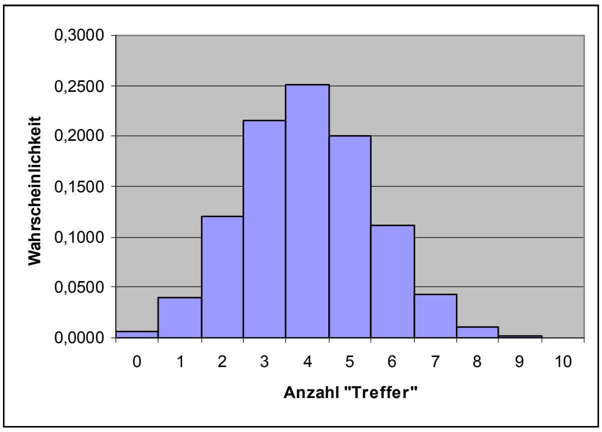
5.1 Überblick Test-Verteilungen

<u>Diskrete Zufallsvariable</u>

Wahrschlk.-Funktion $f(x_i)$ mit $x_i \mapsto W\left(X = x_i\right)$

Verteilungsfunktion $F(x_i)$ mit $x_i \mapsto W\left(X \le x_i\right)$

$$F(x_i) = \sum_i W(X = x_i)$$



$$f_{10\,|\,0,4}(x) = \binom{10}{x} 0,4^x \cdot 0,6^{10-x}$$

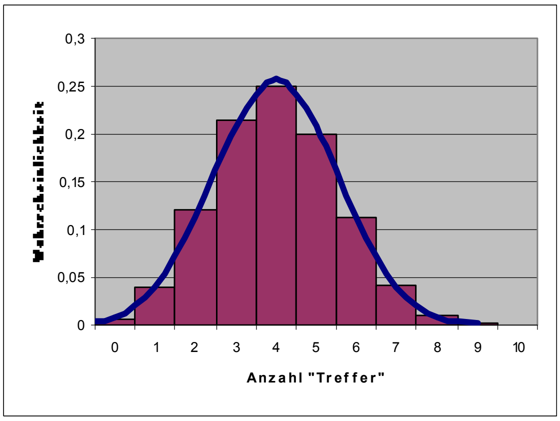
binomialverteilt mit n = 10, p = 0,4

<u>Stetige Zufallsvariable</u>

Dichtefunktion f(x)

Verteilungsfunktion F(x)

$$W(X \le x_2) = \int_{-\infty}^{x_2} f(x)dx = F(x_2)$$ "Flächeninhalte"



$$f_{4|1,55}(x) = \frac{1}{1,55\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-4}{1,55}\right)^2}$$

normalverteilt mit μ = 4, σ = 1,55

<u>❶ Binomialverteilung</u>

$$f_{n,p}(x) = \binom{n}{x} p^x q^{n-x}$$

Für Zufallsversuche <u>**mit**</u> Zurücklegen.

2 mögliche Ereignisse: Treffer A, Niete AΓ
mit den Wahrscheinlichkeiten p, $q = 1 - p$.
x ist die Anzahl der Treffer.

<u>❷ Hypergeometrische Verteilung</u>

$$f_{n,N,M}(x) = \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}$$

Für Zufallsversuche <u>**ohne**</u> Zurücklegen.

2 mögliche Ereignisse: Treffer A, Niete AΓ
mit den Wahrscheinlichkeiten p, $q = 1 - p$.
x ist die Anzahl der Treffer.

<u>❸ POISSON-Verteilung</u> $f_\mu(x) = \frac{\mu^x}{x!} \cdot e^{-\mu}$

für kleine Treffer-Wahrscheinlichkeiten
$p < 0,05$
d.h. seltene Ereignisse,
d.h. sehr unsymmetrische Verteilungen

<u>❹ Fisher – F-Verteilung</u>

$F_{\alpha\,|\,p\,|\,v\,|\,x}$ sind in Tabelle 7.3

Der Zusammenhang quantitativer Daten ist statistisch gesichert, wenn

$$xF_{\text{empirisch}} > xF_{\text{crit.}}$$

und weitere Testverfahren

<u>❺ Student-t-Verteilung</u>

$F_{\alpha\,|\,v}$ (t) sind in Tabelle 7.5

Die Einflussvariable X liefert einen signifikanten Beitrag zur (multiplen) Regression, wenn $t_{\text{empirisch}} > t_{\text{critical}}$

und weitere Testverfahren

<u>❻ chi^2 – χ^2– Verteilung</u>

$F_{\alpha\,|\,v}(\chi^2)$ sind in Tabelle 7.4

Die Merkmale A, B sind abhängig, wenn

$$\chi^2_{\text{empirisch}} > \chi^2_{\text{crit.}}$$

und weitere Testverfahren

<u>❼ Normalverteilung</u>

$$f_{\mu|\sigma}(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \qquad z = \frac{x-\mu}{\sigma}$$

$F_{0\,|\,1}(z) = F_{SN}(z)$ sind in Tabelle 7.6

Für hinreichend große n konvergieren die Verteilungen ❶ bis ❻ gegen die Normalverteilung.

5.2 BERNOULLI-KETTEN

Im Modell für Zufallsversuche "Werfen von Reißnägeln" gibt es zwei mögliche Ereignisse:
"Spitzenlage" ⋌ und "Rückenlage" ⊥ , ⋌ sei ein Treffer, ⊥ sei ein Nicht-Treffer.
Die Treffer-Wahrscheinlichkeit ist p, die für Nicht-Treffer ist $q = 1 - p$.

Bei Zufallsversuchen **mit** Zurücklegen bleiben die Wahrscheinlichkeiten p und q konstant.
Die Ereignisse einer Kette solcher Zufallsversuche sind voneinander unabhängig.
entspricht "Ziehen mit Zurücklegen"
Ketten unabhängiger Zufallsversuche mit jeweils zwei möglichen Ereignissen nennt man BERNOULLI-Ketten. (Gutware | Ausschuss, ja | nein, positiv | negativ, bestanden | nicht best. …)
[BERNOULLI, Jakob, Basel 1713]

Aus den Zufallsversuchen (*Abschnitt 3.9*) kennen wir die Wahrscheinlichkeit dafür, dass ein geworfener Reißnagel auf der Spitze liegt: $p = W(X) = W(⋌) = 0{,}39$ und $q = 1 - p = 0{,}61$.
Die Zufallsvariable X ist die Anzahl der Treffer (Spitzenlage).

Wirft man 2-mal einen Reißnagel, dann ergeben sich die Möglichkeiten:

	⊥⊥	⊥⋌	⋌⊥	⋌⋌	
$X =$	0	1	1	2	
$W(X) =$	$q{\cdot}q$	$q{\cdot}p$	$p{\cdot}q$	$p{\cdot}p$	und $q^2 + 2\,q\,p + p^2 = (q+p)^2 = 1$

Wirft man 3-mal einen Reißnagel, dann ergeben sich die Möglichkeiten:

	⊥⊥⊥	⊥⊥⋌, ⊥⋌⊥, ⋌⊥⊥	⊥⋌⋌, ⋌⊥⋌, ⋌⋌⊥	⋌⋌⋌
$X =$	0	1	2	3
$W(X) =$	$q{\cdot}q{\cdot}q$	$3\,q^2{\cdot}p$	$3\,q{\cdot}p^2$	p^3

und $q^3p^0 + 3\,q^2\,p + 3\,q\,p^2 + p^3 = p^3 + 3\,p^2\,q + 3\,p\,q^2 + q^3 = (p+q)^3 = 1.$

Die Wahrscheinlichkeiten der BERNOULLI-Ketten folgen der Binomialentwicklung $(p+q)^n$.
Z.B. Wahrscheinlichkeit für 2 Treffer $W(X{=}2) = 3\,p^2\,q = 3 \cdot 0{,}39^2 \cdot 0{,}61 = 0{,}278$.
Wiederholung zu Binomialkoeffizienten siehe Abschnitt 5.5

Binomialverteilung: $W(X = x) = f_{Bin,n,p}(x) = \binom{n}{x} p^x q^{n-x} = \binom{n}{x} p^x (1-p)^{n-x}$

p^x Wahrscheinlichkeit der x Treffer, $(1-p)^{n-x}$ Wahrscheinlichkeit der $n - x$ Nicht-Treffer.
(Multiplikationsregel für unabhängige Ereignisse, *Abschnitt 4.2*)
$\binom{n}{x}$ gleichwahrscheinliche Kombinationen (Additionsregel für unabh. Ereignisse, *vgl. 4.2*)
Der Erwartungswert $\mu = np$, das ist unmittelbar einsichtig.
Für die Varianz gilt $\sigma^2 = n\,p\,q$. (siehe 5.4)

Die Binomialreihe für Treffer-Wahrscheinlichkeit p, Stichprobenumfang n:

$$(p+q)^n = \binom{n}{0} p^0 q^n + \binom{n}{1} p^1 q^{n-1} + \binom{n}{2} p^2 q^{n-2} + \ldots + \binom{n}{x} p^x q^{n-x} + \ldots + \binom{n}{n} p^n q^0 = 1$$

Wahrscheinlichkeit für <u>**genau**</u> x Treffer: $f_{n,p}(x) = \binom{n}{x} p^x q^{n-x}$ mit $x = 0, 1, 2, \ldots, n$

Die <u>**Verteilungsfunktion**</u> liefert die aufsummierten Werte der Binomialverteilung:

$$F_{\mathrm{Bin}\,|\,n\,|\,p}(k) = \sum_{x=0}^{k} f_{n,p}(x) = \sum_{x=0}^{k} \binom{n}{x} p^x q^{n-x}$$

Tabelle 7.1

5.3 Binomialverteilung

Beispiel 5.1

In einer Klinik werden in einem Monat n = 50 Geburten registriert.
Die Wahrscheinlichkeit für Mädchen ist $p = 0{,}486$.
a) Wie groß ist die Wahrscheinlichkeit dafür, dass darunter 20 Mädchengeburten sind? *genau!*
b) Wie groß ist die Wahrscheinlichkeit dafür, dass darunter höchstens 20 Mädchengeburten sind?
c) Erwartungswert, erwartete Varianz und erwartete Standardabweichung?
d) Wie ändert sich die erwartete Standardabweichung, wenn man das Stichprobenmittel aus einem Jahr betrachtet?

