

Grundlagen der Inferenzstatistik, Teil 2: Hypothesentesten

- 1 Hypothesen
- 2 α und β -Fehler
- 3 Logik des Hypothesentestens

Einführende Literatur

-  Bortz, J. & Schuster, Ch. (2010). *Statistik für Human- und Sozialwissenschaftler* (7. Auflage). Berlin: Springer. [Kap. 7]
-  Diehl, J. M. & Arbinger, R. (2001). *Einführung in die Inferenzstatistik* (3. Auflage). Eschborn bei Frankfurt: Klotz Verlag. [Kap. 3]
-  Eid, M., Gollwitzer, M. & Schmitt, M. (2010). *Statistik und Forschungsmethoden*. Weinheim: Beltz. [Kap. 8]

Hypothesen

- **Wissenschaftliche Hypothesen (Forschungshypothesen)** sind allgemeingültigere (d.h. über den Einzelfall hinausgehende), theoretische begründete Behauptungen (Aussagen, Annahmen). Sie müssen empirisch überprüfbar und prinzipiell widerlegbar (falsifizierbar) sein.
- Beispiele für **Unterschiedshypothesen** sind:
 - A Frauen sind kreativer als Männer.
 - B Der Anteil der Studierenden, die ihr Bachelor-Studium an der Universität Osnabrück in der Regelstudienzeit absolvieren, unterscheidet sich in den Studienfächern BWL, Biologie und Psychologie.
 - C Das Trainingsprogramm T für Studierende zur Reduzierung von Prüfungsängstlichkeit ist wirksam.
- Beispiele für **Zusammenhangshypothesen** sind:
 - D Je mehr gewalttätige Computerspiele Kinder spielen, desto aggressiver sind sie.
 - E Es besteht ein Zusammenhang zwischen der beruflichen Leistung und der Intention, zu kündigen (und sich einen neuen Arbeitsplatz zu suchen).
- Forschungshypothesen können **gerichtet** oder **ungerichtet** formuliert sein. Bei den obigen Beispielen sind **A, C** und **D** gerichtet, **B** und **E** sind ungerichtet formuliert.

Hypothesen

- Inferenzstatistisch zu überprüfende Hypothesen beziehen sich immer auf eine Population! Auf der Basis der Ergebnisse von Stichproben können wir Hypothesen niemals mit Sicherheit beweisen oder mit Sicherheit widerlegen. Vielmehr sind nur Wahrscheinlichkeitsaussagen möglich, die immer mit einer gewissen Unsicherheit behaftet sind.
- Um Hypothesen einer statistischen Prüfung zugänglich zu machen, müssen die Forschungshypothesen in **statistische Hypothesen** überführt werden.
- Dass, was wir in der Forschungshypothese behaupten, wird dabei (meistens) in der **Alternativhypothese**, abgekürzt H_1 , formalisiert. Das Gegenteil davon wird in der **Nullhypothese**, abgekürzt H_0 , behauptet.
- **Beispiel:** „Frauen und Männer unterscheiden sich in ihrer Kreativität.“

$$H_0: \mu_{\text{Frauen}} = \mu_{\text{Männer}}$$

$$H_1: \mu_{\text{Frauen}} \neq \mu_{\text{Männer}}$$

Der Mittelwert der Kreativität von Frauen und Männern ist in der Population gleich.

Der Mittelwert der Kreativität von Frauen unterscheidet sich von dem Mittelwert der Männer in der Population.

Hypothesen

- Man kann das obige Hypothesenpaar für die Forschungshypothese „Frauen und Männer unterscheiden sich in ihrer Kreativität“ auch wie rechts schreiben, wobei deutlicher wird, warum die Nullhypothese so heißt:

$$\left. \begin{array}{l} H_0: \mu_{\text{Frauen}} = \mu_{\text{Männer}} \\ H_1: \mu_{\text{Frauen}} \neq \mu_{\text{Männer}} \end{array} \right\} \Leftrightarrow \left\{ \begin{array}{l} H_0: \mu_{\text{Frauen}} - \mu_{\text{Männer}} = 0 \\ H_1: \mu_{\text{Frauen}} - \mu_{\text{Männer}} \neq 0 \end{array} \right.$$

- Im obigen Beispiel war die Forschungshypothese **ungerichtet** und die statistische Hypothese **zweiseitig** formuliert. Bei der **gerichteten** Hypothese „Frauen sind kreativer als Männer“ kann man entsprechend **einseitig** die beiden folgenden statistischen Hypothesen formulieren:

$$\left. \begin{array}{l} H_0: \mu_{\text{Frauen}} \leq \mu_{\text{Männer}} \\ H_1: \mu_{\text{Frauen}} > \mu_{\text{Männer}} \end{array} \right\} \Leftrightarrow \left\{ \begin{array}{l} H_0: \mu_{\text{Frauen}} - \mu_{\text{Männer}} \leq 0 \\ H_1: \mu_{\text{Frauen}} - \mu_{\text{Männer}} > 0 \end{array} \right.$$

- Für die ungerichtete **korrelative** Hypothese „Es besteht ein Zusammenhang zwischen der beruflichen Leistung und der Kündigungsintention“ resultiert zweiseitig entsprechend:

$$H_0: \rho = 0$$

Die Korrelation zwischen der beruflichen Leistung und der Kündigungsintention ist in der Population gleich 0.

$$H_1: \rho \neq 0$$

... ist in der Population ungleich 0.

