**Example 13.6.** The speaking rate of four voices was artificially manipulated by means of a rate changer without altering the pitch (Brown, Strong, and Rencher 1973). There were five rates for each voice:

$$
\begin{aligned}
\text{FF} &= 45\% \text{ faster},\\
\text{F} &= 25\% \text{ faster},\\
\text{N} &= \text{normal rate},\\
\text{S} &= 22\% \text{ slower},\\
\text{SS} &= 42\% \text{ slower}.
\end{aligned}
$$

The resulting 20 voices were played to 30 judges, who rated them on 15 paired-opposite adjectives (variables) with a 14-point scale between poles. The following adjectives were used: intelligent, ambitious, polite, active, confident, happy, just, likeable, kind, sincere, dependable, religious, good-looking, sociable, and strong. The results were averaged over the 30 judges to produce 20 observation vectors of 15 variables each. The averaging produced very reliable data, so that even though there were only 20 observations on 15 variables, the factor analysis model fit very well. The correlation matrix is as follows:

$$
\mathbf{R} = \begin{pmatrix}
1.00 & .90 & -.17 & .88 & .92 & .88 & .15 & .39 & -.02 & -.16 & .52 & -.15 & -.79 & -.78 & .73 \\
.90 & 1.00 & -.46 & .93 & .87 & .79 & -.16 & .10 & -.35 & -.42 & .25 & -.40 & .68 & -.60 & .62 \\
-.17 & -.46 & 1.00 & -.56 & -.13 & .07 & .85 & .75 & .88 & .91 & .68 & .88 & .21 & .31 & .25 \\
.88 & .93 & -.56 & 1.00 & .85 & .73 & -.25 & -.02 & -.45 & -.57 & .10 & -.53 & .58 & .84 & .50 \\
.92 & .87 & -.13 & .85 & 1.00 & .91 & .20 & .39 & -.09 & -.16 & .49 & -.10 & .85 & .80 & .81 \\
.88 & .79 & .07 & .73 & .91 & 1.00 & .27 & .53 & .12 & .06 & .66 & .08 & .90 & .85 & .78 \\
.15 & -.16 & .85 & -.25 & .20 & .27 & 1.00 & .85 & .81 & .79 & .79 & .81 & .43 & .54 & .53 \\
.39 & .10 & .75 & -.02 & .39 & .53 & .85 & 1.00 & .84 & .79 & .93 & .77 & .71 & .69 & .76 \\
-.02 & -.35 & .88 & -.45 & -.09 & .12 & .81 & .84 & 1.00 & .91 & .76 & .85 & .28 & .36 & .35 \\
-.16 & -.42 & .91 & -.57 & -.16 & .06 & .79 & .79 & .91 & 1.00 & .72 & .96 & .26 & .28 & .29 \\
.52 & .25 & .67 & .10 & .49 & .66 & .79 & .93 & .76 & .72 & 1.00 & .72 & .75 & .77 & .78 \\
-.15 & -.40 & .88 & -.53 & -.10 & .08 & .81 & .77 & .85 & .96 & .72 & 1.00 & .33 & .32 & .34 \\
.79 & .68 & .21 & .58 & .85 & .90 & .43 & .71 & .28 & .26 & .75 & .33 & 1.00 & .86 & .92 \\
.78 & .60 & .31 & .54 & .80 & .85 & .54 & .69 & .36 & .28 & .77 & .32 & .86 & 1.00 & .82 \\
.73 & .62 & .25 & .50 & .81 & .78 & .53 & .76 & .35 & .29 & .78 & .34 & .92 & .82 & 1.00
\end{pmatrix}
$$

The eigenvalues of $\mathbf{R}$ are 7.91, 5.85, .31, .26, ... , .002, with the scree plot in Figure 13.5. Clearly, by any criterion for choosing $m$, there are two factors.

All four major methods of factor extraction discussed in Section 13.3 produced nearly identical results (after rotation). We give the initial and rotated loadings obtained from the principal component method in Table 13.10.

The two rotated factors were labeled *competence* and *benevolence*. The same two factors emerged consistently in similar studies with different voices and different judges.

The two groupings of variables can also be seen in the correlation matrix. For example, in the first row, the large correlations correspond to the boldface rotated

