Statistik für Informatikstudien

506.004

Übungsblatt 1

26. Jän. 2010

1. [A 1.1] In der folgenden Tabelle sind die CPU–Zeiten [sec] von 25 Jobs, die an einem Großrechner ausgeführt wurden, aufgelistet.
Die Berechnungen in Beispiel 1 sind in `R` *durchzuführen.*

Rechenzeiten [sec] für 25 Computerjobs									
1.17	1.61	1.16	1.38	3.53	1.23	3.76	1.94	0.96	4.75
0.15	2.41	0.71	0.02	1.59	0.19	0.82	0.47	2.16	2.01
0.92	0.75	2.59	3.07	1.40					

(a) Man erstelle ein Stamm-Blatt-Diagramm, ein Histogramm, den Boxplot, die empirische Verteilungsfunktion und den Q-Q-Plot.

(b) Man berechne $\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$, $s_L = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2}$ und vergleiche diese Kenngrößen mit dem Median $\tilde{x}$ und dem Streuungsmaß $s_q = (q_{0.75} - q_{0.25})/1.349$. Wie lauten die Werte für die beiden Schiefemaße $g_1 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^3/s^3$, $g_1^q = (q_{0.95} - \tilde{x})/(\tilde{x} - q_{0.05}) - 1$?

2. [A 1.2] Eine Stichprobenerhebung eines diskreten Merkmals X ergab folgende Verteilung

i	1	2	3	4	5	6	7	8	9	10	11	12
x_i	20	43	53	86	70	54	37	18	10	5	2	2

(a) Zeichnen Sie ein Histogramm, die empirische Verteilungsfunktion und den Boxplot mit Papier und Bleistift.

(b) Bestimmen Sie den Mittelwert $\overline{x}$, die Standardabweichung s, die Schiefe g_1 und die Kurtosis g_2.

3. [A 2.11] Ein Fahrkartenkontrolleur überprüft einen Tag lang auf verschiedenen Grazer Straßenbahnlinien die Fahrkarten von Fahrgästen. Er überprüft jeweils solange, bis er einen Fahrgast ohne gültigen Fahrschein antrifft. Nach Ausstellung eines Strafprotokolls kassiert er von diesem ein Bußgeld und beginnt nach einer Pause mit einer neuen Überprüfung. Die folgenden Zahlen geben an, wieviele Fahrgäste bei 10 solchen Überprüfungen jeweils überprüft wurden, bis ein Bußgeld fällig wurde:

42 50 40 64 30 36 68 42 46 48

Beschreibt die Zufallsvariable X die *Anzahl der Personen*, die überprüft werden, bis ein Fahrgast ohne gültigen Fahrausweis angetroffen wird, so kann angenommen werden, dass

$$P_\theta(X = k) = (1 - \theta)^{k-1}\theta$$

gilt, wobei $\theta \cdot 100\%$ als prozentualer Anteil der Schwarzfahrer unter allen Fahrgästen zu interpretieren ist. Man bestimme aufgrund obiger Messwerte einen Maximum–Likelihood–Schätzwert für θ.

4. [A 2.27] Die Zufallsvariablen $X_1, \ldots, X_n$ seien unabhängig und identisch $N(\mu, 5)$–verteilt mit unbekanntem μ.

(a) Wie groß muss der Stichprobenumfang n mindestens gewählt werden, damit die Länge des zweiseitigen 95%–Konfidenzintervalls für μ *kleiner* als 1.25 wird?

(b) Welches Konfidenzniveau besitzt das Konfidenzintervall, wenn bei $n = 200$ eine Länge von 1.15 gefordert wird?

(c) Welche Länge besitzt das konkrete Konfidenzintervall, das bei $n = 150$ ein Niveau von 0.8 $(\alpha = 0.2)$ aufweist?

5. [A 2.31] Eine Befragung ergab, dass von $n = 50$ männlichen Softwareingenieuren nur 8 regelmässig Zigaretten rauchen. Man konstruiere ein 90% (95%)–Konfidenzintervall für den Anteil p der Raucher unter allen Softwareingenieuren.

6. [A 3.3] Um die Genauigkeit eines neu entwickelten Gerätes zur Messung von Weglängen im Gelände zu überprüfen, wurde eine 1000 Meter lange Strecke zehnmal vermessen.

i	1	2	3	4	5	6	7	8	9	10
x_i	998.4	1001.1	1002.5	1000.5	999.1	997.4	1000.0	999.3	996.2	998.3

(a) Man teste zum Niveau $\alpha = 0.05$, ob das Gerät im Mittel die korrekte Entfernung angibt.

