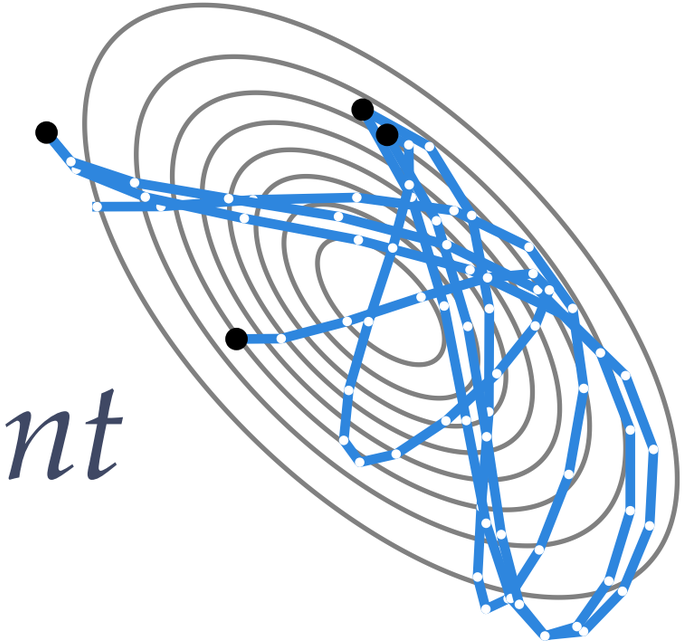
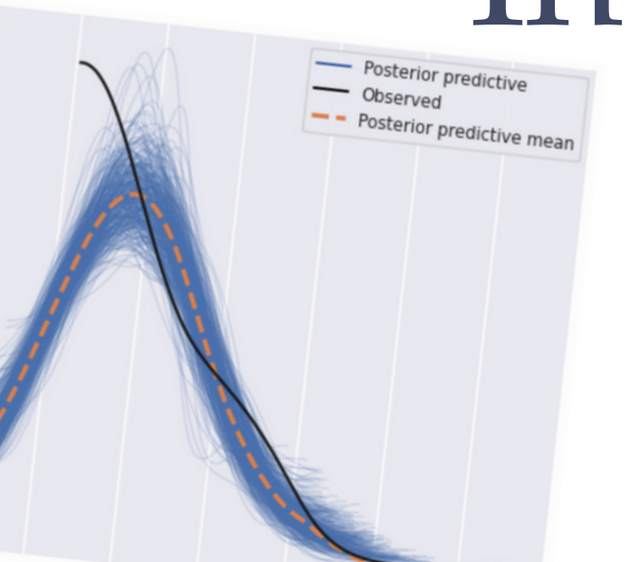
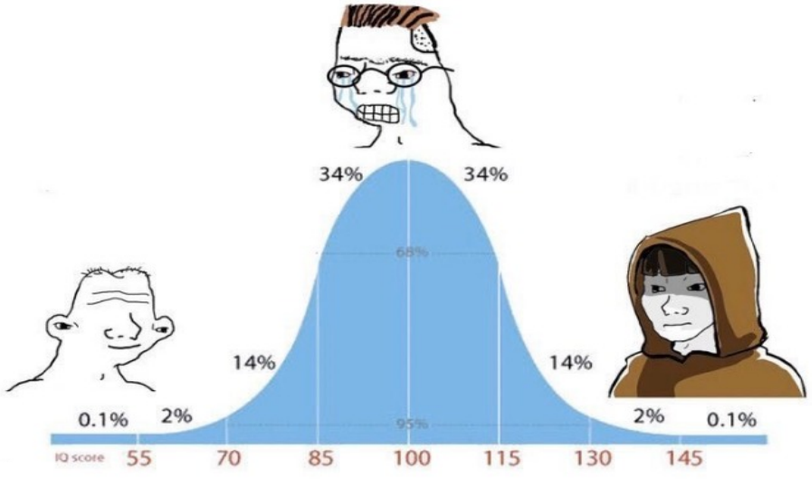


Wahrscheinlich interessant

Statistik für Informatiker:innen





Fachspezifische Bestimmungen mit dem Abschluss Bachelor of Science (Erwerb von 180 ECTS-Punkten)

an der Julius-Maximilians-Universität Würzburg

vom 28. September 2015

In der Fassung der Änderungssatzung vom 21. Mai 2025
le: http://www.uni-wuerzburg.de/amtl_veroeffentlichungen/2025-60)

Fachspezifische Bestimmungen mit dem Abschluss Bachelor of Science (Erwerb von 180 ECTS-Punkten)

an der Julius-Maximilians-Universität Würzburg

vom 28. September 2015

In der Fassung der Änderungssatzung vom 21. Mai 2025

le: http://www.uni-wuerzburg.de/amtl_veroeffentlichungen/2025

Kurzbezeichnung	Version	Modultitel (Deutsch/Englisch)	Art der LV (SWS)	ECTS-Punkte	Dauer (in Semestern)	TN und Auswahl	Bewertung	Art und Umfang der Erfolgsüberprüfung	Prüfungssprache	Zuvor bestandene Module	1) Bonus 2) LV-Sp 3) Prüfu 4) weiter Vorauss 5) Zusat Dauer, 6) Sonst
Schlüsselqualifikationen (20 ECTS-Punkte)											
Allgemeine Schlüsselqualifikationen (5 ECTS-Punkte)											
Neben den nachfolgend aufgeführten Modulen können auch Module aus dem von der JMU angebotenen Pool der allgemeinen Schlüsselqualifikationen (ASQ-Pool) werden.											
10-I-TUT1	**2015-WS**	**Tutorentätigkeit 1** **Tutor Activity 1**	T(2)	2	1-2		B/NB	Endbericht über Tutorentätigkeit im Umfang von 5-10 S.			
10-I-TUT2	**2015-WS**	**Tutorentätigkeit 2** **Tutor Activity 2**	T(2)	2	1-2		B/NB	Endbericht über Tutorentätigkeit im Umfang von 5-10 S.			
10-I-TUT3	**2015-WS**	**Tutorentätigkeit 3** **Tutor Activity 3**	T(2)	2	1-2		B/NB	Endbericht über Tutorentätigkeit im Umfang von 5-10 S.			
Fachspezifische Schlüsselqualifikationen (15 ECTS-Punkte)											
10-I-ASV	**2025-WS**	**Angewandte Statistik und Visualisierung** **Applied Statistics and Visualization**	V(1) + P(2)	3	1		B/NB	a) Portfolioprüfung: (Gesamtumfang ca. 75 h) oder b) Klausur (ca. 60-75 Min.)[1]	Deutsch und/oder Englisch		1) Bonu
10-I-PV	**2025-WS**	**Projektvorstellung** **Project Presentation**	S(3)	2	1		NUM	Präsentation eines selbstentwickelten Projektes analog zu einer Messepräsentation für informatikkundige	Deutsch und/oder Englisch		1) Bonu

Fachspezifische Bestimmungen mit dem Abschluss Bachelor of Science (Erwerb von 180 ECTS-Punkten)

an der Julius-Maximilians-Universität Würzburg

vom 28. September 2015

In der Fassung der Änderungssatzung vom 21. Mai 2025

le: http://www.uni-wuerzburg.de/amtl_veroeffentlichungen/2025

Kurzbezeichnung	Version	Modultitel (Deutsch/Englisch)	Art der LV (SWS)	ECTS-Punkte	Dauer (in Semestern)	TN und Auswahl	Bewertung	Art und Umfang der Erfolgsüberprüfung	Prüfungssprache	Zuvor bestandene Module	1) Bonus 2) LV-Sp 3) Prüfu 4) weiter Vorauss 5) Zusat Dauer, 6) Sonst
Schlüsselqualifikationen (20 ECTS-Punkte)											
Allgemeine Schlüsselqualifikationen (5 ECTS-Punkte)											
Neben den nachfolgend aufgeführten Modulen können auch Module aus dem von der JMU angebotenen Pool der allgemeinen Schlüsselqualifikationen (ASQ-Pool) werden.											
10-I-TUT1	**2015-WS**	**Tutorentätigkeit 1** **Tutor Activity 1**	T(2)	2	1-2		B/NB	Endbericht über Tutorentätigkeit im Umfang von 5-10 S.			
10-I-TUT2	**2015-WS**	**Tutorentätigkeit 2** **Tutor Activity 2**	T(2)	2	1-2		B/NB	Endbericht über Tutorentätigkeit im Umfang von 5-10 S.			
10-I-TUT3	**2015-WS**	**Tutorentätigkeit 3** **Tutor Activity 3**	T(2)	2	1-2		B/NB	Endbericht über Tutorentätigkeit im Umfang von 5-10 S.			
Fachspezifische Schlüsselqualifikationen (15 ECTS-Punkte)											
10-I-ASV	**2025-WS**	**Angewandte Statistik und Visualisierung** **Applied Statistics and Visualization**	V(1) + P(2)	3	1		B/NB	a) Portfolioprüfung: (Gesamtumfang ca. 75 h) oder b) Klausur (ca. 60-75 Min.)[1]	Deutsch und/oder Englisch		1) Bonu
10-I-PV	**2025-WS**	**Projektvorstellung** **Project Presentation**	S(3)	2	1		NUM	Präsentation eines selbstentwickelten Projektes analog zu einer Messepräsentation für informatikkundige	Deutsch und/oder Englisch		1) Bonu

Fachspezifische Bestimmungen mit dem Abschluss Bachelor of Science (Erwerb von 180 ECTS-Punkten)

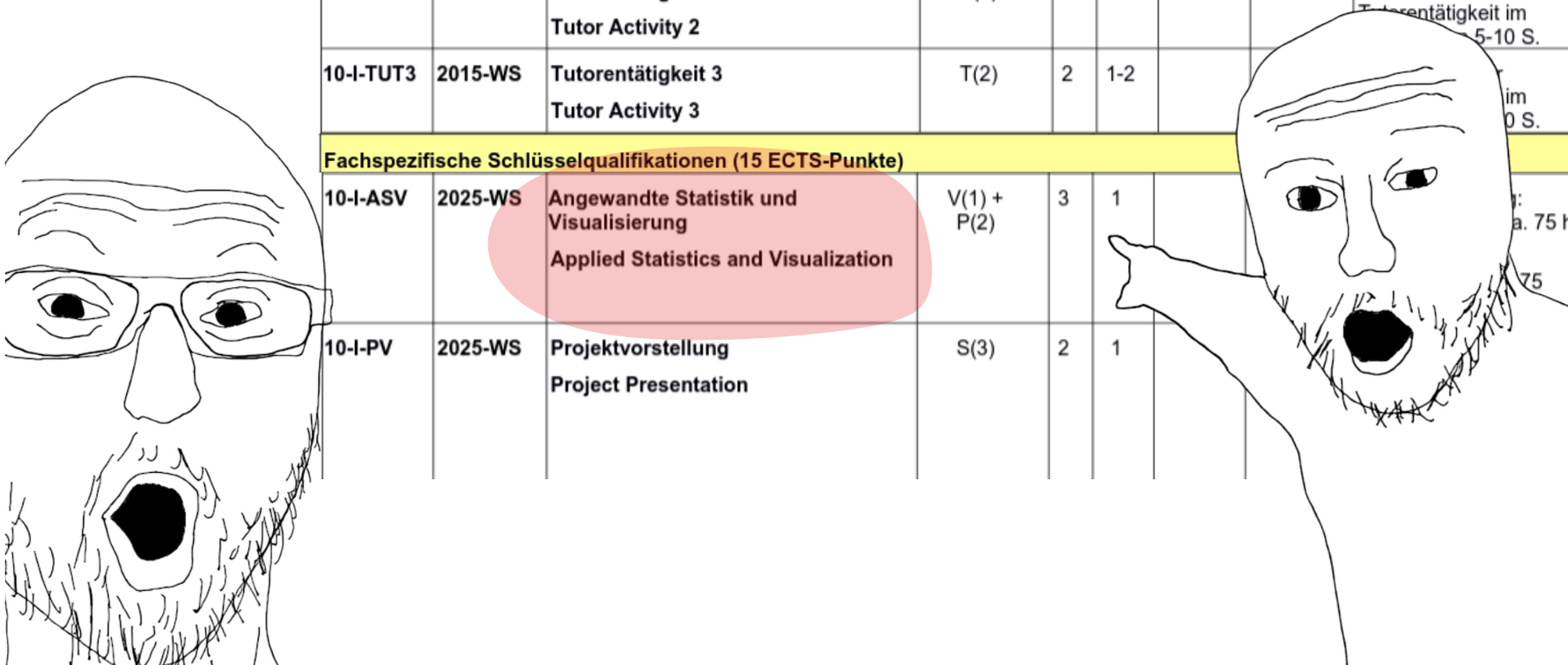
an der Julius-Maximilians-Universität Würzburg

