Grundlagen der Wahrscheinlichkeitstheorie, Teil II

Wahrscheinlichkeitstheorie II: Gliederung

1 Grundlagen von Wahrscheinlichkeitsverteilungen

(Zufallsvariable, Wahrscheinlichkeitsfunktion, Dichte, Verteilungsfunktion, Parameter einer Verteilung, Erwartungswert)

2 Diskrete Wahrscheinlichkeitsverteilungen

(Gleichverteilung, Binomialverteilungen)

3 Stetige Wahrscheinlichkeitsverteilungen

(Normalverteilung, χ^2-, t- und F-Verteilung)

4 Tabellierungen & Umrechnungen

Einführende Literatur

- 🕮 Bortz, J. & Schuster, Ch. (2010). *Statistik für Human- und Sozialwissenschaftler* (7. Auflage). Berlin: Springer. [Kap. 5]
- 🕮 Eid, M., Gollwitzer, M. & Schmitt, M. (2010). *Statistik und Forschungsmethoden*. Weinheim: Beltz. [Kap. 7]
- 🕮 Holling, H. & Gediga, G. (2013). *Statistik – Wahrscheinlichkeitstheorie und Schätzverfahren*. Göttingen: Hogrefe. [Kap. 3 & 4]

Weiterführende Literatur

- 🕮 Kreyszig, E. (1979). *Statistische Methoden und ihre Anwendungen* (7. Auflage). Göttingen: Vandenhoek & Ruprecht.

Wahrscheinlichkeitsverteilungen: Zufallsvariablen

- Eine Zufallsvariable (random variable) ist eine Funktion X, die den Ergebnissen eines Zufallsexperiments reelle Zahlen (Realisationen, Messwerte) zuordnet: $\Omega \to \mathbb{R}, \omega \mapsto X(\omega)$

- Beispiele:

  - Münzwurf: Ω = {Kopf, Zahl} mit $X(\omega) = 0$ bei Kopf und $X(\omega) = 1$ bei Zahl.
  - Auswahl einer Person aus einer Gruppe Ω von Personen und Messung seiner Körpergröße: $X(\omega)$ = Körpergröße von ω.

- Häufig interessieren uns auch Funktionen von Zufallsvariablen, die wieder Zufallsvariablen darstellen.

- Beispiele:

  - Wir würfeln 10 Mal. Die Summe der Augenzahlen ist wieder eine Zufallsvariable.
  - Wir bestimmen die Intelligenzwerte an einer Stichprobe von 100 Schülern und berechnen dann den Mittelwert, der wieder eine Zufallsvariable darstellt.

Wahrscheinlichkeitsverteilungen: Zufallsvariablen

- Man unterscheidet diskrete und stetige Zufallsvariablen.
- Zufallsvariablen X sind diskret, wenn sie nur endlich (oder abzählbar unendlich) viele Werte annehmen können.
- Beispiele: Münzwurf, Studienfach, Präferenz für eine Partei ...

- Zufallsvariablen X sind stetig (oder kontinuierlich), wenn die Zahl der Ergebnisse überabzählbar unendlich sind.
- Stetige Variablen kommen streng genommen nicht vor. So ist z.B. die Menge von Personen Ω, aus der wir eine Person ziehen und ihre Körpergröße erfassen, endlich. Auch kann die Körpergröße nicht beliebig genau gemessen werden.
- Da sich mit stetigen Zufallsvariablen besser operieren lässt, sieht man solche Variablen häufig als (angenähert) stetig an.
- Beispiele: Körpergröße, Intelligenz, ...

Diskrete Wahrscheinlichkeitsverteilungen

- Wahrscheinlichkeitsverteilungen kann man entnehmen, mit welcher Wahrscheinlichkeit Werte von Zufallsvariablen auftreten.

- Bei diskreten Zufallsvariablen ist die Wahrscheinlichkeitsverteilung durch eine Wahrscheinlichkeitsfunktion bestimmt, die angibt, mit welcher Wahrscheinlichkeit die Realisierungen x der Zufallsvariablen X auftreten. Man schreibt: $f(x) = P(X=x)$.

- Wird die Wahrscheinlichkeitsfunktion kumuliert, so spricht man von einer Verteilungsfunktion und schreibt $F(x) = P(X \leq x) = \sum_{x_i \leq x} f(x_i)$.

- Beispiel. Zufallsexperiment: 5-maliger Münzwurf. X = Häufigkeit des Auftretens von Kopf.

Ω	x	$f(x)$	$F(x)$
(Z, Z, Z, Z, Z) →	0	1/32=0.03125	0.03125
(K, Z, Z, Z, Z) →	1	5/32=0.15625	0.18750
(Z, K, Z, Z, Z) ↗	2	10/32=0.3125	0.50000
…	3	10/32=0.3125	0.81250
(K, K, K, K, Z) →	4	5/32=0.15625	0.96875
(K, K, K, K, K) →	5	1/32=0.03125	1

$|\Omega| = 2^5 = 32$

Es gilt also: $X(\omega_1) = 0$, $X(\omega_2) = X(\omega_3) = 1$, … $X(\omega_{32}) = 5$

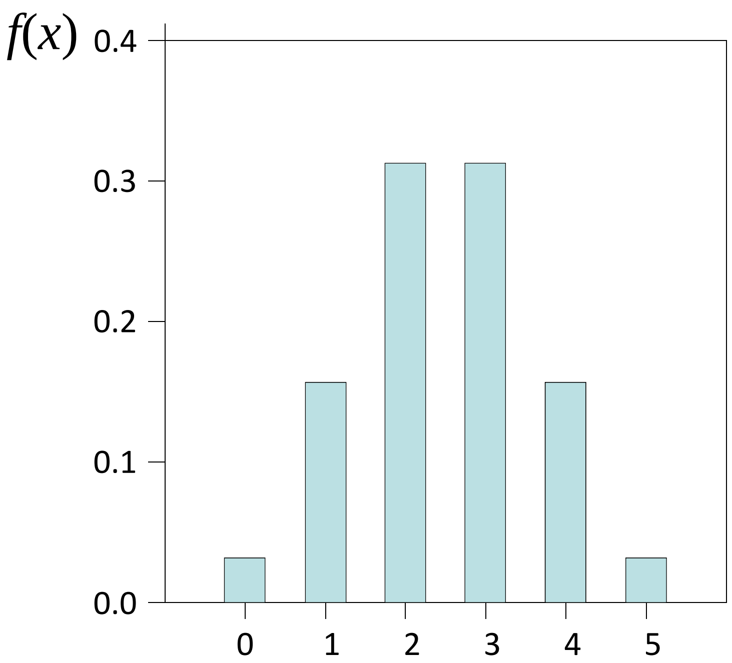
Nach Laplace resultiert: $f(0) = P(X=0) = 1/32 = 0.03125$ usw.

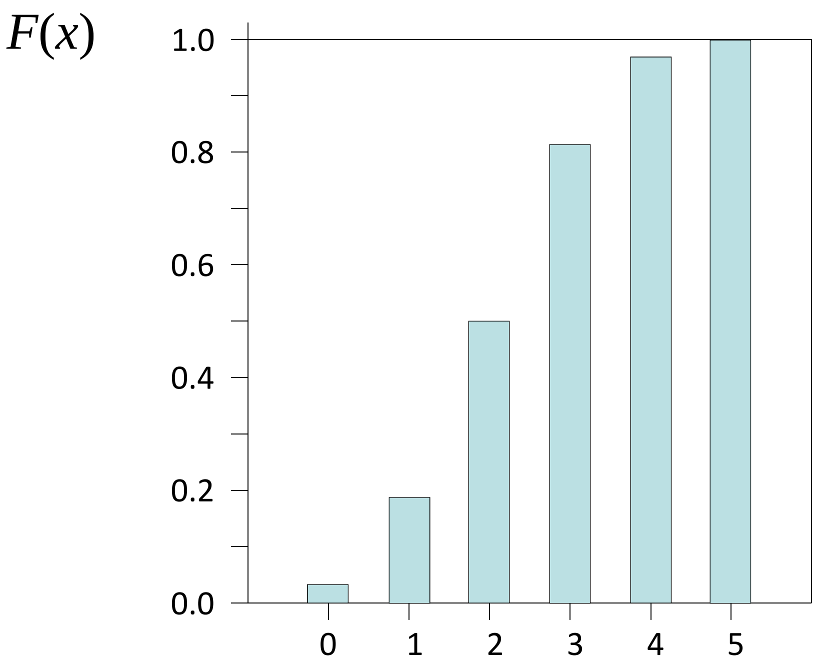
Diskrete Wahrscheinlichkeitsverteilungen

- Wahrscheinlichkeitsfunktion $f(x) = P(X = x)$ und Verteilungsfunktion $F(x) = P(X \leq x)$ einer Zufalls-variable X lassen sich in bekannter Weise auch graphisch darstellen (z.B. als Histogramm):

- Beispiel Zufallsexperiment 5-maliger Münzwurf. X = Häufigkeit des Auftretens von Kopf.

x	$f(x)$	$F(x)$
0	0.03125	0.03125
1	0.15625	0.18750
2	0.3125	0.50000
3	0.3125	0.81250
4	0.15625	0.96875
5	0.03125	1





Parameter diskreter Wahrscheinlichkeitsverteilungen

- Zur Charakterisierung von Verteilungen dienen, wie in der deskriptiven Statistik, bestimmte Kennwerte, die hier als Parameter bezeichnet werden.

- Die zentrale Tendenz einer Verteilung kann durch den Erwartungswert E charakterisiert werden. Für diskrete Zufallsvariablen X mit m Ausprägungen gilt:

$$\mathrm{E}(X) = \sum_{i=1}^{m} x_i \cdot P(X = x_i) = \sum_{i=1}^{m} x_i \cdot f(x_i)$$

- Der Erwartungswert ist dem arithmetischen Mittelwert sehr ähnlich; er entspricht bei diskreten Verteilungen dem Mittelwert aller möglichen k Werte. Anders als das arithmetische Mittel wird er aber nicht anhand von Daten ermittelt, sondern charakterisiert die Lage einer theoretischen Verteilung.

