# CHAPTER 4

# Exploratory Factor Analysis and Principal Components Analysis

Exploratory factor analysis (EFA) and principal components analysis (PCA) both are methods that are used to help investigators represent a large number of relationships among normally distributed or scale variables in a simpler (more parsimonious) way. Both of these approaches determine which, of a fairly large set of items, "hang together" as groups or are answered most similarly by the participants. EFA also can help assess the level of construct (factorial) validity in a dataset regarding a measure purported to measure certain constructs. A related approach, **confirmatory factor analysis,** in which one tests very specific models of how variables are related to underlying constructs (conceptual variables), requires additional software and is beyond the scope of this book so it will not be discussed.

The primary difference, conceptually, between **exploratory factor analysis** and **principal components analysis** is that in EFA one postulates that there is a smaller set of unobserved (latent) variables or constructs underlying the variables actually observed or measured (this is commonly done to assess validity), whereas in PCA one is simply trying to mathematically derive a relatively small number of variables to use to convey as much of the information in the observed/measured variables as possible. In other words, EFA is directed at *understanding* the relations among variables by understanding the constructs that underlie them, whereas PCA is simply directed toward enabling one to derive fewer variables to provide the same information that one would obtain from the larger set of variables.

There are actually a number of different ways of computing factors for factor analysis; in this chapter, we will use only one of these methods, **principal axis factor analysis** (PA). We selected this approach because it is highly similar mathematically to PCA. The primary difference, computationally, between PCA and PA is that in the former the analysis typically is performed on an ordinary correlation matrix, complete with the correlations of each item or variable with itself. In contrast, in PA factor analysis, the correlation matrix is modified such that the correlations of each item with itself are replaced with a "communality"—a measure of that item's relation to all other items (usually a squared multiple correlation). Thus, with PCA the researcher is trying to reproduce all information (variance and covariance) associated with the set of variables, whereas PA factor analysis is directed at understanding only the covariation among variables.



### Conditions for Exploratory Factor Analysis and Principal Components Analysis

There are two main conditions necessary for factor analysis and principal components analysis. The first is that there need to be relationships among the variables. Further, the larger the sample size, especially in relation to the number of variables, the more reliable the resulting factors. Sample size is less crucial for factor analysis to the extent that the communalities of items with the other items are high, or at least relatively high and variable. Ordinary principal axis factor analysis should never be done if the number of items/variables is greater than the number of participants.

### Assumptions for Exploratory Factor Analysis and Principal Components Analysis

The methods of extracting factors and components that are used in this book do not make strong distributional assumptions; normality is important only to the extent that skewness or outliers affect the observed correlations or if significance tests are performed (which is rare for EFA and PCA). The normality of the distribution can be checked by computing the skewness value of each variable. Maximum likelihood estimation, which we will not cover, does require multivariate normality; the variables need to be normally distributed and the joint distribution of all the variables should be normal. Because both principal axis factor analysis and principal components analysis are based on correlations, independent sampling is required and the variables should be related to each other (in pairs) in a linear fashion. The assumption of linearity can be assessed with matrix scatterplots, as shown in Chapter 2. Finally, each of the variables should be correlated at a moderate level with some of the other variables. Factor analysis and principal components analysis seek to explain or reproduce the correlation matrix, which would not be a sensible thing to do if the correlations all hover around zero. Bartlett's test of sphericity addresses this assumption. However, if correlations are too high, this may cause problems with obtaining a mathematical solution to the factor analysis.

- Retrieve your data file: **hsbdataNew.sav.**

## Problem 4.1: Factor Analysis on Math Attitude Variables

In Problem 4.1, we perform a principal axis factor analysis on the math attitude variables. Factor analysis is more appropriate than PCA when one has the belief that there are latent variables underlying the variables or items measured. In this example, we have beliefs about the constructs underlying the math attitude questions; we believe that there are three constructs: *motivation*, *competence*, and *pleasure*. Now, we want to see if the items that were written to index each of these constructs actually do "hang together"; that is, we wish to determine empirically whether participants' responses to the motivation questions are more similar to each other than to their responses to the competence items, and so on. Conducting factor analysis can assist us in validating the data: if the data do fit into the three constructs that we believe exist, then this gives us support for the construct validity of the math attitude measure in this sample. The analysis is considered exploratory factor analysis even though we have some ideas about the structure of the data because our hypotheses regarding the model are not very specific; we do not have specific predictions about the size of the relation of each observed variable to each latent variable, etc. Moreover, we "allow" the factor analysis to find factors that best fit the data, even if this deviates from our original predictions.



4.1 Are there three constructs (*motivation*, *competence*, and *pleasure*) underlying the math attitude questions?

To answer this question, we will conduct a factor analysis using the principal axis factoring method and specify the number of factors to be three (because our conceptualization is that there are three math attitude scales or factors: *motivation*, *competence*, and *pleasure*).

- **Analyze → Dimension Reduction → Factor…** to get Fig. 4.1.
- Next, select the variables *item01* through *item14*. Do not include *item04r* or any of the other reversed items because we are including the unreversed versions of those same items.

