Kennzahlen der beschreibenden Statistik C2BL

Das Skript ist so gestaltet, dass es sowohl mit einem herkömmlichen wissenschaftlichen Taschenrechner als auch computergestützt mit einem Tabellenkalkulationsprogramm bearbeitet werden kann. Beide Methoden ergänzen sich sinnvoll, wenn man zuerst mit dem Taschenrechner arbeitet und dann die Aufgaben anschließend computergestützt bearbeitet.

Die ***beschreibende (deskriptive) Statistik*** hat zum Ziel, Daten durch Tabellen, Kennzahlen und Grafiken zusammenfassend darzustellen. Im Gegensatz zu dieser beschreibenden Aufgabe gibt es auch die ***schließende Statistik***. Hier werden die Daten mit statistischen Methoden interpretiert und Schlüsse gezogen. **Beispiel:** Die Erythrozytenkonzentration zweier Blutproben wurde jeweils durch mehrmalige Messungen bestimmt. Die Mittelwerte der beiden Proben unterscheiden sich geringfügig voneinander. Handelt es sich hierbei um einen zufälligen Unterschied, der darauf basiert, dass nur Stichproben gezogen und nicht das gesamte Blut untersucht wurden? Sind also, wenn man das gesamte Blut untersuchen würde, die wahren Mittelwerte identisch? Mit Methoden der *schließenden Statistik* kann man beurteilen, ob es tatsächlich einen signifikanten Unterschied zwischen den beiden Blutproben gibt.

Beispielaufgabe

Zur Analyse der innerartlichen Variabilität, wurden die Fruchtgewichte einer Pflanzenart an 25 Exemplaren gemessen.

Messwerte (x_i) der Fruchtgewichte in Gramm (g), Stichprobenumfang n = 25								
4,3	4,1	4,8	4	4,6	4,9	5,1	4,1	4,3
3,7	4,8	4,4	4,8	4,9	4,6	4,1	4,7	4,4
4,3	4,9	4,3	5,3	4,3	4,1	4,8		

1. Arithmetisches Mittel

Der Wert $\bar{x}=\frac{1}{n}\cdot\sum x_i$, d.h. die Summe aller Messwerte xi (x1+ x2+ x3+ ...) geteilt durch die Anzahl n aller Messwerte, heißt arithmetisches Mittel.

1.1 *Berechnen Sie das arithmetische Mittel der Fruchtgewichte mit dem Taschenrechner. Rechenhilfe:* $\sum_{i=1}^{i=25} x_i=112{,}6$

****1.2:*** *Berechnen Sie das arithmetische Mittel der Fruchtgewichte mit dem LibreOffice CALC mithilfe der Funktion MITTELWERT. Wie verhält sich die Funktion MITTELWERT, wenn einer der Werte keine Zahl ist? Beschrieben Sie die Auswirkung.*

2. Gewogenes arithmetisches Mittel

Hat man mehrere Stichproben durchgeführt, so kann man aus den einzelnen Mittelwerten einen gemeinsamen Mittelwert bilden. Dabei gehen die Stichprobenmittelwerte entsprechend dem Stichprobenumfang mit unterschiedlicher Gewichtung in den Mittelwert ein:

$$\bar{\bar{x}}=\frac{1}{N}\cdot\sum_{i=1}^{k} n_i\cdot\bar{x}_i$$, wobei

N	Gesamtanzahl über alle Stichproben
$\bar{x}_i$	arithmetisches Mittel der i-ten Stichprobe
ni	Stichprobenumfang der entsprechenden Stichprobe
k	Anzahl der Stichproben

2.1 Für drei Getreidesorten wurde der Ertrag ermittelt. Für Sorte A liegen n1 = 3, für Sorte B n2 = 4 und für Sorte C n3 = 3 Werte vor. Für jede Sorte wurde das arithmetische Mittel ($\bar{x}_i$) gebildet.