- Beispiel Zufallsexperiment: 5-maliger Münzwurf mit $k = 2^5 = 32$ Ergebnissen. X = Häufigkeit des Auftretens von Kopf mit $m = 6$.

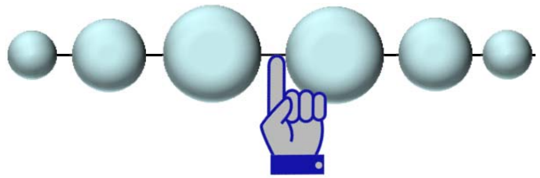
$$\mathrm{E}(X) = 0 \cdot 0.3125 + 1 \cdot 0.15625 + \ldots + 5 \cdot 0.03125 = 2.5$$

x	$f(x)$	$F(x)$
0	0.03125	0.03125
1	0.15625	0.18750
2	0.3125	0.50000
3	0.3125	0.81250
4	0.15625	0.96875
5	0.03125	1

Parameter diskreter Wahrscheinlichkeitsverteilungen

- Man beachte, dass der Erwartungswert nicht das „erwartete“ oder wahrscheinlichste Ergebnis eines Zufallsexperiments ist. Im Beispiel tritt der Erwartungswert nicht einmal als Realisation auf.

- Eine bildliche Vorstellung von der Bedeutung des Erwartungswertes erhält man, wenn man sich Kugeln auf einer Achse vorstellt, wobei x_i der Abstand der i-ten Kugel von einem beliebig gewählten Nullpunkt darstellt und $P(X=x_i)$ dem Gewicht (hier: Größe) der Kugel proportional gewählt wird. An der Stelle der Achse, die man unterstützten müsste, damit sich die Achse im Gleichgewicht befindet (im Schwerpunkt des Systems) liegt der Erwartungswert.



- Als Maß für die Variabilität einer Verteilung kann man deren Varianz bestimmen. Für diskrete Zufallsvariablen X mit m Ausprägungen gilt:

$$\mathrm{Var}(X)=\sum_{i=1}^{m}[x_i-\mathrm{E}(X)]^2\cdot P(X=x_i)$$

- Beispiel wie oben. $m=6$, $\mathrm{E}(X)=2.5$.

$$\begin{aligned}\mathrm{Var}(X) &= (0-2.5)^2\cdot 0.03125+(1-2.5)^2\cdot 0.15625+\ldots\\ &\quad +(5-2.5)^2\cdot 0.03125=1.25\end{aligned}$$

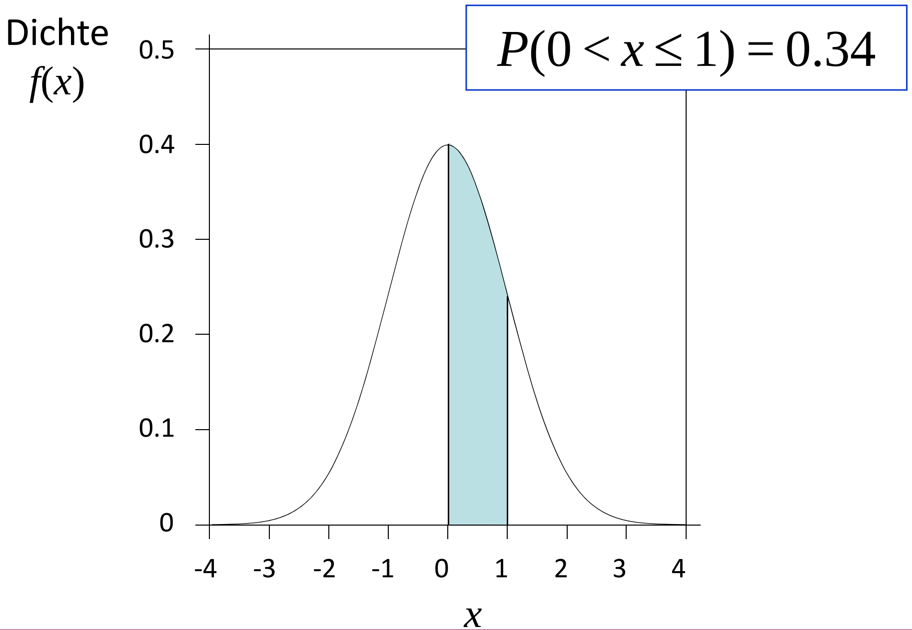
x	$f(x)$	$F(x)$
0	0.03125	0.03125
1	0.15625	0.18750
2	0.3125	0.50000
3	0.3125	0.81250
4	0.15625	0.96875
5	0.03125	1

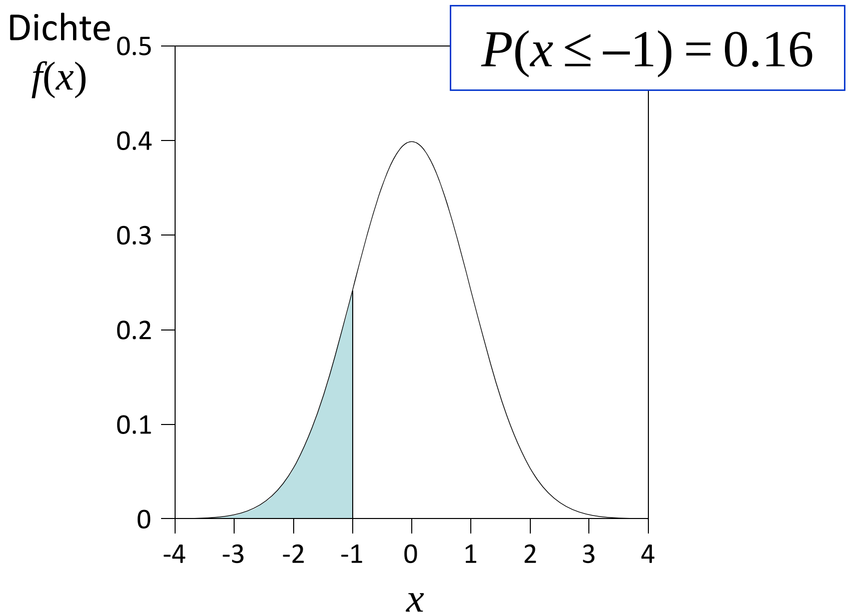
- Einige nützliche Rechenregeln für Erwartungswerte und Varianzen für (sowohl diskrete als auch stetige) Zufallsvariablen X und Y und reelle Konstanten c und d sind:

  - $\mathrm{E}(c) = c$
  - $\mathrm{E}(c \cdot X) = c \cdot \mathrm{E}(X)$
  - $\mathrm{E}(c \cdot X + d \cdot Y) = c \cdot \mathrm{E}(X) + d \cdot \mathrm{E}(Y)$
  - $\mathrm{Var}(X) = \mathrm{E}(X^2) - \mathrm{E}(X)^2$
  - $\mathrm{Var}(c \cdot X + d) = c^2 \cdot \mathrm{Var}(X)$
  - $\mathrm{Var}(c \cdot X + d \cdot Y) = c^2 \cdot \mathrm{Var}(X) + d^2 \cdot \mathrm{Var}(Y) + 2 \cdot c \cdot d \cdot \mathrm{Cov}(X, Y)$
  - $\mathrm{Var}(c \cdot X + d \cdot Y) = c^2 \cdot \mathrm{Var}(X) + d^2 \cdot \mathrm{Var}(Y)$ falls X und Y unabhängig/unkorreliert sind.

Stetige Wahrscheinlichkeitsverteilungen

- Bei stetigen Verteilungen bezeichnet man $f(x)$ nicht mehr als Wahrscheinlichkeit (sfunktion) sondern als Dichte der Zufallsvariablen.
- Die Wahrscheinlichkeit für eine einzelne Realisation einer Zufallsvariablen, also z.B. dem Wert von $x = 1.782543$ Meter Körpergröße, ist $P(X = x) = 0$. (Die Dichten summieren sich also nicht zu 1, was aber von Wahrscheinlichkeiten gefordert wird; vgl. Axiom 2.)
- Wahrscheinlichkeiten existieren immer nur für Intervalle von Dichten, unten am Beispiel der Standardnormalverteilung.





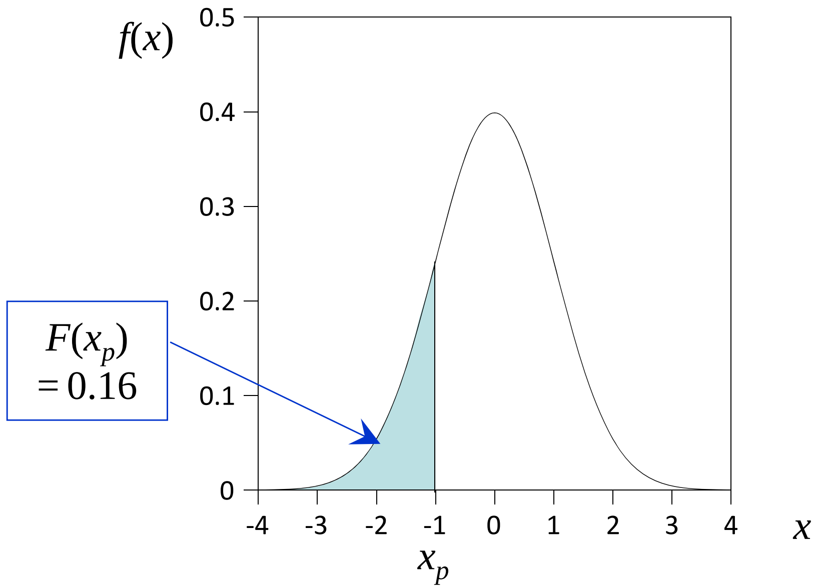
- Formal entspricht der Fläche zwischen zwei Werten a und b dem Integral der Dichten:

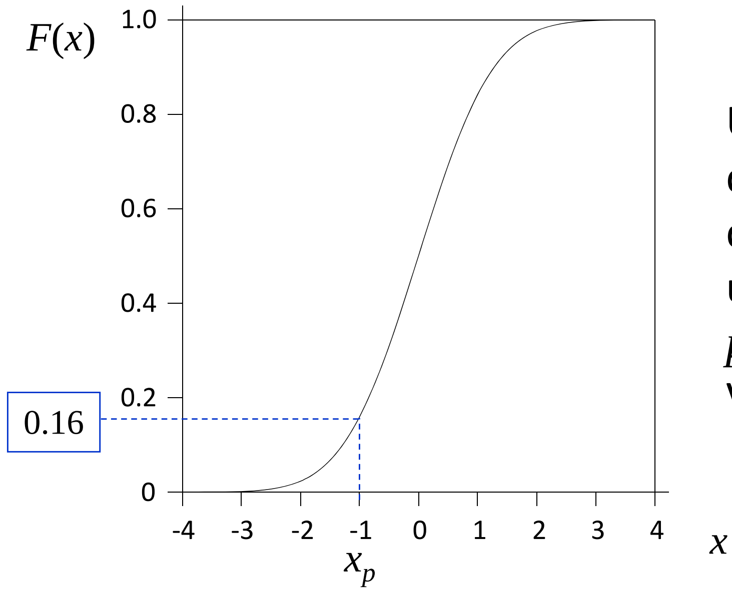
$$P(a < x \leq b) = \int_a^b f(x)dx$$

- Das Integrieren rückt also an die Stelle der Summierung bei diskreten Verteilungen.

- Die Verteilungsfunktion $F(x) = p$ gibt die Wahrscheinlichkeit an, dass höchstens ein Wert von $x = x_p$ auftritt, also dem Flächenanteil der Verteilung unterhalb von x_p. Formal:

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)dt$$





Umgekehrt kann man das Perzentil x_p als den x-Wert angeben, unter dem gegebene p Prozent der Verteilung liegen.

- Auch bei den Erwartungswerten und Varianzen werden dann aus den Summen Integrale, z.B. wird aus dem Erwartungswert bei diskreten Verteilungen

$$\mathrm{E}(X) = \sum_{i=1}^{m} x_i \cdot f(x_i)$$

im stetigen Fall:

$$\mathrm{E}(X) = \int_{-\infty}^{+\infty} x \cdot f(x)dx$$

- Zufallsvariablen weisen in der Statistik immer wieder ähnliche Verteilungen auf. Wir werden die folgenden Verteilungen näher betrachten:

Verteilung	Parameter	Typ	Inferenzstatistik
Gleichverteilung	n	diskret	
Binomialverteilung	n, p		
Normalverteilung	μ, σ	stetig	grundlegend; z.B. Vertrauensintervalle
Standardnormalverteilung	-		
χ^2-Verteilung	df		z.B. Varianzen, Häufigkeiten von nominalskalierten Variablen sowie bei Modellanpassungstests
t-Verteilung	df		z.B. Vergleich von Mittelwerten zweier Gruppen
F-Verteilung	df_1, df_2		z.B. Vergleich von mehr als zwei Mittelwerten, Varianzen, Regressionsanalyse

Wahrscheinlichkeitstheorie II: Gliederung

1 Grundlagen von Wahrscheinlichkeitsverteilungen

(Zufallsvariable, Wahrscheinlichkeitsfunktion, Dichte, Verteilungsfunktion, Parameter einer Verteilung, Erwartungswert)

2 Diskrete Wahrscheinlichkeitsverteilungen

(Gleichverteilung, Binomialverteilungen)

3 Stetige Wahrscheinlichkeitsverteilungen

(Normalverteilung, χ^2-, t- und F-Verteilung)

4 Tabellierungen & Umrechnungen

- Eine diskrete Zufallsvariable ist gleichverteilt (rechteckig verteilt), wenn alle n möglichen Werte mit gleicher Wahrscheinlichkeit realisiert werden, also gilt

$$f(x) = P(x = k) = 1/n$$

- Eine Gleichverteilung weist also einen Parameter n auf. Es gilt:

$$\mathrm{E}(X) = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i$$

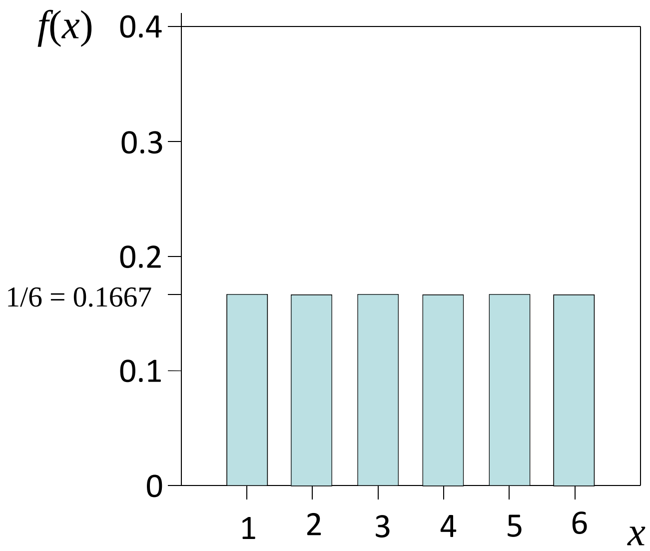
$$\mathrm{Var}(X) = \frac{1}{n} \cdot \sum_{i=1}^{n} [x_i - \mathrm{E}(x)]^2 = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i^2 - \frac{1}{n^2} \cdot \left(\sum_{i=1}^{n} x_i \right)^2$$

- Beispiel Zufallsexperiment: Einmaliges Würfeln.
  X = Augenzahl.

$$P(x = 1) = P(x = 2) = \ldots = P(x = 6) = 1/6$$

$$\mathrm{E}(X) = 1/6 \cdot (1 + 2 + 3 + 4 + 5 + 6) = 3.5$$

$$\mathrm{Var}(X) = 1/6 \cdot (1^2 + \ldots + 6^2) - 1/(6 \cdot 6) \cdot (1 + \ldots + 6)^2 = 2.92$$



Diskrete Wahrscheinlichkeitsverteilungen: Binomialverteilung

- Wird ein Zufallsvorgang unter identischen Bedingungen n mal wiederholt und ist die Wahrscheinlichkeit des Eintretens eines Ereignisses dabei konstant $P(A) = p$, so spricht man von einem Bernoulliexperiment (z.B. die zufällige Ziehung mit Zurücklegen von Kugeln aus einer Urne, in der sich 4 weiße und 6 schwarze Kugeln befinden).

- In einem Bernoulliexperiment ist die Wahrscheinlichkeit, dass bei n Wiederholungen das Ereignis A mit der Wahrscheinlichkeit p genau k mal eintritt (bzw. $n-k$ Mal nicht eintritt)

$$P(X = k) = \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k}$$

für $k = 0, 1, \ldots, n$. Die Wahrscheinlichkeitsfunktion für diese Zufallsvariable X bezeichnet man als Binomialverteilung. Man schreibt auch $X \sim \mathrm{B}(n, p)$. (~ = Tilde = „ist verteilt nach“).

- Durch Aufsummieren erhält man die Verteilungsfunktion, die angibt, wie wahrscheinlich es ist, dass höchstens x mal A eintritt:

$$P(X \le x) = \sum_{k=0}^{x} P(X = k) = \sum_{k=0}^{x} \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k}$$

- Beispiel: Eine Klausur bestehe aus sechs Aufgaben. Zu jeder Aufgabe muss aus drei Antworten die einzig richtige ausgewählt werden („multiple choice"-Aufgabe). Frage: Wie wahrscheinlich ist es, beim zufälligen Ankreuzen (Raten) zwei Aufgaben richtig zu lösen?

Lösung: Einsetzen von $n = 6$, $k = 2$ und $p = 1/3$ ergibt

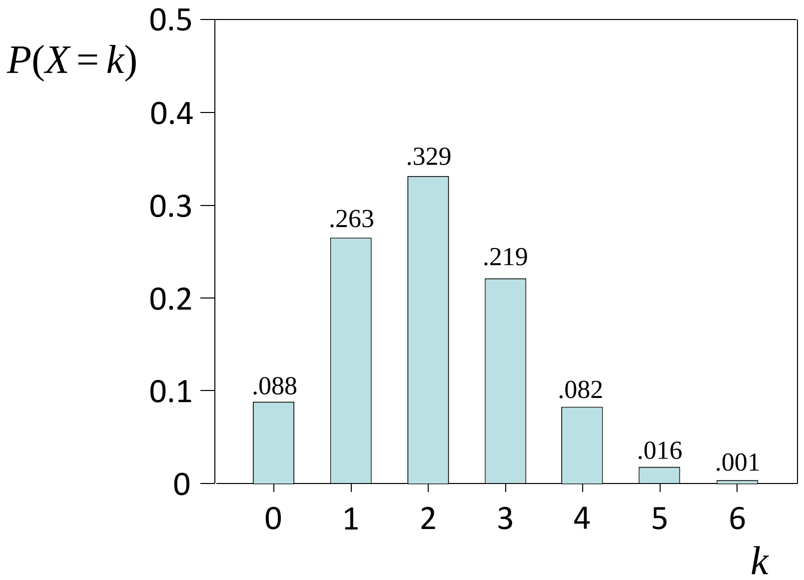
$$P(X = 2) = \binom{6}{2} \cdot (1/3)^2 \cdot (1 - 1/3)^{6-2} = \frac{6!}{2!(6-2)!} \cdot (1/3)^2 \cdot (2/3)^4 = \frac{5 \cdot 6}{1 \cdot 2} \cdot 0.111 \cdot 0.198 = 0.329$$

$$\boxed{\binom{n}{k} = \frac{n!}{k! \cdot (n-k)!}}$$

Die Wahrscheinlichkeit beträgt also knapp 33%.

- Nebenan ist die Wahrscheinlichkeitsverteilung (Wahrscheinlichkeitsfunktion) für $n = 6$ und $p = 1/3$ grafisch dargestellt.

Wahrscheinlichkeitsfunktion

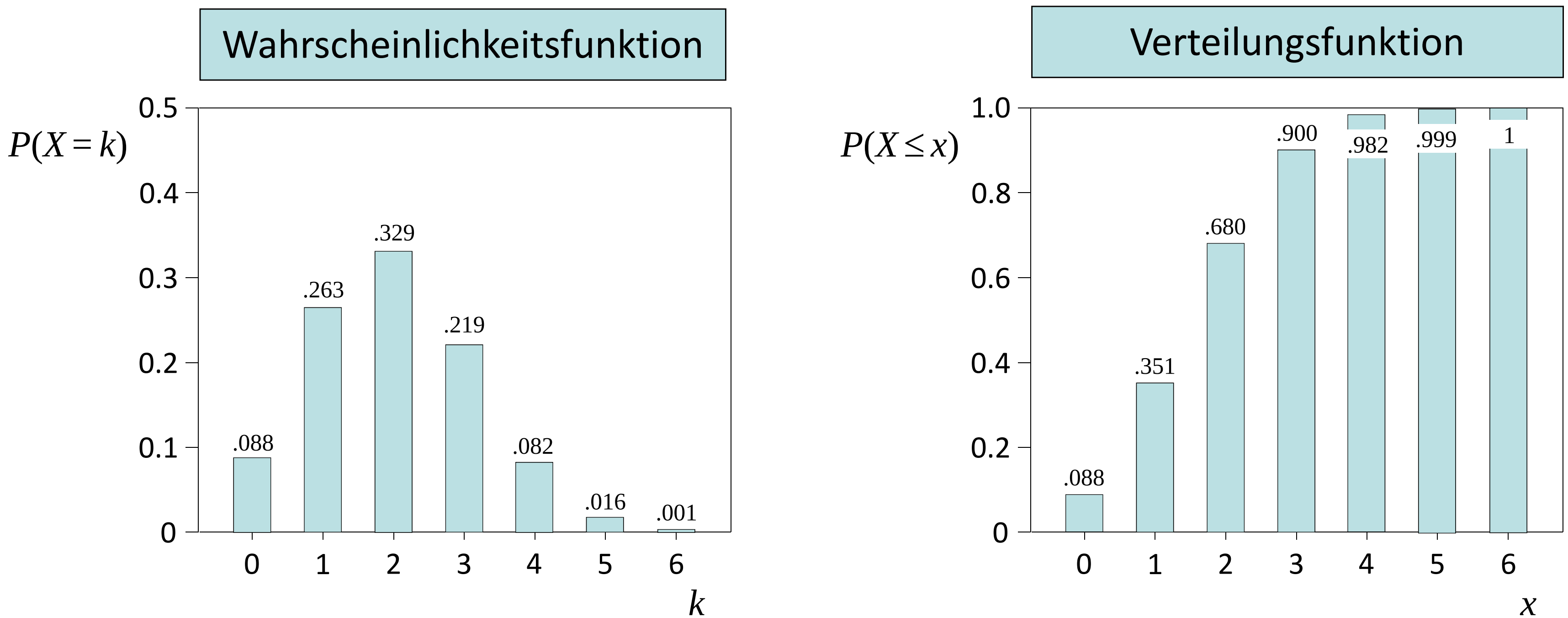


- Beispiel: wie auf vorheriger Folie. Frage jetzt: Wie wahrscheinlich ist es, beim zufälligen Ankreuzen <u>weniger als die Hälfte</u> der Antworten richtig zu lösen (durchzufallen)?

  Gegeben sind $n = 6$ und $p = 1/3$. Die folgende Summe muss bestimmt werden:

$$P(X \leq x) = \sum_{k=0}^{x} P(X = k) = P(X = 0) + P(X = 1) + P(X = 2)$$
$$= 0.088 + 0.263 + 0.329 = 0.680$$

Wenn man rät, ist die Wahrscheinlichkeit zu bestehen, also immerhin 32%.



Diskrete Wahrscheinlichkeitsverteilungen: Binomialverteilung

- Eigenschaften der binomialverteilten Zufallsvariablen $X \sim \mathrm{B}(n, p)$ mit den beiden Parametern n und p sind:
  - Der Erwartungswert ist: $\mathrm{E}(X) = n \cdot p$
  - Die Varianz ist $\mathrm{Var}(X) = n \cdot p \cdot (1 - p)$
  - Mit zunehmendem n nähert sich die Binomialverteilung immer mehr der Normalverteilung an. (Die Approximation ist bereits ab $n > 20$ hinreichend genau.)

- Beispiel wie auf vorheriger Folie. Frage: Wie groß ist die mittlere Anzahl richtiger Antworten, wenn man immer rät?

$$\mathrm{E}(X) = n \cdot p = 6 \cdot 1/3 = 2.0$$

Im Mittel werden also durch Raten 2 von 6 Fragen richtig beantwortet.

Weitere diskrete Wahrscheinlichkeitsverteilungen

- Weitere diskrete Verteilungen (ausgehend vom Beispiel der Binomialverteilung mit einer zufälligen Ziehung von Kugeln mit Zurücklegen aus einer Urne mit weißen und schwarzen Kugeln).

  - Multinominalverteilung: Erweiterung der Binomialverteilung für den Fall mit mehr als zwei Ausgängen.

    am Beispiel: Es gibt auch noch andersfarbige Kugeln und das Ereignis ist ein bestimmtes Muster von Kugeln.

  - Poissonverteilung: Berechnung der Wahrscheinlichkeit, mit der bei n Wiederholungen ein Ereignis k mal eintritt, wenn n sehr groß und p sehr klein ist (seltene Ereignisse).

    am Beispiel: Es interessiert die Ziehung einer weißen Kugel, wenn sich in der Urne 99 schwarze und eine weiße Kugel befinden.

  - Hypergeometrische Verteilung: Wahrscheinlichkeit, mit der bei n Wiederholungen ein Ereignis k mal eintritt, wenn die Objekte nicht zurückgelegt werden (sich also nach jeder Ziehung die Wahrscheinlichkeiten ändern)

    am Beispiel: Die gezogenen Kugeln werden nicht in die Urne zurückgelegt.

Wahrscheinlichkeitstheorie II: Gliederung

1 Grundlagen von Wahrscheinlichkeitsverteilungen

(Zufallsvariable, Wahrscheinlichkeitsfunktion, Dichte, Verteilungsfunktion, Parameter einer Verteilung, Erwartungswert)

2 Diskrete Wahrscheinlichkeitsverteilungen

(Gleichverteilung, Binomialverteilungen)

3 Stetige Wahrscheinlichkeitsverteilungen

(Normalverteilung, χ^2-, t- und F-Verteilung)

4 Tabellierungen & Umrechnungen

- Eine in der Statistik sehr häufig auftretende Verteilung ist die Normalverteilung (Gauß-Verteilung) $X \sim \mathrm{N}(\mu, \sigma)$ mit den Parametern Mittelwert μ und Standardabweichung σ.

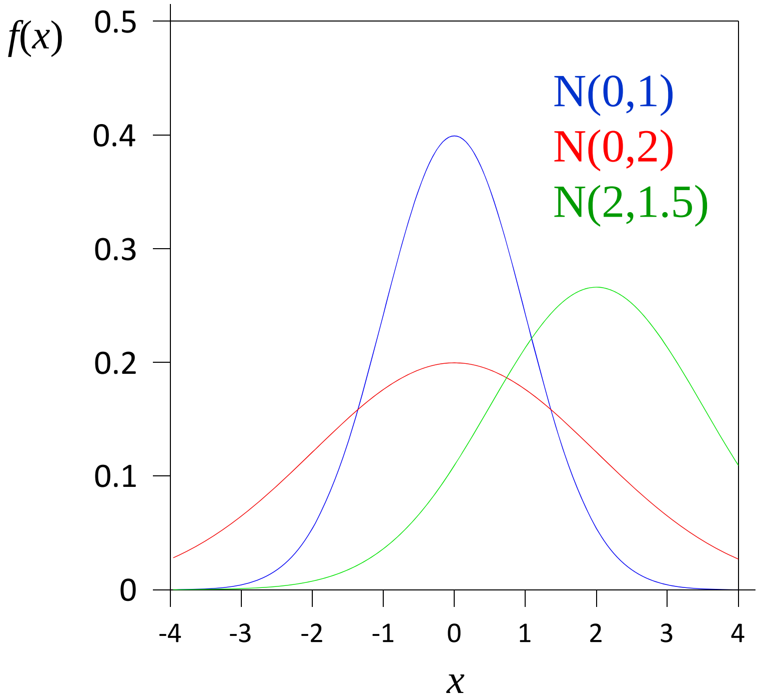
- Die Dichte einer Normalverteilung mit Mittelwert μ und Standardabweichung $\sigma > 0$ ist:

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

- Für den Fall der Standardnormalverteilung mit $\mu = 0$ und $\sigma = 1$ resultiert vereinfacht

$$f(x) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{1}{2}x^2} = \frac{1}{\sqrt{2\pi}} \cdot \exp\left(-\frac{1}{2}x^2\right)$$

alternative Schreibweise

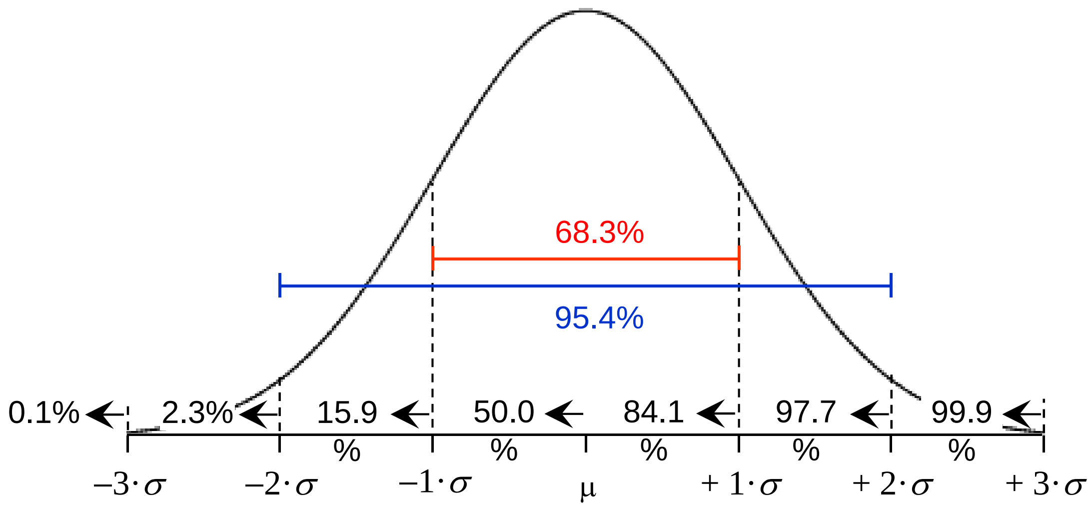


- Normalverteilte Werte mit anderen Mittelwerten oder/und Standardabweichungen lassen sich durch die z-Transformation in standardnormalverteilte Werte umrechnen:

$$Z = \frac{X - \mu}{\sigma}$$

Stetige Wahrscheinlichkeitsverteilungen: Normalverteilung

- Eigenschaften der Normalverteilung:
  - Der Erwartungswert der Normalverteilung ist $E(X) = \mu$.
  - Die Varianz der Normalverteilung ist $\mathrm{Var}(X) = \sigma^2$.
  - Sie ist symmetrisch um μ (= dem Modalwert und dem Median).
  - Die Dichte nähert sich für x gegen $\pm\propto$ asymptotisch der X-Achse an.
  - Zwischen dem Mittelwert $\pm\sigma$ liegen 68.3 Prozent der Verteilung, zwischen $\pm 2\cdot\sigma$ liegen 95.4 und zwischen $\pm 3\cdot\sigma$ liegen 99.7 Prozent der Verteilung.



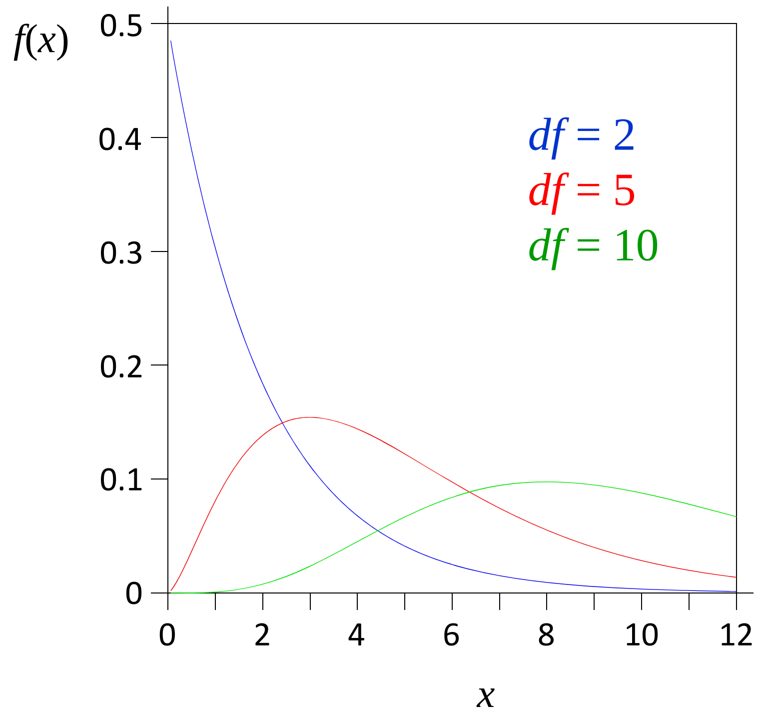
Stetige Wahrscheinlichkeitsverteilungen: χ^2-Verteilung

- Eine χ^2-verteilte (chi) Zufallsvariable $X \sim \chi^2(df)$ erhält man, wenn man mehrere voneinander unabhängige standardnormalverteilte Zufallsvariablen $Z_1, \ldots, Z_k$ quadriert und addiert:

$$X = \sum_{i=1}^{k} Z_i^2$$

- Die χ^2-Verteilung weist einen Parameter, die Anzahl Freiheitsgrade *df* (degrees of freedom) auf, deren Zahl der der aufsummierten Z-Variablen entspricht: $df = k$.

- Die Eigenschaften einer χ^2-Verteilung sind:
  - Der Erwartungswert ist: $\mathrm{E}(X) = df$
  - Die Varianz ist: $\mathrm{Var}(X) = 2 \cdot df$
  - Je kleiner df, desto linkssteiler ist die Verteilung; je größer df, desto mehr nähert sich die χ^2-Verteilung einer Normalverteilung an.

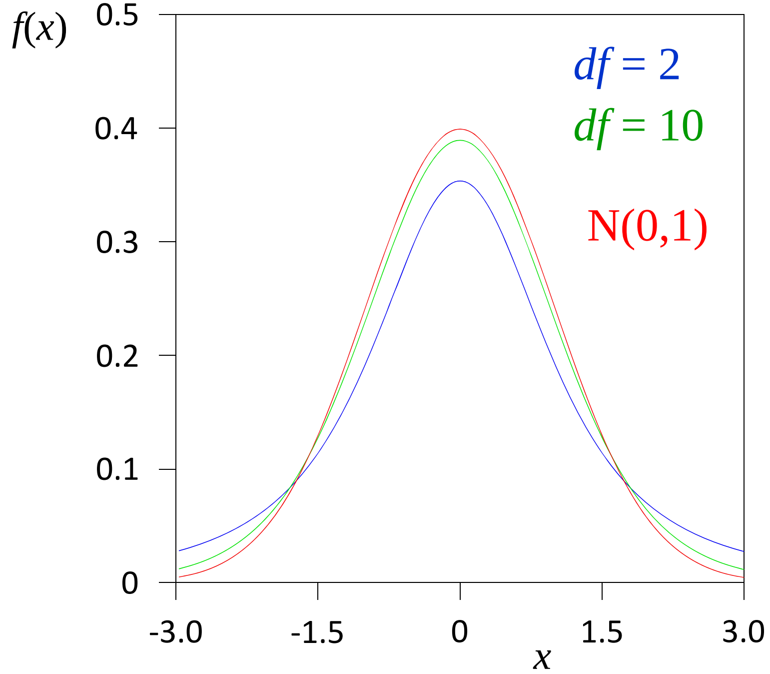


- Eine *t*-verteilte Zufallsvariable $X \sim t(df)$ erhält man, wenn man eine standardnormalverteilte Zufallsvariable Z_0 durch die Wurzel einer dazu unabhängigen χ^2-verteilten Zufallsvariable teilt, die durch die Zahl der Freiheitsgrade dividiert wurde:

$$X = \frac{Z_0}{\sqrt{\dfrac{\chi^2}{df}}} = \frac{Z_0}{\sqrt{\dfrac{1}{k} \cdot \sum_{i=1}^{k} Z_i^2}}$$

- Die *t*-Verteilung weist mit der Anzahl Freiheitsgrade $df = k$ einen Parameter auf.

- Die Eigenschaften einer *t*-Verteilung sind:
  - Der Erwartungswert ist: $\mathrm{E}(X) = 0$, falls $df > 1$
  - Die Varianz ist: $\mathrm{Var}(X) = df/(df-2)$, falls $df > 2$
  - Sie ist symmetrisch um 0 und nähert sich mit wachsenden df immer mehr der Normalverteilung an.



Stetige Wahrscheinlichkeitsverteilungen: F-Verteilung

- Eine F-verteilte Zufallsvariable $X \sim F(df_1, df_2)$ erhält man, wenn man den Quotienten zweier unabhängiger χ^2-verteilter Zufallsvariablen Y_1 und Y_2 bildet, die jeweils durch ihre Freiheitsgrade dividiert wurden:

$$X = \frac{Y_1 / df_1}{Y_2 / df_2} = \frac{\frac{1}{k_1} \cdot \sum_{i=1}^{k_1} Z_{1i}^2}{\frac{1}{k_2} \cdot \sum_{j=1}^{k_2} Z_{2j}^2}$$

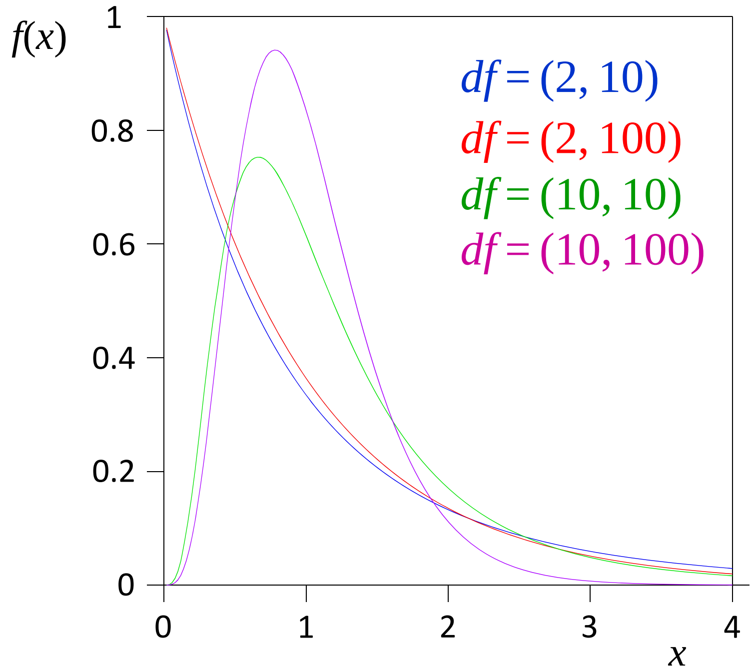
- Die F-Verteilung wird durch zwei Parameter gekennzeichnet: Die Zahl der Freiheitsgrade des Zählers, df_1, und des Nenners, df_2, des Quotienten.

