

Medizinische Hochschule Hannover
Institut für Biometrie

Analytische Statistik in 10 Zügen

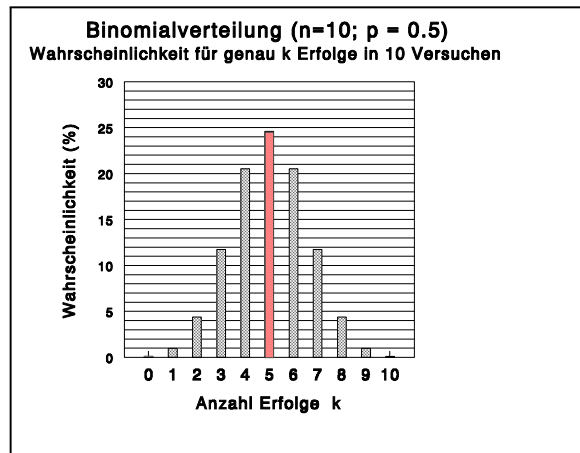
H. Hecker

Contents

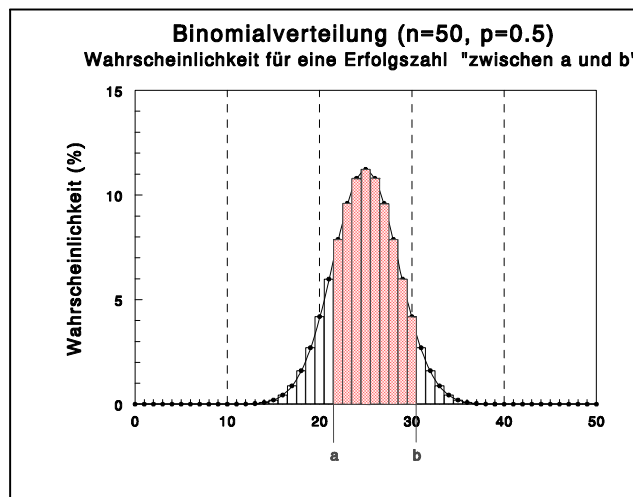
1	Grundannahmen der analytischen Statistik	1
2	Grundgesamtheit und Stichproben: Definition	2
3	Beziehung zwischen Grundgesamtheit und Stichprobe	2
4	Die Zufallsverteilung der Stichprobenmittelwerte: Erwartungswert und Varianz	3
5	Die Zufallsverteilung des Stichprobenmittelwertes: Standardisierung	4
6	Das Dilemma der reinen Wahrscheinlichkeitsrechnung	6
7	Die 5%-Hürde: Aktuelle Stichprobenwerte versus Wahrscheinlichkeitsrechnung	6
8	Konfidenzbereich: Welcher Erwartungswert einer Verteilung ist mit den Stichprobendaten "verträglich"?	6
9	Auf den Punkt gebracht: Sind die Stichprobendaten mit dem Punkt Null verträglich?	7
10	Wie man sich irren kann: Fehler erster und zweiter Art und der P-Wert	8

1. Grundannahmen der analytischen Statistik

- Die einzelne Beobachtung eines Meßwertes (z.B.: RRsys = 180 mmHg) ist das Ergebnis eines Prozesses der (auch) von *zufälligen* (komplexen, unkontrollierbaren) Effekten beeinflusst wird. Man sagt:
Jeder einzelne beobachtete Meßwert y_1, \dots, y_n ist die "Realisierung einer Zufallsvariablen Y "
- Durch die Rahmenbedingungen, unter denen das Experiment (die Beobachtung) durchgeführt wird, werden die *Wahrscheinlichkeiten* festgelegt, mit denen die Zufallsvariable einen bestimmten Wert k (z.B. $k = 5$) annimmt,



oder in ein bestimmtes Intervall $[a, b]$ fällt:



$$P(a \leq Y \leq b) = \text{Wahrscheinl. dafür, daß } Y$$

in das Intervall $[a, b]$ fällt

- Die Gesamtheit dieser Wahrscheinlichkeiten heißt die *Wahrscheinlichkeitsverteilung* der Zufallsvariablen Y . In verkürzter Form kann sie durch einige *Verteilungsparameter* wie *Erwartungswert* $\mu = E(Y)$ und *Varianz* $\sigma^2 = E(Y - \mu)^2$ beschrieben werden.

