SEM 2: Structural Equation Modeling

Week 0 - Expectation algebra and covariance algebra

Sacha Epskamp

SEM 2 2019

Course Overview

- Tuesdays: Lecture
- Thursdays: Unstructured practicals
- Three assignments
  - First two 20% of final grade, last 10% of final grade
- Final project
  - Presentations and report, 50% of final grade
  - Final grade distribution depends on project

Schedule

Week 1 – Introduction to Structural Equation Modeling

- Tuesday May 7 – Lecture
- Thursday May 9 – Practical

Week 2 – Causality and equivalent models

- Tuesday May 14 – Lecture + deadline assignment 1
- Thursday May 16 - Practical

Week 3 – Time-series analysis and network models

- Tuesday May 21 – Lecture + deadline assignment 2
- Thursday May 23 - Free time to work on final project

Week 4 – Wrap-up and presentations

- Tuesday May 28 –Presentations + deadline assignment 3 (end of the day)
- Friday May 31 - deadline final project report

Individual Assignments

Each week, the assignment will be made available 15:00 on Tuesday the latest, and will be due 13:00 the next Tuesday. Each assignment will contribute to 20% or 10% (last week) of your grade.

- Work on the assignments alone.
- Hand in a PDF file and an .R file (in case R was used). If you use Jasp, hand in the Jasp object as well as a screenshot of the options used.
- Make sure your PDF report is as standalone readable as possible. E.g., if you are asked to report a matrix, then report it in the PDF and not just say "look at .R file".
- Assignments are due before 13:00. If you do not hand in an assignment before 13:00, you will get a 1.
- If you encounter any problems, or have any feedback, please let me know before the deadline, as then I can take it into account or help you.

Final Project

Several options:

1. Write a manual for *semPlot* or *psychonetrics* — 1/2; S
2. Extend and/or validate the *psychonetrics* package — 1/2; S
3. Implement support for new packages in *semPlot* — 1/2; S
4. Research an area or a topic of SEM in more detail and teach fellow students about it — 2; P
5. Perform a SEM analysis on your own data and write a report — 1; (P)
6. Any other idea somewhat related to SEM — 1/2; (P)

Legend: 1: one person project, 2: two-person project, P: includes presentation, (P): might include presentation; S: your project is intended to be shared online

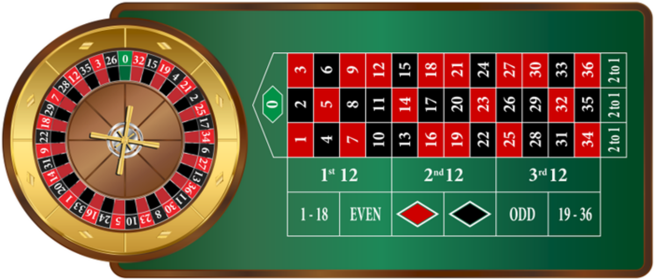
The expected value is defined as:

$$\mathcal{E}(x) = \sum_{k=1}^{n_k} x_k \Pr(x = x_k) \quad \text{if } x \text{ is categorical with } n_k \text{ possible outcomes}$$
$$\mathcal{E}(x) = \int_{\mathbb{R}} x f(x)\, dx \quad \text{if } x \text{ is continuous}$$

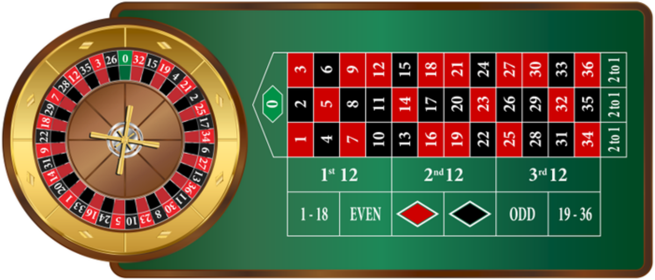
With the following rules:

$$\begin{aligned}
\mathcal{E}(\alpha) &= \alpha \\
\mathcal{E}(\alpha x) &= \alpha \mathcal{E}(x) \\
\mathcal{E}(\alpha x + \beta) &= \alpha \mathcal{E}(x) + \beta \\
\mathcal{E}(x + y) &= \mathcal{E}(x) + \mathcal{E}(y) \\
\mathcal{E}(\alpha x + \beta y) &= \alpha \mathcal{E}(x) + \beta \mathcal{E}(y)
\end{aligned}$$

Where α and β are constants (parameter) and x and y are random variables.