$q = 1 - 0{,}486 = 0{,}514$ ist die Wahrscheinlichkeit für die Geburt eines Jungen.

a) $f_{\text{Bin}\,|50|\,0{,}486}(20) = \binom{50}{20} 0{,}486^{20} \cdot 0{,}514^{30} = \frac{50!}{20!30!} 0{,}486^{20} \cdot 0{,}514^{30} = 0{,}0543 = 5{,}4\%$

b) $F_{\text{Bin}\,|\,50\,|\,0{,}486}(20) = \sum_{x=0}^{20} \binom{50}{x} 0{,}486^{x} \cdot 0{,}514^{50-x} =$

$$\binom{50}{0} 0{,}486^{0} \cdot 0{,}514^{50} + \binom{50}{1} 0{,}486^{1} \cdot 0{,}514^{49} + \ldots \binom{50}{20} 0{,}486^{20} \cdot 0{,}514^{30} = 0{,}141 = 14{,}1\%$$

Das ist sehr aufwendig; man hat zwei Möglichkeiten
(1) ein Numerik-Programm benutzen z.B. in Excel: =BINOMVERT(20;50;0,486;1)
(2) mit Hilfe der Normalverteilung den Näherungswert bestimmen, siehe unten
Tabellen für $F_{\text{Bin}}(k)$ gibt es nur für glatte p und n, hier etwa für p=0,5 und n=50.

c) $\mu = n\,p = 50 \cdot 0{,}486 = 24{,}3$ Mädchen
$\sigma^2 = n\,p\,q = 50 \cdot 0{,}486 \cdot 0{,}514 = 12{,}49$
$\sigma = \sqrt{12{,}49} = 3{,}53$ Mädchen

d) $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{12}} = \frac{3{,}53}{\sqrt{12}} = 1{,}02$ Mädchen

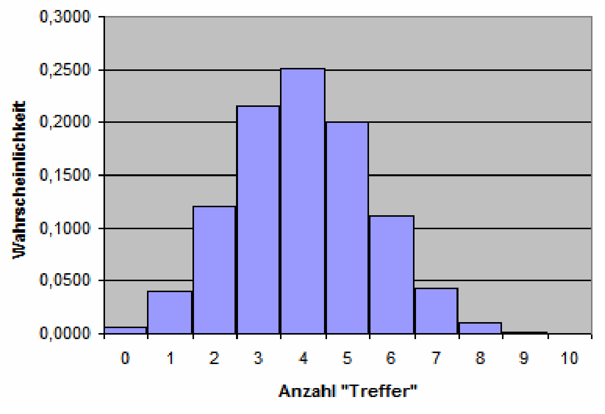
Beispiel 5.2

An einem erosionsgefährdeten Hang sollen 10 Fichtenpflanzen angepflanzt werden.
Die Baumschule gibt eine Anwachswahrscheinlichkeit von p = 40% an.
Wie groß ist die Wahrscheinlichkeit, dass
a) genau 4 b) höchstens 3 c) mindestens 5 d) alle 10 anwachsen?
e) Bestimmen Sie μ, σ^2, σ. f) Zeichnen Sie das Histogramm

Aus Tabelle 7.1

k	$F_{10;\,0,4}$ (k)
0	0,0060
1	0,0464
2	0,1673
3	0,3823
4	0,6331
5	0,8338
6	0,9452
7	0,9877
8	0,9983
9	0,9999
10	1,0000

a) $f(4) = F(4) - F(3) =$
$= 0{,}6331 - 0{,}3823 =$
$= 0{,}2508$
b) $F(3) = 0{,}3823$
c) $W(X \geq 5) = 1 - F(4) =$
$=1 - 0{,}6331 \quad = 0{,}3669$
d) "genau 10" $W(X = 10) =$
$= f(10) = F(10) - F(9) =$
$= 1 - 0{,}9999 = 0{,}0001$
e) $\mu = 10 \cdot 0{,}4 = 4$ Fichten.
$\sigma^2 = 10 \cdot 0{,}4 \cdot 0{,}6 = 2{,}4 < 9$
$\sigma = \sqrt{2{,}4} = 1{,}55$ Fichten



5.4 VARIANZ $\sigma^2 = n\,p\,q$

Varianz einer binomialverteilten Zufallsvariablen

Für die Varianz einer Zufallsvariablen gilt: $\sigma^2 = \sum_i x_i^2 \cdot f(x_i) - \mu^2$

Speziell für Binomialverteilung: $\sigma^2 = \sum_i x_i^2 \cdot \binom{n}{x_i} p^{x_i} q^{n-x_i} - (np)^2$

Bei 1 "Wurf": n = 1 $W(X=0) = \binom{1}{0} p^0 q^1$ und $W(X=1) = \binom{1}{1} p^1 q^0$

$$\sigma^2 = 0^2 \binom{1}{0} p^0 q^1 + 1^2 \binom{1}{1} p^1 q^0 - (1\cdot p)^2 = p - p^2 = p(1-p) = 1pq$$

Bei 2 "Würfen": n = 2 $X = 0;\ 1;\ 2$ $W(X{=}0);\ W(X{=}1);\ W(X{=}2)$

$$\sigma^2 = 0^2 \binom{2}{0} p^0 q^2 + 1^2 \binom{2}{1} p^1 q^1 + 2^2 \binom{2}{2} p^2 q^0 - (2\cdot p)^2 = 1\cdot 2\cdot pq + 4\cdot 1p^2 - 4p^2 = 2pq$$

Bei 3 "Würfen": n = 3 $X = 0;\ 1;\ 2;\ 3$ $W(X{=}0);\ W(X{=}1);\ W(X{=}2);\ W(X{=}3)$

$$\sigma^2 = 0^2 \binom{3}{0} p^0 q^3 + 1^2 \binom{3}{1} p^1 q^2 + 2^2 \binom{3}{2} p^2 q^1 + 3^2 \binom{3}{3} p^3 q^0 - (3\cdot p)^2 = 3pq^2 + 4\cdot 3p^2 q + 9p^3 - 9p^2$$

$$= 3p\left(q^2 + 4pq + 3p^2 - 3p\right) = 3p\left(q^2 + 4pq + 3p(p-1)\right) = 3p\left(q^2 + 4pq - 3pq\right)$$

$$= 3p(q^2 + pq) = 3p(q^2 + (1-q)q) = 3p(q^2 + q - q^2) = 3pq$$

(das ist kein Beweis, der Beweis läuft über die vollständige Induktion)

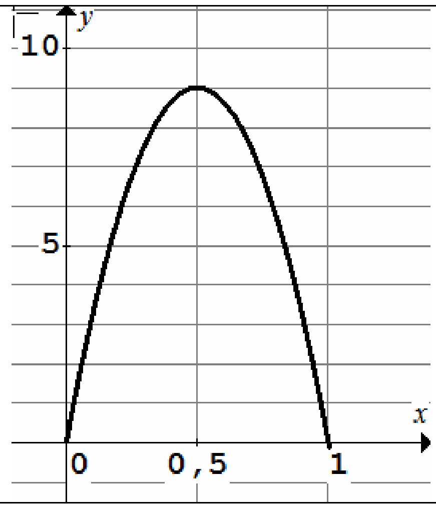
- - -

Allgemein: $\boldsymbol{\sigma^2 = n\,p\,q}$

LAPLACE-Bedingung: wenn npq > 9 kann man Normalverteilung benutzen

Varianz npq > 9. Exakte Symmetrie bei p = q = 0,5.
Je symmetrischer und je größer n, desto besser die Approximation an die Normalverteilung.
Grenze bei npq = np(1-p) = 9 => $n_{kritisch} = 9 / (0{,}5 \cdot 0{,}5) = 36$.
Funktionsklasse $f_n(p) = n\,p\,(1-p) = n\,x\,(1-x) = n\,(x - x^2)$ Parabelschar

z.B. n = 36 $f(x) = 36\,(x - x^2)$



5.5 BINOMIALKOEFFIZIENTEN

Die Binomialkoeffizienten $\binom{n}{x}$ kennen wir aus den Binomischen Formeln $(a+b)^n$.