vom 28. September 2015

In der Fassung der Änderungssatzung vom 21. Mai 2025

le: http://www.uni-wuerzburg.de/amtl_veroeffentlichungen/2025

Kurzbezeichnung	Version	Modultitel (Deutsch/Englisch)	Art der LV (SWS)	ECTS-Punkte	Dauer (in Semestern)	TN und Auswahl	Bewertung	Art und Umfang der Erfolgsüberprüfung	Prüfungssprache	Zuvor bestandene Module	1) Bonus 2) LV-Sp 3) Prüfu 4) weiter Vorauss 5) Zusat Dauer, 6) Sonst
Schlüsselqualifikationen (20 ECTS-Punkte)											
Allgemeine Schlüsselqualifikationen (5 ECTS-Punkte)											
Neben den nachfolgend aufgeführten Modulen können auch Module aus dem von der JMU angebotenen Pool der allgemeinen Schlüsselqualifikationen (ASQ-Pool) werden.											
10-I-TUT1	**2015-WS**	**Tutorentätigkeit 1** **Tutor Activity 1**	T(2)	2	1-2		B/NB	Endbericht über Tutorentätigkeit im Umfang von 5-10 S.			
10-I-TUT2	**2015-WS**	**Tutorentätigkeit 2** **Tutor Activity 2**	T(2)	2	1-2		B/NB	Endbericht über [illegible] 5-10 S.			
10-I-TUT3	**2015-WS**	**Tutorentätigkeit 3** **Tutor Activity 3**	T(2)	2	1-2			[illegible] im [illegible] 0 S.			
Fachspezifische Schlüsselqualifikationen (15 ECTS-Punkte)											
10-I-ASV	**2025-WS**	**Angewandte Statistik und Visualisierung** **Applied Statistics and Visualization**	V(1) + P(2)	3	1			[illegible]: [illegible]a. 75 h] [illegible] 75	Deutsch und/oder Englisch		1) Bonu
10-I-PV	**2025-WS**	**Projektvorstellung** **Project Presentation**	S(3)	2	1				Deutsch und/oder Englisch		1) Bonu



Startseite / Sommersemester 2026 / Grundständige Studiengänge (Bachelor, ...) / Informatik

SS26:Angewandte Statistik und Visualisierung

Kurs Bewertungen Aktivitäten Mehr

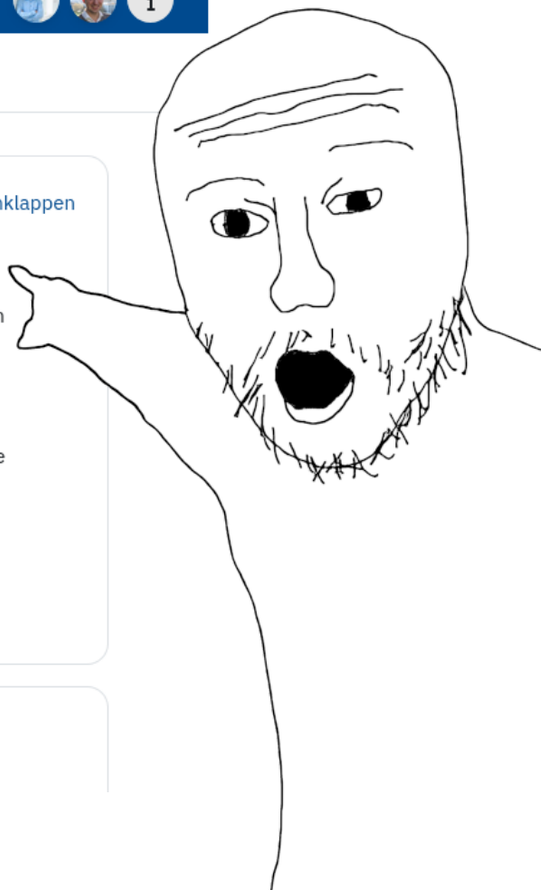
Angewandte Statistik und Visualisierung

Alles einklappen

Das Modul "Angewandte Statistik und Visualisierung" bietet den Studierenden eine Einführung in statistische Methoden und Techniken sowie deren Anwendung in verschiedenen Bereichen. Es konzentriert sich darauf, statistische Konzepte zu verstehen und sie auf praktische Probleme anzuwenden. Das Modul legt auch Wert auf die Fähigkeit der Studierenden, Daten visuell zu präsentieren und zu interpretieren.

- Deskriptive Statistik: Maßzahlen, Streuungsmaße, Verteilungen, Graphiken
- Wahrscheinlichkeitstheorie: Grundbegriffe, diskrete und kontinuierliche Verteilungen
- Schließende Statistik: Schätzungen, Hypothesentests, parametrische und nichtparametrische Tests, Konfidenzintervalle und deren Visualierung
- Regressionsanalyse: z.B. einfache lineare Regression
- Bootstrapping-Verfahren: z.B. für Konfidenzintervalle
- Datenvisualisierungstechniken: Diagramme, Plots und deren Anwendungsbereiche
- Softwareanwendungen: Nutzung von Open Source Software für Statistik und Visualisierung

Ankündigungen und Forum



Ende der Bildschirmpräsentation.
Zum Beenden klicken.

Wo bin ich hier gelandet?

Habe ich gewonnen?

Wo bin ich hier gelandet?

Habe ich gewonnen?

Scientist: *„(...) comparing the mean runtimes on the benchmark, we can conclude that our proposed pruning algorithm is an improvement to the CPLEX baseline (...)“*

Wo bin ich hier gelandet?

Habe ich gewonnen?

Scientist: *„(...) comparing the mean runtimes on the benchmark, we can conclude that our proposed pruning algorithm is an improvement to the CPLEX baseline (...)“*

Reviewer *„The reported difference is marginal and entirely consistent with random variation. Absent any hypothesis test, confidence interval, or effect size, the word 'improvement' here carries no scientific weight. Reject.“*

Wo bin ich hier gelandet?

Habe ich gewonnen?

Scientist: *„(...) comparing the mean runtimes on the benchmark, we can conclude that our proposed pruning algorithm is an improvement to the CPLEX baseline (...)“*

Reviewer *„The reported difference is marginal and entirely consistent with random variation. Absent any hypothesis test, confidence interval, or effect size, the word 'improvement' here carries no scientific weight. Reject.“*

Was Scientist hätte sagen müssen: *„A paired Wilcoxon signed-rank test with* $p = 0.00025$ *concludes that (...)“*

Wo bin ich hier gelandet?

- (Grobe) Motivation: Was muss ich schreiben damit die Reviewer mich nicht haten?
- Part A: Überblick über die klassische (frequentistische) inferenzielle Statistik
  - Hilfe zur Selbsthilfe
  - Theorie vs. Praxis
- Part B: Neue coole bayesian statistics (DIE ZUKUNFT)

Wo bin ich hier gelandet?

- (Grobe) Motivation: Was muss ich schreiben damit die Reviewer mich nicht haten?
- Part A: Überblick über die klassische (frequentistische) inferenzielle Statistik
  - Hilfe zur Selbsthilfe
  - Theorie vs. Praxis
- Part B: Neue coole bayesian statistics (DIE ZUKUNFT)

Was gibt es hier *nicht*?

- Umfangreiches Walk-Through Tutorial
- Deskriptive Statistik
- Stochastik Nachhilfe
- Wie benchmarke ich meinen neuen coolen Algorithmus
- Experimentelles Design, Kausale Inferenz, …

Wo bin ich hier gelandet?

- (Grobe) Motivation: Was muss ich schreiben damit die Reviewer mich nicht haten?
- Part A: Überblick über die klassische (frequentistische) inferenzielle Statistik
  - Hilfe zur Selbsthilfe
  - Theorie vs. Praxis
- Part B: Neue coole bayesian statistics (DIE ZUKUNFT)

Was gibt es hier *nicht*?

- Umfangreiches Walk-Through Tutorial
- Deskriptive Statistik
- Stochastik Nachhilfe
- Wie benchmarke ich meinen neuen coolen Algorithmus
- Experimentelles Design, Kausale Inferenz, …

obligatorisch für KIDS!

Wo bin ich hier gelandet?

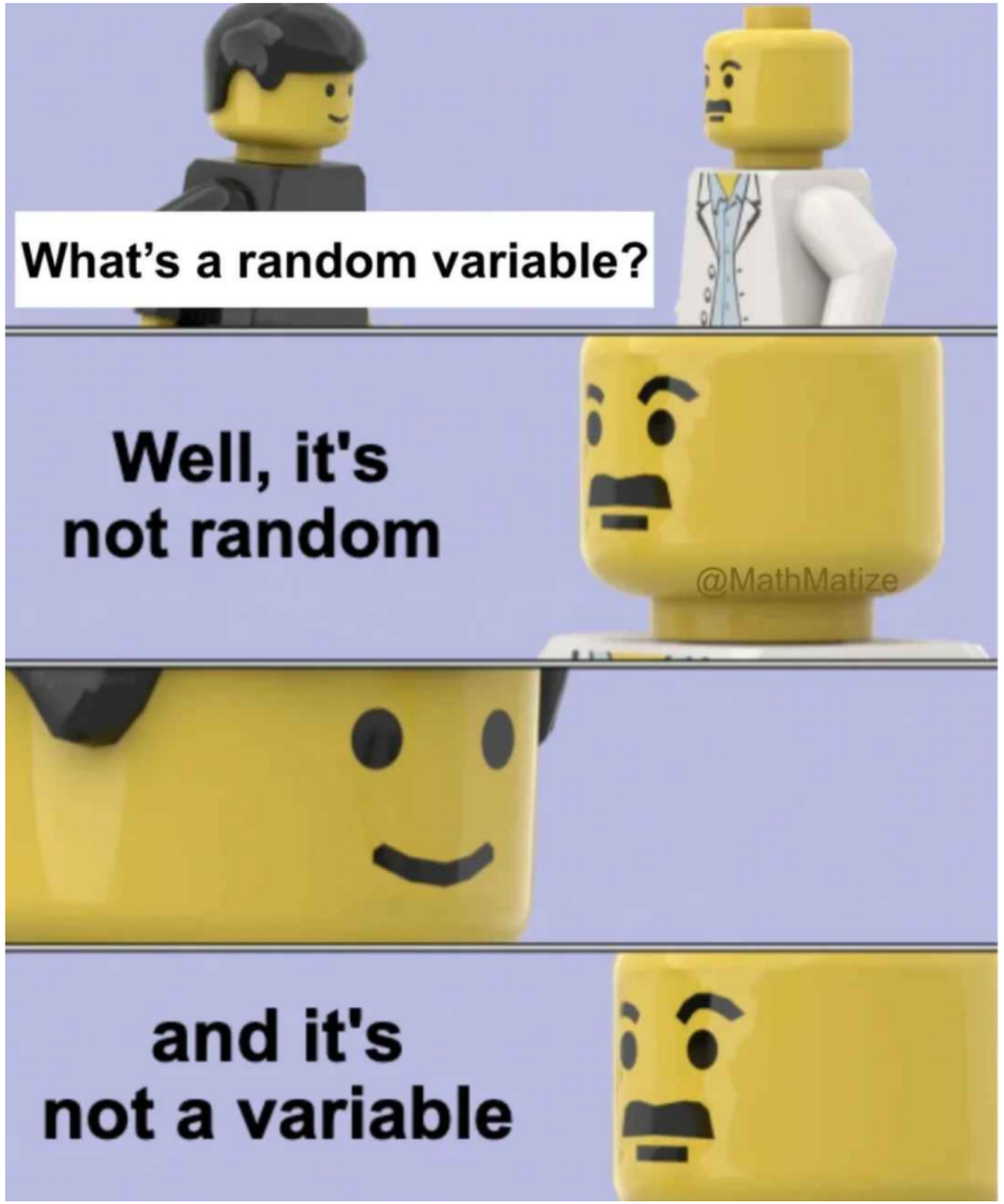
- (Grobe) Motivation: Was muss ich schreiben damit die Reviewer mich nicht haten?
- Part A: Überblick über die klassische (frequentistische) inferenzielle Statistik
  - Hilfe zur Selbsthilfe
  - Theorie vs. Praxis
- Part B: Neue coole bayesian statistics (DIE ZUKUNFT)

Was gibt es hier *nicht*?