Hypothesen

- In allen obigen Beispielen wurde in der H_0 immer behauptet, dass der Parameter in der Population 0 ist (d.h. dass kein Mittelwertsunterschied oder kein Zusammenhang besteht). Dies ist häufig, aber nicht immer, der Fall.
- **Beispiel 1:** Die Forschungshypothese lautet "Abiturienten sind intelligenter sind als die Durchschnittsbevölkerung." (Die mittlere Intelligenz in dem verwendeten Intelligenztest betrage in der Durchschnittsbevölkerung 100.) Dann lautet die einseitige statistische Hypothese:

$$H_0: \mu \leq 100$$

$$H_1: \mu > 100$$

- **Beispiel 2:** Es sei aus Metaanalysen bekannt, dass die Korrelation zwischen der allgemeinen Intelligenz und dem Berufserfolg 0.50 beträgt. Die Forschungshypothese lautet „Die soziale Intelligenz korreliert mit dem Berufserfolg höher oder niedriger als die allgemeine Intelligenz“. Die zweiseitige statistische Hypothese lautet:

$$H_0: \rho = 0.50$$

$$H_1: \rho \neq 0.50$$

Hypothesen

- In allen obigen Beispielen war die Forschungshypothese immer in der Alternativhypothese repräsentiert. Dies ist meist, aber nicht immer, der Fall.
- **Beispiel:** Forschungshypothese: „Männer und Frauen unterscheiden sich nicht in ihrer Kreativität.“ mit der zweiseitigen statistischen Hypothese

$$H_0: \mu_{\text{Frauen}} = \mu_{\text{Männer}}$$

$$H_1: \mu_{\text{Frauen}} \neq \mu_{\text{Männer}}$$

- Die getroffenen Unterscheidungen bei Hypothesen (gerichtet vs. ungerichtet, ein- vs. zweiseitig, Forschungshypothese steckt in der H_0 vs. H_1) haben Konsequenzen in Bezug auf die Hypothesentestung und damit auch für die statistische Entscheidung.

α und β -Fehler

- Am Ende wird auf der Basis der Daten der Stichprobe mittels eines **statistischen Tests** eine Entscheidung über die Hypothesen getroffen: Die H_0 wird entweder zurückgewiesen und die H_1 angenommen oder die H_0 wird beibehalten und die H_1 verworfen.
- Wird die H_1 angenommen, so spricht man auch von einem (**statistisch**) **signifikanten Ergebnis**. Im **Beispiel**: „Frauen sind statistisch signifikant kreativer als Männer“.
- Beide Entscheidungen, die H_1 anzunehmen oder sie zurückzuweisen, können entweder korrekt oder fehlerhaft sein. Die möglichen Konstellationen verdeutlicht die folgende Vierfeldertafel:

Realität: Richtig ist ...	Statistische Entscheidung aufgrund des Tests für ...	
	H_0 (z.B. kein Unterschied)	H_1 (z.B. Unterschied existiert)
H_0 (z.B. kein Unterschied)	$1 - \alpha$ (korrekte Beibehaltung der H_0)	α -Fehler (Fehler 1. Art)
H_1 (z.B. Unterschied existiert)	β -Fehler (Fehler 2. Art)	$1 - \beta$ (korrekte Zurückweisung der H_0)

Power

α und β -Fehler

- Die beiden Fehler, die man beim Hypothesentesten machen kann sind ...
 - der α -Fehler (=Fehler erster Art): Man entscheidet sich für die Alternativhypothese, obwohl sie falsch ist, d.h. die Nullhypothese korrekt ist.
 - der β -Fehler (=Fehler zweiter Art): Man behält die Nullhypothese bei, obwohl die Alternativhypothese richtig ist.
- Am Beispiel der Hypothese „Das Trainingsprogramm T für Studierende zur Reduzierung von Prüfungsängstlichkeit ist wirksam“ (TG = Trainingsgruppe, KG = Kontrollgruppe) mit

$$H_0: \mu_{TG} \geq \mu_{KG}$$

$$H_1: \mu_{TG} < \mu_{KG}$$

würde man also einen α -Fehler begehen, wenn die Entscheidung aufgrund des Testens ergäbe, dass das Trainingsprogramm wirksam ist (sich also für die H_1 entscheiden), obwohl dies gar nicht der Fall ist. Einen β -Fehler würde man begehen, wenn man aufgrund des statistischen Testens an einer Stichprobe zu dem Ergebnis käme, dass das Trainingsprogramm unwirksam wäre (sich also für die H_0 entscheiden), obwohl dies nicht stimmt. Beide Fehler haben hier (und auch sonst) ganz unterschiedliche Konsequenzen.

α und β -Fehler

- Meist liegt ja die Forschungshypothese in der H_1 und behauptet, dass es bestimmte Effekte gibt, also sich Mittelwerte unterscheiden, ein Zusammenhang besteht etc.
- Um zu verhindern, dass fälschlicherweise solche Unterschiede oder Zusammenhänge angenommen werden, wo in Wirklichkeit keine sind, will man beim statistischen Testen meist vor allem das Risiko für einen α -Fehler klein halten.
- Tatsächlich kann man im Rahmen des statistischen Testens darauf Einfluss nehmen, wie groß dieses Fehlerrisiko ist. Man setzt also die Wahrscheinlichkeit, einen α -Fehler zu machen, auch als **Irrtumswahrscheinlichkeit** oder **Signifikanzniveau** bezeichnet, auf einen „kleinen“ Wert fest. Per Konvention wird ein Risiko von 5% (oder 1%) als „klein“ erachtet.
- Das Signifikanzniveau ist vor der Testung (d.h. vor der Kenntnis der Ergebnisse der Studie) zu spezifizieren.
- Man bezeichnet diese Irrtumswahrscheinlichkeit auch als α (-Niveau) und konstruiert die Testprozedur so, dass α nicht größer als z.B. 0.05 ist. Sie ist die Wahrscheinlichkeit, mit der fehlerhafter Weise die H_1 angenommen wird und damit ein statistisch signifikantes Ergebnis resultiert.

α und β -Fehler

- Deutung am **Beispiel**: Wenn unser Training tatsächlich unwirksam ist, und man führt unabhängig voneinander 100 Studien durch, die dies untersuchen, wird man bei einem α von 0.05 in ca. 5 (% der) Studien ein statistisch signifikantes Ergebnis erhalten, also einen α -Fehler begehen.
- Man könnte nun dieses Risiko, fälschlicherweise ein statistisch signifikantes Ergebnis zu erhalten, verringern, indem man α , statt auf 5% auf 1% oder noch kleiner z.B. auf 0.00001% festsetzt. Das Problem ist aber, dass mit der Verringerung des Risikos für einen α -Fehler das Risiko für einen β -Fehler steigt.
- Häufig wird für statistisch signifikante Ergebnisse ein Stern ($\alpha = 0.05$) bzw. zwei Sterne ($\alpha = 0.01$) hochgestellt hinter die zur prüfende Statistik geschrieben oder/und statt α p geschrieben. Zum Beispiel findet man folgende Ergebnisbeschreibungen:
 - „Es besteht ein Zusammenhang zwischen der beruflichen Leistung und der Kündigungsintention, $r = 0.12$, $p < .05$.“
 - oder stattdessen $r = 0.12^*$

Logik des Hypothesentestens

- An folgendem fiktiven Beispiel der Prüfung der Hypothese, dass ein Populationsmittel einen bestimmten Wert aufweist und der (selten realisierten) Annahme, dass dabei die Populationsvarianz bekannt ist, wollen wir das Vorgehen beim statistischen Testen verdeutlichen.

