João Pedro Gonçalves Pacheco

# Exploratory factor analysis

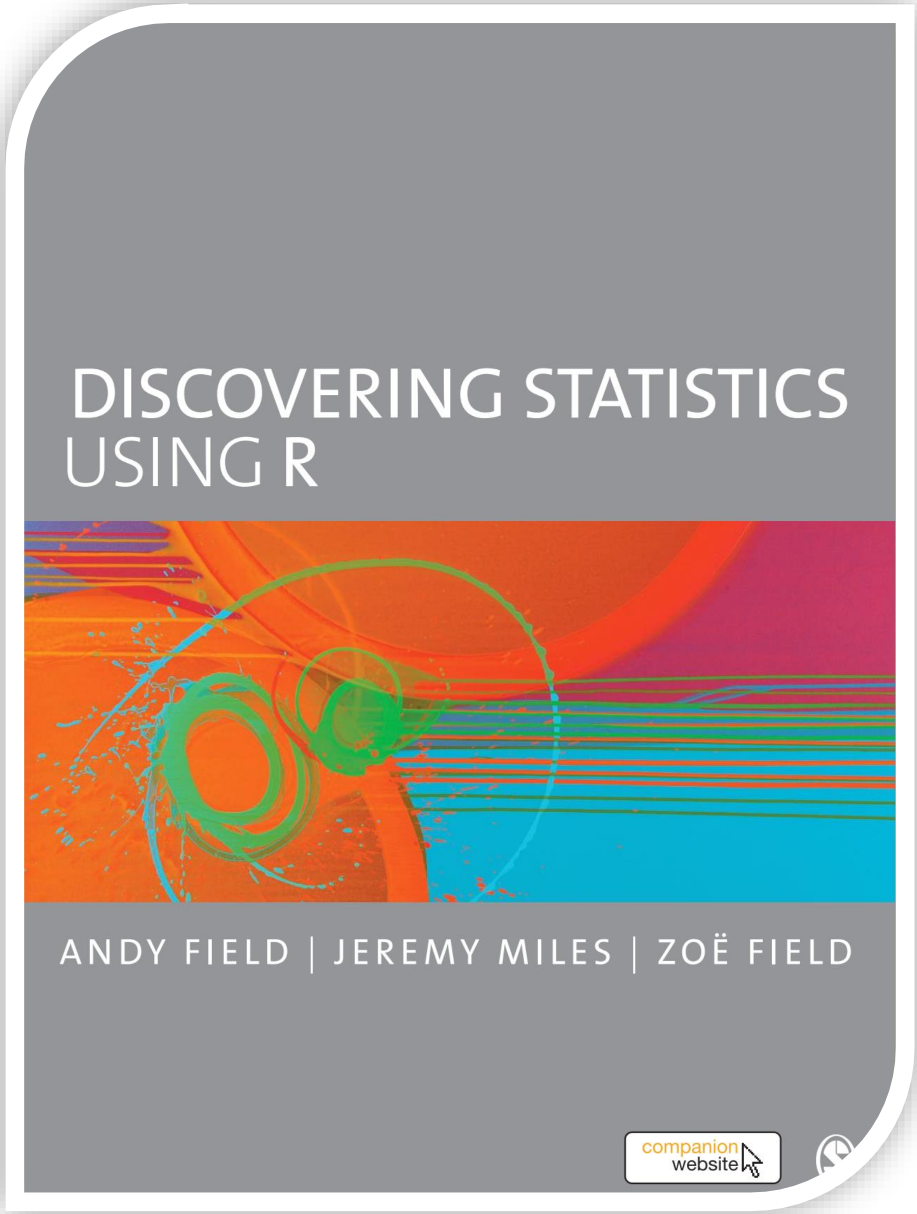
DISCOVERING STATISTICS
USING R
ANDY FIELD | JEREMY MILES | ZOË FIELD
companion website