Der eigentliche Gegenstand einer statistischen Untersuchung ist nicht die Einzelbeobachtung y_i sondern die *Wahrscheinlichkeitsverteilung der Zufallsvariablen Y* .

2. Grundgesamtheit und Stichproben: Definition

Die *Grundgesamtheit* ist die Population aller potentiellen Beobachtungseinheiten für das Experiment Y , zusammen mit Rahmenbedingungen, unter denen das Experiment durchgeführt wird.

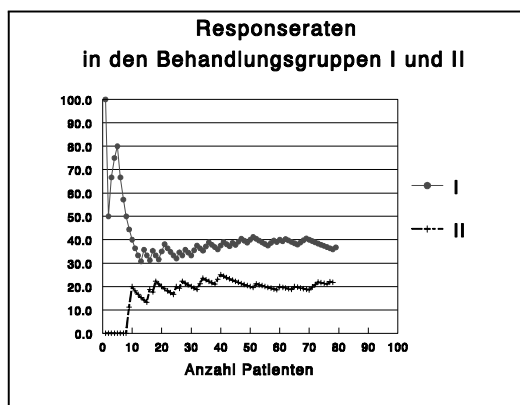
Die *Stichprobe vom Umfang n* ist die Realisation von n unabhängigen Experimenten, das sind n unabhängige Zufallsvariablen Y_1, \dots, Y_n , die alle dieselbe Zufallsverteilung wie Y haben.

Eine solche Stichprobe, also eine größere Anzahl von Beobachtungen, wird benötigt, um stabilere Aussagen über die interessierende "Verteilung in der Grundgesamtheit" zu erhalten.

Experiment: Anzahl "Zahl" nach 10 Münzwürfen			
Grundgesamtheit		Stichprobe (n=20)	
Mögliches Ergebnis	Wahrscheinlichkeit (%)	Ergebnisse: 5 6 0 7 6 4 8 6 3 6 4 7 4 4 4 4 3 4 4 4	
		absolut	Häufigkeit relativ (%)
0	0.1	1	5
1	1.0	0	0
2	4.4	0	0
3	11.7	2	10
4	20.5	9	45
5	24.6	1	5
6	20.5	4	20
7	11.7	2	10
8	4.4	1	5
9	1.0	0	0
10	0.1	0	0

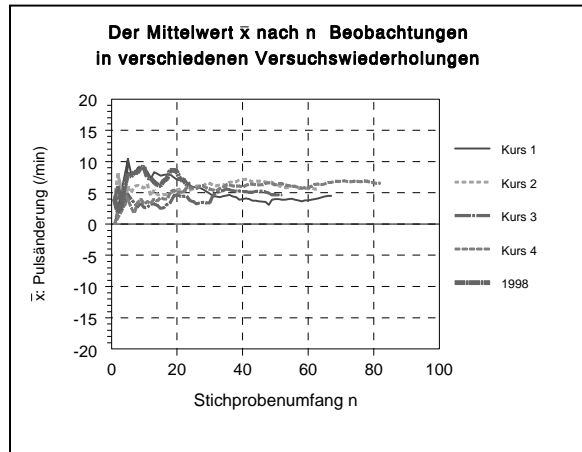
3. Beziehung zwischen Grundgesamtheit und Stichprobe

Die relativen Häufigkeiten $h_n(A)$ nach n Versuchen für das Ereignis A werden als **Schätzungen für die Wahrscheinlichkeit $P(A)$** interpretiert



Analog:

Der Stichprobenmittelwert \bar{y} nach n Versuchen wird als **Schätzung für den Erwartungswert μ** interpretiert.