➤ **Beispiel:** Im Jahre 2009 wurde mit allen Abiturienten in Niedersachsen ein Intelligenztest durchgeführt. Es resultierten normalverteilte Werte mit einem Mittelwert von 108.0 und einer Standardabweichung von 13.0. An einer Stichprobe von 200 Abiturienten soll die Hypothese untersucht werden, ob sich das Intelligenzniveau im Jahre 2010 verändert hat. Dabei wird davon ausgegangen, dass sich die Streuung nicht verändert hat. In der Stichprobe resultiert ein Mittelwert von 110.2 (und eine Standardabweichung von 12.8).

- Man sieht, dass deskriptiv der Mittelwert im Jahr 2010 höher ist als im Jahr 2009. Da dies aber auch auf die gezogene Stichprobe zurückzuführen sein kann (Stichprobenfehler) testen wir inferenzstatistisch. In diesem Fall lautet unsere ungerichtete Forschungshypothese: „Die mittlere Intelligenz im Jahre 2010 ist nicht gleich 108“.

- Die zweiseitigen statistischen Hypothesen lauten:

$H_0: \mu = 108$ (Der Mittelwert der Intelligenz ist im Jahre 2010 in der Population 108.)

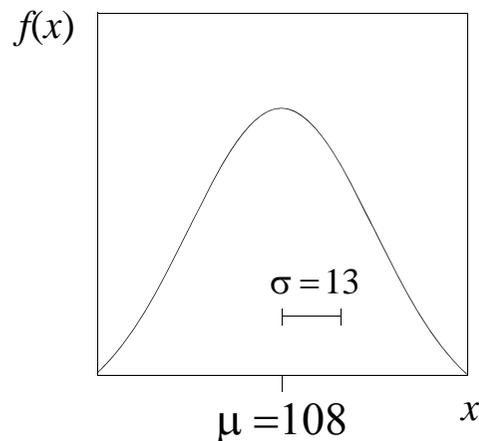
$H_1: \mu \neq 108$ (Der Mittelwert der Intelligenz ist im Jahre 2010 in der Population ungleich 108.)

Logik des Hypothesentestens

- **Beispiel:** Im Jahre 2009 wurde mit allen Abiturienten in Niedersachsen ein Intelligenztest durchgeführt. Es resultierten normalverteilte Werte mit einem Mittelwert von 108.0 und einer Standardabweichung von 13.0. An einer Stichprobe von 200 Abiturienten soll die Hypothese untersucht werden, ob sich das Intelligenzniveau im Jahre 2010 verändert hat. Dabei wird davon ausgegangen, dass sich die Streuung nicht verändert hat. In der Stichprobe resultiert ein Mittelwert von 110.2 (und eine Standardabweichung von 12.8)

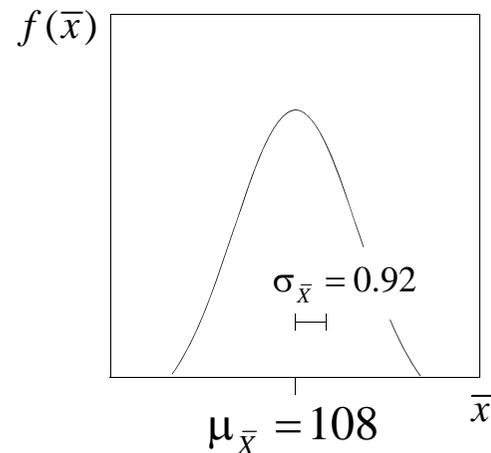
Verteilung der Intelligenz
in der Population 2009

$$X \sim N(\mu, \sigma) \\ = N(108, 13)$$



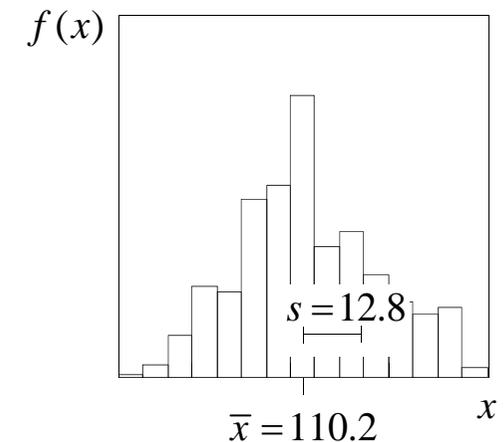
Stichprobenkennwerte-
verteilung des Mittels für
Stichproben der Größe n

$$\bar{X} \sim N(\mu_{\bar{X}}, \sigma_{\bar{X}}) = N(\mu, \sigma / \sqrt{n}) \\ = N(108, 0.92)$$



Verteilung des Merkmals
in der Stichprobe 2010

$$\bar{x} = 110.2 \\ s = 12.8$$



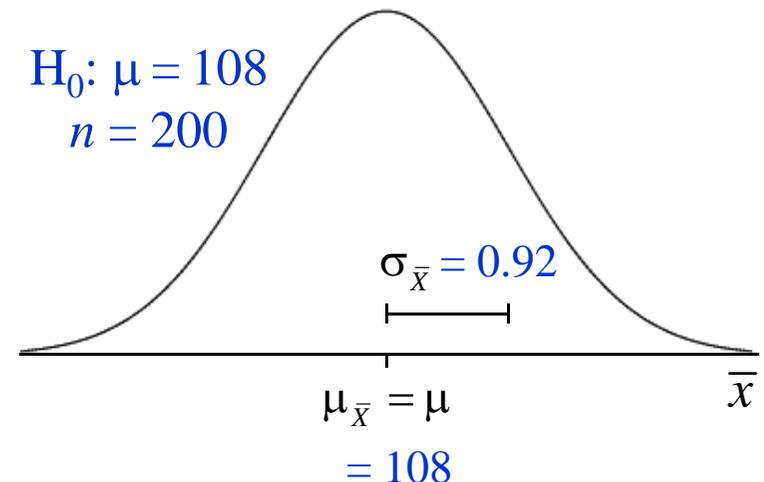
Logik des Hypothesentestens

- Wir wissen bereits, dass bei einer hinsichtlich des Merkmals X normalverteilten Population mit bekannter Streuung σ sich die Stichprobenkennwerteverteilung des Mittelwertes ebenfalls normalverteilt mit Mittelwert μ und $\sigma_{\bar{x}} = \sigma / \sqrt{n}$.

