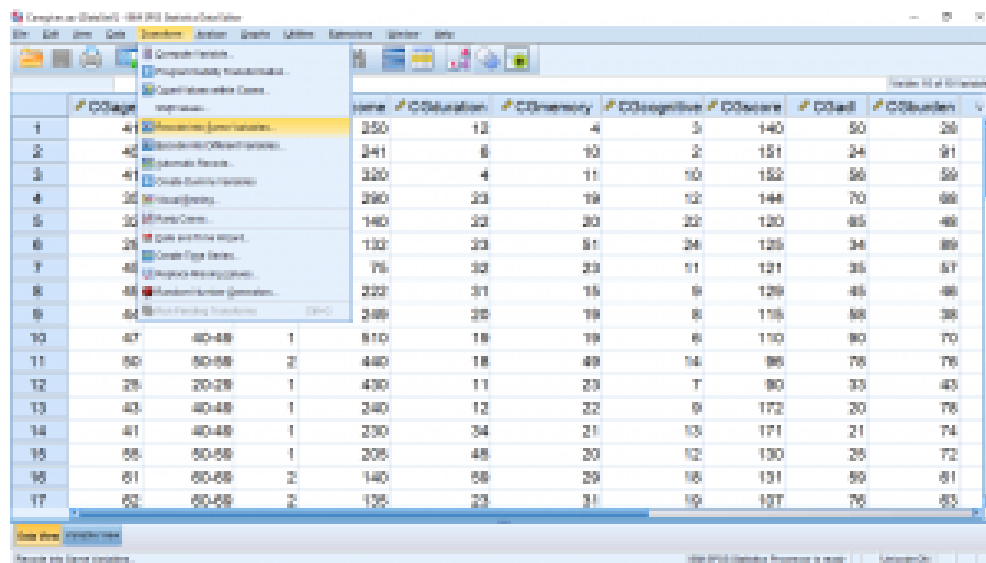


3.4: SPSS Lesson 2- Combining variables and recoding

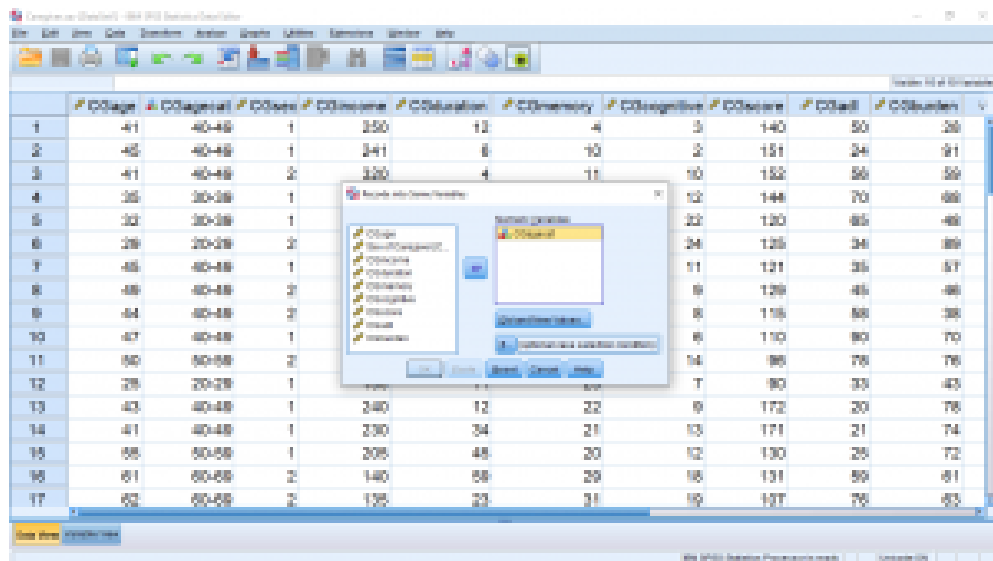
Frequently data collection results in a collection of many variables. This happens, for example, with tests or surveys where people answer questions on a 5 or 7 point *Lickert scale* where questions range from, say, “strongly agree” to “somewhat agree” to ... to “strongly disagree”. A bunch of those questions may refer to, say, happiness and adding up the scores, perhaps averaging them, will lead to a single variable, one dependent variable, that becomes our measurement of happiness. This gives us not only a univariate variable that we can subject to a statistical test but likely gives us a stronger and more reliable measurement of happiness. A problem with combining variables in this way arises if the response “1” for “strongly agree” means happiness for one question (e.g. “I wake up happy”) and sadness in another question (e.g. “I go to bed sad”). In such a situation some of the variables will need to be reverse-scaled or *recoded* before they can be added. Let’s see how to combine and recode variables in SPSS.

