Kapitel 2: Das einfache Regressionsmodell

Moodle

Lehrbuch

Übersicht:

Definition des einfachen Regressionsmodells

Herleitung der gewöhnlichen Kleinste-Quadrate-Schätzer

OLS: Eigenschaften bei beliebigen Stichproben

Maßeinheiten und die funktionale Form

Erwartungswerte und Varianzen der OLS-Schätzer

Definition des einfachen Regressionsmodells

Definition des einfachen Regressionsmodells

Das einfache Regressionsmodell erklärt die Variable y durch die Variable x.

Drei Aspekte:

- Wie modellieren wir die Wirkung von anderen Faktoren auf y?
- Was ist die funktionale Beziehung zwischen y und x?
- Wie können wir ein Ceteris-Paribus Zusammenhang sicherstellen?

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

$$\begin{aligned} y &: \quad \text{Abhängige/erklärte Variable, Regressand} \\ \beta_0 &: \quad \text{Achsenabschnitt (unbekannter Parameter)} \\ \beta_1 &: \quad \text{Steigung (unbekannter Parameter)} \\ x &: \quad \text{Unabhängige/erklärende Variable, Regressor} \\ u &: \quad \text{Störterm (u: unbeobachtbar)} \end{aligned}$$

Interpretation des einfachen Regressionsmodells

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

Das einfache Regressionsmodell erklärt inwiefern y in Veränderungen von x variiert.

Ceteris Paribus-Annahme:

Nehme an, dass sich x verändert ($\Delta x \neq 0$) und alle anderen erklärenden aber unbeobachtbaren Größen konstant bleiben ($\Delta u = 0$).

Dann gilt

$$\frac{\Delta y}{\Delta x} = \beta_1$$

(x_{neu}, y_{neu}) , (x_{alt}, y_{alt}) $y = \beta_0 + \beta_1 x + u$

$\Delta y = y_{neu} - y_{alt}$

$\Delta x = x_{neu} - x_{alt}$

$$\Delta y = y_{neu} - y_{alt} = \overbrace{\cancel{\beta_0} + \beta_1 x_{neu} + u_{neu}}^{y_{neu}} - \overbrace{(\cancel{\beta_0} + \beta_1 x_{alt} + u_{alt})}^{y_{alt}}$$

$$= \beta_1 \underbrace{(x_{neu} - x_{alt})}_{\Delta x} + \underbrace{u_{neu} - u_{alt}}_{\Delta u}$$

$$\Delta y = \beta_1 \Delta x + \Delta u$$

Ceteris Paribus : $\Delta x \neq 0 \Rightarrow \Delta u = 0$

$$\Delta y = \beta_1 \underbrace{\Delta x}_{\neq 0} \iff \frac{\Delta y}{\Delta x} = \beta_1$$

Beispiel: Sojabohnen und Dünger

„Wie wirkt sich die Düngemenge auf den Ernteerfolg aus?“

$$yield = \beta_0 + \beta_1 fertilizer + u \ ,$$

sodass $y = yield$ und $x = fertilizer$.

Der Störterm u beinhaltet andere Größen wie Landqualität, Niederschlag, Parasiten.

Unter der Ceteris Paribus-Annahme gilt:

$$\Delta yield = \beta_1 \Delta fertilizer$$

Beispiel: Eine einfache Lohngleichung

„Wie wirkt sich Bildung auf den Stundenlohn aus?“

$$wage = \beta_0 + \beta_1 educ + u \ ,$$

sodass $y = wage$ und $x = educ$.

Der Störtem u beinhaltet andere Größen wie Berufserfahrung, Talent, etc.

Der unbekannte Parameter β_1 misst (unter der Ceteris Paribus-Annahme), wie sich ein zusätzliches Bildungsjahr auf den Stundenlohn auswirkt.

Diskussion der Linearität des einfachen Regressionsmodells

Die Gleichung

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

impliziert, dass die Variable x immer den gleichen (linearen) Effekt auf y hat.

Diese Unzulänglichkeit kann durch Transformation der Variablen geheilt werden.

Der schwierigste Aspekt ist die Ceteris Paribus Annahme.

Um diese Annahme zu rechtfertigen interpretieren wir zunächst x und u (und damit auch y) als Zufallsvariablen.

Nun können wir eine statistische Annahme an x und u treffen.

Ein (scheinbar nicht-lineares Modell)

$$\tilde{y} = \overset{\text{const}}{A} \cdot \tilde{x}^{\tilde{\beta}_1} \cdot \overset{\text{error}}{\tilde{u}}$$

$$\ln(\tilde{y}) = \ln(A \cdot \tilde{x}^{\tilde{\beta}_1} \cdot \tilde{u}) = \ln(A) + \ln(\tilde{x}^{\tilde{\beta}_1}) + \ln(\tilde{u})$$

$$= \underbrace{\ln(A)}_{=:\beta_0} + \underbrace{\tilde{\beta}_1}_{\beta_1} \underbrace{\ln(\tilde{x})}_{\text{Regressor}} + \underbrace{\ln(\tilde{u})}_{u \text{ Störterm}}$$

Strikte Exogenität

In Kenntnis der erklärenden Variable x gibt es keine zusätzliche Information über den Erwartungswert der unbeobachtbaren Variable u:

$$E[u|x] = E[u]$$

Im Düngerbeispiel:

Die unterschiedlichen Düngermengen werden zufällig auf die einzelnen Ackerflächen verteilt. Demnach enthält die Düngermenge keine Information über die unbeobachtbare Bodenqualität der einzelnen Ackerfläche. (✓)

Im Bildungsbeispiel:

Die unterschiedlichen Bildungsniveaus können durchaus von den zugrundeliegenden intellektuellen Fähigkeiten abhängen. Daher enthält das Bildungsniveau (statistische) Informationen über die zu erwartenden unbeobachtbaren Fähigkeiten. (!)

Annahme: Strikte Exogenität

Im ersten Teil der Vorlesung werden wir die Annahme der **strikten Exogenität der Regressoren** (auch: mean independence) aufrechterhalten:

$$E[u|x] = E[u]$$

Erst in späteren Kapiteln gehen wir genauer auf die Implikationen und Lösungen bei einer Verletzungen dieser Annahme ein.

Ohne Beschränkung der Allgemeinheit: $E[u] = 0$

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

Wir nehmen des Weiteren an, dass der Erwartungswert des unbeobachtbaren Störterms null beträgt:

$$E[u] = 0$$

Diese Annahme erfolgt jedoch ohne Beschränkung der Allgemeinheit!

Wäre $E[u] \neq 0$, so könnten wir das Modell mit $\tilde{u} = u - E[u]$ und $\tilde{\beta}_0 = \beta_0 + E[u]$ modifzieren.

Für das modifizierte Modell

$$y = \tilde{\beta}_0 + \beta_1 x + \tilde{u}$$

wäre $E[\tilde{u}] = E[u - E[u]] = E[u] - E[u] = 0$ erfüllt.

Annahme: Bedingter Null-Erwartungswert

Engl.: zero conditional mean assumption

Mit den beiden Annahmen

$$E[u] = 0$$

und

$$E[u|x] = E[u]$$

folgt direkt:

$$E[u|x] = 0$$

Oft wird direkt diese Gleichung (ohne Vordiskussion) angenommen.

Population Regression Function (PRF)

Bilden wir den bedingten Erwartungswert des Regressionsmodells

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

erhalten wir

$$\begin{aligned} E[y|x] &= E[\beta_0 + \beta_1 x + u|x] \\ &= E[\beta_0|x] + E[\beta_1 x|x] + \underbrace{E[u|x]}_{=0} \end{aligned}$$

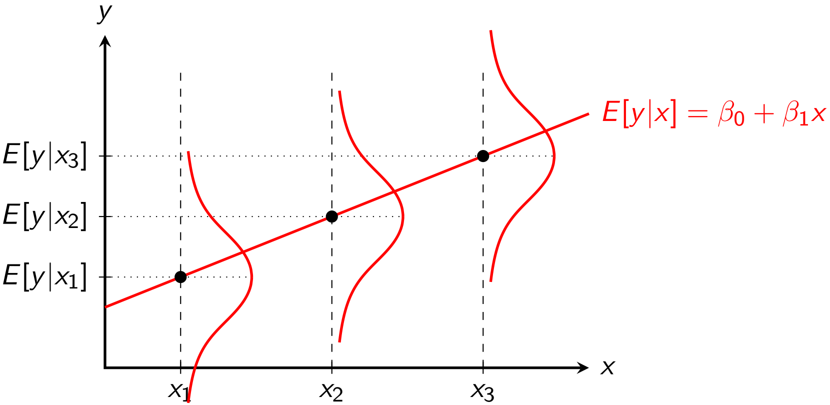
und mit $E[u|x] = 0$ also

$$E[y|x] = \beta_0 + \beta_1 x$$

Eine Erhöhung von x um eine Einheit erhöht den **bedingten Erwartungswert** von y um β_1 Einheiten.

