

Weighting Cases in SPSS

Donald Stephen
Institute of Borneo Studies, Universiti Malaysia Sarawak

Suppose your data are already in a frequency table and you want to use SPSS to analyze them (like exercise 2 below). You could enter the score “1” one time, the score “2” two times, the score “3” three times and so on, but there is a much easier way to enter the data.

1. Below are the exam scores of a group of students. Please construct a bar graph and pie chart based on the data given in the frequency table:

Frequency Table

Group	Count
0-9	1
10-19	2
20-29	3
30-39	4
40-49	5
50-59	4
60-69	3
70-79	2
80-89	2
90-99	1

Manually this is how you enter data

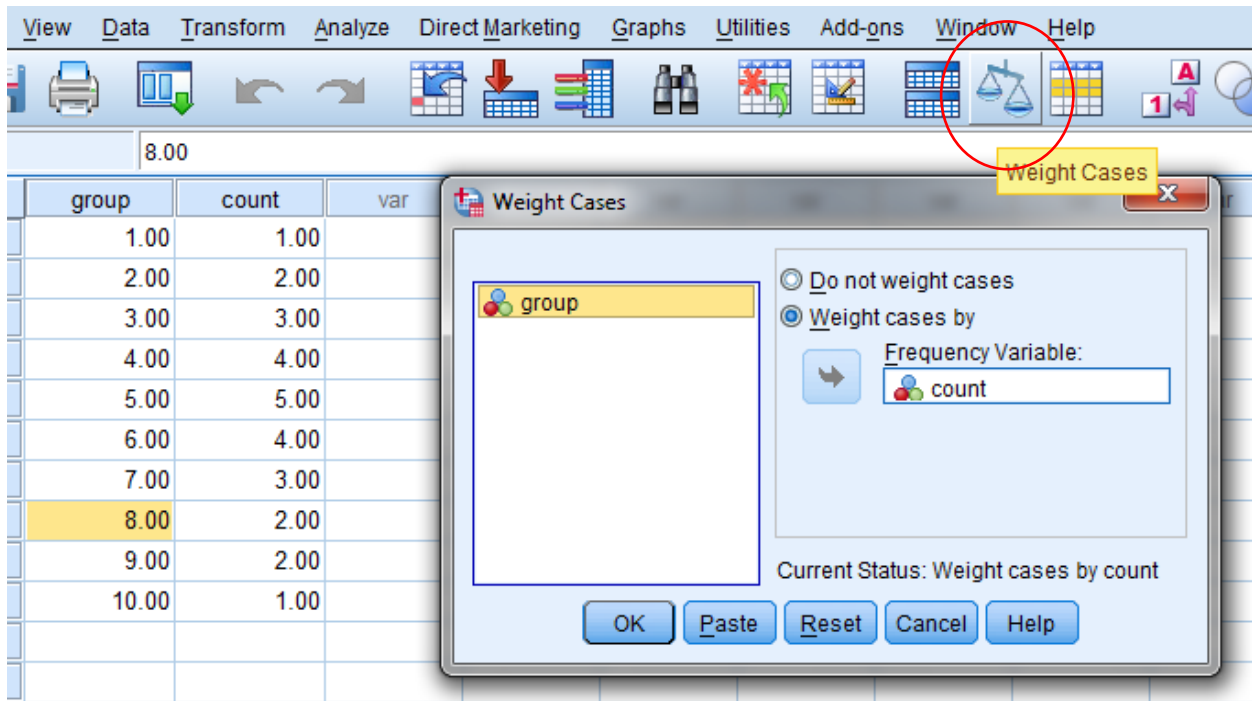
	group	var
1	1.00	
2	2.00	
3	2.00	
4	3.00	
5	3.00	
6	3.00	
7	4.00	
8	4.00	
9	4.00	
10	4.00	
11	5.00	
12	5.00	
13	5.00	
14	5.00	
15	5.00	
16	6.00	
17	6.00	
18	6.00	
19	6.00	
20	7.00	
21	7.00	
22	7.00	
23	8.00	
24	8.00	
25	9.00	
26	9.00	
27	10.00	
28		

But you can also do this:

	group	count	var
1	1.00	1.00	
2	2.00	2.00	
3	3.00	3.00	
4	4.00	4.00	
5	5.00	5.00	
6	6.00	4.00	
7	7.00	3.00	
8	8.00	2.00	
9	9.00	2.00	
10	10.00	1.00	
11			

Create two variables in SPSS, “group” and “count” and enter the data. You data should look like the table above. Of course in both cases you need to define value labels 0-9: value 1, 10-19 value 2 and so on.

