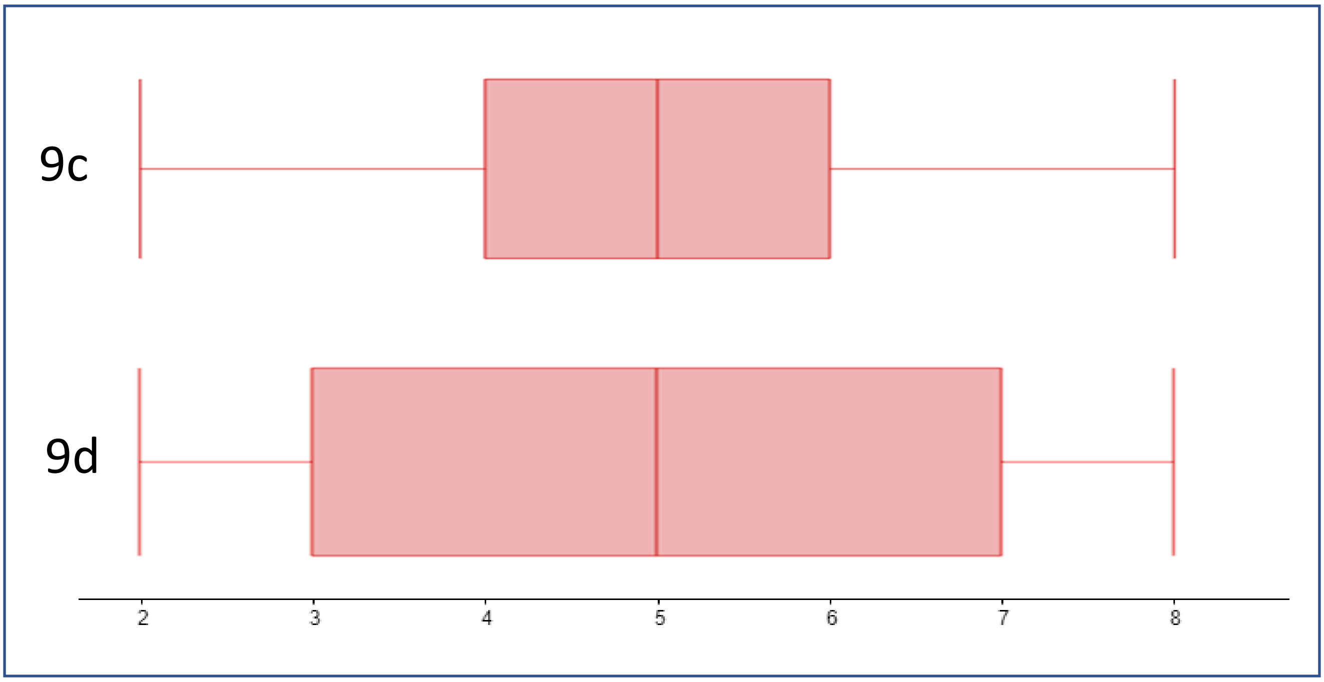
Input 1: Streumaße

Lehrervortrag zur Interpretation von Boxplots mit besonderer Berücksichtigung der Streuung, verbunden mit der Motivation ein Modell zu suchen, welches die Streuung durch einen Zahlenwert beschreibt.
Anmerkung: Die Erstellung der Boxplots kann auch als vorbereitende Hausaufgabe gestellt werden.

Folie:



Heterogene Gruppenarbeit[1]: Berechnungsmodelle für die Abweichungen

In Gruppen soll die Gesamtabweichung nach folgenden Modellen berechnet werden:

- Modell 1: Summe der Differenzen zwischen Auszahlungsbetrag und theoretischem Mittelwert
- Modell 2: Summe der Beträge der Differenzen zwischen Auszahlungsbetrag und theoretischem Mittelwert
- Modell 3: Summe der Beträge der Quadrate der Differenzen zwischen Auszahlungsbetrag und theoretischem Mittelwert

Präsentation und Unterrichtsgespräch: Vorstellen und Vergleichen der Modelle

- Kurze Beschreibung des verwendeten Modells, Nennen des Ergebnisses (mit Einheit!)
- Diskussion der Ergebnisse, Abwägen der Vor- und Nachteile[2], Antizipieren der erzielbaren Effekte mit dem jeweiligen Modell.

Sicherung[3]

Definition der Begriffe Varianz und Standardabweichung; Bezug zu den Modellen 1 – 3 herstellen.