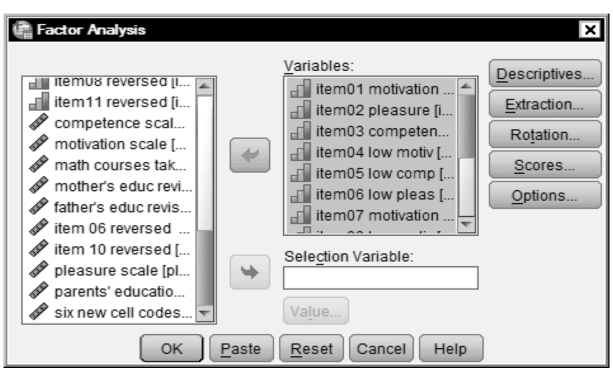



**Fig. 4.1. Factor analysis.**

- Now click on **Descriptives…** to produce Fig. 4.2.
- Then click on the following: **Initial solution** and **Univariate Descriptives** (under **Statistics), Coefficients, Determinant,** and **KMO and Bartlett's test of sphericity** (under **Correlation Matrix**).
- Click on **Continue** to return to Fig. 4.1.

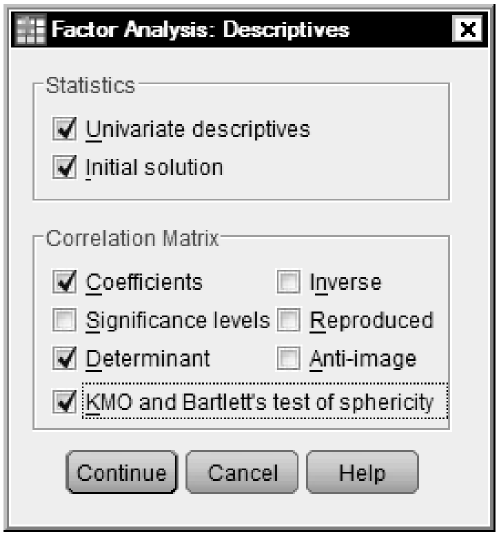



**Fig. 4.2. Factor analysis: Descriptives.**

- Next, click on **Extraction…** This will give you Fig. 4.3.
- Select **Principal axis factoring** from the **Method** pull-down menu.
- *Unclick* **Unrotated factor solution** (under **Display**). We will examine this only in Problem 4.2. We also usually would check the **Scree plot** box. However, again, we will request and interpret the scree plot only in Problem 4.2.
- Click on **Fixed number of factors** under **Extract,** and type **3** in the box. This setting instructs the computer to extract three math attitude factors.
- Click on **Continue** to return to Fig. 4.1.



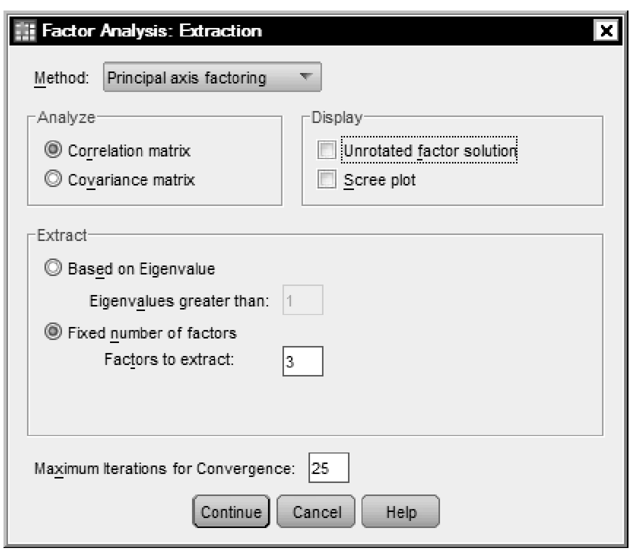



**Fig. 4.3. Extraction method to produce principal axis factoring.**

- Now click on **Rotation…** in Fig. 4.1, which will give you Fig. 4.4.

- Click on **Varimax**, then make sure **Rotated solution** is also checked. Varimax rotation creates a solution in which the factors are orthogonal (uncorrelated with one another), which can make results easier to interpret and to replicate with future samples. If you believe that the factors (latent concepts) are correlated, you could choose **Direct Oblimin,** which will provide an oblique solution allowing the factors to be correlated.
- Click on **Continue**.

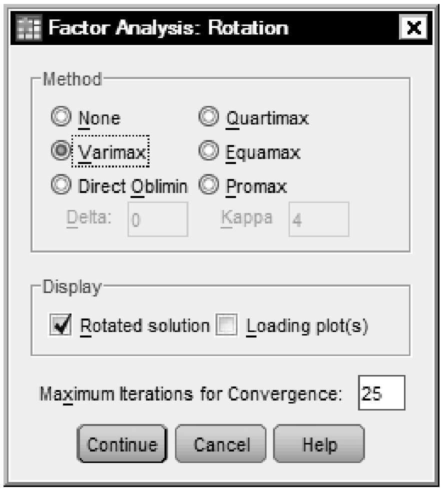



**Fig. 4.4. Factor analysis: Rotation.**

- Next, click on **Options…**, which will give you Fig. 4.5.
- Click on **Sorted by size**.
- Click on **Suppress small coefficients** and type **.3** (point 3) in the **Absolute Value below** box (see Fig. 4.5). Suppressing small factor loadings makes the output easier to read.
- Click on **Continue** then **OK**. Compare Output 4.1 with your output and syntax.