Open the file “Caregiver.sav” from the textbook [Data Sets](#). This dataset is about the different attributes of diamonds such as its color, price, carat, cutting quality etc. Here one of the variables is cut_new which basically represents the cutting quality of diamond and takes values from 1 to 5 depending on the cutting quality with 5 being the best quality. Now let’s assume that we need to reverse scale this variable to use it in other calculations in a meaningful manner. To recode cut_new first open the Transform → Recode in Same Variables... menu :



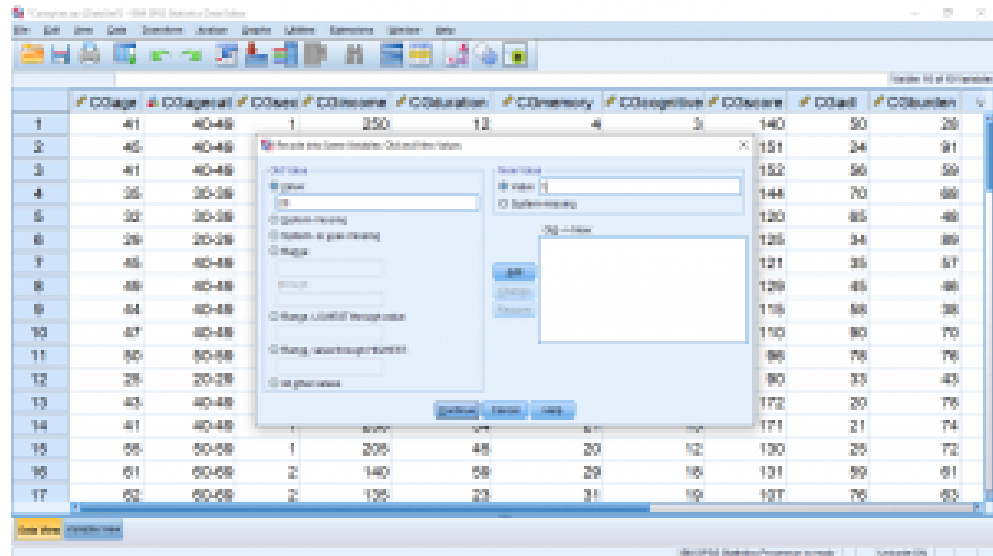
SPSS screenshot © International Business Machines Corporation.

You can choose the Recode into Different Variables... if you want to, instead. That choice will lead to the creation of a new variable that you would use in place of cut_new for your analysis. With our choice of Recode in Same Variables... we will overwrite the old values of cut_new with new ones. (This is a danger if you make a mistake.) Our job is now to map 1 to 5, 2 to 4, 3 to 3, 4 to 2 and 5 to 1, recoding the variable. First move the cut_new variable over in the pop up menu :



SPSS screenshot © International Business Machines Corporation.

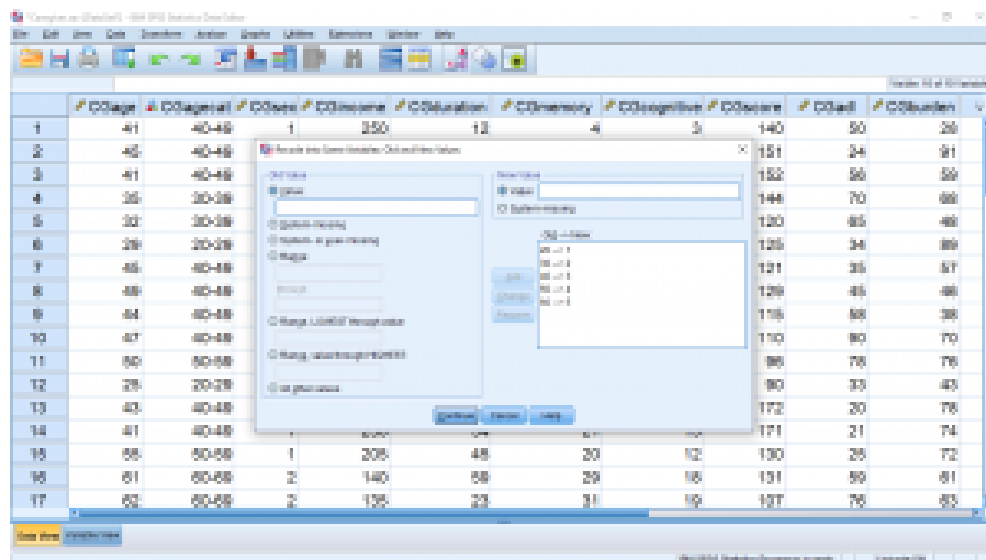
then hit the Old and New Values.. button that will bring up a new pop up menu. Next enter 1 under Old Value and 5 in New Value :



SPSS screenshot © International Business Machines Corporation.

then hit Add :

Continue this way to complete the recoding list :



Now suppose we want to add multiple variables to create a new variable. Let's open the dataset Caregiver from the course website. This dataset is regarding the test scores of students from diverse background in UK. Here we will add the test scores of read, write, math and science to create a new variable totalscore. Pick the Transform → Compute Variable... menu :

	score	CCEvaluation	CCEmergency	CCEognitive	CCEscore	CCEall	CCEurten
1	41	350	13	4	3	140	50
2	40	341	8	10	2	151	24
3	41	320	4	11	10	152	56
4	36	290	23	19	10	144	70
5	30	140	33	30	23	120	65
6	29	132	33	31	24	125	34
7	45	75	33	23	11	121	35
8	48	232	31	15	9	129	45
9	44	249	28	19	8	115	58
10	47	810	18	19	4	110	80
11	50	440	18	49	14	95	78
12	28	430	11	23	7	80	33
13	43	240	12	22	9	172	20
14	41	230	34	21	13	171	21
15	58	205	48	20	12	130	25
16	51	140	55	29	18	131	59
17	52	125	23	31	19	127	75

SPSS screenshot © International Business Machines Corporation.

