

Fachbereich Mathematik
M. Kohler
A. Fromkorth
J. Mehnert



TECHNISCHE
UNIVERSITÄT
DARMSTADT

SS 2009
27.07.2009

Klausur
„Forschungsmethoden I f. BSc. Psychologie
Statistik für BA Päd.
Stat. Datenanal. I f. JBA. Soziologie
Statistik für Human- und Sozialwissenschaften “

Name: | Vorname:
Matrikel-Nr.: | Studiengang:

Aufgabe	1	2	3	4	5	max	Note
Bearbeitet (ankreuzen)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	4 (3) / 5 Aufgaben	
Punktzahl	10	10	10	10	10	40 (30)	
erreichte Punktzahl							

Hinweise

1. TeilnehmerInnen der Modulprüfung „**Statistik für BA Päd.**“ **müssen 3 der 5 Aufgaben bearbeiten**. Für **alle anderen** im Titel genannten Prüfungen ist die Bearbeitung von **4 der 5 Aufgaben** erforderlich. Kreuzen Sie in obiger Tabelle an, welche Aufgaben bewertet werden sollen. **TeilnehmerInnen der Modulprüfung „Forschungsmethoden I f. BSc. Psychologie“ müssen darüber hinaus die zusätzlich ausgeteilte Scheinklausur bearbeiten (und bestehen)**.
2. Als Hilfsmittel ist ein Taschenrechner erlaubt. Sie benötigen außerdem eigenes (unbeschriebenes) Papier und Schreibzeug. Weitere Hilfsmittel sind nicht zugelassen. Insbesondere dürfen Sie keine Mobiltelefone verwenden.
3. Überprüfen Sie die Aufgabenblätter zu Klausurbeginn auf Vollständigkeit. Die Klausuraufgabenstellung umfasst inklusive dieses Deckblatts 4 Seiten und beinhaltet 5 Aufgaben. Hinzu kommt für die TeilnehmerInnen der Klausur „Forschungsmethoden I f. BSc. Psychologie“ die Aufgabenstellung der Scheinklausur.
4. Schreiben Sie auf jedes Blatt deutlich Ihren Namen und Ihre Matrikelnummer.
5. Tragen Sie die genaue Bezeichnung Ihres Studienganges ein. Aus der Bezeichnung muss hervorgehen, an welcher Klausur Sie teilnehmen.
6. Dauer der Modulprüfung „**Statistik für BA Päd.**“ ist **90 Minuten**. Dauer der Modulprüfung „**Forschungsmethoden I f. BSc. Psychologie**“ ist 120 Minuten, hinzu kommen noch 30 Minuten für die Scheinklausur, also **insgesamt 150 Minuten**. Die Prüfungszeit für **alle anderen** im Titel genannten Prüfungen beträgt **120 Minuten**.
7. Es wird auf der Homepage der Vorlesung „Statistik für Human- und Sozialwissenschaften“ (WS 08/09) bekannt gegeben, wenn die Resultate vorliegen. Diese werden dann in einem der Schaukästen im 3. Stock des Mathematikgebäudes (Geb. S2 15) ausgehängt.

1. Aufgabe

(10 Punkte)

- (a) Erläutern Sie **kurz**, was man bei einer Umfrage unter dem Begriff „Verzerrung durch Auswahl“ (sampling bias) und „Verzerrung durch Nicht-Antworten“ (non-response bias) versteht.
- (b) Im Rahmen einer Umfrage sollen Professoren an deutschen Universitäten zum Thema „Der Umgang deutscher Professoren mit Englisch als internationale Wissenschaftssprache“ befragt werden. Dabei sollen ausgewählte Professoren per Email angeschrieben werden. Für die Auswahl der Professoren werden die beiden folgenden Verfahren betrachtet:
- Auswahl aller Mathematik-Professoren an deutschen Universitäten.
 - Für jede deutsche Universität werden rein zufällig 10 Prozent aller Professoren dieser Universität ausgewählt.

Vergleichen Sie diese beiden Vorgehen **kurz** hinsichtlich sampling bias, non-response bias und Aufwand bei der Ermittlung der Email - Adressen.

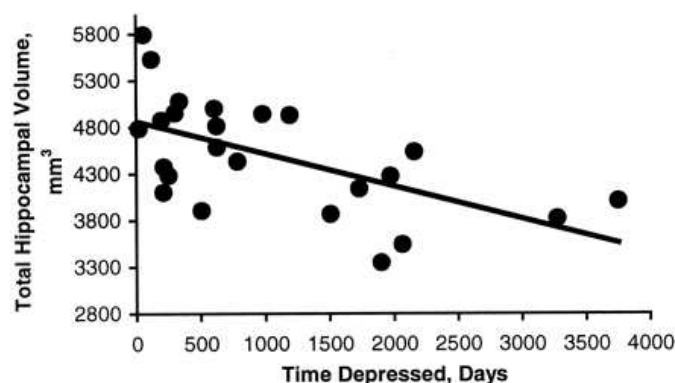
2. Aufgabe

(10 Punkte)

- (a) Das “United States Census” Büro hat mit einer Stichprobe der arbeitenden US-Bevölkerung ermittelt, wie lange der Arbeitsweg dieses Personenkreises ist. Dazu wurde folgende Häufigkeitstabelle erstellt:

Arbeitsweg [in Minuten]	Anzahl der Personen [Absolut]
[0, 30)	81290
[30, 45)	23703
[45, 150)	19096

- Zeichnen Sie hierzu ein Histogramm.
 - Begründen Sie anhand des Histogrammes warum mehr als die Hälfte der Personen weniger als 50 Minuten benötigt um zur Arbeit zu kommen.
- (b) Eine Gruppe von Versuchspersonen wurde nach der Anzahl von Tagen gefragt, an welchen diese depressiv waren. Zudem wurde ihr Gehirnvolumen gemessen. Folgender Graph mit eingezeichneter Regressionsgerade war das Ergebnis:



- Welches Vorzeichen hat die empirische Kovarianz?
- Liegt hier die empirische Korrelation im Intervall $(0, 1]$ oder im Intervall $[-1, 0)$? Warum kann die empirische Korrelation hier nicht gleich 0 sein?