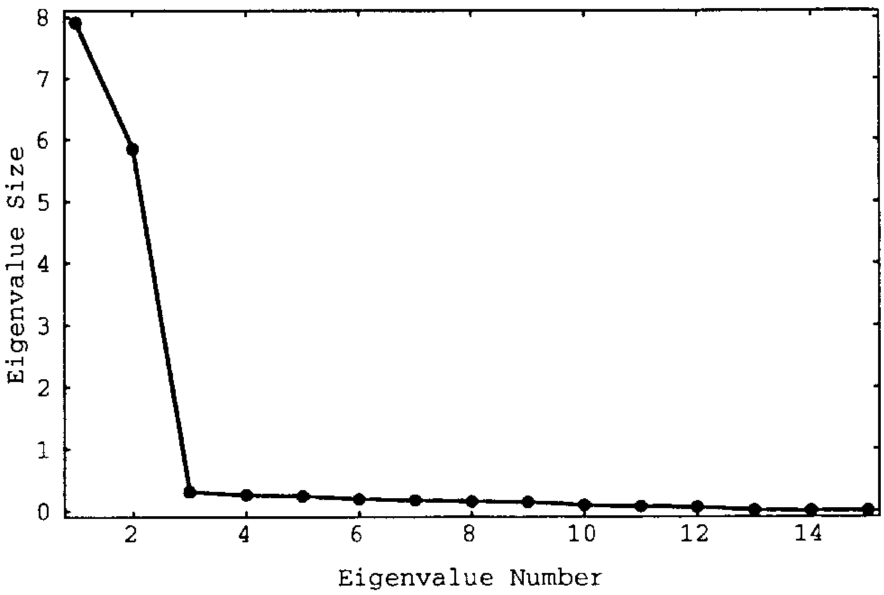


**Figure 13.5.** Scree graph for voice data.

**Table 13.10. Initial and Varimax Rotated Loadings for the Voice Data**

| | Initial Loadings | | Rotated Loadings | | |
|---|---|---|---|---|---|
| Variable | $f_1$ | $f_2$ | $f_1$ | $f_2$ | Communalities |
| Intelligent | .71 | −.65 | **.96** | −.06 | .93 |
| Ambitious | .48 | −.84 | **.90** | −.36 | .94 |
| Polite | .50 | .81 | −.12 | **.95** | .92 |
| Active | .37 | −.91 | **.86** | −.48 | .97 |
| Confident | .73 | −.64 | **.97** | −.04 | .95 |
| Happy | .83 | −.47 | **.94** | .15 | .91 |
| Just | .71 | .58 | .20 | **.89** | .84 |
| Likeable | .89 | .39 | .45 | **.87** | .95 |
| Kind | .58 | .75 | −.02 | **.95** | .89 |
| Sincere | .52 | .82 | −.11 | **.97** | .95 |
| Dependable | .93 | .27 | .56 | **.79** | .94 |
| Religious | .55 | .79 | −.07 | **.96** | .92 |
| Good looking | .91 | −.29 | **.89** | .35 | .91 |
| Sociable | .91 | −.22 | **.84** | .40 | .87 |
| Strong | .91 | −.21 | **.84** | .41 | .86 |
| Variance accounted for | 7.91 | 5.85 | 7.11 | 6.65 | 13.76 |
| Proportion of total variance | .53 | .39 | .47 | .44 | .92 |
| Cumulative proportion | .53 | .92 | .47 | .92 | .92 |

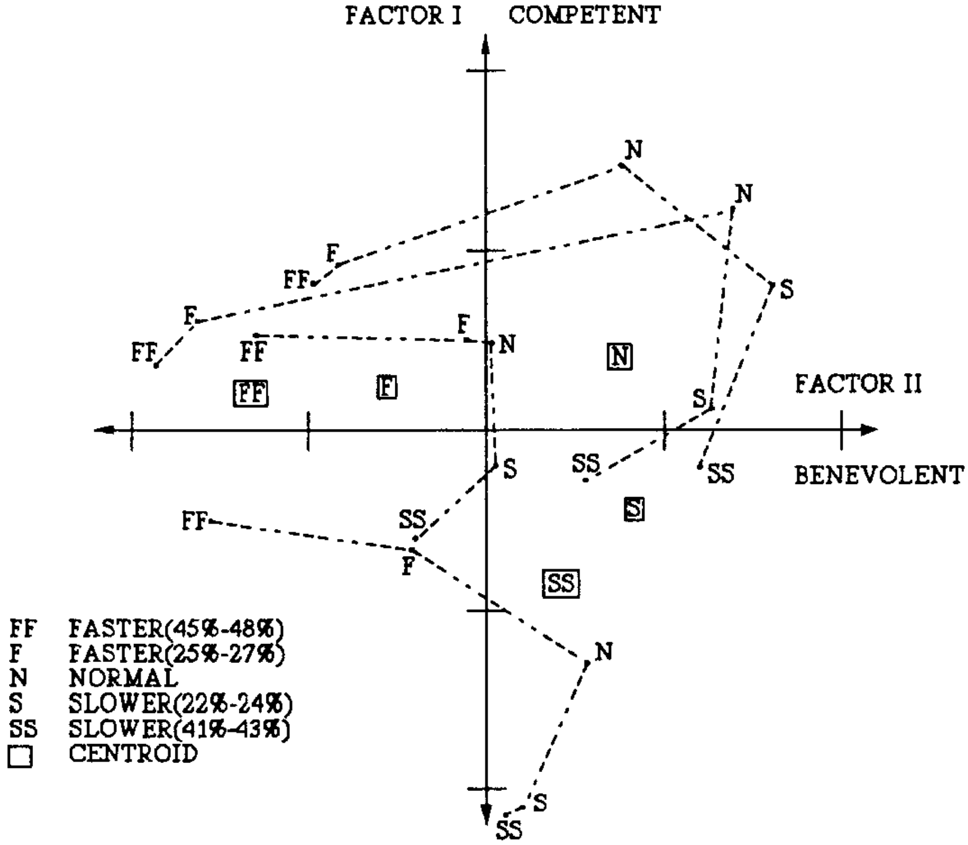


**Figure 13.6.** Factor scores of adjective rating of voices with five levels of manipulated rate.

loadings for $f_1$, whereas in the third row, the large correlations correspond to the boldface rotated loadings for $f_2$.

