3.5 Die Normalverteilung

Seit dem 17. Jahrhundert entwickelten Astronomen immer präzisere Instrumente, um die Positionen von Sternen zu bestimmen.

Gleichzeitig stellten sie fest, dass ihre Messungen stets Fehler enthielten, und sie versuchten zu verstehen, wie die Fehler verteilt waren. Im Jahre 1809 veröffentlichte Carl-Friedrich Gauss seine Methode der kleinsten Fehlerquadrate und stellte sie in Zusammenhang mit einer Verteilung, die seither als die Gauss-Verteilung bekannt ist.

Der britische Astronom John Herrschel zeigte 1850, wie diese Verteilung aus einfachen Annahmen über die ihr zugrunde liegenden Prinzipien entsteht. Wir entwickeln hier, wie sie mit elementaren Argumenten aus diesen Annahmen hergeleitet werden kann.

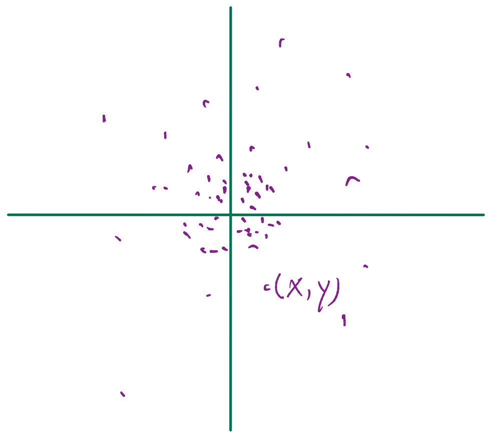
Astronomen bestimmten die Position eines Himmelsobjekts mit Teleskopen, die in horizontaler und vertikaler bewegt werden können. Das Objekt hat dann eine eindeutig bestimmte Position, aber bei jeder Messung findet man eine (etwas) andere Position.

Wir nehmen an, dass Positionen mit (x,y)-Koordinaten beschrieben werden und dass das Objekt, für das wir uns interessieren, sich im Ursprung $(0,0)$ befindet. Die (x,y)-Messungen eines Astronomen können wir als Werte zweier Zufallsvariablen X, Y ansehen, die eine gemeinsame Verteilung haben.

Sei $d(x,y)$ die Dichte dieser gemeinsamen Verteilung. Dieses d ist dann die Dichte der Fehler-Verteilung, weil jede von $(0,0)$ verschiedene Messung fehlerhaft ist.

Was sind vernünftige Annahmen über d?
Herrschel schlug zwei vor:

- Die Wahrscheinlichkeit von Fehlern (x,y) sollte nicht von ihrer Richtung vom Ursprung abhängen, sondern nur von der Entfernung vom Ursprung.



- Fehler in der x-Richtung sollten unabhängig sein von Fehlern in der y-Richtung. (Astronomen benutzten zwei unterschiedliche Mechanismen für die Kalibrierung ihrer Teleskope in jeder Richtung.

Was bedeutet dies mathematisch?

- Der Abstand des Punkts (x,y) vom Ursprung ist $\sqrt{x^2+y^2}$ (Pythagoras!). Daher gibt es eine Funktion

  $g: \mathbb{R}_0^+ \longrightarrow \mathbb{R}_0^+$, so dass gilt

$$d(x,y) = g(\sqrt{x^2+y^2})$$

- Seien f_X, f_Y die marginalen Dichten von d.
  Aus der Unabhängigkeit von X und Y folgt

$$d(x,y) = f_X(x) \cdot f_Y(y)$$

Untersuchen wir zunächst die Beziehung zwischen f_X und f_Y.

Weil d abhängt vom Abstand des Punkts (x,y) vom Ursprung, haben wir

$$d(x,0) = g(\sqrt{x^2+0^2}) = g(\sqrt{0^2+x^2}) = d(0,x)$$

Damit ist

$$f_X(x) \cdot f_Y(0) = d(x,0) = d(0,x) = f_X(0) \cdot f_Y(x)$$

und daher

$$f_Y(x) = \frac{f_Y(0)}{f_X(0)} \cdot f_X(x)$$

Sowohl f_X wie f_Y sind Dichten. Deshalb gilt

$$1 = \int_{\mathbb{R}} f_Y(x)\,dx = \frac{f_Y(0)}{f_X(0)} \int_{\mathbb{R}} f_X(x)\,dx = \frac{f_Y(0)}{f_X(0)} \cdot 1 = \frac{f_Y(0)}{f_X(0)}$$

