Übungsblatt zu Kapitel 2: Das einfache Regressionsmodell

Aufgabe 1

groß: ZVen

klein: Werte der ZVen

Seien X und Y zwei Zufallsvariablen mit der gemeinsamen Dichtefunktion $f_{X,Y}(x,y)$. Die einzelnen Dichtefunktionen seien durch

$$f_X(x) = \int_y f_{X,Y}(x,y)dy \text{ und } f_Y(y) = \int_x f_{X,Y}(x,y)dx$$

gegeben.

Zeige, dass

$$\begin{aligned} E[X+Y] &= \int_x \int_y f_{X,Y}(x,y)(x+y)dydx \\ &= \int_x f_X(x)xdx + \int_y f_Y(y)ydy = E[X] + E[Y] \end{aligned}$$

gilt.

$$E_{X,Y}[X+Y] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x+y) f_{X,Y}(x,y)\, dy\, dx$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x f_{X,Y}(x,y)\, dy\, dx + \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y f_{X,Y}(x,y)\, dy\, dx$$

$$= \int_{-\infty}^{\infty} x \underbrace{\int_{-\infty}^{\infty} f_{X,Y}(x,y)\, dy}_{= f_X(x)}\, dx + \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y f_{X,Y}(x,y)\, dx\, dy$$

$$= \underbrace{\int_{-\infty}^{\infty} x f_X(x)\, dx}_{E_X[X]} + \int_{-\infty}^{\infty} y \underbrace{\int_{-\infty}^{\infty} f_{X,Y}(x,y)\, dx}_{f_Y(y)}\, dy$$

$$= E[X] + \underbrace{\int_{-\infty}^{\infty} y \, f_Y(y) \, dy}_{E[Y]}$$

Also

$$E[X+Y] = E[X] + E[Y]$$

Aufgabe 2

Seien X und Y zwei Zufallsvariablen mit der gemeinsamen Dichtefunktion $f_{X,Y}(x,y)$. Die bedingte Dichte von Y gegeben $X = x$ ist definiert durch:

$$f_{Y|X}(y|x) = f_{X,Y}(x,y)/f_X(x) \text{ (falls } f_X(x) > 0)$$

Zeige, dass

Teleskop-Eigenschaft

$$E[E[Y|X = x]] = E[Y]$$

$$E[Y|X] = \int_{-\infty}^{\infty} y \, f_{Y|X}(y|x) \, dy$$

$$E_x[\,E_y[\,Y \mid X\,]\,] = \int_{-\infty}^{\infty} E_y[Y \mid X]\, f_X(x)\, dx$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y\, f_{Y|X}(y \mid x)\, dy\, f_X(x)\, dx$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y\, \frac{f_{X,Y}(x,y)}{f_X(x)}\, dy\, f_X(x)\, dx$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y\, f_{X,Y}(x,y)\, dy\, dx$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y\, f_{X,Y}(x,y)\, dx\, dy$$

$$= \int_{-\infty}^{\infty} y \underbrace{\int_{-\infty}^{\infty} f_{X,Y}(x,y)\, dx}_{f_Y(y)}\, dy$$

$$= \int_{-\infty}^{\infty} y\, f_Y(y)\, dy = E_Y[Y]$$

$$E[E[Y|X]] = E[Y]$$

Modell

Annahme strikte Exogenität

$$Y_i = \beta_0 + \beta_1 x_i + u_i \quad i = 1, \dots, n$$

$$E[U|X] = E[U] \; (=0)$$

Momente Methode

1. Ergebnis: Schätzer $\hat{\beta}_0, \hat{\beta}_1$ für β_0, β_1 erfüllen

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

Herleitung der 2. Gleichung:

$$\boxed{\underbrace{E[X \cdot U]}_{=0}} = E_x\left[E_u[X \cdot U|X]\right] = E_x\left[x \underbrace{E_u[U|X]}_{=0}\right]$$

$$= E[X \cdot 0] = 0$$

Schätze nun theoretische Moment $E[x \cdot u] = 0$

durch empirisches Moment $\frac{1}{n} \sum_{i=1}^{n} x_i \cdot u_i$

mit $y_i = \beta_0 + \beta_1 x_i + u_i$ folgt: $u_i = y_i - \beta_0 - \beta_1 x_i$

