CHARLOTTE FRESENIUS HOCHSCHULE
UNIVERSITY OF PSYCHOLOGY

Statistik I

Einheit 7: Hypothesen und Hypothesentests

12.12.2023 | Prof. Dr. Stephan Goerigk

Hypothesen und Hypothesentests

Lernziele:

Sie lernen:

- Was ein Signifikanztest ist
- den Unterschied zwischen der Null- und der Alternativhypothese
- das Signifikanzniveau α
- die Abgrenzung zwischen einem einseitigen und einem zweiseitigen Signifikanztest

Hypothesen und Hypothesentests

Wiederholung:

Inferenzstatistik:

- Umfasst alle statistischen Verfahren, die es erlauben, trotz der Informationsunvollständigkeit der Stichprobendaten Aussagen über eine Population zu treffen.
- Wir wissen nun, dass wir einzelne Populationsparameter aus der Stichprobe schätzen können

ABER:

- Das reine Schätzen eines Wertes ist noch keine wissenschaftliche Aussage
- Was für "Aussagen", die wir über die Population treffen, sind gemeint?

Hypothesen und Hypothesentests

Hypothesen

- Statistisch zu prüfende Aussagen: Hypothesen (kennen wir bereits aus QM-1)
- Der Inhalt einer Hypothese muss quantifiziert werden, damit wir sie prüfen können
- Hypothese wird in eine prüfbare Gleichung (oder Ungleichung $\rightarrow$ größer-kleiner Verhältnisse) umgewandelt
    - inhaltlich: Männer sind im Durchschnitt größer als 173 cm
    - numerisch: $\mu > 173$
- Die Entscheidung über die Gültigkeit der Hypothese erfolgt auf Basis unserer Wahrscheinlichkeitsverteilungen ("wie wahrscheinlich ist es, dass...")
    - "Wie wahrscheinlich ist es, dass unter Annahme, dass die Körpergröße normalverteilt ist, der erwartete Mittelwert der Männer größer ist als 173 cm."
- Trifft eine Hypothese zu spricht man oft vom Vorliegen eines **Effekts**

Hypothesen und Hypothesentests

Hypothesentest - To-Do Liste

Zur erfolgreichen Durchführung eines Hypothesentests müssen folgende wichtige Schritte geschehen

1. Austellen von Nullhypothese und Alternativhypothese (Hypothesenpaar)
2. Bestimmung einer zugrundeliegenden Verteilung
3. Festlegung des Annahme- und Ablehnungsbereichs der Nullhypothese (kritischer Wert)
4. Beobachtungswert auf Wahrscheinlichkeitsverteilung abbilden
    - Binomialverteilung $\rightarrow$ Wahrscheinlichkeiten
    - z-Verteilung $\rightarrow$ Mittelwerte, wenn σ in Population bekannt
    - t-Verteilung $\rightarrow$ Mittelwerte, wenn σ in Population nicht bekannt
    - F-Verteilung $\rightarrow$ Varianzen
    - χ^2-Verteilung $\rightarrow$ Häufigkeiten/Proportionen
5. Vergleich kritischer Wert und Teststatistik
6. Entscheidung: Test signifikant oder nicht signifikant

Hypothesen und Hypothesentests

Hypothesen

Statistische Hypothese:

- Entscheidung basiert darauf, ob sich ein beobachteter Wert überzufällig stark von einem vorgegebenen Wert unterscheidet
- Das heißt einfach, dass man überprüft, ob die Abweichung des beobachteten Wertes vom hypothetisierten Wert zu groß ist, als dass sie noch zufällig sein kann.
- Um alle Wahrscheinlichkeiten für einen Ausgang des Hypothesentests abzudecken formuliert man ein Hypothesenpaar
    - H_0: Der hypothetisierte Effekt liegt nicht vor (Werte unterscheiden sich nicht)
    - H_1: Der hypothetisierte Effekt liegt vor (Werte unterscheiden sich)

Hypothesen und Hypothesentests

Nullhypothese und Alternativhypothese

Nullhypothese (H_0):

- Gegenstück zur eigentlichen Untersuchungshypothese, der Alternativhypothese
- Die H_0 stellt meistens den aktuellen Zustand oder anders ausgedrückt den „Standard“ dar gegen den getestet wird

Alternativhypothese (H_1):

- Die Alternativhypothese beinhaltet oft die neue Annahme, den "Effekt".
- Drückt eine "Unterschiedlichkeit" von einem Referenzwert aus

$\rightarrow$ Nur über das komplementäre Hypothesenpaar lässt sich eine komplementäre Gesamtwahrscheinlichkeit abdecken:

$$P_{H_0} + P_{H_1} = 1$$

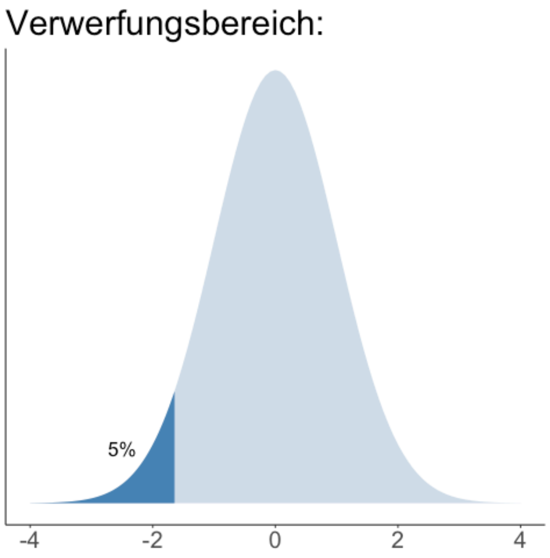
Hypothesen und Hypothesentests

Signifikanzniveau

- Um das Ausmaß der Abweichung zu definieren, welches uns als strengen Wissennschaftler:innen ausreiichend "sicher" erscheint (kritischer Wert) legen wir eine **"Irrtumswahrscheinlichkeit"** fest
- Diese bezeichnet man als **Signifikanzniveau** α.
- Meistens wird $\alpha = .05$ gewählt (häufige Konvention aus der Wissenschaft)
- Das bedeutet, dass wir eine 5% Wahrscheinlichkeit erlauben unsere Hypothese fälschlicherweise anzunehmen
- Die 5% legen einen bestimmten Bereich auf der Wahrscheinlichkeitsverteilung fest (**Verwerfungsbereich**)
    - z-Verteilung
    - t-Verteilung
    - ...

Hypothesen und Hypothesentests

Signifikanzniveau und Verwerfungsbereich



- Die große Fläche der Verteilung entspricht der Annahme unserer H_0
- Der Erwartungswert ist 0 (kein Effekt $\rightarrow$ daher Null-Hypothese)
- Werte jenseits des "kritischen Werts" sind im **Verwerfungsbereich** der H_0 (dunkelblau).
- Diese sind unter Annahme der H_0 ausreichend unwahrscheinlich.
- Wir glauben nicht mehr an einen Zufall!
- Die H_0 wird verworfen.

Hypothesen und Hypothesentests

p-Wert

- In Statistik-Softwareprodukten wird zusammen mit der Teststatistik eines statistischen Tests ein sogenannter **p-Wert** ausgegeben
- Der p-Wert gibt die **Wahrscheinlichkeit für den Fehler erster Art** an, also die Wahrscheinlichkeit, eine gültige H_0 zu verwerfen aufgrund der beobachteten Daten
- **Vorteil** des p-Wertes liegt darin, dass bei der Entscheidung keine Tabelle der Verteilung der Teststatistik benötigt wird
- Wird der zweiseitige p-Wert angegeben und die H_1 ist gerichtet, muss man den p-Wert **durch 2 dividieren** und mit α vergleichen.
- Bei einseitigen Hypothesen ist die zusätzliche überprüfung notwendig, ob die Teststatistik tatsächlich im Verwerfungsbereich der H_0 liegt

Hypothesen und Hypothesentests

CHARLOTTE FRESENIUS HOCHSCHULE
UNIVERSITY OF PSYCHOLOGY

Fehler beim Hypothesentest

- Beim Treffen von Entscheidungen können Menschen nicht nur in ganz alltäglichen Situationen Fehler unterlaufen
- Konkret gibt es bei Hypothesentests **vier Möglichkeiten**, wie die Entscheidung ausfallen kann
  - Fehler 1. Art bzw. α-Fehler: Wenn die Nullhypothese fälschlicherweise verworfen wird und die Alternativhypothese angenommen wird
  - Fehler 2. Art bzw. β-Fehler: Wenn die Nullhypothese fälschlicherweise beibehalten wird, obwohl die Alternativhypothese wahr ist

	H_0 ist wahr	**H_1 ist wahr**
H_0 wird angenommen	1 - α	β (Fehler 2. Art)
H_0 wird verworfen	α (Fehler 1. Art)	1 - β (Power)

Hypothesen und Hypothesentests

Fehler beim Hypothesentest

- Die Wahrscheinlichkeit, einen Fehler 1. Art zu begehen, entspricht immer maximal dem Signifikanzniveau
- Je kleiner α, umso kleiner der Fehler 1. Art $(H_0$ irrtümlich zu verwerfen)
- → α möglichst klein wählen, z.B. $\alpha = 0.001$?
- Entscheiden für Signifikanzniveau ist vergleichbar mit dem Abschluss einer Versicherung. Umso besser der Versicherungsschutz gegen einen α-Fehler, umso höher die Kosten.

Kosten eines kleinen (strengen) Signifikanzniveaus:

- Größerer Fehler 2. Art $(\beta$-Fehler = H_0 irrtümlich beizubehalten)
- geringere Teststärke $1 - \beta$ (Macht oder Power = Wahrscheinlichkeit, H_0 zugunsten einer H_1 zu verwerfen, wenn tatsächlich H_1 gilt)

Hypothesen und Hypothesentests

Beziehungen zwischen Statistischen Fehlern

Dilemma

- Versicherung gegen α-Fehler hat die Kosten eines höheren β-Fehlers und geringerer Macht des Tests
- Versicherung gegen β-Fehler hat die Kosten eines höheren α-Fehlers
- Kompromiss in der Praxis: $\alpha = .05$ oder $\alpha = .01$ je nachdem, welchen Fehler man eher riskieren möchte

Hypothesen und Hypothesentests

Ein- und Zweiseitige Hypothesen

Einseitige Hypothese:

- Der beobachtete Wert ist größer oder kleiner als ein Referenzwert
- Man spricht von einer gerichteten Hypothese
- Beispiel Hypothesenpaar (inhaltlich):
  - H_0: Männer sind durchschnittlich 173 cm groß oder kleiner
  - H_1: Männer sind durchschnittlich größer als 173 cm

Zweiseitige Hypothese:

- Der beobachtete Wert unterscheidet sich von dem Referenzwert
- Man spricht von einer ungerichteten Hypothese
- Beispiel Hypothesenpaar (inhaltlich):
  - H_0: Die Durchschnittsgröße von Männern liegt bei etwa 173 cm
  - H_1: Die Durchschnittsgröße von Männern unterscheidet sich von 173 cm

