Statistik im Klartext

Für Psychologen, Wirtschafts- und Sozialwissenschaftler

3., aktualisierte und erweiterte Auflage

Fabian Heimsch
Rudolf Niederer

Pearson

EXTRAS ONLINE

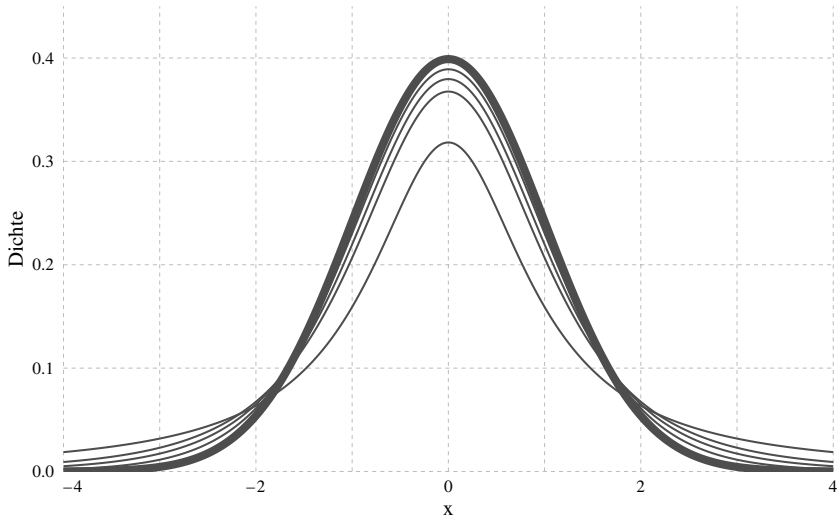

Abbildung 5.3: Standardnormalverteilung (fett) und *t*-Verteilungen mit $df = 1, 3, 5$ und 10 von unten nach oben

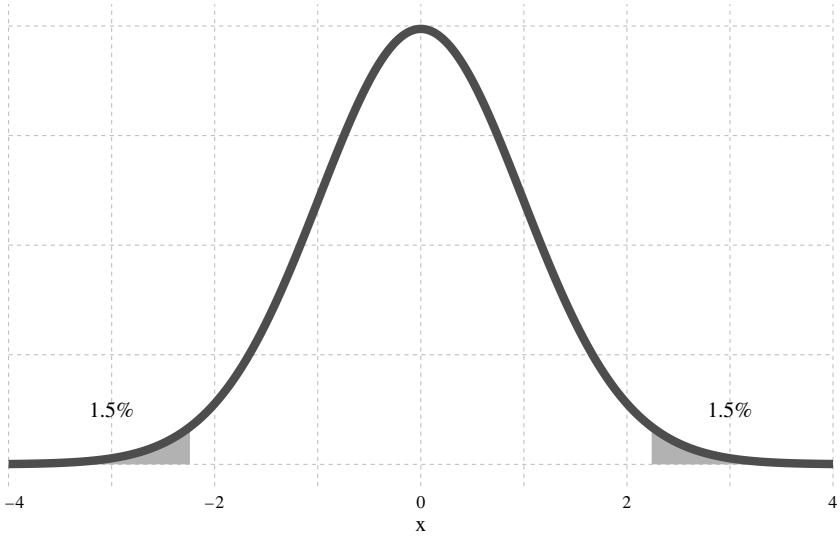

Abbildung 5.4: Zur Illustration des *p*-Werts unter der *t*-Verteilung

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*.

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 63 Freiheitsgraden den kritischen Tabellenwert 1,998. 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 α-Fehler. Die Wahrscheinlichkeit, einen Fehler erster Art zu begehen, ist gleich der Irrtumswahrscheinlichkeit p. Der zweitgenannte Fehler heißt Fehler zweiter Art oder β-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 β-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.

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

	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

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 β 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 α 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.

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 5.5 wiedergegeben.