- Wenn nun die Behauptung richtig ist, dass die Nullhypothese H_0 in der Population gilt, so sieht die normalverteilte Stichprobenkennwerteverteilung wie nebenstehend aus.

- Mit den Daten des Beispiels ist dann

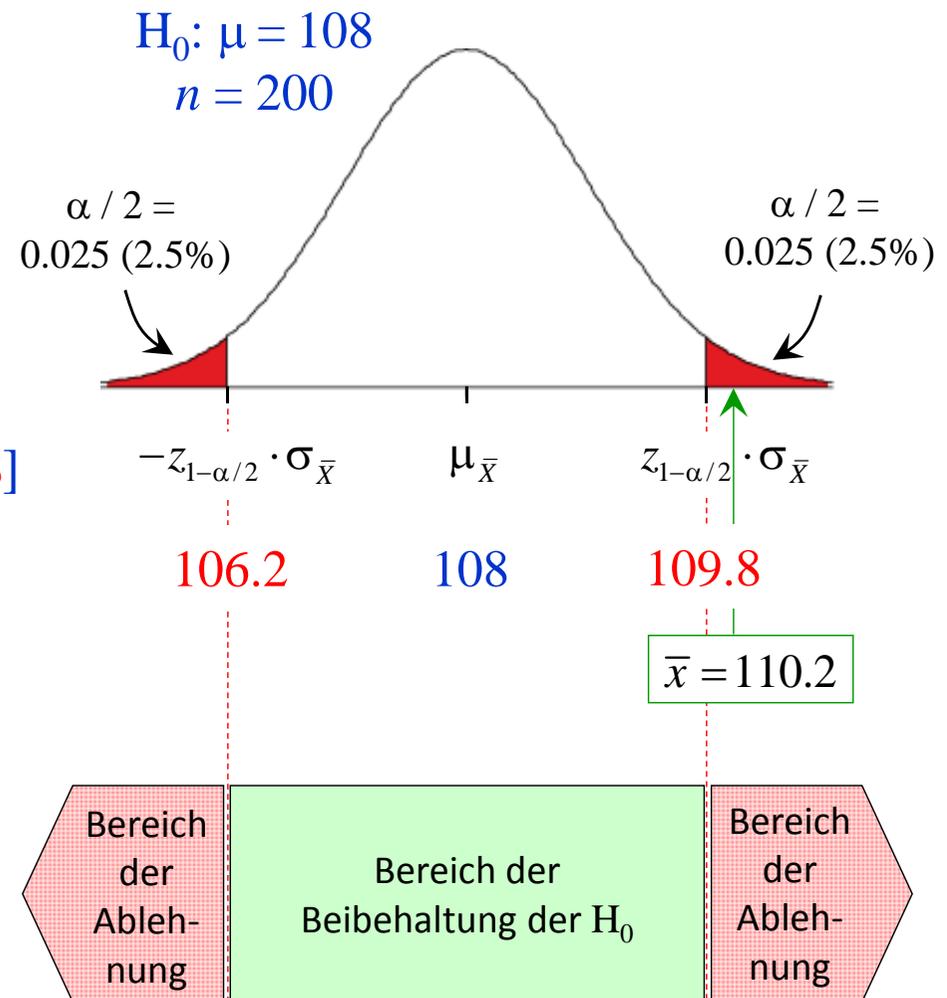
$$\mu_{\bar{x}} = \mu = 108 \quad \sigma_{\bar{x}} = \sigma / \sqrt{n} = 13 / \sqrt{200} = 0.92$$



- Wenn die H_0 richtig ist, ist es wahrscheinlich, dass der Mittelwert der Stichprobe \bar{x} nahe bei $\mu_{\bar{x}}$ liegt. Mit zunehmendem Abstand von $\mu = \mu_{\bar{x}}$ wird es immer unwahrscheinlicher, dass dort ein Stichprobenmittelwert \bar{x} liegt, der aus der Population mit dem Mittelwert μ stammt. Liegt \bar{x} „sehr weit“ von μ entfernt, so würden wir schließen, dass es „sehr unwahrscheinlich“ ist, dass er aus einer Population mit dem Mittelwert μ stammt und daher die Nullhypothese, in diesem Fall $\mu = 108$, zurückweisen. Ein Restrisiko, dass wir dabei die falsche Entscheidung treffen (also einen α -Fehler begehen), bleibt immer bestehen.

Logik des Hypothesentestens

- Welches Restrisiko wir bei einem extrem von $\mu_{\bar{x}}$ abweichenden \bar{x} bereit sind einzugehen, legen wir im Signifikanzniveau α fest.
- Setzen wir $\alpha = 0.05$, so bedeutet dies, dass Mittelwerte, die in den Bereich der 2.5% extremsten Werte auf beiden Seiten der Verteilung fallen, zur Zurückweisung der H_0 führen.
- Im **Beispiel** ist dieser Bereich das Intervall
$$\mu_{\bar{x}} \pm z_{1-\alpha/2} \cdot \sigma_{\bar{x}} = 108 \pm 1.96 \cdot 0.92 \Rightarrow [106.2, 109.8]$$
- Überschreitet der empirisch ermittelte Mittelwert \bar{x} den oberen kritischen Wert des Intervalls oder unterschreitet er den unteren kritischen Wert, so wird die H_0 zurückgewiesen und geschlossen, dass der Mittelwert statistisch signifikant vom Wert 108 abweicht.
- Im **Beispiel** ist dies der Fall: $\bar{x} = 110.2 > 109.8$



Logik des Hypothesentestens

- Rechnerisch vereinfacht sich der Test, wenn wir den Mittelwert z -transformieren und dann die standardnormalverteilte Stichprobenkennwerteverteilung heranziehen.