- Umfangreiches Walk-Through Tutorial
- Deskriptive Statistik
- Stochastik Nachhilfe
- Wie benchmarke ich meinen neuen coolen Algorithmus
- Experimentelles Design, Kausale Inferenz, …

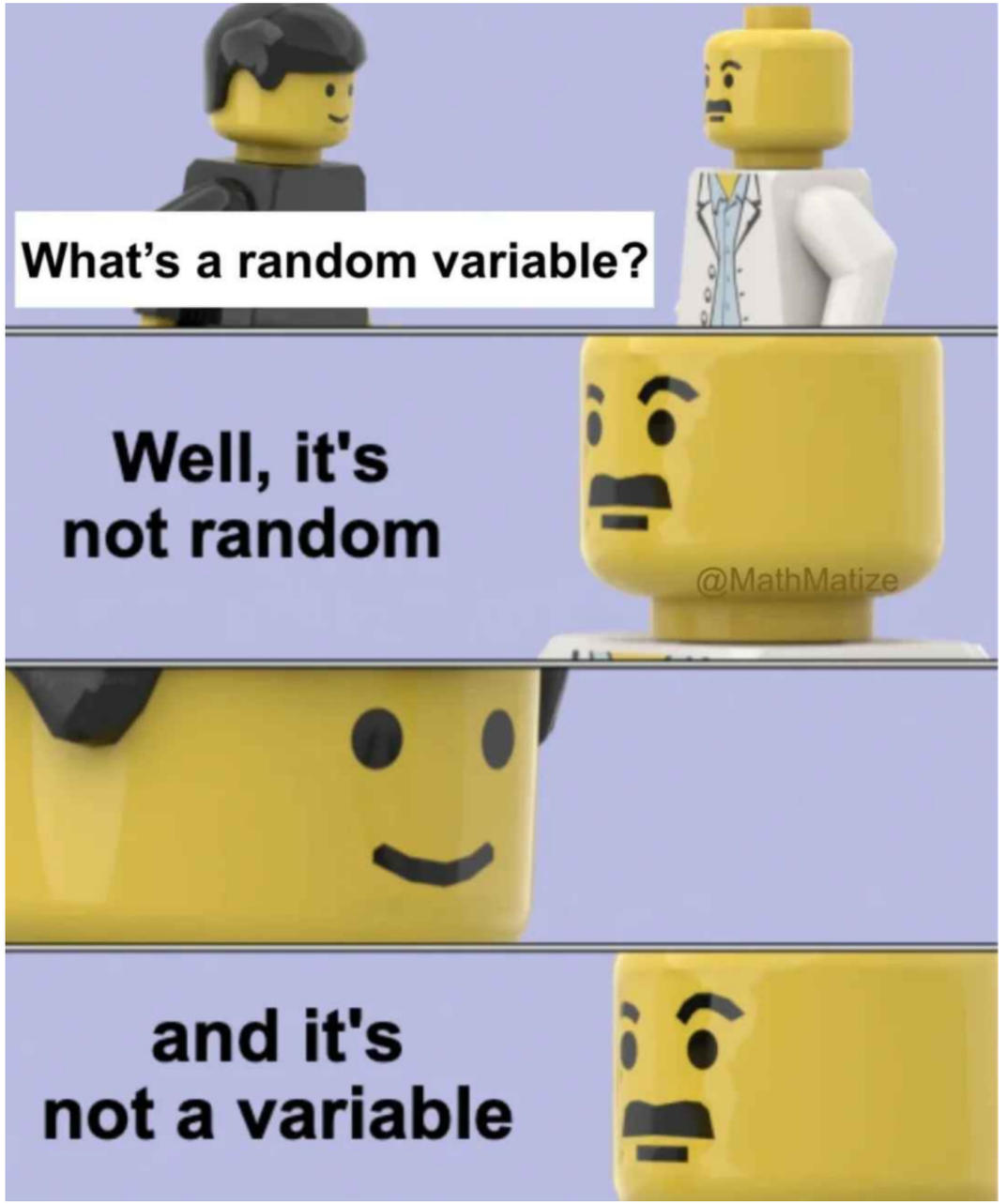
Recap: Zufallsvariablen

Recap: Zufallsvariablen



Recap: Zufallsvariablen

„high school math“



Random variable

Article Talk

From Wikipedia, the free encyclopedia

A **random variable** (also called **random quantity**, **aleatory variable**, or **stochastic variable**) is a mathematical formalization of a quantity or object which depends on random events. The term 'random variable' in its mathematical definition refers to neither randomness nor variability[1]

Random variable

Article Talk

From Wikipedia, the free encyclopedia

A **random variable** (also called **random quantity**, **aleatory variable**, or **stochastic variable**) is a mathematical formalization of a quantity or object which depends on random events. The term 'random variable' in its mathematical definition refers to neither randomness nor variability[1]

[...]

Informally, randomness typically represents some fundamental element of chance, such as in the roll of a die; it may also represent uncertainty, such as measurement error.[2] However, the interpretation of probability is philosophically complicated, and even in specific cases is not always straightforward. The purely mathematical analysis of random variables is independent of such interpretational difficulties, and can be based upon a rigorous axiomatic setup.

Random variable

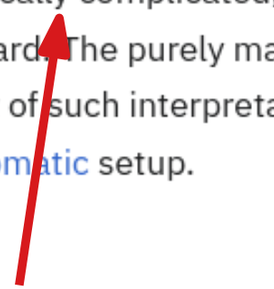
Article Talk

From Wikipedia, the free encyclopedia

A **random variable** (also called **random quantity**, **aleatory variable**, or **stochastic variable**) is a mathematical formalization of a quantity or object which depends on random events. The term 'random variable' in its mathematical definition refers to neither randomness nor variability[1]

[...]

Informally, randomness typically represents some fundamental element of chance, such as in the roll of a die; it may also represent uncertainty, such as measurement error.[2] However, the interpretation of probability is philosophically complicated, and even in specific cases is not always straightforward. The purely mathematical analysis of random variables is independent of such interpretational difficulties, and can be based upon a rigorous axiomatic setup.



foreshadowing

Random variable

Informell: diskrete Zufallsvariable X repräsentiert ein „Messprozess“, welcher zufälligen Phänomenen einen Wert aus $\mathbb{N}$ zuordnet.

Random variable

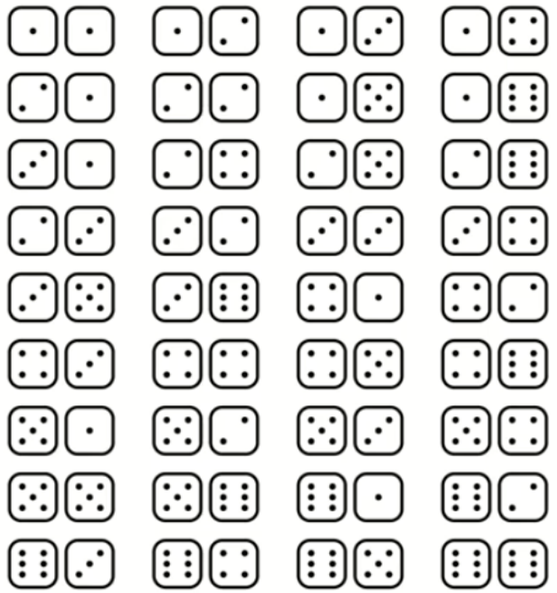
Informell: diskrete Zufallsvariable X repräsentiert ein „Messprozess“, welcher zufälligen Phänomenen einen Wert aus $\mathbb{N}$ zuordnet.

Für $a \in \mathbb{N}$ schreiben wir mit $P(X = a)$ die WK, dass X (in einem experimentellen Messprozess) den Wert a annimmt.

Random variable

Informell: diskrete Zufallsvariable X repräsentiert ein „Messprozess“, welcher zufälligen Phänomenen einen Wert aus $\mathbb{N}$ zuordnet.

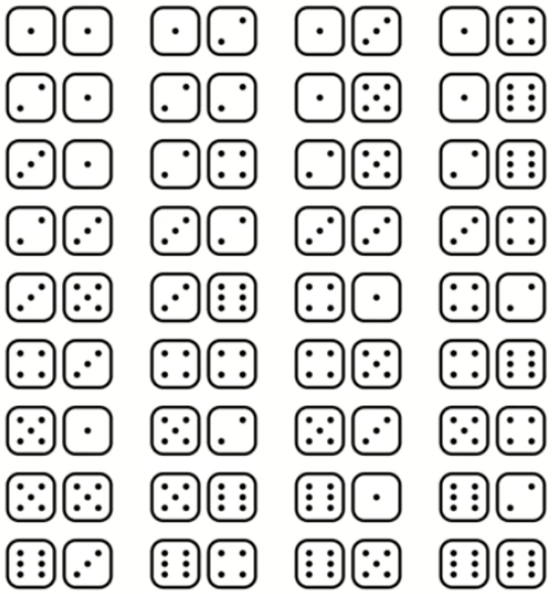
Für $a \in \mathbb{N}$ schreiben wir mit $P(X = a)$ die WK, dass X (in einem experimentellen Messprozess) den Wert a annimmt.

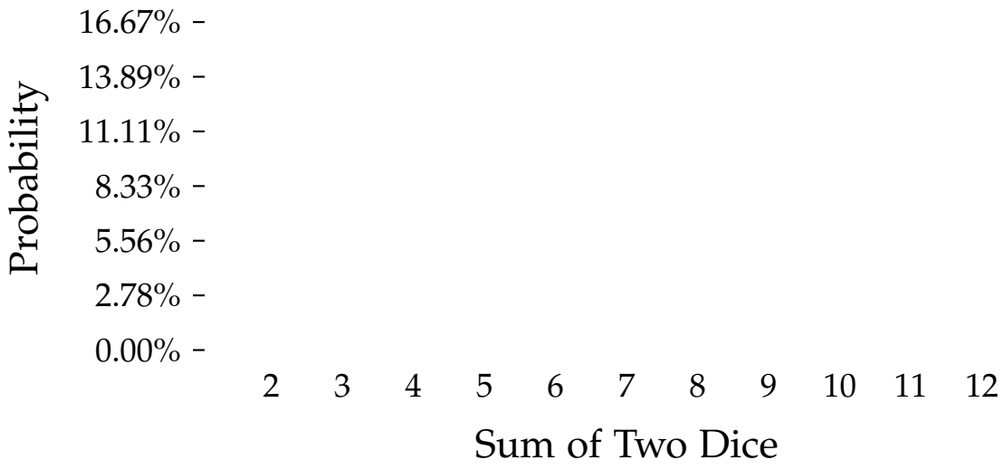


Random variable

Informell: diskrete Zufallsvariable X repräsentiert ein „Messprozess“, welcher zufälligen Phänomenen einen Wert aus $\mathbb{N}$ zuordnet.

Für $a \in \mathbb{N}$ schreiben wir mit $P(X = a)$ die WK, dass X (in einem experimentellen Messprozess) den Wert a annimmt.