Sorte	*i = 1*	*i = 2*	*i = 3*
Stichprobenumfang n_i	*3*	*4*	*3*
Sortenmittelwert $\bar{x}_i$	*2,5*	*1,7*	*1,8*

Berechnen Sie das gewogene arithmetische Mittel mit der Formel und dem Taschenrechner:

3. Modalwert

Der **Modus** oder **Modalwert (D)** ist derjenige Wert, der in einer Beobachtungsreihe am häufigsten auftritt. Ein großer Vorteil ist, dass er nicht nur bei quantitativen (numerischen) Daten, sondern auch bei qualitativen Daten (z.B. Zellart, Wortart) angewendet werden kann. Kommt jeder Wert nur einmal vor, so gibt es keinen Modalwert. Am häufigsten treten monomodale Verteilungen der Stichprobenwerte auf, seltener gibt es zwei (*bimodal*) oder mehrere Werte (*polymodal*) mit auffallenden Häufigkeiten. Die Berechnung ist nur bei umfangreichen Stichproben sinnvoll. Beispiel für einen qualitativen Modalwert:

Leukozytenart	Anzahl (Strichliste)	Summe
Stabkernige neutrophile Granulozyten	////	4
Segmentkernige neutrophile Granulozyten	///// ///// ///// ///// ///// ///// ///// ///// ///// ///// ///// ////	59
Eosinophle Granulozyten	////	4
Basophile Granulozyten	/	1
Lymphozyten	///// ///// ///// ///// ///// /	26
Monozyten	///// /	6

3.1 Geben Sie die beiden Modi (D) bei der Annahme einer bimodalen Verteilung an.

D1 = ... D2 = ...

***3.2:** HINWEIS: Die Funktion MODALWERT kann in LibreOffice CALC nur auf Zahlen angewendet werden. Berechnen Sie damit den Modus/Modalwert bei den Fruchtgewichten. Geben Sie rechts daneben die Häufigkeit des Wertes an, indem Sie mit dem Funktionsassistent die Funktion ZÄHLENWENN nutzen

D = ... Häufigkeit = ...

4. Median

Der **Median (Z)** halbiert die nach der Größe sortierte Folge der Einzelwerte, so dass gleich viele Messwerte unterhalb und oberhalb von Z liegen. Handelt es sich um eine ungerade Anzahl von Messwerten, so entspricht Z dem mittleren Wert. Handelt es sich um eine gerade Anzahl von Messwerten, so muss aus den beiden in der Mitte liegenden Messwerte, das arithmetische Mittel gebildet werden.

***4.1** Ermitteln Sie den Median folgender Klassenarbeitsnoten: 5; 3; 4; 2; 4; 1; 3 sortiert: Z =*

***4.2** Ermittelt Sie den Median folgender Klassenarbeitsnoten: 3; 4; 4; 2; 5; 2 sortiert: Z =*

***4.3** Ermittelt Sie den Median folgender Klassenarbeitsnoten: 3; 4; 4; 2; 6; 2. sortiert: Z =*

***4.4** Anhand der Erkenntnisse aus 4.2 und 4.3: Welchen Vorteil besitzt der Median gegenüber dem arithmetischen Mittel?*

***4.5** Ermitteln Sie mit der entsprechenden Funktion, den MEDIAN der Fruchtgewichte. Z =

5. Harmonisches Mittel ($\overline{x}_H$)

Neben dem arithmetischen Mittel (*siehe oben*), gibt es noch andere Arten der Mittelwertbildung.

Das harmonische Mittel ist der *Kehrwert des Arithmetischen Mittels der Kehrwerte:*

$$\overline{x}_{harm} = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \dots + \frac{1}{x_i}} = \frac{n}{\sum \frac{1}{x_i}}$$

***5.1** In einem Tierversuch an 8 Tieren wurde nach Verabreichung eines Giftstoffs folgende Überlebenszeiten registriert:*

33 Tage; 24 Tage; 16 Tage; noch am Leben; 28 Tage; noch am Leben; 25 Tage; 12 Tage

a) Berechnen Sie die harmonisch gemittelte Überlebenszeit der Ratten.

b) Warum kann das arithmetische Mittel zur Mittelwertbildung hier nicht eingesetzt werden?

