Teil 4: Statistik

Was ist Statistik?

- **Statistik** ist die Anwendung dessen von dem was wir wissen auf das was wir wissen wollen.

Ist der DAX ein gutes Modell der gesamten deutschen Wirtschaft?

Spiegelt die Bevölkerung von Bayern die gesamte deutsche Bevölkerung wieder?

Population vs. Stichprobe

- Diese Begriffe kommen immer wieder vor
- Die **Population** ist jedes Mitglied einer Gruppe, die wir untersuchen wollen
- Die **Stichprobe** ist eine kleine Gruppe von (hoffentlich) zufälligen Mitgliedern dieser Gruppe

Parameter vs. Statistik

- Ein **Parameter** ist ein (echtes) Merkmal einer Population. Oft wollen wir Parameter verstehen.
- Ein **statistischer Wert** ist die Eigenschaft einer Stichprobe. Häufig wenden wir **statistische Rückschlüsse** aus Stichprobe an, mit dem Ziel, die Gesamtpopulation zu beschreiben.

Variablen

- Eine **Variable** ist ein Merkmal, das ein Mitglied der Stichprobe beschreibt.

- Variablen können **diskret** oder **stetig** sein

Alter	Gehalt
Geschlecht	Geburtsort

Stichproben

Stichproben

- Einer der großen Vorteile statistischer Modelle besteht darin, dass eine Stichprobe mit einer hinreichenden Größe willkürlich ausgewählter Elemente (**>30)** fast immer die Population widerspiegelt.
- Die Herausforderung wird sein: wie wählen wir die Elemente zufällig aus, um Verzerrungen (en:. bias) zu vermeiden ?

Stichprobenverzerrungen (Bias)

- Es gibt verschiedene Formen der Verzerrungen:

Auswahlverzerrung (selection bias)

Die vielleicht häufigste Art, bei dieser Verzerrung werden Mitglieder einer Population begünstigt, die eher bereit und in der Lage sind Umfragen zu Beantworten.

Stichprobenverzerrungen (Bias)

- Es gibt verschiedene Formen der Verzerrungen:

Auswahlverzerrung (selection bias)

Verzerrung durch Unterdeckung (Undercoverage Bias):

zu wenige Beobachtungen machen oder ganze Segmente einer Population auslassen

Stichprobenverzerrungen (Bias)

- Es gibt verschiedene Formen der Verzerrungen:

Auswahlverzerrung (selection bias)

Verzerrung durch Selbstselektion (Self-selection Bias):

Personen, die sich beispielsweise freiwillig bereit erklären, können sich erheblich von denen in der Bevölkerung unterscheiden, die dies nicht tun

Stichprobenverzerrungen (Bias)

- Es gibt verschiedene Formen der Verzerrungen:

Auswahlverzerrung (selection bias)

Verzerrung durch Gesundheitszustände (Healthy-user Bias):

Die Stichprobe kann aus einem gesünderen Segment der Gesamtbevölkerung stammen - Menschen, die laufen / joggen, draußen arbeiten, sich gesünderer verhalten usw.

Undercoverage Bias

- Eine Mitarbeiterbefragung in einem Krankenhaus, die tagsüber durchgeführt wurde
- vernachlässigt die Befragung der Teilnehmer, die in der Nachtschicht arbeiten

Self-Selection Bias

- Eine Online-Umfrage über eine Sportmannschaft
- Werden nur Leute beantworten, die sich für dieses Team stärker interessieren

Healthy-User Bias

- Eine Umfrage bei Kunden an einem Obststand durchzuführen, um eine Verbindung zwischen Ernährung und Gesundheit zu untersuchen
- wird vermutlich Menschen erreichen, die auch *andere Dinge* tun, die einen noch größeren Einfluss auf ihre Gesundheit haben.

Stichprobenverzerrungen (Bias)

Überlebensirrtum (survivorship bias)

Wenn sich eine Bevölkerung im Laufe der Zeit weiterentwickelt, kann es daran liegen, dass weniger Mitglieder die Bevölkerung aufgrund von Tod, Vertreibung, Umsiedlung usw. verlassen.

Ein klassisches Rätsel

- Zu Beginn des Ersten Weltkrieges trugen britische Soldaten Stoffmützen.
- Dem Kriegsbüro wurde eine alarmierend hohe Zahl von **Kopfverletzungen** gemeldet, also gaben sie allen Soldaten Metallhelme.

Ein klassisches Rätsel

- Sie waren überrascht, dass die Zahl der **Kopfverletzungen** mit der Verwendung von Metallhelmen **stieg**.
- Wenn die Schwere der Kämpfe vor und nach der Veränderung war gleich, warum sollte also die Anzahl der Kopfverletzungen zunehmen?

Ein klassisches Rätsel

- Antwort: Wir müssen *alle* Daten berücksichtigen!
- Vor dem Wechsel zum Stahlhelm führten viele der Angriffe, die zu **Kopfverletzungen** führten auch zum Tod der Soldaten mit Stoffmützen.

Noch ein Beispiel für das Überlebensirrtum

- Im Zweiten Weltkrieg arbeitete der Statistiker Abraham Wald für Amerikas Statistical Research Group (SRG)

Quelle: https://en.wikipedia.org/wiki/Abraham_Wald

Noch ein Beispiel für das Überlebensirrtum

- Ein Problem, an dem die SRG arbeitete, war die Untersuchung der Verteilung von Flugzeugschäden durch feindliches Feuer, um daraus die beste Platzierung von zusätzlichen Rüstungen abzuleiten.

Quelle: https://en.wikipedia.org/wiki/Abraham_Wald

Noch ein Beispiel für das Überlebensirrtum

- Wald sah das anders - er meinte, dass die Schäden **gleichmäßiger verteilt** seien, und dass Flugzeuge, die sogar noch zurückkehren könnten, an weniger verwundbaren Teilen getroffen worden seien.

Quelle: https://en.wikipedia.org/wiki/Abraham_Wald

Noch ein Beispiel für das Überlebensirrtum

- Wald schlug vor, dass die Navy die Flugzeugteile verstärken solle, an denen die zurückkehrenden Maschinen **nicht beschädigt** seien. Das seien die Teile, die bei einem Treffer das Flugzeug zum Absturz bringen würden.

Quelle: https://en.wikipedia.org/wiki/Abraham_Wald

Stichprobenarten

- Einfache Zufallsstichprobe (Random)
- Geschichtete Zufallsstichprobe (Stratified Random)
- Klumpenstichprobe (Cluster)

Einfache Zufallsstichprobe

- Wie der Name schon sagt, bedeutet **Zufallsstichprobe**, dass jedes Mitglied einer Population die gleiche Chance hat, ausgewählt zu werden.
- Da die Stichproben in der Regel jedoch viel kleiner sind als die Population, besteht die Gefahr, dass damit nicht die gesamte Demografie abgedeckt wird.

Geschichtete Zufallsstichprobe

- Die **geschichtete Zufallsstichprobe** (en.: stratisfied random sampling) gewährleistet, dass verschiedene Gruppen innerhalb einer Population angemessen vertreten sind.
- Dazu wird die Population zunächst anhand von Merkmalen in Segmente ein.
- Mitglieder können nicht zwei Gruppen gleichzeitig angehören.

Geschichtete Zufallsstichprobe

- Als nächstes nehmen wir Stichproben aus jeder Gruppe
- Die Größe jeder Stichprobe basiert auf der Größe der Gruppe im Verhältnis zur Population.

Geschichtete Zufallsstichprobe Beispiel

- Ein Unternehmen möchte eine Umfrage zur Kundenzufriedenheit durchführen
- Du kannst nur 10% ihrer Kunden befragen
- Du willst sicherstellen, dass jede Altersgruppe angemessen vertreten ist

Geschichtete Zufallsstichprobe Beispiel

- Die Aufteilung der Kunden nach Altersgruppen ist wie folgt:

20-29	30-39	40-49	50+	TOTAL
1400	4450	3200	950	10,000

Schicht (stratum)

alle Schichten (strata)

Geschichtete Zufallsstichprobe Beispiel

- Um eine 10%-ige Probe zu erhalten, nimmt man 10% aus jeder Gruppe:

20-29	30-39	40-49	50+	TOTAL
1400	4450	3200	950	10,000
140	**445**	**320**	**95**	**1,000**

Klumpenstichprobe (Clustering)

- Eine dritte - und oft weniger präzise - Methode der Stichprobenerhebung ist das **Clustering**.
- Die Idee besteht darin, die Population in Gruppen aufzuteilen und eine zufällige Auswahl von Gruppen oder *Clustern* zu testen.
- Normalerweise wird dies getan, um Kosten zu reduzieren.

Klumpenstichprobe (Clustering) Beispiel

- Eine Marketingfirma schickt Meinungsforscher zu einer Handvoll Nachbarschaften (anstatt eine ganze Stadt zu abzuklappern)
- Ein Forscher untersucht Fischerboote, die sich an einem bestimmten Tag im Hafen befinden (**Convenience Sampling**).

