

Übungen

Abgabetermin: Mittwoch, 24.06.2009; 09:15 Uhr, BK 41

Aufgabe 1. (5 Punkte)

In Aufgabe 4 auf Blatt 2 wurde bereits die Kontingenztafel des auf die beiden Merkmale **Sex** und **Survived** reduzierten Datensatzes **Titanic** erstellt.

- (a) Erstellen Sie nun den auf die Merkmale **Class** und **Survived** reduzierten Datensatz. Verwenden Sie dazu den **apply**-Befehl. Bestimmen Sie die zugehörigen bedingten relativen Häufigkeiten des reduzierten Datensatzes. Was besagen diese?
- (b) Stellen Sie die reduzierten Daten grafisch dar. Wählen Sie dazu die Darstellungsform, die Ihrer Meinung nach die wichtigen Informationen am besten verdeutlicht. Warum haben Sie diese Art der Darstellung gewählt? Welche Vorteile und Nachteile besitzt sie gegenüber den anderen?
- (c) Erstellen Sie einen Mosaikplot des Gesamtdatensatzes und färben Sie ein Merkmal mithilfe des Arguments **highlighting** ein. Beschreiben Sie die Grafik.

Aufgabe 2. (5 Punkte)

In den Jahren 1931 und 1932 wurden in Minnesota auf sechs verschiedenen Farmen (**site**) jeweils zehn verschiedene Gerstesorten (**variety**) angebaut und die Ernteerträge (**yield**) gemessen. Diese Daten sind in **barley** im Paket **lattice** enthalten.

- (a) Erstellen Sie für beide Jahre eine Zusammenfassung (**summary**) der Daten mit Hilfe der Funktion **by**, deren Verwendung im Hilfetext erläutert wird. Erstellen Sie ebenso Zusammenfassungen der Ertragsdaten **yield** aufgeteilt nach Gerstesorten.
- (b) Berechnen Sie jeweils für beide Jahre die 0.9-Quantile der Ertragsdaten. Welche Farmen bilden die ca. zehn Prozent mit höheren Erträgen?
- (c) Plotten Sie die Art gegen den Ertrag in Abhängigkeit von Ort und Jahr (**variety ~ yield | site+year**) mit Hilfe der Funktion **dotplot**. Die entstandene Grafik ist ein Trellis-Display. Was lässt sich daraus ablesen? Erstellen Sie ein weiteres Trellis-Display mit einer anderen Aufteilung. Was lässt sich aus diesem ablesen?

(Bitte wenden!)

Aufgabe 3. (5 Punkte)

Lesen Sie die auf der Seite des Praktikums eingestellte Liste der 30 umsatzstärksten deutschen Unternehmen (im Jahr 2007 – ohne Banken und Versicherungen) ein. Die Liste umfasst die Umsätze Y (in Millionen Euro) und Beschäftigungszahlen X jeweils aus dem Jahr 2007.

- (a) Zeichnen Sie ein Streudiagramm für die zehn Unternehmen, und interpretieren Sie dieses.
- (b) Man bestimme den Bravais-Pearson- und den Spearman'schen Korrelationskoeffizienten.
- (c) Wie ändern sich die Korrelationskoeffizienten, wenn man in absoluten Euro-Umsätzen rechnet?

Aufgabe 4. (5 Punkte)

In der Datentabelle `airquality` sind die Ergebnisse täglicher Messungen der Luftqualität in New York, Mai bis September 1973 enthalten. Diese umfassen Ozon, Sonneneinstrahlung, Windgeschwindigkeit und Temperatur sowie das Datum.

- (a) Verschaffen Sie sich mit Hilfe von Streudiagrammen einen Überblick über mögliche Zusammenhänge zwischen den meteorologischen Größen. Zwischen welchen Merkmalen könnten Zusammenhänge vorliegen? Welcher Art wären die Zusammenhänge?
- (b) Berechnen Sie die Korrelationskoeffizienten von (Bravais-)Pearson und Spearman für alle sechs möglichen Paarungen verschiedener meteorologischer Merkmale. Vergleichen Sie die beiden Koeffizienten. Passen die Korrelationskoeffizienten zu Ihren Vermutungen aus Teil (a)?