Fachbereich Mathematik
Prof. Dr. Michael Kohler
Andreas Gärtner
Jens Mehnert

TECHNISCHE
UNIVERSITÄT
DARMSTADT

WS 09/10
11./12./13.11.2009

# 2. Übungsblatt zur „Statistik I für Human- und Sozialwissenschaft“

**Aufgabe 5** (3 Punkte)

Zur Evaluierung des Nutzens des Zeitschriftenbestandes hat die Universitätsbibliothek der Universität Stuttgart eine Umfrage durchgeführt. Befragt wurden dabei alle Personen, die entweder die im Lesesaal vorhandenen Zeitschriften oder den elektronischen Zugang über das Internet zu den elektronischen Ausgaben dieser Zeitschriften nutzten. Die Befragung erfolgte durch Ausfüllen eines im Internet bereitgestellten Formulars.
Zur Auswahl der Befragten wurden im Vorfeld der Umfrage unter anderem die beiden folgenden Möglichkeiten diskutiert:

(a) Anschreiben aller Institute der Universität Stuttgart mit der Bitte, alle Mitarbeiter auf die Umfrage aufmerksam zu machen und um Ausfüllen der Fragebögen zu bitten.

(b) Anschreiben aller Institute der Universität Stuttgart mit der Bitte, den Fragebogen durch den Institutsdirektor ausfüllen zu lassen.

Was können Sie über den *sampling bias* und den *non response bias* bei den beiden Umfragearten aussagen ?

**Lösung:**

(a) **sampling bias:** Tritt bei b) auf, da die Institutsdirektoren in der Regel nicht die Nutzer und auch nicht repräsentativ für die Nutzer sind.

(b) **non response bias:** tritt bei a) und b) auf, da davon auszugehen ist, dass nicht jeder Mitarbeiter die Zeit hat diese Umfrage auszufüllen. So gesehen tritt eine Verzerrung durch Nichtbeantworten bei b) wahrscheinlich sogar noch mehr auf, da die Institutsdirektoren in der Regel noch weniger Zeit haben als deren Mitarbeiter.

**Aufgabe 6** (3 Punkte)

Eine Messung des Intelligenzquotienten von Schülern einer Gesamtschule ergab folgende Messreihe:

$$101, 105, 98, 111, 89, 110, 112, 118, 101, 97, 121, 99,$$
$$97, 113, 132, 103, 91, 87, 100, 85, 115, 101, 96, 102$$

Stellen Sie die Messergebnisse in einem Histogramm dar. Verwenden Sie die Intervallunterteilung

$$(80, 85],\ (85, 95],\ (95, 115],\ (115, 135].$$

**Lösung:** Entsprechend der Vorgehensweise aus der Vorlesung erhält man:

| Intervall | Datenpunkte im Intervall | Anzahl der Datenpunkte im Intervall |
|---|---|---|
| $(80, 85]$ | 85 | 1 |
| $(85, 95]$ | 89, 91, 87 | 3 |
| $(95, 115]$ | 101, 105, 98, 111, 110, 112, 101, 97, 99, 97, 113, 103, 100, 115, 101, 96, 102 | 17 |
| $(115, 135]$ | 118, 121, 132 | 3 |

Somit gilt (mit den Bezeichnungen aus der Vorlesung):

$$
\begin{aligned}
I_1 = (80, 85], n_1 = 1, \lambda(I_1) = 85 - 80 = 5 \\
\Rightarrow \text{Wert über Intervall } I_1 = \frac{1}{24 \cdot 5} = \frac{1}{120} = \frac{4}{480} \\
I_2 = (85, 95], n_2 = 3, \lambda(I_2) = 95 - 85 = 10 \\
\Rightarrow \text{Wert über Intervall } I_2 = \frac{3}{24 \cdot 10} = \frac{1}{80} = \frac{6}{480} \\
I_3 = (95, 115], n_3 = 17, \lambda(I_3) = 115 - 95 = 20 \\
\Rightarrow \text{Wert über Intervall } I_3 = \frac{17}{24 \cdot 20} = \frac{17}{480} \\
I_4 = (115, 135], n_4 = 3, \lambda(I_4) = 135 - 115 = 20 \\
\Rightarrow \text{Wert über Intervall } I_4 = \frac{3}{24 \cdot 20} = \frac{3}{480}
\end{aligned}
$$

