Chapter 15

Exploratory Factor Analysis

Factor analysis is a data reduction technique. For instance, factor analysis can be used to identify the underlying components (factors) that explain the correlations among a set of variables. In this way, it is possible to employ a smaller set of measures (the factors) to explain a substantial portion of the total variance that is explained by all the original variables.

As an example, suppose you administer to a sample of adults a 25-item questionnaire asking questions about job satisfaction. Items may focus on rating of relationships with coworkers, intensity of interaction with coworkers, challenge provided by the daily tasks, valuing of the daily tasks, and the like. Factor analysis could be employed to reduce these individual variables to a much smaller set of factors that underlie job satisfaction such as satisfaction with coworkers and job commitment.

In exploratory factor analysis, there is no a priori assumption as to how the variables will combine to make factors. In confirmatory factor analysis, the researcher has a pre-established notion of which variables are associated with a given factor. In this chapter, we will use SPSS to conduct exploratory factor analysis.

15.1 CONDUCTING AN EXPLORATORY FACTOR ANALYSIS

We will examine exploratory factor analysis using data from a large study in which fourth grade teachers rated their students on a series of 15 behaviors (e.g., pays attention in class, thinks school is important, works well with others).[1] Each item was rated on a 5-point scale, with 1 representing "never" and 5 representing "always." We will conduct an exploratory factor analysis to find a few factors to explain most of the variation explained by the 15 behaviors. To do so, open the "behavior.sav" data file and follow the steps below.

1. Click on **Analyze** on the main menu bar.
2. Select **Data Reduction** from the pull-down menu.
3. Select **Factor** from the supplementary pull-down menu to open the Factor Analysis dialog box (see Fig. 15.1).
4. Highlight the variables to be included in the factor analysis (here, q1 through q15) and move them to the Variables box with the **top right arrow button**.
5. Click on **Extraction** to open the Factor Analysis: Extraction dialog box (Fig. 15.2). The default method of extraction is Principal components; the default criterion is to extract all factors with eigenvalues over 1; and the default is to analyze the correlation (rather than the covariance) matrix. Make any necessary changes to the settings and click **Continue.** (In this example, we will maintain the default settings.)
6. Click on **Rotation** to open the Factor Analysis: Rotation dialog box (Fig. 15.3). Select the rotation method desired (if any). In this example, we will select Varimax rotation (the most common form).[2] Click **Continue** to close the dialog box.
7. Click on **Options** to open the Factor Analysis: Options dialog box (Fig. 15.4). Click the **Sorted by size** option for the Coefficient Display Format section, and click on **Continue** to close the dialog box.
8. Click on **OK** to run the procedure.

---

[1] Finn, J. D., Pannozzo, G. M., & Voelkl, K. E. (1995). Disruptive and inattentive-withdrawn behavior and achievement among fourth-graders. Elementary School Journal , 95, 421-434.

[2] For a discussion of extraction and rotation, see Tabachnick, B. G., & Fidell, L. S. (1996). Using Multivariate Statistics. New York: HarperCollins.