Schätzer $\hat{\beta}_0, \hat{\beta}_1$ für β_0, β_1:

$$\hat{u}_i = y_i - (\underbrace{\hat{\beta}_0 + \hat{\beta}_1 x_i}_{\hat{y}_i \text{ Prognose}}) \quad \text{Residuum}$$

$$\frac{1}{n} \sum_{i=1}^{n} x_i \hat{u}_i = \frac{1}{n} \sum_{i=1}^{n} x_i (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) \overset{!}{=} 0 = E[x \cdot u]$$

$$\Leftrightarrow \frac{1}{n} \sum_{i=1}^{n} (x_i y_i - x_i \hat{\beta}_0 - x_i \hat{\beta}_1 x_i) = 0$$

$$\Leftrightarrow \underbrace{\frac{1}{n} \sum_{i=1}^{n} x_i y_i}_{\overline{xy}} - \frac{1}{n} \sum_{i=1}^{n} x_i \hat{\beta}_0 - \frac{1}{n} \sum_{i=1}^{n} \hat{\beta}_1 x_i^2 = 0$$

$$\Leftrightarrow \overline{xy} - \hat{\beta}_0 \underbrace{\frac{1}{n} \sum_{i=1}^{n} x_i}_{\overline{x}} - \hat{\beta}_1 \underbrace{\frac{1}{n} \sum_{i=1}^{n} x_i^2}_{\overline{xx}} = 0$$

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

$$\overline{y} = \hat{\beta}_0 + \hat{\beta}_1 \overline{x} \quad \Leftrightarrow \quad \hat{\beta}_0 = \overline{y} - \hat{\beta}_1 \overline{x}$$

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

$$\overline{xy} = (\overline{y} - \hat{\beta}_1 \overline{x})\, \overline{x} + \hat{\beta}_1 \overline{xx}$$

$$= \overline{y} \cdot \overline{x} + \hat{\beta}_1 (\overline{xx} - \overline{x}\,\overline{x})$$

falls $\overline{xx} \neq \overline{x} \cdot \overline{x}$

$$\Rightarrow \quad \hat{\beta}_1 = \frac{\overline{xy} - \overline{x} \cdot \overline{y}}{\overline{xx} - \overline{x} \cdot \overline{x}}$$

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

Satz von Steiner $\vdots$

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

Stichprobenkovarianz von x & y

S_{xy}

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

Stichprobenvarianz von x

S_{xx}

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

Aufgabe 3

Gegeben seien folgende Daten von acht Personen aus dem College:

Person	1	2	3	4	5	6	7	8
GPA	2,8	3,4	3,0	3,5	3,6	3,0	2,7	3,7
ACT	21	24	26	27	29	25	25	30

Betrachtet sei das Regressionsmodell

$$GPA_i = \beta_0 + \beta_1 ACT_i + u_i$$

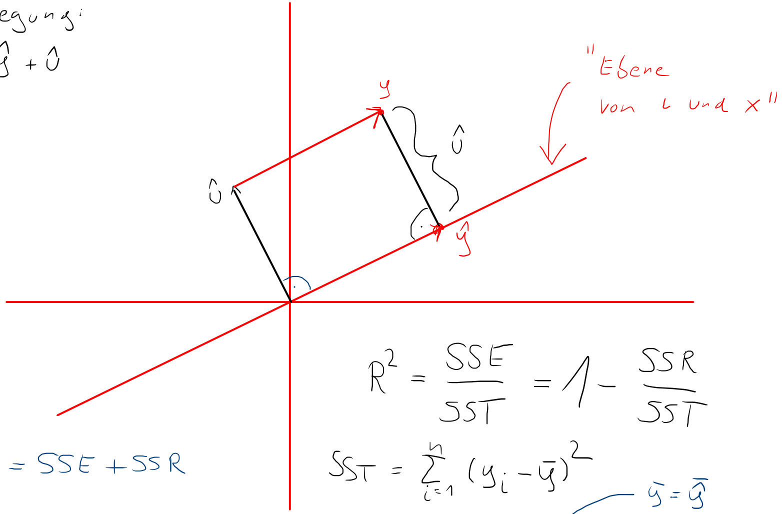
Benutze ein Tabellenkalkulationsprogramm und berechne

