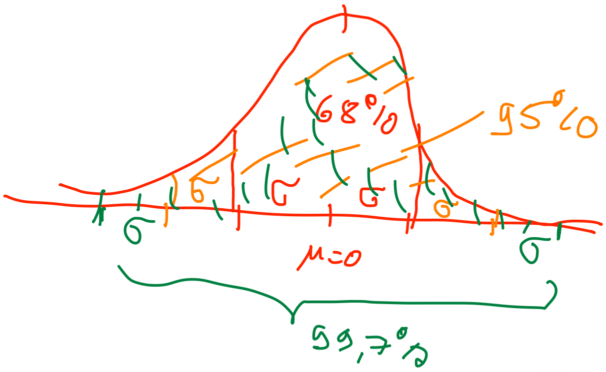
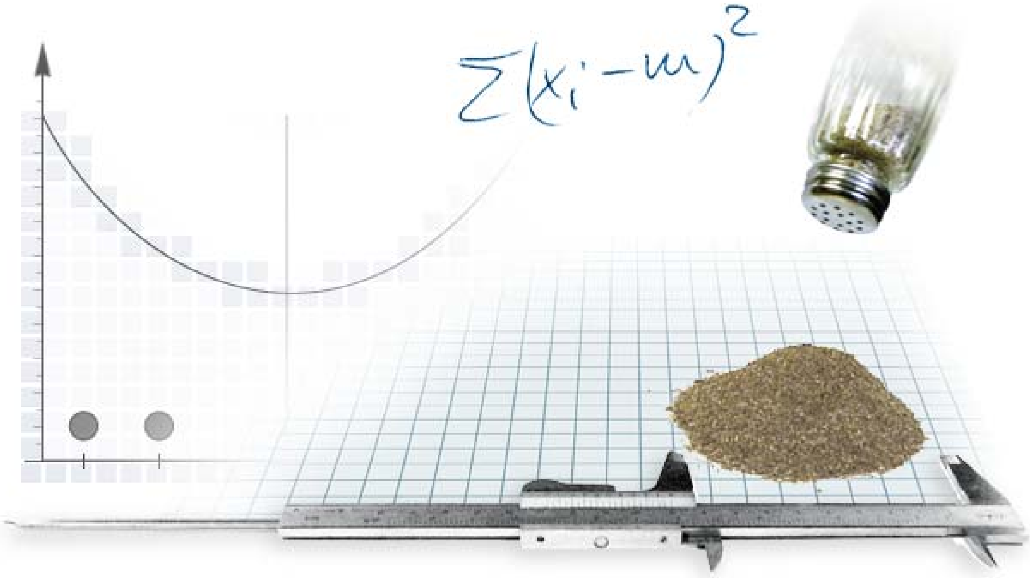


VAR - Varianz und Standardabweichung



Lernziele und Überblick

In dieser Lerneinheit werden Sie Varianz und Standardabweichung kennen lernen.

Lernziele

Nach dem Durcharbeiten dieser Lerneinheit sollen Sie

- den Begriff Streuung erklären können,
- Varianz und Standardabweichung berechnen können,
- wissen, wie sich Varianz und Standardabweichung bei linearen Transformationen verhalten,
- Varianz und Standardabweichung bei klassierten Daten berechnen können.

Gliederung der Lerneinheit

Zeitbedarf und Umfang

Für die Durcharbeitung dieser Lerneinheit benötigen Sie ca. 90 Minuten und für die Übungen mit der Statistiksoftware **R** ca. 180 Minuten.

1 Einleitung

Zur Charakterisierung von Daten reicht die Angabe der mittleren Lage (siehe „*Lerneinheit ARM ‚Arithmetisches Mittel'*", Abschnitt 1.1 Maßzahlen) nicht aus. Wir wissen nämlich nicht, wie stark die Einzelwerte von dem gewählten Lagemaß abweichen. Kennzahlen für diese Abweichung werden als Streuungsmaße bezeichnet. Das Problem erkennen wir sofort an den folgenden Beispielen.

Beispiel

Apfelpreis

Der durchschnittliche Preis für ein Kilogramm Äpfel in verschiedenen Geschäften beträgt 1,5 €.
Das bedeutet nicht unbedingt, dass die einzelnen Preise zwischen 1 und 2 € liegen. Es könnte auch sein, dass man in einigen Geschäften nur 0,70 €, in anderen dagegen 3 € pro Kilogramm bezahlt.

Beispiel

Gehälter

Wie Sie schon in der Übung ARM-03 „Gehälter von acht Angestellten" (siehe „*Lerneinheit ARM*", Abschnitt 2.2) gesehen haben, gibt das durchschnittliche Gehalt der Angestellten in Höhe von 2.425 € keine Informationen darüber, wie viel der einzelne Angestellte wirklich verdient. Es könnte sein, dass alle etwa dieses durchschnittliche Gehalt haben oder dass einige sehr viel mehr, andere dafür viel weniger verdienen. Genau das ist der Fall in diesem Beispiel: ein Angestellter verdient 5.500 € und der andere 1.500 €.

Also, die Lagemaße allein charakterisieren eine Häufigkeitsverteilung nicht ausreichend. Zur Beurteilung einer Häufigkeitsverteilung sollte man auch die Ausdehnung des Wertebereiches und die Verteilung der Häufigkeiten über diesen Bereich hinzuziehen. Darüber hinaus erweist sich gerade die „Streuung" als ein zentraler Begriff der Statistik, insbesondere der induktiven Statistik. Je stärker die Merkmalswerte um den Mittelwert konzentriert sind, desto kleiner ist die Streuung (wenn sich die Werte dicht beieinander befinden), desto repräsentativer ist der Mittelwert, und je schwächer die Merkmalswerte um den Mittelwert konzentriert sind (bei einer starken Variabilität der Daten), desto größer ist die Streuung (wenn sich die Werte nicht dicht beieinander befinden).

Die Streuungsmaße dienen

- der Charakterisierung der Variabilität eines Merkmals
  (der Ausbreitung einer Häufigkeitsverteilung und Homogenität eines statistischen Datensatzes, d. h. der Ähnlichkeit ihrer Einheiten bzw. der Distanz zwischen ihnen);
- der Treffsicherheit einer Prognose oder der Beurteilung der Güte einer Schätzung, z. B. aufgrund einer Stichprobe (in der induktiven Statistik).

Die Streuungen sind neben den Mittelwerten nicht nur wichtige Charakteristika einer eindimensionalen Häufigkeitsverteilung, sondern die Streuung kann für bestimmte Probleme sogar wichtiger als der Mittelwert sein.

2 Streuungsmaße

Sie fragen sich vielleicht, weshalb wir dem Streuungsbegriff ein eigenes Kapitel widmen. Wenn Sie den Streuungsbegriff erfasst haben, werden Sie das verstehen.

Wir erläutern Ihnen das Konzept der Streuung an einem praktischen Beispiel.

Beispiel

Qualität von Nähgarnen

Zur Qualitätsprüfung von Nähgarnen in der Textilindustrie wird die Reißfestigkeit ermittelt. Es liegen Stichproben von Umfang n = 100 zweier Garnsorten (Garn1 und Garn2) vor. Es wird gefordert, dass die Garne mindestens eine Belastung von 4 Einheiten aushalten, bevor sie reißen.

Bei Garn1 waren zwei Proben nicht ausreichend reißfest, bei Garn2 waren es 21.

Betrachteten wir die Mittelwerte der beiden Versuchsreihen (Garn1: 5,01, Garn2: 5,91), kommen wir ins Nachdenken. Wie kommt es, dass bei Garn2 mehr Ausfälle auftreten, obwohl es im Mittel fast eine Einheit reißfester ist als Garn1? Auch die Mediane helfen nicht weiter (Median Garn1: 4,99, Garn2: 5,77). Mittelwerte und Median unterscheiden sich nicht wesentlich.

Woher kommt der offensichtliche Qualitätsunterschied?

Ein Blick auf den Boxplot klärt die Sachlage.

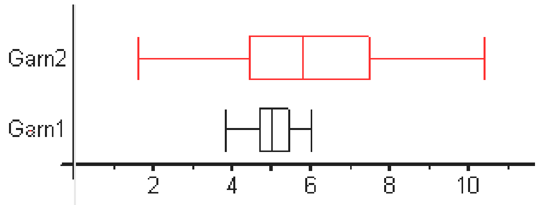


Abb.: Boxplots Reißfestigkeit

Bei Garn2 haben wir es offensichtlich mit einer viel größeren Variabilität der Reißfestigkeit zu tun als bei Garn1. So kommt es, dass mehr Proben den Grenzwert von 4 unterschreiten als bei Garn1, obwohl Garn2 im Mittel eine höhere Belastung verträgt. Der Sachverhalt wird noch klarer, wenn wir die Verteilungsfunktionen zu Rate ziehen.

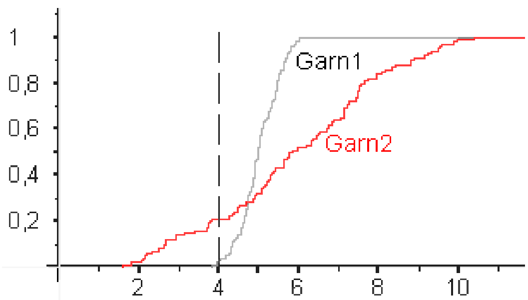


Abb.: Verteilungen Reißfestigkeit

Wir sehen, dass bei Garn2 21 % der Einzelwerte unterhalb der Schranke von 4 Belastungseinheiten liege, während bei Garn1 nur 2 Proben nicht ausreichend belastbar waren. Die unterschiedliche Variabilität der beiden Sorten drückt sich im Schaubild der Verteilungen durch eine unterschiedliche Steilheit der Kurven aus. Je steiler der Anstieg, umso geringer ist die Streuung.

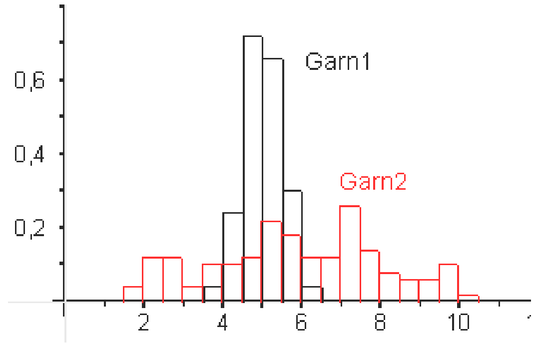
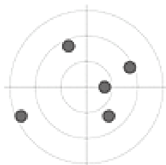


Abb.: Histogramme Reißfestigkeit

Auch die Histogramme zeigen die unterschiedliche Variabilität der beiden Sorten, der Unterschied ist aber schwerer zu erkennen. Die Gestalt der Histogramme hängt außerdem noch von der willkürlichen Wahl der Klassengrenzen ab. Damit wir den Unterschied in der Variabilität quantifizieren können, benötigen wir geeignete Maßzahlen.

