

Übungsblatt 4: Regression I

Aufgabe 1:

- (a) Erläutern Sie kurz die Idee der linearen Einfachregression für das einfache Regressionsmodell

$$y_i = \alpha + \beta x_i + \varepsilon_i \quad .$$

Erklären sie die Begriffe erklärende und abhängige Variable und Störterm. Welche Objekte in der Gleichung sind beobachtbar und welche nicht?

- (b) Zeigen Sie, dass der KQ-Schätzer $\hat{\beta}$ gegeben ist durch

$$\hat{\beta} = \frac{c_{xy}}{s_x^2}$$

wobei c_{xy} die empirische Kovarianz von erklärender und abhängiger Variable und s_x^2 die empirische Varianz der erklärenden Variable ist.

- (c) Leiten Sie außerdem folgenden Ausdruck her:

$$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}$$

- (d) Zeigen Sie in einer geeigneten Graphik, was man unter den Residuen $e_i = y_i - \hat{\alpha} - \hat{\beta}x_i$ versteht.

- (e) Welche Restriktionen implizieren die Bedingung erster Ordnung für das arithmetische Mittel der Residuen $\frac{1}{n} \sum_{i=1}^n e_i$ und die Kovarianz von e und x , $c_{ex} = \frac{1}{n} \sum_{i=1}^n e_i x_i - \bar{e}\bar{x}$.

Aufgabe 2:

- (a) Diskutieren Sie die Annahmen, die dem einfachen Regressionsmodell zugrunde liegen.
- (b) $\hat{\alpha}$ und $\hat{\beta}$ können sowohl als Zufallsvariablen als auch als reelle Zahlen aufgefasst werden. Erläutern Sie.

- (c) Warum kann die KQ-Zielfunktion, berechnet unter der Verwendung der wahren Parameter $\sum_{i=1}^n (y_i - \alpha - \beta x_i)^2$, als auch ihr Minimum $\sum_{i=1}^n (y_i - \hat{\alpha} - \hat{\beta} x_i)^2$, als Zufallsvariable aufgefasst werden?

Aufgabe 3:

Die Gleichung,

$$WAGE_i = \alpha + \beta S_i + \varepsilon_i,$$

stellt den Zusammenhang zwischen dem Einkommen einer Person $WAGE_i$ und der Anzahl an Schuljahren S_i dar. Erläutern Sie die Annahmen des einfachen Regressionsmodells an diesem Beispiel. Sind die Annahmen Ihrer Meinung nach erfüllt?

Aufgabe 4:

Gegeben ist das einfache lineare Regressionsmodell

$$y_i = \alpha + \beta x_i + \varepsilon_i \quad .$$

Ihnen steht der folgende Datensatz zur Verfügung:

x	1	2	3	4	5
y	4	6	7	9	14

- (a) Berechnen Sie die KQ-Schätzer $\hat{\alpha}$ und $\hat{\beta}$.
- (b) Wie verändert sich das Regressionsmodell, wenn $\alpha = 0$? Was bedeutet dies für eine graphische Darstellung der Regressionsgerade? Leiten Sie unter dieser Einschränkung den KQ-Schätzer für $\hat{\beta}$ ab und berechnen Sie den konkreten Wert.
- (c) Werechen Sie $\hat{\alpha}$, wenn $\beta = 0$? Gehen sie vor wie in Aufgabenteil (b).
- (d) Zeichnen Sie die Datenpunkte und die Regressionsgerade aus (a), (b) und (c) in ein Koordinatensystem ein. Prüfen Sie, ob der Punkt (\bar{x}, \bar{y}) auf der Regressionsgeraden liegt. Warum ist dies der Fall? Trifft es auf für die Regressionsgerade aus Aufgabenteil (b) zu?
- (e) Zeichnen Sie die Residuen e_i und die quadrierten Residuen e_i^2 in die Graphik aus Aufgabenteil (d) ein.

Aufgabe 5:

Bei der Analyse eines linearen Zusammenhangs zwischen der abhängigen Variablen y und der erklärenden Variablen x lieferte eine Stichprobe vom Umfang $n = 250$ die Werte:

$$\bar{x} = 60, \bar{y} = 75, \sum(x_i - \bar{x})^2 = 6300, \sum(y_i - \bar{y})^2 = 8150, \sum(x_i - \bar{x})(y_i - \bar{y}) = 6175$$

- (a) Ermitteln Sie die KQ-Schätzer für die Regressionsgerade.
- (b) Ermitteln Sie den KQ-Schätzer für die Regressionsgerade, falls $\beta = 0$ gilt.

Aufgabe 6:

Das amerikanische Forbes Magazine ist eines der erfolgreichsten Wirtschaftsmagazine weltweit. Es ist insbesondere für die Ranglisten zu unterschiedlichen Themen bekannt. Neben der Liste der reichsten Menschen der Welt (The World's Billionaires) veröffentlicht das Magazin auch eine Liste der „America's Best Small Companies“. Für die 60 Unternehmen dieser Liste mit dem höchsten Umsatz wurden die Variablen Alter und Gehalt des CEO (Chief Executive Officer) in tausend US-Dollar gemessen. Die Daten finden Sie auf der Kurshomepage.

- (a) Berechnen Sie die Regressionsgerade mit Hilfe eines Tabellenkalkulationsprogramms.
- (b) Zeichnen Sie sowohl die Daten, als auch die Regressionsgerade in ein Diagramm. Erscheint eine Regression auf Basis dieser Daten sinnvoll?
- (c) Zeichnen Sie außerdem die Residuen in ein neues Diagramm und beschreiben Sie den Plot. Was gilt für die Summe aller Residuen $\sum e_i$?