# Statistik im Klartext

## Für Psychologen, Wirtschafts- und Sozialwissenschaftler

2., aktualisierte und erweiterte Auflage

Fabian Heimsch
Rudolf Niederer
Peter Zöfel

Pearson

EXTRAS ONLINE

# Statistik im Klartext

## Für Psychologen, Wirtschafts- und Sozialwissenschaftler

2., aktualisierte und erweiterte Auflage

Fabian Heimsch
Rudolf Niederer
Peter Zöfel

Pearson

## 5.1 Schätzen

In einer Stadt soll der Mittelwert der Körpergrößen aller erwachsenen männlichen Einwohner ermittelt werden. Theoretisch könnte man die Messung an dieser kompletten Grundgesamtheit vornehmen und dann hieraus den Mittelwert berechnen. Dies wurde an den 12 193 erwachsenen männlichen Einwohnern durchgeführt; die betreffenden Werte sind in der Datei `stadt.sav` gespeichert. Der Mittelwert aller Größenangaben beträgt 175,61 cm.

Damit Sie sich keine Sorgen um die Autoren machen, sei zugegeben, dass diese Messungen nicht real durchgeführt, sondern mit einem Computer simuliert wurden. Auf diese Weise wurden 12 193 normalverteilte Körpergrößenangaben zwischen 142 und 204 cm erzeugt.

Ebenfalls mit einem Computerprogramm wurde nun eine Zufallsstichprobe von 10 Personen gezogen und der Mittelwert gebildet; dieses Verfahren wurde dann mit anderen Stichprobenumfängen wiederholt. Die Ergebnisse sind in Tabelle 5.1 eingetragen.

| Stichprobengröße | Mittelwert |
|---|---|
| 10 | 178,00 |
| 20 | 176,30 |
| 50 | 176,64 |
| 100 | 176,25 |
| 200 | 176,27 |
| 500 | 175,82 |
| 1000 | 175,60 |

**Tabelle 5.1:** Mittelwerte bei steigender Stichprobengröße

Mit steigender Fallzahl wird also der Mittelwert der Grundgesamtheit (175,61) immer besser angenähert. Wie in Abschnitt 6.2.1 erläutert wird, geht man beim Schluss vom Mittelwert der Stichprobe auf den entsprechenden Parameter der Grundgesamtheit so vor, dass man ein *Konfidenzintervall* angibt, innerhalb dessen sich der Mittelwert der Grundgesamtheit mit einer vorgegebenen Wahrscheinlichkeit bewegt. Dabei können Konfidenzintervalle nicht nur für Mittelwerte, sondern auch für Standardabweichungen (siehe Abschnitt 6.2.2) und prozentuale Häufigkeiten (siehe Abschnitt 6.2.3) berechnet werden.

## 5.2 Testen von Hypothesen

Wurde vorangehend der Schluss von den Kennwerten *einer* Stichprobe auf die entsprechenden Parameter der Grundgesamtheit behandelt, soll nun der Fall betrachtet werden, dass zwei (oder mehr) Stichproben vorliegen, deren Kennwerte daraufhin überprüft werden sollen, ob sie zu der gleichen Grundgesamtheit gehören oder nicht. Man spricht in diesem Zusammenhang von *Prüfstatistik*.

Ein Beispiel mag dies erläutern. Insgesamt 129 Patienten mit etwa gleichen Ausgangswerten von erhöhten Cholesterin-Werten wurden mit zwei verschiedenen Medikamenten behandelt. Nach einem halben Jahr wurde der Cholesterinwert erneut festgestellt; dabei ergaben sich die in Tabelle 5.2 enthaltenen Ergebnisse.

| Kollektiv | Mittelwert | Standardabweichung | Fallzahl |
|---|---|---|---|
| Medikament A | 193,0 | 43,1 | 66 |
| Medikament B | 208,0 | 33,5 | 63 |

**Tabelle 5.2:** Kennwerte zweier Stichproben

Die Patienten mit Medikament A haben also einen geringeren durchschnittlichen Cholesterin-Wert. Es gibt nun zwei Möglichkeiten:

1. Der Mittelwertsunterschied ist zufällig zustande gekommen.
2. Der Mittelwertsunterschied ist nicht zufällig zustande gekommen; er ist *signifikant*.

Die Frage der Signifikanz ist das zentrale Thema der analytischen Statistik. Nicht nur Unterschiede von Mittelwerten können auf Signifikanz geprüft werden, sondern zum Beispiel auch Unterschiede von Standardabweichungen, Prozentwerten und Häufigkeitsverteilungen; auch Korrelations- und Regressionskoeffizienten etwa können auf Signifikanz getestet werden, genauer gesagt daraufhin, ob sie sich signifikant von Null unterscheiden (siehe Kapitel 9 bis 11).

