

TutorTube: Frequency in StatCrunch

Spring 2020

Introduction

Hello! In this video, we will be finding relative, cumulative, and relative cumulative frequency in StatCrunch for both raw data and summary data. Finding these frequencies by hand can be very time consuming, but luckily, StatCrunch lets us calculate them much more quickly.

Summary Data

So first, let's start with summary data. If data is presented in summary form, this means that it is already separated into bins and organized into a frequency table, as we can see here.

Category	Frequency
А	1131
В	1087
С	1362
D	589
E	917

Frequency tables are a method of summarizing categorical (or qualitative) data. And here we have qualitative data in the form of the category names, and then the frequency (or count) of each item. So, for example, we know from this table that there were 589 items in the D category, and 917 items in the E category. The question then asks us to find relative frequency, cumulative frequency, and relative cumulative frequency for the data.

Relative Frequency

In order to do this by hand, we would have to find the total count of items, and then apply the formulas for relative frequency and cumulative frequency multiple times. StatCrunch will do this for us very quickly.





Untitled

StatCr	unch 🔹 Ap	plets 🔹 Ed	lit • Data ·	 Stat ▼
Row	Category	Frequency	var3	var4
1	Α	1131		
2	В	1087		
3	C	1362		
4	D	589		
5	E	917		
6				
7				

Figure 1: Data in StatCrunch

Once we have your data entered in StatCrunch like this, we click on "Data," then "Compute," then "Multiple Expressions."

						MySta	tLab Da	ta Set	
StatCru	inch Ap	plets	Edit	Data	Stat	Graph	Help		
Row 1 2 3 4 5 6 7 8 9 10 11 12	Category	Frequer A 1 3 1(C 1) E 9	ncy 131 087 362 589 917	Load Proper Save Export Validat Row Se Compu Simula Sample Arrang Sequel Bin	ties election ite e e nce	Ctrl+S	var5 Expressio Multiple I From Col	var6 on Expressions umn	
12				Rank					
14 15				Indicat Recode	tor e				-
16									

Figure 2: Data > Compute > Multiple Expressions

First, let's calculate the relative frequency. Naming your expressions here is optional, but I'll call the first one "relative freq." Now we just click on "Build."

Compute Multiple E	xpressions	х						
Enter a sequen option is check with the corres Expressions lov in the listing. N spaces or spec	Enter a sequence of name/expression pairs in the listing below. When the corresponding <i>Save</i> option is checked, the results of an expression will be saved to the data table in a new column with the corresponding name. Expressions can be deleted or inserted within the listing. Expressions lower in the listing may refer by name to the results of expressions defined higher in the listing. Names are case-sensitive and should be enclosed in double quotes if they contain spaces or special characters, or if they are entirely numeric.							
Name	Expression	Save						
relative freq =	Build	+ x 🗹						
=	Build	+ x 🗹						
=	Build	+ x 🗹						
	? Cance	Compute!						

Figure 3: Compute Expressions Window

It will open up this window. Now we enter the formula for relative frequency. Recall that formula for relative frequency is:

$Relative \ Frequency = \frac{Frequency}{Sum \ of \ All \ Frequencies}$

And in StatCrunch, this is created by double clicking on the name of the column that contains your frequency numbers (here this is called Frequency). Then click or type the division symbol. And now we scroll down in the functions box until you find the "sum" function. Double click on that. Now find your column name again, and double click again. Then hit "Okay."

Expressi	Expression								
Frequ	Frequency/sum(Frequency)								
7	8	9	+	Columns:	Functions:				
4	5	6	-	Category	substr				
1	2	3	*		substring				
0	1.1	()	1		substring sum				
^	=	.i=	1		tan				
>	>=	<	<=		upperCase upperCaseFirst				
and	or	Clr			ustd				
					var				
				Add Column	Add Function				
Values	589				▼ Add				
History	History: Set								
? Cancel Okay									

Figure 4: The Expression Window: Enter "Frequency/sum(Frequency)"

Ok, so we are taken back to this screen. And you can see it entered our expression into the window. Now we just hit "Compute" and notice that we now have a new column called relative frequency with all of our results automatically calculated. You just have to be sure to round your numbers to the correct number of decimals when giving your answers.

