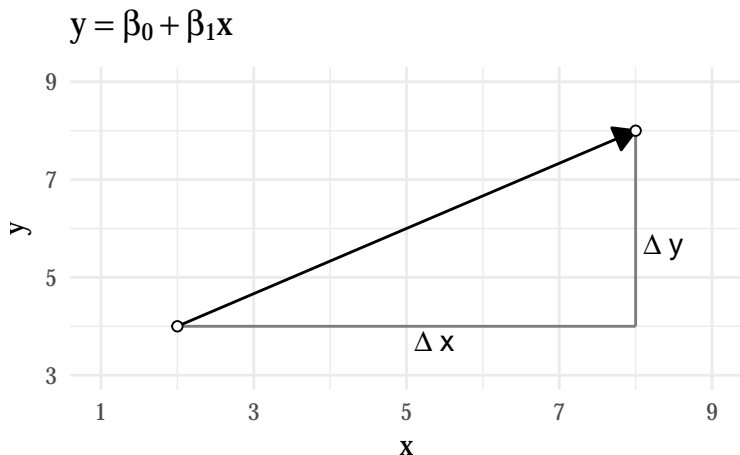


Sitzungsbegleitende Folien vom 06/19/2017

Dag Tanneberg

06/19/2017

Wie bestimmt man eine Gerade?

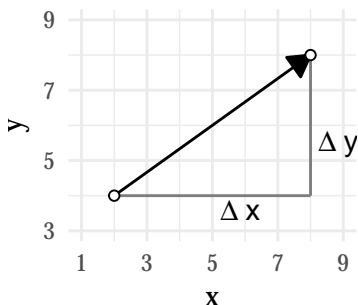


Wie bestimmt man eine Gerade?

geg. $P_1 = (2, 4); P_2 = (8, 8);$

ges. $\beta_0; \beta_1$

$$y = \beta_0 + \beta_1 x$$



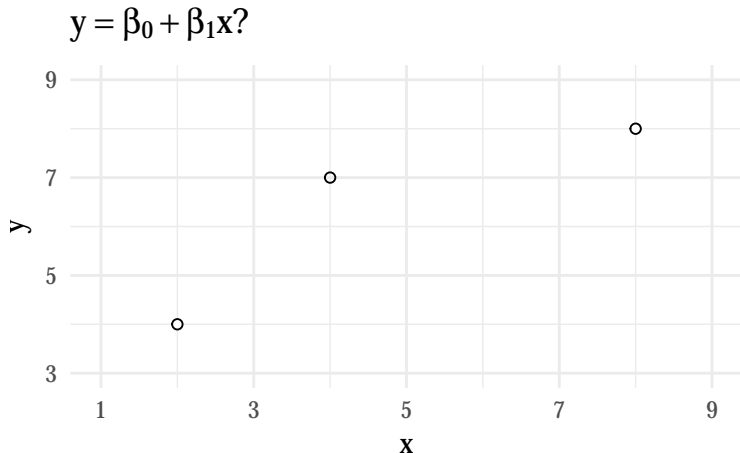
$$\beta_1 = \frac{\Delta y}{\Delta x} = \frac{8 - 4}{8 - 2} = \frac{4}{6} = \frac{2}{3}$$

$$4 = \beta_0 + \beta_1 * 2$$

$$\beta_0 = 4 - \frac{2}{3} * 2 = 4 - \frac{4}{3} = \frac{8}{3}$$

$$y = \frac{8}{3} + \frac{2}{3} * x$$

Wie bestimmt man jetzt eine Gerade?