- Im **Beispiel** ist dann die **Prüfgröße** (empirischer z -Wert)

$$z = \frac{\bar{x} - \mu_{\bar{x}}}{\sigma_{\bar{x}}} = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} = \frac{110.2 - 108.0}{13 / \sqrt{200}} = \frac{2.2}{0.92} = 2.39$$

- Dann ist (wegen der Symmetrie der NV um 0) nur noch zu prüfen, ob der Betrag dieses z -Wertes größer ist als der kritische z -Wert von $z_{crit} = z_{1-\alpha/2}$.

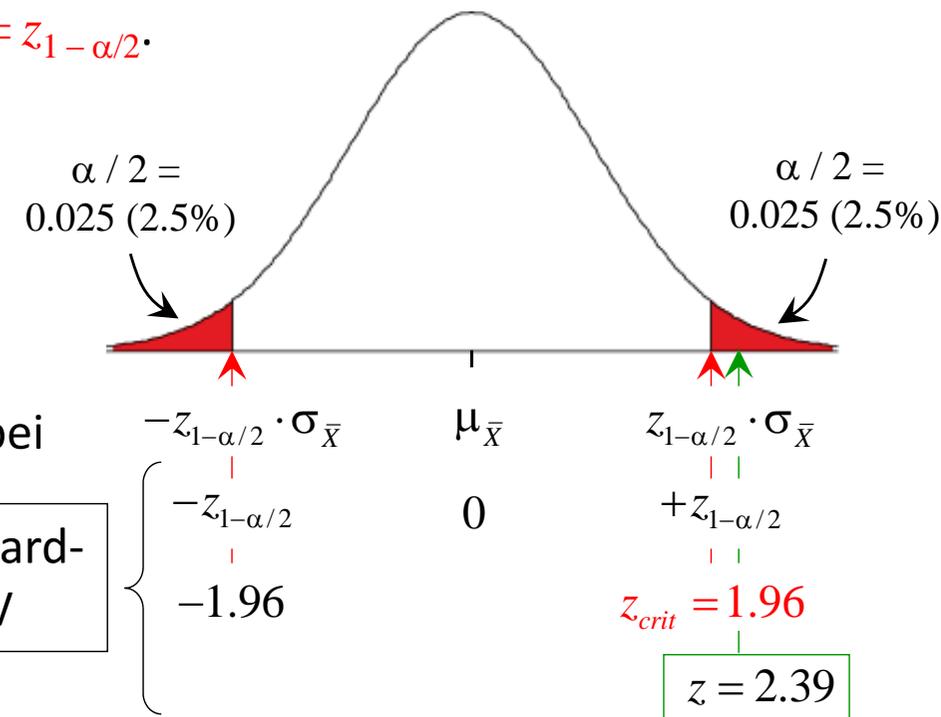
- Für $\alpha = 0.05$ gilt:

$$z_{crit} = z_{1-\alpha/2} = z_{0.975} = 1.96 \quad \text{und}$$

$$|z| = |2.39| = 2.39 > 1.96$$

- Wir schließen also (wie oben), dass der Mittelwert bei einem Signifikanzniveau von 5% statistisch signifikant vom Wert 108 abweicht: Das Intelligenzniveau der Abiturienten hat also statistisch signifikant zugenommen.

Standard-NV



Logik des Hypothesentestens

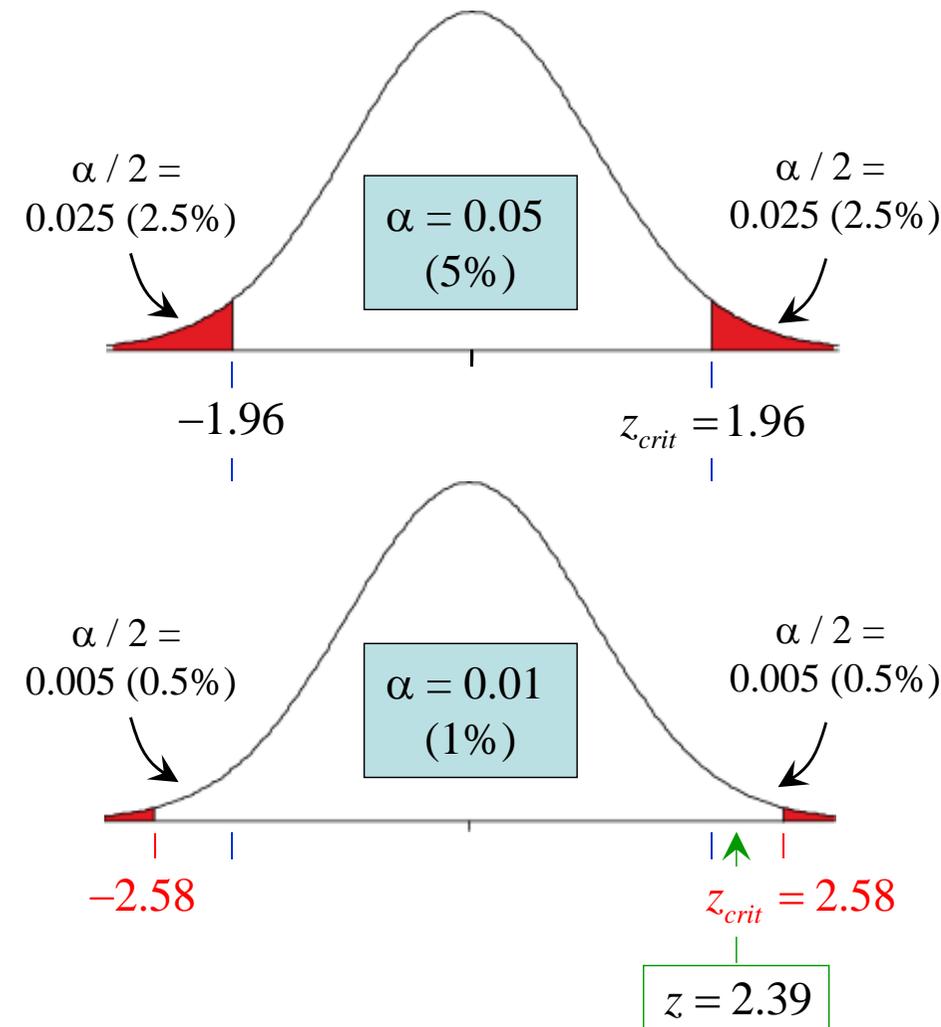
- Was hätte der **z-Test** ergeben, wenn wir das Signifikanzniveau auf $\alpha = 0.01$ festgelegt hätten?
- Dann ist der z -Wert des Stichprobenmittelwerts der gleiche, im **Beispiel**: $z = 2.39$
- Der kritische z -Wert aber hätte sich verändert:

$$z_{crit} = z_{1-\alpha/2} = z_{0.995} = 2.58$$

- Bei diesem kritischen Wert gilt dann:

$$|z| = 2.39 < 2.58$$

- Die Entscheidung fällt hier also anders aus:
Wir behalten die Nullhypothese bei.