Die analytische Statistik gibt objektive Testverfahren an die Hand, nach deren Ergebnis eine Beurteilung möglich ist, ob eine Signifikanz vorliegt oder nicht. Wir betrachten hierzu das gegebene Beispiel und können die beiden folgenden Hypothesen formulieren.

- Hypothese 0 (H0): Der Mittelwertsunterschied ist zufällig zustande gekommen.
- Hypothese 1 (H1): Der Mittelwertsunterschied ist nicht zufällig zustande gekommen.

Die beiden Hypothesen lassen sich auch wie folgt formulieren:

- H0: Die beiden Stichprobenmittelwerte gehören zu der gleichen Grundgesamtheit.
- H1: Die beiden Stichprobenmittelwerte gehören zu verschiedenen Grundgesamtheiten.

Die Hypothese H0 nennt man die Nullhypothese, die Hypothese H1 die Alternativhypothese.

Ob die Nullhypothese beibehalten wird oder zugunsten der Alternativhypothese zu verwerfen ist, wird anhand der betreffenden Prüfstatistik entschieden. Je nach Testsituation entwickelte man hierfür zahlreiche Tests, von denen die wichtigsten im weiteren Verlauf des Buches vorgestellt werden.

Zum Vergleich zweier Mittelwerte $\overline{x}_1$ und $\overline{x}_2$ bei bekannten Standardabweichungen $s_1$ und $s_2$ und bekannten Fallzahlen $n_1$ und $n_2$ gibt es den $t$-Test nach Student (siehe Abschnitt 9.1), den wohl bekanntesten statistischen Test, bei dessen einfacherer Variante zunächst die folgende Prüfgröße berechnet wird:

$$t = \frac{|\overline{x}_1 - \overline{x}_2|}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Zu dieser Prüfgröße $t$ wird noch die so genannte Anzahl der *Freiheitsgrade df* bestimmt (df = degrees of freedom):

$$df = \frac{n_1 + n_2 - 2}{2}$$

Im vorliegenden Beispiel ergeben die Berechnungen

$$t = \frac{|\,193{,}0 - 208{,}0\,|}{\sqrt{\dfrac{43{,}1^2}{66} + \dfrac{33{,}5^2}{63}}} = 2{,}213$$

$$df = \frac{66 + 63 - 2}{2} = 65$$

Zu dieser Prüfgröße $t$ hat W. S. Gosset, der den $t$-Test unter dem Pseudonym Student veröffentlichte, im Jahre 1908 auch die zugehörige Verteilung, die nach ihm benannte Studentsche $t$-Verteilung, entwickelt. Diese Verteilung ist wie die Normalverteilung (siehe Abschnitt 4.3.1) eine symmetrische und eingipflige Verteilung, deren Gestalt von der Anzahl der Freiheitsgrade abhängt und die sich bei hohen Freiheitsgraden der Normalverteilung annähert.

Mit Hilfe dieser Verteilung kann zur Prüfgröße $t$ und zur Anzahl $df$ der Freiheitsgrade eine Wahrscheinlichkeit $p$ bestimmt werden:

$$p = 2 \cdot \frac{\Gamma\left(\dfrac{df+1}{2}\right)}{\Gamma\left(\dfrac{df}{2}\right) \cdot \sqrt{df \cdot \pi}} \cdot \int_t^{\infty} \left(1 + \frac{v^2}{df}\right)^{-\dfrac{df+1}{2}} dv$$

Natürlich wird diese Formel nicht von Hand gerechnet. So braucht auch die Gamma-Funktion, die sich in der Formel wiederfindet, nicht erläutert zu werden. Das Integral ist aber ein Hinweis darauf, dass die Wahrscheinlichkeit einer bestimmten Fläche unter der $t$-Verteilungskurve entspricht.

Gosset war übrigens Angestellter der Guinness-Brauerei und entwickelte $t$-Test und $t$-Verteilung anlässlich der Analyse von Bierproben, womit die öfter diskutierte Frage, ob Bier dumm oder intelligent macht, eindrucksvoll beantwortet wird.

Abbildung 5.2 zeigt den Verlauf der Dichte der Standardnormalverteilung (fett gedruckt) und darunter die dazugehörigen $t$-Verteilungen mit den Freiheitsgraden $df = 1, 2, 3, 5,$ und 10. Erhöht sich die Anzahl Freiheitsgrade $df$ der $t$-Verteilung, so nähert sich diese der Standardnormalverteilung an.

