Deskriptive Statistik - Aufgabe 2

Bundesland	Männer	Frauen
Baden-Württemberg	7,5	7,5
Bayern	6,8	7,5
Berlin-West	14,4	12
Berlin-Ost	11,8	14,2
Brandenburg	10,2	20,8
Bremen	14,6	12,6
Hamburg	11,2	8,1
Hessen	8,2	8,2
Mecklenburg-Vorpommern	12,3	22,1
Niedersachsen	10,3	11,2
Nordrhein-Westfalen	10,7	10,6
Rheinland-Pfalz	8,1	8,8
Saarland	12,7	11,2
Sachsen	9,6	22,3
Sachsen-Anhalt	12,3	23,3
Schleswig-Holstein	9,3	8,7
Thüringen	10,8	22,6

Tabelle: Arbeitslosenquote 1994 (in %)

a) Berechnen Sie für die Geschlechter bezüglich der Arbeitslosenquote die folgenden Zentral- und Streuungsmaße

- Median
- arithmetisches Mittel
- Variationsweite
- durchschnittliche Abweichung
- Varianz
- Standardabweichung

Die Formeln sind der Formelsammlung zu entnehmen.

b) Vergleichen Sie die jeweiligen Zentral- und Streuungsmaße miteinander und heben Sie Stärken und Schwächen hervor.

Lösung – Kurzfassung:

a)

	Männer	Frauen
Median	10,7- Nordrhein-Westfalen	11,2- Niedersachsen
arithmetisches Mittel	10,6	13,6
Variationsweite	7,8	15,8
durchschnittliche Abweichung	1,77	5,12
Standardabweichung	2,19	5,84
Varianz	4,79	34,1

b)

Vergleich der Zentralmaße:

Beim arithmetischen Mittel wird im Gegensatz zum Median der gesamte Stichprobenumfang berücksichtigt, weshalb es ein zuverlässiges Zentralmaß ist. Allerdings ist das arithmetische Mittel ausreißerempfindlich. Der Median ist das nicht, weil er die Stichprobe lediglich in zwei gleich große Hälften teilt.

Vergleich der Streuungsmaße:

Die Variationsweite ist sehr ausreißerempfindlich. Die Varianz ist dies weniger, am wenigsten ausreißerempfindlich ist die Standardabweichung. In Varianz und Standardabweichung fließen im Gegensatz zur Variationsweite alle Werte in die Berechnung ein.

Lösung – Erläuterung:

Männer

1) **Median**: Wenn man Ausprägungen einer Verteilung der Größe nach ordnet, ist der Median die Ausprägung, die man in der Mitte der geordneten Reihe findet. Der Median ist also der „mittlere Wert“ einer Verteilung. Er unterteilt die Reihe in zwei Hälften: die eine Hälfte der Ausprägungen ist kleiner als der Median (oder höchstens gleich groß), die andere Hälfte der Ausprägungen ist größer als der Median (oder höchstens gleich groß).
Falls die Anzahl der Fälle (n) eine gerade Zahl ist, ist das arithmetische Mittel der beiden „mittleren Werte“ zu bilden.
Um den Median zu berechnen, muss die Liste für Männer und Frauen getrennt nach der Arbeitslosenquote geordnet werden.

Geordnete Urliste:

	Bundesland	Männer
1	Bayern	6,8
2	Baden-Württemberg	7,5
3	Rheinland-Pfalz	8,1
4	Hessen	8,2
5	Schleswig-Holstein	9,3
6	Sachsen	9,6
7	Brandenburg	10,2
8	Niedersachsen	10,3
9	Nordrhein-Westfalen	10,7
10	Thüringen	10,8
11	Hamburg	11,2
12	Berlin-Ost	11,8
13	Sachsen-Anhalt	12,3
14	Mecklenburg-Vorpommern	12,3
15	Saarland	12,7
16	Berlin-West	14,4
17	Bremen	14,6

$$Median: \tilde{x} = \frac{n+1}{2} \Rightarrow \tilde{x} = \frac{17+1}{2} = \frac{18}{2} = 9.Wert \quad = 10{,}7 - Nordrhein - Westfalen$$

2) Arithmetisches Mittel: Das arithmetische Mittel ist der durchschnittliche Wert einer Verteilung. Berechnet wird es indem alle Ausprägungen aufaddiert werden und die Summe durch die Anzahl der Fälle geteilt wird.

