Willkommen zur Vorlesung Statistik

Thema dieser Vorlesung:
Inferenzstatistik (schließende Statistik, induktive Statistik)

Prof. Dr. Wolfgang Ludwig-Mayerhofer

Universität Siegen – Philosophische Fakultät, Seminar für Sozialwissenschaften

Inhaltsübersicht

Einführung:
Die Fragestellung der Inferenzstatistik

Was ist Inferenzstatistik?

Die Inferenzstatistik beschäftigt sich mit dem Problem, dass aufgrund von Stichproben gewonnene Ergebnisse nicht unmittelbar auf die Grundgesamtheit (Population, „Statistische Masse“) übertragen werden können.

Die Inferenzstatistik soll hier kurz angerissen werden, damit wir immer im Auge behalten, dass Stichprobendaten mit „Unsicherheit“ behaftet sind. Diese Unsicherheit kann aber berechnet werden!

Bitte beachten Sie: Die Darstellung ist häufig vereinfachend in dem Sinne, dass sie sich auf die wichtigsten Anwendungsfälle und einfache Regeln beschränkt. Sie kann den Komplexitäten der Inferenzstatistik nicht gerecht werden.

Lernziele

Zu den Begriffen aus dem Bereich der Inferenzstatistik, von denen jede/r mal gehört haben muss (und die man verstehen muss, wenn man sich mit sozialwissenschaftlichen Forschungsergebnissen befasst), gehören:

- Verteilungen von Zufallsvariablen, vor allem die (Standard-)Normalverteilung; der Standardfehler

  → Abschnitte Verteilung von Zufallsvariablen sowie Zentraler Grenzwertsatz/Standardfehler
- Konfidenzintervall und Irrtumswahrscheinlichkeit

  → Abschnitt Punkt- und Intervallschätzung
- Signifikanz, Signifikanztest und Signifikanzniveau

  → Abschnitt Statistisches Testen

Vor allem die letzten beiden Punkte tauchen in Texten, die auf Auswertung standardisierter Daten basieren, häufig auf. Der erste Abschnitt dient dem Erwerb eines grundlegenden Verständnisses.

Voraussetzung: Zufallsstichproben

Alle Annahmen der Inferenzstatistik beruhen darauf, dass die vorliegenden Daten eine Zufallsstichprobe aus der Grundgesamtheit darstellen.

Die üblicherweise verwendeten Überlegungen basieren sogar darauf, dass es sich um eine einfache Zufallsstichprobe handelt (Erklärung des Begriffs im nächsten Absatz). In der Forschungspraxis trifft diese Annahme oft nicht zu (z. B. geschichtete Stichproben, Klumpenstichproben, mehrstufige Auswahlen). Aber die grundsätzlichen Überlegungen sind ähnlich, nur die Formeln werden komplizierter.

Eine einfache Zufallsstichprobe ist dadurch charakterisiert, dass die einzelnen Elemente aus der Grundgesamtheit jeweils unabhängig voneinander für die Stichprobe gezogen werden, und zwar mit gleicher (oder zumindest angebbarer) Wahrscheinlichkeit größer als Null. (Mehr zu Stichproben in der Vorlesung „Methoden der empirischen Sozialforschung“).

Zufallsvariablen und ihre Verteilung

Zufallsvariablen: Worum geht es?

Inferenzstatistik befasst sich mit dem Problem, wie wir von einer Stichprobe auf die (unbekannte) Grundgesamtheit schließen können.

Dazu gehen wir zunächst von der umgekehrten Fragestellung aus: Was geschieht, wenn wir viele Stichproben aus einer (zumindest theoretisch) bekannten Grundgesamtheit ziehen?

Von dieser „umgekehrten“ Fragestellung handelt dieser Abschnitt.

Zufallsvorgänge

Überlegungen zur Inferenzstatistik beruhen auf dem Modell des „Zufallsvorgangs“ (oder Zufallsexperiments). Typische Beispiele sind:

- Das Werfen einer Münze:

  Eine Münze landet beim Werfen mit Kopf oder Zahl nach oben. Welches Ereignis eintritt, ist vorher unbekannt. Wir erwarten, dass jedes Ereignis mit einer Wahrscheinlichkeit von 0,5 eintrifft.

- Würfeln:

  Es sind sechs Ereignisse möglich, jedes hat die Wahrscheinlichkeit $1/6 = 0{,}1\overline{666}$.

Die Ergebnisse von Zufallsvorgängen heißen Zufallsvariablen. Im Beispiel: Häufigkeit, mit der die Münze auf „Kopf“ landet; durchschnittliche Augenzahl beim Würfeln.

Stichprobenziehung als Zufallsvorgang

Das Werfen einer Münze oder eines Würfels hat Analogien zur Stichprobenziehung.

- Binäre Variablen: In einer Grundgesamtheit haben die Elemente bspw. die Eigenschaft „allein erziehend" oder „nicht allein erziehend" (Analogie: Kopf oder Zahl beim Münzwerfen; allerdings nicht Anteilswert von je 0,5). Wenn wir zufällig viele Elemente aus der Grundgesamtheit ziehen, ist die Anzahl (oder der Anteil) der allein Erziehenden eine Zufallsvariable.
- Metrische Variablen: In einer Grundgesamtheit haben die Elemente beispielsweise eine Zahl von Geschwistern, die zwischen 0 und (meist maximal) 15 liegt (Analogie: Würfeln; im Unterschied zum Würfeln ist die Wahrscheinlichkeit der verschiedenen möglichen Geschwisterzahlen nicht gleich). Bei zufälliger Stichprobenziehung ist die Summe (oder der Durchschnitt) der Geschwisterzahl eine Zufallsvariable.

Beispiel Münzwurf: Mögliche Ergebnisse

Ein Wurf:

{K}

{Z}

Zwei Würfe:

{K, K}
{K, Z} {Z, K}

{Z, Z}

Vier Würfe:

{K, K, K, K}
{K, K, K, Z} {K, K, Z, K} {K, Z, K, K} {Z, K, K, K}
{K, K, Z, Z} {K, Z, K, Z} {Z, K, K, Z} {K, Z, Z, K} {Z, K, Z, K} {Z, Z, K, K}
{Z, Z, Z, K} {Z, Z, K, Z} {Z, K, Z, Z} {K, Z, Z, Z}
{Z, Z, Z, Z}

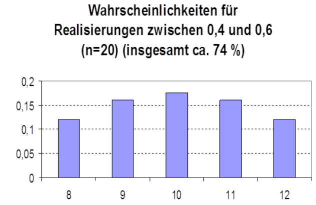
Münzwurf – ein Versuch

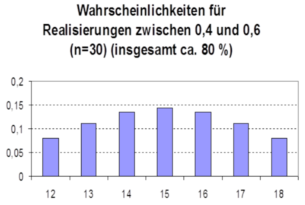
Das Werfen einer Münze oder eines Würfels hat Analogien zur Stichprobenziehung.

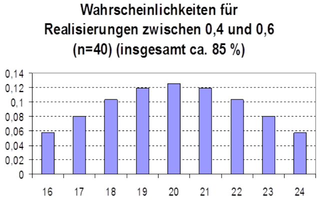
Nehmen Sie eine beliebige Münze und werfen Sie sie zehn Mal.
Bitte notieren Sie nach jedem Wurf, ob die Münze Zahl oder Kopf (Wappen) zeigt.