Stetige Wahrscheinlichkeitsverteilungen: F-Verteilung

- Die Eigenschaften einer F-Verteilung sind
  - Der Erwartungswert ist: $\mathrm{E}(X) = df_2 / (df_2 - 2)$, falls $df_2 > 2$
  - Die Varianz ist:

$$\mathrm{Var}(X) = \frac{2 \cdot df_2^{\,2} \cdot (df_1 + df_2 - 2)}{df_1 \cdot (df_2 - 4) \cdot (df_2 - 2)^2} \quad \text{falls } df_2 > 4$$

  - Sie ist bei kleinen Freiheitsgraden asymmetrisch und linkssteil und wird bei wachsenden df symmetrischer.
  - Das Quadrat einer t-Verteilung entspricht einer F-Verteilung mit einem Zählerfreiheitsgrad: $t^2(df) \sim F(1, df)$



Wahrscheinlichkeitstheorie II: Gliederung

1 Grundlagen von Wahrscheinlichkeitsverteilungen

(Zufallsvariable, Wahrscheinlichkeitsfunktion, Dichte, Verteilungsfunktion, Parameter einer Verteilung, Erwartungswert)

2 Diskrete Wahrscheinlichkeitsverteilungen

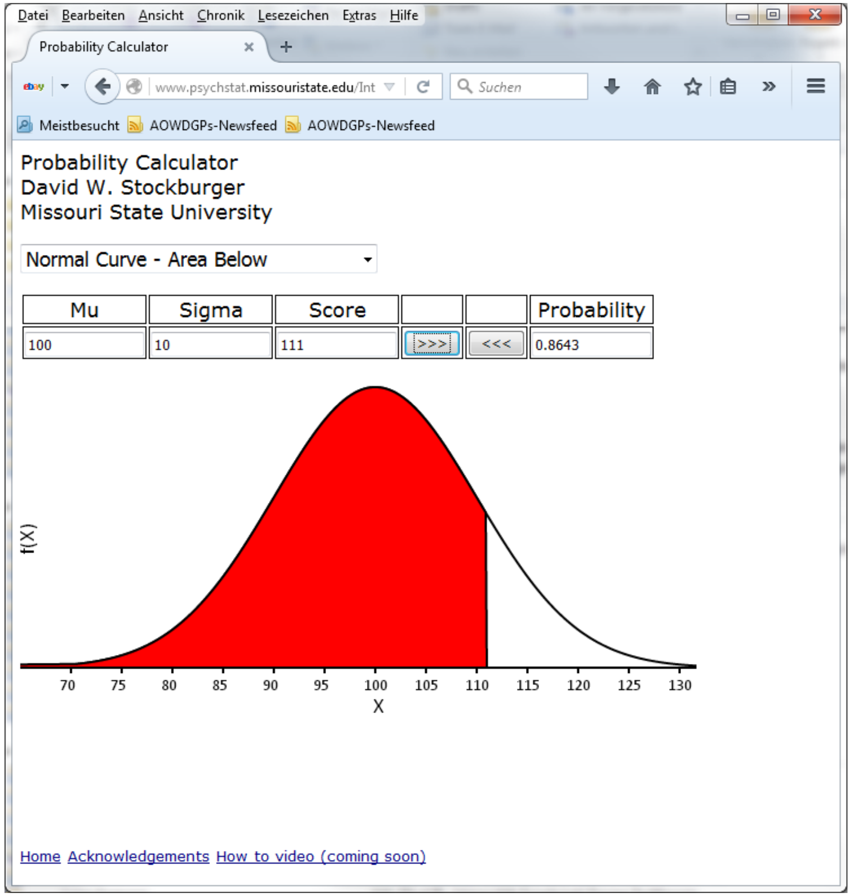
(Gleichverteilung, Binomialverteilungen)

3 Stetige Wahrscheinlichkeitsverteilungen

(Normalverteilung, χ^2-, t- und F-Verteilung)

4 Tabellierungen & Umrechnungen

- Es gibt verschiedene Möglichkeiten, für die behandelten, häufiger auftretenden Wahrscheinlichkeitsverteilungen (in Abhängigkeit von ihren Parametern) gegebene Realisationen der Zufallsvariablen x- in p-Werte „umzurechnen“ vice versa:
  - gedruckte Tabellen, z.B. im Anhang von Statistik-Büchern. Vorteil: ohne Technik nutzbar; Nachteil: nur ausgewählte Werte tabelliert
  - Statistikprogramme (z.B. SPSS) oder Excel
  - Internet-Applikationen (z.B. wie rechts http://www.psychstat.missouristate.edu/IntroBook3/sbkProbabilityCalculator.html)
  - Apps (z.B. frei Randov für Apple iOS)



Wahrscheinlichkeitsverteilungen

- In SPSS können folgende Transformationen für verschiedene Wahrscheinlichkeitsfunktionen mit ihren Parametern durchgeführt werden:
  - Für gegebenes x wird die (kumulative) Wahrscheinlichkeit $p = F(x)$ ausgegeben, mit der ein Wert < x (bei stetigen Funktionen) bzw. $\leq x$ (bei diskreten Funktionen) ist (CDF = Cumulative Distribution Function). Funktionsgruppe: „Verteilungsfunktionen"
  - Für eine gegebene Wahrscheinlichkeit p wird der Wert x, das p-te Quantil, ausgegeben (`IDF` = Invers Distribution Function). Funktionsgruppe: „Quantilfunktionen"
  - Für gegebenes x wird die Dichte (für stetige Funktionen) bzw. Wahrscheinlichkeit (für diskrete Funktionen) $f(x)$ ausgegeben (PDF = Probability Density Function). Funktionsgruppe: „Wahrscheinlichkeitsdichten"
  - Es werden Zufallszahlen ausgegeben, die der Wahrscheinlichkeitsfunktion folgen (RV = Random Values).

- Als Wahrscheinlichkeitsfunktionen stehen in SPSS unter anderem zur Verfügung: Rechtecksverteilung (`UNIFORM`), Binomialverteilung (`BINOM`), Normalverteilung (`NORMAL`), χ^2-Verteilung (`CHISQ`), t-Verteilung (`T`) und F-Verteilung (`F`).

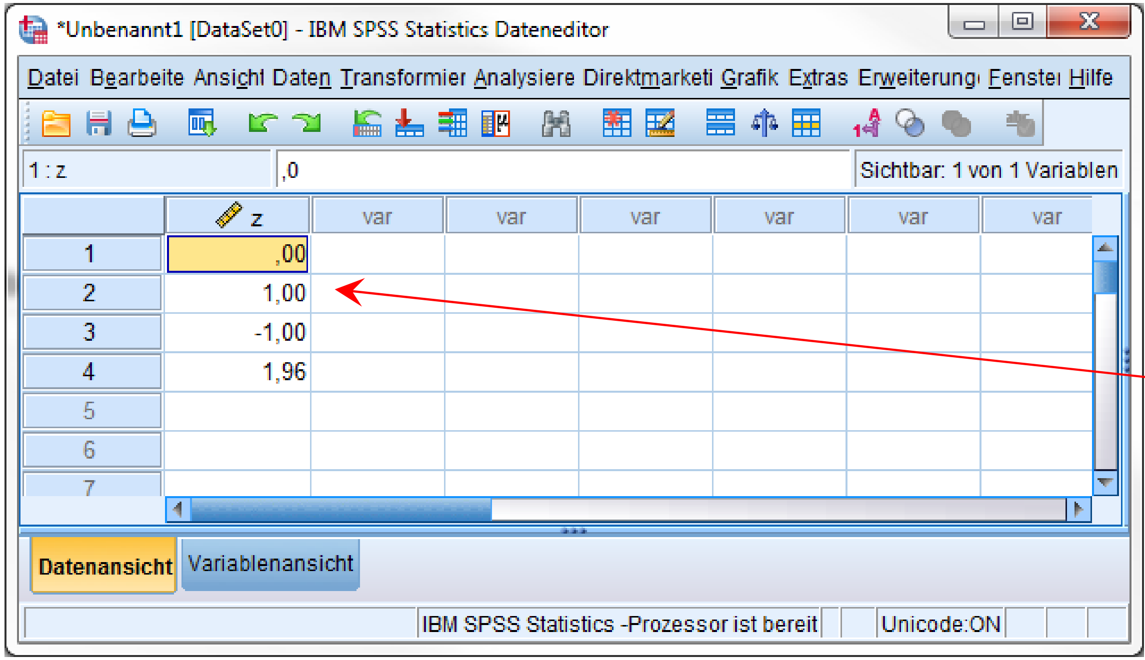
- Die Aufruf der benötigten Funktion setzt sich in SPSS wie folgt zusammen:
  `<Typ>.<Verteilung>(<Parameter>)`

- Die von uns häufiger benötigten Funktionen lauten:

Verteilung	$p \rightarrow x$	$x \rightarrow p$
Rechteck	`IDF.UNIFORM(p,min,max)`	`CDF.UNIFORM(q,min,max)`
Binomial	-	`CDF.BINOM(q,n,p)`
Normal	`IDF.NORMAL(p,m,s)`	`CDF.NORMAL(q,m,s)`
χ^2	`IDF.CHISQ(p,df)`	`CDF.CHISQ(q,df)`
t	`IDF.T(p,df)`	`CDF.T(q,df)`
F	`IDF.F(p,df1,df2)`	`CDF.F(q,df1,df2)`

