1. [12 Punkte]

Ein Unternehmen hat folgende Daten ermittelt:

Monat	Geleistete Arbeitsstunden	Lohnkosten pro Arbeitsstunde
Januar	862.400	20,06
Februar	752.000	23,01
März	756.000	24,47
April	768.000	22,53
Mai	638.400	27,10
Juni	640.000	28,91
Juli	644.000	36,00
August	528.000	32,77
September	588.000	31,80
Oktober	792.000	21,84
November	784.000	22,07
Dezember	676.800	36,33

Wie hoch sind die Lohnkosten pro Arbeitsstunde im Jahresdurchschnitt?

$$h := \begin{pmatrix} 862400 \\ 752000 \\ 756000 \\ 768000 \\ 638400 \\ 640000 \\ 644000 \\ 528000 \\ 588000 \\ 792000 \\ 784000 \\ 676800 \end{pmatrix} \qquad x := \begin{pmatrix} 20.06 \\ 23.01 \\ 24.47 \\ 22.53 \\ 27.1 \\ 28.91 \\ 36 \\ 32.77 \\ 31.8 \\ 21.84 \\ 22.07 \\ 36.33 \end{pmatrix}$$

$$n := \text{länge}(h) = 12$$

$$i := 1..n$$

$$\frac{\sum_{i=1}^{n} \left(h_i \cdot x_i\right)}{\sum_{i=1}^{n} h_i} = 26.64$$

2. [3 Punkte]

Es gebe *n* Ausprägungen des Merkmals *x*. Die Merkmalsausprägungen werden mit i = 1…n indiziert. Wie hoch ist das arithmetische Mittel der Merkmalsausprägungen?

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

3. [4 Punkte]

Es gebe *m* unterschiedliche Ausprägungen des Merkmals *x*, die mit der Häufigkeit *h* vorkommen. Die unterschiedlichen Merkmalsausprägungen und die Häufigkeiten werden mit j = 1…m indiziert. Wie hoch ist das arithmetische Mittel der Merkmalsausprägungen?

$$\bar{x} = \frac{\sum_{j=1}^{m} h_j \cdot x_j}{\sum_{j=1}^{m} h_j}$$

4. [6 Punkte]

Beweisen Sie: Die Summe der Abweichungen vom arithmetischen Mittel ist gleich null.

Behauptung:

$$\sum_{i=1}^{n} \left(\bar{x} - x_i \right) = 0$$

Beweis:

$$\bar{x} = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i \quad \text{ex def.}$$

$$\sum_{i=1}^{n} \left(\bar{x} - x_i \right) = \sum_{i=1}^{n} \bar{x} - \sum_{i=1}^{n} x_i = n \cdot \bar{x} - \sum_{i=1}^{n} x_i = n \cdot \frac{1}{n} \cdot \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} x_i = 0$$

5. [3 Punkte]

In einem bestimmten Zeitraum war der Durchschnittskurs von Aktie A 35,00 mit einer Standardabweichung von 5,00. Bei Aktie B war der Durchschnittskurs im gleichen Zeitraum 9,50 mit einer Standardabweichung von 2,60. Wie lässt sich anhand dieser Daten die Volatilität der Aktien messen, und welche war volatiler? [Quelle: M. Piazolo, Statistik für Wirtschaftswissenschaftler - Daten sinnvoll aufbereiten, analysieren und interpretieren, Karlsruhe 2007, S. 95 f.]

Die Volatilität kann durch den Variationskoeffizienten gemessen werden:

$$\text{Variationskoeffizient} = \frac{\text{Standardabweichung}}{\text{Arithmetisches Mittel}}$$

$\text{Standardabweichung}_A := 5$ $\text{Standardabweichung}_B := 2.6$

$\text{Durchschnittskurs}_A := 35$ $\text{Durchschnittskurs}_B := 9.5$

$$\text{Variationskoeffizient}_A := \frac{\text{Standardabweichung}_A}{\text{Durchschnittskurs}_A} = 0.143$$

$$\text{Variationskoeffizient}_B := \frac{\text{Standardabweichung}_B}{\text{Durchschnittskurs}_B} = 0.274$$

Aktie B ist also volatiler als Aktie A.