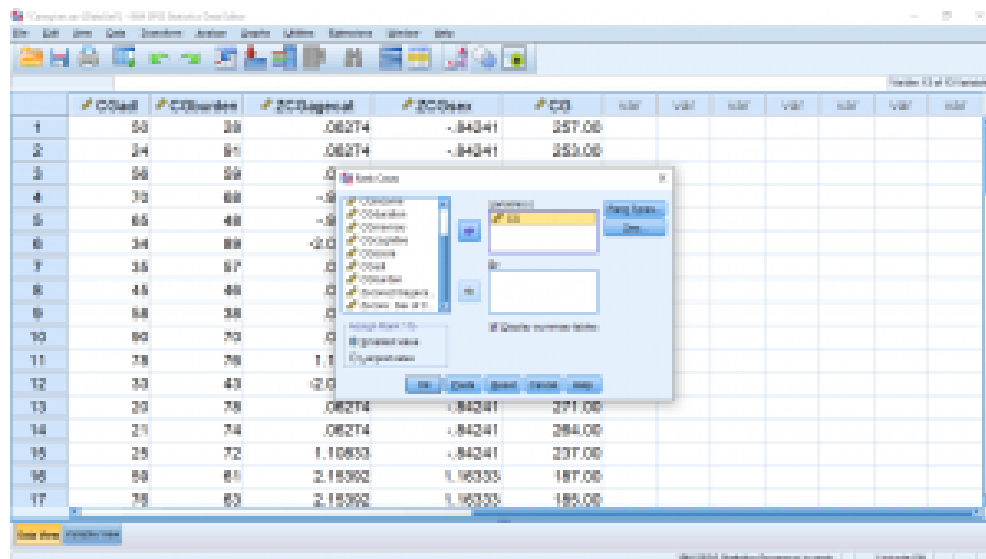
This will bring up a menu which is essentially the calculator feature of SPSS :

	score	CCEvaluation	CCEmergency	CCEognitive	CCEscore	CCEall	CCEurten
1	41	350	13	4	3	140	50
2	40	341	8	10	2	151	24
3	41	320	4	11	10	152	56
4	36	290	23	19	10	144	70
5	30	140	33	30	23	120	65
6	29	132	33	31	24	125	34
7	45	75	33	23	11	121	35
8	48	232	31	15	9	129	45
9	44	249	28	19	8	115	58
10	47	810	18	19	4	110	80
11	50	440	18	49	14	95	78
12	28	430	11	23	7	80	33
13	43	240	12	22	9	172	20
14	41	230	34	21	13	171	21
15	58	205	48	20	12	130	25
16	51	140	55	29	18	131	59
17	52	125	23	31	19	127	75

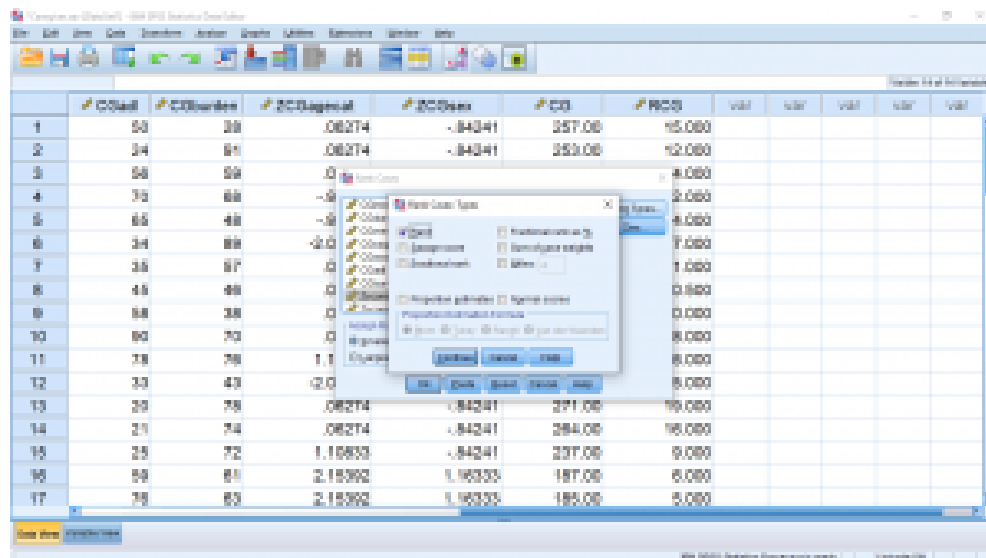
SPSS screenshot © International Business Machines Corporation.

