TutorTube: Histograms in StatCrunch

Spring 2020

Introduction

Hello! In this video I will be discussing two different ways of making histograms in StatCrunch. In the first example, we are given a table of raw data and asked to construct a histogram by finding the best values for our classes (also known as bins). In the second example, we are given a frequency table and asked to create a histogram from pre-established class sizes.

Histogram from Raw Data

30,248	34,196	36,937	40,140
30,430	34,679	37,360	41,157
30,753	34,904	37,723	41,333
32,116	35,169	38,423	52,503
32,975	35,569	38,668	
33,673	35,743	38,960	

Figure 1: Data Values (Pearson, 2020)

For our first example, let's say we are given this table of 25 data values. Our problem wants us to find the optimal class sizes for this data set. StatCrunch's histogram function will automatically format your histogram so that it looks "nice" by creating bins of automatic length. However, some homework problems require you to find your own class sizes using predefined rules, which means we can't always use the histograms StatCrunch gives us automatically. Luckily, if we can find those class sizes, StatCrunch will still create a histogram for us very quickly.

Find Class Width

Our first step is finding the **number of classes** that we need. There is a general rule that says that the number of classes, **k**, should be such that the value of 2^k is just slightly greater than or equal to the number of data values.

$$Number\ of\ Classes = k$$

$$where\ 2^k \geq the\ number\ of\ observations$$

So, for example, we have 25 data values in our list. That means that we need to find a k such that 2^k is greater than or equal to 25. So, if we chose 4 for k, 2^k would be 2^4 which is only 16. But if we try 2^5, we get 32, which is greater than 25.

$$number\ of\ observations = 25$$

$$2^4 = 16 < 25$$

$$2^5 = 32 > 25 \implies number\ of\ classes\ is\ 5$$

That means that we should choose **5** as our number of classes.

The next thing we need it to figure out our **class width**. The formula for this is found by taking the max value in our data set, subtracting the minimum value in our data set, and then dividing by the number of classes that we just found.

$$Class\ Width \approx \frac{Max\ Value - Min\ Value}{Number\ of\ Classes}$$

To find our max and min values in StatCrunch, there are two methods we can use.

First, we can just order the data values by clicking on the tab next to the column name and clicking on "sort values ascending". This sorts our values from lowest to highest, so we can easily pick out that the min value is 30,165 and the max value is 52,440.

Figure 2: Sort Data Values

The second method for finding the max and min is a little more involved, but it is useful if you have a very large data set. First go to "Data," then "Compute," then "Expression." Scroll down until you find the "max" function and then double click. Next, double click on the column name (here "var1"). Now hit "Okay." Then, "Compute." Notice that a new column has been created with the max value of our data set. The same process can be used to find the min.

Applets ▾	Edit ▾	Data ▾
	min(var1)	max(var1)
0165	30165	52440
0533		

Figure 3: Minimum and Maximum

From here, we can plug these values into our class width formula to find the size of our bins. Our max in this example 52,440, minus the min of 30165, divided by the class width of 5, gives us a class width of 4455.

$$Class\ Width \approx \frac{52440 - 30165}{5}$$

$$Class\ Width \approx 4455$$

Find a Starting Value

Next, we need to choose a starting point for our histogram. To find this, we look at our minimum value and round down to the nearest convenient number. This can vary depending on the numbers you are working with, but here our minimum value rounds down easily to 30,000. When choosing a starting value, just make sure that all your data values will be included in the bins that you choose.

Build Histogram

Now we are all ready to build our histogram. Go to "Graph," then "Histogram."

Figure 4: Graph > Histogram

Select the name of the column containing your data values (here "var1"). Next, under "Bin," enter your starting value in the "Start At:" box. So, for our example we enter 30,000.

Now enter your class length in "Width:" So, we will enter our width of 4455. Leave everything else alone, and then hit "Compute!"

Figure 5: Histogram Window

This will create our histogram. Notice that our bins start at 30,000 and all have a width of 4455. Also, the upper limit on the last data set, 56,730, is bigger than the maximum value in our data set. Because of this, we know that all of our values are somewhere in our bins.
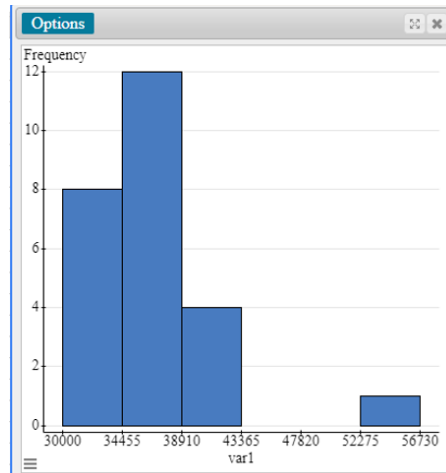
Figure 6: Histogram

Also notice that you can hover your cursor over the bars and it will give you information about the frequency of cases in each bin and the length of the bin itself.