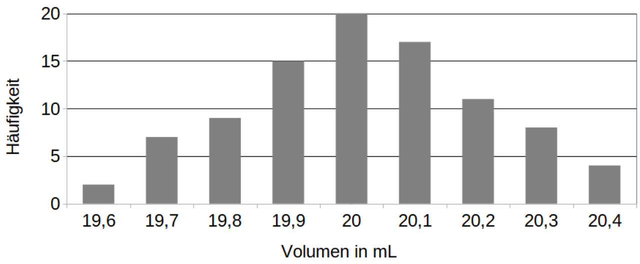
c) Welche weitere der oben beschriebenen Lagemaße, könnten noch angewendet werden? Geben Sie die Ergebnisse an.

***d)** Berechnen Sie das harmonische Mittel mit der entsprechenden Funktion von LO Calc. Hinweis: Da es „noch am Leben“ bzw. ∞ nicht als Zahl gibt, ersten sie die Einträge gegen eine sehr sehr große Zahl.*

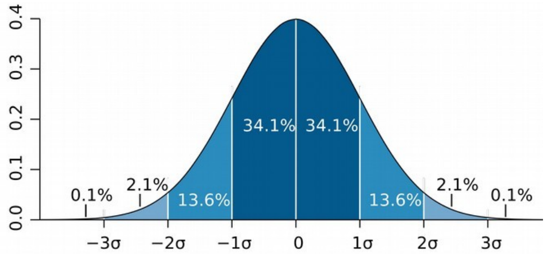
6. Normalverteilung und Standardabweichung

n = 93 Auszubildende aus Laborberufen sollten mit der Vollpipette 20 mL H_2O pipettieren. Durch Auswaage auf der Feinwaage konnte das pipettierte Volumen bestimmt und auf auf eine Nachkommastelle gerundet werden. Er ergaben sich folgende Häufigkeiten. Die Abbildung rechts gibt das Säulendiagramm an.

pipettiertes Volumen in mL	Häufigkeit
19,6	2
19,7	7
19,8	9
19,9	15
20,0	20
20,1	17
20,2	11
20,3	8
20,4	4



Würde man einen viel größeren Umfang an Proben nehmen und auch die tatsächlich pipettierten Volumina genauer auflösen (z.B. auf 3 Nachkommastellen) käme es zu einer Glättung des Säulendiagramms zu einer Kurve. Auch bei sehr vielen anderen statistischen Auswertungen entsteht dabei eine typische Glockenkurve, die Normalverteilung heißt:



Beschreibung der Abb. links: Grafische Darstellung der Normalverteilung. Auf der y-Achse ist die Häufigkeit der Messwerte dargestellt, auf der x-Achse die Abweichung des Messwertes zum tatsächlichen Wert bzw. Mittelwert (hier: $x=\bar{x}=0$). Die Summe aller Häufigkeiten entspricht der Gesamtfläche der Kurve und beträgt 1 (100%). Q: www.wikipedia.de

Die Fläche der Glockenkurve, die zwischen den beiden Wendepunkten (-1σ bis 1σ) beträgt ca. 68,2% (= 34,1% + 34,1%, siehe Abb.) der Gesamtfläche. Nimmt man eine ideale Normalverteilung von Messwerten an, so liegen ca. 68,2% der Messwerte im Intervall $\bar{x}\pm\sigma$. Der Abstand dieser Intervallgrenzen zum Mittelwert wird Standardabweichung (σ oder s) genannt. Verdoppelt man das Intervall, so beträgt die Fläche unterhalb der Kurve ca. 95,4% (= 34,1% + 34,1% + 13,6% + 13,6%). Das bedeutet im Intervall $\bar{x}\pm 2s\ (\text{oder}\ \bar{x}\pm 2\sigma)$, bei Annahme einer Normalverteilung der Messwerte ca. 95,4 % der Messwerte in diesem Wertebereich zu finden sind. Die Standardabweichung ist also ein Maß für die Breite der Normalverteilung. Sie gibt damit an, wie sehr die einzelnen Messwerte rund um den Mittelwert streuen. Sie lässt sich bei Stichproben durch folgende Formel berechnen:

$$s=\sqrt{\frac{\sum (x_i-\bar{x})^2}{n-1}} \quad \text{oder} \quad s=\sqrt{\frac{\sum (x_i^2)-\frac{\left(\sum x_i\right)^2}{n}}{n-1}}$$

s: Standardabweichung der Stichpr. **x_i:** Einzelwerte x_1, x_2 etc.
$\bar{x}$: Mittelwert der Stichpr. **n:** Stichpr.-Umfang

<u>MERKE</u> Bei der Stichprobenstandardabweichung steht im Nenner n-1, deshalb wird sie häufig mit σ(n-1) ODER s_x abgekürzt.

