Frequency Test

How many cases are there in each polatt category?

Before we can transform these variables for our use, we should check the frequencies of the responses in the **polatt** categories. Frequency tabulations allow you to see how many individuals (or cases) fall into each category of the variables you are interested in. In addition, they also let you check for inconsistencies in the information entered into your variables. After you calculate frequencies, you could find that you have a lot of missing data or that missing data has been entered into the variable with numeric codes like "8" or "9," which serve as place holders in variables when information doesn't exist for certain individuals. (Please note that the missing data codes can change between surveys, although 9 or 99 are standard.) This missing data could create problems in your analyses, so it's best to search it out before you begin running tests.

To check the frequencies of the **polatt** variables:

Click on **Analyze**, **Descriptive Statistics**, and **Frequencies**. Find **polatt1-7** in the variable list on the left side of the **Frequencies** dialogue box. Move all six **polatt** variables to the text box on the right side. You can easily search for variable names in dialogue boxes by right-clicking on the list of variables and selecting **Display Variable Names** and **Sort Alphabetically**. (This is a trick you can use in SPSS all the time – it makes finding variables in large datasets much easier!)



When all six **polatt** variables are in the **Variable(s)** text box, click **OK**.

Now you should see six distinct frequency tables in the SPSS Output window, one for each of the **polatt** variables. And each one of these frequency tables has, in addition to the 5 possible question answers, an 8 or a 9 (or both listed) as one of the values of the variable.

	Th	e police in this area would tr	eat you with r	espect if yo	u had contact w	ith them
			Frequency	Percent	Valid Percent	Cumulative Percent
	Valid	Strongly agree	13798	30.0	30.0	30.0
		Tend to agree	25759	56.0	56.0	85.9
		Neither agree nor disagree	3990	8.7	8.7	94.6
		Tend to disagree	1509	3.3	3.3	97.9
		Strongly disagree	600	1.3	1.3	99.2
		9	371	.8	.8	100.0
		Total	46027	100.0	100.0	
	Missing	System	4	.0		
	Total		46031	100.0		

The police in this area treat everyone fairly regardless of who they are

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Strongly agree	6939	15.1	15.1	15.1
	Tend to agree	22785	49.5	49.5	64.6
	Neither agree nor disagree	9547	20.7	20.7	85.3
	Tend to disagree	3964	8.6	8.6	93.9
	Strongly disagree	1184	2.6	2.6	96.5
	8	1	.0	.0	96.5
	9	1607	3.5	3.5	100.0
	Total	46027	100.0	100.0	
Missing	System	4	.0		
Total		46031	100.0		

These are the numeric codes for missing values. We'll need to recode these before we can create our new continuous variable, as these numeric missing values skew our data. If we include them when we calculate our new confidence variable some people who actually did not answer the question will receive a score of 8 or 9, when they should be removed from the analysis. Luckily, it is very easy to recode these missing values.

First, select Transform and Recode into Same Variables.

ta *cse	w_apr1	1mar12_	_nvf.sav	DataSet6] - IB	M SPSS Sta	tistics Data Editor							
File	<u>E</u> dit	View	<u>D</u> ata	<u>T</u> ransform	<u>A</u> nalyze	Direct <u>M</u> arketing	<u>G</u> raphs	s <u>U</u> tilities	Add- <u>o</u> ns	<u>W</u> indow	<u>H</u> elp		
😑 🗄 🖨 🛄		Compute Variable		aa I		4	⊉ 🛄)				
		Na	ime	Shift V:	Shift Voluce				Label				
43	8	comm	att2	Sinit Va	iues			n do you ag	ree or disag	ree that the	police and	l local council .	
43	9	aimpri	3	Recod	e into <u>S</u> ami	e Variables		you think s	hould be the	e highest pr	riority for the	e Criminal Jus.	
44	0	aimpri	4	🜌 <u>R</u> ecod	e into Differ	ent Variables		you think s	hould be the	e SECOND	highest pri	ority for the Cr.	
44	1	secr12	2ma	🛐 <u>A</u> utoma	atic Recode	a		een happen	ing: Someo	ne vandalisi	ing property	/ or a vehicle	
44	2	secr12	?mb	Visual	Binning			een happen	ing: Someo	ne stealing	a vehicle or	r something fr	
44	3	secr12	2mc	提 Optima	I Binning			een happen	ing: Threate	ning or viol	ent behavio	ur including fig.	
44	4	secr12	2md	Prepar	e Data for I	lodelina	•	een happen	ing: Someo	ne being m	ugged or ro	bbed	
44	5	secr12	2me	Rank (een happen	ing: Someo	ne breaking	, into or atte	empting to bre.	
44	6	secr12	?mf	BE Kank C				een happening: Someone shoplifting					
44	7	secr12	ma	Date and		zard		een happen	ing: None o	fthese			

Find **polatt1-7** in the variable list on the left and move the six variables into the text box on the right. **Select Old and New Values**.

