Regressionsanalyse – Übungen: Blatt 1

1. Betrachte ein einfaches lineares Regressionsmodell, SLR, mit normalverteilten Responsevariablen, d.h.
$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i\,, \qquad i = 1, \dots, n\,,$$
$\epsilon_i \overset{\text{iid}}{\sim} \text{Normal}(0, \sigma^2)$.

   Führe die folgenden Punkte ohne Verwendung von Vektoren und Matrizen aus!

   (a) Diskutiere eine suffiziente Statistik für den Parametervektor $(\beta_0, \beta_1, \sigma^2)$.

   (b) Zeige, dass der MLE $\hat\beta_0$ und das arithmetische Mittel $\bar y$ keine unabhängigen Schätzer sind.

   (c) Berechne den Korrelationskoeffizienten der beiden MLEs $\hat\beta_0$ und $\hat\beta_1$ und interpretiere sein Vorzeichen.

   (d) Zeige, dass beide MLEs $\hat\beta_0$ und $\hat\beta_1$ von den Residuen $r_i = y_i - \hat\mu_i$, $i = 1, \dots, n$, unabhängig sind.

   (e) Berechne den Erwartungswert und die Varianz der Residuen und diskutiere deren Verteilung.

   (f) Nimm an, dass unter diesem Modell $\text{SSE}(\hat\beta_0, \hat\beta_1)/\sigma^2 \sim \chi^2_{n-2}$ gilt. Zeige, dass die beiden Statistiken
$$T_j = \frac{\hat\beta_j - \beta_j}{\sqrt{\widehat{\text{var}}(\hat\beta_j)}}\,, \qquad j \in \{0, 1\}\,,$$
   jeweils einer t_{n-2}-Verteilung genügen. Hierbei bezeichnet der Ausdruck $\widehat{\text{var}}(\hat\beta_j)$ die geschätzte Varianz des MLE, in dem der unbekannte Parameter σ^2 durch den Schätzer $\text{SSE}(\hat\beta_0, \hat\beta_1)/(n-2)$ ersetzt ist.

2. Verwende die Variablen `FEV1`, `Height` und `Weight` aus dem Datensatz `AIMU` auf der Homepage zur Lehrveranstaltung. Betrachte ein SLR in dem der Erwartungswert von `FEV1` einmal von `Height` und ein anderes Mal von `Weight` abhängt. Passe diese beiden Modelle in `R` an und interpretiere sämtliche Ergebnisse, die von `summary(.)` jeweils geliefert werden.

3. Betrachte ein einfaches lineares Regressionsmodell mit Intercept und zeige, dass
$$\sum_{i=1}^n (y_i - \hat\mu_i)^2 = \sum_{i=1}^n (y_i - \overline{y})^2 - \hat\beta_1^2 \sum_{i=1}^n (x_i - \overline{x})^2\,.$$

4. Wir diskutieren ein einfaches lineares Regressionsmodell ohne Intercept, d.h.
$$y_i = \beta_1 x_i + \epsilon_i\,, \qquad i = 1, \dots, n\,, \qquad \text{mit} \quad \epsilon_i \overset{iid}{\sim} N(0, \sigma^2)\,.$$

   (a) Leite die Normalgleichung und den Kleinste-Quadrate Schätzer für β_1 her.

   (b) Zeige, dass die Summe der Residuen unter diesem Modell nicht Null ist.

   (c) Finde die Varianzen von $\hat\beta_1$ und von $\hat\mu_i$ unter diesem Modell.

5. Zeige, dass allgemein gilt:
$$\sum_{i=1}^n (x_i - \overline{x})(y_i - \overline{y}) = \sum_{i=1}^n (x_i - \overline{x}) y_i = \sum_{i=1}^n x_i (y_i - \overline{y})\,.$$