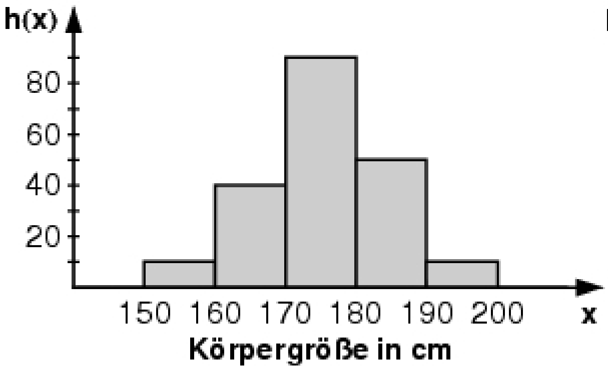
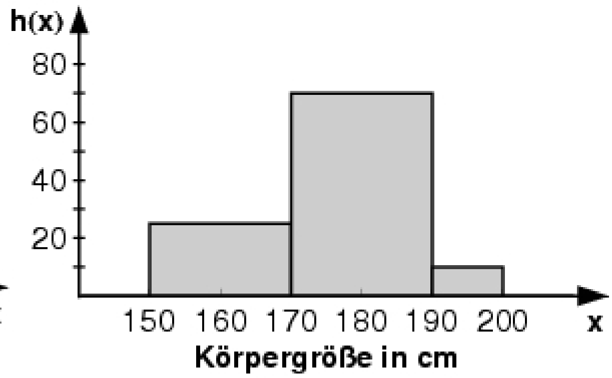
Kurs 40601 Grundlagen der Statistik

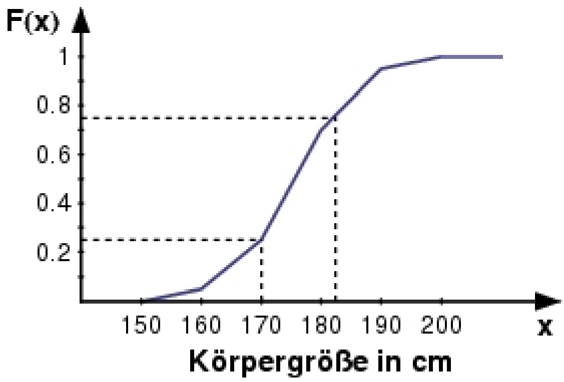
Kurseinheit 1
Lösungskommentare

Aufgabe 1

A,C und D sind richtig.







zu A: Die Klassenhäufigkeiten sind proportional zur Fläche der Rechtecke, so dass sich die Rechteckhöhe rh_j mit $rh_j = \frac{h_j}{b_j}$ angeben lässt. Für die Klassen der ersten Spalte gilt mit der Angabe, dass 10 cm einer Einheit entsprechen, $b_j = 1$ für $j = 1, ..., 5$.

Körpergröße in cm	rh_j	Körpergröße in cm	rh_j
(150;160]	10	(150;170]	$\frac{50}{2} = 25$
(160;170]	40		
(170;180]	90	(170;190]	$\frac{140}{2} = 70$
(180;190]	50		
(190:200]	10	(190;200]	10

Somit stellen beide oberen Grafiken ein mögliches Histogramm dar.

zu B: vgl. A.

zu C: Die Aussage C trifft zu, vgl. die untere Grafik.

zu D: Die Aussage D trifft zu, vgl. die untere Grafik.

Aufgabe 2 A,C und E sind richtig.

zu A:

$$\begin{aligned} \overline{x} &= \frac{1}{n}(x_1 + ...x_n) \\ &= \frac{1}{12}(1+4+5+2+6+5+3+1+2+3+3+1) \\ &= \frac{1}{12} \cdot 36 = 3 \end{aligned}$$

zu B: Die geordnete Reihe der Beobachtungswerte lautet:
1, 1, 1, 2, 2, 3, 3, 3, 4, 5, 5, 6

$$x_{med} = \frac{1}{2}(x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)}) = \frac{1}{2}(x_{(6)} + x_{(7)}) = \frac{1}{2}(3+3) = 3$$

zu C: Die Werte 1 und 3 treten beide am häufigsten auf, und zwar jeder dreimal.

$$h(1) = h(3) = \max_j h(x_j) = 3$$

zu D:

$$\begin{aligned} d &= \frac{1}{n} \cdot \sum_{i=1}^{n} |x_i - x_{med}| \\ &= \frac{1}{12}(2+1+2+1+3+2+0+2+1+0+0+2) = \frac{4}{3} \end{aligned}$$

