Ü02: Univariate Deskriptive Statistik, Teil 2
Lösung

Teil A

1) und 3)

a_j	$h(a_j)$	$f(a_j)$	m_1	m_2	m_3	m_4
5	2	2/6=0,33	5·1·2/6=1,67	5·1,67= 8,33	5· 8,33= 41,67	5· 41,67= 208,33
36/4=9	6–2=4	4/6=0,67	9·1·4/6=6	9·6 =54	9·54 =486	9·486 =4.374
Σ	**6**	**1,00**	**7,67**	**62,33**	**527,67**	**4.582,33**

2) $$m_p = \frac{1}{n}\sum_{i=1}^{n} x_i^p = \frac{1}{n}\sum_{j=1}^{n} a_j^p \cdot h(a_j) = \sum_{j=1}^{n} a_j^p \cdot \frac{h(a_j)}{n} = \sum_{j=1}^{n} a_j^p \cdot f(a_j) = \sum_{j=1}^{n} a_j \cdot \left(a_j^{p-1} \cdot f(a_j)\right) = \sum_{j=1}^{n} a_j \cdot m_{p-1_j}$$

4)

$$s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{1}{n}\sum_{i=1}^{n}\left(x_i^2 - 2\cdot x_i \cdot \bar{x} + \bar{x}^2\right)$$

$$= \frac{1}{n}\sum_{i=1}^{n} x_i^2 - 2\cdot\bar{x}\cdot\underbrace{\frac{1}{n}\sum_{i=1}^{n} x_i}_{\bar{x}} + \frac{1}{n}\sum_{i=1}^{n}\bar{x}^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - 2\cdot\bar{x}^2 + \frac{1}{\cancel{n}}\cdot\cancel{n}\cdot\bar{x}^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \bar{x}^2$$

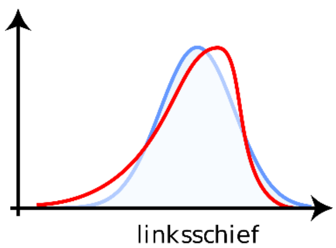
$$= m_2 - m_1^2 = 62,\overline{33} - \left(7,\overline{66}\right)^2 = 62,\overline{33} - 58,\overline{77} = 3,\overline{55}$$

$$s = +\sqrt{s^2} = \sqrt{3,\overline{55}} \approx 1,8856$$

5) $$z_3 = m_3 - 3\cdot m_1 \cdot m_2 + 2\cdot (m_1)^3 = 527,\overline{66} - 3\cdot 7,\overline{66}\cdot 62,\overline{33} + 2\cdot 7,\overline{66}^2 \approx -4,7494$$

$$g_1 = \frac{z_3}{\left(\sqrt{z_2}\right)^3} \approx \frac{4,7494}{1,8856^3} = -0,7084 < 0$$ rechtssteil/linksschief

Durch die dritte Potenz wird das Vorzeichen der Differenz zum Durchschnitt beibehalten und die Abweichung umso stärker gewichtet, je größer diese ist.
Bei Symmetrie heben sich damit die Abweichungen der unterdurchschnittlichen Beobachtungswerte mit denjenigen der überdurchschnittlichen Beobachtungswerte auf.
Hier liegen die unterdurchschnittlichen Beobachtungswerte liegen weiter weg als die überdurchschnittlichen Beobachtungswerte,
so dass die negativen Abweichungen die positiven überkompensieren.
Somit ist die rechte Seite steil und die linke schief.



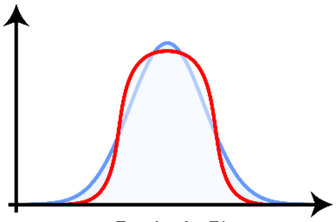
6)

$$z_4 = m_4 - 4\cdot m_1 \cdot m_3 + 6\cdot (m_1)^2 \cdot m_2 - 3\cdot (m_1)^4$$

$$= 4.582,\overline{33} - 4\cdot 7,\overline{66}\cdot 527,\overline{66} + 6\cdot 7,\overline{66}^2 \cdot 62,\overline{33} - 3\cdot 7,\overline{66}^4 \approx 18,9630$$

$$g_2 = \frac{z_4}{z_2^2} \approx \frac{18,9630}{3,55^2} = 1,5 < 3$$ flachgipflig, platykurtisch

$$z = g_2 - 3 = 1,5 - 3 = -1,5 < 0$$ Exzess



Durch die vierte Potenz wird das Vorzeichen der Differenz zum Durchschnitt neutralisiert und die Abweichung umso stärker gewichtet, je größer diese ist.
Hier liegen die kleinen Abweichungen relativ zum „Normalfall“ weiter weg als die großen, so dass die Kurtosis kleiner als der Referenzwert 3 bzw. der Exzess negativ ist.
Somit ist der Gipfel flacher.