- $\hat{\beta}_0$ und $\hat{\beta}_1$ Siehe Aufgabe 3_LS6.xlsx
- $\hat{y}_i$ und $\hat{u}_i$
- R^2

Benutze nun Gretl und bestätige Deine Berechnungen

Zerlegung:

$y = \hat{y} + \hat{u}$



$$R^2 = \frac{SSE}{SST} = 1 - \frac{SSR}{SST}$$

$SST = SSE + SSR$

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

$\bar{y} = \bar{\hat{y}}$

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

$\bar{\hat{u}} = 0$

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

Aufgabe 4

Wir hatten in der Vorlesung hergeleitet:

$$\hat\beta_0 = \bar y - \underbrace{\frac{\overline{xy} - \bar x \bar y}{\overline{xx} - \bar x \bar x}}_{\hat\beta_1} \bar x \text{ und } \hat\beta_1 = \frac{\overline{xy} - \bar x \bar y}{\overline{xx} - \bar x \bar x}$$

Definiere die prognostizierten Werte durch

$$\hat y_i = \hat\beta_0 + \hat\beta_1 x_i = \bar y - \hat\beta_1 \bar x + \hat\beta_1 \cdot x_i$$
$$= \bar y - \hat\beta_1 (\bar x - x_i)$$

Zeige nun, dass $\sum_{i=1}^n \hat y_i = \sum_{i=1}^n y_i$.

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

$$= n \cdot \bar y - \hat\beta_1 \underbrace{\sum_{i=1}^n (\bar x - x_i)}_{=0}$$

$$= \cancel{n} \cdot \frac{1}{\cancel{n}} \sum_{i=1}^n y_i - \hat\beta_1 \cdot 0$$

$$= \sum_{i=1}^n y_i \quad \square$$

Aufgabe 5

Führe diese Aufgabe mit Gretl (oder anderer Statistik-Software) durch.

Bestätige für den Datensatz *ceosal1.xls* die Schätzung

$$\widehat{salary} = 963.191 + 18.501 roe, \;\; R^2 = 0,0132$$

für den Datensatz *wage1.xls* die Schätzung

$$\widehat{wage} = -0,90 + 0,54 \mathrm{educ}$$

und für den Datensatz *vote1.xls* die Schätzung

$$\widehat{voteA} = 26,81 + 0,464 shareA, \;\; R^2 = 0,856$$

Aufgabe 6

Analysiere für diese Aufgabe den Datensatz *ceosal2.xls*.

- Wie hoch ist das durchschnittliche Gehalt *salary*?
- Wie hoch ist die durchschnittliche Beschäftigungsdauer als Manager *ceoten*?
- Wie viele Manager sind in ihrem ersten Beschäftigungsjahr ($ceoten = 0$)?
- Welches ist die längste Beschäftigungsdauer als Manager?
- Schätze das Regressionsmodell

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

Wie sind die Ergebnisse zu interpretieren?

Aufgabe 7

Betrachte das zweimalige Werfen eines fairen Würfels. Diese beiden (unabhängigen) Zufallsvariablen werden als X_1 und X_2 bezeichnet. Definiere durch $Y := X_1 + X_2$ die Augensumme aus beiden Würfen sowie durch $Z := X_1 \cdot X_2$ das Produkt der Augensummen aus den einzelnen Würfen.

a) Berechne $E[Y]$ und $E[Z]$.

b) Berechne $E[Y|X_1]$ sowie $E[Z|X_1]$.

c) Überprüfe jeweils die Eigenschaft $E[Y] = E[E[Y|X_1]]$ bzw. $E[Z] = E[E[Z|X_1]]$.

Falls X_1, X_2 unabhängig : $E[X_1 \cdot X_2] = E[X_1] \cdot E[X_2]$

7 a) $E[Y] = E[X_1 + X_2]$

$= \underbrace{E[X_1]}_{\frac{7}{2}} + \underbrace{E[X_2]}_{\frac{7}{2}} = 7$

$E[Z] = E[X_1 \cdot X_2]$

$\overset{\text{unabh.}}{=} \underbrace{E[X_1]}_{\frac{7}{2}} \cdot \underbrace{E[X_2]}_{\frac{7}{2}} = 49/4 = 12{,}25$

