M24 Statistik 1: Wintersemester 2024 / 2025

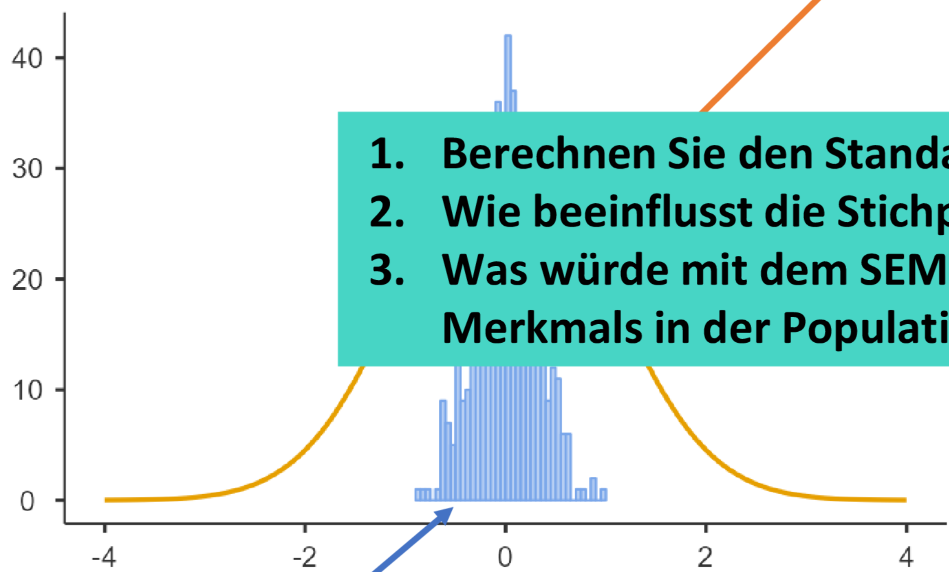
Seminar 08: Signifikanztestung

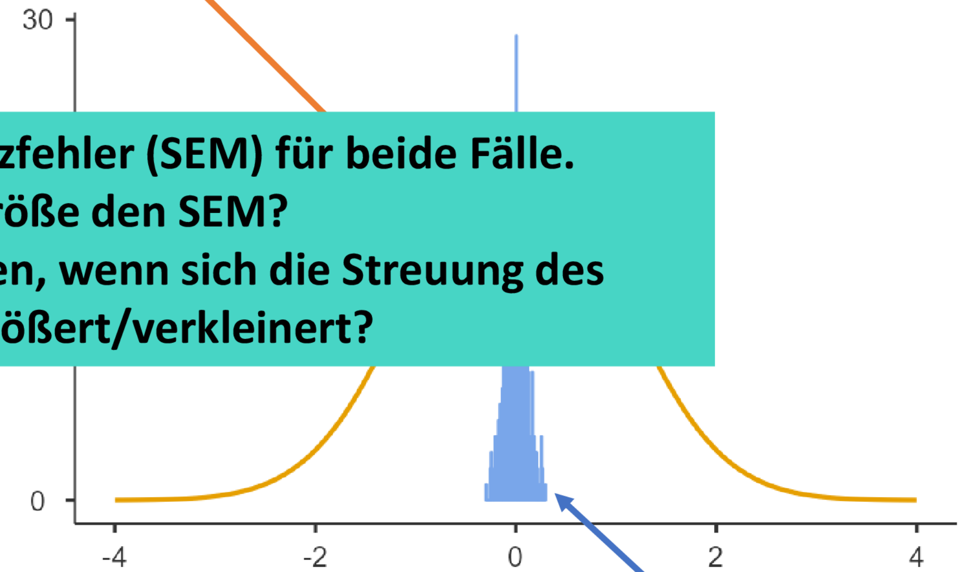
MSc Albert Anoschin & Prof. Matthias Guggenmos

Health and Medical University Potsdam

Wiederholung Stichprobenverteilung

Verteilung des Merkmals in der Population:
$\mu = 0$ und $\sigma = 1$





1. **Berechnen Sie den Standardschätzfehler (SEM) für beide Fälle.**
2. **Wie beeinflusst die Stichprobengröße den SEM?**
3. **Was würde mit dem SEM passieren, wenn sich die Streuung des Merkmals in der Population vergrößert/verkleinert?**

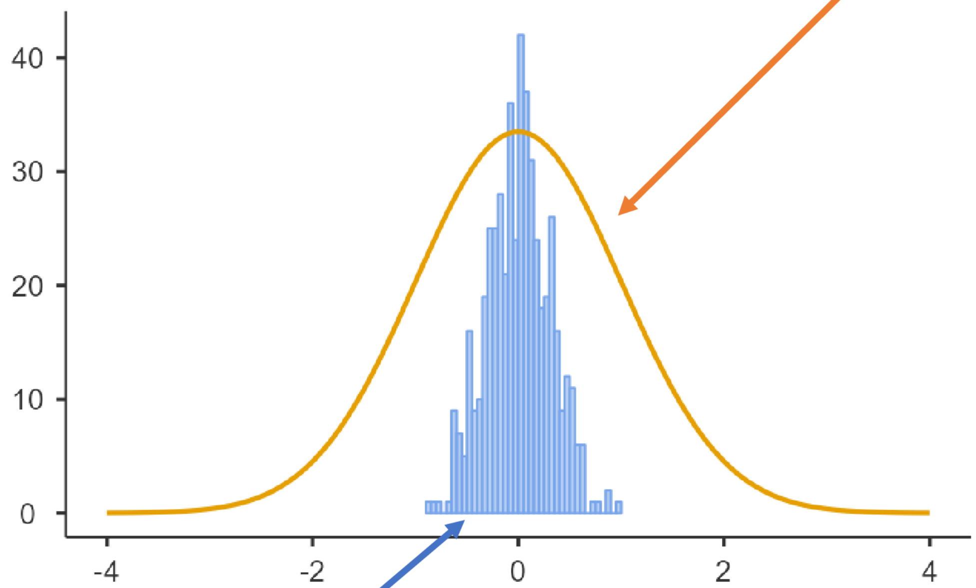
Verteilung der Mittelwerte von 500 Stichproben der Größe $n = 10$

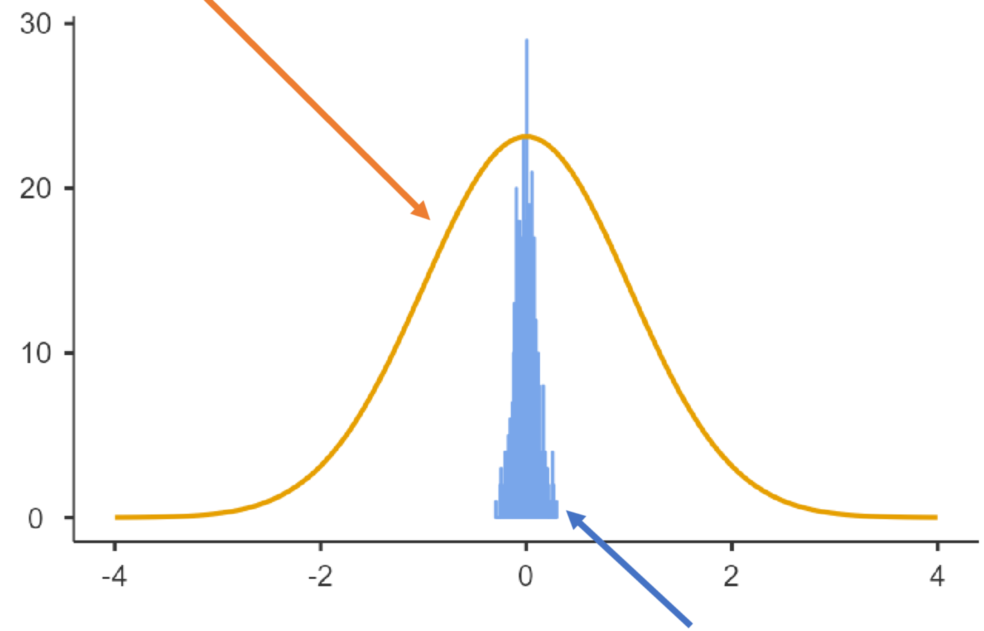
$$SEM = \hat{\sigma}_{\mathrm{sv}} = s\hat{e} = \frac{\hat{\sigma}}{\sqrt{n}}$$

