

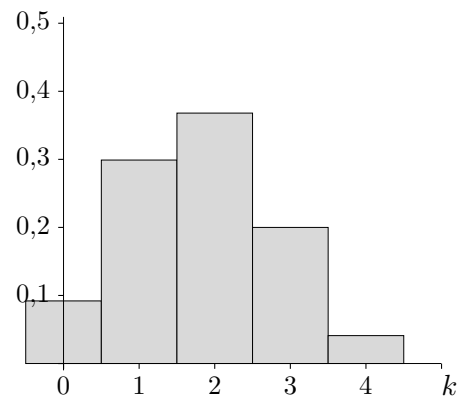
Standardabweichung Daten analysieren

Die Tabelle enthält die Häufigkeit der Krankmeldungen innerhalb eines Jahres der Arbeitnehmer eines Betriebes.

k	0	1	2	3	4
<i>absolute Häufigkeit</i>	92	299	368	200	41

Berechne die relativen Häufigkeiten (Prozentschreibweise) und stelle die Daten grafisch dar.

k	0	1	2	3	4
<i>relative Häufigkeit</i>	9,2%	29,9%	36,8%	20,0%	4,1%



Der Mittelwert \bar{x} beträgt 1,8,
die mittlere lineare Abweichung 0,81
und die Standardabweichung $\sigma = 1,0$.

Die Standardabweichung ist die Wurzel aus der mittleren quadratischen Abweichung.

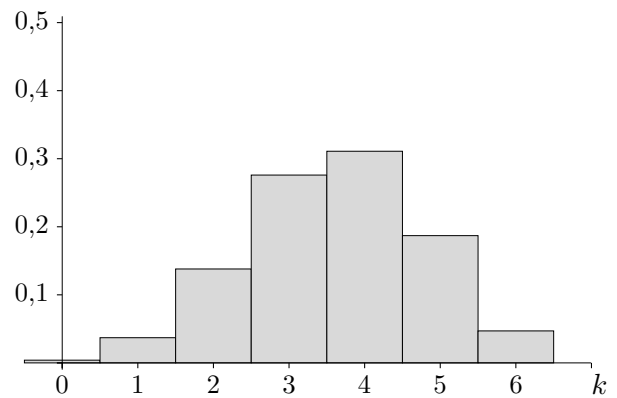
k	0	1	2	3	4
<i>relative Häufigkeit</i>	9,2%	29,9%	36,8%	20,0%	4,1%
<i>Abweichung vom Mittelwert $\bar{x} - k$</i>	1,8	0,8	-0,2	-1,2	-2,2
<i>quadratische Abweichung vom Mittelwert $(\bar{x} - k)^2$</i>	...				

Führe die Berechnungen sowie die grafische Darstellung auch für folgende Daten aus:

k	0	1	2	3	4	5	6
<i>absolute Häufigkeit</i>	4	37	138	276	311	187	47

Daten analysieren Lösung

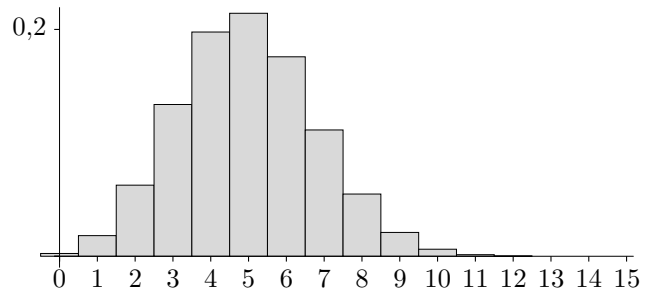
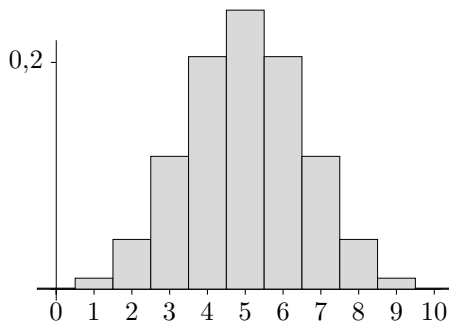
k	0	1	2	3	4	5	6
absolute Häufigkeit	4	37	138	276	311	187	47



Der Mittelwert \bar{x} beträgt 3,6,
die mittlere lineare Abweichung 1,01
und die Standardabweichung $\sigma = 1,2$.