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

$x_i =$ jeweilige Merkmalsausprägung von x_1 (6,8) bis x_{17} (14,6)

n = Anzahl der Fälle = 17, da die Gesamtzahl der Bundesländer 17 beträgt

$$\bar{x} = \frac{6{,}8+7{,}5+8{,}1+8{,}2+9{,}3+9{,}6+10{,}2+10{,}3+10{,}7+10{,}8+11{,}2+11{,}8+12{,}3+12{,}3+12{,}7+14{,}4+14{,}6}{17} = \frac{180{,}8}{17} = 10{,}64$$

3) Variationsweite V: Die Variationsweite erfasst den Abstand zwischen dem minimalen und dem maximalen Wert einer Verteilung. Man erhält den Wert für die Variationsweite, indem man den kleinsten Wert von dem größten Wert subtrahiert.

Variationsweite = größter Wert – kleinster Wert = Vmax - Vmin = 14,6 – 6,8 = 7,8

4) Durchschnittliche Abweichung e: Hierfür wird der Betrag (der **vorzeichenlose** Wert) der einzelnen Abstände der Messwerte vom Mittelwert addiert. Die Summe der Beträge wird dann durch die Anzahl der Fälle dividiert und ergibt die durchschnittliche Abweichung.

$$e = \frac{\sum_{i=1}^{k} f_i \left| x_i - \bar{x} \right|}{n}$$

$$= \frac{\begin{array}{c}|6{,}8-10{,}64|+|7{,}5-10{,}64|+|8{,}1-10{,}64|+|8{,}2-10{,}64|+|9{,}3-10{,}64|+|9{,}6-10{,}64|+|10{,}2-10{,}64|+|10{,}3-10{,}64|+|10{,}7-10{,}64| \\ +|10{,}8-10{,}64|+|11{,}2-10{,}64|+|11{,}8-10{,}64|+|12{,}3-10{,}64|+|12{,}3-10{,}64|+|12{,}7-10{,}64|+|14{,}4-10{,}64|+|14{,}6-10{,}64|\end{array}}{17}$$

$$= \frac{3{,}84+3{,}14+2{,}54+2{,}44+1{,}34+1{,}04+0{,}44+0{,}34+0{,}06+0{,}16+0{,}56+1{,}16+1{,}66+1{,}66+2{,}06+3{,}76+3{,}96}{17}$$

$$= \frac{30{,}16}{17} = 1{,}77$$

5) Standardabweichung:

$$s = \sqrt{\frac{\sum_{i=1}^{k} f_i \left(x_1 - \bar{x} \right)^2}{n}}$$

$$= \sqrt{\frac{\begin{array}{c}(6{,}8-10{,}64)^2+(7{,}5-10{,}64)^2+(8{,}1-10{,}64)^2+(8{,}2-10{,}64)^2+(9{,}3-10{,}64)^2+(9{,}6-10{,}64)^2+(10{,}2-10{,}64)^2+(10{,}3-10{,}64)^2+(10{,}7-10{,}64)^2+ \\ (10{,}8-10{,}64)^2+(11{,}2-10{,}64)^2+(11{,}8-10{,}64)^2+(12{,}3-10{,}6)^2+(12{,}3-10{,}6)^2+(12{,}7-10{,}64)^2+(14{,}4-10{,}64)^2+(14{,}6-10{,}64)^2\end{array}}{17}}$$

$$= \sqrt{\frac{14{,}71+9{,}83+6{,}43+5{,}93+1{,}78+1{,}07+0{,}19+0{,}11+0{,}00+0{,}03+0{,}32+1{,}36+2{,}77+2{,}77+4{,}26+14{,}17+15{,}72}{17}}$$

$$= \sqrt{\frac{81{,}45}{17}} = 2{,}19$$

6) Varianz: $s^2 = \frac{\sum_{i=1}^{k} f_i \left(x_1 - \bar{x} \right)^2}{n}$

Die Varianz (s^2) stellt das Quadrat der Standardabweichung (s) dar. Da die Standardabweichung bereits berechnet wurde lässt sich die Varianz durch eine Quadrierung der Standardabweichung berechnen.

$s^2 = 2{,}19^2 = 4{,}796$

Alternativ ließe sich natürlich auch zuerst die Varianz berechnen und im Anschluss daran die Standardabweichung als Wurzel der Varianz ziehen. Die Ergebnisse wären die gleichen, abgesehen von Unterschieden durch unterschiedliche Rundung.