Man entnimmt sie rekursiv aus dem PASCAL'schen Koeffizientenschema (siehe unten) oder bestimmt sie mit einer der beiden folgenden Formeln:

$$\binom{n}{x} = \frac{n \cdot (n-1) \cdot (n-2) \cdot \ldots \cdot (n-x+1)}{x!} = \frac{n!}{x!(n-x)!}$$

Bei den meisten Taschenrechnern gibt es dazu die Taste [nCr]: Number of Combinations $\binom{n}{r}$

$$\begin{aligned}
(a+b)^0 &= 1 && = 1\,a^0b^0 \\
(a+b)^1 &= a + b && = 1\,a^1b^0 + 1\,a^0b^1 \\
(a+b)^2 &= a^2 + 2ab + b^2 && = 1\,a^2b^0 + 2\,a^1b^1 + 1\,a^0b^2 \\
(a+b)^3 &= a^3 + 3\,a^2b + 3\,a\,b^2 + b^3 && = 1\,a^3b^0 + 3\,a^2b^1 + 3\,a^1b^2 + 1\,a^0b^3 \\
(a+b)^4 &= && = 1\,a^4b^0 + 4\,a^3b^1 + 6\,a^2b^2 + 4\,a^1b^3 + 1\,a^0b^4
\end{aligned}$$

Binomialkoeffizienten: $\binom{4}{x} = \binom{4}{0};\binom{4}{1};\binom{4}{2};\binom{4}{3};\binom{4}{4}$

$6\,a^2b^2$ bedeutet, dass die Kombination a^2b^2 bei $(a+b)^4$ sechs mal vorkommt.

- ☐ $\binom{4}{2} = \frac{4 \cdot 3 \cdot \cancel{2 \cdot 1}\text{kürzen}}{2 \cdot 1 \cdot \cancel{2 \cdot 1}\text{kürzen}} = 6$ (1.Formel) $\binom{4}{2} = \frac{4!}{2! \cdot (4-2)!} = \frac{24}{2 \cdot 2} = 6$ (2. Formel).

  mit Taschenrechner: 4[nCr]2

- ☐ $\binom{75}{3} = \frac{75 \cdot 74 \cdot 73}{3 \cdot 2 \cdot 1} = 67525.$ 75! nicht mit Taschenrechner!

5.6 Ziehen ohne Zurücklegen

Beispiel 5.3 (*nach Lambacher-Schweizer S.135*)

In einer Urne befinden sich $M = 64$ schwarze Kugeln und $N - M = 36$ weiße Kugeln.
"Treffer" sei das Ziehen einer schwarzen Kugel, d.h.

$p = \frac{64}{100}$ und $q = \frac{100-64}{100}$ allgemein: $p = \frac{M}{N}$ und $q = 1 - \frac{M}{N}$

Es werden $n = 5$ Kugeln nacheinander gezogen. Stichprobenumfang $n = 5$.
Wie groß ist die Wahrscheinlichkeit unter den 5 Kugeln 3 schwarze zu ziehen, $W(X=3)$?

Ziehen **mit** Zurücklegen BERNOULLI-Kette, $p = 0{,}64$ ist konstant, X ist binomialverteilt:

$$W(X=3) = f_{5;\,0,64}(3) = \binom{5}{3} \cdot 0{,}64^3 \cdot 0{,}36^2 = 10 \cdot \frac{64}{100} \cdot \frac{64}{100} \cdot \frac{64}{100} \cdot \frac{36}{100} \cdot \frac{36}{100} = 0{,}3397$$

allgemein: $W(X=x) = f_{n,p}(x) = \binom{n}{x} \cdot \left(\frac{M}{N}\right)^x \cdot \left(1 - \frac{M}{N}\right)^{n-x}$ (Binomialverteilung)

Ziehen **ohne** Zurücklegen: die Treffer-Wahrscheinlichkeit ***p* ändert sich nach jedem Zug**.

$$W(X=3) = \binom{5}{3} \cdot \frac{64}{100} \cdot \frac{63}{99} \cdot \frac{62}{98} \cdot \frac{36}{97} \cdot \frac{35}{96} = 0{,}3486$$

allgemein: $W(X=x) = f_{n,N,M}(x) = \binom{n}{x} \cdot \underbrace{\frac{M}{N} \cdot \frac{M-1}{N-1} \cdot \ldots}_{x \text{ fallende Faktoren}} \quad \underbrace{\frac{N-M}{N-x} \cdot \frac{N-M-1}{N-x-1} \cdot \ldots}_{n-x \text{ fallende Faktoren}}$

Dieser Ausdruck lässt sich zusammenfassen zu $W(X=x) = f_{Hyp|n,N,M}(x) = \dfrac{\binom{M}{x} \cdot \binom{N-M}{n-x}}{\binom{N}{n}}$

Das ist die hypergeometrische Wahrscheinlichkeits-Funktion.
Wenn man die Symbole p und q aus der Binomialverteilung benutzt, ergeben sich

Erwartungswert $\mu = n\,p$. Varianz $\sigma^2 = n\,p\,q \cdot \frac{N-n}{N-1}$. $\frac{1000-5}{1000-1} \approx 1$

Standardabweichung $\sigma_{\text{Hyp}} = \sqrt{n \cdot p \cdot q} \cdot \sqrt{\frac{N-n}{N-1}}$ => Korrekturfaktor $\sqrt{\frac{N-n}{N-1}}$

Die Funktionswerte $f_{\text{Hyp}\,|\,n\,|\,N\,|\,M}(x)$ sind von 3 Parametern abhängig *n, N, M.*
Sie lassen sich deshalb schwer tabellieren.

Wenn der **Auswahlsatz** $\frac{n}{N} \le 0{,}05$ kann man die Hypergeometrische Verteilung mit Hilfe der Binomialverteilung approximieren (annähern).
Dies ist eine **Approximationsbedingung**.

Unter der weiteren Bedingung $npq > 9$ kann man sogar die Normalverteilung benutzen.
Wenn die Auswahlsatz-Bedingung $n/N \le 0{,}05$ verletzt ist, dann muss man den Korrekturfaktor $\sqrt{\frac{N-n}{N-1}}$ verwenden… siehe unten

5.7 Hypergeometrische Verteilung

Beispiel 5.4

Bei der Produktion von Elektromotoren seien im Durchschnitt 5% Ausschuss. $p=0,05$
Die Fertigungsmenge umfasst N=200 Stück.
Wie groß ist die Wahrscheinlichkeit dafür, dass bei einer Entnahme von n=20 Stück

a) genau 5 Stück Ausschuss b) alle verwendbar c) höchstens 2 Stück Ausschuss sind?
d) Erwartungswert und erwartete Varianz und Standardabweichung?
e) Inwiefern ist es hier zulässig, die Binomialverteilung zu verwenden?
f) Aufgabenteil c) auch mit Binomialverteilung zum Vergleich.

1. Das ist eine Stichprobe **ohne Zurücklegen**, das ist der Normalfall!
2. N=200. n=20. n/N = 0,1 > 0,05 => f_{Hyp}. M = p·N= 0,05·200 = 10 Stück. nCr

a) "genau 5" $W(X=5)=f_{20;200;10}(5)=\dfrac{\binom{10}{5}\cdot\binom{190}{15}}{\binom{200}{20}}=\dfrac{\frac{10!}{5!5!}\cdot\frac{190!}{15!175!}}{\frac{200!}{20!180!}}=\dfrac{252\cdot 6,58\cdot 10^{21}}{1,61\cdot 10^{27}}=0,001$

b) "0 Stück Ausschuss" $W(X=0)=f_{20;200;10}(0)=\dfrac{\binom{10}{0}\cdot\binom{190}{20}}{\binom{200}{20}}=\dfrac{\frac{10!}{0!10!}\cdot\frac{190!}{20!170!}}{\frac{200!}{20!180!}}$ =0,3398

c) $W(X=0)+W(X=1)+W(X=2)=0,3398+\dfrac{\binom{10}{1}\cdot\binom{190}{19}}{\binom{200}{20}}+\dfrac{\binom{10}{2}\cdot\binom{190}{18}}{\binom{200}{20}}=$

$$=0,3398+\frac{\frac{10!}{1!9!}\cdot\frac{190!}{19!171!}}{\frac{200!}{20!180!}}+\frac{\frac{10!}{2!8!}\cdot\frac{190!}{18!172!}}{\frac{200!}{20!180!}}=0,3398+0,3974+0,1975=0,9347$$

d) μ = n·p = 20·0,05 = 1 Stück Ausschuss

$\sigma^2=n\,p\,q\cdot\dfrac{N-n}{N-1}=20\cdot 0,05\cdot 0,95\cdot\dfrac{200-20}{200-1}=0,95\cdot 0,9045=0,859$.

$\sigma=\sqrt{0,859}=0,927$ Stück Ausschuss

e) Das ist nicht zulässig, weil für den Auswahlsatz gilt n/N = 0,1 > 0,05.

f) $F_{\text{Bin}\,|\,20\,|\,0,05}(2)=\sum_{x=0}^{2}\binom{20}{x}0,05^x\cdot 0,95^{20-x}=$

$$\binom{20}{0}0,05^0\cdot 0,95^{20}+\binom{20}{1}0,05^1\cdot 0,95^{19}+\binom{20}{2}0,05^2\cdot 0,95^{18}=$$

$$=0,3585+0,3774+0,1887=0,9245$$

5.8 Näherung nach Poisson

Wenn Ereignisse selten auftreten, dann ist die Treffer-Wahrscheinlichkeit p klein.
Für kleine p kann man einen Grenzwert für $n\to\infty$ bestimmen.