Wir schließen daraus, dass

$$f_Y(x) = f_X(x) \qquad \text{f.a. } x \in \mathbb{R}$$

Wir haben gesehen, dass X und Y dieselbe Dichte haben, die wir als f schreiben. Da $d(x,y) = g(\sqrt{x^2+y^2})$ ist haben wir

$$g(\sqrt{x^2+y^2}) = f(x) \cdot f(y) \quad , \quad \text{f.a. } x,y \in \mathbb{R}$$

Sind x, y nicht-negativ, so ist $x = \sqrt{x^2}$ und $y = \sqrt{y^2}$.
Wir können dann die Gleichung oben schreiben als

$$g(\sqrt{x^2+y^2}) = f(\sqrt{x^2}) \cdot f(\sqrt{y^2}).$$

Wir sehen also, dass die Funktion $g(\sqrt{\cdot})$ Summen von Quadraten abbildet auf Produkte von Werten dieser Quadrate unter $f(\sqrt{\cdot})$.

Für nicht-negative x gilt außerdem

$$g(x) = g(\sqrt{x^2+0}) = f(x) \cdot f(0) = K \cdot f(x)$$

mit $K = f(0)$, oder

$$f(x) = \frac{1}{K} g(x).$$

Aus der Gleichung

$$g(\sqrt{x^2+y^2}) = f(\sqrt{x^2})\cdot f(\sqrt{y^2})$$

schließen wir

$$(*)\quad g(\sqrt{x^2+y^2}) = f(\sqrt{x^2})\cdot f(\sqrt{y^2}) = \frac{1}{K}\,g(\sqrt{x^2})\cdot\frac{1}{K}\,g(\sqrt{y^2}) \quad \text{f.a. } x\in\mathbb{R}$$

Jede nichtnegative Zahl ist das Quadrat einer nichtnegativen Zahl. Deshalb bedeutet $(*)$, dass für alle $u,v\in\mathbb{R}_0^+$ gilt

$$g(\sqrt{u+v}) = \frac{1}{K^2}\,g(\sqrt{u})\cdot g(\sqrt{v}).$$

Multiplizieren wir dies mit $\frac{1}{K^2}$, erhalten wir

$$\frac{1}{K^2}\,g(\sqrt{u+v}) = \frac{1}{K^2}\,g(\sqrt{u})\cdot\frac{1}{K^2}\,g(\sqrt{v}).$$

Mit $h(u) := \frac{1}{K^2}\,g(\sqrt{u})$ ergibt dies

$$h(u+v) = h(u)\cdot h(v) \qquad \text{f.a. } u,v\in\mathbb{R}_0^+$$

Aus

$$h(u+v) = h(u)\cdot h(v) \qquad \text{f.a. } u,v\in\mathbb{R}_0^+$$

schließen wir, gestützt auf unsere Resultate über Exponentialfunktionen,

$$h(u) = a^u \qquad \text{für ein } a>0$$

Wegen $h(u) = \frac{1}{K^2}\cdot g(\sqrt{u})$ haben wir

$$\frac{1}{K^2}\,g(\sqrt{u}) = a^u \quad , \text{ f.a. } u\geq 0 .$$

Es gilt auch $g(x) = K\cdot f(x)$. Daher ist für $x\geq 0$

$$a^x = \frac{1}{K^2}\,g(\sqrt{x}) = \frac{1}{K^2}\cdot K f(\sqrt{x}) = \frac{1}{K}\,f(\sqrt{x})$$

$$\Rightarrow f(\sqrt{x}) = K\cdot a^x$$

$$\Rightarrow f(x) = f(\sqrt{x^2}) = K\cdot a^{x^2}, \qquad \text{f.a. } x\geq 0$$

Wir haben also

$$f(x) = K \cdot a^{x^2} \qquad \text{f.a. } x \geq 0$$

Was ist mit negativen x? Beachte, dass

$$f(x) \cdot f(0) = g\left(\sqrt{x^2 + 0^2}\right) = g\left(\sqrt{(-x)^2 + 0^2}\right)$$
$$= f(-x) \cdot f(0),$$

und damit

$$f(x) = f(-x) \qquad \text{f.a. } x \in \mathbb{R}$$

Somit ist

$$f(x) = K \cdot a^{x^2} \qquad \text{f.a. } x \in \mathbb{R}$$

Somit haben unsere marginalen Dichten $f_X = f_Y = f$ die Gestalt

$$f(x) = K \cdot a^{x^2}.$$

Was bedeutet dies für K und a?