---

[1] Arbeitsaufträge auf den Folgeseiten
[2] Vgl. auch Material der ZPG VIII zur Normalverteilung (Auszug auf Seite 6)
[3] Definition in Anlehnung an das Formeldokument (Seite 5)

Konzeptionsgruppe Abitur 2024

Modell 1

Dieses Modell berücksichtigt die positiven und negativen Abweichungen

Betrag in €	2	3	4	5	6	7	8
9a (27 Spiele)	2	3	5	7	5	3	2
9b (28 Spiele)	1	5	6	6	5	3	2
9c (29 Spiele)	2	4	3	8	7	3	2
9d (27 Spiele)	4	4	4	3	5	5	2
9e (29 Spiele)	3	4	6	6	4	4	2
9f (31 Spiele)	2	5	5	8	4	4	3
9g (29 Spiele)	0	5	5	7	7	3	2

Beispielrechnung für die Klasse 9a: Abweichungen nach unten bzw. oben vom theoretischen Mittelwert 5 €

Betrag in €	2	3	4	5	6	7	8
Abweichung in €	$2-5=-3$	$3-5=-2$	$4-5=-1$	$5-5=0$	$6-5=1$	$7-5=2$	$8-5=3$
Häufigkeit H	2	3	5	7	5	3	2
Gesamtabweichung in €	$2\cdot(-3)=-6$	$3\cdot(-2)=-6$	$5\cdot(-1)=-5$	$7\cdot 0=0$	$1\cdot 5=5$	$3\cdot 2=6$	$2\cdot 3=6$
Summe der Gesamtabweichungen	$-6+(-6)+(-5)+0+5+6+6=\mathbf{0}$						

a) Berechnet die Summe der Gesamtabweichungen für die restlichen Klassen, ihr könnt dabei arbeitsteilig vorgehen.
b) Berechnet die durchschnittliche Abweichung je Spiel.
c) Stellt eine Formel für die Berechnung der Summe der Gesamtabweichungen auf.
d) Stellt eine Formel für die Berechnung der durchschnittlichen Abweichung je Spiel auf.
e) Überlegt euch Vor- und Nachteile eures Modells.

Modell 2

Dieses Modell berücksichtigt die Beträge der Abweichungen.

Betrag in €	2	3	4	5	6	7	8
9a (27 Spiele)	2	3	5	7	5	3	2
9b (28 Spiele)	1	5	6	6	5	3	2
9c (29 Spiele)	2	4	3	8	7	3	2
9d (27 Spiele)	4	4	4	3	5	5	2
9e (29 Spiele)	3	4	6	6	4	4	2
9f (31 Spiele)	2	5	5	8	4	4	3
9g (29 Spiele)	0	5	5	7	7	3	2

Beispielrechnung für die Klasse 9a: Abweichungen vom theoretischen Mittelwert 5 €

Betrag in €	2	3	4	5	6	7	8
Abweichung in €	$\lvert 2-5\rvert = 3$	$\lvert 3-5\rvert = 2$	$\lvert 4-5\rvert = 1$	$\lvert 5-5\rvert = 0$	$\lvert 6-5\rvert = 1$	$\lvert 7-5\rvert = 2$	$\lvert 8-5\rvert = 3$
Häufigkeit H	2	3	5	7	5	3	2
Gesamtabweichung in €	$2 \cdot 3 = 6$	$3 \cdot 2 = 6$		$7 \cdot 0 = 0$	$5 \cdot 1 = 5$	$3 \cdot 2 = 6$	$2 \cdot 3 = 6$
Summe der Gesamtabweichungen	$6 + 6 + 5 + 0 + 5 + 6 + 6 =$ **34**						

a) Berechnet die Summe der Gesamtabweichungen für die restlichen Klassen, ihr könnt dabei arbeitsteilig vorgehen.
b) Berechnet die durchschnittliche Abweichung je Spiel.
c) Stellt eine Formel für die Berechnung der Summe der Gesamtabweichungen auf.
d) Stellt eine Formel für die Berechnung der durchschnittlichen Abweichung je Spiel auf.
e) Überlegt euch Vor- und Nachteile eures Modells.

Modell 3

Dieses Modell berücksichtigt die Quadrate der Abweichungen.