Fill in the menu as shown above. You can move variable names into the Numeric Expression box by double clicking on the variable name, by clicking on the variable name and the arrow or by simply typing it. There are fancier ways to get a sum of variables expression in Numeric Expression, but we will keep it simple for now. The target variable name is totalscore which, after you hit OK, shows up as a new variable, ready for statistical analysis, in the last column in the Data View window :

Let's do a couple of (descriptive) analysis with this new variable. Let's take Caregiver as our dataset. Suppose we want to find the median of the totalscore values. To do this task by hand, we need to put the data in order from smallest to largest. This is tedious but SPSS can do it with a couple of mouse clicks (yes, yes SPSS can compute the median directly but whatever). There are a couple of approaches in SPSS to ordering, or ranking, data. One is to compute the rank, that is, give rank 1 to the lowest value, 2 to the next lowest up to n for the highest value. Pick Transform \rightarrow Rank Cases and move totalscore into the Variable(s) box :



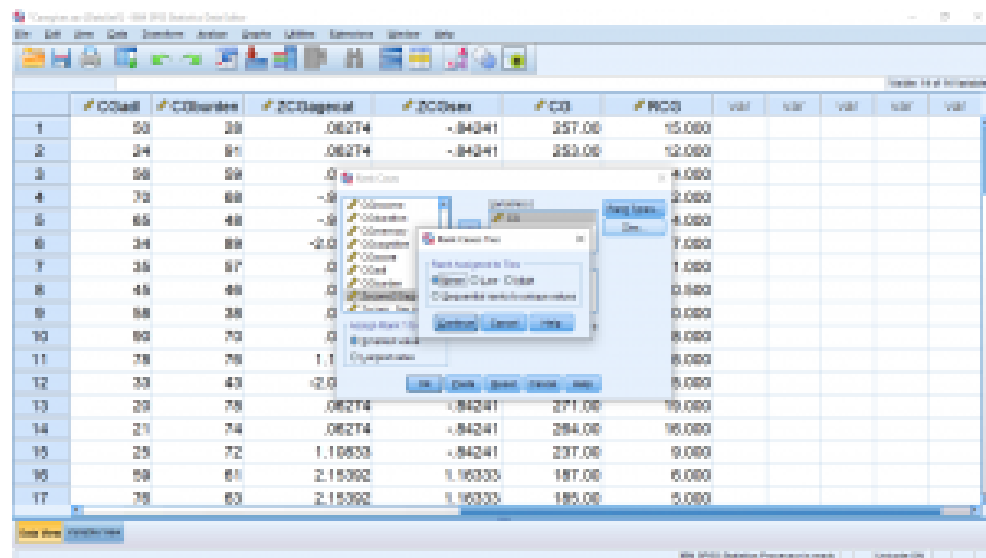
This is a new menu for us, so let's take a look at the submenus. First, the Rank Types menu :



	# Cases	# Columns	# Rows	# Columns	# Rows	# Columns	# Rows	# Columns	# Rows	# Columns	# Rows
1	50	30	.06274	-.84241	257.06	15.000					
2	34	31	.06274	-.84241	253.06	12.000					
3	50	30	.0			4.000					
4	70	30	-.5			2.000					
5	65	40	-.5			4.000					
6	34	30	-.2			7.000					
7	35	37	.0			1.000					
8	45	40	.0			0.000					
9	60	30	.0			0.000					
10	60	70	.0			8.000					
11	70	70	1.1			8.000					
12	30	40	-.2			8.000					
13	20	70	.06274	-.84241	271.06	19.000					
14	21	74	.06274	-.84241	264.06	16.000					
15	25	72	1.16933	-.84241	237.06	9.000					
16	50	61	2.15392	1.16933	187.06	6.000					
17	70	63	2.15392	1.16933	185.06	5.000					

SPSS screenshot © International Business Machines Corporation.

Pretty fancy. Much too advanced for our use, so let's leave that one be, hit Continue. Next look at Ties...



	# Cases	# Columns	# Rows	# Columns	# Rows	# Columns	# Rows	# Columns	# Rows	# Columns	# Rows
1	50	30	.06274	-.84241	257.06	15.000					
2	34	31	.06274	-.84241	253.06	12.000					
3	50	30	.0			4.000					
4	70	30	-.5			2.000					
5	65	40	-.5			4.000					
6	34	30	-.2			7.000					
7	35	37	.0			1.000					
8	45	40	.0			0.000					
9	60	30	.0			0.000					
10	60	70	.0			8.000					
11	70	70	1.1			8.000					
12	30	40	-.2			8.000					
13	20	70	.06274	-.84241	271.06	19.000					
14	21	74	.06274	-.84241	264.06	16.000					
15	25	72	1.16933	-.84241	237.06	9.000					
16	50	61	2.15392	1.16933	187.06	6.000					
17	70	63	2.15392	1.16933	185.06	5.000					

SPSS screenshot © International Business Machines Corporation.