Aus den Bedingungen für eine Dichte schließen wir

$$f \geq 0 \qquad \text{und} \qquad \int_{\mathbb{R}} f(x)\, dx = 1.$$

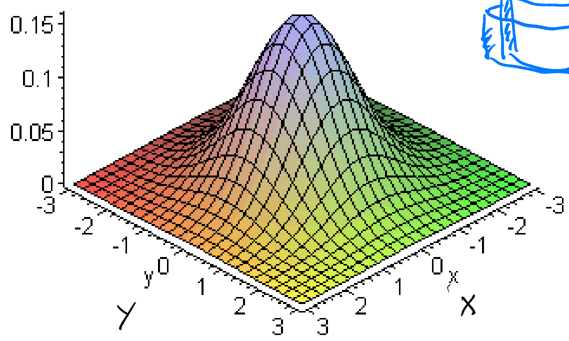
Die erste Bedingung ist offensichtlich erfüllt, falls $K > 0$ ist, da $a^{x^2} > 0$ ist f.a. $x \in \mathbb{R}$.

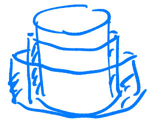
Die zweite Bedingung impliziert, dass $a < 1$ ist, weil sonst $\lim_{x \to \infty} a^{x^2} = \infty$ gälte. Sei daher $\alpha := \log \frac{1}{a}$. Dann ist $\alpha > 0$ und

$$f(x) = K e^{-\alpha x^2}$$

Bivariate Normal

$d(x,y) = f(x) \cdot f(y)$





Die Konstanten K und α sind miteinander verbunden durch die Bedingung

$$K \int_{\mathbb{R}} e^{-\alpha x^2} \, dx = 1$$

Wir möchten gerne verstehen, wie man zu einem gegebenen K das α findet und umgekehrt.

Das ist allerdings schwer, weil die Stammfunktionen von e^{x^2} sich nicht durch einen elementaren Ausdruck darstellen lassen.

Allerdings sind wir nicht an der Verteilungsfunktion (pdf)

$$\int_{-\infty}^{x} K e^{-\alpha y^2} \, dy$$

interessiert, sondern an dem Integral

$$I := \int_{-\infty}^{\infty} e^{-\alpha x^2} \, dx,$$

denn $K = \frac{1}{I}$.

Ohne Stammfunktion von $e^{-\alpha x^2}$ ist auch I schwer zu berechnen.

Ursprünglich waren wir allerdings nicht interessiert an der Dichte

$$f(x) = K e^{-\alpha x^2},$$

sondern an $d(x,y) = f(x) \cdot f(y)$.

Was können wir schließen aus der Bedingung

$$1 = \int_{\mathbb{R}} \int_{\mathbb{R}} d(x,y)\, dx\, dy = K^2 \int_{\mathbb{R}} \int_{\mathbb{R}} e^{-\alpha x^2} e^{-\alpha y^2}\, dx\, dy$$

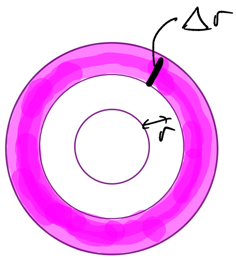
$$= K^2 I_2 \; ?$$

Wir konzentrieren uns zunächst auf I_2:

$$I_2 = \int_{\mathbb{R}} \int_{\mathbb{R}} e^{-\alpha x^2} e^{-\alpha y^2}\, dx\, dy = \int_{\mathbb{R}} \int_{\mathbb{R}} e^{-\alpha(x^2+y^2)}\, dx\, dy$$

Der Integrand hängt nur ab vom Abstand r des Arguments vom Ursprung:
Wenn (x,y) auf einem Kreis mit Radius r liegt,
hat der Integrand den Wert $e^{-\alpha r^2}$.