Hypothesen und Hypothesentests

CHARLOTTE FRESENIUS HOCHSCHULE
UNIVERSITY OF PSYCHOLOGY

Ein- und Zweiseitige Hypothesen

Einseitige Hypothese:

- Der beobachtete Wert ist größer oder kleiner als ein Referenzwert
- Man spricht von einer gerichteten Hypothese
- Beispiel Hypothesenpaar (statistisch):
  - H_0: $\mu \leq 173$ cm
  - H_1: $\mu > 173$ cm

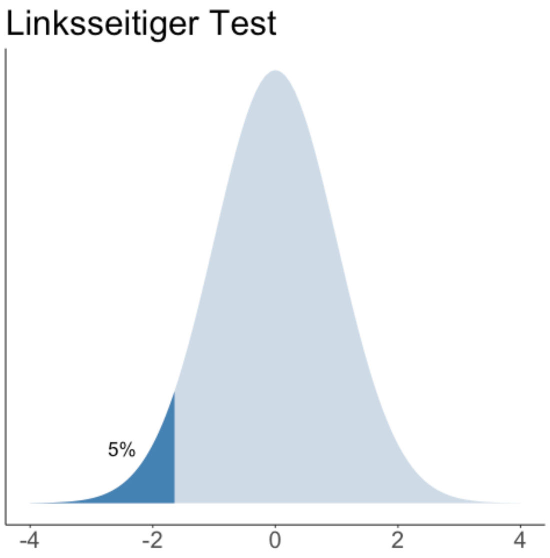
Zweiseitige Hypothese:

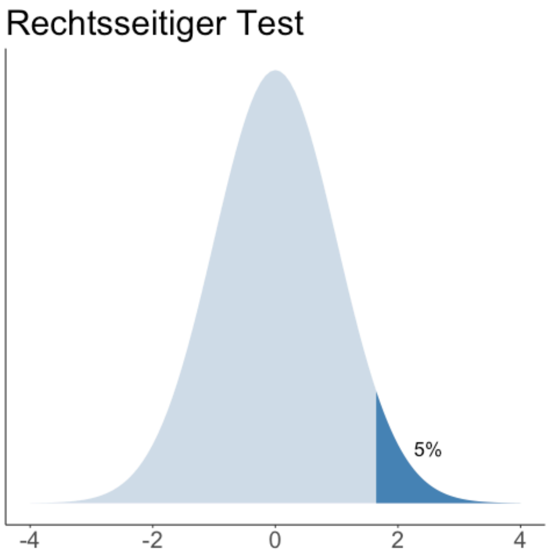
- Der beobachtete Wert unterscheidet sich von dem Referenzwert
- Man spricht von einer ungerichteten Hypothese
- Beispiel Hypothesenpaar (statistisch):
  - H_0: $\mu = 173$ cm
  - H_1: $\mu \neq 173$ cm

Hypothesen und Hypothesentests

Ein- und Zweiseitige Hypothesen - Graphisch

$\rightarrow$

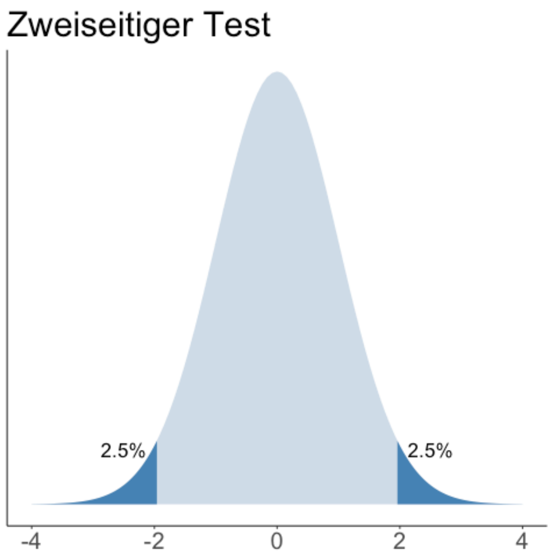




Gesamte "Unsicherheit" $(\alpha = 0.05)$ wird auf eine Seite gesetzt (z.B. kleiner vs. größer als 0).

Hypothesen und Hypothesentests

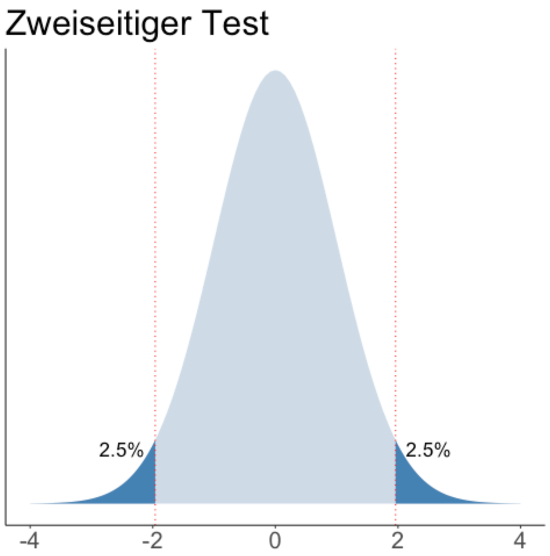
Ein- und Zweiseitige Hypothesen - Graphisch



$\rightarrow$ "Unsicherheit" $(\alpha = 0.05)$ wird auf beide Seiten aufgeteilt (unterschiedlich von 0).