7) Die (linerare) Transformation $\frac{x_i - \bar{x}}{s}$ bewirkt eine **Standardisierung** des Datensatzes.

Das Abziehen des arithmetischen Mittels bewirkt eine Verschiebung,
so dass der resultierende Durchschnitt gleich 0 ist (**Zentrierung**).

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i = \frac{1}{n}\sum_{i=1}^{n}\frac{x_i - \bar{x}}{s} = \frac{\underbrace{\frac{1}{n}\sum_{i=1}^{n} x_i}_{\bar{x}} - \frac{1}{n}\sum_{i=1}^{n}\bar{x}}{s} = \frac{\bar{x} - \frac{1}{\cancel{n}}\cdot\cancel{n}\cdot\bar{x}}{s} = \frac{\bar{x}-\bar{x}}{s} = 0$$

Das Dividieren durch die Standardabweichung bewirkt
ein Dehnen (s > 1) bzw. ein Stauchen (s < 1),
so dass die resultierende Varianz bzw. Standardabweichung gleich 1 ist (**Normierung**).

$$s_y^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2 = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i - \bar{x}}{s_x} - 0\right)^2 = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2}{s_x^2} = \frac{\cancel{s_x^2}}{\cancel{s_x^2}} = 1$$

Nach der Standardisierung sind alle Datensätze hinsichtlich arithmetischem Mittel und Varianz identisch und darüber hinaus auch noch dimensionslos. Somit unterscheiden sich nur noch hinsichtlich der höheren Momente. Der Grad der Schiefe bzw. Wölbung kann dann für die einzelnen Datensätze verglichen werden.

8)

$$\bar{y} = \frac{y_1 + y_2}{2} = 2{,}5 \quad\Leftrightarrow\quad y_1 = 5 - y_2 = 5 - (2{,}5 \pm 1{,}5) = 2{,}5 \mp 1{,}5 = \begin{cases}1\\4\end{cases}$$

$$s_y^{\,2} = \frac{(y_1-\bar{y})^2 + (y_2-\bar{y})^2}{2} = \frac{\left((5-y_2) - \frac{y_1+y_2}{2}\right)^2 + \left(y_2 - \frac{y_1+y_2}{2}\right)^2}{2}$$

$$= \frac{\left(\frac{10 - 2\cdot y_2 - \left((5\cancel{-y_2}) \cancel{+y_2}\right)}{2}\right)^2 + \left(\frac{2\cdot y_2 - \left((5\cancel{-y_2})\cancel{+y_2}\right)}{2}\right)^2}{2} = \frac{\left(\frac{\cancel{5}^{2{,}5} - \cancel{2}\cdot y_2}{\cancel{2}}\right)^2 + \left(\frac{\cancel{2}\cdot y_2 - \cancel{5}^{2{,}5}}{\cancel{2}}\right)^2}{2}$$

$$= \frac{\cancel{2}\cdot(y_2 - 2{,}5)^2}{\cancel{2}} = (y_2 - 2{,}5)^2 \overset{!}{=} 2{,}25 \qquad y_2 = \pm\sqrt{2{,}25} + 2{,}5 = 2{,}5 \pm 1{,}5 = \begin{cases}4\\1\end{cases}$$

a_j	$h(a_j)$
1	1
4	1
Σ	**2**

Generell gilt: Kennt man die ersten n Anfangsmomente,
so kann die Urliste (bis auf die Reihenfolge) rekonstruiert werden.

