Deskriptive Statistik - Aufgabe 3

Die Übernachtungszahlen in der Fremdenverkehrsgemeinde "Bachstadt" für die Monate Dezember und März zeigen auf den ersten Blick schon deutliche Unterschiede in den einzelnen Ortschaften. Wie sieht ein entsprechender Vergleich der beiden Monate mit Hilfe von geeigneten Zentral- und Streuungsmaßen aus?

Name des Ortes	**Übernachtungen März**	**Übernachtungen Dezember**
Umsiedelbach	150	500
Kirchbach	100	100
Kleinplantschbach	50	50
Dortmusikantenbach	100	50
Kuckmündebach	100	50
Unterkarnickelbach	100	50
Oberkarnickelbach	100	50
Forellenbach	150	50
Mündebach	50	50
Großplantschbach	100	50

<u>Lösung – Kurzfassung:</u>

	März	Dezember
arithmetisches Mittel	100	100
durchschnittliche Abweichung	20	80
Standardabweichung	31,62	134,16
Variationsweite	100	450

Hinweis: Die Daten beschreiben keine Stichprobe, sondern eine Grundgesamtheit, nämlich die Bettenbelegung in der gesamten Gemeinde oder allen Ortsteilen. Abweichend von der Formelsammlung sind die gängigen Streuungsmaße daher nur mit dem Nenner n zu rechnen.

Vergleich

Die beiden Datenreihen sind sehr gut vergleichbar, da sie den gleichen arithmetischen Mittelwert aufweisen. Im Dezember gab es also durchschnittlich genauso viele Übernachtungen pro Ortsteil wie im März.

Die wesentlichen Unterschiede müssen also über die Streuungsmaße beschrieben werden: Hier zeigt sich, dass die Streuung im Dezember wesentlich größer ist - sowohl die durchschnittliche Abweichung als auch die Standardabweichung der Dezember-Datenreihe sind im Vergleich rund vier mal so groß.

Hinweis: Die direkte Vergleichbarkeit der Standardabweichungen ist hier nur gegeben, weil beide Datenreihen den gleichen Mittelwert haben, um den sie schwanken. Bei Datenreihen mit verschiedenen Mittelwerten (Regelfall) empfiehlt sich immer eine Berechnung des Variationskoeffizienten.

Die Daten für Dezember sind also offenbar weniger einheitlich als im März. Das kann heißen, dass viele Werte weiter vom Mittelwert entfernt liegen oder einzelne Ausreißer für eine breitere Streuung sorgen. Der ‚Range', also die Variationsweite weist auf Letzteres hin. Der kleinste Wert V_{min} ist identisch, es ist also offenbar der Maximalwert, der hier zu Buche schlägt.

Deskriptive Interpretation:
Auf den ersten Blick bewegen sich die Übernachtungen im Wintermonat Dezember nicht überraschend fast durchgängig auf niedrigerem Niveau, mit Ausnahme des Ortsteils Umsiedelbach. Die größere Streuung erscheint also etwas irreführend.
Tilgt man diesen Ausreißer aus der Datenreihe, erhält man für den Dezember einen Durchschnittswert von 55,6 verglichen mit 94,4 im März; auch hier nahm Umsiedelbach einen Maximalwert ein und wurde außer Acht gelassen. Die Unterschiede werden so deutlicher.
Der Ortsteil Umsiedelbach hat also entweder den Wintersport in der Gemeinde ankurbeln können oder ist sehr erfolgreich dabei, das touristische ‚Winterloch' mit Tagungstouristen oder einem attraktiven Weihnachtsangebot (z.B. Betriebsfeiern mit Übernachtung) zu überbrücken. Diese Vermutungen könnten aber nur mit weiteren Daten, z.B. zu den anderen Wintermonaten untermauert werden.

Lösung – Erläuterung:

März

1) 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 (150) bis x_{17} (100)

n= Anzahl der Fälle = 10, da die Gesamtzahl der Orte 10 beträgt

$$\bar{x} = \frac{150+100+50+100+100+100+100+150+50+100}{10} = 100$$

2) Durchschnittliche Abweichung 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.

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

$\left(\text{Das Symbol } |...| \text{ bedeutet, dass das Vorzeichen des Ausdrucks nicht berücksichtigt wird.}\right)$

$$e = \frac{|150-100|+|100-100|+|50-100|+|100-100|+|100-100|+|100-100|+|100-100|+|150-100|+|50-100|+|100-100|}{10}$$

$$e = \frac{50+0+50+0+0+0+0+50+50+0}{10} = \frac{200}{10} = 20$$

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

$$s = \sqrt{\frac{(150-100)^2+(100-100)^2+(50-100)^2+(100-100)^2+(100-100)^2+(100-100)^2+(100-100)^2+(150-100)^2+(50-100)^2+(100-100)^2}{10}}$$

$$= \sqrt{\frac{2500+0+2500+0+0+0+0+2500+2500+0}{10}} = \sqrt{\frac{10000}{10}} = 31{,}62$$

4) Variationsweite

$V_{max} - V_{min}$ = 150-50 =100

5) Variationskoeffizient

$$V = \frac{s}{\bar{x}} \bullet 100\% = \frac{\text{Standardabweichung}}{\text{arithmetisches Mittel}} \bullet 100\%$$

$$V = \frac{31{,}62}{100} \bullet 100\% = 31{,}62\%$$

Sonderfall: Wie o.g. eigentlich hier nicht nötig. Der Variationskoeffizient nimmt hier nur deshalb den gleichen Wert an wie die Standardabweichung, weil das arithmetische Mittel den Wert 100 annimmt.