Betrag in €	2	3	4	5	6	7	8
9a (27 Spiele)	2	3	5	7	5	3	2
9b (28 Spiele)	1	5	6	6	5	3	2
9c (29 Spiele)	2	4	3	8	7	3	2
9d (27 Spiele)	4	4	4	3	5	5	2
9e (29 Spiele)	3	4	6	6	4	4	2
9f (31 Spiele)	2	5	5	8	4	4	3
9g (29 Spiele)	0	5	5	7	7	3	2

Beispielrechnung für die Klasse 9a: Abweichungen nach unten bzw. oben vom theoretischen Mittelwert 5 €

Betrag in €	2	3	4	5	6	7	8
Abweichung in €	$(2-5)^2 = 9$	$(3-5)^2 = 4$	$(4-5)^2 = 1$	$(5-5)^2 = 0$	$(6-5)^2 = 1$	$(7-5)^2 = 4$	$(8-5)^2 = 9$
Häufigkeit H	2	3	5	7	5	3	2
Gesamtabweichung in €	$2 \cdot 9 = 18$	$3 \cdot 4 = 12$	$5 \cdot 1 = 5$	$7 \cdot 0 = 0$	$5 \cdot 1 = 5$	$3 \cdot 4 = 12$	$2 \cdot 9 = 18$
Summe der Gesamtabweichungen	6 + 6 + 5 + 0 + 5 + 6 + 6 = **34**						

a) Berechnet die Summe der Gesamtabweichungen für die restlichen Klassen, ihr könnt dabei arbeitsteilig vorgehen.
b) Berechnet die durchschnittliche Abweichung je Spiel.
c) Stellt eine Formel für die Berechnung der Summe der Gesamtabweichungen auf.
d) Stellt eine Formel für die Berechnung der durchschnittlichen Abweichung je Spiel auf.
e) Überlegt euch Vor- und Nachteile eures Modells.

III) Streuung eines Datensatzes messbar machen – Varianz und Standardabweichung

Als Maß für die **Streuung** (Abweichungen der einzelnen Daten eines Datensatzes vom Mittelwert) gibt es in der Statistik zwei Größen: Die **Varianz *V*** und die **Standardabweichung *σ***.

Berechnung der Varianz und der Standardabweichung mithilfe der absoluten Häufigkeiten H eines Datensatzes aus n Daten mit den Werten d_1, d_2, … , d_n und dem Mittelwert $\overline{m}$::

1) Abweichungen zum Mittelwert berechnen und Quadrieren
2) Summe der Quadrate bilden und diese Summe durch die Anzahl n dividieren.
3) Für die Standardabweichung: die Wurzel aus diesem Wert ziehen.

Berechnung der Varianz und der Standardabweichung mithilfe der relativen Häufigkeiten h eines Datensatzes aus n Daten mit den Werten d_1, d_2, … , d_n und dem Mittelwert $\overline{m}$:

$$V = (d_1 - \overline{m})^2 \cdot h_1 + (d_2 - \overline{m})^2 \cdot h_2 + \cdots + (d_3 - \overline{m})^2 \cdot h_n$$

$$\sigma = \sqrt{V}$$

Berechnung der Varianz und der Standardabweichung mithilfe der Wahrscheinlichkeiten p einer Zufallsgröße *X* mit den Werten x_1, x_2, … , x_n und dem Erwartungswert E(X):

$$V(X) = \left(x_1 - E(X)\right)^2 \cdot P(X = x_1) + \left(x_2 - E(X)\right)^2 \cdot P(X = x_2) + \cdots + \left(x_n - E(X)\right)^2 \cdot P(X = x_n)$$

$$\sigma(X) = \sqrt{V(X)}$$

Nachbereitung:

Erläutern der Vorteile der Standardabweichung als Streumaß gegenüber der Varianz.

Übungen: s. Aufgabensammlung

WTR-Einsatz: s. Material der ZPG VIII zur Normalverteilung
(Auszug auf Seite 7: Casio FX-87DE X; Auszug auf Seite 8: TI-30X Plus MathPrint)

Varianz und Standardabweichung

Beispiel:

Zwei Maschinen A und B schneiden Stahlstifte auf die Länge 10 cm zu. Von jeder Maschine wurden 30 Stifte entnommen und nachgemessen.

Länge in cm	7	8	9	10	11	12	13
Anzahl A	1	3	7	8	7	3	1
Anzahl B	0	2	8	10	8	2	0

Mittelwert:

Aufgrund der Symmetrie der Daten gilt $\overline{x_A} = 10\,\text{cm}$ und $\overline{x_B} = 10\,\text{cm}$.

Man hat also denselben Mittelwert, offensichtlich aber verschiedene Streuungen um den Mittelwert.

Wie kann man nun diese Streuung der Daten erfassen?