2.1 Varianz und Standardabweichung



Zu den bekanntesten und am häufigsten benutzten Streuungsmaßen zählen die Standardabweichung und ihr Quadrat, die Varianz. Diese zwei wichtigen Streuungsmaße lernen Sie in dieser Lerneinheit kennen.

Die Definition von Standardabweichung und Varianz beruht auf folgender Idee: Die Abweichungen $x_i - \overline{x}$ messen, wie stark die Daten um ihren Mittelwert $\overline{x}$ streuen. Dabei treten sowohl positive wie negative Abweichungen $x_i - \overline{x}$ auf, so dass die Summe aller Abweichungen keine geeignete Maßzahl für die Streuung ist. Das arithmetische Mittel sorgt gerade dafür, dass die Summe der Abweichungen stets Null ist:

$$\sum (x_i - \overline{x}) = 0$$

(vgl. Lerneinheit *„ARM - Arithmetisches Mittel“* Abschnitt 3.1 Schwerpunkteigenschaft).

Statt der Summe der Abweichungen wird die Summe der quadrierten Abweichungen zur Berechnung der Variabilität verwendet. Einerseits haben wir durch das Quadrieren die Gewähr, dass alle Abweichungen in die Berechnung der Variabilität eingehen. Andererseits wissen wir, dass die Summe der quadrierten Abweichungen vom arithmetischen Mittel minimal ist:

$$\sum_{i=1}^{n} (x_i - c)^2 \geq \sum_{i=1}^{n} (x_i - \overline{x})^2$$

(vgl. Lerneinheit *„ARM - Arithmetisches Mittel“*, Abschnitt 3.2 Minimumeigenschaft).

Die quadrierten Abweichungen vom arithmetischen Mittel nehmen also eine Sonderstellung ein. Wenn wir jetzt noch eine Mittelwertbildung vornehmen, damit die Streuung nicht von der Anzahl der betrachteten Werte abhängig ist, haben wir einen vernünftigen Vorschlag für ein Streuungsmaß: die Varianz.

Definition

Varianz

Ist X ein mindestens intervallskaliertes Merkmal, dann heißt das Mittel der quadratischen Abweichungen der beobachteten Merkmalswerte (reelle Zahlen) x_1 , x_2 , ..., x_n von ihrem arithmetischen Mittel Varianz (engl.: variance - Veränderung, Streuung, Varianz) der Beobachtungsreihe.

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

$$\boxed{\sigma^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \mu)^2 \qquad \mu = E(X) \qquad \sigma^2 = Var(X)}$$

$s^2 = \hat{\sigma}^2$ besser als Schätzer $1/n-1$ $\hat{\mu} = \overline{x}$

Definition

Standardabweichung

Die Standardabweichung ist die Wurzel aus der Varianz.

$$s = \sqrt{s^2}$$

$$s = \frac{1}{\sigma} = \sqrt{\frac{1}{\sigma^2}}$$

Hinweis

Sie müssten sich eigentlich über den Wert n - 1 als Nenner wundern. Tun Sie es!
Wir werden später noch erklären, warum n - 1 in der Statistik oft dem Wert n als Nenner vorgezogen wird.

Wie immer zeigen wir Ihnen zunächst ein ganz einfaches Rechenbeispiel.

Beispiel

Wieder ganz einfach!

Wir betrachten die Zahlen 1, 2, 3. Die Abweichungen vom arithmetischen Mittel sind -1, 0, 1, und wir erhalten somit $\frac{(1+0+1)}{(3-1)} = 1$ als Varianz. Die Standardabweichung können Sie jetzt sicher selbst angeben. ;-)

Reißfestigkeit Streuung

Ohne Nebenrechnung geben wir Ihnen die Werte für Varianz und Standardabweichung der beiden Garnstichproben an:

Beispiel

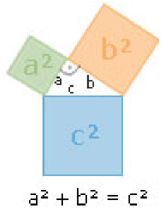
	n	s	s^2
Garn1	100	0,49	0,24
Garn2	100	2,17	4,69

Tab.: Beispiel für Varianz und Standardabweichung von zwei Garnsorten

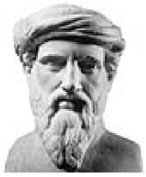
Die berechneten Werte für die Standardabweichung zeigen deutlich die unterschiedlichen Streuungen der beiden Garnsorten.

2.2 Quadratsummenzerlegung

In der Mathematik treten häufig Summen von Quadraten auf. Wir kennen alle den Satz des Pythagoras über die Seitenlängen eines rechtwinkligen Dreiecks.



Längen von Vektoren berechnet man mit Hilfe des Pythagoras, Quadratsummen gehören einfach zum Handwerkszeug. Unter anderem wird deshalb für theoretische Berechnungen die Varianz verwendet, die Standardabweichung ist als Streuungsmaß aber anschaulicher und interpretierbarer.



Allgemein lassen sich Quadratsummen auf einfache Weise zerlegen. Diese Zerlegung kann zur vereinfachten Berechnung der Varianz benutzt werden.

Quadratsummen-zerlegung

Für beliebige Datenreihen $x_1, \ldots, x_n$ und eine beliebige reelle Zahl a gilt:

$$\sum (x_i - a)^2 = n \cdot (\overline{x} - a)^2 + \sum (x_i - \overline{x})^2.$$

Steiner'scher Verschiebungssatz

In der Physik wird eine analoge Zerlegung als Steiner'scher Verschiebungssatz verwendet. Dieser gibt Auskunft über das Trägheitsmoment eines Körpers, der nicht um seinen Schwerpunkt, sondern einen beliebigen Punkt rotiert.

Wir zeigen im Folgenden die Gültigkeit der angegebenen Zerlegung.
Zunächst fügen wir in die Summe den Term $0 = -\overline{x} + \overline{x}$ ein.

$$\sum (x_i - a)^2 = \sum (x_i - \overline{x} + \overline{x} - a)^2$$

Die Anwendung der Binomischen Formel ergibt

$$\sum (x_i - a)^2 = \sum (x_i - \overline{x})^2 + 2 \sum (x_i - \overline{x})(\overline{x} - a) + \sum (\overline{x} - a)^2$$

Damit ist schon alles gezeigt, denn der mittlere Term ergibt Null. Wir können $(\overline{x} - a)$ ausklammern, und die Summe der Abweichungen vom Mittelwert ergibt bekanntlich 0 . Der letzte Ausdruck ist die Summe über die Konstante $(\overline{x} - a)^2$ und ergibt das n-fache der Konstanten.

Mit der Gültigkeit der angegebenen Zerlegung haben wir übrigens auch gezeigt, dass die Summe der Abweichungsquadrate vom arithmetischen Mittel minimal ist. Überlegen Sie sich bitte, weshalb das gilt.

Eine nützliche Anwendung erhalten wir im Spezialfall a = 0, dann gilt nämlich

Quadratsummen-zerlegung

Für beliebige Datenreihen $x_1, \ldots, x_n$ gilt:

$$\sum x_i^2 = n \cdot \overline{x}^2 + \sum (x_i - \overline{x})^2.$$

Die angegebene Formel erhalten Sie sofort, wenn Sie den Wert a = 0 in die allgemeine Quadratsummenzerlegung einsetzen.

Beispiel

Probieren geht über Studieren

Für die Zahlen 1, 2, 3 ergibt sich 1 + 4 + 9 = 14 als Quadratsumme.
Wir haben bereits gesehen, dass die Summe der Abweichungsquadrate vom arithmetischen Mittel den Wert 2 ergibt. Wir erhalten also mit n = 3 und $\overline{x}$ = 2 nach der angegebenen Formel den Wert 3 · 4 + 2 = 14.
Stimmt!

2.3 Berechnung der Varianz

Damit Sie die Berechnung der Varianz in Zukunft selbst durchführen können, haben wir ein einfaches Beispiel in Tabellenform dargestellt.

Punktzahlen Klausuren

In der folgenden Tabelle sind die zur Rechnung benötigten Terme aufgeführt.

Index i	Punkte x_i	Abweichungen $(x_i - \overline{x})$	Quadrate $(x_i - \overline{x})^2$
1	87	3	9
2	78	-6	36
3	97	13	169
4	67	-17	289
5	91	7	49
6	92	8	64
7	76	-8	64
Summe	**588**	**0**	**680**

Beispiel

Wir bestimmen zunächst das arithmetische Mittel: $\overline{x} = \frac{588}{7} = 84.$
Offensichtlich haben wir keinen Rechenfehler gemacht, denn die Summe die Abweichungen ergibt 0.

Tab.: Beispiel für Klausurpunkte und Abweichungen

Aus der Summe der Abweichungsquadrate ergibt sich nun die Varianz:
$s^2 = \frac{680}{(7-1)} = 113{,}33$ und die Standardabweichung $s = \sqrt{s^2} = \sqrt{113{,}33} = 10{,}65.$

n-1 Freiheitsgrade

Wenn Sie in der obigen Tabelle dafür sorgen, dass die Summe der Abweichungen immer den Wert Null hat, dann ist bei gegebenen n - 1 Abweichungen die n-te Abweichung schon festgelegt.

Äquivalent dazu legt die Kenntnis von n - 1 Datenpunkten und des arithmetischen Mittels den n-ten Datenpunkt fest. Man sagt deshalb, dass für die Abweichungen nur n - 1 Freiheitsgrade zu Verfügung stehen. Dieses ist eine mögliche Begründung für die Wahl von n - 1 als Nenner bei der Berechnung der Varianz.

In der mathematischen Statistik sagt man, dass die Varianz der Grundgesamtheit somit unverzerrt geschätzt wird.

Die von uns definierte Form der Varianz wird in der induktiven Statistik bevorzugt und ist deshalb oft in statistischen Softwarepaketen und auf vielen Taschenrechnern einprogrammiert. Die Formel ist auch so in der Statistiksoftware **R** implementiert.

2.4 Varianz und lineare Transformationen

Wir haben in Lerneinheit *„ARM Arithmetisches Mittel"* gesehen, dass das arithmetische Mittel mit linearen Transformationen vertauschbar ist.

Wie verhält es sich mit der Varianz bei linearen Transformationen?

