

Musterlösung zur Klausur Statistik

TIT11

Oettinger 03.2013

Zeit: 90Min.

Insgesamt erreichbare Punktzahl: 100.

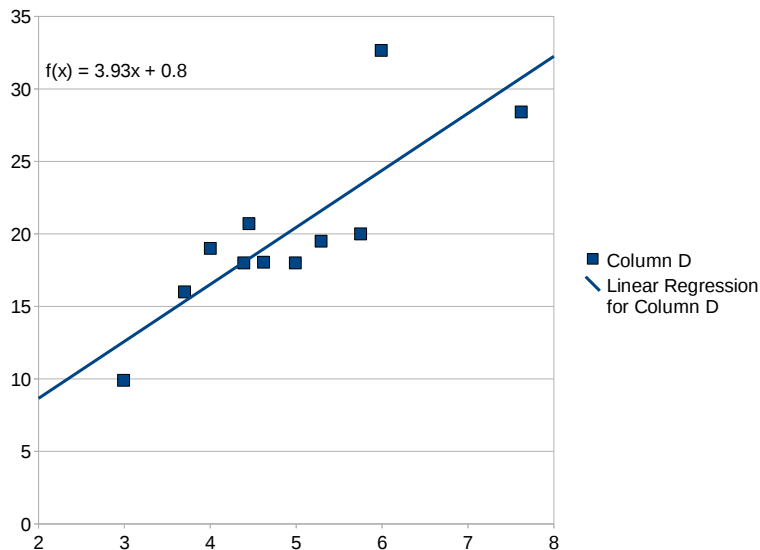
Aufgabe 1

- (a) Der Median entspricht dem 50%-Quantil oder 0,5-Quantil - richtig.
- (b) Für eine unimodale, nicht symmetrische Verteilung gilt stets, dass der Median und der Modus sich unterscheiden - falsch.
- (c) Die Varianz kann nur positive Werte annehmen - richtig.
- (d) Das arithmetische Mittel kann auch negative Werte annehmen - falsch.
- (e) Ein Merkmal ist entweder nominal oder stetig, d.h. es gibt kein Merkmal, das gleichzeitig nominal und stetig ist - richtig, alle stetigen Merkmale sind kardinal skaliert.
- (f) Der Pearson-Koeffizient bewegt sich als anständiger Koeffizient zwar *betragsmäßig* zwischen 0 und 1, er enthält aber auch Information über die Steigung der Geraden. Falsch, der Pearson-Koeffizient bewegt sich zwischen -1 und 1.

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

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

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

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

Aufgabe 3

Die Summe der Vorfälle 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:

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

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 Daten lassen sich über die Variationskoeffizienten vergleichen (keine Berechnung gefordert: der Variationskoeffizient der weltweiten Daten ist $v_y = 0,214$).

Aufgabe 4

Ergänzte Tabelle der absoluten Häufigkeiten:

Y X	1	2	3	4	Summe
1	2	4	1	9	16
2	17	34	10	85	146
3	3	6	2	15	26
Summe	22	44	13	109	188

Die beiden Merkmale sind nicht statistisch unabhängig, da die relativen Häufigkeiten in den Spalten sowie der Randspalte unterschiedlich sind.

Tabelle der relativen Häufigkeiten:

Y X	1	2	3	4	Summe
1	1/11	1/11	1/13	9/109	4/47
2	17/22	17/22	10/13	85/109	73/94
3	3/22	3/22	2/13	15/109	13/94
Summe	1	1	1	1	1

Die bedingte Verteilung ist

$$f(x_i|Y = 2) = 1/11; 17/22; 3/22$$

Für die Varianz wir die zweite Zeile der Tabelle benutzt, das arithmetische Mittel ist

$$\frac{1}{n} \sum_{i=1}^n x_i = \frac{1}{146} (17 \cdot 1 + 34 \cdot 2 + 10 \cdot 3 + 85 \cdot 4) = 3,12$$

Die Varianz

$$s^2(Y|X = 2) = \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2$$

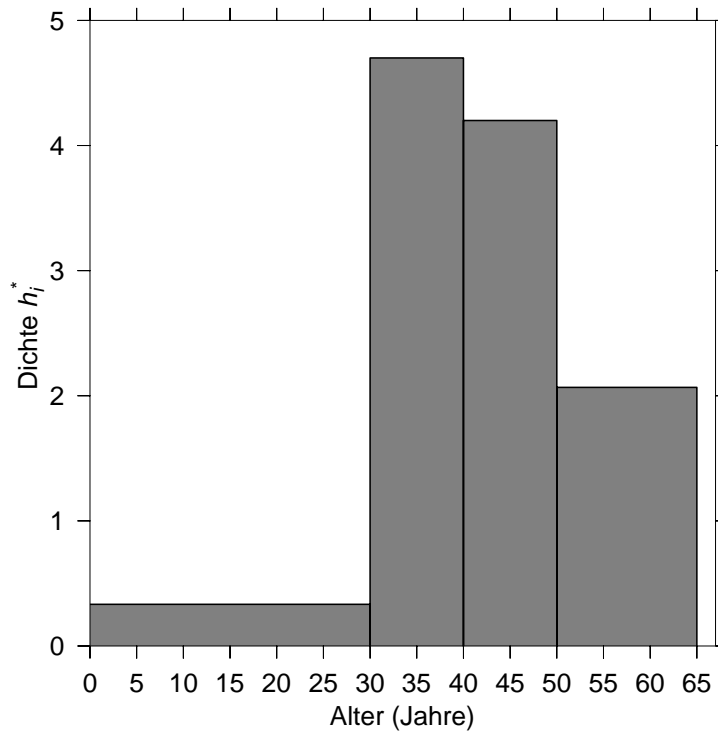
$$= \frac{1}{146} (17 \cdot (1 - 3,12)^2 + 34 \cdot (2 - 3,12)^2 + 10 \cdot (3 - 3,12)^2 + 85 \cdot (4 - 3,12)^2) = 1,27$$

Aufgabe 5

Daten zur Altersverteilung:

Alter in Jahren von ... bis unter ...	Absolute Häufigkeit	relative Häufigkeit f_i	kumulierte rel. Häufigkeit F_i	Klassenbreite Δ_i	Dichte $h_i^* = h_i / \Delta_i$
bis 30	10	0,076	0,076	30	0,333
30 - 40	47	0,362	0,438	10	4,7
40 - 50	42	0,323	0,761	10	4,2
50 - 65	31	0,238	1	15	20,67

(a) Histogramm der Altersverteilung:
Altersverteilung



(b) Berechnung des Durchschnittsalters (in Jahren):

$$\bar{x} = \frac{1}{130}(10 \cdot 15 + 47 \cdot 35 + 42 \cdot 45 + 31 \cdot 57,5) = 42,06$$

(c) Berechnung des Medians:

50% der befragten Personen werden in der 3.Klasse erreicht. Der Median lässt sich also wie folgt berechnen:

$$F(\bar{x}_Z) = x_3^u + (x_3^o - x_3^u) \frac{F(\bar{x}_Z) - F(x_3^u)}{F(x_3^o) - F(x_3^u)} = x_3^u + (x_3^o - x_3^u) \frac{F(0,5) - F(x_3^u)}{F(x_3^o) - F(x_3^u)}$$
$$40 + 10 \cdot \frac{0,5 - 57/130}{(99 - 57)/130} = 40 + 10 \cdot \frac{8}{42} = 41,9$$

Aufgabe 6

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.