Zentraler Grenzwertsatz

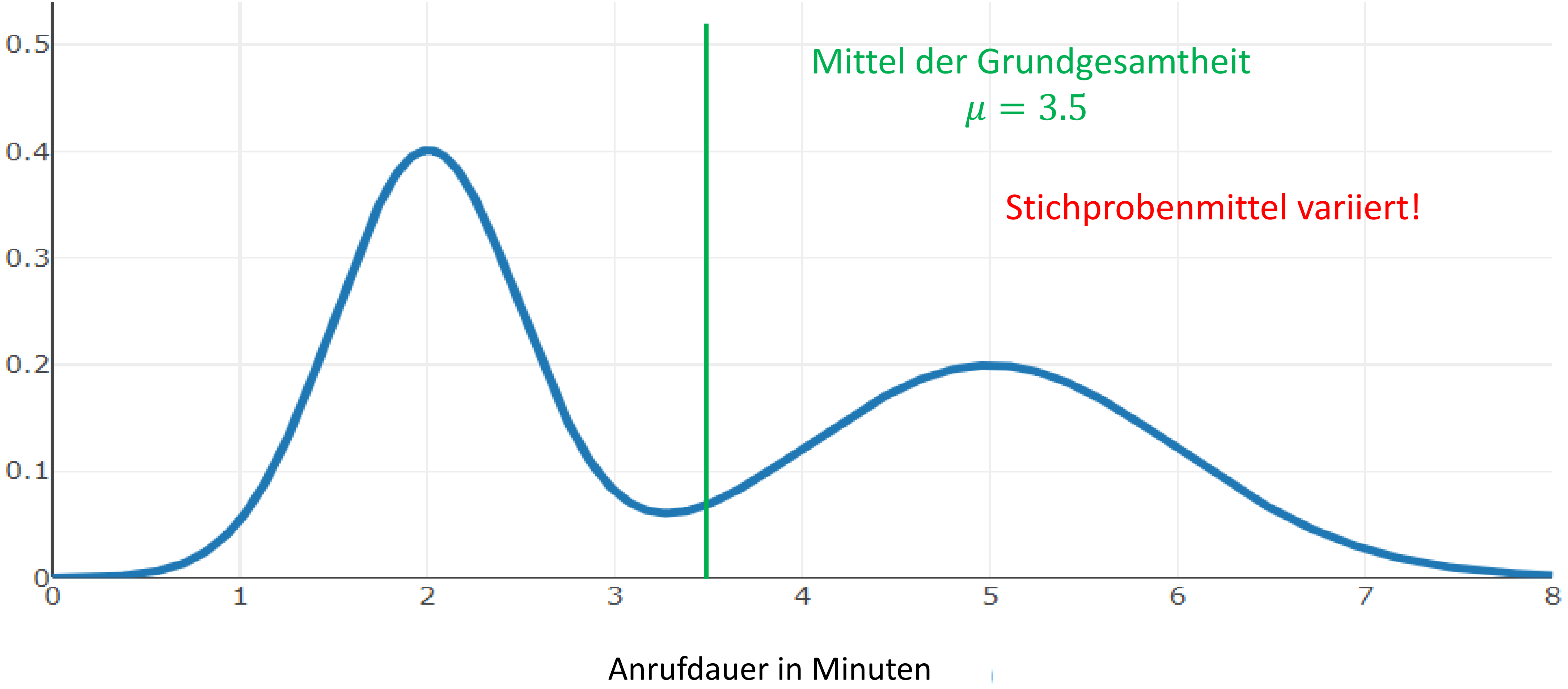
Zentraler Grenzwertsatz

- Was die Stichprobenerhebung zu einem so guten statistischen Werkzeug macht, ist der **zentrale Grenzwertsatz**
- Erinnern wir uns daran, dass ein Stichprobenmittel oft vom Bevölkerungsdurchschnitt abweicht
- Der zentrale Grenzwertsatz berücksichtigt eine große Anzahl von Stichproben.

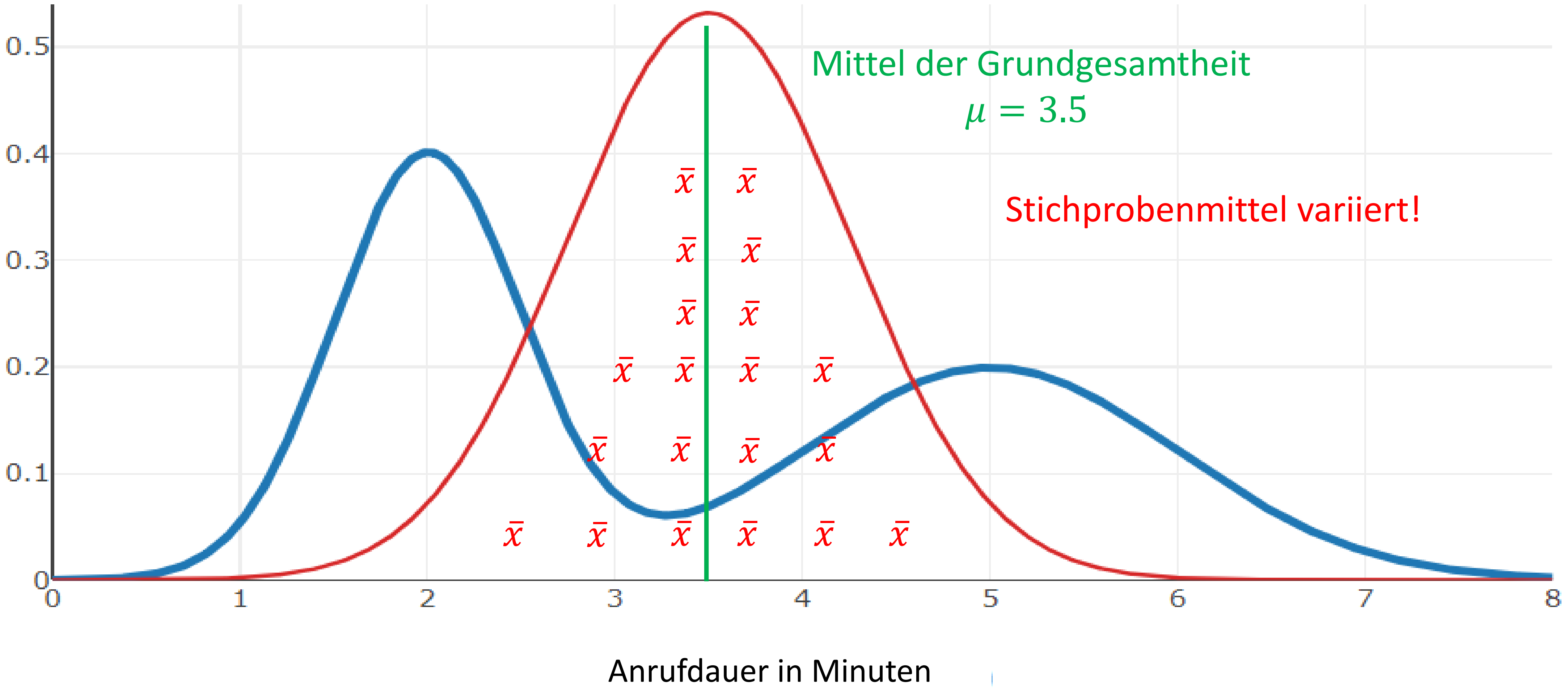
Zentraler Grenzwertsatz

- Der zentrale Grenzwertsatz besagt, dass sich die Verteilung der Stichprobenmittelwerte mehrerer Stichproben mit wachsendem Stichprobenumfang einer *Normalverteilung* annähert
- auch wenn die Grundgesamtheit *selbst nicht normal verteilt* ist
- das heißt, 95% aller Stichprobenmittel sollten innerhalb von 2σ des Populationsmittels liegen

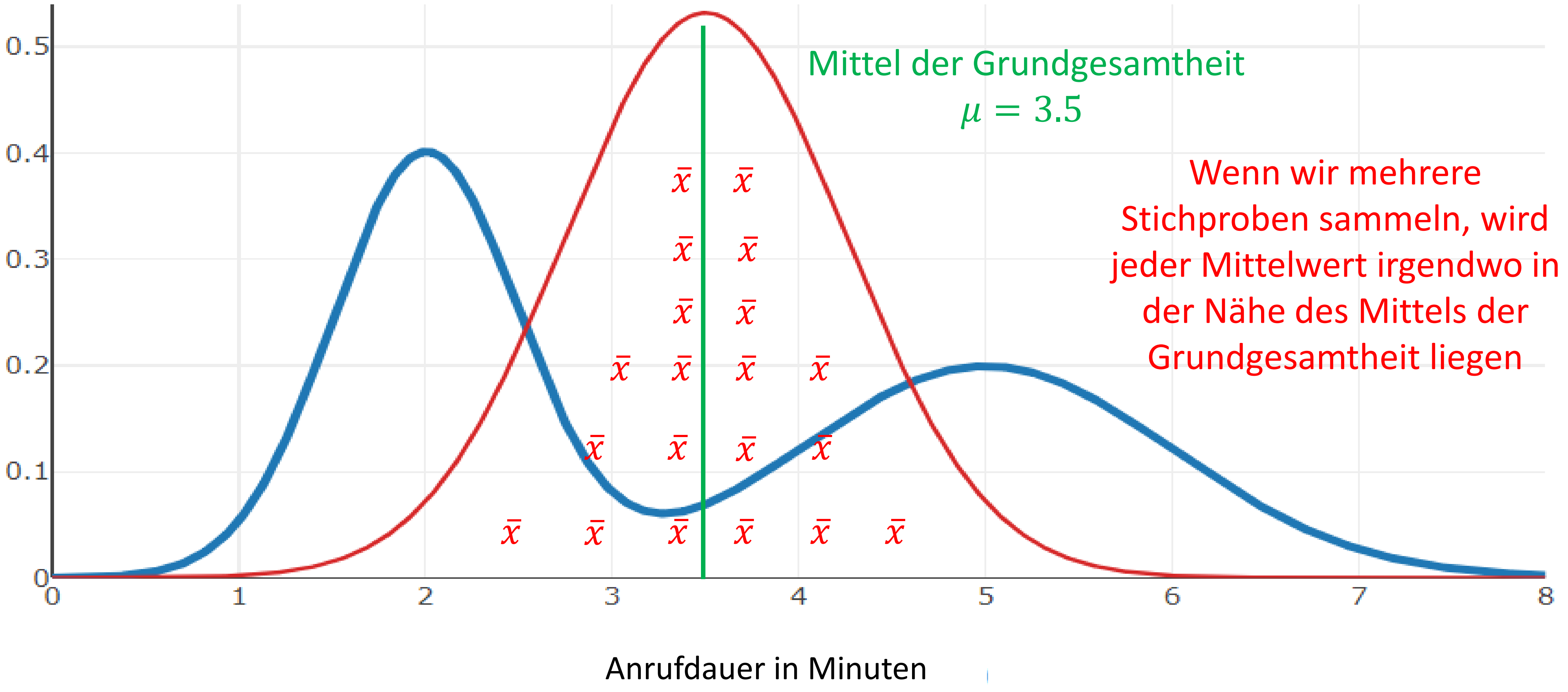
Zentraler Grenzwertsatz



Zentraler Grenzwertsatz



Zentraler Grenzwertsatz



Zentraler Grenzwertsatz in Wikipedia

- Für diejenigen, die neugierig sind, ist der vollständige Beweis des zentralen Grenzwertsatzes auf Wikipedia zu finden:

https://de.wikipedia.org/wiki/Zentraler_Grenzwertsatz

WIKIPEDIA
The Free Encyclopedia

Der Zentrale Grenzwertsatz der Statistik bei identischer Verteilung [Bearbeiten | Quelltext bearbeiten]

Sei $X_1, X_2, X_3, \ldots$ eine Folge von Zufallsvariablen, die auf demselben Wahrscheinlichkeitsraum mit dem Wahrscheinlichkeitsmaß P alle dieselbe Wahrscheinlichkeitsverteilung aufweisen und unabhängig sind (u.i.v. = *unabhängig und identisch verteilt*, engl. i.i.d. = *independent and identically distributed*). Sei weiter angenommen, dass sowohl der Erwartungswert μ als auch die Standardabweichung $\sigma > 0$ existieren und endlich sind.

