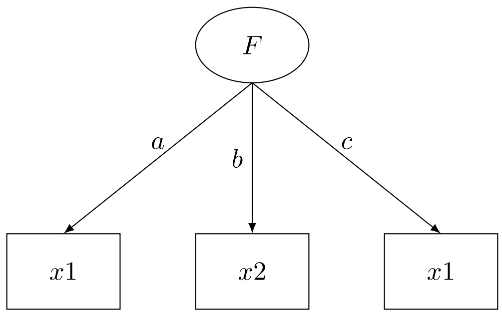
Deriving Trace Plots

The Expected Slope Between Two Indicators, Given the Latent Variable

Suppose we have the factor analysis model shown below.



Assume both indicators and the latent variable are standard normal variables. In this situation, we define the observed variables as follows:

$$
\begin{aligned}
x_1 &= aF + e_1 \\
x_2 &= bF + e_2 \\
x_3 &= cF + e_3
\end{aligned}
$$

These can be re-expressed in terms of F as follows:

$$
\begin{aligned}
F &= \frac{x_1 - e_1}{a} \\
F &= \frac{x_2 - e_2}{b} \\
F &= \frac{x_3 - e_3}{c}
\end{aligned}
$$

Suppose we wish to plot the fitted line between x_1 and x_2 implied by the latent variable model. To do so, we can re-express x_2 in terms of x_1:

$$
\begin{aligned}
x_2 &= bF + e_2 \\
&= b(\frac{x_1 - e_1}{a}) + e_2 \qquad (1)
\end{aligned}
$$

Computing the expectation, we get

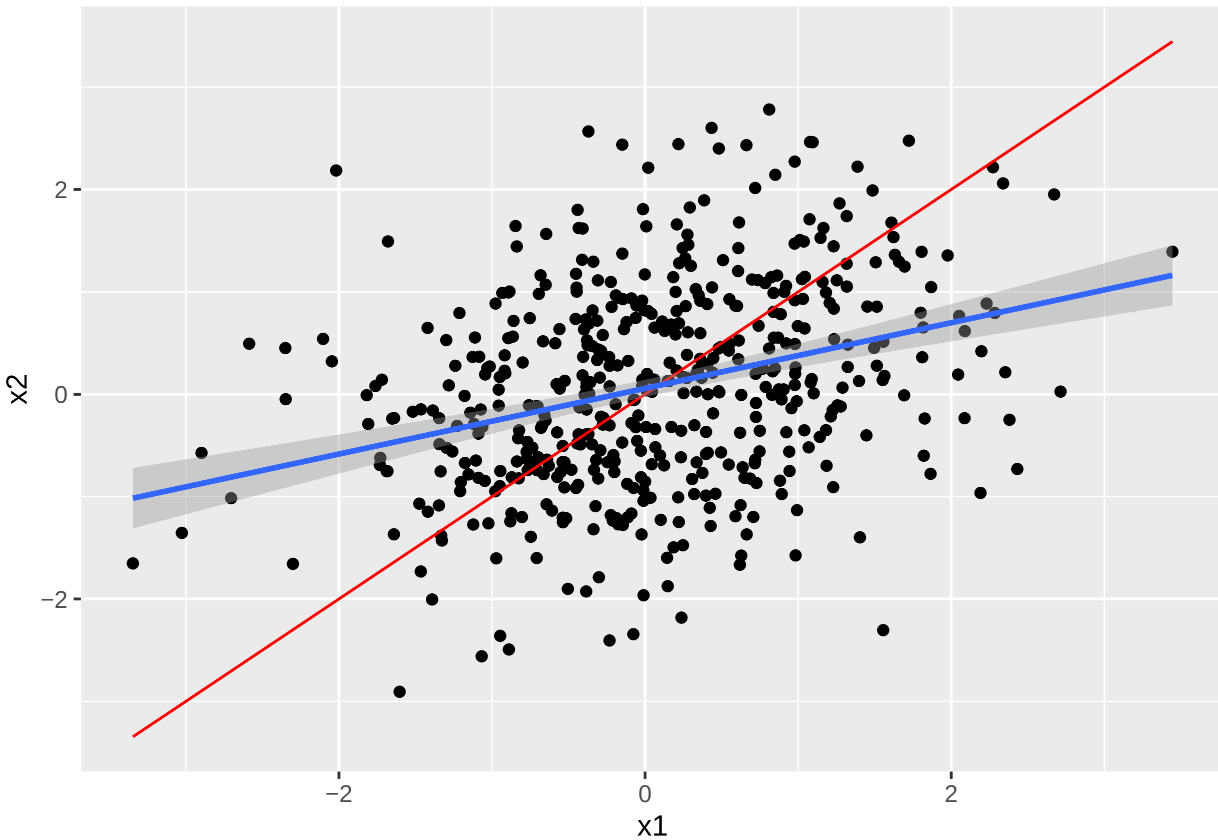


Figure 1: Scatterplot of the x1/x2 relationship. The red line shows the implied fit between the two variables under the assumption of perfect reliability. The blue line shows the actual fitted relationship.

$$\begin{aligned} E[x_2|x_1] &= E\left[b(\frac{x_1 - e_1}{a}) + e_2\right] \\ &= \frac{b}{a}E[(x_1 - e_1)] + E[e_2] \\ &= \frac{b}{a}E[x_1] - E[e_1] + E[e_2] \\ &= \frac{b}{a}E[x_1] \end{aligned} \tag{2}$$

Also, since these variables are standardized, Equation 2 is both the slope of the line predicting x_2 from x_1, as well as the correlation coefficient, ρ.

Recall that latent variable models explicitly model the unreliability in indicators. As such, 2 is the "corrected" estimate of the correlation between x_1 and x_2. As such, the regression line generated from Equation 2 will visually overestimate the strength of the relationship, as demonstrated below:

For example, of b and a are both 0.5, Equation 2 suggests the slope is one (indicating a perfect correlation between these standardized variances). This is clearly not the case. As such, the slope needs to be "uncorrected" for unreliability.

Recall the standard correction for unreliability

$$\rho = \frac{r}{\sqrt{\rho_{xx}\rho_{yy}}} \tag{3}$$

To "uncorrect" this estimate, we simply solve for r

$$r = \rho\sqrt{\rho_{xx}\rho_{yy}} \tag{4}$$

We can then multiply Equation 4 by Equation 2:

$$\begin{aligned} E[x_2|x_1] &= \sqrt{\rho_{xx}\rho_{yy}}\frac{b}{a}E[x_1] \\ &= \sqrt{a^2b^2}\left(\frac{b}{a}\right)E[x_1] \\ &= a*b\left(\frac{b}{a}\right)E[x_1] \\ &= b^2E[x_1] \end{aligned} \tag{5}$$

Using Equation 5, we now get a line that closely approximates the actual relationship between the two variables: