Musterlösung zur Übungsklausur Statistik

TI 23 / Oettinger / 3.2025

Aufgabe 1

(a) Richtig: das arithmetische Mittel reagiert im Gegensatz zum Median empfindlich auf Ausreißer.

(b) Falsch: die Varianz und damit auch die Standardabweichung kann natürlich auch Null werden (wenn alle Abweichungen vom arithmetischen Mittel zu Null verschwinden) - dann ist eigentlich keine Statistik nötig.

(c) Richtig: es handelt sich um eine Summe quadrierter Größen.

(d) Falsch: der Modus ist die am häufigsten auftretende Merkmalsausprägung in einer Stichprobe.

(e) Falsch: es handelt sich um ein mittleres Wachstum, es kann auch negativ sein.

Aufgabe 2

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

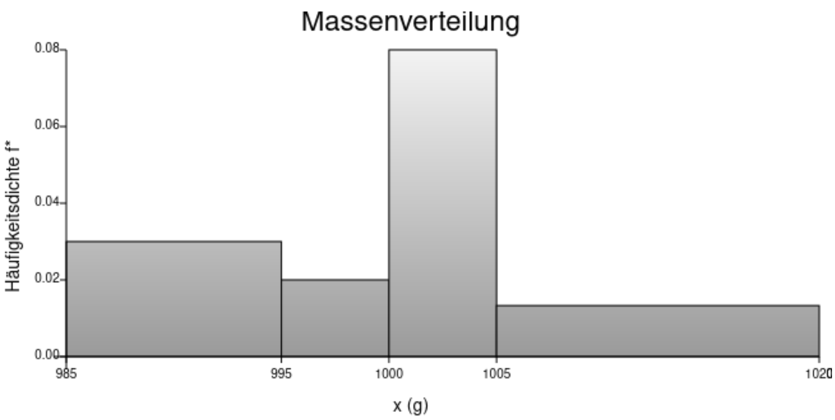


Abbildung 1: Histogramm zum Gewicht des Quallengelees

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

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

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

Die arithmetischen Mittel für die Siedetemperatur T_S und den Druck sind $\bar{x} = 95,6$°C und $\bar{y} = 652,806$ mm Hg. Damit lassen sich die benötigten Daten berechnen:

T_S (°C)	Druck	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$	$(x_i - \bar{x})(y_i - \bar{y})$
90.3	528.07	-5.3	-124.736	28.09	15559.069696	661.1008
92.2	568.96	-3.4	-83.846	11.56	7030.151716	285.0764
93	588.01	-2.6	-64.796	6.76	4198.521616	168.4696
93.8	606.81	-1.8	-45.996	3.24	2115.632016	82.7928
94.1	610.11	-1.5	-42.696	2.25	1822.948416	64.044
95.9	674.88	0.3	22.074	0.09	487.261476	6.6222
98.6	723.65	3	70.844	9	5018.872336	212.532
98.1	705.1	2.5	52.294	6.25	2734.662436	130.735
99.9	758.95	4.3	106.144	18.49	11266.548736	456.4192
100.1	763.52	4.5	110.714	20.25	12257.589796	498.213

Die Kovarianz (nicht gefragt, in °C · mm Hg) ist

$$s_{xy} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) = 256,60$$

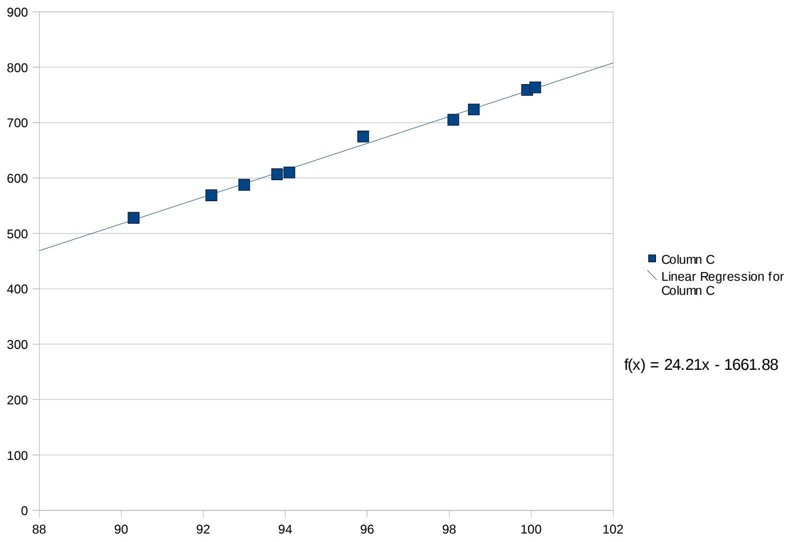
Lineare Regression: sei $y(x) = a \cdot x + b$:

$$a = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2} = 24,21 \text{ mm Hg / °C}$$

und

$$b = \bar{y} - a\bar{x} = -1661,88 \text{ mm Hg.}$$

Die Standardabweichungen der beiden Größen sind $s_x = 3,26$°C und $s_y = 79,05$ mm Hg. Plot der Daten:



Aufgabe 4

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 4

Redezeiten von Politikern, alle Angaben und Ergebnisse in Minuten:

Politiker	A	B	C	D	E	F
Redezeit	6	8	10	12	20	10

(a) Geordneter Vektor der Merkmalswerte: (6, 8, 10, 10, 12, 20). Median ist

$$\bar{x}_Z = \frac{10+10}{2} = 10.$$

(b) $\bar{x} = \frac{1}{6} \cdot (6+8+10+10+12+20) = 11.$

(c) $s_w =$max$(x_i)-$min$(x_i) = 20 - 6 = 14.$

(d) mittlere absolute Abweichung

$$d_{\bar{x}} = \frac{1}{6} \sum_{i=1}^{n} \mid x_i - \bar{x} \mid$$

$$= \frac{1}{6} (\mid 6-11 \mid + \mid 8-11 \mid +2 \cdot \mid 10-11 \mid + \mid 12-11 \mid + \mid 20-11 \mid) = 10/3$$

Aufgabe 5

a) Die benötigten Daten:

Einkommen in Talern	Klassenbreite in Tausend Talern	Häufigkeit	relativ	kumuliert	Dichte
]0-500]	0,5	9	0,09	0,09	18
]500-1000]	0,5	13	0,13	0,22	26
]1000-1500]	0,5	32	0,32	0,54	64
]1500-2000]	0,5	41	0,41	0,95	82
]2000-3000]	1,0	3	0,03	0,98	3
]3000-5000]	2,0	2	0,02	1,0	1

Histogramm der Einkommen:

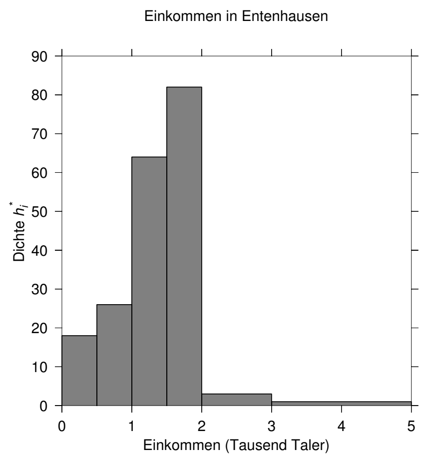


Abbildung 2: Einkommensverteilung in Entenhausen

b) arithmetisches Mittel (in Tausend Taler):

$$\bar{x} = \frac{1}{100}(0,25 \cdot 9 + 0,75 \cdot 13 + 1,25 \cdot 32 + 2,5 \cdot 3 + 4,0 \cdot 2) = \frac{139,25}{100} = 1,39$$

50% der befragten Personen werden in der 3.Klasse erreicht, der Median (in Tausend Taler) ist

$$F(\bar{x}_Z) = x_3^u + (x_3^o - x_3^u)\frac{F(\bar{x}_Z) - F(x_3^u)}{F(x_3^o) - F(x_3^u)} = x_3^u + (x_3^o - x_3^u)\frac{F(0,5) - F(x_3^u)}{F(x_3^o) - F(x_3^u)}$$

$$1 + (1,5 - 1) \cdot \frac{0,5 - 0,22}{0,54 - 0,22} = 1,44.$$

Der Median ist etwas größer - die Verteilung scheint leicht rechtssteil.

