Klausur
Grundlagen der Statistik
(WS 2001 / 02)

Bemerkungen:

- Die Klausur besteht aus **5 Aufgaben**, die alle zu lösen sind.
- Lösen Sie die Aufgaben **nur** auf den ausgeteilten Lösungsbögen.
- Vermerken Sie auf **jedem Blatt** Ihren Namen und Ihre Matrikelnummer.
- Schreiben Sie bitte **lesbar**.
- Achten Sie darauf, dass Ihr **Lösungsweg nachvollziehbar** ist.

Aufgabe 1 *(10 Punkte)*

Bewerten Sie jede dieser fünf Aussagen mit *RICHTIG* oder *FALSCH* und **begründen** Sie Ihre Entscheidung. Für (richtige) Bewertungen ohne Begründung gibt es **keine** Punkte.

1. Falls für die Zufallsvariablen X und Y gilt: $Var(X) = Var(Y) = 1$, so folgt daraus, dass ihre Korrelation gleich ihrer Kovarianz ist.

   Lösung: RICHTIG!

   Es gilt:

   $$\rho(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var(X) \cdot Var(Y)}}$$

   Eingesetzt:

   $$\rho(X,Y) = \frac{Cov(X,Y)}{\sqrt{1 \cdot 1}} = Cov(X,Y)$$

   Der Korrelationskoeffizient ist die "standardisierte" Kovarianz.

2. Das Konfidenzintervall zum Niveau $\alpha = 0,01$ für den unbekannten Parameter θ sei $[3;7]$. Dann gilt mit einer Wahrscheinlichkeit von $99\,\%$, dass $3 \leq \theta \leq 7$ ist.

   Lösung: FALSCH!
   Ein realisiertes konfidenzintervall überdeckt den Parameter oder überdeckt ihn nicht. Realisierten Konfidenzintervallen können keine Wahrscheinlichkeitsaussagen zugeordnet werden. Dies ist nur bei der Konfidenz*strategie* möglich.

3. Die zufällige Variable X sei binomial verteilt $(X \sim B(n,\pi))$. Dann wird $Var(X)$ um so größer, je größer der Parameter π ist $(\pi \in [0;1])$.

   Lösung: Falsch!
   Es gilt:

$$Var(X) = n \cdot \pi \cdot (1-\pi)$$

   Darum also:

$$\max_{\pi} Var(X) = \frac{n}{4} \Rightarrow \pi = \frac{1}{2}$$

   Die Varianz von X ist am größten, wenn $\pi = \frac{1}{2}$.

4. Beim Durchführen eines statistischen (Hypothesen-) Tests darf man H_0 und H_1 erst nach Realisation der Stichprobe aufstellen.

   Lösung: Falsch!
   Die Hypothesen eines Tests *müssen* vor Ziehung der Stichprobe feststehen. Die Stichprobe liefert die Prüfgröße, die je nach Test-Niveau zur Ablehnung oder Nicht-Verwerfung der Nullhypothese führt.

5. Likelihoodfunktionen geben die Wahrscheinlichkeit dafür an, dass ein (unbekannter) Parameter einen bestimmten Wert annimmt. Dabei ist sie maximal für den wahren Wert des Parameters.

   Lösung: Falsch!
   Der ML-Schätzer gibt den Wert für den unbekannten Parameter an, der *bezüglich des Stichproben-Ergebnisses* am plausibelsten erscheint.

Aufgabe 2 *(20 Punkte)*

Fotografin F. lässt seit drei Jahren ihre Bilder bei einem Fotolabor entwickeln. Von diesem bekommt sie monatlich eine Rechnung Sie erteilt Ihnen den Auftrag, die Labor-Kosten der Jahre 1999 - 2001 statistisch auszuwerten.

Die Jahresabrechnung für das Jahr 2001 (in DM) liegt Ihnen dazu in folgender Form vor:

Monat (i)	1	2	3	4	5	6	7	8	9	10	11	12
Betrag (x_i)	353	209	203	125	126	244	179	215	150	283	237	217

1. Geben Sie das arithmetische Mittel, die empirische Standardabweichung, den Median und den Interquartilsabstand der Daten für das Jahr 2001 an.