Zunächst halten wir fest, dass sich die Varianz bei einer Translation (Verschiebung) der Daten nicht ändert. Gilt nämlich $y = a + x$, dann folgt das auch für den Mittelwert, und die Abweichungen vom Mittelwert bleiben unverändert: $y - \overline{y} = x - \overline{x}$

Was geschieht bei Multiplikation mit einer Konstanten b: $y = bx$?
Dann gilt: $y - \overline{y} = b(x - \overline{x})$

Damit wissen wir schon alles für die folgende Regel:

Regel

Varianzberechnung bei linearen Transformationen

Bei einer linearen Transformation eines Merkmals X der Form $Y = a + b \cdot X$ gilt:

$$s_Y^2 = b^2 \cdot s_x^2$$

Berechnen wir die Standardabweichung, dann geht das Vorzeichen von b möglicherweise verloren und es gilt:

Regel

Standardabweichung bei linearen Transformationen

Bei einer linearen Transformation eines Merkmals X der Form $Y = a + b \cdot X$ gilt:

$$s_Y = |b| \cdot s_X$$

Das folgende Beispiel soll Sie von der Nützlichkeit der angegebenen Regeln überzeugen:

Beispiel

Statistik und Gehaltserhöhung

Die 70 Beschäftigen einer Firma erhalten ein monatliches Durchschnittsgehalt von $\overline{x}$ = 2000 € mit einer Standardabweichung von s = 600 €. Aufgrund einer Produktivitätssteigerung hat der Geschäftsführer entschieden, das Monatsgehalt jedes Beschäftigen um 3 % anzuheben. Es soll in Zukunft jedes Jahr jedem Beschäftigen eine Prämie in Höhe 600 € pro Jahr gewährt werden.

Wie ändern sich Mittelwert, Varianz und Standardabweichung der Gehälter der Beschäftigen?

Es liegt eine Lineartransformation vor:
Die „alten" Gehälter $x_i (i = 1, 2, \ldots, 70)$ werden zu den „neuen" Gehältern y_i transformiert nach Maßgabe der Transformation $y_i = a + bx_i$ mit $a = \frac{600}{12} = 50$ und $b = 1{,}03.$

Nach Formeln $\overline{y} = a + b\overline{x}$ erhält man das arithmetische Mittel
$\overline{y} = 50 + 1{,}03 \cdot 2000 = 2110$ (in Euro).

Die Varianz und die Standardabweichung errechnen sich über die Formeln
$s_Y^2 = b^2 \cdot s_x^2$ bzw. $s_Y = |b| \cdot s_X$
$s_Y^2 = 1{,}03^2 \cdot 600^2 = 618^2 = 381924$ und $s_Y = 618$ (in Euro)

2.5 Übungen: Varianz und Standardabweichung I

Nach dem Motto „Grau ist alle Theorie“ können Sie sich jetzt mit Varianz und Standardabweichung vertraut machen: Sie rechnen die Werte einfach für verschiedene Beispiele aus. Zur Not helfen Ihnen dabei unsere ausführlichen Lösungen.

Berechnen

Übung VAR-01

Klausurergebnisse

Eine Studentin erreicht bei sieben Klausuren folgende Punktzahlen:
87, 78, 97, 67, 91, 92 und 79.

Berechnen Sie die Varianz und Standardabweichung der Punktzahlen.

Lösung (Siehe Anhang)

Bearbeitungszeit: 5 Minuten

Berechnen

Übung VAR-02

Klausuren von zwei Studierenden

Die Punktzahlen, die die erste Studentin bei sieben Klausuren erreichte, waren 84, 73, 97, 67, 88, 93 und 79; der zweite Student erreichte: 82, 67, 85, 97, 87, 89, 88.

Berechnen Sie die Varianz und die Standardabweichung der Punktzahl von beiden Studierenden.

Lösung (Siehe Anhang)

Bearbeitungszeit: 5 Minuten

2.6 Übungen: Varianz und Standardabweichung II

Berechnen

Übung VAR-03

Bibliotheken in europäischen Ländern

Berechnen Sie die Varianz und die Standardabweichung der (fiktiven) Anzahl von öffentlichen Bibliotheken der zehn europäischen Länder mit der größten Anzahl von Bibliotheken.

Land	Jahr	Bibliotheken
Bulgarien	2022	4237
Deutschland	2020	11817
Frankreich	2020	1620
Rumänien	2022	2953
Italien	2020	51500
Polen	2022	9230
Tschechische Republik	2022	6076
Ungarn	2021	2833
Spanien	2020	6768
Slowakei	2020	2696

Interpretieren Sie das Ergebnis.

Lösung (Siehe Anhang)

Bearbeitungszeit: 10 Minuten

Berechnen

Übung VAR-04

Einwohner je Bibliothek

Berechnen Sie die Varianz und die Standardabweichung der Einwohner je Bibliothek für die in der folgenden Tabelle aufgeführten acht europäischen Länder. Diese Länder haben die meisten öffentlichen Bibliotheken in Europa.

Interpretieren Sie bitte Ihr Ergebnis.

Land	Jahr	Bevölkerung (in 1000)	Bibliotheken	Einwohner je Bibliothek
Bulgarien	2022	8190,9	4237	1933,2
Deutschland	2020	82163,5	11817	6953,0
Frankreich	2020	58518,4	1620	36122,5
Rumänien	2022	22455,4	2953	7604,3
Italien	2020	148195,0	51500	2877,6
Polen	2022	38653,6	9230	4187,8
Tschechische Republik	2022	10289,6	6076	1693,5
Ungarn	2021	10091,8	2833	3562,2

Lösung (Siehe Anhang)

Bearbeitungszeit: 5 Minuten

Berechnen

Übung VAR-05

Unterrichtsstunden pro Schüler/in in den Bundesländern

Berechnen Sie Varianz und Standardabweichung für die beiden Variablen.

Interpretieren Sie das Ergebnis.

Was sind die Schüler den Ländern wert?

	Bundesländer	Ausgaben pro Schüler/in im Jahr (€)	Verhältnis von Schüler/innen zu Lehrenden an Gymnasien
1	Baden-Württemberg	4703,89	16,7
2	Bayern	4908,40	17,1
3	Berlin	5010,66	17,4
4	Brandenburg	4090,34	18,7
5	Bremen	5828,73	18,1
6	Hamburg	6340,02	16,1
7	Hessen	4397,11	20,4
8	Mecklenburg-Vorpommern	3988,08	18,3
9	Niedersachsen	4601,63	17,1
10	Nordrhein-Westfalen	4397,11	20,7
11	Rheinland-Pfalz	4499,37	19,3
12	Saarland	4397,11	20,7
13	Sachsen	3988,08	16,8
14	Sachsen-Anhalt	4499,37	15,7
15	Schleswig-Holstein	4703,89	16,8
16	Thüringen	4703,89	16,2

schuelerinnen.txt

Lösung (Siehe Anhang)

Bearbeitungszeit: 10 Minuten

Berechnen

Übung VAR-06

Grundstückspreise

Die Inhaberin des Weinfachgeschäftes Maestro hat 10 neue Angebote für Grundstücke in Veneto bekommen. Die Grundstücke haben folgende Preise:

Angebot	Preis
1	10500
2	5000
3	4300
4	12000
5	14000
6	10500
7	6200
8	5000
9	7300
10	10500

Berechnen Sie die Varianz und die Standardabweichung der Grundstückspreise.

Lösung (Siehe Anhang)

Bearbeitungszeit: 15 Minuten

3 Varianzberechnung aus Häufigkeitsverteilungen

Wie wir schon in Abschnitt 2 Streuungsmaße gesehen haben, ist die Varianz eigentlich eine Eigenschaft der Verteilung der Daten. Je steiler die Verteilungsfunktion verläuft, umso geringer ist die Streuung.

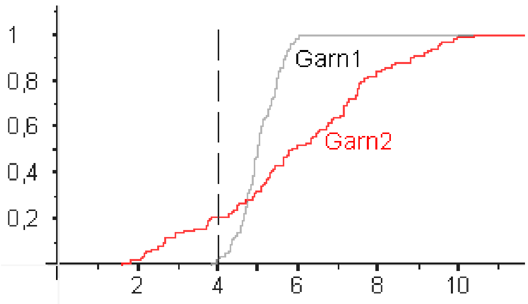


Abb.: Reißfestigkeit Verteilung

Garn1: s = 0,49, Garn2: s = 2,17

Der Zusammenhang zwischen Verteilung und Streuung wird verdeutlicht, wenn wir als Berechnungsgrundlage die Häufigkeitsverteilung unserer Daten verwenden. Dadurch ändert sich natürlich nicht der Wert der Varianz, aber Sie werden es leichter haben, wenn Sie sich in der induktiven Statistik mit der theoretischen Varianz einer Grundgesamtheit beschäftigen wollen.

In den folgenden Abschnitten werden wir Sie deshalb mit Berechnungsformeln, die sich für Häufigkeitsverteilungen eignen, vertraut machen.

3.1 Formeln zur Varianzberechnung

In diesem Abschnitt zeigen wir Ihnen verschiedene Varianten der Varianzberechnung, die sich durch Anwendung der Quadratsummenzerlegung herleiten lassen. Außerdem stellen wir Ihnen die Formeln für gewogene Werte vor. Zunächst noch einmal sicherer Grund:

Messwerte:

Aus den n ursprünglichen Messwerten$x_1, x_2, \ldots, x_n$ berechnet man die Varianz als Mittel der Abweichungsquadrate:

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

Wenn die Daten als Häufigkeitstabelle vorliegen, benötigen wir weniger Rechenschritte.

Häufigkeitstabelle:

Aus den k voneinander verschiedenen Merkmalswerten $x_i (i = 1, \ldots, k)$ einer Häufigkeitstabelle berechnet man die Varianz als ein gewogenes quadratisches Mittel mit absoluten Häufigkeiten: $s^2 = \frac{1}{N-1}\left(\sum_{i=1}^{k}(x_i - \overline{x})^2 n_i\right)$

oder als gewogenes quadratisches Mittel mit relativen Häufigkeiten:

$$s^2 = \frac{N}{N-1}\left(\sum_{i=1}^{k}(x_i - \overline{x})^2 h_i\right)$$

wobei N gleich der Gesamtanzahl ist.

In der nachstehenden Tabelle haben wir alle gängigen Formeln für die Varianz zusammengefasst. Die Verwendung von N bei tabellierten Daten ist üblich, deshalb verwenden wir für die Formeln mit gruppierten Daten N. Natürlich stimmen n und N überein.

Tab.: Formeln zur Varianzberechnung

Varianz	aus ursprünglichen Messwerten	aus einer Häufigkeitstabelle	
		mit absoluten Häufigkeiten	mit relativen Häufigkeiten
Definition	$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2$	$s^2 = \frac{1}{N-1}\left(\sum_{i=1}^{k}(x_i - \overline{x})^2 n_i\right)$	$s^2 = \frac{N}{N-1}\left(\sum_{i=1}^{k}(x_i - \overline{x})^2 h_i\right)$
Entwickelte Form	$s^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} x_i^2 - n \cdot \overline{x}^2\right)$	$s^2 = \frac{1}{N-1}\left(\sum_{i=1}^{k} x_i^2 n_i - n\overline{x}^2\right)$	$s^2 = \frac{N}{N-1}\left(\sum_{i=1}^{k} x_i^2 h_i - \overline{x}^2\right)$
Entwickelte Form für Tabellen-rechnung	$s^2 = \frac{1}{n-1} \cdot \left[\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}\right]$	$s^2 = \frac{1}{N-1} \cdot \left[\sum_{i=1}^{k} x_i^2 n_i - \frac{\left(\sum_{i=1}^{k} x_i n_i\right)^2}{n}\right]$	$s^2 = \frac{N}{N-1} \cdot \left(\sum_{i=1}^{k} x_i^2 h_i - \left(\sum_{i=1}^{k} x_i h_i\right)^2\right)$

3.2 Varianz bei klassierten Daten

Sie erinnern sich (?), wie wir in Lerneinheit *„ARM - Arithmetisches Mittel“* näherungsweise das arithmetische Mittel aus der Häufigkeitsverteilung klassierter Werte berechnet haben. Dabei haben wir die Formeln für gewichtete Berechnungen verwendet. Dasselbe Vorgehen wollen wir Ihnen auch für die Varianz an einem Beispiel erläutern.

