



Prof. Antony Unwin, Alexander Pilhöfer
Lehrstuhl für Rechnerorientierte Statistik und Datenanalyse
Institut für Mathematik
Universität Augsburg
<http://stats.math.uni-augsburg.de/>

Statistik I

Übungsblatt 10

Abgabe: Dienstag 26. Juni 2012, bis spätestens 12.00 Uhr; Briefkasten: Statistik I oder per email an die Übungsleiter

Die Aufgaben können auch in 2er-Gruppen bearbeitet und abgegeben werden!

- Drücken Sie die Korrelation ausschließlich durch Kovarianzen aus! **(1P)**
 - Im χ^2 -Unabhängigkeitstest der Variablen "liest BILD" (nie/selten/oft/täglich) und "Ausbildung" (7 Stufen: Universität, Abitur, Berufsausbildung,...) wird eine χ^2 -Verteilung mit wievielen Freiheitsgraden verwendet? **(1P)**
 - Ist der Test aus Aufgabe 1 b) perfekt geeignet? **(1P)**
 - Was ist der grundlegende Unterschied zwischen einem Unabhängigkeitstest und einem Anpassungstest. Welche Gemeinsamkeiten liegen vor?**(1P)**
- Mensa (5P)** Während der Stosszeiten betreten pro Sekunde drei Personen die Mensa. Bei einer Beobachtung über eine Minute wurden die folgenden Häufigkeiten von Anzahlen Personen, die die Mensa innerhalb einer Sekunde betraten, gemessen:

Anzahlen	0	1	2	3	4	5	6	7	8	> 8
Häufigkeiten	3	5	16	14	11	6	4	0	1	0

Nehmen Sie an, dass die Anzahl Personen, die pro Sekunde die Mensa betreten, Poisson-verteilt ist mit $\lambda = 3$. Berechnen Sie in R die Beiträge der einzelnen X^2 -Komponenten und überprüfen Sie mittels eines χ^2 -Anpassungstests, ob die Annahme der Poisson-Verteilung (mit $\lambda = 3$) verworfen werden darf.

3. Ford (10P)

Im Rahmen eines Wettbewerbs hat die Firma Ford verschiedene Messungsgrößen während der Autofahrt verschiedener Fahrer aufgezeichnet. Ziel war es, anhand anderer, leicht messbarer Größen vorherzusagen, wann ein Fahrer nicht mehr wach/aufmerksam ist (zB Sekundenschlaf). Es könnte dann ein Weck- oder Warnsystem oder gar ein Autopilot eingreifen. Der Datensatz *Ford* auf der Webseite enthält diese Messungen für einen solchen Testfahrer.

Verwenden Sie zur Bearbeitung sowohl R als auch Mondrian.

- Ein häufig zu beobachtender erster Schritt in der Datenanalyse ist die Berechnung der Korrelationsmatrix. Berechnen Sie diese für die Variablen mit den Bezeichnungen *P* und *V*.
- Stellen Sie die Korrelationsmatrix mit dem Paket `corrgram` in R grafisch dar. Orientieren Sie sich an der Hilfe, um verschiedene Optionen zu testen. Halten Sie die Visualisierung für gelungen?
- Betrachten Sie folgende Variablenpaare und beschreiben Sie deren Zusammenhang.
 - $V2 - V4$: Besteht hier ein leicht negativer, linearer Zusammenhang?

- $P1 - P4$: Die Korrelation ist fast 0. Sind die beiden stochastisch unabhängig?
 - $P5 - V11$: Eine starke positive Assoziation?
 - $V2 - V5$: Welchen Test würden Sie hier empfehlen?
- (d) Betrachten Sie das Variablenpaar $V1$ vs. $V6$. Ist die Korrelation hier sinnvoll? Würden Sie einer Modellierung $V6 = a + b \cdot V1 + \epsilon$, wobei ϵ eine zufälliger, normalverteilter Störterm ist, zustimmen?
- (e) Berechnen Sie für die Variablenpaare $P1 - P2$, $V1 - V6$ und $V2 - V4$ die Korrelation der Distanzen aller Beobachtungspaare (`dist(ford$V2)`) und vergleichen Sie diese mit den Korrelationen.
- (f) Welche Variablen würden Sie empfehlen, um die Aufmerksamkeit des Fahrers (mittels eines Modells) vorherzusagen?
- (g) **(Extra)** Erstellen Sie ein Streudiagramm der Variablen $V1$ und $V6$ sowie ein Histogramm der Variable `ObsNum` mit sehr vielen Bins. Aktivieren Sie über das Menü oder STRG+M sogenannte *Selection Sequences*. Erzeugen Sie im Histogramm ein schmales Selektionsfeld, mit dem Sie über das Histogramm fahren. beobachten Sie dabei die Achterbahnfahrt im Streudiagramm. Welche Streckenabschnitte sind am heftigsten? (d.h. wann verliert der Fahrer das Bewusstsein?).
- (h) **(Extra)** Finden Sie mehr über die Funktionsweise von *Selection Sequences* heraus.

4. Bevölkerungsexplosion (5P)

Auf der Webseite der Vorlesung finden Sie unter der Rubrik Übungsblätter einen Artikel zum Thema Bevölkerungsentwicklung. Lesen Sie diesen Artikel und nehmen Sie dazu kritisch Stellung. Gehen Sie auch auf die getroffenen Aussagen im Text ein und inwiefern Sie in den Grafiken (korrekt) visualisiert sind. Finden Sie die wichtigsten Kritikpunkte zu den grafischen Darstellungen. Was hätten Sie anders (besser) gemacht?