Musterlösung zur Nachklausur Statistik

TIT14

Oettinger 10.2016

Zeit: 90Min.

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

Aufgabe 1

(a) Im Allgemeinen sind Modus und Median verschieden - falsch.

(b) Bei einer symmetrischen Verteilung sind arithmetisches Mittel und Median oft gleich - falsch.

(c) Die Varianz ist eine Summe quadratischer Größen - richtig.

(d) Der Gini-Koeffizient bewegt sich als anständiger Koeffizient natürlich zwischen 0 und 1. Richtig.

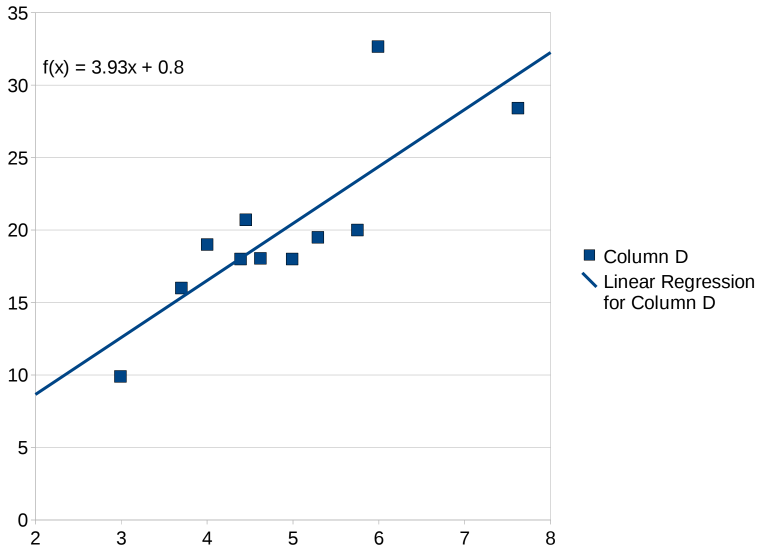
(e) Der Quartilsabstand ist ein Maß für die Streuung - richtig.

(f) Maschinenausfälle besitzen keine Lebensdauer - es handelt sich um eine Bewegungsmasse - falsch.

Aufgabe 2

a) Alle Werte sind zulässig.

b) Daten mit Ausgleichsgerade



c) Anpassung einer Geraden $y = a \cdot x + b$ über lineare Regression: Tabelle benötigter Daten

City	Hamburger	Kinokarten				
	x_i	y_i	$(x_i - \bar{x})$	$(y_i - \bar{y})$	$(x_i - \bar{x})(y_i - \bar{y})$	$(x_i - \bar{x})^2$
Toyo	5.99	32.66	1.1	12.64	13.9	1.21
London	7.62	28.41	2.73	8.39	22.9	7.45
New York	5.75	20	0.86	-0.02	-0.02	0.74
Sydney	4.45	20.71	-0.44	0.69	-0.3	0.19
Chicago	4.99	18	0.1	-2.02	-0.2	0.01
San Francisco	5.29	19.5	0.4	-0.52	-0.21	0.16
Boston	4.39	18	-0.5	-2.02	1.01	0.25
Atlanta	3.7	16	-1.19	-4.02	4.78	1.42
Toronto	4.62	18.05	-0.27	-1.97	0.53	0.07
Rio de Janeiro	2.99	9.9	-1.9	-10.12	19.23	3.61
Friedrichshafen	4	19	-0.89	-1.02	0.91	0.79

Die Steigung der Geraden ist

$$a = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2} = 3,93$$

der Achsenabschnitt ist

$$b = \bar{y} - a\bar{x} = 0,80$$

d) der Pearson-Koeffizient ist

$$r_{xy} = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2} \cdot \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2}} \tag{1}$$

$$= \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2} \cdot \sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}} = 0,82 \tag{2}$$

Der Wert nahe eins legt einen relativ guten linearen Zusammenhang zwischen den Daten nahe.