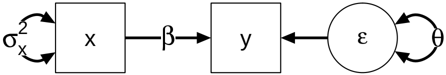
If you bet on a number and the ball falls on that number, you win 35 times your bet! Should you play this game?



If you bet on a number and the ball falls on that number, you win 35 times your bet! Should you play this game?
Suppose we bet 1 Euro, the expected value is:

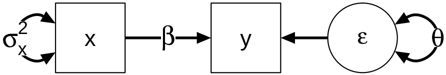
$$\frac{1}{37} \times 35 - \frac{36}{37} \times 1 = -0.027$$

So you expect to lose 2.7 cent on average for every bet.



$$y_i = \tau + \beta x_i + \varepsilon_i$$
$$x \sim N(\mu_x, \sigma_x)$$
$$\varepsilon \sim N(0, \theta)$$

We wish to express the variances and covariances related to endogenous (y) variables in terms of parameters and variances and covariances of the exogenous (x) variables.



$$
\begin{aligned}
y_i =& \tau + \beta x_i + \varepsilon_i \\
x &\sim N(\mu_x, \sigma_x) \\
\varepsilon &\sim N(0, \theta)
\end{aligned}
$$

We wish to express the variances and covariances related to endogenous (y) variables in terms of parameters and variances and covariances of the exogenous (x) variables.
We can now derive:

$$
\begin{aligned}
\mathcal{E}(y) &= \mathcal{E}(\tau + \beta x + \varepsilon) \\
&= \mathcal{E}(\tau) + \mathcal{E}(\beta x) + \mathcal{E}(\varepsilon) \\
&= \tau + \beta \mathcal{E}(x) \\
&= \tau + \beta \mu_x
\end{aligned}
$$

The covariance is defined as:

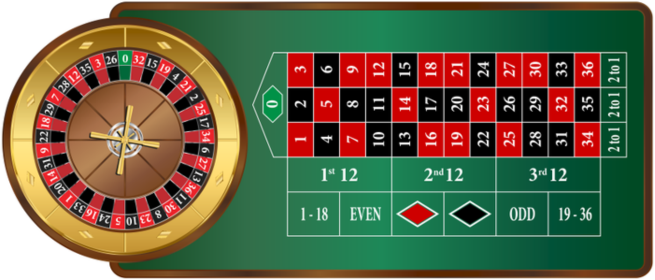
$$\mathrm{Cov}(x, y) = \mathcal{E}\left((x - \mu_x)(y - \mu_y)\right),$$

in which $\mu_x = \mathcal{E}(x)$ and $\mu_y = \mathcal{E}(y)$. The following rules can be derived:

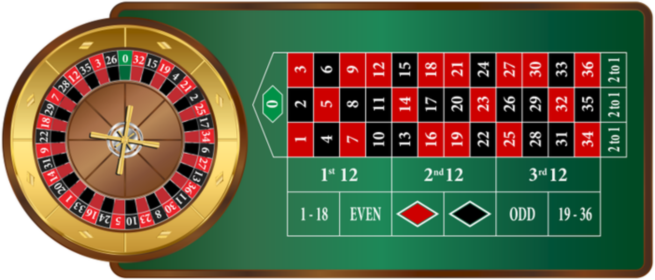
$$\begin{aligned}
\mathrm{Cov}(x, \alpha) &= 0 \\
\mathrm{Cov}(x, y) &= \mathrm{Cov}(y, x) \\
\mathrm{Cov}(\alpha x, \beta y) &= \alpha\beta\mathrm{Cov}(x, y) \\
\mathrm{Cov}(x + y, Z) &= \mathrm{Cov}(x, Z) + \mathrm{Cov}(y, Z) \\
\mathrm{Cov}(x, \alpha y + \beta Z) &= \alpha\mathrm{Cov}(x, y) + \beta\mathrm{Cov}(x, Z)
\end{aligned}$$