Recode into Same Varial	bles
	Numeric <u>V</u> ariables:
plfeasm policeconf1 polinf1 polinf2 polmap3	 polatt1 polatt2 polatt3 polatt5 polatt6
polmap4 polvola polvolb polvolc polvolc	Old and New Values
ОК	Paste Reset Cancel Help

We want to code the values "8" and "9" in **polatt1-7** as "system missing," so that they are not used as actual data points in our analysis of our new continuous variable **policeconf1**. So, in the **Old Value** text box on the upper left, enter **8**, and click **System Missing** under the **New Value** header on the right. Click **Add**, which will **enter 8** \rightarrow **SYSMIS** into the **Old** \rightarrow **New** text box.

Recode into Same Variables: Old and New Values Old Value Value: System- or user-missing Range: through Range, LOWEST through value: Range, value through HIGHEST:	New Value Value System-missing Old> New: Add Change Remove
Range, value through HIGHEST: All other values	
Continue	Cancel Help

Repeat these steps for "9." Your **Recode into Same Variables: Old and New Values** dialogue box should look like this:

Recode into Same Variables: Old and New Values	Nagana party to be St. 1 Manage Same
Old Value	New Value Value System-missing Old> New: > SYSMIS > SYSMIS> S
Continue	Cancer help

Click **Continue**. And then click **OK**. Your SPSS Output should let you know that you've now recoded "8" and "9."

RECODE polatt1 polatt2 polatt3 polatt5 polatt6 polatt7 (8=SYSMIS) (9=SYSMIS). EXECUTE.

Now that we've checked for (and corrected) data inconsistencies in the **polatt** variables, we can use them to create one new variable we'll use as our dependent variable for this section.

To begin building this new variable:

Click Transform and Compute Variable.

📬 csew_apr11mar12_nvf.sav [DataS-x5] - 1004-5055 Statistics Data Editor								
<u>F</u> ile <u>E</u> dit	<u>V</u> iew <u>D</u> ata	Transform Analyze Direct Marketing Graph	ns <u>U</u> tilities Add- <u>o</u> ns <u>W</u> indow <u>H</u> elp					
😂 🖩 🖨 🛽		Compute Variable	🖌 📗 🖾 🚍 🖧 🚍 📲 🛇					
	Name	Shift Values	Label	Values				
438	commatt2		-ch do you agree or disagree that the police and local council	{1, Strongly				
439	aimpri3	Recode into Same Variables	o you think should be the highest priority for the Criminal Jus	{1, Meeting				
440	440 aimpri4 Recode into Different Variables		o you think should be the SECOND highest priority for the Cr	{1, Meeting				
441	secr12ma	Mutomatic Recode	seen happening: Someone vandalising property or a vehicle	{0, no Some				
442	secr12mb	Visual <u>B</u> inning	seen happening: Someone stealing a vehicle or something fr	{0, no Some				
443	secr12mc	🔀 Optimal Binning	seen happening: Threatening or violent behaviour including fig	{0, no Threa				

Because we are using **polatt1-7** to create a new continuous variable concerning confidence in the police, we can enter **policeconf1** into the **Target Variable** box on the upper left. This will be the name of our new continuous variable.



Next, find **polatt1-polatt7** in the variable list on the left.

Move **polatt1** over to the **Numeric Expression** text box using the blue arrow or by double-clicking the variable name. Using the keypad in the dialogue box, enter a + sign after **polatt1**. Then, repeat these steps until all of the **polatt** variables are in the **Numeric Expression** box, each followed by a + symbol. (There will be no polatt4. This variable does not exist in this dataset – it could have been removed from the survey questionnaire for 2011-2012.) Your dialogue box should look like this:

tai Compute Variable	A		— ×
Target Variable: policeconf1 Type & Label plfeask plfeask plfeasm plfeasm plfeasm	Vumeric Expression: polat1 + polat2 + polat3 + po	ilatt5 + polatt6 + polatt7	Function group:
polatt polatt polatt polatt5 polatt5 polatt5 polatt7 polatt7 polatt7 polatt7 polatt7 polatt7 polatt7 polatt7 polatt7 polatt7	+ < > / - <= >= 4 + = ~= 1 / & 1 () + ~ ()	8 9 5 6 2 3 0 . Delete	Arithmetic CDF & Noncentral CDF Conversion Current Date/Time Date Arithmetic Date Creation
polvola polvolb polvolc polvolc polvolc polvole polvole polvole polvole polvole			

Click **OK**. The SPSS Output window should appear, telling you that you have computed **polatt1+polatt2+polatt3+polatt5+polatt6+polatt7**:

```
COMPUTE policeconf1=polatt1 + polatt2 + polatt3 + polatt5 + polatt6 + polatt7.
EXECUTE.
```

Now, we'll run a frequency statistic to see if our new continuous variable has been successfully created. Go to **Analyze**, **Descriptive Statistics**, and **Frequencies**. In the **Frequencies** dialogue box that appears, scroll to the very end of the list of variables. Our new variable, **policeconf1**, should be at the very end of the list, because it's the latest variable to be created.

Frequencies	and the second second	×					
 Reporting rate Alt Reporting rate Alt Respondent oth HRP other countr Respondent oth HRP other nation 	Variable(s):	<u>Statistics</u> Charts <u>F</u> ormat <u>B</u> ootstrap					
✓ Dispray inequency tables OK Paste Reset Cancel Help							

Move **policeconf1** to the **Variable(s)** text box. Click **OK**. The SPSS Output window should open, displaying a list of value frequencies for **policeconf1**.

This list of value frequencies begins with 6.00, which makes sense because we have just combined six variables in which the first possible value was 1.00 (as in 1=strongly agree). This first value, 6.00, reflects all the respondents who answered "1=strongly agree" to all six **polatt** variables.

Our list of value frequencies ends at 30.00, because respondents had only 5 response choices (1=strongly agree to 5=strongly disagree) over only 6 questions. The highest possible combined variable value would be 5.00 multiplied by 6 possible questions.

policeconf1							
	Frequency	Percent	Valid Percent	Cumulative Percent			

Practical Applications of Statistics in the Social Sciences – University of Southampton 2014

				-	
_	6.00	1882	4.1	4.4	4.4
	7.00	1372	3.0	3.2	7.6
	8.00	1402	3.0	3.3	10.9
	9.00	1745	3.8	4.1	15.0
	10.00	2189	4.8	5.1	20.2
	11.00	3153	6.8	7.4	27.6
	12.00	7933	17.2	18.6	46.2
	13.00	4813	10.5	11.3	57.5
	14.00	4115	8.9	9.7	67.1
	15.00	2932	6.4	6.9	74.0
	16.00	2368	5.1	5.6	79.6
	17.00	1957	4.3	4.6	84.2
	18.00	1776	3.9	4.2	88.3
Valid	19.00	1057	2.3	2.5	90.8
	20.00	943	2.0	2.2	93.0
	21.00	646	1.4	1.5	94.6
	22.00	608	1.3	1.4	96.0
	23.00	444	1.0	1.0	97.0
	24.00	345	.7	.8	97.8
	25.00	250	.5	.6	98.4
	26.00	188	.4	.4	98.9
	27.00	136	.3	.3	99.2
	28.00	96	.2	.2	99.4
	29.00	80	.2	.2	99.6
	30.00	174	.4	.4	100.0
	Total	42604	92.6	100.0	
Missina	System	3427	7.4		
	Total	46031	100.0		

Here we have variable values from 6.00 to 30.00, which is exactly what we might expect after computing our six categorical **polatt** variables into one continuous **policeconf1** variable. We can take 6.00 to mean "strongly agree" and 30.00 to mean "strongly disagree."

We can go back to **Variable View**, and scroll to the end of the list of variables. We should find **policeconf1** at the very end, because it is the newest variable in the dataset. Now we can enter in a label for our variable by double-clicking on the **Label** cell and typing in something like, "**I have confidence in the police**," a statement that our respondents either agree or disagree with, in varying degrees.

Summary

In this section you have studied the variables related to your research question – confidence in the police – and investigated if there was missing data. You have also created a new variable which measures confidence by combining six related variables, after first recoding the missing data so it doesn't affect your results. This new variable can be treated as continuous. Now we can explore this new confidence variable.

***Note: as we are making changes to a dataset we'll continue using for the rest of this section, please make sure to save your changes before you close down SPSS. This will save you having to repeat sections you've already completed!