Population Regression Function (PRF)

Für einen gegebenen Wert x ist $E[y|x] = \beta_0 + \beta_1 x$ der Gravitationspunkt der bedingten Dichte von y.



Die Gleichung $E[y|x] = \beta_0 + \beta_1 x$ beschreibt den erwarteten Wert von y. Der tatsächliche Wert von y streut um $E[y|x]$.

Herleitung der gewöhnlichen Kleinste-Quadrate-Schätzer

Herleitung von gewöhnlichen Schätzern

Um die unbekannten Parameter β_0 und β_1 des einfachen linearen Regressionsmodells zu schätzen, benötigen wir neben den Annahmen

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

und

$$E[u|x] = 0$$

Daten: die **Zufallsstichprobe**

$$\begin{array}{ccl} (x_1, y_1) & : & \text{Erste Beobachtung} \\ (x_2, y_2) & : & \text{Zweite Beobachtung} \\ & \vdots & \\ (x_n, y_n) & : & n\text{-te Beobachtung} \end{array}$$

Herleitung von gewöhnlichen Schätzern

Sei

$$\{(x_i, y_i) \in \mathbb{R}^2 : i = 1, \ldots, n\}$$

eine Zufallsstichprobe des Umfangs $n \in \mathbb{N}$.

Für $i = 1, \ldots, n$:
x_i: Wert der erklärenden Variable der i-ten Beobachtung
y_i: Wert der erklärten Variable der i-ten Beobachtung

Das Regressionsmodell lässt sich nun schreiben als

$$y_i = \beta_0 + \beta_1 x_i + u_i,\ \ i = 1, \ldots, n$$

u_i: Störterm für Beobachtung i.
Enthält alle Einflussfaktoren auf y_i, die nicht in x_i enthalten sind.

Die Momentenmethode (Method of Moments, MM)

Seien n Paare von Zufallsvariablen (A_i, B_i) unabhängig und identisch verteilt mit theoretischen Momenten $E[A|\boldsymbol{\theta}], E[B|\boldsymbol{\theta}], E[AB|\boldsymbol{\theta}], E[A^2|\boldsymbol{\theta}]$ und $E[B^2|\boldsymbol{\theta}]$, wobei $\boldsymbol{\theta}$ ein Vektor von unbekannten Parametern darstellt.

Dann gilt mit dem Gesetz der großen Zahlen

$$\frac{1}{n}\sum_{i=1}^{n} A_i \underset{n\to\infty}{\longrightarrow} E[A|\boldsymbol{\theta}],\ \frac{1}{n}\sum_{i=1}^{n} A_i B_i \underset{n\to\infty}{\longrightarrow} E[AB|\boldsymbol{\theta}],\ \text{usw.}$$

Die Momentenmethode approximiert nun die theoretischen Momente durch die empirischen Momente und schätzt hier die unbekannten Parameter $\boldsymbol{\theta}$ durch die implizit definierten Schätzer $\hat{\boldsymbol{\theta}}$:

$$\frac{1}{n}\sum_{i=1}^{n} A_i = E[A|\hat{\boldsymbol{\theta}}],\ \frac{1}{n}\sum_{i=1}^{n} A_i B_i = E[AB|\hat{\boldsymbol{\theta}}],\ \text{usw.}$$

Die Momentenmethode mit $E[u|x] = 0$

Für $y_i = \beta_0 + \beta_1 x_i + u_i$, $i = 1, \dots, n$, sei angenommen, dass u_i unabhängig und identisch verteilt seien mit $E[u_i|x_i] = 0$.

Zunächst schreiben wir $u_i = y_i - \beta_0 - \beta_1 x_i$.

Das theoretische Moment $E[u_i|x_i] = E[y_i|x_i] - \beta_0 - \beta_1 x_i$ enthält die unbekannten Parameter β_0 und β_1. Für diese definieren wir die Schätzer $\hat{\beta}_0$ und $\hat{\beta}_1$ implizit durch die Schätzung des theoretischen Moments durch das empirische Moment

$$\frac{1}{n}\sum_{i=1}^{n} u_i \approx \frac{1}{n}\sum_{i=1}^{n} y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i = 0 = E[u_i|x_i]$$

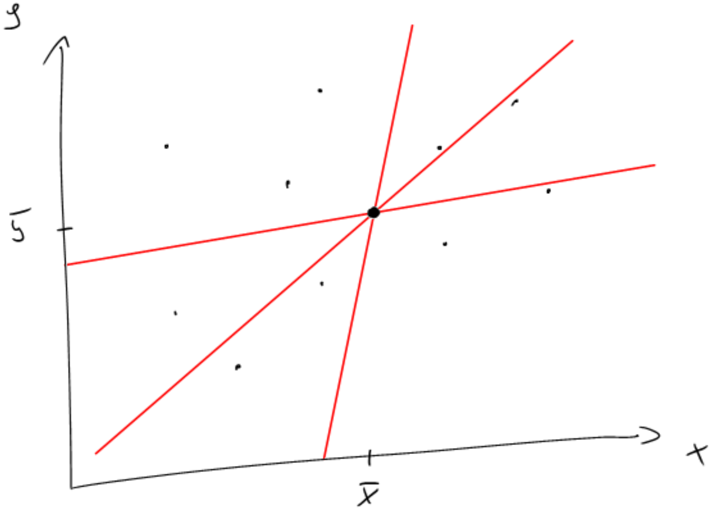
Mit $\bar{y} := \frac{1}{n}\sum_{i=1}^{n} y_i$ und $\bar{x} := \frac{1}{n}\sum_{i=1}^{n} x_i$:

$$\Rightarrow \bar{y} = \hat{\beta}_0 + \hat{\beta}_1 \bar{x}$$

$$\frac{1}{n}\sum_i y_i - \frac{1}{n}\underbrace{\sum_i \hat{\beta}_0}_{n \cdot \hat{\beta}_0} - \frac{1}{n}\sum_i \hat{\beta}_1 x_i = 0$$

$$\frac{1}{n}\sum_i y_i - \hat{\beta}_0 - \hat{\beta}_1 \frac{1}{n}\sum_i x_i = 0$$

$$\bar{y} - \hat{\beta}_0 - \hat{\beta}_1 \bar{x} = 0$$



y
ȳ
x̄
x

Die Momentenmethode mit $E[u|x] = 0$

Für das Moment $E[xu]$ gilt aufgrund der Annahme $E[u|x] = 0$:

$$E[xu] = E[x \underbrace{E[u|x]}_{=0}] = 0$$

Die Eigenschaft $E[xu] = 0$ wird auch mit **Orthogonalität** von x und u bezeichnet.

Mit $u_i = y_i - \beta_0 - \beta_1 x_i$ schätzen wir $E[xu]$ durch

$$\frac{1}{n}\sum_{i=1}^{n} x_i u_i \approx \frac{1}{n}\sum_{i=1}^{n} x_i \left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right) = E[xu] = 0$$

Mit $\overline{xy} := \frac{1}{n}\sum_{i=1}^{n} x_i y_i$ und $\overline{xx} := \frac{1}{n}\sum_{i=1}^{n} x_i^2$:

$$\Rightarrow \overline{xy} = \hat{\beta}_0 \bar{x} + \hat{\beta}_1 \overline{xx}$$

Die Momentenmethode mit $E[u|x] = 0$

Die Momentenmethode liefert also über die Annahme $E[u|x] = 0$ zwei Gleichungen für die beiden Schätzer $\hat{\beta}_0$ und $\hat{\beta}_1$:

$$\bar{y} = \hat{\beta}_0 + \hat{\beta}_1 \bar{x}$$

und

$$\overline{xy} = \hat{\beta}_0 \bar{x} + \hat{\beta}_1 \overline{xx}$$

Unter der Voraussetzung $\overline{xx} \neq \bar{x}\bar{x}$ ergibt sich:

$$\hat{\beta}_0 = \bar{y} - \frac{\overline{xy} - \bar{x}\bar{y}}{\overline{xx} - \bar{x}\bar{x}} \bar{x}$$

und

$$\hat{\beta}_1 = \frac{\overline{xy} - \bar{x}\bar{y}}{\overline{xx} - \bar{x}\bar{x}}$$

Die gewöhnliche Kleinste-Quadrate-Methode (OLS)

Eine andere Schätzmöglichkeit besteht in der Vorgabe einer Zielfunktion.

Die gebräuchlichste ist:

Minimiere die Summe der quadrierten Abweichungen zwischen der erklärten Variablen und ihren durch die Schätzung vorhergesagten Werten.

Dies ist das gewöhnliche Kleinste-Quadrate-Prinzip (**OLS, ordinary least squares**).

Die gewöhnliche Kleinste-Quadrate-Methode (OLS)

Schätze die unbekannten Parameter β_0 und β_1 des Regressionsmodells

$$y_i = \beta_0 + \beta_1 x_i + u_i, \; i = 1, \ldots, n$$

zunächst durch beliebige Zahlen b_0 und b_1.

