Musterlösung zur Klausur Statistik

TIT/TIM/TIS 18

Oettinger 06.2020

Zeit: 90Min.

Insgesamt erreichbare Punktzahl: 105, 100%: 100 Punkte.

Aufgabe 1

(a) Der Median ist der Wert in der Mitte der Stichprobe, er entspricht dem 50%-Quantil - falsch.

(b) Bei einer engipfligen, symmetrischen Vertilung liegt der Modus genau in der Mitte der Stichprobe, genau wie der Median - richtig (falls die Stichprobe groß genug ist).

(c) Nominale Merkmalswerte können nicht geordnet werden - die Bestimmung des Medians ist nicht möglich - falsch.

(d) Die Varianz kann nur positive Werte annehmen - richtig, sie ist eine Summe quadrierter Größen.

(e) Das arithmetische Mittel kann auch negative Werte annehmen, beispielsweise bei negativen Merkmalswerten - falsch.

Aufgabe 2

Vollständige gemeinsame Häufigkeitstabelle der beiden Merkmale X und Y und relative Werte:

Y / X	1	2	3	Summe
1	8	2	6	16
2	16	4	12	32
3	8	2	6	16
Summe	32	8	24	64

Y / X	1	2	3	$f_X(X)$
1	1/4	1/4	1/4	1/4
2	1/2	1/2	1/2	1/2
3	1/4	1/4	1/4	1/4
$f_Y(Y)$	1	1	1	1

- Die Spalten sowie die Randverteilung der rechten Tabelle sind identisch $\Longrightarrow$ die Merkmale X und Y sind stochastisch unabhängig.
- Bei $f(x_i|Y = 2)$ handelt es sich um relative Häufigkeiten bzw. Wahrscheinlichkeiten (Die Summe muss eins ergeben!):
  $f(x_i|Y = 2) = (1/4, 1/2, 1/4)$
- Varianz $s^2(X|Y = 2) = \sum_{i=1}^{n}(x_i - \bar{x})^2 \cdot f_i$. Benötigt wird das arithmetische Mittel

$$\bar{x} = \sum_{i=1}^{n} f_i \cdot x_i = (1/4 \cdot 1 + 1/2 \cdot 2 + 1/4 \cdot 3) = 2$$

oder

$$\bar{x} = \frac{1}{1/2}(1/8 \cdot 1 + 1/4 \cdot 2 + 1/8 \cdot 3) = 2 \cdot (\frac{1+4+3}{8}) = 2$$

$$s^2(X|Y = 2) = \sum_{i=1}^{3}(x_i - \bar{x})^2 \cdot f_i = (1-2)^2 \cdot \frac{1}{4} + (2-2)^2 \cdot \frac{1}{2} + (3-2)^2 \cdot \frac{1}{4}$$
$$= \frac{1}{4} + 0 + \frac{1}{4} = \frac{1}{2}$$

Aufgabe 3

Die Summe der Vorfälle in China ist 63, $n = 7$ und damit

$$\bar{x} = 63/7 = 9.$$

Zur Bestimmung des Medians werden die Daten in Form eines geordneten Vektors dargestellt:

$$\{x_i\} = (5, 7, 7, 8, 11, 12, 13)$$

Der Median $\bar{x}_Z$ ist der Wert x_4, also $\bar{x}_Z = 8$.

Die Varianz berechnet sich wie folgt:

$$s_x^2 = \frac{1}{7}(5^2 + 7^2 + 7^2 + 8^2 + 11^2 + 12^2 + 13^2) - 9^2$$
$$= 7,714,$$

damit ergibt sich die Standardabweichung

$$s_x = |\sqrt{s_x^2}| = 2,777$$

und der Variationskoeffizient

$$v_x = \frac{s_x}{\bar{x}} = \frac{2,878}{8} = 0,309.$$

Die Summe der Vorfälle weltweit ist 266, die Anzahl ist ebenfalls 7, also

$$\bar{x} = 266/7 = 38.$$

Der geordnete Vektor der Daten ist

$$\{x_i\} = (27, 31, 34, 35, 41, 46, 52),$$

der Median ist der vierte Wert $\bar{x}_Z = 35$. Die Varianz

$$s_x^2 = \frac{1}{7}(27^2 + 31^2 + 34^2 + 35^2 + 41^2 + 46^2 + 52^2) - 38^2$$
$$= 66,29$$

die Standardabweichung

$$s_x = |\sqrt{s_x^2}| = 8.14$$

und der Variationskoeffizient

$$v_x = \frac{s_x}{\bar{x}} = \frac{8,14}{38} = 0,214.$$

Die Streuung der weltseit erhobenen Daten ist geringer, also finden die Reisunfälle weltweit regelmäßiger statt.