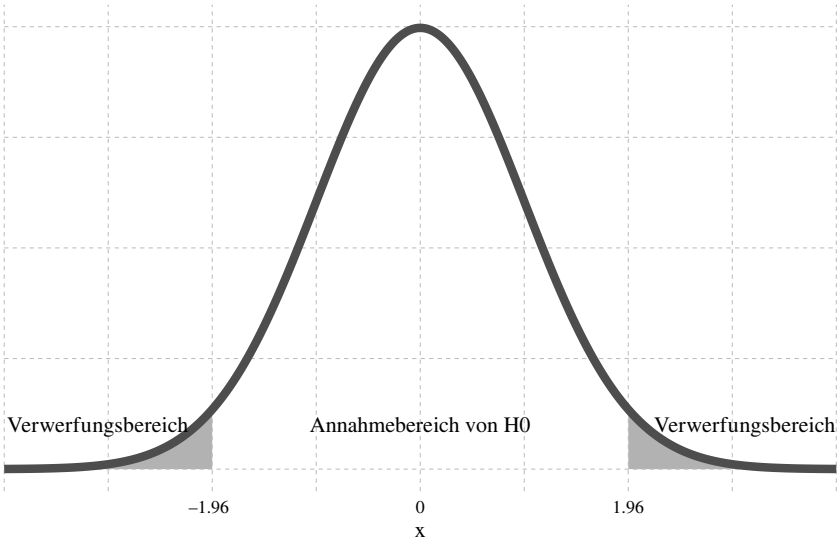

Abbildung 5.5: Annahme- und Ablehnungsbereich unter der Dichtefunktion der Standardnormalverteilung

Die Gesamtfläche unter der Kurve ist 1 und die Fläche $\Phi(z)$ von $-\infty$ bis 1,96 beträgt nach der z-Tabelle 0,975 (Tabelle 1 im Anhang). 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.

Ein Kunde bestellt ein Seil, dessen Reißqualität in kg mit $\mu = 3\,600$ 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 3 690 kg heraus.

Er führt die folgende z-Transformation durch:

$$z = \frac{3\,690 - 3\,600}{80} = 1{,}13$$

Mithilfe der z-Tabelle ermittelt er hierzu folgenden Ablehnungsbereich:

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

Aufgrund dieses Werts ($> 0{,}05$) 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 ($\mu = 3\,600$).
H1:	Die gewählte Kiste ist die falsche ($\mu \neq 3\,600$).

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 ($\mu = 3\,600, \sigma = 80$) eine solche mit $\mu = 3\,800$ und gleicher Standardabweichung. In diesem Fall kann man die Alternativhypothese H1 mit $\mu > 3\,600$ oder noch präziser mit $\mu = 3\,800$ 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.

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.

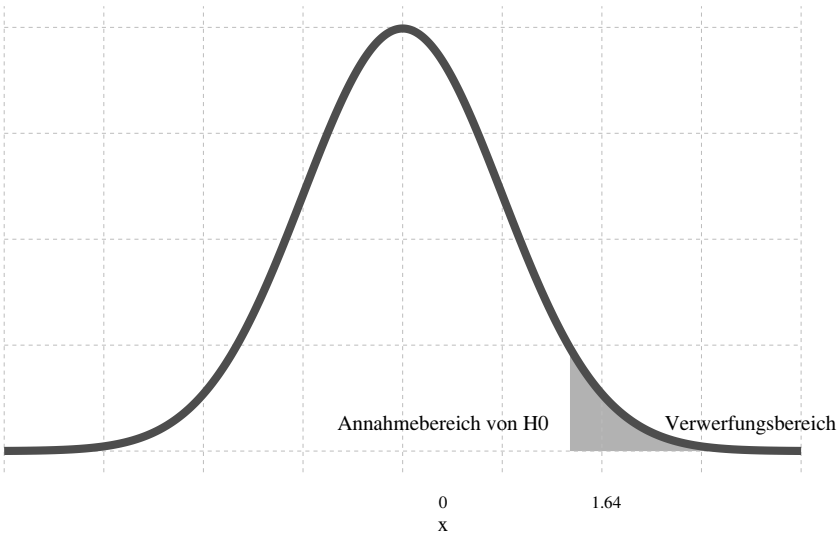

Abbildung 5.6: Annahme- und Ablehnungsbereich bei einseitiger Fragestellung

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 α-Fehler bezeichnet. Er entspricht im vorliegenden Beispiel der dunkel schraffierten Fläche in Abbildung 5.7.