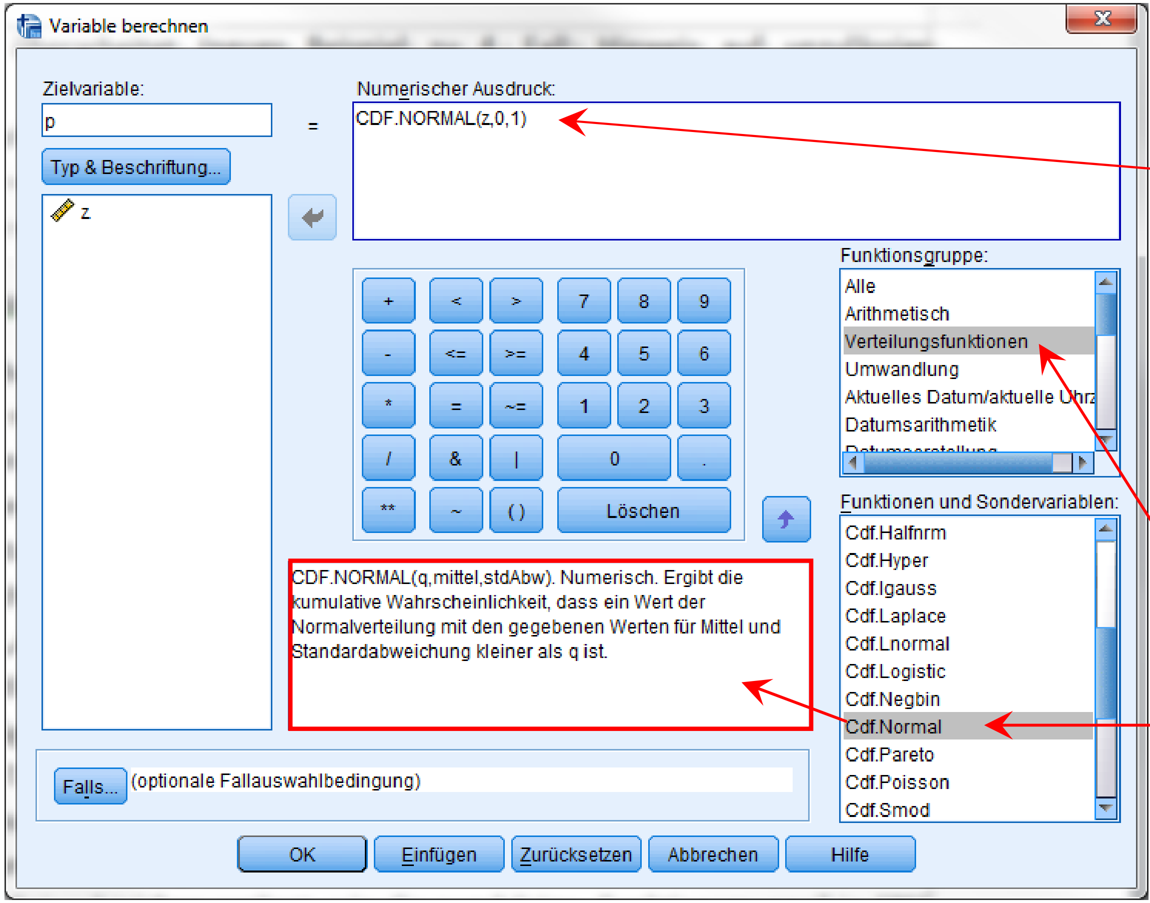
- Es bezeichnen: `q` $= x$ und `p` $= p$; `df` = Freiheitsgrade df. Bei Normalverteilung: `m` = Mittelwert, `s` = Standardabweichung. Bei Binomialverteilung: `n` = n und `p` = p.

- Die Wahrscheinlichkeits-Transformationen lassen sich unter `Transformieren/Variablen_berechnen…` ausführen. Bevor man damit startet, muss man zunächst eine Variable (mit einem beliebigen Namen) im Dateneditor anlegen und dort die Werte eingeben, die man transformiert haben möchte.



Im Beispiel wollen wir für verschiedene z-Werte die p-Werte der Standardnormalverteilung ermitteln.
Dazu haben wir in einer leeren Datendatei eine numerische Variable z angelegt und die Werte dort eingetragen.

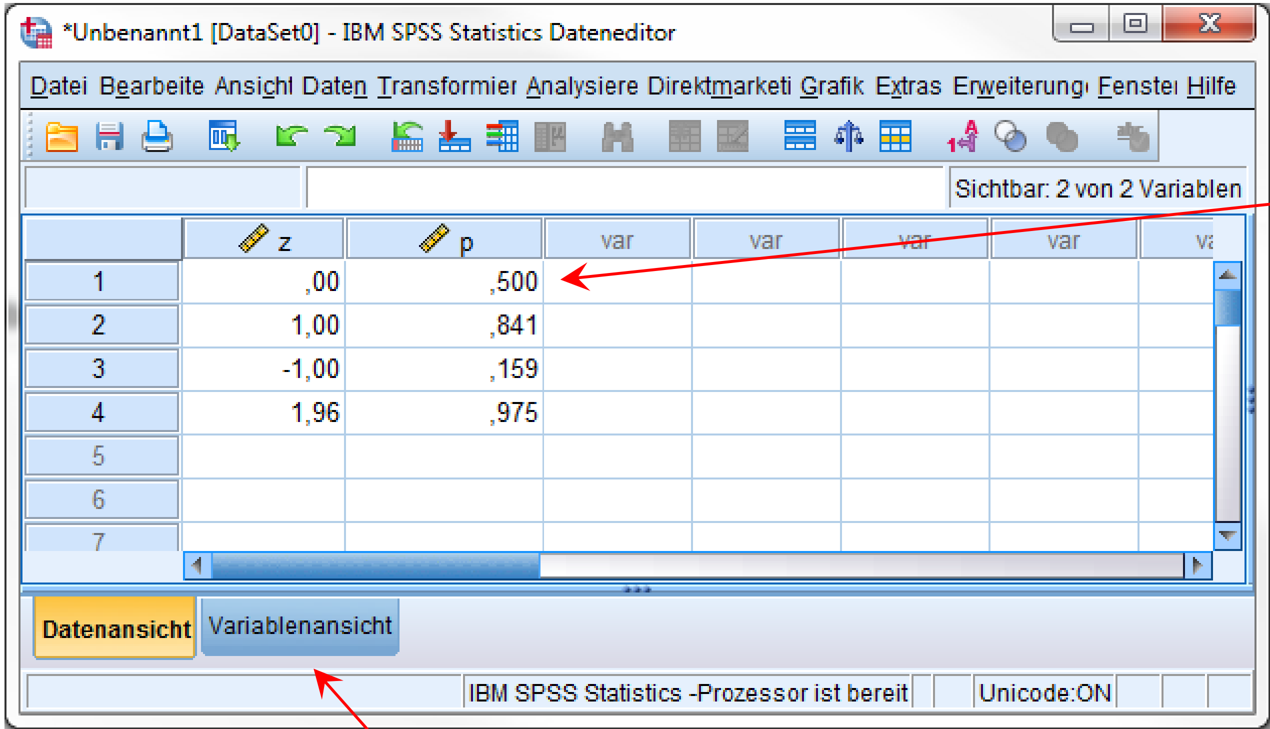
- Im Dialog unter `Transformieren/Variablen_berechnen` geben wir zunächst als »Zielvariable« eine beliebige neue Variable an (hier: `p`) und anschließend unter »Numerischer Ausdruck« die benötigte Funktion mit ihren Argumenten.



In unserem Fall ist das die Funktion `CDF.NORMAL(q,m,s)`, in die nun für die Platzhalter `m=0` und `s=1` (der Standard-NV) sowie die Variable `z` für `q` eingesetzt wird, die die Werte enthält, also: `CDF.NORMAL(z,0,1)`

Wählt man als »Funktionsgruppe« „Verteilungsfunktionen“ an, dann erscheinen unter »Funktionen und Sondervariablen« alle zur Verfügung stehen Funktionen, deren Aufruf links daneben erläutert wird

- Nach der Aktivierung der Berechnung mittels (OK) wechseln wir in den Dateneditor. Dort ist nun die neue Variable p von SPSS hinzugefügt worden, die die gewünschten Werte enthält.



Die Flächenanteile für den z-Wert von 0 lautet also $p = 0.5$ und der für $z = 1$ lautet $p = 0.841$. (Man sieht, dass für $z = -1$ der p-Wert $1 - 0.841 = 0.159$ beträgt.)

Im Reiter ⌊Variablenansicht⌋ kann man die Zahl der Nachkommastellen der neuen Variablen noch hochsetzen (z.B. wie hier auf 3).

Wahrscheinlichkeitsverteilungen

- Tabellen zur Umrechnung von p in x-Werte (oder umgekehrt) finden sich für die gängigen Wahrscheinlichkeitsverteilungen im Anhang (fast) jeden Statistikbuches.
- Ein entsprechender Tabellenband wird für Übungszwecke auch als pdf-Datei in Stud.IP eingestellt. In der Klausur wird ausschließlich dieser Tabellenband verwendet.
- Die Tabellenbände sind nicht immer gleich aufgebaut, gleich beschriftet und enthalten auch nicht immer genau die gleichen Wertebereiche.
- Um die erforderlichen Werte ablesen zu können muss man wissen,
  (1) um welche Verteilung es sich handelt,
  (2) welche Parameter die Verteilung aufweisen soll und
  (3) ob x gegeben und p gesucht ist oder umgekehrt.

Anhang

Aus: Eid, Gollwitzer und Schmitt (2010)

Anhang A: Tabellen

1 Binominalverteilung
2 Standardnormalverteilung
3 Zentrale t-Verteilung
4 Wilcoxon-Vorzeichen-Rangtest
5 Zentrale χ^2-Verteilung
6 Kritische Werte für den Kolmogorov-Smirnov-Test und den Lilliefors-Test
7 Wilcoxon-Rangsummen-Test
8 Zentrale F-Verteilung
9 Kritische Werte für die Differenz $n_K - n_D$

Wahrscheinlichkeitsverteilungen

- **Beispiel**: Wie wahrscheinlich ist es, in einem normalverteilten Intelligenztest mit einem Mittelwert von 100 und einer Standardabweichung von 15 einen Wert von 90 oder weniger zu erhalten?