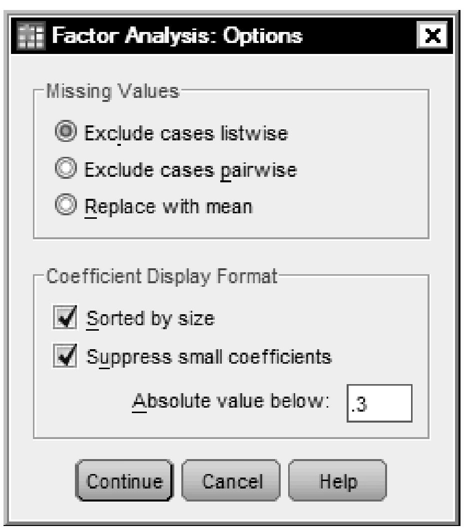



**Fig. 4.5. Factor analysis: Options.**

## Output 4.1: Factor Analysis for Math Attitude Questions

```
FACTOR
  /VARIABLES item01 item02 item03 item04 item05 item06 item07 item08 item09 item10 item11 item12
item13 item14
  /MISSING LISTWISE
  /ANALYSIS item01 item02 item03 item04 item05 item06 item07 item08 item09 item10 item11 item12
item13 item14
```

```
/PRINT UNIVARIATE INITIAL CORRELATION DET KMO EXTRACTION ROTATION
/FORMAT SORT BLANK(.3)
/CRITERIA FACTORS(3) ITERATE(25)
/EXTRACTION PAF
/CRITERIA ITERATE(25)
/ROTATION VARIMAX
/METHOD=CORRELATION.
```

## Factor Analysis

**Descriptive Statistics**

| | Mean | Std. Deviation | Analysis N |
|---|---|---|---|
| item01 motivation | 2.99 | .918 | 71 |
| item02 pleasure | 3.58 | .822 | 71 |
| item03 competence | 2.82 | .915 | 71 |
| item04 low motiv | 2.21 | .909 | 71 |
| item05 low comp | 1.61 | .948 | 71 |
| item06 low pleas | 2.44 | .996 | 71 |
| item07 motivation | 2.77 | 1.072 | 71 |
| item08 low motiv | 1.96 | .917 | 71 |
| item09 competence | 3.32 | .770 | 71 |
| item10 low pleas | 1.41 | .748 | 71 |
| item11 low comp | 1.38 | .763 | 71 |
| item12 motivation | 2.99 | .837 | 71 |
| item13 motivation | 2.68 | .807 | 71 |
| item14 pleasure | 2.86 | .723 | 71 |

**Interpretation of Output 4.1**

The factor analysis program generates a variety of tables depending on which options you have chosen. The first table includes **Descriptive Statistics** for each variable and the **Analyses N,** which in this case is 71 because several items have one or more participants missing. It is especially important to check the Analysis N when you have a small sample, scattered missing data, or one variable with lots of missing data. In the latter case, it may be wise to run the analysis without that variable.



**Correlation Matrix[a]**

| | | item01 motivation | item02 pleasure | item03 competence | item04 low motiv | item05 low comp | item06 low pleas | item07 motivation | item0 mo |
|---|---|---|---|---|---|---|---|---|---|
| Correlation | item01 motivation | 1.000 | .484 | .626 | -.305 | -.745 | -.165 | .461 | |
| | item02 pleasure | .484 | 1.000 | .389 | -.166 | -.547 | -.312 | .361 | |
| | item03 competence | .626 | .389 | 1.000 | -.348 | -.743 | -.209 | .423 | |
| | item04 low motiv | -.305 | -.166 | -.348 | 1.000 | .363 | .323 | -.596 | |
| | item05 low comp | -.745 | -.547 | -.743 | .363 | 1.000 | .260 | -.538 | |
| | item06 low pleas | -.165 | -.312 | -.209 | .323 | .260 | 1.000 | -.268 | |
| | item07 motivation | .461 | .361 | .423 | -.596 | -.538 | -.268 | 1.000 | |
| | item08 low motiv | -.340 | -.176 | -.248 | .576 | .276 | .192 | -.606 | |
| | item09 competence | .209 | .219 | .328 | -.120 | -.351 | -.131 | .228 | |
| | item10 low pleas | .071 | -.389 | .027 | .102 | .130 | .217 | -.169 | |
| | item11 low comp | -.441 | -.401 | -.513 | .398 | .605 | .418 | -.331 | |
| | item12 motivation | .186 | .116 | .165 | -.391 | -.187 | -.044 | .347 | |
| | item13 motivation | .187 | .028 | .170 | -.334 | -.169 | .001 | .361 | |
| | item14 pleasure | .040 | .475 | .068 | -.063 | -.166 | -.469 | .180 | |

a. Determinant = .001

Should be greater than .0001. If very close to zero, collinearity is too high. If zero, no solution is possible.

Indicates how each question is associated (correlated) with each of the other questions. Only part of the matrix is included so font would not be too small to read.

**Interpretation of Output 4.1 continued**

The second table is part of a **correlation matrix** showing how each of the 14 items is associated with each of the other 13. Note that some of the correlations are high (e.g., + or −.60 or greater) and some are low (i.e., near zero). Relatively high correlations indicate that two items are associated and will probably be grouped together by the factor analysis. Items with low correlations (e.g., ≤.20) usually will not have high loadings on the same factor.

One assumption is that the **determinant** (located under the correlation matrix) should be more than .0001. Here, it is .001 so this assumption is met. If the determinant is zero, then a factor analytic solution cannot be obtained, because this would require dividing by zero, which would mean that at least one of the items can be understood as a linear combination of some set of the other items.

**KMO and Bartlett's Test**

Tests of assumptions.

| | | |
|---|---|---|
| Kaiser-Meyer-Olkin Measure of Sampling Adequacy. | | .770 |
| Bartlett's Test of Sphericity | Approx. Chi-Square | 433.486 |
| | df | 91 |
| | Sig. | .000 |

This is greater than .70 indicating sufficient items for each factor.

This is significant (less than .05), indicating that the correlation matrix is significantly different from an identity matrix, in which correlations between variables are all zero.