Verteilung der Mittelwerte von 500 Stichproben der Größe $n = 100$

Wiederholung Stichprobenverteilung

Verteilung des Merkmals in der Population:
μ = 0 und σ = 1





Verteilung der Mittelwerte von 500 Stichproben der Größe *n* = 10

$$SEM = \frac{1}{\sqrt{10}} = 0.32$$

$$SEM = \frac{\hat{\sigma}}{\sqrt{n}}$$

Verteilung der Mittelwerte von 500 Stichproben der Größe *n* = 100

$$SEM = \frac{1}{\sqrt{100}} = 0.1$$

Wiederholung: Standardfehler des Mittelwerts

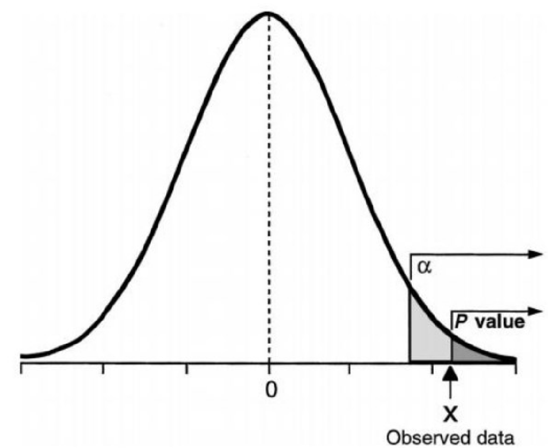
- Der Standardfehler (SEM) ist ein Maß für die Streuung der Stichprobenmittelwerte $\bar{x}$ um den wahren Populationsmittelwert μ.
- Der SEM bezeichnet die Streuung der theoretischen Stichprobenverteilung und kann als „Präzisionsmaß“ für die Schätzung des wahren Populationsparameters μ auf Grundlage der vorhandenen Stichprobe interpretiert werden.
- Die Größe des SEM wird beeinflusst durch die (geschätzte) Streuung des Merkmals in der Population $\hat{\sigma}$ und die Größe der gezogenen Stichprobe n.

$$SEM = \hat{\sigma}_{\text{SV}} = \hat{se} = \frac{\hat{\sigma}}{\sqrt{n}}$$

Prinzip der Hypothesentestung

- Für die Signifikanztestung wird die Stichprobenverteilung herangezogen.
- Das Signifikanzniveau definiert den *Ablehnungsbereich* unter der Stichprobenverteilung der Nullhypothese H_0.
  ➢ Bei einem Signifikanzniveau von $\alpha = 0.05$ entspricht der Ablehnungsbereich 5% der Fläche unter der Verteilung
- Liegt unser Testwert* in diesem Bereich, wird die H_0 abgelehnt
  ➢ Die Wahrscheinlichkeit, den Testwert unter Annahme der H_0 zu messen liegt in dem Fall unter 5%.
- Der *p*-Wert gibt Auskunft darüber, wie wahrscheinlich unser beobachteter Testwert (oder ein noch extremerer Testwert) unter Annahme der H_0 ist.

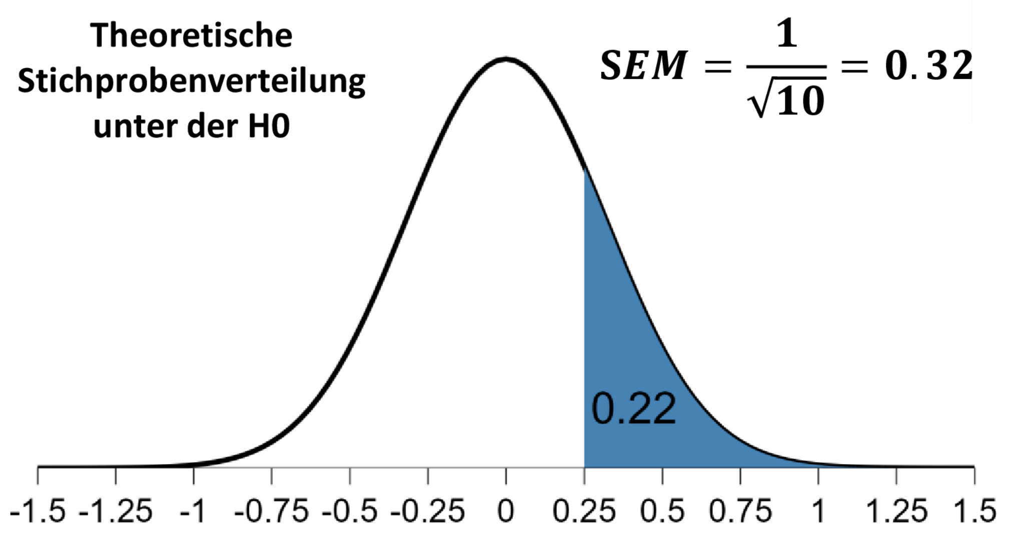
Verteilung der Mittelwertdifferenzen unter der H0

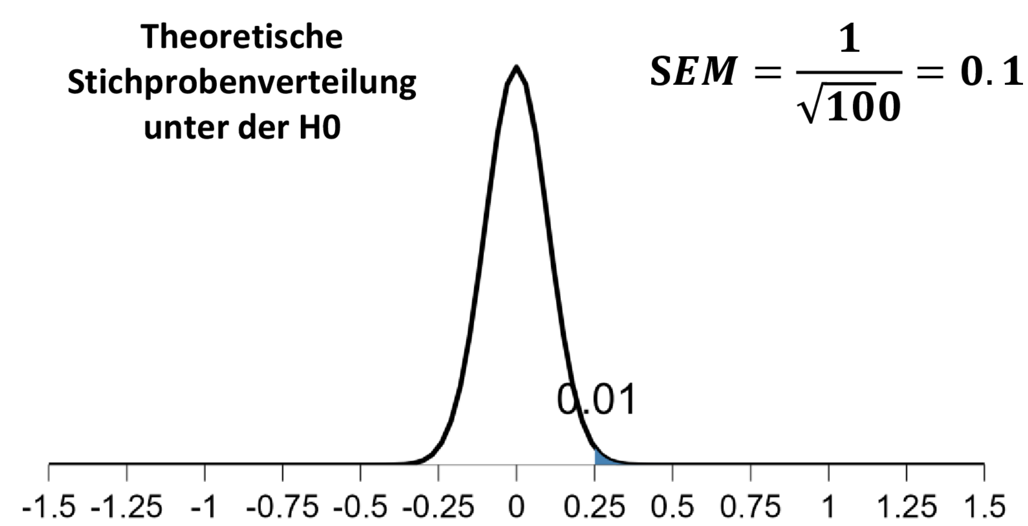


*Mittelwertdifferenz

Gleiche Effekte bei unterschiedlichen Stichprobengrößen

Für die Signifikanztestung wird die Stichprobenverteilung herangezogen. Daher hat der SEM einen entscheidenden Einfluss auf das Ergebnis eines Signifikanztests!