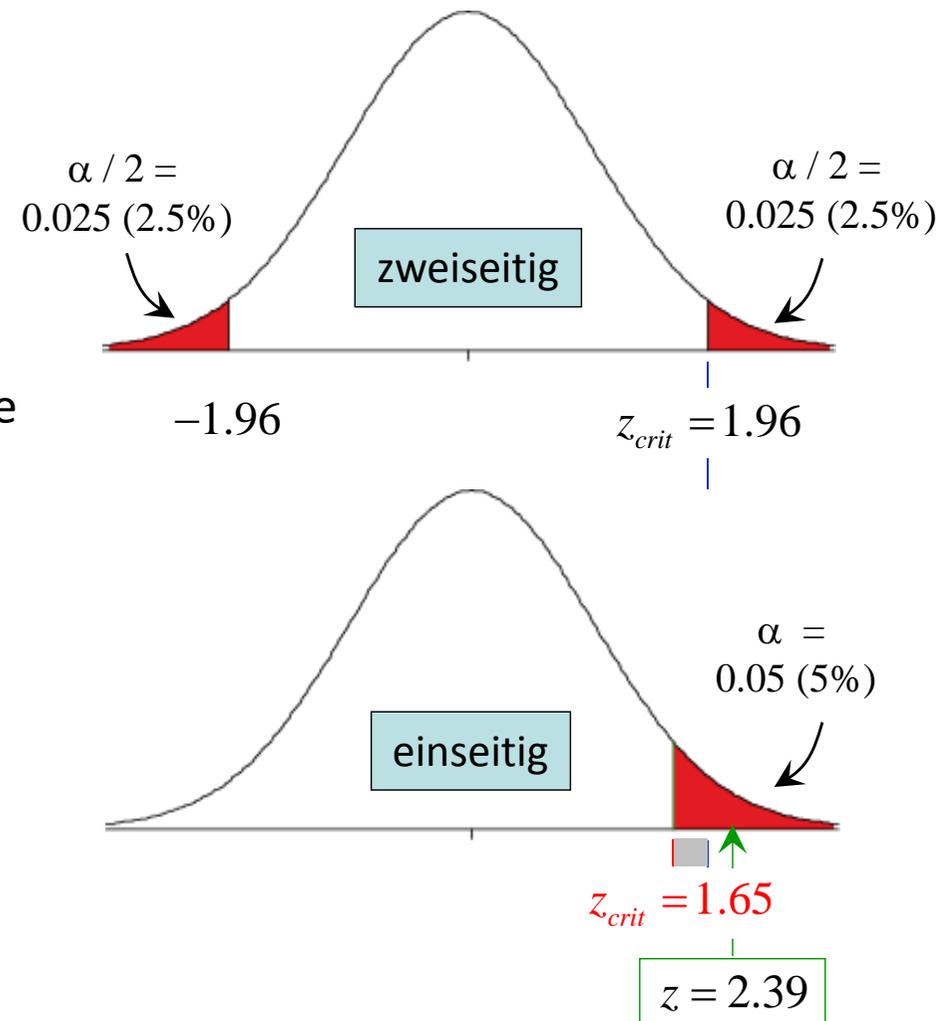
Logik des Hypothesentestens

- Bei der Interpretation von statistisch signifikanten und insignifikanten Ergebnissen ergibt sich eine gewisse Asymmetrie in der Deutung der Befunde.
- Resultiert ein statistisch signifikantes Ergebnis (da z.B. wie oben $|z| > z_{crit}$), so weisen wir die H_0 zurück, die H_0 wird verworfen. Wir schließen, dass die Nullhypothese falsch ist (und nehmen die Alternativhypothese an und schließen also z.B., dass das Intelligenzniveau der Abiturienten zugenommen hat).
- Ergibt sich ein statistisch insignifikantes Ergebnis, so ist die Situation schwieriger: Hier können wir in der bisher dargestellten Logik des Signifikanztestens nicht sagen, dass die Nullhypothese stimmt oder richtig ist (also z.B. dass beide Jahrgänge gleich intelligent sind), sondern formulieren vorsichtiger, dass wir die **Nullhypothese beibehalten** (und die Alternativhypothese zurückweisen). Man konnte die H_0 weder belegen noch zurückweisen.
- Das Problem ist dabei, dass wir die Wahrscheinlichkeit, die Nullhypothese fälschlicherweise beibehalten zu haben, β , (im Gegensatz zu α) nicht kontrolliert haben. Diese Wahrscheinlichkeit kann unakzeptabel hoch sein.
- Um auch eine Nullhypothese bestätigen zu können, benötigen wir eine erweiterte Testlogik, auf die wir später zurückkommen.

- Wie soll man entscheiden, welches Signifikanzniveau α man wählen soll?
- Zunächst ist es sinnvoll, α „klein“ zu wählen, wenn man vor allem verhindern will, dass Effekte behauptet werden, die gar nicht existieren.
 - Die in den Standard-Testungen am häufigsten anzutreffende Festlegung von $\alpha = 0.05$ ist eine reine Konvention. Eine inhaltliche oder statistische Begründung für die Wahl von genau dem Wert 0.05 (statt z.B. 0.04, 0.07 oder 0.55555) gibt es nicht.
 - Wählt man eine kleinere Irrtumswahrscheinlichkeit als 0.05, z.B. 0.01, so steigt die Wahrscheinlichkeit eines β -Fehlers und die Chance, einen Effekt zu finden, wenn es ihn gibt (d.h. die Power $1 - \beta$), sinkt.
 - Hält man das Risiko eines β -Fehlers aus inhaltlichen Gründen für schwerwiegender, so wird man das Signifikanzniveau eher größer wählen, z.B. $\alpha = 0.10$ oder 0.20. In diesem Fall verringert sich das Risiko eines β -Fehlers und die Power erhöht sich.
 - Eine besondere Situation besteht, wenn die Forschungs- oder „Wunsch“-Hypothese in der H_0 liegt. Wenn wir nun das α klein wählen, steigt die Chance, unsere Hypothese zu bestätigen. Weil wir nur mit einer kleinen Wahrscheinlichkeit irrtümlich unsere Forschungshypothese bestätigen wollen (was hier dem β -Fehler entspricht), wählen wir unter solchen Bedingungen in der Regel ein größeres α , z.B. 0.10 oder 0.20.

