VERTEILUNGEN

1. Binomialverteilung

Voraussetzungen

1. Resultat eines einzelnen Versuches ist entweder Erfolg oder Mißerfolg
2. Die Wahrscheinlichkeit für Erfolg (Mißerfolg) ist in jedem Versuch dieselbe, die Versuche sind nicht voneinander abhängig (Unabhängigkeit).

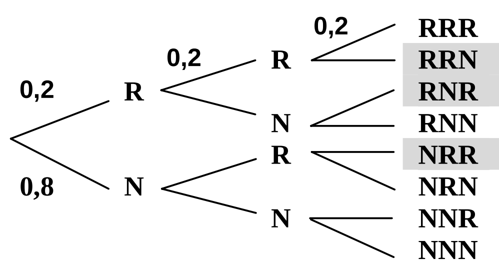
Die Binomialverteilung ist eine Verteilung für diskrete Daten. Typisches Beispiel: Das Werfen einer Münze, oder die Anzahl von Kaninchen in einem Feld.

Bernoulli-Versuch: (Nach Jakob Bernoulli) Durch oftmaliges Wiederholen eines solchen Experimentes konvergiert die relative Häufigkeit gegen die Wahrscheinlichkeit p. (Gesetz der großen Zahlen)

Beispiel.:
Aus einem Hasenstall mit 5 Hasen wird 3 mal hintereinander blind 1 Hase „gezogen“. Im Stall sind alle Hasen in jeder Hinsicht „gleich“, außer in der Farbe: Es gibt 2 blaue, 2 grüne und einen roten Hasen. Wie groß ist die Wahrscheinlichkeit, 2 mal den roten Hasen zu „ziehen“?

Jedes Mal ziehen ist die Wahrscheinlichkeit für rot 1/5, die Gegenwahrscheinlichkeit – keinen roten Hasen zu ziehen – ist 4/5; Aus 3 Ziehungen gibt es insgesamt 8 rot/nicht-rot unterschiedliche Kombinationen.
R = Roter Hase, N = Nicht roter Hase.

Baumdiagramm:



Aus dem Baumdiagramm ist ersichtlich daß es insgesamt 8 verschiedene 3-stellige N-R Kombinationen gibt (eigentlich 2^n Variationen !), von denen bei 3 das Ereignis R genau zweimal vorkommt. In mathematisch statistischer Notation heißt das, es gibt $\binom{3}{2} = 3$ mögliche Kombinationen. Also: Die Wahrscheinlichkeit für das Eintreffen eines einzelnen Ereignisses ist p.p.q, bzw. 0,2.0,2.0,8 = 0,032. Alle 3 Ereignisse treten dann mit einer Gesamtwahrscheinlichkeit von p(X=2) = 3.0,032 = 0,096 ein.

Erwartungswert (= Durchschnitt): $E(X) = \mu = n.p$

Varianz $V(X) = \sigma^2 = n.p.(1-p)$

Variationskoeffizient (CV) $CV = \frac{\sqrt{V(X)}}{E(X)}$

(CV = **C**oefficent of **V**ariation)

E(X) = 3 x 0,2 = 0,6
V(X) = 3 x 0,2 x 0,8 = 0,48

Bei kleinen Werten für n und für p <> 0,5 ist die Binomialverteilung unsymmetrisch, je nachdem ob $p < 0,5$ linksschief, oder bei $p > 0,5$ rechtsschief. Für „große“ n wird die Verteilung zunehmend symmetrischer und konvergiert schließlich für $n \to \infty$ gegen die symmetrische Normalverteilung.

Allgemein: Die Wahrscheinlichkeit bei n Versuchen k Erfolge zu erzielen, ist:

$$p(X = k) = \binom{n}{k} p^k q^{n-k}$$

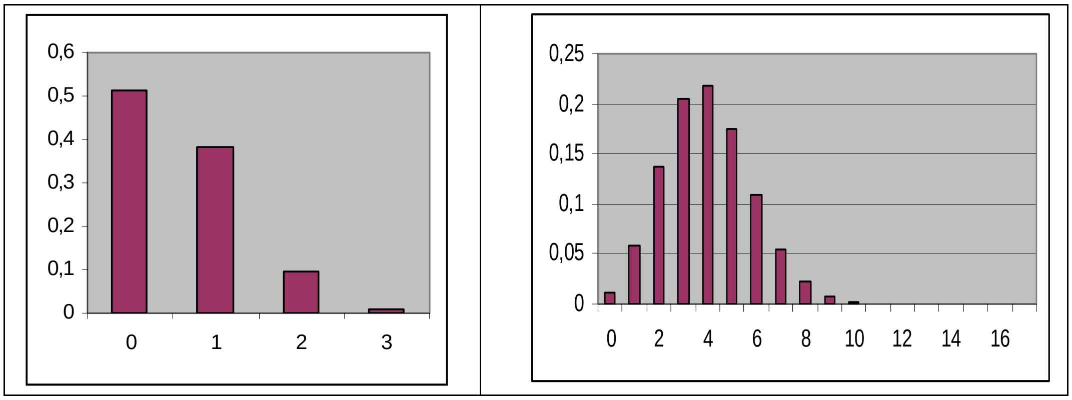


Abb: Auswirkung der Anzahl von Experimenten auf die Symmetrie der Verteilung. p = 0,2, linkes Bild: n = 4, rechtes Bild n = 17.