Für binomialverteilte Zufallsvariable gilt: $\mu = n\cdot p \Rightarrow p=\frac{\mu}{n}$ und $q = 1-p = 1-\frac{\mu}{n}$

$$f_{n;p}(x)=\binom{n}{x}p^x q^{n-x}=\binom{n}{x}\cdot\left(\frac{\mu}{n}\right)^x\cdot\left(1-\frac{\mu}{n}\right)^{n-x}$$

$$=\frac{n(n-1)(n-2)...(n-x+1)}{x!}\cdot\frac{\mu^x}{n^x}\cdot\left(1-\frac{\mu}{n}\right)^{n-x}=\frac{n(n-1)(n-2)...(n-x+1)}{n^x}\cdot\frac{\mu^x}{x!}\cdot\left(1-\frac{\mu}{n}\right)^{n-x}$$

$$=\frac{n(n-1)(n-2)...(n-x+1)}{n\cdot n\cdot n\cdot ...\cdot n}\cdot\frac{\mu^x}{x!}\cdot\left(1-\frac{\mu}{n}\right)^{n-x}$$

$$=\frac{\mu^x}{x!}\cdot\frac{n}{n}\cdot\frac{n-1}{n}\cdot\frac{n-2}{n}\cdot...\cdot\frac{n-x+1}{n}\cdot\left(1-\frac{\mu}{n}\right)^{n}\cdot\left(1-\frac{\mu}{n}\right)^{-x}$$

$$=\frac{\mu^x}{x!}\cdot\underbrace{\left(1-\frac{\mu}{n}\right)^{n}}_{❷}\cdot 1\cdot\underbrace{\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)...\left(1-\frac{x-1}{n}\right)\cdot\left(1-\frac{\mu}{n}\right)^{-x}}_{❶}\xrightarrow{n\to\infty}\frac{\mu^x}{x!}\cdot e^{-\mu}$$

❶ $\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)...\left(1-\frac{x-1}{n}\right)\cdot\left(1-\frac{\mu}{n}\right)^{-x}$ streben alle für $n\to\infty$ gegen den Grenzwert 1.

❷ $\lim\limits_{n\to\infty}\left(1-\frac{\mu}{n}\right)^{n}=e^{-\mu}$ $\left[\text{vgl. } \lim\limits_{n\to\infty}\left(1+\frac{1}{n}\right)^{n}=e.\quad \lim\limits_{n\to\infty}\left(1+\frac{x}{n}\right)^{n}=e^{x}.\quad \lim\limits_{n\to\infty}\left(1-\frac{x}{n}\right)^{n}=e^{-x}\right]$

Die Poisson-Verteilung hat

die Wahrscheinlichkeitsfunktion $f_{\mathrm{Poi}|\mu}(x)=\frac{\mu^x}{x!}\cdot e^{-\mu}=\frac{\mu^x}{x!\;e^{\mu}}$

und die Verteilungsfunktion $F_{\mathrm{Poi}\,|\,\mu}(k)=\sum_{x=0}^{k}\frac{\mu^x}{x!}\cdot e^{-\mu}$ *Tabelle 7.2*

der Erwartungswert ist $\mu = n\cdot p$

erwartete Varianz $\sigma^2=\mu$ weil $\sigma^2=npq=\mu\cdot(1-p)=\mu\cdot\left(1-\frac{\mu}{n}\right)\xrightarrow{n\to\infty}\mu$

nur **<u>einen</u>** Parameter, nämlich μ. Die Funktionswerte $f_{\mathrm{Poi}}(x)$ lassen deshalb einfacher berechnen und tabellieren als $f_{\mathrm{Bin}}(x)$ und $f_{\mathrm{Hyp}}(x)$.
[Poisson, Siméon Denis, Paris 1837]

Wenn die <u>Approximationsbedingungen</u> erfüllt sind, wird man die Hypergeometrische Verteilung und die Binomialverteilung durch die Poisson-Verteilung approximieren:

<u>Approximationsbedingungen:</u>

Übergang von $f_{\mathrm{Hyp}}(x)$ zu $f_{\mathrm{Bin}}(x)$, wenn $\frac{n}{N}\le 0{,}05$ mit $p=\frac{M}{N}$. $\mu=n\frac{M}{N}$

Übergang von $f_{\mathrm{Bin}}(x)$ zu $f_{\mathrm{Poi}}(x)$, wenn $\frac{n}{p}\ge 1500$ (d.h. p klein, n groß)

Hypergeom.Vert. $\xrightarrow{n/N\le 0{,}05}$ Binomial-Vert. $\xrightarrow{n/p\ge 1500}$ Poisson-Vert.

5.9 POISSON-VERTEILUNG

Beispiel 5.5

In einem Unternehmen mit einer sehr großen Anzahl N sehr groß Telefonaten beträgt die Anzahl der falschen Verbindungen p=1 % . Wie groß ist die Wahrscheinlichkeit, dass bei einer Stichprobe von n=200 Telefonaten genau x=3 falsche Verbindungen entstehen?

1. Das ist eine Stichprobe ohne Zurücklegen.
   "sehr große Anzahl der Telefonate" => $n/N = 200/N < 0{,}05$ => man kann f_{Bin} benutzen:
   $$W(X=3) = f_{\text{Bin, 200; 0,01}}(3) = \binom{200}{3} \cdot 0{,}01^3 \cdot 0{,}99^{197} = 0{,}181 \qquad \frac{200 \cdot 199 \cdot 198}{1 \cdot 2 \cdot 3}$$
2. Man kann sogar die POISSON-Verteilung benutzen, weil $n/p = 200/0{,}01 = 20000 > 1500$
   $\mu = n\,p = 200 \cdot 0{,}01 = 2$ glatte Zahl => Tabelle
   $$W(X=3) = f_{\text{Poi},2}(3) = \frac{2^3}{3!\; e^2} = 0{,}1804 \qquad W(X=3)=W(X\le 3)-W(X\le 2)=0{,}8571-0{,}6767$$
3. Wie groß muss N sein, damit man f_{Bin} benutzen darf? Grenze bei $200/N = 0{,}05$ => $N \ge 4000$.
4. Wie groß muss Stichprobenumfang n mindestens sein, damit man die POISSON-Verteilung verwenden kann? Grenze bei n/0,01=1500 => $n = 15$.

Beispiel 5.6

In einem Betrieb werden viele, N groß Transistoren hergestellt, von denen im Durchschnitt 3% fehlerhaft sind. Wie groß ist die Wahrscheinlichkeit, dass in einer Sendung von n=100 Stück Treffer ≙ fehlerhaft,

e) Mit welcher Verteilung kann man hier die Wahrscheinlichkeiten bestimmen?
   Prüfen Sie die Approximationsmöglichkeiten genau.

a) keine Ausschuss-Stücke x=0, b) genau 5 Ausschuss-Stücke,
c) höchstens 4 Ausschuss-Stücke,
d) mindestens 8 Ausschuss-Stücke enthalten sind?

e) Das ist eine Stichprobe ohne Zurücklegen. N ist sehr groß,
   (1) wegen $n/N < 0{,}05$ kann man f_{Bin} verwenden.
   (2) p klein?"Ausschuss" ist ein seltenes Ereignis: wegen 100/0,03=3333 > 1500
   => kann man f_{Poi} verwenden.
   $\mu = 100 \cdot 0{,}03 = 3$, für diesen Erwartungswert liegt eine Tabelle vor => *Tabelle 7.2*

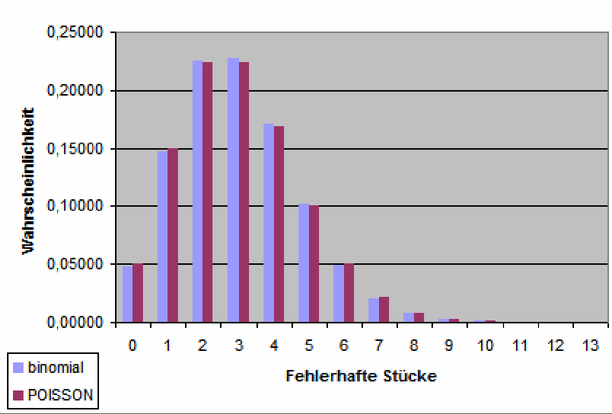
a) $W(X=0) = f_3(0) = F_3(0) = \frac{3^0}{0!} e^{-3} = 0{,}0498 \approx 5\%$

b) $W(X=5) = \sum_{x=0}^{5} \frac{3^x}{x!} e^{-3} - \sum_{x=0}^{4} \frac{3^x}{x!} e^{-3} = F_3(5) - F_3(4) = 0{,}9161 - 0{,}8153 = 0{,}1008 \approx 10\%$

c) $W(X \le 4) = \sum_{x=0}^{4} \frac{3^x}{x!} e^{-3} = F_3(4) = 0{,}8153$

d) $W(X \ge 8) = 1 - F_3(7) = 1 - 0{,}9881 = 0{,}0119$

	binomial		POISSON	
x, k	$f_{100;0,03}(x)$	$F(k)=W(X\le k)$	$f_3(x)$	$F(k)=W(X\le k)$
0	0,04755	0,04755	0,04979	0,04979
1	0,14707	0,19462	0,14936	0,19915
2	0,22515	0,41978	0,22404	0,42319
3	0,22747	0,64725	0,22404	0,64723
4	0,17061	0,81785	0,16803	0,81526
5	0,10131	0,91916	0,10082	0,91608
6	0,04961	0,96877	0,05041	0,96649
7	0,02060	0,98938	0,02160	0,98810
8	0,00741	0,99678	0,00810	0,99620
9	0,00234	0,99913	0,00270	0,99890
10	0,00066	0,99979	0,00081	0,99971
11	0,00017	0,99995	0,00022	0,99993
12	0,00004	0,99999	0,00006	0,99998
13	0,00001	1,00000	0,00001	1,00000