9) $s_x^2 = 3{,}\overline{55} > s_y^2 = 2{,}25$

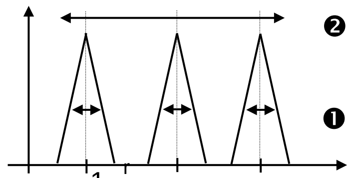
Die absolute Streuung ist bei den bisherigen Vorbereitungszeiten größer als bei den zukünftigen.
Nur Merkmale mit gleicher Einheit können hinsichtlich der absoluten Streuung verglichen werden.

$$v_x = \frac{s_x}{\bar{x}} \approx \frac{1{,}8856}{7{,}\overline{66}} = 0{,}2459 < v_y = \frac{1{,}5}{2{,}5} = 0{,}6$$

Die relative Streuung ist bei den zukünftigen Vorbereitungszeiten größer als bei den bisherigen.
Alle Merkmale können hinsichtlich der relativen Streuung verglichen werden,
da der Variationskoeffizient dimensionslos ist.

10)

Gesamtvarianz von r disjunkten Datensätzen
mit jeweils n_j Beobachtungswerten,
einem arithmetischen Mittel $\bar{x}_j$
und einer Varianz s_j^2



$$s_{ges}^2 = \frac{1}{n}\sum_{j=1}^{r} n_j \cdot s_j^2 + \frac{1}{n}\sum_{j=1}^{r} n_j \cdot \left(\bar{x}_j - \bar{x}_{ges}\right)^2 \quad \text{mit} \quad \bar{x}_{ges} = \frac{1}{n}\sum_{j=1}^{r} n_j \cdot \bar{x}_j$$

Streuung ❶ *innerhalb* ❷ *zwischen* *den Datensätzen*

$$\bar{x}_{ges} = \frac{n_x \cdot \bar{x} + n_y \cdot \bar{y}}{n_x + n_y} = \frac{6 \cdot 7{,}\overline{66} + 2 \cdot 2{,}5}{6+2} = 6{,}375$$

$$s_{ges}^2 = \frac{n_x \cdot s_x^2 + n_y \cdot s_y^2}{n_x + n_y} + \frac{n_x \cdot \left(\bar{x} - \bar{x}_{ges}\right)^2 + n_y \cdot \left(\bar{y} - \bar{x}_{ges}\right)^2}{n_x + n_y}$$

$$= \frac{6 \cdot 3{,}\overline{55} + 2 \cdot 2{,}25}{6+2} + \frac{6 \cdot \left(7{,}\overline{66} - 6{,}375\right)^2 + 2 \cdot \left(2{,}5 - 6{,}375\right)^2}{6+2} \approx 3{,}2292 + 5{,}005 = 8{,}2344$$

$$s_{ges} = +\sqrt{s_{ges}^2} \approx \sqrt{8{,}2344} \approx 2{,}86956443$$

Teil B

1) Merkmalsträger: Einwohner in Land A bzw. B
   Merkmale $X_{A\text{ bzw. }B}$: „Vermögen in Land A bzw. B in Mio. Geldeinheiten"

2) Land A:

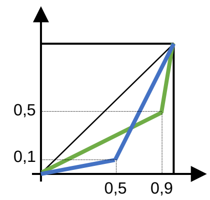
		4)							5)	
v_L	u_L	f_j	h_j	$m_j \cdot h_j/V$	$m_j \cdot h_j$	m_j	g_{j-1}	g_j	$u_j \cdot v_{j+1}$	$v_j \cdot u_{j+1}$
0,10	0,50	0,50	5	0,10	12	2,4	0	4,8	0,50	0,10
1,00	1,00	0,50	5	0,90	108	21,6	4,8	38,4	1,00	1,00
		1,00	10	1,00	120				1,50	1,10

$$G_A = 1{,}5 - 1{,}1 = 0{,}4$$

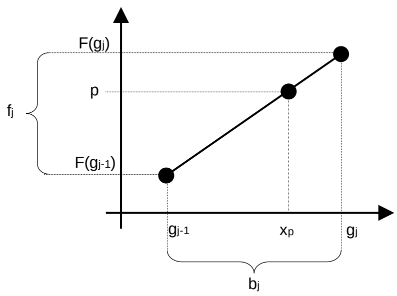
Land B:

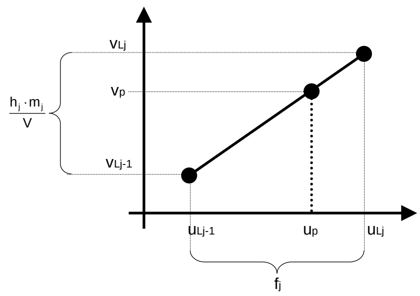
		4)							5)	
v_L	u_L	f_j	h_j	$m_j \cdot h_j/V$	$m_j \cdot h_j$	m_j	g_{j-1}	g_j	$u_j \cdot v_{j+1}$	$v_j \cdot u_{j+1}$
0,50	0,90	0,90	18	0,50	90	5	0	10	0,90	0,50
1,00	1,00	0,10	2	0,50	90	45	10	80	1,00	1,00
		1,00	20		180	50			1,90	1,50

$$G_B = 1{,}9 - 1{,}5 = 0{,}4$$



3)





$$p = F(g_{j-1}) + f_j \cdot \frac{x_p - g_{j-1}}{b_j}$$

$$x_p = g_{j-1} + \frac{p - F(g_{j-1})}{f_j} \cdot b_j$$

$$v_p = v_{L_{j-1}} + \frac{u_p - u_{L_{j-1}}}{f_j} \cdot \frac{h_j \cdot m_j}{V}$$

$$u_p = u_{L_{j-1}} + \frac{v_p - v_{L_{j-1}}}{\frac{h_j \cdot m_j}{V}} \cdot f_j$$

a) Land A

$$v_{0,25} = 0{,}00 + \frac{0{,}25 - 0{,}00}{0{,}50} \cdot 0{,}10 = 0{,}05$$

$$v_{0,75} = 0{,}10 + \frac{0{,}75 - 0{,}50}{0{,}50} \cdot 0{,}9 = 0{,}55$$

$$(v_{0,75} - v_{0,25}) \cdot V = (0{,}55 - 0{,}05) \cdot 120 = 60$$

Land B

$$v_{0,25} = 0{,}00 + \frac{0{,}25 - 0{,}00}{0{,}90} \cdot 0{,}50 = 0{,}13\overline{88}$$

$$v_{0,75} = 0{,}00 + \frac{0{,}75 - 0{,}00}{0{,}90} \cdot 0{,}50 = 0{,}41\overline{66}$$

$$(v_{0,75} - v_{0,25}) \cdot V = (0{,}41\overline{66} - 0{,}13\overline{88}) \cdot 180 = 50$$

b) Land A

$$u_{0,25} = 0{,}50 + \frac{0{,}25 - 0{,}10}{0{,}90} \cdot 0{,}50 = 0{,}58\overline{33}$$

$$u_{0,75} = 0{,}50 + \frac{0{,}75 - 0{,}10}{0{,}90} \cdot 0{,}50 = 0{,}86\overline{11}$$

$$\left(u_{0,75} - u_{0,25}\right) \cdot n = \left(0{,}86\overline{11} - 0{,}58\overline{33}\right) \cdot 10 = 2{,}\overline{77}$$

Land B

$$u_{0,25} = 0{,}00 + \frac{0{,}25 - 0{,}00}{0{,}50} \cdot 0{,}90 = 0{,}45$$

$$u_{0,75} = 0{,}90 + \frac{0{,}75 - 0{,}50}{0{,}50} \cdot 0{,}10 = 0{,}95$$

$$\left(u_{0,75} - u_{0,25}\right) \cdot n = \left(0{,}95 - 0{,}45\right) \cdot 20 = 10$$

4) s. oben

5) a)

$$G = 2 \cdot \left(\frac{1}{2} - \left(\underbrace{\frac{u_1 \cdot v_1}{2}}_{F1} + \underbrace{\left(u_2 - u_1\right) \cdot v_1}_{F2} + \underbrace{\frac{\left(u_2 - u_1\right) \cdot \left(v_2 - v_1\right)}{2}}_{F3} \right) \right)$$

$$= \cancel{2} \cdot \left(\frac{1 - \left(u_1 \cdot v_1 + 2 \cdot \left(u_2 \cdot v_1 - u_1 \cdot v_1\right) + \left(u_2 \cdot v_2 - u_2 \cdot v_1 - u_1 \cdot v_2 + u_1 \cdot v_1\right)\right)}{\cancel{2}} \right)$$

$$= 1 \cancel{- u_1 \cdot v_1} - \cancel{2} \cdot u_2 \cdot v_1 \cancel{+ 2 \cdot u_1 \cdot v_1} - u_2 \cdot v_2 \cancel{+ u_2 \cdot v_1} + u_1 \cdot v_2 \cancel{- u_1 \cdot v_1}$$