$\rightarrow$ Prognostizierte Werte $\hat{y}_i$ (**fitted values**) mit

$$\hat{y}_i = b_0 + b_1 x_i, \; i = 1, \ldots, n$$

Definiere **Residuen** $\hat{u}_i$ mit

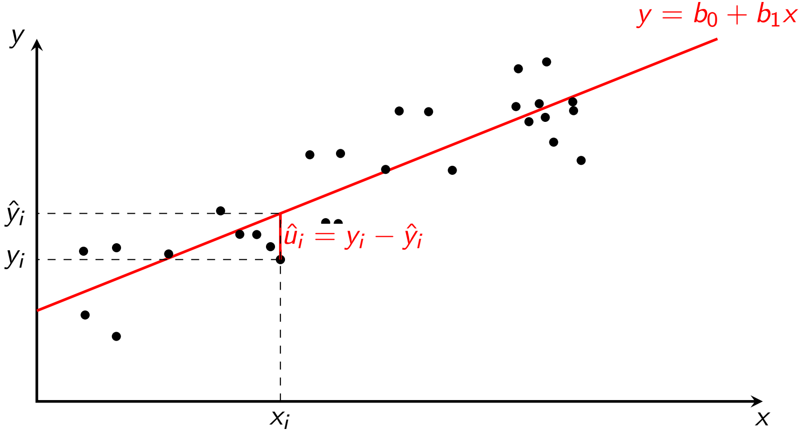
$$\hat{u}_i = y_i - \hat{y}_i = y_i - b_0 - b_1 x_i$$

Je besser die Zahlen b_0 und b_1, desto „kleiner“ die Residuen.

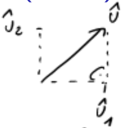
Prognose und Residuen

Vermutete Parameter: b_0, b_1

Schätzfehler (hängt von b_0, b_1 ab): $\hat{u}_i = y_i - \hat{y}_i = y_i - b_0 - b_1 x_i$



Die gewöhnliche Kleinste-Quadrate-Methode (OLS)



$\hat{u}_1^2 + \hat{u}_2^2 = ||\hat{u}||^2$

$\Rightarrow ||\hat{u}|| = \sqrt{\hat{u}_1^2 + \hat{u}_2^2}$

Der Vektor $\hat{u} = (\hat{u}_1, \hat{u}_2, \ldots, \hat{u}_n)$ hat die Länge

$$||\hat{u}|| = \sqrt{\hat{u}_1^2 + \hat{u}_2^2 + \ldots + \hat{u}_n^2}$$

Wähle nun b_0 und b_1, sodass die Norm $||\hat{u}||$ minimal wird!

$\Rightarrow$ Zielfunktion $S(b_0, b_1) = \sum_{i=1}^{n} (\underbrace{y_i - (\overbrace{b_0 + b_1 x_i}^{\hat{y}_i})}_{\hat{u}_i})^2$

$$\min_{b_0, b_1 \in \mathbb{R}} \sum_{i=1}^{n} (y_i - b_0 - b_1 x_i)^2$$

$$\min_{b_0, b_1 \in \mathbb{R}} \sum_{i=1}^{n} \big(y_i - \underbrace{(b_0 + b_1 x_i)}_{\hat{y}_i}\big)^2 = S(b_0, b_1)$$

BEO

$$\frac{\partial S}{\partial b_0} = \sum_{i=1}^{n} 2\left(y_i - b_0 - b_1 x_i\right) \cdot (-1)$$

$$= -2 \sum_{i=1}^{n} y_i - b_0 - b_1 x_i \overset{!}{=} 0 \quad \Big\| :(-2)$$

$$\Leftrightarrow \sum_{i=1}^{n} y_i - b_0 - b_1 x_i = 0$$

$\vdots$ (siehe Momente Methode)

$$\Leftrightarrow \bar{y} = b_0 + b_1 \bar{x}$$

$$\min_{b_0, b_1 \in \mathbb{R}} \sum_{i=1}^{n} \Big(y_i - \underbrace{(b_0 + b_1 x_i)}_{\hat{y}_i} \Big)^2 = S(b_0, b_1)$$

$$\frac{\partial S}{\partial b_1} = \sum_{i=1}^{n} 2 \left(y_i - b_0 - b_1 x_i \right)(-x_i)$$

| BEO

$$= -2 \sum_{i=1}^{n} \left(y_i - b_0 - b_1 x_i \right) x_i \overset{!}{=} 0 \quad \| :(-2)$$

$$\Leftrightarrow \sum_{i=1}^{n} \left(y_i - b_0 - b_1 x_i \right) x_i = 0$$

$\vdots$ (siehe MM)

$$\Leftrightarrow \overline{xy} = b_0 \overline{x} + b_1 \overline{xx}$$

Hesse matrix

$$H = \begin{pmatrix} \dfrac{\partial^2 S}{(\partial b_0)^2} & \dfrac{\partial^2 S}{\partial b_0\, \partial b_1} \\[2em] \dfrac{\partial^2 S}{\partial b_1\, \partial b_0} & \dfrac{\partial^2 S}{(\partial b_1)^2} \end{pmatrix} = \begin{pmatrix} 2n & 2n\bar{x} \\[2em] 2n\bar{x} & 2n\,\overline{xx} \end{pmatrix}$$

$$\frac{\partial^2 S}{(\partial b_0)^2} = -2\sum_{i=1}^{n}(-1) = 2\sum_{i=1}^{n} 1 = 2n$$

$$\frac{\partial^2 S}{\partial b_0\, \partial b_1} = -2\sum_{i=1}^{n}(-x_i) = n2\underbrace{\frac{1}{n}\sum_{i=1}^{n} x_i}_{\bar{x}} = 2\cdot n\cdot \bar{x}$$

$$\frac{\partial^2 S}{(\partial b_1)^2} = -2\sum_{i=1}^{n}(-x_i\cdot x_i) = n2\underbrace{\frac{1}{n}\sum_{i=1}^{n} x_i\cdot x_i}_{=\overline{xx}} = 2\cdot n\cdot \overline{xx}$$

$$H = 2n \begin{pmatrix} 1 & \overline{x} \\ \overline{x} & \overline{xx} \end{pmatrix}$$ positiv definit? $\checkmark$

$$\det(2n) = 2n \underline{\underline{> 0}} \quad \checkmark$$

$$\det\left(2n \begin{pmatrix} 1 & \overline{x} \\ \overline{x} & \overline{xx} \end{pmatrix}\right) = (2n)^2 \det \begin{pmatrix} 1 & \overline{x} \\ \overline{x} & \overline{xx} \end{pmatrix}$$

$$= (2n)^2 \left(1 \cdot \overline{xx} - \overline{x} \cdot \overline{x}\right) > 0 \quad \checkmark$$

$\Rightarrow$ S streng konvex

$\Rightarrow$ $\hat{\beta}_0, \hat{\beta}_1$ Minimumstelle von $S(b_0, b_1)$

Die gewöhnliche Kleinste-Quadrate-Methode (OLS)

$$\min_{b_0, b_1 \in \mathbb{R}} \sum_{i=1}^{n} (y_i - b_0 - b_1 x_i)^2$$

Die optimalen Schätzer $\hat{\beta}_0$ und $\hat{\beta}_1$ erfüllen die Bedingungen erster Ordnung:

$$\begin{aligned}
\sum_{i=1}^{n} 2(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)(-1) &= 0 \\
\sum_{i=1}^{n} 2(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)(-x_i) &= 0
\end{aligned}$$

Hessematrix:

$$H = \begin{pmatrix} 2n & 2\sum_{i=1}^{n} x_i \\ 2\sum_{i=1}^{n} x_i & 2\sum_{i=1}^{n} x_i^2 \end{pmatrix} = 2n \begin{pmatrix} 1 & \bar{x} \\ \bar{x} & \overline{xx} \end{pmatrix}$$

Die gewöhnliche Kleinste-Quadrate-Methode (OLS)

Die Hessematrix ist positiv definit, falls $\overline{xx} > \bar{x}\bar{x}$.

Die Bedingungen erster Ordnung sind äquivalent zu

$$\begin{aligned} \bar{y} &= \hat{\beta}_0 + \hat{\beta}_1 \bar{x} \\ \overline{xy} &= \hat{\beta}_0 \bar{x} + \hat{\beta}_1 \overline{xx} \end{aligned}$$

Unter der Voraussetzung $\overline{xx} > \bar{x}\bar{x}$ ergibt sich also auch hier:

$$\hat{\beta}_0 = \bar{y} - \frac{\overline{xy} - \bar{x}\bar{y}}{\overline{xx} - \bar{x}\bar{x}} \bar{x}$$

und

$$\hat{\beta}_1 = \frac{\overline{xy} - \bar{x}\bar{y}}{\overline{xx} - \bar{x}\bar{x}}$$

Orthogonale Projektion

Das Regressionsmodell

$$y_i = \beta_0 + \beta_1 x_i + u_i,\ \ i = 1, \ldots, n$$

liefert mit $E[u|x] = 0$ die funktionale Beziehung

$$E[y_i|x] = \beta_0 + \beta_1 x_i,\ \ i = 1, \ldots, n$$

Geometrische Interpretation:

$$\begin{pmatrix} E[y_1|x] \\ E[y_2|x] \\ \vdots \\ E[y_n|x] \end{pmatrix} = \beta_0 \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} + \beta_1 \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

Der Vektor $E[\mathbf{y}|x]$ ist eine Linearkombination des Einsvektors $\boldsymbol{\iota}$ und des Regressorvektors $\mathbf{x}$.