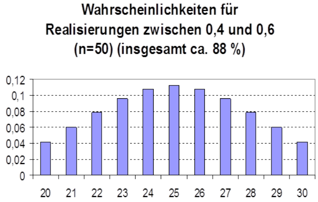
Beispiel: Zunehmende Zahl von Münzwürfen

Mit zunehmendem n liegen größere Anteile der Realisierungen nahe am wahren Wert (hier: zwischen 0,4 und 0,6) → „Gesetz der großen Zahl“.









Die Gaußsche Normalverteilung

Die Normalverteilung ist eine stetige Zufallsvariable.

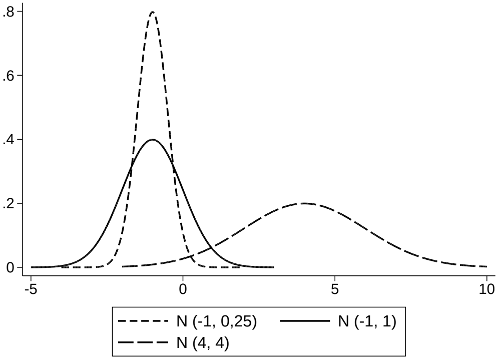
Ihre Dichteverteilung kann durch folgende (nicht auswendig zu lernende) Formel beschrieben werden:

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

Dabei ist μ der Erwartungswert („Mittelwert") der Verteilung, σ^2 ihre Varianz und π die Kreiszahl 3,14. . . . Eine normalverteilte Variable ist vollständig durch ihren Erwartungswert und ihre Varianz gekennzeichnet; man bezeichnet eine entsprechende Normalverteilung oft kurz mit N(μ, σ^2).

Die Gaußsche Normalverteilung

Es gibt viele Normalverteilungen mit unterschiedlichen Mittelwerten μ und Varianzen σ^2. Für alle gilt: Sie sind symmetrisch um μ, unimodal, mehr oder weniger glockenförmig. Die Wahrscheinlichkeitsdichte strebt asymptotisch gegen 0, wenn x gegen $-\infty$ bzw. $+\infty$ strebt.



Die Bedeutung der Normalverteilung

Viele (aber beileibe nicht alle . . .) empirische Merkmale sind normalverteilt, teilweise werden sie als normalverteilte Merkmale konstruiert (z. B. Intelligenztests oder andere psychometrische Tests) (siehe Vorlesung „Streuungsmaße").

Auch Stichprobenkennwerte sind unter bestimmten Bedingungen normalverteilt. Damit ist gemeint: Wenn wir den Vorgang der Stichprobenziehung immer wieder durchführen, erhalten wir zwar vermutlich jedesmal einen anderen Stichprobenkennwert (Mittelwert, Anteilswert . . .), die Verteilung dieser Kennwerte folgt jedoch einer Normalverteilung.

Diese Einsicht kann man für Schlüsse von der Stichprobe auf die Grundgesamtheit ausnützen. Doch zunächst befassen wir uns ausführlicher mit der Normalverteilung und der Verteilung von Stichprobenkennwerten.

Die Standardnormalverteilung

Durch Standardisierung wird die Normalverteilung in die Standardnormalverteilung N (0;1) überführt.

Unter Standardisierung (auch z-Transformation) versteht man folgende Transformation:

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

Es wird von jedem einzelnen Wert von X der Mittelwert von X (hier geschrieben als μ) abgezogen und das Resultat durch σ dividiert. Eine standardisierte Variable hat einen Mittelwert von 0 und eine Standardabweichung von 1.

Die Dichtefunktion der Normalverteilung vereinfacht sich so zu dem (nicht zu lernenden) Ausdruck

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

Normalverteilung und Standardnormalverteilung

Für die Normalverteilung gilt:

- Ca. 68 % der Werte liegen in einem Bereich von $\pm 1\,\sigma$ symmetrisch um den Mittelwert.
- Gut 95 % der Werte liegen in einem Bereich von $\pm 2\,\sigma$ symmetrisch um den Mittelwert.
- Ca. 99,7 % der Werte liegen in einem Bereich von $\pm 3\,\sigma$ symmetrisch um den Mittelwert.

Für die Standardnormalverteilung gilt entsprechend:

- Ca. 68 % der Werte liegen in einem Bereich von ± 1 symmetrisch um den Mittelwert.
- Gut 95 % der Werte liegen in einem Bereich von ± 2 symmetrisch um den Mittelwert.
- Ca. 99,7 % der Werte liegen in einem Bereich von ± 3 symmetrisch um den Mittelwert.

Einige Quantile der Standardnormalverteilung

Eine stetige Verteilung kann u. a. durch die Angabe von Quantilen charakterisiert werden.

Die folgende Tabelle zeigt beispielsweise, dass der Wert des 0,01-Quantils (erstes Perzentil) $-2{,}326$ beträgt. Ein Prozent der Werte einer standardnormalverteilten Variablen ist also kleiner oder gleich $-2{,}326$ (genauer gesagt: liegt im Bereich von $-\infty$ bis $-2{,}326$), 99 Prozent sind größer oder gleich $-2{,}326$.

z	Quantil	z	Quantil
-3,000	0,0013	1,000	0,841
-2,326	0,01	1,282	0,90
-1,96	0,025	1,645	0,95
-1,645	0,05	1,96	0,975
-1,282	0,10	2,326	0,99
-1,000	0,159	3,000	0,9986
0	0,5		

Mehr zum „Lesen“ der Standardnormalverteilung

90 Prozent aller Werte, die symmetrisch links und rechts von 0 liegen, liegen zwischen $-1{,}645$ und $+1{,}645$.

95 Prozent aller Werte, die symmetrisch um 0 liegen, liegen zwischen $-1{,}96$ und $+1{,}96$.

Umgekehrt: Die 10 Prozent der extremsten Werte (am weitesten vom Mittelwert entfernt) verteilen sich auf die 5 Prozent der niedrigsten und 5 Prozent der höchsten Werte; sie liegen also in den Bereichen $-\infty$ bis $-1{,}645$ sowie $+1{,}645$ bis $+\infty$.

Die 5 Prozent der extremsten Werte verteilen sich auf die 2,5 Prozent der niedrigsten und 2,5 Prozent der höchsten Werte; sie liegen also in den Bereichen $-\infty$ bis $-1{,}96$ sowie $+1{,}96$ bis $+\infty$.

Häufigkeitsaussagen als Wahrscheinlichkeitsaussagen

Wenn es um Zufallsvorgänge geht, lassen sich Aussagen darüber, wie häufig etwas geschieht, auch als Wahrscheinlichkeitsaussagen formulieren. Wir können also die Aussagen der vorherigen Folie wie folgt reformulieren:

- Mit einer Wahrscheinlichkeit von 90 Prozent liegen die Werte symmetrisch links und rechts von 0 im Bereich zwischen $-1{,}645$ und $+1{,}645$.
- Mit einer Wahrscheinlichkeit von ca. 95 Prozent liegen die Werte symmetrisch links und rechts von 0 im Bereich zwischen $-1{,}96$ und $+1{,}96$.
- Umgekehrt: Mit einer Wahrscheinlichkeit von 5 Prozent liegt ein Wert zwischen $-\infty$ und $-1{,}645$; ebenso liegt ein Wert mit einer Wahrscheinlichkeit von 5 Prozent zwischen $+1{,}645$ und $+\infty$.
- Mit einer Wahrscheinlichkeit von je 2,5 Prozent liegt ein Wert entweder zwischen $-\infty$ und $-1{,}96$ oder zwischen $+1{,}96$ und $+\infty$.