Die Berechnung dieser exakten Wahrscheinlichkeit ist erst seit der Entwicklung entsprechender Computerprogramme möglich geworden. Im gegebenen Beispiel erhalten wir $p = 0{,}03$, was folgendermaßen zu deuten ist:

*Die Wahrscheinlichkeit – unter der Annahme, die Nullhypothese sei richtig – dass das gegebene Untersuchungsergebnis oder ein noch extremeres auftritt, beträgt* $p = 0{,}03$. *Dieser Wert wird in der Statistik als p-Wert bezeichnet.*

Da wir in Kapitel 3 gelernt haben, dass sich Wahrscheinlichkeiten stets zwischen den Werten 0 und 1 bewegen, werden wir die Wahrscheinlichkeit von 0,03 als sehr klein einstufen. Mittelwertsunterschiede und überhaupt alle Untersuchungsergebnisse, die mit einer solch kleinen Wahrscheinlichkeit behaftet sind, nennt man daher *signifikant*.

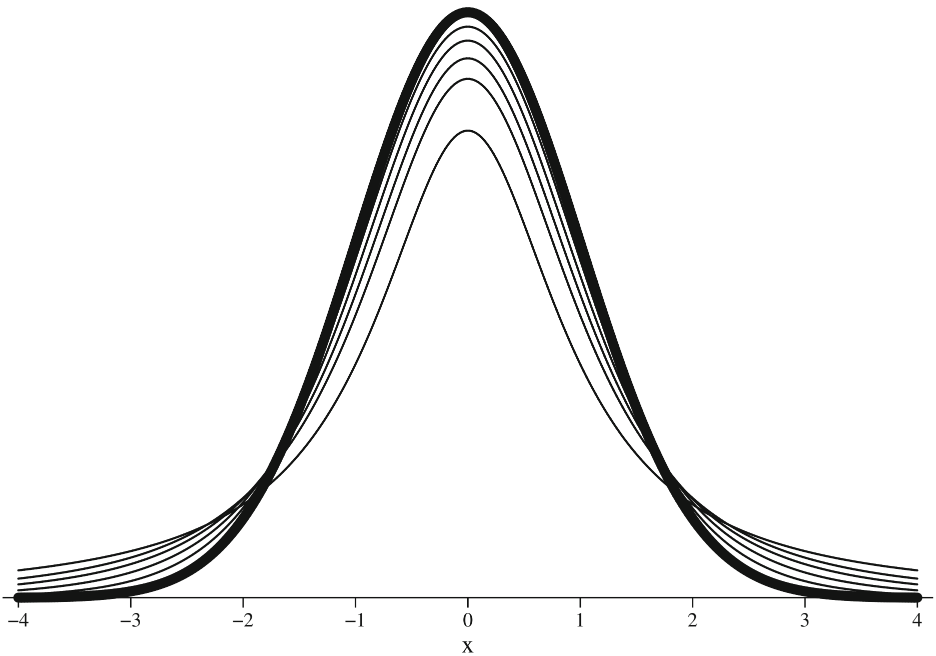


Abbildung 5.2: Standardnormalverteilung (fett) und $t$-Verteilungen mit $df = 1, 2, 3, 5$ und 10

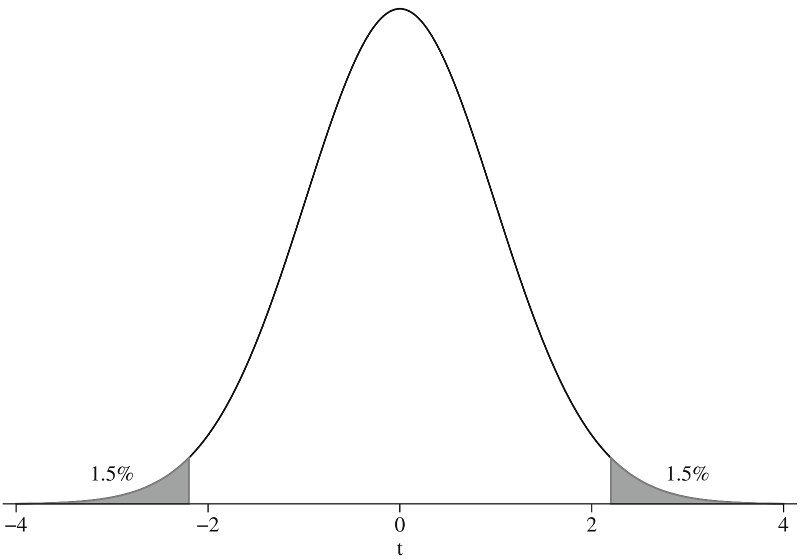


Abbildung 5.3: Zur Illustration des $p$-Wertes unter der $t$-Verteilung

Man spricht in diesem Zusammenhang von der *Irrtumswahrscheinlichkeit*. Dabei gibt es klassischerweise drei Signifikanzstufen:

| | | |
|---|---|---|
| $p \leq 0{,}05$ | signifikant | * |
| $p \leq 0{,}01$ | sehr signifikant | ** |
| $p \leq 0{,}001$ | höchst signifikant | *** |

Die im gegebenen Beispiel ermittelte Irrtumswahrscheinlichkeit von $p = 0{,}03$ bedeutet also, dass sich die Mittelwerte des Cholesterins bei den Medikamenten A und B signifikant voneinander unterscheiden; d. h. die Nullhypothese H0 wird verworfen. Man nimmt die Alternativhypothese H1 an und die Wahrscheinlichkeit, dass der Entscheid, die Nullhypothese zu verwerfen, falsch war, ist $p = 0{,}03$

Es sei an dieser Stelle der Hinweis gegeben, dass die Tatsache der Signifikanz nicht unbedingt auch mit einer fachlichen (hier: medizinischen) Bedeutsamkeit einhergehen muss. Testen Sie zum Beispiel eine neue Diät und stellen fest, dass alle Versuchspersonen ihr Gewicht in einem Monat um 100 Gramm reduzierten, so wird dieses wohl ein höchst signifikantes Ergebnis, medizinisch aber nicht bedeutsam und daher unbefriedigend sein, denn Signifikanz ist nicht gleich Relevanz.

In früherer computerloser Zeit, als es nicht möglich war, die Irrtumswahrscheinlichkeit $p$ aus der Prüfgröße und der Anzahl der Freiheitsgrade exakt zu berechnen, behalf man sich mit tabellierten Grenzwerten (so genannten kritischen Werten), wobei üblicherweise die kritischen Werte zu $\alpha = 0{,}05$, $\alpha = 0{,}01$ und $\alpha = 0{,}001$ tabelliert wurden. Solche Tabellen finden Sie in Anhang A.

Der $t$-Tabelle entnehmen Sie zum Signifikanzniveau $\alpha = 0{,}05$ und zu 65 Freiheitsgraden den kritischen Tabellenwert 1,997. Dieser wird vom berechneten $t$-Wert (2,213) überschritten, was Signifikanz auf diesem Niveau bedeutet.

## 5.3 Fehler erster und zweiter Art

Hat man Nullhypothese und Alternativhypothese formuliert, so kann man beim Überprüfen dieser Hypothesen mit einem passenden statistischen Test offenbar zwei Fehler machen:

- Die Nullhypothese wird verworfen, obwohl sie richtig ist.
- Die Nullhypothese wird beibehalten, obwohl sie falsch ist.

Der erstgenannte Fehler heißt Fehler erster Art oder $\alpha$-Fehler. Die Wahrscheinlichkeit, einen Fehler erster Art zu begehen, ist gleich der Irrtumswahrscheinlichkeit $p$. Der zweitgenannte Fehler heißt Fehler zweiter Art oder $\beta$-Fehler; die Wahrscheinlichkeit, einen solchen Fehler zu begehen, ist allenfalls bei präzise bekannter Alternativhypothese berechenbar, wie dies am Schluss des folgenden Abschnitts gezeigt wird. Es lässt sich auf alle Fälle sagen, dass die Gefahr, einem $\beta$-Fehler zu erliegen, umso kleiner ist, je deutlicher die berechnete Irrtumswahrscheinlichkeit $p$ die Signifikanzgrenze übersteigt.

Zur Verdeutlichung sei noch einmal das Schema in Tabelle 5.3 betrachtet. Hier sind die Verhältnisse in der Wirklichkeit (H0 wahr, H0 falsch) den Ergebnissen des Signifikanztests (H0 abgelehnt, H0 beibehalten) gegenübergestellt.

| | H0 wahr | H0 falsch |
|---|---|---|
| H0 abgelehnt | Fehler 1. Art | richtige Entscheidung |
| H0 beibehalten | richtige Entscheidung | Fehler 2. Art |

**Tabelle 5.3:** Fehler erster und zweiter Art

Haben Sie sich die übliche Signifikanzgrenze von $\alpha = 0{,}05$ vorgegeben und erzielen bei Ihrem Signifikanztest, zum Beispiel beim $t$-Test, ein $p = 0{,}07$, so müssen Sie also die Nullhypothese beibehalten. Die Gefahr, dass Sie das fälschlicherweise tun und somit einen Fehler zweiter Art begehen, wird aber recht groß sein. Erzielen Sie hingegen ein $p = 0{,}9$, so wird die Gefahr, die Nullhypothese fälschlicherweise beizubehalten, eher gering sein.

