

## 04a: Composite Variables and Reverse Scoring

### 1. Manifest and Latent Variables

Recall the earlier presentation on manifest and latent variables. Summarized:

- Manifest variables = loosely described, are those that can be directly observed or measured
  - Examples
    - height
    - weight
    - age
    - income
- Latent variables = not easily observed or measured; constructed through composite variables as measured through scales and indexes
  - Examples
    - Stress
    - general self-efficacy
    - workplace autonomy
    - life satisfaction
    - test anxiety
- Constructs often used to measure latent variables
  - created by taking composite scores from
  - indicators that are designed to measure a latent variable
  - indicator is an instrument item that provides an indication about one's position or level on some attribute, attitude, etc.
- Example
  - Test Anxiety, composed of two dimensions
  - dimension 1: physiological (somatic, emotionality) reactions
    - sweating
    - headache
    - upset stomach
    - rapid heartbeat
    - feeling of dread
  - dimension 2: negative cognition, thoughts
    - expecting failure
    - negative thoughts
    - frustration
    - comparing oneself to others negatively
    - feelings of inadequacy
    - self-condemnation
  - Indicators of physiological reaction
    1. Before or during tests you feel your heart start to beat faster.
    2. You get upset stomachs while taking tests.
    3. When taking a test, you get a feeling of dread.
  - Indicators of negative cognition
    4. While taking tests you think about how poorly you are doing.
    5. You expect failure or poor grades when taking tests.
    6. You become frustrated during testing.
  - Indicator response options

- 1 = Not at all like me
  - 7 = Very much like me
- One student's responses
  1. Heart beats faster = 2
  2. Upset stomach = 3
  3. Feel dread = 2
  4. Think of poor performance = 2
  5. Expect failure = 1
  6. Frustrated = 1
- Composite score on one student's responses
  - $\text{Sum} = 2+3+2+2+1+1 = 11$
  - Mean is  $11 / 6 = 1.83$
  - Composite test anxiety is 1.83 on 1 to 7 scale

## 2. Composite Scores

- Composite scores are constructed scores
  - Summing responses across items or indicators (not a good option, explained below)
  - Mean of responses across items or indicators (good option)
  - Weighted composite from factor analysis or similar analysis (usually sample dependent) is sometimes used; weighted means some items account for more of the composite score than others; this requires more complex statistics or theoretical guidance; using equally weighted composite scores – like taking the sum of all items or the mean of all items – works well in many cases. Weights described below.
- Sometimes called scale scores, but this can be confusing since scale scores generally are understood to be scores with predefined mean and standard deviation (standard score, Z score)
- Index (Summary) vs. Scale
  - Scale: items designed to measure the same construct (e.g., test anxiety); theoretically items should be correlated
  - Index: items used to form a composite score, but items do not have to measure the same construct; items may be theoretically unrelated
    - Example: Socio-economic Status (SES)
      - Educational level
      - Occupational Prestige
      - Wealth
    - Dow Jones Industrial Average: composed of 30 major companies
- Sum scored can be misleading
  - Example from test anxiety, student has maximum anxiety
    - 1. Heart beats faster = 7 (on scale from 1 to 7)
    - 2. Upset stomach = 7 (on scale from 1 to 7)
    - 3. Feel dread = 7 (on scale from 1 to 7)
  - Minimum and maximum summed scores are
    - $1+1+1 = 3$
    - $7+7+7 = 21$
  - Respondent's sum =  $7+7+7 = 21$ , which is top score possible for sum of these three items
  - Item 2 has missing data
    - 1. Heart beats faster = 7 (on scale from 1 to 7)
    - 2. Upset stomach = missing (on scale from 1 to 7)
    - 3. Feel dread = 7 (on scale from 1 to 7)