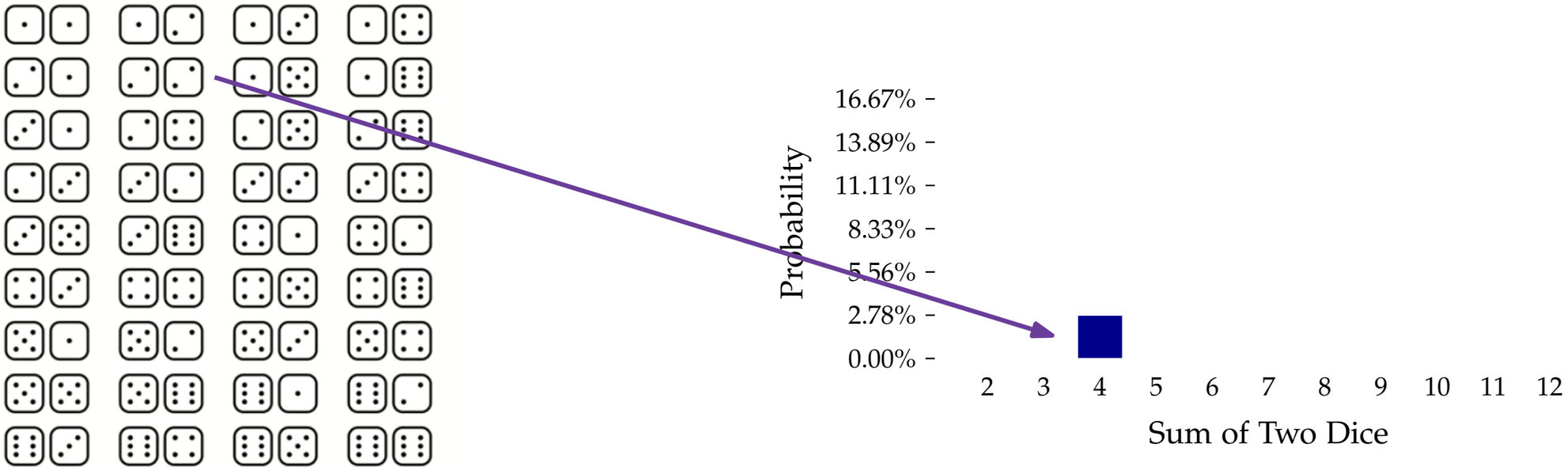




Random variable

Informell: diskrete Zufallsvariable X repräsentiert ein „Messprozess“, welcher zufälligen Phänomenen einen Wert aus $\mathbb{N}$ zuordnet.

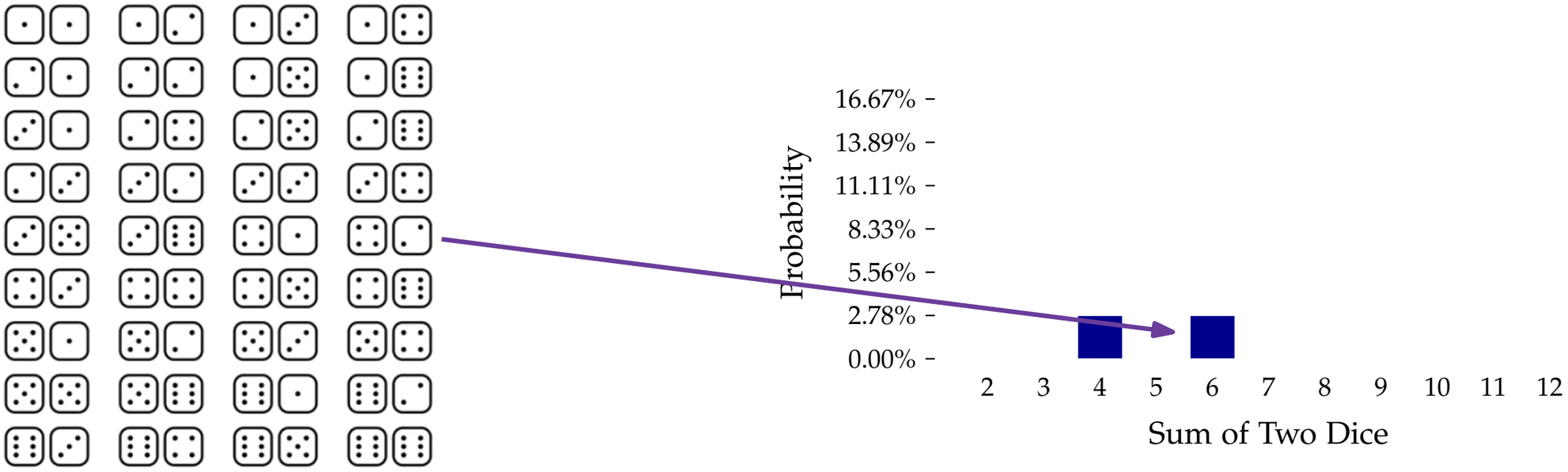
Für $a \in \mathbb{N}$ schreiben wir mit $P(X = a)$ die WK, dass X (in einem experimentellen Messprozess) den Wert a annimmt.



Random variable

Informell: diskrete Zufallsvariable X repräsentiert ein „Messprozess“, welcher zufälligen Phänomenen einen Wert aus $\mathbb{N}$ zuordnet.

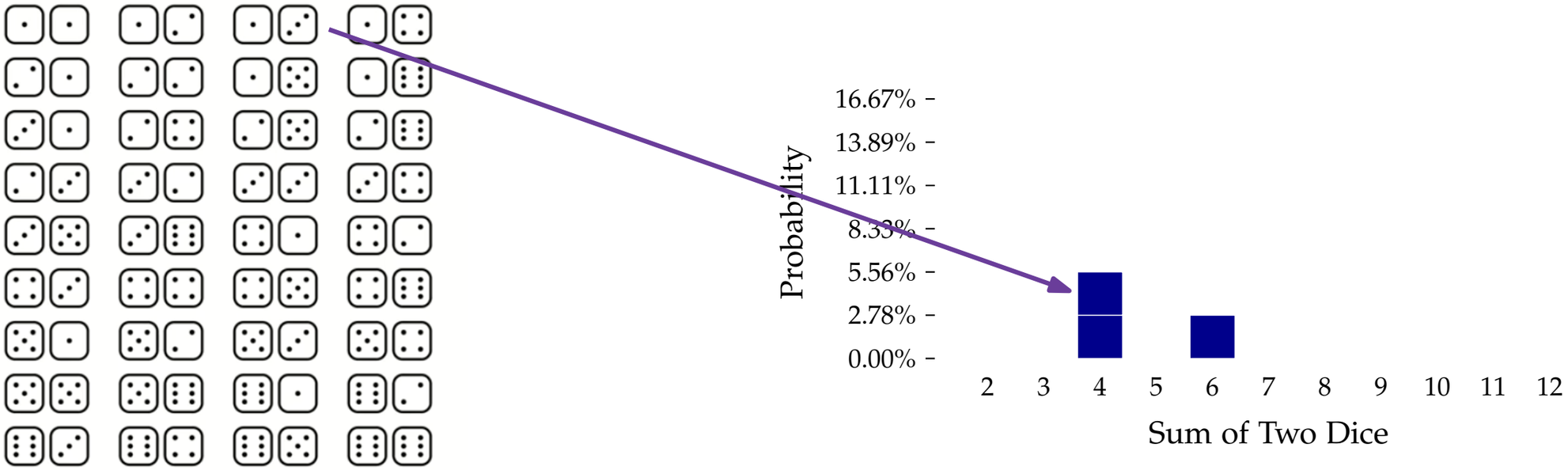
Für $a \in \mathbb{N}$ schreiben wir mit $P(X = a)$ die WK, dass X (in einem experimentellen Messprozess) den Wert a annimmt.



Random variable

Informell: diskrete Zufallsvariable X repräsentiert ein „Messprozess“, welcher zufälligen Phänomenen einen Wert aus $\mathbb{N}$ zuordnet.

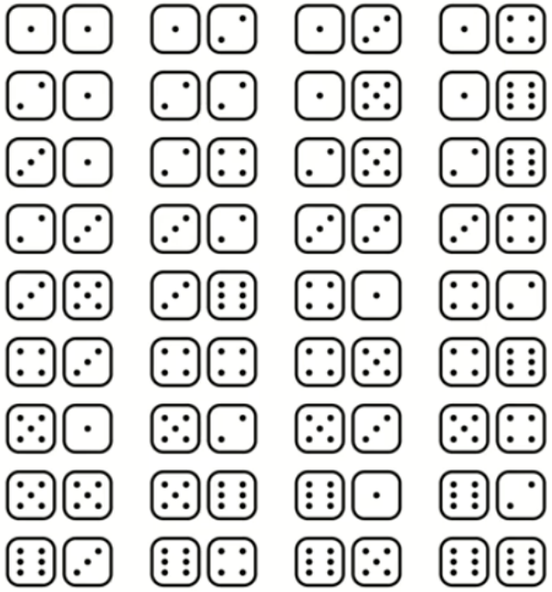
Für $a \in \mathbb{N}$ schreiben wir mit $P(X = a)$ die WK, dass X (in einem experimentellen Messprozess) den Wert a annimmt.

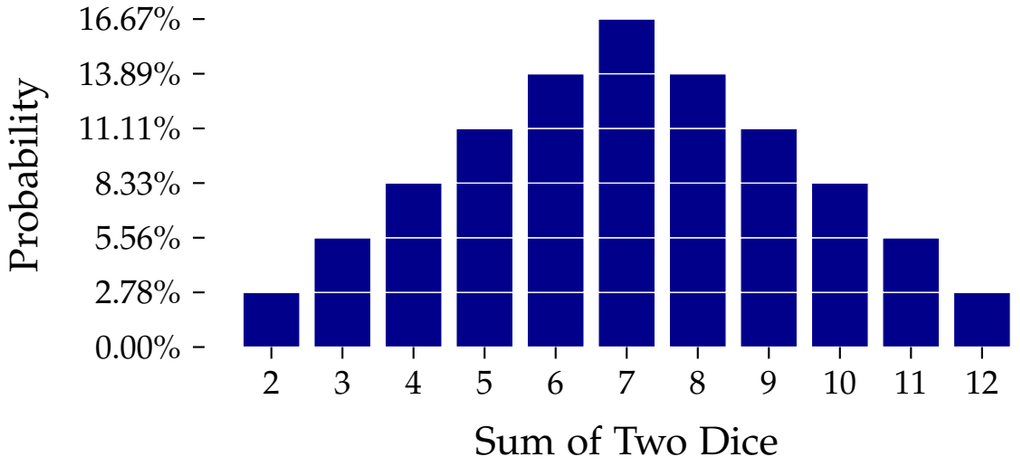


Random variable

Informell: diskrete Zufallsvariable X repräsentiert ein „Messprozess“, welcher zufälligen Phänomenen einen Wert aus $\mathbb{N}$ zuordnet.

Für $a \in \mathbb{N}$ schreiben wir mit $P(X = a)$ die WK, dass X (in einem experimentellen Messprozess) den Wert a annimmt.