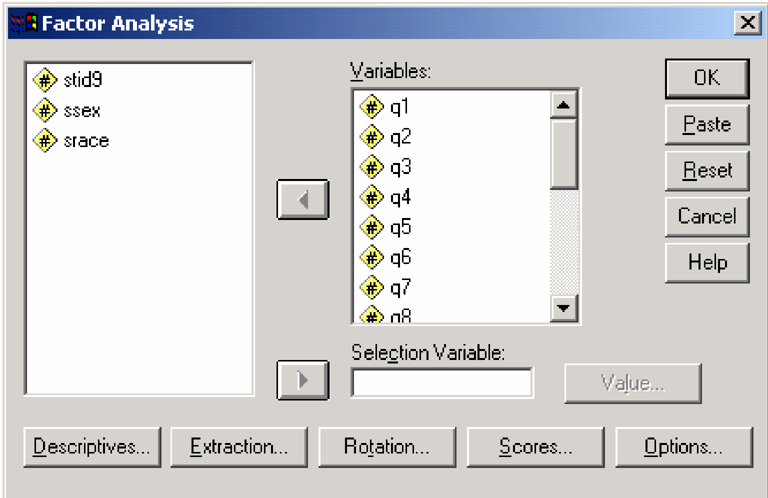

Figure 15.1 Factor Analysis Dialog Box

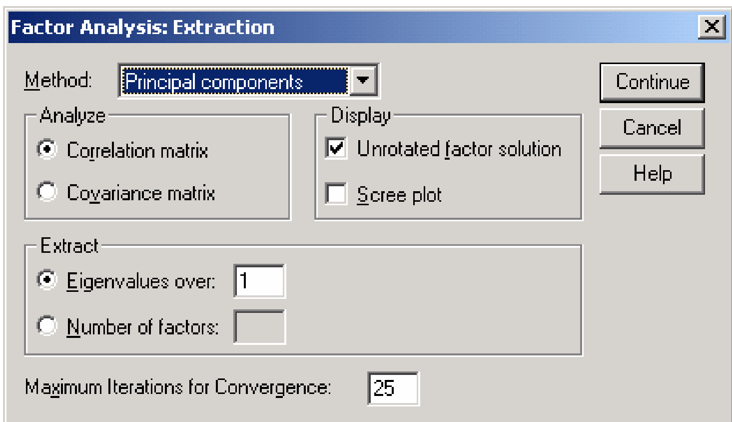

Figure 15.2 Factor Analysis: Extraction Dialog Box

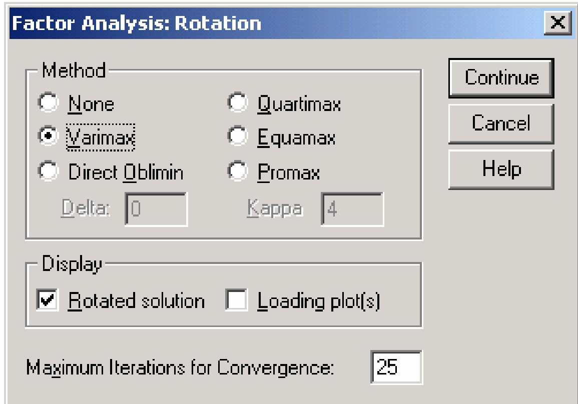

Figure 15.3 Factor Analysis: Rotation Dialog Box

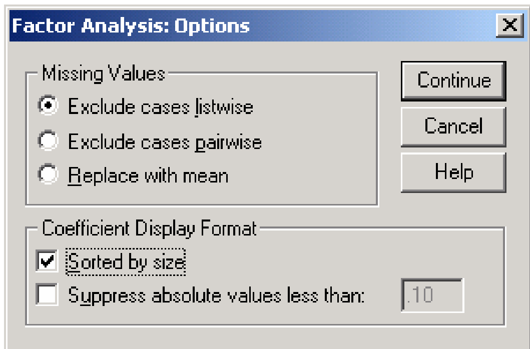

Figure 15.4 Factor Analysis: Options Dialog Box

15.2 INTERPRETING THE RESULTS OF THE FACTOR ANALYSIS PROCEDURE

The output of the factor analysis procedure we completed in Section 15.1 should appear as in Figure 15.5. The first table of the output reports communalities, which indicate the amount of variance in each variable that is accounted for. Because we have selected principal components analysis as the extraction method and opted to analyze the correlation matrix, the initial communalities for each variable are always 1.000. The extraction column indicates the variance in each variable accounted for by the factors (components) extracted as a result of the factor analysis.

Communalities

	Initial	Extraction
PAYS ATTENTION IN CLASS	1.000	.731
COMPLETES HOMEWORK ON TIME	1.000	.786
WORKS WELL W OTHERS	1.000	.610
TRIES TO DO WORK WELL	1.000	.749
PARTICIPATES IN DISCUSSIONS	1.000	.654
COMPLETES SEAT WORK	1.000	.780
THINKS SCHOOL IS IMPORTANT	1.000	.741
DOES EXTRA WORK	1.000	.644
MAKES EFFORT	1.000	.777
ASKS QUESTIONS	1.000	.699
TRIES TO FINISH DIFFICULT WORK	1.000	.742
RAISES HAND TO TALK	1.000	.613
SEEKS REFERENCE MATERIAL	1.000	.621
DISCUSSES OUTSIDE OF CLASS	1.000	.652
ATTENDS EXTRACURRICULAR ACTIVITIES	1.000	.332

Extraction Method: Principal Component Analysis.