<u>Stichprobengröße *n* = 10</u>
Wahrscheinlichkeit für die Beobachtung einer Mittelwertdifferenz von ≥ 0.25 bei Annahme der H0: ***p* = 0.22**

<u>Stichprobengröße *n* = 100</u>
Wahrscheinlichkeit für die Beobachtung einer Mittelwertdifferenz von ≥ 0.25 bei Annahme der H0: ***p* = 0.01**

Inferenzstatistik: Mittelwertdifferenzen

Unterschied einer Gruppe von einem Referenzwert (Einzelmessung)	Unterschiede zwischen Gruppen in einer abhängigen Variable	Unterschiede zwischen zwei Zeitpunkten in einer abhängigen Variable

↔ IQ = 100

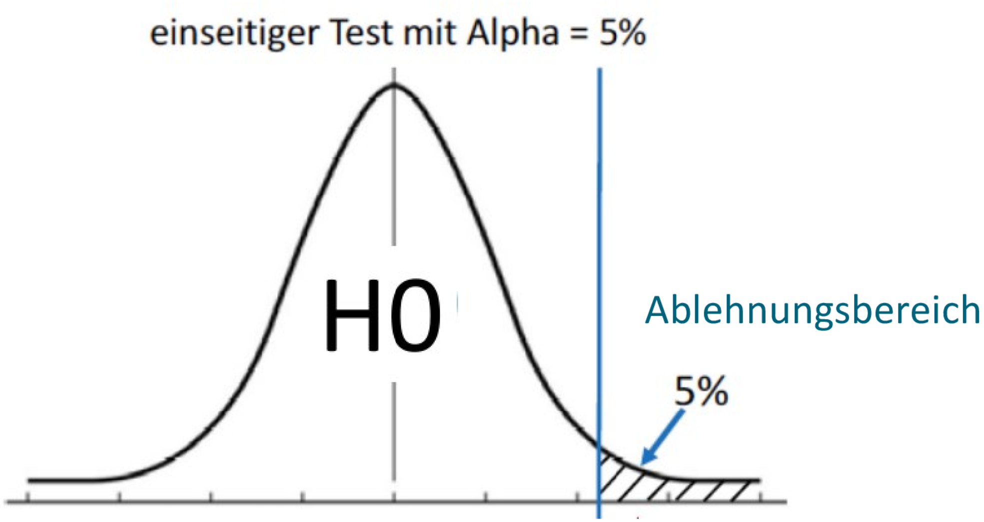
a

b

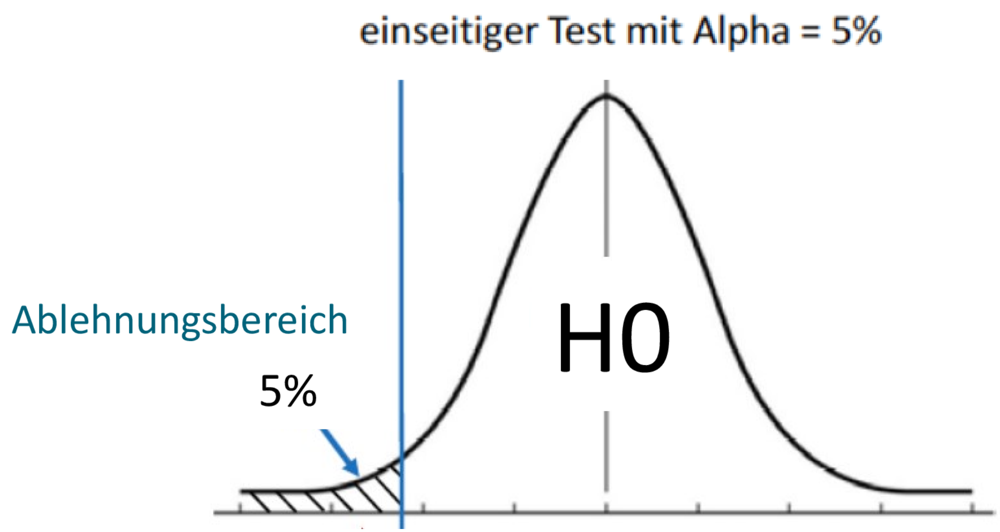
t1 → t2

Gerichtete Hypothesen: Einseitiges Testen

H1: Psychologiestudenten haben einen IQ > 100



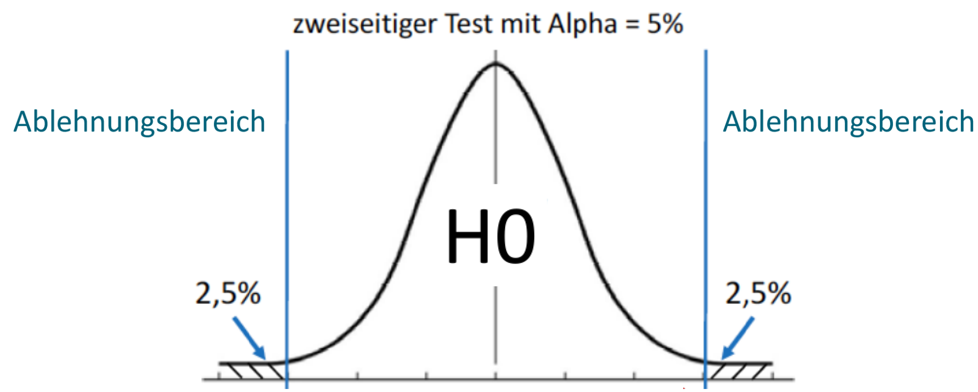
H1: Psychologiestudenten haben einen IQ < 100



