

Aufgabe 1

Teil A

Nora soll für den Online-Shop „TU Campus Merch“ eine deskriptive Datenanalyse über den Umsatz des Oktobers diesen Jahres erstellen. Dafür hat sie alle Bestellungen nach ihrer Bestellhöhe (in EUR) gruppiert und ausgewertet. Sie geht davon aus, dass sich die Bestellwerte gleichmäßig innerhalb einer Gruppe verteilt sind. Zur besseren Übersicht hat Nora folgende Tabelle mit bekannter Notation aus ihrer Statistik 1 Vorlesung erstellt:

j	von g_{j-1} bis g_j	f_j	$F(x)$	h_j	b_j	\tilde{f}_j	m_j	$m_j \cdot h_j$
1	[0; 20)	0,15	0,15	15	20	0,0075	10	150
2	[20; 50)	0,30	0,45	30	30	0,0100	35	1050
3	[50; 150)	0,45	0,90	45	100	0,0045	100	4500
4	[150; 300]	0,10	1,00	10	150	0,0007	225	2250
Σ	-	1,00	-	$n = 100$	-	-	-	7950

- a) Berechnen Sie die empirischen Quartile, welche notwendig für die Erstellung eines Boxplots sind. Zeichnen Sie den Boxplot sowie den modifizierten Boxplot. Runden Sie auf ganze Zahlen.

Für das p -te (emp.) Quantil x_p gilt: $x_p := g_{j-1} + [p - F(g_{j-1})] \frac{1}{\tilde{f}_j}$ mit $\frac{1}{\tilde{f}_j} = \frac{b_j}{f_j}$, also

$$x_{0,25} = 20 + [0,25 - 0,15] \cdot \frac{1}{0,01} = 30,$$

$$x_{0,5} = 50 + [0,5 - 0,45] \cdot \frac{1}{0,0045} \approx 61,$$

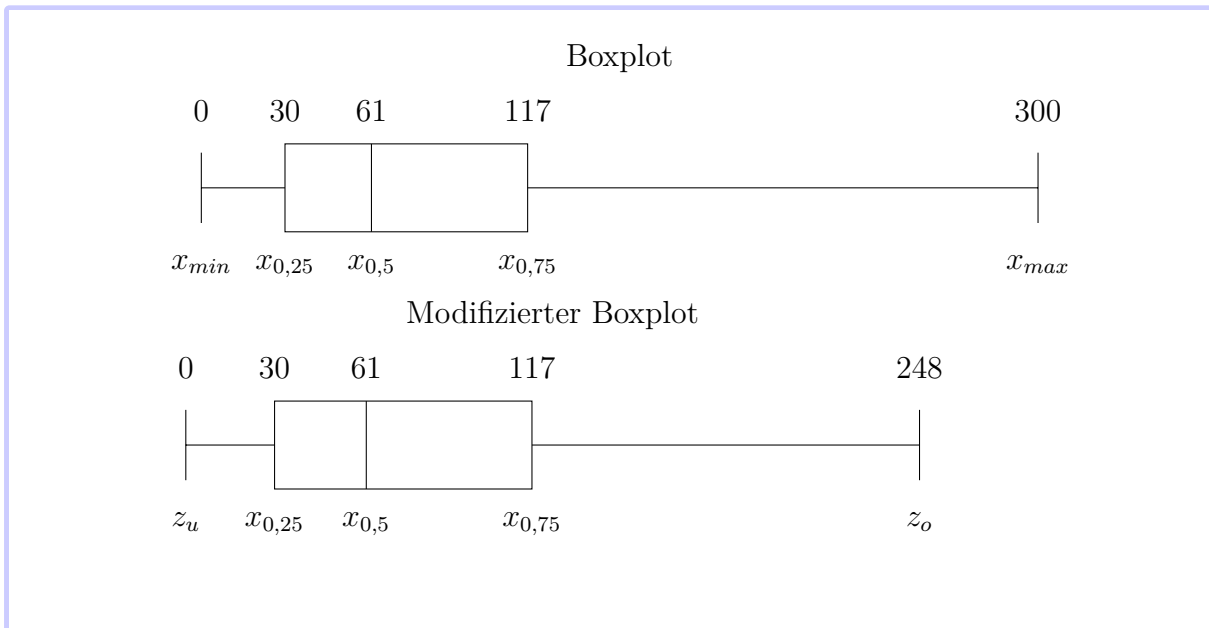
$$x_{0,75} = 50 + [0,75 - 0,45] \cdot \frac{1}{0,0045} \approx 117.$$