Orthogonale Projektion

Falls die beiden n-Vektoren $\boldsymbol{\iota}$ und $\mathbf{x}$ linear unabhängig sind, spannen sie eine Ebene im n-Dimensionalen Raum auf.

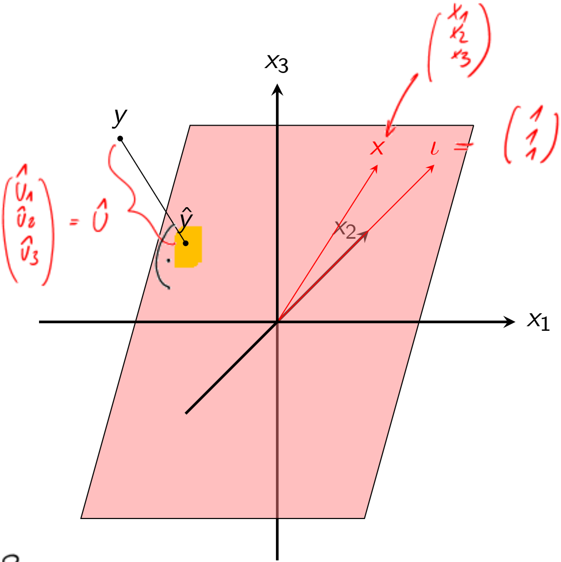
Die n-Vektoren $\boldsymbol{\iota}$ und $\mathbf{x}$ sind genau dann linear unabhängig, falls es kein $\lambda \in \mathbb{R}$ gibt, sodass $\mathbf{x} = \lambda\boldsymbol{\iota}$. Dies ist äquivalent zu $\overline{xx} \neq \bar{x}\bar{x}$.

Jedes Paar (b_0, b_1) definiert einen Punkt $b_0\boldsymbol{\iota} + b_1\mathbf{x}$ auf der von $\boldsymbol{\iota}$ und $\mathbf{x}$ aufgespannten Ebene.

Gesucht ist derjenige Punkt $\hat{\mathbf{y}} = \hat{\beta}_0\boldsymbol{\iota} + \hat{\beta}_1\mathbf{x}$, welcher in der Ebene am nächsten zum Vektor $\mathbf{y}$ ist.

Dieser Punkt $\hat{\mathbf{y}}$ ist die orthogonale Projektion von $\mathbf{y}$ auf die durch $\boldsymbol{\iota}$ und $\mathbf{x}$ aufgespannte Ebene.

Orthogonale Projektion



$\hat{u} \cdot x = 0$

$\hat{u} \cdot \iota = 0$

Orthogonale Projektion

Der Vektor von $\hat{\mathbf{y}}$ zu $\mathbf{y}$, also $\mathbf{y} - \hat{\mathbf{y}}$ muss demnach orthogonal zu den Vektoren $\boldsymbol{\iota}$ und $\mathbf{x}$ sein.

Es muss also gelten:

$$\begin{aligned}
\boldsymbol{\iota}'(\underbrace{\mathbf{y} - \hat{\mathbf{y}}}_{0}) &= \sum_{i=1}^{n} 1(y_i - \hat{y}_i) = \sum_{i=1}^{n} \hat{u}_i = 0 \Leftrightarrow \bar{y} = b_0 + b_1 \bar{x} \\
\mathbf{x}'(\underbrace{\mathbf{y} - \hat{\mathbf{y}}}_{0}) &= \sum_{i=1}^{n} x_i(y_i - \hat{y}_i) = \sum_{i=1}^{n} x_i \hat{u}_i = 0 \Leftrightarrow \overline{xy} = b_0 \bar{x} + b_1 \overline{xx}
\end{aligned}$$

Wie wir wissen implizieren diese beiden Bedingungen

$$\hat{\beta}_0 = \bar{y} - \frac{\overline{xy} - \bar{x}\bar{y}}{\overline{xx} - \bar{x}\bar{x}}\bar{x}$$

und

$$\hat{\beta}_1 = \frac{\overline{xy} - \bar{x}\bar{y}}{\overline{xx} - \bar{x}\bar{x}}$$

Diskussion: $\overline{xx} > \bar{x}\bar{x}$

Der Ausdruck $\overline{xx} - \bar{x}\bar{x}$ ist die **Stichprobenvarianz von** x:

$$\overline{xx} - \bar{x}\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

Die Stichprobenvarianz von x_i ist positiv, falls mindestens eine Beobachtung x_i vom Stichprobendurchschnitt $\bar{x}$ abweicht, falls also nicht alle Beobachtungen x_i identisch sind.

Wären alle Beobachtungen x_i identisch, hätte **x** keinen Erklärungsgehalt für die Variation von **y**.

Diskussion: $\overline{xy} - \bar{x}\bar{y}$

$\overline{xy} - \bar{x}\bar{y}$ ist die **Stichprobenkovarianz von x und y**:

$$\overline{xy} - \bar{x}\bar{y} = \frac{1}{n}\sum_{i=1}^{n} x_i y_i - \left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)\left(\frac{1}{n}\sum_{i=1}^{n} y_i\right) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$$

Wäre y_i eine lineare Funktion von x_i, also z.B.

$$y_i = \beta_0 + \beta_1 x_i \ ,$$

dann würde gelten: $\bar{y} = \beta_0 + \beta_1 \bar{x}$ und

$$\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) = \beta_1 \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \beta_1\left(\overline{xx} - \bar{x}\bar{x}\right)$$

und

$$\hat{\beta}_1 = \frac{\overline{xy} - \bar{x}\bar{y}}{\overline{xx} - \bar{x}\bar{x}} = \beta_1 \frac{\overline{xx} - \bar{x}\bar{x}}{\overline{xx} - \bar{x}\bar{x}} = \beta_1$$

Der Korrelationskoeffizient

Der (theoretische) Korrelationskoeffizient:

$$\rho_{xy} = \frac{Cov(x, y)}{\sqrt{Var(x)}\sqrt{Var(y)}}$$

Der empirische Korrelationskoeffizient ersetzt die theoretischen Momente durch die empirischen Momente:

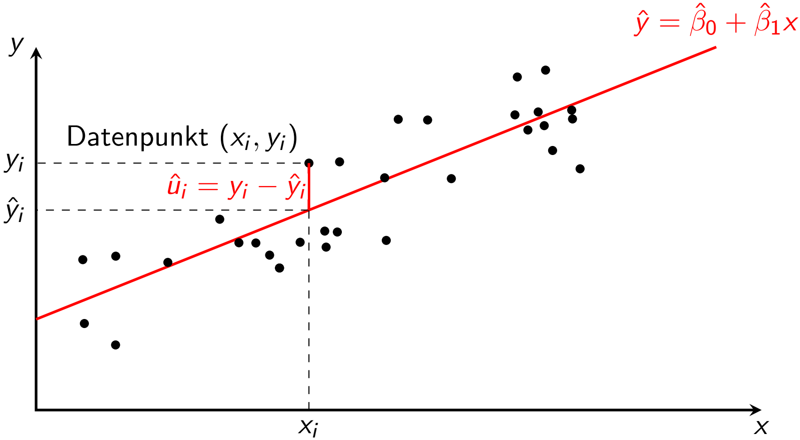
$$\hat{\rho}_{x,y} := \frac{\overline{xy} - \bar{x}\bar{y}}{\sqrt{(\overline{xx} - \bar{x}\bar{x})}\sqrt{(\overline{yy} - \bar{y}\bar{y})}}$$

Mit $\hat{\beta}_1 = \frac{\overline{xy} - \bar{x}\bar{y}}{\overline{xx} - \bar{x}\bar{x}}$ ergibt sich

$$\hat{\beta}_1 = \hat{\rho}_{x,y} \frac{\sqrt{\overline{yy} - \bar{y}\bar{y}}}{\sqrt{\overline{xx} - \bar{x}\bar{x}}}$$

Die gewöhnliche Regressionsanalyse misst also nur Korrelation und nicht Kausalität!

Die OLS Regressionsgerade



Die Regressionsgerade $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$ schätzt die unbekannte Population Regression Function $E[y|x] = \beta_0 + \beta_1 x$.