Aufgabe 3 A,B und C sind richtig.

zu A:

x_j	h_j	f_j	$x_j h_j$	$g_j = \frac{x_j h_j}{\sum_{k=1}^{m} x_k h_k}$	F_j	G_j
1	120	0.6	120	0.267	0.6	0.267
3	50	0.25	150	0.333	0.85	0.6
6	30	0.15	180	0.4	1.0	1.0
$\sum$	200	1.0	450	1.0		

$$\begin{aligned} \text{LKM} &= \left[\sum_{j=1}^{m} (F_{j-1} + F_j) \cdot g_j \right] - 1 \\ &= \frac{0.6 \cdot 120 + 1.45 \cdot 150 + 1.85 \cdot 180}{450} - 1 \\ &= \frac{72 + 217.5 + 333}{450} - 1 = \frac{622.5}{450} - 1 = 0.383 \end{aligned}$$

zu B: 60% der kleinsten Betriebe werden von den kleinen Betrieben erreicht, welche einen Gesamtumsatz g_1 von 0.267, d.h. 26.7%, besitzen.

zu C: Die mittleren Betriebe besitzen einen Gesamtumsatz g_2 von 0.333, d.h. 33.3%.

zu D: 40% der größten Betriebe werden von den mittleren und den großen Betrieben erreicht, welche einen Gesamtumsatz von $g_2 + g_3 = 0.333 + 0.4 = 0.733$, d.h. 73.3% besitzen.

Aufgabe 4 B,C und D sind richtig.

zu A:

	$y_1 = 1$	$y_2 = 6$	$y_3 = 9$	$\sum$
$x = x_2$	3	4	1	8

$$\overline{y} = \frac{1}{8} \sum_{j=1}^{3} y_j h(y_j) = \frac{1}{8}(1 \cdot 3 + 6 \cdot 4 + 9 \cdot 1) = 4.5$$

zu B:

$$f(y_1|x_1) = \frac{f(y_1, x_1)}{f(x_1)} = \frac{4}{4} = 1$$

zu C:

$$\begin{aligned} \tilde{s}^2 &= \frac{1}{n}\sum_{j=1}^{3}(y_j - \overline{y})^2 h(y_j) \\ &= \frac{1}{8}\big((1-4.5)^2 \cdot 3 + (6-4.5)^2 \cdot 4 + (9-4.5)^2 \cdot 1\big) \\ &= 8.25 \end{aligned}$$

zu D:

$$f(y_2|x_2) = \frac{f(y_2, x_2)}{f(x_2)} = \frac{4}{8} = 0.5$$

zu E:

$$f(x_3) = \frac{h(x_3)}{n} = \frac{2}{18} = \frac{1}{9}$$

Aufgabe 5 A,C und D sind richtig.

A Aus $f(x_2, y_1) = 0.5$ folgt $f(x_1, y_1) = 0.5$ und somit $h(x_1, y_1) = h(x_2, y_1) = 15$.

B Aus $f(x_1, y_2) = 0.5$ folgt $f(x_2, y_2) = 0.5$ und somit $h(x_1, y_2) = h(x_2, y_2)$. Aus $h(x_1, y_1) = 15$ und $h(x_2, y_1) = 15$ und $n = 50$ folgt $h(x_1, y_2) + h(x_2, y_2) = 20$ und somit $h(x_1, y_2) = 10$.

C $h(x_2, y_2) = h(x_1, y_2) = 10$

$$f(x_2, y_2) = \frac{h(x_2, y_2)}{n} = \frac{10}{50} = 0.2$$

D Da $f(x_i|y_i) = 0.5$ für alle i und j gilt, sind X und Y empirisch unabhängig.

E Die Aussage E ist falsch, siehe A - D.

Aufgabe 6 A ist richtig.

zu A: Der Korrelationskoeffizient nach Bravais-Pearson ist ein Maß für die Linearität des Zusammenhangs zwischen zwei Merkmalen. Vollständige Linearität liegt bei $r = 1$ und $r = -1$ vor; $r = -0.9$ deutet auf annähernde Linearität hin.

zu B: Auch -0.9 ist vom Betrag kleiner als 1 ($|-0.9| = 0.9 < 1$).

zu C: Aus $r = -0.9$ folgt, dass die Regressionsgerade eine negative Steigung besitzt. Folglich kann C nicht stimmen.

zu D: Der Wert -0.9 liegt wesentlich näher bei -1 als bei 0, womit der Zusammenhang zwischen den Merkmalen annähernd linear ist.

