Musterlösung Aufgabe 7: Univariate statistische Kennzahlen für die Streuung

In einer Mensa gibt es drei Gerichte zur Auswahl, G1 (Lachsfilet), G2 (Schnitzel) und G3 (Linseneintopf). Die Gerichte seien absteigend nach dem Preis geordnet (G1 ist teurer als G2, G2 ist teurer als G3). An einem bestimmten Tag wählen genau 50 Studierende Gericht G1, 150 wählen G2 und 100 wählen G3.

(a) Berechnen Sie für die Variable Preis *Simpson's D* per Hand.

(b) Fassen Sie die Variable Preis als ordinale Variable auf und berechnen Sie für diese Variable die Kennzahl *Leti's D* per Hand. Vergleichen Sie den Wert mit dem Wert aus Teilaufgabe (a) und interpretieren Sie die beiden Werte.

Lösung:

(a)

$$J = 3;$$

$$f_1 = \frac{50}{300} = \frac{1}{6}, \quad f_2 = \frac{150}{300} = \frac{1}{2}, \quad f_3 = \frac{100}{300} = \frac{1}{3};$$

$$\begin{aligned} D &= \frac{J}{J-1}\left(1 - (f_1^2 + f_2^2 + f_3^2)\right) = \frac{3}{2}\left(1 - \left((1/6)^2 + (1/2)^2 + (1/3)^2\right)\right) \\ &= \frac{3}{2}\left(1 - \frac{1+9+4}{36}\right) = \frac{3}{2}\left(1 - \frac{14}{36}\right) = \frac{3}{2}\frac{22}{36} = \frac{11}{12} \end{aligned}$$

(b)

$$F_1 = f_1 = \frac{1}{6}, \quad F_2 = f_1 + f_2 = \frac{1}{6} + \frac{1}{2} = \frac{2}{3}$$

$$D = \frac{4}{J-1}\sum_{i=1}^{2} F_N(x(j)) \cdot (1 - F_N(x(j))) = \frac{4}{2}\left(\frac{1}{6}\frac{5}{6} + \frac{2}{3}\frac{1}{3}\right) = 2\frac{8}{36} = \frac{13}{18}$$

Simpsons's D hat einen großen Wert, da alle Klassen relativ gut vertreten sind. Es werden nur die relativen Häufigkeiten betrachtet. Bei Leti's D ist der Wert etwas kleiner, da die Reihenfolge jetzt eine Rolle spielt und die mittlere Klasse am stärksten besetzt ist.

Musterlösung Aufgabe 8: Univariate statistische Kennzahlen

Geben Sie alle Werte für x_3 an, für die der Variationskoeffizient den Wert 1 hat.

$$x_1 = 0, \quad x_2 = 2, \quad x_3 = ?$$

Lösung:

Es muss gelten:

$$\frac{\hat{s}}{\bar{x}} = 1 \quad \Longleftrightarrow \quad \bar{x}^2 = \hat{s}^2 \qquad \text{(da in diesem Fall } \hat{s} > 0). \tag{1}$$

Betrachte Datensatz $\{0,\ 2,\ x\}$:

$$\begin{aligned}
\left(\frac{x+2}{3}\right)^2 &= \frac{1}{2}\left(\left(\frac{0}{3} - \frac{x+2}{3}\right)^2 + \left(\frac{3\cdot 2}{3} - \frac{x+2}{3}\right)^2 + \left(\frac{3\cdot x}{3} - \frac{x+2}{3}\right)^2\right) \\
\Longleftrightarrow \quad 2\,(x+2)^2 &= (-x-2)^2 + (6-x-2)^2 + (3x-x-2)^2 \\
&= (x+2)^2 + (x-4)^2 + (2x-2)^2 \\
\Longleftrightarrow \quad 2x^2 + 8x + 8 &= x^2 + 4x + 4 + x^2 - 8x + 16 + 4x^2 - 8x + 4 \\
\Longleftrightarrow \quad 0 &= 4x^2 - 20x + 16 \\
\Longleftrightarrow \quad 0 &= x^2 - 5x + 4 \\
\Longleftrightarrow \quad 0 &= (x-1)\,(x-4)
\end{aligned}$$

Lösungen sind somit $x_3 = 1$ und $x_3 = 4$.