Beispiele der einfachen Regression

CEO-Gehälter und Eigenkapitalrentabilität

Population Regression Function

$$salary = \beta_0 + \beta_1 roe + u$$

$salary$: Gehalt in tausend \$

roe : durchschnittliche Eigenkapitalrentabilität (return on equity)

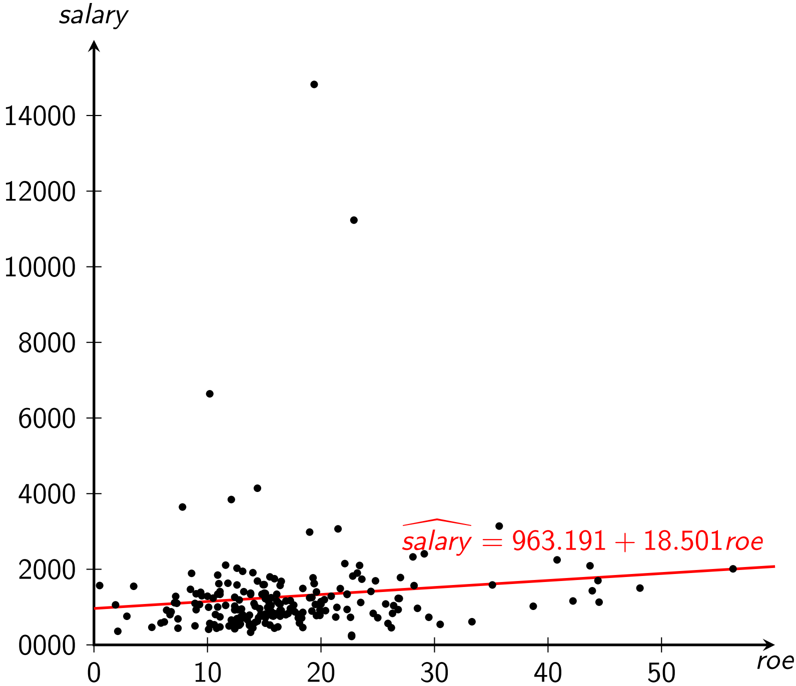
Regressionsgerade:

$$\widehat{salary} = 963.191 + 18.501 roe$$

Datensatz: CEOSAL1 ($n = 209$)

Die Regressionsgerade hängt von der Stichprobe ab.

Die Population Regression Function ist unbekannt!


salary
14000
12000
10000
8000
6000
4000
2000
0000
0
10
20
30
40
50
roe
$\widehat{salary} = 963.191 + 18.501 roe$

Beispiele der einfachen Regression

Stundenlohn und Bildung

Population Regression Function

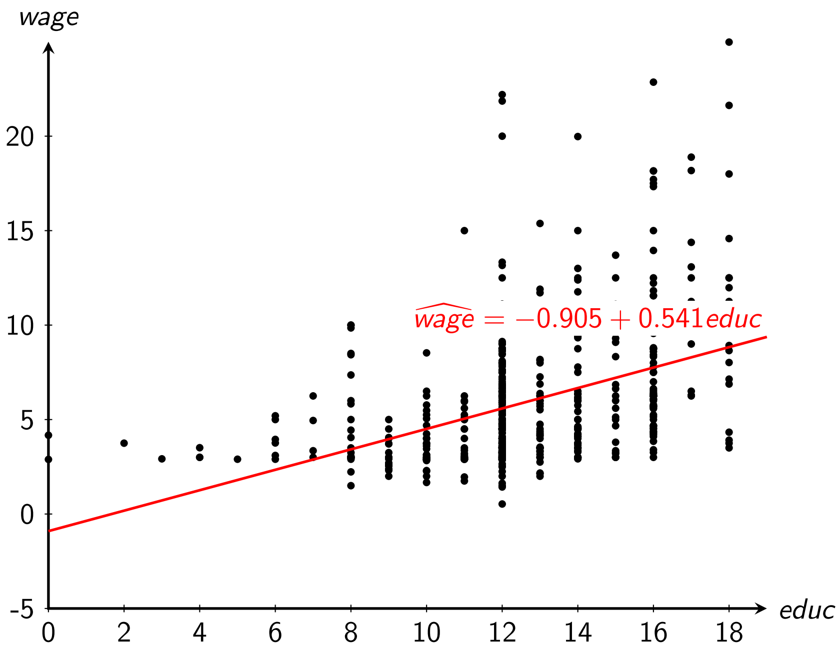
$$wage = \beta_0 + \beta_1 educ + u$$

$$\begin{aligned} wage &: \quad \text{Stundenlohn in \$} \\ educ &: \quad \text{Bildungsjahre} \end{aligned}$$

Regressionsgerade:

$$\widehat{wage} = -0.90 + 0.54 educ$$

Datensatz: WAGE1 ($n = 526$)


wage
20
15
10
5
0
-5
0
2
4
6
8
10
12
14
16
18
educ
$\widehat{wage} = -0.905 + 0.541 educ$

Beispiele der einfachen Regression

Wahlergebnisse und Wahlkampfausgaben

Population Regression Function

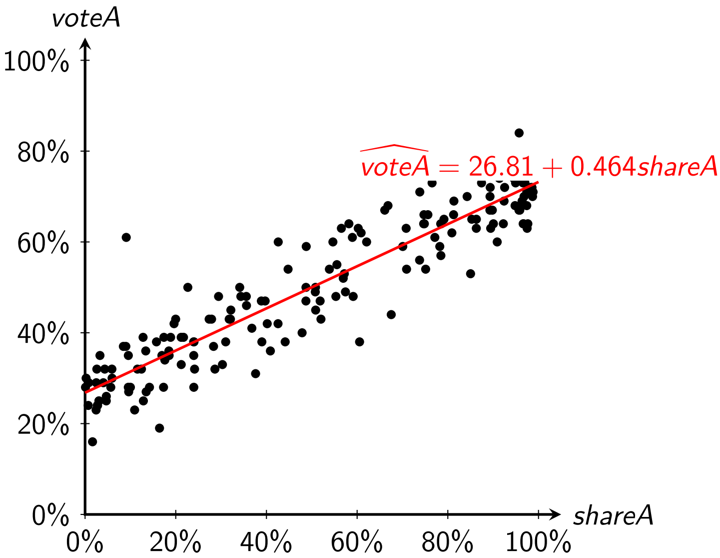
$$voteA = \beta_0 + \beta_1 shareA + u$$

$$\begin{aligned} votaA &: \quad \text{\% Kandidat:in A} \\ shareA &: \quad \text{Anteil Wahlkampfausgaben für Kandidat:in A} \end{aligned}$$

Regressionsgerade:

$$\widehat{voteA} = 26.81 + 0.464 shareA$$

Datensatz: VOTE1 ($n = 173$)


voteA
100%
80%
60%
40%
20%
0%
0%
20%
40%
60%
80%
100%
shareA
$\widehat{voteA} = 26.81 + 0.464 shareA$

OLS: Eigenschaften bei beliebigen Stichproben

Prognose und Residuen

Für eine beliebige[1] Stichprobe $\{(x_i, y_i)\}_{i=1}^n$ liefern die OLS-Schätzer $\hat{\beta}_0$ und $\hat{\beta}_1$ die Prognose

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$$

und damit die Residuen

$$\hat{u}_i = y_i - \hat{y}_i = y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i$$

für $i = 1, \ldots, n$

[1] Die Stichprobe muss verschiedene Werte x_i enthalten.

Beispiel: Managergehälter (Daten: CEOSAL1)

obsno	*roe*	*salary*	$\widehat{salary}$	$\hat{u}$
1	14.1	1095	1224.058	-129.0581
2	10.9	1001	1164.854	-163.8542
3	23.5	1122	1397.969	-275.9692
4	5.9	578	1072.348	-494.3484
5	13.8	1368	1218.508	149.4923
6	20.0	1145	1333.215	-188.2151
7	16.4	1078	1266.611	-188.6108
8	16.3	1094	1264.761	-170.7606
9	10.5	1237	1157.454	79.54626
10	26.3	833	1449.773	-616.7726
11	25.9	567	1442.372	-875.3721
12	26.8	933	1459.023	-526.0231
13	14.8	1339	1237.009	101.9911
14	22.3	937	1375.768	-438.7678
15	56.3	2011	2004.808	6.191895

Algebraische Eigenschaften der OLS-Statistiken

1. Die Summe der OLS-Residuen ergibt null:

$$\sum_{i=1}^{n} \hat{u}_i = 0 \qquad \Leftrightarrow \qquad \hat{u} \cdot \iota = 0$$

2. Die Stichprobenkovarianz des Regressors und der OLS-Residuen ist null:

$$\sum_{i=1}^{n} x_i \hat{u}_i = 0 \qquad \Leftrightarrow \qquad \hat{u} \cdot x = 0$$

3. Die Durchschnitte des Regressors und des Regressanden liegen auf der Regressionsgeraden:

$$\bar{y} = \hat{\beta}_0 + \hat{\beta}_1 \bar{x}$$

Diese drei Gleichungen folgen direkt aus der Definition der OLS-Schätzer $\hat{\beta}_0$ und $\hat{\beta}_1$.

zu 2.