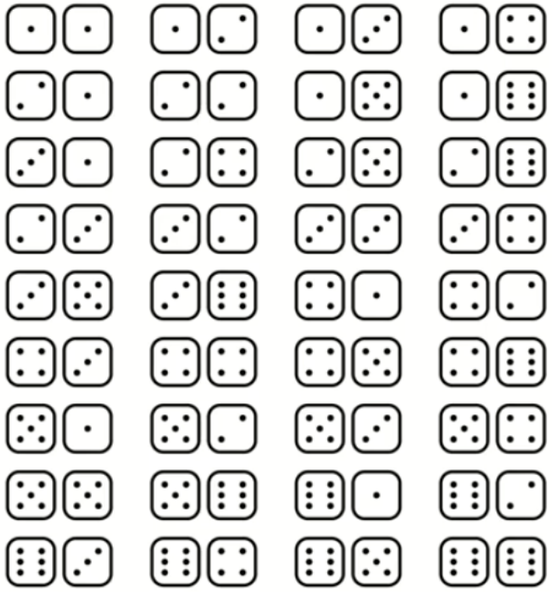


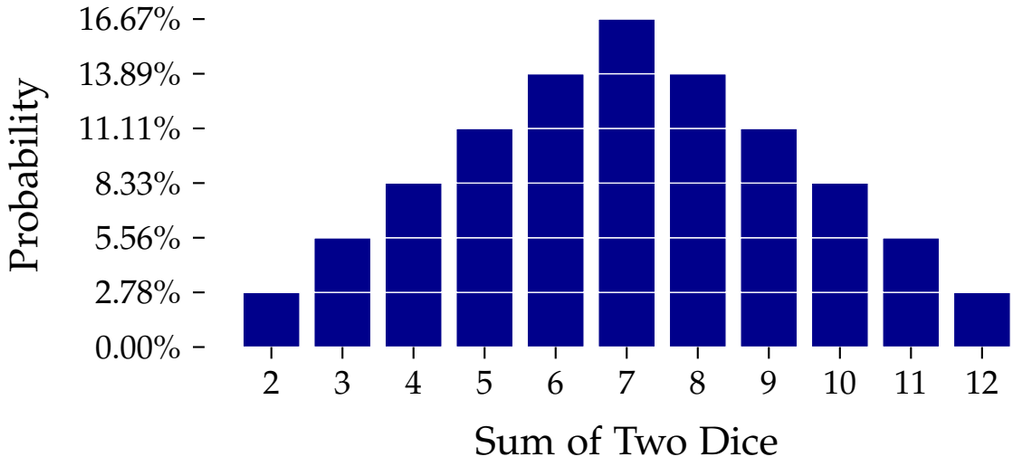


Random variable

Informell: diskrete Zufallsvariable X repräsentiert ein „Messprozess“, welcher zufälligen Phänomenen einen Wert aus $\mathbb{N}$ zuordnet.

Für $a \in \mathbb{N}$ schreiben wir mit $P(X = a)$ die WK, dass X (in einem experimentellen Messprozess) den Wert a annimmt.





Allgemeiner: Wir schreiben $X \sim f$ wenn $P(X = a) = f(a) \quad = \frac{6-|7-a|}{36}$

Random variable

kontinuierliche

Informell: ~~diskrete~~ Zufallsvariable X repräsentiert ein „Messprozess“, welcher zufälligen Phänomenen einen Wert aus ~~$\mathbb{N}$~~ $\mathbb{R}$ zuordnet.

$P(a \leq X \leq b)$

Für $a \in \mathbb{N}$ schreiben wir mit $P(X = a)$ die WK, dass X (in einem experimentellen Messprozess) den Wert a ~~annimmt~~ zwischen a und b annimmt.

Allgemeiner: Wir schreiben $X \sim f$ wenn ~~$P(X = a) = f(a)$~~ $P(a \leq X \leq b) = \int_a^b f(x)\,\mathrm{d}x$

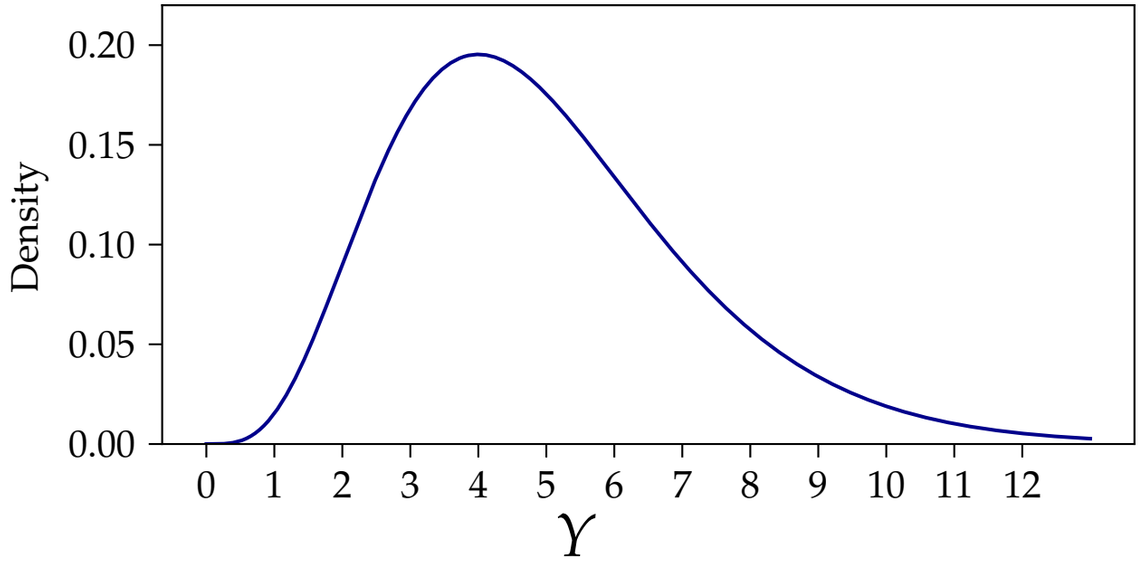
Random variable

kontinuierliche

Informell: ~~diskrete~~ Zufallsvariable X repräsentiert ein „Messprozess“, welcher zufälligen Phänomenen einen Wert aus ~~$\mathbb{N}$~~ $\mathbb{R}$ zuordnet.

$P(a \leq X \leq b)$

Für $a \in \mathbb{N}$ schreiben wir mit $P(X = a)$ die WK, dass X (in einem experimentellen Messprozess) den Wert a ~~annimmt~~ zwischen a und b annimmt.



$Y \sim \text{Erlang}(5, 1)$

(Wartezeit bis 5 Requests rein kommen, bei mittl. 1 Request pro Sekunde)

Allgemeiner: Wir schreiben $X \sim f$ wenn ~~$P(X = a) = f(a)$~~ $P(a \leq X \leq b) = \int_a^b f(x)\,\mathrm{d}x$

Random variable

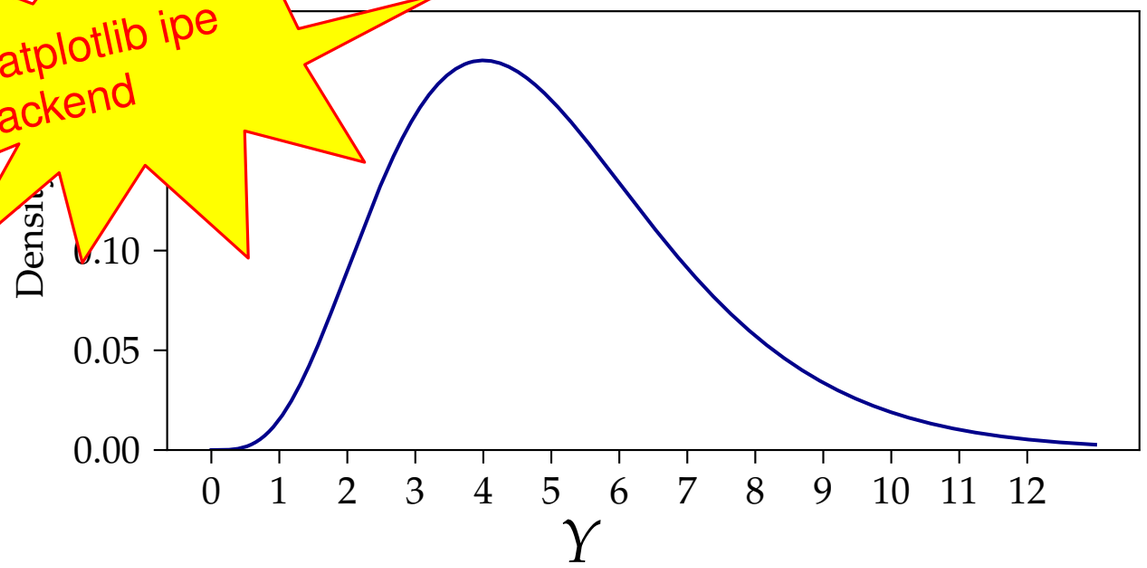
kontinuierliche

Informell: ~~diskrete~~ Zufallsvariable X repräsentiert ein „Messprozess“, welcher zufälligen Phänomenen einen Wert aus ~~$\mathbb{N}$~~ $\mathbb{R}$ zuordnet.

$P(a \leq X \leq b)$

Für $a \in \mathbb{N}$ schreiben wir mit ~~$P(X = a)$~~ die WK, dass X (in einem experimentellen Messprozess) den Wert a ~~annimmt~~ zwischen a und b annimmt

matplotlib ipe backend



$Y \sim \text{Erlang}(5, 1)$

(Wartezeit bis 5 Requests rein kommen, bei mittl. 1 Request pro Sekunde)

Allgemeiner: Wir schreiben $X \sim f$ wenn ~~$P(X = a) = f(a)$~~ $P(a \leq X \leq b) = \int_a^b f(x)\,\mathrm{d}x$

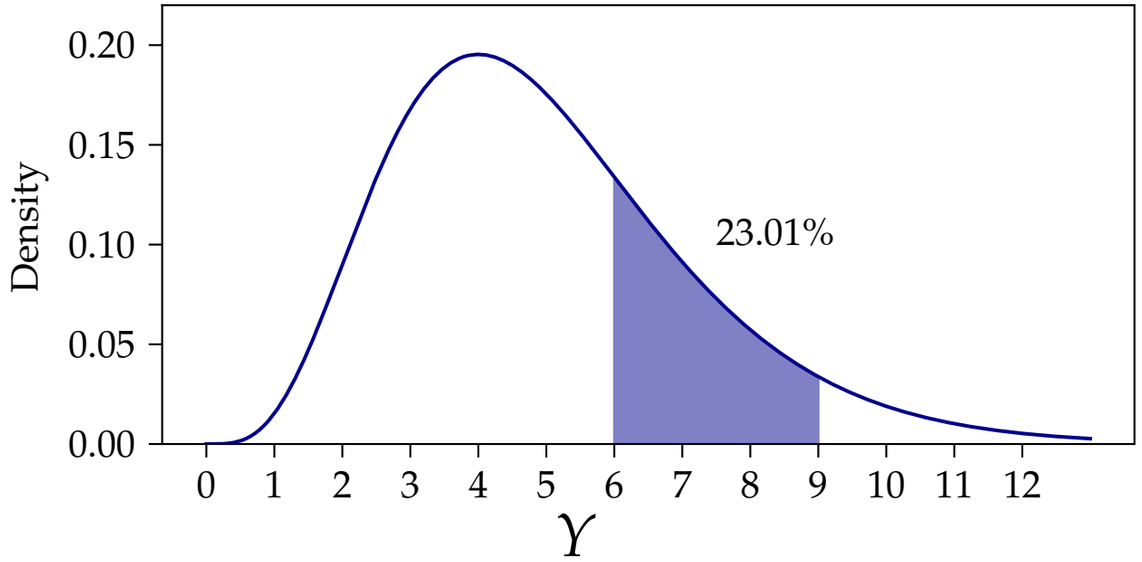
Random variable

kontinuierliche

Informell: ~~diskrete~~ Zufallsvariable X repräsentiert ein „Messprozess“, welcher zufälligen Phänomenen einen Wert aus ~~$\mathbb{N}$~~ $\mathbb{R}$ zuordnet.

