

## Exercise 1 - Restaurantbesuche

Sie möchten untersuchen, ob man durch den Betrag, den Restaurantgäste für Essen ausgeben, ungefähr den Betrag vorhersagen kann, den sie für Getränke ausgeben. Sie zeichnen dafür über einen Monat hinweg die entsprechenden Beträge in Euro auf und erhalten daraus folgende Regressionsgleichung:

$$\hat{y} = 10 + 1.5x$$

Sie ermitteln zudem das Bestimmtheitsmaß von  $R^2 = 0.8$ .

- (a) Interpretieren Sie diese Regressionsgleichung unter Berücksichtigung des Bestimmtheitsmaßes im Hinblick auf die untersuchte Fragestellung. Erläutern Sie dabei die inhaltliche Bedeutung der hier ermittelten Werte der Regressionskoeffizienten
- (b) Stellen Sie diese Regressionsgerade in einem Diagramm/Koordinatensystem schematisch dar. Achten Sie vor allem auf die Steigung und den Schnittpunkt der Regressionsgeraden und die Benennung/Beschriftung der Achsen.
- (c) Wie würde ein R-Befehl für jene konkrete Aufgabe aussehen? Was wäre die erklärende Variable, was wäre die Zielgröße?

## Exercise 2 - Korrelationskoeffizient und Heteroskedastizität

1. Welche der folgenden Antworten ist bzgl. der Korrelation bzw. des Korrelationskoeffizienten  $r$  richtig und welche falsch? Begründe jeweils:
  - a) Ist  $r=1$  so liegen alle Werte auf einer steigenden Gerade
  - b) Ist  $r=0$ , dann besteht keine lineare Beziehung zwischen den beiden Variablen.
  - c)  $r$  misst Zusammenhang zwischen einer quantitativen Und qualitativen Variable
  - d)  $r$  liegt immer zwischen  $0 \leq r \leq 1$
  - e)  $r$  kann auch unendlich groß oder klein werden.
  - f)  $r$  gibt die Stärke der Beziehung zwischen X und Y an.
  - g)  $r$  ist standardisiert
  - h)  $r$  ist nicht symmetrisch
2. Was versteht man unter Heteroskedastizität der Residuen? Versuchen Sie diese grafisch "schematisch" darzustellen, einmal für den Fall der Hetero- und einmal im Fall der Homoskedastizität.

## Exercise 3 - R Lineare Regression mit Wetterdaten

Benutzen Sie den Datensatz "Seehöhe". Dieser Datensatz beinhaltet aktuelle Daten aus den öffentlich zugänglichen ZAMG Wetterdaten. Bearbeiten Sie die folgenden Aufgabenstellungen in R:

- (a) Berechnen Sie die Koeffizienten eines einfachen linearen Regressionsmodells, wobei die Seehöhe als Einfluss- und die Temperatur als Zielvariable angesehen werden soll. Interpretieren Sie das Ergebnis!

- (b) Ist die Annahme eines linearen Regressionsmodells angemessen? Um diese Frage zu beantworten, erstellen Sie bitte einen Residuen-Plot in R (d.h. Modell-Residuen auf der y-Achse und die erklärende Variable auf der x-Achse auftragen).

## **Exercise 4 - R: Hippocampi Daten**

In diesem Beispiel geht es um Daten, die zwar “künstlich” (d.h. simuliert) sind, aber den realen Verhältnissen in gewisser Weise entsprechen. Die damit verbundenen Forschungsfragen sind auf jeden Fall auch in der aktuellen klinischen Forschung relevant! Es geht dabei um Personen, die gesund sind oder bestimmte kognitive Defizite haben, die - möglicherweise - als “Vorstufe” von (Alzheimer-)Demenz angesehen werden können (im R-Code sind etwas genauere Informationen zu finden). Bearbeiten Sie bitte die folgenden Aufgabenstellungen – die bewusst sehr allgemein formuliert sind, denn es geht darum, dass Sie nun einmal selbstständig versuchen, Ihr bisher gelerntes Wissen anzuwenden:

1. Lesen Sie die Kurzbeschreibung der Daten im R-Code. Verschaffen Sie sich einen ersten Überblick, indem Sie die Daten (oder einen Teil davon) anzeigen lassen und von Ihnen selbstgewählte, geeignete deskriptive Auswertungen machen.
2. Außerdem sind auch diverse Zusammenhänge zwischen den Variablen des Datensatzes von Interesse (z.B. zwischen dem gesamten intrakraniellen Volumen und dem Hippocampusvolumen). Führen Sie entsprechende deskriptive Analysen aus, und “experimentieren” Sie mit verschiedenen einfachen und multiplen linearen Regressionsmodellen. Im Code sind einige Beispiele dazu genannt.