b) $E[Y \mid X_1] = E[X_1 + X_2 \mid X_1]$

$= \underbrace{E[X_1 \mid X_1]}_{} + \underbrace{E[X_2 \mid X_1]}_{}$

$= X_1 + \underbrace{E[X_2]}_{\frac{7}{2}} = X_1 + \frac{7}{2}$

$$E[Z \mid X_1] = E[X_1 \cdot X_2 \mid X_1]$$

$$= X_1 \cdot E[X_2 \mid X_1]$$

$$= X_1 \cdot \underbrace{E[X_2]}_{\frac{7}{2}} = \frac{7}{2} \cdot X_1$$

c) zu zeigen $E_{X_1}[E_Y[Y \mid X_1]] = E[Y]$

$$E[E[Y \mid X_1]] = E[X_1 + \tfrac{7}{2}] = \underbrace{E[X_1]}_{\frac{7}{2}} + \frac{7}{2} = 7 = E[Y]$$

$$E_{X_1}[E_Z[Z \mid X_1]] = E[\tfrac{7}{2} X_1] = \frac{7}{2} \cdot \underbrace{E[X_1]}_{\frac{7}{2}} = \frac{49}{4} = E[Z]$$

Aufgabe 8

$\widehat{sleep} = \hat{\beta}_0 + \hat{\beta}_1 \, totwork$

Analysiere für diese Aufgabe den Datensatz *sleep75.xls* und betrachte das Modell

$$sleep = \beta_0 + \beta_1 totwrk + u$$

$totwork_i = 0$

$\Rightarrow \widehat{sleep}_i = \hat{\beta}_0$

sleep: Schlaf-Minuten pro Woche
totwrk: Arbeits-Minuten pro Woche

- Führe eine Regression durch und betrachte den Output.
- Welche Bedeutung hat der Achsenabschnitt $\hat{\beta}_0$? $= 3586 \hat{=} 8{,}5\,h/Tag$
- Wie groß ist R^2 und was bedeutet dies? $R^2 = 0{,}1$ $\quad$ $(SSR = 1{,}25e{+}08$
- Wenn *totwrk* um 2 Stunden steigt, wie verändert sich der prognostizierte Wert von *sleep*?

$= 1{,}25 \cdot 10^8$

$= \underbrace{125.xxx.xxx}_{}$

$R^2 = 0{,}1$: $\sim 90\%$ von $SST)$

$\Delta \widehat{sleep} = \hat{\beta}_1 \Delta totwork = (-0{,}15) \cdot 120 = -18$

Aufgabe 9

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

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

$S_x = \sqrt{S_{xx}}$

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

$S_x = 2{,}8449$ (Gretl)

$S_{xx} = 8{,}0935$

- Generiere eine Stichprobe von 500 Beobachtungen x_i von unabhängigen uniform verteilten Zufallsvariablen auf [0,10]. gretl: `x=randgen(u,0,10)`
- Berechne den Stichprobendurchschnitt $\bar{x}$ (5,0603) und die Stichprobenvarianz $\overline{xx}$ (33,684) $- \bar{x}\bar{x}$. $= 33{,}684 - (5{,}0603)^2 = 8{,}077$
- Generiere eine Stichprobe von 500 Fehlern u_i von unabhängigen normalverteilten Zufallsvariablen mit Erwartungswert 0 und Varianz 36. gretl: `u=randgen(n,0,6)`
- Generiere eine Stichprobe von 500 Beobachtungen y_i mit

$$y_i = 1 + 2x_i + u_i$$

- Schätze anhand OLS die Parameter β_0 und β_1 des Modells

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

Aufgabe 10

SLR3: $\overline{xx} - \bar{x}\bar{x} > 0$

SLR4: $E[u|x] = 0$

Betrachte das ökonometrische Modell

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

mit SLR 1, SLR 3 und SLR 4 und ermittle die OLS-Residuen $\hat{u}_i$, $i = 1, \ldots, n$.