Aufgabe 7 B und C sind richtig.

zu A: Aus dem zweiten Korb wird gezogen, wenn die Zahlen $1, 3, 5, 7$ ausgewählt werden. Daher beträgt die Wahrscheinlichkeit $\frac{4}{7}$.

zu B: Unter der Voraussetzung, dass aus dem ersten Korb gezogen wird, beträgt die Wahrscheinlichkeit, eine Birne zu ziehen, $\frac{6}{12} = 0.5$.

zu C: $\boldsymbol{P}(\text{Birne}) = \boldsymbol{P}(\text{Birne} \cap \text{Korb 1}) + \boldsymbol{P}(\text{Birne} \cap \text{Korb 2}) = \frac{1}{2} \cdot \frac{3}{7} + \frac{1}{4} \cdot \frac{4}{7} = \frac{3}{14} + \frac{2}{14} = \frac{5}{14}$

zu D: $\boldsymbol{P}(\text{Apfel}) = \boldsymbol{P}(\text{Apfel} \cap \text{Korb 1}) + \boldsymbol{P}(\text{Apfel} \cap \text{Korb 2}) = \frac{1}{2} \cdot \frac{3}{7} + \frac{3}{4} \cdot \frac{4}{7} = \frac{3}{14} + \frac{6}{14} = \frac{9}{14}$

Aufgabe 8 B und D sind richtig.

zu A: Die Funktion $f_X(x)$ entspricht einer Dichtefunktion, wenn $\int_{-\infty}^{\infty} f_X(\xi)\, d\xi = 1$ gilt.

$$\begin{aligned}
\int_{-\infty}^{\infty} f_X(\xi)\, d\xi &= \int_0^{4a^2} \frac{1}{2a}\, d\xi \\
&= \left[\frac{\xi}{2a}\right]_0^{4a^2} = \frac{4a^2}{2a} = 2a \overset{!}{=} 1 \text{ für } a = \frac{1}{2}
\end{aligned}$$

zu B:

$$\begin{aligned}
F_X(x) &= \int_{-\infty}^{x} f_X(\xi)\, d\xi = \int_0^{x} \frac{1}{2a}\, d\xi \\
&= \left[\frac{\xi}{2a}\right]_0^{x} = \frac{x}{2a} \text{ für } 0 \leq x \leq 4a^2 \\
F_X(0) &= 0 \\
F_X(4a^2) &= 2a = 1 \text{ für } a = \frac{1}{2} \text{ vgl. A.}
\end{aligned}$$

Die Verteilungsfunktio $F_X(x)$ kann daher wie folgt angegeben werden:

$$F_X(x) = \begin{cases} 0 & \text{für} \quad x < 0 \\ \frac{x}{2a} & \text{für} \quad 0 \leq x \leq 4a^2 \\ 1 & \text{für} \quad x > 4a^2 \end{cases}$$

zu C: vgl. B.

zu D: $P(X \leq 0.5) = F_X(0.5) = \frac{0.5}{2a} = 0.5$ für $a = \frac{1}{2}$, vgl. A.

Aufgabe 41 $r = [1.895; 1.905]$

Da es sich um Änderungsraten handelt, wird hier das geometrische Mittel verwendet.

$$\begin{aligned}
1 + r &= \sqrt[4]{\Pi_{i=1}^{4}(1 + r_i)} = \sqrt[4]{1.1 \cdot 1.4 \cdot 0.7 \cdot 1} \\
&= \sqrt[4]{1.078} = 1.019 \\
r &= 1.9\%
\end{aligned}$$

Aufgabe 42 $\boldsymbol{P} = 0.444$

$$\begin{aligned}
\boldsymbol{P}(R \cup 4) &= \boldsymbol{P}(R) + \boldsymbol{P}(4) - \boldsymbol{P}(R \cap 4) \\
\boldsymbol{P}(R \cap 4) &= \frac{\text{Zahl der günstigen Fälle}}{\text{Zahl der möglichen Fälle}} = \frac{1}{18} \\
\boldsymbol{P}(R \cup 4) &= = \frac{1}{3} + \frac{1}{6} - \frac{1}{18} = \frac{8}{18} = 0.444
\end{aligned}$$