$P(a \leq X \leq b)$

Für $a \in \mathbb{N}$ schreiben wir mit $P(X = a)$ die WK, dass X (in einem experimentellen Messprozess) den Wert a ~~annimmt~~. zwischen a und b annimmt

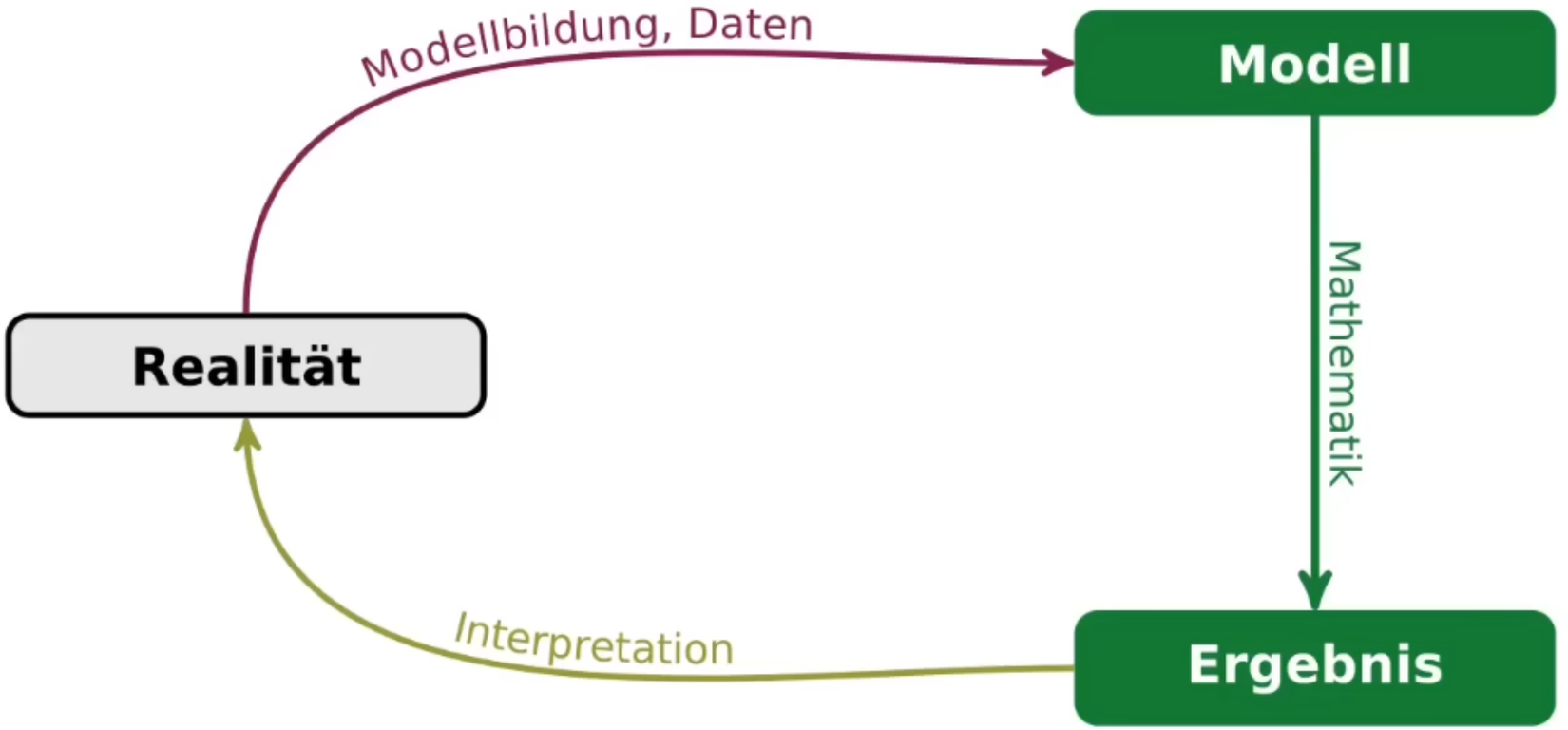


$Y \sim \text{Erlang}(5, 1)$

(Wartezeit bis 5 Requests rein kommen, bei mittl. 1 Request pro Sekunde)

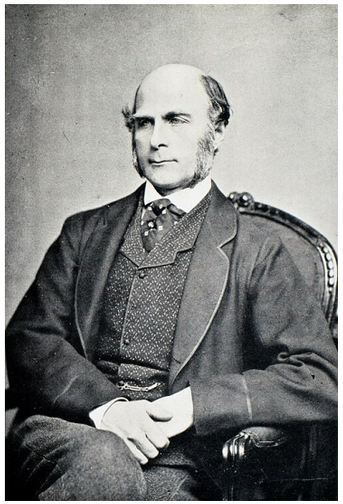
Allgemeiner: Wir schreiben $X \sim f$ wenn ~~$P(X = a) = f(a)$~~ $P(a \leq X \leq b) = \int_a^b f(x)\,\mathrm{d}x$

„ALL MODELS ARE WRONG, BUT SOME ARE USEFUL.“



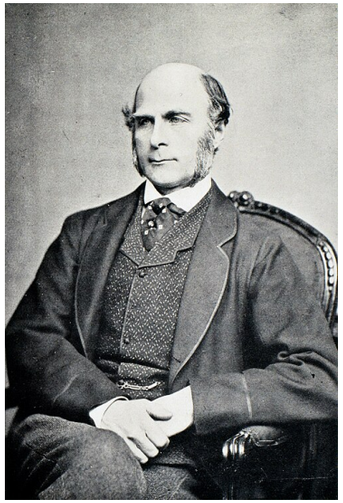
If people do not believe that mathematics is simple,
it is only because they do not realize how complicated life is.
[John von Neumann]

Francis Galton (1822-1911)



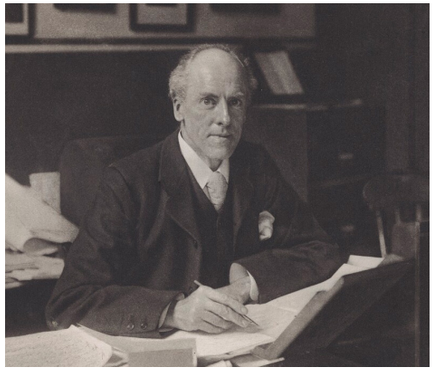
- ‚self-funded researcher‘
- Cousin von Darwin
- Erste Anwendung von stat. Methoden in Sozialforschung
- Eugeniker

Francis Galton (1822-1911)



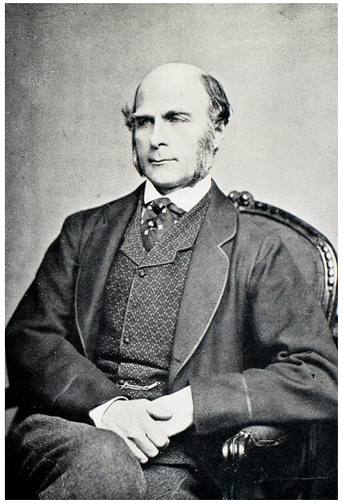
- ‚self-funded researcher‘
- Cousin von Darwin
- Erste Anwendung von stat. Methoden in Sozialforschung
- Eugeniker

Karl Pearson (1857-1936)



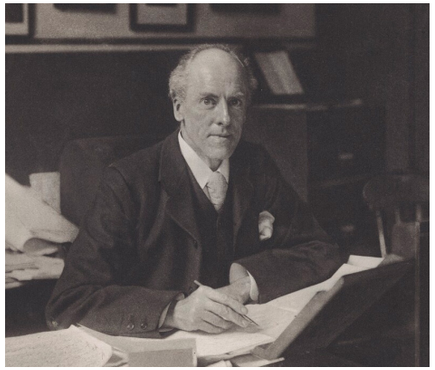
- Biologe
- selbsterklärter Sozialist
- Etablierte stat. Tests (insb. χ^2)
- Eugeniker

Francis Galton (1822-1911)



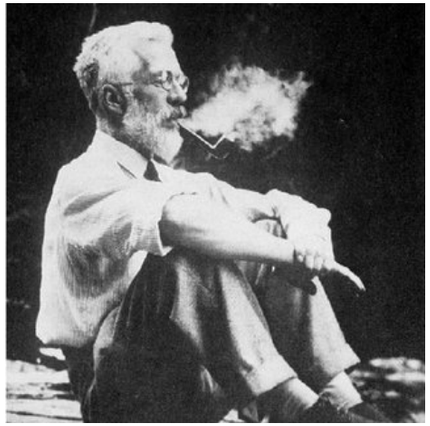
- ‚self-funded researcher‘
- Cousin von Darwin
- Erste Anwendung von stat. Methoden in Sozialforschung
- Eugeniker

Karl Pearson (1857-1936)



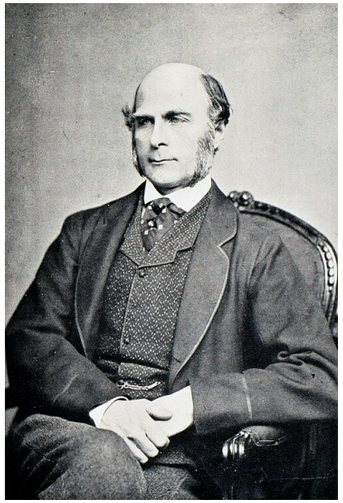
- Biologe
- selbsterklärter Sozialist
- Etablierte stat. Tests (insb. χ^2)
- Eugeniker

Ronald Fisher (1890-1962)



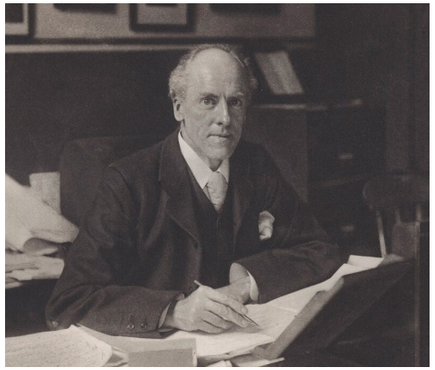
- Biologe / Agrarwissenschaftler
- Systematisiert experimentelles Design, ANOVA
- Eugeniker
- Lobbyist für die Tabakindustrie

Francis Galton (1822-1911)



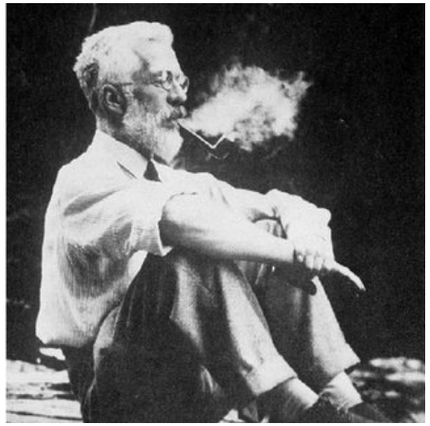
- ‚self-funded researcher‘
- Cousin von Darwin
- Erste Anwendung von stat. Methoden in Sozialforschung
- Eugeniker

Karl Pearson (1857-1936)



- Biologe
- selbsterklärter Sozialist
- Etablierte stat. Tests (insb. χ^2)
- Eugeniker

Ronald Fisher (1890-1962)



- Biologe / Agrarwissenschaftler
- Systematisiert experimentelles Design, ANOVA
- Eugeniker
- Lobbyist für die Tabakindustrie

Wiliam Gosset (1876-1937)

- Ging in die Industrie
- Bierbrauer bei Guinness
- Student’s t-Test
- „loved sailing and fishing“

Wahrscheinlichkeit frequentistisch

Wahrscheinlichkeit frequentistisch

Empirische Beobachtung: Beim wiederholten Münzwurf stabilisiert sich die relative Häufigkeit von „Kopf“

Wahrscheinlichkeit frequentistisch

Empirische Beobachtung: Beim wiederholten Münzwurf stabilisiert sich die relative Häufigkeit von „Kopf“

$$\lim_{n \to \infty} \frac{n_S}{n} = c$$

Wahrscheinlichkeit frequentistisch

Empirische Beobachtung: Beim wiederholten Münzwurf stabilisiert sich die relative Häufigkeit von „Kopf“

Wahrscheinlichkeit $:=$ $\lim_{n\to\infty} \frac{n_S}{n} = c$

Wahrscheinlichkeit frequentistisch

Empirische Beobachtung: Beim wiederholten Münzwurf stabilisiert sich die relative Häufigkeit von „Kopf“

Wahrscheinlichkeit := $\lim_{n\to\infty} \frac{n_S}{n} = c$

neat: WK ist empirisch messbar

Wahrscheinlichkeit frequentistisch

Empirische Beobachtung: Beim wiederholten Münzwurf stabilisiert sich die relative Häufigkeit von „Kopf“

Wahrscheinlichkeit $:=$ $\lim_{n\to\infty} \frac{n_S}{n} = c$

neat: WK ist empirisch messbar

Definition: Wahrscheinlichkeit ist die (theoretische)
relative Häufigkeit eines Ereignisses
unter langfristiger Wiederholung
von Experimenten unter gleichen Bedingungen.

