

Konkretes Durchführen einer Inferenzstatistik

Die Frage ist, welche inferenzstatistischen Schlüsse bei einer *kontinuierlichen* Variablen - Beispiel: Reaktionszeit gemessen in ms - von der Stichprobe auf die Grundgesamtheit gezogen werden können. Zur Gegenüberstellung: Bei einer diskreten Variablen mit nur zwei möglichen Ausprägungen verwendeten wir hierzu die Binomialverteilung. Wie können wir nun die bereits bei der Binomialverteilung angewandten inferenzstatistischen Schlüsse auf eine kontinuierliche Variable übertragen? Auch bei einer kontinuierlichen Variablen sind drei Begriffe auseinander zuhalten: Stichprobe, Grundgesamtheit, Stichprobenkennwertverteilung.

1) Stichprobe: Folgendes konkrete empirische Ergebnis liege vor: Bei 64 Messungen von Reaktionszeiten ($n = 64$) wurde eine durchschnittliche Reaktionszeit von 105 ms erzielt.

2) Als nächstes brauchen wir eine Hypothese. Wir gehen von der Annahme aus, in "Wahrheit" sei der Mittelwert in der Population = 100 ms. Die Standardabweichung σ in der Population sei $\sigma = 20$ ms.

3) Die zentrale Frage lautet nun: Inwiefern lässt sich unsere Hypothese, dass nämlich in Wahrheit die Reaktionszeit 100 ms beträgt, durch unsere empirisch erhobene Stichprobe bestätigen? Anders ausgedrückt: Wie wahrscheinlich ist es, bei einem $\mu = 100$ rein zufällig einen Mittelwert von 105 ms zu bekommen? Von der Binomialverteilung her wissen wir bereits: Je unwahrscheinlicher diese Abweichung ist, umso weniger werden wir an der Nullhypothese festhalten. Die Frage wiederum, wie wahrscheinlich bzw. unwahrscheinlich unser Stichprobenergebnis bei Geltung unserer Hypothese ist, lässt sich nur beantworten, wenn wir die *Stichprobenkennwertverteilung der Mittelwerte* kennen.

Laut zentralem Grenzwerttheorem ist diese Stichprobenverteilung bei größeren Stichprobenumfängen (in der Praxis: ab $n \geq 30$) *normalverteilt*. Aufgrund *dieser* Verteilungsform der Stichprobenverteilung der Mittelwerte lässt sich nun die Wahrscheinlichkeit ausrechnen, bei einem μ von 100 rein zufällig ein empirisches Ergebnis von 105 ms oder mehr zu bekommen.

Um das zu verstehen, müssen wir uns nochmals einige Eigenschaften der Normalverteilung vergegenwärtigen: Bei einer Normalverteilung befinden sich ca. 68 % aller Werte im Bereich von ± 1 s. Im Bereich von ± 2 s befinden sich ca. 95.5 % aller Werte. Um die Bedeutung dieser Aussage zu verstehen, muss man sich vergegenwärtigen, dass die Normalverteilung genau genommen eine ganze *Familie* von Normalverteilungen umfasst. Die verschiedenen Normalverteilungen unterscheiden sich durch ihren Mittelwert und ihre Standardabweichung. Nehmen wir einmal an, der Mittelwert einer Verteilung sei 100 und die Standardabweichung sei 2, so liegen im Bereich von 98 (=100-2) und 102 (=100+2) 68% aller Fälle!! Nehmen wir weiters an, der Mittelwert sei wiederum 100, die Standardabweichung sei aber 4. Wieviel Prozent aller Werte liegen zwischen 100 ± 4 ? Natürlich wiederum 68%! Für *alle* Normalverteilungen gilt: Im Bereich von ± 1 s liegen 68 % aller Werte. Das ist deshalb so, weil die *Form der Verteilung* vom Mittelwert und der Standardabweichung unabhängig ist. Normalverteilungen mit *verschiedenen* Mittelwerten und *verschiedenen* Standardabweichungen haben die *gleiche Form*. Man kann daher bei allen Normalverteilungen angeben, wie viel Prozent aller Werte im Bereich von Einheiten, gemessen in Standardabweichungen, liegen.