2. Normalverteilung

Ist X eine normalverteilte Zufallsvariable mit dem Mittelwert μ und der Varianz σ^2 dann ist die Dichtefunktion (probability density function) von X gegeben durch:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \; x \in R,$$

$$Schreibweise: X \approx N(\mu, \sigma^2)$$

<u>Standard-Normalverteilung</u>

Durch eine Transformation kann eine Normalverteilung in die Standardisierte Normalverteilung übergeführt werden. Diese wird üblicherweise mit „Z“ gekennzeichnet.
Ist $X \approx N(\mu,\sigma^2)$, dann ist die transformierte Zufallsvariable

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

nomalverteilt mit dem Mittelwert 0 und der Varianz 1:

$$Z \approx (0,1)$$

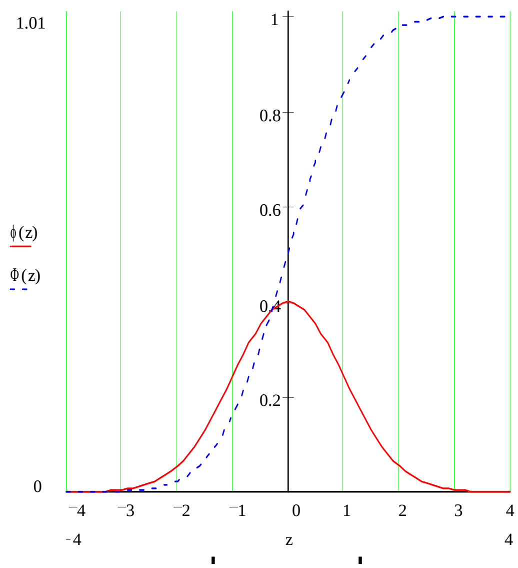
Z heißt die Standard Normal Verteilung mit der *Dichtefunktion* $\phi(z)$

$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{\frac{-z^2}{2}}, \quad z \in R$$

Die *kummulative Verteilungsfunktion* (Summenkurve, c.p.d.f. cumulative probability density function) ist dann definiert als:

$$\Phi(z) = P(Z \leq z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$$

Diese Funktion kann nicht einfach integriert werden, die Werte für Φ(z) sind aber in statistischen Tafeln tabelliert und können aus diesen entnommen werden. Natürlich stellen auch alle Statistikprogramme, Mathematikprogramme, die meisten Tabellenkalkulationsprogramme und viele Taschenrechner diese Funktionen zur Verfügung.



Standard-Normalverteilung mit μ = 0, σ = 1

Gesamtfläche der Normalverteilungskurve, die zwischen zwei z-Werten liegt.	
Intervall [-z, + z]	Fläche P[-z < Z < +z]
-3σ, +3σ	0,9974
-2,56σ, +2,56σ	0,99
-1,96σ, +1,96σ	0,95
-1σ, + 1σ	0,68

Wahrscheinlichkeitsdichte ϕ(z), Summenhäufigkeit Φ(z) und Wahrscheinlichkeiten

Standardnormalverteilung:

Normalverteilung: Warum wird diese standardisiert, und welche Vorteile ergeben sich daraus?

Nehmen wir dazu folgendes an: X sei eine Zufallsvariable mit Mittelwert μ_x und Varianz σ_x^2. Wir bilden jetzt eine neue Zufallsvariable Y als Linearkombination aus X und suchen Konstanten α, β (β > 0), sodaß gilt:

Y = α + βX

α, β sind so zu wählen, daß $\mu_y = 0$ und $\sigma_y^2 = 1$ wird.

$$\begin{aligned}
\mu_Y &= E(\alpha + \beta X) = \alpha + E(\beta X) \\
&= \alpha + \beta E(X) \\
\sigma_Y^2 &= E((Y-\mu_Y)^2) = E((\alpha + \beta X - \alpha - \beta\mu_X)^2) \\
&= \beta^2 E((X-\mu_X)^2) \\
&= \beta^2\sigma^2
\end{aligned}$$

$$\mu = 0 \ und \ \sigma_Y{}^2 = 1, \ wenn$$

$$\alpha + \beta\mu_x = 0$$
$$\beta^2\sigma_x^2 = 1$$

Die Lösung der beiden Gleichungen führt zu dem Ergebnis:

$$\alpha = \frac{-\mu_x}{\sigma_x}, \ \beta = \frac{1}{\sigma_x}$$

und daraus wird die Transformation der Zufallsvariable X in Y:

$$\begin{aligned}
Y &= \alpha + \beta X \\
&= -\frac{\mu_x}{\sigma_x} + \frac{X}{\sigma_x} \\
&= \frac{X - \mu_X}{\sigma_X}
\end{aligned}$$

