

## SPSS DEMONSTRATION [GSS18SSDS-B]

### Producing Z Scores With SPSS

In this chapter, we have discussed the theoretical normal curve, *Z* scores, and the relationship between raw scores and *Z* scores. The SPSS Descriptives procedure can calculate *Z* scores for any distribution. We'll use it to study the distribution of education in the GSS18SSDS-B file. Locate the Descriptives procedure in the Analyze menu, under Descriptive Statistics, then click on *Descriptives*. We can select one or more variables to place in the Variable(s) box; for now, we'll just place EDUC (years of education) in this box. Check the box in the bottom left corner (save standardized values as variables), to create standardized values, or *Z* scores, as new variables. Any new variable is placed in a new Column in the Data View window and will then be available for additional analyses. Click on *OK* to run the procedure.

The output from Descriptives (Figure 4.14) is brief, listing the mean and standard deviation for EDUC, plus the minimum and maximum values and the number of valid cases.

Although not indicated in the Output window, SPSS has created a new *Z*-score variable for EDUC. To see this new variable, switch to the Data View screen. Then, go to the last column, scrolling all the way to the right (Figure 4.15). By default, SPSS appends a *Z* to the variable name, so the new variable is called ZEDUC. The first case in the file has a *Z* score of .09379, so the education years for this person must be above the mean of 13.71. If we locate the respondent's EDUC score, we see that the score for this person was 14 (not pictured), or above the mean as we expected. There may be missing values for ZEDUC. Missing values include those that the question did not apply to, those that outright refused to provide occupational information, or those data entry errors that resulted in an invalid answer.

If the data file is saved, the new *Z*-score variable will be saved along with the original data and then can be used in analyses. In addition, if we have SPSS calculate the mean and standard deviation of ZEDUC, we find that they are equal to 0 and 1.00, respectively.

**Figure 4.14 Descriptives for Education**

	N	Minimum	Maximum	Mean	Std. Deviation
Highest year of school completed	1498	0	20	13.71	3.039
Valid N (listwise)	1498				

**Figure 4.15 The Creation of the Z-Score Variable for EDUC**

 WORKBLK S	 WORKWH TS	 ZEDUC
6	4	.09379
0	0	-1.22235
4	4	.75186
4	4	.75186
0	0	1.40993
0	0	-.23524
0	0	-.56428
2	3	-1.88042
4	4	-.56428
4	4	1.73897

## SPSS PROBLEMS [GSS18SSDS-B]

- S1. The majority of variables that social scientists study are not normally distributed. This doesn't typically cause problems in analysis when the goal of a study is to calculate means and standard deviations—as long as sample sizes are greater than about 50. (This will be discussed in later chapters.) However, when characterizing the distribution of scores in one sample, or in a complete population (if this information is available), a nonnormal distribution can cause complications. We can illustrate this point by examining the distribution of age in the GSS data file.
- Create a histogram for AGE. Click on *Analyze*, *Descriptive Statistics*, *Frequencies* (select AGE), *Charts*, and *Histograms* (select *Show normal curve on histogram*). How does the distribution of AGE deviate from the theoretical normal curve?
  - Calculate the mean and standard deviation for AGE in this sample, using either the *Frequencies* or *Descriptives* procedure.

- c. Assuming the distribution of AGE is normal, calculate the number of people who should be 25 years of age or less.
  - d. Use the Frequencies procedure to construct a table of the percentage of cases at each value of AGE. Compare the theoretical calculation in (c) with the actual distribution of age in the sample. What percentage of people in the sample are 25 years old or less? Is this value close to what you calculated?
- S2. In the SPSS Demonstration, we examined the distribution of EDUC (years of school completed).
- a. What is the equivalent  $Z$  score for someone who has completed 18 years of education?
  - b. Use the Frequencies procedure to find the percentile rank for a score of 18.
  - c. Does the percentile rank that you found from Frequencies correspond to the  $Z$  score for a value of 18? In other words, is the distribution for years of education normal? If so, then the  $Z$  score that SPSS calculates should be very close, after transforming it into an appropriate area, to the percentile rank for that same score.
  - d. Create histograms for EDUC and the new variable ZEDUC. Explain why they have the same shape.
- S3. Run separate analyses for men versus women (SEX) based on the variables EDUC and ZEDUC (click on *Data, Split File, Organize Output by Groups*, and select SEX). Answer Questions S2a–d separately for men and women.

## EXCEL DEMONSTRATION [GSS18SSDS-E]

### Producing $Z$ Scores With Excel

We will use the CHILDS (number of children) to demonstrate how to use Excel to produce  $Z$  scores. Copy the CHILDS data from the protected Data View sheet and paste it into a new Excel sheet. Our copied data were placed in our new sheet from A1 to A136. In cell B2 of this new sheet, we've typed "Z Scores (CHILDS)." We also expanded the B column so that the full text in the cell B2 is visible. It's important to be organized when working with data.

Figure 4.16

	A	B	C	D	E	F
1	CHILDS	Z Scores (CHILDS)				
2	1			<b>CHILDS, GSS18SSDS-E</b>		
3	1					
4	3			Mean	2.10447761	
5	3			Standard Error	0.14419087	
6	4			Median	2	
7	2			Mode	3	
8	0			Standard Deviation	1.66912996	
9	2			Sample Variance	2.78599484	
10	0			Kurtosis	-0.1128752	
11	4			Skewness	0.56087322	
12	2			Range	7	
13	1			Minimum	0	
14	3			Maximum	7	
15	1			Sum	282	
16	3			Count	134	
17	3				-7.605E-52	
18	4					

Before we can ask Excel to calculate a *Z* score for each respondent in our data set, we need to find the mean and standard deviation for CHILDS. We will use the *Data Analysis* function to obtain these values. For assistance working with *Data Analysis*, see the Excel Demonstrations in Chapters 3 and 4. Our Descriptive Statistics table begins in cell D2. In addition to other values, it includes the mean (2.10) and standard deviation (1.67) for CHILDS (see Figure 4.16). We are now ready for Excel to calculate *Z* scores of our variable.

In cell B2, we will write an Excel equation that will calculate the *Z* score of CHILDS for each respondent in our data set (see Figure 4.17). The formula is as follows (=A2-\$E\$4/\$E\$8).

- = asks Excel to run our equation
- A2 is the location of the value for our first participant
- indicates subtraction

Figure 4.17

	A	B	C	D	E	F	G
1	CHILDS	Z Scores (CHILDS)					
2	1	-0.260823098		<b>CHILDS, GSS18SSDS-E</b>			
3	1						
4	3			Mean	2.10447761		
5	3			Standard Error	0.14419087		
6	4			Median	2		
7	2			Mode	3		
8	0			Standard Deviation	1.66912996		
9	2			Sample Variance	2.78599484		
10	0			Kurtosis	-0.1128752		
11	4			Skewness	0.56087322		
12	2			Range	7		
13	1			Minimum	0		
14	3			Maximum	7		
15	1			Sum	282		
16	3			Count	134		
17	3				-7.605E-52		
18	4						
19	2						

**\$E\$4** E4 indicates where in our Excel sheet the mean value appears. The dollar sign before and after the “E” in “E4” lets Excel know that it should always use the value in E4 as the mean regardless of where we place the equation. *Note:* E4 is where we’ve placed the mean in our sheet. You may find the mean in a different cell in your sheet. Adjust the cell letter and number accordingly.

**\$E\$8** E8 indicates where in our Excel sheet the standard deviation value appears. The dollar signs before and after E8 let Excel know that it should always use the value in E8 as the standard deviation regardless of where we place the equation. *Note:* E8 is where we’ve placed the standard deviation in our sheet. You may find the standard deviation in a different cell in your sheet. Adjust the cell letter and number accordingly.

The Z score for a respondent who has one child is  $-0.26$  (rounded). You can click on that cell and drag the equation you wrote down to cell B136. Doing such will ask Excel to calculate the Z score of CHILDS for every respondent in our data set (see Figure 4.18).

Figure 4.18

	A	B	C	D	E	F
1	CHILDS	Z Scores (CHILDS)				
2	1	-0.260823098		<b>CHILDS, GSS18SSDS-E</b>		
3	1	-0.260823098		Mean	2.10447761	
4	3	1.739176902		Standard Error	0.14419087	
5	3	1.739176902		Median	2	
6	4	2.739176902		Mode	3	
7	2	0.739176902		Standard Deviation	1.66912996	
8	0	-1.260823098		Sample Variance	2.78599484	
9	2	0.739176902		Kurtosis	-0.1128752	
10	0	-1.260823098		Skewness	0.56087322	
11	4	2.739176902		Range	7	
12	2	0.739176902		Minimum	0	
13	1	-0.260823098		Maximum	7	
14	3	1.739176902		Sum	282	
15	1	-0.260823098		Count	134	
16	3	1.739176902				
17	3	1.739176902			-7.605E-52	
18	4	2.739176902				
19	2	0.739176902				
20	3	1.739176902				
21	1	-0.260823098				
22	5	3.739176902				
23	1	-0.260823098				
24	0	-1.260823098				
25	5	3.739176902				
26	1	-0.260823098				
27	0	-1.260823098				
28	1	-0.260823098				
29	2	0.739176902				
30	3	1.739176902				
31	2	0.739176902				
32	4	2.739176902				
33	0	-1.260823098				
34	3	1.739176902				

## EXCEL PROBLEM [GSS18SSDS-E]

- E1. For this Excel problem, we will be working with CHLIDDEL (ideal number of children).
- Use the Excel *Data Analysis* function to produce descriptive statistics for CHLIDDEL.
  - What is the value of the mean? The standard deviation?
  - Have Excel produce the *Z* score of CHLIDDEL for all participants.  
*Note:* Excel will calculate a *Z* score for every respondent, including those with missing CHLIDDEL data (missing data are represented in each cell with gray shading).
  - What is the *Z* score for someone who feels the ideal number of children is 2?