**Communalities**

| | Initial |
|---|---|
| item01 motivation | .660 |
| item02 pleasure | .542 |
| item03 competence | .598 |
| item04 low motiv | .562 |
| item05 low comp | .772 |
| item06 low pleas | .382 |
| item07 motivation | .607 |
| item08 low motiv | .533 |
| item09 competence | .412 |
| item10 low pleas | .372 |
| item11 low comp | .591 |
| item12 motivation | .499 |
| item13 motivation | .452 |
| item14 pleasure | .479 |

Extraction Method: Principal Axis Factoring.

These initial communalities represent the relation between the variable and all other variables (i.e., the squared multiple correlation between the item and all other items) before rotation. If many or most communalities are low (< .30), a small sample size is more likely to distort results.

**Interpretation of Output 4.1 continued**

The **Kaiser-Meyer-Olkin (KMO)** measure should be greater than .70 and is inadequate if less than .50. The KMO test tells us whether or not enough items are predicted by each factor. Here it is .77 so that is good. The **Bartlett test** should be significant (i.e., a significance value of less than .05); this means that the variables are correlated highly enough to provide a reasonable basis for factor analysis as in this case.

The **Communalities** table shows the **Initial** commonalities before rotation. See the call out box for more interpretation. Note that all the initial communalities are above .30, which is good.

Eigenvalues refer to the variance accounted for, in terms of the number of "items' worth" of variance each explains. So, Factor 1 explains almost as much variance as in five items.

Percent of covariation among items accounted for by each factor before and after rotation.

**Total Variance Explained**

| Factor | Initial Eigenvalues Total | Initial Eigenvalues % of Variance | Initial Eigenvalues Cumulative % | Rotation Sums of Squared Loadings Total | Rotation Sums of Squared Loadings % of Variance | Rotation Sums of Squared Loadings Cumulative % |
|---|---|---|---|---|---|---|
| 1 | 4.888 | 34.916 | 34.916 | 3.017 | 21.549 | 21.549 |
| 2 | 2.000 | 14.284 | 49.200 | 2.327 | 16.621 | 38.171 |
| 3 | 1.613 | 11.519 | 60.719 | 1.784 | 12.746 | 50.917 |
| 4 | 1.134 | 8.097 | 68.816 | | | |
| 5 | .904 | 6.459 | 75.275 | | | |
| 6 | .716 | 5.113 | 80.388 | | | |
| 7 | .577 | 4.125 | 84.513 | | | |
| 8 | .461 | 3.293 | 87.806 | | | |
| 9 | .400 | 2.857 | 90.664 | | | |
| 10 | .379 | 2.710 | 93.374 | | | |
| 11 | .298 | 2.126 | 95.500 | | | |
| 12 | .258 | 1.846 | 97.346 | | | |
| 13 | .217 | 1.551 | 98.897 | | | |
| 14 | .154 | 1.103 | 100.000 | | | |

Extraction Method: Principal Axis Factoring.

Half of the variance is accounted for by the first three factors.

### Interpretation of Output 4.1 continued

The **Total Variance Explained** table shows how the variance is divided among the 14 possible factors. Note that four factors have **eigenvalues** (a measure of explained variance) greater than 1.0, which is a common criterion for a factor to be useful. When the eigenvalue is less than 1.0 the factor explains less information than a single item would have explained. Most researchers would not consider the information gained from such a factor to be sufficient to justify keeping that factor. Thus, if you had not specified otherwise, the computer would have looked for the best four-factor solution by "rotating" four factors. Because we specified that we wanted only three factors rotated, only three will be rotated, as seen on the right side of the table under **Rotation Sums of Squared Loadings**.

For this and other analyses in this chapter, we will use an *orthogonal* rotation (varimax). This means that the final factors will be at right angles with each other. As a result, we can assume that the information explained by one factor is independent of the information in the other factors. Note that if we create scales by summing or averaging items with high loadings from each factor, these *scales* will *not* necessarily be uncorrelated; it is the best-fit *vectors* (factors) that are orthogonal.

**Factor Matrix** [a]

a. 3 factors extracted. 12 iterations required.

**Rotated Factor Matrix**[a]

| | Factor | | |
|---|---|---|---|
| | 1 | 2 | 3 |
| item05 low comp | -.897 | | |
| item03 competence | .780 | | |
| item01 motivation | .777 | | |
| item11 low comp | -.572 | | .355 |
| item12 motivation | | .721 | |
| item13 motivation | | .667 | |
| item08 low motiv | | -.619 | |
| item04 low motiv | | -.601 | |
| item07 motivation | .412 | .585 | |
| item09 competence | | .332 | |
| item14 pleasure | | | -.797 |
| item10 low pleas | | | .580 |
| item02 pleasure | .487 | | -.535 |
| item06 low pleas | | | .515 |

Extraction Method: Principal Axis Factoring.
Rotation Method: Varimax with Kaiser Normalization.

a. Rotation converged in 5 iterations.

The items cluster into these three groups defined by the highest loading on each item.



**Interpretation of Output 4.1 continued**

Factors are rotated so that they are easier to interpret. Rotation makes it so that, as much as possible, different items are explained or predicted by different underlying factors, and each factor explains more than one item. This is a condition called simple structure. Although this is the *goal* of rotation, in reality, this is not always achieved. One thing to look for in the **Rotated Matrix** of factor loadings is the extent to which simple structure is achieved.

The **Rotated Factor Matrix** table is key for understanding the results of the analysis. Factors are rotated so that they are easier to interpret. Rotation makes it so that, as much as possible, different items are explained or predicted by different underlying factors, and each factor explains more than one item. This is a condition called simple structure. Although this is the *goal* of rotation, in reality, this is not always achieved. One thing to look for in the **Rotated Matrix** of factor loadings is the extent to which simple structure is achieved.