Hypothesen und Hypothesentests

Ein- und Zweiseitige Hypothesen - Graphisch



- Erwartungswert: wahrscheinlichster Wert unter Annahme der H_0

Beispiel:

- H_0: $\mu = 173$
- wenn $\mu = 173$ dann $\mu - 173 = 0$

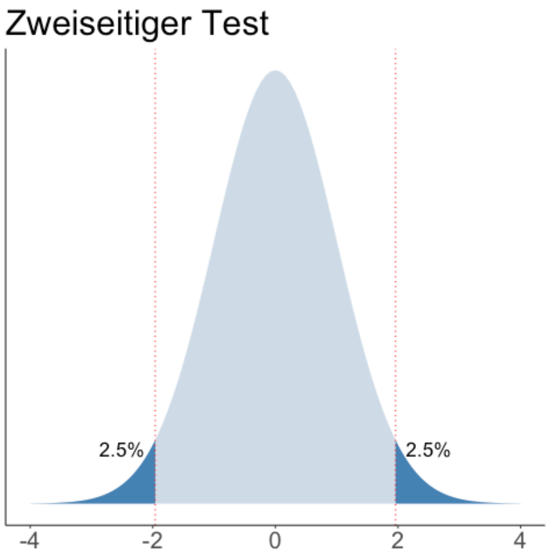
- Erwartungswert unter Annahme der H_0 = 0

- dunkelblaue Fläche: Verwerfungsbereich H_0

- hellblaue Fläche: Annahmebereich H_0

Hypothesen und Hypothesentests

Ein- und Zweiseitige Hypothesen - Graphisch



- rote Linie: Kritischer Wert
- Um zu glauben, dass $\mu \neq 173$ (H_1) muss der beobachtete Wert ausreichend weit vom Erwartungswert der $H_0 = 0$ wegliegen
- Als Schwelle/Entscheidungsgrundlage definiert man einen kritischen Wert
- Dieser liegt oft bei dem Wert, der unter Annahme der Wahrscheinlichkeitsverteilung 5% Auftretenswahrscheinlichkeit hat

In Worten: Der beobachtete Wert ist unter Annahme der H_0 nur 5% wahrscheinlich, somit ist es auf Basis der Beobachtung unwahrscheinlich, dass die H_0 zutrifft.

Hypothesen und Hypothesentests

Ein- und Zweiseitige Hypothesen

Festlegung auf eine Formulierung

- Wahl des Hypothesenpaars sollte a priori erfolgen
  - vor der eigenen Untersuchung
  - ohne Berücksichtigung der aktuellen Daten
  - aufgrund inhaltlicher Kriterien
- Spezialfall einseitige H_1:
  - Richtung basiert auf einer von den aktuellen Daten unabhängigen Vorinformation

Hypothesen und Hypothesentests

Beispiel 1: Binomialtest (einseitiges Testen)

Ein Therapieverfahren zur Behandlung von Angststörungen soll nach 6 Montaten eine Rückfallquote von maximal 30% haben.

- Das Verfahren soll an einer Zufallsstichprobe von $N = 100$ Patient:innnen geprüft werden

Aufstellen des Hypothesenpaares:

- H_0: $p \geq .3$ (mindestens 30%)
- H_1: $p < .3$ (unter 30%)

Fragen:

- Wie viele Patienten in der Stichprobe erleiden einen Rückfall
- Ist dieses Ergebnis **überzufällig** und somit auf die Population (aller Angstpatient:innen) verallgemeinerbar

Hypothesen und Hypothesentests

Beispiel 1: Binomialtest (einseitiges Testen)

Beobachtung in der Stichprobe (Schätzer):

- In der Stichprobe erleiden $n = 28$ Patienten einen Rückfall
- In absoluten Zahlen spricht das erst einmal für das Verfahren $(\frac{28}{100} = 28\% < 30\%)$

Frage, die der Hypothesentest stellt:

- Ist $n = 28$ (28%) ausreichend weit weg von 30%, sodass wir das Ergebnis verallgemeinern können?

Hypothesen und Hypothesentests

Beispiel 1: Binomialtest (einseitiges Testen)

Bestimmung einer passenden Wahrscheinlichkeitsverteilung:

- Art (und i.d.R.) Name des Tests basiert auf zugrunde liegender Verteilung
- Unsere Zufallsvariable X (s.h. Hypothese) ist in Wahrscheinlichkeiten angegeben $(p = .3)$
- Wahrscheinlichkeitsverteilung zur Wahrscheinlichkeit von Wahrscheinlichkeiten (p) : Binomialverteilung

Auswahl Hypothesentest:

- Wir wählen als Hypothesentest des **Binomialtest**
- Zur Entscheidung über unsere Hypothesen werden wir eine **Binomialverteilungstabelle** nutzen
- Merke:
  - Bei gerichteten (einseitigen) Formulierungen wie "mindestens" oder "höchstens" nutzen wir die **kumulierte Binomialverteilungstabelle**
  - Bei ungerichteten Formulierungen wie "genau" nutzen wir die **einfache Binomialverteilungstabelle**

Hypothesen und Hypothesentests

Beispiel 1: Binomialtest (einseitiges Testen)

Parameter für den Binomialtest:

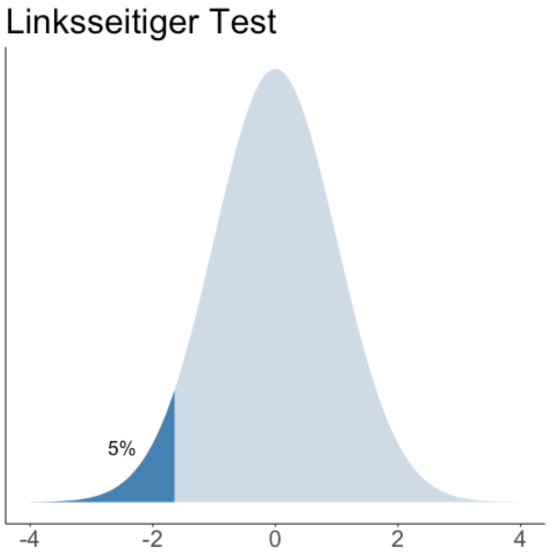
- $p = 30\% = .3$
- $N = 100$
- Signifikanzniveau $\alpha = 5\% = .05$
- $k = 28$

Ist die Rückfallquote niedrig genug, um die Nulllhypothese zu verwerfen?

