1. Does the regression line shown on the scatterplot seem to fit the data? Which statistical measure obtained above reflects how well the regression line fit the data?
2. In one or two complete sentences, state the relationship between the two variables of your study based on the results obtained above.
3. Are there any outliers? If so, which point is an outlier? If there are outliers, should you exclude them from the analysis of the data?