Stichprobenkovarianz von x und $\hat{u}$:

$$\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})(\hat{u}_i-\bar{\hat{u}}) \qquad = \frac{1}{n}\underbrace{\sum_{i=1}^{n}\hat{u}_i}_{=0} \Rightarrow \bar{\hat{u}}=0$$

$$= \frac{1}{n}\sum_{i=1}^{n}\left(x_i\cdot\hat{u}_i - \bar{x}\cdot\hat{u}_i - x_i\underbrace{\bar{\hat{u}}}_{=0} + \bar{x}\cdot\underbrace{\bar{\hat{u}}}_{=0}\right)$$

$$= \frac{1}{n}\sum_{i=1}^{n}x_i\,\hat{u}_i - \frac{1}{n}\sum_{i=1}^{n}\bar{x}\,\hat{u}_i \quad = \frac{1}{n}\sum_{i=1}^{n}x_i\cdot\hat{u}_i - \bar{x}\cdot\frac{1}{n}\underbrace{\sum_{i=1}^{n}\hat{u}_i}_{=0}$$

$$= \frac{1}{n}\sum_{i=1}^{n}x_i\,\hat{u}_i$$

OLS-Zerlegung

$\hat{u}_i = y_i - \hat{y}_i$

Wir können die Beobachtungen y_i in einen erklärten und einen unerklärten Teil zerlegen:

$$y_i = \hat{y}_i + \hat{u}_i = \cancel{\hat{y}_i} + y_i - \cancel{\hat{y}_i} \quad \checkmark$$

Aus den Eigenschaften

$$\sum_{i=1}^{n} \hat{u}_i = 0 \text{ und } \sum_{i=1}^{n} x_i \hat{u}_i = 0$$

folgt, dass der erklärte Teil $\hat{y}_i$ und der unerklärte Teil $\hat{u}_i$ unkorreliert sind:

$$\sum_{i=1}^{n} (\hat{y}_i - \bar{y})(\hat{u}_i - \bar{\hat{u}}) = \sum_{i=1}^{n} \hat{y}_i \hat{u}_i = \hat{\beta}_0 \sum_{i=1}^{n} \hat{u}_i + \hat{\beta}_1 \sum_{i=1}^{n} x_i \hat{u}_i = 0$$

OLS-Zerlegung

Ebenso können wir die Stichprobenvarianz der Beobachtungen in einen erklärten und in einen unerklärten Teil zerlegen.

Total Sum of Squares

$$SST = \sum_{i=1}^{n} (y_i - \bar{y})^2$$

Explained Sum of Squares

$$SSE = \sum_{i=1}^{n} (\hat{y}_i - \bar{\hat{y}})^2$$

Residual Sum of Squares

$$SSR = \sum_{i=1}^{n} (\hat{u}_i - \bar{\hat{u}})^2$$

Es gilt:

$$SST = SSE + SSR$$

Das Bestimmheitsmaß / Coefficient of Determination

Die Güte der Anpassung einer OLS-Regression ist der erklärte Anteil der Stichprobenvarianz der erklärten Variable.

Je größer SSE im Vergleich zu SSR, desto besser die Güte der Anpassung.

Mit $SST = SSE + SSR$ definieren wir das **Bestimmtheitsmaß** R^2:

$$R^2 = SSE/SST = 1 - SSR/SST$$

Das Bestimmtheitsmaß nimmt Werte zwischen null und eins an und kann wie folgt interpretiert werden:

Der Regressor erklärt $100 \cdot R^2\%$ der Streuung des Regressanden.

Das Bestimmtheitsmaß der Beispiele

CEO-Gehälter und Eigentkapitalrentabilität

$$\widehat{salary} = 963.191 + 18.501 roe$$

$$n = 209, \; R^2 = 0.0132$$

Die Regression erklärt nur 1.3% der Streuung der Gehälter!

Wahlergebnisse und Wahlkampfausgaben

$$\widehat{voteA} = 26.81 + 0.464 shareA$$

$$n = 173, \; R^2 = 0.856$$

Der Anteil der Wahlkampfausgaben erklärt 86% der Streuung der Wahlergebnisse!

Maßeinheiten und die funktionale Form

Lineare Transformationen (Veränderung der Maßeinheiten)

Multiplikation der Variablen des linearen Regressionsmodells mit Konstanten impliziert eine entsprechende Transformation der Schätzer:

Seien $\tilde{x}_i = c \cdot x_i$ und $\tilde{y}_i = d \cdot y_i$ für $c, d \in \mathbb{R}, \; c, d \neq 0, \; i = 1, \ldots, n$.

Dann gilt:

$$\tilde{y}_i = \tilde{\beta}_0 + \tilde{\beta}_1 \tilde{x}_i$$

$\Leftrightarrow$

$$dy_i = \tilde{\beta}_0 + \tilde{\beta}_1 c x_i \qquad \| :d$$

$\Leftrightarrow$

$$y_i = \underbrace{\frac{1}{d}\tilde{\beta}_0}_{\hat{\beta}_0} + \underbrace{\frac{c}{d}\tilde{\beta}_1}_{\hat{\beta}_1} x_i$$

dass

$$\tilde{\beta}_0 = d\hat{\beta}_0 \text{ und } \tilde{\beta}_1 = \frac{d}{c}\hat{\beta}_1$$

Das Bestimmtheitsmaß R^2 ist invariant!

Die Semi-logarithmische Form

$$\frac{w_1}{w_1} \quad \ln(w_2) - \ln(w_1) = \cancel{\beta_0} + \beta_1 e_2 + u_2 - \cancel{\beta_0} - \beta_1 e_1 - u_1$$

$$\ln\left(1+\frac{w_2-w_1}{w_1}\right) = \ln\left(1+\frac{w_2}{w_1}-1\right) = \ln\left(\frac{w_2}{w_1}\right) \quad = \beta_1(e_2-e_1) + u_2 - u_1 = \beta_1 \cdot \Delta e + \Delta u$$

Die Semi-logarithmische Form misst die prozentuale Veränderung des Regressanden bei absoluter Veränderung des Regressors.

Das ökonometrische Modell

$$\ln(wage) = \beta_0 + \beta_1 educ + u$$

misst (falls $\Delta u = 0$):

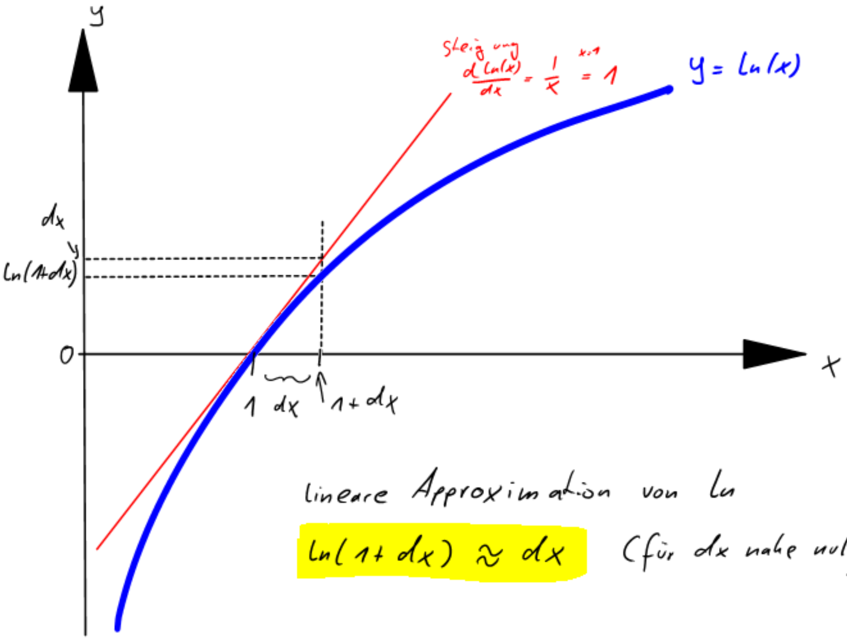
relative Veränderung von wage

$$\frac{\Delta wage}{wage} \approx \beta_1 \Delta educ$$

absolute Veränderung von educ

Eine Erhöhung der Bildung um ein Jahr geht einher mit einem Anstieg des Lohnes um $100 \cdot \beta_1$ %.

$$\underbrace{\ln\left(1+\frac{\Delta w}{w}\right)}_{\approx \frac{\Delta w}{w}} = \beta_1 \Delta e + \underbrace{\Delta u}_{=0 \text{ (c.p.)}}$$



lineare Approximation von ln

$\ln(1+dx) \approx dx$ (für dx nahe null)

Log-Stundenlohn und Bildung

Population Regression Function

$$\ln(wage) = \beta_0 + \beta_1 educ + u$$

$\ln(wage)$	:	natürlicher Logarithmus des Stundenlohns in $
$educ$	:	Bildungsjahre

Regressionsgerade:

$$\widehat{\ln(wage)} = 0.584 + \overbrace{0.083}^{\hat{\beta}_1} educ$$

$$n = 526, \;\; R^2 = 0.186$$

Im Vergleich:

Wird *wage* anstelle von $\ln(wage)$ erklärt, beträgt $R^2 = 0,165$.