<u>Verteilung des Mittelwertes:</u>

Ist $\overline{X}$ der Mittelwert einer Stichprobe vom Umfang n gezogen aus einer Normalverteilung N(μ,σ²), dann ist die Verteilung des Mittelwertes $\overline{X}$ gegeben durch :

$$\overline{X} \approx N(\mu, \frac{\sigma^2}{n})$$

Diese Verteilung ändert sich sehr rasch, wenn n größer wird, die Streuung um den Mittelwert (=Standardabweichung) nimmt mit 1/√n ab. Im Klartext bedeutet dies: Wird die Probengröße vervierfacht, dann wird die Standardabweichung halbiert. Durch Standardisierung einer normalverteilten Zufallsvariable (s.o.) erhielten wir eine neue Zufallsvariable Z ~ N(0,1). In gleicher Weise können wir auch $\overline{X}$ standardisieren und erhalten dadurch:

$$\frac{(\overline{X} - \mu)}{\sigma / \sqrt{n}} = Z \approx N(0,1)$$

Dieses Ergebnis kann zur Berechnung der Wahrscheinlichkeiten für $\overline{X}$ verwendet werden

Dazu ein kleines Beispiel:

Nehmen wir an, aus einer Normalverteilung mit dem Mittelwert 20 (z.B. das Gewicht kleiner Mäuse in g) und der Varianz 12 wird eine (Mäuse-)Probe von 15 Individuen gezogen.

a) Berechne die Wahrscheinlichkeit, daß der Durchschnittswert der Stichprobe zwischen 19 und 21 liegt.

b) Finde Werte y_1 und y_2, so daß die Wahrscheinlichkeit, daß der Durchschnitt der Stichprobe kleiner oder gleich y_1 oder mindestens y_2 ist, jeweils gleich 0.025 sein soll.

Aus der Standardisierung ergeben sich folgende z – Werte:

ad a)
n = 15
$\sigma^2 = 12$

$\bar{x}_1 = 21$	$z_1 = (21-20)/(12/15)^{\wedge 0,5}$	$z_1 = 1,118$	$\Phi(z_1) = 0,868224$
$\bar{x}_2 = 19$	$z_2 = (19-20)/(12/15)^{\wedge 0,5}$	$z_1 = -1,118$	$\Phi(z_2) = 0,131776$

$$P(19 \le \bar{x} \le 21) = P(-1{,}118 \le Z \le 1{,}118) = \Phi(1{,}118) - \Phi(-1{,}118) = 0{,}736$$

Die Wahrscheinlichkeit, dass der Durchschnittswert der Stichprobe zwischen 19 und 21 liegt, beträgt 73,6%

ad b)

$P(\bar{x} \le y_1) = 0{,}025 \; und \;\; P(x \ge y_2) = 0{,}025$; d.h. $\Phi(z_1) =$ 0,025 und $\Phi(z_2)$= 0,975; Aus den Tabellen oder aus einer Excelfunktion entnehmen wir $z_1 = -1{,}96$ und $z_2 = 1{,}96$; Wir können daher $\bar{x}$ berechnen aus der Beziehung: $z = \dfrac{\bar{x} - \mu}{\sigma / \sqrt{n}}$; $\bar{x} = \mu \pm z\,\sigma / \sqrt{n}$

$$\bar{x}_1 = 20 - 1{,}96.\sqrt{12/15} \;=\; 18{,}247$$
$$\bar{x}_2 = 20 + 1{,}96.\sqrt{12/15} \;=\; 21{,}753$$

$y_1 = 18{,}247$ und $y_2 = 21{,}753$

<u>Normalverteilung als Grenzfall der Binomialverteilung:</u>

Für einen 30-maligen Münzwurf sei die Wahrscheinlichkeit zu bestimmen, mindestens 12 mal Zahl zu erzielen. Bei Verwendung der Binomialverteilung sind dafür die Wahrscheinlichkeiten für P(X=12) + P(X=13)+....+P(X=30) zu addieren um das gewünschte Ergebnis zu erhalten. Die Summe dieser Wahrscheinlichkeiten beträgt 0,8998. Unter Verwendung der Näherung mit Normalverteilung wird

$\mu = n.p = 30.0{,}5 = 15$
$\sigma = (n.p.(1-p))^{0,5} = 2{,}7386.$

Durch die Normalisierung ergibt sich für den Z-Wert: $z=(x-\mu)/\sigma = (12-15)/2{,}7386 = -1{,}095$. Aus der z-Tabelle entnehmen wir, daß der Bereich von $-\infty$ bis –1,095 einer Wahrscheinlichkeit von 0,1368 entspricht. Die Gegenwahrscheinlichkeit – unser gesuchter Wert – ist demnach 0,8632. Hier wird eine kleine Diskrepanz zum Ergebnis der Binomialverteilung deutlich: Die Wahrscheinlichkeit der Binomialverteilung ist etwas größer.

Diese kleine Diskrepanz entsteht durch die Umwandlung der *diskreten* Binomial in die *stetige* Normalverteilung. Bei der Binomialverteilung umfaßt jeder Wert eine ganze Einheit, die bei der Normalverteilung auf einen Punkt schrumpft. Der Wert zwölf bedeutet im wesentlichen die Fläche zwischen 11,5 und 12,5. Wir müssen daher bei der Normalverteilung nach der Wahrscheinlichkeit p(X>=11,5) fragen. Dann erhalten wir Z=(11,5-15)/2,7386 = -1,278. Das entspricht einer Wahrscheinlichkeit von:
$\Phi(-1{,}278) = 0{,}1007$ bzw von $\Phi(\geq 1{,}278) = 1-0{,}1007 = 0{,}8993$, ein sehr nahe bei 0,8998 gelegener Wert. Diese „Kontinuitätskorrektur“ liefert akzeptable Resultate, wenn sowohl erwartete Erfolge n.p und Mißerfolge n.(1-p) mindestens 5 betragen. Die Anpassung der Binomialverteilung an die Normalverteilung wird umso genauer, je größer n und je kleiner der Unterschied zwischen p und (1-p) ist.