4. Die Zufallsverteilung der Stichprobenmittelwerte: Erwartungswert und Varianz

Wie die Einzelbeobachtungen y_1, y_2, \dots, y_n , so ist auch der Stichprobenmittelwert \bar{y} von zufälligen Effekten abhängig: Aus den n unabhängigen Zufallsvariablen Y_1, Y_2, \dots, Y_n wird der Stichproben-Mittelwert als *neue Zufallsvariable* gebildet:

$$\bar{Y} = \frac{1}{n}(Y_1 + Y_2 + \dots + Y_n).$$

Wenn man die Verteilung des Einzelexperimentes Y kennt, so kann man daraus (im Prinzip) auch die Verteilung des Stichprobenmittelwertes berechnen. Für *Erwartungswert* und *Varianz* gilt *immer* (unabhängige Wiederholungen desselben Experimentes vorausgesetzt):

Wenn die Einzelexperimente Y_i den Erwartungswerte $E(Y_i) = \mu$ und die Varianz $var(Y_i) = \sigma^2$ haben so ist

$$E(\bar{Y}) = \mu$$

Der Stichprobenmittelwert \bar{Y} ist ein "erwartungstreuer" Schätzer (eine erwartungstreue "Schätzfunktion") für den Erwartungswert μ .

Und:

$$var(\bar{Y}_n) = \frac{1}{n}\sigma^2 \quad \text{und}$$

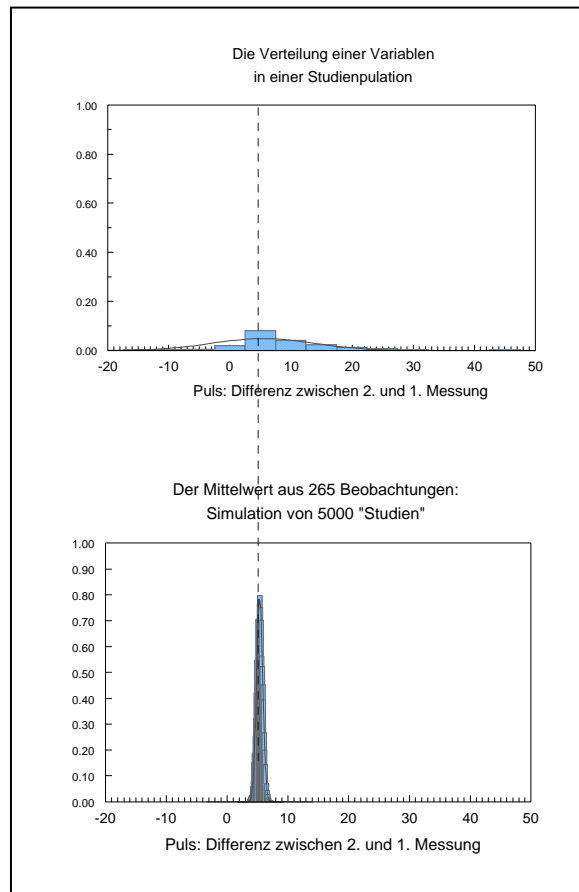
$$\sigma(\bar{Y}_n) = \sqrt{\frac{1}{n}\sigma^2} = \frac{\sigma}{\sqrt{n}}$$

(Die "Eins durch Wurzel n" — Regel)

$\sigma(\bar{Y}_n)$ ist der "Standardfehler des Mittelwertes" (engl. S.E.M. = "Standard Error of the Mean"). Er wird geschätzt durch

$$S.E.M. = \sqrt{\frac{s_Y^2}{n}} = \frac{s_Y}{\sqrt{n}}$$

Beispiel:

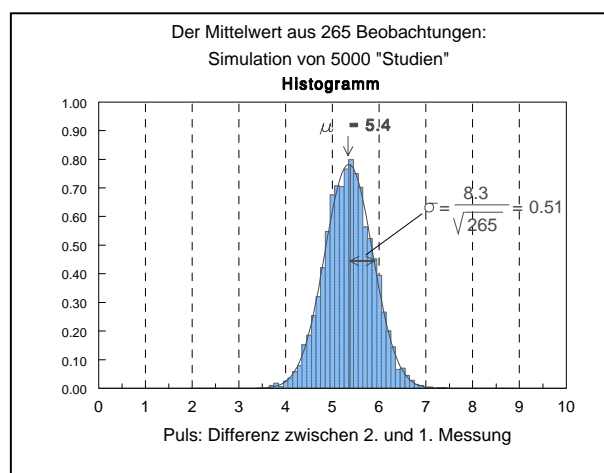


5. Die Zufallsverteilung des Stichprobenmittelwertes: Standardisierung

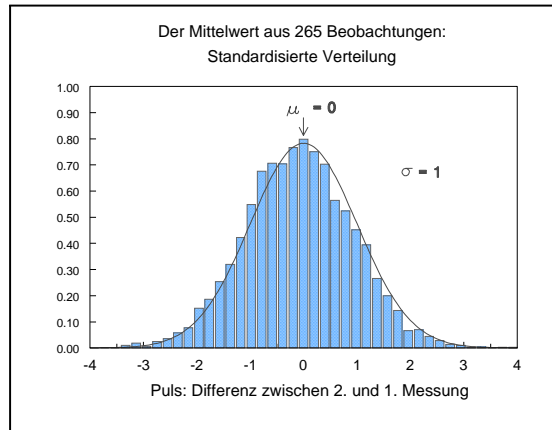
Nach dem *Zentralen Grenzwertsatz* ist für große Stichprobenumfänge der standardisierte Mittelwert "ungefähr" standard-normalverteilt:

$$Z_n = \frac{\bar{Y}_n - \mu}{\frac{1}{\sqrt{n}}\sigma} \approx N(0, 1)$$

Die Verteilung des Mittelwertes:



Nach Standardisierung:



Geht man davon aus, daß die Varianz σ^2 der Grundgesamtheit "hinreichend" genau durch die empirische Varianz s_Y^2 geschätzt wird, so kann man darin σ durch s_Y ersetzen, d.h. auch

$$\frac{\bar{Y}_n - \mu}{\frac{s_Y}{\sqrt{n}}} = \frac{\bar{Y}_n - \mu}{S.E.M.}$$

ist "etwa" standard-normalverteilt.

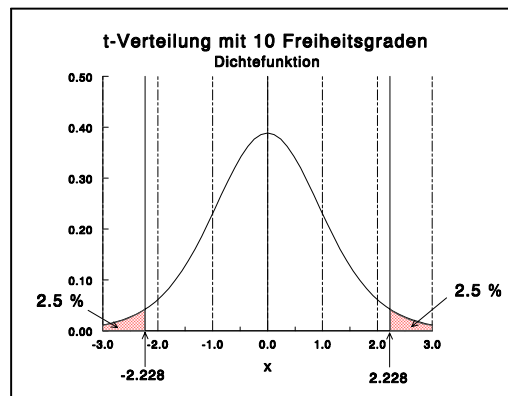
Sind alle "Einzelversuche" Y_i schon selber *normalverteilt* mit Erwartungswert μ und Varianz σ^2 :

$$Y_i \sim N(\mu, \sigma^2) \quad (i = 1, 2, \dots, n)$$

so ist die Verteilung von

$$T_n := \frac{\bar{Y}_n - \mu}{\frac{s_Y}{\sqrt{n}}}$$

die (Student'sche) "t-Verteilung mit $n-1$ Freiheitsgraden".



Das q -Quantil der t -Verteilung mit $n - 1$ Freiheitsgraden wird mit $t_{n-1,q}$ bezeichnet. Aufgrund der Symmetrie zum Wert 0 gilt (speziell für $q = \alpha/2$):

$$P(|T_n| \leq t_{n-1,1-\alpha/2}) = 1 - \alpha, \quad (5.1)$$

d.h. T_n ist mit der Wahrscheinlichkeit $1 - \alpha$ Werte dem Betrage nach höchstens gleich dem oberen $\alpha/2$ -Quantil an.