Testen Sie also zum Beispiel zwei Mittelwerte auf signifikanten Unterschied und erhalten ein $p$ knapp oberhalb der Signifikanzgrenze, so wäre eine Formulierung der Art „Die beiden Mittelwerte unterscheiden sich nicht“ unangemessen; besser wäre eine vorsichtigere Formulierung wie „Beim Vergleich der beiden Mittelwerte wurde die Signifikanzgrenze knapp verfehlt“. Für Irrtumswahrscheinlichkeiten $p \leq 0{,}1$ verwendet man auch hin und wieder die Formulierung „Tendenz zur Signifikanz“.

Nehmen Sie an, ein Hersteller testet den Erfolg eines von ihm neu entwickelten Medikaments und vergleicht diesen mit dem Erfolg eines bestehenden Medikaments. Liefert der betreffende Signifikanztest keinen signifikanten Unterschied, obwohl in Wirklichkeit einer besteht, so geht das Risiko dieses nicht erkannten Unterschieds zu Lasten des Produzenten, so dass man das Risiko, einen solchen Fehler zweiter Art zu begehen, auch als *Produzentenrisiko* bezeichnet.

Zeigt der betreffende Signifikanztest hingegen einen signifikant besseren Erfolg des neuen Medikaments an, obwohl ein solcher in Wirklichkeit nicht besteht, so geht das Risiko dieses fälschlicherweise erkannten Unterschieds zu Lasten des Konsumenten, so dass man das Risiko, einen solchen Fehler erster Art zu begehen, auch als *Konsumentenrisiko* bezeichnet.

Ist $\beta$ die Wahrscheinlichkeit dafür, dass ein bestehender Unterschied nicht erkannt wird, so ist $1 - \beta$ die Wahrscheinlichkeit dafür, dass ein bestehender Unterschied auch aufgezeigt wird. Diesen Wert bezeichnet man als *Teststärke* (auch: *Power, Güte, Trennschärfe.*)

In Zusammenhang mit diesen Begriffen wurden Verfahren entwickelt, um den für einen geplanten Test optimalen Stichprobenumfang abzuschätzen. Dieser soll dann bei vorgegebenem $\alpha$ eine maximale Teststärke $1 - \beta$ garantieren. Da diese Verfahren recht kompliziert und überdies von vielen Unwägbarkeiten begleitet sind, wollen wir erst im Kapitel Varianzanalyse näher darauf eingehen.

## 5.4 Einseitige und zweiseitige Fragestellung

Im Allgemeinen wird über die Richtung der Alternativhypothese von vornherein keine Aussage zu machen sein. Beim vorgestellten Beispiel der beiden Patientengruppen (mit Medikament A bzw. B) war nicht abzusehen, welche der beiden Gruppen gegebenenfalls höhere Cholesterinwerte aufweist. In allen diesen Fällen ist *zweiseitig* zu testen. Dies ist die normale Testform und im weiteren Verlauf dieses Buches wird auch stets so getestet, ohne dass jeweils besonders darauf hingewiesen wird.

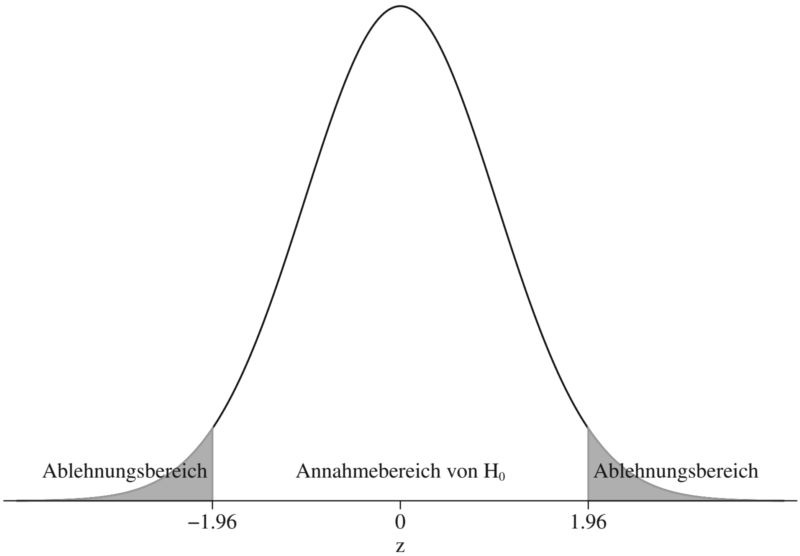


Abbildung 5.4: Annahme- und Ablehnungsbereich unter der Dichtefunktion der Standardnormalverteilung