Der Variationskoeffizient als relatives Streuungsmaß ist hier wegen des unterschiedlichen Kursniveaus geeigneter als ein absolutes Streuungsmaß.

6. [2 Punkte]

Warum wird der Median auch als 50%-Quantil bezeichnet?

Der Median ist der Zentralwert einer Häufigkeitsverteilung, d.h. links vom Median liegen genau so viele Merkmalsausprägungen wie rechts davon.

Quantile unterteilen die geordneten Merkmalswerte in definierte Teile. Das 50%-Quantil unterteilt also die Merkmalswerte in 50%, die links davon liegen und 50% rechts davon, ebenso wie der Median.

7. [15 Punkte]

In einer Firma werden folgende Monatseinkommen erzielt:

Mitarbeiter 1	1.500,00
Mitarbeiter 2	1.800,00
Mitarbeiter 3	2.200,00
Mitarbeiter 4	2.500,00
Chef	8.000,00

Bereiten Sie die Daten so auf, dass daraus eine Lorenzkurve gezeichnet werden kann und zeichnen Sie diese.

$\text{ORIGIN} \equiv 1$

$x := \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}$ Anzahl der Merkmalsträger

$y := \begin{pmatrix} 1500 \\ 1800 \\ 2200 \\ 2500 \\ 8000 \end{pmatrix}$ Merkmalsausprägung

$n := \text{länge}(x) = 5$ Anzahl der unterschiedlichen Merkmalsauprägungen

$i := 1..n$ Index der unterschiedlichen Merkmalsausprägungen

x_i = Anzahl der Träger der Merkmalsausprägung i

y_i = Merkmalsausprägung i

$$xka(i) := \frac{\sum_{i=1}^{i} x_i}{\sum_{i=1}^{n} x_i}$$ Kumulierter Anteil der Merkmalsträger

$xka(i) =$

0.2
0.4
0.6
0.8
1

$$yka(i) := \frac{\sum_{i=1}^{i} y_i}{\sum_{i=1}^{n} y_i}$$

Kumulierter Anteil der Merkmalsausprägungen

$yka(i) =$

0.094
0.206
0.344
0.5
1

$x0 := 0, 0.001 .. xka(1)$

Abschnitt der x-Achse zwischen 0 und xka(1)