Was ist nun eine Standardnormalverteilung? Wir bekommen sie, indem wir die Werte in einer Normalverteilung *z-transformieren*. Eine Z-Transformation ist nichts anderes als eine *lineare Transformation* der Daten. Unsere Daten im Falle der Stichprobenkennwertverteilung der Mittelwerte sind alle potentiell möglichen Stichprobenmittelwerte - einschließlich des Mittelwertes unserer empirisch erhobenen Stichprobe. Der Mittelwert dieser Kennwertverteilung ist gleich dem Populationsmittelwert gemäß der Nullhypothese (in unserem konkreten Beispiel beträgt $\mu_0 = 100$). Die Standardabweichung in der Stichprobenkennwertverteilung der Mittelwerte ist der Standardfehler des Mittelwertes, $\sigma_{\bar{x}}$.

Kommen wir nun zur Z-Transformation. Die Formel dafür lautet:

$$z = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}}$$

Der z-Wert in seiner allgemeinen Verwendung sagt uns, wie viel Einheiten von Standardabweichungen ein bestimmter Wert vom Mittelwert seiner Verteilung abweicht. Im Falle der Kennwertverteilung der Mittelwerte sagt uns die z-Transformation im speziellen, wie viele Standardabweichungen ein bestimmter Mittelwert vom Populationsmittelwert (dem Mittelwerte aller Mittelwerte) abweicht.

Für die z-transformierten Mittelwerte gilt aufgrund der oben angegebenen Formel: Ihr Mittelwert ist gleich dem Mittelwert der ursprünglichen Kennwertverteilung abzüglich des Wertes μ . Nachdem der Mittelwert der ursprünglichen Verteilung aber μ war, muss nach dieser Operation der neue Mittelwert gleich 0 sein (zur Erinnerung: Der Mittelwert der lineartransformierten Werte ist gleich der Lineartransformation des Mittelwerts der ursprünglichen Werte).

Die Division durch $\sigma_{\bar{x}}$ ergibt einen um $1/\sigma_{\bar{x}}$ veränderten neuen Standardfehler des Mittelwertes. Nachdem der Standardfehler der Kennwertverteilung gleich $\sigma_{\bar{x}}$ ist, ist der Standardfehler der z-transformierten Kennwertverteilung = 1!

Um nun die Wahrscheinlichkeit zu berechnen, bei einem μ von 100 rein zufällig ein empirisches Ergebnis von 105 ms oder mehr zu bekommen, müssen wir den empirisch ermittelten Stichprobenmittelwert z-transformieren, also:

$$z = \frac{105 - 100}{\sigma_{\bar{x}}}$$

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{20}{\sqrt{64}} = \frac{20}{8} = 2.5$$

z ist daher: $5 / 2,5 = 2!$

2 Standardabweichungen weicht unser empirisch erhobenes Ergebnis von einem $\mu = 100$ ab!

Aus der z-Tabelle lesen wir für einen z-Wert von 2 einen Flächenanteil von 0,9772 ab. Dies ist die Fläche in einer Standardnormalverteilung von $-\infty$ bis zu einem z-Wert von 2.

Dieser Flächenanteil ist der *gleiche* wie in einer Normalverteilung mit einem Mittelwert von 100 und einer Standardabweichung von 2,5.

Das bedeutet wiederum: Die Wahrscheinlichkeit aus einer Stichprobenkennwertverteilung mit einem $\mu = 100$ und einem $\sigma_{\bar{x}} = 2,5$ einen Stichprobenmittelwert *bis zu* 105 zu bekommen, ist die gleiche Wahrscheinlichkeit wie aus einer Standardnormalverteilung einen z-Wert *bis zu* 2 zu bekommen.

Dementsprechend ist dann die Überschreitungswahrscheinlichkeit, einen Stichprobenmittelwert von 105 oder noch größer zu bekommen: $1 - 0,9772 = 0,0228!$

Nun zur statistischen Hypothesenprüfung:

Statistische Hypothesenprüfung - ein - und zweiseitige Hypothesen

Bei der Hypothesenprüfung sind zwei verschiedene Typen von Alternativhypothesen zu unterscheiden: Entweder geht man davon aus, dass sich die Alternativhypothese in einer ganz bestimmten Richtung von der Nullhypothese unterscheidet.