<u>ACHTUNG. VERWECHSLUNGSGEFAHR:</u> Die Formel zur Standardabw. von <u>Grundgesamtheiten (σ)</u>, liefert falsches Ergebnis!

Der ***Variationskoeffizient(=relative Standardabweichung***, ***s% oder VK)*** ist prozentuale Standardabweichung bezogen auf den Mittelwert ($\bar{x}$): **VK = $s_{\%}$ = • 100% • s/$\bar{x}$**

***6.1** Die Stichprobe von Zeugnisnoten sind: 2, 3, 2, 1, 4 Berechnen Sie Standardabweichung und Variationskoeffizient (VK).*
***a)** mit Taschenrechner und Formeln* ***b)** mit Taschenrechner und Tipp (s.u.)* **c)** mit CALC: Funktionen STABW und ANZAHL.

TIPP: Mit einem Taschenrechner kann man s_x (σ(n-1) in Sekundenschnelle berechnen und Prüfungszeit einsparen!

CASIO-Modelle: z.B. fx991 oder ClassWiz-Modelle und NoName-Klone:

- ① **Daten eingeben**. Je nach Modell $\boxed{\text{MODE}}$ oder $\boxed{\text{MENU}}$ drücken → *„STAT“* wählen → *„1-VAR“* (Statistiken mit 1 Variable, also keine Werte*paare*)→ Werte in (X)-Spalte eingeben. Dateneingabe durch drücken von $\boxed{\text{AC}}$ (= „all clear“) beenden.
- ② **Statistik-Funktionen aufrufen:** $\boxed{\text{SHIFT}}$ $\boxed{1}$ („STAT“) **<u>ODER (je nach Modell)</u>** $\boxed{\text{OPTN}}$ → „1 Variable“ (1-VAR)→ **„<u>xσn-1</u>“** ODER s_x wählen (ACHTUNG! Nicht die FALSCHE Standardabweichung wählen! vgl. Bemerkung oben.) → $\boxed{=}$: Standardabweichung wird unter **„s_x“** ODER **„<u>x σn-1</u>“** angezeigt (evtl. scrollen!). Auch andere Größen der beschreibenden Statistik finden sich in der Auswahl, z.B. Mittelwert ($\bar{x}$).

weitere Casio-Modelle mit Zusatz MS, z.B. fx-82MS:

- ① **Daten eingeben**. $\boxed{\text{SHIFT}}$ $\boxed{\text{MODE}}$ „Scl“ auswählen und mit $\boxed{=}$ bestätigen (→ Statistikspeicher wird geleert). Jetzt Einzelwerte eintippen, jeweils anschließendes Drücken von $\boxed{\text{M+}}$ in Speicher überführen. Dateneingabe durch $\boxed{\text{AC}}$ beenden.
- ② **Statistik-Funktionen aufrufen:** $\boxed{\text{SHIFT}}$ $\boxed{2}$ („S-VAR“) → **„<u>xσn-1</u>“** ODER s_x wählen (ACHTUNG! Nicht die FALSCHE Standardabweichung wählen! vgl. Bemerkung oben.) → $\boxed{=}$: Standardabweichung durch **„s_x“** ODER **„<u>x σn-1</u>“** auswählbar

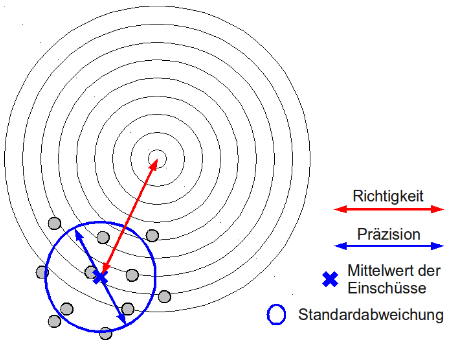
Texas Instruments z.B. TI30X Plus (u.v weitere TI-Modelle):