**FIGURE 17.2**
An *R*-matrix

| | Talk 1 | Social Skills | Interest | Talk 2 | Selfish | Liar |
|---|---|---|---|---|---|---|
| Talk 1 | 1.000 | | | | | |
| Social Skills | .772 | 1.000 | | | | |
| Interest | .646 | .879 | 1.000 | | | |
| Talk 2 | .074 | −.120 | .054 | 1.000 | | |
| Selfish | −.131 | .031 | −.101 | .441 | 1.000 | |
| Liar | .068 | .012 | .110 | .361 | .277 | 1.000 |

Factor 1

Factor 2

The coordinate of a variable along a classification axis is known as a **factor loading**.

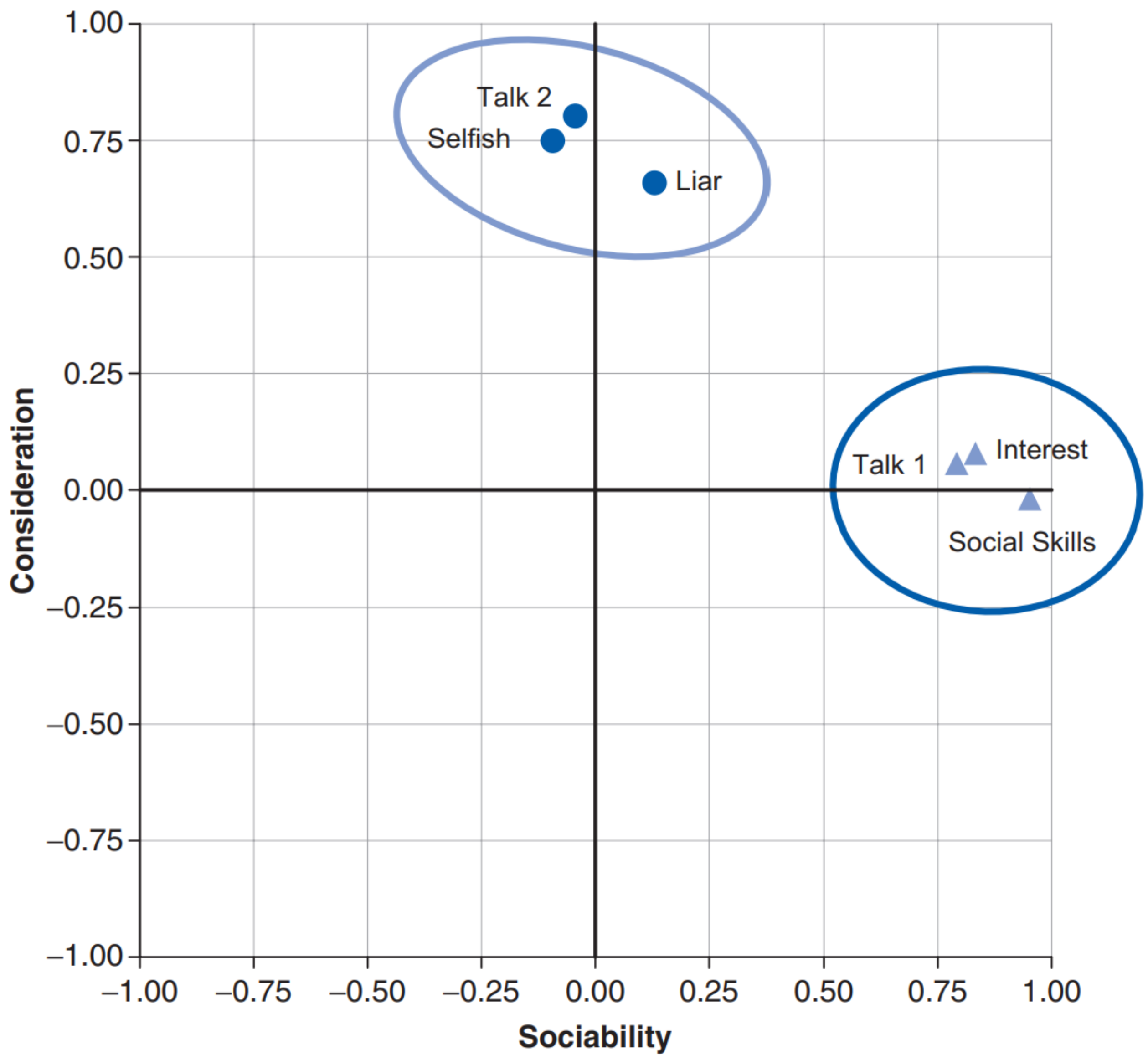


**FIGURE 17.3**
Example of a factor plot

The following equation reminds us of the equation describing a linear model and then applies this to the scenario of describing a factor:

$$Y_i = b_1X_{1i} + b_2X_{2i} + \ldots + b_nX_{ni} + \varepsilon_i$$

$$\text{Factor}_i = b_1 \text{ Variable}_{1i} + b_2 \text{ Variable}_{2i} + \ldots + b_n \text{ Variable}_{ni} + \varepsilon_i \tag{17.1}$$

# factor scores

$$
\begin{aligned}
\text{Sociability} &= 0.87\text{Talk1} + 0.96\text{Social Skills} + 0.92\text{Interest} \\
&\quad + 0.00\text{Talk2} - 0.10\text{Selfish} + 0.09\text{Liar} \\
&= (0.87 \times 4) + (0.96 \times 9) + (0.92 \times 8) + (0.00 \times 6) \\
&\quad - (0.10 \times 8) + (0.09 \times 6) \\
&= 19.22 \\
\text{Consideration} &= 0.01\text{Talk1} - 0.03\text{Social Skills} + 0.04\text{Interest} \\
&\quad + 0.82\text{Talk2} + 0.75\text{Selfish} + 0.70\text{Liar} \\
&= (0.01 \times 4) - (0.03 \times 9) + (0.04 \times 8) + (0.82 \times 6) \\
&\quad + (0.75 \times 8) + (0.70 \times 6) \\
&= 15.21
\end{aligned}
\qquad (17.4)
$$

# communality

The total variance for a particular variable will have two components: some of it will be shared with other variables or measures (**common variance**) and some of it will be specific to that measure (**unique variance**).

We tend to use the term unique variance to refer to variance that can be reliably attributed to only one measure. However, there is also variance that is specific to one measure but not reliably so; this variance is called error or **random variance**.

The proportion of common variance present in a variable is known as the **communality**.