Alternative zeitsparende Berechnungsmöglichkeit für Streuungsmaße

Eine alternative zeitsparende Berechnungsmöglichkeit der durchschnittlichen Abweichung, der Standardabweichung und der Varianz stellt die Erstellung der folgenden Tabelle dar. Bei der Berechnung werden die gleichen Formeln wie oben verwendet. Die ersten beiden Spalten der Tabelle (hier „Bundesland“ und „Männer“) werden aus der Tabelle der Aufgabenstellung übernommen. Die Tabelle wird dann um die Spalte $\left|x_i - \bar{x}\right|$ (Abweichung der gemessenen Merkmalsausprägungen vom arithmetischen Mittel) erweitert. Am unteren Ende der Spalte summiert man die einzelnen Werte und kann die Summe in die Formel der durchschnittlichen Abweichung einsetzen. Die Summe wird in dieser Formel lediglich noch durch n geteilt und man erhält die durchschnittliche Abweichung.

Um die Varianz oder die Standardabweichung zu berechnen erweitert man die Tabelle zusätzlich um die Spalte $\left(x_1 - \bar{x}\right)^2$. Auch hier summiert man die einzelnen Werte und setzt die Summe in die Formel der Varianz bzw. der Standardabweichung ein.
Man sieht, dass die Ergebnisse der beiden Berechnungsmöglichkeiten identisch sind.

Bundesland	**Männer**	$\left\|x_i - \bar{x}\right\|$	$\left(x_i - \bar{x}\right)^2$
Bayern	6,8	3,84	14,71
Baden-Württemberg	7,5	3,14	9,83
Rheinland-Pfalz	8,1	2,54	6,43
Hessen	8,2	2,44	5,93
Schleswig-Holstein	9,3	1,34	1,78
Sachsen	9,6	1,04	1,07
Brandenburg	10,2	0,44	0,19
Niedersachsen	10,3	0,34	0,11
Nordrhein-Westfalen	10,7	0,06	0,00
Thüringen	10,8	0,16	0,03
Hamburg	11,2	0,56	0,32
Berlin-Ost	11,8	1,16	1,36
Sachsen-Anhalt	12,3	1,66	2,77
Mecklenburg-Vorpommern	12,3	1,66	2,77
Saarland	12,7	2,06	4,26
Berlin-West	14,4	3,76	14,17
Bremen	14,6	3,96	15,72
		$\sum$ 30,16	$\sum$ 81,45

4) Durchschnittliche Abweichung e: $e = \frac{\sum_{i=1}^{k} f_i \left|x_i - \bar{x}\right|}{n} = \frac{30{,}16}{17} = 1{,}77$

(die Zahl 30,16 kann in der Tabelle am unteren Ende der dritten Spalte abgelesen werden)

5) Standardabweichung s: $s = \sqrt{\frac{\sum_{i=1}^{k} f_i \left(x_1 - \bar{x}\right)^2}{n}} = \sqrt{\frac{81{,}45}{17}} = 2{,}19$

(die Zahl 81,45 kann in der Tabelle am unteren Ende der vierten Spalte abgelesen werden)

6) Varianz: $s^2 = \frac{\sum_{i=1}^{k} f_i \left(x_1 - \bar{x}\right)^2}{n} = \frac{81{,}45}{17} = 4{,}791$

Alternative Berechnung der Varianz: Quadrierung der Standardabweichung: $s^2 = 2{,}19^2 = 4{,}796$

(Geringe Unterschiede entstehen durch unterschiedliche Rundung!)

Frauen

1) Median

	Bundesland	Frauen
1	Baden-Württemberg	7,5
2	Bayern	7,5
3	Hamburg	8,1
4	Hessen	8,2
5	Schleswig-Holstein	8,7
6	Rheinland-Pfalz	8,8
7	Nordrhein-Westfalen	10,6
8	Saarland	11,2
9	Niedersachsen	11,2
10	Berlin-West	12
11	Bremen	12,6
12	Berlin-Ost	14,2
13	Brandenburg	20,8
14	Mecklenburg-Vorpommern	22,1
15	Sachsen	22,3
16	Thüringen	22,6
17	Sachsen-Anhalt	23,3

$$Median: \tilde{x} = \frac{N+1}{2} \Rightarrow \tilde{x} = \frac{17+1}{2} = \frac{18}{2} = 9.\,Wert = 11{,}2 - Niedersachsen$$

2) Arithmetisches Mittel $\bar{x}$: $\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$

$$\bar{x} = \frac{7{,}5+7{,}5+8{,}1+8{,}2+8{,}7+8{,}8+10{,}6+11{,}2+11{,}2+12+12{,}6+14{,}2+20{,}8+22{,}1+22{,}3+22{,}6+23{,}3}{17} = \frac{231{,}7}{17} = 13{,}6$$