Hypothesen und Hypothesentests

Beispiel 1: Binomialtest (einseitiges Testen)

Festlegung kritischer Wert (Annahme und Ablehnungsbereich der Nullhypothese):



Fällt das Stichprobenergebnis in den Ablehnungsbereich:

- H_0: $p \geq .3$ (verwerfen)
- H_1: $p < .3$ (annehmen)

Kritischer Wert kann in der kumulierte Binomialverteilungstabelle nachgesehen werden (**Link zur Tabelle**)

Wir wollen für $n = 100$ und die Wahrscheinlichkeit $p = .3$ den letzten Wert k finden, der unter Annahme der BV eine Auftretenswahrscheinlichkeit kleiner als unser Signifikanzniveau $(\alpha = .05)$ hat

CHARLOTTE FRESENIUS HOCHSCHULE
UNIVERSITY OF PSYCHOLOGY

Hypothesen und Hypothesentests

Beispiel 1: Binomialtest (einseitiges Testen)

Festlegung kritischer Wert (Annahme und Ablehnungsbereich der Nullhypothese):

- Kritischer Wert kann in der kumulierten Binomialverteilungstabelle nachgesehen werden (**Link zur Tabelle, S.6**)
- Wir wollen für $n = 100$ und die Wahrscheinlichkeit $p = .3$ den letzten Wert k finden, für den der dazugehörige Wert in der Tabelle unterhalb des Signifikanznveaus $\alpha = .05$ gilt

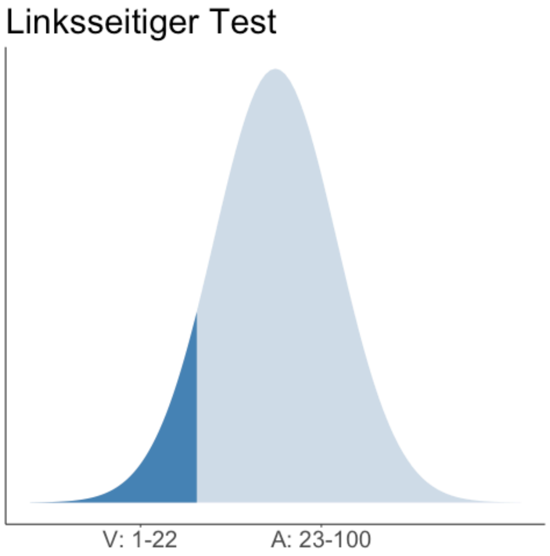
$\rightarrow$ Dieser Wert ist unser kritischer Wert

- In unserem Beispiel ist der kritische Wert $k = 22$ bei einer Wahrscheinlichkeit von $p = .0479$
- Hätten wir $k = 22$ unterschritten, könnten wir mit 95% Sicherheit glauben, dass das Programm max. 30% Rückfallrate hat.

Hypothesen und Hypothesentests

Beispiel 1: Binomialtest (einseitiges Testen)

Festlegung kritischer Wert (Annahme und Ablehnungsbereich der Nullhypothese):



Es ergibt sich:

- Verwerfungsbereich der H_0 von 1-22
- Annahmebereich der H_0 von 23-100
- Unsere Stichprobe ergab 28 Rückfälle
- Dieser Wert liegt im Annahmebereich der H_0
- Test ist nicht signifikant $\rightarrow$ H_0 beibehalten

Interpretation:

- Numerisch weniger Rückfälle als 30%
- ABER: Wir können nicht mit 95% Wahrscheinlichkeit sagen, dass $p < .3$

Hypothesen und Hypothesentests

Ein-Stichproben z-Test (Gauß-Test)

- Hypothesen über μ einer normalverteilten Variable, wobei σ^2 (Populationsvarianz) bekannt ist
- Mögliche Hypothesen:
  - H_0: $\mu = \mu_0$; H_1: $\mu \neq \mu_0$ (ungerichtet)
  - H_0: $\mu \leq \mu_0$; H_1: $\mu > \mu_0$ (gerichtet)
  - H_0: $\mu \geq \mu_0$; H_1: $\mu < \mu_0$ (gerichtet)
- $\mu =$ Populationsmittelwert; $\mu_0 =$ hypothetischer Populationsparameter
- Prüft anhand des Mittelwerts einer Stichprobe ob der Erwartungswert in der entsprechenden Population gleich einem vorgegebenen Wert ist (dem unter H_0 erwarteten μ_0).
- Vergleich eines Stichprobenmittelwertes mit dem hypothetischen Populationsparameter μ_0.