Ein Kreis mit Breite Δr und Radius r hat ungefähr eine Fläche $2\pi r \cdot \Delta r$ und trägt ungefähr den Wert



$$e^{-\alpha r^2} \cdot 2\pi r \cdot \Delta r$$

zum Integral bei. Geht $\Delta r \to 0$, ergibt dies

$$I_2 = \int_0^\infty 2\pi r\, e^{-\alpha r^2}\, dr.$$

Dies Integral können wir auswerten.

I_2 kann mit der Substitutionsregel ausgewertet werden:

Hier sind f und g nur Symbole, nicht die Funktionen von vorher

$$I_2 = \int_0^\infty 2\pi r\, e^{-\alpha r^2}\, dr \qquad = C \int_0^\infty f(g(r)) \cdot g'(r)\, dr$$

$$= -\frac{\pi}{\alpha} \int_0^\infty \overbrace{(-e^{-\overbrace{\alpha r^2}^{g}})}^{f}\, \overbrace{2\alpha r}^{g'}\, dr$$

$$= -\frac{\pi}{\alpha} \int_{g(0)}^{g(\infty)} f(z)\, dz$$

$$= -\frac{\pi}{\alpha} \int_{g(0)}^{g(\infty)} -e^{-z}\, dz$$

$$= -\frac{\pi}{\alpha} \left[e^{-z} \right]_{g(0)}^{g(\infty)} = -\frac{\pi}{\alpha} \left[e^{-z} \right]_0^\infty$$

$$= -\frac{\pi}{\alpha} (0 - 1) = \frac{\pi}{\alpha}$$

Ausgangspunkt für unsere Rechnungen war die Bedingung

$$K^2 I_2 = 1 .$$

Wir wissen jetzt

$$K^2 \frac{\pi}{\alpha} = 1 \quad \text{und daher} \quad K = \sqrt{\frac{\alpha}{\pi}} .$$

Somit ist

$$f(x) = \frac{\sqrt{\alpha}}{\sqrt{\pi}}\, e^{-\alpha x^2}$$

die Dichte von X und Y.

Mittelwert und Varianz von f:

$$f(x) = \frac{\sqrt{\alpha}}{\sqrt{\pi}} \quad e^{-\alpha x^2}$$

Mittelwert: Offenbar ist f symmetrisch um 0, das heißt $f(x) = f(-x)$.
Also ist der Mittelwert μ, der Schwerpunkt der Verteilung, gleich 0.

Varianz: Wir wenden partielle Integration an:

$$\int f g' = fg - \int f' g$$

$$\sigma^2 = \int_{\mathbb{R}} (x-\mu)^2 f(x)\, dx = \int_{\mathbb{R}} x^2 f(x)\, dx$$

$$= K \int_{\mathbb{R}} x^2 e^{-\alpha x^2} dx = K \int_{\mathbb{R}} \underbrace{\left(-\frac{1}{2\alpha} x\right)}_{f} \underbrace{\left(-2\alpha x \cdot e^{-\alpha x^2}\right)}_{g'} dx$$

$$= K \left(\left[\underbrace{\left(-\frac{1}{2\alpha} x\right)}_{f} \underbrace{\left(e^{-\alpha x^2}\right)}_{g} \right]_{-\infty}^{\infty} - \int_{\mathbb{R}} \underbrace{-\frac{1}{2\alpha}}_{f'} \underbrace{e^{-\alpha x^2}}_{g} dx \right)$$

$$= \frac{1}{2\alpha} K \int_{\mathbb{R}} e^{-\alpha x^2} dx = \frac{1}{2\alpha}$$

Allgemeine Form der normalen Dichte (mit $\mu = 0$)

$$\sigma^2 = \frac{1}{2\alpha} \quad \Longrightarrow \quad \alpha = \frac{1}{2\sigma^2}$$

$$\Longrightarrow \quad K = \frac{\sqrt{\alpha}}{\sqrt{\pi}} = \frac{1}{\sqrt{2\sigma^2}} \cdot \frac{1}{\sqrt{\pi}} = \frac{1}{\sqrt{2\pi}\,\sigma}$$

Also

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

Dies ist eine Dichte mit Mittelwert $\mu = 0$
Varianz σ^2

Allgemeine Form der normalen Dichte mit beliebigem Mittelwert

Angenommen, der Stern, den wir beobachten, ist nicht an der Stelle $(0,0)$, sondern an der Stelle (μ,ν).
Dann hängt die Fehlerdichte ab vom Abstand von diesem Punkt, das heißt von

$$\sqrt{(x-\mu)^2+(y-\nu)^2}$$

In diesem Fall haben die marginalen Dichten die Form

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

oder die analoge Form mit ν.