(1) Zunächst liegt es nahe, diese Streuung der Daten durch die Summe ihrer Abweichungen $x_i - \overline{x}$ vom Mittelwert zu bestimmen.

$$S_A = \frac{1}{30}\cdot\left[(7-10)\cdot 1+(8-10)\cdot 3+(9-10)\cdot 7+(10-10)\cdot 8+(11-10)\cdot 7+(12-10)\cdot 3+(13-10)\cdot 1\right]=0$$

Analog erhält man für $S_B = 0$.

Da sich die positiven und negativen Abweichungen gegenseitig aufheben, ergibt die Summe 0 und ist daher kein verwertbares Maß für die mittlere Abweichung vom Mittelwert.

(2) Das Problem ließe sich beseitigen, indem man über die Beträge der Abweichungen $\left|x_i - \overline{x}\right|$ aufsummiert.

$$S_A = \frac{1}{30}\cdot\Big[|7-10|\cdot 1+|8-10|\cdot 3+|9-10|\cdot 7+|10-10|\cdot 8+|11-10|\cdot 7$$
$$+|12-10|\cdot 3+|13-10|\cdot 1\Big] = \frac{32}{30} \approx 1{,}067$$

Analog erhält man für $S_B = \frac{24}{30} \approx 0{,}8$

(3) Eine andere Möglichkeit, die Streuung zu berechnen, ist die Summe der quadratischen Abweichungen $\left(x_i - \overline{x}\right)^2$. Diese hat den Vorteil, dass dabei die größeren Abweichungen stärker gewichtet werden.

$$S_A{}^2 = \frac{1}{30}\cdot\Big[(7-10)^2\cdot 1+(8-10)^2\cdot 3+(9-10)^2\cdot 7+(10-10)^2\cdot 8+(11-10)^2\cdot 7$$
$$+(12-10)^2\cdot 3+(13-10)^2\cdot 1\Big] \approx 1{,}866$$

Analog erhält man für $S_B{}^2 \approx 1{,}066$ und damit $S_A \approx 1{,}366$ und $S_B \approx 1{,}033$

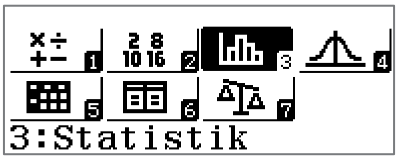
Hinweise zum WTR-Einsatz (CASIO FX-87DE X)

1. Eingabe von Daten / Ermitteln der Kenngrößen μ und σ

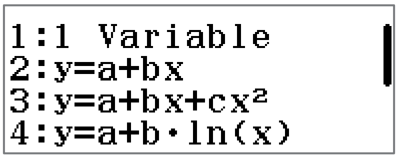
Kann ein Datensatz als normalverteilt angenommen werden, so entspricht der Mittelwert dem Erwartungswert. Für die Standardabweichung bietet der WTR zwei Kenngrößen an:

- σx: die aus dem Datensatz errechnete Standardabweichung
- sx: eine aus der Analyse des Datensatzes empirisch ermittelte Standardabweichung

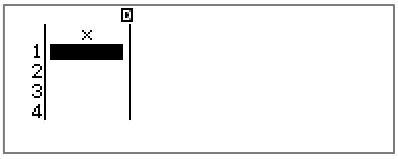
Aufrufen des Statistik-Menus `3:Statistik`



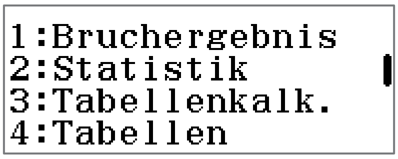
Untermenu `1:Variable`
Es öffnet sich ein Bildschirm mit einer Spalte (Liste).

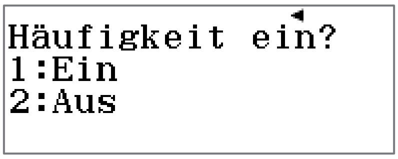


Hier kann nun der Datensatz eingegeben werden.

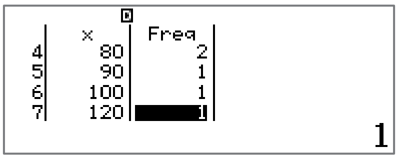


Sollen Daten sowie die zugehörigen Häufigkeiten eingegeben werden, muss zuvor in `SHIFT` `SETUP` (eventuell mit der Pfeiltaste nach unten scrollen) unter `2:Statistik` bei `Häufigkeit ein?` `1:Ein` ausgewählt werden.