Hypothesen und Hypothesentests

Ein-Stichproben z-Test (Gauß-Test)

- Ziel: Prüfen, ob der Populationsmittelwert (μ) sich von hypothetischen Populationsparameter (μ_0) unterscheidet
- Methode: Da wir Populationsmittelwert (μ) nicht kennen, nutzen wir den Stichprobenmittelwert $(\bar{x})$
- Es geht also um den **Unterschied** zwischen μ und μ_0
  - mathematisch können wir diesen als Differenz ausdrücken: $\bar{x} - \mu_0$.
- Wenn H_0: $\mu = \mu_0$ und H_1: $\mu \neq \mu_0$, dann müsste $\bar{x} - \mu_0 \neq 0$ sein.
  - In Worten: Wenn ein Unterschied besteht, muss die Differenz $\neq 0$ sein

Hypothesen und Hypothesentests

Ein-Stichproben z-Test (Gauß-Test)

- Wenn ein Unterschied besteht, muss die Differenz $\bar{x} - \mu_0 \neq 0$ sein.
- Dabei reicht uns ein rein numerischer Unterschied nicht
  - z.B. 1 vs. 1.000000001 ist ein Unterschied, aber dieser ist sehr klein und ggf. nicht sehr bedeutsam
- Wir wollen eine **verlässliche** Aussage machen (Unterschied muss über gewisse Unsicherheit erhaben sein)
- **numerischer Unterschied $\neq$ signifikanter (verlässlicher) Unterschied**
- Ziel: Wir müssen für den Unterschied eine Wahrscheinlichkeit angeben können
- Für normalverteilte Variablen können wir dafür die z-Tabelle nutzen

Hypothesen und Hypothesentests

Ein-Stichproben z-Test (Gauß-Test)

Bestimmung des z-Werts

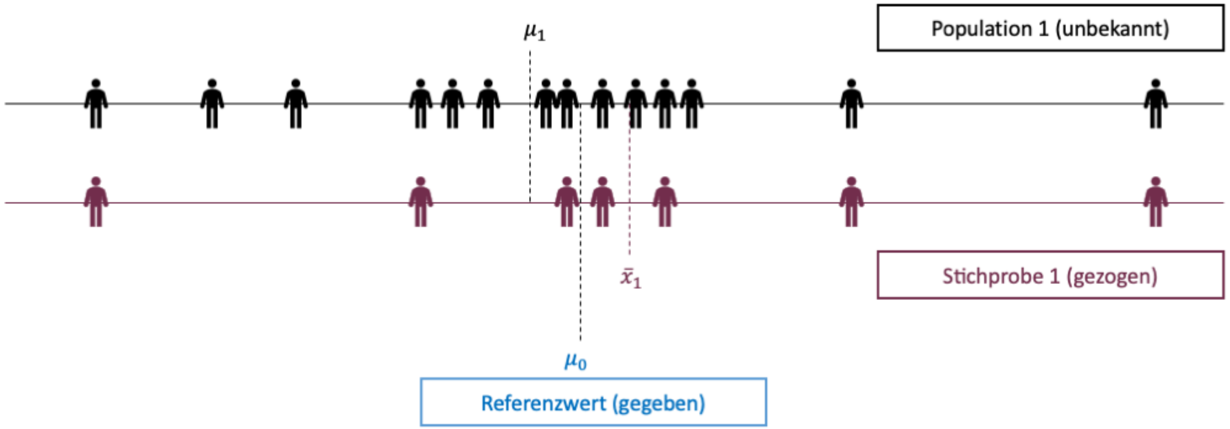
- Wir wandeln den Effekt, über den wir unsere Hypothese aufstellen $(\bar{x} - \mu_0)$ in eine Teststatistik um:

$$z_{emp} = \sqrt{n} \cdot \frac{\bar{x} - \mu_0}{\sigma}$$

- Da der z-Wert aus den beobachteten Daten berechnet wurde, nenne wir ihn auch empirischen z-Wert (z_{emp})
- z : standardnormalverteilter Wert, für den Wahrscheinlichkeit in Tabelle nachgesehen werden kann (Ziel)
- $\bar{x}$: Mittelwert in Stichprobe (beobachtet)
- μ_0 : hypothetischer Populationsparameter aus Fragestellung
- σ : Populationsstandardabweichung (bekannt)
- n : Stichprobengröße (muss berücksichtigt werden $\rightarrow$ größere Stichprobe, mehr Verlässlichkeit)

Hypothesen und Hypothesentests

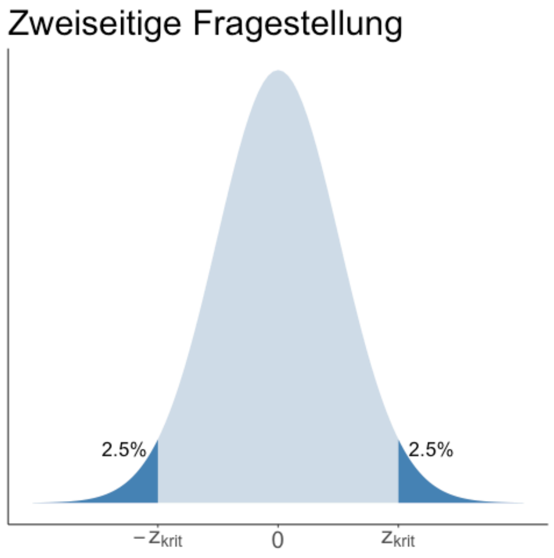
Ein-Stichproben z-Test (Gauß-Test)



Hypothesen und Hypothesentests

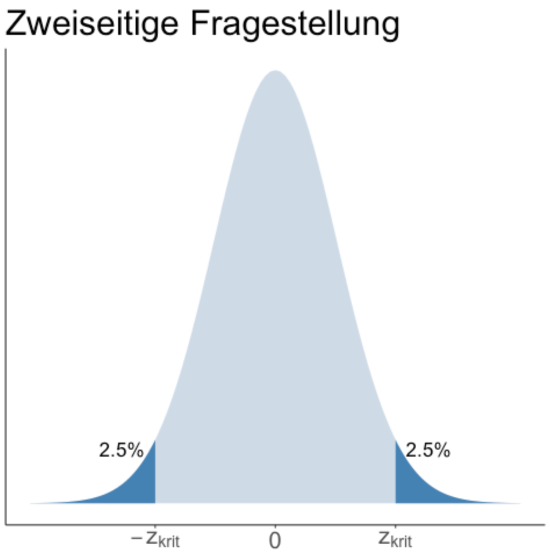
Ein-Stichproben z-Test (Gauß-Test)