Ungerichtete Hypothesen: Zweiseitiges Testen

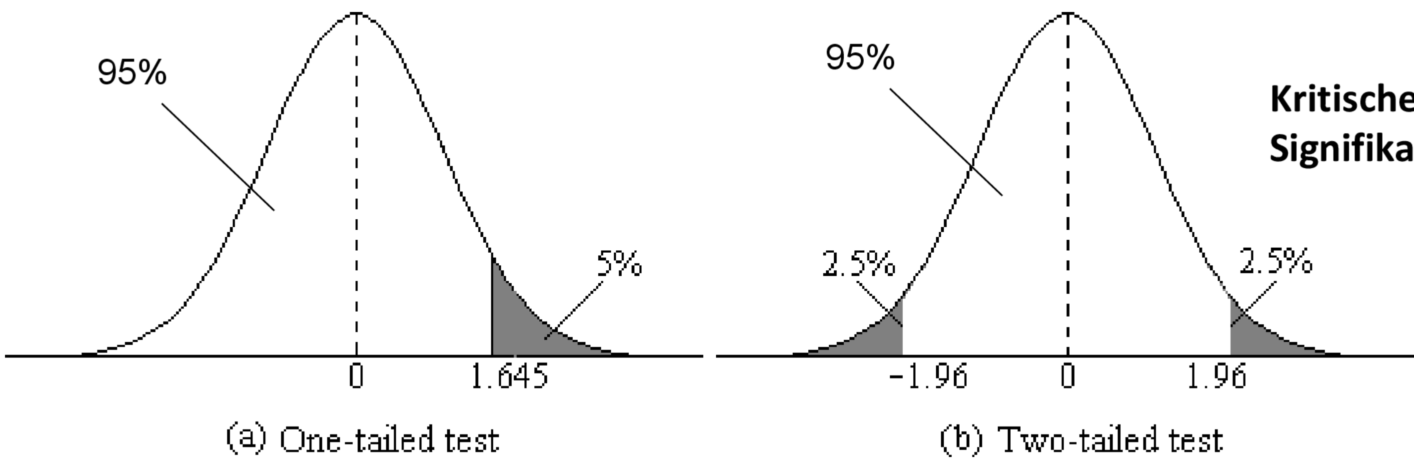
- Bei einem Signifikanzniveau von $\alpha = 0.05$ muss die Irrtumswahrscheinlichkeit insgesamt 5% betragen.
- Bei ungerichteten Hypothesen umfasst der Ablehnungsbereich somit 2.5% auf der linken Seite der Verteilung und 2.5% auf der rechten Seite der Verteilung.
- Zweiseitige Testung ist in Statistikprogrammen wie JASP die Standardeinstellung!

H1: Psychologiestudenten haben einen IQ ≠ 100
H0: Psychologiestudenten haben einen IQ = 100



Bestimmung des Ablehnungsbereichs

- Um zu ermitteln, ob der Testwert (Mittelwertdifferenz) signifikant ist, wird er in eine standardisierte Prüfgröße überführt **(empirischer z-Wert oder t-Wert)**.
- Der „Beginn“ des Ablehnungsbereichs wird durch einen **kritischen z-Wert** oder **kritischen t-Wert** definiert.
- Der empirische Prüfwert muss mind. gleich groß oder größer als der kritische Wert sein, um als signifikant zu gelten.



Kritische z-Werte für einseitige und zweiseitige Signifikanztestung mittels z-Test

Beispiel: z-Test

Der z-Test kann nur für Vergleiche bei **bekannten** Populationsparametern θ verwendet werden. Er basiert auf der z-Standardisierung, die Sie bereits kennengelernt haben.

Stichproben mittelwert

Mittelwert in der Population

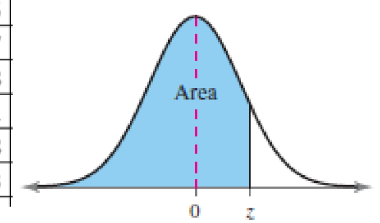
$$z = \frac{\boxed{\bar{x}} - \boxed{\mu}}{\boxed{\sigma / \sqrt{n}}}$$

Standardfehler

Sie möchten prüfen, ob Psychologen extravertierter sind als Mediziner. Sie wissen, dass Mediziner in der Population eine Extraversion von $\mu = \mathbf{5}$ mit einer Streuung von $\sigma = \mathbf{2}$ aufweisen. In Ihrer Stichprobe von **150 Psychologen** messen Sie eine Extraversion von $\mathbf{\bar{x} = 5.3}$. Ist diese Mittelwertdifferenz signifikant?

1. Bestimmung des Standardfehlers SEM.
2. Bestimmung der empirischen Prüfgröße z.
3. Bestimmung des Ablehnungsbereichs: Ab wann signifikant?
4. Ablesen der Fläche $P(x \leq z)$ aus z-Tabelle.
5. Bestimmung des *p*-Werts für die beobachtete Prüfgröße.
6. Bestimmung der Effektstärke d.

z	0	0,01	0,02	0,03
0	0,5	0,504	0,508	0,512
0,1	0,5398	0,5438	0,5478	0,5517
0,2	0,5793	0,5832	0,5871	0,591
0,3	0,6179	0,6217	0,6255	0,6293
0,4	0,6554	0,6591	0,6628	0,6664
0,5	0,6915			
0,6	0,7257			
0,7	0,758			
0,8	0,7881			
0,9	0,8158			
1	0,8413			



Beispiel: z-Test

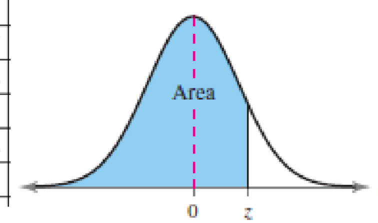
Der z-Test kann nur für Vergleiche bei **bekannten** Populationsparametern θ verwendet werden. Er basiert auf der z-Standardisierung, die Sie bereits kennengelernt haben.

$$z = \frac{\overbrace{\boxed{\bar{x}}}^{\text{Stichproben mittelwert}} - \overbrace{\boxed{\mu}}^{\text{Mittelwert in der Population}}}{\underbrace{\boxed{\sigma / \sqrt{n}}}_{\text{Standardfehler}}}$$

Sie möchten prüfen, ob Psychologen extravertierter sind als Mediziner. Sie wissen, dass Mediziner in der Population eine Extraversion von $\mu = \mathbf{5}$ mit einer Streuung von $\sigma = \mathbf{2}$ aufweisen. In Ihrer Stichprobe von **150 Psychologen** messen Sie eine Extraversion von $\bar{\mathbf{x}} = \mathbf{5.3}$. Ist diese Mittelwertdifferenz signifikant?