The factor scores were of primary interest in this study. The goal was to ascertain the effect of the rate manipulations on the two factors, that is, to determine the perceived change in competence and benevolence when the speaking rate is increased or decreased.

The two factor scores were obtained for each of the 20 voices; these are plotted in Figure 13.6, where a consistent effect of the manipulation of speaking rate on all four voices can clearly be seen. Decreasing the speaking rate causes the speaker to be rated less competent; increasing the rate causes the speaker to be rated less benevolent. The mean vectors (centroids) are also given in Figure 13.6 for the four speakers. □

## 13.7 VALIDITY OF THE FACTOR ANALYSIS MODEL

For many statisticians, factor analysis is controversial and does not belong in a toolkit of legitimate multivariate techniques. The reasons for this mistrust include the following: the difficulty in choosing $m$, the many methods of extracting factors, the many rotation techniques, and the subjectivity in interpretation. Some statisticians also criticize factor analysis because of the indeterminacy of the factor loading matrix $\mathbf{\Lambda}$ or $\hat{\mathbf{\Lambda}}$, first noted in Section 13.2.2. However, it is the ability to rotate that gives factor analysis its utility, if not its charm.

The basic question is whether the factors really exist. The model (13.11) for the covariance matrix is $\mathbf{\Sigma} = \mathbf{\Lambda}\mathbf{\Lambda}' + \mathbf{\Psi}$ or $\mathbf{\Sigma} - \mathbf{\Psi} = \mathbf{\Lambda}\mathbf{\Lambda}'$, where $\mathbf{\Lambda}\mathbf{\Lambda}'$ is of rank $m$. Many populations have covariance matrices that do not approach this pattern unless $m$ is large. Thus the model will not fit data from such a population when we try to impose a small value of $m$. On the other hand, for a population in which $\mathbf{\Sigma}$ is reasonably close to $\mathbf{\Lambda}\mathbf{\Lambda}' + \mathbf{\Psi}$ for small $m$, the sampling procedure leading to $\mathbf{S}$ may obscure this pattern. The researcher may believe there are underlying factors but has difficulty collecting data that will reveal them. In many cases, the basic problem is that $\mathbf{S}$ (or $\mathbf{R}$) contains both structure and error, and the methods of factor analysis cannot separate the two.

A statistical consultant in a university setting or elsewhere all too often sees the following scenario. A researcher designs a long questionnaire, with answers to be given in, say, a five-point semantic differential scale or Likert scale. The respondents, who vary in attitude from uninterested to resentful, hurriedly mark answers that in many cases are not even good subjective responses to the questions. Then the researcher submits the results to a handy factor analysis program. Being disappointed in the results, he or she appeals to a statistician for help. They attempt to improve the results by trying different methods of extraction, different rotations, different values of $m$, and so on. But it is all to no avail. The scree plot looks more like the foothills than a steep cliff with gently sloping debris at the bottom. There is no clear value of $m$. They have to extract 10 or 12 factors to account for, say, 60% of the variance, and interpretation of this large number of factors is hopeless. If a few underlying dimensions exist, they are totally obscured by both systematic and random errors in marking the questionnaire. A factor analysis model simply does not fit such a data set, unless a large value of $m$ is used, which gives useless results.

It is not necessarily the "discreteness" of the data that causes the problem, but the "noisiness" of the data. The specified variables are not measured accurately. In some cases, discrete variables yield satisfactory results, such as in Examples 13.3.1, 13.3.2, 13.5.2a, and 13.5.2b(a), where a 12-year-old girl, responding carefully to a semantic differential scale, produced data leading to an unambiguous factor analysis. On the other hand, continuous variables do not guarantee good results [see Example 13.7(a)].

In cases in which some factors are found that provide a satisfactory fit to the data, we should still be tentative in interpretation until we can independently establish the existence of the factors. If the same factors emerge in repeated sampling from the same population or a similar one, then we can have confidence that application of the model has uncovered some real factors. Thus it is good practice to repeat the experiment to check the stability of the factors. If the data set is large enough, it could be split in half and a factor analysis performed on each half. The two solutions could be compared with each other and with the solution for the complete set.

If there is replication in the data set, it may be helpful to average over the replications. This was done to great advantage in Example 13.6, where several judges rated the same voices. Averaging over the judges produced variables that apparently possessed very low noise. Similar experimentation with different judges always pro-