Es geht jedoch auch einfacher mit: STAT | CALC 1: 1-Var Stats L1, L2

Standardabweichung



Die beiden Wahrscheinlichkeits-Verteilungen haben denselben Erwartungswert $\mu = 5$, jedoch sind sie unterschiedlich um den Erwartungswert konzentriert. Dieser Streuung wollen wir uns zuwenden.

Sei die Verteilung ($n = 4$)

k	a_1	a_2	a_3	a_4
$P(X = k)$	p_1	p_2	p_3	p_4

gegeben. Dann ist der Erwartungswert

$$\mu = E(X) = a_1 p_1 + a_2 p_2 + a_3 p_3 + a_4 p_4.$$

Die Abweichung von a_1 zum Erwartungswert beträgt $|a_1 - \mu|$, sie wird mit der Wahrscheinlichkeit p_1 angenommen. Der Erwartungswert für die Abweichung ist dann

$$d = |a_1 - \mu| p_1 + |a_2 - \mu| p_2 + |a_3 - \mu| p_3 + |a_4 - \mu| p_4$$

Mit der Umformung $|a_1 - \mu| = \sqrt{(a_1 - \mu)^2}$ können die Beträge ersetzt werden, dies ergibt:

$$d = \sqrt{(a_1 - \mu)^2} \cdot p_1 + \sqrt{(a_2 - \mu)^2} \cdot p_2 + \sqrt{(a_3 - \mu)^2} \cdot p_3 + \sqrt{(a_4 - \mu)^2} \cdot p_4$$

Weit bedeutsamer als Maß für die Streuung ist die sogenannte Standardabweichung:

$$\sigma = \sqrt{(a_1 - \mu)^2 \cdot p_1 + (a_2 - \mu)^2 \cdot p_2 + (a_3 - \mu)^2 \cdot p_3 + (a_4 - \mu)^2 \cdot p_4}$$

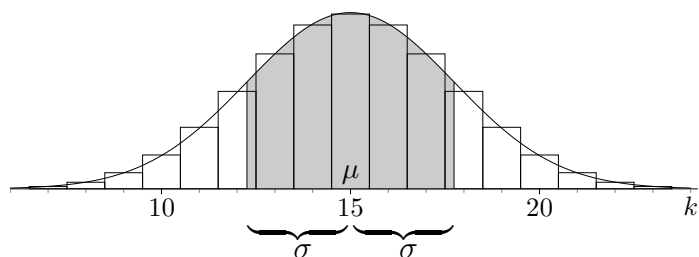
Hierdurch liegt keine Ungenauigkeit ($d \neq \sigma$) vor, vielmehr wird mit σ ein eigenständiger Abstand zum Erwartungswert festgelegt, der eine anschauliche Bedeutung besitzt.

$$n = 30$$

$$p = 0,5$$

$$\sigma = \sqrt{n \cdot p \cdot q}, \quad \mu = n p$$

(z. B. 30-maliges Werfen einer Münze, X Anzahl von "Zahl")



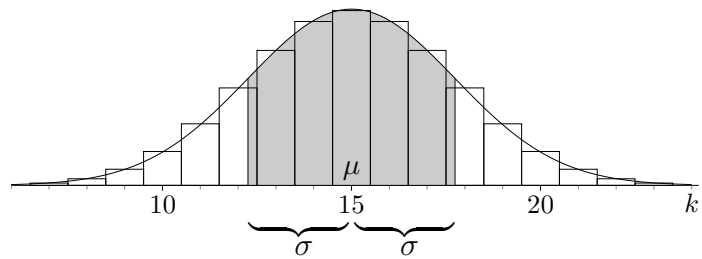
Wird durch das Histogramm einer Binomialverteilung eine Ausgleichskurve gelegt (Gauß'sche Glockenkurve), so gibt σ , wie wir später sehen werden, die Entfernung der Wendepunkte der Kurve vom Erwartungswert μ an. σ kann zudem leicht bestimmt werden: $\sigma = \sqrt{n \cdot p \cdot (1 - p)}$.

Standardabweichung Fortsetzung

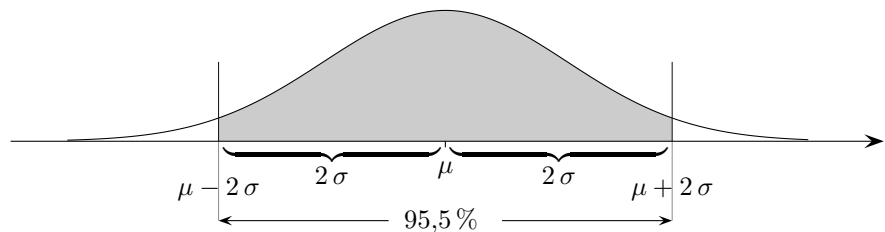
$$n = 30$$

$$p = 0,5$$

$$\sigma = \sqrt{n \cdot p \cdot q}, \quad \mu = np$$



Es lässt sich unter anderem zeigen, dass für Binomialverteilungen (falls $\sigma > 3$) das Stichprobenergebnis mit 95,5%-iger Wahrscheinlichkeit innerhalb der 2σ -Umgebung liegt und mit 99,7%-iger Wahrscheinlichkeit innerhalb der 3σ -Umgebung. Liegt ein Stichprobenergebnis außerhalb der 2σ -Umgebung, so bestehen ernsthafte Zweifel an der zugrundeliegenden Wahrscheinlichkeit p .



Als Varianz wird

$$V(X) = (a_1 - \mu)^2 \cdot p_1 + (a_2 - \mu)^2 \cdot p_2 + (a_3 - \mu)^2 \cdot p_3 + (a_4 - \mu)^2 \cdot p_4$$

bezeichnet. Dann ist

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

Zeige für $n = 4$: $V(X) = a_1^2 p_1 + a_2^2 p_2 + a_3^2 p_3 + a_4^2 p_4 - \mu^2 \quad (= E(X^2) - \mu^2)$

Berechne damit die Standardabweichung

k	0	1	2	3
$P(X = k)$	27,46%	44,36%	23,89%	4,29%

Lösung:
 $\sigma = 0,826$

Roofls

Standardabweichung GTR

Sei die Verteilung ($n = 4$)

k	a_1	a_2	a_3	a_4
$P(X = k)$	p_1	p_2	p_3	p_4

gegeben. Dann kann der Erwartungswert

$$\mu = E(X) = a_1p_1 + a_2p_2 + a_3p_3 + a_4p_4.$$

mit dem GTR bestimmt werden.

In die Listen L1 und L2 werden die Werte a_1, \dots und p_1, \dots eingetragen und `2nd LIST | MATH 3: mean(L1, L2)` aufgerufen.

L1	L2	L3
a_1	p_1	
a_2	p_2	
...	...	

Um die Standardabweichung zu ermitteln, wird zunächst die sogenannte Varianz bestimmt und aus diesem Wert die Wurzel gezogen.

$$V(X) = (a_1 - \mu)^2 \cdot p_1 + (a_2 - \mu)^2 \cdot p_2 + (a_3 - \mu)^2 \cdot p_3 + (a_4 - \mu)^2 \cdot p_4$$

Die Varianz ist der Erwartungswert der quadratischen Abweichungen $(a_1 - \mu)^2, \dots$. Daher werden diese Werte in L3 mit $(L1 - \mu)^2$ eingefügt und `2nd LIST | MATH 3: mean(L3, L2)` aufgerufen.

L1	L2	L3
a_1	p_1	
a_2	p_2	
...	...	

↙ $(L1 - \mu)^2$

Es geht jedoch auch einfacher mit: `STAT | CALC 1: 1-Var Stats L1, L2`

Beispiel:

k	0	1	2	4
$P(X = k)$	$\frac{3}{7}$	$\frac{1}{7}$	$\frac{1}{7}$	$\frac{2}{7}$

$$E(X) = 1,57$$

$$V(X) = 2,82$$

$$\sigma = 1,68$$