Wir sagen, dass ein ZV mit dieser Dichte normalverteilt ist nach $\mathcal{N}(\mu,\sigma^2)$.
Im Falle von $\mathcal{N}(0,1)$ sprechen wir von der Standard-Normalverteilung, mit Dichte

$$\phi(x)=\frac{1}{\sqrt{2\pi}}\, e^{-\frac{x^2}{2}}$$

Die Verteilungsfunktion von $\mathcal{N}(0,1)$

Die Verteilungsfunktion der Standard-Normalverteilung wird bezeichnet als Φ und erfüllt

$$\Phi(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{y^2}{2}}\,dy.$$

Die Funktion Φ kann nicht elementar dargestellt werden (das heißt, es gibt keine Formel). Sie kann näherungsweise berechnet werden durch numerische Integration. Implementieren gibt es in statistischen Bibliotheken (R, Java Packages). Es gibt auch Tabellen.

Oft möchte man, gegeben eine Wahrscheinlichkeit p, das x finden, so dass gilt

$$\Phi(x)=P[X\le x]=p,$$

also

$$x=\Phi^{-1}(p).$$ Umkehrfunktion (inverse function)

Tabellen der Normalverteilung

Tabellen sind das traditionelle Mittel, Werte von Φ zu finden. Um Redundanz zu vermeiden, enthalten sie nur Werte $\Phi(x)$ für $x \geq 0$, also mit $\Phi(x) \geq 0{,}5$.

Die Symmetrie von ϕ spiegelt sich in Φ dadurch wider, dass

$$\Phi(-x) = 1 - \Phi(x) \qquad x \geq 0,$$

denn für eine $\mathcal{N}(0,1)$-verteilte ZV Z haben wir

$$\begin{aligned} \Phi(-x) = P[Z \leq -x] &= P[Z \geq x] \\ &= 1 - P[Z \leq x] \\ &= 1 - \Phi(x) \end{aligned}$$

Eigenschaften der Normalverteilung

Wir sagen, dass X normalverteilt ist, wenn

$$X \sim \mathcal{N}(\mu, \sigma^2) \qquad \text{für ein } \mu \in \mathbb{R} \text{ und ein } \sigma \in \mathbb{R}^+.$$

Satz: Seien X, Y normalverteilt und unabhängig, und seien $a, b \in \mathbb{R}$. Dann gilt:
- $aX + b$
- $X + Y$

sind normalverteilt.

Beweis (Idee): Ist $X \sim f$ (Dichte f), dann ist $aX + b \sim g$ mit $g(y) = f\left(\frac{y-b}{a}\right)$, denn $y = ax + b \Rightarrow x = \frac{y-b}{a}$

Überprüfe: Ist f Dichte einer Normalverteilung, dann auch g.

Der zweite Teil ist schwieriger zu zeigen und man braucht die Faltung.

Korollar: Seien $X \sim \mathcal{N}(\mu_X, \sigma_X^2)$, $Y \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$, $a, b \in \mathbb{R}$.
Dann gilt

- $aX + b \sim \mathcal{N}(a\mu_X + b, a^2\sigma_X^2)$
- $X + Y \sim \mathcal{N}(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2)$

Wir schreiben $\mathcal{N}(0,1)$-verteilte ZVen als Z.

Satz: Seien $Z \sim \mathcal{N}(0,1)$, $X \sim \mathcal{N}(\mu, \sigma^2)$. Dann gilt

- $\sigma Z + \mu \sim \mathcal{N}(\mu, \sigma^2)$
- $\dfrac{X - \mu}{\sigma} \sim \mathcal{N}(0,1)$

Beispiel 61: Wir wollen Signale 0, 1 über einen Kanal mit Rauschen (noise) senden. Wir kodieren

0	als	-2
1	als	2.

Der Empfänger (receiver) sieht Signale der Form

$$E = x + R, \quad R \sim \mathcal{N}(0,1)$$

und dekodiert nach den Regeln

$E \geq 0{,}5$	wird	1
$E < 0{,}5$	wird	0

Wie groß ist die Fehlerwahrscheinlichkeit in jedem Fall?

