

Math 1040 Project 1 - Introduction to StatCrunch

This exercise is designed to give you a brief introduction to the StatCrunch software that accompanies your textbook. Log-in to MyStatLab and locate StatCrunch. (You can find the button on the left-hand side for “Tools for Success”, then click on the link for StatCrunch.) The program opens in a new window where you may edit your data.

First try loading data from an existing file. From the list of chapters at the left side of your screen choose Chapter 2 and the dataset 2 2 33.txt. This file contains the values from problem 33 in section 2.2 “Average Income”. Per capita disposable income for the 50 states and the District of Columbia are recorded for the year 2006. Note that nearly every dataset in the textbook is accessible through this program, either on the left-hand side, or by using the pull-down data menu.

Start by sorting the data. Click on “Data”, “Sort columns” and the column you want to sort (“Income”) by clicking on it. The program default is to sort in ascending order, but you can choose descending on this screen. For now, leave the box “un-checked” and click “Sort Columns”. This will add a new column to the spreadsheet window with the sorted data. Rename this column “Sorted Incomes” by clicking up in the column title. You will need to delete the current title before renaming it.

To create a histogram for disposable income, click on “Graphics” on the tool bar at the top of your screen. Click on “Histogram”. Choose the column you would like a histogram for (“Income”), use the computer defaults and click on “Create Graph!”. You should now have a histogram. Notice that the computer selected the class limits and class width. Try changing them by clicking “Options”, “Edit”, and then selecting the “Next>” button. Try changing the class width (“Binwidth”) to 8000 or 3000. Just click on “Create Graph!” again to see how your changes look. Notice that even though the data values are the same, the histogram looks different when you change the number of classes. Play with the other options in the histogram window. What happens if you change from a frequency histogram to a relative frequency histogram? Once you know how the options work, pick a class width and class start that you like, label both axes, give it a nice **title with your name**, and copy the histogram into the word file you will submit.

Next we will explore another feature of StatCrunch. Close your histogram window and return to the data window. From the tool bar at the top of your screen choose “Stat” and click on “Summary Stats”, then “Columns”. Choose Income (the per capita disposable income), leave the default setting in the first window, and click “Calculate”. This calculates the measures of center and variation and gives the 5-number summary for the data set. This is going to be one of your favorite StatCrunch tools, so don’t forget how to access this information! Hit “Options”, “Edit”, and then “Next” to see how you can select or de-select certain measures. Copy the descriptive statistics for the income data into your file and then close the window.

Next, what happens if you want to use your own data? Clear the income data by going to the “Data” menu and “Data table” then “Delete columns”. Select the column titles and hit “Delete”.

Now place your cursor in Row1, Column 1. Type in the following data values, hitting enter after each number: 38, 25, 55, 47, 62, 40, 41, 24, 29, 33, 36, 50. To review what you have learned, create a histogram and evaluate the descriptive statistics for this data. Copy the graph and descriptive statistics into your file.

Finally, let’s use StatCrunch to create scatter plots and compute correlation coefficients and regression equations. Clear the data again and then load the data from chapter 4, section 4.1, problem 27. This dataset has data from the National Center for Health Statistics on Age and Cholesterol levels for 17 males. To create a scatterplot, go to “Graphics”, “Scatter Plot”. We want to see if a man’s age is a good predictor of his cholesterol level, so select age as the x variable and cholesterol as the y variable, then hit “Create Graph!”. To see the linear regression line and find the correlation coefficient, go to the “Stat” menu, then “Regression” and “Simple Linear”. Choose Age for the X-Variable and Cholesterol for the Y-Variable and then click “Next>” twice. Now on the Graphics options menu, click the box next to “Plot the fitted line”. (Notice that this is also where you can select to see a residuals plot – “Residuals vs. X-values”.) Finish by hitting “Calculate”. The first page gives you the correlation coefficient, R-squared, and the values for the intercept and slope of the regression line. Hit “Next>” to see the Scatter Plot with the fitted regression line. Copy this graph into your file.

Turn-in:

1. The income histogram
2. The income summary statistics
3. Histogram and descriptive statistics for the given numerical data
4. Scatter plot with fitted regression line