Das Prinzip des ein- und zweiseitigen Testens soll anhand eines hoffentlich einsichtigen Beispiels gezeigt werden, welches sich auf die Standardnormalverteilung gründet (siehe Abschnitt 4.3.1). Die Dichtefunktion dieser Verteilung ist in Abbildung 5.4 wiedergegeben.

Die Gesamtfläche unter der Kurve ist 1 und die Fläche $\Phi(z)$ von $-\infty$ bis 1,96 beträgt nach der $z$-Tabelle (Tabelle 1) 0,975. Das bedeutet wegen der Symmetrie der Dichtefunktion, dass der schraffierte Teil der Fläche unter der Kurve 0,05 oder 5 % der Gesamtfläche beträgt. Fällt ein $z$-Wert in diesen Bereich, d. h., ist ein $z$-Wert dem Betrag nach größer als 1,96, so gehört er zu den randständigen 5 % der Werte.

Die schraffierte Fläche unter der Standardnormalverteilungskurve nennt man daher auch den Ablehnungsbereich, genauer gesagt, den Ablehnungsbereich auf der 5 %-Stufe.

Eine Firma möge Hanfseile herstellen, die sich durch ihre Reißlast unterscheiden. Dabei sei in den einzelnen Sorten diese Reißlast normalverteilt mit unterschiedlichen Mittelwerten und Standardabweichungen.

Eine Kunde bestellt ein Seil, dessen Reißqualität in kg mit $\mu = 3600$ und $\sigma = 80$ beschrieben ist. Die Tausorten liegen in verschiedenen Kisten, die versehentlich nicht beschriftet sind. Der Mitarbeiter, der die Bestellung bearbeitet, greift in eine der Kisten und zieht ein Seil mit einer Reißlast von 3690 kg heraus.

Er führt die folgende $z$-Transformation durch:

$$z = \frac{3690 - 3600}{80} = 1{,}13$$

Mit Hilfe der $z$-Tabelle ermittelt er hierzu folgenden Ablehnungsbereich:

$$2 \cdot (1 - 0{,}871) = 0{,}258$$

Aufgrund dieses Wertes (> 0,05) behält er die Hypothese, die richtige Kiste gefunden zu haben, bei.

Im vorliegenden Fall stehen die beiden folgenden Hypothesen zur Disposition:

| | |
|---|---|
| H0: | Die gewählte Kiste ist die richtige ($\mu = 3600$). |
| H1: | Die gewählte Kiste ist die falsche ($\mu \neq 3600$). |

Dabei wird in der Alternativhypothese über die Art des Missgriffs nichts ausgesagt: Die gewählte Kiste kann sowohl Taue mit kleineren als auch mit größeren Reißlasten enthalten.

In solchen Fällen, wo von vornherein keine Angaben über die Richtung der Alternativhypothese zu machen sind, muss man *zweiseitig* testen. Der Ablehnungsbereich liegt dabei zu gleichen Teilen an beiden Enden der Standardnormalverteilungskurve.

Es werde nun angenommen, die Firma stelle nur zwei Arten von Hanfseilen her, und zwar neben der schon erwähnten ($\mu = 3600, \sigma = 80$) eine solche mit $\mu = 3800$ und gleicher Standardabweichung. In diesem Fall kann man die Alternativhypothese H1 mit $\mu > 3600$ oder noch präziser mit $\mu = 3800$ angeben. Entsprechend liegt der Ablehnungsbereich auf nur einer Seite der Standardnormalverteilungskurve, und zwar auf der rechten. Man spricht in diesem Fall von *einseitiger* Fragestellung.

Der kritische $z$-Wert für den 5 %-Ablehnungsbereich liegt in diesem Fall nicht bei 1,96 wie beim zweiseitigen Test, sondern gemäß $z$-Tabelle bei 1,64.