- Sum = 7+missing+7 = 14, which is toward middle range of 3 to 21, so misleading level of anxiety presented
- Mean score can ignore missing responses so it reflects better level of anxiety, also mean lies within original scale units so easier to interpret
  - Example from test anxiety, student has maximum anxiety
    - 1. Heart beats faster = 7 (on scale from 1 to 7)
    - 2. Upset stomach = 7 (on scale from 1 to 7)
    - 3. Feel dread = 7 (on scale from 1 to 7)
  - Scale ranges from 1 to 7
  - Respondent's mean =  $7+7+7 = 21/3 = 7.0$  top score possible
  - Item 2 has missing data
    - 1. Heart beats faster = 7 (on scale from 1 to 7)
    - 2. Upset stomach = missing (on scale from 1 to 7)
    - 3. Feel dread = 7 (on scale from 1 to 7)
  - Mean of available data =  $7+7 = 14 / 2 = 7.0$ , top score possible
  - Mean of all items =  $7+7 = 14 / 3 = 4.66$ , misleading score
  - Be sure calculation of mean uses available data, not all possible scores

**3. Missing Data**

- This is a very brief and unsophisticated introduction to missing data
- See this link for references to more detailed examinations of missing data
  - [https://en.wikipedia.org/wiki/Missing\\_data](https://en.wikipedia.org/wiki/Missing_data)
- Missing at random is ok if only a few observations missing here and there
- Mean substitution: practice of replacing missing values with mean values, that is what occurs when one calculates the mean score as a composite score
- Check pattern of missing data to note whether pattern appears to be non-random, systematic

Example 1

Respondent	Sex	Race	Education	TA1	TA2	TA3	TA4	TA5	TA6
1	0	2	1	1	2	3	3	2	1
2	1	3	2	2	2	2	3	4	2
3	0	3	2	7	7	6	5	7	6
4	1	2	2	4	5	4	2	3	2
5	0	1	3	2	5	6	6	4	5
6	0	1	1	1	1	1	1	1	1
7	0	1	2	3	4	6	3	5	6
8	1	2	4	6	7	6	7	7	7
9	1	3	3	2	1	4	5	1	2
10	1	3	4	2	4	4	4	3	1

No Missing Data for Test Anxiety Questionnaire

## Example 2

Respondent	Sex	Race	Education	TA1	TA2	TA3	TA4	TA5	TA6
1	0	2	1	1	2	3	3	-----	1
2	1	3	2	2	2	2	3	4	2
3	0	3	2	-----	7	6	5	7	6
4	1	2	2	4	5	4	2	3	2
5	0	1	3	2	5	6	6	4	5
6	0	1	1	1	1	1	-----	1	1
7	0	1	2	3	4	6	3	5	6
8	1	2	4	6	7	6	7	7	7
9	1	3	3	2	1	4	5	1	2
10	1	3	4	2	-----	4	4	3	1

Some Missing Data for Test Anxiety Questionnaire

Missing Seems Random

Mean Replacement Acceptable

## Example 3

Respondent	Sex	Race	Education	TA1	TA2	TA3	TA4	TA5	TA6
1	0	2	1	1	2	-----	3	2	1
2	1	3	2	2	2	2	3	4	2
3	0	3	2	7	7	-----	5	7	6
4	1	2	2	4	5	-----	-----	-----	-----
5	0	1	3	2	5	6	6	4	5
6	0	1	1	1	1	-----	1	-----	1
7	0	1	2	3	4	6	3	5	6
8	1	2	4	6	7	-----	7	7	7
9	1	3	3	2	1	4	5	1	2
10	1	3	4	2	4	-----	-----	3	1

Some Missing Data for Test Anxiety Questionnaire

Look at TA3 item for wording, perhaps offensive or too personal or maybe difficult to see on questionnaire

Missing Seems Systematic

Mean Replacement Not Acceptable

#### 4. Outliers

- Outliers are scores or score combinations that produce observations that are very difficult from rest of sample
- Outliers can influence statistical results so should be examined and fixed, accepted, or removed depending upon findings of case study of outlier
- Discussion that follows is an unsophisticated review of outliers; see link below for more detailed treatments
  - <https://en.wikipedia.org/wiki/Outlier>
- Checking for outliers
  - Frequency Display
  - Z Scores
  - Scatterplot
  - Boxplot
  - Histograms