We will assign the average (mean) rank to ties in our classes. To understand the ties options, think of two people in a race who cross the finish line at exactly the same time, a tie. With the mean rank, they both come in 1.5 place. With lowest, they both come in 1st place, with highest, they both come in 2nd place. Hit Continue, the OK and a new variable Rtotalscore will be formed in the Data View menu :

	#Cases	#Cases	#Cases	#Cases	#Cases	#Cases	#Cases	#Cases	#Cases	#Cases
1	50	50	.06274	-.84241	257.06	15.060	15.800			
2	34	51	.06274	-.84241	253.06	12.060	13.800			
3	56	56	.06274	1.16333	341.06	34.060	34.800			
4	70	66	-.98266	-.84241	321.06	32.060	32.800			
5	65	48	-.98266	-.84241	182.06	4.060	4.800			
6	34	66	-2.03845	1.16333	267.06	7.060	7.800			
7	35	57	.06274	-.84241	159.06	1.060	1.800			
8	45	48	.06274	1.16333	246.06	10.800	10.800			
9	66	38	.06274	1.16333	276.06	20.060	20.800			
10	60	70	.06274	-.84241	636.06	48.060	48.800			
11	78	76	1.16333	1.16333	633.06	48.060	48.800			
12	33	43	-2.03845	-.84241	480.06	45.060	45.800			
13	20	78	.06274	-.84241	271.06	19.060	19.800			
14	21	74	.06274	-.84241	264.06	16.060	16.800			
15	25	72	1.16333	-.84241	237.06	9.060	9.800			
16	59	61	2.15392	1.16333	187.06	6.060	6.800			
17	78	63	2.15392	1.16333	185.06	5.060	5.800			

SPSS screenshot © International Business Machines Corporation.

Here the variable RCG ranks the total score of the students. But it's very difficult from this data view to identify which students' rank the highest or lowest, let alone who falls in the middle to find the median. This is not quite what we are after to easily get the median. Ranking will become useful on Psy 234 (in Chapter 16), but it's not that useful for us now. What we need, is to shuffle the numbers around from lowest to highest (of course we can do that directly). To shuffle pick Data → Sort Cases :

	#Cases	#Cases	#Cases	#Cases	#Cases	#Cases	#Cases	#Cases	#Cases	#Cases
1	50	50	.06274	-.84241	257.06	15.060	15.800			
2	34	51	.06274	-.84241	253.06	12.060	13.800			
3	56	56	.06274	1.16333	341.06	34.060	34.800			
4	70	66	-.98266	-.84241	321.06	32.060	32.800			
5	65	48	-.98266	-.84241	182.06	4.060	4.800			
6	34	66	-2.03845	1.16333	267.06	7.060	7.800			
7	35	57	.06274	-.84241	159.06	1.060	1.800			
8	45	48	.06274	1.16333	246.06	10.800	10.800			
9	66	38	.06274	1.16333	276.06	20.060	20.800			
10	60	70	.06274	-.84241	636.06	48.060	48.800			
11	78	76	1.16333	1.16333	633.06	48.060	48.800			
12	33	43	-2.03845	-.84241	480.06	45.060	45.800			
13	20	78	.06274	-.84241	271.06	19.060	19.800			
14	21	74	.06274	-.84241	264.06	16.060	16.800			
15	25	72	1.16333	-.84241	237.06	9.060	9.800			
16	59	61	2.15392	1.16333	187.06	6.060	6.800			
17	78	63	2.15392	1.16333	185.06	5.060	5.800			

SPSS screenshot © International Business Machines Corporation.

which brings up, after moving over the RCG total score variable :

#Case#	#Columns#	#ZC00agesat	#ZC00sex	#C0	#RC0	#RAN0001	#var	#var	#var
1	50	59	.06274	-.84241	257.06	15.060	15.800		
2	34	61	.06274	-.84241	12.060	12.060	12.800		
3	56	59	.06274			34.060	34.800		
4	70	68	-.98268			32.060	32.800		
5	65	48	-.98268			4.060	4.800		
6	34	69	-2.03845			7.060	7.800		
7	35	57	.06274			1.060	1.800		
8	45	48	.06274			10.860	10.800		
9	68	38	.06274			20.060	20.800		
10	60	70	.06274			48.060	48.800		
11	78	76	1.16833			48.060	48.800		
12	33	43	-2.03845			45.060	45.800		
13	30	79	.06274			19.060	19.800		
14	21	74	.06274	-.84241	284.06	16.060	16.800		
15	29	72	1.16833	-.84241	237.06	9.060	9.800		
16	59	61	2.16392	1.16333	187.06	6.060	6.800		
17	79	63	2.16392	1.16333	185.06	5.060	5.800		

SPSS screenshot © International Business Machines Corporation.