Einer der wichtigsten Sätze der Inferenzstatistik:
Der zentrale Grenzwertsatz

Der zentrale Grenzwertsatz

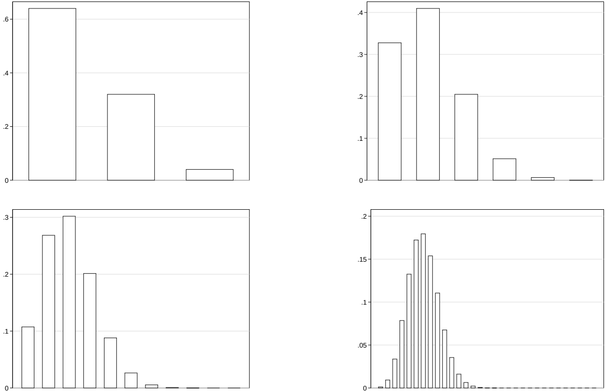
Die Verteilung der Summe von n unabhängigen standardisierten Zufallsvariablen, die alle die identische Wahrscheinlichkeitsverteilung haben, nähert sich mit steigender Stichprobengröße der Standardnormalverteilung an.

Daraus folgt u. a., dass Mittelwerte und Anteilswerte von Zufallsstichproben bei „hinreichend großem" n einer Normalverteilung folgen. (Was „hinreichend groß" heißt, kann unterschiedlich sein, mindestens gilt $n \geq 30$).

Achtung: Die (Standard-) Normalverteilung gilt ganz exakt nur bei bekannter Grundgesamtheit; bei unbekannter Grundgesamtheit wird sie durch andere Verteilungen angenähert, die allerdings bei großen Stichproben sich asymptotisch der Standardnormalverteilung annähern.

Der zentrale Grenzwertsatz illustriert

Eine diskrete und recht schief verteilte Variable (Binomialverteilung mit $\pi = 0{,}2$) wird mit zunehmendem n einer Normalverteilung ähnlicher (n = 2, 5, 10 und 30):



Folgen aus dem zentralem Grenzwertsatz

Der zentrale Grenzwertsatz besagt, dass wir die Verteilung von Stichprobenkennwerten anhand der Standardnormalverteilung beschreiben können.

Das heißt z. B.: Wenn wir viele (große) Stichproben ziehen und daraus jeweils den Mittelwert einer Variablen berechnen, so liegen beispielsweise 95 Prozent aller standardisierten Mittelwerte zwischen $-1{,}96$ und $+1{,}96$. Ähnliches gilt auch für (standardisierte) Anteilswerte.

Für die *nicht* standardisierten Variablen gilt: Wenn wir viele (große) Stichproben ziehen und daraus jeweils den Mittelwert einer Variablen berechnen, so liegen beispielsweise 95 Prozent aller Mittelwerte zwischen $-1{,}96$ und $+1{,}96$ Standardabweichungen um den wahren Mittelwert. Ähnliches gilt auch für Anteilswerte.

Der Standardfehler

Achtung: Der Ausdruck „Standardabweichung“ auf der letzten Folie bezog sich auf die Standardabweichung von *Stichprobenkennwerten*. Um Verwechslungen mit der Standardabweichung empirisch gemessener Variablen vorzubeugen, bezeichnet man die Standardabweichung von Stichprobenkennwerten mit einem eigenen Begriff, dieser lautet Standardfehler (englisch: Standard Error, davon die häufig verwendete Abkürzung S. E.).

Der Standardfehler gibt uns also an, wie stark die Stichprobenkennwerte (die wir bei Ziehen vieler Stichproben erhalten) um den wahren Wert (den Wert in der Grundgesamtheit) schwanken. Wenn die Verteilung des Standardfehlers bekannt ist, gilt noch mehr: Wir können sagen, wie wahrscheinlich es ist, dass die Stichprobenkennwerte in einer bestimmten Entfernung vom wahren Wert liegen.

Wie schon erwähnt, ist die Verteilung des Standardfehlers für Mittel- und Anteilswerte in großen Stichproben bekannt – es ist die (Standard-)Normalverteilung.

Der Standardfehler von Mittelwerten

Wenn wir viele Stichproben ziehen und daraus jeweils den Mittelwert eines uns interessierenden Merkmals schätzen, so beträgt der Standardfehler (S. E.), also das Ausmaß der Streuung der Stichproben mittelwerte um den wahren Wert der Grundgesamtheit:

$$S.E._{Mittelwert} = \sqrt{\frac{\sigma_x^2}{n}} = \frac{\sigma_x}{\sqrt{n}}$$

Der Standardfehler hängt also ab

- von der Streuung des Merkmals in der Grundgesamtheit, und
- vom Stichprobenumfang n.

Der Standardfehler von Anteilswerten

Wenn wir viele Stichproben ziehen und daraus jeweils den Anteilswert eines uns interessierenden Merkmals schätzen, so beträgt der Standardfehler (S. E.):

$$S.E._{Anteilswert} = \sqrt{\frac{\pi_1 \cdot (1 - \pi_1)}{n}} = \frac{\sqrt{\pi_1 \cdot (1 - \pi_1)}}{\sqrt{n}}$$

Wenn π_1 nicht in Anteilswerten sondern in Prozentwerten ausgedrückt wird, (also z. B. 40 Prozent statt 0,4), muss $(1 - \pi_1)$ durch $(100 - \pi_1)$ ersetzt werden!

Der Standardfehler hängt also ab

- vom Anteilswert π_1, den das Merkmal in der Grundgesamtheit aufweist, und
- wiederum vom Stichprobenumfang n.

Standardfehler bei der Schätzung von Anteilswerten

Angenommen, eine Partei hat einen Stimmenanteil von 40 Prozent, als Anteilswert: 0,4. In Abhängigkeit von der Stichprobengröße gelten dann (beispielsweise) folgende Standardfehler für den Anteilswert:

n	S. E.
96	0,0500
192	0,0354
384	0,0250
768	0,0177
1536	0,0125

Um den Standardfehler zu halbieren, muss der Stichprobenumfang vervierfacht werden. Dies ist Folge der Tatsache, dass im Nenner von S. E. die Wurzel aus n steht („Wurzel-n-Gesetz“).

Die Streuung von Anteils-/Mittelwerten in Stichproben

Wenn wir große Stichproben ziehen (wie in der Sozialforschung üblich), folgt die Verteilung der Stichprobenkennwerte einer Normalverteilung. Wir können dann in Anwendung der oben formulierten Verteilungsgesetze für wiederholte Stichprobenziehungen sagen (in Klammern: Beispiel der vorherigen Folie für n = 384):

- Ca. 68 % der Werte liegen in einem Bereich von $\pm 1\ S.E.$ um den wahren Anteils- oder Mittelwert
(Bereich von $0{,}4 - 0{,}025$ bis $0{,}4 + 0{,}025$, also 0,375 bis 0,425)
- Gut 95 % der Werte liegen in einem Bereich von $\pm 2\ S.E.$ um den wahren Anteils- oder Mittelwert
(Bereich von $0{,}4 - 2 \cdot 0{,}025$ bis $0{,}4 + 2 \cdot 0{,}025$, also 0,35 bis 0,45)
- Ca. 99,7 % der Werte liegen in einem Bereich von $\pm 3\ S.E.$ um den wahren Anteils- oder Mittelwert.
(Bereich von $0{,}4 - 3 \cdot 0{,}025$ bis $0{,}4 + 3 \cdot 0{,}025$, also 0,325 bis 0,475)

Häufigkeitsaussagen als Wahrscheinlichkeitsaussagen

Aussagen darüber, wie häufig etwas geschieht, lassen sich auch als Wahrscheinlichkeitsaussagen formulieren. Wir können also die Aussagen der vorherigen Folie wie folgt reformulieren:

- Mit einer Wahrscheinlichkeit von ca. 68 % liegen die Werte in Stichproben in einem Bereich von $\pm 1\ S.E.$ um den wahren Anteils- oder Mittelwert
- Mit einer Wahrscheinlichkeit von ca. 95 % liegen die Werte in Stichproben in einem Bereich von $\pm 2\ S.E.$ um den wahren Anteils- oder Mittelwert.
  Sozialwissenschaftler interessieren sich am häufigsten für diese Wahrscheinlichkeit und formulieren exakt: „... in einem Bereich von $\pm 1{,}96\ S.E.$ um den wahren Anteils- oder Mittelwert.“
- Mit einer Wahrscheinlichkeit von ca. 99,7 % liegen die Werte in Stichproben in einem Bereich von $\pm 3\ S.E.$ um den wahren Anteils- oder Mittelwert.

