Fachbereich Mathematik
Prof. Dr. Michael Kohler
Dipl.-Math. Andreas Fromkorth
Dipl.-Inf. Jens Mehnert

TECHNISCHE
UNIVERSITÄT
DARMSTADT

WS 08/09
10.03.09

Klausur „Statistik I für Human- und Sozialwissenschaft“

Name: .... | Vorname: ....

Matrikel-Nr.: .... | Fachrichtung: ....

Aufgabe	1	2	3	4	5	$\sum$	Note
Punktzahl	10	10	10	10	10	50	
erreichte Punktzahl							

ACHTUNG: Die Prüfungsdauer hängt vom Studienfach ab

Fach	Klausurlänge in Stunden	Zusätzliche Scheinklausur ?	Aufgaben zu bearbeiten
Psychologie	$2,5$	Ja	4 von 5
Soziologie	2	Nein	4 von 5
Pädagogik	$1,5$	Nein	3 von 5
Nebenfächler	2	Nein	4 von 5

Zugelassene Hilfsmittel sind:

- Taschenrechner aller Art,
- Fremdsprachenwörterbücher.

Zudem sind alle Resultate und **Zwischenschritte** zu **begründen**. **Formeln aus der Vorlesung** dürfen ohne Begründung verwendet werden.

Sofern Sie mehr Aufgaben bearbeiten als für Ihr Studienfach vorgeschrieben sind, so notieren Sie bitte auf diesem Blatt, welche Aufgaben gewertet werden sollen.

1. Aufgabe (10 Punkte)

Um den Einfluss der Klassengröße auf Leistungen von Schülern zu ermitteln, wurde in den USA eine Studie mit 11.600 Schülern durchgeführt. Diese wurden dabei zufällig in zwei gleich große Gruppen unterteilt, und die Schüler der ersten Gruppe wurden in Klassen der Klassengröße 13 bis 17 Schüler, die der zweiten Gruppe in Klassen der Klassengröße 22 bis 25 Schüler unterteilt. Nach vier Jahren wurden die Leistungen der Schüler im Rahmen verschiedener Tests in beiden Gruppen ermittelt und miteinander verglichen.

(a) Um was für eine Studie handelt es sich im Sinne der Vorlesung?

(b) Erläutern Sie <u>kurz</u> den Begriff des "konfundierenden Faktors".

(c) Inwiefern kann man aus einer solchen Studie auf kausale Zusammenhänge zwischen der Leistung der Schüler und der Klassengröße zurückschließen?

(d) Bei obiger Studie wechselten ca. 10% der Schüler in Klassen der jeweils anderen Klassengröße. Inwiefern beeinflusst das die Interpretation der Ergebnisse der Studie?

2. Aufgabe (10 Punkte)

Abbildung 1

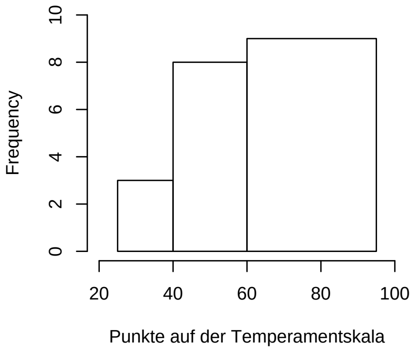


Abbildung 2

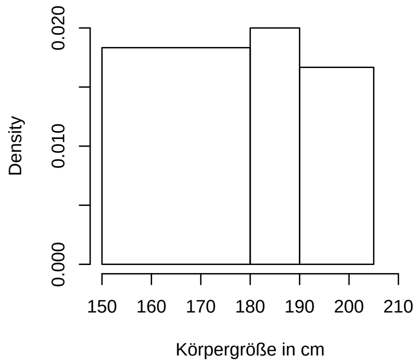


Abbildung 3

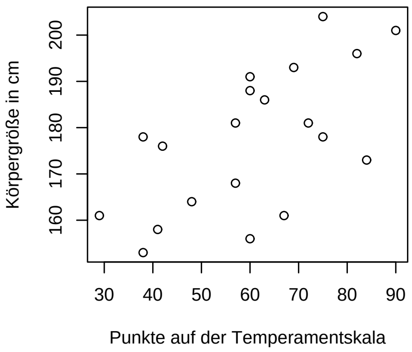
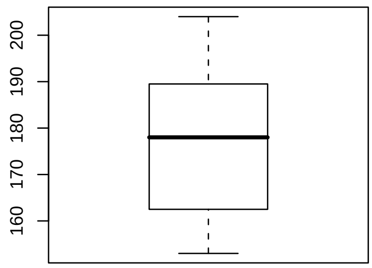


Abbildung 4



Im Rahmen einer psychologischen Studie wird Personen ein Temperamentwert zugeordnet und die Körpergröße in cm gemessen.

(a) Das Säulendiagramm in Abbildung 1 beschreibt die Werte von $n = 20$ Personen auf der Temperamentskala.

(a_1) Inwieweit ist die Darstellung in diesem Säulendiagramm irreführend?