Betrachten wir nun die n-te Teilsumme dieser Zufallsvariablen $S_n = X_1 + X_2 + \cdots + X_n$. Der Erwartungswert von S_n ist $n\mu$ und die Varianz ist $n\sigma^2$. Bildet man daraus die standardisierte Zufallsvariable

$$Z_n = \frac{S_n - n\mu}{\sigma\sqrt{n}},$$

dann besagt der Zentrale Grenzwertsatz, dass die Verteilungsfunktion von Z_n für $n \to \infty$ punktweise gegen die Verteilungsfunktion der Standardnormalverteilung $N(0,1)$ konvergiert. Dies entspricht genau dem Begriff der Konvergenz in Verteilung in der Stochastik. Ist $\Phi(z)$ die Verteilungsfunktion von $N(0,1)$, dann bedeutet dies, dass für jedes reelle z

$$\lim_{n\to\infty} P(Z_n \leq z) = \Phi(z).$$

In etwas anderer Schreibweise erhält man

$$\lim_{n\to\infty} P\left(\frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \leq z\right) = \Phi(z),$$

wobei

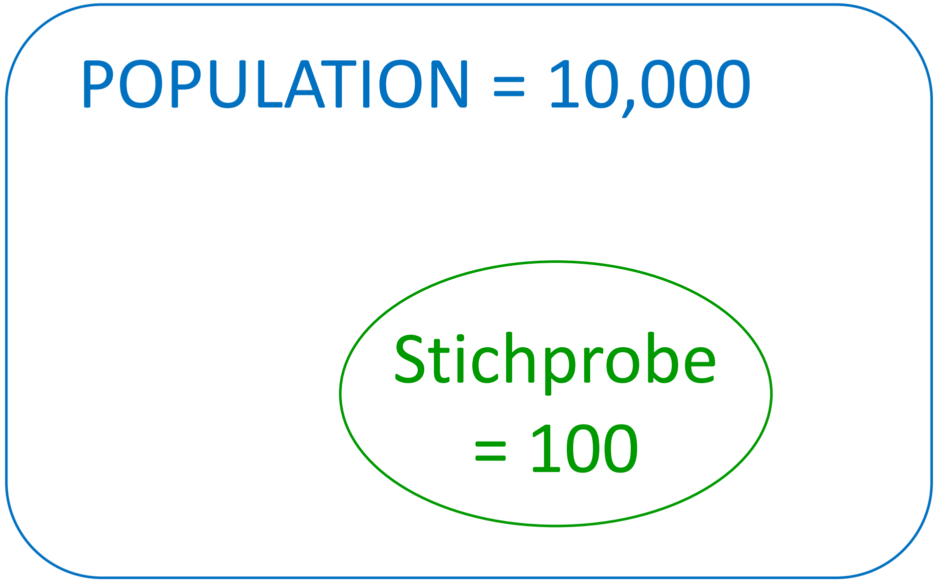
$$\overline{X}_n = \frac{S_n}{n} = \frac{X_1 + \cdots + X_n}{}$$

Standardfehler

Standardfehler

- Ein kurzes Review zur Terminologie
- Sagen wir, wir haben eine **Population** von Wählern
- Es ist unrealistisch, die gesamte Bevölkerung zu befragen, also befragen wir eine **Stichprobe**
- Wir berechnen eine **Schätzwert** aus dieser Stichprobe, mit der wir einen **Parameter** der Population schätzen können

Standardfehler



N = # Grundgesamtheit

P = Parameter der Grundgesamtheit

σ = Standardabweichung der Grundgesamtheit

n = # Stichprobenumfang

$\hat{p}$ = Stichprobenschätzwert

$SE_{\hat{p}}$ = Standardfehler

(Standard Error of the Mean)

Standardfehler

POPULATION = 10,000

Stichprobe = 100

- Wenn die Bevölkerung Australiens durchschnittlich 1,75m groß ist
- und für unsere 100-Personen-Stichprobe ist die durchschnittliche Größe 1,76m beträgt
- Dann

$$P = 1{,}75\text{m}$$
$$\hat{p} = 1{,}76m$$
$$SE_{\hat{p}} = Standard\ Error\ of\ the\ Mean$$

Standardfehler des Mittelwerts

- Die Standardabweichung der Grundgesamtheit beschreibt, wie groß die individuellen Werte vom Mittelwert der Gesamtheit abweichen.
- Der Standardfehler des Mittelwerts beschreibt, wie weit ein Stichprobenmittel vom Mittelwert der Grundgesamtheit abweichen kann.

Standardfehler des Mittelwerts

- Wenn die Standardabweichung der Grundgesamtheit σ bekannt ist, kann Standardfehler des Mittelwerts der Stichprobe wie folgt berechnet werden:

$$SE_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

Standardfehler Übung

- Ein IQ-Test hat eine durchschnittliche Punktzahl von 100 mit einer Standardabweichung von 15 Punkten.
- Wenn eine Stichprobe von 10 Tests einen Mittelwert von 104 hat, können wir annehmen, dass sie die allgemeinen Bevölkerung repräsentieren?

Standardfehler Übung

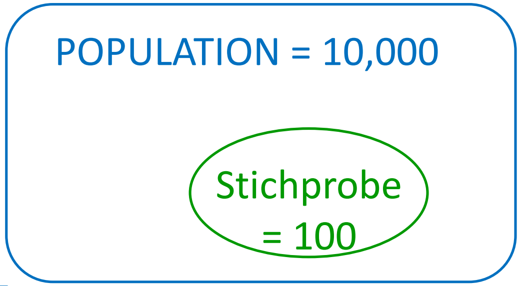
- Stichprobe von 10 **IQ-Testergebnissen**:

$$n = 10 \quad \overline{x} = 104 \quad \sigma = 15$$

$$SE_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{15}{\sqrt{10}} = 4.743$$

- 68% der 10er-Stichproben werden voraussichtlich zwischen 95,257 und 104,743 Punkten liegen

Konfidenzintervall



"Wir können sagen dass der Populationsparameter mit einer **Konfidenz** von 95% in einem **Konfidenzintervall** von plus-oder-minus zwei Standarderror von dem Samplemittelwert (Schätzwert) liegt."

N = # Grundgesamtheit

P = Parameter der Grundgesamtheit

σ = Standardabweichung der Grundgesamtheit

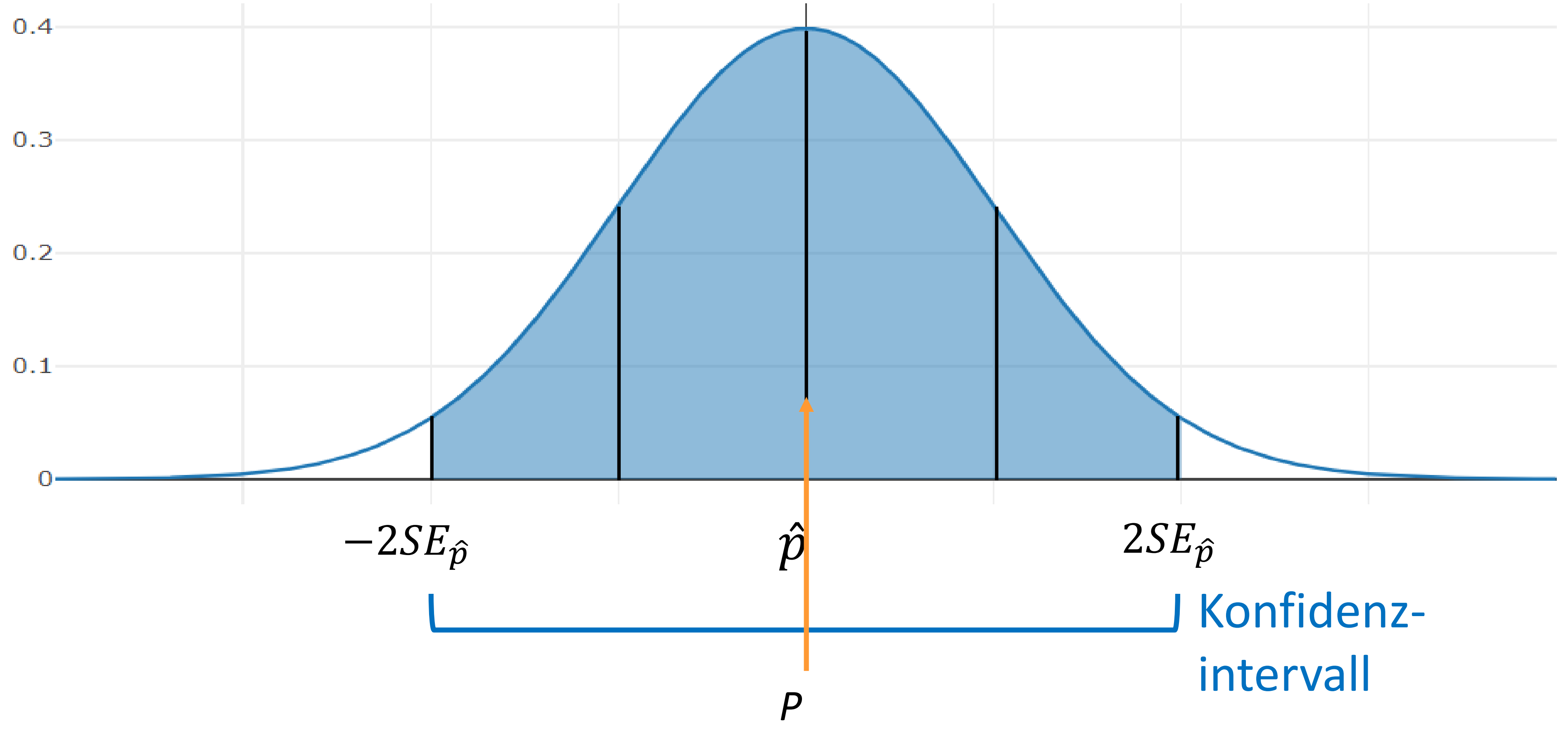
n = # Stichprobenumfang

$\hat{p}$ = Stichprobenschätzwert

$SE_{\hat{p}}$ = Standardfehler

(Standard Error of the Mean)

Konfidenzintervall



Punktschätzung

- Im vorigen Beispiel ist die Stichprobenschätzung $\boldsymbol{\hat{p}}$ eine **Punktschätzung** des Parameters $\boldsymbol{P}$ der Grundgesamtheit.