In der folgenden Tabelle ist eine Häufigkeitsverteilung der Körpergewichte von 100 Studenten einer Fachhochschule vorgestellt. Lassen Sie sich nicht verwirren, in der Mathematik werden nicht immer die gleichen Symbole verwendet. Hier bezeichnen wir die Klassenmitte mit ${x_i}^*$ und nicht mit m_i wie in den vorherigen Formeln.

Beispiel

Körpergewichte, klassiert

Wir berechnen die Varianz und die Standardabweichung der Körpergewichte von 100 Studenten näherungsweise aus den klassierten Daten.

Die Körpergewichte sind nur als ganze kg erfasst und als Klassengrenzen ergeben sich wieder ganze Zahlen, z. B. Mitte Klasse 1 = 58, Klasse 2 = 61 usw. Bitte verwenden Sie diese Klassenmitten bei Ihren Berechnungen.

Index i	Klasse K_i	Klassenmitte ${x_i}^*$	absolute Klassenhäufigkeit n_i	Produkte ${x_i}^* \cdot n_i$
1	[57; 60)	58	2	116
2	[60; 63)	61	5	305
3	[63; 66)	64	16	1024
4	[66; 69)	67	42	2814
5	[69; 72)	70	25	1750
6	[72; 75)	73	7	511
7	[75; 78)	76	3	228
		Summe:	**100**	**6748**

Tab.: Beispiel für klassierte Daten von Körpergewichten

Zur Berechnung werden die benötigten Hilfsgrößen nach und nach in eine Tabelle eingetragen. Das Ergebnis sehen Sie gleich.

Das in Näherung bestimmte arithmetische Mittel hat den Wert $\overline{x} = 67.48$.

Wie man das berechnet, sollten Sie noch wissen, ansonsten schauen Sie bitte noch einmal in Lerneinheit *„ARM - Arithmetisches Mittel“*, Abschnitt 4 nach.

Die folgende Tabelle wurde zur besseren Darstellung auf dem Bildschirm in zwei Teile aufgespalten.

Index i	Klasse K_i	Klassenmitte ${x_i}^*$	Abweichungen $(x_i - \overline{x})$	Quadrate $(x_i - \overline{x})^2$
1	[57; 60)	58	-9,48	89,8704
2	[60; 63)	61	- 6,48	41,9904
3	[63; 66)	64	- 3,48	12,1104
4	[66; 69)	67	- 0,48	0,2304
5	[69; 72)	70	2,52	6,3504
6	[72; 75)	73	5,52	30,4704
7	[75; 78)	76	8,52	72,5904

Index i	**absolute Klassenhäufigkeit** n_i	**Produkte** $({x_i}^* - \overline{x})^2 \cdot n_i$	**relative Klassenhäufigkeit** h_i	**Produkte** $({x_i}^* - \overline{x})^2 \cdot h_i$
1	2	179,7408	0,02	1,797408
2	5	209,9520	0,05	2,099520
3	16	193,7664	0,16	1,937664
4	42	9,6768	0,42	0,096768
5	25	158,7600	0,25	1,587600
6	7	213,2928	0,07	2,132928
7	3	217,7712	0,03	2,177712
	$N = \sum_{i=1}^{k} n_i = 100$	$\sum_{i=1}^{k} (x_i^* - \overline{x})^2 n_i = 1182{,}9600$	$\sum_{i=1}^{k} h_i = 1$	$\sum_{i=1}^{k} (x_i^* - \overline{x})^2 h_i = 11{,}8296$

Tab.: Beispiel – Hilfgrößen zur Berechnung der Varianz und Standardabweichung

In der Tabelle finden wir alle Zwischenergebnisse, die wir zur Varianzberechnung benötigen.

Wir erhalten $s^2 = \frac{1}{N-1} \sum_{i=1}^{k} (x_i^* - \overline{x})^2 n_i = \frac{1182{,}96}{99} = 11{,}9491$

Die Standardabweichung hat somit den Wert s $= \sqrt{11{,}9491} = 3{,}46\, kg.$

Ergebnisinterpretation

Die einzelnen Körpergewichte streuen (in Näherung) durchschnittlich um 3,46 kg um ihre (in Näherung bestimmte) durchschnittlichen Körpergewichte von 67,48 kg.

Hinweis

Die Berechnung der Standardabweichung und der Varianz über relative Häufigkeiten führt zum selben Ergebnis. Überprüfen Sie das doch bitte selbst.

In diesem konkreten Beispiel ist zu beachten, dass sowohl das arithmetische Mittel als auch die Varianz und die Standardabweichung aufgrund der Datenaggregation nur näherungsweise bestimmt werden können. Eine genaue Bestimmung der jeweiligen Verteilungsparameter setzt die Existenz von Urlistendaten voraus.

3.3 Varianzberechnung: Kodierte Werte

Wir zeigen Ihnen die Anwendung der Quadratsummenzerlegung anhand eines Beispiels klassierter Werte, bei dem wir von der Möglichkeit der Kodierung Gebrauch machen.

In Abschnitt 4.1 der Lerneinheit *„ARM - Arithmetisches Mittel"* haben wir die mittlere Miete von 99 Wohnungen aus der Häufigkeitstabelle kodierter Werte näherungsweise bestimmt.
Anstelle der ursprünglichen Klassenmitten m_k des Merkmals X wurden kodierte Werte u_i verwendet $u_i = \frac{(m_k-500)}{200}$, die Rücktransformation ergab dann den gesuchten Mittelwert.

Zur Berechnung der Varianz der kodierten Werte benutzen wir folgende Formel, die sich aus der Quadratsummenzerlegung ergibt:

$$s_u^2 = \left(\frac{1}{(N-1)}\right) \cdot \left[\sum f_i \cdot u_i^2 - \frac{\left(\sum f_i \cdot u_i\right)^2}{N}\right],$$

wobei $N = \sum f_i$ die Gesamtzahl der Werte angibt.

Unsere Regel für Varianzberechnung bei linearer Transformation liefert dann die gesuchte Varianz der ursprünglichen Werte als ${s_x}^2 = 200^2 \cdot {s_u}^2$.

Anhand der folgenden Tabelle sollte Ihnen die Anwendung der Formel klar werden.

Index i	**Klasse** K_i	**Klassenmitte (kodiert)** u_i	**Häufigkeit** f_i	**Produkte** $f_i \cdot u_i$	**Produkte** $f_i \cdot {u_i}^2$
1	K_1= [0; 200)	- 2	0	0	0
2	K_2= [200; 400)	- 1	43	- 43	- 43
3	K_3= [400; 600	0	38	0	0
4	K_4 = [600; 800)	1	9	9	9
5	K_5 = [800; 1000)	2	3	6	12
6	K_6 = [1000; 1200)	3	4	12	36
7	K_7 = [1200; 1400)	4	1	4	16
8	K_8 = [1400; 1600)	5	1	5	25
		Summe:	**99**	**- 7**	**141**

Tab.: Beispiel von kodierten Werten zur Varianzberechnung

Wir hatten schon in Abschnitt 4.3 der Lerneinheit *„ARM - Arithmetisches Mittel"* gesehen, wie sich das arithmetische Mittel der Mieten mit Hilfe der kodierten Werte berechnen lässt, es gilt:
$\overline{x} = 500 + 200 \cdot -\frac{7}{99} = 486.85\,€$

Die Berechnung der Varianz der kodierten Werte ergibt nach der oben angegebenen Formel
$s_u^2 = \left(\frac{1}{98}\right) \cdot \left(141 - \frac{49}{99}\right) = 1.197$

Da $s_x = 200 \times s_U$ gilt, erhalten wir für die Standardabweichung der Mieten den Wert
$s_x = 200 \cdot \sqrt{1{,}197} = 239.48\,€$.

Zum Vergleich noch einmal die Berechnungen auf Grund der Originalwerte:
Mittelwert: 495 €, Standardabweichung: 228,5 €.

Sie sehen, die Näherung durch die Klassierung der Daten liefert keine exakten Werte, hier jedoch ausreichend gute.

Zusammenfassung

- Die Angabe der Streuung ist für die Beschreibung von Daten unerlässlich.
- Die Standardabweichung ist eine geeignete Maßzahl für die Streuung der Daten.
- Die Varianz wird mit Hilfe der Abweichungsquadrate vom arithmetischen Mittel berechnet.
- Die Quadratsummenzerlegung der Daten kann zur Varianzberechnung verwendet werden.
- Bei linearen Transformationen lässt sich die Varianz der transformierten Werte leicht berechnen. Verschiebungen ändern nicht die Varianz.
- Die Varianz kann bei klassierten Daten näherungsweise durch gewichtete Berechnung ermittelt werden.
- Das Verfahren der Kodierung lässt sich auch für Varianzen anwenden.

Sie sind am Ende dieser Lerneinheit angelangt. Auf den folgenden Seiten finden Sie noch die Übungen zur Wissensüberprüfung und die Übungen mit der Statistiksoftware R.

Wissensüberprüfung

Mit den folgenden Übungen können Sie ihr Wissen überprüfen. Versuchen Sie zuerst, die Aufgaben selbst zu lösen, bevor Sie sich die Antworten anzeigen lassen.

Multiple Choice

Übung VAR-07

Welche Aussagen sind falsch und welche richtig?

	Richtig	Falsch	Auswertung
Je stärker die Merkmalswerte um den Mittelwert konzentriert sind, desto kleiner ist die Streuung.			
Es gibt negative Standardabweichung.			
Die Standardabweichung hat nicht die gleiche Maßeinheit wie die Werte selbst.			
Die Varianz kann negativ sein.			
Die Varianz misst die Streuung um das arithmetische Mittel mit der gleichen Maßeinheit.			

Multiple Choice

Übung VAR-08

Welche Aussagen sind falsch und welche richtig?

	Richtig	Falsch	Auswertung
Die Standardabweichung ist stets kleiner als die Varianz.			
Werden die Einzelwerte mit 2 multipliziert, ergibt sich die zweifache Varianz.			
Werden die Einzelwerte um 2 erhöht, erhöht sich die Standardabweichung um 2.			