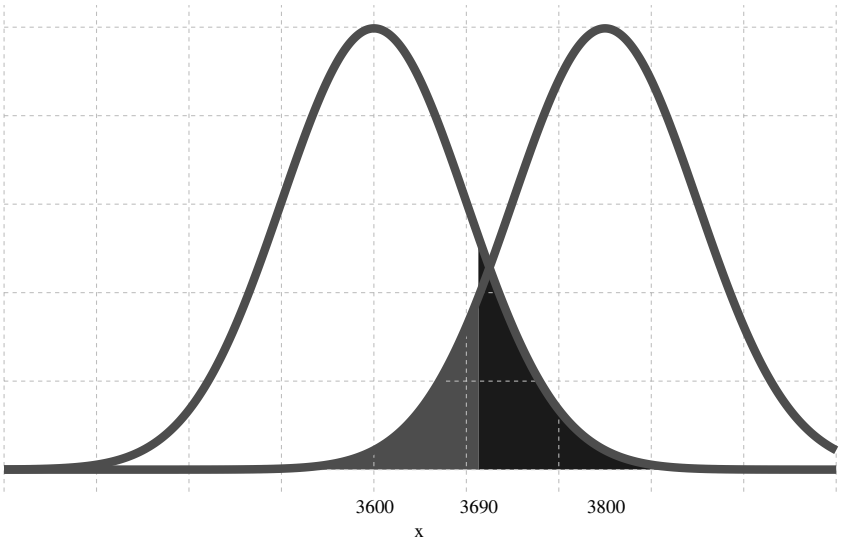

Abbildung 5.7: Fehler erster und zweiter Art bei einseitiger Fragestellung

α ist die Wahrscheinlichkeit dafür, die Nullhypothese fälschlicherweise abzulehnen. Die hellgrau schraffierte Fläche β 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 $\mu = 3\,800$ folgenden z-Wert:

$$z = \frac{3\,690 - 3\,800}{80} = -1{,}38$$

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 Sie alle untereinander mit einer Kreuztabelle und anschließendem Chi-Quadrat-Test in Beziehung setzen wollen, um signifikante Zusammenhänge aufzuspüren.

Haben Sie also fünfzig Variablen, von denen Sie jede mit jeder kreuzen wollen, so ergibt das, wenn Sie redundante Beziehungen auslassen,

$$\frac{50 \cdot 49}{2} = 1\,225$$

Vergleiche. Führen Sie jeweils den Chi-Quadrat-Test (Kapitel 12) aus und geben Sie die übliche Signifikanzschranke $p \leq 0{,}05$ vor, dann bedeutet das, dass von vornherein 5 % der Vergleiche ein signifikantes Ergebnis liefern werden. Bei 1 225 durchgeführten Vergleichen wären dies 61 Vergleiche, die von vornherein mit signifikantem Ergebnis zu erwarten sind.

Haben Sie nun zum Beispiel insgesamt 92 Signifikanzen aufgedeckt, so ist bei jeder dieser Signifikanzen die Gefahr, einen Fehler erster Art (α-Fehler) zu begehen, sehr groß. Ihre Ergebnisse sind daher wertlos.

Wir wollen diese Problematik noch einmal anhand einer Computersimulation verdeutlichen. Es erfolgte eine Simulation von bis zu 50 000 Stichproben mit jeweils 100 normalverteilten Werten, die dann per Zufall in zwei gleich große Gruppen eingeteilt wurden. Diese wurden bezüglich ihrer Mittelwerte mit dem *t*-Test nach Student miteinander verglichen. Die Ergebnisse sind Tabelle 5.5 zu entnehmen.