Hypothesentests

Hypothesentests

- Ein **Hypothesentest** ist die Anwendung statistischer Methoden auf reale Fragen.
- Wir beginnen mit einer Annahme, der sogenannten **Nullhypothese**
- Wir führen ein Experiment durch, um diese Nullhypothese zu testen

Hypothesentests

- Basierend auf den Ergebnissen des Experiments können wir die Nullhypothese nun entweder **annehmen** oder **ablehnen**.

- Wenn die Nullhypothese zurückgewiesen wird, dann sagen wir, dass die Daten eine andere, sich gegenseitig ausschließende **Alternativhypothese** unterstützen.

- Wir "**beweisen**" nie eine Hypothese!

Hypothesenbildung

- Wie gestalten wir die Frage, die unsere Nullhypothese bildet?
- Zu Beginn der Untersuchung wird angenommen, dass die Nullhypothese wahr ist.
- Wenn die Daten die Nullhypothese nicht unterstützen, können wir uns nach einer alternativen Hypothese schauen.

Hypothesenbildung

- Wenn etwas, das als wahr angenommen wird, getestet wird, kann die Nullhypothese diese Annahme widerspiegeln:

- Aussage: "*Unser Produkt hat ein durchschnittliches Versandgewicht von 3,5 kg„*

- **Nullhypothese:** **Durchschnittsgewicht = 3,5kg**
- **Alternativhypothese:** **Durchschnittsgewicht ≠ 3,5 kg**

Hypothesenbildung

- Wenn wir eine Behauptung testen wollen, von der wir uns wünschen, dass sie wahr ist, wir dies aber nicht annehmen können, dann testen wir deren Gegenteil:

- Behauptung: *"Dieser Vorbereitungskurs verbessert die Testergebnisse„*

- **Nullhypothese:** **alte Punkte ≥ neue Punkte**
- **Alternativhypothese:** **alte Punkte < neue Punkte**

Hypothesenbildung

- Die Nullhypothese sollte eine Gleichung (=, ≤, ≥) enthalten:

  durchschnittliches Versandgewicht = 3.5kg $\boxed{H_0: \mu = 3.5}$

- Die alternative Hypothese sollte ein Antonym (≠, <,>) enthalten:

  durchschnittliches Versandgewicht ≠ 3,5 kg $\boxed{H_1: \mu \neq 3.5}$

Hypothesenbildung

- Die Nullhypothese sollte eine Gleichung (=, ≤, ≥) enthalten:

  alte Punkte **≥** neue Punkte

  $$\boxed{H_0\colon\ \mu_0 \geq \mu_1}$$

- Die alternative Hypothese sollte ein Antonym (≠, <,>) enthalten:

  alte Punkte **<** neue Punkte

  $$\boxed{H_1\colon\ \mu_0 < \mu_1}$$

Hypothesentests

- Was also lässt uns die Nullhypothese zurückweisen oder nicht zurückweisen?

Hypothesentests

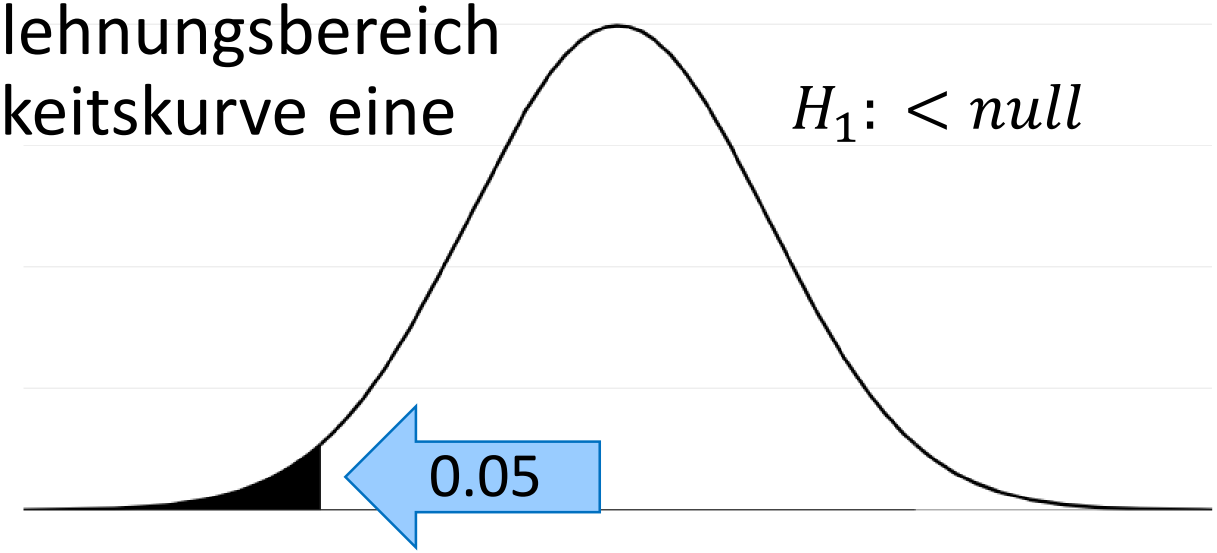
- Wir führen eine Untersuchung durch und speichern die Ergebnisse.
- **Unter der Annahme, dass unsere Nullhypothese gültig** ist und die Wahrscheinlichkeit, die getroffene Annahme zu beobachten sehr klein ist (innerhalb von 0,05), lehnen wir die Nullhypothese ab.

- Hier ist 0,05 unser **Signifikanzniveau**

$$\alpha = 0{,}05$$

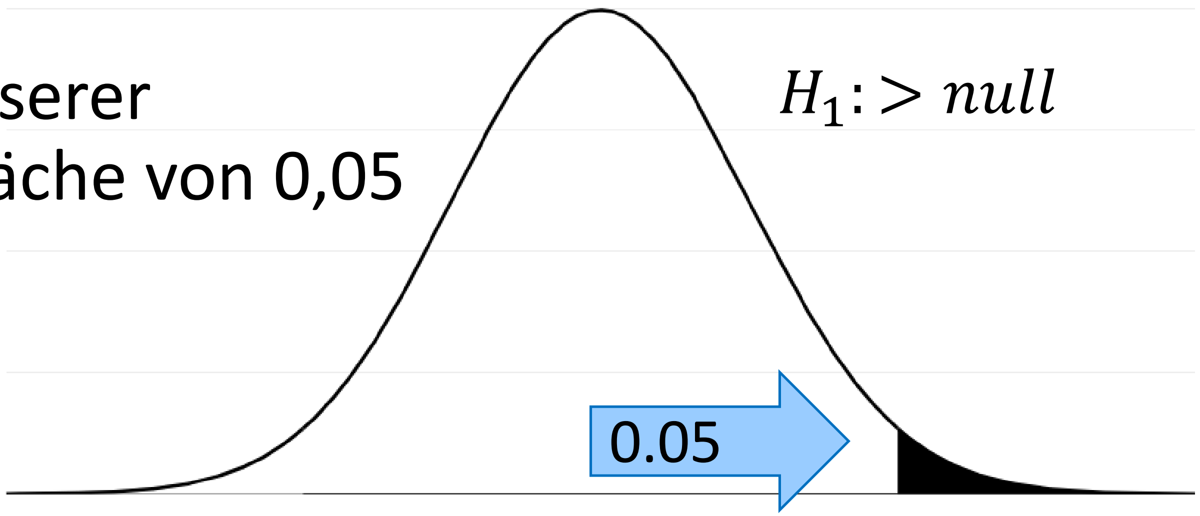
Hypothesentests Ablehnungsbereiche (Tails)

- Das Signifikanzniveau α ist das Gebiet innerhalb des Ablehnungsbereichs (Tail) unserer Nullhypothese.
- Wenn α = 0,05 und die Alternativhypothese *kleiner* als Null ist, dann hat der linke Ablehnungsbereich (left-tail) unserer Wahrscheinlichkeitskurve eine Fläche von 0,05



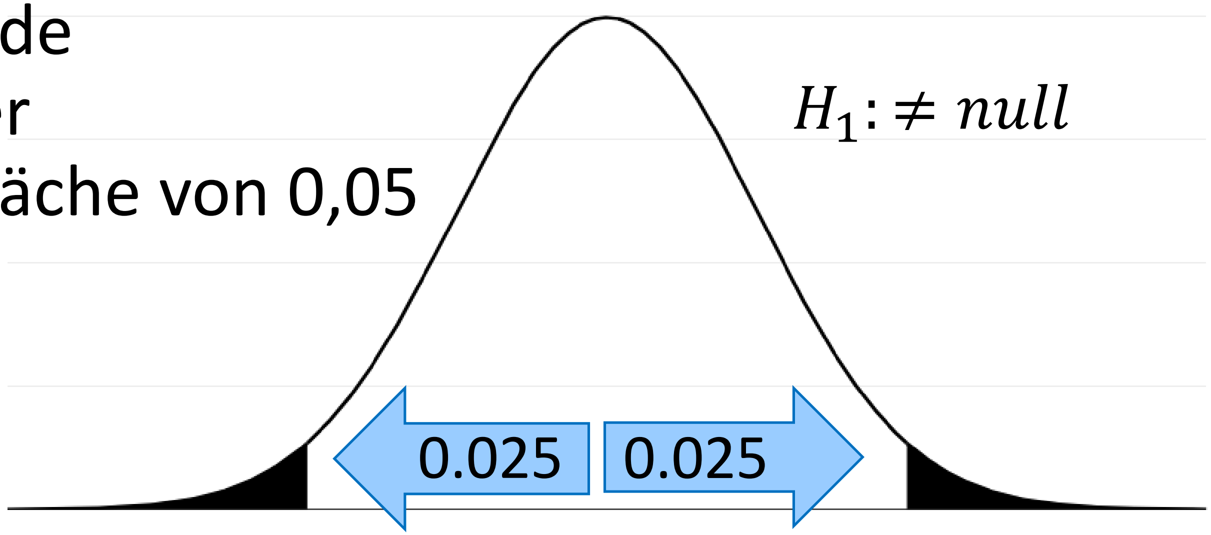
Hypothesentests Ablehnungsbereiche (Tails)