   Rechenhilfe : $\sum_{i=1}^{12} x_i^2 = 584649$

   Lösung: $\overline{x} = \frac{\sum_{i=1}^{12} x_i}{12} = \frac{(353+209+203+125+126+244+179+215+150+283+237+217)}{12} = \frac{2541}{12} = 211,75$

   $s_x = \sqrt{\frac{1}{12}\sum_{i=1}^{12} x_i - \overline{x}} = \sqrt{\frac{1}{12}\sum_{i=1}^{12} x_i^2 - \overline{x}^2} = \sqrt{\frac{1}{12} \cdot 584649 - (211,75)^2} \approx 62,31$

   Daten sortiert:

Reihenfolge	1	2	3	4	5	*6*	*7*	8	9	10	11	12
Betrag	125	126	150	179	203	*209*	*215*	217	237	244	283	353

   $x_{med} = \frac{209+215}{2} = 212$ oder $x_{med} = x_{(6)} = 209$

   $x_{0,25} = x_{(3)} = 150$

   $x_{0,75} = x_{(9)} = 237$

   $IQR = x_{0,75} - x_{0,25} = x_{(9)} - x_{(3)} = 237 - 150 = 87$

2. Fertigen Sie den Boxplot der Monatsbeträge für das Jahr 2001 an. Charakterisieren Sie danach die Form der Verteilung.

   Lösung:

   Die Verteilung ist links-schief bzw. rechts-steil. Es sind keine extremen Ausreißer zu verzeichnen.

3. Ermitteln Sie das arithmetische Mittel und die empirische Standardabweichung *bezogen auf die Anzahl* der entwickelten Bilder für das Jahr 2001.
   Die Entwicklung eines Bildes kostete 0,12 DM. Jeden Monat wurde zusätzlich eine Pauschale von 24,60 DM für die Entwicklung aller Filme erhoben (die bereits im Monatspreis enthalten ist).

   Lösung: Die Umrechnung stellt eine einfache lineare Transformation dar. Die transformierten Daten werden mit y bezeichnet.

   $\overline{y} = \frac{\overline{x}-24,6}{0,12} = \frac{211,75-24,6}{0,12} \approx 1559,58$

   $s_y = \frac{s_x}{0,12} = \frac{62,31}{0,12} = 519,25$

Für die Jahre 1999 und 2000 liegen die monatlichen Rechnungsbeträge (in DM) nur in gruppierter Form vor:

Gruppe j von... - unter...	Beobachtungen h_j
50 - 150	8
150 - 200	8
200 - 250	2
250 - 500	4
500 - 800	2

4. Fügen Sie die Daten für das Jahr 2001 hinzu und erstellen Sie das gemeinsame Histogramm über alle drei Jahre. Charakterisieren Sie die Form der Verteilung und vergleichen Sie Ihr Ergebnis mit dem aus Aufgabenteil 2.

   Lösung: Für alle drei Jahre ergeben sich folgende absolute Häufigkeiten:

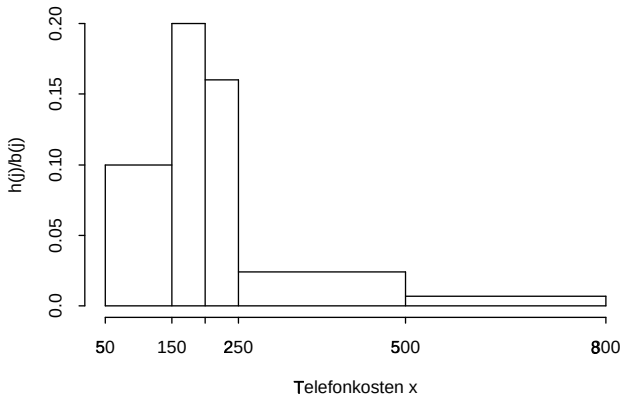
Gruppe j	Beobachtungen h_j
50 - 150	$8 + 2 = 10$
150 - 200	$8 + 2 = 10$
200 - 250	$2 + 6 = 8$
250 - 500	$4 + 2 = 6$
500 - 800	$2 + 0 = 2$

   Nach dem Prinzip der Flächentreue ergibt sich die Höhe l_j für jede Gruppe:

Lösung:

Gruppe	$l_j = \frac{h_j}{b_j}$
50 - 150	0,1
150 - 200	0,2
200 - 250	0,16
250 - 500	0,024
500 -800	0,0067

Damit ergibt sich das folgende Histogramm:



Die Form der Verteilung hat sich durch die zusätzlichen Beobachtungen geändert. Nun ist sie links-steil, bzw. rechts-schief.