5.10 DISKRETE VERTEILUNGEN

Aufgabe Diskrete Verteilungen

gegeben: Sachverhalt wie in den obigen Beispielen,
(versteckte) Hinweise auf "Ziehen ohne Zurücklegen", "große *N*",
Tabellen und Approximationsbedingungen in der Formelsammlung.

gesucht / *Schritte*:

1. Passendes Modell aufgrund der Approximationsbedingungen
   entscheiden, ob f_{Hyp} *,* f_{Bin} *,* f_{Poi} *hier benutzt werden können/sollen/müssen, dazu die Liste der Approximationsbedingungen als Checklist einsetzen.*
2. Wahrscheinlichkeiten für einzelne Ereignisse oder Intervalle
   aus der Binomial-, hypergeometrischen oder/und POISSON-Verteilung.
   Vergleich zu Ergebnissen mit anderen Verteilungen.
   Formelmäßige Ansätze für die konkret gegebenen Zahlen.
   Prüfen, ob eine geeignete Tabelle vorhanden ist,
   Wahrscheinlichkeiten durch Rechnen anhand der Formeln oder mit Hilfe der Tabelle bestimmen,
   W(X ist "mindestens" k) = 1 – W(X ist "höchstens" k–1)
   Vergleichsrechnung mit anderer Wahrscheinlichkeitsverteilung durchführen (über Tabelle oder Berechnen der betreffenden Werte).
   Ansätze, indem man die konkreten Zahlen in die Formeln einsetzt.
3. Erwartungswert, erwartete Varianz, erwartete Standardabweichung
   p, q, n, evtl. N in die entsprechenden Formeln einsetzen

5.11 Normalverteilung

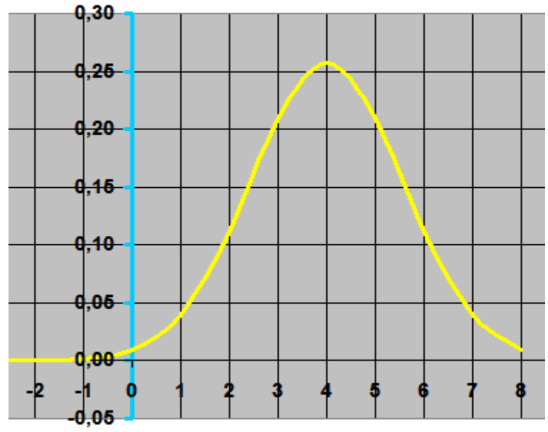
Die meisten Merkmale X lassen sich mit Hilfe der Normalverteilung untersuchen.
Das gilt besonders für Zufallsvariable X, die durch Überlagerung vieler Zufallsvariablen entstehen.
Für normalverteilte Zufallsvariable gilt

die Dichtefunktion $f_{\text{Norm}|\mu,\sigma}(x) = \dfrac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$

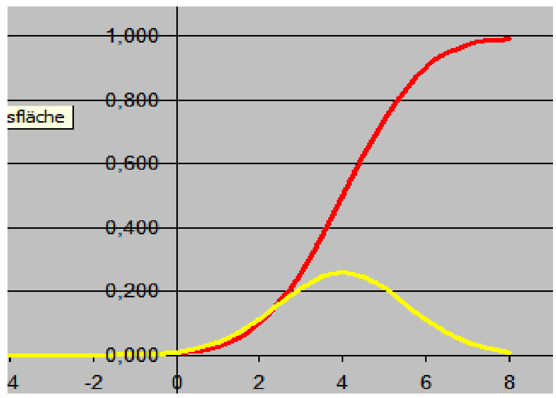
die Verteilungsfunktion $W(X \le x_2) = F_{\text{Norm}|\mu,\sigma}(x_2) = \dfrac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x_2} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx$

[De Moivre, Abraham, 1733 London, Gauss, Carl Friedrich, 1826 Göttingen]

→ *Excel / Normalverteilung*



Dichtefunktion $f_{\text{Norm}\,|\,4\,|\,1{,}55}$



Verteilungsfunktion $F_{\text{Norm}\,|\,4\,|\,1{,}55}$

Eigenschaften der Dichtefunktion $f_{\text{Norm}\,|\,\mu\,|\,\sigma}$

1. Zwei Parameter Erwartungswert μ, Standardabweichung σ
2. Die Dichten $f_{\mu,\sigma}(x)$ sind symmetrisch zur Achse $x = \mu$.
3. Der Hochpunkt liegt bei $(\mu \mid f(\mu))$, das typische zentrale Maximum
4. Die Wendepunkte bei $x_{1;2} = \mu \pm \sigma$.
5. Die Dichten $f_{\mu,\sigma}(x)$ konvergieren für $x \to \pm\infty$ gegen null.
6. Mit zunehmender "Streuung" σ wird der Funktionsgraph breiter und flacher.
7. Die Verteilungsfunktion $F_{\text{Norm}\,|\,\mu\,|\,\sigma}$ ist S-förmig, monoton steigend, $\lim\limits_{x\to\infty} F(x) = 1$.

Die **Dichtefunktion** der standardisierten Normalverteilung $f_{\text{SN}}(z) = \dfrac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} = 0{,}4e^{-\frac{1}{2}z^2}$

erhält man mit $\mu = 0$ und $\sigma = 1$ $\qquad f_{SN|0|1}(x) = \dfrac{1}{1\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-0}{1}\right)^2} = \dfrac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$

Die Dichten sind symmetrisch zur $f(z)$-Achse, Hochpunkt $(0 \mid 0{,}4)$, Wendepunkte $(1 \mid \pm 0{,}24)$.

Die **Verteilungsfunktion** $W(Z \le z_2) = F_{SN}(z_2) = \dfrac{1}{\sqrt{2\pi}} \int_{-\infty}^{z_2} e^{-\frac{1}{2}z^2} dz$ *Tabelle 7.6*

sie ist parameterfrei, eine einzige Funktion, eine einzige Tabelle!

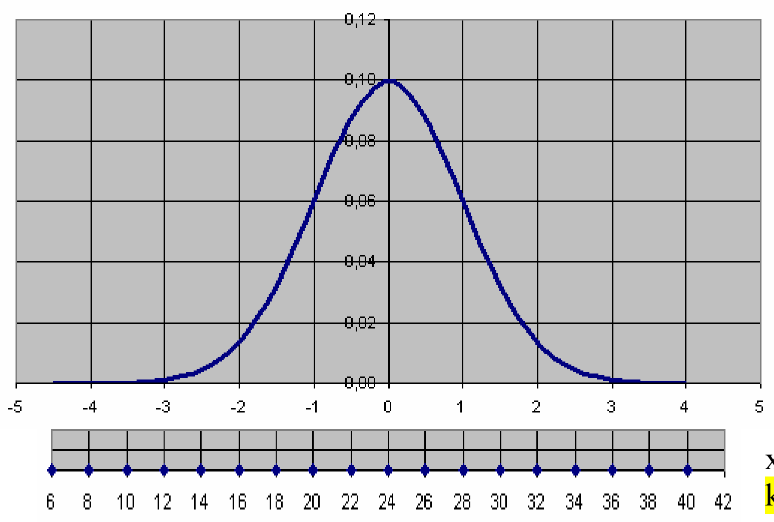
Mit den Umrechnungen $z = \dfrac{x-\mu}{\sigma}$ bzw. $x = \mu + z \cdot \sigma$

schließt man von konkreten x-Werten auf standardisierte z-Werte und umgekehrt.
Oft schreibt man auch $\varphi(z)$ statt $f_{\text{SN}}(z)$ und $\Phi(z)$ statt $F_{\text{SN}}(z)$.

5.12 "24-STUNDEN-LICHTER"

Beispiel 5.7

Die Brenndauer handelsüblicher Kerzen von Typ "24-Stunden-Licht" ist näherungsweise normalverteilt mit einem Erwartungswert μ = 24 Stunden und einer Standardabweichung von σ = 4 Stunden. Stetige Zufallsvariable!



Standardisierung:
μ = 24
σ = 4

$$\boxed{z=\frac{x-\mu}{\sigma}=\frac{x-24}{4}}$$

x = μ + z · σ

Achse z

x-Achse: Brenndauer x [Stunden]
konkrete Achse

Wie groß ist die Wahrscheinlichkeit dafür, dass eine solche Kerze
a) höchstens 30 Stunden brennt?
b) höchstens 16 Stunden brennt?
c) mehr als 16 Stunden brennt? (Gegenwahrscheinlichkeit von b)
d) mindestens 20 Stunden brennt? Stetige Zufallsvariable!
e) zwischen 24 – 1σ und 24 + 1σ Stunden brennt, also im 1-σ-Bereich liegt ?
f) zwischen 20 und 26 Stunden brennt?
g) zwischen 24–1,96 σ und 24+1,96 σ Stunden brennt? (d.h. im 95%-Bereich liegt?)
[24-7,84 ; 24 + 7,84] = [16,16 h ; 31,84 h]
h) Wir messen die Brenndauer bei 25 Kerzen, Erwartungswert und erwartete Standardabweichung des Stichprobenmittels $\bar{X}$ sind zu berechnen.

