1. Reasonably large communalities for all the measured variables (e.g., $h^2 > .10$ for all variables; however, this is an arbitrary standard).

2. A reasonably high proportion of the total variance $p$ (where $p$ corresponds to the number of z scores on measured variables) should be explained by the retained components.

3. The retained component or components should reproduce all the correlations between measured variables reasonably well; for example, we might require the residuals for each predicted correlation to be less than .05 in absolute value (however, this is an arbitrary criterion).

4. For all the preceding points to be true, we will also expect to see that many of the measured variables have reasonably large loadings (e.g., loadings greater than .30 in absolute magnitude) on at least one of the retained components.

Note that SPSS does not provide statistical significance tests for any of the estimated parameters (such as loadings), nor does it provide confidence intervals. Judgments about the adequacy of a one- or two-component model are not made based on statistical significance tests, but by making arbitrary judgments whether the model that is limited to just one or two components does an adequate job of reproducing the communalities (the variance in each individual measured $x$ variable) and the correlations among variables (in the $R$ correlation matrix).

18.11 ♦ Principal Components Versus Principal Axis Factoring

As noted earlier, the most widely used method in factor analysis is the PAF method. In practice, PC and PAF are based on slightly different versions of the $R$ correlation matrix (which includes the entire set of correlations among measured $X$ variables). PC analyzes and reproduces a version of the $R$ matrix that has 1s in the diagonal. Each value of 1.00 corresponds to the total variance of one standardized measured variable, and the initial set of $p$ components must have sums of squared correlations for each variable across all components that sum to 1.00. This is interpreted as evidence that a $p$-component PC model can reproduce all the variances of each standardized measured variable. In contrast, in PAF, we replace the 1s in the diagonal of the correlation matrix $R$ with estimates of communality that represent the proportion of variance in each measured $X$ variable that is predictable from or shared with other $X$ variables in the dataset. Many programs use multiple regression to obtain an initial communality estimate for each variable; for example, an initial estimate of the communality of $X_1$ could be the $R^2$ for a regression that predicts $X_1$ from $X_2, X_3, \ldots, X_p$. However, after the first step in the analysis, communalities are defined as sums of squared factor loadings, and the estimation of communalities with a set of factor loadings that can do a reasonably good job of reproducing the entire $R$ correlation matrix typically requires multiple iterations in PAF.

For some datasets, PC and PAF may yield similar results about the number and nature of components or factors. The conceptual approach involved in PAF treats each $X$ variable as a measurement that, to some extent, may provide information about the same small set of factors or latent variables as other measured $X$ variables, but at the same time, each $X$ variable may also be influenced by unique sources of error. In PAF, the analysis of data
structure focused on shared variance and not on sources of error that are unique to individual measurements. For many applications of factor analysis in the behavioral and social sciences, the conceptual approach involved in PAF (i.e., trying to understand the shared variance in a set of X measurements through a small set of latent variables called factors) may be more convenient than the mathematically simpler PC approach (which sets out to represent all of the variance in the X variables through a small set of components). Partly because of the conceptual basis (PAF models only the shared variance in a set of X measurements) and partly because it is more familiar to most readers, PAF is more commonly reported in social and behavioral science research reports than PC. The next two empirical examples illustrate application of PAF to nine items for the data in Table 18.1.

18.12 PAF of Nine Items, Two Factors Retained, No Rotation

Analysis 3 differs from the first two analyses in several ways. First, Analysis 3 includes nine variables (rather than the set of three variables used in earlier analyses). Second, PAF is used as the method of extraction in Analysis 3. Finally, in Analysis 3, two factors were retained based on the sizes of their eigenvalues.

Figure 18.20 shows the initial Factor Analysis dialog window for Analysis 3, with nine self-rated characteristics included as variables (e.g., nurturant, affectionate, . . . , aggressive). None of the additional descriptive statistics (such as reproduced correlations) that were requested in Analyses 1 and 2 were also requested for Analysis 3. To perform the extraction as a PAF, the Extraction button was used to open the Factor Analysis: Extraction window that appears in Figure 18.21. From the pull-down menu near the top of this window, Principal axis factoring was selected as the method of extraction. The decision about the number of factors to retain was indicated by clicking the radio button for Eigenvalues over; the default minimum size generally used to decide which factors to retain is 1, and that was not changed. Under the Display heading, a box was checked to request a scree plot; a scree plot summarizes information about the magnitudes of the eigenvalues across all the factors, and sometimes the scree plot is examined when making decisions about the number of factors to retain. The Rotation button was used to open the Factor Analysis: Rotation window that appears in Figure 18.22. The default (indicated by a radio button) is no rotation (“None,” under the heading for method), and that was not changed. The box for Loading plots under the heading Display was checked to request a plot of the factor loadings for all nine variables on the two retained (but not rotated) factors. The Options button opened up the window for Factor Analysis: Options in Figure 18.23; in this box, under the heading for Coefficient Display Format, a check was placed in the check box for the Sorted by size option. This does not change any computed results but it arranges the summary table of factor loadings so that variables that have large loadings on the same factor are grouped together, and this improves the readability of the output, particularly when the number of variables included in the analysis is large. The syntax for Analysis 3 that resulted from the menu selections just discussed appears in Figure 18.24. The results of this PAF analysis of nine variables appear in Figures 18.25 through 18.31.

The information that is reported in the summary tables for Analysis 3 includes loadings and sums of squared loadings. The complete $9 \times 9$ matrix that contains the correlations of all nine measured variables with all nine factors does not appear as part of the SPSS output; the tables that do appear in the printout summarize information that is