Total Variance Explained

	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
Component	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	8.724	58.157	58.157	8.724	58.157	58.157	6.223	41.484	41.484
2	1.407	9.382	67.539	1.407	9.382	67.539	3.908	26.055	67.539
3	.793	5.285	72.824						
4	.563	3.755	76.578						
5	.506	3.376	79.954						
6	.476	3.171	83.125						
7	.382	2.548	85.673						
8	.329	2.192	87.865						
9	.322	2.146	90.012						
10	.307	2.048	92.060						
11	.290	1.934	93.994						
12	.263	1.756	95.750						
13	.248	1.650	97.401						
14	.220	1.464	98.864						
15	.170	1.136	100.000						

Extraction Method: Principal Component Analysis.

Figure 15.5 Output for Factor Analysis with Behavior Variables

The second table reports information on Initial Eigenvalues, Extraction Sums of Squared Loadings, and (if appropriate) Rotation Sum of Squared Loadings. Eigenvalues are measures of the variance in all the variables that are accounted for by a given factor. In the initial solution, there are as many factors (components) as there are variables. In our example, the first component accounts for 8.724 of the total variance; the second, 1.407, etc. The sum of the individual eigenvalues is equivalent to the variance in all of the variables (here, 15). The "% of variance" column represents the ratio of the variance explained by the individual component to the total variance. In our example, the first component accounts for 58.157% of the total variance (8.724 ÷15 = 58.157%). The initial solution is the complete solution; it explains 100% of the variance in the original variables. (This can be seen in the cumulative percent column; note that the cumulative percent for the fifteenth component is 100.00%.)

Component Matrix[a]

	Component	
	1	2
MAKES EFFORT	.869	-.149
THINKS SCHOOL IS IMPORTANT	.847	-.152
TRIES TO DO WORK WELL	.835	-.227
COMPLETES SEAT WORK	.832	-.297
TRIES TO FINISH DIFFICULT WORK	.831	-.228
PAYS ATTENTION IN CLASS	.831	-.203
COMPLETES HOMEWORK ON TIME	.829	-.315
DOES EXTRA WORK	.787	.155
RAISES HAND TO TALK	.750	.225
SEEKS REFERENCE MATERIAL	.738	.276
ASKS QUESTIONS	.737	.395
PARTICIPATES IN DISCUSSIONS	.729	.351
WORKS WELL W OTHERS	.720	-.303
ATTENDS EXTRACURRICULAR ACTIVITIES	.411	.404
DISCUSSES OUTSIDE OF CLASS	.555	.586

Extraction Method: Principal Component Analysis.

a. 2 components extracted.

Rotated Component Matrix[a]

	Component	
	1	2
COMPLETES HOMEWORK ON TIME	.856	.229
COMPLETES SEAT WORK	.848	.246
TRIES TO DO WORK WELL	.810	.304
TRIES TO FINISH DIFFICULT WORK	.807	.301
PAYS ATTENTION IN CLASS	.793	.321
MAKES EFFORT	.792	.387
THINKS SCHOOL IS IMPORTANT	.776	.372
WORKS WELL W OTHERS	.762	.175
DISCUSSES OUTSIDE OF CLASS	.107	.800
ASKS QUESTIONS	.367	.752
PARTICIPATES IN DISCUSSIONS	.386	.711
SEEKS REFERENCE MATERIAL	.438	.655
RAISES HAND TO TALK	.477	.621
DOES EXTRA WORK	.548	.586
ATTENDS EXTRACURRICULAR ACTIVITIES	9.697E-02	.568

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.

a. Rotation converged in 3 iterations.

Figure 15.5 Output for Factor Analysis with Behavior Variables, *continued*

One of the goals of factor analysis is parsimony, so we must balance explaining a substantial percentage of variation with limiting the number of factors (components) selected. The Extraction Sum of Squared Loadings section displays information regarding the factors/components extracted. Here, SPSS has extracted two factors/components, based on the extraction criterion we chose —

eigenvalues greater than 1. These two components explain 67.539% of the variance in the original variables. Because we selected principal components analysis as the method of extraction, the "Total," "% of Variance," and "Cumulative %" columns for this section are identical to those of the first two components in the Initial Eigenvalues section.

