Musterlösung zur Klausur Statistik

TI22

Oettinger 03.2024

Zeit: 90Min.

Insgesamt erreichbare Punktzahl: 106, 100%: 100 Punkte.

Aufgabe 1

(15 Punkte)

(a) Nominale Merkmale besitzen keine natürliche Rangfolge, eine Symmetrie kann nicht definiert werden - falsch.

(b) Der Gini-Koeffizient eintspricht der doppelten Fläche zwischen der Lorenz-Kurve und der Winkelhalbierenden D, die Fläche unter D kann maximal $1/2$ sein - richtig.

(c) Der Median kann je nach Symmetrie einer Häufigkeitsverteilung im Vergleich zum arithmetischen Mittel größer, gleich oder kleiner sein - falsch.

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

(e) Das arithmetische Mittel kann natürlich größer als der Median sein - die Stichprobe ist dann meist linkssteil - richtig.

Aufgabe 2

(14 Punkte) Daten zur Gewichtsverteilung mit Häufigkeitsdichte:

i	Klasse $(x_i^u; x_i^o]$	h_i	Mitte	Breite	Dichte h_i^*	rel. Häufigkeit f_i	kumuliert F_i
1	(985; 995]	15	990	10	1,5	0,3	0,3
2	(995; 1000]	5	997,5	5	1	0,1	0,4
3	(1000; 1005]	20	1002,5	5	4	0,4	0,8
4	(1005; 1020]	10	1012,5	15	0,666	0,2	1

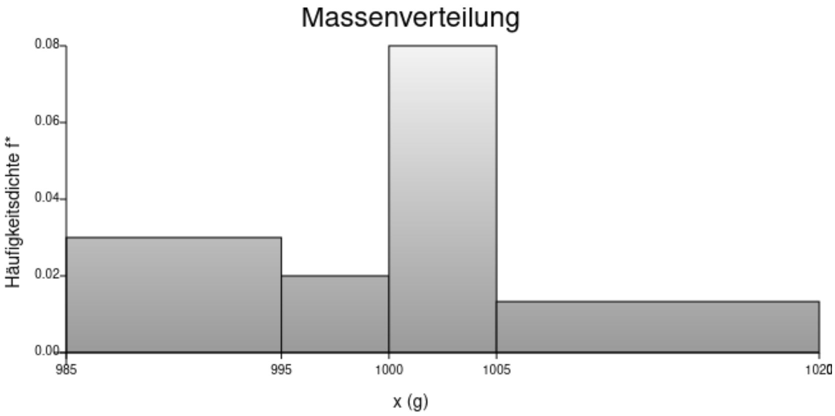


Abbildung 1: Histogramm zum Gewicht des Quallengelees

Die Näherung für das arithmetische Mittel (in Gramm) über die Klassenmitten ist

$$\bar{x} \approx \frac{1}{50}\,(15 \cdot 990 + 5 \cdot 997,5 + 20 \cdot 1002,5 + 10 \cdot 1012,5) = 1000,25.$$

Die Hälfte der untersuchten Pakete (25) wird in der dritten Klasse erreicht ($i = 3$), der Median kann ebenfalls geschätzt werden (in Gramm)

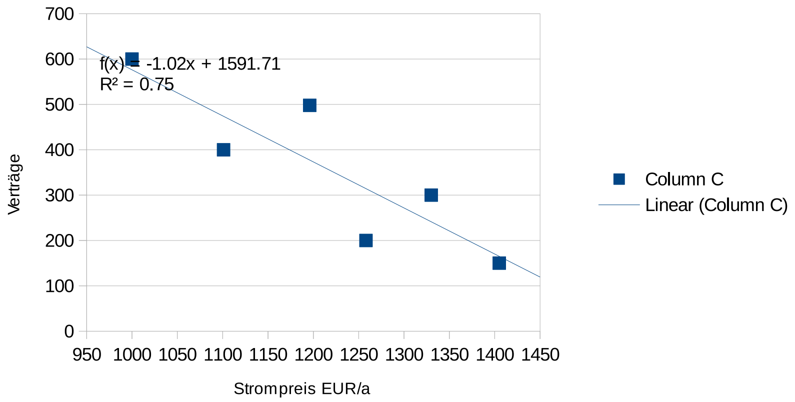
$$\bar{x_Z} \approx x_3^u + (x_3^o - x_3^u)\frac{0,5 - F(x_3^u)}{F(x_3^o) - F(x_3^u)} = 1000 + 5 \cdot \frac{0,5 - 0,4}{0,8 - 0,4} = 1001,25$$

Als typischer Wert sind beide geeignet - in den meisten Fällen gibt der Median die Verhältnisse der Verteilung aber etwas besser wieder.