Aufgabe 4

(a) Jede Antwort ist zulässig

(b) Statistische Einheit sind Luigis Kunden, Merkmale sind Speisen und Getränke. Die Merkmalsausprägungen sind die vom Kunden bestellten Gerichte und Getränke, beispielsweise 'Rotwein' oder 'Pizza'.

(c) Darstellung der Daten im Stab- oder Balkendiagramm. Grafik nicht gefordert!

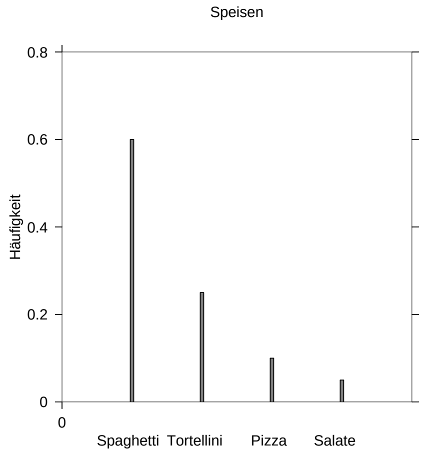

Abbildung 1: Stabdiagramm als Beispiel für die grafische Darstellung

(d) Ereignis A: Kunde bestellt Spaghetti, $P(A) = 0,6$.
Ereignis B: Kunde bestellt Rotwein, $P(B) = 0,7$.
Ereignis C: Kunde bestellt Spaghetti und Rotwein, $P(C) = P(A) \cdot P(B) = 0,6 \cdot 0,7 = 0,42 = 42\%$.
Ereignis D: Kunde bestellt Spaghetti, aber keinen Rotwein, $P(D) = P(A) \cdot (1 - P(B)) = 0,6 \cdot (1 - 0,7) = 0,6 \cdot 0,3 = 0,18 = 18\%$.

(e) Die relative Häufigkeit für eine Spaghettilänge größer als 80cm ist $0,15 + 0,05 = 0,2 = 20\%$.
Das arithmetische Mittel (in cm) muss mit den Klassenmitten gerechnet werden (wegen der Angabe relativer Häufigkeiten wird nicht durch die Gesamtzahl geteilt):

$$\bar{x} = 0,15 \cdot 30 + 25 \cdot 50 + 40 \cdot 70 + 15 \cdot 90 + 5 \cdot 115 = 64,25$$

(f) Der Median (in cm) entfällt auf die 3.Klasse $]60;80]$:

$$\bar{x}_Z = x_3^u + (x_3^o - x_3^u) \cdot \frac{F(\bar{x}_Z) - F(x_3^u)}{F(x_3^o) - F(x_3^u)}$$

$$= 60 + (80 - 60) \cdot \frac{0,5 - 0,4}{0,8 - 0,4} = 65.$$

(g) Da die Klassen unterschiedlich breit sind, wird die Dichte der relativen Häufigkeiten aufgetragen. Daten zur grafischen Darstellung:

Länge (cm)	Klasenbreite (in m) Δ_k	rel. Häufigkeit f_k	Dichte $f_k^* = \frac{f_k}{\Delta_k}$
$(20; 40]$	$0,20$	0,15	0,75
$(40; 60]$	$0,20$	0,25	1,25
$(60; 80]$	$0,20$	0,4	2,0
$(80; 100]$	$0,20$	0,15	0,75
$(100; 130]$	$0,30$	0,05	0,167

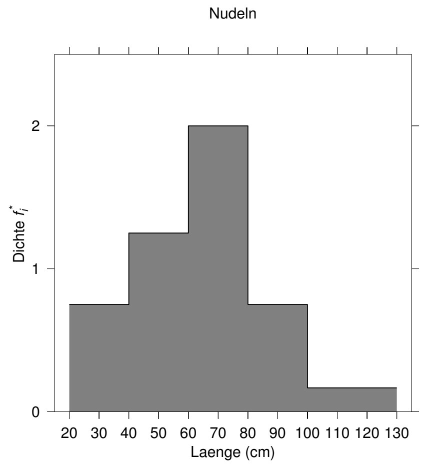

Abbildung 2: Histogramm der Verteilung der Nudellänge

Aufgabe 5

Geeignete Mittelwerte.

1. Wenn Ted eine mittlere Geschwindigkeit von 60 km/h fahren will, benötigt er für die insgesamt 8 km Weg eine Zeit von 8/60 h = 8 min. Da er aber für den Rückweg von 4km bereits eine Zeit von 4/30 h = 8min einplant, kann er die geplante Durchschnittsgeschwindigkeit nicht erreichen.

2. Geometrisches Mittel:
$\bar{x}_G = \sqrt[3]{(1 + 0,1) \cdot (1 + 0,15) \cdot (1 - 0,0005)} - 1 = 8,13\%$

3. Insgesamt befragte Personen: 100 + 1000 = 1100. Für die Abschaffung sind 60 + 380 = 440. Also sind 440/1100 = 40% dafür.