Die Log-Logarithmische Form

Die Log-logarithmische Form misst die prozentuale Veränderung des Regressanden bei prozentualer Veränderung des Regressanden.

Im ökonometrische Modell

$salary = e^{\beta_0} \cdot sales^{\beta_1} \cdot e^{u}$

$$\ln(salary) = \beta_0 + \beta_1 \ln(sales) + u$$

ist β_1 die Elastizität von *salary* in Bezug auf *sales*:

Wenn sich *sales* um 1% ändert, ändert sich *salary* um β_1%.

Die OLS-Regression ergibt

$$\ln(\widehat{salary}) = 4.822 + 0.257 \ln(sales)$$

$$n = 209, \;\; R^2 = 0.211$$

Diese Form wird auch **Modell mit konstanten Elastizitäten** genannt.

Nichtlineare Transformationen

"absolute Veränderung"

Zusammenfassung:

"relative Veränderung"

Modell	Regressand	Regressor	Interpretation β_1
Level-Level	y	x	$\Delta y = \beta_1 \Delta x$
Level-Log	y	$\ln(x)$	$\Delta y \approx \beta_1 \frac{\Delta x}{x}$
Log-Level	$\ln(y)$	x	$\frac{\Delta y}{y} \approx \beta_1 \Delta x$
Log-Log	$\ln(y)$	$\ln(x)$	$\frac{\Delta y}{y} \approx \beta_1 \frac{\Delta x}{x}$

Erwartungswerte und Varianzen der OLS-Schätzer

Die Schätzer $\hat{\beta}_0$ und $\hat{\beta}_1$ des einfachen linearen Regressionsmodells sind Funktionen der Daten der Stichprobe.

Wenn wir die Stichprobe als Zufallsgröße betrachten, so sind auch $\hat{\beta}_0$ und $\hat{\beta}_1$ Zufallsvariablen und haben statistische Eigenschaften.

Um diese Eigenschaften herzuleiten, benötigen wir einige Annahmen.

Die Abkürzung *SLR* steht hierbei für „simple linear regression“.

Annahme SLR 1

Linearität in den Parametern

Im ökonometrischen Modell steht die abhängige Variable y in Relation zu der unabhängigen Variable x und dem Störterm u:

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

wobei β_0 und β_1 unbekannte aber feste Parameter sind.

Bemerkungen:

- Erlaubt, dass wir die Variablen nichtlinear transformieren.
- Fordert additive Separabilität.

Annahme SLR 3

Variation der erklärenden Variable

Die Stichprobenwerte $x_1, x_2, \ldots, x_n$ sind nicht alle identisch.

Bemerkung:

Mit SLR 3 folgt: $\overline{xx} - \bar{x}\bar{x} > 0$

Annahme SLR 4

Bedingter Nullerwartungswert

Bei gegebenen Werten x beträgt der bedingte Erwartungswert des Störterms $u_i = y_i - \beta_0 - \beta_1 x_i$, $i = 1, \ldots, n$, null:

$$E[u_i|x] = 0 \ , \ i = 1, \ldots, n$$

Bemerkung:

Mit SLR 1 folgt: $E[y_i|x] = \beta_0 + \beta_1 x_i, \ i = 1, \ldots, n.$

Erwartungstreue

Die beiden Schätzer

$$\begin{aligned}\hat{\beta}_0 &= \bar{y} - \frac{\overline{xy} - \bar{x}\bar{y}}{\overline{xx} - \bar{x}\bar{x}}\bar{x} \\ \hat{\beta}_1 &= \frac{\overline{xy} - \bar{x}\bar{y}}{\overline{xx} - \bar{x}\bar{x}}\end{aligned}$$

sind erwartungstreu:

Theorem 2.1

Unter den Annahmen SLR 1, SLR 3 und SLR 4 gilt

$$E[\hat{\beta}_0] = \beta_0 \text{ und } E[\hat{\beta}_1] = \beta_1$$

für beliebige Werte β_0 und β_1.

Beweis Theorem 2.1 für $E[\hat{\beta}_1] = \beta_1$

$$
\begin{aligned}
E[\hat{\beta}_1|x] &\overset{SLR\ 3}{=} E\left[\left.\frac{\sum_{i=1}^n (x_i - \bar{x}) y_i}{\sum_{i=1}^n (x_i - \bar{x})^2}\right| x\right] \\
&= \frac{\sum_{i=1}^n (x_i - \bar{x}) E[y_i|x]}{\sum_{i=1}^n (x_i - \bar{x})^2} \\
&\overset{SLR\ 1,4}{=} \frac{\sum_{i=1}^n (x_i - \bar{x})(\beta_0 + \beta_1 x_i)}{\sum_{i=1}^n (x_i - \bar{x})^2} \\
&= \beta_0 \frac{\sum_{i=1}^n (x_i - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2} + \beta_1 \frac{\sum_{i=1}^n (x_i - \bar{x}) x_i}{\sum_{i=1}^n (x_i - \bar{x})^2} = \beta_1
\end{aligned}
$$

Mit der Teleskopeigenschaft für Erwartungswerte gilt

$$
E[\hat{\beta}_1] = E[E[\hat{\beta}_1|x]] = \beta_1
$$

Beweis Theorem 2.1 für $E[\hat{\beta}_0] = \beta_0$

$$\begin{aligned}
E[\hat{\beta}_0|x] &= E\left[\bar{y} - \hat{\beta}_1\bar{x}\middle| x\right] \\
&= E[\bar{y}|x] - E[\hat{\beta}_1|x]\bar{x} \\
&= \frac{1}{n}\sum_{i=1}^{n} E[y_i|x] - \beta_1\bar{x} \\
&\overset{SLR\ 1,4}{=} \beta_0 + \beta_1\bar{x} - \beta_1\bar{x} = \beta_0
\end{aligned}$$

Mit der Teleskopeigenschaft für Erwartungswerte gilt

$$E[\hat{\beta}_0] = E[E[\hat{\beta}_0|x]] = \beta_0$$

Annahmen

Annahme SLR 5

Homoskedastizität

Für beliebige gegebene Regressordaten x ist die bedingte Varianz des Störterms u_i, $i = 1, \ldots, n$ identisch:

$$Var(u_i|x) = \sigma^2 > 0 \ , \ i = 1, \ldots, n$$

Bemerkungen:

Da $Var(u_i|x)$ nicht von x abhängt, gilt außerdem $Var(u_i) = \sigma^2$.

Ist SLR 5 verletzt, heißt der Störterm **heteroskedastisch**.

Annahme SLR 5b

Wir fordern zusätzlich: $Cov(u_i, u_j|x) = 0$ für $i, j = 1, \ldots, n$, $i \neq j$

Definition Varianz einer Zufallsvariablen u_i

$$Var(u_i) = E[(u_i - E[u_i])^2]$$

Falls $E[u_i] = 0 \Rightarrow Var(u_i) = E[u_i^2]$

$$= E[u_i^2 - 2u_i E[u_i] + E[u_i]^2]$$

$$= E[u_i^2] - 2E[u_i]E[u_i] + E[u_i]^2$$

$$= E[u_i^2] - E[u_i]^2$$

Rechenregel

$$Var(a + b u_i) = E[(a + b u_i - E[a + b u_i])^2]$$

$$= E\left[\left(\cancel{a} + b v_i - \cancel{a} - b E[v_i]\right)^2\right]$$

$$= E\left[\left(b v_i - b E[v_i]\right)^2\right]$$

$$= E\left[b^2 \left(v_i - E[v_i]\right)^2\right]$$

$$= b^2 \underbrace{E\left[\left(v_i - E[v_i]\right)^2\right]}_{Var(v_i)}$$

$$Var(a + b v_i) = b^2 Var(v_i)$$

Definition Kovarianz zweier Zufallsvariablen U, V

$$\text{Cov}(U,V) = E[(U-E[U])(V-E[V])]$$

$$= \dots$$

$$= E[U \cdot V] - E[U] \cdot E[V]$$

Rechenregel

$$\text{Cov}(a+bU, c+dV) = b \cdot d \cdot \text{Cov}(U,V)$$

$$\text{Var}(a+bU+c+dV)$$

Zwischenrechnung: $E[a+b \cdot U+c+d \cdot V] = a+bE[U]+c+d \cdot E[V]$

$$= E[(a+bU+c+dV - E[a+bU+c+dV])^2]$$

$$= E[(\cancel{a}+b \cdot U+\cancel{c}+dV-\cancel{a}-bE[U]-\cancel{c}-dE[V])^2]$$

$$= E\left[\left(\underline{b} \cdot u - \underline{b} E[u] + d v - d E[v] \right)^2 \right]$$