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

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

Betrachte nun das Modell

neuer Störterm

$$\hat{u}_i = \alpha_0 + \alpha_1 \cdot x_i + v_i \;, \; i = 1, \ldots, n$$

wobei wieder SLR 1, SLR 3 und SLR 4 gelte.

$$\hat{\alpha}_1 = \frac{\overline{x\hat{u}} - \bar{x} \cdot \overline{\hat{u}}}{\overline{xx} - \bar{x} \cdot \bar{x}} = 0$$

$\overline{\hat{u}} = \underbrace{\frac{1}{n}\sum_{i=1}^{n} \hat{u}_i}_{=0}$

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

Wie lauten die OLS-Schätzer $\hat{\alpha}_0$ und $\hat{\alpha}_1$?

$$\hat{\alpha}_0 = \underbrace{\overline{\hat{u}}}_{=0} - \underbrace{\hat{\alpha}_1}_{=0} \cdot \bar{x} = 0$$

Wie interpretierst Du das Ergebnis?

Wir können $\hat{u}$ nicht durch x erklären.

Aufgabe 11

Betrachte das ökonometrische Modell (SLR 1)

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

a) Zeige, dass $\hat{\beta}_0 = \sum_{i=1}^{n} a_i^0 \cdot y_i$ und $\hat{\beta}_1 = \sum_{i=1}^{n} a_i^1 \cdot y_i$ gilt mit

$$a_i^0 = \frac{1}{n} - \bar{x} \cdot a_i^1 \text{ und } a_i^1 = \frac{x_i - \bar{x}}{\sum_{j=1}^{n}(x_j - \bar{x})^2}$$

b) Zeige, dass $\sum_{i=1}^{n} a_i^0 = 1$ und $\sum_{i=1}^{n} a_i^0 \cdot x_i = 0$.
Benutze hierfür die in der Vorlesung gezeigten Ergebnisse
$\sum_{i=1}^{n} a_i^1 = 0$ und $\sum_{i=1}^{n} a_i^1 \cdot x_i = 1$.

c) Zeige mit den Ergebnissen aus a) und b) und den Annahmen SLR 3 & 4, dass $\hat{\beta}_0$ erwartungstreu für β_0 und dass $\hat{\beta}_1$ erwartungstreu für β_1 ist.

Lineare Funktion

$f: \mathbb{R}^n \to \mathbb{R}$ ist linear, falls

- Homogenität $f(\lambda x_1, \dots, \lambda x_n) = \lambda f(x_1, \dots, x_n)$
- Additivität $f(x_1 + y_1, \dots, x_n + y_n) = f(x_1, \dots, x_n) + f(y_1, \dots, y_n)$

für alle $x, y \in \mathbb{R}^n$, $\lambda \in \mathbb{R}$

$\Leftrightarrow$

$$f(x_1, \dots, x_n) = a_1 x_1 + \dots + a_n x_n = \sum_{i=1}^{n} a_i x_i \quad , a_1, \dots, a_n \in \mathbb{R}$$

$$\hat{\beta}_1 = \frac{\overline{xy} - \bar{x}\cdot\bar{y}}{\overline{xx} - \bar{x}\cdot\bar{x}} = \frac{\frac{1}{n}\sum_{i=1}^{n} x_i \cdot y_i - \left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)\left(\frac{1}{n}\sum_{i=1}^{n} y_i\right)}{\frac{1}{n}\sum_{i=1}^{n} x_i \cdot x_i - \left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)}$$

(Steiner)

$$= \frac{\cancel{\frac{1}{n}}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\cancel{\frac{1}{n}}\sum_{j=1}^{n}(x_j - \bar{x})^2} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})y_i}{\boxed{\sum_{j=1}^{n}(x_j - \bar{x})^2}} = \sum_{i=1}^{n} \boxed{\frac{(x_i - \bar{x})y_i}{\sum_{j=1}^{n}(x_j - \bar{x})^2}}$$

d_i^1

Zwischenrechnung

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

$$= \sum_{i=1}^{n}(x_i - \bar{x})y_i - \bar{y}\underbrace{\sum_{i=1}^{n}(x_i - \bar{x})}_{=0}$$

$$= \sum_{i=1}^{n}(x_i - \bar{x})y_i$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = \frac{1}{n}\sum_{i=1}^{n} y_i - \bar{x} \underbrace{\sum_{i=1}^{n} a_i^1 \cdot y_i}_{\hat{\beta}_1}$$

$$= \sum_{i=1}^{n} \left(\frac{1}{n} y_i - \bar{x} \cdot a_i^1 \cdot y_i \right)$$

$$= \sum_{i=1}^{n} \overbrace{\left(\frac{1}{n} - \bar{x} \cdot a_i^1 \right)}^{a_i^0} y_i$$