(a_2) Stellen Sie die Daten in einem Histogramm so dar, dass die Flächeninhalte der einzelnen Balken (aufgrund von Problemen beim Ablesen der Werte in Abbildung 1 eventuell nur ungefähr) proportional zur Anzahl der Datenpunkte in den zugrundeliegenden Intervallen sind.

(b) Das Histogramm in Abbildung 2 beschreibt die Körpergröße von $n = 20$ Personen. Bestimmen Sie mit Hilfe dieses Histogramms (approximativ) die Anzahl der Personen, die eine Körpergröße von höchstens 180 cm haben.

(c) Im Streudiagramm in Abbildung 3 sind die Körpergröße gegen den Wert auf der Temperamentskala abgetragen. Ist die Korrelation zwischen diesen Werten größer oder kleiner als Null? Begründen Sie ihre Antwort.

(d) Der Boxplot in Abbildung 4 beschreibt die Körpergröße in cm. Wie groß ist (ungefähr) der Median und wie groß ist (ungefähr) der Interquartilabstand dieser Daten?

3. Aufgabe (10 Punkte)

(a) Die Funktion $f : \mathbb{R} \to \mathbb{R}$ ist definiert durch

$$f(x) = \begin{cases} 2 - 2 \cdot x & \text{für } 0 \leq x \leq 1 \\ 0 & \text{für x<0 oder } x > 1. \end{cases}$$

Skizzieren Sie f und begründen Sie (z.B. anhand der Skizze), dass f eine Dichte ist.

(b) Die Zufallsvariable X sei stetig verteilt mit der Dichte f aus Teil a). Wie können Sie

$$P[X \geq \frac{1}{2}]$$

anhand der in Teil a) angefertigten Skizze bestimmen?
Bestimmen Sie diesen Wert.

(c) Bestimmen Sie den Erwartungswert der Zufallsvariable X aus Teil b).

(d) Bestimmen Sie die Varianz der Zufallsvariable X aus Teil b).

(e) Begründen Sie mit Hilfe der Rechenregeln für den Erwartungswert und aus der Vorlesung bekannten Formeln:
Für eine beliebige reelle Zufallsvariable Z, für die $V(Z)$ existiert, gilt immer:

$$V(Z) = E(Z(Z-1)) + E(Z) - (E(Z))^2.$$

4. Aufgabe (10 Punkte)

Bei der Shell Jugendstudie 2006 wurden 1231 Mädchen befragt. Dabei gaben 55 Prozent der befragten Mädchen 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 $X_1, \ldots, X_n$ mit $0 < V(X_1) < \infty$ folgt, ein approximatives zweiseitiges Konfidenzintervall zum Konfidenzniveau 0.95 für den Anteil der Mädchen her, die das Abitur anstreben. Beachten Sie dabei, dass für $x_1, \ldots, x_n \in \{0, 1\}$ gilt:

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

wobei

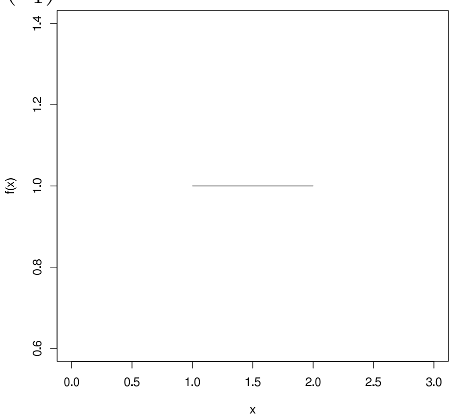
$$\bar{x}=\frac{1}{n}\sum_{j=1}^{n}x_j.$$

5. Aufgabe (10 Punkte)

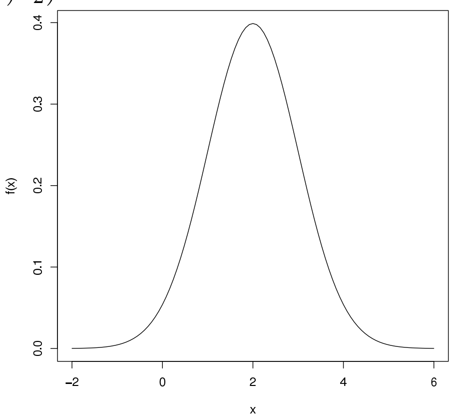
(a) Angenommen, die Lotto-Zahlen beim Lotto 6 aus 49 werden unbeeinflusst voneinander jede Woche rein zufällig aus den Zahlen von 1 bis 49 ausgewählt.
Ist es dann wahrscheinlicher, dass nächste Woche die Zahlen $1, 9, 10, 15, 32, 37$ gezogen werden, oder dass nächste Woche die gleichen Zahlen wie in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

(b) Welche der beiden Dichten

(b$_1$)



)b$_2$)



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

(c$_1$) empirischen Gesetzes der großen Zahlen,

(c$_2$) starken Gesetzes der großen Zahlen,

(c$_3$) zentralen Grenzwertsatzes.