Umgekehrt ist daher

$$P(|T_n| > t_{n-1,1-\alpha/2}) = \alpha, \quad (5.2)$$

und speziell für $\alpha = 0.05 = 5\%$:

$$P(|T_n| > t_{n-1,0.975}) = 0.05 = 5\%: \quad (5.3)$$

T_n ist mit der Wahrscheinlichkeit von 5% dem Betrage nach größer als das obere 97.5%-Quantil der t -Verteilung mit $n - 1$ Freiheitsgraden.

6. Das Dilemma der reinen Wahrscheinlichkeitsrechnung

Wahrscheinlichkeitsrechnung ist reine Mathematik. Die numerische Berechnung von Wahrscheinlichkeiten ist immer nur unter bestimmten *Annahmen und Voraussetzungen* möglich (z.B. wenn man der Erwartungswert einer Verteilung kennt). Aber über eben diese Voraussetzungen weiß man in der Regel nichts.

Man kann z.B. berechnen, daß (für $n=11$): $\left| \frac{\bar{Y}_n - \mu}{\frac{s_Y}{\sqrt{n}}} \right|$ mit 95% Wahrscheinlichkeit ≤ 2.28 ausfallen muß. Aber was nutzt einem das, wenn man μ nicht kennt?

7. Die 5%–Hürde: Aktuelle Stichprobenwerte versus Wahrscheinlichkeitsrechnung

Grundsatz:

Akzeptiere eine Annahme über die Verteilung von Y dann, aber auch nur dann, wenn die aktuelle Beobachtung aus der Stichprobe unter dieser Annahme nicht als "zu unwahrscheinlich" erscheint!

Übliche Grenze hierfür: 5% Wahrscheinlichkeit (die "5%–Hürde").

8. Konfidenzbereich: Welcher Erwartungswert einer Verteilung ist mit den Stichprobendaten "verträglich"?

Mit Wahrscheinlichkeit 95% fällt (für $n=11$) die Testgröße $T_n := \frac{\bar{Y}_n - \mu}{\frac{s_Y}{\sqrt{n}}}$ zwischen die Grenzen -2.228 und $+2.228$ (siehe Abb. oben), d.h.:

$$\frac{\bar{Y}_n - \mu}{\frac{s_Y}{\sqrt{n}}} \leq 2.228 \quad \text{und} \quad \frac{\bar{Y}_n - \mu}{\frac{s_Y}{\sqrt{n}}} \geq -2.228$$

1. Daher ist jeder Wert μ mit den vorliegenden Daten \bar{y}_n und s_y (Mittelwert und Streuung) noch "vereinbar", der in diesem Sinne "nahe genug am Stichprobenmittelwert \bar{y}_n liegt", für den also beide Ungleichungen erfüllt sind. Löst man diese nun nach μ auf, so ist dies gleichbedeutend mit:

$$\bar{y}_n - 2.228 \frac{s_y}{\sqrt{n}} \leq \mu \quad \text{und} \quad \mu \leq \bar{y}_n + 2.228 \frac{s_y}{\sqrt{n}}$$

2. Man erhält also das Intervall mit den Grenzen

$$\bar{y}_n - 2.228 \frac{s_y}{\sqrt{n}} \quad \text{und} \quad \bar{y}_n + 2.228 \frac{s_y}{\sqrt{n}} .$$

Bei dieser Konstruktion passiert es nur mit einer (Fehler-) Wahrscheinlichkeit von 5%, daß man mit dem angegebene Intervall den den "wahren" Parameter μ nicht erfaßt.

3. Allgemeine Version: Das von der Stichprobe abhängige, also *zufällige Intervall* mit den Grenzen

$$\bar{Y}_n - t_{n-1, 1-\alpha/2} \frac{s_Y}{\sqrt{n}} \quad \text{und} \quad \bar{Y}_n + t_{n-1, 1-\alpha/2} \frac{s_Y}{\sqrt{n}} \quad (8.1)$$

überdeckt den wahren Erwartungswert μ der Verteilung von Y in der Grundgesamtheit mit der Wahrscheinlichkeit von $1 - \alpha$.