Because we selected varimax rotation, there is a third section — "Rotation Sum of Squared Loadings." The variance accounted for by individual rotated factors/components may differ from that accounted for by individual unrotated factors/components. However, the final cumulative percent of variance accounted for is equivalent (here, 67.539%).

The next two tables, the Component Matrix and the Rotated Component Matrix represent factor loadings for each of the original variables on the original (unrotated) and rotated solution, respectively. For the unrotated solution, factor loadings are the correlations between the specific item and unrotated factor. For the rotated solution, factor loadings are the partial correlation between the item and the rotated factor. These correlations are helpful in determining the structure underlying the individual items. That is, you can examine variables (here, responses to the questionnaire items) that have high loadings (correlations or partial correlations) on a particular factor, and attempt to discern a common thread among them.

The correlations (loadings) from the rotated solution are sometimes easier to interpret for this purpose. For instance, varimax rotation tends to simplify the factors. That is, items tend to load heavily on only one factor. Thus, interpretation is simplified because it is easier to determine which variable is associated with which factor. In addition, because we selected "Sorted by size" (see step 7 above), the variables are already somewhat grouped. That is, the first 8 questionnaire items (beginning with "completes homework on time" and ending with "works well with others" load heavily on the first component, and the remaining 7 items ("discusses outside of class" through "attends extracurricular activities") load heavily on the second component.

The final task is to determine what the items within the task have in common. The first component seems to relate to academic work and behavior, and the second component seems to relate to social interactions and initiative.

15.3 SCALE RELIABILITY

An important characteristic of educational and psychological measurements is their reliability. Reliability indicates the amount of variation to expect in the measurement from one occasion to another. Indexes of reliability — reliability coefficients — can be computed that range from 0.0 (the measure has no reliability and may be expected to vary a great deal) to 1.0 (the measure has perfect reliability and will be consistent from one occasion to another. This is the scale of all the reliability coefficients discussed here.

The reliability of a measure can be determined in several ways. One way is to administer the measurement instrument at two different points in time — preferably close together — and to correlate the results from the two administrations. This is called "test-retest reliability" and can be done with SPSS using the correlation procedure described in Chapter 5. This approach can be taken with any measurement instrument having one or more items.

For multi-item measurement scales, more common indexes of reliability reflect the "internal consistency" of the scale. To compute internal consistency coefficients, the measure only needs to be administered once. The indexes reflect the extent to which scores are consistent from one portion of the measurement instrument to another. One internal consistency measure — the "split half" coefficient — is obtained by scoring two halves of the measurement scales and then correlating the results on one half with the results on the other half. The halves may be the first half of the instrument and the second half or, preferably, the odd-numbers items (1, 3, 5, ...) and the even-numbered items (2, 4, 6, ...). The correlation between scores on the two halves, however, only gives an estimate of the reliability of half of the test. It must be adjusted to estimate the reliability of the full-length test. The Spearman-Brown prophecy formula accomplishes this; it is applied automatically by SPSS when the split-half method is chosen.

A commonly used coefficient of internal consistency — Cronbach's coefficient alpha — is based on the consistency of responses from one item to another. It is often preferable to split-half reliability because it avoids making an arbitrary decision about what the halves of the measurement instrument are.

Coefficient alpha is interpreted in the same way as other reliability coefficients, with 0.0 indicating that the instrument has no reliability and 1.0 indicating the highest possible reliability. When you use the Reliability Analysis procedure in SPSS, the output includes item-by-item statistics indicating how each item contributes to the internal consistency of the measuring instrument. Items that do not contribute much to the reliability of the total measurement can be rewritten or deleted to improve the measurement scale. Also, the output indicates what the reliability of the total measurement scale would be if all items were "standardized" to the same metric (item mean = 0.0; item standard deviation = 1.0). If this value is much higher than the alpha-value for unstandardized items, it may be worthwhile taking this step when scoring the total scale for further use.

WARNING: SPSS only computes coefficient alpha correctly when all items are scored in the same direction. Thus, if high scores indicate high levels of the construct on some items and low levels of the construct on other items, one or the other set of items must be reverse scored. For example, suppose that you are rating student behavior and one item is "student studies hard" with response categories 1 = never, 2 = sometimes, 3 = always. Perhaps another item is "student is disruptive" with response categories 1 = never, 2 = sometimes, 3 = always. In order for coefficient alpha to be computed correctly by SPSS, one or the other item needs to be reversed. You can accomplish this by recoding the second item (recode "1" to "3" and "3" to "1") or by a transformation (scores on the transformed variable are "4 - the original item score") (see Chapter 1). In either case the revised item scores are used in the Reliability analysis.