$$= \left(u_1 \cdot v_2 + \underbrace{1}_{u_2 = v_3 = 1} \right) - \left(u_2 \cdot v_1 + \underbrace{u_2 \quad \cdot v_2}_{u_2 = u_3 = 1} \right) = \left(u_1 \cdot v_2 + u_2 \cdot v_3\right) - \left(v_1 \cdot u_2 + v_2 \cdot u_3\right)$$

$$= \sum_{j=1}^{k=2} u_j \cdot v_{j+1} - \sum_{j=1}^{k=2} v_j \cdot u_{j+1}$$

b) s. oben

Die Stärke der Konzentration ist in beiden Ländern gleich.
Allerdings wirkt sich die Konzentration unterschiedlich aus.
In Land A gibt es sehr viele, die sehr wenig haben (bzw. weniger als in Land B).
In Land B gibt es sehr wenige, die sehr viel haben (bzw. mehr als in Land A).

6)

$$\bar{x}_A = \frac{1}{n} \sum_{j=1}^{k} m_j \cdot h_j = \frac{120}{10} = 12$$

$$s_A^2 = \sum_{j=1}^{k} \left(m_j - \bar{x}_A\right)^2 \cdot f_j = \left(2{,}4 - 12\right)^2 \cdot 0{,}5 + \left(21{,}6 - 12\right)^2 \cdot 0{,}5 = 92{,}16$$

$$s_A = +\sqrt{s_A^2} = +\sqrt{92{,}16} = 9{,}6$$

$$v_A = \frac{9{,}6}{12} = 0{,}8$$

$$\bar{x}_B = \frac{1}{n} \sum_{j=1}^{k} m_j \cdot h_j = \frac{180}{20} = 9$$

$$s_B^2 = \sum_{j=1}^{k} \left(m_j - \bar{x}_A\right)^2 \cdot f_j = \left(5 - 9\right)^2 \cdot 0{,}9 + \left(45 - 9\right)^2 \cdot 0{,}1 = 144$$

$$s_B = +\sqrt{s_B^2} = +\sqrt{144} = 12$$

$$v_B = \frac{12}{9} = 1{,}\overline{33}$$

7) $$H=\sum_{i=1}^{n}p_i^2=\sum_{i=1}^{n}\left(\frac{x_i}{\sum_{i=1}^{n}x_i}\right)^2=\sum_{i=1}^{n}\left(\frac{x_i}{n\cdot\frac{1}{n}\sum_{i=1}^{n}x_i}\right)^2=\sum_{i=1}^{n}\left(\frac{x_i}{n\cdot\bar{x}}\right)^2=\frac{1}{(n\cdot\bar{x})^2}\sum_{i=1}^{n}x_i^2$$

$$s^2=\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2=\frac{1}{n}\sum_{i=1}^{n}x_i^2-\bar{x}^2$$

$$\sum_{i=1}^{n}x_i^2=n\cdot\left(s^2+\bar{x}^2\right)$$

$$H=\frac{1}{(n\cdot\bar{x})^2}\sum_{i=1}^{n}x_i^2=\frac{1}{n^2\cdot\bar{x}^2}\cdot n\cdot\left(s^2+\bar{x}^2\right)=\frac{1}{n^{\cancel{2}}\cdot\bar{x}^2}\cdot\cancel{n}\cdot\left(s^2+\bar{x}^2\right)=\frac{s^2+\bar{x}^2}{n\cdot\bar{x}^2}=\frac{\frac{s^2}{\bar{x}^2}+1}{n}=\frac{v^2+1}{n}$$

$$H_A=\frac{v_A^2+1}{n}=\frac{0{,}8^2+1}{10}=0{,}164 \qquad H_B=\frac{v_B^2+1}{n}=\frac{1{,}\overline{33}^2+1}{20}=0{,}13\overline{88}$$

8) Gesamteffekt: $H_A=0{,}164>0{,}13\overline{88}=H_B$
Land A stärker konzentriert

Anzahleffekt: $n_A=10<20=n_B$
Land A stärker konzentriert

Merkmalseffekt: $H'_A=\frac{H_A}{n_B}=\frac{0{,}164}{20}=0{,}0082<0{,}01388=\frac{0{,}13\overline{88}}{10}=\frac{H_B}{n_A}=H'_B$
Land B stärker konzentriert

Anzahl- und Merkmalseffekt wirken entgegengesetzt,
wobei der Anzahleffekt den Merkmalseffekt überkompensiert.