Als Faustregel gilt: Approximation der Binomialverteilung mit der Normalverteilung, falls $n \geq \frac{9}{p(1-p)}$
Für die Anpassung ist eine Kontinuitätskorrektur (Stetigkeitskorrektur) erforderlich.

2.1. Summe von unabhängigen Variablen

Die Verteilungsfunktion für den Mittelwert:

$$\overline{X} \sim N(\mu, \frac{\sigma^2}{n}),$$

ist eigentlich die Verteilungsfunktion einer Summe von Zufallsvariablen X_i, mit gleichen Varianzen $\sigma^2_1 = \sigma^2_2 \ldots\ldots \sigma^2_n$. Die Verallgemeinerung dieses Zusammenhanges auf eine beliebige lineare Funktion von unabhängigen normal verteilten Zufallsvariablen lautet dann folgendermaßen:

Sind $X_i \sim N(\mu,\sigma^2_i)$, $i = 1,2,\ldots\ldots,n$ unabhängige normal verteilte Zufallsvariablen, dann ist die gewichtete Summe $Y = \Sigma a_i X_i$, wobei die a_i Konstanten sind, normalverteilt mit dem Mittelwert $\Sigma a_i \mu_i$ und der Varianz $\Sigma a^2 \sigma^2_i$

$$Y = N(\Sigma a_i \mu_i, \Sigma a_i^2 \sigma^2_i)$$

<u>Beispiel:</u>

Im Chemielabor werden die gleichen Chemikalien von 2 verschiedenen Herstellern verwendet. (Z.B) KCl –Behälter werden von beiden Herstellern in Einheiten von 2 kg geliefert; Von der Einwaage des Herstellers 1 ist bekannt, daß sie normalverteilt ist mit $\mu = 2.005$ kg, $\sigma^2 = 12g^2$, für Hersteller 2 gelten $\mu = 2.006$kg, $\sigma^2 = 15g^2$. Von jedem Hersteller wird zufällig ein Behälter ausgewählt. Wie groß ist die Wahrscheinlichkeit, daß der zweite Behälter schwerer als 4 g wie der erste ist ?

Lösung:
X_1 ist das Gewicht des Behälters von Hersteller 1, X_2 das Gewicht des zweiten Herstellers und

$X_1 \sim N(2.005,12)$,
$X_2 \sim N(2.006,15)$.

Die gesuchte Wahrscheinlichkeit ist:

$$P(X_2 > X_1 + 4) = P(X_1 - X_2 < -4)$$

Mittelwert MW uns Varianz S² sind daher mit den Koeffizienten $a_1 = 1$ und $a_2 = -1$:

$$MW = 1*2005 – 1*2006 = -1;$$
$$S^2 = 1*12 + 1*15 = 27$$

Die neue Zufallsvariable $Y = X_1\text{-}X_2 \sim N(-1,27)$.

Daher ist $P(X_1 - X_2 < -4) = P(Z < \frac{-4-(-1)}{\sqrt{27}}) = -0.58$

$\approx \Phi(-0.58)$

$\approx$

<u>Wo liegen die Probleme in der praktischen Anwendung ?</u>

Die Anwendung der Normalverteilung, bzw. der Standardnormalverteilung, setzt (stillschweigend) voraus, daß σ und μ bekannt sind. In der Realität ist das aber nur ganz selten der Fall, weil die *experimentell* bestimmten Parameter Mittelwert $\bar{x}$ und Standardabweichung s, im allgemeinen nur *Schätzungen* der Parameter μ und σ der Normalverteilung sind. Vor allem bei kleiner Stichprobengröße ist dieser Umstand relevant und daher zu berücksichtigen. Genau das tut ....

3. Die t - Verteilung

Bestimmung des Mittelwertes μ für unbekanntes σ².

Bei der Normalverteilung behandelten wir die möglichen Schlußfolgerungen über den Mittelwert einer Normalverteilung bei bekannter Varianz σ². In der Praxis ist allerdings σ² kaum bekannt, daher ist eine Verteilung zu finden, in der dieser Umstand berücksichtigt wird.