Wahrscheinlichkeit frequentistisch

Empirische Beobachtung: Beim wiederholten Münzwurf stabilisiert sich die relative Häufigkeit von „Kopf“

Wahrscheinlichkeit $:=$ $\lim_{n\to\infty} \frac{n_S}{n} = c$

neat: WK ist empirisch messbar

Definition: Wahrscheinlichkeit ist die (theoretische)
relative Häufigkeit eines Ereignisses
unter langfristiger Wiederholung
von Experimenten unter gleichen Bedingungen.

Frequentistisches Programm: Die Mechanismen der Umwelt sind unbekannt aber konstant – Experimentelle Daten sind zufällig.
Unsere Inferenzprozeduren sollen derart sein, dass sie „wahrscheinlich“ korrekte Ergebnisse liefern.

Wahrscheinlichkeit frequentistisch

Empirische Beobachtung: Beim wiederholten Münzwurf stabilisiert sich die relative Häufigkeit von „Kopf“

Wahrscheinlichkeit $:=$ $\lim_{n\to\infty} \frac{n_S}{n} = c$

neat: WK ist empirisch messbar

Definition: Wahrscheinlichkeit ist die (theoretische)
relative Häufigkeit eines Ereignisses
unter langfristiger Wiederholung
von Experimenten unter gleichen Bedingungen.

[weird multiverse shit???]

Frequentistisches Programm: Die Mechanismen der Umwelt sind unbekannt aber konstant – Experimentelle Daten sind zufällig.
Unsere Inferenzprozeduren sollen derart sein, dass sie „wahrscheinlich“ korrekte Ergebnisse liefern.

Wahrscheinlichkeit frequentistisch

Empirische Beobachtung: Beim wiederholten Münzwurf stabilisiert sich die relative Häufigkeit von „Kopf“

Wahrscheinlichkeit $:=$ $\lim_{n\to\infty} \frac{n_S}{n} = c$

neat: WK ist empirisch messbar

Definition: Wahrscheinlichkeit ist die (theoretische) **relative Häufigkeit** eines Ereignisses unter langfristiger Wiederholung von Experimenten unter gleichen Bedingungen.

[singuläre Events haben können keine WK zugeordnet werden]

[weird multiverse shit???]

Frequentistisches Programm: Die Mechanismen der Umwelt sind unbekannt aber konstant – Experimentelle Daten sind zufällig. Unsere Inferenzprozeduren sollen derart sein, dass sie „wahrscheinlich“ korrekte Ergebnisse liefern.

Wahrscheinlichkeit frequentistisch

Em ...wiederholter
d...

„Morgen regnet es wahrscheinlich“

„Es existierte wahrscheinlich Leben auf dem Mars“

Wahrscheinlichkeit := $\lim_{n\to\infty} \frac{n_S}{n} = c$

neat: WK ist empirisch messbar

Definition: Wahrscheinlichkeit ist die (theoretische) **relative Häufigkeit** eines Ereignisses unter langfristiger Wiederholung von Experimenten unter gleichen Bedingungen.

[singuläre Events haben können keine WK zugeordnet werden]

[weird multiverse shit???]

Frequentistisches Programm: Die Mechanismen der Umwelt sind unbekannt aber konstant – Experimentelle Daten sind zufällig. Unsere Inferenzprozeduren sollen derart sein, dass sie „wahrscheinlich“ korrekte Ergebnisse liefern.

Wahrscheinlichkeit frequentistisch

Em ...wiederholter
d...

„Morgen regnet es wahrscheinlich“

„Es existierte wahrscheinlich Leben auf dem Mars“

Wahrscheinlichkeit := $\lim_{n\to\infty} \frac{n_S}{n} = c$

neat: WK ist empirisch messbar

Definition: Wahrscheinlichkeit ist die (theoretische) **relative Häufigkeit** eines Ereignisses unter langfristiger Wiederholung von Experimenten unter gleichen Bedingungen.

[singuläre Events haben können keine WK zugeordnet werden]

[weird multiverse shit???]

Frequentistisches Programm: Die Mechanismen der Umwelt sind unbekannt aber konstant – Experimentelle Daten sind zufällig. Unsere Inferenzprozeduren sollen derart sein, dass sie „wahrscheinlich“ korrekte Ergebnisse liefern.

Null Hypothesis Significance Testing

Idee: Behauptung (Hypothese) über die Eigenschaft von Populationen, welche durch empirische Untersuchungen auf Stichprobe überprüft werden muss.

Null Hypothesis Significance Testing

Idee: Behauptung (Hypothese) über die Eigenschaft von Populationen, welche durch empirische Untersuchungen auf Stichprobe überprüft werden muss.

Beispiel: *Neuer Algorithmus B ist schneller als Algorithmus A.*

Null Hypothesis Significance Testing

Idee: Behauptung (Hypothese) über die Eigenschaft von Populationen, welche durch empirische Untersuchungen auf Stichprobe überprüft werden muss.

Beispiel: *Neuer Algorithmus B ist schneller als Algorithmus A.*

vs. *Neuer Algorithmus B ist NICHT schneller als Algorithmus A.*

Null Hypothesis Significance Testing

Idee: Behauptung (Hypothese) über die Eigenschaft von Populationen, welche durch empirische Untersuchungen auf Stichprobe überprüft werden muss.

Beispiel: *Neuer Algorithmus B ist schneller als Algorithmus A.*
Gegenhypothese H_1 („**exciting**“)

vs. *Neuer Algorithmus B ist NICHT schneller als Algorithmus A.*
Nullhypothese H_0 („**langweilig, status quo**“)

Null Hypothesis Significance Testing

Idee: Behauptung (Hypothese) über die Eigenschaft von Populationen, welche durch empirische Untersuchungen auf Stichprobe überprüft werden muss.

Beispiel: *Neuer Algorithmus B ist schneller als Algorithmus A.*
Gegenhypothese H_1 („**exciting**“)

vs. *Neuer Algorithmus B ist NICHT schneller als Algorithmus A.*
Nullhypothese H_0 („**langweilig, status quo**“)

Ziel: Test, um aus Datenlage möglichst fehlerfrei
H_0 zu verwerfen oder H_0 nicht zu verwerfen.

zugunsten H_1 zugunsten H_0

Null Hypothesis Significance Testing

Idee: Behauptung (Hypothese) über die Eigenschaft von Populationen, welche durch empirische Untersuchungen auf Stichprobe überprüft werden muss.

Beispiel: *Neuer Algorithmus B ist schneller als Algorithmus A.*
Gegenhypothese H_1 („**exciting**“)

vs. *Neuer Algorithmus B ist NICHT schneller als Algorithmus A.*
Nullhypothese H_0 („**langweilig, status quo**“)

Ziel: Test, um aus Datenlage möglichst fehlerfrei
H_0 zu verwerfen oder H_0 nicht zu verwerfen.

zugunsten H_1 zugunsten H_0

Testergebnis \ Realität	H_0	H_1
H_0 verwerfen		
H_0 nicht verwerfen		

Null Hypothesis Significance Testing

Idee: Behauptung (Hypothese) über die Eigenschaft von Populationen, welche durch empirische Untersuchungen auf Stichprobe überprüft werden muss.

Beispiel: *Neuer Algorithmus B ist schneller als Algorithmus A.*
Gegenhypothese H_1 („**exciting**“)

vs. *Neuer Algorithmus B ist NICHT schneller als Algorithmus A.*
Nullhypothese H_0 („**langweilig, status quo**“)

Ziel: Test, um aus Datenlage möglichst fehlerfrei
H_0 zu verwerfen oder H_0 nicht zu verwerfen.

zugunsten H_1 zugunsten H_0

Testergebnis \ Realität	H_0	H_1
H_0 verwerfen	FP	TP
H_0 nicht verwerfen	TN	FN

Null Hypothesis Significance Testing

Idee: Behauptung (Hypothese) über die Eigenschaft von Populationen, welche durch empirische Untersuchungen auf Stichprobe überprüft werden muss.

Beispiel: *Neuer Algorithmus B ist schneller als Algorithmus A.*
Gegenhypothese H_1 („**exciting**“)

vs. *Neuer Algorithmus B ist NICHT schneller als Algorithmus A.*
Nullhypothese H_0 („**langweilig, status quo**“)

Ziel: Test, um aus Datenlage möglichst fehlerfrei
H_0 zu verwerfen oder H_0 nicht zu verwerfen.

zugunsten H_1 zugunsten H_0

Testergebnis \ Realität	H_0	H_1
H_0 verwerfen	FP	TP ✅
H_0 nicht verwerfen	TN ✅	FN

Null Hypothesis Significance Testing

Idee: Behauptung (Hypothese) über die Eigenschaft von Populationen, welche durch empirische Untersuchungen auf Stichprobe überprüft werden muss.

Beispiel: *Neuer Algorithmus B ist schneller als Algorithmus A.*
Gegenhypothese H_1 („**exciting**“)

vs. *Neuer Algorithmus B ist NICHT schneller als Algorithmus A.*
Nullhypothese H_0 („**langweilig, status quo**“)

Ziel: Test, um aus Datenlage möglichst fehlerfrei
H_0 zu verwerfen oder H_0 nicht zu verwerfen.

zugunsten H_1 zugunsten H_0

Testergebnis \ Realität	H_0	H_1
H_0 verwerfen	FP ☹️	TP ✅
H_0 nicht verwerfen	TN ✅	FN

Null Hypothesis Significance Testing

Idee: Behauptung (Hypothese) über die Eigenschaft von Populationen, welche durch empirische Untersuchungen auf Stichprobe überprüft werden muss.

Beispiel: *Neuer Algorithmus B ist schneller als Algorithmus A.*
Gegenhypothese H_1 („**exciting**“)

vs. *Neuer Algorithmus B ist NICHT schneller als Algorithmus A.*
Nullhypothese H_0 („**langweilig, status quo**“)

Ziel: Test, um aus Datenlage möglichst fehlerfrei
H_0 zu verwerfen oder H_0 nicht zu verwerfen.

zugunsten H_1 — zugunsten H_0

Testergebnis \ Realität	H_0	H_1
H_0 verwerfen	FP ☹️	TP ✅
H_0 nicht verwerfen	TN ✅	FN

Typ-1-Fehler

Null Hypothesis Significance Testing

Idee: Behauptung (Hypothese) über die Eigenschaft von Populationen, welche durch empirische Untersuchungen auf Stichprobe überprüft werden muss.

Beispiel: *Neuer Algorithmus B ist schneller als Algorithmus A.*
Gegenhypothese H_1 („**exciting**“)

vs. *Neuer Algorithmus B ist NICHT schneller als Algorithmus A.*
Nullhypothese H_0 („**langweilig, status quo**“)

Ziel: Test, um aus Datenlage möglichst fehlerfrei
H_0 zu verwerfen oder H_0 nicht zu verwerfen.

zugunsten H_1 — zugunsten H_0

Testergebnis \ Realität	H_0	H_1
H_0 verwerfen	FP ☹️	TP ✅
H_0 nicht verwerfen	TN ✅	FN ☹️

Typ-1-Fehler

Typ-2-Fehler

Null Hypothesis Significance Testing

Idee: Behauptung (Hypothese) über die Eigenschaft von Populationen, welche durch empirische Untersuchungen auf Stichprobe überprüft werden muss.

Beispiel: *Neuer Algorithmus B ist schneller als Algorithmus A.*
Gegenhypothese H_1 („**exciting**“)

vs. *Neuer Algorithmus B ist NICHT schneller als Algorithmus A.*
Nullhypothese H_0 („**langweilig, status quo**“)

Ziel: Test, um aus Datenlage möglichst fehlerfrei H_0 zu verwerfen oder H_0 nicht zu verwerfen.

zugunsten H_1 — zugunsten H_0

Testergebnis \ Realität	H_0	H_1
H_0 verwerfen	FP ☹️	TP ✅
H_0 nicht verwerfen	TN ✅	FN ☹️

Typ-1-Fehler

Typ-2-Fehler

Der Plan:

- H_0 aufstellen.
- Daten **X** sammeln
- Widerspruchsbeweis: Wir nehmen H_0 an
- Falls **X** unwahrscheinlich unter H_0 ist, dann kann H_0 nicht stimmen und wird verworfen

Null Hypothesis Significance Testing

Idee: Behauptung (Hypothese) über die Eigenschaft von Populationen, welche durch empirische Untersuchungen auf Stichprobe überprüft werden muss.