Tests	$p \leq 0{,}05$		$p \leq 0{,}01$		$p \leq 0{,}001$	
	n	%	*n*	%	*n*	%
100	3	3,00	2	2,00	0	0,00
200	9	4,50	3	1,50	0	0,00
500	26	5,20	9	1,80	1	0,20
1 000	56	5,60	12	1,20	2	0,20
2 000	111	5,55	19	0,95	2	0,10
5 000	254	5,08	45	0,90	7	0,14
10 000	505	5,05	86	0,86	11	0,11
20 000	988	4,94	172	0,86	28	0,14
50 000	2 428	4,86	449	0,90	69	0,14

Tabelle 5.5: Anzahlen signifikanter Testergebnisse

Die Tabelle enthält die absoluten und prozentualen Häufigkeiten der auf dem betreffenden Niveau signifikanten Ergebnisse, wobei diese Häufigkeiten gut unseren Erwartungen entsprechen.

Um einen Ausweg aus dieser Problematik zu finden, gibt es mehrere Vorschläge. Der beste Vorschlag ist sicher der, diesen Unfug einfach zu lassen. Formulieren Sie nur einzelne sachlogisch fundierte Hypothesen, denen Sie dann mit passenden Tests nachgehen.

Allerdings liegt es in der Natur des Menschen, aus wissenschaftlicher Neugierde Zusammenhänge aufspüren zu wollen, an die vorher niemand gedacht hat. So ist es meist doch zu verlockend, eben mal in Sekunden einige hundert oder gar tausend Tests zu rechnen. Eine Möglichkeit besteht dann darin, das Signifikanzniveau schärfer zu fassen und zum Beispiel bei $\alpha = 0{,}001$ festzulegen. Bei tausend Tests ist schließlich von vornherein nur ein solch höchst signifikantes Ergebnis zu erwarten, was sicherlich zu vernachlässigen ist, wenn Sie viele solcher Resultate erhalten haben.

Einen ähnlichen Ausweg bietet die *Bonferroni-Korrektur*. Wollten Sie ursprünglich mit der Signifikanzschranke $p = 0{,}05$ testen, so sollte bei dieser Korrektur und insgesamt n Signifikanztests diese Schranke auf $0{,}05/n$ herabgesetzt werden. Für eine große Zahl von Tests ist dies aber kein praktikabler Weg.

Elegante Lösungen gibt es für den Fall, dass eine größere Anzahl von Vergleichen dadurch zustande kommt, dass durch eine Gruppierungsvariable verschiedene Gruppen entstehen, die dann paarweise miteinander verglichen werden sollen. In diesem Fall ist eine einfaktorielle Varianzanalyse (siehe Kapitel 13) oder der *H*-Test nach Kruskal und Wallis (siehe Kapitel 10.3) vorzuschalten.

5.6 Prüfverteilungen

In Abschnitt 5.2 wurde die *t*-Verteilung nach Student (W. S. Gosset) erläutert. Weitere bedeutsame Verteilungen sind die Standardnormalverteilung nach Gauß (siehe auch Abschnitt 4.3.1), die *F*-Verteilung nach Fisher und die χ^2-Verteilung nach Pearson. Diese Verteilungen sollen nun im Einzelnen vorgestellt werden.

Standardnormalverteilung

Eine normalverteilte Prüfgröße, stets *z* genannt, wird zum Beispiel berechnet beim *U*-Test nach Mann und Whitney, beim Wilcoxon-Test (Kapitel 10) und bei der Absicherung des Rangkorrelationskoeffizienten nach Kendall (Kapitel 11).

Aus der Prüfgröße *z* berechnet sich die Irrtumswahrscheinlichkeit *p* nach der folgenden Formel:

$$p = \frac{2}{\sqrt{2 \cdot \pi}} \cdot \int_{z}^{\infty} \mathrm{e}^{-\frac{v^2}{2}} \mathrm{d}v$$

Für die *z*-Werte von 0 bis 3,49 sind in Schritten von 0,01 die den *z*-Werten zugeordneten *p*-Werte in der *z*-Tabelle aufgeführt.

t-Verteilung

Eine *t*-verteilte Prüfgröße wird berechnet beim *t*-Test nach Student, beim *t*-Test für abhängige Stichproben sowie bei der Absicherung des Produkt-Moment-Korrelationskoeffizienten, der Rangkorrelation nach Spearman, der partiellen Korrelation und der Regressionskoeffizienten (Kapitel 11).

