Aufgabe 1

(10 Punkte)
Welche der folgenden Aussagen sind richtig?

(a) Richtig, die Varianz ist eine Summe quadratischer Größen.

(b) Falsch, die Abweichung ordinaler Merkmale vom Median ist nicht definiert - also auch keine mittlere Abweichung.

(c) Richtig, die untersuchten Merkmale besitzen keine Lebensdauer.

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

(e) Richtig, der Quantilsabstand steigt mit der Breite der Verteilung an.

Aufgabe 2

(25 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 Computer (ist die Produktivität angestiegen?).

b) Benötigt werden das Arithmetische Mittel und der Median:

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

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

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

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

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

nach der Einführung des Computers

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

Die mittlere absolute Abweichung vom arithmetischen Mittel ist

$$d_{\bar{x}} = \frac{1}{n}\sum_{i=1}^{m} h_i(x_i - \bar{x}),$$

$= 1,25$ ohne Computer bzw.

$= 2,4$ mit Computer.

Die mittlere absolute Abweichung vom Median ist

$$d_{\bar{x}_Z} = \frac{1}{n}\sum_{i=1}^{m} h_i(x_i - \bar{x}_Z),$$

$= 1,25$ ohne Computer bzw.

$= 2,2$ mit Computer.

Aufgabe 3

(11 Punkte)

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

a) Arithmetisches Mittel (in km/h):

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

b) Harmonisches Mittel (in km/h):

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

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

$$\frac{7 \cdot 10\text{km}}{\frac{10\text{km}}{15\text{km/h}} + \frac{10\text{km}}{16{,}5\text{km/h}} + \frac{10\text{km}}{17{,}5\text{km/h}} + \frac{10\text{km}}{18\text{km/h}} + \frac{10\text{km}}{18\text{km/h}} + \frac{10\text{km}}{20\text{km/h}} + \frac{10\text{km}}{22\text{km/h}}}$$

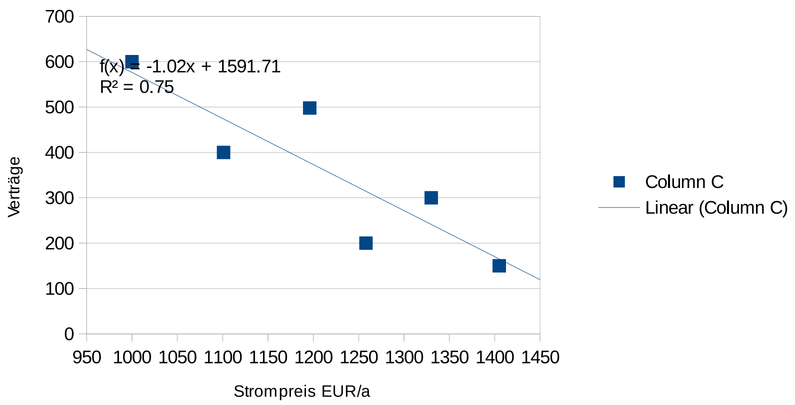
Die Anwendung des harmonischen Mittels ist hier korrekt.

d) Vielleicht!

Aufgabe 4

(28 Punkte)

a) Daten mit Ausgleichsgerade



Anpassung einer Geraden $y = a \cdot x + b$ über lineare Regression: die arithmetischen Mittel sind $\bar{x} = 1215$ und $\bar{y} = 358$, die Tabelle der 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

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

der Achsenabschnitt ist

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

b) Die erwartete Zahl der Verträge ist bei einem linearen Zusammenhang von $y = 1591,71 - 1,02 \cdot x$

$$y(1200) = 367,71.$$

Es werden also etwa 368 Verträge erwartet.

c) der Pearson-Koeffizient ist

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

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

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

Aufgabe 5

(22 Punkte)
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

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

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

Die Varianz

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

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

Aufgabe 6

(9 Punkte)

a) Geometrisches Mittel:
$\bar{x}_G = \sqrt[3]{(1+0,2) \cdot (1+0,15) \cdot (1-0,0005)} - 1 = 9,68\%$

b) Der Mittelwert berechnet sich nach

$$\bar{x} = \frac{1}{30}(10 \cdot 1 + 11 \cdot 2 + 8 \cdot 3 + 4) = \frac{60}{30} = 2,0.$$

Für die Kandidaten, die nicht bestanden haben, kann keine Note angegeben werden, sie werden zur Berechnung *nicht* herangezogen.

c) Es wurden insgesamt $n = 11$ Hotels untersucht, auch das Hotel ohne Stern kann mitgezählt werden (es gibt nur die eine Möglichkeit 0). also ist das arithmetische Mittel

$$\bar{x} = \frac{1}{11}(5 \cdot 3 + 2 \cdot 2 + 3 \cdot 1 + 0) = \frac{22}{11} = 2,0.$$

d) 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.