- Das Signifikanzniveau α ist das Gebiet innerhalb des Ablehnungsbereichs (Tail) unserer Nullhypothese.
- Wenn α = 0,05 und die Alternativhypothese *größer* als Null ist, dann hat der rechte Ablehnungsbereich (right-tail) unserer Wahrscheinlichkeitskurve eine Fläche von 0,05



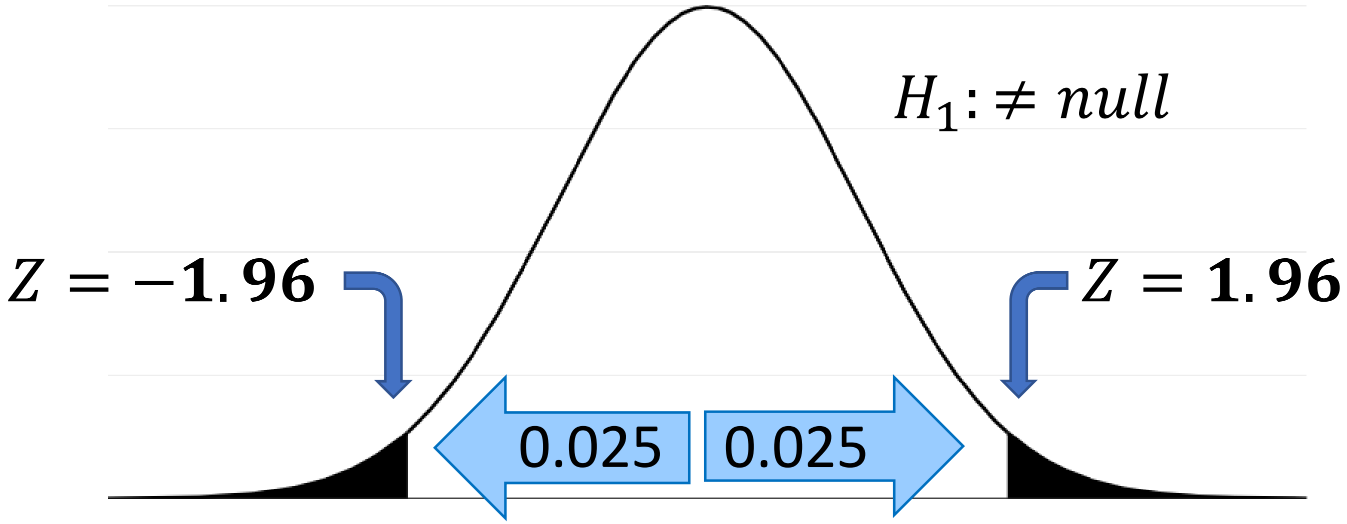
Hypothesentests Ablehnungsbereiche (Tails)

- Das Signifikanzniveau α ist das Gebiet innerhalb des Ablehnungsbereichs (Tail) unserer Nullhypothese.
- Wenn α = 0,05 und die Alternativhypothese *ungleich* Null ist, dann haben beide Ablehnungsbereich (tails) unserer Wahrscheinlichkeitskurve eine Fläche von 0,05



Hypothesentests Ablehnungsbereiche (Tails)

- Diese Bereiche legen unsere **kritischen Werte** oder Z-Werte fest:



Mittelwerttest vs. Häufigkeitstest

- In den nächsten zwei Lektionen werden wir einige vollständige Beispiele für Hypothesentests behandeln.

- Es gibt zwei Haupttypen von Tests:
  - **Mittelwerttest (test of means)**
  - **Häufigkeitstest (test of proportion)**

Mittelwerttest vs. Häufigkeitstest

- Jede dieser beiden Arten von Tests hat ihre eigene Teststatistik zur Berechnung.
- Lasst uns die Ausgangssituation für jeden Test besprechen, bevor wir die Beispiele in den kommenden Lektionen durchgehen..

Mittelwerttest vs. Häufigkeitstest

- **Mittelwert**

  Wenn wir einen **durchschnittlichen** oder spezifischen Wert in einer Population suchen, haben wir es mit Mittelwerten zu tun

- **Häufigkeiten**

  Wann immer wir etwas wie **"35%"** oder „**hauptsächlich**" sagen, handelt es sich um Häufigkeiten

Teststatistiken

- Wenn wir mit dem Mittelwert arbeiten:

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

Setzt voraus, dass wir die Standard-abweichung der Population kennen

- Wenn wir mit Häufigkeiten arbeiten:

$$Z = \frac{\hat{p} - p}{\sqrt{\frac{p \cdot q}{n}}} = \frac{\hat{p} - p}{\sqrt{\frac{p \cdot (1 - p)}{n}}}$$

Hypothesentests P-Wert-Test

1. In einem traditionellen Test:

- nimm das Signifikanzniveau α
- Verwende es, um den kritischen Wert zu bestimmen
- Vergleichst du die Teststatistik mit dem kritischen Wert

2. In einem P-Wert-Test:

- nimm die Teststatistik
- verwende sie, um den P-Wert zu bestimmen
- vergleiche den P-Wert mit dem Signifikanzniveau α

Hypothesentests P-Wert-Test

- *“If the P-value is low, the null must go!”*
- "Wenn der P-Wert niedrig ist, muss die Null gehen!„

lehne H_0 ab

- *“If the P-value is high, the null must fly!”*
- "Wenn der P-Wert hoch ist, muss die Null fliegen!„

nehme H_0 an

Hypothesentest Beispielübung #1

Hypothesentest Beispielübung #1 - Mittelwert

- Für dieses nächste Beispiel arbeiten wir auf der linken Seite der Wahrscheinlichkeitsverteilung mit negativen z-Werten
- Wir zeigen, wie man den Hypothesentest mit der traditionellen Methode und dann mit der P-Wert-Methode durchführt.

Hypothesentest Beispielübung #1 - Mittelwert

- Ein Unternehmen möchte seine Website-Leistung verbessern.
- Derzeit haben Seiten eine mittlere Ladezeit von 3,125 Sekunden mit einer Standardabweichung von 0,700 Sekunden.
- Sie beauftragen ein Beratungsunternehmen, um die Ladezeiten zu verbessern.

$$\boxed{\begin{aligned} \mu &= 3.125 \\ \sigma &= 0.700 \end{aligned}}$$

Hypothesentest Beispielübung #1 - Mittelwert

- Das Management möchte ein Konfidenzniveau von 99% haben
- Ein Beispiellauf von 40 neuen Seiten hat eine mittlere Ladezeit von 2,875 Sekunden.
- Sind diese Ergebnisse statistisch schneller als zuvor?

$$\mu = 3.125$$
$$\sigma = 0.700$$
$$\alpha = 0.01$$
$$n = 40$$
$$\bar{x} = 2.875$$

Hypothesentest Beispielübung #1 - Mittelwert

1. Die Nullhypothese

$$H_0: \mu \geq 3.125$$

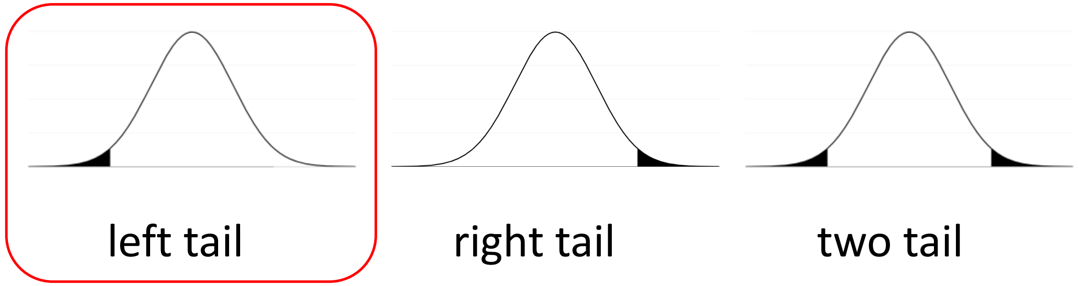
2. Die Alternativhypothese:

$$H_1: \mu < 3.125$$

3. Das Signifikanzniveau

$$\alpha = 0.01$$

4. Ermittle den Testtyp:



Hypothesentest Beispielübung #1 - Mittelwert

TRADITIONELLE METHODE:

5. Teststatistik (Prüfgröße):

$$Z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{2.875 - 3.125}{0.7/\sqrt{40}} = -2.259$$

6. Kritischer Wert:

Z-Tabelle bei 0.01 $z = -2.325$

$\mu = 3.125$
$\sigma = 0.700$
$\alpha = 0.01$
$n = 40$
$\bar{x} = 2.875$

$Z = -2.259$
$z = -2.325$

Hypothesentest Beispielübung #1 - Mittelwert

TRADITIONELLE METHODE:

7. Die Nullhypothese kann angenommen werden

$$\mu = 3.125$$
$$\sigma = 0.700$$
$$\alpha = 0.01$$
$$n = 40$$
$$\bar{x} = 2.875$$

$$Z = -2.259$$
$$z = -2.325$$

Bei $-2{,}259 > -2{,}325$, liegt die Teststatistik außerhalb des Ablehungsbereichs

Wir können nicht sagen, dass die neuen Webseiten statistisch schneller sind.