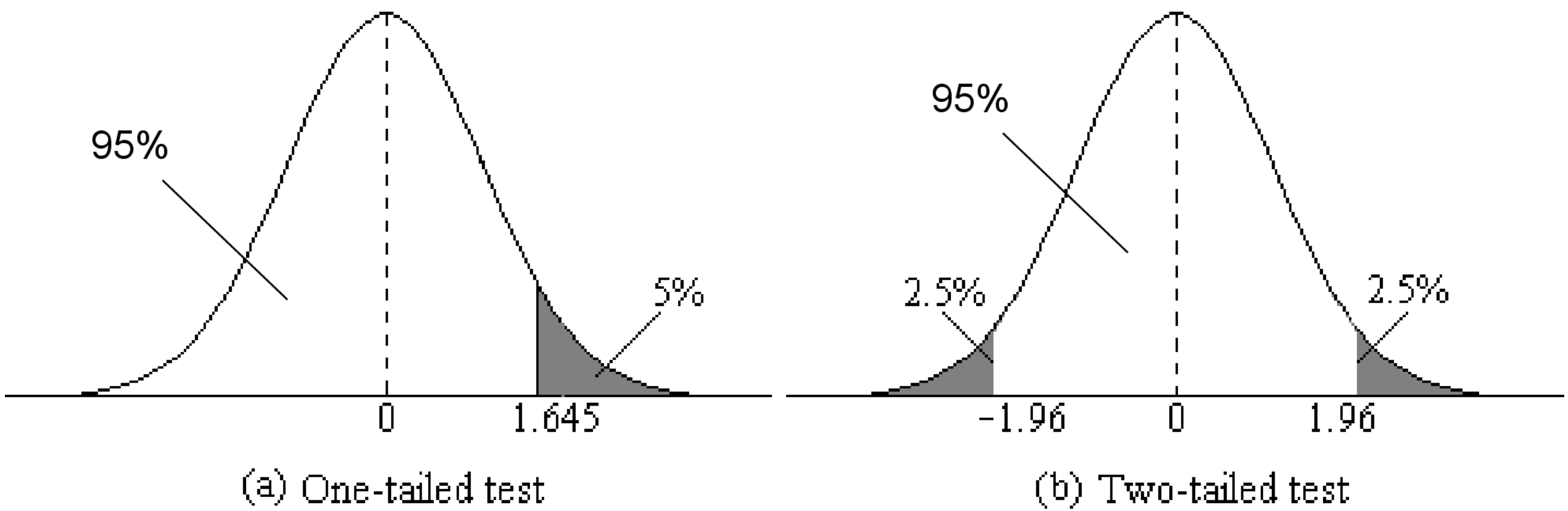
1. Bestimmung des Standardfehlers $\mathbf{SEM = 2/sqrt(150) = 0.163}$.
2. Bestimmung der empirischen Prüfgröße $\mathbf{z = 0.3/0.163 = 1.84}$.
3. Bestimmung des Ablehnungsbereichs: Ab wann signifikant? **Bei rechtsseitiger Testung muss die Fläche links vom empirischen z-Wert mind. 95% betragen!**
4. Ablesen der Fläche $P(x \leq z)$ aus z-Tabelle. **Wenn > 95% , dann signifikant! Abgelesen: 96.71%.**
5. Bestimmung des *p*-Werts für die beobachtete Prüfgröße. $\mathbf{p = 1 - P(x \leq z) = 0.033}$.
6. Bestimmung der Effektstärke $\mathbf{d = 0.3/2 = 0.15}$.

z	0	0,01	0,02	0,03
0	0,5	0,504	0,508	0,512
0,1	0,5398	0,5438	0,5478	0,5517
0,2	0,5793	0,5832	0,5871	0,591
0,3	0,6179	0,6217	0,6255	0,6293
0,4	0,6554	0,6591	0,6628	0,6664
0,5	0,6915			
0,6	0,7257			
0,7	0,758			
0,8	0,7881			
0,9	0,8158			
1	0,8413			



„Quick and Dirty Methode“: Vergleich mit kritischem z-Wert

- Für einen rechtsseitigen Test ist der **kritische z-Wert** $\mathbf{z_{crit}} = \mathbf{1.645}$.
- Der Ablehnungsbereich ist also $\mathbf{z} \geq \mathbf{1.645}$.
- Aufgrund der Symmetrie der Verteilung ist der kritische z-Wert für einen linksseitigen Test $\mathbf{z} = -\mathbf{1.645}$.
- Für einen zweiseitigen Test ist der **kritische z-Wert** $\mathbf{z_{crit}} = \mathbf{1.96}$.
- Der Ablehnungsbereich befindet sich dann in den Bereichen $-\mathbf{z} \leq -\mathbf{1.96}$ und $\mathbf{z} \geq \mathbf{1.96}$.

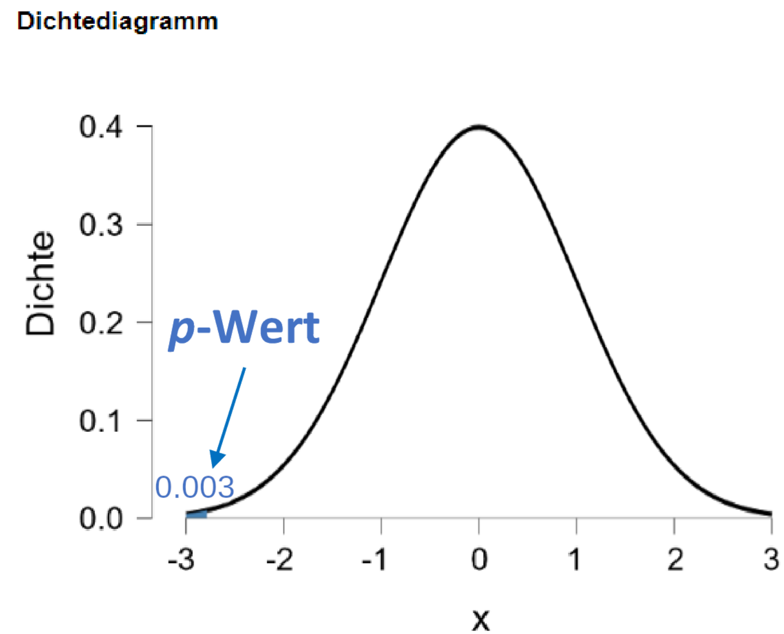


JASP Übung z-Test

Sie möchten prüfen, ob Psychologen **weniger** gewissenhaft sind als eine Normstichprobe. In der Normstichprobe (altersgematchte Population mit hohem Bildungsniveau) ergab sich ein **Mittelwert 3.91** und eine **Standardabweichung von 0.83**. Sie führen einen z-Test durch.

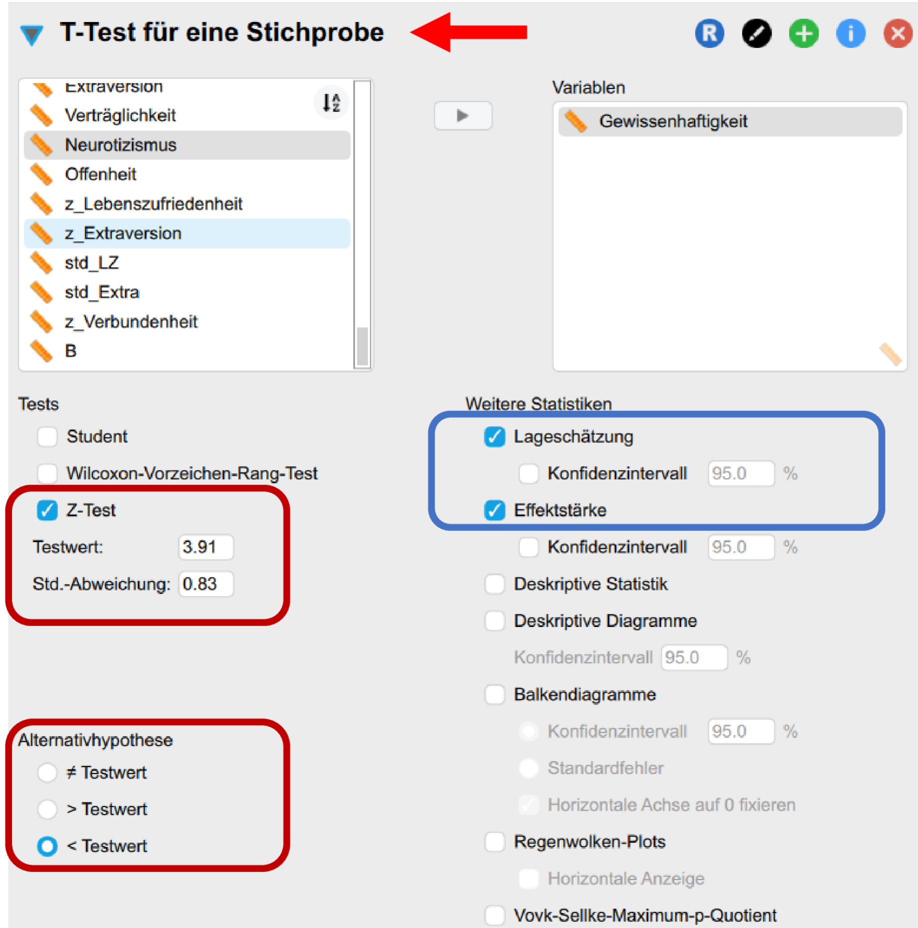
1. Bestimmen Sie den Mittelwert der Gewissenhaftigkeit in unserem Datensatz.
2. Bestimmen Sie die empirischen Prüfgröße z.
3. Bestimmen Sie die Lage des Ablehnungsbereichs.
4. Bestimmen Sie Ihren p-Wert mittels des JASP-Moduls ***Verteilungen → Normal***.
5. Bestimmen Sie die Effektstärke d.
6. Prüfen Sie Ihre Ergebnisse, indem Sie einen z-Test über das Menü ***T-Tests → T-Test für eine Stichprobe → Z-Test*** rechnen.