Wir können beispielsweise vermuten, dass sich die Reaktionszeit bei Einnahme des Medikaments *verlangsamt* (Dass also nach Einnahme des Medikaments die Reaktionszeit größer wird). In diesem Falle verwenden wir eine Alternativhypothese mit einer bestimmten Richtung. Eine ähnlich gerichtete Alternativhypothese kam bereits im Beispiel des Falschspielers zum Tragen. Dort war die Alternativhypothese, dass der Spieler überzufällig häufig eine sechs würfeln kann. Nun besagt aber die Unschuldbehauptung - die Nullhypothese - vorerst nur, dass alle Abweichungen des statistischen Kennwerts in der Stichprobe (im Falle der Binomialverteilung ist dies der k-Wert und im Falle der Stichprobenverteilung der Mittelwert ist dies der Mittelwert der Stichprobe) vom Populationsmittelwert rein zufälliger Natur ist. Gehen wir nur von dieser Nullhypothese aus, so sind sowohl extreme Abweichungen im äußerst linken Bereich als auch extreme Abweichungen im äußerst rechten Bereich der Verteilung gleich unwahrscheinlich. Es würde uns beispielsweise ebenso sonderbar vorkommen, wenn ein Spieler bei 1000-maligem Würfeln nur einmal eine sechs würfeln könnte oder wenn unsere Probanden nach Einnahme des Medikaments plötzlich ein durchschnittliches Reaktionszeitvermögen von 20 ms aufweisen könnten.

Ist die Richtung der Alternative nicht bekannt, so spricht man von einer *zweiseitigen* Alternativhypothese.

Fassen wir den Unterschied zwischen ein- und zweiseitigen Hypothesen schematisch wie folgt zusammen:

1) einseitige Hypothese

a)

$H_0 : \mu = \mu_0$ (unsere Stichprobe stammt aus einer Population mit einem Populationsmittelwert μ gleich dem Mittelwert laut Geltung der Nullhypothese)

$H_1 : \mu > \mu_0$ (unsere Stichprobe stammt aus einer Population mit einem Populationsmittelwert μ größer als der laut Nullhypothese angenommene Mittelwert)

b)

$H_0 : \mu = \mu_0$

$H_1 : \mu < \mu_0$

Die grundsätzliche Logik für die statistische Entscheidungsfindung ist die gleiche wie sie uns bereits in Zusammenhang mit der Binomialverteilung begegnet ist.

Die Fläche rechts von einem Stichprobenmittelwert von 105 in unserer Kennwertverteilung der Mittelwerte beträgt, wie wir bereits wissen, 0,0228. Dies ist die Überschreitungswahrscheinlichkeit, aus einer Population mit einem Mittelwert von 100 und einer Standardabweichung von 2,5 rein zufällig einen Stichprobenmittelwert von *105 oder größer* zu erhalten. Warum wir an der Überschreitungswahrscheinlichkeit und nicht an der Einzelwahrscheinlichkeit, gerade einen Stichprobenmittelwert von 105 zu bekommen, interessiert sind, hat den folgenden Grund:

Halten wir einen Stichprobenmittelwert von 105 zu unwahrscheinlich, um noch die Nullhypothese bestätigen zu können, so impliziert dies, dass wir bei noch größeren Stichprobenmittelwerten (also 106, 107 usw.) die Nullhypothese erst recht verwerfen müssen!

Denn die Fläche in der Verteilung der Stichprobenmittelwerte wird ja an ihrem äußeren rechten (oder linken) Rand immer kleiner.

Um also das Risiko abschätzen zu können, die Nullhypothese fälschlicherweise zu verwerfen, benötigen wir die Wahrscheinlichkeit nicht nur für den Stichprobenmittelwert von 105 unter Geltung der Nullhypothese, sondern die Gesamtwahrscheinlichkeit für alle möglichen Stichprobenmittelwerte ab 105. Dies ist eben die Überschreitungswahrscheinlichkeit.