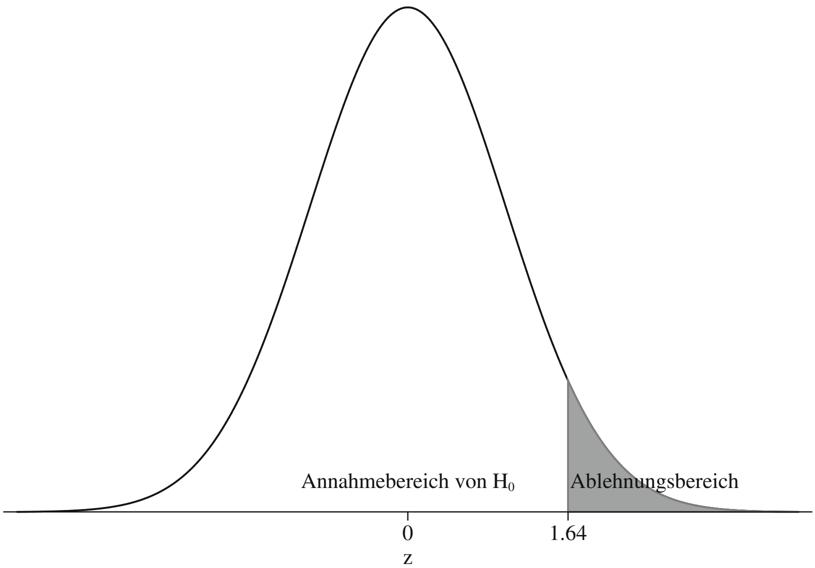


Abbildung 5.5: Annahme- und Ablehnungsbereich bei einseitiger Fragestellung

Dem im gegebenen Beispiel berechneten $z$-Wert von 1,13 entspricht ein Ablehnungsbereich von

$$1 - 0{,}871 = 0{,}129$$

Sie können diesen Wert in der $z$-Tabelle auch direkt der Spalte $\Phi(-z)$ entnehmen. Das ist die Fläche am linken Ende der Standardnormalverteilungskurve, die aus Symmetriegründen gleich der gesuchten Fläche am rechten Ende ist.

Der Ablehnungsbereich ist gleichbedeutend mit der Irrtumswahrscheinlichkeit (siehe Abschnitt 5.2). Die sich beim einseitigen Test ergebende Irrtumswahrscheinlichkeit ist also kleiner als die beim zweiseitigen Test (nämlich halb so groß). Das bedeutet, dass beim einseitigen Test die Nullhypothese eher abgelehnt wird als beim zweiseitigen Test.

Ist die Richtung der Alternativhypothese vorgegeben, steht also zum Beispiel bei einem Mittelwertvergleich von vornherein fest, welche Gruppe höhere Werte aufweisen wird, kann man *einseitig* testen. Diese Zusatzinformation erlaubt es eher, signifikante Unterschiede aufzudecken.

Die im letzten Rechenbeispiel sich ergebende Irrtumswahrscheinlichkeit $p = 0{,}129$ wurde in Abschnitt 5.4 als Fehler erster Art oder $\alpha$-Fehler bezeichnet. Er entspricht im vorliegenden Beispiel der dunkel schraffierten Fläche in Abbildung 5.6.

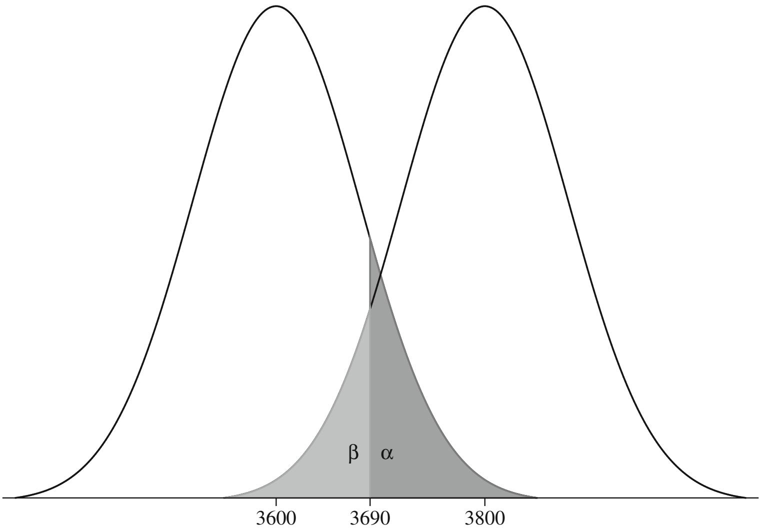


Abbildung 5.6: Fehler erster und zweiter Art bei einseitiger Fragestellung

$\alpha$ ist die Wahrscheinlichkeit dafür, die Nullhypothese fälschlicherweise abzulehnen. Die waagerecht schraffierte Fläche $\beta$ ist dementsprechend die Wahrscheinlichkeit dafür, die Alternativhypothese fälschlicherweise abzulehnen, also die Nullhypothese fälschlicherweise beizubehalten. Das wurde in Abschnitt 5.3 als Fehler zweiter Art bezeichnet.