Schätzen und Testen

In den folgenden Abschnitten kommen wir zur Anwendung auf das Problem der empirischen Sozialforschung: Wir haben eine Stichprobe vorliegen und wollen anhand ihrer Aussagen über die (unbekannte) Grundgesamtheit machen. Wir unterscheiden zwischen:

- Schätzen (Aussagen über Größen in der Grundgesamtheit)
- Testen (Prüfen von Hypothesen).

Punkt- und Intervallschätzung: Schätzung von Parametern der Grundgesamtheit aus einer Stichprobe

Punkt- und Intervallschätzung

Beim Schätzen geht es um Aussagen über die Grundgesamtheit: Was ist der Mittelwert der Variablen X? Was ist der Anteilswert der Ausprägung 1 der Variablen Y? Usw. Dabei unterscheiden wir zwischen

- Punktschätzung: Berechnung *eines*, und zwar des besten Schätzwertes für die Grundgesamtheit.
  Da allerdings auch der beste Schätzwert meist mit sehr großer Unsicherheit behaftet ist, sollte man auch immer eine
- Intervallschätzung durchführen. Hierbei errechnet man ein Intervall, das mit großer Wahrscheinlichkeit den wahren Wert (den Wert der Grundgesamtheit) einschließt.

Punktschätzung

Terminologie: Eine für die Stichprobe berechnete Größe (Mittelwert, Anteilswert, ...) heißt *Kennwert* (kurz für: Stichprobenkennwert). Der wahre Wert (Wert in der Grundgesamheit) wird als *Parameter* bezeichnet.

Will man nicht eine bestimmte Größe bezeichnen, sondern bezieht sich allgemein auf verschiedene Parameter bzw. Kennwerte, verwendet man meist das Symbol θ (griech. theta). Wieder unterscheiden wir:

$\theta =$ Bekannter Parameter der Grundgesamtheit,

$\widehat{\theta} =$ Ein Parameter, der aus einer Stichprobe geschätzt wird.

Oft ist der Wert, der die Stichprobendaten kennzeichnet, identisch mit $\widehat{\theta}$ (er ist also der beste Schätzwert für den Parameter). Das gilt etwa für Mittel- und Anteilswerte. Eine Ausnahme haben wir schon kennengelernt: die Varianz.

Auf Probleme der Punktschätzung gehen wir hier nicht weiter ein; gegebenenfalls werden die Formeln einfach ohne Begründung eingeführt.

Intervallschätzung

Die Wahrscheinlichkeit, dass ein Schätzwert $\widehat{\theta}$ *genau* dem wahren Parameter entspricht, ist ziemlich gering – auch wenn es sich um den besten möglichen Schätzwert handelt. Denn wir wissen ja, dass die Kennwerte streuen (also: mal diesen, mal jenen Wert annehmen) können, je nach den Zufälligkeiten der Stichprobenziehung.

Wenn aber Zufallsstichproben gezogen wurden, kann man unter Heranziehen des zentralen Grenzwertsatzes angeben, wie weit die Stichprobenkennwerte um den Parameter streuen; genauer gesagt kann man für gegebene Wahrscheinlichkeiten angeben, in welchem Bereich die Stichprobenkennwerte liegen (siehe den vorherigen Teil der Vorlesung). Daraus ergibt sich eine verblüffend einfache Lösung für das Problem der Intervallschätzung – also für die Aufgabe, ein Intervall zu finden, das mit (großer) Wahrscheinlichkeit den wahren Parameter enthält.

Das Beispiel 95-Prozent-Konfidenzintervall

Um ein Intervall anzugeben, welches den Wert in der Grundgesamtheit wahrscheinlich enthält, müssen wir festlegen, was wir unter „wahrscheinlich“ verstehen wollen.

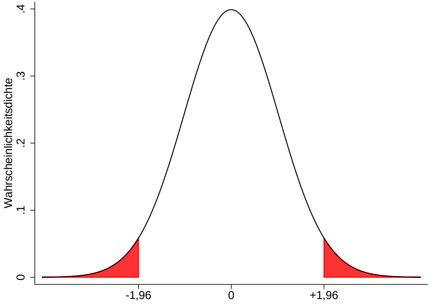
In den Sozialwissenschaften hat sich als Standard eine Wahrscheinlichkeit von 95 Prozent etabliert. Gesucht wird also ein Intervall, das mit 95-prozentiger Wahrscheinlichkeit den wahren Wert der Grundgesamtheit enthält. Man bezeichnet es auch als „95-Prozent-Konfidenzintervall“ (Konfidenz = Vertrauen [dass das Intervall den wahren Wert enthält]).

Das *Gegenteil* dieser Wahrscheinlichkeit heißt Irrtumswahrscheinlichkeit. Sie wird mit dem griech. Buchstaben α (alpha) bezeichnet. Bei einem 95-Prozent-Konfidenzintervall ist $\alpha = 5$ Prozent (als Anteil: 0,05).

Der Wert $100 - \alpha$ wird als Vertrauenswahrscheinlichkeit, Konfidenzniveau oder auch Überdeckungswahrscheinlichkeit (Fahrmeir et al.) bezeichnet.

Beispiel: Ein α von 5 Prozent bedeutet, dass das Intervall den gesuchten Wert der Grundgesamtheit mit 95-prozentiger Wahrscheinlichkeit enthält.

Der Ausgangspunkt: Das Wahrscheinlichkeitsintervall



Erinnerung: Ziehen wir aus einer Grundgesamtheit viele (große) Stichproben, so liegen die Stichprobenkennwerte $\widehat{\theta}$ (Mittelwerte, Anteilswerte, ...) mit 95-prozentiger Wahrscheinlichkeit im Bereich $\pm$1,96 S. E. um den wahren Wert der Grundgesamtheit. Diesen Bereich nennen wir Wahrscheinlichkeitsintervall.

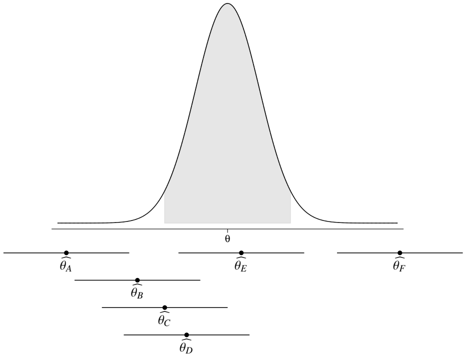
Das Konfidenzintervall

Für die einzelnen Stichproben gilt dann:

- Wenn der Stichprobenkennwert $\widehat{\theta}$ im (95-Prozent-) Wahrscheinlichkeitsintervall liegt, heißt das, dass er nicht mehr als 1,96 S. E. vom wahren Wert entfernt liegt. Somit enthält das Intervall von $\widehat{\theta} - 1{,}96$ S. E. bis $\widehat{\theta} + 1{,}96$ S. E. den wahren Wert der Grundgesamtheit.
- Liegt der Stichprobenkennwert $\widehat{\theta}$ nicht im Wahrscheinlichkeitsintervall, so enthält das Intervall von $\widehat{\theta} - 1{,}96$ S. E. bis $\widehat{\theta} + 1{,}96$ S. E. nicht den wahren Wert der Grundgesamtheit.

