Lineare Modelle – SoSe 2025 **Übungsblatt 1**

Sabine Hoffmann, Christina Sauer, Cornelia Gruber

Aufgabe 1: Lineare Einfachregression

Der auf der Moodle-Homepage verfügbare Datensatz `RunningAgg.Rds` enthält Informationen zu einstündigen Laufeinheiten von 345 Hobbyläufern. Für die einzelnen Läufer sind die durchschnittliche Pace) (`pace`, Kehrwert der Geschwindigkeit) in min/km sowie die zugehörige durchschnittliche Herzfrequenz (`HR`) in bpm (beats per minute) gegeben.

(a) Lesen Sie den Datensatz in **R** ein und visualisieren Sie den Zusammenhang zwischen der Laufgeschwindigkeit (`pace`) und der Herzfrequenz (`HR`). Passen sie ein Lineares Modell für diesen Zusammenhang an und interpretieren Sie die Parameterschätzer $\hat{\beta}_0$ und $\hat{\beta}_1$.

(b) Warum könnte es sinnvoller sein, den Zusammenhang zwischen der Geschwindigkeit (in km/h) und der Herzfrequenz zu untersuchen? Passen Sie das Modell mit der neuen Einflussgröße erneut an. Vergleichen Sie die Anpassung der beiden Modelle.

(c) Leiten Sie die Umrechnungsformel für die Kleinste-Quadrate-Schätzer nach einer linearen Variablentransformation der folgenden Form her:

$$
\begin{aligned}
x_i \to t_i &= a_0 + a_1 x_i, \quad \text{mit } a_1 \neq 0 \\
y_i \to u_i &= b_0 + b_1 y_i, \quad \text{mit } b_1 \neq 0
\end{aligned}
$$

(d) Wie ändern sich die Parameterschätzungen, wenn die Geschwindigkeit in Meilen pro Stunde ($1\ mi = 1.61\ km$) und die Herzfrequenz in Schläge pro Sekunde (bps) umgerechnet werden? Berechnen Sie hierzu die neuen Schätzungen anhand der Ergebnisse aus Teilaufgabe (b). Hätten hierfür auch die Ergebnisse aus Teilaufgabe (a) verwendet werden können?

(e) Betrachten Sie nun erneut das Modell aus Teilaufgabe (b). Zentrieren Sie die Einflussvariable `speed` und passen Sie das Modell mit der zentrierten Variable neu an. Interpretieren Sie die geschätzten Parameter.

(f) Berechnen Sie die Parameterschätzer nun direkt aus den Parameterschätzern des Modells aus Teilaufgabe (c). Vergleichen Sie dazu die geschätzten Parameter aus Teilaufgabe (a).

(g) Betrachten Sie nun das Regressionsmodell $\widetilde{HR}_i = \gamma_0 + \gamma_1 \ \widetilde{speed}_i \ + \varepsilon_i$, wobei $\widetilde{HR}$ und $\widetilde{speed}$ jeweils die standardisierten Versionen der entsprechenden Variablen darstellen. Vergleichen Sie den Koeffizientenschätzer $\hat{\gamma}_1$ mit dem Bravais-Pearson-Korrelationskoeffizienten zwischen den beiden Variablen. Was fällt dabei auf?

Aufgabe 2: Eigenschaften der Koeffizientenschätzer

(a) Betrachten Sie das einfache lineare Regressionsmodell (siehe Vorlesung, Folie 19)

$$Y_i \quad = \quad \beta_0 + \beta_1 x_i + \epsilon_i; \; i = 1, \ldots, n$$

mit den Annahmen

$$\begin{aligned} E(\epsilon_i) &= 0; \; i = 1, \ldots, n \\ \mathrm{Var}(\epsilon_i) &= \sigma^2; \; i = 1, \ldots, n \\ \{\epsilon_i \mid i = 1, \ldots, n\} & \quad \text{stochastisch unabhängig} \\ \epsilon_i &\sim N(0, \sigma^2); \; i = 1, \ldots, n. \end{aligned}$$

Leiten Sie die Varianzen $\mathrm{Var}(\hat{\beta}_0)$ und $\mathrm{Var}(\hat{\beta}_1)$ der Koeffizientenschätzer $\hat{\beta}_0$ und $\hat{\beta}_1$ her (Gleichungen (1.13) und (1.14) aus der Vorlesung Folie 26).

(b) Veranschaulichen Sie anhand einer Simulation in **R**, dass die Koeffizientenschätzer des in Teilaufgabe (b) beschriebenen einfachen linearen Regressionsmodells folgende Verteilungen besitzen:

$$\begin{aligned} \hat{\beta}_0 &\sim N(\beta_0, \mathrm{Var}(\hat{\beta}_0)) \\ \hat{\beta}_1 &\sim N(\beta_1, \mathrm{Var}(\hat{\beta}_1)). \end{aligned}$$

Verwenden Sie hierzu den folgenden Regressionszusammenhang zur Durchführung von $N = 10.000$ Simulationen:

$$Y_i = -2 + 3.5 \cdot x_i + \epsilon_i \quad \text{mit } \epsilon_i \sim N(0, 10).$$

Aufgabe 3: Quadratsummenzerlegung

In einer Erfassung verschiedener Merkmale von Baseball-Spielern der Major League Baseball (MLB) wurden unter anderem die Körpergröße (in cm) sowie das Gewicht (in kg) der Spieler erhoben. Für 10 zufällig ausgewählte Spieler liegen folgende Werte vor:

Spieler i	1	2	3	4	5	6	7	8	9	10
Größe x_i	198	188	196	190	180	183	196	196	193	183
Gewicht y_i	104	84	107	95	76	79	109	94	113	93

Hinweis: Für die Teilaufgaben (b)-(d) kann **R** als Taschenrechner verwendet werden.

(a) Zeichnen Sie die Information in ein Streudiagramm. Welcher Zusammenhang ist zwischen Körpergröße und Gewicht erkennbar?

(b) Berechnen Sie die Gesamtstreuung (SST) der Variable `Gewicht`.

(c) Bestimmen Sie per Hand die Koeffizientenschätzer $\hat{\beta}_0$ und $\hat{\beta}_1$ für ein lineares Regressionsmodell, das das Gewicht in Abhängigkeit der Körpergröße modelliert.

(d) Berechnen Sie mit Hilfe der in Teilaufgabe (c) bestimmten Parameter die Streuung, die durch das Modell erklärt wird (SSM). Bestimmen Sie anschließend das Bestimmtheitsmaß R^2 und interpretieren Sie dieses.

(e) Überprüfen Sie Ihre Berechnungen, indem Sie die Schätzung des Regressionsmodells nun in **R** mit der Funktion `lm` durchführen.