Gegeben: $N(100, 15)$, $x = 90$; Gesucht: kumulative Wahrscheinlichkeit p

Problem 1: Es gibt nur Tabellen für die <u>Standard</u>normalverteilung, also $N(0, 1)$.

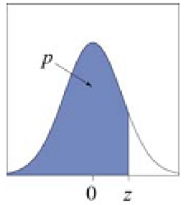
Lösung: Umrechnung in einen z-Wert ergibt: $z = (90 – 100) / 15 \approx –0.67$

Problem 2: Es sind nur Werte $x \geq 0$ tabelliert. (Die x-Werte sind in dieser Tabelle mit „z“ bezeichnet)

Lösung: Da die Normalverteilung symmetrisch ist, gilt $p = P(X \leq x) = 1 – P(X \leq –x)$. Wir entnehmen also für den x-Wert von 0.67 den Wert 0.7486. Die gesuchte kumulative Wahrscheinlichkeit ist dann $p = 1 – 0.7486 = 0.2514$.

Standardnormalverteilung

Die Tabelle enthält die Werte der Verteilungsfunktion
$p := F(z) = P(Z \leq z)$ für $z \geq 0$.



z	p	z	p	z	p	z	p	z	p
0.00	0.5000	0.46	0.6772	0.92	0.8212	1.38	0.9162	1.84	0.9671
0.01	0.5040	0.47	0.6808	0.93	0.8238	1.39	0.9177	1.85	0.9678
0.02	0.5080	0.48	0.6844	0.94	0.8264	1.40	0.9192	1.86	0.9686
0.03	0.5120	0.49	0.6879	0.95	0.8289	1.41	0.9207	1.87	0.9693
0.04	0.5160	0.50	0.6915	0.96	0.8315	1.42	0.9222	1.88	0.9699
0.05	0.5199	0.51	0.6950	0.97	0.8340	1.43	0.9236	1.89	0.9706
0.06	0.5239	0.52	0.6985	0.98	0.8365	1.44	0.9251	1.90	0.9713
0.07	0.5279	0.53	0.7019	0.99	0.8389	1.45	0.9265	1.91	0.9719
0.08	0.5319	0.54	0.7054	1.00	0.8413	1.46	0.9279	1.92	0.9726
0.09	0.5359	0.55	0.7088	1.01	0.8438	1.47	0.9292	1.93	0.9732
0.10	0.5398	0.56	0.7123	1.02	0.8461	1.48	0.9306	1.94	0.9738
0.11	0.5438	0.57	0.7157	1.03	0.8485	1.49	0.9319	1.95	0.9744
0.12	0.5478	0.58	0.7190	1.04	0.8508	1.50	0.9332	1.96	0.9750
0.13	0.5517	0.59	0.7224	1.05	0.8531	1.51	0.9345	1.97	0.9756
0.14	0.5557	0.60	0.7257	1.06	0.8554	1.52	0.9357	1.98	0.9761
0.15	0.5596	0.61	0.7291	1.07	0.8577	1.53	0.9370	1.99	0.9767
0.16	0.5636	0.62	0.7324	1.08	0.8599	1.54	0.9382	2.00	0.9772
0.17	0.5675	0.63	0.7357	1.09	0.8621	1.55	0.9394	2.01	0.9778
0.18	0.5714	0.64	0.7389	1.10	0.8643	1.56	0.9406	2.02	0.9783
0.19	0.5753	0.65	0.7422	1.11	0.8665	1.57	0.9418	2.03	0.9788
0.20	0.5793	0.66	0.7454	1.12	0.8686	1.58	0.9429	2.04	0.9793
0.21	0.5832	0.67	0.7486	1.13	0.8708	1.59	0.9441	2.05	0.9798
0.22	0.5871	0.68	0.7517	1.14	0.8729	1.6	0.9452	2.06	0.9803
0.23	0.5910	0.69	0.7549	1.15	0.8749	1.61	0.9463	2.07	0.9808

Wahrscheinlichkeitsverteilungen

- Problem 3: Die Umrechnung in einen z-Wert $z = (90 - 100) / 15 \approx -0.67$ hatten wir nur mit einer Genauigkeit von zwei Nachkommastellen durchgeführt. Auf drei Nachkommastellen genau wäre der Wert -0.667. Wenn wir diesen Wert in der Tabelle nachschlagen wollen, ist der Wert nicht tabelliert.

Lösung: Wir können der Tabelle nur entnehmen, dass für den x-Wert von 0.667 die kumulative Wahrscheinlichkeit p zwischen 0.7454 und 0.7486 liegt.

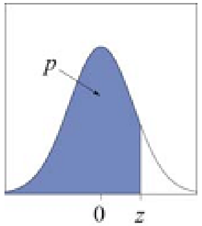
Wir können nun entweder p auch nur gerundet angeben, also $p = 0.75$ (bzw. $1 - 0.75 = 0.25$) oder

angeben, dass mindestens $1 - 0.7486 = 0.2514$ einen kleineren Wert aufweisen (konservative Angabe) oder

den p-Wert per Computerprogramm genauer bestimmen.

Standardnormalverteilung

Die Tabelle enthält die Werte der Verteilungsfunktion $p := F(z) = P(Z \leq z)$ für $z \geq 0$.



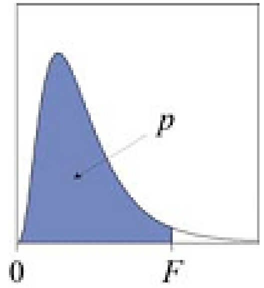
z	p	z	p	z	p	z	p	z	p
0.00	0.5000	0.46	0.6772	0.92	0.8212	1.38	0.9162	1.84	0.9671
0.01	0.5040	0.47	0.6808	0.93	0.8238	1.39	0.9177	1.85	0.9678
0.02	0.5080	0.48	0.6844	0.94	0.8264	1.40	0.9192	1.86	0.9686
0.03	0.5120	0.49	0.6879	0.95	0.8289	1.41	0.9207	1.87	0.9693
0.04	0.5160	0.50	0.6915	0.96	0.8315	1.42	0.9222	1.88	0.9699
0.05	0.5199	0.51	0.6950	0.97	0.8340	1.43	0.9236	1.89	0.9706
0.06	0.5239	0.52	0.6985	0.98	0.8365	1.44	0.9251	1.90	0.9713
0.07	0.5279	0.53	0.7019	0.99	0.8389	1.45	0.9265	1.91	0.9719
0.08	0.5319	0.54	0.7054	1.00	0.8413	1.46	0.9279	1.92	0.9726
0.09	0.5359	0.55	0.7088	1.01	0.8438	1.47	0.9292	1.93	0.9732
0.10	0.5398	0.56	0.7123	1.02	0.8461	1.48	0.9306	1.94	0.9738
0.11	0.5438	0.57	0.7157	1.03	0.8485	1.49	0.9319	1.95	0.9744
0.12	0.5478	0.58	0.7190	1.04	0.8508	1.50	0.9332	1.96	0.9750
0.13	0.5517	0.59	0.7224	1.05	0.8531	1.51	0.9345	1.97	0.9756
0.14	0.5557	0.60	0.7257	1.06	0.8554	1.52	0.9357	1.98	0.9761
0.15	0.5596	0.61	0.7291	1.07	0.8577	1.53	0.9370	1.99	0.9767
0.16	0.5636	0.62	0.7324	1.08	0.8599	1.54	0.9382	2.00	0.9772
0.17	0.5675	0.63	0.7357	1.09	0.8621	1.55	0.9394	2.01	0.9778
0.18	0.5714	0.64	0.7389	1.10	0.8643	1.56	0.9406	2.02	0.9783
0.19	0.5753	0.65	0.7422	1.11	0.8665	1.57	0.9418	2.03	0.9788
0.20	0.5793	0.66	0.7454	1.12	0.8686	1.58	0.9429	2.04	0.9793
0.21	0.5832	0.67	0.7486	1.13	0.8708	1.59	0.9441	2.05	0.9798
0.22	0.5871	0.68	0.7517	1.14	0.8729	1.6	0.9452	2.06	0.9803
0.23	0.5910	0.69	0.7549	1.15	0.8749	1.61	0.9463	2.07	0.9808

Wahrscheinlichkeitsverteilungen

➢ Beispiel: Wir suchen das 95. Perzentil des F-Wertes mit $df_1 = 2$ Zähler- und $df_2 = 3$ Nennerfreiheitsgraden.

F-Verteilung

Die Tabelle enthält die F-Werte der zentralen F-Verteilung für ausgewählte Quantile p (in den Zeilen) in Abhängigkeit von den Zählerfreiheitsgraden df_1 (in den Spalten) und Nennerfreiheitgraden df_2 (in den Zeilen).



		df_1									
df_2	p	1	2	3	4	5	6	7	8	9	10
1	0.90	39.86	49.50	53.59	55.83	57.24	58.20	58.91	59.44	59.86	60.19
	0.95	161.45	199.50	215.71	224.58	230.16	233.99	236.77	238.88	240.54	241.88
2	0.90	8.53	9.00	9.16	9.24	9.29	9.33	9.35	9.37	9.38	9.39
	0.95	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40
	0.99	98.50	99.00	99.17	99.25	99.30	99.33	99.36	99.37	99.39	99.40
3	0.90	5.54	5.46	5.39	5.34	5.31	5.28	5.27	5.25	5.24	5.23
	0.95	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79
	0.99	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.35	27.23
4	0.90	4.54	4.32	4.19	4.11	4.05	4.01	3.98	3.95	3.94	3.92
	0.95	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96
	0.99	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.66	14.55
5	0.90	4.06	3.78	3.62	3.52	3.45	3.40	3.37	3.34	3.32	3.30

➢ Lösung: Dem Ausschnitt aus der Tabelle entnimmt man einen F-Wert von 9.55.