$$y0(x0) := \frac{yka(1)}{xka(1)} \cdot x0$$

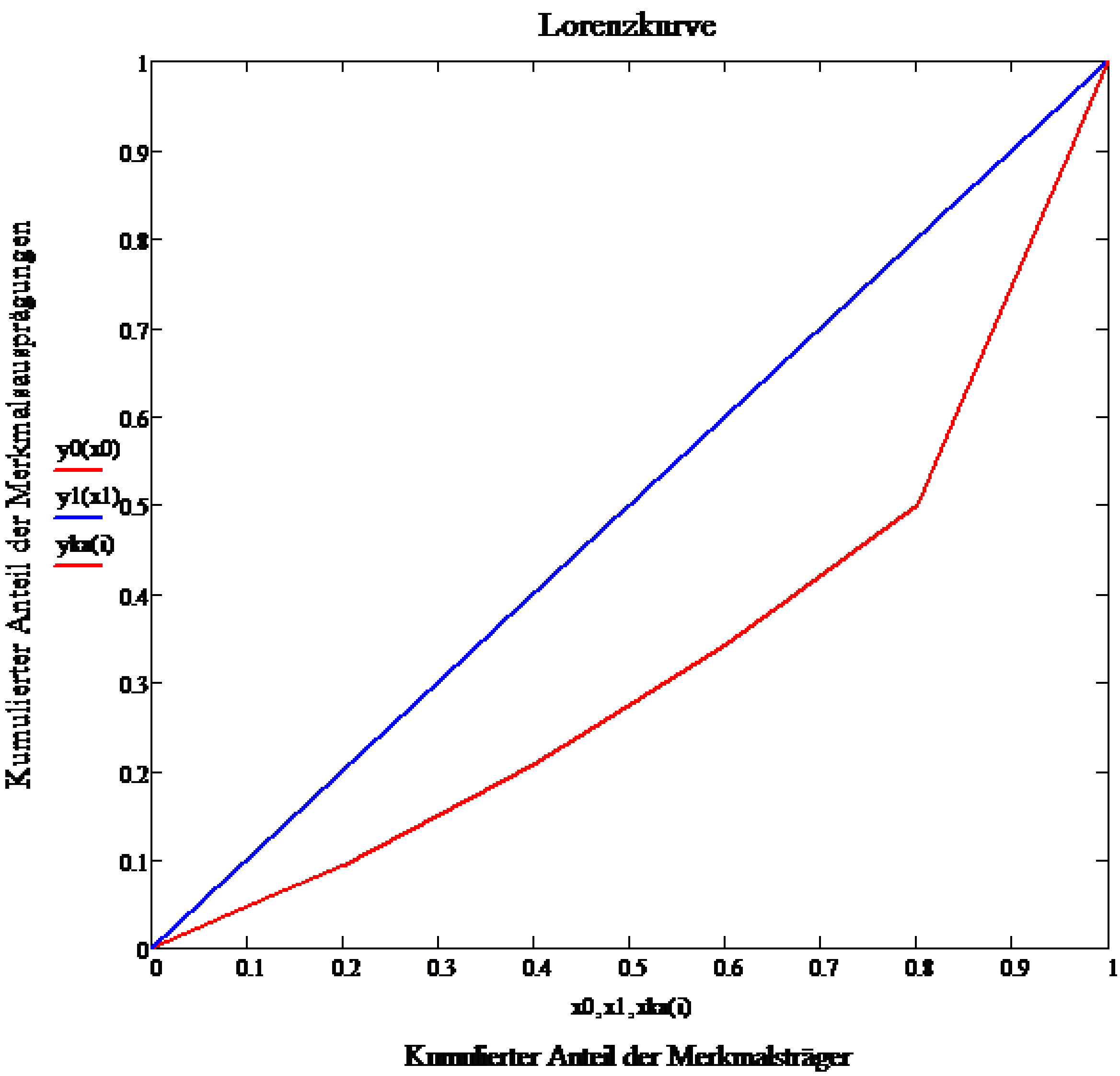
Funktion der Lorenzkurve zwischen 0 und xka(1)

$x1 := 0, 0.001 .. 1$

Abschnitt der x-Achse für die Gleichverteilungsgerade

$y1(x1) := x1$

Funktion der Gleichverteilungsgeraden



8. [5 Punkte]

Gegeben seien n Wertepaare der Merkmalsausprägungen x_i und y_i, wobei i = 1…n. Für jedes i gelte $x_i - \overline{x} = y_i - \overline{y}$.

Beweisen Sie, dass unter dieser Voraussetzung für den Korrelationskoeffizienten r gilt:

$$r = \frac{\frac{1}{n}\cdot\sum_{i=1}^{n}\left(x_i-\overline{x}\right)\cdot\left(y_i-\overline{y}\right)}{\sqrt{\frac{1}{n}\cdot\sum_{i=1}^{n}\left(x_i-\overline{x}\right)^2}\cdot\sqrt{\frac{1}{n}\cdot\sum_{i=1}^{n}\left(y_i-\overline{y}\right)^2}} = 1$$

$$\frac{1}{n}\cdot\frac{\sum_{i=1}^{n}\left[\left(x_i-x_d\right)\cdot\left(x_i-x_d\right)\right]}{\sqrt{\frac{1}{n}\cdot\sum_{i=1}^{n}\left(x_i-x_d\right)^2}\cdot\sqrt{\frac{1}{n}\cdot\sum_{i=1}^{n}\left(x_i-x_d\right)^2}} \rightarrow 1$$

$$\frac{\sum_{i=1}^{n}(x_i-\overline{x})\cdot(x_i-\overline{x})}{\sqrt{\sum_{i=1}^{n}(x_i-\overline{x})^2}\cdot\sqrt{\sum_{i=1}^{n}(x_i-\overline{x})^2}}=\frac{\sum_{i=1}^{n}(x_i-\overline{x})^2}{\left[\sum_{i=1}^{n}(x_i-\overline{x})^2\right]^{0,5}\cdot\left[\sum_{i=1}^{n}(x_i-\overline{x})^2\right]^{0,5}}=\frac{\sum_{i=1}^{n}(x_i-\overline{x})^2}{\sum_{i=1}^{n}(x_i-\overline{x})^2}=1$$

9. [20 Punkte]

Zwischen der Produktionsmenge *x* und den Kosten *K* eines Unternehmens wird folgender Zusammenhang festgestellt:

x	0	1	2	3	4
K	98,00	147,00	178,00	197,00	210,00

a) Wie stark ist die Korrelation zwischen den Kosten und der Produktionsmenge?
b) Welche Regressionsgerade ergibt sich?
c) Wie hoch sind die Kosten für die Mengen 6 und 7 aufgrund der Regressionsgeraden?