Where α and β are constants (parameter) and x, y, and Z are random variables. Furthermore, because the variance $\mathrm{Var}(x) = \mathrm{Cov}(x, x)$, we can derive:

$$\begin{aligned}
\mathrm{Var}(x + y) &= \mathrm{Var}(x) + \mathrm{Var}(y) + 2\mathrm{Cov}(x, y) \\
\mathrm{Var}(\beta x) &= \beta^2\mathrm{Var}(x)
\end{aligned}$$

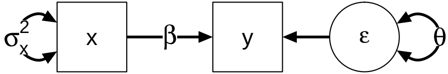


Given that $\mathrm{Var}(x) = \mathcal{E}\left((x-\mu_x)^2\right)$ and $\mathcal{E}\left(x\right) = \mu_x = -0.027$ for betting 1 Euro in Roulette. What is the variance and standard deviation of our bet?

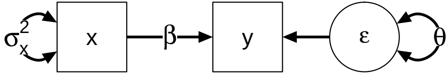


Given that $\mathrm{Var}(x) = \mathcal{E}\left((x - \mu_x)^2\right)$ and $\mathcal{E}\,(x) = \mu_x = -0.027$ for betting 1 Euro in Roulette. What is the variance and standard deviation of our bet?

$$\begin{aligned}
\mathrm{Var}\,(\text{bet 1 Euro}) &= \frac{1}{37} \times (35 - -0.027)^2 + \frac{36}{37} \times (-1 - -0.027)^2 \\
&= 34.14 \\
\mathrm{SD}\,(\text{bet 1 Euro}) &= \sqrt{\mathrm{Var}\,(\text{bet 1 Euro})} = 5.84
\end{aligned}$$



$$y_i = \tau + \beta x_i + \varepsilon_i$$
$$x \sim N(\mu_x, \sigma_x)$$
$$\varepsilon \sim N(0, \theta)$$

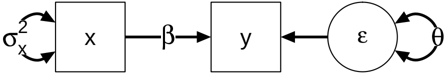


$$y_i = \tau + \beta x_i + \varepsilon_i$$
$$x \sim N(\mu_x, \sigma_x)$$
$$\varepsilon \sim N(0, \theta)$$

We can now derive:

$$\begin{aligned}\mathrm{var}\left(y\right) &= \mathrm{var}\left(\tau + \beta x + \varepsilon\right)\\ &= \mathrm{var}\left(\beta x + \varepsilon\right)\end{aligned}$$

Because $\boldsymbol{\tau}$ is a constant.



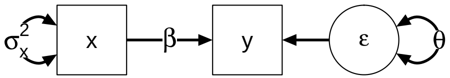
$$y_i = \tau + \beta x_i + \varepsilon_i$$
$$x \sim N(\mu_x, \sigma_x)$$
$$\varepsilon \sim N(0, \theta)$$

We can now derive:

$$\begin{aligned}\operatorname{var}(y) &= \operatorname{var}(\tau + \beta x + \varepsilon) \\ &= \operatorname{var}(\beta x + \varepsilon)\end{aligned}$$

Because $\boldsymbol{\tau}$ is a constant. next:

$$\begin{aligned}\operatorname{Var}(y) &= \operatorname{Cov}(\beta x + \varepsilon, \beta x + \varepsilon) \\ &= \operatorname{Cov}(\beta x, \beta x + \varepsilon) + \operatorname{Cov}(\varepsilon, \beta x + \varepsilon) \\ &= \operatorname{Cov}(\beta x, \beta x) + \operatorname{Cov}(\beta x, \varepsilon) + \operatorname{Cov}(\varepsilon, \beta x) + \operatorname{Cov}(\varepsilon, \varepsilon)\end{aligned}$$



$$y_i = \tau + \beta x_i + \varepsilon_i$$
$$x \sim N(\mu_x, \sigma_x)$$
$$\varepsilon \sim N(0, \theta)$$