Aufgabe 3 *(20 Punkte)*

Assistent A. soll das Skript des Professors "Grundlagen der Statistik" auf Fehler prüfen. Da A. schon lange für den Professor arbeitet weiß er, dass er bei den ihm vorgelegten Texten mit durchschnittlich 1 Fehler je Seite rechnen muss. Man weiß, dass die Fehler unabhängig voneinander auftreten.

1. Die zufällige Variable X sei die Anzahl der Tippfehler auf einer beliebigen Seite des Skripts. Geben Sie ihre Verteilung und Parameter an.

   Lösung:

   $$X \sim Pois\left(\lambda\right), \text{ mit } \lambda = 1$$

2. Stellen Sie die Likelihood-Funktion für den Erwartungswert der Anzahl Fehler pro Seite auf und leiten sie allgemein für eine Stichprobe von n Seiten den ML-Schätzer her.

   Lösung:

   $$L\left(\lambda|k_i\right) = \prod_{i=0}^{n} \lambda^{k_i} \cdot e^{-\lambda}$$

   $$\frac{\partial lnL\left(\lambda|k_i\right)}{\partial \lambda} = \sum_{i=0}^{n} k_i \frac{1}{\lambda} - 1$$

   $$0 = \sum_{i=0}^{n} k_i \frac{1}{\lambda_{ML}} - 1$$

   $$1 = \sum_{i=0}^{n} k_i \frac{1}{\lambda_{ML}} = \frac{\sum k_i}{\sum \lambda_{ML}} = \frac{\sum k_i}{n \cdot \lambda_{ML}}$$

   $$\lambda_{ML} = \frac{\sum k_i}{n} = \overline{k}$$

3. Zeigen Sie, dass dieser ML-Schätzer erwartungstreu ist.

   Lösung:

   $$E\left(\overline{k}\right) = E\left(\frac{1}{n} \cdot \sum_{i=0}^{n} k_i\right) = \frac{1}{n} \cdot n \cdot E\left(k_i\right) = \frac{n}{n} \cdot \lambda = \lambda$$

   Der Erwartungswert des Mittelwertes der Beobachtungen entspricht dem "wahren" Parameter.

4. Berechnen Sie die Wahrscheinlichkeit, dass auf einer beliebigen Seite die A. liest mindestens ein Tippfehler enthalten ist.

   Lösung:

   $$P\left(X \geq 1\right) = 1 - P\left(X = 0\right)$$

   $$P\left(X = 0\right) = e^{-1} = 0,368$$

   $$P\left(X \geq 1\right) = 1 - 0,368 = 0,632$$

Fehler pro Seite	0	1	2	3	4
Anzahl der Seiten	20	10	15	40	15

5. Im vorliegenden Skript wurden folgende Fehlerzahlen gefunden:
   Berechnen Sie den ML-Schätzwert.

   Lösung:
   $$\lambda_{ML} = \frac{0 \cdot 20 + 1 \cdot 10 + 2 \cdot 15 + 3 \cdot 40 + 4 \cdot 15}{20 + 10 + 15 + 40 + 15} = \frac{220}{100} = 2,2$$

6. Interpretieren Sie Ihr Ergebnis.

   Lösung: Das Stichprobenergebnis lässt eine höhere Fehlerrate pro Seite plausibler erscheinen. Für die Person, die das Skript geschrieben hat, gilt offenbar ein anderes λ.

7. Wie groß ist die Wahrscheinlichkeit, dass im neuen Skript zur Vorlesung "Bayesianische Statistik", welches der Professor gerade schreibt und das 100 Seiten umfassen soll, kein Fehler enthalten ist. Welche Zufallsvariable haben Sie verwendet.

   Lösung:
   $$Y \sim B(n, \pi), \text{ mit } n = 100 \text{ und } \pi = 0,368$$
   $$P(Y = 0) = 0,386^{100} \approx 0$$

Aufgabe 4 *(20 Punkte)*

Wissenschaftler W. ist stolz, dass er nun schon seit mehreren Jahren sein Idealgewicht von $\mu^* = \mathbf{85\,kg}$ halten kann. Als er am 1. Februar von einer längeren Forschungsreise aus Island zurückkehrt, vermutet er, dass die dortigen Ernährungsgewohnheiten Einfluss auf sein Gewicht hatten.