Übungen mit der Statistiksoftware R

Statistiksoftware R

Die in der Lerneinheit behandelten Themen können Sie anhand der folgenden Übungsaufgaben auch mit der Statistiksoftware **R** bearbeiten. Um die Übungen zu bearbeiten, sollte die Software „**R**“ auf Ihrem Rechner installiert sein.

Installationshinweise [Manuals | **R** Installation and Administration]

Versuchen Sie, die auf dieser Seite gestellten Aufgaben mit Hilfe von **R** oder **R-Studio** zu lösen. Ihre Ergebnisse können Sie mit der Lösungsdatei vergleichen. Speichern Sie dazu die R-Datei mit der Lösung auf Ihrem Rechner und öffnen Sie die Datei mit **R**. Überprüfen Sie vor dem Öffnen, ob das Arbeitsverzeichnis von **R** mit dem Speicherort der Datei übereinstimmt.

Für einige Übungen stehen auch Musterlösungen für das Programm Excel bereit.

Berechnen

Übung VAR-09a

Sieben Klausuren

Die Punktzahlen, die eine Studentin bei sieben Klausuren erreichte, waren 87, 78, 97, 67, 91, 92 und 79.

1. Berechnen Sie Varianz und Standardabweichung der Punktzahlen.

Lösung mit R und Excel (Siehe Anhang)

Bearbeitungszeit: 10 Minuten

Berechnen

Übung VAR-09b

Zwei Studierende

Die Punktzahlen, die ein Student bei sieben Klausuren erreichte, waren
84, 73, 97, 67, 88, 93 und 79.

Seine Kommilitonin erreichte:
82, 67, 85, 97, 87, 89 und 88.

1. Berechnen Sie Varianz und Standardabweichung der Punktzahlen von beiden.
2. Interpretieren Sie das Ergebnis

Lösung mit R und Excel (Siehe Anhang)

Bearbeitungszeit: 10 Minuten

Berechnen

Übung VAR-09c

Wiederholungsprüfung

Die folgende Liste beinhaltet die Anzahl der Wiederholungsprüfungen im Fach Statistik von 117 Studierenden einer Berliner Hochschule, die im Verlauf des WS 2022 ihre Prufung absolvierten.

0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 1 0 0 2 0 0 0 0 0 0 0 1
0 0 0 1 0 0 0 0 0 1 0 0 1 0 0 3 0 0 1 0 0 0 0 2 0 1 0 1 0 1 0 0 0 1 0 0 1 1
0 1 0 0 1 0 0 0 0 0 1 0 0 0 1 0 2 1 1 0 0 0 0 0 1 0 1 1 1 1 0 2 0 1 1 2 1 1
0 1 0

Aufgabe

1. Charakterisieren Sie die Verteilung des Erhebungsmerkmals mit Hilfe der Varianz und der Standardabweichung.
2. Interpretieren Sie diese Verteilungsmaßzahlen.

Lösung mit R (Siehe Anhang)

Bearbeitungszeit: 10 Minuten

Berechnen

Übung VAR-09d

Körpergewichte

In der folgenden Tabelle ist eine Häufigkeitsverteilung in Klassen der Körpergewichte von 100 Studierenden einer Fachhochschule vorgestellt.

x_i	**Klasse** K_i **(kg)**	**Klassenmitte** ${x_i}^*$ **(kg)**	**absolute Klassenhäufigkeit** n_i
1	[57; 60)	58	2
2	[60; 63)	61	5
3	[63; 66)	64	16
4	[66; 69)	67	42
5	[69; 72)	70	25
6	[72; 75)	73	7
7	[75; 78)	76	3

1. Berechnen Sie Varianz und Standardabweichung der Körpergewichte der 100 Studierenden.

Lösung mit R und Excel (Siehe Anhang)

Bearbeitungszeit: 10 Minuten

Zusätzliche Übungsaufgaben

Berechnen

Übung VAR-10

Aufgabe zum Aufwärmen

Berechnen Sie die Varianz und die Standardabweichung der folgenden Daten.

a. 5, 0, 8, 6, 5, 9
b. 3, 3, 3, 3, 3, 3
c. 1, 3, 1, 3, 1, 3

Lösung (Siehe Anhang)

Bearbeitungszeit: 6 Minuten

Berechnen

Übung VAR-11

Gewicht eines Briefes

Mit einer elektronischen Waage wurde wiederholt das Gewicht eines Briefes (in g) gemessen. Man erhielt folgende Daten:

22,94	22,90	22,92	22,76	22,80	22,85	22,84	22,86

Berechnen Sie den Mittelwert und die Standardabweichung.

Lösung (Siehe Anhang)

Bearbeitungszeit: 3 Minuten

Berechnen

Übung VAR-12

VAR für die Häufigkeitsverteilung

Berechnen Sie für die Häufigkeitsverteilung die Varianz und die Standardabweichung.

x_i	0	1	2	3	4	5
h_i	0,2	0,325	0,25	0,15	0,05	0,025

Lösung (Siehe Anhang)

Bearbeitungszeit: 3 Minuten

Berechnen

Übung VAR-13

Verkaufszahlen für Sportschuhe

Die folgende Häufigkeitstabelle zeigt die Verkaufszahlen eines Sportfachgeschäftes für die Sportschuhe im Laufe eines Jahres:

Jan	Feb	Mär	Apr	Mai	Jun	Jul	Aug	Sep	Okt	Nov	Dez
2	0	4	12	24	54	43	35	48	35	8	1

Berechnen Sie den Mittelwert und die Standardabweichung.

Lösung (Siehe Anhang)

Bearbeitungszeit: 3 Minuten

Berechnen

Übung VAR-14

Gewichtsverteilung einer Gruppe Studierender

Bestimmen Sie für folgende Tabelle mit der klassierten Gewichtsverteilung (kg) einer Gruppe von Studierenden die Varianz und die Standardabweichung.

x_i	48	53	58	63	68	73	78	83
n_i	8	8	16	14	9	12	2	4

Lösung (Siehe Anhang)

Bearbeitungszeit: 3 Minuten

Berechnen

Übung VAR-15

Reaktionstest

In einer Fahrschule wurden bei einem Reaktionstest mit 10 Personen die folgenden Reaktionszeiten in Sekunden gemessen:

0,70	0,64	0,61	0,65	0,78	0,59	0,63	0,64	0,63	0,61

Berechnen Sie die Varianz und die Standardabweichung für die Reaktionszeit.

Lösung (Siehe Anhang)

Bearbeitungszeit: 3 Minuten

Berechnen

Übung VAR-16

Fehltage

Für das 10. Schuljahr einer Klasse der Gesamtschule wurde eine Erhebung zur Anzahl der Fehltage der einzelnen Schüler/innen gemacht:

18, 19, 5, 9, 2, 15, 48, 14, 0, 0, 22, 0, 32, 0, 16, 0, 12, 6, 20, 9

Bestimmen Sie die Varianz und die Standardabweichung der Fehltage.

Lösung (Siehe Anhang)

Bearbeitungszeit: 3 Minuten

Berechnen

Übung VAR-17

Durchschnittsnoten

Für 20 Schüler/innen eines Gymnasiums wurde eine Erhebung zu den Durchschnittsnoten auf den Zeugnissen gemacht:

2,4	2,6	2,0	3	2,1	2,5	2,6	2,4	3,3	2,3
3,3	2,9	2,4	3,0	3,4	2,7	2,4	2,8	2,5	2,9

Bestimmen Sie die Varianz und die Standardabweichung für die Durchschnittsnoten.

Lösung (Siehe Anhang)

Bearbeitungszeit: 3 Minuten

Berechnen

Übung VAR-18

Gewichtsexperiment

Ein Student des Studiengangs Mathematik stellt sich einen Monat lang jeden Morgen auf seine elektronische Waage und misst sein Gewicht bzw. die Masse seines Körpers und bekommt folgende Werte:

80, 81, 80, 80, 83, 82, 82, 80, 82, 81, 81, 83, 84, 84, 85, 85, 83,
84, 82, 82, 81, 80, 80, 83, 83, 82, 80, 80, 81, 82

Bestimmen Sie die Standardabweichung des Monatsgewichtes.

Lösung (Siehe Anhang)

Bearbeitungszeit: 3 Minuten

Berechnen

Übung VAR-19

Varianz bei klassierten Daten

Gegeben ist die folgende Häufigkeitsverteilung eines klassierten Merkmals:

Klasse	[10; 20)	[20; 30)	[30; 40)	[40; 50)
Häufigkeit	12	23	20	50

Bestimmen Sie die Varianz und die Standardabweichung des klassierten Merkmals.

Lösung (Siehe Anhang)

Bearbeitungszeit: 3 Minuten

Appendix

Lösung für Übung VAR-01

Klausurergebnisse

Arithmetisches Mittel der Punktzahlen: $\overline{x}$ = 84,4.
Die Varianz berechnet man als ein einfaches arithmetisches Mittel aus den quadrierten Abweichungen der Ausprägungen x_i vom arithmetischen Mittel der einzelnen Datensätze:

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

$$= \frac{(87-84{,}4)^2 + (78-84{,}4)^2 + (97-84{,}4)^2 + (67-84{,}4)^2 + (91-84{,}4)^2 + (92-84{,}4)^2 + (79-84{,}4)^2}{6}$$

$$= \frac{6{,}76 + 40{,}96 + 158{,}76 + 302{,}76 + 43{,}56 + 57{,}76 + 29{,}16}{6} = \frac{639{,}72}{6} = 106{,}62$$

Standardabweichung:

$$s = \sqrt{106{,}62} = 10{,}3$$

Lösung für Übung VAR-02

Klausuren von zwei Studierenden

Bei der Auswertung der Klausurergebnisse erreichten die Studierenden folgende Werte

1. Studentin:
arithmetisches Mittel der Punktzahlen: $\overline{x}_1 = 83$.

Varianz: $s_1^2 = \frac{1}{n-1}\sum\limits_{i=1}^{n}(x_i - \overline{x})^2 = 115{,}67$

Standardabweichung: $s_1 = \sqrt{115{,}67} = 10{,}75$

2. Student:
arithmetisches Mittel der Punktzahlen: $\overline{x}_2 = 85$.

Varianz: $s_2^2 = \frac{1}{n-1}\sum\limits_{i=1}^{n}(x_i - \overline{x})^2 = 84{,}33$

Standardabweichung: $s_1 = \sqrt{84{,}33} = 9{,}18$

Interpretation:
Im Durchschnitt weichen die beobachteten n = 7 Punktzahlen der Klausuren von ihrem Durchschnitt bei der ersten Studentin $\overline{x}_1$ = 83 bzw. bei dem zweiten Studenten $\overline{x}_2$ = 85 um s_1 = 10,75 bzw. s_2 = 9,18 nach oben und unten ab.