Aufgabe 3

(27 Punkte)

a) Daten mit Ausgleichsgerade



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

Preis	Verträge	$(x_i - \bar{x})$	$(y_i - \bar{y})$	$(x_i - \bar{x})(y_i - \bar{y})$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$
1000	600	-215	242	-52030	46225	58564
1101	400	-114	42	-4788	12996	1764
1196	498	-19	140	-2660	361	19600
1258	200	43	-158	-6794	1849	24964
1330	300	115	-58	-6670	13225	3364
1405	150	190	-208	-39520	36100	43264

Die Steigung der Geraden ist

$$a = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = -1,02$$

der Achsenabschnitt ist

$$b = \bar{y} - a\bar{x} = 1591,71$$

c) der Pearson-Koeffizient ist

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

$$= \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2} \cdot \sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}} = -0,87 \tag{2}$$

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

Aufgabe 4

(27 Punkte)
Die benötigten Daten zur Aufgabe:

Zahl der Tabellen x_i	Tage früher	Tage jetzt	kumuliert früher	kumuliert jetzt
1	60	5	60	5
2	160	10	220	15
3	110	25	330	40
4	0	20	330	60
5	60	0	390	60
6	50	0	440	60
8	0	40	440	100

a) Die Aussage ist: Lohnt sich der Droide (ist die Produktivität angestiegen?).

b) Benötigt werden das Arithmetische Mittel und der Median:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{m} h_i x_i$$

$$\bar{x} = \frac{1}{440}(60 + 2 \cdot 160 + 3 \cdot 110 + 4 \cdot 0 + 5 \cdot 60 + 6 \cdot 50 + 8 \cdot 0) = 2,977 \text{ ohne Droide,}$$

$$\bar{x} = \frac{1}{100}(5 + 2 \cdot 10 + 3 \cdot 25 + 4 \cdot 20 + 5 \cdot 0 + 6 \cdot 0 + 8 \cdot 40) = 5 \text{ mit Droide.}$$

Der Median lässt sich aus den Daten in der Tabelle ablesen, für die Daten ohne den Droiden

$$\bar{x}_Z = \frac{x_{220} + x_{221}}{2} = 2,5,$$

nach der Einführung des Droiden

$$\bar{x}_Z = \frac{x_{50} + x_{51}}{2} = 4.$$

Die mittlere absolute Abweichung vom arithmetischen Mittel ist

$$\begin{aligned} d_{\bar{x}} &= \frac{1}{n}\sum_{i=1}^{m} h_i |x_i - \bar{x}|, \\ &= 1,25 \text{ ohne Droide bzw.} \\ &= 2,4 \text{ mit Droide.} \end{aligned}$$

Die mittlere absolute Abweichung vom Median ist

$$\begin{aligned} d_{\bar{x}_Z} &= \frac{1}{n}\sum_{i=1}^{m} h_i |x_i - \bar{x}_Z|, \\ &= 1,25 \text{ ohne Droide bzw.} \\ &= 2,2 \text{ mit Droide.} \end{aligned}$$

Aufgabe 5

(14 Punkte)
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: das mittlere Wachstum ist

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

   Ist die Größe der Kultur zu Beginn N_0, beträgt sie nach 5 Tagen

$$N_5 = (1+\bar{q})^5 N_0 = (1,0813)^5 N_0 = 1,48 N_0,$$

   das entspricht einem Wachstum von von 48%.

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

Aufgabe 6

(8 Punkte)
Nominale/ordinale/kardinale Merkmale:

(a) Körpergröße: kardinal (Zahl)

(b) Farbe: nominal (keine Rangfolge)

(c) Krawattenlänge: kardinal (Zahl)

(d) Qualität von Vorlesungen: ordinal (keine Zahl, aber mit Rangfolge)