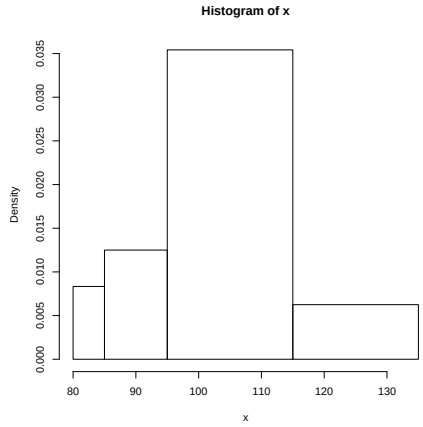



Abbildung 1: Histogramm aus Aufgabe 6

**Aufgabe 7** (3 Punkte)

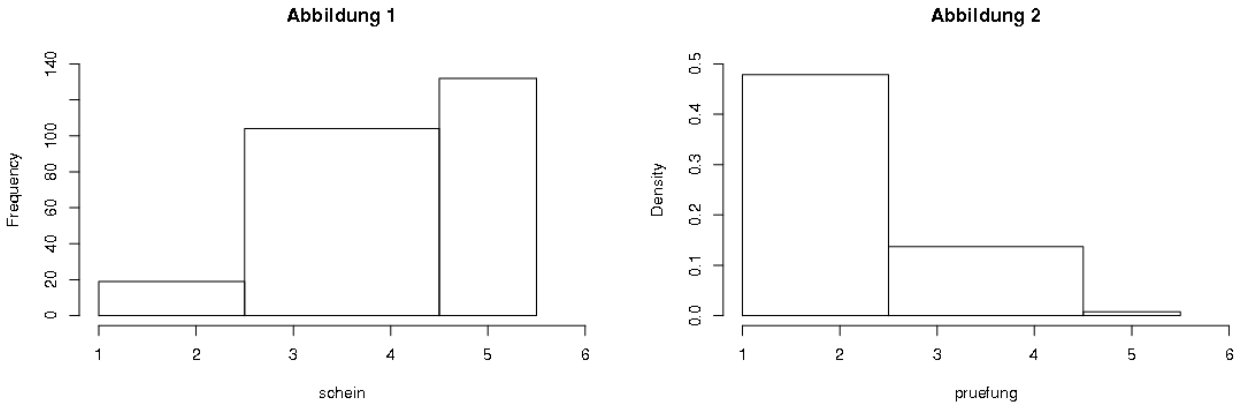
In der Tageszeitung "Darmstädter Echo" vom 08.10.2009 war unter der Überschrift Fast jeder Dritte wird im Winter trübsinnig das Folgende zu lesen:

> Fast jeder dritte Deutsche leidet im Winter unter Stimmungsschwankungen, Konzentrationsschwäche und Müdigkeit, wie eine Forsa-Umfrage im Auftrag der Techniker Krankenkasse ergab. Frauen sind stärker betroffen als Männer. 36 Prozent gaben an, in der dunklen Jahreszeit in ein Stimmungstief zu fallen, dagegen nur jeder vierte Mann. Forsa befragte 1026 Deutsche ab 18 Jahren.
>
> Ausgelöst werde das Tief hauptsächlich durch Lichtmangel, sagt der Psychologe York Scheller. Ohne Licht schüttet seinen Angaben zufolge der Körper weniger vom Glückshormon Serotonin aus.

Welche Aussage können Sie aufgrund dieses Artikels über den sampling bias bzw. den non-response bias bei dieser Umfrage machen? Begründen Sie ihre Antwort.