Was wir noch benötigen, ist eine Konvention, ab der die Nullhypothese verworfen wird. Dies ist das so genannte Signifikanzniveau. Ist unsere berechnete Überschreitungswahrscheinlichkeit ≤ 0.05 , so verwerfen wir die Nullhypothese.

Wir sagen, wir haben in diesem Falle ein „signifikantes“ (gewissermaßen auffälliges) Ergebnis.

Zum gleichen Ergebnis kommen wir auch, indem wir den errechneten z-Wert für das Stichprobenergebnis von 105 mit dem so genannten kritischen z-Wert vergleichen. Unter dem kritischen z-Wert wird jener z-Wert am äußersten linken oder rechten Rand der Verteilung verstanden, dessen Überschreitungswahrscheinlichkeit einer Fläche von 0,05 entspricht.

Dieser Flächenanteil ist der *Verwerfungsbereich* der Nullhypothese.

Dies ist ein z von +1,65 bzw. von -1,65

Ist unser errechnetes $z \geq +1,65$ oder $\leq -1,65$, so verwerfen wir die Nullhypothese. Dies lässt sich auch so zusammenfassen:

Ist $|z_{\text{errechnet}}| \geq |z_{\text{kritisch}}| \rightarrow H_1$

2) zweiseitige Hypothese

$$H_0 : \mu = \mu_0$$

Dies liest sich so: Der Populationsmittelwert der Stichprobe ist gleich dem Populationsmittelwert μ_0 laut unserer Theorie.

$$H_1 : \mu \text{ ungleich } \mu_0$$

Die zweiseitige Alternativhypothese besagt nur, dass das Medikament in irgendeiner Richtung die Reaktionszeit beeinflusst, bzw. ungeachtet dessen, ob die Reaktionszeit beschleunigt oder verlangsamt wird. In diesem Falle wird nur getestet, ob sich die Reaktionszeit prinzipiell bei Einnahme des Medikaments von der normalen Reaktionszeit unterscheidet. Auch bei einer zweiseitigen Hypothesenprüfung gehen wir zunächst - wie bereits bekannt - von der Stichprobenverteilung der Mittelwerte aus, die wir z-transformieren.

Im Unterschied zur einseitigen Hypothese liegt der Verwerfungsbereich der Nullhypothese in den beiden extrem linken und rechten Seiten der Verteilung. Bleiben wir bei einem Signifikanzniveau von insgesamt $p \leq 0,05$, also von 5%, so verteilen sich die Wahrscheinlichkeiten des Verwerfungsbereichs der Nullhypothese links und rechts der Verteilung auf je 2,5%.

Daraus folgt: *Wir entscheiden uns bei einer zweiseitigen Hypothesenprüfung für die Alternativhypothese, wenn die Überschreitungswahrscheinlichkeit unseres Stichprobenmittelwertes kleiner als 0,025 ist.*

Zum gleichen Ergebnis kommen wir auch, indem wir den errechneten z-Wert für das Stichprobenergebnis von 105 mit dem so genannten kritischen z-Wert vergleichen. Unter dem kritischen z-Wert wird im Falle einer zweiseitigen Hypothesenprüfung jener z-Wert am äußerst linken oder rechten Rand der Verteilung verstanden, dessen Überschreitungswahrscheinlichkeit einer Fläche von 0,025 entspricht.

Denn im Bereich von

$$\mu - 1.96 \cdot \sigma_{\bar{x}} \leq \bar{x} \leq \mu + 1.96 \cdot \sigma_{\bar{x}}$$

liegen 95% aller Stichprobenmittelwerte

(Zur Erinnerung: $\mu = \bar{\bar{x}}$)

Wie kommen wir nun zu der Behauptung, dass in den Schranken von $\pm 1,96$ 95% aller Stichprobenmittelwerte liegen?

Betrachten wir dazu vorerst nur die positive Schranke von $+1,96$. Oberhalb dieser Schranke (inklusive der Schranke) liegen 2,5 % aller Stichprobenmittelwerte. Das sind 0,025. Das bedeutet aber zugleich, dass unterhalb von 1,96 97,5 % aller Mittelwerte = 0,975 liegen.