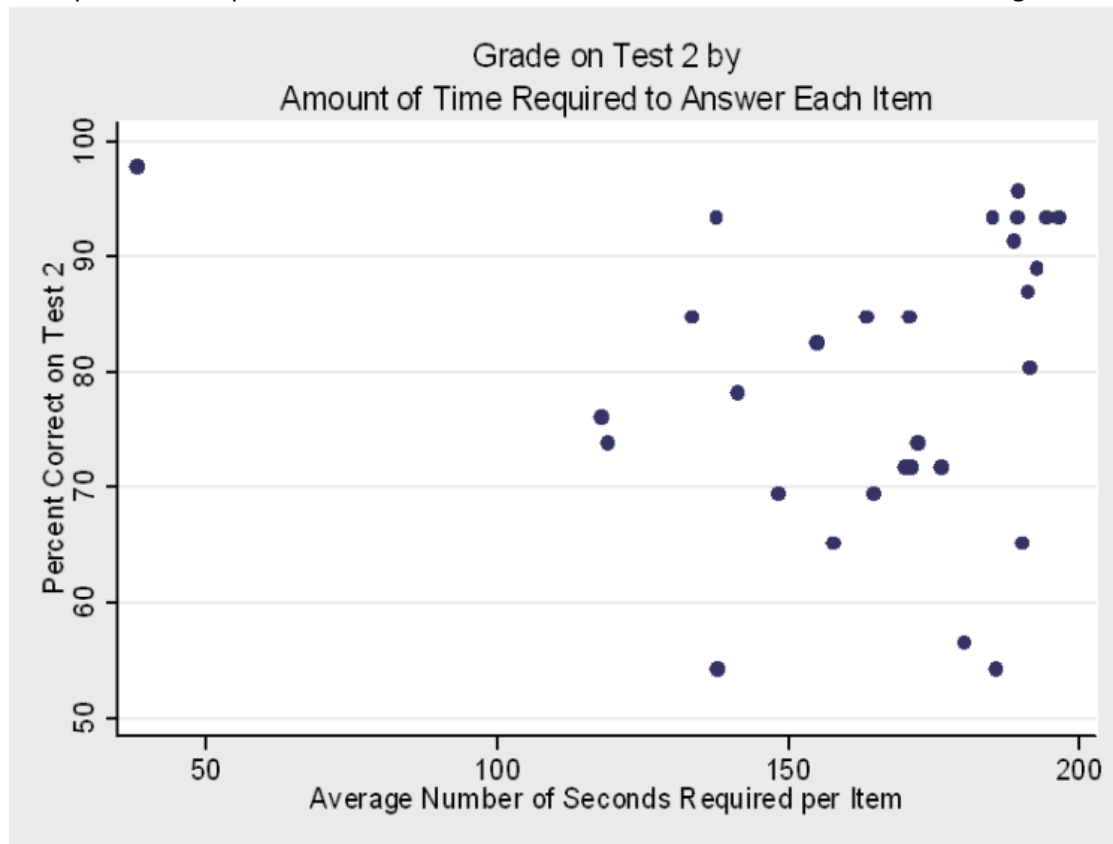
Example 4: Frequency Display of TA1 (scale min and max 1 to 7)

**TA1 - Heart Beats Faster During Tests**

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid 1.00	5	23.8	23.8	23.8
2.00	4	19.0	19.0	42.9
3.00	1	4.8	4.8	47.6
4.00	4	19.0	19.0	66.7
5.00	3	14.3	14.3	81.0
7.00	3	14.3	14.3	95.2
77.00	1	4.8	4.8	100.0
Total	21	100.0	100.0	