**Lösung:** Es läßt sich weder eine Aussage über den sampling bias noch über den non-response bias machen, da der Artikel weder Information über die Art der Erhebung der Umfrage noch über die Auswahl der befragten Personen enthält.

**Aufgabe 8** (3 Punkte)




**a)** Das Säulendiagramm in Abbildung 1 beschreibt die Noten von $n = 255$ Studenten in der Scheinklausur zur Vorlesung "Statistik".

**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 Noten von $n = 255$ Studenten in der Diplom-Vorprüfung zur Vorlesung "Statistik". Bestimmen Sie mit Hilfe dieses Histogramms (approximativ) die Anzahl der Studenten, die in der Prüfung eine Note besser als 2.5 hatten (d.h., die als Note eine 1.0, 1.3, 1.7, 2.0 oder 2.3 hatten).

**Lösung:**

**a) a$_1$)** Die graphische Darstellung ist irreführend, da die Klassen, bzw. die Länge der Intervalle nicht alle gleich lang sind. Vergleicht man beispielsweise den Flächeninhalt der mittleren mit der rechten Klasse, so entsteht der Eindruck, dass die mittlere Klasse fast anderthalb mal so viele Datenpunkte enthält wie die rechte Klasse und das ist falsch!

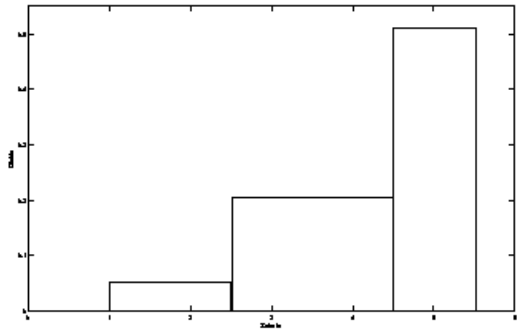
**a$_2$)** Aus der Abbildung lesen wir ab: Anzahl der Studenten in $I_1 = [1, 2.5)$ ungefähr 20, in $I_2 = [2.5, 4.5]$ ungefähr 105 und in $I_3 = [4.5, 5.5)$ ungefähr 130. Damit berechnen sich die Werte, die über $I_1$, $I_2$ und $I_3$ abgetragen werden zu

$$\frac{n_1}{n \cdot \lambda(I_1)} = \frac{20}{255 \cdot 1.5} = 0.052$$

$$\frac{n_2}{n \cdot \lambda(I_2)} = \frac{105}{255 \cdot 2} = 0.206$$

$$\frac{n_3}{n \cdot \lambda(I_3)} = \frac{130}{255 \cdot 1} = 0.510$$

und es ergibt sich folgendes Histogramm:



**b)** Beim Histogramm gibt der Flächeninhalt (FI) einer Klasse $j$ den prozentualen Anteil der Datenpunkte (PAD) im zugrunde liegenden Intervall ($I_j$) an. Somit lässt sich die Anzahl der Datenpunkte in Klasse $j$ wie folgt berechnen:
1. Möglichkeit:

$$\text{PAD in } I_j = \text{FI von } I_j$$
$$\Rightarrow \text{Anzahl der Datenpunkte in } I_j = \text{GD x PAD in } I_j$$

mit GD = Gesamtzahl der Datenpunkte.
2. Möglichkeit:

$$\text{Höhe von } I_j = \frac{n_j}{n \cdot \lambda(I_j)}$$
$$\Rightarrow \quad n_j = n \cdot \lambda(I_j) \cdot \text{Höhe von } I_j$$

mit $n_j$ = Anzahl der Datenpunkte im $j$-ten Intervall und $\lambda(I_j)$ = Länge des $j$-ten Intervalls. Als Höhe des Intervalls $I_1$ liest man 0.48 ab. Damit ergibt sich

$$n_1 = n \cdot \lambda(I_1) \cdot \text{Höhe von } I_1 = 255 \cdot 1.5 \cdot 0.48 = 183.6,$$

also $n_1 \approx 184$.