

Lab: SPSS Recoding and Management of Data

- A. Review
- B. Recoding Variables
- C. Calculating New Variables

On your own:

- Recode raw scores into Pass/Not Pass, where "Pass" = 70 and above, and "Not Pass" = anything 69 or below
- Create Bar graph of Gender
- Create Histogram of Score

B. Recoding Variables

Recoding variables is necessary when we want to collapse raw data into categories or reassign new codes to old ones to improve our management of data. This can be done for categorical and continuous variables, BUT you as a researcher must decide whether the recoding is conceptually useful for your project. Usually you keep continuous variables in their raw form.

Always enter your data is as raw a form as possible and if you recode, do so into a different variable to retain the original information.

*** Example of ethnicity (self-identified ethnic group of origin)

Variable = Ethnic			Recoded Into:	Ethnic2		
Code (Value)	Ethnicity (Value Label)	Frequency		Code (Value)	Recoded Ethnicity (Value Label)	Frequency
1	African	21		1	White	137
2	American	4		2	Black	34
3	American Indian	1		3	Latino	56
4	Biracial	12		4	Asian	37
5	Black	4		5	Other	29
6	Brazilian	8				
7	Canadian	87				
8	Caucasian	12				
9	Chinese	4				
10	English	1				
11	Ethiopian	5				
12	French	1				
13	German	5				
14	Indian	5				
15	Japanese	2				
16	Korean	39				
17	Latino	12				
18	Mexican	22				
19	Middle Eastern	2				
20	Native American	4				
21	Filipino	8				
22	Russian	1				
23	South African	1				
24	South American	9				
25	Vietnamese	23				
	White					
		N = 293				N = 293

*** Our SPSS Example: Recoding raw data from Chapter 6 into grade categories

Conceptually, you are RECODING data in your "score" variable so that:

- 90 thru Highest = 1 (which will be labeled an "A")
- 80 thru 89 = 2 (which will be labeled a "B")
- 70 thru 79 = 3 (which will be labeled a "C")
- 60 thru 69 = 4 (which will be labeled a "D")
- Lowest thru 59 = 5 (which will be labeled an "F")

INTO a new variable called "grade"

You will then create value labels for this new variable "grade" so that

- 90-100 is an "A" (value label associated with 1)
- 80-89 is a "B" (value label associated with 2)
- 70-79 is a "C" (value label associated with 3)
- 60-69 is a "D" (value label associated with 4)
- 0-59 is an "F" (value label associated with 5)

Let's actually do this recoding with our Chapter 6 data.

- Open your saved Chapter 6 data set from last time
- Click **Transform, Recode, Into Different Variables**. A dialogue box named "Recode into Different Variables" should appear
- Select your "score" variable and click the arrow to pop it into the middle Input Variable → Output Variable box
- Type "grade" in the Name box of the Output Variable Section. Label it "grade categories"
- Click **Old and New Values**, a dialogue box "Recode into Different Variables: Old and New Values" should appear
- Look at the above notes regarding grade categories and figure out the procedures about how to create these new values
- When you are done recoding, save your data with a different file name: **chapter6a**

SCORE

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid 33.00	1	3.3	3.3	3.3
53.00	1	3.3	3.3	6.7
55.00	1	3.3	3.3	10.0
60.00	1	3.3	3.3	13.3
68.00	1	3.3	3.3	16.7
70.00	2	6.7	6.7	23.3
73.00	2	6.7	6.7	30.0
75.00	1	3.3	3.3	33.3
76.00	1	3.3	3.3	36.7
77.00	1	3.3	3.3	40.0
78.00	1	3.3	3.3	43.3
81.00	1	3.3	3.3	46.7
82.00	1	3.3	3.3	50.0
83.00	1	3.3	3.3	53.3
85.00	1	3.3	3.3	56.7
86.00	1	3.3	3.3	60.0
87.00	1	3.3	3.3	63.3
88.00	2	6.7	6.7	70.0
89.00	4	13.3	13.3	83.3
91.00	2	6.7	6.7	90.0
92.00	1	3.3	3.3	93.3
93.00	1	3.3	3.3	96.7
98.00	1	3.3	3.3	100.0
Total	30	100.0	100.0	
Total	30	100.0		

grade categories

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid A	5	16.7	16.7	16.7
B	12	40.0	40.0	56.7
C	8	26.7	26.7	83.3
D	2	6.7	6.7	90.0
F	3	10.0	10.0	100.0
Total	30	100.0	100.0	
Total	30	100.0		

C. Calculating New Variables

Calculating new variables is necessary when we want to use a formula to convert raw scores into values to match our operational definitions.

*** Example of SES. SES could be operationally defined as:

(father's years of education + mother's years of education) x person's annual income

*** Calculating scores for self-esteem using the Rosenberg Self-Esteem Inventory (or any other multiple item instruments)?

ID#	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6	Item 7	Item 8	Item 9	Item 10
1	1	3	4	1	3	4	1	2	1	3
2	2	2	2	2	2	2	2	4	2	1
3	3	3	1	3	3	1	3	2	3	2
4	2	2	4	2	2	4	2	1	2	3
5	1	1	3	1	1	3	1	2	2	1

If these instruments have multiple items or formulas which lead to a summary score, then you can use SPSS to calculate these scores instead of doing them by hand.

Item 1 + Item 2 + Item 3 + Item 4 + Item 5 +
Item 6 + Item 7 + Item 8 + Item 9 + Item 10 = Total Self Esteem score

*** SPSS Example: Recoding raw data from Chapter 6 data into percentages. Conceptually, you are calculating the percentage of each score by:

Score / 100

Or, computing percent = score / 100 .

Let's actually do this recoding with our Chapter 6 data (see last page).

- Enter the data and set up your variables (good for practice!)
- Click **Transform, Compute**. A dialogue box named "Compute Variable" should appear
- In the Target Variable box, type in "percent"
- Select your "score" variable and click the arrow to pop it into the Numeric Expression box
- Continue typing in the Numeric Expression box so that it reads:
score / 100
- Notice the other mathematical commands available to you. These may be helpful when computing more complicated values
- Click **OK**, notice what happens
- Let's assume that this score for the midterm is worth 50% of the student's final grade, and the instructor wanted not just the percentage of the midterm score, but its weight towards the final grade. Conceptually, what would you have to do to your formula?

Student	Gender	Score
1	1	87
2	1	53
3	1	92
4	1	70
5	1	78
6	1	73
7	1	91
8	1	60
9	1	77
10	1	82
11	1	85
12	1	33
13	1	88
14	1	98
15	1	88
16	2	89
17	2	73
18	2	91
19	2	76
20	2	75
21	2	89
22	2	81
23	2	83
24	2	68
25	2	86
26	2	55
27	2	89
28	2	89
29	2	70
30	2	93

- How many variables do we have?
What type of variables are they?
- When we enter data, it is important to convert our categorical information into numbers whenever possible. In this case, our “Gender” variable will use the following **values** and **value labels**:
 - 1 = Male
 - 2 = Female

The numbers you assign do not matter—as long as you know (and SPSS knows) what they mean
- Type in your data (just the numbers; ignore the column headings for now) first in the “Data View” part of the SPSS Data Editor found by clicking the “Data View” tab at the bottom of the SPSS screen. The “Data View” is a primitive spread sheet.
- We will design our variables using the “Variable View” part of the SPSS Data Editor found by clicking the “Variable View” tab at the bottom of the SPSS screen.