3) Variationsweite V: 23,3 – 7,5 = 15,8

4) Durchschnittliche Abweichung e:

$$e = \frac{\sum_{i=1}^{k} f_i \left| x_i - \bar{x} \right|}{n}$$

$$= \frac{\begin{array}{l}|7{,}5-13{,}6|+|7{,}5-13{,}6|+|8{,}1-13{,}6|+|8{,}2-13{,}6|+|8{,}7-13{,}6|+|8{,}8-13{,}6|+|10{,}6-13{,}6|+|11{,}2-13{,}6|+|11{,}2-13{,}6|+\\ |12-13{,}6|+|12{,}6-13{,}6|+|14{,}2-13{,}6|+|20{,}8-13{,}6|+|22{,}1-13{,}6|+|22{,}3-13{,}6|+|22{,}6-13{,}6|+|23{,}3-13{,}6|\end{array}}{17}$$

$$= \frac{6{,}13+6{,}13+5{,}53+5{,}43+4{,}93+4{,}83+3{,}03+2{,}43+2{,}43+1{,}63+1{,}03+0{,}57+7{,}17+8{,}47+8{,}67+8{,}97+9{,}67}{17}$$

$$= \frac{87{,}05}{17} = 5{,}12$$

5) Standardabweichung s: $s=\sqrt{\frac{\sum_{i=1}^{k} f_i\left(x_1-\bar{x}\right)^2}{n}}$

$$=\sqrt{\frac{(7{,}5-13{,}6)^2+(7{,}5-13{,}6)^2+(8{,}1-13{,}6)^2+(8{,}2-13{,}6)^2+(8{,}7-13{,}6)^2+(8{,}8-13{,}6)^2+(10{,}6-13{,}6)^2+(11{,}2-13{,}6)^2+(11{,}2-13{,}6)^2+(12-13{,}6)^2+(12{,}6-13{,}6)^2+(14{,}2-13{,}6)^2+(20{,}8-13{,}6)^2+(22{,}1-13{,}6)^2+(22{,}3-13{,}6)^2+(22{,}6-13{,}6)^2+(23{,}3-13{,}6)^2}{17}}$$

$$s=\sqrt{\frac{580{,}18}{17}}=5{,}84$$

6) Varianz: $s^2=\frac{\sum_{i=1}^{k} f_i\left(x_1-\bar{x}\right)^2}{n}$

s² = 5,84² = 34,11

Alternative zeitsparende Berechnungsmöglichkeit

Bundesland	**Frauen**	$\lvert x_i-\bar{x}\rvert$	$\left(x_i-\bar{x}\right)^2$
Baden-Württemberg	7,5	6,13	37,57
Bayern	7,5	6,13	37,57
Hamburg	8,1	5,53	30,57
Hessen	8,2	5,43	29,48
Schleswig-Holstein	8,7	4,93	24,30
Rheinland-Pfalz	8,8	4,83	23,32
Nordrhein-Westfalen	10,6	3,03	9,18
Saarland	11,2	2,43	5,90
Niedersachsen	11,2	2,43	5,90
Berlin-West	12	1,63	2,65
Bremen	12,6	1,03	1,06
Berlin-Ost	14,2	0,57	0,33
Brandenburg	20,8	7,17	51,42
Mecklenburg-Vorpommern	22,1	8,47	71,75
Sachsen	22,3	8,67	75,18
Thüringen	22,6	8,97	80,47
Sachsen-Anhalt	23,3	9,67	93,52
		$\sum$ 87,05	$\sum$ 580,18

4) Durchschnittliche Abweichung e: $e=\frac{\sum_{i=1}^{k} f_i\lvert x_i-\bar{x}\rvert}{n}=\frac{87{,}05}{17}=5{,}12$

5) Standardabweichung s: $s=\sqrt{\frac{\sum_{i=1}^{k} f_i\left(x_1-\bar{x}\right)^2}{n}}=\sqrt{\frac{580{,}18}{17}}=5{,}84$

6) Varianz: $s^2=\frac{\sum_{i=1}^{k} f_i\left(x_1-\bar{x}\right)^2}{n}=\frac{580{,}18}{17}=34{,}13$

Alternative Berechnung der Varianz: Quadrierung der Standardabweichung: s²= 5,84²= 34,11

(geringe Unterschiede entstehen durch unterschiedliche Rundung!)