Ist $\bar{X}$ der Mittelwert einer zufälligen Stichprobe von der Größe n aus einer Normalverteilung, dann sind auch bei unbekanntem σ² beide folgenden Aussagen richtig:

$$1)\ \bar{X} \approx N(\mu, \frac{\sigma^2}{n}) \quad \text{und} \quad 2)\ Z = \frac{(\bar{X}-\mu)}{\sigma/\sqrt{n}} \approx N(0,1)$$

Diese Zusammenhänge können aber nicht genützt werden den Mittelwert zu bestimmen, weil σ unbekannt ist. σ ist daher durch S, die Standardabweichung der Stichprobe, zu ersetzen. Zur Erinnerung:

$$S^2 = \frac{\sum (X_i - \bar{X})^2}{n-1}$$

Der Erwartungswert von S² ist gleich: $E(S^2) = \sigma^2$. Es kann daher eine neue Zufallsvariable T definiert werden als:

$$T = \frac{\overline{X} - \mu}{S / \sqrt{n}}$$

Wenn die Verteilung von T bekannt ist, dann kann diese zur Bestimmung des Mittelwertes μ verwendet werden. Nachdem T aus Z durch den Ersatz der Konstanten σ durch eine Variable S erhalten wurde, ist es nicht verwunderlich, daß T nicht normalverteilt ist. Tatsächlich gibt es nicht nur eine Verteilung von T, sondern eine ganze „Familie" davon, eine Verteilung für jedes n, (n = 2,3,4…).
Diese Verteilungen werden nicht mit n indiziert, wie man erwarten könnte. Die Indizierung erfolgt mit dem Freiheitsgrad ν, wobei ν = n-1. Die übliche Bezeichungsweise

$$T \sim t_{(\nu)}$$

bringt zum Ausdruck, daß die Zufallsvariable T t-verteilt ist mit dem Freiheitsgrad ν. Nach ihrem Entdecker wird diese Verteilung ***Student t-Verteilung*** bezeichnet. Sie wurde zuerst 1908 abgleitet von W.S.Gosset, der unter dem Pseudonym Student publizierte.

Jede Verteilung von ν, ν = 1,2,3…., ist wie Z symmetrisch um Null, aber in Hinblick auf die Abhängigkeiten von 2 Statistiken $\overline{X}$ und S, ist sie variabler als Z und die p.d.f hat weitere Flanken als Z. Wenn ν größer wird, dann nähert sich die t-Verteilung der Standard-Normalverteilung. Aus den statistischen Tafeln sind die Werte für die t-Verteilung zu entnehmen:

Beispiel:

1) Annahme ν = 8; bestimme einen Wert c, sodaß gilt $P(T \geq c) = 0.05$
2) Annahme ν = 7; bestimme einen Wert c, sodaß gilt $P(|T| \geq c) = 0.01$

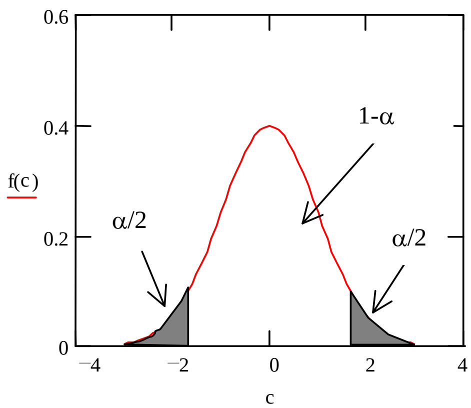
ad 1): Wenn $P(T \geq c) = 0.05$, dann ist $P(T < c) = 0.95$, daher ist q = 0.95. Aus der Tabelle wird für q = 0.95 und ν = 8 ein Wert von c = 1.8595 entnommen.

ad 2) Wenn $|T| \geq c$, dann ist $T \geq c$ oder $T \leq -c$. Daher ist $P(|T| \geq c) = P(T \geq c) + P(T \leq -c)$. Die T-Verteilung ist symmetrisch um 0, daher sind diese beiden Wahrscheinlichkeiten gleich 0.005 und

$q = P(T \leq c) = 1-P(T > c)$; Aus der Tabelle entnimmt man dafür:
c = 3.4995

3) a/ ν = 5, Bestimme ein c sodaß $P(T > c) = 0.01$
b/ ν = 30, Bestimme ein c sodaß $P(T < -c)= 0.05$
c/ ν = 10, Bestimme ein c sodaß $P(|T| < c) = 0.90$

Vertrauensbereiche für μ



Aus der Tabelle werden die Werte für c für gewählte Wahrscheinlichkeiten entnommen. Für feste Werte von ν und α gilt:

$$P(-c < T < c) = 1-\alpha$$

wobei $T \sim t_{(\nu)}$. Anders ausgedrückt bedeutet heißt dies, daß ein t-Wert für eine Wahrscheinlichkeit von α/2 für beide Seiten der Verteilung gefunden werden kann.

Nehmen wir an, $\overline{X}$ sei der Mittelwert einer Stichprobe aus einer Normalverteilung mit dem Mittelwert μ und S^2, dem Erwartungswert der Standardabweichung σ. Dann ist nach obiger Definition

$$T = \frac{\overline{X} - \mu}{S / \sqrt{n}}$$

T ist $t_{(\nu)}$ verteilt mit $\nu = n-1$. Daher ist :

$$1-\alpha = P\left(-c < \frac{(\overline{X} - \mu)}{S / \sqrt{n}} < c\right)$$
$$= P\left(\overline{X} - \frac{c.S}{\sqrt{n}} < \mu < \overline{X} + \frac{c.S}{\sqrt{n}}\right)$$

Damit erhält man ein Zufallsintervall, $(\overline{X} - c.S/\sqrt{n},\ \overline{X} + c.S/\sqrt{n})$ mit der Wahrscheinlichkeit 1 - α, daß der Mittelwert μ derPopulation in ihm enthalten ist,.
Für die Entnahme einer Probe ist das beobachtete Intervall $(\overline{x} - c.S/\sqrt{n},\ \overline{x} + c.S/\sqrt{n})$ daher ein 100.(1-α)% Konfidenzbereich für μ. Mit einer Wahrscheinlichkeit von 100.(1-α)% ist der Mittelwert daher in diesem Intervall enthalten.