Keep the ascending button selected (sort from lowest to highest), then hit OK to sort the file :

#Case#	#Columns#	#ZC00agesat	#ZC00sex	#C0	#RC0	#RAN0001	#var	#var	#var
1	35	57	.06274	-.84241	139.06	1.060	1.800		
2	44	62	1.16833	-.84241	142.06	2.060	2.800		
3	60	65	.06274	1.16333	150.06	3.060	3.800		
4	65	48	-.98268	-.84241	162.06	4.060	4.800		
5	78	63	2.16392	1.16333	185.06	5.060	5.800		
6	59	61	2.16392	1.16333	187.06	6.060	6.800		
7	34	69	-2.03845	1.16333	207.06	7.060	7.800		
8	45	103	-.98268	1.16333	236.06	8.060	8.800		
9	29	72	1.16833	-.84241	237.06	9.060	9.800		
10	48	48	.06274	1.16333	246.06	10.860	10.800		
11	61	71	.06274	-.84241	246.06	10.860	10.800		
12	34	61	.06274	-.84241	253.06	12.060	12.800		
13	60	67	.06274	1.16333	264.06	13.860	13.800		
14	47	64	.06274	-.84241	264.06	13.860	13.800		
15	60	28	.06274	-.84241	267.06	15.060	15.800		
16	21	74	.06274	-.84241	264.06	16.060	16.800		
17	65	64	-.98268	1.16333	285.06	17.060	17.800		

SPSS screenshot © International Business Machines Corporation.

Everything is sorted now. (Note how useful the id variable is now. If that wasn't there, we'd lose track of who's data was what.) Now if we scroll down, we will find that the middle two total test scores are both 210. Thus the median of total score is 210.

As a final analysis of the Caregiver data, suppose we wanted some descriptive statistics for the male students separate from the female students. To do this we use the "split file" feature of SPSS. Select Data → Split File to get

	COGtest	COGscore	COGagecat	COGsex	COG	RCOG	RANCOG	var	var	var
1	35	57	.06274	-.84241	109.06	1.000	1.800			
2	44	62	1.16833	-.84241	142.06	2.000	3.800			
3	62	65	.061			3.000	3.800			
4	65	48	-.981			4.000	4.800			
5	78	63	2.161			5.000	5.800			
6	68	61	2.161			6.000	6.800			
7	34	68	-2.008			7.000	7.800			
8	45	103	-.981			8.000	8.800			
9	25	72	1.161			9.000	9.800			
10	45	48	.061			10.000	10.800			
11	61	71	.061			10.000	10.800			
12	24	61	.061			12.000	12.800			
13	60	67	.06274	-.84241	204.06	13.000	13.800			
14	47	64	.06274	-.84241	204.06	13.000	13.800			
15	60	28	.06274	-.84241	207.06	15.000	15.800			
16	21	74	.06274	-.84241	204.06	15.000	15.800			
17	65	64	-.06268	1.16333	205.06	17.000	17.800			

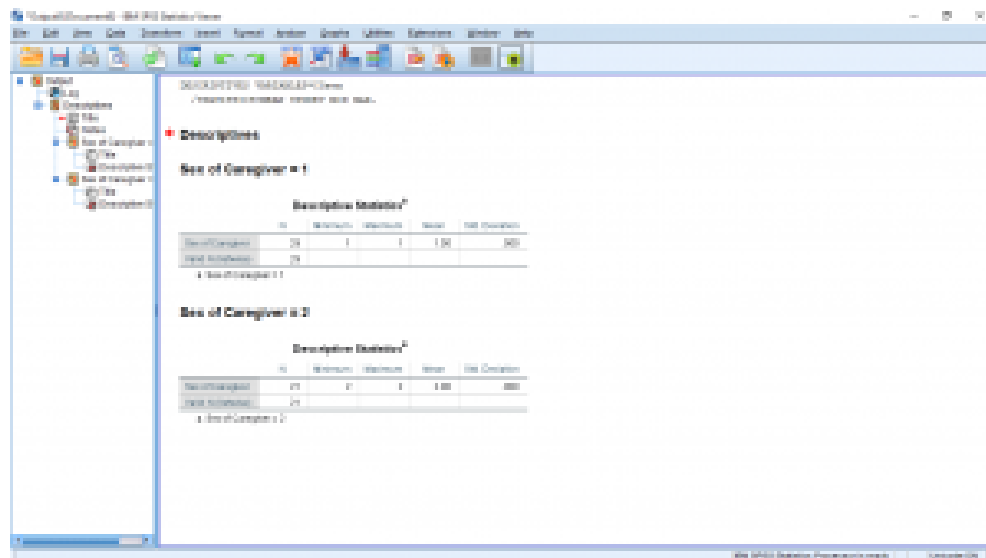
SPSS screenshot © International Business Machines Corporation.