- Zur Erinnerung: Warhscheinlichkeit, die H_0 abzulehnen, obwohl sie in Wirklichkeit gilt, heißt α-Fehler oder Fehler 1. Art
- z-Wert ist signifikant, wenn seine Auftretenswahrscheinlichkeit kleiner ist als das gewählte α $(H_0$ verwerfen)
- Für die Signifikanzprüfung kann der z-Wert (z_{emp}) mit dem kritischen z-Wert (z_{krit}) verglichen werden (in z-Tabelle nachsehen)
- Die Wahl des Signikanzniveaus ist von inhaltlichen Überlegungen abhängig und wird oft als $\alpha = .05$ gewählt.



Hypothesen und Hypothesentests

Ein-Stichproben z-Test (Gauß-Test)



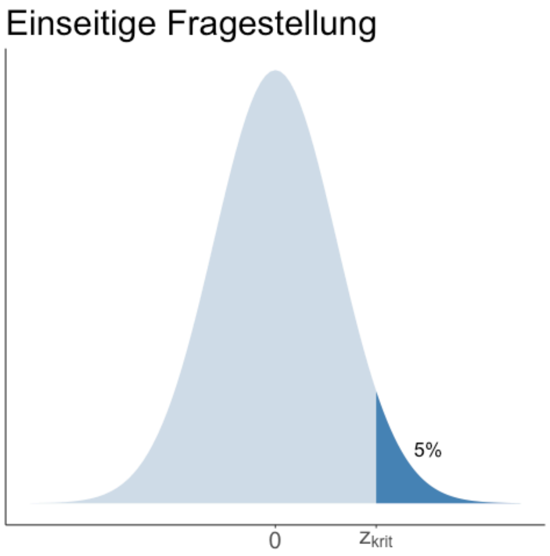
- Signifikanzniveau $\alpha = .05$ muss auf beide Seiten aufgeteilt werden
- Damit $\alpha = .05$ erreicht wird, darf z_{krit} nur 2.5% der Fläche abschneiden
- Auftretenswahrscheinlichkeit von z_{emp} muss kleiner als 2.5%
- Ist Betrag von z_{emp} größer als z_{krit} so ist Test signifikant

CHARLOTTE FRESENIUS HOCHSCHULE
UNIVERSITY OF PSYCHOLOGY

Hypothesen und Hypothesentests

Ein-Stichproben z-Test (Gauß-Test)

- Mittelwertsdifferenz muss in vorhergesagte Richtung auftreten
- Gesamte 5% liegen auf vorhergesagter Seite der Verteilung
- Folge: Gleiche empirische Mittelwertsdifferenz wird bei einseitigen Hypothesen leichter signifikant (Betrag von z_{krit} ist kleiner, bzw. Ablehnungsbereich ist größer).



Hypothesen und Hypothesentests

Ein-Stichproben z-Test (Gauß-Test) - Beispiel

- Weicht der Mittelwert einer Zufallsstichprobe aus dem Irak, $\bar{x}_I$, in einem in Deutschland entwickelten sprachfreien Intelligenztest signifikant vom Populationsparameter in Deutschland, $\mu_D = 100$, ab?

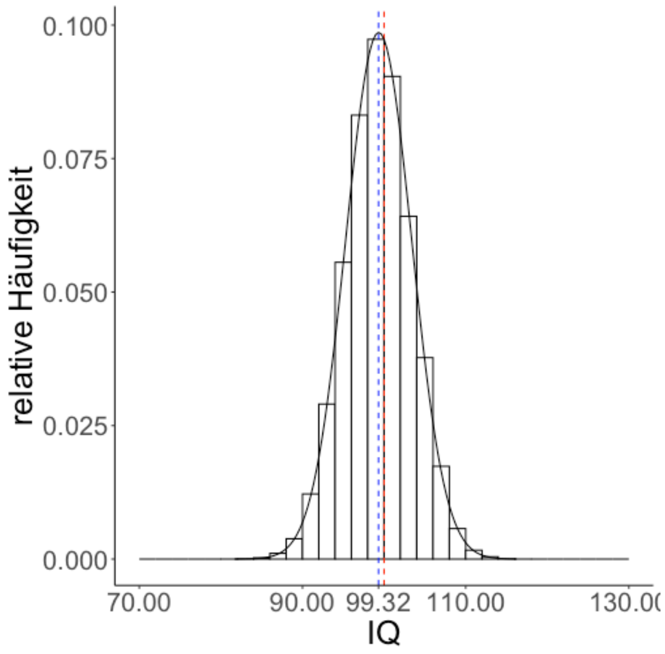
- WICHTIG: Nur die iranische Stichprobe wurde in der Studie rekrutiert und gemessen. Der deutsche Referenzwert wird vorgegeben.