4. Die Poissonverteilung

Die Binomialverteilung arbeitet meist mit kleinen n und nicht allzu großen Unterschieden von p und q. Für große n und geringe Ereigniswahrscheinlichkeiten kann die Binomialverteilung durch die Poissonverteilung ersetzt werden.

Als Faustregel wird allgemein angegeben: Approximation der Binomialverteilung mit der Poissonverteilung, falls $n \geq 50$ und $p \leq 0.05$ ist.

Binomialverteilung: $\mu = n.p$; $\sigma^2 = n.p.(1-p)$

Poissonverteilung: Für $p << q$ geht σ^2 in μ^2 über. Die Poissonverteilung ist daher durch einen einzigen Parameter, μ, bestimmt!
Sei X eine poissonverteilte Zufallsvariable mit Erwartungswert μ, dann ist die Wahrscheinlichkeit, daß k Ereignisse eintreten gegeben durch:

$$p(X = k) = \frac{e^{-\mu}.\mu^k}{k!}$$

<u>Beispiele:</u>
- Allgemein für seltene Ereignisse.
- Die Verteilung von seltenen Pflanzen oder Tieren.
- Ein viel zitiertes Beispiel ist die Anzahl der Soldaten, die jährlich in preussischen Armeekorps durch Pferde zu Tode getreten wurden.
- In der Physik folgt der radioaktive Zerfall einer Poissonverteilung. Auch hier ist der Zerfall eines Elements ein seltenes Ereignis

<u>Übungsbeispiele</u>

1. Nehmen wir an (eine Schätzung!), die Größe von Männern sei normalverteilt mit einem Durchschnittswert von 178 cm und einer Standardabweichung von 12 cm. Wie groß ist der Anteil von Männern, die größer als 2 m sind?

2. Frauen sind wohl etwas kleiner im Schnitt; Nehmen wir an, der Durchschnittswert liege 10 cm unter dem der Männer, die Standardabweichung sei aber gleich groß. Wie groß ist dann der Anteil der Frauen, die kleiner sind als der Durchschnittswert der Männer.

3. Wie groß ist unter den gegebenen Annahmen die Wahrscheinlichkeit, daß eine Frau zwischen 160 und 175 cm groß ist?

4. (Annahme!) Die Verkehrsüberwachung auf Autobahnen hat gezeigt, daß die Geschwindigkeit der KFZ eine normalverteilte ZV ist. 90 % fahren langsamer als 145 $km.h^{-1}$, aber nur 5 % fahren langsamer als 125 $km.h^{-1}$. Wie hoch ist die Durchschnittsgeschwindigkeit und wie häufig wird die „amtlich festgelegte Höchstgeschwindigkeit“ von 130 $km.h^{-1}$ übertreten?

5. Nehmen wir an, aus einer Normalverteilung mit dem Mittelwert 20 (z.B. das Gewicht kleiner Mäuse in g) und der Varianz 12 wird eine (Mäuse-)Probe von 15 gezogen. Berechne die Wahrscheinlichkeit, daß der Durchschnittswert zwischen 17 und 23 ist

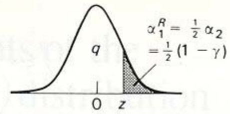
6. Eine Stichprobe vom Umfang $n = 10$ wird gezogen aus einer Grundgesamtheit mit $\mu = 180$ und $\sigma = 12$.
   a) Bestimme die Wahrscheinlichkeit, daß der Mittelwert der Probe zwischen 178 und 182 liegt.
   b) Über welchen Wert x liegt der Durchschnittswert mit einer Wahrscheinlichkeit von 0.01?

7. In einem Herstellungsprozeß werden Bolzen erzeugt, die in dazupassende Löcher montiert werden. Die normalverteilten Bolzen haben einen Mittelwert von 7mm mit einer Standardabweichung von 0.05mm. Die Löcher sind ebenfalls normalverteilt mit einem Mittelwert von 7.2mm und einer Standardabweichung von 0.01mm.

   a) Mit welcher Wahrscheinlichkeit paßt ein zufällig ausgewählter Bolzen in ein zufällig ausgewähltes Loch ?
   b) Wenn ein Werkstück 50 Löcher hat, wie groß ist dann die Wahrscheinlichkeit, daß mindestens 1 Bolzen nicht paßt.
      (ps: 0.01mm ist eine hohe Genauigkeit in der Herstellung von Bohrungen, wie ändern sich die Antworten in a und b, wenn die Standardabweichung 0.1mm beträgt)

Statistische Tabellen für die Normal- und t-Verteilung. (Aus Neave: Statistical Tables)

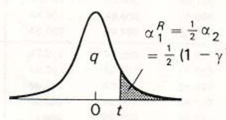
Percentage points of the normal distribution