Musterlösung Aufgabe 9: Bivariate Daten: Zusammenhangsmaße

Für die gewählten Gerichte aus Aufgabe 7 stehe zusätzlich eine Variable Geschlecht zur Verfügung, die angibt, ob das Gericht von einer männlichen oder eine weiblichen Person gewählt wurde. Dies führt auf die folgende Kontingenztafel K der beiden Variablen *Gericht* und *Geschlecht*:

		Gericht		
		G1 (Lachsfilet)	G2 (Schnitzel)	G3 (Linseneintopf)
Geschlecht	männlich	20	90	40
	weiblich	30	60	60

(a) Berechnen Sie die (univariaten) relativen Randhäufigkeitsverteilungen und die bedingten Verteilungen, jeweils für die beiden Variablen Gericht und Geschlecht.

(b) Berechnen Sie für die Kontingenztafel K den Wert der χ^2-Größe.

(c) Welche Einträge in der Kontingenztafel sind größer als der jeweils erwartete Wert unter Unabhängigkeit der beiden Variablen? Wie ist dies zu interpretieren?

Lösung:

(a) Absolute Häufigkeiten und absolue Randhäufigkeiten:

		Gericht			
		G1	G2	G3	Summe
Geschlecht	männlich	20	90	40	150
	weiblich	30	60	60	150
	Summe	50	150	100	300

Relative Häufigkeiten und relative Randhäufigkeiten:

		Gericht			
		G1	G2	G3	Summe
Geschlecht	männlich	2/30	9/30	4/30	1/2
	weiblich	3/30	6/30	6/30	1/2
	Summe	1/6	3/6	2/6	1

Bedingte Verteilung für Gericht:

		Gericht			
		G1	G2	G3	Summe
Geschlecht	männlich	2/15	9/15	4/15	1
	weiblich	1/5	2/5	2/5	1

Bedingte Verteilung für Geschlecht:

		Gericht		
		G1	G2	G3
Geschlecht	männlich	2/5	3/5	2/5
	weiblich	3/5	2/5	3/5
	Summe	1	1	1

(b) Berechnung der erwarteten Werte unter Unabhängigkeit:

		Gericht		
		G1	G2	G3
Geschlecht	männlich	150 · 50 /300 = 25	150 · 150 /300 = 75	150 · 100 /300 = 50
	weiblich	150 · 50 /300 = 25	150 · 150 /300 = 75	150 · 100 /300 = 50

Berechnung der χ^2-Größe:

$$\begin{aligned}\chi^2 &= (20-25)^2/25 + (90-75)^2/75 + (40-50)^2/50 \\ &\quad +(30-25)^2/25 + (60-75)^2/75 + (60-50)^2/50 \\ &= 25/25 + 225/75 + 100/50 + 25/25 + 225/75 + 100/50 \\ &= 1+3+2+1+3+2 \;=\; 12\end{aligned}$$

$$\begin{aligned}C &= \sqrt{\frac{\chi^2}{\chi^2+N}\,\frac{\min(J,K)}{\min(J,K)-1}} \qquad \text{mit} \quad \chi^2 = 12,\ N = 300 \\ &= \sqrt{\frac{12}{12+300}\,\frac{2}{1}} = \sqrt{\frac{24}{312}} = \sqrt{\frac{1}{13}} \approx 0.277\end{aligned}$$

Man erhält mit 0.277 einen relativ kleinen Wert, d.h. es ist gibt einen schwachen Zusammenhang.

(c) Für die weiblichen Personen sind die Einträge für G1 (Lachsfilet) und G3 (Linseneintopf) etwas größer, für die männlichen Personen der Eintrag für G2 (Schnitzel). Männer entscheiden sich also im direkten Vergleich in dieser Population etwas öfter für Fleisch und Frauen öfter für die anderen Gerichte.