$$a)\ z=\frac{30-24}{4}=1,5\Rightarrow W(X\le 30)=F_{SN}(1,5)=0,9332 \qquad \mathrm{F_{SN}}(z)-\text{Spalte!}$$

$$\text{b) } z=\frac{16-24}{4}=-2\Rightarrow W(X\le 16)=F_{SN}(-2)=0,0228 \qquad \mathrm{F_{SN}}(-z)-\text{Spalte}$$

c) Gegenwahrscheinlichkeit in der Gegenspalte: $1-F_{SN}(-2)=F_{SN}(2)=0,9772$

$$\text{d) } z=\frac{20-24}{4}=-1\Rightarrow W(X\ge 20)=1-W(X\le 20)=1-F_{SN}(-1)=1-0,1587=0,8413$$

e) $1\sigma-$Intervall: $D(1)=0,6827=W(20\le X\le 28)=F_{SN}(1)-F_{SN}(-1)=0,8413-0,1587$

f) $\mu-4\le X\le \mu+2$ nicht symmetrisch, nicht mit D(z) $z_1=\frac{20-24}{4}=-1.\ z_2=\frac{26-24}{4}=0,5$

$$W(20\le X\le 26)=W(X\le 26)-W(X\le 20)=$$
$$=F_{SN}(0,5)-F_{SN}(-1)=0,6915-0,1587=0,5328$$

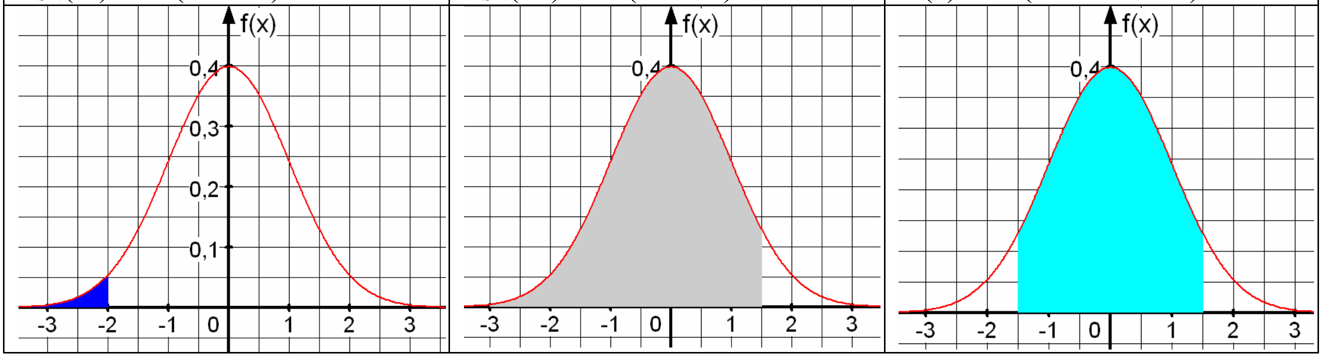
g) $1,96\sigma-$Intervall: $D(1,96)=0,9500$

h) $\mu_{\bar{X}}=\mu_x=24$ [Stunden] $\qquad \sigma_{\bar{X}}=\frac{\sigma}{\sqrt{n}}=\frac{4}{5}=0,8$ [Stunden]

5.13 Tabelle ablesen

Ablesebeispiele

$F_{SN}(-z) = W(X \le 16)$	$F_{SN}(+z) = W(X \le 30)$	$D(z) = W(18 \le X \le 30)$



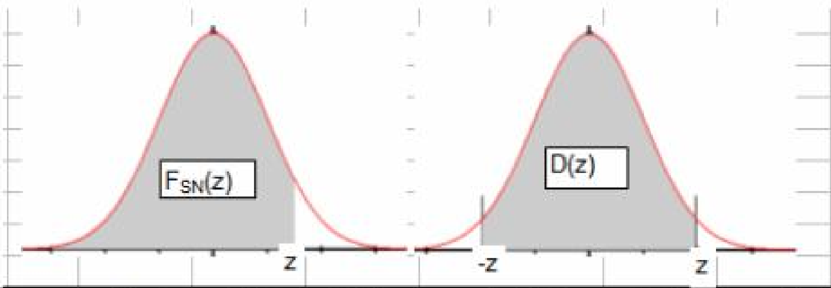
(7.6)

Standardnormal-verteilung

$F_{SN}(-z)$, $F_{SN}(z)$, $D(z)$

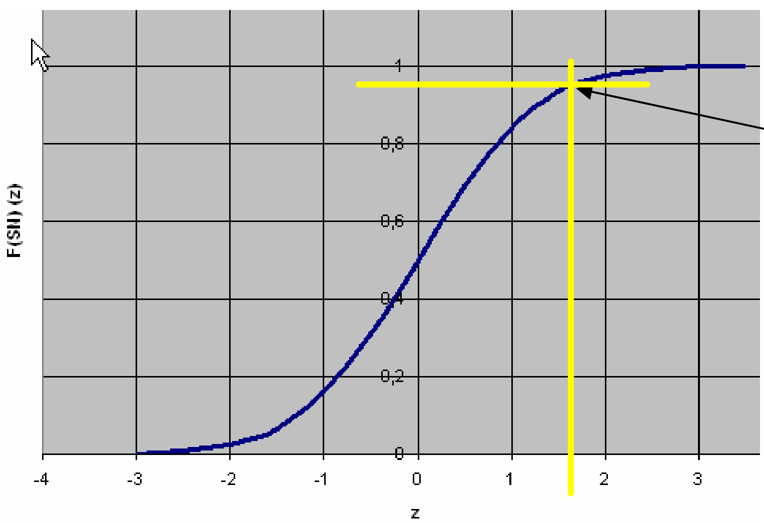
$F_{SN}(-z) = 1 - F_{SN}(z)$

$D(z) = 2\,F_{SN}(z) - 1$



z	F(-z)	F(z)	*D(z)*	z	F(-z)	F(z)	*D(z)*	z	F(-z)	F(z)	*D(z)*
1,50	0,0668	0,9332	*0,8664*	2,00	0,0228	0,9772	*0,9545*	2,50	0,0062	0,9938	*0,9876*
1,51	0,0655	0,9345	*0,8690*	2,01	0,0222	0,9778	*0,9556*	2,51	0,0060	0,9940	*0,9879*
1,52	0,0643	0,9357	*0,8715*	2,02	0,0217	0,9783	*0,9566*	2,52	0,0059	0,9941	*0,9883*
1,53	0,0630	0,9370	*0,8740*	2,03	0,0212	0,9788	*0,9576*	2,53	0,0057	0,9943	*0,9886*
1,54	0,0618	0,9382	*0,8764*	2,04	0,0207	0,9793	*0,9586*	2,54	0,0055	0,9945	*0,9889*
1,55	0,0606	0,9394	*0,8789*	2,05	0,0202	0,9798	*0,9596*	2,55	0,0054	0,9946	*0,9892*

<u>Verteilungsfunktion der standardisierten Normalverteilung</u> $F_{SN}(z) = \dfrac{1}{\sqrt{2\pi}} \displaystyle\int_{-\infty}^{z_2} e^{-\frac{1}{2}z^2}\,dz$



a) für $z > 0$ als Tabelle $F_{SN}(z)$ tabelliert, z.B. $F_{SN}(1{,}7) = 0{,}9554$

b) $F_{SN}(z)$ konvergiert für $z \to \infty$ nach $1 = 100\%$.

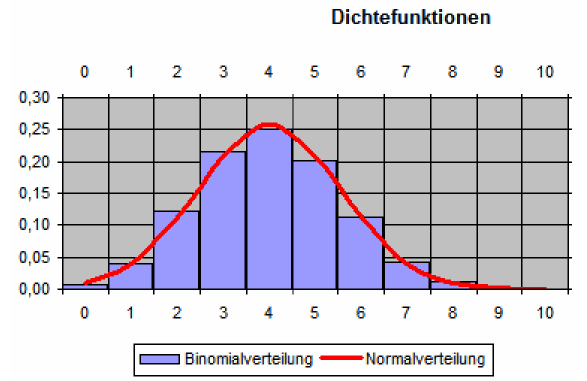
c) Wendepunkt bei $W(0 \mid 0{,}5)$

5.14 STETIGKEITSKORREKTUR

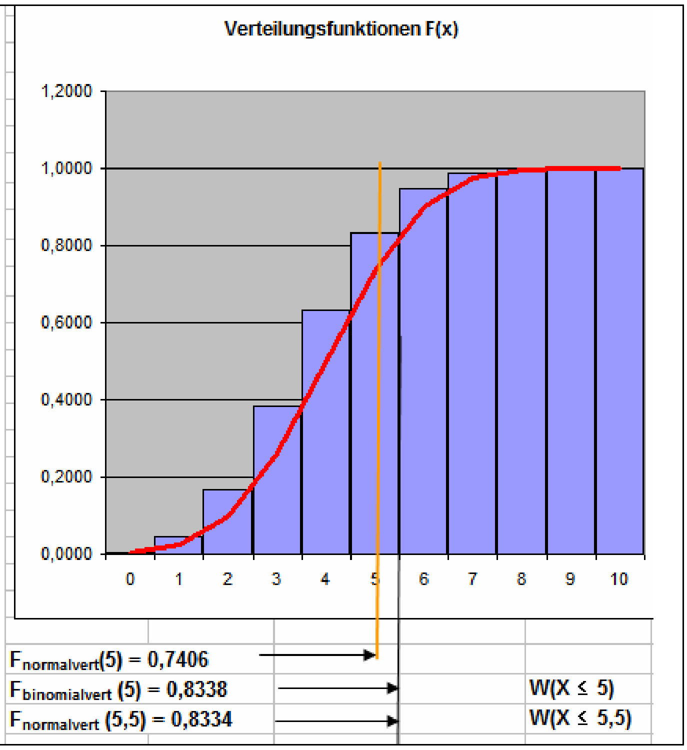
Für einigermaßen normalverteilte stetige Zufallsvariablen benutzt man die Normalverteilung. Unter bestimmten Approximationsbedingungen geht das auch für diskrete Zufallsvariable.