Note that the analysis has sorted the 14 math attitude questions (*item01* to *item14*) into three somewhat overlapping groups of items, as shown by the circled items. The items are sorted so that the items that have the highest loading (not considering whether the correlation is positive or negative) from factor 1 (four items in this analysis) are listed first, and they are sorted from the one with the highest factor weight or loading (i.e., *item05*, with a loading of −.897) to the one with the lowest loading from <u>that first factor</u> (*item11*). Actually, every item has some loading from every factor, but we requested for loadings less than |.30| to be excluded from the output, so there are blanks where low loadings exist. (|.30| means the absolute value, or value without considering the sign).

Next, the six items that have their highest loading from factor 2 are listed from highest loading (*item12*) to lowest (*item9*). Finally, the four items on which factor 3 loads most highly are listed in order. Loadings resulting from an orthogonal rotation are correlation coefficients between each item and the factor, so they range from −1.0 through 0 to + 1.0. A negative loading just means that the question needs to be interpreted in the opposite direction from the way it is written for that factor (e.g., *item05* "I am a little slow catching on to new topics in math" has a negative loading from the competence factor, which indicates that the people scoring higher on this item are lower in competence). Usually, factor loadings lower than |.30| are considered low, which is why we suppressed loadings less than |.30|. On the other hand, loadings of |.40| or greater are typically considered high. This is just a guideline, however, and one could set the criterion for "high" loadings as low as .30 or as high as .50. Setting the criterion lower than .30 or higher than .50 would be very unusual.

The investigator should examine the content of the items that have high loadings from each factor to see if they fit together conceptually and can be named. Items 5, 3, and 11 were intended to reflect a perception of *competence* at math, so the fact that they all have strong loadings from the same factor provides some support for their being conceptualized as pertaining to the same construct. On the other hand, *item01* was intended to measure *motivation* for doing math, but it is highly related to this same *competence* factor. In retrospect, one can see why this item could also be interpreted as competence. The item reads, "I practice math skills until I can do them well." Unless one felt one could do math problems well, this would not be true. Likewise, *item02*, "I feel happy after solving a hard problem," although intended to measure *pleasure* at doing math (and having its strongest loading there), might also reflect competence at doing math, in that, again, one could not endorse this item unless one had solved hard problems, which one could only do if one were good at math. Note that item02 loaded almost as highly (.49) on the competence factor (#1) as on the low pleasure factor (#3) so it loaded highly on two factors. On the other hand, *item09*, which was originally conceptualized as a competence item, had no really strong loadings.

Every item has a weight or loading from every factor, but in a "clean" factor analysis almost all of the loadings that are not in the circles that we have drawn on the **Rotated Factor Matrix** will be low (blank or less than |.40|). The fact that both Factors 1 and 3 load highly on *item02* and fairly highly on *item11*, and the fact that Factors 1 and 2 both load highly on *item07* is common but undesirable, in that one wants only one factor to predict each item.



**Factor Transformation Matrix**

| Factor | 1 | 2 | 3 |
|---|---|---|---|
| 1 | .747 | .552 | -.370 |
| 2 | -.162 | .692 | .704 |
| 3 | .645 | -.466 | .606 |

Extraction Method: Principal Axis Factoring.
Rotation Method: Varimax with Kaiser Normalization.

We will ignore this; it was used to convert the initial factor matrix into the rotated factor matrix.

## Example of How to Write About Problem 4.1

### Results

Principal axis factor analysis with varimax rotation was conducted to assess the underlying structure for the 14 items of the Math Attitude Questionnaire. (The assumption of independent sampling was met. The assumptions of normality, linear relationships between pairs of variables, and the variables' being correlated at a moderate level were checked.) Three factors were requested, based on the fact that the items were designed to index three constructs: motivation, competence, and pleasure. After rotation, the first factor accounted for 21.5% of the variance, the second factor accounted for 16.6%, and the third factor accounted for 12.7%. Table 4.1 displays the items and factor loadings for the rotated factors, with loadings less than .40 omitted to improve clarity.

Table 4.1
*Factor Loadings from Principal Axis Factor Analysis with Varimax Rotation for a Three-Factor Solution for Math Attitude Questions (N =71)*

| Item | Factor Loading | | | |
|---|---|---|---|---|
| | 1 | 2 | 3 | Communality |
| Slow catching on to new topics | −.90 | | | .77 |
| Solve math problems quickly | .78 | | | .60 |
| Practice math until do well | .78 | | | .66 |
| Have difficulties doing math | −.57 | | | .59 |
| Try to complete math even if takes long | | .72 | | .50 |
| Explore all possible solutions | | .67 | | .45 |
| Do not keep at it long if problem challenging | | −.62 | | .53 |
| Give up easily instead of persisting | | −.60 | | .56 |
| Prefer to figure out problems without help | .41 | .59 | | .61 |
| Really enjoy working math problems | | | −.80 | .48 |
| Smile only a little when solving math problem | | | .58 | .37 |
| Feel happy after solving hard problem | .49 | | −.54 | .54 |
| Do not get much pleasure out of math | | | .52 | .38 |
| Eigenvalues | 3.02 | 2.33 | 1.78 | |
| % of variance | 21.55 | 16.62 | 12.75 | |

*Note*. Loadings < .40 are omitted.