Logik des Hypothesentestens

- Welchen Einfluss hat es auf den statistischen Test, wenn unsere Forschungshypothese nicht ungerichtet, sondern gerichtet formuliert ist? Im **Beispiel** lautet die Forschungshypothese dann z.B. „Die Intelligenz der Abiturienten ist von 2009 nach 2010 angestiegen“.
- Die einseitigen statistischen Hypothesen lauten dann: $H_0: \mu \leq 108$ und $H_1: \mu > 108$.
- Dann ist der z -Wert des Stichprobenmittels wieder gleich, im **Beispiel**: $z = 2.39$
- In diesem einseitigen Fall liegt die α entsprechende Fläche vollständig auf der der Hypothese entsprechenden Seite (im Bsp. oberhalb von μ).
- Entsprechend ergibt sich ein kritischer Wert von $z_{crit} = z_{1-\alpha} = z_{0.95} = 1.65$ und entsprechend $|z| = 2.39 > 1.65$ und damit in diesem Beispiel die gleiche Entscheidung wie im zweiseitigen Fall.



Logik des Hypothesentestens

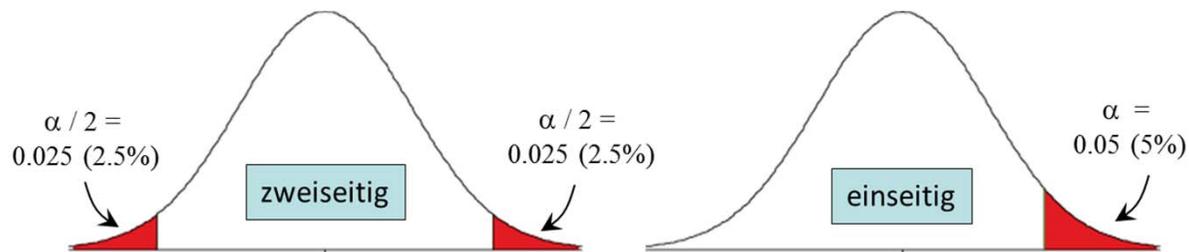
- Bisher sind wir immer davon ausgegangen, dass eine ungerichtete Forschungshypothese zu einer zweiseitigen statistischen Hypothese führt und eine gerichtete Forschungshypothese zu einer einseitigen.
- Tatsächlich lässt sich aus einer **ungerichteten** Forschungshypothese immer nur eine **zweiseitige** statistische Hypothese formulieren und prüfen.
- In der Forschungspraxis tritt aber auch der Fall auf, dass eine **gerichtete** Forschungshypothese **zweiseitig** getestet wird.
- So könnte man z.B. die Forschungshypothese „Frauen sind kreativer als Männer“ zweiseitig testen: $H_0: \mu_F = \mu_M$ und $H_1: \mu_F \neq \mu_M$. Es könnte dann folgende Ergebnisse geben:
 - Frauen sind statistisch signifikant kreativer als Männer. Damit wird die Forschungshypothese gestützt.
 - Frauen und Männer unterscheiden sich nicht statistisch signifikant in ihrer Kreativität. Dies spricht gegen die Forschungshypothese.
 - Männer sind statistisch signifikant kreativer als Frauen. Obwohl sich hier ein statistisch signifikanter Unterschied zeigt, spricht das Ergebnis gegen die Forschungshypothese.

Logik des Hypothesentestens

- Liegt eine gerichtete Hypothese vor, so sprechen folgende Argumente dagegen bzw. dafür, diese **zweiseitig** zu testen:
 - Wenn man die Richtung des Effektes begründet ableiten kann, sollte man diese Information auch in die Testung einfließen lassen.
 - Die Power ist bei der einseitigen Testung höher, d.h. wenn ein Effekt in die Richtung besteht, die in der Hypothese formuliert wurde, wird er mit größer Wahrscheinlichkeit auch gefunden (d.h. der Effekt statistisch signifikant). (Fällt auf der Folie zwei Seiten vorher der z -Wert des Stichprobenmittels in den grauen Bereich zwischen 1.65 und 1.96, so ergibt sich ein statistisch signifikantes Ergebnis nur bei einseitiger Testung .)
 - + Liegt der Effekt in die falsche Richtung, so muss (egal wie groß er ist) immer die H_0 beibehalten werden. Wenn wir also gerichtet vorhergesagt haben, dass Frauen kreativer als Männer sind, und es zeigt sich anschließend anders herum, dann dürfen wir nur schließen, dass Frauen nicht statistisch signifikant kreativer sind als Männer.
 - + Im Sinne der Theorie macht es fast immer einen Unterschied, ob der Effekt 0 ist oder in die Gegenrichtung der Hypothese geht. Das sollte der Test auch widerspiegeln.
 - + Die Richtung des Effektes ist häufig so schwierig eindeutig zu begründen, dass man immer mit beiden Richtungen rechnen sollte.