15.4 COMPUTING SPLIT-HALF COEFFICIENT ESTIMATES

Recall that split-half coefficients represent the consistency in responding between the first half and second half of the items listed on the reliability analysis. Choose the items to include in each half so that the halves are as equivalent as possible. A common method of splitting the items is to assign the odd numbered items to one half, and the even numbered items to the other half.

Let's calculate the internal consistency of the behavior scale described in Section 15.1. To compute a split-half coefficient:

1. Click on **Analyze** on the main menu bar.
2. Click on **Scale** from the pull-down menu.
3. Click on **Reliability Analysis** to open the Reliability Analysis dialog box (see Fig. 15.6).
4. Highlight the odd numbered variables (q1, q3, q5 to q15) to be included in the first half of the reliability analysis and move them into the Items box with the **right arrow button**.
5. Highlight the even numbered variables (q2, q4, q6 to q14) to be included in the second half of the reliability analysis and move them into the Items box with the **right arrow button**.
6. Click on **Split-half** in the drop down menu labeled Model.
7. Click on **OK** to run the procedure.

The results of the analysis are shown in Figure 15.7. The reliability coefficient is calculated using the Spearman-Brown formula. Because there was an odd number of items in the scale, the split produced an unequal number of items in each half, so the unequal-length Spearman-Brown reliability coefficient is reported. Notice that the value of the coefficient is .956 indicating that the scale has good internal consistency. Additional statistics such as descriptives and inter-item correlations may be obtained by clicking on the **Statistics** button in the reliability analysis dialog box.

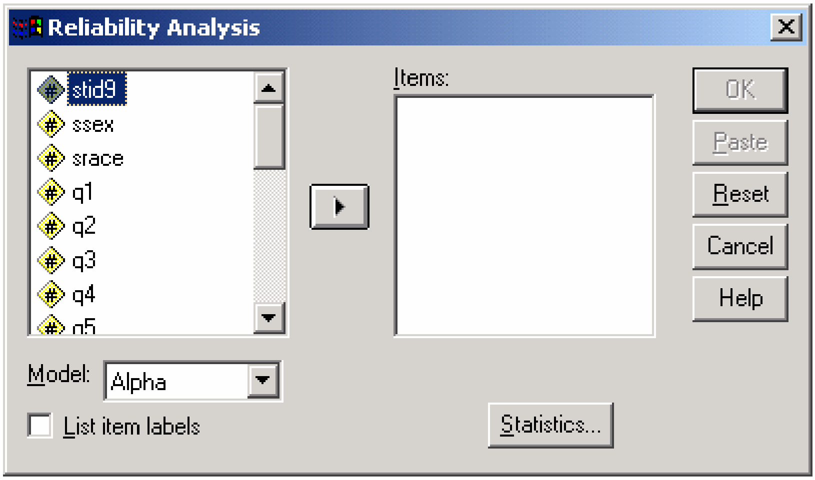

Figure 15.6 Reliability Analysis Dialog Box

Reliability Statistics

Cronbach's Alpha	Part 1	Value	.896
		N of Items	8[a]
	Part 2	Value	.890
		N of Items	7[b]
	Total N of Items		15
Correlation Between Forms			.916
Spearman-Brown Coefficient	Equal Length		.956
	Unequal Length		.956
Guttman Split-Half Coefficient			.954

a. The items are: PAYS ATTENTION IN CLASS, WORKS WELL W OTHERS, PARTICIPATES IN DISCUSSIONS, THINKS SCHOOL IS IMPORTANT, MAKES EFFORT, TRIES TO FINISH DIFFICULT WORK, SEEKS REFERENCE MATERIAL, ATTENDS EXTRACURRICULAR ACTIVITIES.

b. The items are: ATTENDS EXTRACURRICULAR ACTIVITIES, COMPLETES HOMEWORK ON TIME, TRIES TO DO WORK WELL, COMPLETES SEAT WORK, DOES EXTRA WORK, ASKS QUESTIONS, RAISES HAND TO TALK, DISCUSSES OUTSIDE OF CLASS.