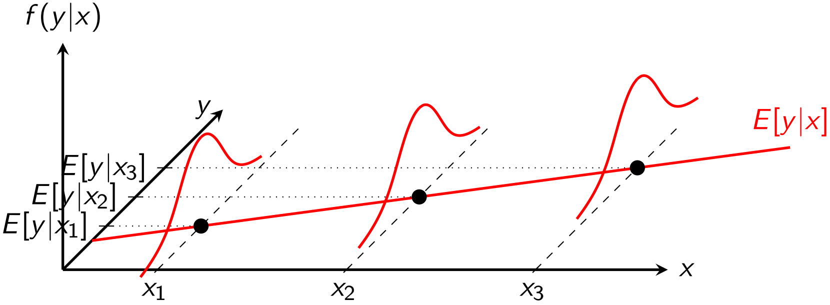
$$= E\left[\left(b(u - E[u]) + d(v - E[v]) \right)^2 \right]$$

$$= E\left[(b \cdot (u - E[u]))^2 + 2 b (u - E[u]) d (v - E[v]) + (d \cdot (v - E[v]))^2 \right]$$

$$= b^2 E\left[(u - E[u])^2 \right] + 2bd E\left[(u - E[u])(v - E[v]) \right] + d^2 E\left[(v - E[v])^2 \right]$$

$$= \boxed{ \begin{aligned} & b^2 Var(u) + 2bd Cor(u, v) + d^2 Var(v) \\ = \; & Var(a + b \cdot u + c + d \cdot v) \end{aligned} }$$

Bedingte Verteilung von y unter Homoskedastizität



Aus $Var(u_i|x) = \sigma^2$ für $i = 1, \ldots, n$ und SLR 1 folgt

$$Var(y_i|x) = Var(\beta_0 + \beta_1 x_i + u_i|x_i) = Var(u_i|x_i) = \sigma^2 \text{ für } i = 1, \ldots, n$$

Theorem 2.2

Stichprobenvarianzen der OLS-Schätzer

Unter den Annahmen SLR 1, SLR 3, SLR 4 und SLR 5+b gilt:

$$Var(\hat{\beta}_0|x) = \sigma^2 \frac{1}{n} \frac{\overline{xx}}{\overline{xx} - \bar{x}\bar{x}}$$

$$Var(\hat{\beta}_1|x) = \sigma^2 \frac{1}{n} \frac{1}{\overline{xx} - \bar{x}\bar{x}}$$

Beweis Theorem 2.2 für $Var(\hat{\beta}_1|x)$

$$
\begin{aligned}
Var(\hat{\beta}_1|x) &\overset{SLR\ 3}{=} Var\left(\frac{\sum_{i=1}^n (x_i - \bar{x})y_i}{\sum_{i=1}^n (x_i - \bar{x})^2}\right) \\
&\overset{SLR\ 1}{=} Var\left(\frac{\sum_{i=1}^n (x_i - \bar{x})(\beta_0 + \beta_1 x_i)}{\sum_{i=1}^n (x_i - \bar{x})^2} + \frac{\sum_{i=1}^n (x_i - \bar{x})u_i}{\sum_{i=1}^n (x_i - \bar{x})^2}\right) \\
&\overset{SLR\ 2,5b}{=} \frac{\sum_{i=1}^n (x_i - \bar{x})^2 Var(u_i|x)}{\left(\sum_{i=1}^n (x_i - \bar{x})^2\right)^2} \\
&\overset{SLR\ 5}{=} \frac{\sum_{i=1}^n (x_i - \bar{x})^2 \sigma^2}{\left(\sum_{i=1}^n (x_i - \bar{x})^2\right)^2} \\
&= \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2} = \frac{1}{n}\sigma^2 \frac{1}{\overline{xx} - \bar{x}\bar{x}}
\end{aligned}
$$

Je stärker der Regressor streut und je größer die Stichprobe ist, desto genauer ist $\hat{\beta}_1$.

Beweis Theorem 2.2 für $Var(\hat{\beta}_0|x)$ Schritt 1

$$
\begin{aligned}
Var(\bar{y}|x) &\overset{SLR\ 1}{=} Var\left(\frac{1}{n}\sum_{i=1}^{n}\beta_0+\beta_1 x_i+u_i\middle| x\right)\\
&= Var\left(\frac{1}{n}\sum_{i=1}^{n}u_i\middle| x\right)\\
&\overset{SLR\ 5b}{=} \frac{1}{n^2}\sum_{i=1}^{n}Var(u_i|x)\\
&\overset{SLR\ 5}{=} \frac{1}{n}\sigma^2
\end{aligned}
$$

Beweis Theorem 2.2 für $Var(\hat{\beta}_0|x)$ Schritt 2

$$
\begin{aligned}
Cov(\bar{y}, \hat{\beta}_1|x) \quad & \overset{SLR\ 1}{=} \quad Cov\left(\frac{1}{n}\sum_{i=1}^{n} y_i, \frac{\sum_{i=1}^{n}(x_i-\bar{x})y_i}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\middle| x\right) \\
& = \quad \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{x_j-\bar{x}}{\sum_{k=1}^{n}(x_k-\bar{x})^2} Cov(y_i, y_j|x) \\
& \overset{SLR\ 5b}{=} \quad \frac{1}{n}\sum_{i=1}^{n}\frac{x_i-\bar{x}}{\sum_{k=1}^{n}(x_k-\bar{x})^2} Cov(y_i, y_i|x) \\
& \overset{SLR\ 5}{=} \quad \sigma^2 \frac{1}{n}\frac{\sum_{i=1}^{n} x_i-\bar{x}}{\sum_{i=1}^{n}(x_i-\bar{x})^2} = 0
\end{aligned}
$$

Beweis Theorem 2.2 für $Var(\hat{\beta}_0|x)$ Schritt 3

$$\begin{aligned}
Var(\hat{\beta}_0|x) &= Var(\bar{y} - \hat{\beta}_1\bar{x}|x) \\
&= Var(\bar{y}|x) - 2Cov(\bar{y}, \hat{\beta}_1|x)\bar{x} + Var(\hat{\beta}_1|x)\bar{x}\bar{x} \\
&= \frac{1}{n}\sigma^2 - 2 \cdot 0 + \frac{1}{n}\sigma^2 \frac{1}{\overline{xx} - \bar{x}\bar{x}}\bar{x}\bar{x} \\
&= \frac{1}{n}\sigma^2 \frac{\overline{xx}}{\overline{xx} - \bar{x}\bar{x}}
\end{aligned}$$

Schätzung der Varianz des Störterms

Bei einer gegebenen Stichprobe sind die Ausdrücke

$$\overline{xx} = \frac{1}{n}\sum_{i=1}^{n} x_i^2$$

und

$$\bar{x}\bar{x} = \left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)^2$$

bekannt.

Bei σ^2 handelt es sich aber um einen unbekannten Parameter, der zu schätzen ist.

Theorem 2.3

Unverzerrter Schätzer für σ^2

Sei der Schätzer für σ^2 definiert durch:

$$\hat{\sigma}^2 = \frac{1}{n-2} \sum_{i=1}^{n} \hat{u}_i^2$$

Unter den Annahmen SLR 1, SLR 3, SLR 4 und SLR 5+b gilt:

$$E[\hat{\sigma}^2 | x] = \sigma^2$$

Begründung für Theorem 2.3

Der Beweis ist etwas aufwendiger und wird später nachgeholt.

Warum gilt nicht $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} \hat{u}_i^2$?

Durch die beiden Gleichungen

$$\sum_{i=1}^{n} \hat{u}_i = 0$$

$$\sum_{i=1}^{n} x_i \hat{u}_i = 0$$

werden zwei Freiheitsgrade „verbraucht“.

Daher teilen wir die Summe der quadrierten Residuen durch $n-2$.

Schätzer für die Varianzen der Schätzer

Mit dem Schätzer

$$\hat{\sigma}^2 = \frac{1}{n-2}\sum_{i=1}^{n}\hat{u}_i^2$$

ergibt sich

$$\widehat{Var}(\hat{\beta}_0|x) = \hat{\sigma}^2\frac{1}{n}\frac{\overline{xx}}{\overline{xx}-\bar{x}\bar{x}}$$

$$\widehat{Var}(\hat{\beta}_1|x) = \hat{\sigma}^2\frac{1}{n}\frac{1}{\overline{xx}-\bar{x}\bar{x}}$$

Mit den Standardfehlern der Schätzer der Regressionskoeffizienten

$$se(\hat{\beta}_0) = \sqrt{\widehat{Var}(\hat{\beta}_0|x)} \text{ und } se(\hat{\beta}_1) = \sqrt{\widehat{Var}(\hat{\beta}_1|x)}$$

können wir später Intervallschätzungen für die Regressionskoeffizienten vornehmen.

Begriffe aus Kapitel 2

- einfaches Regressionsmodell
- Regressor, Regressand, Störterm
- strikte Exogenität
- Population Regression Funktion (PRF)
- gewöhnliche Kleinste Quadrate Schätzer
- Momentenmethode, orthogonale Projektion
- Prognose und Residuen
- Stichproben(ko)varianz, Korrelationskoeffizient, Bestimmtheitsmaß
- Erwartungstreue
- Homoskedastizität, Heteroskedastizität