Hypothesentest Beispielübung #1 - Mittelwert

P-WERT METHODE:

5. Teststatistik:

$$Z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{2.875 - 3.125}{0.7/\sqrt{40}} = -2.259$$

6. P-Wert:

Z-Tabelle bei -2,26 $P = 0{,}0119$

$\mu = 3.125$
$\sigma = 0.700$
$\alpha = 0.01$
$n = 40$
$\bar{x} = 2.875$

$Z = -2.259$
$P = 0.0119$

Hypothesentest Beispielübung #1 - Mittelwert

P-WERT METHODE:

7. Die Nullhypothese kann angenommen werden

$$\text{Bei } 0{,}0119 > 0{,}01, \text{liegt der P} - \text{Wert}$$
$$\text{über dem Signifikanzniveau } \alpha$$

Wir können nicht sagen, dass die neuen Webseiten statistisch schneller sind.

$\mu = 3.125$
$\sigma = 0.700$
$\alpha = 0.01$
$n = 40$
$\bar{x} = 2.875$

$Z = -2.259$
$z = -2.325$

Hypothesentest
Beispielübung #2

Hypothesentest Beispielübung #2 - Häufigkeit

- Eine Videospielfirma befragt 400 ihrer Kunden und stellt fest, dass 58% der Stichprobe Teenager sind.
- Ist es daher richtig zu behaupten, dass die meisten Kunden des Unternehmens Teenager sind?

Hypothesentest Beispielübung #2 - Häufigkeit

1. Setze die Nullhypothese: $H_0: P \leq 0.50$

2. Lege die Alternativhypothese fest: $H_1: P > 0.50$

3. Berechne die Teststatistik (Prüfgröße):

$$Z = \frac{\hat{p}-p}{\sqrt{\frac{p \cdot q}{n}}} = \frac{0.58-0.50}{\sqrt{\frac{0.50(1-0.50)}{400}}} = \frac{0.08}{0.025} = \mathbf{3.2}$$

Hypothesentest Beispielübung #2 - Häufigkeit

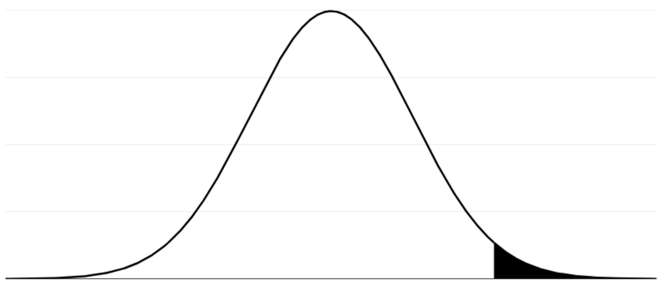
4. Lege ein Signifikanzniveau fest: $\alpha = 0.05$

5. Entscheide dich, welcher Ablehnungsbereich betroffen ist:

   $H_1 : P > 0.50$ **bedeutet einen Test des rechten Bereichs**

6. Schlage den kritischen Wert nach:

   $Z = \mathbf{1.645}$

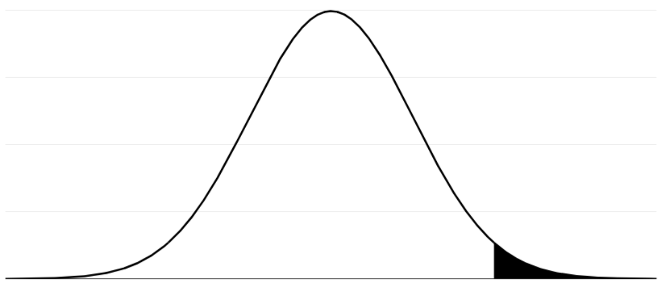


Kritischer Wert = 1.645

Teststatistik = 3.2

Hypothesentest Beispielübung #2 - Häufigkeit

7. Basierend auf der Stichprobe lehnen wir die Nullhypothese ab und unterstützen die Behauptung, dass die meisten Kunden Teenager sind.



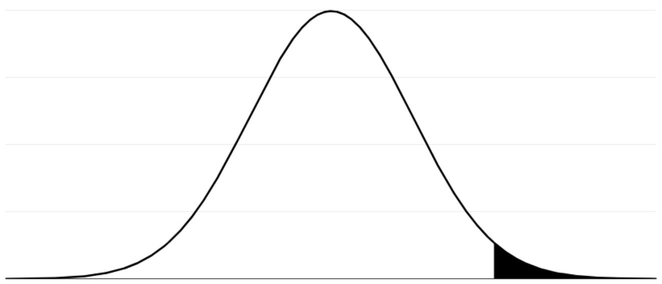
Kritischer Wert = 1.645

Teststatistik = 3.2

Hypothesentest Beispielübung #2 - Häufigkeit

- **HINWEIS: Die Größe der Stichprobe ist wichtig!**

Wenn wir mit einer Stichprobengröße von **40** statt 400 begonnen hätten, wäre unsere Teststatistik nur **1,01**, und wir würden die Nullhypothese nicht verwerfen.



Kritischer Wert = 1.645

Teststatistik = 3.2

Fehler erster (Typ I) und zweiter Art (Typ II)

Typ I und Typ II Fehler

- Oft werden in medizinischen Bereichen (und anderen wissenschaftlichen Gebieten) Hypothesentests verwendet, um gegen Ergebnisse zu testen, bei denen die "Wahrheit" bereits bekannt ist.
- Zum Beispiel, wird ein neuer diagnostischer Test für Krebspatienten getestet, deren Diagnose bereits mit anderen Mitteln festgestellt wurde.

Typ I und Typ II Fehler

- In dieser Situation wissen wir ja bereits, ob die Nullhypothese wahr oder falsch ist.
- In diesen Situationen, in denen die "Wahrheit" bereits bekannt ist, wissen wir, dass es möglich ist, mit unseren Ergebnissen Fehler zu erzeugen.

Typ I und Typ II Fehler

- Diese Art der Analyse tritt so häufig auf, dass diese Fehler bereits spezifische Namen haben:
- Fehler erster Art (Typ I Error)
- Fehlertyp zweiter Art (Typ II Error)

Typ I und Typ II Fehler

- Wenn wir eine Nullhypothese **ablehnen**, die jedoch richtig ist und unterstützt werden sollte, haben wir einen **Fehler erster Art** begangen

$$\boldsymbol{H_0}: es\ brennt\ nicht$$

Ich betätige den Feueralarm, um zu prüfen, dass es wirklich kein Feuer gab.

Typ I und Typ II Fehler

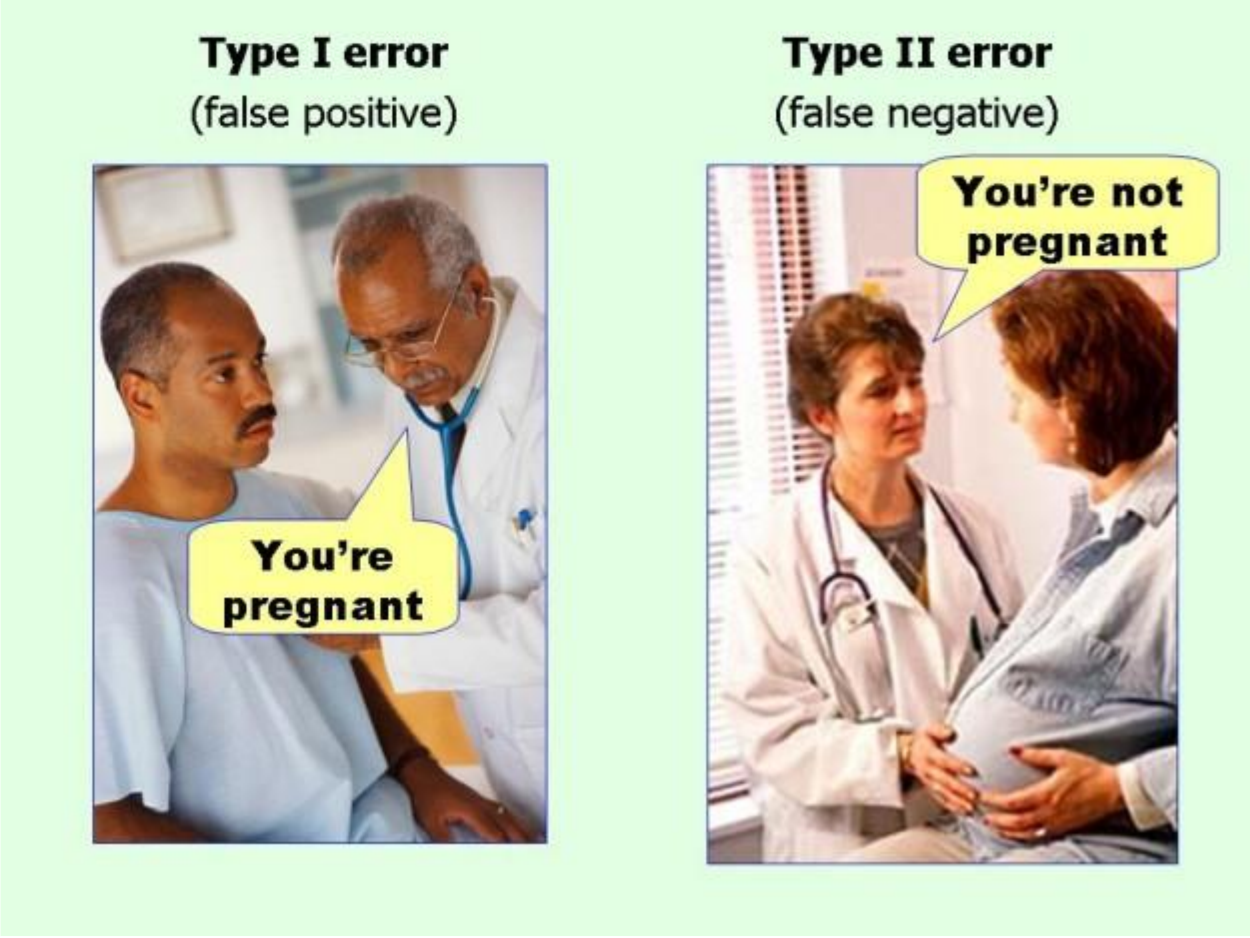
- Wenn wir eine Nullhypothese **nicht ablehnen**, die jedoch falsch ist und abgelehnt werden sollte, haben wir einen **Fehler zweiter Art** begangen

$$\boldsymbol{H_0}: es\ brennt\ nicht$$

Ich betätige den Feueralarm nicht, nur um festzustellen, dass es brennt.