Der Fehler zweiter Art lässt sich nur bei genauer Kenntnis der Alternativhypothese berechnen. Im vorliegenden Beispiel berechnet man bei Kenntnis des alternativen Mittelwerts $\mu = 3800$ folgenden $z$-Wert:

$$z = \frac{3690 - 3800}{80} = -1{,}38$$

Aus der $z$-Tabelle entnimmt man hierzu

$$\beta = \Phi(-1{,}38) = 0{,}084$$

Zusammenfassend lässt sich sagen, dass im vorliegenden Fall das Risiko, die Nullhypothese abzulehnen, obwohl sie richtig ist, in Prozenten ausgedrückt 12,9 % beträgt. Das Risiko hingegen, sie beizubehalten, obwohl sie falsch ist, beträgt 8,4 %. Das erstgenannte Risiko geht offenbar zu Lasten des Hanfseilproduzenten, der bei Ablehnung der Nullhypothese noch einmal in eine Kiste greifen muss; das zweitgenannte geht zu Lasten des Kunden, der bei fälschlicherweise Beibehaltung der Nullhypothese mit einem Hanfseil der falschen Sorte beliefert würde. Die entsprechenden Begriffe Produzenten- und Konsumentenrisiko wurden bereits in Abschnitt 5.3 eingeführt.

Die Tabelle 5.4 verdeutlicht, in welcher Weise im vorliegenden Beispiel ein sich verändernder Fehler erster Art einen Fehler zweiter Art nach sich zieht.

| Fehler erster Art | Fehler zweiter Art |
|---|---|
| 50,0 % | 0,6 % |
| 40,0 % | 1,2 % |
| 30,0 % | 2,4 % |
| 20,0 % | 4,9 % |
| 10,0 % | 11,2 % |
| 5,0 % | 19,6 % |
| 1,0 % | 43,1 % |
| 0,1 % | 72,3 % |

**Tabelle 5.4:** Fehler erster und zweiter Art

Ob man lieber einen Fehler erster oder zweiter Art in Kauf nehmen möchte, ist von Fall zu Fall zu entscheiden und hängt von der jeweiligen Testsituation ab.

Ist es folgenschwer, die Nullhypothese fälschlicherweise abzulehnen, wie dies in der Regel bei den in Kapitel 9 vorgestellten Signifikanztests der Fall ist, wird man den Fehler erster Art klein halten wollen. Ist es hingegen folgenschwer, die Nullhypothese fälschlicherweise beizubehalten, wie etwa im vorliegenden Beispiel, wird man bestrebt sein, den Fehler zweiter Art, das Konsumentenrisiko, zu drücken.

## 5.5 Die Gefahr der Alpha-Inflation

Das Gute an der früheren computerlosen Zeit war für uns Statistiker, dass jeder Test, den wir durchführten, vorher gut durchdacht sein wollte. Zu groß war nämlich die Rechenarbeit selbst bei einfachen Tests, als dass es jemandem in den Sinn gekommen wäre, einfach „nur mal so“ drauflos zu testen, nach dem Motto „Irgendwo wird schon was Signifikantes sein“.

Man hatte eine bestimmte Fragestellung, die es zu untersuchen galt, notierte Nullhypothese und Alternativhypothese und rechnete den passenden Test. Von der Rechenarbeit erschöpft, hielt man inne und ging dann gegebenenfalls daran, in Ruhe die nächste Fragestellung abzuklären.

Heute, im Zeitalter immer schnellerer Computer und ausgefeilterer Statistikprogramme, ist nach erfolgter Dateneingabe die Rechenarbeit meist eine Sache von Sekundenbruchteilen. Haben Sie dann etwa hundert Variablen und dazu eine Gruppenvariable wie das Geschlecht, verschiedene Altersklassen oder Ähnliches, dann verführt das schnell dazu, einfach mal alle Variablen auf Gruppenunterschiede durchzutesten. Oder Sie haben fünfzig nominal- und ordinalskalierte Variablen, die