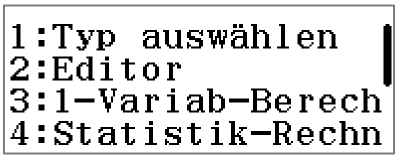




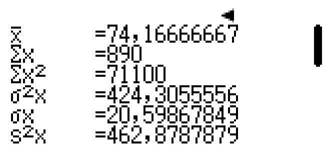
Zurück im Statistik-Menu hat man nun zwei Spalten: In die erste Spalte gibt man die Daten ein, in die zweite die jeweiligen Häufigkeiten.



Das Löschen der Daten erfolgt über `OPTN` `2:Editor` und `2:Alles löschen`.



Zur Ausgabe der Kenngrößen gelangt man über `OPTN` `3:1-Variab-Berech`



$\bar{x}$	Mittelwert
$\sum x$ / $\sum x^2$	Summe aller Daten / Summe aller Datenquadrate
$\sigma^2 x$ / σx	Varianz / Standardabweichung (aus Datensatz ermittelt)
$s^2 x$ / sx	Varianz / Standardabweichung (empirisch ermittelt)
n	Gesamtzahl der Daten
min(X) /max(X)	Minimum / Maximum
Q_1 / Q_3	unteres Quartil / oberes Quartil
Med	Median

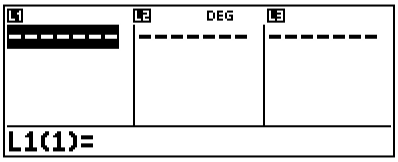
Hinweise zum WTR-Einsatz (TI-30X Plus MathPrint)

1. Eingabe von Daten / Ermitteln der Kenngrößen μ und σ

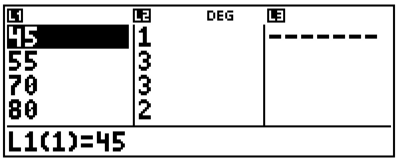
Kann ein Datensatz als normalverteilt angenommen werden, so entspricht der Mittelwert dem Erwartungswert. Für die Standardabweichung bietet der WTR zwei Kenngrößen an:

- σx: die aus dem Datensatz errechnete Standardabweichung
- sx: eine aus der Analyse des Datensatzes empirisch ermittelte Standardabweichung

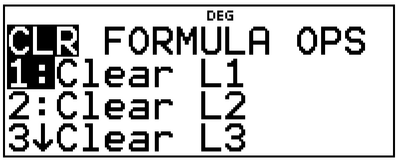
Aufrufen der Listen zur Eingabe von Datensätzen über die Taste data . Es öffnet sich ein Bildschirm mit drei Spalten (Listen).



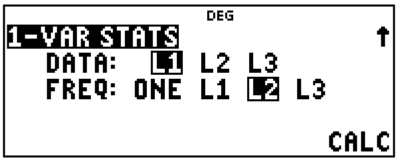
In die erste Spalte gibt man die Daten ein, in die zweite die jeweilige Häufigkeit.



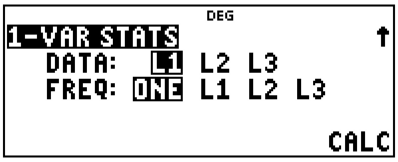
Das Löschen der Daten erfolgt durch erneutes Betätigen der Taste data und Auswahl jener Listen, deren Inhalt gelöscht werden soll.



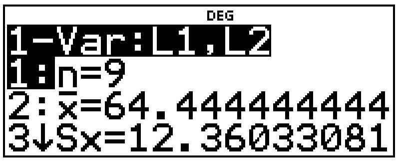
Zur Ausgabe der Kenngrößen gelangt man über 2nd stat-reg . Man wählt die Option 2:1-VAR STATS und dort unter DATA die Liste aus, in welcher die Daten stehen und unter FRQ die Liste aus, welche die entsprechenden Häufigkeiten enthält.



Es gibt auch die Möglichkeit, den gesamten Datensatz ohne Häufigkeiten einzugeben, dann ist unter FRQ die Option ONE zu wählen.



Bestätigen von CALC liefert dann die Kenngrößen, die jeweils durch Betätigen der Eingabetaste Enter in Liste 3 gespeichert werden können.



1:n	Gesamtzahl der Daten
2:$\bar{x}$	Mittelwert
3:sx	Standardabweichung (empirisch ermittelt)
4:σx	Standardabweichung(aus Datensatz ermittelt)
5:$\sum x$ 6:$\sum x^2$	Summe aller Daten Summe aller Datenquadrate
7:min(X)	Minimum
8:Q_1	unteres Quartil
9:Med	Median
…:Q_3	oberes Quartil
…:max(X)	Maximum