where the gender variable has been moved into the “Groups Based on” box — you will need to click on the “Organize output by groups” button also. We’ll also leave the “Sort the file by grouping variables” (gender in this case), this will shuffle the file yet again, putting all the males and females together. So, when you hit OK the result is

	COGagecat	COGsex	COGscore	COGeducation	COGmemory	COGcognitive	COGscore	COGtest	COGscore
19	35- 30-39	1	550	23	19	12	144	70	68
20	53- 30-39	1	320	28	21	7	121	39	68
21	37- 30-39	1	320	28	41	29	140	50	47
22	60- 60-69	1	380	25	55	3	120	30	57
23	40- 40-49	1	380	28	37	25	179	70	62
24	29- 30-39	1	370	28	68	30	162	57	67
25	25- 30-39	1	430	11	23	7	80	33	43
26	31- 30-39	1	480	38	24	6	101	41	71
27	47- 40-49	1	610	18	18	6	110	60	70
28	38- 30-39	1	660	27	40	29	144	61	60
29	44- 40-49	1	610	28	27	1	104	49	52
30	40- 40-49	2	95	28	32	23	109	62	65
31	62- 60-69	2	135	23	31	19	107	76	63
32	61- 60-69	2	140	68	29	18	131	69	61
33	29- 30-39	2	132	23	61	24	125	34	69
34	35- 30-39	2	175	51	41	20	127	45	103
35	45- 40-49	2	222	31	15	9	129	45	46

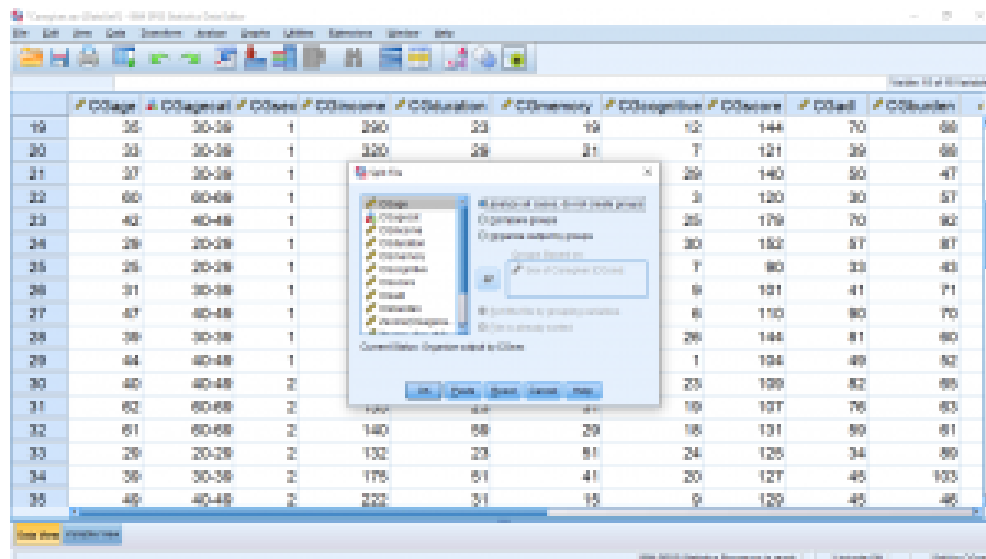
SPSS screenshot © International Business Machines Corporation.

Now the file is sorted into Male and Female (the 1-A button at the top has been pressed). Also note that “Split by gender” appears on the lower right corner of the Data View window. Now let’s do a simple descriptive statistics analysis of the total score variable. The output looks like :



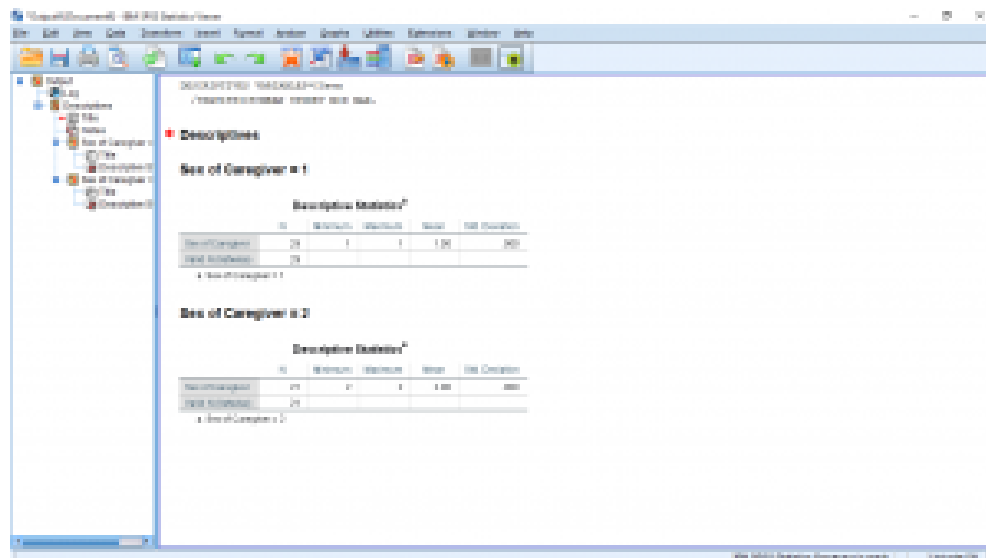
SPSS screenshot © International Business Machines Corporation.