Übung: z-Test (linksseitig)



empirischer z-Wert

z-Test mit JASP



T-Test für eine Stichprobe ▼

	Z	p	Mittelwertsdifferenz	Cohens d	Std.-Fehler Cohens d
Gewissenhaftigkeit	−2.767	0.003	−0.557	−0.671	0.268

Hinweis. Beim Z-Test wird die Effektstärke mit Cohens *d* angegeben (basierend auf der eingegebenen Standardabweichung der Grundgesamtheit).
Hinweis. Beim Z-Test gibt die Alternativhypothese an, dass der Mittelwert kleiner als 3.91 ist.
Hinweis. Für die Z-Test ist die Schätzung der Lagedifferenz durch die Stichprobenmittelwertdifferenz *d* gegeben.
Hinweis. Z-Test.

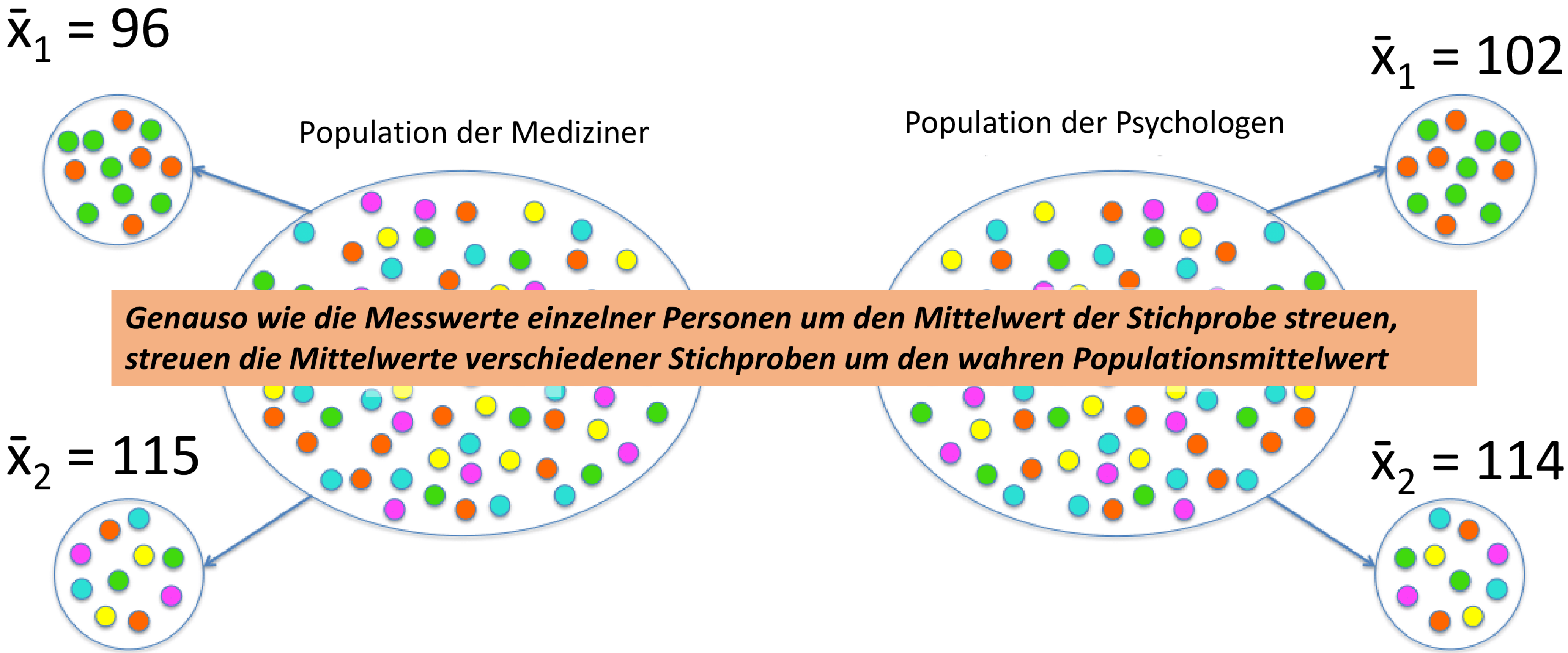
Hypothesentestung

Hypothese H_1: Psychologen unterscheiden sich in der Intelligenz von Medizinern.

Hypothese H_0: Psychologen und Mediziner sind gleich intelligent.

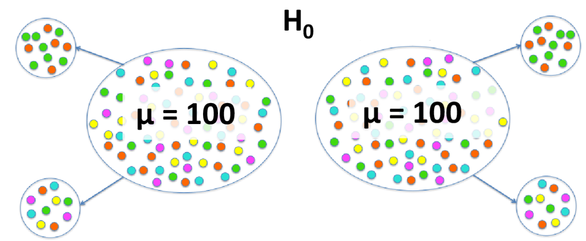
- Sie ziehen eine Stichprobe von $\mathbf{n = 30}$ Psychologen und $\mathbf{n = 30}$ Medizinern und führen mit jeder Person einen Intelligenztest durch
- In der Gruppe der Psychologen erhalten Sie den **Stichprobenmittelwert** $\bar{\mathbf{x}} = \mathbf{104}$, in der Gruppe der Mediziner erhalten Sie $\bar{\mathbf{x}} = \mathbf{100}$.
- Wie lässt sich feststellen, ob der beobachtete Unterschied *statistisch bedeutsam* ist?

Weshalb benötigen wir Signifikanztests?



Genauso wie die Messwerte einzelner Personen um den Mittelwert der Stichprobe streuen, streuen die Mittelwerte verschiedener Stichproben um den wahren Populationsmittelwert

Weshalb benötigen wir Signifikanztests?



$$\bar{x} = \hat{\mu}$$

- Die wahren Populationsmittelwerte sind i.d.R. nicht bekannt.
- Zieht man Stichproben aus der Population, streuen die Mittelwerte der Stichproben $\bar{x}_1, \bar{x}_2, \ldots$ zufällig um den wahren Populationsmittelwert μ.
- Ein deskriptiver Unterschied zwischen dem Mittelwert einer Stichprobe von Psychologen und dem Mittelwert einer Stichprobe von Medizinern kann aufgrund dieser Streuung auch zufällig bedingt sein.
- Der Signifikanztest sagt uns, wie wahrscheinlich der beobachtete Mittelwertunterschied ist *unter der Annahme, dass es in Wahrheit auf Populationsebene keinen Unterschied gibt* (H_0) → **p-Wert**.