c) Für die Lorenzkurve wird das gesamte Einkommen benötigt (in Tausend Talern):

$$q_{ges} = 0,25 \cdot 9 + 0,75 \cdot 13 + 1,25 \cdot 32 + +2,5 \cdot 3 + 4,0 \cdot 2 = 139,25$$

Jetzt kann der Anteil der Personen in jeder Klasse am Gesamteinkommen berechnet werden:

$$q_k = h_k \cdot (x_k^o + x_k^u)/2 \cdot \frac{1}{q_{ges}}.$$

Einkommen in Talern	relative Häufigkeit f_k	kumuliert F_k	Anteil q_k	kumuliert Q_k
]0-500]	0,09	0,09	0,016	0,016
]500-1000]	0,13	0,22	0,070	0,102
]1000-1500]	0,32	0,54	0,287	0,389
]1500-2000]	0,41	0,95	0,515	0,904
]2000-3000]	0,03	0,98	0,054	0,958
]3000-5000]	0,02	1,0	0,057	1

Lorenz-Plot:

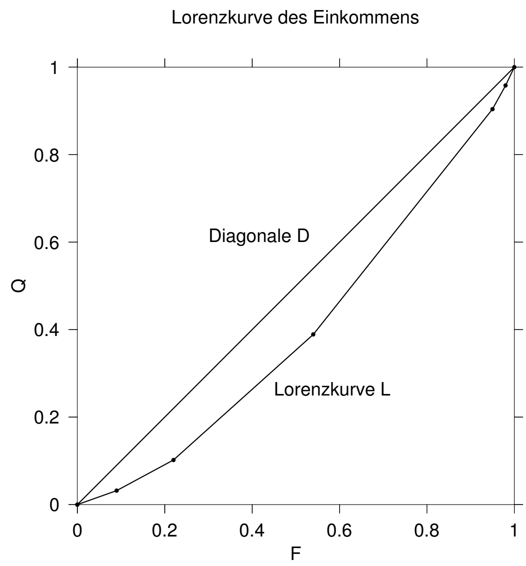


Abbildung 3: Lorenzkurve der Einkommensverteilung in Entenhausen

Der Gini-Koeffizient entspricht der doppelten Fläche zwischen den beiden Kurven. Man erkennt, dass die relative Konzentration (und damit der Gini-Koeffizient) nicht besonders groß ist.

Aufgabe 6

Landtagswahlen: die Aussage des Herrn Trumpheimer kann durch Vergleich der Variationskoeffizienten überprüft werden. Benötigt werden die arithmetischen Mittel (in %)

$$\bar{x}_A = \frac{1}{7}(5,6 + 6,3 + 6,6 + 6,9 + 7,1 + 7,6 + 6,1) = 6,6$$

$$\bar{x}_B = \frac{1}{7}(40,4 + 41,9 + 47,9 + 40,4 + 48,9 + 41,4 + 42,9) = 43,4$$

sowie die Varianzen und die daraus berechneten Standardabweichungen

$$s_A^2 = \frac{1}{7}\left(5,6^2 + 6,3^2 + 6,6^2 + 6,9^2 + 7,1^2 + 7,6^2 + 6,1^2\right) - 6,6^2$$

$$= 0,383$$

$$\Rightarrow s_A = \sqrt{s_A^2} = 0,619$$

$$
\begin{aligned}
s_B^2 &= \frac{1}{7}\left(40,4^2 + 41,9^2 + 47,9^2 + 40,4^2 + 48,9^2 + 41,4^2 + 42,9^2\right) - 43,4^2 \\
&= 10,714 \\
&\Rightarrow s_B = \sqrt{s_B^2} = 3,273.
\end{aligned}
$$

Die Variationskoeffizienten $v_x = \frac{s_x}{\bar{x}}$ für die beiden Stichproben sind

$$v_A = \frac{s_A}{\bar{x}_A} = \frac{0,619}{6,6} = 0,094$$

und

$$v_B = \frac{s_B}{\bar{x}_B} = \frac{3,273}{43,4} = 0,075.$$

Damit wird klar, dass Herr Trumpheimer falsch liegt, die Verteilung der Stimmenanteile für seine Partei A ist deutlich breiter.