Aufgabe 3

Alle Angaben in min:

(a) Geordneter Vektor der Verspätungen: (2, 6, 8, 8, 10, 10, 12, 20). Der Umfang der Stichprobe ist $n = 8$, das untere Quartil (das 0,25-Quantil) ist

$$x_{0,25} = \frac{x_{n/4} + x_{n/4+1}}{2} = \frac{6+8}{2} = 7.$$

(b) Mittelwert $\bar{x} = \frac{1}{8}(2 + 6 + 2 \cdot 8 + 2 \cdot 10 + 12 + 20) = 9,5$.

(c) Die Spannweite ist $\max(\{x_i\}) - \min(\{x_i\}) = 20 - 2 = 18$.

(d) Die mittlere absolute Abweichung vom arithmetischen Mittel ist

$$d_{\bar{x}} = \frac{1}{8}\left(|2 - 9,5| + |6 - 9,5| + 2 \cdot |8 - 9,5| + 2 \cdot |10 - 9,5| + |12 - 9,5|\right.$$
$$\left. + |20 - 9,5|\right) = 3,5$$

Die mittlere quadratische Abweichung (die Varianz) hat den Wert

$$s^2 = \frac{1}{8}\left((2 - 9,5)^2 + (6 - 9,5)^2 + 2 \cdot (8 - 9,5)^2 + 2 \cdot (10 - 9,5)^2 + (12 - 9,5)^2\right.$$
$$\left. + (20 - 9,5)^2\right) = 23,75$$

Aufgabe 4

Tag	1	2	3	4	5	6	7
km (x)	15	16,5	17,5	18	18	20	22

a) Arithmetisches Mittel:

$$\bar{x} = \frac{15 + 16,5 + 17,5 + 18 + 18 + 20 + 22}{7} = \frac{127}{7} = 18,1429$$

b) Harmonisches Mittel:

$$\bar{x}_H = \frac{7}{\frac{1}{15} + \frac{1}{16,5} + \frac{1}{17,5} + \frac{1}{18} + \frac{1}{18} + \frac{1}{20} + \frac{1}{22}} = 17,9037$$

c) Die Durchschnittsgeschwindigkeit erhält man als Quotienten der gesamten zurückgelegten Strecke und der gesamten benötigten Zeit, also

$$\frac{15km + 16,5km + 17,5km + 18km + 18km + 20km + 22km}{7h}$$

Die Anwendung des arithmetischen Mittels ist hier korrekt.

Aufgabe 5

(a) Jede Antwort ist zulässig

(b) Statistische Einheit sind Giovannis 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!

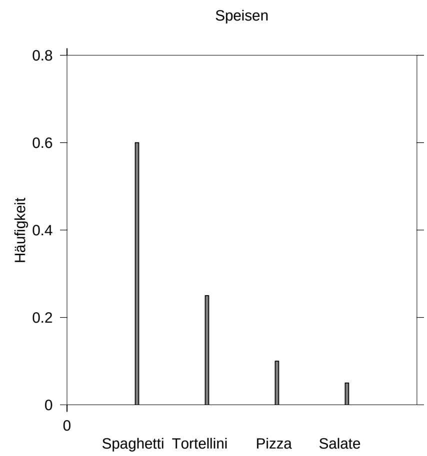


Abbildung 1: Stabdiagramm als Beispiel für die grafische Darstellung

(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

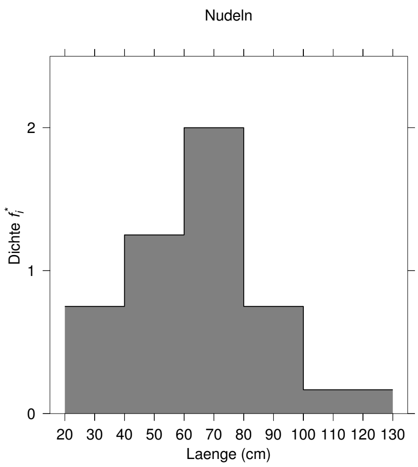


Abbildung 2: Histogramm der Verteilung der Nudellänge