$$a_i^1 = \frac{x_i - \bar{x}}{\sum_j (x_j - \bar{x})^2}$$

$$\sum_{i=1}^{n} a_i^1 = \sum_{i=1}^{n} \frac{x_i - \bar{x}}{\boxed{\sum_j (x_j - \bar{x})^2}} = \frac{1}{\sum_{j=1}^{n} (x_j - \bar{x})^2} \cdot \underbrace{\sum_{i=1}^{n} (x_i - \bar{x})}_{=0} = 0$$

$$\sum_{i=1}^{n} a_i^1 \cdot x_i = \sum_{i=1}^{n} \frac{x_i - \bar{x}}{\boxed{\sum_j (x_j - \bar{x})^2}} \cdot x_i = \frac{1}{\sum_j (x_j - \bar{x})^2} \sum_{i=1}^{n} (x_i - \bar{x}) \cdot x_i = 1$$

Zwischenergebnis $\sum_i (x_i - \bar{x})(x_i - \bar{x}) = \sum_{i=1}^{n} (x_i - \bar{x}) x_i$

$$a_i^0 = \frac{1}{n} - \bar{x} a_i^1$$

$$\sum_{i=1}^{n} a_i^0 = \sum_{i=1}^{n} \left(\frac{1}{n} - \bar{x} a_i^1 \right) = \sum_{i=1}^{n} \frac{1}{n} - \sum_{i=1}^{n} \bar{x} a_i^1$$

$$= n \cdot \frac{1}{n} - \bar{x} \underbrace{\sum_{i=1}^{n} a_i^1}_{=0} = 1$$

$$\sum_{i=1}^{n} a_i^0 \cdot x_i = \sum_{i=1}^{n} \left(\frac{1}{n} - \bar{x} a_i^1 \right) x_i = \sum_{i=1}^{n} \frac{1}{n} x_i - \sum_{i=1}^{n} \bar{x} a_i^1 \cdot x_i$$

$$= \underbrace{\frac{1}{n} \sum_{i=1}^{n} x_i}_{\bar{x}} - \bar{x} \underbrace{\sum_{i=1}^{n} a_i^1 \cdot x_i}_{=1} = 0$$

$$\hat{\beta}_0 = \sum_{i=1}^{n} a_i^0 \cdot y_i \quad \text{mit} \quad \sum_i a_i^0 = 1 \quad \sum_i a_i^0 \cdot x_i = 0$$

$$\hat{\beta}_1 = \sum_{i=1}^{n} a_i^1 y_i \quad \text{mit} \quad \sum_i a_i^1 = 0 \quad \sum_i a_i^1 x_i = 1$$

$$E[y_i | X] \qquad y_i = \beta_0 + \beta_1 x_i + u_i \quad , i = 1, \dots, n$$

$$= E[\beta_0 + \beta_1 x_i + u_i | X] = E[\beta_0 | X] + E[\beta_1 x_i | X] + \underbrace{E[u_i | X]}_{=0}$$

$$= \beta_0 + \beta_1 x_i$$

$$E[\hat{\beta}_1 | X] = E\left[\sum_{i=1}^{n} a_i^1 y_i | X\right] = \sum_{i=1}^{n} a_i^1 E[y_i | X]$$

$$= \sum_{i=1}^{n} a_i^1 (\beta_0 + \beta_1 x_i)$$

$$= \sum_{i=1}^{n} a_i^1 \beta_0 + a_i^1 \beta_1 x_i$$

$$= \sum_{i=1}^{n} a_i^1 \beta_0 + \sum_{i=1}^{n} a_i^1 \beta_1 x_i = \beta_0 \underbrace{\sum_{i=1}^{n} a_i^1}_{=0} + \beta_1 \underbrace{\sum_{i=1}^{n} a_i^1 \cdot x_i}_{=1}$$

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

$$E_x[E_y[\hat{\beta}_1 | X]] = E[\beta_1] = \beta_1 \quad \checkmark$$

$$E[\hat{\beta}_0 | X] = E\left[\sum_{i=1}^{n} a_i^0 y_i\right] = \sum_{i=1}^{n} a_i^0 E[y_i | X]$$

$$= \sum_{i=1}^{n} a_i^0 (\beta_0 + \beta_1 x_i) = \dots = \beta_0 \underbrace{\sum_{i=1}^{n} a_i^0}_{=1} + \beta_1 \underbrace{\sum_{i=1}^{n} a_i^0 x_i}_{=0}$$

$$E[\hat{\beta}_0] = E[E[\hat{\beta}_0 | X]] = E[\beta_0] = \beta_0 \quad \checkmark$$

