Übungsblatt zu Kapitel 2: Das einfache Regressionsmodell

Aufgabe 1

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.

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

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

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 $\hat{\beta}_1$
- $\hat{y}_i$ und $\hat{u}_i$
- R^2

Benutze nun Gretl und bestätige Deine Berechnungen

Aufgabe 4

Wir hatten in der Vorlesung hergeleitet:

$$\hat{\beta}_0 = \bar{y} - \frac{\overline{xy} - \bar{x}\bar{y}}{\overline{xx} - \bar{x}\bar{x}}\bar{x} \text{ und } \hat{\beta}_1 = \frac{\overline{xy} - \bar{x}\bar{y}}{\overline{xx} - \bar{x}\bar{x}}$$

Definiere die prognostizierten Werte durch

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

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

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

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

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. $E[Z] = E[E[Z|X_1]]$.

Aufgabe 8

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

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

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

- Führe eine Regression durch und betrachte den Output.
- Welche Bedeutung hat der Achsenabschnitt $\hat{\beta}_0$?
- Wie groß ist R^2 und was bedeutet dies?
- Wenn *totwrk* um 2 Stunden steigt, wie verändert sich der prognostizierte Wert von *sleep*?

Aufgabe 9

- 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 $\bar{x}$ und die Stichprobenvarianz $\overline{xx} - \bar{x}\bar{x}$.
- 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

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 $\hat{u}_i$, $i = 1, \ldots, n$.

Betrachte nun das Modell

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

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

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

Wie interpretierst Du das Ergebnis?

Aufgabe 11

Betrachte das ökonometrische Modell (SLR 1)

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

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.

Aufgabe 12

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

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

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?