# factor extraction: eigenvalues and scree plot

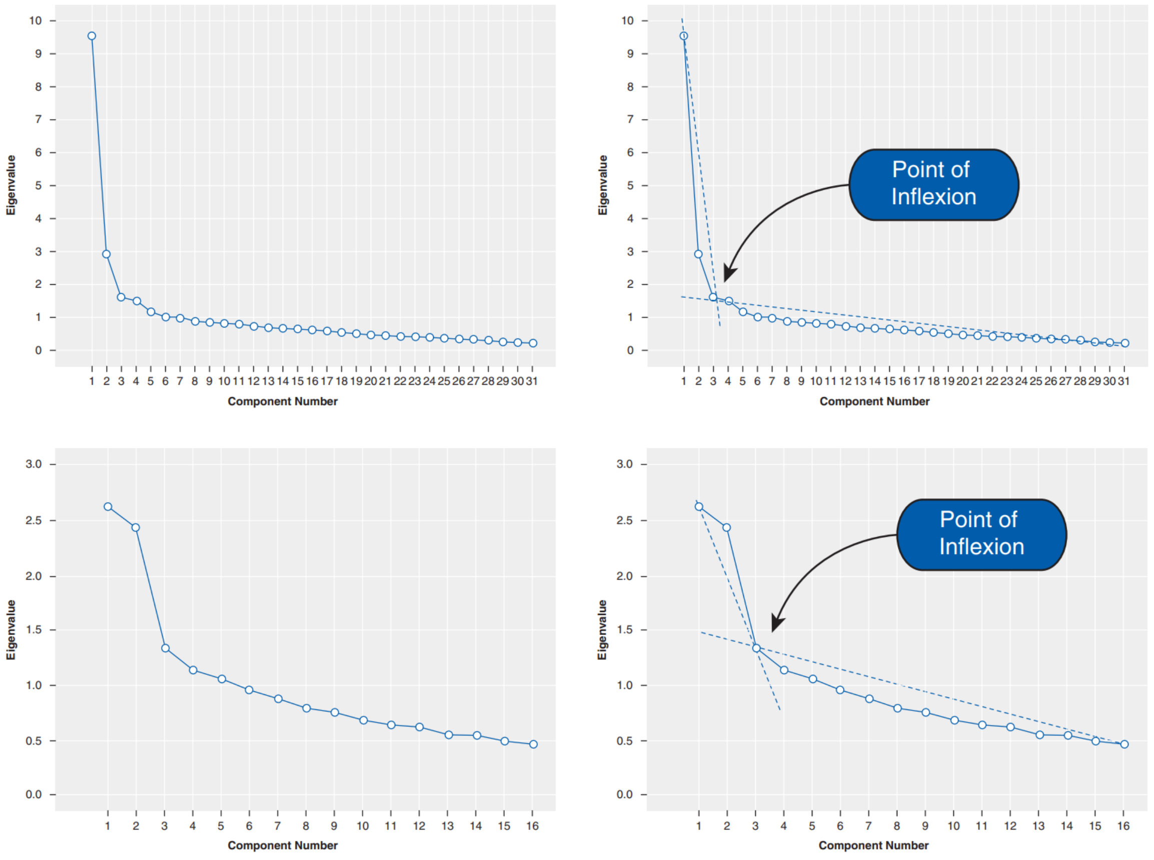


**FIGURE 17.4**
Examples of scree plots for data that probably have two underlying factors

# rotation:
# oblique and orthogonal

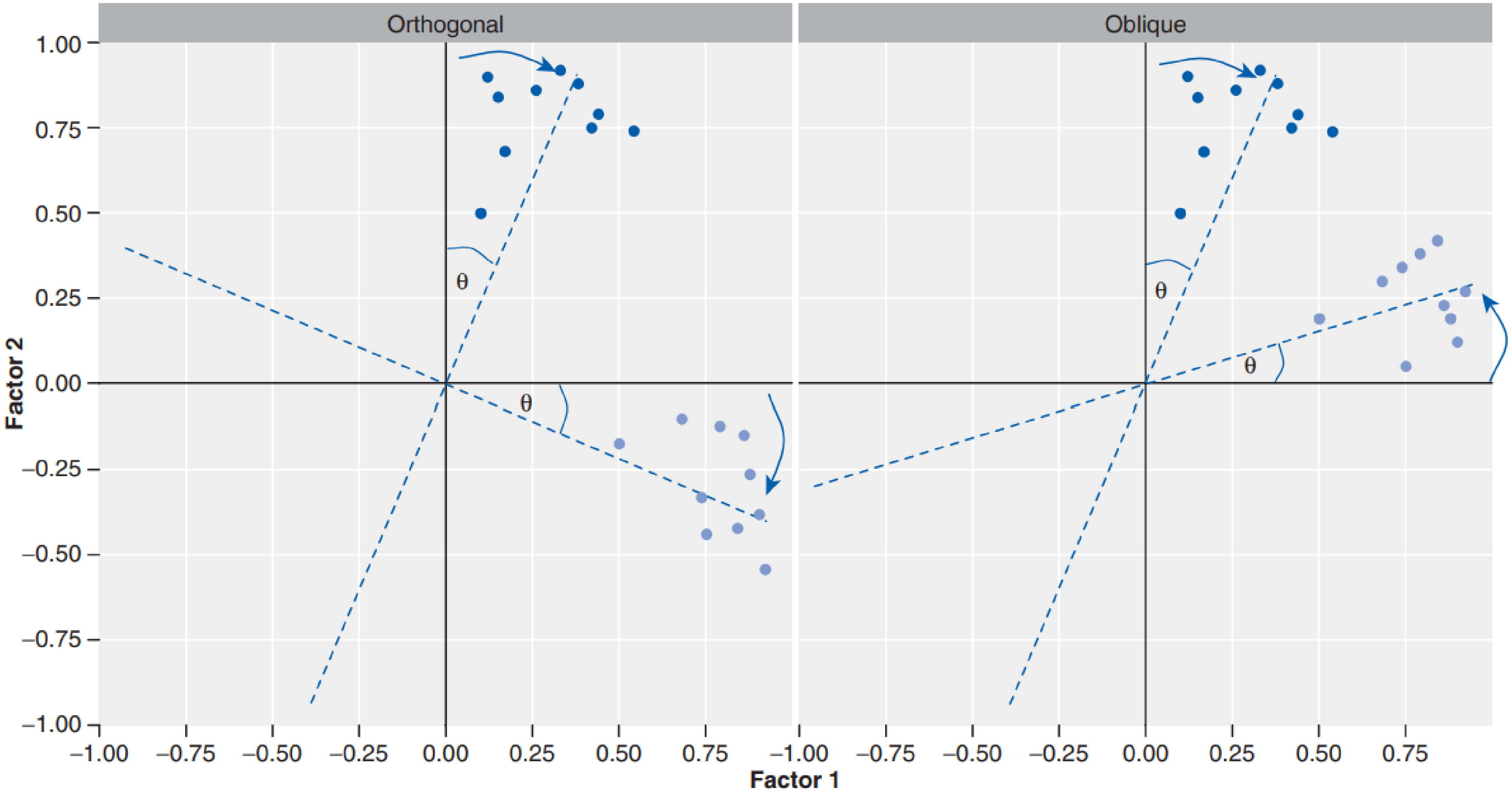


**FIGURE 17.5** Schematic representations of factor rotation. The left graph displays orthogonal rotation whereas the right graph displays oblique rotation (see text for more details). $\theta$ is the angle through which the axes are rotated

# factorable?
# sample size

**Kaiser-Meyer-Olkin (KMO)** measure of sampling adequacy
The KMO can be calculated for individual and multiple variables and
The KMO statistic varies between 0 and 1.
A value of 0 indicates that the sum of partial correlations is large relative to the sum of correlations, indicating, factor analysis is likely to be inappropriate.

Kaiser (1974) recommends accepting values greater than .5 as barely acceptable (values below this should lead you to either collect more data or rethink which variables to include).
Furthermore, values between .5 and .7 are mediocre, values between .7 and .8 are good, values between .8 and .9 are great and values above .9 are superb

# factorable?
# Correlations (too low)

If the variables in our correlation matrix did not correlate at all, then our correlation matrix would be an **identity matrix**.

**Bartlett's test** examines whether the population correlation matrix resembles an identity matrix.

A non-significant Bartlett's test is cause for concern.

# factorable?
# Correlations (too high)

If correlations are too high, we might have **multicollinearity**.

Multicollinearity can be detected by looking at the **determinant** of the R-matrix, denoted |R|.

One simple heuristic is that the determinant of the R-matrix should be greater than 0.00001.

# factorable?
# distribution

If continuous data -> variables should have normal distributions

If Likert scales -> **tetrachoric**

If dichotomous scales -> **polychoric**

João Pedro Gonçalves Pacheco

# Exploratory factor analysis