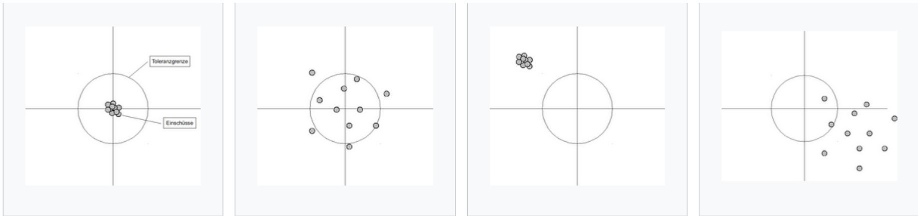
- ① **Daten eingeben:** → $\boxed{\text{data}}$. In Spalte *„L1“* eingeben. jew. mit „enter“ bestätigen) Evtl. vorhandene Daten können durch erneutes drücken von $\boxed{\text{data}}$ und *„clear“* gelöscht werden.
- ② **Statistik-Funktionen aufrufen:** $\boxed{\text{2nd}}$ $\boxed{\text{data}}$ („star-reg“) → „1-VAR *Stats*“ wählen → „L1“ und „FRQ: ONE“ (Häufigkeit überall 1) wählen, zu „CALC“ scrollen und $\boxed{\text{enter}}$. $\bar{x}$ und S_x ablesen. **Nicht die FALSCHE Standardabweichung „σ“ ablesen!**

Ihr Taschenrechner nicht dabei? ⇒ bei **youtube** unter Modellname suchen!

7. Systematischer und zufälliger Fehler, Richtigkeit und Präzision

- **Systematischer Fehler:** Messfehler der sich auch bei vielen (theoretisch: unendlich) wiederholten Messungen im Mittel nicht aufhebt. Häufig ist der wahre Wert nicht bekannt, so dass eine exakte Quantifizierung nicht möglich ist.
- **Zufälliger Fehler:** Messfehler die sich bei vielen (unendlich) wiederholten Messungen im Mittel ausgleicht.
- **Präzision (precision):** Ein Messung ist dann präzise, wenn die einzelnen Messwerte bei gleichbleibenden Untersuchungsobjekt nur wenig streuen. Eine quantitative Kennzahl für die Präzision ist die Standardabweichung und der Variationskoeffizient.
- **Richtigkeit (accuracy):** Ein Messung ist um so richtiger, je größer die Übereinstimmung zwischen dem Mittelwert mit dem Erwartungswert (wahrer Wert) übereinstimmt. Häufig ist der wahre Wert nicht bekannt, so dass eine exakte Quantifizierung nicht möglich ist.





Richtigkeit und Präzision gut (gute Genauigkeit) | Richtigkeit gut, aber schlechte Präzision (schlechte Genauigkeit) | Präzision gut, aber schlechte Richtigkeit (schlechte Genauigkeit) | Präzision und Richtigkeit schlecht, die Genauigkeit natürlich auch

Quelle beider Bilder: www.wikipedia.de

***7.1** In einem Versuch über die Genauigkeit von Volumenmessgeräten wurden 15 mal 20 mL Wasser mit der Vollpipette und mit dem Messzylinder pipettiert und die Masse des Wassers auf der Feinwaage gemessen. Folgende Werte wurden erhalten.*

	Messzylinder	**Vollpipette**
Mittelwert (aus 15 Messungen)	19,7737 g	19,7210 g
Standardabweichung	0,1984 g	0,0556 g
Variationskoeffizient	1,00%	0,28%

Diskutieren Sie folgende Fragen mithilfe der oben gegebenen Definitionen!

- *Welches Messgerät war präziser?*

- *Warum weichen die Mittelwerte von 20,00 g ab?*

- *Welche Aussage macht die Standardabweichung, welche Aussage die relative Standardabweichung?*