Die Abbildung auf der nächsten Seite zeigt einige Beispiele.

Konfidenzintervalle visualisiert



Die Intervalle um $\widehat{\theta}_A$, $\widehat{\theta}_B$ und $\widehat{\theta}_F$ schließen den wahren Wert θ nicht ein. $\widehat{\theta}_A$, $\widehat{\theta}_B$ und $\widehat{\theta}_F$ gehören zu den 5 Prozent der „unwahrscheinlichsten" (mehr als 1,96 S.E. vom wahren Wert entfernten) Stichprobenkennwerte.

Konfidenzintervalle – erstes Fazit

Am Beispiel der (meist gewählten) Irrtumswahrscheinlichkeit von 5 Prozent kann man also sagen:

Die 5 Prozent der „unwahrscheinlichsten" Stichprobenkennwerte (also derjenigen, die mehr als $-1{,}96$ oder $+1{,}96$ S. E. vom wahren Wert – dem Wert der Grundgesamtheit – entfernt liegen) führen zu einem Konfidenzintervall, das den wahren Wert *nicht* einschließt.

Die übrigen 95 Prozent der Stichprobenkennwerte führen zu einem Konfidenzintervall, das den wahren Wert einschließt.

Ergo: Mit einer Wahrscheinlichkeit von 95 Prozent enthält ein Konfidenzintervall um einen (einzelnen) gegebenen Stichprobenkennwert den wahren Wert.

Achtung: Ob nun im konkreten Fall das Konfidenzintervall den wahren Wert enthält oder nicht, wissen wir nicht – denn wir wissen ja nicht, ob der geschätzte Parameter $\widehat{\theta}$ im Wahrscheinlichkeitsintervall liegt oder nicht.

Konfidenzintervalle for practical purposes

Für eine grobe („quick and not so dirty after all“) Abschätzung des 95-Prozent-Konfidenzintervalls kann man das Intervall

$$\bar{x} - 2 \cdot S.E. \leqslant \bar{x} \leqslant \bar{x} + 2 \cdot S.E.$$

für Mittelwerte bzw.

$$p_1 - 2 \cdot S.E. \leqslant p_1 \leqslant p_1 + 2 \cdot S.E.$$

für Anteilswerte verwenden.

Diese Faustregel kann man für Stichproben ab $n = 100$, andere sagen: bereits ab $n = 30$, verwenden.

Berechnung von S. E. bei der Intervallschätzung

Es bleibt noch ein kleines Problem zu lösen: Woher sollen wir bei unbekannter Grundgesamtheit die Standardfehler kennen?

Die Lösung lautet: Wir schätzen diese Werte einfach aus der Stichprobe! Es gilt also:

$$S.E._{Mittelwert} = \frac{\widehat{\sigma}_x}{\sqrt{n}} = \frac{s_x}{\sqrt{n-1}}$$

$$S.E._{Anteilswert} = \sqrt{\frac{p_1 \cdot (1 - p_1)}{n}}$$

$\widehat{\sigma}_x$ = aus der Stichprobe geschätzte Standardabweichung von X
s_x = Standardabweichung von X, berechnet für vorliegende Daten
p_1 = Anteilswert der interessierenden Ausprägung in der Stichprobe

Wird p_1 als Prozentwert ausgedrückt, muss $(100 - p_1)$ verwendet werden.

Beispiel: Konfidenzintervall für einen Anteilswert

Das Intervall $\pm 1{,}96$ S.E. um den geschätzten Anteilswert (im Beispiel: 0,4) enthält mit 95-prozentiger Wahrscheinlichkeit den wahren Wert:

n	S. E.	Konfidenzintervall
96	0,0500	0,302 bis 0,498
192	0,0354	0,330 bis 0,469
384	0,0250	0,351 bis 0,449
768	0,0177	0,365 bis 0,435
1536	0,0125	0,376 bis 0,425

Beispiel: Konfidenzintervall für einen Mittelwert

Gegeben: Alter in einer Stichprobe: $n = 100$, ar. Mittel $= 42$.
Standardabweichung $\widehat{\sigma}_x$ (Schätzwert für Grundgesamtheit) $= 11{,}1$.
Konfidenzniveau wird festgelegt auf 95 Prozent.

$$\text{Untergrenze: } 42 - 1{,}96 \cdot \frac{11{,}1}{\sqrt{100}} \approx 39{,}8$$

$$\text{Obergrenze: } 42 + 1{,}96 \cdot \frac{11{,}1}{\sqrt{100}} \approx 44{,}2$$

Wegen $n = 100$ kann problemlos die Standardnormalverteilung verwendet werden.

Konfidenzintervalle mit anderen Konfidenzniveaus

Allgemein gilt für große Stichproben:

- Die linke (untere) Grenze des Konfidenzintervalls liegt bei $\widehat{\theta} -$ S. E. $\cdot$ $(1 - \alpha/2)$-Wert
- Die rechte (obere) Grenze des Konfidenzintervalls liegt bei $\widehat{\theta} +$ S. E. $\cdot$ $(1 - \alpha/2)$-Wert

$(1 - \alpha/2)$-Wert: Wert einer geeigneten Zufallsvariablen (hier: der Standardnormalverteilung), der dem Quantil $(1 - \alpha/2)$ entspricht (α muss hier als Anteilswert ausgedrückt werden, also z. B. 0,05 statt 5 Prozent!).

Wichtige Fälle:

- Konfidenzniveau 90 Prozent ($\alpha = 10$ Prozent): 1,645
- Konfidenzniveau 95 Prozent ($\alpha = 5$ Prozent): 1,96
- Konfidenzniveau 99 Prozent ($\alpha = 1$ Prozent): 2,576

In Tabellen (oder implementiert in Software) finden sich Werte für beliebige andere Konfidenzintervalle.

Konfidenzintervalle bei kleinen Stichproben

Bei kleinen Stichproben (je nach Standpunkt: $n < 100$ bis $n < 30$, für Anteilswerte wird auch genannt: $n < 60$) gelten die bisher diskutierten Regeln nicht.

- Für Mittelwerte wird statt der Standardnormalverteilung die t-Verteilung herangezogen. Diese berücksichtigt, dass der Standardfehler nur eine Schätzung darstellt (die umso ungenauer ist, je kleiner die Stichprobe!). Die t-Verteilung sieht ähnlich aus wie die SNV, nur sind ihre Enden breiter, so dass die Quantilwerte entsprechend größer werden ($n = 20$: 97,5-Prozent-Quantilwert $\approx 2{,}09$).
- Für Anteilswerte gibt es diverse Lösungsvorschläge, die hier nicht besprochen werden müssen. Achtung: Auch für sehr extreme Anteilswerte (etwa $p_1 < 0{,}05$ bzw. $p_1 > 0{,}95$) gelten andere Regeln.