Signifikanzniveau

- Wenn der *p*-Wert ein bestimmtes Signifikanzniveau unterschreitet ($p < 0.05$), lehnen wir die Nullhypothese H_0 ab und werten dies als Evidenz für die Alternativhypothese H_1.
- Wir legen *a priori* fest, ab welchem *p*-Wert die Nullhypothese abgelehnt wird: **Signifikanzniveau** oder **Alpha-Fehlerwahrscheinlichkeit** (α)) genannt
- In der Psychologie hat sich ein Signifikanzniveau von $\alpha = \mathbf{0.05}$ eingebürgert

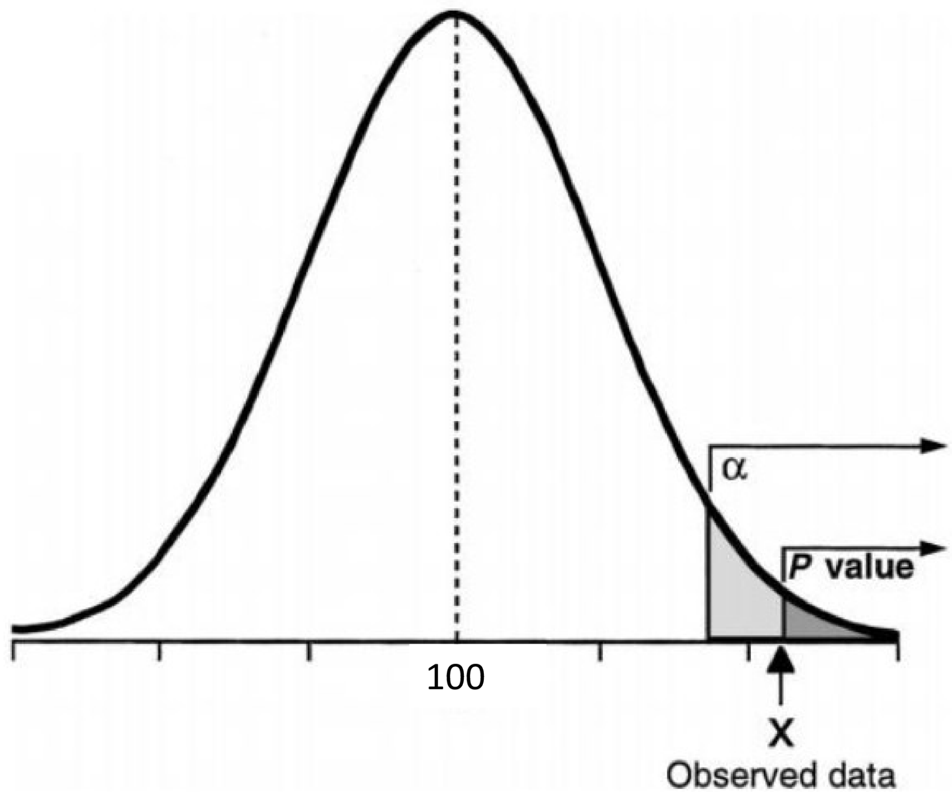
Fazit:

- Die Alternativhypothese H_1 wird nicht geprüft!
- Wenn wir einen signifikanten Unterschied von $p < 0.05$ finden, nehmen wir in Kauf, dass wir uns mit 5%-iger Wahrscheinlichkeit irren und einen Effekt vermuten, wo keiner ist!

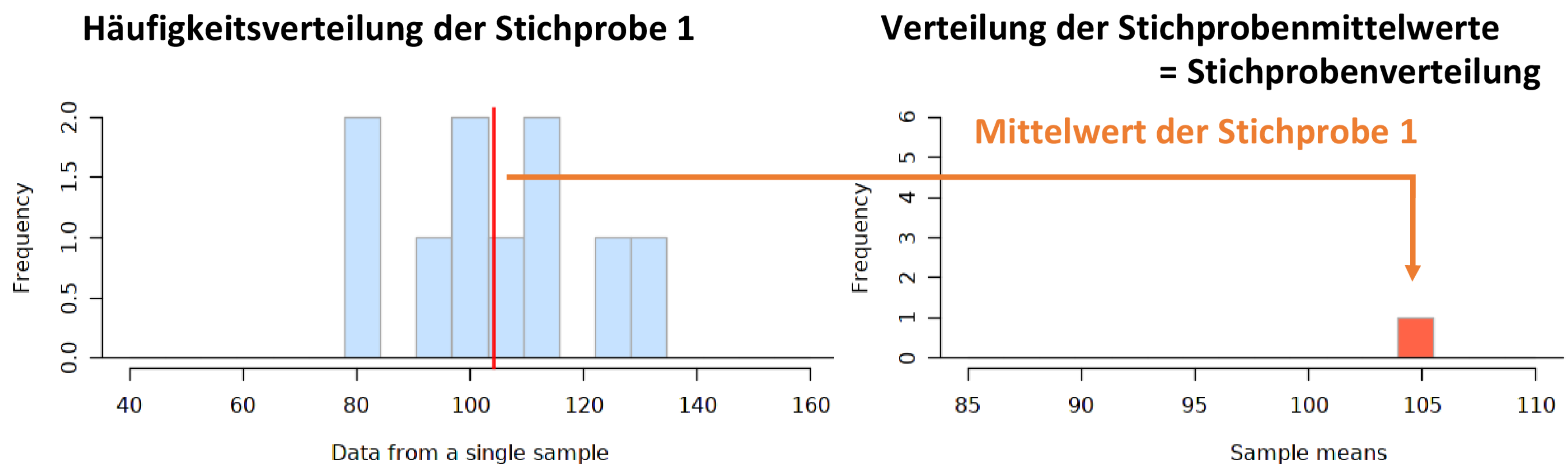
Was sagt der p-Wert aus?