The first factor, which seems to index competence, had strong loadings on the first four items. Two of the items indexed low competence and had negative loadings. The second factor, which seemed to index motivation, had high loadings on the next five items in Table 4.1. “I prefer to figure out the problem without help” had its highest loading from the second factor but had a cross-loading over .4 on the competence factor. The third factor, which seemed to index low pleasure from math, loaded highly on the last four items in the table. “I feel happy after solving a hard problem” had its highest loading from the pleasure factor but also had a strong loading from the competence factor.

## Problem 4.2: Principal Components Analysis on Achievement Variables

Principal components analysis is most useful if one simply wants to reduce a relatively large number of variables to a smaller number of variables that still capture the same information. In this problem we will look at the initial (unrotated) solution as well as the rotated solution because we might want to use the first, unrotated, principal component to summarize all of the variables if it explains most of the variance rather using multiple, rotated components. This would especially be true if the scree plot suggests a large drop-off after the first component in variance explained (eigenvalues), so we will look at the scree plot too.

4.2 Run a principal components analysis to see how the five “achievement” variables cluster. These variables are *grades in h.s., math achievement, mosaic pattern test, visualization test,* and *scholastic aptitude test – math*.

- Click on **Analyze → Dimension Reduction → Factor...**
- First press **Reset**.
- Next select the variables *grades in h.s., math achievement, mosaic pattern test, visualization test,* and *scholastic aptitude test – math*, similar to what we did in Fig. 4.1.

- In the **Descriptives** window (Fig. 4.2), check **Univariate descriptives, Initial solution, Coefficients, Determinant**, and **KMO and Bartlett's test of sphericity**. Click on **Continue.**
- In the **Extraction** window (Fig. 4.3), use the default **Method** of **Principal components**. Be sure that **unrotated factor solution** and **Eigenvalues over 1** checked. Also, request a **Scree plot** (to see if one component would do a good job in summarizing the data or if a different number of components would be preferable to the default based on the criterion of components with eigenvalues over 1).
- Click on **Continue.**
- In the **Rotation** window (Fig. 4.4), check **Varimax**. Under **Display**, check **Rotated solution** and **Loading plot(s)**.
- Click on **Continue** and then **OK**.

We have requested a principal components analysis for the extraction and some different options for the output to contrast with the earlier one. Compare Output 4.2 with your syntax and output.

## Output 4.2: Principal Components Analysis for Achievement Scores

```
FACTOR
  /VARIABLES grades mathach mosaic visual satm
  /MISSING LISTWISE
  /ANALYSIS grades mathach mosaic visual satm
  /PRINT UNIVARIATE INITIAL CORRELATION DET KMO EXTRACTION ROTATION
  /PLOT EIGEN ROTATION
  /CRITERIA MINEIGEN(1) ITERATE(25)
  /EXTRACTION PC
  /CRITERIA ITERATE(25)
  /ROTATION VARIMAX
  /METHOD=CORRELATION .
```



## Factor Analysis

**Descriptive Statistics**

| | Mean | Std. Deviation | Analysis N |
|---|---|---|---|
| grades in h.s. | 5.68 | 1.570 | 75 |
| math achievement test | 12.5645 | 6.67031 | 75 |
| mosaic, pattern test | 27.413 | 9.5738 | 75 |
| visualization test | 5.2433 | 3.91203 | 75 |
| scholastic aptitude test - math | 490.53 | 94.553 | 75 |

**Correlation Matrix**[a]

| | | grades in h.s. | math achievement test | mosaic, pattern test | visualization test | scholastic aptitude test - math |
|---|---|---|---|---|---|---|
| Correlation | grades in h.s. | 1.000 | .504 | -.012 | .127 | .371 |
| | math achievement test | .504 | 1.000 | .213 | .423 | .788 |
| | mosaic, pattern test | -.012 | .213 | 1.000 | .030 | .110 |
| | visualization test | .127 | .423 | .030 | 1.000 | .356 |
| | scholastic aptitude test - math | .371 | .788 | .110 | .356 | 1.000 |

a. Determinant = .210

**KMO and Bartlett's Test**

| | | |
|---|---|---|
| Kaiser-Meyer-Olkin Measure of Sampling Adequacy. | | .615 |
| Bartlett's Test of Sphericity | Approx. Chi-Square | 111.440 |
| | df | 10 |
| | Sig. | .000 |

This is acceptable but mediocre. Because KMO is >.5, but it indicated there may not be enough items for one of the components.

### Interpretation of 4.2

As in Problem 4.1, the **Descriptive Statistics** table provides the mean and SD for each item. The Analysis N is important because it tells you how many students have scores on all five of these variables; in this case there is no missing data so the N is 75. The **Correlation Matrix** shows how each of the five items is related to the other four; note that the mosaic scores are very weakly correlated with the other four variables (-.012 to .213).

In terms of assumptions, the **Determinant** is much larger than zero so that is good. The **KMO** is .615 so mediocre and may be a problem. The Bartlett test is significant ($p < .001$), which is good and indicates that the correlation s are not near zero.