Zum Vergleich werden in der folgenden Tabelle verschiedene Streuungsmaße berechnet: Spannweite, Quartilsabstände, mittlere absolute Abweichung vom Median, mittlere absolute Abweichung vom arithmetischen Mittel, Varianz und Standardabweichung. Die zusätzlichen Streuungsmaße werden Sie in der Lerneinheit *„ASM Alternative Streuungsmaße“* kennenlernen.

Vergleichstabelle der Klausurenergebnisse der beiden Studierenden

Klausurergebnisse	**Studentin 1 (67, 73, 79, 84, 88, 93, 97)**	**Student 2 (67, 82, 85, 87, 88, 89, 97)**
Spannweite R	30	30
Quartilsabstand $Q_{0.5}$	20	7
mittlere abs. Abweichung vom Median MA $\overline{x}_{0.5}$	8,43	5,71
mittlere abs. Abweichung vom arithm. Mittel MA $\overline{x}$	8,57	6,0
Varianz s^2;	115,67	84,33
Standardabweichung s	10,75	9,18

Lösung für Übung VAR-03

Bibliotheken in europäischen Ländern

Land	x_i	$x_i - \overline{x}$	$(x_i - \overline{x})^2$	x_i^2
Italien	51500	41527	1724491729	2652250000
Deutschland	11817	1844	3400336	139641489
Polen	9230	- 743	552049	85192900
Spanien	6768	- 3205	10272025	45805824
Tschechische Republik	6076	- 3897	15186609	36917776
Bulgarien	4237	- 5736	32901696	17952169
Rumänien	2953	- 7020	49280400	8720209
Ungarn	2833	- 7140	50979600	8025889
Slowakei	2696	- 7277	52954729	7268416
Frankreich	1620	- 8353	69772609	2624400
Summe	**99730**	**0**	**2009791782**	**3004399072**

Arithmetisches Mittel der Anzahl der öffentlichen Bibliotheken: $\overline{x} = 9973.$

Berechnungsweise aus Daten:

Varianz: $s_1^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2 = \frac{2009791782}{10-1} = \frac{2009791782}{9} = 223310198$

Standardabweichung: $s = \sqrt{223310198} = 14944$

Andere (aus den Daten entwickelte) Form der Berechnungsweise (handlicher und weniger anfällig gegenüber möglichen Rundungsungenauigkeiten):

$$s^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} x_i^2 - n \cdot \overline{x}^2\right) = \frac{1}{10-1}\left(3004399072 - 10 \cdot (9973)^2\right) = 223310198$$

Interpretation:
Im Durchschnitt weichen die beobachteten Anzahlen der öffentlichen Bibliotheken von ihrem Durchschnitt $\overline{x}$ = 9973 um s = 14944 nach oben und unten ab.

Lösung mit R

bibliotheken_loesung.R

```r
# Einlesen der fiktiven Daten
biblio <-
  data.frame(
    Land = c(
      "Italien",
      "Deutschland",
      "Polen",
      "Bulgarien",
      "Spanien",
      "Tschechische Republik",
      "Rumaenien",
      "Ungarn",
      "Slowakei",
      "Frankreich"
    ),
    Bibliotheken = c(51500, 11817, 9230, 4237, 6768,
                     6076, 2953, 2833, 2696, 1620)
  )

#  Varianz
var(biblio$Bibliotheken)

#  Standardabweichung
sd(biblio$Bibliotheken)
```

Lösung für Übung VAR-04

Einwohner je Bibliothek

Arithmetisches Mittel der Anzahl der Einwohner je öffentliche Bibliothek: $\overline{x}$ = 8.116,8.

Berechnungsweise aus Daten:
n: 8
Mittelwert: 8116,8
Varianz: 132.723.235,7
Standardabweichung: 11.520,6

Interpretation
Im Durchschnitt weicht die beobachtete Anzahl der Einwohner je Bibliothek von ihrem Durchschnitt $\overline{x}$ = 8.116,8 um s = 11.520,6 nach oben bzw. nach unten ab.

Lösung für Übung VAR-05

Unterrichtsstunden pro Schüler/in in den Bundesländern

Entwickelte Form der Berechnungsweise aus Daten:

Varianz der Ausgaben pro Schüler/in im Jahr:

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

$$= \frac{1}{15}\left(40195853{,}6 + \ldots + 15904782{,}09 - 16 \cdot (4691{,}1)^2\right)$$

$$= \frac{357956629 - 352103458}{15} = 390211{,}38$$

Standardabweichung: $s = \sqrt{s^2} = \sqrt{390211{,}38} = 624{,}67$

Varianz des Verhältnisses von Schülern und Schülerinnen zu Lehrenden an Gymnasien:

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

$$= \frac{1}{15}\left(259{,}21 + \ldots + 282{,}24 - 16 \cdot (17{,}78)^2\right)$$

$$= \frac{5089{,}42 - 5055{,}21}{15} = 2{,}28$$

Standardabweichung: $s = \sqrt{s^2} = \sqrt{2{,}28} = 1{,}51$

Interpretation:
Die Ausgaben pro Schüler/in im Jahr:
s^2 = 390.211,38 , d. h. die mittlere quadratische Abweichung vom Mittelwert 4.691,10 € beträgt 390.211,38.
s = 624,67, d. h. die Standardabweichung vom Mittelwert beträgt 624,67 €.

Das Verhältnis von Schülern und Schülerinnen zu Lehrenden an Gymnasien:
s^2 = 2,28, d. h. die mittlere quadratische Abweichung der durchschnittlichen Anzahl der Schüler und Schülerinnen pro Lehrende/m vom Mittelwert 17,78 beträgt 2,28 (Schüler/innen pro Lehrer/in)^2.
s = 1,51, d. h. die Standardabweichung der durchschnittlichen Anzahl der Schüler und Schülerinnen pro Lehrende/m vom Mittelwert beträgt 1,51 Schüler/innen pro Lehrer/in.

Lösung mit R

schuelerinnen_loesung.R
schuelerinnen.txt

```r
# Einlesen der Daten aus der Datei schuelerinnen.txt
schule<-read.table("schuelerinnen.txt",sep="\t",header=TRUE)

#  Varianz für die Kosten
round(var(schule$Ausgaben),2)

#  Varianz für das Verhältnis
round(var(schule$Verhaeltnis),2)

#  Standardabweichung für die Kosten
round(sd(schule$Ausgaben),2)

#  Standardabweichung für das Verhältnis
round(sd(schule$Verhaeltnis),2)
```

Lösung für Übung VAR-06

Grundstückspreise

Als Erstes wird eine Häufigkeitstabelle erstellt, in der die absoluten und relativen Häufigkeiten dargestellt werden.

Index i	Merkmalsausprägungen x_i	absolute Häufigkeit n_i	relative Häufigkeit h_i
1	4,3	1	0,1
2	5,0	2	0,2
3	6,2	1	0,1
4	7,3	1	0,1
5	10,5	3	0,3
6	12,0	1	0,1
7	14,0	1	0,1

Das arithmetische Mittel der Preise der Grundstücke ist: $\overline{x}$ = 8,53.

Im Folgenden sind 9 Lösungsvarianten für die Berechnung der Varianz und der Standardabweichung ausführlich dargestellt. Alle führen zum gleichen Ergebnis (Das soll ja auch so sein!). Aber schauen Sie sich die Berechnungen doch einfach einmal an.

1. Aus den Daten berechnen sich die Varianz und die Standardabweichung wie folgt:

Varianz:

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

$$= \frac{1}{10-1}\left[(10{,}5-8{,}53)^2+(5-8{,}53)^2+\ldots+(7{,}3-8{,}53)^2+(10{,}5-8{,}53)^2\right]$$

$$= \frac{1}{9}(3{,}8809+12{,}4609+\ldots+1{,}5129+3{,}8809) = \frac{1}{9}\cdot 103{,}361 = 11{,}48$$

Standardabweichung: $s = \sqrt{s^2} = \sqrt{11{,}48} = 3{,}389$

2. Aus einer Häufigkeitstabelle berechnen sich Varianz und Standardabweichung als gewogenes quadratisches Mittel mit absoluten Häufigkeiten zu:

Varianz:

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

$$= \frac{1}{10-1}\Big[(4{,}3-8{,}53)^2\cdot 1+(5-8{,}53)^2\cdot 2+(6{,}2-8{,}53)^2\cdot 1+(7{,}3-8{,}53)^2\cdot 1$$

$$+(10{,}5-8{,}53)^2\cdot 3+(12{,}0-8{,}53)^2\cdot 1+(14{,}0-8{,}53)^2\cdot 1\Big]$$

$$= \frac{1}{9}(17{,}8929+\ldots+29{,}9209) = \frac{1}{9}\cdot 103{,}361 = 11{,}48$$

Standardabweichung: $s = \sqrt{s^2} = \sqrt{11{,}48} = 3{,}389$

3. Berechnungsweise aus einer Häufigkeitstabelle als gewogenes quadratisches Mittel mit relativen Häufigkeiten:

Varianz:

$$s^2 = \frac{N}{N-1} \sum_{i=1}^{n} (x - \overline{x})^2 h_i$$

$$= \frac{10}{9} \Big((4{,}3 - 8{,}53)^2 \cdot 0{,}1 + (5 - 8{,}53)^2 \cdot 0{,}2 + (6{,}2 - 8{,}53)^2 \cdot 0{,}1 + (7{,}3 - 8{,}53)^2 \cdot 0{,}1$$

$$+ (10{,}5 - 8{,}53)^2 \cdot 0{,}3 + (12{,}0 - 8{,}53)^2 \cdot 0{,}1 + (14{,}0 - 8{,}53)^2 \cdot 0{,}1 \Big)$$

$$= 1{,}111 \cdot (17{,}8929 \cdot 0{,}1 + \ldots + 29{,}9209 \cdot 0{,}1) = 11{,}48$$

Standardabweichung: $s = \sqrt{s^2} = \sqrt{11{,}48} = 3{,}389$

4. Entwickelte Berechnungsweise aus den Daten:

Varianz:

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

$$= \frac{10{,}5^2 + 5^2 + \ldots + 10{,}5^2 - 10 \cdot (8{,}53)^2}{9}$$

$$= \frac{1}{9}(1917{,}67 - 10 \cdot 72{,}7609) = 11{,}4846$$

Diese Formel ist vor allem zur schnellen Berechnung von Varianz s^2 geeignet.

Man bildet die Summe der quadrierten Werte x_i^2, mittelt diese und zieht das bereits vorher berechnete quadrierte arithmetische Mittel $\overline{x}^2$ ab.