We can now derive:

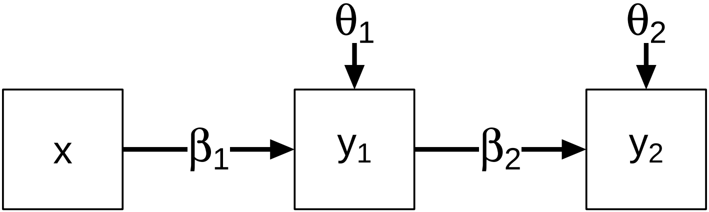
$$\begin{aligned}\mathrm{var}(y) &= \mathrm{var}(\tau + \beta x + \varepsilon) \\ &= \mathrm{var}(\beta x + \varepsilon)\end{aligned}$$

Because $\boldsymbol{\tau}$ is a constant. next:

$$\begin{aligned}\mathrm{Var}(y) &= \mathrm{Cov}(\beta x + \varepsilon, \beta x + \varepsilon) \\ &= \mathrm{Cov}(\beta x, \beta x + \varepsilon) + \mathrm{Cov}(\varepsilon, \beta x + \varepsilon) \\ &= \mathrm{Cov}(\beta x, \beta x) + \mathrm{Cov}(\beta x, \varepsilon) + \mathrm{Cov}(\varepsilon, \beta x) + \mathrm{Cov}(\varepsilon, \varepsilon)\end{aligned}$$

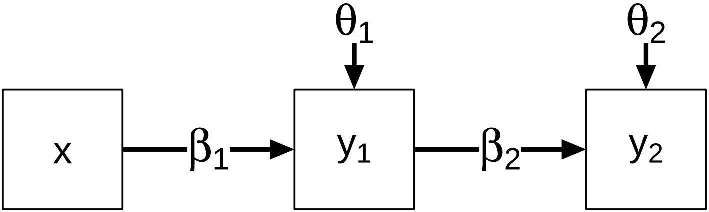
But since x is not correlated with the residuals, $\mathrm{Cov}(x, \varepsilon) = 0$ and thus:

$$\begin{aligned}\mathrm{Var}(y) &= \beta^2 \mathrm{Cov}(x, x) + \mathrm{Cov}(\varepsilon, \varepsilon) \\ &= \beta^2 \mathrm{Var}(x) + \mathrm{Var}(\varepsilon) \\ &= \beta^2 \sigma_x^2 + \theta^2\end{aligned}$$



$$y_{i1} = \beta_1 x_i + \varepsilon_{i1}$$
$$y_{i2} = \beta_2 y_{i1} + \varepsilon_{i2} = \beta_2(\beta_1 x_i + \varepsilon_{i1}) + \varepsilon_{i2}$$

Number of parameters (ignoring mean structure): 2 regressions $+2$ residual variances $+1$ exogenous variance (not drawn) $= 5$, number of observations: 3 variances and 3 covariances. 1 degree of freedom!



Implied covariance between x and y_2:

$$\begin{aligned}
\mathrm{Cov}\left(x, y_2\right) &= \mathrm{Cov}\left(x, \beta_2(\beta_1 x_i + \varepsilon_{i1}) + \varepsilon_{i2}\right) \\
&= \mathrm{Cov}\left(x, \beta_2\beta_1 x + \beta_2\varepsilon_1 + \varepsilon_2\right) \\
&= \mathrm{Cov}\left(x, \beta_2\beta_1 x\right) + \mathrm{Cov}\left(x, \beta_2\varepsilon_1\right) + \mathrm{Cov}\left(x, \varepsilon_2\right) \\
&= \beta_1\beta_2 \mathrm{Cov}\left(x, x\right) \\
&= \beta_1\beta_2\sigma_x^2
\end{aligned}$$

On Canvas you will find some more practice exercises that you can work on for the remainder of the practical!