If you are doing method #2, click on **Data**, select **Weight Cases**. Select **Weight cases by** and put in the “count” variable. Click OK. (Alternately, you can get to weight case option by clicking the weight case shortcut icon that looks like weight scale)



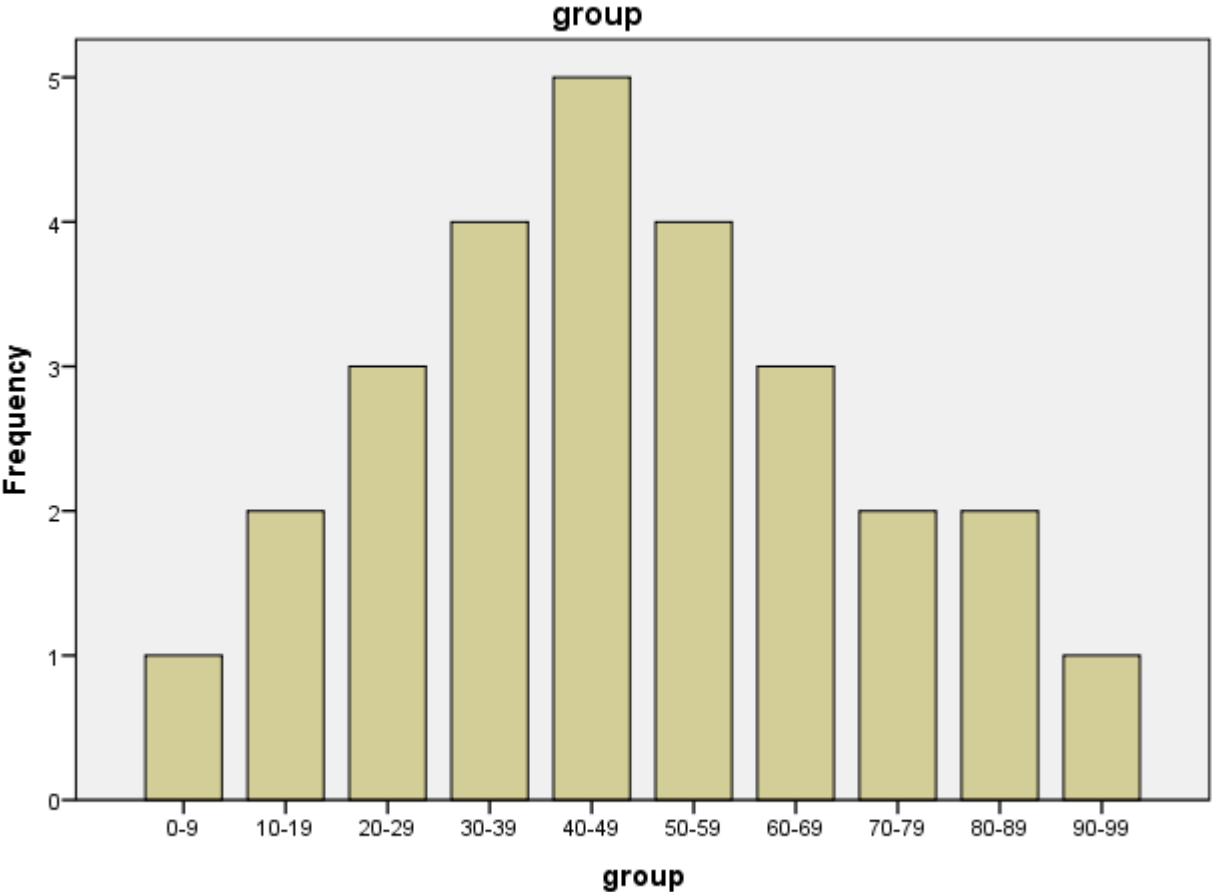
To answer the question, go back to your data and you are ready to do histogram.