Sie können mit den Definitionen aus Aufgabe 8 folgende Formeln verwenden:

Kovarianz von x und y: $s_{xy}=\frac{1}{n}\cdot\sum_{i=1}^{n}(x_i-\overline{x})\cdot(y_i-\overline{y})$

Standardabweichung von x: $s_x=\sqrt{\frac{1}{n}\cdot\sum_{i=1}^{n}(x_i-\overline{x})^2}$

Standardabweichung von y: $s_y=\sqrt{\frac{1}{n}\cdot\sum_{i=1}^{n}(y_i-\overline{y})^2}$

Korrelationskoeffizient: $r=\frac{s_{xy}}{s_x\cdot s_y}$

Steigung der Regressionsgeraden: $b=\frac{s_{xy}}{{s_x}^2}$

Achsenabschnitt der Regressionsgeraden: $a=\overline{y}-b\cdot\overline{x}$

Regressionsgerade: $y=a+b\cdot x$

$\mathrm{ORIGIN}\equiv 1$

$$x:=\begin{pmatrix}0\\1\\2\\3\\4\end{pmatrix}$$

$n:=\mathrm{länge}(x)=5$

$i:=1..n$

$$x_d := \frac{1}{n} \cdot \sum_{i=1}^{n} x_i = 2$$

$$K := \begin{pmatrix} 98 \\ 147 \\ 178 \\ 197 \\ 210 \end{pmatrix}$$

$$K_d := \frac{1}{n} \cdot \sum_{i=1}^{n} K_i = 166$$

$x_i - x_d =$	$(x_i - x_d)^2 =$	$K_i - K_d =$	$(K_i - K_d)^2 =$	$(x_i - x_d) \cdot (K_i - K_d) =$
-2	4	-68	4624	136
-1	1	-19	361	19
0	0	12	144	0
1	1	31	961	31
2	4	44	1936	88

$$\sum_i (x_i - x_d)^2 = 10 \qquad \sum_i (K_i - K_d)^2 = 8026 \qquad \sum_i \left[(x_i - x_d) \cdot (K_i - K_d)\right] = 274$$

$$\frac{1}{n} \cdot \sum_i (x_i - x_d)^2 = 2 \qquad \frac{1}{n} \cdot \sum_i (K_i - K_d)^2 = 1605.2 \qquad \frac{1}{n} \cdot \sum_i \left[(x_i - x_d) \cdot (K_i - K_d)\right] = 54.8$$

$$\sqrt{\frac{1}{n} \cdot \sum_i (x_i - x_d)^2} = 1.414214 \qquad \sqrt{\frac{1}{n} \cdot \sum_i (K_i - K_d)^2} = 40.064947$$

$$s_{xy} := \frac{1}{n} \cdot \sum_{i=1}^{n} \left[(x_i - x_d) \cdot (K_i - K_d)\right] = 54.8$$

$$s_x := \sqrt{\frac{1}{n} \cdot \sum_i (x_i - x_d)^2} = 1.414$$

$$s_y := \sqrt{\frac{1}{n} \cdot \sum_i (K_i - K_d)^2} = 40.065$$

a)

$$r := \frac{s_{xy}}{s_x \cdot s_y} = 0.967$$ Hohe Korrelation

$$\mathrm{korr}(x, K) = 0.967$$

b)

$$b := \frac{s_{xy}}{s_x^2} = 27.4$$

$$a := K_d - b \cdot x_d = 111.2$$

$$K(x) := a + b \cdot x$$

c)

$$K(6) = 275.6$$

$$K(7) = 303$$

10. [2 Punkte]

Tatsächlich folgen die Kosten aus Aufgabe 9 der Funktion $K(x) = x^3 - 12x^2 + 60x + 98$.