Beispiel 5.2 noch einmal

An einem erosionsgefährdeten Hang sollen 10 Fichtenpflanzen angepflanzt werden. Die Baumschule gibt eine Anwachswahrscheinlichkeit von $p = 40\%$ an. Wie groß ist die Wahrscheinlichkeit, dass höchstens 5 anwachsen?



Man kann die Mittelpunkte der Rechteckoberkanten im Histogramm zur Binomialverteilung verbinden. Diese "Hüllkurve" ist eine Approximation der Dichtefunktion der Normalverteilung.
Ganz entsprechend gilt das für die Verteilungsfunktionen $F_{\text{Bin}}(k)$ und $F_{\text{Norm}}(x)$



$$F_{\text{Bin}\,|\,10\,|\,0,4}(5) = 0{,}8338 \quad = \sum_{k=0}^{5} \binom{10}{k} 0{,}4^k 0{,}6^{10-k} \qquad \mu = 10 \cdot 0{,}4 = 4 \qquad \sigma = \sqrt{npq}$$

$$\mu = 10 \cdot 0{,}4 = 4. \;\; \sigma^2 = 10 \cdot 0{,}4 \cdot 0{,}6 = 2{,}4. \;\; \sigma = 1{,}55. \quad z = \frac{5-4}{1{,}55} = 0{,}645. \;\; F_{\text{SN}}(0{,}645) = 0{,}7406$$

mit Stetigkeitskorrektur: $$z = \frac{x+0{,}5-\mu}{\sigma} = \frac{5+0{,}5-4}{1{,}55} = \frac{5{,}5-4}{1{,}55} = 0{,}967 \quad F_{\text{SN}}(0{,}967) = 0{,}8334$$

Wenn man diskrete Zufallsvariable mit der Normalverteilung untersucht, verbessert die Stetigkeitskorrektur den Näherungswert.

Standardnormalvariable z mit Stetigkeitskorrektur: $$z = \frac{x+0{,}5-\mu}{\sigma}$$

weiter Beispiel 5.2

Eigentlich darf man für eine binomialverteilte Zufallsvariable nur dann die Normalverteilung als Näherung benutzen, wenn gilt Varianz $\sigma^2 = npq > 9$. $\sigma > 3$. LAPLACE-Bedingung.
Hier ist $\sigma^2 = 10 \cdot 0{,}4 \cdot 0{,}6 = 2{,}4 < 9$.
Mit der Stetigkeitskorrektur ist trotzdem eine recht gute Näherung erreicht worden!

□ Nun sollen an dieser Stelle 40 Fichtenpflanzen angepflanzt werden.
Wie groß ist die Wahrscheinlichkeit, dass höchstens 20 Pflanzen anwachsen?
$\sigma^2 = n\,p\,q = 40 \cdot 0{,}4 \cdot 0{,}6 = 9{,}6 > 9$ approximierbar

$$W(X \le 20) = F_{SN}(z) = F_{SN}\left(\frac{20+0{,}5-16}{3{,}098}\right) = F_{SN}(1{,}45) = 0{,}9265 \quad \text{zum Vergleich: } F_{\text{Bin}}(20) = 0{,}926$$

5.15 Aufgaben zur Normalverteilung

Aufgabe Normalverteilung

gegeben: Sachverhalt mit diskreter (binomialverteilter) oder stetiger Zufallsvariablen. Tabellen und Approximationsbedingungen in der Formelsammlung.

gesucht / *Schritte:*

1. Approximationsbedingungen für die Normalverteilung prüfen
   stetige Zufallsvariable: ohne weiteres anwenden
   diskrete Zufallsvariable:
   $$\text{Hypergeom.Vert.} \xrightarrow{n/N \leq 0{,}05} \text{Binomial-Vert.} \xrightarrow{n/p \geq 1500} \text{Poisson-Vert.}$$
   $$\text{Binomial.Vert} \xrightarrow{n\,p\,q > 9} \text{Normalverteilung}$$
   dann mit Stetigkeitskorrektur rechnen

2. Wahrscheinlichkeiten für bestimmte Intervalle
   Formelmäßige Ansätze für die konkret gegebenen Zahlen.
   jeweils benötigte Standard-Normalvariable z brechnen aus:
   $$z = \frac{x-\mu}{\sigma} \,\big|\,\underline{\text{stetig}} \quad \text{bzw.} \quad z = \frac{x+0{,}5-\mu}{\sigma} \,\big|\,\underline{\text{diskret}}$$ *berechnen,*
   damit $F_{SN}(z)$ oder $D_{SN}(z)$ berechnen,
   wie in den Beispielen 5.7 und 5.2 (Abschnitt 5.14)
   W(X ist "mindestens" k) = 1 – W(X ist "höchstens" k)

3. μ, σ^2, σ für das Stichprobenmittel
   $$\mu_{\bar{X}} = \mu_x \,,\; \sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$$ *anwenden*

4. Eigenschaften der Normalverteilung und der Standard-Normalverteilung in Zusatzfragen
   überlegt man sich am leichtesten, indem man sich die Funktionsgraphen zu den Dichtefunktionen $f_{Norm\,|\,\mu\,|\,\sigma}(x)$ bzw. $f_{SN}(z)$ vorstellt oder skizziert.

5.16 Approximationsbedingungen

Übergang von der

Hypergeometrischen V.	zur Binomial-V.,	wenn $n/N \leq 0{,}05$
Binomial-V.	zur Poisson-V.,	wenn $n/p \geq 1500$
Hypergeometrischen V.	zur Poisson-V.,	wenn $n/N \leq 0{,}05$ und $n/p \geq 1500$

Binomial-V	zur Normalverteilung, wenn $\sigma^2 = n\,p\,q > 9$ (Laplace-Bedingung)
Hypergeometrischen V.	zur Normalverteilung, wenn $n/N \leq 0{,}05$ und $\sigma^2 = n\,p\,q > 9$
Poisson-V.	zur Normalverteilung, wenn $\mu = \sigma^2 > 9$
Student-t-V.	zur Normalverteilung, wenn $n > 30$, bei normalverteilter Grundgesamtheit wenn $n > 50$, bei unbekannter Verteilung der Grundgesamtheit

→ *Excel / Normalverteilung / Zeile 117*

Anpassung und Korrekturfaktoren

Diskrete Zufallsvariable X $\quad z = \dfrac{x + 0{,}5 - \mu}{\sigma}$ (Stetigkeitskorrektur)

Stichprobenmittel $\sigma_{\bar{x}}$ $\quad \sigma_{\bar{x}} = \dfrac{\sigma}{\sqrt{n}} \Rightarrow z = \dfrac{\bar{x} - \mu}{\sigma}\sqrt{n}$

$n/N > 0{,}05$: $\quad \sigma_{\text{korrigiert}} = \sigma \cdot \sqrt{\dfrac{N-n}{N-1}}$

σ unbekannt, n < 30 bzw. n < 50 $\quad t = \dfrac{x - \mu}{s}$ bzw. $t = \dfrac{\bar{x} - \mu}{s}\sqrt{n}$

σ wird durch die Standardabweichung s aus der Stichprobe ersetzt (*Thema 6*)

Binomial
Hypergeometrisch
Poisson

Standard-Normalverteilung
$f_{SN}(z)$, $F_{SN}(z)$

unter den Approximationsbedingungen:

$n/N \leq 0{,}05$
$\sigma^2 = n\,p\,q > 9$

4 Anpassungsprobleme

1. Umrechnung konkret – standardisiert
$$z = \frac{x - \mu}{\sigma}$$
2. Für Stichprobenmittelwerte $\bar{x}$ statt Einzelobjekte x
$$z = \frac{\bar{x} - \mu}{\sigma} \cdot \sqrt{n}$$
3. Für ohne Zurücklegen bei n/N > 0,05
$$z = \frac{\bar{x} - \mu}{\sigma \cdot \sqrt{\frac{N-n}{N-1}}} \cdot \sqrt{n}$$
4. σ unbekannt, ersatzweise s aus der Stichprobe nutzen:
$$t = \frac{\bar{x} - \mu}{s} \cdot \sqrt{n}$$

5.17 CHI²-, STUDENT-T-, FISHER-F- VERTEILUNG

a) Chi²-, χ^2-Verteilung [Pearson, Karl, chi², London, ca. 1914]

Eine stetige Zufallsvariable X ist **χ^2-verteilt** mit ν Freiheitsgraden, wenn ihre Dichtefunktion

$$f_\nu(\chi^2) = \frac{1}{2^{\frac{\nu}{2}} \cdot \Gamma(\frac{1}{2}\nu)} \cdot (\chi^2)^{\frac{\nu}{2}-1} \cdot e^{-\frac{\chi^2}{2}} \qquad \text{mit} \quad \Gamma(x) = \int_0^\infty e^{-t} \cdot t^{x-1} dt \quad \text{(Gammafunktion)}$$

→ *Tabelle 7.4*

Es seien $X_1, X_2, \ldots, X_\nu$ ν unabhängige standardnormalverteilte Zufallsvariable.

Dann ist die Zufallsvariable $C = X_1^2 + X_2^2 + \ldots + X_\nu^2$ χ^2-verteilt.