Das Intervall (8.1) wird "Konfidenzintervall für den Erwartungswert μ zum Niveau $1 - \alpha$ " genannt.

9. Auf den Punkt gebracht: Sind die Stichprobendaten mit dem Punkt Null vertäglich?

Allgemein:

Wie groß ist die Wirkung des Medikamentes?

Wie groß sind die Wirkungen der Behandlungen A und B im Vergleich; wie groß ist die Differenz ihrer Wirkungen?

Wie groß ist die Korrelation zwischen dem Alter und dem Test-Score im Zahlenverbindungstest?

Speziell:

Ist die Wirkung des Medikamentes gleich **null**?

Ist die Differenz der Wirkungen von Behandlung A und B gleich **null**?

Ist dies Korrelation zwischen Alter und Zahlenverbindungstest gleich **null**?

Beispiel Z-Test:

Wenn die Stichprobe X_1, X_2, \dots, X_n von Messungen der Blutdruckveränderung ("vor Behandlung" minus "nach Behandlung") aus einer normalverteilten Grundgesamtheit stammen mit jeweils dem (unbekannten) Erwartungswert μ und der (bekannten) Streuung $\sigma = 10$ mmHg, so ist der Stichprobenmittlerwert

$$S(X_1, X_2, \dots, X_n) := \bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$$

ebenfalls normalverteilt mit demselben Erwartungswert μ und der Streuung $\frac{\sigma}{\sqrt{n}}$.

Wenn man dann annimmt, daß die Nullhypothese $H_0 : \mu = 0$ richtig wäre (angenommen also: Keine Veränderung des Blutdrucks zu erwarten), so kann man *vorausberechnen*, mit welchen Wahrscheinlichkeiten die Testgröße $S = \bar{X}$ nach z.B. $n = 49$ Versuchen welche Werte annehmen wird: $S \sim N(0, (\frac{10}{\sqrt{49}})^2)$, d.h. S ist normalverteilt mit Erwartungswert 0 und Standardabweichung $\frac{10}{\sqrt{49}} = \frac{10}{7} = 1.4286$. Um die Berechnung dieser Verteilung auf die *Standardnormalverteilung* zurückzuführen, transformiert man S noch mit Hilfe der Z-Transformation und erhält dann die standard-normalverteilte Testgröße Z :

$$Z = \frac{S - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{\bar{X}_n}{1.4286} \sim N(0, 1)$$

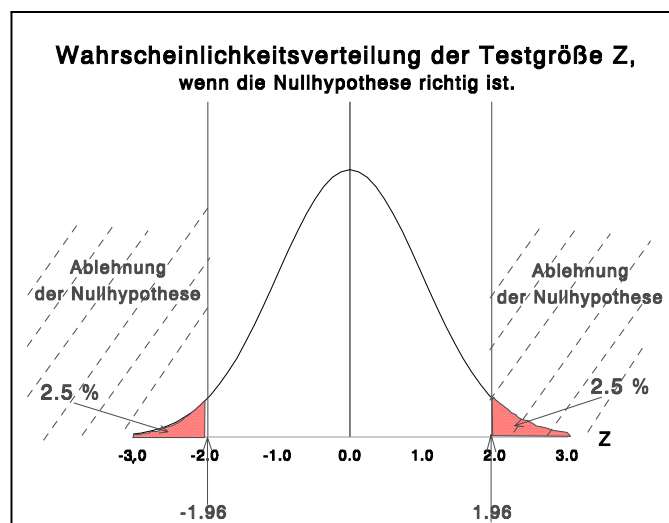


Abb.5.3: Ablehnbereich und "kritischer Wert" einer Testgröße

Es wird nun vorab festgelegt, daß die Nullhypothese abgelehnt wird, wenn Z "stark positiv" oder "stark negativ" ausfällt, genauer: wenn $Z < -1.96$ oder $Z > 1.96$ wird. Damit ist der *Ablehnbereich*

festgelegt:

$$\text{Lehne die Nullhypothese ab, wenn } |Z| = \frac{|\bar{X} - \mu|}{\frac{\sigma}{\sqrt{n}}} = \frac{|\bar{X}_n|}{1.4286} \geq 1.96$$

Der sogenannte "kritische Wert" $\lambda = 1.96$ für $|Z|$ ist dabei so gewählt, daß die Wahrscheinlichkeit für eine Ablehnung von H_0 nicht größer als 5%.