To unsplit the file, go back to Data → Split File and hit the “Analyze all cases, do not create groups” button. This will remove the “Split” message from the lower right corner and when the descriptive statistics is run again, you will get :



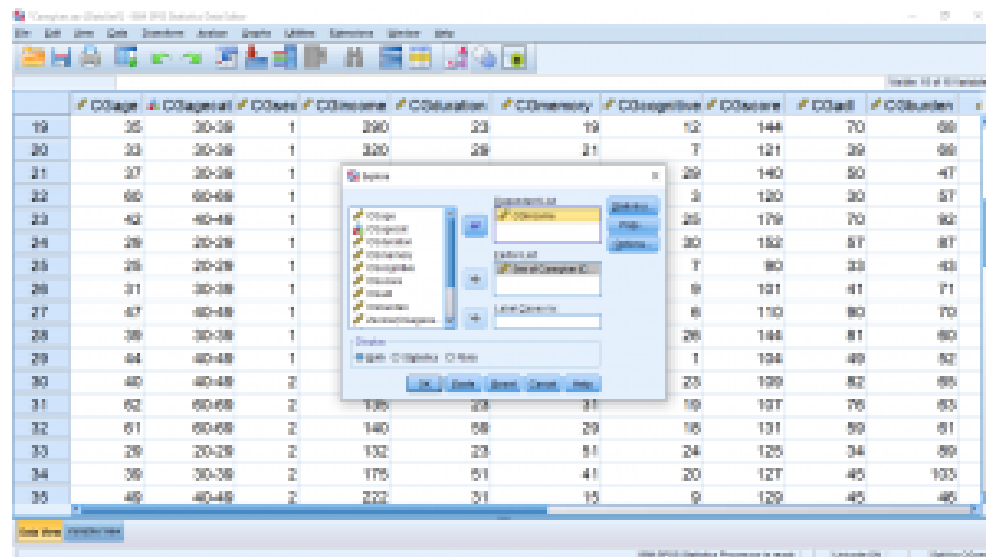
SPSS screenshot © International Business Machines Corporation.

From here, with the file unsplit, we can use gender as a factor to get separate descriptive statistics for males and female. Select Analyze → Explore and use gender as the factor, which results in :



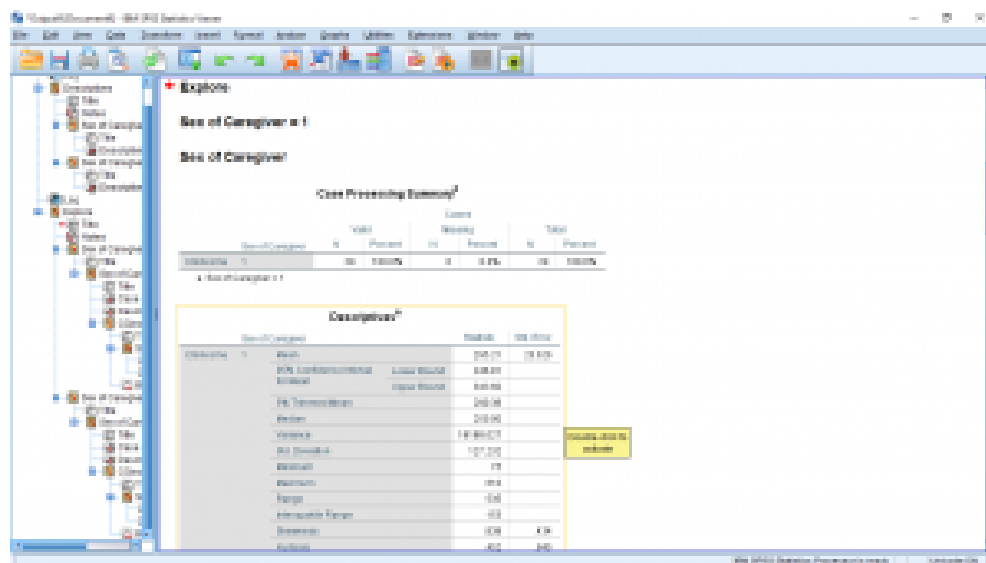
SPSS screenshot © International Business Machines Corporation.

From here, with the file unsplit, we can use gender as a factor to get separate descriptive statistics for males and female. Select Analyze → Explore and use gender as the factor :



SPSS screenshot © International Business Machines Corporation.

The result is :



SPSS screenshot © International Business Machines Corporation.

This page titled [3.4: SPSS Lesson 2- Combining variables and recoding](#) is shared under a [CC BY-NC-SA 4.0](#) license and was authored, remixed, and/or curated by [Gordon E. Sarty](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.