group	count	var	var	var	var	var	var	var
1.00	1.00							
2.00	2.00							
3.00	3.00							
4.00	4.00							
5.00	5.00							
6.00	4.00							
7.00	3.00							
8.00	2.00							
9.00	2.00							
10.00	1.00							

Analyze -> descriptive statistics -> frequencies -> put in "group" into **variable** click "chart" -> bar charts

Compute mean and SD if you want

Regardless of whether you use method #1 or #2, you should get this



Cases weighted by count

QUESTION 3 EXERCISE 1

3. The data below is obtained from one section of a questionnaire that investigate the level of satisfaction of students with services in CAIS. The variable is " Book Collection" = BC

X1 = The collection of books in CAIS is good
 X2 = The collection of books in CAIS is up to date
 X3 = Many of the books are lost
 X4 = The books are well arranged
 X5 = Many of the books are torn and dirty
 X6 = Books not found can be easily obtained through interlibrary loan

the subject we want to measure.

"BC"	VS (1)	S (2)	N (3)	DS (4)	VDS (5)	Total Sample
X1	4	5	3	3	5	20
X2	2	4	4	3	7	20
X3	1	2	5	2	10	20
X4	4	4	2	3	7	20
X5	2	5	3	5	5	20
X6	6	4	2	6	2	20

(1) – Very satisfied (5) = Very Dissatisfied

a) Computing the overall mean and standard deviation from the data set above.
 b) Identify which of the sub question (X1, X2,.....) needs to be recoded.

IMPORTANT: ANSWER 3(b) FIRST, THEN 3(a).

ans 3(a) :X3,X5

Typically, this is how we enter this data:

	X1	X2	X3	X4	X5	X6	X3_recoded	X5_recoded
1	1.00	1.00	1.00	1.00	1.00	1.00	5.00	5.00
2	1.00	1.00	2.00	1.00	1.00	1.00	4.00	5.00
3	1.00	2.00	2.00	1.00	2.00	1.00	4.00	4.00
4	1.00	2.00	3.00	1.00	2.00	1.00	3.00	4.00
5	2.00	2.00	3.00	2.00	2.00	1.00	3.00	4.00
6	2.00	2.00	3.00	2.00	2.00	1.00	3.00	4.00
7	2.00	3.00	3.00	2.00	2.00	2.00	3.00	4.00
8	2.00	3.00	3.00	2.00	3.00	2.00	3.00	3.00
9	2.00	3.00	4.00	3.00	3.00	2.00	2.00	3.00
10	3.00	3.00	4.00	3.00	3.00	2.00	2.00	3.00
11	3.00	4.00	5.00	4.00	4.00	3.00	1.00	2.00
12	3.00	4.00	5.00	4.00	4.00	3.00	1.00	2.00
13	4.00	4.00	5.00	4.00	4.00	4.00	1.00	2.00
14	4.00	5.00	5.00	5.00	4.00	4.00	1.00	2.00
15	4.00	5.00	5.00	5.00	4.00	4.00	1.00	2.00
16	5.00	5.00	5.00	5.00	5.00	4.00	1.00	1.00
17	5.00	5.00	5.00	5.00	5.00	4.00	1.00	1.00
18	5.00	5.00	5.00	5.00	5.00	4.00	1.00	1.00
19	5.00	5.00	5.00	5.00	5.00	5.00	1.00	1.00
20	5.00	5.00	5.00	5.00	5.00	5.00	1.00	1.00