Anwendungen:

(1) Chi²-Unabhängigkeitstest, zum Prüfen nominaler Merkmale auf Unabhängigkeit *(4.4)*

(2) Chi²-Anpassungstest, zum Prüfen, ob eine Zufallsvariable normalverteilt ist *(5.17)*

(3) Varianztest, zum Testen, ob $s^2_{\text{Stichprobe}}$ signifikant von $\sigma^2_{\text{Grundgesamtheit}}$ abweicht.

b) t-Verteilung (STUDENT-t-Verteilung) [GOSSET, WILLIAM, "STUDENT", Dublin, Irland, 1908]

Eine stetige Zufallsvariable X ist **t-verteilt** mit ν Freiheitsgraden, wenn ihre Dichtefunktion

$$f(t) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\pi \cdot \nu} \cdot \Gamma\left(\frac{\nu}{2}\right)} \cdot \left(1 + \frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}} \qquad \text{mit} \quad \Gamma(x) = \int_0^\infty e^{-t} \cdot t^{x-1} dt \qquad \text{(Gammafunktion)}$$

→ *Tabelle 7.5*

Es sei X eine standardnormalverteilte Zufallsvariable und C eine χ^2-verteilte Zufallsvariable mit ν Freiheitsgraden. X und C sind unabhängige Zufallsvariable.

Dann ist die Zufallsvariable $T = \frac{X}{\sqrt{\frac{C}{\nu}}}$ student-t-verteilt mit ν Freiheitsgraden.

Anwendungen:

(1) Multiple Regressionsanalyse, zum Prüfen, ob ein Einflussfaktor einen signifikanten Beitrag zur Regression liefert. *(Abschnitt 1.14)*

(2) Hypothesentests, zum Prüfen, ob $x\gamma_{\text{Stichprobe}}$ signifikant von $\mu_{\text{Grundgesamtheit}}$ abweicht, anstelle der Normalverteilung, wenn $\sigma_{\text{Grundgesamtheit}}$ unbekannt ist. *(Abschnitt 6.3)*

c) FISHER -F-Verteilung [FISHER, RONALD, Rothamsted, GB, 1918]

Die Zufallsvariable X sei χ^2-verteilt mit p Freiheitsgraden und die von X unabhängige Zufallsvariable Y sei χ^2-verteilt mit ν Freiheitsgraden, dann ist die Zufallsvariable Z

$$Z = \frac{\frac{1}{p}X}{\frac{1}{\nu}Y}$$ FISHER-F-verteilt mit den Freiheitsgraden p und ν

→ *Tabelle 7.3*

Anwendungen:

(1) Korrelationsanalyse, zum Prüfen, ab ein signifikanter Zusammenhang besteht *(1.11)*.

(2) Varianzanalysen, z.B. zum Testen, ob $s^2_{\text{Stichprobe A}}$ signifikant von $s^2_{\text{Stichprobe B}}$ abweicht.

5.18 IST X NORMALVERTEILT ?

Schon oft hieß es "normalverteilte" Zufallsvariable.
Es ist zu prüfen, ob eine Zufallsvariable "einigermaßen" normalverteilt ist.
Wir werten eine möglichst große Stichprobe aus => k Klassen, Häufigkeiten n_i, xγ, s.
Wir vergleichen diese beobachteten Häufigkeiten n_i mit den theoretischen Häufigkeiten u_i.
Genau so sind wir beim Chi²-Unabhängigkeitstest vorgegangen *(Abschnitt 4.4)*.
Als Testgröße bilden wir wieder die Summe der normierten Abweichungsquadrate

$$chi^2_{empirisch} = \chi^2_{empirisch} = \sum_{\mathrm{i}=1}^{\mathrm{k}} \frac{(n_\mathrm{i} - u_\mathrm{i})^2}{u_\mathrm{i}}$$ und testen mit dem Prüfmaß χ^2_{crit}.

Der Test heißt Chi²-Anpassungstest, weil die Güte der Anpassung einer theoretischen Verteilung an eine empirische Verteilung überprüft wird.
Die Nullhypothese H_0 lautet: die zu prüfende Grundgesamtheit gehorcht der Normalverteilung.
Wir sprechen besser von **Verteilungshypothese**, die zu bestätigen oder abzulehnen ist.
Auf ähnliche Weise kann man prüfen, ob eine Zufallsvariable einer anderen theoretischen Verteilung gehorcht, etwa ob sie POISSON-verteilt ist, oder ob sie jene Dichtefunktion für die Verspätung X der U-Bahn hat…

Beispiel 5.8 *(nach Bleymüller S.129)*
Es soll geprüft werden, ob die "Lebensdauer" X eines bestimmten Akku-Typs normalverteilt ist.
Es wurde eine Stichprobe im Umfang $n = 80$ Stück ausgewertet. Signifikanzniveau $\alpha = 0{,}05$.
Die stetige Zufallsvariable Lebensdauer wurde in 7 Merkmalsklassen gegliedert: $k = 7$.
Als mittlere Lebensdauer ergibt sich $\bar{x} = 3{,}41$ [Jahre], die Standardabweichung ist $s = 0{,}7$ [Jahre].

1. Zunächst eine Arbeitstabelle mit der Häufigkeitsverteilung der k Klassen $[x_i^{unten} ; x_i^{oben}]$.
   $\bar{x}$ und s sind schon gegeben, die Klassenmitten x_i^* und die Spalte $x_i^2 \cdot n_i$ benötigen wir nicht.
   → *Excel / Anpassungstest*

2. Wir benötigen die Wahrscheinlichkeiten $F_{SN}(z_i)$ für die Intervalle $-\infty < X \le x_i^{oben}$.
   Dazu bestimmen wir die Standardnormalvariablen $z_\mathrm{i} = \frac{x_\mathrm{i}^{oben} - \bar{x}}{s}$.
   Für diese z_i lesen wir aus der Tabelle F_{SN} die Werte $F_{SN}(z_i)$ ab.

3. Wir benötigen die Wahrscheinlichkeiten (= theoretische relative Häufigkeiten) w_i für die Klassen i. Wir erhalten die w_i durch Differenzenbildung $w_i = F_{SN}(z_i) - F_{SN}(z_{i-1})$.
   Dabei setzen wir $F_{SN}(z_0) = 0$.

4. Wir benötigen die theoretischen absoluten Häufigkeiten $u_i = n \cdot w_i$. <= $w_i = u_i / n$
   Beim Chi²-Anpassungstest wird vorausgesetzt, dass in jeder Klasse $u_i \ge 5$ ist.
   Wenn dies nicht der Fall ist, dann müssen Klassen entsprechend zusammengefasst werden.

5. Zum Vergleichen der empirischen Häufigkeiten n_i mit den theoretischen Häufigkeiten u_i bildet man die Summe der normierten Abweichungsquadrate $\chi^2_{empirisch} = \sum_{\mathrm{i}=1}^{\mathrm{k}} \frac{(n_\mathrm{i} - u_\mathrm{i})^2}{u_\mathrm{i}}$

6. Zum Ablesen des Prüfmaßes χ^2_{crit} benötigen wir den Sicherheitsgrad $1 - \alpha$
   und die Freiheitsgrade: $\nu = k - p - 1$.
   k ist die Anzahl der Klassen, p ist die Anzahl der Parameter ($x\gamma$, s) , die aus der Stichprobe ermittelt wurden. Wenn $\mu_{\mathrm{Grundgesamtheit}}$ gegeben ist, dann ist $k = 1$.

7. Die Verteilungshypothese bestätigen wir, wenn $\chi^2_{\mathrm{empirisch}} \le \chi^2_{\mathrm{crit.}}$ (H_0 wir beibehalten).

5.19 CHI² - ANPASSUNGSTEST

Aufgabe Anpassungstest

gegeben: Häufigkeitsverteilung mit k Klassen und den Häufigkeiten n_i.
$\bar{x}$, s aus der Stichprobe oder μ, σ aus der Grundgesamtheit.
Verteilungshypothese. Signifikanzsniveau (Irrtums-Wahrscheinlichkeit) α.

gesucht: Es ist zu prüfen, ob die untersuchte Zufallsvariable normalverteilt ist.
Arbeitstabelle und Zwischenschritte sind anzugeben.

Schritte: Wie auf der vorangehenden Seite dargestellt.

a) Arbeitstabelle mit Spalte x_i^{oben}

b) Spalte Standardnormalvariablen $z_i = \dfrac{x_i^{oben} - \bar{x}}{s}$

c) Spalte $F_{SN}(z_i)$

d) Spalte $w_i = F_{SN}(z_i) - F_{SN}(z_{i-1})$

e) Spalte $u_i = n \cdot w_i$ Klassen zusammenfassen, wenn $u_i \geq 5$

f) Spalte $\dfrac{(n_i - u_i)^2}{u_i}$

g) Summe bilden: $\chi^2_{empirisch} = \sum_{i=1}^{k} \dfrac{(n_i - u_i)^2}{u_i}$

h) Freiheitsgrade berechnen $\nu = k - p - 1$.
p ist 0, 1 oder 2 je nachdem wie viele Parameter aus der Stichprobe kommen.

i) χ^2_{crit} *in Tabelle 7.4 ablesen:* $\chi^2_{crit\,|\,1-\alpha\,|\,\nu}$. *(bei Sicherheitsgrad 1 – α)*

j) Verteilungshypothese bestätigen, wenn $\chi^2_{empirisch} \leq \chi^2_{crit.}$ *, andernfalls ablehnen*