

Übung 10

Explorative Datenanalyse und Visualisierung

Wintersemester 2019
S. Döhler, B. Nedic (FBMN, h_da)

Name:

Punkte:

Aufgabe 22. Auf der Moodle-Seite zur Vorlesung finden Sie 4 verschiedene zweidimensionale Datensätze.

- Ermitteln Sie in jedem Fall \bar{x} , \bar{y} , s_x , s_y und r_{xy} . Was fällt Ihnen auf?
- Ermitteln Sie in allen 4 Fällen das Regressionsmodell und die Regressionsgerade. Benutzen Sie hierfür den `lm`-Befehl.
- Schreiben Sie eine Funktion `plot.regression`. Als Argumente sollen dieser Funktion übergeben werden:
 - **daten:** Ein Datensatz
 - **model:** Das zugehörige `lm`-Objekt
 - **header:** Eine Überschrift
 - **x.lim, y.lim:** Wertebereich der x - und y -Achse

Als Output soll die Funktion einen Scatterplot der Daten (auf den übergebenen Wertebereichen) mit zugehöriger Regressionsgerade sowie eines Titels liefern.

- Plotten Sie die 4 Scatterplots inklusive ihrer Regressionsgeraden in einem gemeinsamen Grafikpanel. Benutzen Sie die Funktion `plot.regression`.
- Führen Sie für die 4 Datensätze jeweils eine Residuenanalyse durch und interpretieren Sie die Ergebnisse.

Interpretieren Sie die Ergebnisse.

Lösung

Anmerkungen/Korrektur

Aufgabe 23. Arbeiten Sie weiter an dem Datensatz `UmfrageBis2019.csv` (s. Aufgabe 21).

- a) Untersuchen Sie, ob und wenn ja, welcher Zusammenhang zwischen den Merkmalen 'Fussballfan' und 'Geschlecht' besteht. Erzeugen Sie hierzu eine Vierfeldertafel und einen Mosaicplot mit geeigneten R-Befehlen.
- b) Untersuchen Sie, die gleiche Fragestellung wie in a) bei den Merkmalen 'Musikalitaet' und 'Geschlecht'. Was unterscheidet das Merkmal 'Musikalitaet' vom Merkmal 'Fussballfan'? Wie sollte man das in der Analyse berücksichtigen?
- c) Untersuchen Sie, die gleiche Fragestellung wie in a) bei den Merkmalen 'Musikalitaet' und 'Haarfarbe'.
- d) In welchen der 3 obigen Situation scheint der Mosaicplot am ehesten auf Unabhängigkeit der Merkmale hinzudeuten? Begründen Sie Ihre Antwort!

Lösung

Anmerkungen/Korrektur
