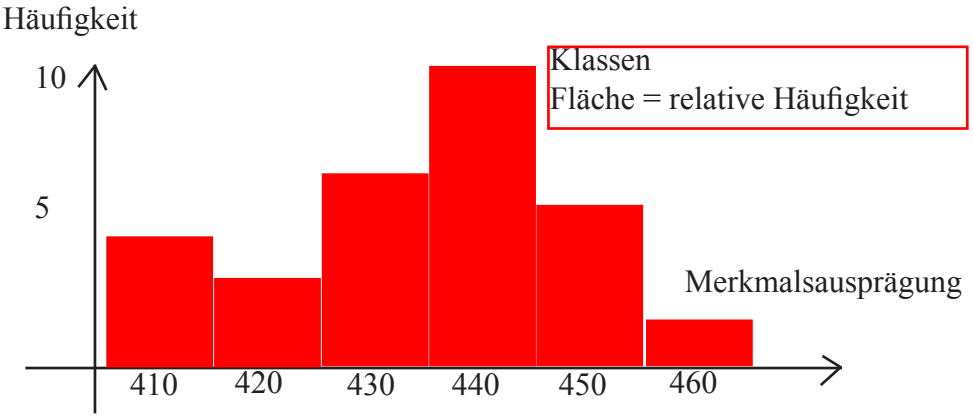
Merkmalsausprägung x_i	410	420	430	440	450	460
Häufigkeit n_i (wieviel mal)	4	3	6	10	5	2



TI - 89:
APPS -> DATA MATRIX EDITOR -> NEW
Typ : Data
C1 : Merkmalsausprägung x;
C2 : alle n_i Werte;
F5 ->
Calculation Types : one Var
x : C1
Frequent Category: Yes
Frequ : C2

bei relativer Häufigkeit:
y Achsen Höhe =
Höhe der Klasse in % oder Merkmalsausprägung / Breite der Klasse in x

$$y_i = a + b * x_i$$
$$\bar{y} = a + b * \bar{x}$$

Stichprobenumfang n
absolute Häufigkeit n_i : $n_1 = 4,\ n_2 = 3....$
relative Häufigkeit $h(x_i) = \frac{n_i}{n} = \frac{4}{30}$

empirischer Mittelwert $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i =$

$$\frac{1}{30} * (4*410 + 3*420 + 6*430 + 10*440 + 5*450 + 2*460)$$

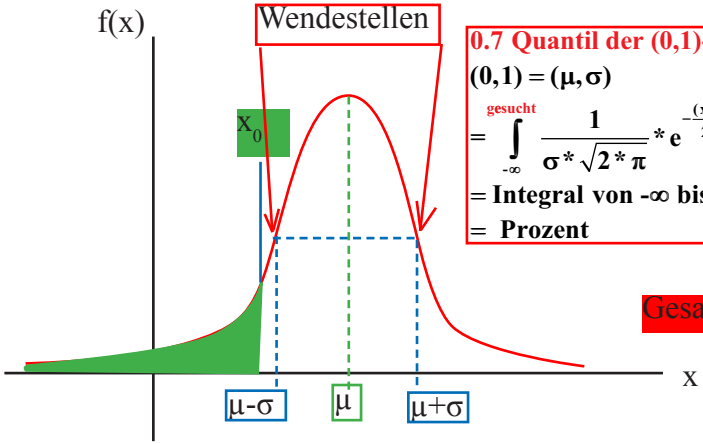
bei k Klassen $\bar{x} = \frac{1}{n}\sum_{i=1}^{k} n_i * x_i$ $\quad x_i$ **= Stabmitten**

empirische Standartabweichung $s = \sqrt{\frac{1}{n-1} * \sum_{i=1}^{k} n_i * (\bar{x} - x_i)^2}$

Varianz $= s^2$

Dichtefunktion, Glockenkurve, Gaussche Kurve oder Normalverteilung:

Fläche entspricht der relativen Häufigkeit:
Integration von $-\infty$ bis x_0



0.7 Quantil der (0,1)-Normverteilung:

$$(0,1) = (\mu,\sigma) = \int_{-\infty}^{\text{gesucht}} \frac{1}{\sigma * \sqrt{2*\pi}} * e^{-\frac{(x-\mu)^2}{2*\sigma^2}} * dx = 0.7$$

= Integral von $-\infty$ bis gesucht = Fläche von 0.7
= Prozent

Gesamtfläche immer 1

$$f(x) = \frac{1}{\sigma * \sqrt{2*\pi}} * e^{-\frac{(x-\mu)^2}{2*\sigma^2}}$$

relative Häufigkeit = Fläche unter der Kurve = 1:

f(x) hat das absolute Maximum bei $x = \mu$

$f(\mu) = \frac{1}{\sigma * \sqrt{2*\pi}}$ **d.h der Maximalwert wird keiner, wenn σ erhhöht wird**

f(x) hat Wendestellen bei $x = \mu \pm \sigma$ $f''(x)=0$ $f'''(x) \neq 0$

$$\text{Mittelwert} = \bar{x} = \mu = E(X) = \int_{-\infty}^{\infty} x * f(x) * dx$$

$$\text{Standartabweichung} = s = \sigma = V(X) = \int_{-\infty}^{\infty} (x - E(X))^2 * f(x) * dx$$

$$\text{Varianz} = v = \sqrt{V(X)}$$

relative Häufigkeit = Fläche unter der Kurve = 1:
Ti 89 hat Numerikprobleme:
nSolve = Solve
nInt = ∫

$\phi < 0.5 \quad \rightarrow \quad \phi = 1 - \phi$

$\phi(u) = 0.008 \quad \rightarrow \quad \phi(-u) = 1 - 0.008$

$\phi(-0.34) = 1 - \phi(0.34)$

$$u = \frac{x_0\text{-Mittelwert}}{\text{Standartabweichung}} = \frac{x_0 - \mu}{\sigma}$$

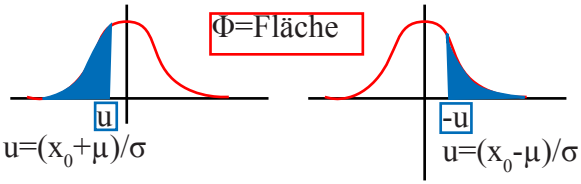
für negative u $\Phi(u) = 1 - \Phi(-u)$

falls nicht in Liste, $\Phi = 0$ oder $\Phi = 1$

1.62

1.6 = Zeile

2 = Spalte



	0	1	2	3	4	5	6	7	8	9
0	0.5	0.504	0.508	0.512	0.516	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.591	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.648	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.67	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.695	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.719	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.758	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.791	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.834	0.8365	0.8389
1.	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.877	0.879	0.881	0.883
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.898	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.937	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.975	0.9756	0.9761	0.9767
2.	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.983	0.9834	0.9838	0.9842	0.9846	0.985	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.989
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.992	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.994	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.996	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.997	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.998	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.999	0.999
3.1	0.999	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998
3.5	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998
3.6	0.9998	0.9998	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999

Statistik / Stochastik

Zufallsversuch -> Das werfen einer Münze

Stichprobenraum Ω -> Die Menge aller möglichen Ausgänge, oder Ergebnisse eines Zufallsversuches

Ausfall ω -> Ein Ergebnis

Ereignis -> Teilmenge des Stichprobenraumes Ω

Beispiel Würfel:

$\omega_1 = \text{"Kopf"} \quad \omega_2 = \text{"Zahl"}$

$\Omega = \{\text{"Kopf","Zahl"}\}$

Ereignis A={2,4,6}

Assitionsregel=p(A)= $\sum_{\omega \in A} p(\omega)$

Gegenereignis $= p\left(\overline{A}\right) = 1 - p(A)$

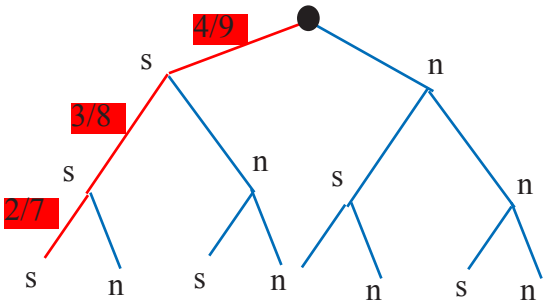
Laplace – Gerät -> Zufallsgerät, bei dem alle Ausfälle dieselbe Wahrscheinlichkeit besitzen

$p(A) = \frac{\text{Anzahl günstige Fälle}}{\text{Anzahl mögliche Fälle}}$

Mehrstufige Zufallsversuche:

In einem Baum gilt:
Die Wahrschienlichkeit eines Pfades ist gleich dem Produkt aller Wahrscheinlichkeiten längs des Pfades

Reisegruppe mit 4 Schmugglern und 5 ehrliche Leute. Wie gross ist die Wahrscheinlichkeit, dass von 3 Stichproben 3 Schmuggler ertappt werden?
Schmuggler = s
nicht Schmuggler = n



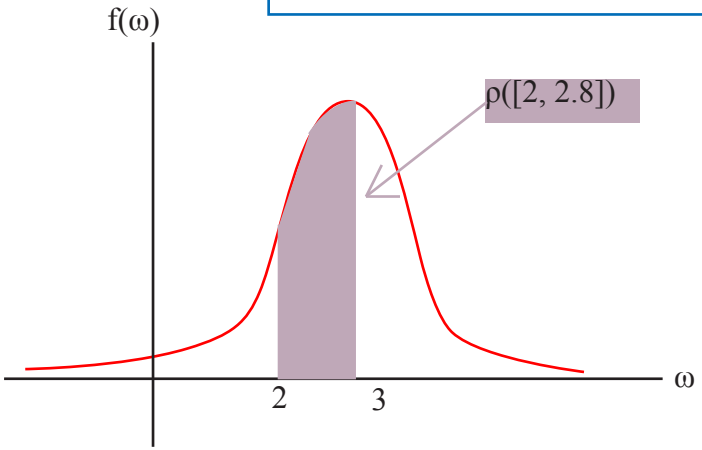
3 mal Schmuggler hintereinender= (4/9)*(3/8)*(2/7)=1/21

Wahrheitstabelle:

ω	sss	ssn	snn	sns.....
ρ(ω)	1/21	...	...	...

← Quersumme = 1

Die Funktion f(ω) entspricht also dem Begriff der Dichtfunktion = Wahrschienlichkeit



Bei Binomenalverteilung:

$$E(X) = n * p$$

$$V(X) = n * p * (1-p)$$

6* Würfel, 6 verschiedene Zahlen:

1-2-3-4-5-6 Pfad egal

$$1 * \frac{5}{6} * \frac{4}{6} * \frac{3}{6} * \frac{2}{6} * \frac{1}{6}$$

bei 12 Würfeln genau 6*eins:

$$p = \binom{12}{6} * \left(\frac{1}{2}\right)^6 * \left(\frac{1}{2}\right)^6 \quad \binom{n}{k} * p^k * (1-p)^{n-k}$$

Aus 12 Plätzen 6 wählen

Geordnete Stichproben ohne Zurücklegen:

Aus n Objekten sind k herauszugreifen und in einer Folge anzuordnen.
Wieviele Möglichkeiten gibt es??
Anzahl Möglichkeiten des Versuches = Anzahl Pfade = x

$$p(\Omega) = 1$$

$$\frac{x * (n-k)!}{n!} = 1$$

$$x = \frac{n!}{(n-k)!}$$

-> Es gibt $\frac{n!}{(n-k)!}$ Möglichkeiten aus n Objekten k herauszugreifen und anzuordnen

Permutationen :

Eine Anordnung von n Kugeln (allgemein : Elemente) in einer bestimmten Reihenfolge heisst Permutation. Für die Anzahl der möglichen Permutationen gilt dann:

1.Alle n Kugeln sind voneineander verschieden:

$$p(n) = n!$$

2.Ungeordnerter Fall (Reihenfolge egal)

totale Anzahl ungeordnete Falle $x * k! = \frac{n!}{(n-k)!}$

$$x = \frac{n!}{(n-k)! * k!} = \binom{n}{k}$$ lies "n tief k" (Binomenalkoeffizient)

->Es gibt $\binom{n}{k}$ Möglichkeiten aus einer Menge mit n Elementen eine Teilmenge von k Elementen herauszugreifen

3.Unter den n Kugeln befinden sich jeweils $n_1, n_2, ..., n_k$ einander gleiche:

$$p(n;\ n_1, ..., n_k) = \frac{n!}{n_1! n_2! ... n_k!}$$

günstige Fälle:

10 Transistren, 4 defekt, 3nehmen
die Wahrscheinlickeit, dass 2 defekt sind unter den 3 genommenen:

$$\binom{4}{2} * \binom{6}{1} = \binom{\text{4 defekte}}{\text{2 davon}} * \binom{\text{6 gute}}{\text{1 genommen}}$$

bei mehreren Stufen: $\binom{n}{k} * \binom{n_2}{k_2}$

bei n=k: n!

bei Mannschaften: /2

Beispiel Geordnete Stichprobe ohne Zurücklegung:
Aus n Objekten sind k herauszugreifen und in einer Folge anzuordnen.

Beispiel ungeordnete Stichprobe ohne Zurücklegung:
Wieviele Möglichkeiten gibt es, in einem Verein mit 70 mitgliedern einen fünfköpfigen Vorstand zu wählen?
Oder:
Im Lotto einen viere erzielen....
Oder:
Frau Maier will ihre 5 Kinder in einer Reihe anordnen. Auf wieviel Arten kann sie das tun?

Es gibt p(3)=3! = 6 verschiedene Möglichkeiten, 3 verschiedenfarbige Kugeln anzuordnen:

In einer Urne befinden sich 5 Kugeln, 3 weisse und 2 rote. Sie lassen sich auf

$$\mathbf{p(5;\ \ 3,2) = \frac{5!}{3!2!} = 10}$$

verschiedene Arten anordnen

Beim Zahlenlotto werden aus 45 Zahlen 6 gezogen:

1.Wieviele Möglcihkeiten gibt es, einen 4-er zu erzielen?

mache 2.Stufingen Versuch:

1.Stufe 4 Zahlen aus Urne 1 ziehen $\binom{6}{4}$ **Möglichkeiten**

2.Stufe 2 zahlen aus Urne 2 ziehen $\binom{39}{2}$ **Möglichkeiten**

Gesamt :

$$\binom{6}{4} * \binom{39}{2} \text{ Möglichkeiten} = \frac{6!}{(6-4)! * 4!} * \frac{39!}{(39-2)! * 2!}$$

2.Wie gross ist die Wahrscheinlichkeit für einen Vierer?

a) entweder über Bäume

$$\text{b)} p(\text{"vierer"}) = \frac{\text{Anzahl der günstigen Fälle}}{\text{Anzahl der möglichen Fälle}} = \frac{\text{Anzahl Vierer}}{\text{Anzahl der möglcihen Fälle}}$$

$$= \frac{\binom{6}{4} * \binom{39}{2} \text{ Möglichkeiten}}{\binom{4\square}{6}}$$

Produkteregel:

In einem Zweistufenversuch seien n_1 die Anzahl der Möglichkeiten wie die erste Stufe ausgehen kann und n_2 die Anzahl der Möglichkeiten für die zweite Stufe. Wenn man für jeden Ausgang der ersten Stufe alle Möglichkeiten der zweiten Stufe offen hat, so berechnet sich die totale Anzahl der Möglichkeiten des Gesamtversuchs als Produkt:

$n_1 * n_2$

Zufallsvariablen und ihre verteilung:

$$\sum_{i=1}^{n} p(X = x_i) * x_i$$

Wahrscheinlichkeit, Wert

Bsp.

Beim Würfeln wird bei geraden Zahlen das doppelte ausgezahlt, bei ungeraden verliert man das dreifache der Zahl

X(ω)=Gewinn bei Ausfall ω

Verteilung von X:

$X=x_i$	-3	4	-9	8	-15	12
$p(X=x_i)$	1/6	1/6	1/6	1/6	1/6	1/6

$$E(X) = (-3)*\frac{1}{6} + 4*\frac{1}{6} - 9*\frac{1}{6} + 8*\frac{1}{6} - 15*\frac{1}{6} + 12*\frac{1}{6}$$

$$= -\frac{1}{2}$$ **Durchschnittlicher Verlust pro Spiel ist 0.5 Fr., wenn man oft spielt**

Unter einem n-stufigen Bernoulli-Versuch versteht man die n-Fache Durchführung eines Zufalsversuchs, dessen Stichprobenraum nur aus 2 Elementen besteht:

$\Omega_0 = \{0,1\}$

Die 1 wird als Erfolg bezeichnet und

$p = p(1)$ **als Erfolgswahrscheinlichkeit**

Sei X die Anzahl der Erfolge ine einem n-Stufigen Bernoulliversuch.

Dann gilt für die Verteilungsfunktion

(k plätze aus n) bsp. von 12 Münzenwürfen (n), 3*Kopf(k) :

$$p(X = x_i) = p(X = k) = \binom{n}{k} * p^k * (1-p)^{n-k}$$

oder bei mehreren $$\sum_{k=anfang}^{ende} \binom{n}{k} * p^k * (1-p)^{n-k}$$

Bsp :

Eine Münze wird 120 mal geworfen. Wie gross ist die Wahrscheinlichkeit, dass zwischen 48 und 60 mal Kopf ergibt?

$$\sum_{k=48}^{60} \binom{120}{k} * 0.5^k * (1-0.5)^{120-k}$$

Statistik / Stochastik

Beschreibende Statistik:

Merkmalsausprägung x_i

relative Häufigkeit $\frac{n_i}{n}$

empirischer Mittelwert $\bar{x} = \frac{1}{n} * \sum_{i=1}^{k} n_i * x_i$

empirische Varianz $s^2 = \frac{1}{n-1} * \sum_{i=1}^{k} n_i * \left(\bar{x} - x_i\right)^2$

Wahrscheinlichkeitsrechnung:

Wert x_i einer Zufallsgrösse X

$p(X = x_i)$, X nimmt Wert x_i an

$E(X) = \sum_{k=1}^{n} p(X = x_i) * x_i$, falls nur n Werte

$\int_{-\infty}^{\infty} x * f(x) * dx$, falls die Werte von X alle relle Zahlen durchlaufen

Nimmt eine Zufallsvariable X nur endlich viele Werte an mit den Wahrscheinlichkeiten $p(X = x_i)$ und dem Erwartungswert $m = E(X)$, dann:

Varianz der Zufallsvariable X:

$V(X) = \sum_{i=1}^{n} (x_i - m)^2 * p(X = x_i)$ $\quad s = \sqrt{V(X)}$ heisst Standartabweichung von X

$V(X) = n * p * (1 - p)$

Anzahl Stufen in diesem Bernoulliveruch n

Erfolgswahrscheinlichkeit p

Füher n-stufigen Versuch durch-sagen wir N mal.

-> Grundversuch wird $n * N$ mal durchgeführt

-> habe ca. $n * N * p$ Erfolge erzielt

-> durchschnittliche Anzahl Erfolge pro Bernoulliversuch

» $n * p$ wird ums genauer, je grösser N

W = Die Menge aller möglichen Ausgänge $\quad W = \{0,1\}$

X = Anzahl Erfolge

$X = x_i$ Verteilung von X

$P(x = x_i)$ = Wahrscheinlichkeit

$E(X)$ = Durchschnittliche Anzahl Erfolge, wenn der Bernoulliversuch sehr oft durchgeführt wird

$= n * p$

man berechne die Wahrscheinlichkeit dafür, dass bei 120 Würfen mit einem Laplace-Würfel 19 oder 20 oder 21 mal die sechs erscheint:

$n * p * (1 - p) > 9$

1.Methode:

$$p(a \le X \le b) = \sum_{Anfang}^{Ende} \binom{n}{k} * p^k * (1-p)^{n-k} \quad = \quad \sum_{n=19}^{21} \binom{120}{k} * \left(\frac{1}{6}\right)^k * \left(\frac{5}{6}\right)^{120-k}$$

2.Methode :

$$\mu = n * p = 120 * \frac{1}{6} = 20$$

$$\sigma = \sqrt{n * p * (1-p)} = \sqrt{120 * \frac{1}{6} * \frac{5}{6}} = 4.0824$$

$$p(a \le X \le b) = \frac{1}{\sigma * \sqrt{2 * \pi}} * \int_{18.5}^{21.5} e^{-\frac{(x-\mu)^2}{2*\sigma^2}} * dx$$

Boissonverteilung nur, wenn n* gleicher Versuch mit gleicher Ausgangslage (Zurücklegung).

Anzahl Pfade $\binom{4}{2}$ **4*gezogen und 2*Ass**

$Y = X * \beta + \varepsilon$

Die binominalverteilte Zufallsvariable X mit

$n*p*(1-p)>9$:

$\mu = n*p$

und

$\sigma = \sqrt{n*p*(1-p)}$

ist für den Fall, wo $n*p*(1-p)>9$ ist, angenähert normalverteilt.

Eine Zufallsvariable X heisst normalverteilt, wenn sie jede reelle Zahl als Wert annehmen kann und für jedes Interval [a,b] gilt:

$$p(a \le X \le b) = \frac{1}{\sigma * \sqrt{2*\pi}} * \int_a^b e^{-\frac{(x-\mu)^2}{2*\sigma^2}} * dx$$

für eine normalverteilte Zufallsvariable X gilt:

$E(X) = \mu = n*p \qquad V(X) = \sigma^2$

$n*p*(1-p) \le 9$:

Faustregel : $(n>10 \text{ und } p<0.05)$ **oder** $(n \ge 100 \text{ und } p \le 0.1)$

Die Poissonverteilung ist, wenn $n*p*(1-p) \le 9$,

Die Wahrscheinlichkeit sehr klein ist, aber die Anzahl der betrachteten Ereignisse sehr gross. Man spricht von "seltenen Ereignissen".

Die binomenalverteilte Zufallsvariable X mit $E(X) = \mu = n*p$ gilt:

$V(X) = \mu$

$p(X) = k$ strebt mit wachsendem n einer festen Zahl zu:

$$p = \lim_{x \to \infty} \binom{n}{k} * p^k * (1-p)^{n-k} = \frac{\mu^k}{k!} * e^{-\mu}$$

$X = k$	0	1	2	3	...	k	...	¥
-----	-	-	-	----	-	-	-	-
$p(X=k)$				$\frac{m^k}{k!} * e^{-m}$				

→ Verteilung

Bsp.

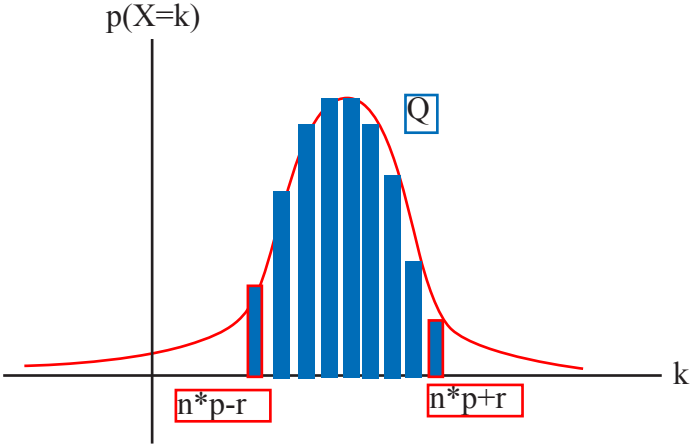
Eine fabrik produziert Schrauben, die mit einer Wahrscheinlichkeit von $p = 0.005$ defekt sind. Wie gross ist die Wahrscheinlichkeit, in einer Schachtel von 2000 Schrauben 10 oder mehr defekte zu finden?

X = Anzahl defekte Scharauben aus 2000-Stufigem Bernoulli Versuch

$m = n*p = 2000*0.005$

$p(X \ge 10)$ **mind. 10 defekte** $= 1 - p(X \ge 10)$

$$= 1 - \sum_{k=0}^{9} \frac{m^k}{k!} * e^{-m}$$



Ungleichungen :

Nie beidseitiges Dividieren oder multiplizieren mit negativen Zahlen!

$-b < a < b \quad \rightarrow \quad a^2 < b^2$

$a*p^2+b*p+c \le 0$

Rechung :

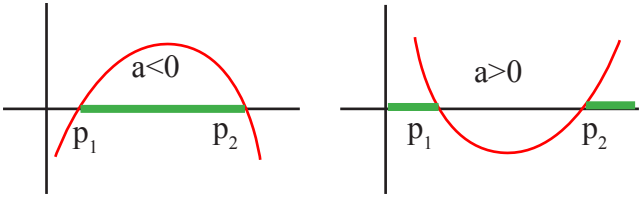
$a*p^2+b*p+c = 0$

ergibt p_1, p_2

$$n*p-2*\sqrt{n*p*(1-p)} \le X \le n*p+2*\sqrt{n*p*(1-p)} \quad \text{¦ *n ¦-p}$$

$$-\frac{2}{n}*\sqrt{n*p*(1-p)} \le \frac{X}{n}-p \le 2*\sqrt{n*p*(1-p)} \quad \text{¦ ^2}$$

$$\left(\frac{X}{n}-p\right)^2 \le \frac{4*n*p*(1-p)}{n^2}$$



Vertrauenswahrcheinlichkeit, Erfahrungswahrscheinlichkeit:

Die X - Werte liegen also in einem Intervall $[n*p-r, n*p+r]$ für einen geeigneten Radius r. Man sucht nun ein r so, dass die Wahrscheinlichkeit dafür, dass X im Bereich $n*p-r \le X \le n*p+r$ liegt, gerade Q ist.

Algebraisch:

$p(m-r \le X \le m+r) = Q$

$E(X) = n*p$

$V(X) = n*p*(1-p)$

Achtung:

Fonfidenzintervall $\frac{X}{n} = \text{Prozent}$

$p_1 \le p \le p_2$

n geeignet? e. ungünstig?

1.Konfidentintervall bestimmen

2.ungünstigste Fälle von p_2 n bestimmen -> ev. n zu klein

Wir lassen nicht jede Zahl für Q zu, sonder nur:

$$p(m-2*s \le X \le m+2*s) = f\left(\frac{(n*p+2*s)-n*p}{s}\right) - f\left(\frac{(n*p-2*s)-n*p}{s}\right) = 0.9544 \text{ aus Liste}$$

$$p(m-3*s \le X \le m+3*s) = f\left(\frac{(n*p+3*s)-n*p}{s}\right) - f\left(\frac{(n*p-3*s)-n*p}{s}\right) = 0.9974 \text{ aus Liste}$$

X und n ist bekannt ergibt nach lösen der Ungeleichung:

$$\left(\frac{X}{n}-p\right)^2 \le m*\frac{p*(1-p)}{n}$$

dabei ist $m = 4$ bei $Q = 0.954$ $m = 9$ bei $Q = 0.997$

$$\left(\frac{X}{n}-p\right)^2 - m*\frac{p*(1-p)}{n} = 0$$ ergibt $p_1 = 0.0029$ $p_2 = 0.0428$

bsp.

Von 1000 werkstücken einer Sendung erwiesen sich 30 als defekt.

$-(\text{Rechnung mit } n = 1000 \text{ und } X = 30)$

$p_1 = 0.0029$ $p_2 = 0.0428$

Achtung, um sicher zu sein Radius vergrössern! $\rightarrow p_1 = 0.002,\ p_2 = 0.043$

zu 95.4% kann man sicher sein, dass zwischen 2% und 4.3% der Werkstücke kapputt sind!

$0.002 \le p \le 0.043$

Allgemein:

Eine Stichprobe sei etstanden durch einen n - stufigen Bernoulliversuch mit unbekannter Erfolgswahrscheinlichkeit p nd X sei die Anzahl der erreichten Erfolge.

Ist $n*p(1-p) > 9$, so bilden alle jene p, welche die Ungleichung

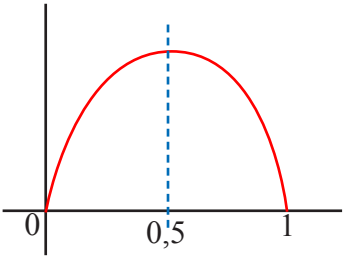
$\left(\frac{X}{n}-p\right)^2 \le m*\frac{p*(1-p)}{n}$ erfüllen,

das Vertrauensintervall für p zur Vertrauenswahrscheinlichkeit Q oder kurz Q% - Vertrauensintervall oder Konfidenzintervall.

Der notwendige Stichprobenumfang:

"Genauigkeit von 3% mit einer Vertrauenswahrscheinlichkeit von 95,4%"

$$\left|\frac{X}{n}-p\right| \le 0.03 \qquad \left(\frac{X}{n}-p\right)^2 \le 0.03^2$$



$$\left(\frac{X}{n}-p\right)^2 \le \frac{4*p*(1-p)}{n} \le 0.03^2$$

$$\frac{4*p*(1-p)}{0.03^2} \le n \qquad n(p) \quad \text{mit} \quad 0 \le p \le 1$$

$$n(p) = \frac{4*p*(1-p)}{0.03^2} = \text{Parabel} \qquad \text{"schlechtester Fall" bei 0,5 weil}$$

$$\frac{4*0,5*(1-0.5)}{0,03^2} \approx 1111,11$$

Für eine Genauigkeit von 3% und einer Vertrauenswahrscheinlichkeit von 95,4% sind mindestens 1112 Befragungen nötig.

Regression

univariante Statistik = nur ein Merkmal wird untersucht
bivariante Statistik = zwei Merkmale werden untersucht
multivariante Statistik = mehr als 2 Merkmale werden untersucht

Abweichung $\varepsilon_i = y_i - F(x_i) \qquad y_i = \beta_0 + \beta_1 * x_i + \varepsilon_i$

$$SSE = \sum_{i=1}^{n} (y_i - f(x_i))^2$$

in Matrizenform:

$$(1 \quad x_i) * \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} + \varepsilon_i = 1 * \beta_0 + x_i * \beta_1 + \varepsilon_i$$

$$SSE = \sum_{i=1}^{n} y_i - (\beta_0 + x_i * \beta_1)^2 = \sum_{i=1}^{n} \varepsilon_i^2 \qquad |\varepsilon|^2 \text{ muss minimal sein!!!}$$

$$\frac{\partial SSE}{\partial \beta_0} = -2 * \sum_{i=1}^{n} (y_i - (\beta_0 + \beta_1 * x_i)) = 0 \qquad \text{¦:-2}$$

$$\sum_{i=1}^{n} y_i = n * \beta_0 + \beta_1 * \sum_{i=1}^{n} x_i$$

$$\frac{\partial SSE}{\partial \beta_1} = -2 * \sum_{i=1}^{n} x_i * (y_i - (\beta_0 + \beta_1 * x_i)) = 0 \qquad \text{¦:-2}$$

$$\sum_{i=1}^{n} y_i * x_i = \beta_0 * \sum_{i=1}^{n} x_i + \beta_1 * \sum_{i=1}^{n} x_i^2$$

Mit Matrizen dargestellt :

linke Seite :
$$\begin{pmatrix} \sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} x_i * y_i \end{pmatrix} = \begin{pmatrix} 1 & 1 & \dots & 1 \\ x_1 & x_2 & \dots & x_n \end{pmatrix} * \begin{pmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{pmatrix} \qquad X^T \qquad Y$$

rechte Seite :
$$\begin{pmatrix} n * b_0 + b_1 * \sum_{i=1}^{n} x_i \\ b_0 * \sum_{i=1}^{n} x_i + b_1 * \sum_{i=1}^{n} x_i^2 \end{pmatrix} = \begin{pmatrix} n & \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} x_i^2 \end{pmatrix} * \begin{pmatrix} b_0 \\ b_1 \end{pmatrix}$$

$$= \begin{pmatrix} 1 & 1 & \dots & 1 \\ x_1 & x_2 & \dots & x_n \end{pmatrix} * \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \dots & \dots \\ 1 & x_n \end{pmatrix} * \begin{pmatrix} b_0 \\ b_1 \end{pmatrix} \qquad X^T \; X \; b$$

$$\rightarrow X^T * Y = X^T * X * b$$

$$b = (X^T * X) * X^T * Y$$

TI-89 :

$\{x_1, x_2, \dots, x_n\}$ STO A

$\{y_1, y_2, \dots, y_n\}$ STO B

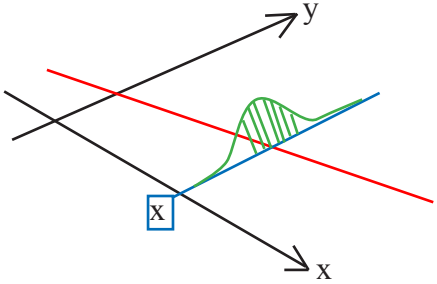
LinReg A, B

QaudReg A, B

ExpReg A, B

danach ShowStat

Damit das Werkzeug "lineare Regression" sinnvolll angewendet werden kann auf Datenpunkte (x,y), sollte das Residuum $\varepsilon = y - (\beta_1 + \beta_2 * x)$ and der Stelle x normalverteilt sein mit Mittelwert 0 und einer Varianz σ^2, welche unabhängig von der Stelle x ist.



Regression bei multivarianten Daten :

$$Y = \begin{pmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{pmatrix} \qquad b = \begin{pmatrix} b_1 \\ b_2 \\ \dots \\ b_p \end{pmatrix} \qquad X = \begin{pmatrix} 1 & x_{11} & \dots & x_{1p} \\ 1 & x_{21} & \dots & x_{2p} \\ \dots & \dots & & \dots \\ 1 & x_{n1} & \dots & x_{np} \end{pmatrix}$$

Achtung : bei x^2 $\quad X = \begin{pmatrix} 1 & x_1 & x_1^2 \\ \dots & \dots & \dots \\ 1 & x_n & x_n^2 \end{pmatrix}$ $\qquad$ bei $\frac{1}{x}$ $X = \begin{pmatrix} 1 & x_1 & \frac{1}{x_1} \\ \dots & \dots & \dots \\ 1 & x_n & \frac{1}{x_n} \end{pmatrix}$

$$b = \left(X^T * X\right) * X^T * Y$$

Die Funktion kann nicht linear sein, die Regression ist es aber!

Trick :

Daten (x_i / y_i) $\qquad$ **Modell** $y = a * e^{b*x}$ **(Funktion aussuchen, die etwa passt)**

$y = a * e^{b*x} \qquad | \ln( )$

$\ln(y) = \ln(a) + b * x \quad \rightarrow$ **ist linear**

$\ln(y) = a^* + b^* * x$

Rechne lineare Regression für transformierte Daten

Rücktransormation : $\quad a = e^{a^*}, \quad b = b^*$

Beispiel :

$n = 2$

$y_1 = 1, \quad y_2 = 3$

$\hat{y}_1 = -1, \quad \hat{y}_2 = -3$

$\bar{y} = 2$

$\hat{\bar{y}} = -2$

$$R^2 = \frac{\sum_{i=1}^{n} \left(\hat{y}_i - \hat{\bar{y}}\right)^2}{\sum_{i=1}^{n} \left(y_i - \bar{y}\right)^2} = \frac{(-1+2)^2 + (-3+2)^2}{(1-2)^2 + (3-2)^2} = \frac{2}{2} = 1$$

Nur aus der Formel für R^2 kann nicht geschlossen werden:

$\left(R^2 = 1\right) \neq \left(y_i \approx \hat{y}_i\right)$

Wie gut ist der Fit??

Als Mass für den Fehler :

$SS_{tot} = \sum_{i=1}^{n} \left(y_i - \bar{y}\right)^2$ **ist bis auf den Vorfaktor** $\frac{1}{n-1}$ **de Varianz!**

Durch x verursachte Varianz von y oder erklärte Varianz von y :

$$\frac{1}{n-1} * \sum_{i=1}^{n} \left(\hat{y}_i - \hat{\bar{y}}\right)^2$$

Es seien Datenpunkte $\left(x_{i1}, x_{i2}, \ldots, x_{ip}, y_i\right)$ **gegeben mit** $i = 1, 2, \ldots, n$

für eine Regression $\hat{y}_i = f\left(x_{i1}, x_{i2}, \ldots, x_{ip}\right)$ **durchgeführt worden, dann**

heisst der Term :

$$R^2 = \frac{\sum_{i=1}^{n} \left(\hat{y}_i - \hat{\bar{y}}\right)^2}{\sum_{i=1}^{n} \left(y_i - \bar{y}\right)^2} = \frac{\text{erklärte Varianz}}{SS_{tot}} = \text{Bestimmtheitsmass oder Determinationskoeffizient}$$

der Regression.

$\bar{y}$ **= empirischer Mittelwert der** y_i

$\hat{\bar{y}}$ **= empirischer Mittelwert der** $\hat{y}_i$

je besser die Näherung $y_i \approx \hat{y}_i$**, deso näher liegt** R^2 **bei 1.**

Es seien Datenpunkte $\left(x_i, y_i\right)$ **gegeben mit** $i = 1, 2, \ldots, n$ **für die eine Regression**

$\hat{y}_i = b_0 + b_1 * x_i$ **durchgeführt worden ist, dann gilt :**

$$\sum_{i=1}^{n} \left(\hat{y}_i - \hat{\bar{y}}\right)^2 = \sum_{i=1}^{n} \left(y_i - \bar{y}\right)^2 - SSE$$

$$R^2 = 1 - \frac{SSE}{\sum_{i=1}^{n} \left(y_i - \bar{y}\right)^2}$$

$$SSE = \left(1 - R^2\right) * \sum \left(y_i - \bar{y}\right)^2$$

Die Qualität des Fits ist umso besser, je kleiner der SSE.

$1 - R^2$ **macht den SSE klein, wenn** R^2 **nahe bei 1**

Mehr noch :

$1 - R^2 \geq 0$ **da** $SSE \geq 0$ **und** $SS_{tot} \geq 0$ $\qquad 0 \leq R^2 \leq 1$

Korrelationskoeffizient :

$$r = \frac{\sum_{i=1}^{n} \left(x_i - \bar{x}\right) * \left(y_i - \bar{y}\right)}{\sqrt{\sum_{i=1}^{n} \left(x_i - \bar{x}\right)^2 * \sum_{i=1}^{n} \left(y_i - \bar{y}\right)^2}}$$

r^2 = **Bestimmtheitsmass**

r ist positiv, wenn die Regressionsgerade steigend ist

r ist negativ, wenn die Regressionsgerade fallend ist