$q=\Phi(z)$	α_1^R	α_2	γ	z
0.50				0.0000
0.60	40%			0.2533
0.70	30%			0.5244
0.80	20%	40%	60%	0.8416
0.85	15%	30%	70%	1.0364
0.90	10%	20%	80%	1.2816
0.91	9%	18%	82%	1.3408
0.92	8%	16%	84%	1.4051
0.93	7%	14%	86%	1.4758
0.94	6%	12%	88%	1.5548
0.950	5.0%	10.0%	90.0%	1.6449
0.952	4.8%	9.6%	90.4%	1.6646
0.954	4.6%	9.2%	90.8%	1.6849
0.956	4.4%	8.8%	91.2%	1.7060
0.958	4.2%	8.4%	91.6%	1.7279
0.960	4.0%	8.0%	92.0%	1.7507
0.962	3.8%	7.6%	92.4%	1.7744
0.964	3.6%	7.2%	92.8%	1.7991
0.966	3.4%	6.8%	93.2%	1.8250
0.968	3.2%	6.4%	93.6%	1.8522
0.970	3.0%	6.0%	94.0%	1.8808
0.971	2.9%	5.8%	94.2%	1.8957
0.972	2.8%	5.6%	94.4%	1.9110
0.973	2.7%	5.4%	94.6%	1.9268
0.974	2.6%	5.2%	94.8%	1.9431
0.975	2.5%	5.0%	95.0%	1.9600
0.976	2.4%	4.8%	95.2%	1.9774
0.977	2.3%	4.6%	95.4%	1.9954
0.978	2.2%	4.4%	95.6%	2.0141
0.979	2.1%	4.2%	95.8%	2.0335
0.980	2.0%	4.0%	96.0%	2.0537
0.981	1.9%	3.8%	96.2%	2.0749
0.982	1.8%	3.6%	96.4%	2.0969
0.983	1.7%	3.4%	96.6%	2.1201
0.984	1.6%	3.2%	96.8%	2.1444
0.985	1.5%	3.0%	97.0%	2.1701
0.986	1.4%	2.8%	97.2%	2.1973
0.987	1.3%	2.6%	97.4%	2.2262
0.988	1.2%	2.4%	97.6%	2.2571
0.989	1.1%	2.2%	97.8%	2.2904
0.990	1.0%	2.0%	98.0%	2.3263
0.991	0.9%	1.8%	98.2%	2.3656
0.992	0.8%	1.6%	98.4%	2.4089
0.993	0.7%	1.4%	98.6%	2.4573
0.994	0.6%	1.2%	98.8%	2.5121
0.995	0.5%	1.0%	99.0%	2.5758
0.996	0.4%	0.8%	99.2%	2.6521
0.997	0.3%	0.6%	99.4%	2.7478
0.998	0.2%	0.4%	99.6%	2.8782
0.999	0.1%	0.2%	99.8%	3.0902
0.9995	0.05%	0.1%	99.9%	3.2905
0.9999	0.01%	0.02%	99.98%	3.7190
0.99995	0.005%	0.01%	99.99%	3.8906
0.99999	0.001%	0.002%	99.998%	4.2649
0.999995	0.0005%	0.001%	99.999%	4.4172
0.999999	0.0001%	0.0002%	99.9998%	4.7534
0.9999995	0.00005%	0.0001%	99.9999%	4.8916
0.9999999	0.00001%	0.00002%	99.99998%	5.1993
0.99999995	0.000005%	0.00001%	99.99999%	5.3267
0.99999999	0.000001%	0.000002%	99.999998%	5.6120

The following notation is used in this and subsequent tables. q represents a quantile, i.e. q and the tabulated value z are related here by $\mathrm{Prob}(Z \leqslant z) = q = \Phi(z)$; e.g. $\Phi(1.9600) = q = 0.975$, where $z = 1.9600$. α_1, α_1^L and α_1^R denote significance levels for one-tailed or one-sided critical regions. Sometimes α_1^L and α_1^R values, corresponding to critical regions in the left-hand and right-hand tails, need to be tabulated separately; in other cases one may easily be obtained from the other. Here we have included only α_1^R, since α_1^L values are obtained using the symmetry of the normal distribution. Thus if a 5% critical region in the right-hand tail is required, we find the entry corresponding to $\alpha_1^R = 5\%$ and obtain $Z \geqslant 1.6449$. Had we required a 5% critical region in the left-hand tail it would have been $Z \leqslant -1.6449$. α_2 gives critical regions for two-sided tests; here $|Z| \geqslant 1.9600$ is the critical region for the two-sided test at the $\alpha_2 = 5\%$ significance level. Finally, γ indicates confidence levels for confidence intervals – so a 95% confidence interval here is derived from $|Z| \leqslant 1.9600$. For example with a large sample $X_1, X_2, \ldots, X_n$ we know that $(\bar{X} - \mu)/(s/\sqrt{n})$ has approximately a standard normal distribution, where $\bar{X} = \Sigma X_i/n$ and the adjusted sample standard deviation s is given by $s = \{\Sigma(X_i - \bar{X})^2/(n-1)\}^{1/2}$. So a 95% confidence interval for μ is derived from $|(\bar{X} - \mu)/(s/\sqrt{n})| \leqslant 1.9600$, which is equivalent to $\bar{X} - 1.96s/\sqrt{n} \leqslant \mu \leqslant \bar{X} + 1.96s/\sqrt{n}$.