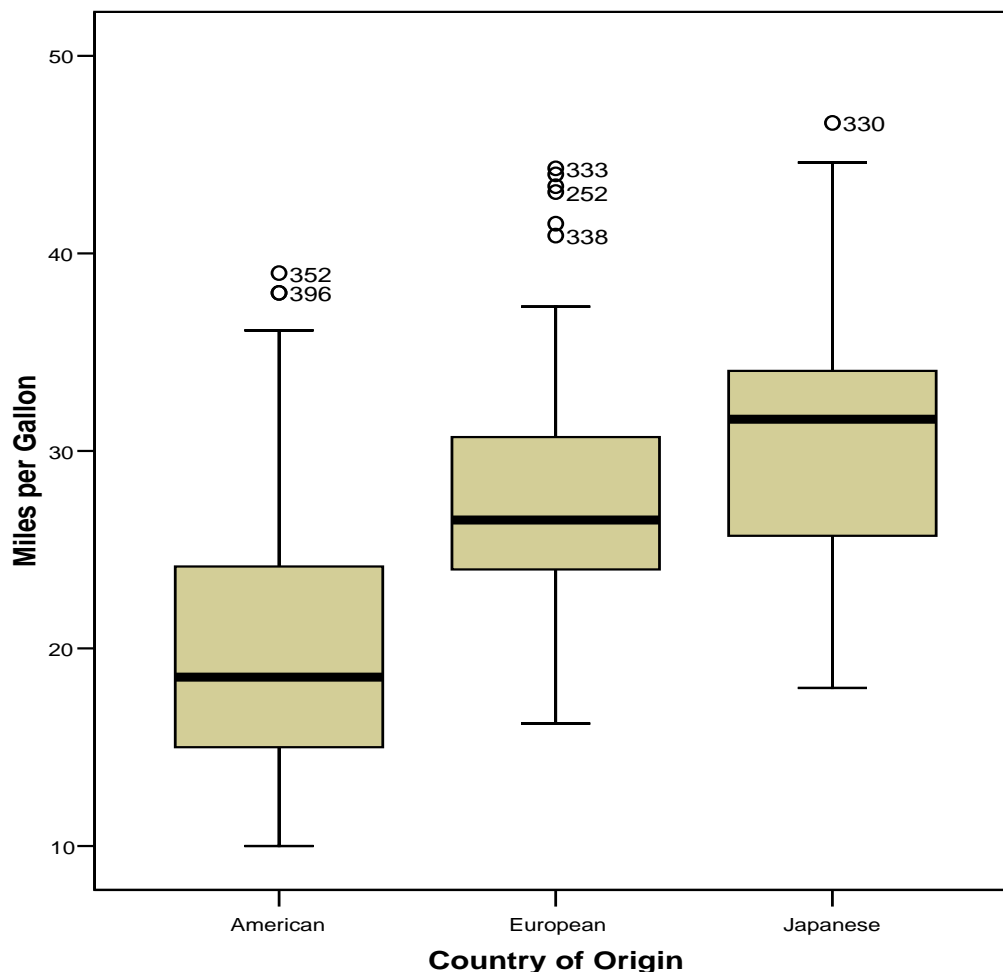
Recall that the scale for test anxiety is 1 to 7; note one score above is 77 – likely a data entry error – but this score can have a large effect on scoring and analysis so must be corrected.

Example 5: Scatterplot of Tests 2 Grades and Seconds to Answer Each Item on Average



Note the outlier in the upper, left corner. This graph makes it easy to identify someone who is likely cheating. All other students took 120 seconds or more, on average, to complete test items. The individual with a score near 100 took less than 45 seconds on average to answer each item.

Example 6: Boxplot showing MPG for Cars from US, Europe, and Japan.



The boxplot identifies potential outliers by noting individual observations with scores above the 90<sup>th</sup> percentile. In this example the points are labeled with their case numbers (e.g., 352) making it easy to check the data file for data entry or other errors.