**Communalities**

| | Initial | Extraction |
|---|---|---|
| grades in h.s. | 1.000 | .493 |
| math achievement test | 1.000 | .869 |
| mosaic, pattern test | 1.000 | .949 |
| visualization test | 1.000 | .330 |
| scholastic aptitude test - math | 1.000 | .748 |

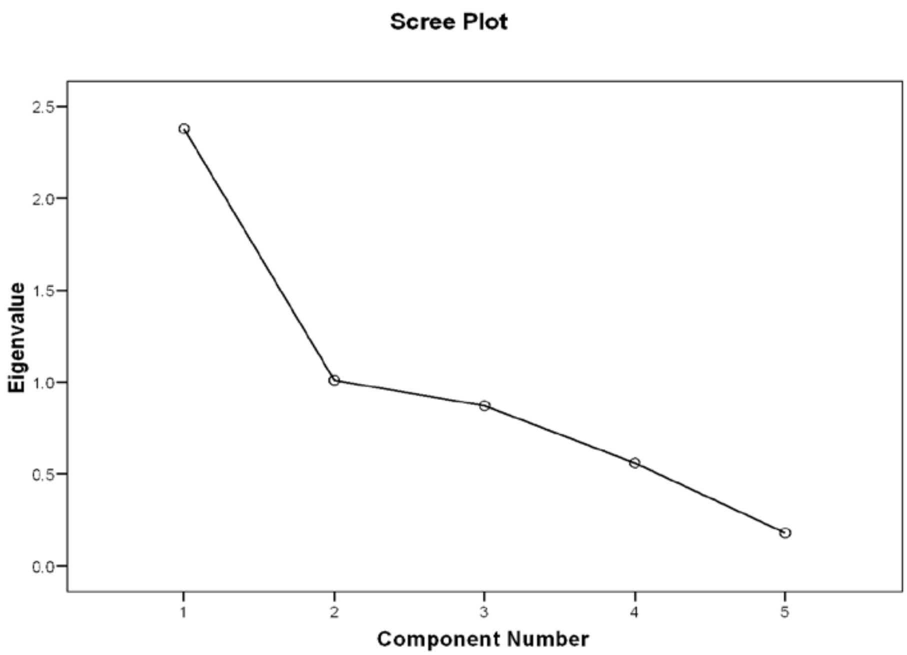
Extraction Method: Principal Component Analysis.

Note that 46% of the variance is explained by the first component.

**Total Variance Explained**

| Component | Initial Eigenvalues | | | Extraction Sums of Squared Loadings | | | Rotation Sums of Squared Loadings | | |
|---|---|---|---|---|---|---|---|---|---|
| | Total | % of Variance | Cumulative % | Total | % of Variance | Cumulative % | Total | % of Variance | Cumulative % |
| 1 | 2.379 | 47.579 | 47.579 | 2.379 | 47.579 | 47.579 | 2.340 | 46.805 | 46.805 |
| 2 | 1.010 | 20.198 | 67.777 | 1.010 | 20.198 | 67.777 | 1.049 | 20.972 | 67.777 |
| 3 | .872 | 17.437 | 85.214 | | | | | | |
| 4 | .560 | 11.197 | 96.411 | | | | | | |
| 5 | .179 | 3.589 | 100.000 | | | | | | |

Extraction Method: Principal Component Analysis.




The Scree plot shows that after the first two components, differences between the eigenvalues decline (the curve flattens), and they are less than 1.0. This again supports a two-component solution.

**Component Matrix[a]**

| | Component | |
|---|---|---|
| | 1 | 2 |
| grades in h.s. | .624 | -.322 |
| math achievement test | .931 | .044 |
| mosaic, pattern test | .220 | .949 |
| visualization test | .571 | -.056 |
| scholastic aptitude test - math | .865 | -.020 |

Extraction Method: Principal Component Analysis.

a. 2 components extracted.

This unrotated matrix should not be interpreted; it provides information about how the loadings change when the solution is rotated. However, the first unrotated component provides the simplest summary of the variables. In this case, it appears that if one used the first component only as the basis for creating summary scores, such scores would not include mosaic pattern score, which does not have a high loading for the first component.



### Interpretation of 4.2 continued

The **Total Variance Explained** table shows that there are two components with initial Eigenvalues more than 1.0, although the Eigenvalue for the second component is barely over 1 at 1.01. The first component explains 47.58% of the total variance, but because this is less than 50%, we probably want to rotate more than one component, as shown on the right hand side of this Total Variance Explained table.

The **Scree Plot** shows the initial Eigenvalues. Note that both the scree plot and the eigenvalues support the conclusion that these five variables can be reduced to two components. Note that the scree plot flattens out after the second component. However, the second component is very poorly defined, relating only to one variable. Thus, one may decide to use only one summary variable, based on all variables except *mosaic*, or to redo the PCA after omitting *mosaic*.  It usually is best for components to be defined by at least four variables.

The unrotated **Component Matrix** should not be interpreted. However, if you want to compute only one variable that provides the most information about this set of variables, a linear combination of the variables with high loadings from the first component of the unrotated matrix would be used.

**Rotated Component Matrix**[a]

| | Component | |
|---|---|---|
| | 1 | 2 |
| grades in h.s. | .669 | -.213 |
| math achievement test | .911 | .200 |
| mosaic, pattern test | .057 | .972 |
| visualization test | .573 | .041 |
| scholastic aptitude test - math | .856 | .126 |

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

a. Rotation converged in 3 iterations.