Alternative zeitsparende Berechnungsmöglichkeit

Eine alternative zeitsparende Berechnungsmöglichkeit der durchschnittlichen Abweichung und der Standardabweichung stellt die Erstellung der folgenden Tabelle dar. Bei der Berechnung werden die gleichen Formeln wie oben verwendet. Die ersten beiden Spalten der Tabelle (hier „Name des Ortes“ und „März“) 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 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 Standardabweichung ein.
Man sieht, dass die Ergebnisse der beiden Berechnungsmöglichkeiten identisch sind, es ist also egal, welche der beiden Berechnungsmöglichkeiten gewählt wird.

Name des Ortes	März	$\left\|x_i - \bar{x}\right\|$	$\left(x_i - \bar{x}\right)^2$
Umsiedelbach	150	50	2500
Kirchbach	100	0	0
Kleinplantschbach	50	50	2500
Dortmusikantenbach	100	0	0
Kuckmündebach	100	0	0
Unterkarnickelbach	100	0	0
Oberkarnickelbach	100	0	0
Forellenbach	150	50	2500
Mündebach	50	50	2500
Großplantschbach	100	0	0
		$\sum$ 200	$\sum$ 10.000

2) Durchschnittliche Abweichung e: $e = \frac{\sum_{i=1}^{k} f_i \left|x_i - \bar{x}\right|}{n}$ $\quad e = \frac{200}{10} = 20$

(die Zahl 200 kann in der Tabelle am unteren Ende der dritten Spalte abgelesen werden)

3) Standardabweichung s: $s = \sqrt{\frac{\sum_{i=1}^{k} f_i \left(x_1 - \overline{x}\right)^2}{n}} = \sqrt{\frac{10000}{10}} = 31{,}62$

(die Zahl 10000 kann in der Tabelle am unteren Ende der vierten Spalte abgelesen werden)

Dezember

1) Arithmetisches Mittel:

$$\overline{x} = \frac{\sum_{i=1}^{n} x_i}{n} \qquad \overline{x} = \frac{500+100+50+50+50+50+50+50+50+50}{10} = 100$$

2) Durchschnittliche Abweichung:

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

$$e = \frac{|500-100|+|100-100|+|50-100|+|50-100|+|50-100|+|50-100|+|50-100|+|50-100|+|50-100|+|50-100|}{10}$$

$$e = \frac{400+0+50+50+50+50+50+50+50+50}{10} = \frac{800}{10} = 80$$

3) Standardabweichung:

Zur Berechnung der Standardabweichung wird die Spalte $\left(x_i - \overline{x}\right)^2$ erstellt und unten die Summe abgelesen.

$$s = \sqrt{\frac{\sum_{i=1}^{k} f_i \left(x_1 - \overline{x}\right)^2}{n}} = \sqrt{\frac{180000}{10}} = 134{,}16$$

$$s = \sqrt{\frac{(500-100)^2+(100-100)^2+(50-100)^2+(50-100)^2+(50-100)^2+(50-100)^2+(50-100)^2+(50-100)^2+(50-100)^2+(50-100)^2}{10}}$$

$$= \sqrt{\frac{160.000+0+2500+2500+2500+2500+2500+2500+2500+2500}{10}} = \sqrt{\frac{180.000}{10}} = 134{,}16$$

4) Variationsweite

$V_{max} - V_{min}$ = 500-50 = 450

5) Variationskoeffizient

$$V = \frac{s}{\bar{x}} \bullet 100\% = \frac{\text{Standardabweichung}}{\text{arithmetisches Mittel}} \bullet 100\%$$

$$V = \frac{134{,}16}{100} \bullet 100\% = 134{,}16\%$$

Sonderfall: Wie o.g. eigentlich hier nicht nötig. Der Variationskoeffizient nimmt hier nur deshalb den gleichen Wert an wie die Standardabweichung, weil das arithmetische Mittel den Wert 100 annimmt.

Alternative zeitsparende Berechnungsmöglichkeit

Name des Ortes	Dezember	$\lvert x_i - \bar{x}\rvert$	$(x_i - \bar{x})^2$
Umsiedelbach	500	400	160.000
Kirchbach	100	0	0
Kleinplantschbach	50	50	2.500
Dortmusikantenbach	50	50	2.500
Kuckmündebach	50	50	2.500
Unterkarnickelbach	50	50	2.500
Oberkarnickelbach	50	50	2.500
Forellenbach	50	50	2.500
Mündebach	50	50	2.500
Großplantschbach	50	50	2.500
		$\sum$ 800	$\sum$ 180.000

2) Durchschnittliche Abweichung:

$$e = \frac{\sum_{i=1}^{k} f_i \lvert x_i - \bar{x}\rvert}{n} = \frac{800}{10} = 80$$

(die Zahl 800 kann in der Tabelle am unteren Ende der dritten Spalte abgelesen werden)

3) Standardabweichung:

Zur Berechnung der Standardabweichung wird die Spalte $(x_i - \bar{x})^2$ erstellt und unten die Summe abgelesen.

$$s = \sqrt{\frac{\sum_{i=1}^{k} f_i (x_1 - \bar{x})^2}{n}} = \sqrt{\frac{180.000}{10}} = 134{,}16$$

(die Zahl 180.00 kann in der Tabelle am unteren Ende der vierten Spalte abgelesen werden)