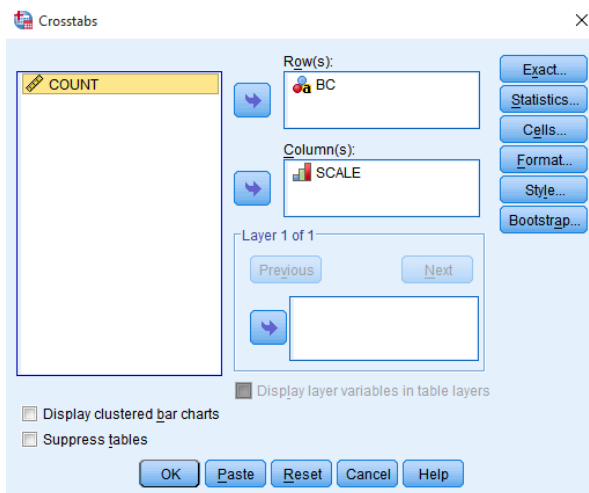
However you can also use weight cases method. This is how you enter:

	BC	SCALE	COUNT
1	X1	VS	4.00
2	X1	S	5.00
3	X1	N	3.00
4	X1	DS	3.00
5	X1	VDS	5.00
6	X2	VS	2.00
7	X2	S	4.00
8	X2	N	4.00
9	X2	DS	3.00
10	X2	VDS	7.00
11	X3	VS	10.00
12	X3	S	2.00
13	X3	N	5.00
14	X3	DS	2.00
15	X3	VDS	1.00
16	X4	VS	4.00
17	X4	S	4.00
18	X4	N	2.00
19	X4	DS	3.00
20	X4	VDS	7.00
21	X5	VS	5.00
22	X5	S	5.00
23	X5	N	3.00
24	X5	DS	5.00
25	X5	VDS	2.00
26	X6	VS	6.00
27	X6	S	4.00
28	X6	N	2.00
29	X6	DS	6.00
30	X6	VDS	2.00

You can use value label for BC and SCALE. This one I activated the value label button so you don't see numbers for BC and SCALE

Important: note that I have already reverse the count for X3 and X5 so no need to recode variable.

Now we can run our crosstab and verify your data. Click **Analyze > Descriptive Statistics > Crosstabs.**



click ok

BC * SCALE Crosstabulation

Count

		SCALE					Total
		VS	S	N	DS	VDS	
BC	X1	4	5	3	3	5	20
	X2	2	4	4	3	7	20
	X3	10	2	5	2	1	20
	X4	4	4	2	3	7	20
	X5	5	5	3	5	2	20
	X6	6	4	2	6	2	20
Total		31	24	19	22	24	120

Compare with the original data you will notice the score for 3 and 5 is reversed. Others maintain. This step is just to make sure you have the correct data.

To answer the question 3(a) [find mean and SD], just do mean for SCALE.

Analyze->descriptive statistics ->frequencies, then put in SCALE.

(You don't need to do any horizontal mean if you use this method, because the variable SCALE already include ALL questions X1-X6)

Regardless of your method, your M should be the same. Just that you will have different frequency table.

Difference in SD is because you are computing horizontal mean in method 1, so your dispersion is based on the horizontal mean, not the scale 1-5. SD for method 2 is based on dispersion of likert scale.

Statistics		
Mean		
N	Valid	20
	Missing	0
Mean		2.8667
Std. Deviation		.57634

Method 1: Composite score
(mean)

Statistics		
SCALE		
N	Valid	120
	Missing	0
Mean		2.8667
Std. Deviation		1.48908

Method 2: weight

Mean					
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	2.00	1	5.0	5.0	5.0
	2.17	2	10.0	10.0	15.0
	2.33	3	15.0	15.0	30.0
	2.50	2	10.0	10.0	40.0
	2.67	2	10.0	10.0	50.0
	2.83	2	10.0	10.0	60.0
	3.17	1	5.0	5.0	65.0
	3.50	5	25.0	25.0	90.0
	3.67	2	10.0	10.0	100.0
	Total		20	100.0	100.0

SCALE					
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	VS	31	25.8	25.8	25.8
	S	24	20.0	20.0	45.8
	N	19	15.8	15.8	61.7
	DS	22	18.3	18.3	80.0
	VDS	24	20.0	20.0	100.0
Total		120	100.0	100.0	

End

Disclaimer: This note is for educational purpose only and author does not make profit from this. SPSS statistics is a property of IBM. Support the software by purchasing legal copy of it.