Standardabweichung: $s = \sqrt{s^2} = \sqrt{11{,}48} = 3{,}389$

5. Entwickelte Berechnungsweise aus einer Häufigkeitstabelle als gewogenes quadratisches Mittel mit absoluten Häufigkeiten:

Varianz:

$$s^2 = \frac{1}{N-1} \left(\sum_{i=1}^{k} x_i^2 n_i - n\overline{x}^2 \right)$$

$$= \frac{1}{9} \Big((4{,}3^2 \cdot 1 + 5^2 \cdot 2 + \ldots + 14{,}0^2 \cdot 1) - 10(8{,}53)^2 \Big)$$

$$= \frac{1}{9}(1917{,}67 - 10 \cdot 72{,}7609) = 11{,}4846$$

Standardabweichung:$s = \sqrt{s^2} = \sqrt{11{,}48} = 3{,}389$

6. Entwickelte Berechnungsweise aus einer Häufigkeitstabelle als gewogenes quadratisches Mittel mit relativen Häufigkeiten:

Varianz:

$$s^2 = \frac{N}{N-1}\left(\sum_{i=1}^{k} x_i^2 h_i - \overline{x}^2\right)$$

$$= \frac{10}{9}\Big((4{,}3^2 \cdot 0{,}1 + 5^2 \cdot 0{,}2 + \ldots + 14{,}0^2 \cdot 0{,}1) - (8{,}83)^2\Big)$$

$$= 1{,}111 \cdot (83{,}0970 - 72{,}7609) = 11{,}48$$

Standardabweichung: $s = \sqrt{s^2} = \sqrt{11{,}48} = 3{,}389$

Um die Auswirkungen von Rundungsfehlern insbesondere bei der Entwicklung von Software zu begrenzen, empfiehlt es sich, $\sqrt{s^2}$ nach den folgenden drei Formeln zu berechnen.

7. Entwickelte Berechnungsweise für Tabellenrechnung aus Daten:

Varianz:

$$s^2 = \frac{1}{n-1}\left[\sum_{i=1}^{n} x_i^2 - \frac{(\sum\limits_{i=1}^{n} x_i)^2}{n}\right]$$

$$= \frac{1}{9}\left[(10{,}5^2 + 5{,}0^2 + \ldots + 14{,}0^2) - \frac{(10{,}5 + 5{,}0 + \ldots 14{,}0)^2}{10}\right]$$

$$= \frac{1}{9}(830{,}970 - 727{,}609) = 11{,}48$$

Standardabweichung: $s = \sqrt{s^2} = \sqrt{11{,}48} = 3{,}389$

8. Entwickelte Berechnungsweise für Tabellenrechnung aus einer Häufigkeitstabelle als gewogenes quadratisches Mittel mit absoluten Häufigkeiten:

Varianz:

$$s^2 = \frac{1}{N-1}\left[\sum_{i=1}^{k} x_i^2 n_i - \frac{(\sum\limits_{i=1}^{k} x_i n_i)^2}{N}\right]$$

$$= \frac{1}{9}\left(4{,}3^2 + 5{,}0^2 + \ldots + 14{,}0^2 - \frac{(4{,}3 \cdot 1 + 5{,}0 \cdot 2 + \ldots 14{,}0 \cdot 1)}{10}\right)$$

$$= \frac{1}{9}(830{,}970 - 727{,}609) = 11{,}48$$

Standardabweichung:$s = \sqrt{s^2} = \sqrt{11{,}48} = 3{,}389$

9. Entwickelte Berechnungsweise für Tabellenrechnung aus einer Häufigkeitstabelle als gewogenes quadratisches Mittel mit relativen Häufigkeiten:

Varianz:

$$s^2 = \frac{N}{N-1}\left(\sum_{i=1}^{k} x_i^2 h_i - \left(\sum_{i=1}^{k} x_i h_i\right)^2\right)$$

$$= \frac{10}{9}\left((4{,}3^2 \cdot 0{,}1 + 5{,}0^2 \cdot 0{,}2 + \ldots + 14{,}0^2 \cdot 0{,}1) - (4{,}3 \cdot 0{,}1 + 5{,}0 \cdot 0{,}2 + \ldots + 14{,}0 \cdot 0{,}1)^2\right)$$

$$= 1{,}111 \cdot (83{,}0970 - 72{,}7609) = 11{,}48$$

Standardabweichung: $s = \sqrt{s^2} = \sqrt{11{,}48} = 3{,}389$

Interpretation:

$\sqrt{s^2} = 11{,}48$, d. h. die mittlere quadratische Abweichung der Grundstückspreise vom Mittelwert 8,53 beträgt 11,48 (Tausend €)^2.

s = 3,389, d. h. die mittlere standardisierte Abweichung der Grundstückspreise vom Mittelwert beträgt 3,22 Tausend €.

Lösung mit R

grundstuecke_loesung.R

```
# Einlesen der Daten
angebote<-c(10.5,5,4.3,12,14,10.5,6.2,5,7.3,10.5)

#  Varianz
var(angebote)

#  Standardabweichung
sd(angebote)
```

Lösung für Übung VAR-09a

Sieben Klausuren

1. Die Varianz ist 106,619, die Standardabweichung beträgt 10,32565

Lösung mit R

sklausuren_loesung.R

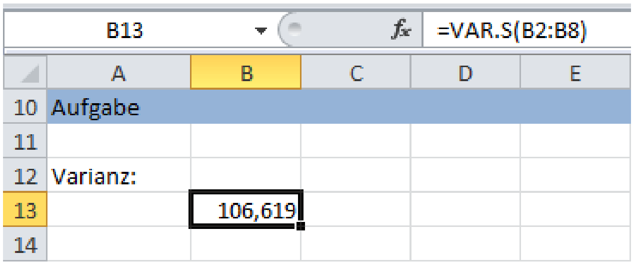
```r
#  Sieben Klausuren - Aufgabe

#  Einlesen der Ergebnisse in einen Vektor
punkte<-c(87,78,97,67,91,92,79)

#  Varianz
var(punkte)

#  Standardabweichung
sd(punkte)
```

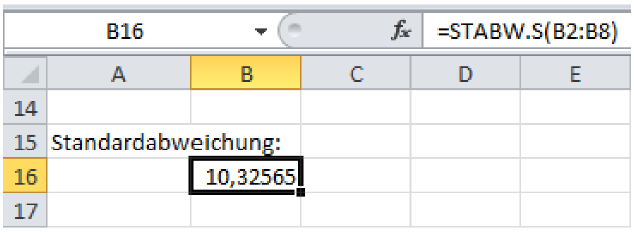
Lösung mit Excel

WMS_VAR_09_Sieben_Klausuren.xlsx (20 KB)

***Aufgabe 1**: Die Punktzahlen, die eine Studentin bei sieben Klausuren erreichte, waren 87, 78, 97, 67, 91, 92*
und 79. Berechnen Sie die Varianz und die Standardabweichung der Punktzahlen.

Um die Größen Varianz und Standardabweichung der Punktzahlen zu berechnen, verwenden wir die Funktionen `VAR.S` und `STABW.S`. In den nachfolgenden Bildern ist das genaue Vorgehen zu sehen.





Das Ergebnis ist folgendes:

Varianz:	
	106,62
Standardabweichung:	
	10,33

Lösung für Übung VAR-09b

Zwei Studierende

1. Student (1): Die Varianz ist 115,67, die Standardabweichung beträgt 10,75
   Studentin (2): Die Varianz ist 84,33, die Standardabweichung beträgt 9,18
2. Die durchschnittliche Abweichung der Punktzahlen vom Mittelwert beträgt etwa 11 Punkte bei dem Studenten (1) und etwa 9 Punkte bei der Studentin (2). Das heißt, dass die Studentin (2) etwas konstantere Ergebnisse bei den Klausuren erzielt hat.

Lösung mit R

2studierende_loesung.R

```
#  Einlesen der Punktzahlen Student = (1), Studentin = (2)
punkte1 <- c(84, 73, 97, 67, 88, 93, 79)
punkte2 <- c(82, 67, 85, 97, 87, 89, 88)
punkte1
punkte2

#  Berechnung der Varianz und der Standardabweichung für den Studenten (1):
round(var(punkte1), 2)
round(sd(punkte1), 2)

#  Berechnung der Varianz und der Standardabweichung für die Studentin (2):
round(var(punkte2), 2)
round(sd(punkte2), 2)w
```

Lösung mit Excel

WMS_VAR_09_Zwei_Studierende.xlsx (9 KB)

Aufgabe 1: *Die Punktzahlen, die der Student (1) bei sieben Klausuren erreichte, waren 84, 73, 97, 67, 88, 93 und 79. Seine Kommilitonin (2) erreichte: 82, 67, 85, 97, 87, 89, 88. Berechnen Sie die Varianz und die Standardabweichung der Punktzahl von beiden und interpretieren Sie die Ergebnisse.*

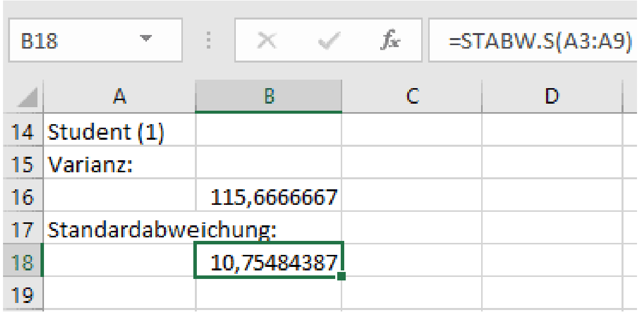
Um die Größen Standardabweichung und Varianz zu berechnen, verwenden wir wieder die Funktionen `VAR.S` und `STABW.S`. Dafür schreiben wir die Punktzahlen zunächst in der folgenden Form auf.

	A	B
1	Punkte	
2	Student (1)	Studentin (2)
3	84	82
4	73	67
5	97	85
6	67	97
7	88	87
8	93	89
9	79	88

Dann tragen wir in die Zellen für den ersten Student folgendes ein:

B16 =VAR.S(A3:A9)

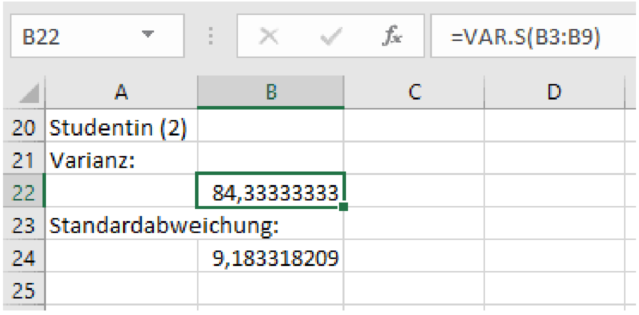
	A	B	C	D
14	Student (1)			
15	Varianz:			
16		115,6666667		
17	Standardabweichung:			
18		10,75484387		
19				



B18 | =STABW.S(A3:A9)

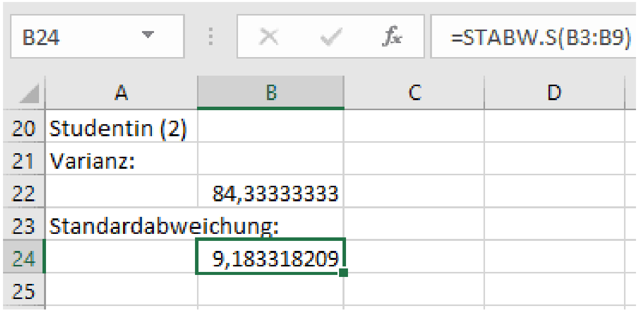
	A	B	C	D
14	Student (1)			
15	Varianz:			
16		115,6666667		
17	Standardabweichung:			
18		10,75484387		
19				

Für den zweiten Student gehen wir genauso vor:



B22 | =VAR.S(B3:B9)

	A	B	C	D
20	Studentin (2)			
21	Varianz:			
22		84,33333333		
23	Standardabweichung:			
24		9,183318209		
25				



B24 | =STABW.S(B3:B9)

	A	B	C	D
20	Studentin (2)			
21	Varianz:			
22		84,33333333		
23	Standardabweichung:			
24		9,183318209		
25				

Daraus erhalten wir das Ergebnis:

Student (1): Die Varianz ist 115,67, die Standardabweichung beträgt 10.75
Studentin (2): Die Varianz ist 84,33, die Standardabweichung beträgt 9,18

Lösung für Übung VAR-09c

Wiederholungsprüfung

1. Die Varianz beträgt 0,3723843
   Die Standardabweichung beträgt: 0,610233
2. Interpretation: Die mittlere Abweichung der Anzahl an Wiederholungen zum Mittelwert liegt bei etwa 0,6.
   Die mittlere quadratische Abweichung liegt in etwa bei 0,4.