Sender: 0 als -2
1 als 2

$E = x + R$

Empfänger: $E \geq 0{,}5$ als 1
$E < 0{,}5$ als 0

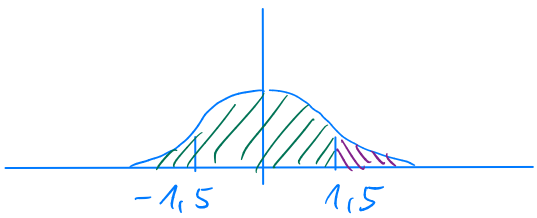
Fehler beim Empfang von 1:

$$P[E < 0{,}5 \mid S = 2] = P[x + R < 0{,}5 \mid x = 2]$$

$$= P[R < -1.5] = P[R > 1.5] = 1 - P[R \leq 1.5]$$

Mit R: pnorm(-1.5)

Tabelle: $1 - \Phi(1{,}5)$



Sender: 0 als -2
1 als 2

$E = x + R$

Empfänger: $E \geq 0{,}5$ als 1
$E < 0{,}5$ als 0

Fehler beim Empfang von 2:

$$P[E \geq 0{,}5 \mid S = -2] = P[x + R \geq 0{,}5 \mid x = -2]$$

$$= P[-2 + R \geq 0{,}5] = P[R \geq 2{,}5]$$

$$= 1 - P[R < 2.5]$$

Mit R: pnorm(-1.5, lower.tail = FALSE)

Tabelle: $1 - \Phi(2{,}5)$

Beispiel 62: Angenommen die Größe europäischer Männer ist normalverteilt mit Mittelwert $\mu = 177.6$ cm und Standardabweichung $\sigma = 4$ cm.

- Wie groß ist die Wahrscheinlichkeit, dass von zwei zufällig gewählten Männern der ältere mindestens 2 cm größer ist als der jüngere (unter der Annahme, dass Alter und Größe unabhängig sind)?

Beispiel 62: Angenommen die Größe europäischer Männer ist normalverteilt mit Mittelwert $\mu = 177.6$ cm und Standardabweichung $\sigma = 4$ cm.

- Wie groß ist die Wahrscheinlichkeit, dass von zwei zufällig gewählten Männern der ältere mindestens 2 cm größer ist als der jüngere (unter der Annahme, dass Alter und Größe unabhängig sind)?

Sei die ZV X die Größe europäischer Männer und seien X_1, X_2 zwei unabhängige Kopien von X.

Sei $D := X_1 - X_2$ die Differenz der Größen zweier zufälliger Männer. Wir fragen nach

$$P[D \geq 2]$$

Wir wissen

$$X_1, X_2 \sim N(\mu, \sigma^2) \quad \Rightarrow \quad -X_2 \sim N(-\mu, \sigma^2)$$

$$\Rightarrow D = X_1 - X_2 \sim N(\mu - \mu, \sigma^2 + \sigma^2) = N(0, 2\sigma^2)$$

Dann gilt

$$P[D \geq 2] = P\left[\frac{1}{\sqrt{2}\sigma} D \geq \frac{2}{\sqrt{2}\sigma}\right] = P\left[Z \geq \frac{2}{\sqrt{2}\sigma}\right]$$

$$= 1 - P\left[Z \leq \frac{2}{\sqrt{2}\sigma}\right]$$

$$= 1 - \Phi\left(\frac{2}{\sqrt{2}\,4}\right)$$

$$\approx 1 - \Phi(0{,}3536) = 0{,}3632$$

Die 68 – 95 – 99,7 – Regel

Sei $Z \sim N(0,1)$. Dann gilt

$$P[-1 \leq Z \leq 1] = 0{,}68$$

$$P[-2 \leq Z \leq 2] = 0{,}95$$

$$P[-3 \leq Z \leq 3] = 0{,}997$$

Das heißt

68 % / 95 % / 99,7 % aller Werte liegen innerhalb 1 / 2 / 3 Standardabweichungen vom Mittelwert

Für ein $X \sim N(\mu, \sigma^2)$ gilt

$$P[\mu - \sigma \leq X \leq \mu + \sigma] = 0{,}68$$

$$P[\mu - 2\sigma \leq X \leq \mu + 2\sigma] = 0{,}95$$

$$P[\mu - 3\sigma \leq X \leq \mu + 3\sigma] = 0{,}997$$