			MvS	tati ah D	ata Set			
StatCru	unch Ap	plets Edit	Data	Stat Gra	ph Help]		
Row	Category	Frequency	relative freq	var4	var5	var6	var7	var8
1	1	A 1131	0.22237515					
2	E	3 1087	0.21372395					
3	(1362	0.26779394	Optio	ons			× ×
4	[589	0.1158081					
5		E 917	0.18029886	New	column, rela	tive freq, add	ded to data t	able!
6								
7								
8								
9								
10								
11								
12								
10								

Figure 5: Relative Frequency Results

Cumulative Frequency

Next let's find Cumulative Frequency. Again, click on "Data," then "Compute," then "Multiple Expressions." This time I'll call the expression "cumu freq" and again we click on "Build."

And recall that cumulative frequency is found by adding each of the frequency values in each row. And StatCrunch has a function that does this automatically. So look in the functions box again, but this time we are going to look for a function called "cumsum." Double click on that. Now double click on "Frequency" again, this tells StatCrunch to find the cumulative sum of all of our frequency values.

Expressio	n				×					
cumsu	cumsum(Frequency <mark>)</mark>									
7 4 1 0 ^ > and	7 8 9 + 4 5 6 - 1 2 3 * 0 . 0 / ^ = != ! > >= < <= and or Clr			Columns: Category Frequency relative freq	Concat contains cor cos count cov cumsum dbeta dbinom					
					dcauchy dchiso					
				Add Column	Add Function					
Values:	589				▼ Add					
History	Freq	uency/	sum(F	requency)	▼ Set					
				?	Cancel Okay					

Figure 6: Cumulative Frequency Formula

Ok, so now we just click "Okay" and then "Compute." And we can see that a new column has been created with the cumulative frequency values for our data.

Untitle	d				
StatCr	unch - Ap	plets 🔹 Ed	it 🔹 Data 🕇	Stat 🕶	Graț
Row	Category	Frequency	relative freq	cumu freq	
1 👻	A	1131	0.22237515	1131	
2	В	1087	0.21372395	2218	
3	C	1362	0.26779394	3580	
4	D	589	0.1158081	4169	
5	E	917	0.18029886	5086	
6					
7					
8					
0					

Figure 7: Cumulative Frequency Results

Relative Cumulative Frequency

Finally, let's find relative cumulative frequency for this dataset. This process is very similar to the process for finding regular cumulative frequency that we just did. First, start by clicking on "Data" then "Compute," "Multiple Expressions." This time I'll call our column "rel cum freq." Ok, now click on "Build" again. So remember that relative cumulative frequency simply takes the relative frequencies in each row and adds them up, totaling up to a final value of 1. This process requires the same "cumsum" function that we used before.

Expressio	n				×					
cumsu	cumsum("relative freq")									
7 4 1 0 ^ > and	7 8 9 + 4 5 6 - 1 2 3 * 0 . 0 / ^ = != ! > >= <		+ - * / ! <=	Columns: Category Frequency relative freq cumu freq	Functions: Corr Cos Count Cov Cumsum dbeta dbinom dcauchy dchisq					
				Add Column	dchisq df Add Function					
Values:	0.11	58081(06685	502	▼ Add					
History	cum	sum(Fi	requen	cy)	▼ Set					
				?	Cancel Okay					

Figure 8: Relative Cumulative Formula

But this time, instead of using the "Frequency" column we will select the "relative frequency" variable instead. This tells StatCrunch to compute the cumulative sum of all of our relative frequency values, which is what we want. Now click on "Okay" and then on "Compute."