Abschließendes zur Intervallschätzung

Beachten Sie: Die Aussage „Das Konfidenzintervall enthält (oder überdeckt) mit 95-prozentiger Wahrscheinlichkeit den wahren Wert“ ist eine Aussage über das Schätzverfahren. Sie besagt: Auf lange Sicht liefert das Verfahren in 95 Prozent der Fälle ein Intervall, das den wahren Wert enthält (echte Zufallsstichproben vorausgesetzt). Ein konkretes Konfidenzintervall enthält entweder den wahren Wert oder es enthält ihn nicht – welches davon der Fall ist, können wir nicht wissen.

So kann man nicht sagen (im Beispiel „Alter“ ein paar Folien weiter vorne): „Das Ergebnis bedeutet, dass 95 Prozent der Stichproben Werte zwischen 39,8 und 44,2 enthalten“. Man kann nur sagen: Beim Intervall von 39,8 bis 44,2 besteht eine 95-Prozent-Chance, dass es den wahren Wert enthält.

Noch etwas Anderes: In einigen Lehrbüchern finden Sie auch Hinweise auf einseitige Konfidenzintervalle; diese werden hier nicht behandelt.

Statistisches Testen:
Prüfen von Annahmen über die Grundgesamtheit anhand einer Stichprobe

Statistisches Testen: Fragestellung

Ausgangslage und Problem

In einer Stichprobe findet sich ein Zusammenhang zwischen zwei Merkmalen oder, was das gleiche ist, ein Unterschied zwischen zwei oder mehr Gruppen hinsichtlich eines relevanten Merkmals.

Bei Stichproben besteht jedoch immer der Verdacht, dass der Zusammenhang nur zufällig (aufgrund von Stichprobenschwankungen) aufgetreten ist, dass er also in der Grundgesamtheit gar nicht besteht.

Wie können wir diesem Verdacht begegnen?

Statistisches Testen: Grundidee einer Lösung

Die Beweisführung ist indirekt und liefert keine Gewissheit, sondern nur Plausibilität.

1. Wir gehen von der Annahme aus, dass der Zusammenhang in der Grundgesamtheit tatsächlich nicht besteht.
2. Wir errechnen, wie wahrscheinlich oder unwahrscheinlich das Stichprobenergebnis unter dieser Annahme wäre.
3. Wenn sich zeigt, dass das Stichprobenergebnis im Lichte dieser Annahme unwahrscheinlich ist, dann schließen wir, dass die Annahme (vermutlich) falsch ist.
4. Das ist aber gleichbedeutend damit, dass vermutlich die gegenteilige Annahme zutrifft, also die, dass der Zusammenhang auch in der Grundgesamtheit besteht.

Ein im Licht der Annahme, dass kein Zusammenhang besteht, unwahrscheinliches Ergebnis nennt man (statistisch) signifikant. Daher spricht man auch von Signifikanztests.

Vorgehen bei Signifikanztests im Überblick

1. Die Annahme, dass in Wahrheit (also in der Grundgesamtheit) kein Zusammenhang besteht, heißt Nullhypothese. Sie stellt das Gegenteil unserer eigenen (einen Zusammenhang vermutenden) Forschungshypothese dar, die auch Alternativhypothese genannt wird.
2. Um zu errechnen, wie (un)wahrscheinlich das beobachtete Stichprobenergebnis im Lichte der Nullhypothese ist, benötigen wir eine geeignete Prüfgröße (Teststatistik).
3. Außerdem müssen wir entscheiden, wie unwahrscheinlich das Stichprobenergebnis im Lichte der Nullhypothese sein soll, damit wir es als Beleg gegen die Nullhypothese akzeptieren (sog. Signifikanzniveau.
4. Nun können wir die Teststatistik berechnen und uns auf dieser Grundlage für oder gegen die Nullhypothese (und damit gegen oder für die Forschungshypothese) entscheiden.

Vorgehen bei Signifikanztests, Schritt 1

- *Beispiel*: Es soll (anhand von Stichprobendaten) geprüft werden, ob und gegebenenfalls wie sich die Löhne von Frauen und Männern unterscheiden.
- Die uns inhaltlich interessierende Vermutung, dass ein Lohnunterschied besteht (und gegebenenfalls: welche Richtung er hat), heißt Alternativhypothese (Kürzel: H_1 oder H_A), manchmal auch als Forschungshypothese bezeichnet.
- Als „Gegenannahme“ formuliert man eine Nullhypothese (H_0), deren genaue Formulierung von der Alternativhypothese abhängt (siehe nächste Folie).

Vorgehen bei Signifikanztests, Schritt 1

Man kann zwei Arten von Forschungshypothesen unterscheiden:

- Gerichtete (oder einseitige) Hypothesen, Beispiel: Frauen verdienen weniger als Männer (allgemein: die Richtung eines Unterschieds oder eines Zusammenhangs wird angegeben).

  Im Beispiel: H_A: $\mu_F < \mu_M$; H_0: $\mu_F \geq \mu_M$

- Ungerichtete (oder zweiseitige) Hypothesen: Frauen verdienen nicht das Gleiche wie Männer (allgemein: Es wird angenommen, dass ein Unterschied oder ein Zusammenhang besteht, über dessen Richtung wird jedoch nichts ausgesagt).

  Im Beispiel: H_A: $\mu_F \neq \mu_M$; H_0: $\mu_F = \mu_M$

μ_F, μ_M: Mittelwerte von Frauen bzw. Männern in der Grundgesamtheit (zur Erinnerung: μ = „mü“)!

Vorgehen bei Signifikanztests, Schritt 2

Als nächstes ist die Prüfgröße (Teststatistik) festzulegen.

Hierbei handelt es sich um eine aus den Stichprobendaten zu berechnende Kennzahl, die Aufschluss darüber gibt, ob die Daten mit der Nullhypothese vereinbar (also: im Lichte der Nullhypothese einigermaßen wahrscheinlich) sind oder nicht.

Es gibt eine ganze Menge solcher Teststatistiken. Allgemein gilt: Welche Teststatistik wir in welchem Fall anwenden, übernehmen wir aus den Statistik-Lehrbüchern.

In Schritt 4 lernen wir exemplarisch eine Teststatistik kennen, die Teststatistik für einen Mittelwertunterschied.

Vorgehen bei Signifikanztests, Schritt 2

Der „Sinn" von Teststatistiken:

Eine Teststatistik ist eine Transformation des Stichprobenergebnisses in eine (den Statistikern!) bekannte Verteilung einer Zufallsvariablen. Eine wichtige Verteilung kennen wir schon: Die Standardnormalverteilung. Eine andere wichtige Verteilung, die t-Verteilung, geht bei großen Fallzahlen in die Standardnormalverteilung über.

Beispiel: In einer Stichprobe bestehe zwischen zwei Gruppen ein Einkommensunterschied von € 800. Kann das Zufall (also: wahrscheinlich) sein, wenn in der Grundgesamtheit kein Unterschied besteht? Hierfür fehlt uns ein Maßstab.

Wenn wir den Unterschied in die Standardnormalverteilung (oder die t-Verteilung) überführen können, haben wir einen solchen Maßstab. Wenn wir annehmen, dass es keinen Unterschied gibt (ungerichtete Hypothese!), sind am wahrscheinlichsten Werte um Null herum. Werte am Ende der Verteilung (z.B. kleiner als $-1{,}96$ oder größer als $+1{,}96$) sind recht unwahrscheinlich.

Vorgehen bei Signifikanztests, Schritt 3

Festlegung des Signifikanzniveaus:

Auch statistisches Testen geschieht unter Unsicherheit; welche Unsicherheit wir akzeptieren, müssen wir (wie bei Konfidenzintervallen) selbst festlegen.