Figure 15.7 Output for Split-Half Reliability Analysis

15.5 COMPUTING CRONBACH ALPHA COEFFICIENT

Cronbach alpha is another coefficient that can be used to determine reliability of a scale. To compute coefficient alpha for the 15-item behavior scale:

1. Click on **Analyze** on the main menu bar.

Reliability Statistics

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.945	.946	15

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
PAYS ATTENTION IN CLASS	50.35	141.862	.786	.673	.940
COMPLETES HOMEWORK ON TIME	50.17	140.249	.778	.746	.940
WORKS WELL W OTHERS	50.18	144.029	.661	.535	.943
TRIES TO DO WORK WELL	50.40	138.176	.787	.694	.940
PARTICIPATES IN DISCUSSIONS	50.56	140.602	.697	.578	.942
COMPLETES SEAT WORK	50.04	141.090	.782	.749	.940
THINKS SCHOOL IS IMPORTANT	50.16	139.077	.807	.693	.939
DOES EXTRA WORK	51.49	137.128	.755	.608	.940
MAKES EFFORT	50.38	138.141	.830	.735	.939
ASKS QUESTIONS	50.76	141.420	.710	.599	.941
TRIES TO FINISH DIFFICULT WORK	50.32	138.041	.782	.696	.940
RAISES HAND TO TALK	50.38	140.880	.716	.568	.941
SEEKS REFERENCE MATERIAL	51.17	138.807	.705	.561	.942
DISCUSSES OUTSIDE OF CLASS	51.31	144.055	.524	.413	.946
ATTENDS EXTRACURRICULAR ACTIVITIES	50.75	150.259	.379	.189	.949

Figure 15.8 Output for Coefficient Alpha Reliability Analysis

2. Select **Scale** from the pull-down menu.
3. Click on **Reliability Analysis** to open the Reliability Analysis dialog box.
4. Highlight the variables (q1 to q15) to be included in the reliability analysis and move them into the Items box with the **right arrow button**.
5. Click on **Statistics** to open the Reliability Analysis: Statistics dialog box.
6. Click on **Item** and **Scale** and **Scale if item deleted** in the Descriptives for area.
7. Click on **Correlations** in the Inter-item area.
8. Click on **Continue**.
9. Click on **Alpha** in the drop down menu labeled Model.
10. Click on **OK** to run the procedure.

A portion of the output is shown in Figure 15.8. The Reliability Statistics Table indicates that Cronbach's alpha for the 15-item scale is .945, indicating good reliability. The Item-Total Statistics table displays information about the scale as if it were calculated without each item. This allows the analyst to gain some information on how individual items contribute to the whole.

For instance, the last column indicates the value of Cronbach's alpha if each single item is deleted. In this example, omitting the item that measures the extent to which the student "thinks school is important" would result in a reliability coefficient of .939. This is a decrease from the calculation with all items (.945). Thus, we consider this item — thinks school is important — to contribute positively to the construct the scale measures. On the other hand, if the reliability were calculated without reference to the extent to which the student "attends extracurricular activities," the reliability would increase slightly, to .949.

Chapter Exercises

15.1 The "crime.sav" data file contains information on crime rates per 100,000 in the 50 states in the US. Use this datafile to conduct an exploratory factor on US crime rates, with the default extraction methods and criterion, and selecting varimax rotation. On the basis of the factor analysis, answer the following:

- **a.** What percent of total variance is accounted for by the component(s)?
- **b.** What is the number of eigenvalues greater than 1?
- **c.** How many components are extracted?
- **d.** If there is more than one component, use the rotated component matrix to interpret the components. What meaning would you attach to the factor(s)?

15.2 Based on the results of the factor analysis of classroom behaviors conducted in this chapter, conduct a reliability analysis using the eight variables that loaded heavily on the first factor, academic work. These variables are: Q1, Q2, Q3, Q4, Q6, Q7, Q9, Q11.

a. What is the coefficient alpha?

b. Does the reliability of the academic scale increase or decrease if you exclude the item "completes homework on time" (Q11)? Hint: use the options in the Reliability Analysis: Statistics dialog box.

15.3 Using the "fire.sav" data file, calculate the reliability coefficient for a fitness scale based on the items: stair clime time, body drag time, and obstacle course time. Discuss the internal consistency of the fitness scale.