Wenn diese Funktion als Regressionsfunktion vorgegeben wird, wie hoch ist dann der Korrelationskoeffizient zwischen den Kosten gemäß Aufgabe 9 und den Werten gemäß der Regressionsfunktion?

$$K := \begin{pmatrix} 98 \\ 147 \\ 178 \\ 197 \\ 210 \end{pmatrix} \qquad K1(x) := x^3 - 12x^2 + 60x + 98 \qquad K1 := \begin{pmatrix} K1(0) \\ K1(1) \\ K1(2) \\ K1(3) \\ K1(4) \end{pmatrix} \qquad K1 = \begin{pmatrix} 98 \\ 147 \\ 178 \\ 197 \\ 210 \end{pmatrix}$$

$$\mathrm{korr}(K, K1) = 1$$

11. [4 Punkte]

Die folgende Tabelle zeigt den durchschnittlichen Bruttomonatsverdienst der Arbeitnehmer im Kredit- und Versicherungsgewerbe in den Jahren 2001 bis 2010:
[Quelle: Institut der deutschen Wirtschaft Köln (Hrsg.), Deutschland in Zahlen - Ausgabe 2011 -, Köln 2011, S. 56]

Jahr	Monatsgehalt
2001	3.321
2002	3.427
2003	3.525
2004	3.619
2005	3.692
2006	3.761
2007	3.865
2008	4.012
2009	4.097
2010	4.258

Wie hoch ist die durchschnittliche jährliche Wachstumsrate der Bruttomonatsverdienste zwischen 2001 und 2010? [Angabe in % mit zwei Stellen nach dem Komma]

$$\mathrm{Gehalt}_{2001} := 3321$$

$$\mathrm{Gehalt}_{2010} := 4258$$

$$w_n := \left(\frac{\mathrm{Gehalt}_{2010}}{\mathrm{Gehalt}_{2001}}\right)^{\frac{1}{2010-2001}} - 1 = 2.80\%$$

12. [2 Punkte]

Der Verbraucherpreisindex ist von 94,5 im Jahre 2001 auf 108,2 im Jahre 2010 gestiegen. Wie hoch war die durchschnittliche jährliche Preissteigerungsrate? [Angabe in % mit zwei Stellen nach dem Komma]

$$\mathrm{Index}_{2001} := 94.5$$

$$\mathrm{Index}_{2010} := 108.2$$

$$w_p := \left(\frac{\mathrm{Index}_{2010}}{\mathrm{Index}_{2001}}\right)^{\frac{1}{2010-2001}} - 1 = 1.52\%$$

13. [4 Punkte]

Wie hoch ist die reale Wachstumsrate der Gehälter aus Aufgabe 11 zwischen 2001 und 2010, wenn man die Preisentwicklung aus Aufgabe 12 zugrunde legt? [Angabe in % mit zwei Stellen nach dem Komma]

$$w_n := 2.8\%$$

$$w_p := 1.52\%$$

$$w_r := \frac{1 + w_n}{1 + w_p} - 1 = 1.26\%$$

14. [4 Punkte]

Die Wahrscheinlichkeit, dass bei einem Versicherungsnehmer in einem Geschäftsjahr der Schadenfall eintritt, sei w_s. Ein Kumulrisiko wird nicht in Betracht gezogen, d.h. die Wahrscheinlichkeit für den Schadeneintritt bei einem Versicherungsnehmer wird als unabhängig vom Schadeneintritt bei einem anderen Versicherungsnehmer angesehen. Wie wahrscheinlich ist es, dass in einem Geschäftsjahr der Schadenfall bei *i* Versicherungsnehmern derselben Risikogruppe eintritt?

$w_s^{\,i}$

15. [4 Punkte]

Die Wahrscheinlichkeit, dass bei einem Versicherungsnehmer in einem Geschäftsjahr der Schadenfall nicht eintritt, sei $1 - w_s$. Ein Kumulrisiko wird nicht in Betracht gezogen, d.h. die Wahrscheinlichkeit für den Schadeneintritt bei einem Versicherungsnehmer wird als unabhängig vom Schadeneintritt bei einem anderen Versicherungsnehmer angesehen. Wie wahrscheinlich ist es, dass der Schadenfall bei n – *i* Versicherungsnehmern derselben Risikogruppe in einem Geschäftsjahr nicht eintritt?

$(1-w_s)^{n-i}$