Eine „Unsicherheit“ bei Signifikanztests ist das Risiko (quantifiziert als die Wahrscheinlichkeit), die Nullhypothese abzulehnen (und damit unsere Forschungshypothese, die Alternativhypothese, vorläufig zu akzeptieren), obwohl die Nullhypothese tatsächlich zutrifft. Dieses Risiko heißt Signifikanzniveau. Oft wird auch hier von Irrtumswahrscheinlichkeit gesprochen, was aber missverständlich sein kann.

Das Signifikanzniveau sollte möglichst niedrig sein, z. B. 0,05 (als Prozentwert: 5 Prozent) oder noch geringer. In den Sozialwissenschaften wird in der Regel der Wert von 0,05 gewählt.

Vorgehen bei Signifikanztests, Schritt 3

Signifikanzniveau, Ablehnungsbereich und kritischer Wert:

Das Signifikanzniveau entspricht dem Bereich der Werte einer Zufallsvariablen, die (im Lichte der Nullhypothese) „extrem“ und daher unwahrscheinlich sind. Dieser Bereich heißt auch Ablehnungsbereich. Der Wert (bzw. die Werte), der/die den Ablehnungsbereich (oder die Ablehnungsbereiche) von den übrigen Werten trennt/trennen, heißt auch kritischer Wert/kritische Werte.

Ein Beispiel folgt auf den nächsten beiden Folien.

Vorgehen bei Signifikanztests, Schritt 3

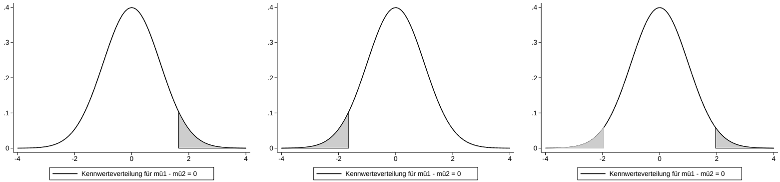
Ablehnungsbereich: Angenommen, die Teststatistik folgt einer Standardnormalverteilung (SNV).

Bei einer ungerichteten Hypothese und einem Signifikanzniveau von 5 % beschreiben jeweils die „äußersten“ 2,5 Prozent der Verteilung den Ablehnungsbereich – also die Werte zwischen $-\infty$ und $-1{,}96$ und die Werte zwischen $+1{,}96$ und $+\infty$.

Bei einer gerichteten (einseitigen) Hypothese und gleicher Irrtumswahrscheinlichkeit beschreiben die obersten (oder untersten) 5 % der SNV den Ablehnungsbereich – also die Werte $\geq +1{,}645$ (wenn die Hypothese einen positiven Unterschied behauptet) oder $\leq -1{,}645$ (bei negativem Unterschied).

Vorgehen bei Signifikanztests, Schritt 3 visualisiert

Je nach Art und Richtung der Forschungshypothese liegt der Ablehnungsbereich im oberen Bereich ($\mu_1 > \mu_2$, linke Graphik), im unteren Bereich ($\mu_1 < \mu_2$, mittlere Graphik) oder im oberen und unteren Bereich ($\mu_1 \neq \mu_2$, rechte Graphik). In letzterem Fall entspricht der Ablehnungsbereich auf jeder Seite $\alpha/2$.



Die Graphiken zeigen das Beispiel der vorherigen Seite: SNV und $\alpha = 0{,}05$.

Vorgehen bei Signifikanztests, Schritt 4

Beispiel für eine Teststatistik: Unterschied zwischen zwei Mittelwerten $\bar{x}_1 - \bar{x}_2$.

Die entsprechende Teststatistik lautet:

$$T = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\left(\frac{\widehat{\sigma}_1^2}{n_1} + \frac{\widehat{\sigma}_2^2}{n_2}\right)}}$$

mit n_1 und n_2 = Umfang der Stichprobe von Gruppe 1 und Gruppe 2 sowie $\widehat{\sigma}_1^2$ und $\widehat{\sigma}_2^2$ = geschätzte Varianz der beiden Gruppen.

Der Mittelwertunterschied wird also durch folgende Größe dividiert, die den Standardfehler des Mittelwertunterschieds darstellt:

$$\sqrt{\left(\frac{\widehat{\sigma}_1^2}{n_1} + \frac{\widehat{\sigma}_2^2}{n_2}\right)}$$

Vorgehen bei Signifikanztests, Schritt 4

Der Test auf einen Mittelwertunterschied im Beispiel:

Unsere Stichprobe enthalte 100 Männer und 100 Frauen. Das Durchschnittseinkommen der Männer $\bar{x}_1$ betrage € 2 800 bei einer Standardabweichung $\widehat{\sigma}_1$ von 1 500, das Einkommen der Frauen $\bar{x}_2$ betrage € 2 000 bei einer Standardabweichung $\widehat{\sigma}_2$ von 1 000.

Nicht vergessen: in die Formel muss die Varianz, also die quadrierte Standardabweichung eingesetzt werden!

$$T = \frac{2\,800 - 2\,000}{\sqrt{\left(\frac{2\,250\,000}{100} + \frac{1\,000\,000}{100}\right)}} = \frac{800}{180{,}277} = 4{,}4376.$$

Die T-Statistik folgt einer t-Verteilung; diese kann jedoch bei größeren Stichproben (wie sie hier zweifelsfrei vorliegen) durch die Standardnormalverteilung angenähert werden.

Der Signifikanztest im Beispiel, zusammengefasst:

1. H_1 lautet: $\mu_1 > \mu_2$; es handelt sich also um eine einseitige Forschungshypothese mit der H_0: $\mu_1 \leq \mu_2$
2. Da es sich um einen Mittelwertunterschied handelt, ist die geeignete Teststatistik die auf den vorhergehenden Folien dargestellte Teststatistik für Mittelwertdifferenzen.[1]
3. Als Signifikanzniveau haben wir $\alpha = 0{,}05$ festgelegt; bei einer positiven gerichteten Hypothese beträgt damit der kritische Wert bei der hier zugrundegelegten Standardnormalverteilung 1,645.
4. Da $4{,}4376 > 1{,}645$ (anders gesagt: die errechnete Teststatistik liegt im Ablehnungsbereich), wird die H_0 (vorläufig) verworfen und die H_1 (vorläufig) akzeptiert – unter dem Vorbehalt, dass dies auch irrtümlich geschehen könnte.

[1] Genau genommen gilt die Formel nur, wenn die Varianzen in beiden Gruppen unterschiedlich groß sind, wie es hier der Fall ist.

Empirische Signifikanz

- Statistikprogramme geben im allgemeinen nicht an, ob ein bestimmtes Signifikanzniveau überschritten wird oder nicht, sondern sie geben exakt den Quantilswert an, der der jeweiligen Prüfgröße entspricht (also z.B.: „$T = 4{,}4376$; $p = 0{,}00000455$“ oder wahrscheinlicher: „$p = 0{,}000$“). Der p-Wert wird manchmal als empirisches Signifikanzniveau bezeichnet. Man achtet dann darauf, ob gilt: $p < \alpha$. Da $0{,}00000455 < 0{,}05$, kommen wir auch hier zu dem Schluss, dass wir die Forschungshypothese akzeptieren können.
- Wichtig: Ein p-Wert von 0,000 ist ein gerundeter Wert und bedeutet: das empirische Signifikanzniveau ist $<0{,}0005$. Einen p-Wert von exakt 0 gibt es nicht.
- Ebenfalls ist zu beachten, dass manche Statistikpakete nicht (oder nur nach Aufforderung durch die Benutzer) zwischen einseitigen und zweiseitigen Hypothesen unterscheiden.