## 5. Composite Score Weights

- Mean or summed composite scores assume all items have equal weight, e.g.
  - $\text{Sum} = \text{TA1} + \text{TA2} + \text{TA3} + \text{TA4} + \text{TA5} + \text{TA6}$
  - $\text{Sum} = (\text{TA1} \times 1) + (\text{TA2} \times 1) + (\text{TA3} \times 1) + (\text{TA4} \times 1) + (\text{TA5} \times 1) + (\text{TA6} \times 1)$
- Possible to differentially weight items based upon theory or empirical logic, e.g.
  - Negative cognition (TA 4 5 & 6) is twice as debilitating as physiological reactions (TA 1 2 & 3) with test anxiety, so
  - $\text{Sum} = (\text{TA1} \times 1) + (\text{TA2} \times 1) + (\text{TA3} \times 1) + (\text{TA4} \times 2) + (\text{TA5} \times 2) + (\text{TA6} \times 2)$
  - $\text{Sum} = (\text{TA1} \times .11) + (\text{TA2} \times .11) + (\text{TA3} \times .11) + (\text{TA4} \times .22) + (\text{TA5} \times .22) + (\text{TA6} \times .22)$
- Empirical methods of weighting (regression, factor analysis) are sample specific, so weights might change with each sample
- Structural equation modeling is partially sold on finding optimal weight of latent variable indicators, which means using weights to form constructs
- Bobko et al. (2007) found weighting items provided little benefit over unit weighting (i.e., unit weight is weighting all items the same; unit means a weight of 1.00).

- Hendrickson et al. (2008) found some benefit for weighting, but it was small, they write: “The increase in reliability by using the maximum reliability weighting scheme was as much as 0.029” and for validity results were inconclusive and small.
- In short, unit weighting is easier, is not sample dependent, and loses little in terms of reliability and validity especially when number of items per construct increases.

## 6. Reversed Scores

- Reverse scoring is necessary for those items that take opposite responses from other items designed to measure the same construct; reverse coded items are sometimes referred to as negative valence and non-reverse coded items are referenced as positive valence
- Some argue that reversed items can be useful to help keep respondents alert and break response set
- Failure to reverse score items that are reverse coded can result in lower reliability and validity, and produces scores that cannot be interpreted
- Critical that you explain to readers
  - Which items reverse scored
  - Which items used to form construct, reversed or non-reversed
  - Interpretation of construct scores, e.g., 7 = high anxiety and 1 = low, or if reversed, 7 = low anxiety and 1 = high
- Identify reversed items by assuming role of respondent who has extreme position on construct
- Example: Test Anxiety – assume you have extreme test anxiety, which item elicits a response that is different from the others?

	Not true of me						Very true of me
	1	2	3	4	5	6	7
1. Before or during tests you feel your heart start to beat faster.	1	2	3	4	5	6	7
2. You get upset stomachs while taking tests.	1	2	3	4	5	6	7
3. While taking tests you get a feeling of confidence that you will do well tests.	1	2	3	4	5	6	7
4. While taking tests you think about how poorly you are doing.	1	2	3	4	5	6	7

- Formula for reverse scoring: Reversed Score = (minimum score) + (maximum score) – actual score
- Example of reverse scoring with 1 to 5 scale:

Original Score	Formula	Reversed Score
1	$1 + 5 - 1 =$	5
2	$1 + 5 - 2 =$	4
3	$1 + 5 - 3 =$	3
4	$1 + 5 - 4 =$	2
5	$1 + 5 - 5 =$	1

- Example of reverse scoring with -3 to +3 scale:

Original Score	Formula	Reversed Score
-3	$-3 + 3 - (-3) =$	3
-2	$-3 + 3 - (-2) =$	2
-1	$-3 + 3 - (-1) =$	1
0	$-3 + 3 - (0) =$	0
1	$-3 + 3 - (1) =$	-1
2	$-3 + 3 - (2) =$	-2
3	$-3 + 3 - (3) =$	-3

- Never delete original variable, instead, form a new reversed variable, e.g., TA3 becomes TA3R (R indicates reversed version)
- Check correlations of original and reversed to ensure reversed scores correct, if correct Pearson  $r = -1$
- Worked Example: Academic Control – one’s ability to control or determine their academic behavior and outcomes in college.
  - Which item or items should be reverse scored?
  - Goal is to have higher scores indicate more academic control