Die *t*-Verteilung und die formelmäßige Berechnung der Irrtumswahrscheinlichkeit aus der Prüfgröße *t* und der Anzahl der Freiheitsgrade wurden bereits in Abschnitt 5.2 erläutert.

Für die drei klassischen Signifikanzniveaus und verschiedene Anzahlen von Freiheitsgraden sind kritische Tabellenwerte in der *t*-Tabelle aufgeführt. Signifikanz auf dem betreffenden Niveau liegt vor, wenn die berechnete Prüfgröße *t* den betreffenden kritischen Tabellenwert übersteigt.

χ^2-Verteilung

χ^2-verteilte Prüfgrößen werden berechnet bei den verschiedenen Chi-Quadrat-Tests, dem *H*-Test nach Kruskal und Wallis, dem Friedman-Test (Kapitel 10), dem Bartlett-Test (Kapitel 13) und Cochran's *Q*.

Die Kurven zur χ^2-Verteilung sind linksgipflig. Die Irrtumswahrscheinlichkeit berechnet sich aus der Prüfgröße χ^2 und der Anzahl df der Freiheitsgrade nach der folgenden Formel:

$$p = \frac{1}{2^{\frac{df}{2}} \cdot \Gamma(\frac{df}{2})} \cdot \int_{\chi^2}^{\infty} v^{\frac{df-2}{2}} \cdot \mathrm{e}^{-\frac{v}{2}} \mathrm{d}v$$

Für die drei klassischen Signifikanzniveaus und für verschiedene Anzahlen von Freiheitsgraden sind kritische Tabellenwerte in der χ^2-Tabelle aufgeführt. Signifikanz liegt vor, wenn die berechnete Prüfgröße χ^2 den betreffenden kritischen Tabellenwert übersteigt.

Abbildung 5.8 zeigt den Verlauf der χ^2-Verteilung für unterschiedliche Freiheitsgrade (df).

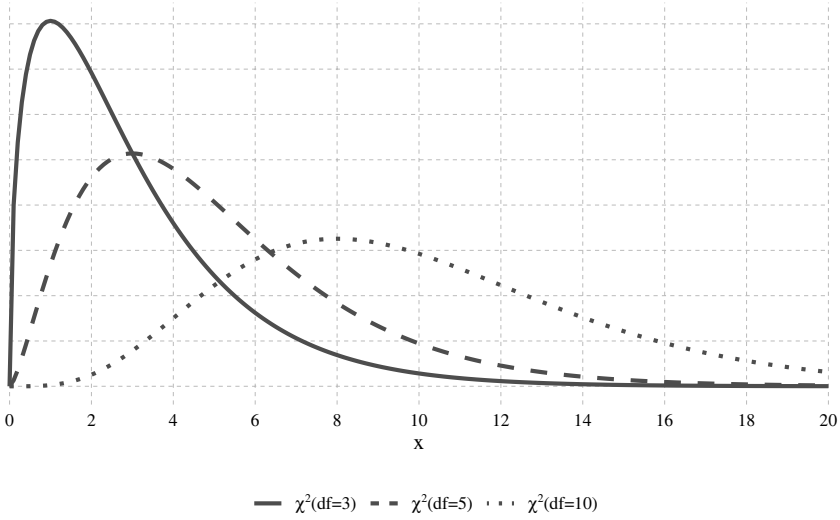

Abbildung 5.8: χ^2-Verteilungen für unterschiedliche Freiheitsgrade

F-Verteilung

F-verteilte Prüfgrößen werden berechnet bei den verschiedenen Formen der Varianzanalyse (Kapitel 13), dem Scheffé-Test, dem *F*-Test, dem Levene-Test und dem Hartley-Test.

