



Lösungsvorschläge zum 3. Übungsblatt zur „Statistik I für Human- und Sozialwissenschaft“

Lösung zur Aufgabe 9

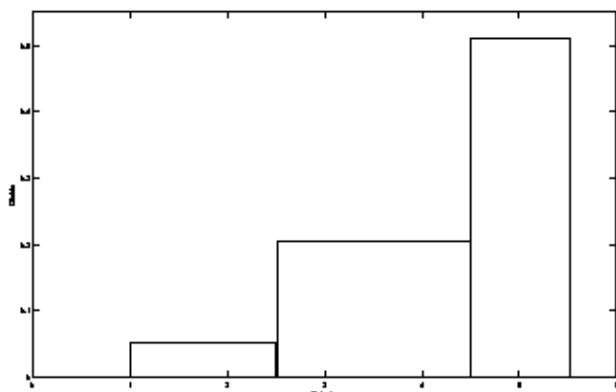
(3 Punkte)

- (a) Wird in der Abbildung die rechte obere mit der rechten unteren Graphik verglichen, dann ist festzustellen, dass der relative Anteil bei den Noten zwischen 4 und 6 in der Kontrollgruppe deutlich höher ist, als in der nicht ausgew./betracht. Stud., d.h. die betrachtete Fläche ist größer. Somit lässt sich die Aussage feststellen.
- (b) Betrachtet man die Graphik „Noten Studiengruppe“, dann ist der relative Anteil bei den Noten zwischen 4 und 6 deutlich niedriger als bei der Graphik „Noten Kontrollgruppe“ (im gleichen Notenintervall). Daher lässt sich folgern, dass das Anbieten des Zusatzkurses zu einer Verringerung der Durchfallquote bei den als durchfallgefährdet eingestuften StudentInnen geführt hat.

Lösung zur Aufgabe 10

(3 Punkte)

- a) a₁) 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!
- a₂) Anzahl der Studenten in der Klasse 1, d.h. im Intervall $I_1 (= [1, 2.5)) \approx 20$, in $I_2 \approx 105$ und in $I_3 \approx 130$. Somit 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:

$$\begin{aligned} \text{PAD in } I_j &= \text{FI von } I_j \\ \Rightarrow \text{Anzahl der Datenpunkte in } I_j &= \text{GD} \times \text{PAD in } I_j \end{aligned}$$

mit GD = Gesamtzahl der Datenpunkte

2. Möglichkeit

$$\begin{aligned} \text{Höhe von } I_j &= \frac{n_j}{n \cdot \lambda(I_j)} \\ \Rightarrow n_j &= n \cdot \lambda(I_j) \cdot \text{Höhe von } I_j \end{aligned}$$

mit n_j = Anzahl der Datenpunkte im j -ten Intervall und $\lambda(I_j)$ = Länge des j -ten Intervalls.
Hier: $n_j \approx 182$

Lösung zur Aufgabe 11

(3 Punkte)

Wir bezeichnen die einzelnen Datenwerte mit x_1, \dots, x_{20} . Die aufsteigend sortierten Werte mit $x_{(1)}, \dots, x_{(20)}$.

- (a) Da 20 Datenwerte gegeben sind, berechnet sich der Median durch

$$\frac{x_{(10)} + x_{(11)}}{2} = \frac{110.9 + 112.7}{2} = 111.8$$

Für das arithmetische Mittel gilt

$$\bar{x} = \sum_{i=1}^{20} x_i = \frac{1}{20} (131.8 + 106.7 + \dots + 112.7) = 107.21$$

Um die empirische Varianz zu berechnen kann man entweder die Formel aus der Vorlesung verwenden, d.h.

$$s^2 = \frac{1}{19} \left((131.8 - 107.21)^2 + (106.7 - 107.21)^2 + \dots + (112.7 - 107.21)^2 \right)$$

oder man berechnet zuerst

$$\sum_{i=1}^{19} x_i^2 = 131.8^2 + \dots + 112.7^2$$

und verwendet dann die Formel

$$s^2 = \frac{n}{n-1} \left(\frac{1}{n} \sum_{i=1}^n x_i^2 - \left(\frac{1}{n} \sum_{i=1}^n x_i \right)^2 \right).$$

In beiden Fällen erhält man

$$s^2 = 728.2799$$

(b) Um den Boxplot zu bestimmen, benötigen wir noch das 1. und das 3. Quartil. Unsere Messreihe hat 20 Datenpunkte. Diese sind alle verschieden. Wir suchen jetzt diejenigen Werte aus der Datenreihe für die gilt:

- i. mindestens 25 % der Datenpunkte sind kleiner oder gleich dem ausgewählten Wert und
- ii. mindestens 75 % der Datenpunkte sind größer oder gleich dem ausgewählten Wert.

Das 1. Quartil liegt also zwischen $x_{(5)}$ und $x_{(6)}$ und ist somit der Mittelwert dieser beiden Werte (also gleich $\frac{84.3+86.7}{2} = 85.5$). Entsprechend geht man für das 3. Quartil vor. Hier suchen wir diejenigen Werte, für die gilt:

- i. mindestens 75 % der Datenpunkte sind kleiner oder gleich dem ausgewählten Wert und
- ii. mindestens 25 % der Datenpunkte sind größer oder gleich dem ausgewählten Wert.

Das 3. Quartil liegt also zwischen $x_{(15)}$ und $x_{(16)}$ und ist somit der Mittelwert dieser beiden Werte (also gleich $\frac{119.2+123.4}{2} = 121.3$). Als Interquartilabstand erhalten wir also $121.3 - 85.5 = 35.8$.

Das oben beschriebene Verfahren, die Quartile zu berechnen entspricht der im Buch von Prof. Kohler beschriebenen Vorgehensweise. In der Vorlesung wurde dies etwas vereinfacht: Als 1. Quartil nimmt man

$$x(\lceil \frac{n}{4} \rceil) = x(\lceil \frac{20}{4} \rceil) = x_{(5)} = 84.3$$

und als 3. Quartil

$$x(\lceil \frac{3 \cdot n}{4} \rceil) = x(\lceil \frac{3 \cdot 20}{4} \rceil) = x_{(15)} = 119.2.$$

IQR verändert sich entsprechend. In der Klausur werden beide Vorgehensweisen akzeptiert (es sollte aber aus der Lösung ersichtlich werden, welche der beiden Definitionsmöglichkeiten verwendet wird).

Als Boxplot ergibt sich