Aufgabe 12

$Cor(u_i, u_j | x) = 0$

Benutze das Modell und die Ergebnisse aus Aufgabe 11 und zeige unter SLR 5 und SLR 5b, dass

$Var(u_i | x) = \sigma^2$

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

und

$$Cov(\hat{\beta}_0, \hat{\beta}_1|x) = -\sigma^2 \frac{1}{n} \frac{\bar{x}}{\overline{xx} - \bar{x}\bar{x}}$$

Benutze hierbei das in der Vorlesung gezeigte Ergebnis $\sum_{i=1}^{n} \left(a_i^1\right)^2 = \frac{1}{n} \frac{1}{\overline{xx} - \bar{x}\bar{x}}$.

Erstelle die Varianz-Kovarianz-Matrix des Vektors $\binom{\hat{\beta}_0}{\hat{\beta}_1}$:

$$\Sigma\left(\hat{\beta}_0, \hat{\beta}_1\right) = \begin{pmatrix} Var(\hat{\beta}_0) & Cov(\hat{\beta}_0, \beta_1) \\ Cov(\hat{\beta}_1, \beta_0) & Var(\hat{\beta}_1) \end{pmatrix}$$

Wie interpretierst Du diese Matrix?

$$\hat{\beta}_1 = \sum_{i=1}^{n} a_i^1 y_i \qquad \text{mit} \quad a_i^1 = \frac{x_i - \bar{x}}{\sum_{j=1}^{n} (x_j - \bar{x})^2}$$

$$Var(\hat{\beta}_1 | x) = Var\left(\sum_{i=1}^{n} a_i^1 y_i \,|\, x\right)$$

$$= \sum_{i=1}^{n} (a_i^1)^2 \underbrace{Var(y_i | x)}_{\sigma^2} + \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} a_i^1 \cdot a_j^1 \cdot \underbrace{Cor(y_i, y_j | x)}_{=0}$$

Zwischenrechnung $Var(y_i | x)$

$$Var(y_i | x) \overset{SLR1}{=} Var(\beta_0 + \beta_1 x_i + u_i | x) = Var(u_i | x) \overset{SLR5}{=} \sigma^2$$

ZR $\quad Cor(y_i, y_j | x) \quad i \neq j$

$$Cor(y_i, y_j | x) \overset{SLR1}{=} Cor(\underbrace{\beta_0 + \beta_1 x_i}_{\text{konstant}} + u_i, \underbrace{\beta_0 + \beta_1 x_j}_{\text{Konst}} + u_j | x)$$

$$= Cor(u_i, u_j | x) \overset{SLR56}{=} 0$$

$$\hookrightarrow \operatorname{Var}(\hat{\beta}_1 \mid x) = \sum_{i=1}^{n} (a_i^1)^2 \cdot \sigma^2 = \sigma^2 \cdot \sum_{i=1}^{n} (a_i^1)^2$$

Zwischenrechnung

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

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

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

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

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

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

$$\mathrm{Var}(\hat{\beta}_0 \mid x) = \mathrm{Var}\left(\sum_{i=1}^{n} a_i^0 \cdot y_i \mid x\right)$$

$$= \mathrm{Var}(\bar{y} - \hat{\beta}_1 \bar{x} \mid x)$$

$$= \underbrace{\mathrm{Var}(\bar{y} \mid x)}_{\frac{1}{n}\cdot\sigma^2} - 2\bar{x}\,\underbrace{\mathrm{Cor}(\bar{y}, \hat{\beta}_1 \mid x)}_{=0} + \bar{x}^2\,\underbrace{\mathrm{Var}(\hat{\beta}_1 \mid x)}_{\sigma^2 \frac{1}{n} \frac{1}{\overline{xx} - \bar{x}\bar{x}}}$$

$$\mathrm{Var}(\bar{y} \mid x) = \mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{n} y_i \mid x\right) = \left(\frac{1}{n}\right)^2 \mathrm{Var}\left(\sum_{i=1}^{n} y_i \mid x\right)$$