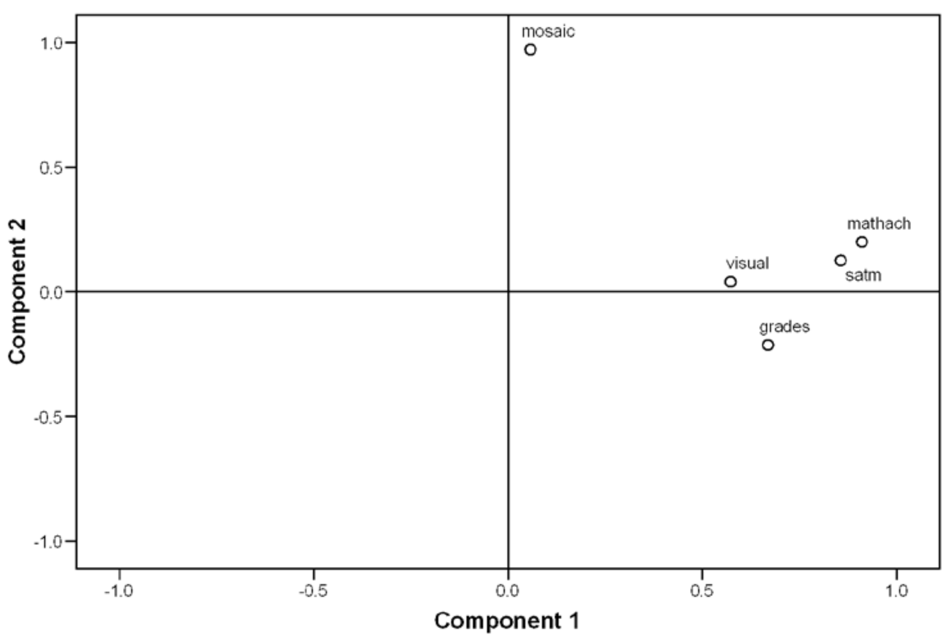
Even after rotation, mosaic is predicted by its own component, which does not have strong loadings on any of the other variables.

**Component Transformation Matrix**

| Component | 1 | 2 |
|---|---|---|
| 1 | .986 | .168 |
| 2 | -.168 | .986 |

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

**Component Plot in Rotated Space**




**Interpretation of Output 4.2 continued**

The **Rotated Component Matrix**, which contains all the loadings (even those < .3) for each component, is similar to the rotated factor matrix in Output 4.1. The **Component Plot in rotated Space** gives one a visual representation of the loadings plotted in a 2-dimensional space. The plot shows how closely related the items are to each other and to the two components. This plot of the component loadings shows that *math achievement*, *SATmath*, *grades in h.s.*, and *visualization test* all load highly and positively on the first component. *Mosaic* has a laoding near zero on the first component, but loads highly on the second.

Also, note that the default setting we used does not sort the variables in the **Rotated Component Matrix** by magnitude of loadings and does not suppress low loadings. Thus, you have to organize the table yourself; that is, *math achievement, scholastic aptitude test, grades in h.s.,* and *visualization*, in that order, have high Component 1 loadings, and *mosaic* is the only variable with a high loading for Component 2.

Researchers usually give names to rotated components in a fashion similar to that used in EFA; however, there is no assumption that this indicates a variable that underlies the measured items. Often, a researcher will aggregate (add or average) the items that define (have high loadings for) each component and use this composite variable in further research. Actually, the same thing is often done with EFA factor loadings; however, the implication of the latter is that this composite variable is an index of the underlying construct.

**Example of How to Write About Problem 4.2**

**Results**

Principal components analysis with varimax rotation was conducted to assess how five "achievement" variables clustered. These variables were *grades in h.s., math achievement*, *mosaic pattern test*, *visualization test,* and *scholastic aptitude test – math*. (The assumption of independent sampling was met. The assumptions of normality, linear relationships between pairs of variables, and the variables being correlated at a moderate level were checked and *mosaic pattern test* did not meet the assumptions, in that it was correlated at a low level with each of the other variables.) Two components were rotated, based on the eigenvalues over 1 criterion and the scree plot. After rotation, the first component accounted for 47% of the variance, and the second component accounted for 21% of the variance. Table 4.2 displays the items and component loadings for the rotated components, with loadings less than .30 omitted to improve clarity. Results suggest, in keeping with zero-order correlations, that *mosaic pattern test* scores are not substantially related to the other measures and should not be aggregated with them but that the other measures form a coherent component.



Table 4.2
*Component Loadings for the Rotated Components (N = 75)*

| Item | Component Loading | | |
|---|---|---|---|
| | 1 | 2 | Communality |
| Grades in high school | .67 | | .49 |
| Math achievement | .91 | | .87 |
| Visualization test | .57 | | .33 |
| Scholastic aptitude test – math | .86 | | .75 |
| Mosaic pattern test | | .97 | .95 |
| Eigenvalues | 2.38 | 1.01 | |
| % of variance | 46.81 | 20.97 | |

Note. Loadings < .25 are omitted.

## Interpretation Questions

4.1 Using Output 4.1: (a) Are the factors in Output 4.1 close to the conceptual composites (motivation, pleasure, competence) indicated in Chapter 1? (b) How might you name the three factors in Output 4.1? (c) Why did we use factor analysis rather than principal components analysis for this exercise?

4.2 Using Output 4.2: (a) Were any of the assumptions that were tested violated? Explain. (b) Describe the main aspects of the correlation matrix, the rotated component matrix, and the plot in Output 4.2.

4.3 What does the plot in Output 4.2 tell us about the relation of *mosaic* to the other variables and to component 1? How does this plot relate to the rotated component matrix?

## Extra SPSS Problems

4.1 Using the *judges.sav* data file, do exploratory factor analysis to see if the seven variables (the judges' countries) can be grouped into two categories: former communistic block countries (Russia, China, and Romania) and non-communist countries (U.S., South Korea, Italy, and France). What, if any, assumptions were violated?



4.2 Using the *satisf.sav* data file, see if the six satisfaction scales can be reduced to a smaller number of variables.

4.3 Using the *love.sav* data file, see if the four love questions can be grouped into one category. What, if any, assumptions were violated?

4.4 Using the *1991 U.S. General Social Survey.sav* data file, do exploratory factor analysis to see if the health variables (hlth1 to hlth9) and the work variables (work1 to work9) fall into two categories: health and work. Were any assumptions violated?