Lösung mit R

wpruefung_loesung.R

```
# Einlesen der Daten
absolventenliste <-
  c(
    0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,1,
    0,0,0,0,0,0,0,1,0,0,2,0,0,0,0,0,0,0,1,
    0,0,0,1,0,0,0,0,0,1,0,0,1,0,0,3,0,0,1,
    0,0,0,0,2,0,1,0,1,0,1,0,0,0,1,0,0,1,1,
    0,1,0,0,1,0,0,0,0,0,1,0,0,0,1,0,2,1,1,
    0,0,0,0,0,1,0,1,1,1,1,0,2,0,1,1,2,1,1,
    0,1,0
  )

#  Varianz
var(absolventenliste)

#  Standardabweichung
sd(absolventenliste)
```

Lösung für Übung VAR-09d

Körpergewichte

1. Die Varianz beträgt 11,94909
   Die Standardabweichung beträgt: 3,456746

Lösung mit R

kgewichte_loesung.R

```
#  Einlesen der Klassenmitten
km<-c(58.5,61.5,64.5,67.5,70.5,73.5,76.5)

#  Häufigkeiten
hkeit<-c(2,5,16,42,25,7,3)

#  Liste der Klassenmitten und deren Häufigkeit erstellen
gewichte<-rep(km,hkeit)

#  Varianz
var(gewichte)

#  Standardabweichung
sd(gewichte)
```

Lösung mit Excel

WMS_VAR_09_Koerpergewichte.xlsx (10 KB)

Aufgabe 1: *Berechnen Sie Varianz und Standardabweichung der Körpergewichte der 100 Studierenden.*

Liegen die Daten klassiert vor, so ist die Berechnung von Standardabweichung und Varianz etwas aufwändiger. Zunächst berechnen wir die Klassenmitten und damit das Produkt aus Klassenmitte und Klassenhäufigkeit. Wir erhalten folgende Tabelle.

D16 =B16*C16

	A	B	C	D	E
13	Ergänzen der Tabelle um näherungsweises arithmetisches Mittel zu bestimmen:				
14					
15	Klasse	Häufigkeit	Klassenmitte	Produkt von Klassenmitte und Klassenhäufigkeit	
16	57 bis 60kg	2	58,5	117	
17	60 bis 63kg	5	61,5	307,5	
18	63 bis 66kg	16	64,5	1032	
19	66 bis 69kg	42	67,5	2835	
20	69 bis 72kg	25	70,5	1762,5	
21	72 bis 75kg	7	73,5	514,5	
22	75 bis 78kg	3	76,5	229,5	

Mit der ergänzten Tabelle können wir das arithmetische Mittel berechnen.

D25 =D24/B24

	A	B	C	D
24	Summe:	100		6798
25	arithmetisches Mittel:			67,98
26				

Mithilfe des arithmetischen Mittels können wir die Tabelle nun um eine Spalte ergänzen, die das Produkt aus quadrierter Abweichung und der Klassenhäufigkeit enthält. Die Summe der Werte in dieser Spalte berechnen wir in Zelle E36, indem wir die Funktion `SUMME` verwenden.

E28 =(C28-D$25)^2*B28

	A	B	C	D	E	F	G	H
27	Klasse	Häufigk	Klassenmitte	Produkt von Klassenm	Produkt von quadrierter Abweichung und Klassenhäufigkeit			
28	57 bis 60kg	2	58,5	117	179,7408			
29	60 bis 63kg	5	61,5	307,5	209,952			
30	63 bis 66kg	16	64,5	1032	193,7664			
31	66 bis 69kg	42	67,5	2835	9,6768			
32	69 bis 72kg	25	70,5	1762,5	158,76			
33	72 bis 75kg	7	73,5	514,5	213,2928			
34	75 bis 78kg	3	76,5	229,5	217,7712			
35								
36	Summe:				1182,96			

Die Varianz berechnet sich nun aus der Summe, die durch die Anzahl an Beobachtungen (um 1 reduziert) geteilt wird. Was in der Zelle stehen muss, ist im folgenden Bild zu sehen.

B39 =E36/(B24-1)

	A	B	C	D
38	Varianz:			
39		11,9490909		
40	Standardabweichung:			
41		3,45674571		
42				

Die Standardabweichung ergibt sich aus der Varianz durch das Ziehen der Wurzel – wie in Zelle B41 zu sehen. Dafür verwenden wir die Funktion `WURZEL`.

B41 | =WURZEL(B39)

	A	B	C	D
38	Varianz:			
39		11,9490909		
40	Standardabweichung:			
41		3,45674571		
42				

Übersichtlich sieht das Ergebnis so aus:

Varianz:	
	11,95
Standardabweichung:	
	3,46

Lösung für Übung VAR-10

Aufgabe zum Aufwärmen

a.)

$$\overline{x} = \frac{1}{6}(5+0+8+6+5+9) = 5{,}5$$

$$Var_x = \frac{1}{5}\left[(5-5{,}5)^2 + (0-5.5)^2 + (8-5{,}5)^2 + (6-5{,}5)^2 + (5-5{,}5)^2 + (9-5{,}5)^2\right]$$

$$= \frac{1}{5} \cdot 49{,}25 = 9{,}9$$

$$s_x = \sqrt{9{,}9} = 3{,}15$$

b.)

$$\overline{x} = \frac{1}{6}(3+3+3+3+3+3) = 3$$

$$Var_x = \frac{1}{5}\left[(3-3)^2 + (3-3)^2 + (3-3)^2 + (3-3)^2 + (3-3)^2 + (3-3)^2\right]$$

$$= \frac{1}{5} \cdot 0 = 0$$

$$s_x = 0$$

c.)

$$\overline{x} = \frac{1}{6}(1+3+1+3+1+3) = 2$$

$$Var_x = \frac{1}{5}\left[(1-2)^2 + (3-2)^2 + (1-2)^2 + (3-2)^2 + (1-2)^2 + (3-2)^2\right]$$

$$= \frac{1}{5} \cdot 6 = 1{,}2$$

$$s_x = \sqrt{1{,}2} = 1{,}1$$

Lösung für Übung VAR-11

Gewicht eines Briefes

$$\overline{x} = \frac{1}{8}(22{,}94 + 22{,}90 + 22{,}92 + 22{,}76 + 22{,}80 + 22{,}85 + 22{,}84 + 22{,}86)$$

$$= 22{,}859$$

$$Var_x = \frac{1}{7}\sum(x_i - 22{,}86)^2 = 0{,}004$$

$$s_x = \sqrt{0{,}004} = 0{,}06$$

Lösung für Übung VAR-12

VAR für die Häufigkeitsverteilung

i	x_i	h_i	$x_i \cdot h_i$	$\overline{x}$	$(x_i - \overline{x})^2 h_i$
1	0	0,200	0,000	1,6	0,512
2	1	0,325	0,325	1,6	0,117
3	2	0,250	0,500	1,6	0,004
4	3	0,150	0,450	1,6	0,294
5	4	0,050	0,200	1,6	0,288
6	5	0,025	0,125	1,6	0,289
		1	**1,6**		**1,54**

$$\overline{x} = \sum x_i \cdot h_i = 1{,}6$$

$$Var_x = \frac{N}{N-1}\left[\sum(x_i - \overline{x})^2 h_i\right] = 1{,}848$$

$$s^2 = \sqrt{1{,}848} = 1{,}359$$

Lösung für Übung VAR-13

Verkaufszahlen für die Sportschuhe

$$\overline{x} = \frac{266}{12} = 22{,}167$$

$$Var_x = \frac{4427{,}668}{11} = 402{,}52$$

$$s = \sqrt{402{,}52} = 20{,}06$$

Lösung für Übung VAR-14

Gewichtsverteilung einer Gruppe Studierender

$\overline{x} = 62{,}93$

$Var_x = \frac{6474{,}66}{72} = 89{,}93$

$s = \sqrt{89{,}92} = 9{,}48$

Lösung für Übung VAR-15

Reaktionstest

$\overline{x} = 0{,}648$

$Var = 0{,}003$

$s = \sqrt{0{,}003} = 0{,}055$

Lösung für Übung VAR-16

Fehltage

$\overline{x} = 12{,}35$

$Var = 152{,}34$

$s = \sqrt{152{,}34} = 12{,}34$

Lösung für Übung VAR-17

Durchschnittsnoten

$\overline{x} = 2{,}68$

$Var = 0{,}15$

$s = \sqrt{0{,}15} = 0{,}39$

Lösung für Übung VAR-18

Gewichtsexperiment

$\overline{x} = 81{,}87$

$Var = 2{,}46$

$s = \sqrt{2{,}46} = 1{,}57$

Lösung für Übung VAR-19

Varianz bei klassierten Daten

Klasse	Klassen-mittte	n_i	$n_i \cdot x_i$	$(x_i - \overline{x})$	$(x_i - \overline{x})^2$	$(x_i - \overline{x})^2 n_i$
[10; 20)	15	12	180	- 20,29	411,68	4940,21
[20; 30)	25	23	575	- 10,29	105,88	2435,33
[30; 40)	35	20	700	- 0,29	0,08	1,68
[40; 50)	45	50	2250	9,71	94,28	4714,21
Σ		**105**	**3705**			**12091,43**

$\overline{x} = 35{,}29$

$$Var = \frac{12091{,}43}{104} = 116{,}26$$

$s = \sqrt{116{,}26} = 10{,}78$