Zu diesem Zweck misst er an den folgenden 5 Tagen sein Gewicht. Dabei erhält er folgende Werte:

Datum	Gewicht in kg
2. 2.	89
3. 2.	88
4. 2.	88
5. 2.	89
6. 2.	90

Hinweis: Nehmen Sie an, dass Ws täglich ermitteltes Gewicht (G_i) zufällig um sein wahres Gewicht (μ_W) schwankt:

$$G_i = \mu_w + \varepsilon_i \text{ mit } \varepsilon_i \overset{i.i.d.}{\sim} N\left(0\,;\,\sigma^2\right)$$

1. Schätzen Sie Ws mittleres Gewicht und die Varianz der Messungen aus den gegebenen Daten.

   Lösung:

$$\overline{X}_w = \frac{89 + 88 + 88 + 89 + 90}{5} = 88,8\,\text{kg}$$

$$\hat{\sigma}^2 = \frac{1}{n-1}\sum_{\text{i}=1}^{5}\left(X_i - \overline{X}_W\right)^2 = \frac{1}{4}\left[(0,2)^2 + (-0,8)^2 + (-0,8)^2 + (0,2)^2 + (1,2)^2\right] = 0,7$$

Mit Hilfe eines statistischen Tests will W entscheiden, ob er von einer signifikanten Veränderung seines Gewichts ausgehen muss. Der Test soll dabei ein Niveau von $\alpha = \mathbf{0{,}01}$ haben.

2. Stellen Sie die Hypothesen zu diesem Test auf und begründen Sie die Hypothesenwahl.

   Lösung:

$$H_0:\ \mu_W = \mu^* = 85\,\text{kg}$$

$$H_1:\ \mu_W \neq \mu^* = 85\,\text{kg}$$

   Bei diesem Test handelt es sich um einen zweiseitigen Test, da die (positive oder negative) Veränderung des Gewichts untersucht werden soll. Demnach gibt es nur eine Möglichkeit zur Wahl der Hypothesen. Eine Ablehnung von H_0 untermauert die Annahme, dass W nicht mehr sein Idealgewicht von $\mu^* = 85\,\text{kg}$ besitzt.

3. Führen Sie den Test durch. Ermitteln Sie dazu die Prüfgröße und deren Verteilung. Fällen Sie anschließend die Testentscheidung. Interpretieren Sie Ihre Testentscheidung.

   Lösung: Die Prüfgröße ist in diesem Fall:

   $$\overline{X}_W = 88,8$$

   Für die Verteilung der Prüfgröße gilt:

   $$t_{PG} \sim t_{(n-1)}$$

   Für den Annahmebereich gilt folglich:

   $$\text{AB} = \left[\mu^* - \frac{\hat{\sigma}}{\sqrt{n}} \cdot t_{(n-1)}^{(1-\frac{\alpha}{2})}; \mu^* + \frac{\hat{\sigma}}{\sqrt{n}} \cdot t_{(n-1)}^{(1-\frac{\alpha}{2})}\right]$$

   Das gesuchte Quantil der t-Verteilung lautet:

   $$t_{(n-1)}^{(1-\frac{\alpha}{2})} = t_{(4)}^{(0,995)} = 4,6041$$

   Für den Annahmebereich gilt ergibt sich:

   $$\text{AB} = \left[85 - \frac{\sqrt{0,7}}{\sqrt{5}} \cdot 4,6041; 85 + \frac{\sqrt{0,7}}{\sqrt{5}} \cdot 4,6041\right]$$

   $$\text{AB} = [83,28\,;\,86,72]$$

   Die Prüfgröße liegt außerhalb des Annahmebereichs. H_0 muss also verworfen werden.

   Es ist von einer signifikanten Veränderung von Ws Köpergewicht auszugehen. Die 5 Messungen sprechen gegen einen Erwartungswert des Köpergewichts von 85 kg. W hat offenbar zugenommen.