Beispiel (Z-Test für $n = 49$, $\sigma = 10$, $\alpha = 0.05 = 5\%$; Nullhypothese $H_0 : \mu = 0$):

\bar{X}	Testgröße Z	Kriterium:	Entscheidung:
-4.3	-3.01	$ -3.01 = 3.01 \geq 1.96$	Ablehnung der H_0
-1.5	-1.05	$ -1.05 = 1.05 < 1.96$	H_0 nicht widerlegt
1.8	1.26	$ 1.26 = 1.26 < 1.96$	H_0 nicht widerlegt
2.3	1.75	$ 1.75 = 1.75 < 1.96$	H_0 nicht widerlegt
3.3	2.31	$ 2.31 = 2.31 \geq 1.96$	Ablehnung von H_0

10. Wie man sich irren kann: Fehler erster und zweiter Art und der P-Wert

Ein Signifikanztest zum Niveau α ist ein Verfahren, welches in Abhängigkeit von den Daten einer Stichprobe über die Ablehnung oder die Annahme einer Nullhypothese H_0 entscheidet; ist dabei H_0 tatsächlich richtig, so wird höchstens mit der Wahrscheinlichkeit α gegen H_0 entschieden.

Die Kontrolle des Fehlers 1. Art (auch "α-Fehler" genannt) schützt ("bis auf α ") vor der fälschlichen Annahme einer Nullhypothese, schützt also beispielsweise

- vor der Zulassung und Verordnung eines Medikamentes, dessen Wirkung (im Vergleich zum Placebo) gleich null ist,
- vor der (zusätzlichen) Anwendung eines diagnostischen Eingriffs, dessen Ergebnis mit den untersuchten Krankheiten keinen Zusammenhang hat,
- vor der öffentlichen Behauptung eines Zusammenhangs zwischen X und Y , obwohl beide stochastisch unabhängig voneinander sind.

Die Anwendung eines Testes ist in diesem Sinn eine *Sicherheitsmaßnahme*. Tatsächlich aber wird der Test und das damit eingeleitete Entscheidungsverfahren *mit der Intention durchgeführt, die Nullhypothese zu widerlegen*: Man sucht

- nach wirksamen Medikamenten und will ihre Wirksamkeit, wenn sie denn vorhanden ist, nachweisen;
- nach Maßnahmen, welche ein Diagnoseverfahren echt verbessern können;
- nach tatsächlich vorhandenen Zusammenhängen, wenn man nach den Ursachen oder Entstehungsbedingungen einer Krankheit forscht.

Während man sich also einerseits davor hüten möchte, bei der Behauptung von Zusammenhängen irgendwelchen Zufälligkeiten aufzusitzen, sollen andererseits aber *tatsächlich vorhandene Zusammenhänge* auch erkannt und nachgewiesen werden. Ein Signifikanztest sollte also im folgenden Sinne eine hohe "Teststärke" haben:

Teststärke (Güte, "power") = **Wahrscheinlichkeit, eine tatsächlich vorliegende Abweichung von der Nullhypothese zu entdecken**

Eine hohe Teststärke ist damit gleichbedeutend mit einer niedrigen Wahrscheinlichkeit für den folgenden Fehler:

Fehler 2. Art: Beibehaltung der Nullhypothese, obwohl sie falsch ist

Die Wahrscheinlichkeit für den Fehler 2. Art wird i.a. mit β bezeichnet und der Fehler 2. Art dementsprechend auch der " β -Fehler" genannt. Es gilt somit

$$\beta = 1 - \text{Güte ("power")} \text{ des Tests}$$

Angenommen, im vorliegenden Beispiel ist der Stichprobenmittelwert $\bar{x} = 3.3$ und der Wert der Testgröße daher gleich 2.31. Man berechne dann, wie groß (unter der Nullhypothese) die Wahrscheinlichkeit dafür ist, daß diese oder eine noch größere Abweichung von der Nullhypothese auftreten kann. Das heißt hier: Wie wahrscheinlich ist es, daß $|Z| > 2.31$ wird? Ist diese Wahrscheinlichkeit kleiner als $\alpha = 5\%$, wird die Nullhypothese abgelehnt, im anderen Fall wird sie beibehalten. Im vorliegenden Fall beträgt die Wahrscheinlichkeit $2 \times 1.044\% = 2.088\%$, die Nullhypothese wird also abgelehnt. In der folgenden Abb. 5.4 sind diese Beziehungen graphisch dargestellt:

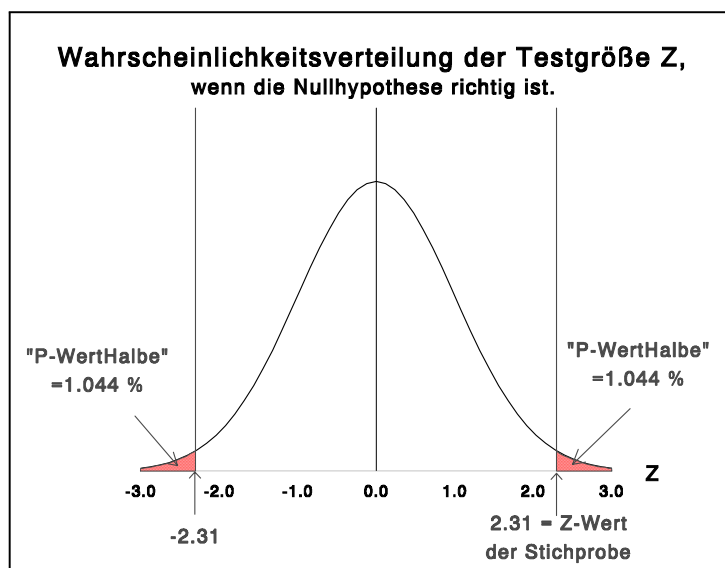


Abb. 5.4: Der P-Wert einer Stichprobe im Z-Test

Die so berechnete Wahrscheinlichkeit wird als der "*P-Wert*" des Tests oder als "*Abweichwahrscheinlichkeit*" bezeichnet. In den gängigen Statistik-Programmsystemen wird üblicherweise dieser P-Wert (häufig auch unter der Bezeichnung "*Signifikanz*") ausgedrückt. Dadurch wird die Kenntnis der "*kritischen Werte*" λ überflüssig. Die zum vorliegenden Beispiel entsprechende Tabelle sieht dann so aus:

\bar{X}	Testgröße Z	P-Wert	Kriterium:	Entscheidung:
-4.3	-3.01	0.26 %	$0.26 \% \leq 5 \%$	Ablehnung von H_0
-1.5	-1.05	29.37 %	$29.37 \% > 5 \%$	H_0 nicht widerlegt
1.8	1.26	20.77 %	$20.77 \% > 5 \%$	H_0 nicht widerlegt
2.3	1.75	8.01 %	$8.01 \% > 5 \%$	H_0 nicht widerlegt
3.3	2.31	2.09 %	$2.09 \% \leq 5 \%$	Ablehnung von H_0

Über seine Funktion als Entscheidungskriterium hinaus ist der P-Wert aber auch von eigenem Interesse: Aufgrund seiner Konstruktion kann er als "*Abweichwahrscheinlichkeit*" interpretiert werden:

Der P-Wert gibt an, mit welcher Wahrscheinlichkeit bei Gültigkeit der Nullhypothese H_0 eine Abweichung von H_0 auftreten kann, die so groß ist oder noch größer als aktuell beobachtet.