Daten:

- $n = 108$, Testpunkte normalverteilt,
- $\bar{x}_I = 99.32$
- $\sigma_I = 15$
- $\alpha = 0.05$

Hypothesen:

- H_0: $\mu_I = 100$
- H_1: $\mu_I \neq 100$;



Hypothesen und Hypothesentests

Ein-Stichproben z-Test (Gauß-Test)

Beispiel 1:

- $n = 108$, Testpunkte normalverteilt, $\bar{x}_I = 99.32$, $\sigma_I = 4.03$
- H_0: $\mu_I = 100$; H_1: $\mu_I \neq 100$; $\alpha = 0.05$

Teststatistik:

$$z_{emp} = \sqrt{n} \cdot \frac{\bar{x} - \mu_0}{\sigma}$$

$$z_{emp} = \sqrt{108} \cdot \frac{99.32 - 100}{15} = -0.47$$

In z-Tabelle z_{krit} bei $p = 1 - \frac{\alpha=.05}{2} = 0.975$ (durch 2 teilen, da 2-seitige Fragestellung)

- $z_{krit} = -1.96$
- $-0.47 > -1.96 \rightarrow H_0$ beibehalten
- Interpretation: Der Mittelwert der Zufallsstichprobe aus dem Irak unterscheidet sich nicht signifikant vom deutschen Referenzwert.

Hypothesen und Hypothesentests

Macht eines Tests (Teststärke, aka Power)

- $1 - \beta$ = Power
- Wahrscheinlichkeit dafür, dass ein statistischer Test zugunsten einer H_1 entscheidet, wenn diese gilt (d.h. die H_0 richtigerweise zu verwerfen)
- Die Teststärke hängt von verschiedenen Faktoren ab:
  - Signifikanzniveau α (je kleiner, desto weniger Power)
  - Ein- oder zweiseitige Alternativhypothese (bei einseitiger H_1 höhere Power)
  - Größe des Effekts (des Unterschiedes zwischen Beobachtungs- und Referenzwerts)
  - Streuung der Variable X in der Population (je größer desto weniger Power)
  - Stichprobenumfang n (je größer, desto mehr Power)
  - Verwendeter statistischer Test: je höher der verwendete Informationsgehalt der Daten, umso höher seine Macht

Hypothesen und Hypothesentests

Statistische Signifikanz und praktische Relevanz

- Mit genügend großem Stichprobenumfang kann man praktisch jede H_0 verwerfen
- Festlegung einer Effektgröße ε (z.B. praktisch relevanter Unterschied) und einer spezifischen H_1
- Muss aufgrund inhaltlicher Überlegungen vor Datenerhebung festgelegt werden (Poweranalyse)
- Vorteil: notwendiger Stichprobenumfang, der bei gegebenem α, gewünschtem $1 - \beta$ und bestimmter Effektgröße eine eindeutige Entscheidung ermöglicht, bestimmbar

Hypothesen und Hypothesentests

Kumulierung des α-Fehlers

Führt man mehrere statistische Tests durch, hat man bei jedem dieser Tests das Risiko eines α-Fehlers

Beispiel:

- Jemand führt drei statistische Tests durch um eine Hypothese zu prüfen
- Annahme: Unabhängigkeit der einzelnen Tests
- Wahrscheinlichkeit, dass mindestens 1 Test signifikant wird $= 1 - (1 - 0.05)^3 = 0.14$
- 14% Fehlerwahrscheinlichkeit statt 5%

Hypothesen und Hypothesentests

Kumulierung des α-Fehlers

Führt man k unabhängige statistische Tests zur überprüfung einer H_0 durch, kommt es zu einer Kumulierung des α-Fehler Risikos

$$\alpha' = 1 - (1 - \alpha)^k$$

Lösung:

- Lässt es sich nicht vermeiden, mehrere statistische Tests zur Prüfung einer Hypothese durchzuführen, muss eine α-Fehler Korrektur vorgenommen werden
- Der Test wird somit strenger (Grenze für signifikantes Ergebnis niedriger)
- α-Fehler Adjustierung nach Boferroni:

$$\alpha_{adj} = 1 - (1 - \alpha)^{1/k}$$

Beispiel: 3 Tests

$$\alpha_{adj} = 1 - (1 - 0.05)^{1/3} = 0.017$$

Hypothesen und Hypothesentests

Darstellung des Ergebnisses eines Hypothesentests

- Voraussetzung: Formulierung der H_0 und H_1
- Angabe von α
- Angabe von N
- Angabe des p-Wertes
- Angabe der Teststatistik (+ df wo nötig)
- Teststärke und Effektgröße
- Korrekte Formulierung bei nicht-signifikantem Ergebnis:
    - H_0 wird beibehalten (korrekt)
    - H_0 wurde bewiesen (falsch)
- Falsche Interpretationen:
- "Je kleiner der p-Wert, umso größer der Effekt."
    - Der p-Wert ist keine Effektgröße.
    - Der p-Wert gibt die Wahrscheinlichkeit für das Zutreffen der H_0 an.
    - Der p-Wert ist die $P(beobachtetes Ergebnis|H_0)$

Take-aways

- Zur Prüfung der Geltung einer Hypothese (ja/nein) rechnen wir **Signifikanztests**
- **Auswahl des Tests** richtet sich nach der Wahrscheinlichkeitsverteiltung der Testgrößen (Wahrscheinlichkeit, Mittelwerte, Häufigkeiten…)
- Zur Prüfung muss ein Hypothesenpaar aus Nullhypothese und Alternativhypothese aufgestellt werden.
- Je nach Vorwissen können wir Nullhypothese und Alternativhypothese **gerichtet und ungerichtet** formulieren.
- Als Entscheidungskriterium definieren wir einen **kritischen Wert** (oft Wert mit Auftretenswahrscheinlichkeit 5% unter Annahme der Nullhypothese)
- Ist der Test signifikant wird die Nullhypothese **verworfen** und die Alternativhypothese vorläufig angenommen (s.h. Popper)