	Strongly Disagree	Disagree	Mix of Disagree and Agree	Agree	Strongly Agree
43. There is little I can do about my performance in college/university.	1	2	3	4	5
44. The more effort I put into my courses, the better I do in them.	1	2	3	4	5
45. How well I do in my courses is often the “luck of the draw.”	1	2	3	4	5
46. I have a great deal of control over my academic performance in my courses.	1	2	3	4	5
47. When I do poorly in a course, it’s usually because I haven’t given it my best effort.	1	2	3	4	5

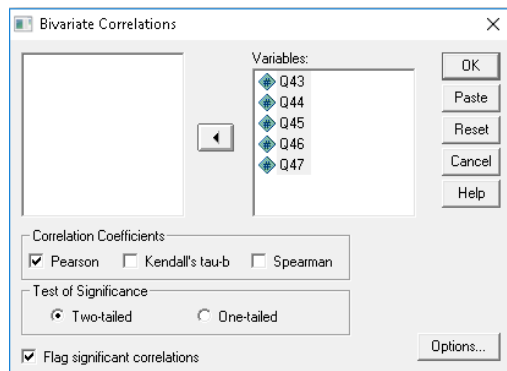
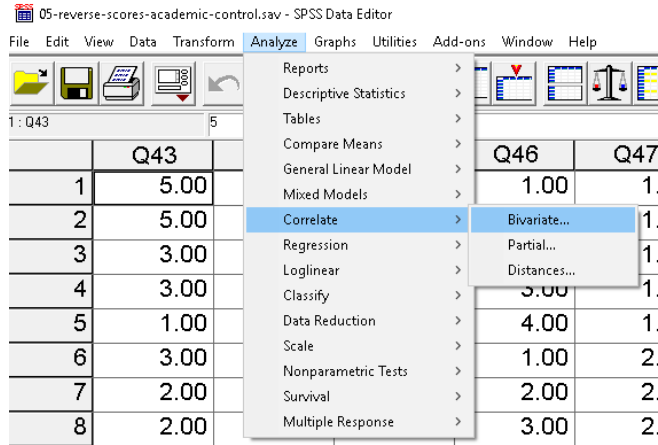
- Worked example of Academic Control
  - Data File:

<http://www.bwgriffin.com/gsu/courses/edur9131/2018spr-content/05-composite-scores/05-reverse-scores-academic-control.sav>

Steps in Computing Reversed Items

1. Assess correlations among items to ensure items to be reversed correlate negatively with positive items designed to measure the same construct.



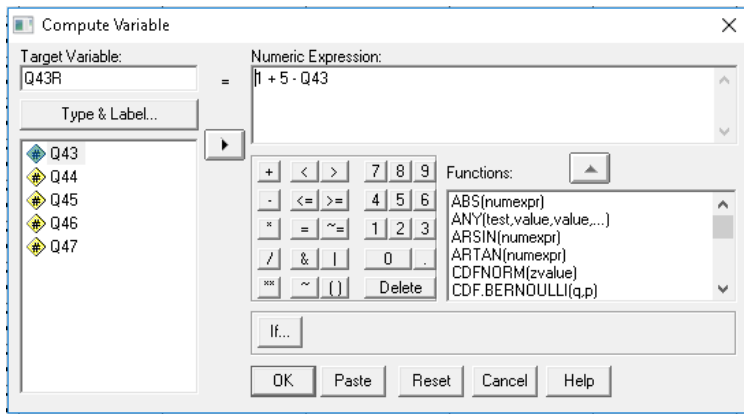
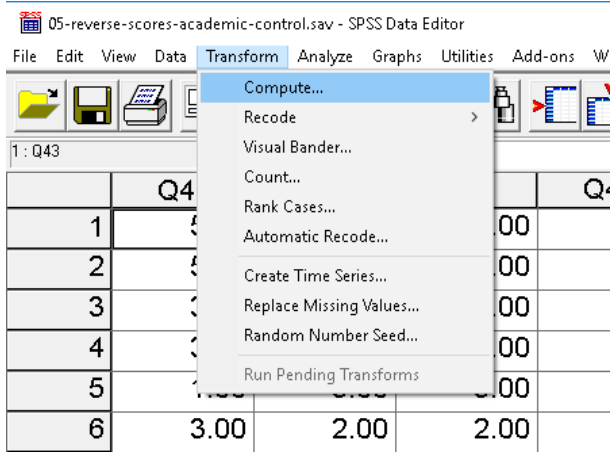