3. Aufgabe

(10 Punkte)

Seien $a, b \in \mathbb{R}$ mit $a < b$. Dann ist die Dichte der Gleichverteilung auf dem Intervall $[a, b]$ gegeben durch

$$f: \mathbb{R} \rightarrow \mathbb{R}, \quad f(x) = \begin{cases} \frac{1}{b-a} & \text{falls } x \in [a, b] \\ 0 & \text{sonst} \end{cases}$$

- (a) Sei X eine auf dem Intervall $[1, 2]$ gleichverteilte Zufallsvariable. Bestimmen Sie den Erwartungswert und die Varianz von X .
- (b) Sei Y eine auf dem Intervall $[2, 4]$ gleichverteilte Zufallsvariable. Bestimmen Sie den Erwartungswert von Y^2 .
- (c) Bestimmen Sie den Erwartungswert von $2X + 3Y^2$, wobei X und Y wie in den Aufgabenteilen (a) bzw. (b) definiert sind.

Hinweis Die Varianz einer Zufallsvariable X ist definiert als

$$\mathbf{V}(X) = \mathbf{E} \left((X - \mathbf{E}X)^2 \right).$$

4. Aufgabe

(10 Punkte)

Bei der Shell Jugendstudie 2006 wurden 1301 Jungen befragt. Dabei gaben 47 Prozent der befragten Jungen an, das Abitur anzustreben.

Leiten Sie aus dem zentralen Grenzwertsatz, aus dem

$$P \left[\left| \frac{\sqrt{n}}{\sqrt{V(X_1)}} \cdot \left(\frac{1}{n} \sum_{i=1}^n X_i - EX_1 \right) \right| \leq 1.96 \right] \approx 0.95$$

für unabhängig identisch verteilte reelle Zufallsvariablen X_1, \dots, X_n mit $0 < V(X_1) < \infty$ folgt, ein approximatives zweiseitiges Konfidenzintervall zum Konfidenzniveau 0.95 für den Anteil der Jungen her, die das Abitur anstreben. Beachten Sie dabei, dass für $x_1, \dots, x_n \in \{0, 1\}$ gilt:

$$\begin{aligned} \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 &= \frac{n}{n-1} \left(\frac{1}{n} \sum_{i=1}^n x_i^2 - (\bar{x})^2 \right) \\ &= \frac{n}{n-1} \left(\bar{x} - (\bar{x})^2 \right), \end{aligned}$$

wobei

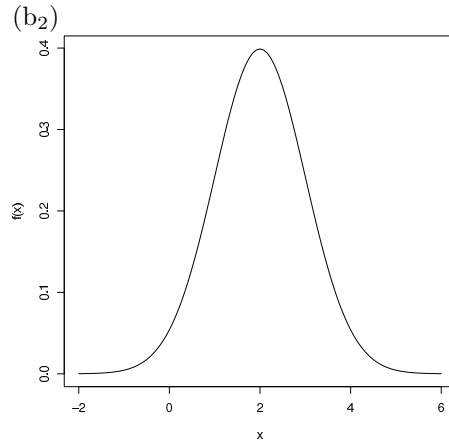
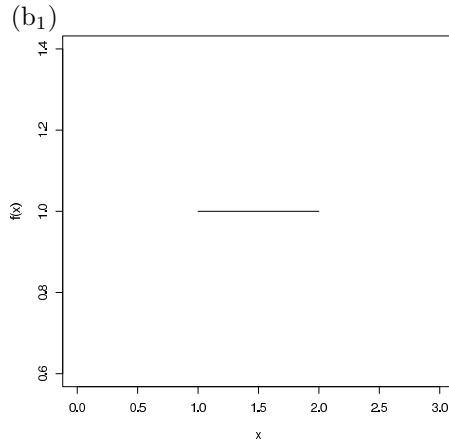
$$\bar{x} = \frac{1}{n} \sum_{j=1}^n x_j.$$

5. Aufgabe

(10 Punkte)

- (a) Angenommen, die Lotto-Zahlen beim Lotto 6 aus 49 werden unbeeinflusst voneinander jede Woche rein zufällig aus den Zahlen von 1 bis 49 ausgewählt. Ist es dann wahrscheinlicher, dass nächste Woche die Zahlen 1, 9, 10, 15, 32, 37 gezogen werden, oder dass nächste Woche die gleichen Zahlen wie in dieser Woche gezogen werden?

- (b) Welche der beiden Dichten



gehört zu einer Normalverteilung? Was können Sie über den Erwartungswert dieser Normalverteilung aussagen?

- (c) Erläutern Sie jeweils kurz (und evtl. auch anschaulich) die Aussagen des
- (c₁) empirischen Gesetzes der großen Zahlen,
 - (c₂) starken Gesetzes der großen Zahlen,
 - (c₃) zentralen Grenzwertsatzes.