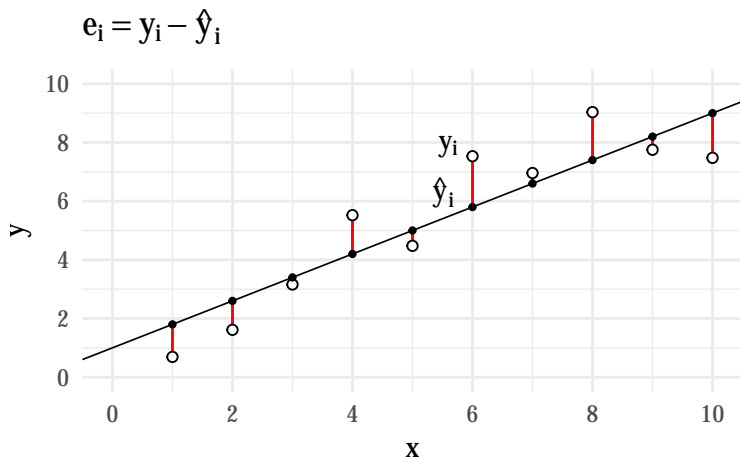
Wie bestimmt man jetzt eine Gerade?

- eindeutige Lösung ist nicht möglich
- Schulwissen versagt bereits bei 3 Datenpunkten
- Näherungslösungen werden benötigt

Was ist eine Näherungslösung?

- Verfahren, das Voraussagen \hat{y} über y trifft
- Voraussage und Beobachtung stimmen i.d.R. nicht überein
- Differenz zwischen Beobachtung und Vorhersage \equiv Residuum
 - formal: $e = y - \hat{y}$
- Interpretationen für Residuen sind vielfältig
 - Messfehler für y
 - Übersehene, systematische Einflüsse auf y
 - Echte Zufälligkeit in der Beziehung $y \sim x$, etc.
- Näherungslösungen kann man über ihre Residuen vergleichen

Was ist eine Näherungslösung?



Was ist eine gute Näherungslösung?

- Ansatz: $y = f(x, \beta, \epsilon) = f(x, \beta) + \epsilon = x\beta + \epsilon = \beta_0 + \beta_1 x + \epsilon$
 - 1 y ist eine lineare Funktion von x
 - 2 β_1 gibt Auskunft über die Stärke dieser Beziehung
 - 3 ϵ kennzeichnet einen unvermeidbaren Vorhersagefehler
- Ziel: β schätzen, sodass ϵ in der Summe minimal wird

$$\min_{\hat{\beta}} \sum_{i=1}^N \omega_i |e_i|$$

Was leistet das OLS-Verfahren?

- Ordinary least squares (OLS): $\omega \sim$ Größe der Residuen

$$\begin{aligned} & \min_{\hat{\beta}} \sum_{i=1}^N |e_i| \cdot |e_i| \\ &= \min_{\hat{\beta}} \sum_{i=1}^N e_i^2 \\ &= \min_{\hat{\beta}} \sum_{i=1}^N (y_i - \hat{y}_i)^2 \\ &= \min_{\hat{\beta}} \sum_{i=1}^N (y_i - (\beta_0 + \beta_1 x_i))^2 \end{aligned}$$

Beispiel ohne erklärende Variable

$$\min_{\hat{y}} \sum_{i=1}^N (y_i - \hat{y})^2$$

First order condition:

$$0 = \frac{\partial}{\partial \hat{y}} \sum_{i=1}^N (y_i - \hat{y})^2 = -2 \sum_{i=1}^N (y_i - \hat{y}) = (-2 \sum_{i=1}^N (y_i)) + 2N\hat{y}$$

$$\hat{y} = \frac{1}{N} \sum_{i=1}^N (y_i) = \bar{y}$$

Zusammenfassend

- 1 Ist die Beziehung zwischen zwei oder mehr Variablen nicht eindeutig bestimmt, dann müssen Näherungslösungen verwendet werden. Dabei entstehen Residuen $e = y - \hat{y}$.
- 2 Diese Residuen kann man nutzen, um unterschiedliche Näherungslösungen miteinander zu vergleichen.
- 3 Die lineare Regression ist eine Näherungslösung. Sie behauptet, eine abhängige Variable y ließe sich auf eine gewichtete Summe unabhängiger Variablen $X\beta$ zurückführen.
- 4 Schätzt man ein lineares Regressionsmodell im OLS verfahren, dann sucht man Regressionsgewichte β , die die quadrierte Summe der Residuen minimieren.
- 5 Unter Geltung weiterer Annahmen (Gauss-Markov-Annahmen) sind diese Koeffizienten *blue*, d.h. *best unbiased linear estimators*.

Genug Trockenschwimmen

Auf nach STATA!