Weiter gilt:

$$d_q = x_{0,75} - x_{0,25} = 117 - 30 = 87,$$

$$z_u = \max(x_{\min}; x_{0,25} - 1,5 \cdot d_q) = \max(0; -100,5) = 0,$$

$$z_o = \min(x_{\max}; x_{0,75} + 1,5 \cdot d_q) = \min(300; 247,5) \approx 248.$$



- b) Berechnen Sie die durchschnittliche Bestellhöhe (arithmetisches Mittel). Von welcher Annahme gehen Sie bei dieser Berechnung aus?

Annahme: *Symmetrische Verteilung* der Beobachtungen innerhalb einer Gruppe j um die Gruppenmitte m_j (für alle Gruppen j) $\Rightarrow \bar{x}_j = m_j$

$$\bar{x}_{TU} = \frac{1}{n} \sum_{j=1}^4 m_j \cdot h_j = \frac{1}{100} \cdot 7950 = 79,5$$

- c) Welche Aussage kann Nora bezüglich der Schiefe der empirischen (relativen) Verteilung mithilfe
- i) der Lageregel,
 - ii) des empirischen Quartilkoeffizienten der Schiefe
- treffen. Warum sind unterschiedliche Ergebnisse in i) und ii) möglich?

i) $\tilde{x} = 61 < 79,5 = \bar{x}_{TU} \Rightarrow$ Die empirische Verteilung ist rechtsschief/linkssteil.

ii) $g_{0,25} := \frac{(x_{0,75} - x_{0,5}) - (x_{0,5} - x_{0,25})}{x_{0,75} - x_{0,25}} = \frac{(117 - 61) - (61 - 30)}{117 - 30} = \frac{25}{87} \approx 0,2874 > 0$

\Rightarrow Die empirische Verteilung ist rechtsschief/linkssteil.

Der Quartilkoeffizient der Schiefe berücksichtigt lediglich den *mittleren Datenkörper*, d.h. alle $x \in [x_{0,25}; x_{0,75}]$, während die Lageregel auch Beobachtungen außerhalb berücksichtigt. Dadurch kann es zu unterschiedlichen Ergebnissen kommen.

- d) Wie weit kann das arithmetische Mittel der Beobachtungen der letzten Gruppe maximal von jenem Wert abweichen, den Sie als Mittelwert dieser Gruppe angenommen haben?

Das arithmetische Mittel kann maximal um ± 75 EUR vom angenommenen Wert abweichen.

Teil B

Nora beschafft sich zum Vergleich Informationen über den Umsatz des Online-Shops „HU Campus Merch“. Die verfügbaren Kennzahlen stammen jedoch aus dem Oktober des Vorjahres; die Betreiber teilen Nora mit, dass dieses Jahr im Oktober die gleichen Bestellungen getätigt wurden, jedoch mit einem 20% Anstieg des Bestellwertes über alle Bestellungen hinweg.

$$95 \text{ Bestellungen} \mid \bar{x}_{HU, \text{Vorjahr}} = 80 \text{ (EUR)} \mid s_{HU, \text{Vorjahr}}^2 = 1600 \text{ (EUR)}^2$$

- e) Nora möchte nun anhand der Vorjahreswerte und des Anstiegs der Bestellwerte um 20% das arithmetische Mittel und die Standardabweichung für den Online-Shop „HU Campus Merch“ für den Oktober dieses Jahres berechnen.

$$\begin{aligned} \bar{x}_{HU} &= 1,2 \cdot \bar{x}_{HU, \text{Vorjahr}} = 1,2 \cdot 80 = 96 \\ s_{HU}^2 &= 1,2^2 \cdot s_{HU, \text{Vorjahr}}^2 = 2304 \Rightarrow s_{HU} = \sqrt{2304} = 48 \end{aligned}$$

Teil C

Nora bittet ebenso um Stellungnahme des Online-Shops „FU Campus Merch“. Dieser konnte im ganzen Oktober jedoch lediglich fünf Bestellungen verzeichnen. Die Bestellwerte werden wie folgt übermittelt.