Mehr über Signifikanztests

Zu sagen, dass ein Ergebnis „signifikant“ (im statistischen Sinne) ist, bedeutet nur: Die Wahrscheinlichkeit, dass ein Stichprobenergebnis wie das vorliegende zustande kommen würde, wenn in Wirklichkeit gar kein Unterschied (oder Zusammenhang) bestünde, ist ziemlich gering (nämlich so klein wie das gewählte Signifikanzniveau).

Das Stichprobenergebnis ist „unwahrscheinlich“ also nur gemessen an der H_0, die eine rein fiktive Größe ist. Eine Aussage, wie wahrscheinlich oder unwahrscheinlich unser Stichprobenergebnis ist, kann nicht gemacht werden. (Erst recht sagen die Ergebnisse nichts darüber, wie wahrscheinlich oder unwahrscheinlich die H_0 ist.)

Auch ein signifikantes Ergebnis steht unter Vorbehalt, da es ja nur „unwahrscheinlich“ (im Lichte der H_0) ist; die H_0 kann also nicht mit Sicherheit abgelehnt werden.

Achtung Sprachregelung!

- Ein statistisch signifikantes Ergebnis bedeutet, dass man die H_0 ablehnen kann; oft sagt man auch: die H_0 wird verworfen.
- Durch die Ablehnung der H_0 hat man eine *indirekte* Bestätigung der Forschungshypothese H_A: Weil es unplausibel (wenn auch keineswegs mit Gewissheit falsch!) ist, dass die H_0 zutrifft, die H_0 aber das Gegenteil der H_A behauptet, folgt aus der Ablehnung der H_0, dass die Daten eher für die H_A sprechen.
- Da man nur über eine indirekte Bestätigung der H_A verfügt, sagen manche ForscherInnen, dass man nicht davon sprechen dürfe, dass die H_A „bestätigt" sei.
- Allerdings ist es nicht konsequent, einerseits die H_0 abzulehnen, andererseits dies nicht als Entscheidung für die H_A zu interpretieren. Ich selbst spreche davon, dass man die H_A „akzeptieren" könne (immer zumindest gedanklich mit dem Zusatz „unter Unsicherheit" oder „vorläufig"), habe aber auch nichts dagegen, wenn man sagt, die H_A sei „bestätigt" (immer unter Unsicherheit).

Warnung vor Internetquellen

„**1.3.2 Irrtumswahrscheinlichkeit und Signifikanzniveau**

Unter der Irrtumswahrscheinlichkeit p versteht man die zahlenmäßig ausgedrückte Wahrscheinlichkeit, dass sich ein Ergebnis einer statistischen Analyse substantiell vom tatsächlichen Ergebnis der Grundpopulation unterscheidet. Falsch!

In der Statistik arbeitet man meist mit den drei folgenden Signifikanzniveaus oder -grenzen:

$p \leq 0{,}05$: signifikant (Irrtumswahrscheinlichkeit kleiner als 5 %) (…)“

Quelle: Erwin Ebermann, Universität Wien, Institut für Kultur- und Sozialanthropologie, unter `http://www.univie.ac.at/ksa/elearning/cp/quantitative/quantitative-8.html`

Fehler 1. und 2. Art

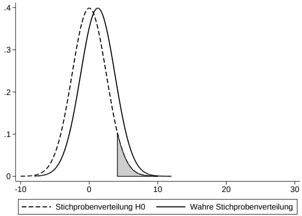
Es gibt tatsächlich *zwei* Quellen des „Risikos“ bei Signifikanztests: Neben dem Fehler, die H_0 zu verwerfen, obwohl sie in Wahrheit zutrifft (auch Fehler 1. Art oder α-Fehler genannt), kann man auch den Fehler begehen, die H_0 nicht zu verwerfen, obwohl sie in Wahrheit falsch ist (auch β-Fehler genannt).

		Tatsächlich gilt:	
		H_0	H_A
Entscheidung aufgrund der *Stichprobe*:	H_0	richtig	β-Fehler
	H_A	α-Fehler	richtig

Getestet wird in der Regel der α-Fehler!

Der Fehler 2. Art

Über die Wahrscheinlichkeit des Fehlers 2. Art kann man ohne weiteres Wissen keine Aussagen machen. Man kann nur sagen, dass er sich umgekehrt zum Fehler 1. Art verhält: Wenn ich das Risiko, eine wahre H_0 zu verwerfen, minimiere (durch kleines α), steigt das Risiko, eine falsche H_0 beizubehalten. Genauere Aussagen ließen sich nur machen, wenn man wüsste, wie groß der Unterschied (Zusammenhang) in der Grundgesamtheit wirklich ist!



Wahrscheinlichkeit für Fehler 2. Art: Links sehr groß, rechts sehr gering!

Warnung vor Überschätzung von Signifikanztests

Die Ergebnisse von Signifikanztests hängen ganz wesentlich von der Stichprobengröße ab! Mit großen Stichproben kann man fast alle Zusammenhänge „beweisen", mit kleinen Stichproben kann man „beweisen", dass Zusammenhänge nicht bestehen.

Erinnern wir uns an die erste Stunde, Daten zum CHE-Ranking: Eine Universität der Mittelgruppe hatte schlechtere Ergebnisse als Universitäten der Schlussgruppe. Wie ist das möglich?

Universitäten wurden grundsätzlich der Mittelgruppe zugeschlagen, wenn sie keine statistisch signifikant besseren oder schlechteren Ergebnisse als der Durchschnitt erzielten. Je weniger Studierende einer Universität an der Befragung teilnahmen, desto größer also die Wahrscheinlichkeit, dass die Universität zur Mittelgruppe zählte (selbst bei schlechten Ergebnissen). Wenn dagegen viele Studierende einer Universität teilnahmen, konnten auch schon kleinere Abweichungen von Mittelwert statistisch abgesichert werden.

Warnung vor Überschätzung von Signifikanztests

Wenn SozialwissenschaftlerInnen mit Blick auf statistische Datenauswertungen sagen, etwas (ein Zusammenhang, ein Unterschied, ein Einfluss, ein Regressionskoeffizient) sei (evtl. explizit: statistisch) „signifikant“, so heißt das nur:

„Ein Zusammenhang (Unterschied, Einfluss . . .), der in einer Stichprobe gefunden wurde, wäre ziemlich unwahrscheinlich, wenn in der Grundgesamtheit kein Zusammenhang (o. ä.) bestünde – daher nehmen wir (mit einem gewissen Risiko, uns zu irren) an, dass auch in der Grundgesamtheit ein Zusammenhang besteht.“

Das sagt absolut nichts darüber, ob der Zusammenhang bzw. Unterschied stark (groß, deutlich, wichtig) ist!

Warnung vor Überschätzung von Signifikanztests

Das Wort „signifikant“ kann auch in anderen (nicht-statistischen) Kontexten auftreten. *Hier* soll es im allgemeinen so etwas wie „wichtig“, „bedeutsam“, „bemerkenswert“ oder dgl. bedeuten.

Es ist also genau darauf zu achten, in welchem Zusammenhang das Wort „signifikant“ gebraucht wird.

Außerdem kommt es (immer noch) vor, dass auch Wissenschaftler, die Ergebnisse statistischer Auswertungen vorstellen, leichtfertig statistisch signifikante mit inhaltlich signifikanten Einflüssen verwechseln.