(b) Kann die Hypothese, dass die Varianz σ^2 den Wert $\sigma_0^2 = 4\,[m^2]$ nicht unterschreitet, verworfen werden $(\alpha = 0.05)$?

7. [A 3.19] Zum Vergleich zweier Spanplattenlieferungen A und B von zwei unterschiedlichen Lieferanten wurde an je 10 Platten beider Lieferungen die Biegefestigkeit ermittelt. Dabei ergaben sich folgende Werte [in kp/cm^2]:

A	198	195	200	194	192	196	199	195	197	196
B	196	202	199	194	196	198	207	200	203	199

(a) Kann auf dem Niveau $\alpha = 0.05$ die Gleichheit der Biegefestigkeit zwischen den Lieferungen nicht verworfen werden (Annahme $\sigma_A = \sigma_B$)?

(b) Kann die für den Test in (a) erforderliche Hypothese der Gleichheit der Varianzen zum Niveau $\alpha = 0.05$ aufrechterhalten werden?

8. [A 3.28] $n = 10$ Patienten werden an zwei aufeinanderfolgenden Nächten Schlaftabletten verabreicht. In einer Nacht bekamen sie eine Tablette vom Typ A in der anderen eine Tablette vom Typ B, wobei für die Patienten nicht bekannt war, welche Tablette in welcher Nacht gegeben wurde. Folgende Schlafdauern in Stunden wurden registriert:

Patient	1	2	3	4	5	6	7	8	9	10
Pille A	6.5	7.2	5.1	4.0	8.3	7.5	7.0	4.5	6.4	8.0
Pille B	5.8	5.8	4.5	3.5	8.4	5.1	7.7	4.2	5.1	5.4

Läßt sich ein statistisch signifikanter Unterschied in den Schlafdauern feststellen? Man formuliere eine geeignete Hypothese und teste diese zu den Niveaus von $\alpha = 0.05$ und $\alpha = 0.01$.

Besprechungstermin:

Di. 26. 01. 2010 12:00 - 13:30 HS G
Di. 26. 01. 2010 14:15 - 15:45 HS L

Übungstermine:

Gruppe 1: Di. 26. 01. 2010 13:30 - 14:15 HS G
Gruppe 2: Di. 26. 01. 2010 14:30 - 15:00 HS G
Gruppe 3: Di. 26. 01. 2010 14:15 - 15:45 HS L

Bei Fragen wenden Sie sich bitte an unseren wissenschaftlichen Mitarbeiter oder an unsere StudienassistentInnen:

Moritz Jirak `m0ritz@yahoo.com`
Markus Zahrnhofer `markus.zahrnhofer@student.tugraz.at`
Markus Kügerl `kuegerl@student.TUGraz.at`
Brigitte Pfeiler `b.pfeiler@student.TUGraz.at`
Lisa Stadlmüller `lisa86@sbox.TuGraz.at`

Lösungen:

1. (b) $\bar{x} = 1.63$, $s_L = 1.17$, $\tilde{x} = 1.38$, $s_q = 0.99$, $g_1 = 0.86$, $g_1^q = 0.91$
2. (b) $\bar{x} = 33.33$, $s = 28.22$, $g_1 = 0.41$, $g_2 = -1.32$
3. $\hat{\theta} = \frac{1}{\bar{x}} = 0.0215$
4. (a) $n \geq 246$
   (b) Konfidenzniveau $100 \cdot 0.896\%$, $\alpha = 0.104$
   (c) 1.046
5. $90\% : [0.075; 0.245]$, $95\% : [0.058; 0.262]$
6. (a) $t = -1.237$. Man kann H_0 zum geg. Niveau nicht verwerfen.
   (b) $y = 7.62$. Man kann H_0 zum geg. Niveau nicht verwerfen.
7. (a) $t = -2.238$. Verwerfe H_0 zum geg. Niveau.
   (b) $q = 0.3897$. Man kann H_0 zum geg. Niveau nicht verwerfen.
8. $t = 2.733$.
   $\alpha = 0.05$: Verwerfe H_0.
   $\alpha = 0.01$: Man kann H_0 nicht verwerfen.