Die Kurven zur *F*-Verteilung sind linksgipflig und von zwei Freiheitsgraden df_1 und df_2 abhängig. Die Irrtumswahrscheinlichkeit berechnet sich aus der Prüfgröße F und den Anzahlen der Freiheitsgrade df_1 und df_2 nach der folgenden Formel:

$$p = \frac{\Gamma(\frac{df_1+df_2}{2})}{\Gamma(\frac{df_1}{2}) \cdot \Gamma(\frac{df_2}{2})} \cdot df_1^{\frac{df_1}{2}} \cdot df_2^{\frac{df_2}{2}} \cdot \int_{F}^{\infty} v^{\frac{df_1-2}{2}} \cdot (df_1 \cdot v + df_2)^{-\frac{df_1+df_2}{2}} \mathrm{d}v$$

Für die drei klassischen Signifikanzniveaus und verschiedene Anzahlen von Freiheitsgraden sind kritische Tabellenwerte in der *F*-Tabelle aufgeführt. Signifikanz liegt

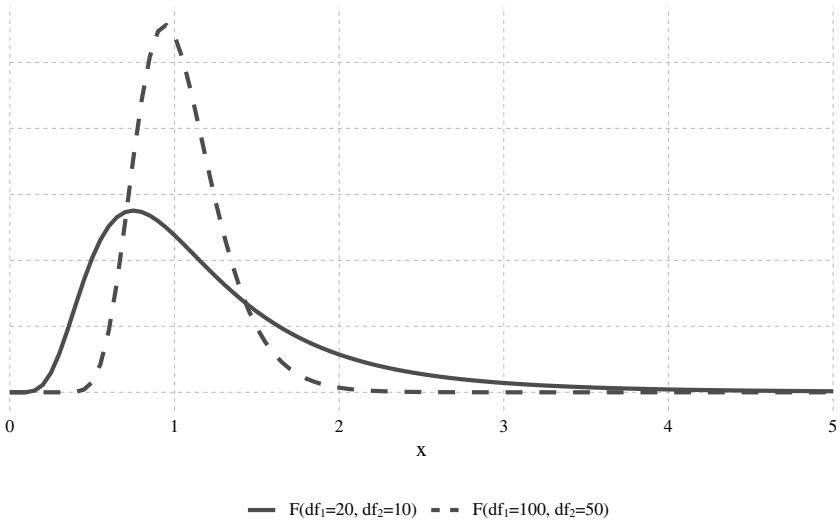

Abbildung 5.9: *F*-Verteilungen für unterschiedliche Freiheitsgrade

vor, wenn die berechnete Prüfgröße F den betreffenden kritischen Tabellenwert übersteigt.

ZUSAMMENFASSUNG

- Die analytische Statistik befasst sich mit dem Schluss von der Stichprobe auf die Grundgesamtheit.
- Nullhypothese und Alternativhypothese sind zu formulieren und mit einem geeigneten statistischen Testverfahren zu überprüfen.
- Das Ergebnis eines statistischen Tests ist die Irrtumswahrscheinlichkeit p. Sie entscheidet darüber, ob ein Testergebnis signifikant ist. Unterschreitet die Irrtumswahrscheinlichkeit p das Signifikanzniveau α, so ist das Ergebnis signifikant. Die häufigst verwendeten Signifikanzniveaus sind $\alpha = 0{,}1$, $\alpha = 0{,}05$ und $\alpha = 0{,}01$
- Die wichtigsten Verteilungen der Prüfstatistik sind die Standardnormalverteilung, die t-Verteilung, die F-Verteilung und die χ^2-Verteilung.
- Wird die Nullhypothese verworfen, obwohl sie richtig ist, spricht man von einem Fehler erster Art; wird sie beibehalten, obwohl sie falsch ist, liegt ein Fehler zweiter Art vor.
- Ist die Richtung der Alternativhypothese vorgegeben, kann einseitig getestet werden.
- Beim kritiklosen Ausführen sehr vieler Tests besteht die Gefahr der Alpha-Inflation.

Copyright