Dies entspricht der Fläche in der Standardnormalverteilung von -unendlich bis 1,96. Diesen Flächenanteil entnehmen wir der Standardnormalverteilungstabelle. Oberhalb von $z = 1,96$ liegen daher $1 - 0,975$ aller Fälle = 0,025.

Unterhalb der Schranke von -1,96 (links von -1,96) liegt (man beachte: die Standardnormalverteilung ist symmetrisch!) ein Flächenanteil von 0,025.

Beide Flächenanteile zusammen ergeben 0,05, was einem Signifikanzniveau von 5% entspricht.

Zusammenfassung (Faustregeln):

- 1) Wir erheben eine Stichprobe mit dem Mittelwert von 105 ms.
- 2) Der Mittelwert und die Standardabweichung in der Population sind bekannt. $\mu=100$; $\sigma = 20$.

3) Wir berechnen den Standardfehler:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{20}{\sqrt{64}} = 20/8 = 2.5$$

4) Wir berechnen die Abweichung des beobachteten Mittelwerts von dem Mittelwert laut Annahme der Nullhypothese mit Hilfe der z-Transformation.

$$z = (105-100)/2,5 = 2 \quad [= z_{\text{errechnet}}]$$

5) Wir entscheiden uns entweder für eine einseitige oder für eine zweiseitige Hypothesenprüfung

a) einseitig:

Der Fläche, die am äußerst rechten bzw. äußerst linken Rand der Verteilung 5% abschneidet (je nach Richtung der Hypothesen), entspricht ein kritischer z-Wert bei einem 5% Signifikanzniveau von 1,65 bzw. von - 1,65

Ist $|z_{\text{errechnet}}| \geq |1,65| \rightarrow$ Entscheidung für H_1

b) zweiseitig:

Der Fläche, die an den beiden linken und rechten Rändern der Verteilung insgesamt 5% abschneidet, entspricht ein kritischer z-Wert von 1,96

Ist $|z_{\text{errechnet}}| \geq |1,96| \rightarrow$ Entscheidung für H_1

Nun noch eine letzte, allerdings wichtige Nachbemerkung zur Hypothesenprüfung: Da inferenzstatistische Tests üblicherweise nicht per Hand, sondern mit Hilfe von Statistikprogrammen berechnet werden, erfolgt die Hypothesenprüfung in der Praxis nicht über den Vergleich der z-Werte "z-kritisch" und "z-errechnet".

Die Ausgabelisten der Computerprogramme geben immer direkt die Überschreitungswahrscheinlichkeit an. Dies ist die Wahrscheinlichkeit, die beobachtete Mittelwertsabweichung (in unserem Beispiel!) oder eine noch größere zu bekommen. Statt aber diese Überschreitungswahrscheinlichkeit im Falle einer zweiseitigen Hypothesenprüfung mit $\alpha/2$ zu vergleichen (Regel: Wenn $p \leq \alpha/2 \rightarrow H_1$), wird die Überschreitungswahrscheinlichkeit verdoppelt (da $2p \leq \alpha \equiv p \leq \alpha/2$). Durch diese Verdoppelung der Überschreitungswahrscheinlichkeit wird erreicht, dass der Benutzer bei der Signifikanzprüfung im Falle eines zweiseitigen Tests die

vom Programm ausgegebene Überschreitungswahrscheinlichkeit direkt mit dem Signifikanzniveau α vergleichen kann (und eben nicht mit $\alpha/2$).

(Man beachte: Will man einseitig testen, so muss man die vom Programm ausgegebene – verdoppelte - Überschreitungswahrscheinlichkeit *halbieren!*)

Für eine Abweichung von 5ms vom Populationsmittelwert von $\mu = 100$ erhält man beispielsweise als Ergebnis:

two-tailed error probability = 0,0456 (= 2 x 0,0228)

Diese Überschreitungswahrscheinlichkeit wird direkt mit dem Signifikanzniveau von $p=0,05$ verglichen

Ist diese error probability $\leq 0,05$ -> Entscheidung für H_1 !