Percentage points of the Student *t* distribution



q	0.95	0.975	0.99	0.995
α_1^R	5%	2.5 %	1%	0.5%
α_2	10%	5%	2%	1%
γ	90%	95%	98%	99%
ν				
1	6.3138	12.7062	31.8205	63.6567
2	2.9200	4.3027	6.9646	9.9248
3	2.3534	3.1824	4.5407	5.8409
4	2.1318	2.7764	3.7469	4.6041
5	2.0150	2.5706	3.3649	4.0321
6	1.9432	2.4469	3.1427	3.7074
7	1.8946	2.3646	2.9980	3.4995
8	1.8595	2.3060	2.8965	3.3554
9	1.8331	2.2622	2.8214	3.2498
10	1.8125	2.2281	2.7638	3.1693
11	1.7959	2.2010	2.7181	3.1058
12	1.7823	2.1788	2.6810	3.0545
13	1.7709	2.1604	2.6503	3.0123
14	1.7613	2.1448	2.6245	2.9768
15	1.7531	2.1314	2.6025	2.9467
16	1.7459	2.1199	2.5835	2.9208
17	1.7396	2.1098	2.5669	2.8982
18	1.7341	2.1009	2.5524	2.8784
19	1.7291	2.0930	2.5395	2.8609
20	1.7247	2.0860	2.5280	2.8453
21	1.7207	2.0796	2.5176	2.8314
22	1.7171	2.0739	2.5083	2.8188
23	1.7139	2.0687	2.4999	2.8073
24	1.7109	2.0639	2.4922	2.7969
25	1.7081	2.0595	2.4851	2.7874
26	1.7056	2.0555	2.4786	2.7787
27	1.7033	2.0518	2.4727	2.7707
28	1.7011	2.0484	2.4671	2.7633
29	1.6991	2.0452	2.4620	2.7564
30	1.6973	2.0423	2.4573	2.7500
31	1.6955	2.0395	2.4528	2.7440
32	1.6939	2.0369	2.4487	2.7385
33	1.6924	2.0345	2.4448	2.7333
34	1.6909	2.0322	2.4411	2.7284
35	1.6896	2.0301	2.4377	2.7238
36	1.6883	2.0281	2.4345	2.7195
37	1.6871	2.0262	2.4314	2.7154
38	1.6860	2.0244	2.4286	2.7116
39	1.6849	2.0227	2.4258	2.7079
40	1.6839	2.0211	2.4233	2.7045
42	1.6820	2.0181	2.4185	2.6981
44	1.6802	2.0154	2.4141	2.6923
46	1.6787	2.0129	2.4102	2.6870
48	1.6772	2.0106	2.4066	2.6822
50	1.6759	2.0086	2.4033	2.6778
55	1.6730	2.0040	2.3961	2.6682
60	1.6706	2.0003	2.3901	2.6603
65	1.6686	1.9971	2.3851	2.6536
70	1.6669	1.9944	2.3808	2.6479
75	1.6654	1.9921	2.3771	2.6430
80	1.6641	1.9901	2.3739	2.6387
85	1.6630	1.9883	2.3710	2.6349
90	1.6620	1.9867	2.3685	2.6316
95	1.6611	1.9853	2.3662	2.6286
100	1.6602	1.9840	2.3642	2.6259
125	1.6571	1.9791	2.3565	2.6157
150	1.6551	1.9759	2.3515	2.6090
175	1.6536	1.9736	2.3478	2.6042
200	1.6525	1.9719	2.3451	2.6006
∞	1.6449	1.9600	2.3263	2.5758

The t distribution is mainly used for testing hypotheses and finding confidence intervals for means, given small samples from normal distributions. For a single sample, $(\bar{X} - \mu)/(s/\sqrt{n})$ has the t distribution with $\nu = n - 1$ degrees of freedom (see notation above). So, e.g. if $n = 10$, giving $\nu = 9$, the $\gamma = 95\%$ confidence interval for μ is $\bar{X} - 2.2622s/\sqrt{10} \leqslant \mu \leqslant \bar{X} + 2.2622s/\sqrt{10}$. Given two samples of sizes n_1 and n_2, sample means $\bar{X}_1$ and $\bar{X}_2$, and adjusted sample standard deviations s_1 and s_2, $(\bar{X}_1 - \bar{X}_2)/\{s\sqrt{(1/n_1) + (1/n_2)}\}$ has the t distribution with $\nu = n_1 + n_2 - 2$ degrees of freedom, where $s = [\{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2\}/(n_1 + n_2 - 2)]^{1/2}$. So if the population means are denoted μ_1 and μ_2, then to test $H_0: \mu_1 = \mu_2$ against $H_1: \mu_1 > \mu_2$ at the 5% level, given samples of sizes 6 and 10, the critical region is $(\bar{X}_1 - \bar{X}_2)/(s\sqrt{\frac{1}{6} + \frac{1}{10}}) \geqslant 1.7613$, using $\nu = 6 + 10 - 2 = 14$ and $\alpha_1^R = 5\%$. As with the normal distribution, symmetry shows that α_1^L values are just the α_1^R values prefixed with a minus sign.