

Aufgaben zur Vorlesung „Werkzeuge der empirischen Forschung“

Aufg. 43) (3 P.) Erzeugen Sie 10000 unabhängige normalverteilte Zufallszahlen, und untersuchen Sie, ob diese wirklich normalverteilt sind (Q-Q-Plot, Anpassungstests). Verwenden Sie `set.seed`.

Wie kann man testen, ob sie auch wirklich unabhängig sind?

Aufg. 44) (10 P.) (Clusteranalyse)

a) (1 P.) Erzeugen Sie jeweils 50 Paare bivariat normalverteilter Zufallsvariablen, jeweils mit Varianz $\sigma^2 = 1$ und den Erwartungswerten $(\mu_x = 0, \mu_y = 0)$, $(\mu_x = 4, \mu_y = 8)$ und $(\mu_x = 8, \mu_y = 0)$, den Kovarianzen Null, und speichern Sie diese in einem `data.frame`. Verwenden Sie auch hier `set.seed`.

Hinweis: Verwenden Sie `mvrnorm` aus dem Paket `MASS`.

b) (2 P.) Führen Sie mit diesen Daten zwei hierarchische Clusteranalysen durch. Die Methoden seien Single Linkage bzw. Ward. Interpretieren Sie die Ergebnisse.

c) (3 P.) Zeichnen Sie den Clusterbaum und erzeugen Sie einen `data.frame` der je Zeile das Paar aus **a)** enthält sowie den zugeordneten Cluster. Nutzen Sie dazu `cutree`, um die drei Cluster zu unterscheiden. Stellen Sie die drei Cluster in einem Scatterplot mit unterschiedlichen Farben dar.

d) (1 P.) Welche hierarchischen Clustermethoden gibt es noch, und wodurch unterscheiden sie sich?

e) (3 P.) Generieren Sie 50 weitere Paare von Zufallsvariablen (X, Y) , wobei X und Y unabhängig und normalverteilt seien, wobei $\mu_X = \mu_Y = 0.25$ und $\sigma_X^2 = \sigma_Y^2 = 4$ und wiederholen Sie beide Clusteranalysen mit allen 200 Beobachtungspaaren.

Aufg. 45) (7 P.) (logistische Regression, Datei `challenger.dat`)

Modellieren Sie die Wahrscheinlichkeit des Defektes eines Dichtungsringes in Abhängigkeit von der Außentemperatur mit Hilfe einer logistischen Regression.

a) (2 P.) Schätzen Sie die Parameter im logistischen Regressionsmodell. Rechnen Sie das Modell mit allen Beobachtungen, die **vor** der Challenger-Katastrophe aufgenommen wurden, d.h. **ohne** den Temperaturwert von 30 °F.

b) (2 P.) Schätzen Sie die Wahrscheinlichkeit dafür, dass bei einer Außentemperatur von 30 °F ein Dichtungsring ausfällt.

c) (2 P.) Zeichnen Sie eine Funktion, die die geschätzte Wahrscheinlichkeit für den Ausfall eines Dichtungsringes in Abhängigkeit von der Außentemperatur angibt.

d) (1 P.) Diskutieren Sie die Ergebnisse.