Typ I und Typ II Fehler

H_0: *nicht schwanger*
H_1: *schwanger*



Studentsche t-Verteilung

Zweck eines t-Test

- Erinnern wir uns: wenn wir Z-Werte mit einer Normalverteilung verwendeten, dann müssen wir die Standardabweichung (Sigma) der Population kennen, um Z zu berechnen.
- Aber was ist, wenn wir in der realen Welt die Standardabweichung der Population nicht kennen?

Studentsche t-Verteilung

- Entwickelt von William Sealy Gossett während er in der Guinness Brauerei arbeitete
- Veröffentlicht unter dem Pseudonym "Student" weil Guinness ihm verbot, seinen Namen zu verwenden.
- Ziel war es, die beste Gerste aus kleinen Proben auszuwählen, obwohl die Standardabweichung der Population unbekannt war!

Zweck eines t-Test

- Unter Verwendung der T-Tabelle bestimmt der Studentsche-T-Test, ob es einen signifikanten Unterschied zwischen zwei Datensätzen gibt
- Aufgrund von Varianz und Ausreißern reicht es nicht aus, nur die Mittelwerte zu vergleichen
- Ein t-Test berücksichtigt auch die Stichprobenvarianz

Typen vom t-Test

- **Einstichproben t-Test (one-sample t-test)**

Testet die Nullhypothese, dass der Mittelwert der Population gleich einem spezifischen Wert μ ist, basierend auf dem Stichprobenmittelwert.

Typen vom t-Test

- **Einstichproben t-Test Beispiel**

Wir möchten prüfen, ob eine Stichprobe von Schülern die gleichen durchschnittlichen Testergebnisse wie die Gesamtschülerzahl aufweist.

Typen vom t-Test

- **Zweistichproben-t-Test (indipendent two-sample t-test)**

Prüft die Nullhypothese, dass zwei Stichprobenwerte $\bar{x}_1$ und $\bar{x}_2$ gleich sind.

Typen vom t-Test

- **Zweistichproben-t-Test (indipendent two-sample t-test) Beispiel**

Wir möchten überprüfen, ob die durchschnittlichen Testergebnisse von zwei separaten Stichproben von Schülern einen statistisch signifikanten Unterschied aufweisen?

Typen vom t-Test

- **Gepaarter/abhängiger t-Test (Dependent, paired-sample t-test)**

Wird verwendet, wenn die Stichproben abhängig sind:

- eine Probe wurde zweimal getestet (wiederholte Messungen)
- zwei Proben wurden verglichen oder „gepaart“

Typen vom t-Test

- **Gepaarter/abhängiger t-Test (Dependent, paired-sample t-test)**

Wir möchten prüfen, ob die gleiche Gruppe von Schülern vor dem Vorbereitungskurs und nach dem Vorbereitungskurs ihre Testergebnisse verbessern konnte?

Wir müssen beachten, dass wir dieselbe Stichprobe von Schülern verwenden (diese also abhängig ist)

t-Test

Wie bei Z-Statistik berechnen wir die t-Statistik (Zufallsvariable).

Einstichproben-t-Test

- Kalkuliere die T-Statistik

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

$\bar{x}$ = Stichprobenmittel

μ = Mittelwert Population

s = Standardfehler

n = Stichprobengröße

Einstichproben-t-Test

- Kalkuliere die T-Statistik

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

Wir kennen die Standardabweichung der Population nicht, daher basieren wir diese Metrik auf der Stichprobe.

$\bar{x}$ = Stichprobenmittel

μ = Mittelwert Population

s = Standardfehler

n = Stichprobengröße

Einstichproben-t-Test

- Wie bei den Z-Werten vergleichen wir die t-Werte nun mit einer t-Wert-Tabelle.
- Diese Werte hängen ab von:
  - Den Freiheitsgrade (basierend auf der Stichprobengröße n)
  - Dem gewählten Signifikanzniveau (Standard 0,05)

Einstichproben-t-Test

- **Vergleiche mit einem t-Wert**

$$t \lesseqgtr t_{n-1,\alpha}$$

t = t-Wert

$t_{n-1,\alpha}$ = kritischer t-Wert

$n-1$ = Freiheitsgrad

α = Signifikanzniveau

Zweistichproben-t-Test

- Die Berechnung der t-Werte unterscheidet sich für die folgenden Szenarien geringfügig:
    - gleiche Stichprobengröße, gleiche Varianz
    - ungleiche Stichprobengrößen, gleiche Varianz
    - **gleiche oder ungleiche Stichprobengrößen, ungleiche Varianz**

Zweistichproben-t-Test

- Wenn man mit zwei Stichproben arbeitet und versucht, sie mit einem t-Test zu vergleichen, ist es oft nützlich, sich den t-Test als Verhältnis von Signal (Stichprobenmittel) zu Rauschen (Stichprobenvariabilität) vorzustellen (Signal-Rausch-Verhältnis oder Störabstand).

Zweistichproben-t-Test

- Kalkuliere die t-Statistik:

$$t = \frac{Signal}{Rauschen} = \frac{Stichprobenmitteldifferenz}{Stichprobenvariabilität} = \frac{\overline{x_1} - \overline{x_2}}{\sqrt{\frac{{s_1}^2}{n_1} + \frac{{s_2}^2}{n_2}}}$$

$\overline{x_1},\ \overline{x_2}$	= Stichprobenmittel
${s_1}^2, {s_2}^2$	= Stichprobenvarianz
$n_1,\ n_2$	= Stichprobengröße

Zweistichproben-t-Test

- **Vergleiche mit einem t-Wert**

$$t \lesseqgtr t_{df,\alpha}$$

Da wir zwei potentiell ungleich große Proben mit unterschiedlichen Varianzen haben, ist die Bestimmung der Freiheitsgrade ein wenig komplizierter.

t = t-Wert
$t_{df,\alpha}$ = kritischer t-Wert
df = Freiheitsgrade
α = Signifikanzlevel

Freiheitsgrade im Zweistichproben-t-Test

- Welch-Satterthwaite Formel für df (Freiheitsgrade)

$$df = \frac{\left(\frac{{s_1}^2}{n_1}+\frac{{s_2}^2}{n_2}\right)^2}{\frac{1}{n_1-1}\left(\frac{{s_1}^2}{n_1}\right)^2+\frac{1}{n_2-1}\left(\frac{{s_2}^2}{n_2}\right)^2}$$

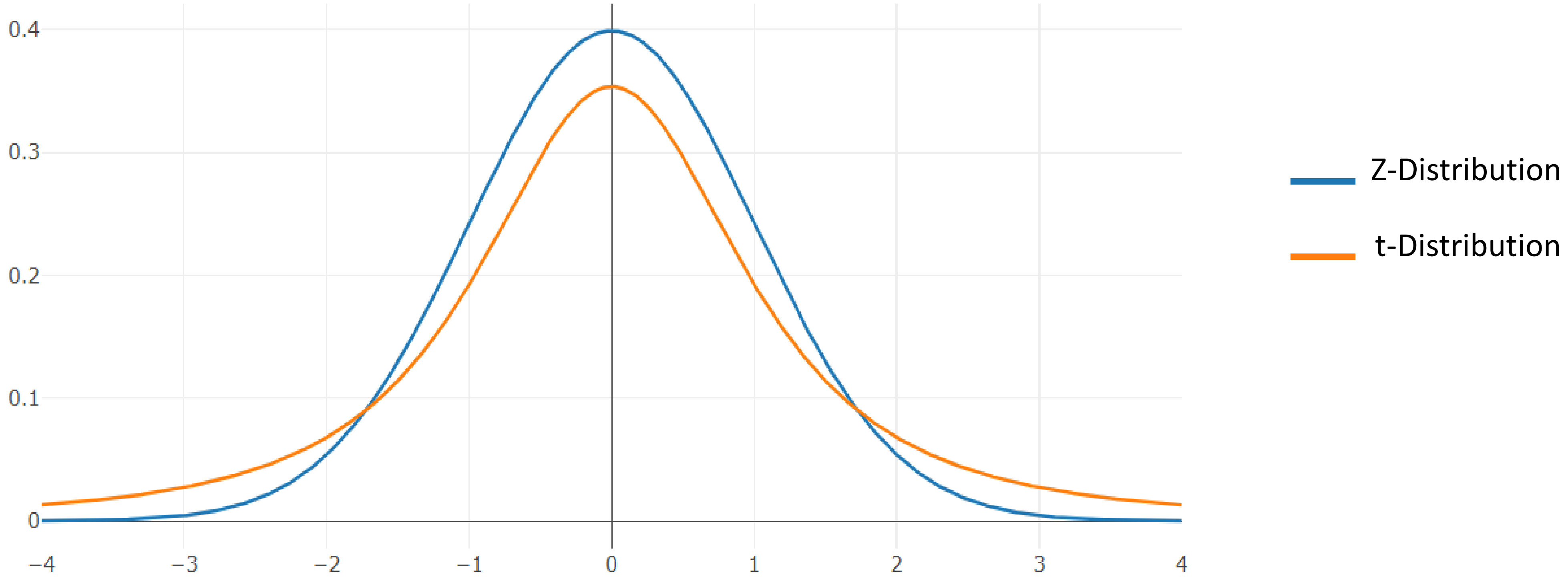
Freiheitsgrade

- Die allgemeine Formel lautet:

$$df = n_1 + n_2 - 2$$

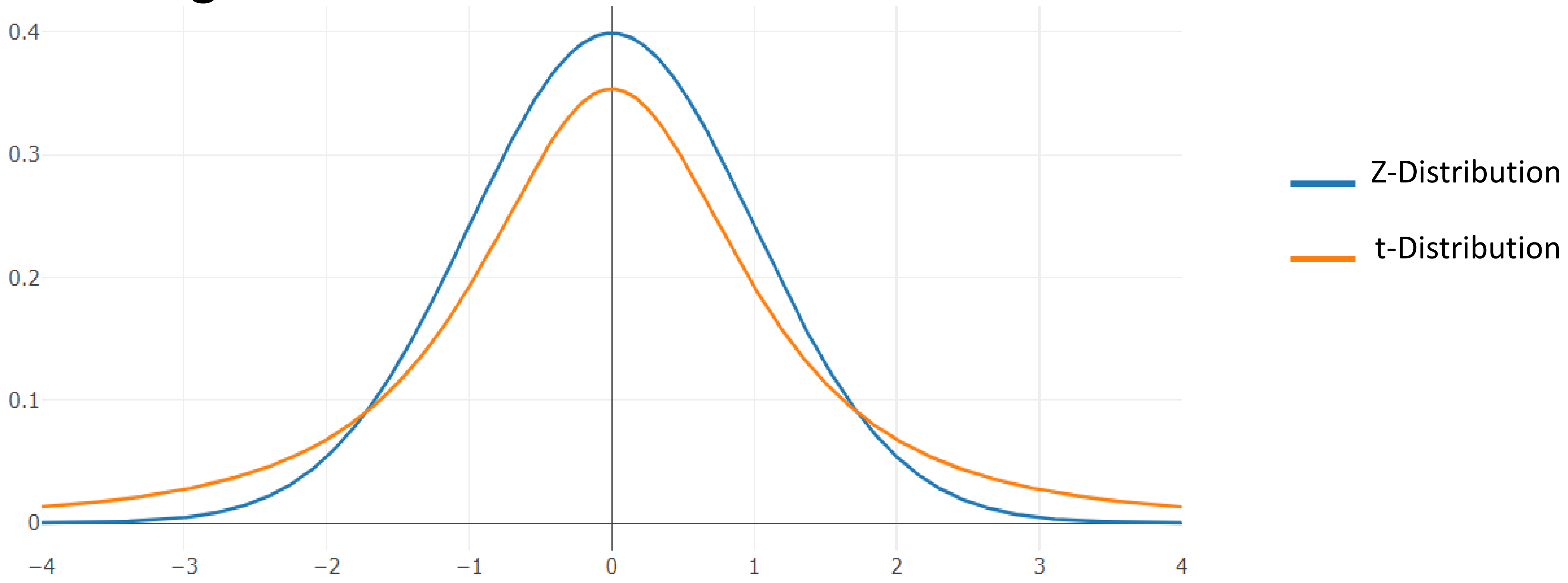
Studentsche t-Verteilung

t-Distributionen haben dickere Ausläufer als normale Z-Distributionen



Studentsche t-Verteilung

Sie nähern sich einer normalen Verteilung an, bei steigenden Freiheitsgraden



Studentsche t-Verteilung
Beispielübung

t-Test Beispiel

- Ein Automobilhersteller hat zwei Werke, die das gleiche Auto produzieren.

t-Test Beispiel

- Sie sind gezwungen, eines der Werke zu schließen.

t-Test Beispiel

- Das Unternehmen möchte nun wissen, ob es einen signifikanten Produktionsunterschied zwischen den beiden Werken gibt.

Studentischer t-Test Beispiel

- Die tägliche Produktion über 10 Tage ist wie folgt:

Werk A	Werk B
1184	1136
1203	1178
1219	1212
1238	1193
1243	1226
1204	1154
1269	1230
1256	1222
1156	1161
1248	1148

Studentischer t-Test Beispiel

- Zuerst vergleichen wir die Stichproben

$$\boldsymbol{\bar{x}_A - \bar{x}_B} = 1222 - 1186 = \mathbf{36}$$

Anhand dieser Stichproben sieht es so aus, als ob Anlage A 36 Autos mehr pro Tag produziert als Anlage B.

	Werk A	Werk B
	1184	1136
	1203	1178
	1219	1212
	1238	1193
	1243	1226
	1204	1154
	1269	1230
	1256	1222
	1156	1161
	1248	1148
	$\bar{x}_A$	$\bar{x}_B$
Mittel	1222	1186

Studentischer t-Test Beispiel

- Sind 36 Autos mehr genug, um zu sagen, dass die Werke sich unterscheiden?

	Werk A	Werk B
	1184	1136
	1203	1178
	1219	1212
	1238	1193
	1243	1226
	1204	1154
	1269	1230
	1256	1222
	1156	1161
	1248	1148
	$\bar{x}_A$	$\bar{x}_B$
Mittel	1222	1186

$$H_0: X_A \leq X_B$$

$$H_1: X_A > X_B$$

einseitiger Test

$$(10 + 10 - 2) = 18\, degrees\ of\ freedom$$

Studentischer t-Test Beispiel

Berechne die Varianz

A	(x-1222)	$(x-1222)^2$
1184	-38	1444
1203	-19	361
1219	-3	9
1238	16	256
1243	21	441
1204	-18	324
1269	47	2209
1256	34	1156
1156	-66	4356
1248	26	676
		11232

$$s^2 = \frac{\Sigma(x-\bar{x})^2}{n-1}$$

$\Sigma(x-1222)^2$	11232
$\frac{\Sigma(x-1222)^2}{9}$	**1248**

Werk A	Werk B
1184	1136
1203	1178
1219	1212
1238	1193
1243	1226
1204	1154
1269	1230
1256	1222
1156	1161
1248	1148

	$\bar{x}_A$	$\bar{x}_B$
Mittel	1222	1186
Varianz	1248	1246

Studentischer t-Test Beispiel

Berechne den t-Wert

$$= \frac{\overline{x_1} - \overline{x_2}}{\sqrt{\frac{{s_1}^2}{n_1} + \frac{{s_2}^2}{n_2}}}$$

$$= \frac{36}{\sqrt{\frac{1248}{10} + \frac{1246}{10}}} = \frac{36}{15.792}$$

= 2,28

	Werk A	Werk B
	1184	1136
	1203	1178
	1219	1212
	1238	1193
	1243	1226
	1204	1154
	1269	1230
	1256	1222
	1156	1161
	1248	1148
	$\bar{x}_A$	$\bar{x}_B$
Mittel	1222	1186
Varianz	1248	1246

t-Test Beispiel

- Suche nun den kritischen Wert in der t-Tabelle
- einseitiger Test (one-tail)
- 95% Vertrauen
- 18 Freiheitsgrade

cum. prob	$t_{.90}$	$t_{.95}$	$t_{.975}$	$t_{.99}$	$t_{.995}$
one-tail	0.10	0.05	0.025	0.01	0.005
two-tails	0.20	0.10	0.05	0.02	0.01
17	1.333	1.740	2.110	2.567	2.898
18	1.330	1.734	2.101	2.552	2.878
19	1.328	1.729	2.093	2.539	2.861

- kritischer Wert = 1,734

Studentischer t-Test Beispiel

Vergleiche nun unseren t-Wert (2,28) mit dem kritischen Wert (1,734):

$$2.28 > 1.734$$

Da unser berechneter t-Wert *größer* als der kritische Wert ist, lehnen wir die Nullhypothese ab.

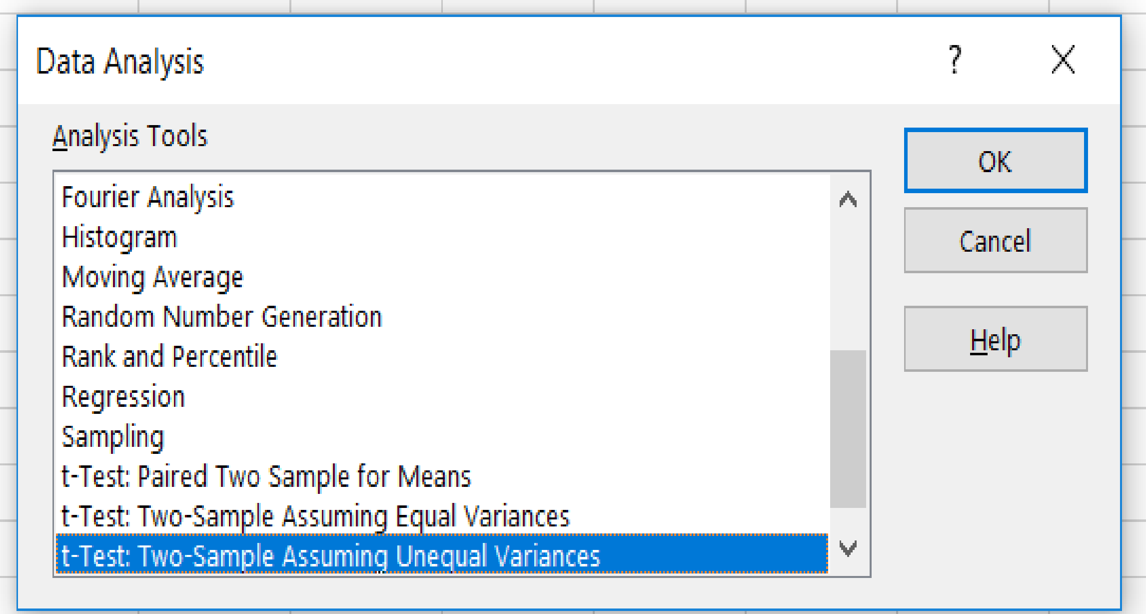
Werk A	Werk B
1184	1136
1203	1178
1219	1212
1238	1193
1243	1226
1204	1154
1269	1230
1256	1222
1156	1161
1248	1148

t-Test Beispiel

- Wir glauben demnach mit 95%-iger Sicherheit, dass Anlage A mehr Autos pro Tag produziert als Anlage B.
- Wir beschließen damit, das Werk B zu schließen.

t-Test in Excel

	A	B	C
1	t-Test: Two-Sample Assuming Unequal Variances		
2			
3		*Variable 1*	*Variable 2*
4	Mean	1186	1222
5	Variance	1246	1248
6	Observations	10	10
7	Hypothesized Mean Difference	0	
8	df	18	
9	t Stat	-2.279577051	
10	P(T<=t) one-tail	0.017522528	
11	t Critical one-tail	1.734063607	
12	P(T<=t) two-tail	0.035045056	
13	t Critical two-tail	2.10092204	
14			



t-Test in Python

```
>>> from scipy.stats import ttest_ind
>>> a = [1184, 1203, 1219, … 1248]
>>> b = [1136, 1178, 1212, … 1148]
>>> ttest_ind(a,b).statistic
2.2795770510504845
>>> ttest_ind(a,b).pvalue/2
0.017522528133638322
```

Als nächstes:
ANOVA