StatCr	unch 🗸 Ap	plets 🔹 Ed	it 🔹 Data 🖣	Stat 🕶	Graph 👻 H	lelp 🕶
Row	Category	Frequency	relative freq	cumu freq	rel cum freq	varó
1	А	1131	0.22237515	1131	0.22237515	
2	В	1087	0.21372395	2218	0.4360991	
3	С	1362	0.26779394	3580	0.70389304	
4	D	589	0.1158081	4169	0.81970114	
5	E	917	0.18029886	5086	1	
6						

Figure 9: Relative Cumulative Frequency Results

Ok, and now after we click out (it'll sometimes popup these boxes and we can click out of those) we'll see our new column with the relative cumulative frequency values. Notice too that they add up to 1. This will always be true with relative cumulative frequency.

Raw Data

Okay, so now let's consider a second example. Say that instead of summary data, you were instead given a table of raw values and asked to find the same information.

StatCr	unch 🗸	Ар	plets 🕶	Ed	it •	Data 🔻	Stat •	Gra
Row	Grade	s	var2		,	var3	var4	
1		Α						
2		Α						
3		Α						
4		Α						
5		Α						
6		Α						
7		Α						
8		Α						
9		Α						
10		Α						
11		Α						
12		Α						
13		Α						
14		В						
15		В						

Figure 10: Raw Data

In this example, we have a column of data that contains letter grades for a certain class. We want to know the frequency, cumulative frequency, and relative cumulative frequency for each of these grades. In other words, we want to be able to answer questions like: How many A's, B's, and C's did we have? And if we chose a random student, what is the probability that their grade would be an A or a B? We need frequency counts and relative frequency in order to find the answers. Luckily, computing these values for raw data is much simpler than it is for summary data.

Ар	plets 🔹 Ed	lit 🔹 Data	 Stat - 	Graph 🕶	Help	-
s	var2	var3	Calc	ulators	•	var6
Α			Sum	mary Stats	•	
Α			Table	25		_
Α	Frequence	сy	Table			
Α	Continge	ncy	Z St	ats	•	
Α	Outcome	•	T Sta	ats	•	
Α			Prop	ortion Stats	•	
Α			Varia	ance Stats	•	
A			Rear	ression		_
A			- Kegi		, ,	
Α			ANO	VA	•	
Α			Non	parametrics	•	
Α			Good	dness-of-fit	•	
Α			Cont	rol Charts	•	
В			Deer	manle		
В			Resa	imple	P	
В			Time	e Series	•	
В						

Figure 11: Select Stat > Tables > Frequency

So we just need to go to "Stats," and then "Tables," and then "Frequency." In the Frequency window, select the name of the column containing your data values. Here that's "Grades." Now under the "Statistics" box, we are going to click and drag to highlight all of the options. Then under "Output" click the box next to "Store in data table." Okay, now we have everything we need, and we can just hit "Compute."

ency Table		×
Grades Grades		
Where:		
optional	Build	
optional Statistic(s): Frequency Relative frequency Percent of total Cumulative frequency Cumulative relative frequency Order by: Value Ascending		
optional Output: ✓ Store in data table		
? Cancel	Comp	ute!

Figure 12: Frequency Window

StatCr	unch 🕶	Арр	olets 👻 Ed	lit 🕶	Data 🖥	✓ Stat •	Graph 👻 H	lelp ∙
Row	Grades		Frequency	Relat	ive Freque	Percent of Tota	Cumulative Fre	Cumulative Re
1		А	13	0.1	19402985	19.402985	13	0.19402985
2		В	17	0.2	25373134	25.373134	30	0.44776119
3		С	19	0.2	28358209	28.358209	49	0.73134328
4		D	10	0.1	14925373	14.925373	59	0.88059701
5		F	8	0.1	11940299	11.940299	67	1
6								
-								

Figure 13: Final Results

Ok, so now we can see that StatCrunch has created a bunch of new columns for us. We have one listing the Frequency values, one listing Cumulative

Frequency, one listing the Relative Frequency values, and one listing Relative Cumulative Frequency values. And notice again that Relative Cumulative Frequency adds up to 1. And also, notice that the Cumulative Frequency values, here, total to 67. And that is the total number of grades that we had in our original data set.

Outro

Ok, so I hope you found this video helpful. If you are a UNT student, there will be some more links to other resources in the video description. Thank you for watching!

References

All figures were produced using StatCrunch 2019, Pearson Education Inc.