$$= \left(\frac{1}{n}\right)^2 \left(\sum_{i=1}^{n} \underbrace{\mathrm{Var}(y_i \mid x)}_{SLR5 = \sigma^2} + \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} \underbrace{\mathrm{Cor}(y_i, y_j \mid x)}_{SLR5_b = 0}\right)$$

$$= \left(\frac{1}{n}\right)^2 \sum_{i=1}^{n} \sigma^2 = \left(\frac{1}{n}\right)^2 \cdot n \cdot \sigma^2 = \frac{1}{n}\sigma^2$$

$$\mathrm{Cor}(\bar{y}, \hat{\beta}_1 | X) = \mathrm{Cor}\left(\frac{1}{n}\sum_{i=1}^{n} y_i \,,\, \sum_{i=1}^{n} a_i^1 \cdot y_i \,|\, X\right)$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{n} \mathrm{Cor}\left(\frac{1}{n} y_i \,,\, a_j^1 y_j \,|\, X\right)$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{1}{n} \cdot a_j^1 \, \mathrm{Cor}(y_i \,,\, y_j \,|\, X)$$

$$= \sum_{i=1}^{n} \frac{1}{n} a_i^1 \, \mathrm{Cor}(y_i \,,\, y_i \,|\, X)$$ (da $\mathrm{Cor}(y_i, y_j | X) = 0$ für alle $i \neq j$)

$$= \frac{1}{n} \sum_{i=1}^{n} a_i^1 \underbrace{\mathrm{Var}(y_i | X)}_{\sigma^2}$$

$$= \frac{1}{n} \sigma^2 \underbrace{\sum_{i=1}^{n} a_i^1}_{=0} = 0$$

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

$$= \frac{1}{n}\sigma^2 \left(\frac{\overline{xx} - \bar{x}\bar{x}}{\overline{xx} - \bar{x}\bar{x}} + \frac{\bar{x}\cdot\bar{x}}{\overline{xx} - \bar{x}\bar{x}} \right)$$

$$= \frac{1}{n}\sigma^2 \frac{\overline{xx}}{\overline{xx} - \bar{x}\bar{x}}$$

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

$$Cor(\hat{\beta}_0, \hat{\beta}_1 | X)$$

$$= Cor(\bar{y} - \bar{x}\hat{\beta}_1, \hat{\beta}_1 | X)$$

$$= \underbrace{Cor(\bar{y}, \hat{\beta}_1 | X)}_{=0} - \bar{x} \underbrace{Cor(\hat{\beta}_1, \hat{\beta}_1 | X)}_{Var(\hat{\beta}_1 | X)}$$

$$= -\bar{x}\sigma^2 \cdot \frac{1}{n} \frac{1}{\overline{xx} - \bar{x}\cdot\bar{x}}$$

Varianz-Kovarianz-Matrix von $\hat{\beta}_0, \hat{\beta}_1$

$$\Sigma(\hat{\beta}_0, \hat{\beta}_1 | x) = \begin{pmatrix} Var(\hat{\beta}_0 | x) & Cov(\hat{\beta}_0, \hat{\beta}_1) \\ Cov(\hat{\beta}_1, \hat{\beta}_0) & Var(\hat{\beta}_1 | x) \end{pmatrix}$$

$$= \begin{pmatrix} \sigma^2 \frac{1}{n} \frac{\overline{xx}}{\overline{xx} - \bar{x} \cdot \bar{x}} & -\bar{x} \sigma^2 \frac{1}{n} \frac{1}{\overline{xx} - \bar{x} \cdot \bar{x}} \\ -\bar{x} \sigma^2 \frac{1}{n} \frac{1}{\overline{xx} - \bar{x}\bar{x}} & \sigma^2 \frac{1}{n} \frac{1}{\overline{xx} - \bar{x}\bar{x}} \end{pmatrix} = \sigma^2 \frac{1}{n} \frac{1}{\overline{xx} - \bar{x}\bar{x}} \begin{pmatrix} \overline{xx} & -\bar{x} \\ -\bar{x} & 1 \end{pmatrix}$$

Ü: Berechne $\begin{pmatrix} 1 & \overline{x} \\ \overline{x} & \overline{xx} \end{pmatrix}^{-1}$