Bestellung	1	2	3	4	5
Bestellhöhe (in EUR)	20	120	120	170	2000

- f) Welche Mittelwerte können Sie in diesem Fall sinnvoll angeben? Geben Sie diese an.

$$\bar{x}_{FU} = 486 \quad \tilde{x}_{FU} = 120 \quad \hat{x}_{FU} = 120$$

Eine E-Mail korrigiert die Höhe der fünften Bestellung auf 20 EUR.

g) Welchen Mittelwert können Sie nun nicht mehr eindeutig angeben?

Der Modus \hat{x}_{FU} kann nicht mehr eindeutig angegeben werden.

h) Berechnen Sie aus den korrigierten Bestellhöhen das arithmetische Mittel und die Standardabweichung.

$$\begin{aligned}\bar{x}_{FU} &= 90 \\ s_{FU}^2 &= \frac{1}{5} \cdot [(20 - 90)^2 \cdot 2 + (120 - 90)^2 \cdot 2 + (170 - 90)^2] = 3600 \\ s_{FU} &= \sqrt{3600} = 60\end{aligned}$$

i) Vergleichen Sie die Streuung der Bestellhöhen von „HU Campus Merch“ und „FU Campus Merch“ mithilfe des Variationskoeffizienten.

$$\begin{aligned}V_{HU} &= \frac{s_{HU}}{\bar{x}_{HU}} = \frac{48}{96} = \frac{1}{2} \\ V_{FU} &= \frac{s_{FU}}{\bar{x}_{FU}} = \frac{60}{90} = \frac{2}{3}\end{aligned}$$

$V_{HU} < V_{FU} \Rightarrow$ Streuung der Bestellhöhen bei „FU Campus Merch“ größer als bei „HU Campus Merch“ nach dem Variationskoeffizienten.

j) Berechnen Sie die durchschnittliche Bestellhöhe (arithmetisches Mittel) in EUR für alle drei Shops zusammen.

$$\bar{x}_{Ges} = \bar{x}_{TU} \cdot \frac{100}{200} + \bar{x}_{HU} \cdot \frac{95}{200} + \bar{x}_{FU} \cdot \frac{5}{200} = 79,5 \cdot \frac{100}{200} + 96 \cdot \frac{95}{200} + 90 \cdot \frac{5}{200} = 87,6$$

k) Berechnen Sie die empirische Varianz der Bestellhöhe in EUR² für alle drei Shops zusammen. Nutzen Sie: $s_{TU}^2 = 4000$

Es gilt: $s_{Ges}^2 = s_{Ges,int}^2 + s_{Ges,ext}^2 = \frac{1}{n_{Ges}} \sum_{j=1}^r n_j s_j^2 + \frac{1}{n_{Ges}} \sum_{j=1}^r n_j (\bar{x}_j - \bar{x}_{Ges})^2$, also

$$\begin{aligned}s_{Ges}^2 &= \frac{1}{200} (100 \cdot 4000 + 95 \cdot 2304 + 5 \cdot 3600) + \frac{1}{200} (100 \cdot (79,5 - 87,6)^2 \\ &\quad + 95 \cdot (96 - 87,6)^2 + 5 \cdot (90 - 87,6)^2) \\ &= \underbrace{3184,4}_{\text{Streuung innerhalb der Shops}} + \underbrace{66,465}_{\text{Streuung zwischen den Shops}} = 3250,865\end{aligned}$$

- 1) Welcher Anteil der Gesamtvarianz wird durch die Unterschiede zwischen den drei Shops verursacht? Runden Sie auf vier Nachkommastellen.

$$\frac{66,465}{3250,865} \approx 0,0204$$