

Stochastik II — Statistik I (Statistik und Datenanalyse, SS 2009)

Übungsblatt 9

Abgabe: Dienstag 23. Juni 2009, bis spätestens 10.00 Uhr; Briefkasten: Stochastik II

1. Zeigen Sie, dass für den χ^2 -Unabhängigkeitstests im Fall einer Vierfeldertafel

	$S1$	$S2$	Σ
$Z1$	n_{11}	n_{12}	$n_{1.}$
$Z2$	n_{21}	n_{22}	$n_{2.}$
Σ	$n_{.1}$	$n_{.2}$	n

für die Teststatistik $X^2 = \sum_{i=1}^2 \sum_{j=1}^2 \frac{(n_{ij} - e_{ij})^2}{e_{ij}}$ gilt: $X^2 = \frac{(n_{21}n_{12} - n_{11}n_{22})^2 (n_{1.} + n_{2.})}{n_{.1}n_{.2}(n_{11} + n_{12})(n_{1.} + n_{2.} - n_{11} - n_{12})}$.

2. Bei einem Schulfest an 5005 ungefähr gleichaltrigen Kindern wurden die Ergebnisse nach Geschlecht getrennt aufgeführt. Es ergaben sich die folgenden Daten:

	Mädchen	Jungen
Schulreif	2350	2337
Nicht schulreif	121	197

- (a) Berechnen Sie per Hand einen χ^2 -Unabhängigkeitstest und entscheiden Sie zum Test-Niveau 0.05, ob es Assoziationen zwischen Geschlecht und Reife gibt.
- (b) Kann man letzteres Ergebnis auch schon direkt den Daten ansehen?
3. 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 \mathbb{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.

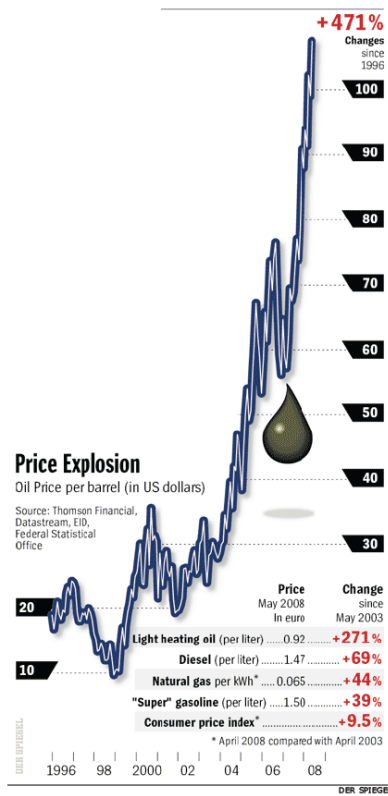
4. (a) An zehn Patienten wurden zwei Schlafmittel erprobt, wobei folgende zusätzliche Schlafzeiten in Stunden beobachtet wurden:

Patient	1	2	3	4	5	6	7	8	9	10
Medikament 1	0.7	-1.6	-0.2	-1.2	-0.1	3.4	3.7	0.8	0.0	2.0
Medikament 2	1.9	0.8	1.1	0.1	-0.1	4.4	5.5	1.6	4.6	3.4

Führen Sie einen geeigneten t -Test durch, um zu überprüfen, ob die beiden Medikamente verschiedene Schlafwirkungen haben.

- (b) Der Datensatz *AgentOrange* (auf der Internetseite der Lehrveranstaltung) enthält Messungen von Dioxin (auch bekannt als Agent Orange) für 646 Vietnam-Veteranen und für 97 Veteranen, die nicht im Vietnam waren. Sind die Mittelwerte der beiden Gruppen signifikant voneinander verschieden? Was für einen Einfluss haben die Ausreißer? Kann man die Varianzen als gleich ansehen?

5. Kommentieren Sie die folgende Graphik:



Spiegel Online, 29.05.2008

- Was ist das Thema der Graphik und welche Informationen stellt sie bereit?
- Was stellen die beiden Skalen dar? Wie beurteilen Sie deren Skalierungen und Umsetzungen in der Karte?
- Was stellt die Tabelle unten dar? Diskutieren Sie das Zahlenmaterial in der Graphik.
- Sollten weitere Informationen in der Graphik gezeigt werden, wenn ja welche?
- Beurteilen Sie die Graphik in Bezug auf Suggestivität, Übersichtlichkeit und Ästhetik.

6. Einzeiler-Aufgaben.

(a) Bestimmen Sie den p -Wert, wenn gilt $X^2 = 5.501$ und $X^2 \sim \chi_1^2$.

(b) Was besagt der folgende R-Code

```
> (x1<-chisq.test(Survived, Sex))
```

Pearson's Chi-squared test with Yates' continuity correction

data: Survived and Sex

X-squared = 454.4998, df = 1, p-value < 2.2e-16

im Fall des **Titanic** Datensatzes?