**Correlations**

		Q43	Q44	Q45	Q46	Q47
Q43	Pearson Correlation	1	-.439**	.582**	-.483**	-.364**
	Sig. (2-tailed)		.000	.000	.000	.000
	N	313	313	313	313	313
Q44	Pearson Correlation	-.439**	1	-.382**	.562**	.451**
	Sig. (2-tailed)	.000		.000	.000	.000
	N	313	313	313	313	313
Q45	Pearson Correlation	.582**	-.382**	1	-.432**	-.296**
	Sig. (2-tailed)	.000	.000		.000	.000
	N	313	313	313	313	313
Q46	Pearson Correlation	-.483**	.562**	-.432**	1	.500**
	Sig. (2-tailed)	.000	.000	.000		.000
	N	313	313	313	313	313
Q47	Pearson Correlation	-.364**	.451**	-.296**	.500**	1
	Sig. (2-tailed)	.000	.000	.000	.000	
	N	313	313	313	313	313

\*\* . Correlation is significant at the 0.01 level (2-tailed).

2. Use Compute to make reverse scored items using the reverse score formula

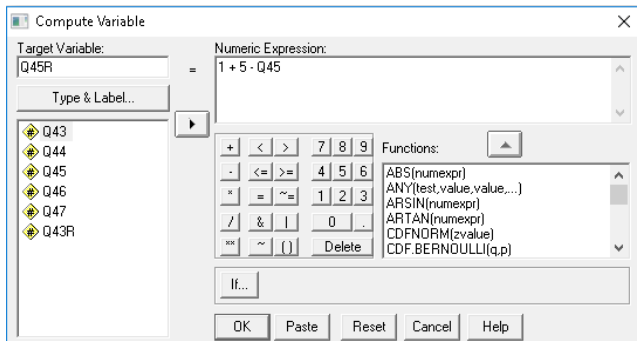


05-reverse-scores-academic-control.sav - SPSS Data Editor

File Edit View Data Transform Analyze Graphs Utilities Add-ons Window Help

	Q43	Q44	Q45	Q46	Q47	Q43R	var
1	5.00	1.00	3.00	1.00	1.00	1.00	
2	5.00	4.00	3.00	2.00	1.00	1.00	
3	3.00	5.00	3.00	3.00	1.00	3.00	
4	3.00	3.00	4.00	3.00	1.00	3.00	
5	1.00	3.00	3.00	4.00	1.00	5.00	
6	3.00	2.00	2.00	1.00	2.00	3.00	
7	2.00	2.00	2.00	2.00	2.00	4.00	
8	2.00	4.00	2.00	3.00	2.00	4.00	
9	2.00	4.00	2.00	3.00	2.00	4.00	
10	2.00	4.00	2.00	3.00	2.00	4.00	
11	2.00	4.00	2.00	3.00	2.00	4.00	
12	2.00	4.00	2.00	3.00	2.00	4.00	
13	4.00	5.00	2.00	3.00	2.00	2.00	
14	2.00	2.00	3.00	3.00	2.00	4.00	

Now for item 45



Check on reverse scoring

**Correlations**

		Q43	Q43R	Q45	Q45R
Q43	Pearson Correlation	1	-1.000**	.582**	-.582**
	Sig. (2-tailed)		.000	.000	.000
	N	313	313	313	313
Q43R	Pearson Correlation	<u>-1.000**</u>	1	-.582**	.582**
	Sig. (2-tailed)	<u>.000</u>		.000	.000
	N	313	313	313	313
Q45	Pearson Correlation	.582**	-.582**	1	-1.000**
	Sig. (2-tailed)	.000	.000		.000
	N	313	313	313	313
Q45R	Pearson Correlation	-.582**	.582**	<u>-1.000**</u>	1
	Sig. (2-tailed)	.000	.000	<u>.000</u>	
	N	313	313	313	313

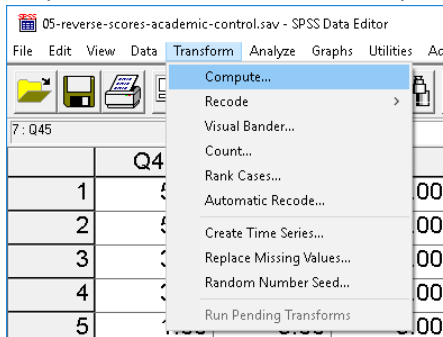
\*\* Correlation is significant at the 0.01 level (2-tailed).

## 7. SPSS Example for Composite Scores

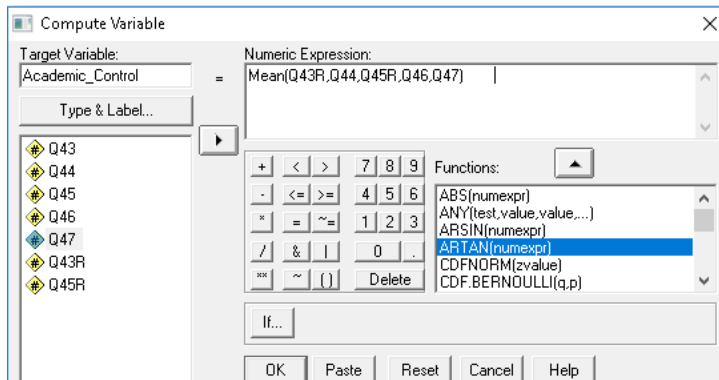
- Form mean of academic control
- Use reversed scores

Steps in Computing Composite Variables

1. Use Compute command to create composite scores.



2. Use Mean(variable1, variable2, variable3, etc.) function in Compute to create new variable. Use the positive and reversed items, do not use the original items that were reversed. See below.



3. Check accuracy. For the first person, the responses to items are:

Q43R = 1  
 Q44 = 1  
 Q45R = 3  
 Q46 = 1  
 Q47 = 1

Mean across these items is  $1+1+3+1+1 = 7/5 = 1.40$  which agrees with the mean for the first respondent below.

	Q43	Q44	Q45	Q46	Q47	Q43R	Q45R	Academic_Control
1	5.00	1.00	3.00	1.00	1.00	1.00	3.00	1.40
2	5.00	4.00	3.00	2.00	1.00	1.00	3.00	2.20
3	3.00	5.00	3.00	3.00	1.00	3.00	3.00	3.00
4	3.00	3.00	4.00	3.00	1.00	3.00	2.00	2.40
5	1.00	3.00	3.00	4.00	1.00	5.00	3.00	3.20
6	3.00	2.00	2.00	1.00	2.00	3.00	4.00	2.40
7	2.00	2.00	2.00	2.00	2.00	4.00	4.00	2.80
8	2.00	4.00	2.00	3.00	2.00	4.00	4.00	3.40
9	2.00	4.00	2.00	3.00	2.00	4.00	4.00	3.40
10	2.00	4.00	2.00	3.00	2.00	4.00	4.00	3.40
11	2.00	4.00	2.00	3.00	2.00	4.00	4.00	3.40
12	2.00	4.00	2.00	3.00	2.00	4.00	4.00	3.40
13	4.00	5.00	2.00	3.00	2.00	2.00	4.00	3.20
14	2.00	2.00	3.00	3.00	2.00	4.00	3.00	2.80

**8. Combining Difference Scales to Form Composites (to be added)**

**9. Factor Scores (to be added)**

**10. Gain Scores (to be added)**

**11. Normalized Gain Scores (to be added)**

## References

Bobko, P., Roth, P., & Buster, M. (2007). The usefulness of unit weights in creating composite scores. *Organizational Research Methods*, 10, 289–709.

Hendrickson A, Patterson B, & Melican G (2008). The Effect of Using Different Weights for Multiple-Choice and Free-Response Item Sections. Presentation at the National Council for Measurement in Education, New York.