Logik des Hypothesentestens

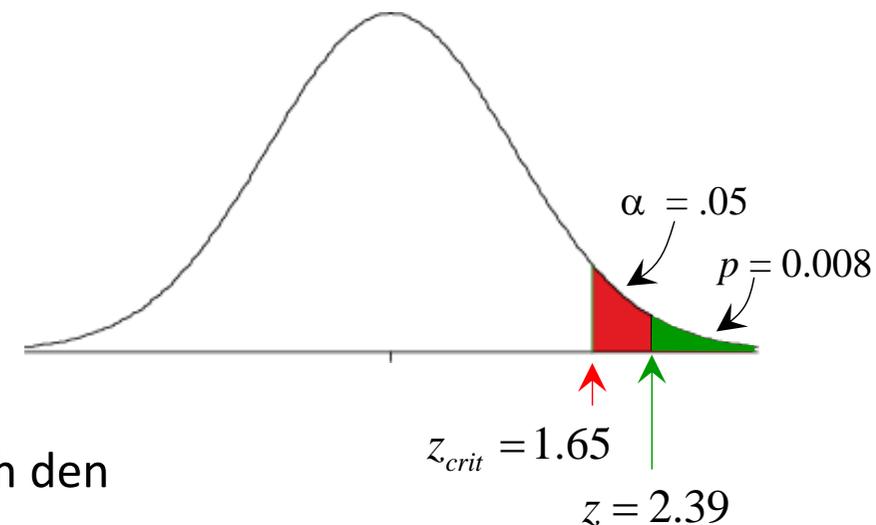
- Die Frage, ob man bei gerichteten Hypothesen ein- oder zweiseitig testet, ist zu entscheiden, **bevor** man sich den Effekt in den Daten angesehen hat (also mit der Festlegung von α).
- Insbesondere ist es unzulässig ...
 - in den Daten zu schauen, in welche Richtung der Effekt fällt und dann den kompletten Ablehnungsbereich α auf die betreffende Seite zu legen. (Die faktische Irrtumswahrscheinlichkeit beträgt dann $2 \cdot \alpha$.)
 - zunächst eine gerichtete Hypothese einseitig zu testen und, falls sich aus den Daten die andere Richtung ergibt, umzuschwenken und nun zweiseitig zu testen. (Die faktische Irrtumswahrscheinlichkeit beträgt dann $1.5 \cdot \alpha$.)



- Meist werden in der Forschung (und den entsprechenden Publikationen) auch gerichtete Hypothesen zweiseitig getestet. Wenn dies nicht explizit angegeben ist, kann man in der Regel von zweiseitiger Testung ausgehen.

Logik des Hypothesentestens

- Häufig wird (z.B. von Statistikprogrammen) die Wahrscheinlichkeit angegeben, die das Auftreten der beobachteten Prüfgröße oder einer noch extremen unter der Annahme der Gültigkeit der H_0 angibt. Diese Wahrscheinlichkeit wird als **p -Wert** oder **Überschreitungswahrscheinlichkeit** bezeichnet.
- **Beispiel:** Für die Prüfgröße $|z| = 2.39$ muss dazu der Flächenanteil rechts von diesem Wert in der Standardnormalverteilung gesucht werden. Es resultiert $p = 1 - 0.992 = 0.008$.
- Wurde einseitig getestet und liegt der Mittelwert in die in der H_1 postulierten Richtung, so liegt dann ein statistisch signifikantes Ergebnis vor, wenn gilt $p < \alpha$.
- Wurde zweiseitig getestet, so liegt dann ein statistisch signifikantes Ergebnis vor, wenn gilt $p < \alpha / 2$.
- Im **Beispiel** resultiert für $\alpha = 0.05$ bei ein- und zweiseitigem Testen ein statistisch signifikantes Ergebnis und für $\alpha = 0.01$ nur bei einseitigem Testen ein statistisch signifikantes Resultat (immer vorausgesetzt die Richtung stimmt beim einseitigen Testen).
- p und α sind nicht zu verwechseln! α wird apriori durch den Forscher festgelegt und p anhand der Daten bestimmt.



- Wie hängen Hypothesentesten und Intervallschätzung zusammen?
- Bestimmen wir in unserem **Beispiel** das Konfidenzintervall um den Mittelwert (unter der Bedingung normalverteilter Population und bekanntem σ), so resultiert bei $\alpha = 0.05$:

$$\bar{x} \pm z_{1-\alpha/2} \cdot \sigma_{\bar{x}} = 110.2 \pm 1.96 \cdot 0.92 \Rightarrow [108.4, 112.0]$$

- Die zweiseitigen Hypothesen bei ungerichteter Fragestellung lauten:
 - $H_0: \mu = 108$ (Der Mittelwert der Intelligenz in der Population ist 108)
 - $H_1: \mu \neq 108$ (Der Mittelwert der Intelligenz in der Population ist ungleich 108)
- Tatsächlich zeigt sich, dass der (unter gleichen Bedingungen, d.h. normalverteilte Population und bekanntes σ) durchgeführte z -Test bei gleichem α genau dann zu einem statistisch signifikanten Ergebnis führt, wenn (wie hier) das Konfidenzintervall den in der H_0 behaupteten Wert nicht enthält. Wir könnten also bei dieser Fragestellung die Hypothesentestung auch mittels der Berechnung des Konfidenzintervalls durchführen.

- Ergänzung: Analog zur einseitigen Hypothesentestung, z.B.

$H_0: \mu \leq 108$ (Der Mittelwert der Intelligenz in der Population ist kleinergleich 108)

$H_1: \mu > 108$ (Der Mittelwert der Intelligenz in der Population ist größer als 108)

lassen sich auch einseitige Konfidenzintervalle bestimmen.

- In unserem Beispiel wird dann aus dem zweiseitigen Konfidenzintervall

$$\bar{x} \pm z_{1-\alpha/2} \cdot \sigma_{\bar{x}} = 110.2 \pm 1.96 \cdot 0.92 \Rightarrow [108.4, 112.0]$$

das einseitige Konfidenzintervall

$$\bar{x} - z_{1-\alpha} \cdot \sigma_{\bar{x}} = 110.2 - 1.65 \cdot 0.92 \Rightarrow [108.7, \infty[$$

- Die Interpretation ist analog der oben dargestellten Logik möglich: Das Konfidenzintervall enthält den in der H_0 behaupteten Wert von 108 nicht; diese wird also zurückgewiesen.
- (Würde die $H_1: \mu < 108$ lauten, so würde analog die obere Grenze des einseitigen Konfidenzintervalls mittels $\bar{x} + z_{1-\alpha} \cdot \sigma_{\bar{x}}$ bestimmt; die untere Grenze wäre $-\infty$.)

Logik des Hypothesentestens

- Der Ablauf des bisher beschriebenen Hypothesentestens (manchmal auch als „null hypothesis significance testing“, NHST, bezeichnet) lässt sich wie folgt schematisch zusammenfassen:

