

A 1 (10 Punkte)

Betrachten Sie den Datensatz Auto aus der Bibliothek ISLR (gegebenenfalls vorher mit dem Befehl `install.packages("ISLR")` installieren, dann mit `library(ISLR)` laden). Wir wollen mit den Daten des Objekts Auto arbeiten. Sie bekommen eine Übersicht des Objekts mit dem Befehl `head(Auto)`. Erstellen Sie mit dem Befehl `'hist(Daten)'` Histogramme der Variablen

- Miles per Gallon .
- Horsepower.
- Year (Baujahr).

Verwenden Sie die `lm`-Funktion um eine einfache lineare Regression durchzuführen mit `mpg` (miles per gallon) als abhängiger Variablen und `horsepower` als Prädiktor. Beantworten Sie folgende Fragen mit Begründung:

- Existiert ein linearer Zusammenhang zwischen `mpg` und `horsepower`?
- Wie stark ist der lineare Zusammenhang? Berechnen Sie dazu bitte die Korrelation.
- Ist die Beziehung zwischen abhängiger und unabhängiger Variablen positiv oder negativ?
- Bestimmen Sie die Prognose des Regressionsmodells für `horsepower= 104`.
- Erzeugen Sie einen Plot der Daten mit der eingezeichneten Regressionsgeraden (in roter Farbe, verwenden Sie die Funktion `abline`).

A 2 (5 Punkte)

Richtig oder falsch? Der Regressionskoeffizient b der Regressionsgeraden $y = a + b \cdot x + \varepsilon$ zeigt wie viele Einheiten. . .

- der Wert in x sich im Durchschnitt ändert, wenn der Wert in y um eine Einheit ansteigt;
- die Werte in y sich im Durchschnitt ändern, wenn der Wert in x um eine Einheit ansteigt;
- y ansteigt, wenn x um eine Standardabweichung ansteigt.
- Der Regressionskoeffizient b ist positiv, wenn die Variablen der Regression gleichsinnig variieren, negativ wenn sie gegensinnig variieren.

A 3 (10 Punkte)

In dieser Aufgabe untersuchen wir den Einfluss der Variablen `prestige` auf die Variable `income` im Datensatz `Prestige` in der Bibliothek `'car'`. Speziell wollen wir sehen, ob die Güte der Anpassung besser wird, wenn man die Variablen logarithmiert, also folgende Anpassungen untersucht:

$$\text{income}_i = \beta_0 + \beta_1 \text{prestige}_i + \epsilon_i$$

bzw.

$$\text{income}_i = \beta_0 + \beta_1 \log(\text{prestige}_i) + \epsilon_i$$

oder

$$\log(\text{income}_i) = \beta_0 + \beta_1 \log(\text{prestige}_i) + \epsilon_i.$$

Installieren Sie zunächst in R Studio unter Packages die Bibliothek 'car' und laden Sie diese mit `library(car)`. Die folgenden Fragen sind mittels linearer Regression zu beantworten.

- a)** Untersuchen Sie den Einfluss von `prestige` auf `income`. Interpretieren Sie den geschätzten Regressionskoeffizienten β_1 , dessen statistische Signifikanz, und das R^2 .
- b)** Untersuchen Sie den Einfluss von `prestige` auf `log(income)`. Interpretieren Sie den geschätzten Regressionskoeffizienten β_1 , dessen statistische Signifikanz, und das R^2 .
- c)** Untersuchen Sie den Einfluss von `log(prestige)` auf `log(income)`. Interpretieren Sie den geschätzten Regressionskoeffizienten β_1 , dessen statistische Signifikanz, und das R^2 .