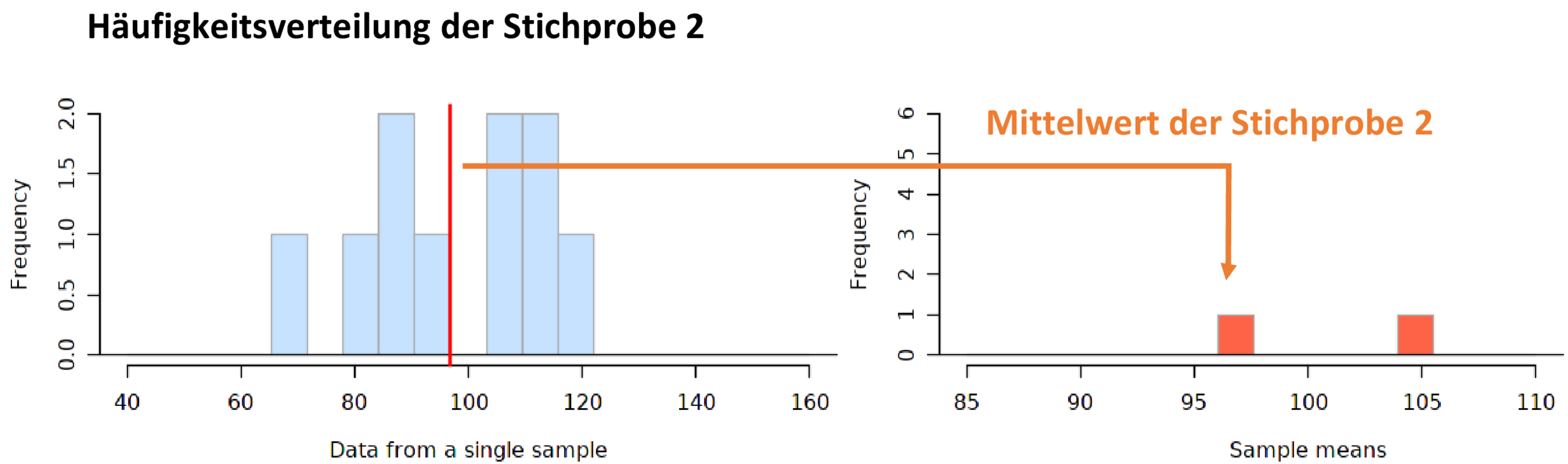
- Der p-Wert gibt uns nur Auskunft über die Wahrscheinlichkeit der beobachteten (oder noch extremerer) Daten **unter Annahme der Nullhypothese:** $\mathbf{P(D \mid H_0)}$.
- Er gibt uns ***keine Auskunft*** über:
  - Die Wahrscheinlichkeit, dass die Nullhypothese wahr ist $P(D \mid H_0)$.
  - Die Wahrscheinlichkeit, dass die Alternativhypothese wahr ist $1 - P(H_1)$.
  - Die Größe des Effekts (→ Effektstärken).
  - Die Replizierbarkeit des Effektes.

Theoretische Verteilung der Stichprobenmittelwerte unter der H0



Demonstration der Stichprobenverteilung

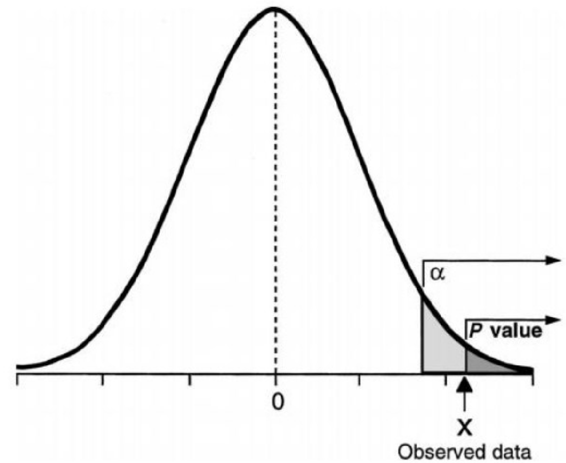




Die Stichprobenverteilung

- Zöge man sehr viele Stichproben von mindestens $n = 30$ aus der Population, würden die Stichprobenmittelwerte **normalverteilt** um den wahren Populationsmittelwert streuen. Diese Streuung bezeichnet man als **Standardfehler (SEM = Standard Error of Mean)**. Dabei ist es unerheblich, ob das gemessene Merkmal selbst in der Population normalverteilt ist.
- Die Verteilung der Stichprobenmittelwerte um den wahren Populationsmittelwert nennt man auch „**Stichprobenverteilung**".
- Die Stichprobenverteilung lässt sich theoretisch beschreiben: Dies ist sinnvoll, *weil eine Stichprobe genügt, um den Populationsmittelwert und die Streuung der Stichprobenmittelwerte um den Populationsmittelwert zu schätzen*!

Theoretische Stichprobenverteilung unter der H0



$$SEM = \frac{\hat{\sigma}}{\sqrt{n}}$$

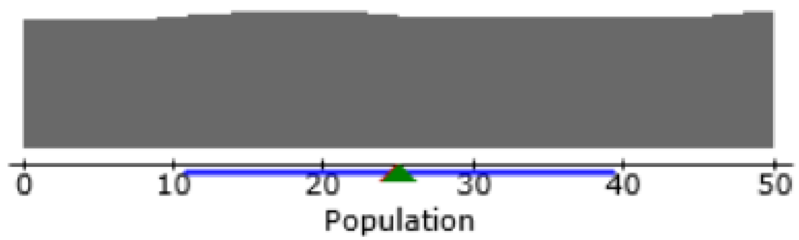
Demonstrationen der Stichprobenverteilung

http://shiny.calpoly.sh/Sampling_Distribution/

https://t.ly/gpXPu

Demonstration

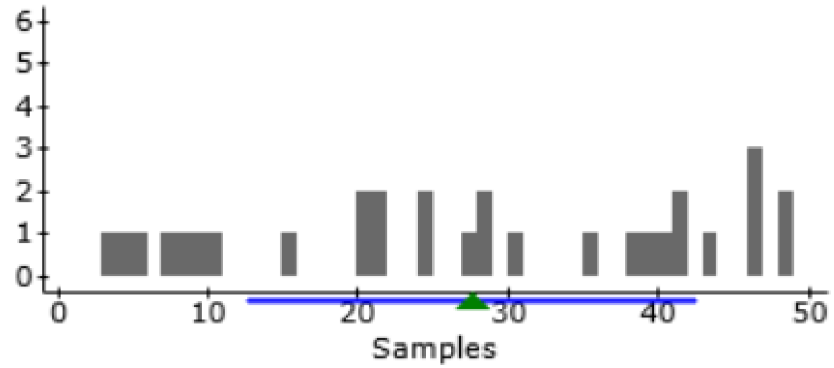
Verteilung des Merkmals in der Population



Population

Mean	25.0666
Median	24.9216
Std. dev.	14.3842

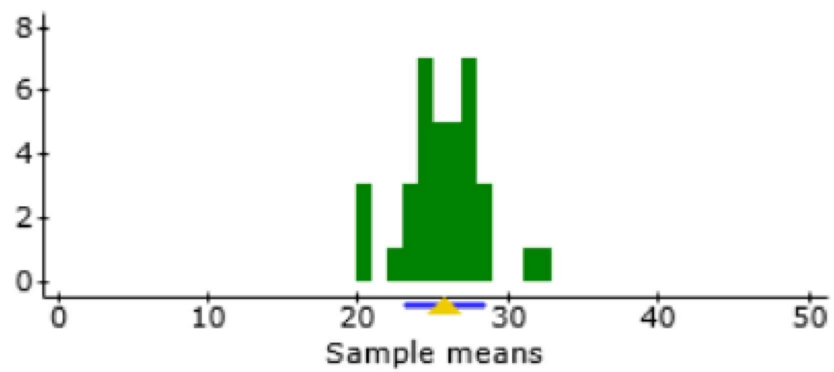
Empirische Häufigkeits-verteilung einer Stichprobe (n = 30)



Samples

Sample size	30
Mean	27.6218
Median	28.0285
Std. dev.	14.9607

Stichprobenverteilung der Mittelwerte (hier: für 36 Stichproben)



Sample means

# of Samples	36
Mean	25.7857
Median	25.8242
Std. dev.	2.6756