Technische Universität Dortmund
Fakultät Statistik
Jun.-Prof. Dr. Andreas Groll
M.Sc. Hendrik van der Wurp

Wintersemester 2018/2019
7. November 2018

Übungen zur Vorlesung Wahrscheinlichkeitsrechnung und Mathematische Statistik (für Informatiker)

Blatt 2 - Lösungsvorschlag

Aufgabe 5: (per Hand)

Betrachten Sie die Ergebnisse der Modulprüfung *Statistik VIII* mit den Noten

$$D = \{3.7, 3.3, 1.7, 2, 3.3, 2.7\}.$$

Berechnen Sie

- den Stichprobenumfang,
- den Modus,
- den Median,
- das arithmetische Mittel,
- das untere und obere Quartil sowie
- das 15% und 85%-Quantil.

Lösungsvorschlag

Sortierte Stichprobe: $\{1.7, 2, 2.7, 3.3, 3.3, 3.7\}$

- Stichprobenumfang: $n = 6$
- Modus: 3.3 mit Häufigkeit 2
- Median mit n gerade: $\frac{x_{\left(\frac{6}{2}\right)}+x_{\left(\frac{6}{2}+1\right)}}{2} = \frac{x_{(3)}+x_{(4)}}{2} = \frac{2.7+3.3}{2} = 3$
- Arithmetisches Mittel: $\frac{1}{6}(3.7 + 3.3 + 1.7 + 2 + 3.3 + 2.7) = \frac{16.7}{6} \approx 2.78$
- unteres Quartil = 0.25-Quantil. $np = 6 \cdot 0.25 = 1.5$ nicht ganzzahlig.

$$Q_{0.25} = x_{(j)}, \qquad \text{mit } j = \lceil np \rceil = 2 \qquad \text{führt zu } Q_{0.25} = x_{(2)} = 2$$

- oberes Quartil = 0.75-Quantil. $np = 6 \cdot 0.75 = 4.5$ nicht ganzzahlig.

$$Q_{0.75} = x_{(j)}, \qquad \text{mit } j = \lceil np \rceil = 5 \qquad \text{führt zu } Q_{0.75} = x_{(5)} = 3.3$$

- Die 0.15- und 0.85-Quantile mit $np = 0.9$ und $np = 5.1$ respektive, beides nicht ganzzahlig.

$$Q_{0.15} = x_{(j)}, \qquad \text{mit } j = \lceil np \rceil = 1 \qquad \text{führt zu } Q_{0.15} = x_{(1)} = 1.7$$

$$Q_{0.85} = x_{(j)}, \qquad \text{mit } j = \lceil np \rceil = 6 \qquad \text{führt zu } Q_{0.85} = x_{(6)} = 3.7$$

Aufgabe 6: (per Hand)

An einem Kneipenabend der Orientierungswoche müssen sich die 120 Teilnehmer jeweils für eine Biersorte entscheiden. Die neu-Studierenden haben sich in folgender Aufteilung für die drei Biere entschieden:

- *Dortmunder Oniun*: 40 Studierende
- *Dortmunder Kornen*: 60 Studierende
- *Dortmunder Hasna*: 20 Studierende

Die Biere seien zudem in obiger Reihenfolge absteigend nach dem Preis sortiert.

(a) Berechnen Sie für die Variable *Biersorte* die Kennzahl *Simpson's D* sowohl in Standardform als auch normiert.

(b) Fassen Sie die Variable Preis als ordinale Variable auf und berechnen Sie für diese Variable die Kennzahl *Leti's D* (Standardform und normiert). Vergleichen Sie die Werte mit den Werten aus Teilaufgabe (a) und interpretieren Sie die Ergebnisse.

Lösungsvorschlag

(a) Seien für die Biersorten

$$\begin{aligned} f_{\text{Oniun}} &= f_1 = \frac{40}{120} = \frac{1}{3} \\ f_{\text{Kornen}} &= f_2 = \frac{60}{120} = \frac{1}{2} \\ f_{\text{Hasna}} &= f_3 = \frac{20}{120} = \frac{1}{6} \end{aligned}$$

Und damit

$$\begin{aligned} D &= 1 - \sum_{j=1}^{3} f_j^2 = 1 - \left(\left(\frac{1}{3}\right)^2 + \left(\frac{1}{2}\right)^2 + \left(\frac{1}{6}\right)^2 \right) = 1 - \left(\frac{1}{9} + \frac{1}{4} + \frac{1}{36} \right) \approx 0.61 \\ D_z &= \frac{3\left(1 - \sum_{j=1}^{3} f_j^2\right)}{3-1} = \frac{3}{2} D \approx 0.917 \end{aligned}$$

(b) Preise: $P_{\text{Hasna}} < P_{\text{Kornen}} < P_{\text{Oniun}}$ ordinal. F_N sei die empirische Verteilungsfunktion mit

$$\begin{aligned} F_N[x(1)] &= F_N(x_{\text{Hasna}}) = \frac{1}{6} \\ F_N[x(2)] &= F_N(x_{\text{Kornen}}) = \frac{1}{6} + \frac{1}{2}. \end{aligned}$$

Und damit

$$\begin{aligned} D_L &= \sum_{j=1}^{2} F_N[x(j)] \cdot (1 - F_N[x(j)]) \\ &= \underbrace{\frac{1}{6} \cdot \left(1 - \frac{1}{6}\right)}_{\text{für } j=1} + \underbrace{\left(\frac{1}{6} + \frac{1}{2}\right) \cdot \left(1 - \left(\frac{1}{6} + \frac{1}{2}\right)\right)}_{\text{für } j=2} \approx 0.36 \end{aligned}$$

$$D_{LZ} = \frac{4}{3-1} \sum_{j=1}^{2} F_N[x(j)] \cdot (1 - F_N[x(j)]) = \frac{4}{2} D_L = 2 D_L \approx 0.72$$

Beide Kennzahlen nehmen Streuung wahr. *Simpsons D* deutet auf eine sehr starke Streuung hin, welche bei $f_1 = f_2 = f_3 = \frac{1}{3}$ maximal wäre. *Letis D* nimmt weniger Streuung wahr. Dies liegt daran, dass die Reihenfolge bei *Leti* berücksichtigt wird. Eine maximale Streuung im Sinne von *Letis D* läge bei $f_1 = f_3 \approx 0.5$ und $f_2 \approx 0$ vor. In diesem Fall:

$$
\begin{aligned}
D &= 1 - \left(\frac{1}{2^2} + 0^2 + \frac{1}{2^2}\right) && = 0.5 \\
D_z &= \frac{3}{2} \cdot D = \frac{3}{2} \cdot 0.5 && = 0.75 \\
D_L &= \frac{1}{2} \cdot \frac{1}{2} + \frac{1}{2} \cdot \frac{1}{2} && = 0.5 \\
D_{LZ} &= \frac{4}{2} D_L = 2 D_L && = 1
\end{aligned}
$$

Simpsons D und *Letis D* messen unterschiedliche Ansätze von Streuung. Letzteres hat aufgrund der Ordinalität einen höheren Informationsgehalt.