Histogram from a Frequency Table

Class/Bin	Frequency
30000-34455	8
34455-38910	12
38910-43365	4
43365-47820	0
47820-52275	0
52275-56730	1

Figure 7: Frequency Table

For our second example, let's say that we are instead given a frequency table and asked to make a histogram from this summarized data instead. Unlike the previous problem, we won't have to choose the class sizes or starting points because they are already given.
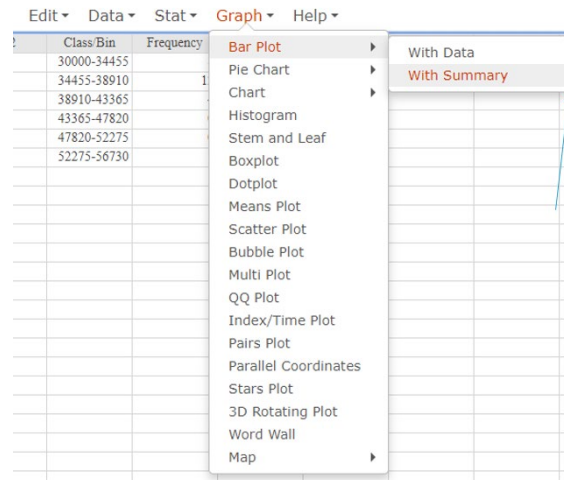
Figure 8: Graph > Bar Plot > With Summary

First off, we want to enter this data into StatCrunch as it is written in the table. One column contains the bins and the other contains the frequencies. Now, instead of going to histogram, we go to "Graph," then "Bar Plot," then "With Summary." This will open up a new window.

In the "Categories in:" section, select the name of the column containing your Classes/Bins. Next, in the "Counts in:" section, select the column name containing your frequencies (here it's called Frequency). Leave everything else alone, and select "Compute!"

Figure 9: Bar Plot Window

This will create our graph. Notice that this isn't a traditional histogram because the bars themselves don't touch each other. But all of our information is here, and we can get a good idea of the shape of the distribution.
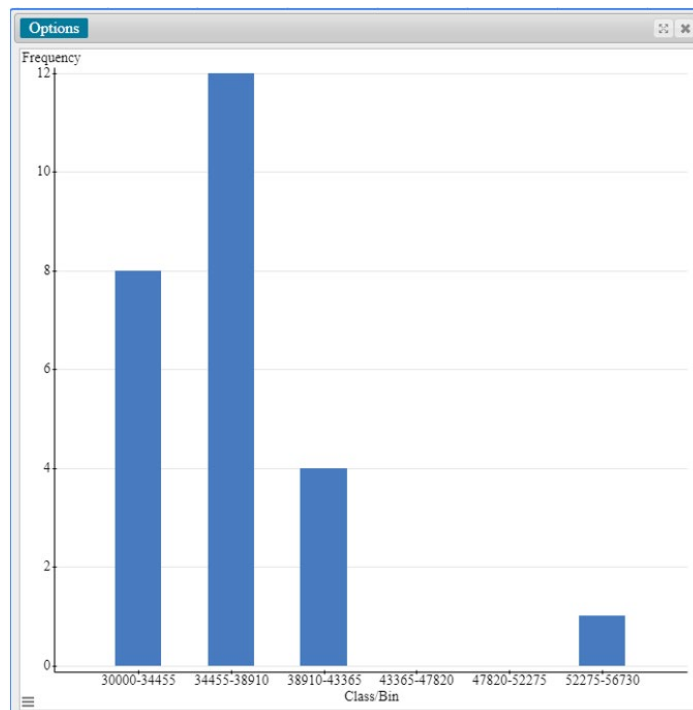
Figure 10: Bar Plot / Histogram from Summary

Outro

And that is all there is to it. I hope you found this video helpful. If you are UNT student, you will find some links to other services and resources in the description of this video. Thanks for watching!

References

Pearson. (2020). MyLab: Statistics. *Pearson Higher Education Inc.*

*All calculations in this video were performed with Pearson StatCrunch 2020 software.