   ALTERNATIVE:

   Die standardisierte Prüfgröße lautet:

   $$t_{PG} = \frac{\overline{X}_W - \mu^*}{\hat{\sigma}} \cdot \sqrt{n}$$

   $$t_{PG} = \frac{88,8 - 85}{\sqrt{0,7}}\sqrt{5} \approx 10,16$$

   Der Annahmebereich lautet dann:

   $$\text{AB} = [-4,6041\,;\,4,6041]$$

   Die standardisierte Prüfgröße liegt außerhalb dieses Bereichs.

4. Vergleichen Sie die Konstruktion des Annahmebereichs des oben durchgeführten Tests mit der Konstruktion eines Konfidenzintervalls für Ws Gewicht. Welche Gemeinsamkeiten, welche Unterschiede gibt es?

Lösung: Konfidenzintervall und Signifikanztest werden auf sehr ähnliche Weisen berechnet. Hauptunterschied: Mittelpunkt eines Konfidenzintervalls ist der aus der Stichprobe vorgegebene Schätzwert; Mittelpunkt des Annahmebereichs eines zweiseitigen Tests ist dagegen der Parameterwert unter der Nullhypothese.

Aufgabe 5 *(20 Punkte)*

An der Gauß-Universität muss ein Student der Informatik eine Klausur in Mathematik und eine im Fach Statistik schreiben. Die Resultate des letzten Semester ergaben, dass 63% der Studenten die Statistik-Klausur bestanden haben, 58% die Mathematik-Klausur und 68% der Studenten wenigstens eine der beiden Klausuren bestanden haben.

1. Wie groß ist die Wahrscheinlichkeit, dass ein zufällig befragter Student in beiden Klausuren erfolgreich war und wie groß , dass er in keiner der beiden Klausuren erfolgreich war.

   Lösung:

   $$P(S \cup M) = 0,68$$
   $$P(S \cap M) = P(S) + P(M) - P(S \cup M) = 0,63 + 0,58 - 0,68 = 0,53$$

   Die Wahrscheinlichkeit, dass der Student in beiden Klausuren erfolgreich war beträgt 0,53.

   $$P\left(\overline{S} \cap \overline{M}\right) = P\left(\overline{S \cup M}\right) = 0,32$$

   Die Wahrscheinlichkeit, dass der Student in beiden Klausuren nicht erfolgreich war beträgt 0,32.

2. Wie groß ist die Wahrscheinlichkeit, dass ein befragter Student, der in der Statistik-Klausur keinen Erfolg hatte, die Mathematik-Klausur erfolgreich bestanden hat?

   Lösung:

   $$P\left(\overline{S}\right) = 0,37$$

   $$P\left(M|\overline{S}\right) = 1 - P\left(\overline{M}|\overline{S}\right)$$
   $$P\left(\overline{M}|\overline{S}\right) = \frac{P\left(\overline{S} \cap \overline{M}\right)}{P\left(\overline{S}\right)} = \frac{0,32}{1 - 0,63} = 0,86$$
   $$P\left(M|\overline{S}\right) = 1 - P\left(\overline{M}|\overline{S}\right) = 1 - 0,86 = 0,14$$

3. In diesem Semester haben 73% der zur Statistik-Klausur angemeldeten Studenten die Übung zur Statistik besucht. Aus Erfahrung weiß man, dass ein Student der die Übung besucht hat mit einer Wahrscheinlichkeit von 81% die Klausur besteht, ein Student, der die Übung nicht besucht hat jedoch nur mit 31%iger Wahrscheinlichkeit erfolgreich ist. Mit welchem Anteil an bestandenen Statistik-Klausuren kann die Gauß -Universität in diesem Semester rechnen?

   Lösung:

   $$P(U) = 0,73$$
   $$P\left(\overline{U}\right) = 1 - 0,73 = 0,27$$
   $$P(S|U) = 0,81$$
   $$P\left(S|\overline{U}\right) = 0,31$$
   $$P(S) = P(U) \cdot P(S|U) + P\left(\overline{U}\right) \cdot P\left(S|\overline{U}\right) = 0,73 \cdot 0,81 + 0,27 \cdot 0,31 = 0,675$$

   Die Universität kann mit einem Anteil von 0,65 rechnen.

4. Haben Sie in diesem Semester die Übung besucht?
   Lösung: Wahrscheinlich nicht.