Beispiel: *Neuer Algorithmus B ist schneller als Algorithmus A.*
Gegenhypothese H_1 („**exciting**“)

vs. *Neuer Algorithmus B ist NICHT schneller als Algorithmus A.*
Nullhypothese H_0 („**langweilig, status quo**“)

Ziel: Test, um aus Datenlage möglichst fehlerfrei H_0 zu verwerfen oder H_0 nicht zu verwerfen.

zugunsten H_1 — zugunsten H_0

Testergebnis \ Realität	H_0	H_1
H_0 verworfen	FP ☹	TP ✅
H_0 nicht verwerfen	TN ✅	FN ☹

$P(\text{Typ-1-Fehler}) \stackrel{!}{=} 5\%$

Typ-2-Fehler

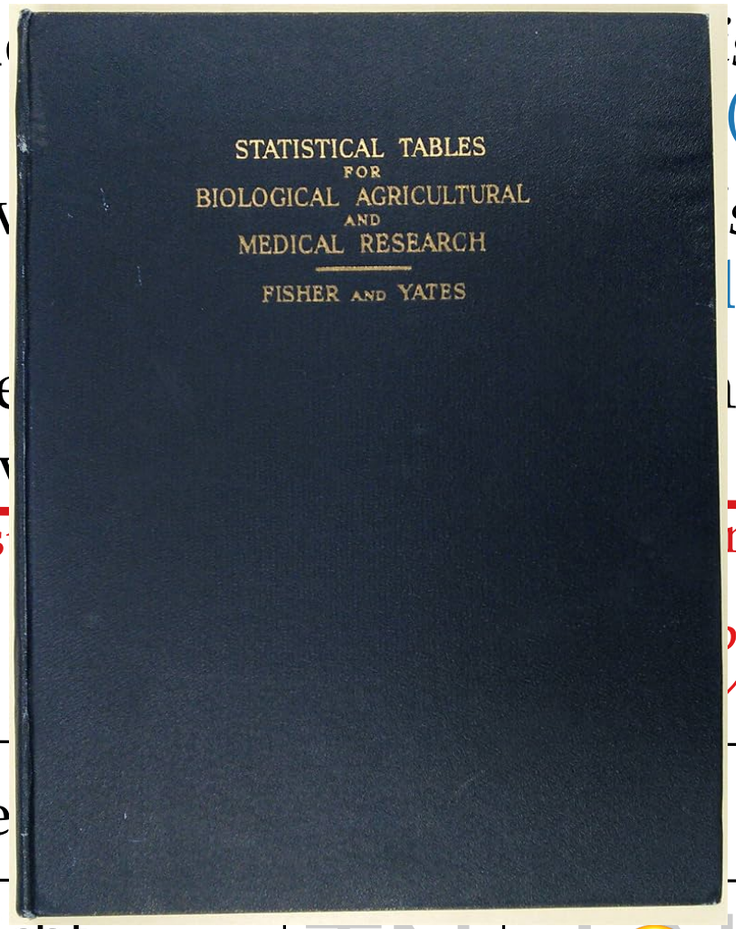
Der Plan:

- H_0 aufstellen.
- Daten **X** sammeln
- Widerspruchsbeweis: Wir nehmen H_0 an
- Falls **X** unwahrscheinlich unter H_0 ist, dann kann H_0 nicht stimmen und wird verworfen

Null Hypothesis Significance Testing

Idee: Behauptung (Hypothese) über die Eigenschaft von Populationen, welche durch empirische Untersuchungen auf Stichprobe überprüft werden muss.

Beispi



st schneller als Algorithmus A.
(„**exciting**“)

st NICHT schneller als Algorithmus A.
langweilig, status quo“)

Ziel: Te öglichst fehlerfrei
H_0 zu v zu verwerfen.
zugunst nsten H_0

?(Typ-1-Fehler) $\stackrel{!}{=} 5\%$

Testergebnis		
H_0 ve		
H_0 nicht verwerfen	✅ TN	☹️ FN

Typ-2-Fehler

Der Plan:

- H_0 aufstellen.
- Daten **X** sammeln
- Widerspruchsbeweis: Wir nehmen H_0 an
- Falls **X** unwahrscheinlich unter H_0 ist, dann kann H_0 nicht stimmen und wird verworfen

Null Hypothesis Significance Testing

Idee: Behauptung (Hypothese) über die Eigenschaft von Populationen, welche durch empirische Untersuchungen auf Stichprobe überprüft werden muss.

Beispi[redacted]*st schneller als Algorithmus A.*

TABLE III. DISTRIBUTION OF t

	Probability.												
n	·9	·8	·7	·6	·5	·4	·3	·2	·1	·05	·02	·01	·001
1	·158	·325	·510	·727	1·000	1·376	1·963	3·078	6·314	12·706	31·821	63·657	636·619
2	·142	·289	·445	·617	·816	1·061	1·386	1·886	2·920	4·303	6·965	9·925	31·598
3	·137	·277	·424	·584	·765	·978	1·250	1·638	2·353	3·182	4·541	5·841	12·924
4	·134	·271	·414	·569	·741	·941	1·190	1·533	2·132	2·776	3·747	4·604	8·610
5	·132	·267	·408	·559	·727	·920	1·156	1·476	2·015	2·571	3·365	4·032	6·869
6	·131	·265	·404	·553	·718	·906	1·134	1·440	1·943	2·447	3·143	3·707	5·959
7	·130	·263	·402	·549	·711	·896	1·119	1·415	1·895	2·365	2·998	3·499	5·408
8	·130	·262	·399	·546	·706	·889	1·108	1·397	1·860	2·306	2·896	3·355	5·041
9	·129	·261	·398	·543	·703	·883	1·100	1·383	1·833	2·262	2·821	3·250	4·781
10	·129	·260	·397	·542	·700	·879	1·093	1·372	1·812	2·228	2·764	3·169	4·587
11	·129	·260	·396	·540	·697	·876	1·088	1·363	1·796	2·201	2·718	3·106	4·437
12	·128	·259	·395	·539	·695	·873	1·083	1·356	1·782	2·179	2·681	3·055	4·318
13	·128	·259	·394	·538	·694	·870	1·079	1·350	1·771	2·160	2·650	3·012	4·221
14	·128	·258	·393	·537	·692	·868	1·076	1·345	1·761	2·145	2·624	2·977	4·140
15	·128	·258	·393	·536	·691	·866	1·074	1·341	1·753	2·131	2·602	2·947	4·073
16	·128	·258	·392	·535	·690	·865	1·071	1·337	1·746	2·120	2·583	2·921	4·015
17	·128	·257	·392	·534	·689	·863	1·069	1·333	1·740	2·110	2·567	2·898	3·965
18	·127	·257	·392	·534	·688	·862	1·067	1·330	1·734	2·101	2·552	2·878	3·922

lgorithmus A.

Ziel:
H_0 zu
zugu

Testergebnis

	H_0	H_0 nicht verwerfen
		✅

Der Plan:

- H_0 aufstellen.
- Daten **X** sammeln
- Widerspruchsbeweis: Wir nehmen H_0 an
- Falls **X** unwahrscheinlich unter H_0 ist, dann kann H_0 nicht stimmen und wird verworfen

Null Hypothesis Significance Testing

Idee: Behauptung (Hypothese) über die Eigenschaft von Populationen, welche durch empirische Untersuchungen auf Stichprobe überprüft werden muss.

Beispiel: *Neuer Algorithmus B ist schneller als Algorithmus A.*
Gegenhypothese H_1 („**exciting**“)

vs. *Neuer Algorithmus B ist NICHT schneller als Algorithmus A.*
Nullhypothese H_0 („**langweilig, status quo**“)

Ziel: Test, um aus Datenlage möglichst fehlerfrei H_0 zu verwerfen oder H_0 nicht zu verwerfen.

zugunsten H_1 — zugunsten H_0

Testergebnis \ Realität	H_0	H_1
H_0 verworfen	FP ☹	TP ✅
H_0 nicht verwerfen	TN ✅	FN ☹

$P(\text{Typ-1-Fehler}) \stackrel{!}{=} 5\%$

Typ-2-Fehler

Der Plan:

- H_0 aufstellen.
- Daten **X** sammeln
- Widerspruchsbeweis: Wir nehmen H_0 an
- Falls **X** unwahrscheinlich unter H_0 ist, dann kann H_0 nicht stimmen und wird verworfen

Ein Signifikanztest hier: Binomialtest

Ein Signifikanztest hier: Binomialtest

Szenario: Unser Server degradiert manchmal unter Hochlast-Anfrage S.

Ein Signifikanztest hier: Binomialtest

Szenario: Unser Server degradiert manchmal unter Hochlast-Anfrage S.

$$P(S \text{ langsam verarbeitet}) = 23\%$$

$$P(S \text{ schnell verarbeitet}) = 77\%$$

Ein Signifikanztest hier: Binomialtest

Szenario: Unser Server degradiert manchmal unter Hochlast-Anfrage S.

$$P(S \text{ langsam verarbeitet}) = 23\%$$
$$P(S \text{ schnell verarbeitet}) = 77\%$$

Spielen auf dem Server ein neues Update auf.
Wollen wissen, ob es jetzt besser ist.

Ein Signifikanztest hier: Binomialtest

Szenario: Unser Server degradiert manchmal unter Hochlast-Anfrage S.

$$P(S \text{ langsam verarbeitet}) = 23\%$$
$$P(S \text{ schnell verarbeitet}) = 77\%$$

Spielen auf dem Server ein neues Update auf.
Wollen wissen, ob es jetzt besser ist.

Experiment: Führen Anfrage S 100 mal auf dem neuen Server aus.

Ein Signifikanztest hier: Binomialtest

Szenario: Unser Server degradiert manchmal unter Hochlast-Anfrage S.

$$P(S \text{ langsam verarbeitet}) = 23\%$$
$$P(S \text{ schnell verarbeitet}) = 77\%$$

Spielen auf dem Server ein neues Update auf.
Wollen wissen, ob es jetzt besser ist.

Experiment: Führen Anfrage S 100 mal auf dem neuen Server aus.
Ergebnis: 100 Anfragen, davon 15 langsam.

Ein Signifikanztest hier: Binomialtest

Szenario: Unser Server degradiert manchmal unter Hochlast-Anfrage S.

$$P(S \text{ langsam verarbeitet}) = 23\%$$
$$P(S \text{ schnell verarbeitet}) = 77\%$$

Spielen auf dem Server ein neues Update auf.
Wollen wissen, ob es jetzt besser ist.

Experiment: Führen Anfrage S 100 mal auf dem neuen Server aus.
Ergebnis: 100 Anfragen, davon 15 langsam.

Frage: Ist der Server jetzt stabiler?

Ein Signifikanztest hier: Binomialtest

Szenario: Unser Server degradiert manchmal unter Hochlast-Anfrage S.

$$P(S \text{ langsam verarbeitet}) = 23\%$$
$$P(S \text{ schnell verarbeitet}) = 77\%$$

Spielen auf dem Server ein neues Update auf.
Wollen wissen, ob es jetzt besser ist.

Experiment: Führen Anfrage S 100 mal auf dem neuen Server aus.
Ergebnis: 100 Anfragen, davon 15 langsam.

Frage: Ist der Server jetzt stabiler?

H_0: Server genauso stabil wie vorher
H_1: Server stabiler/instabiler geworden

Ein Signifikanztest hier: Binomialtest

$$P(X_i = 1) = \theta$$
$$P(X_i = 0) = 1 - \theta$$

1. Modell definieren

$$X_1, X_2, \ldots, X_{100} \sim \text{Bernoulli}(\theta)$$

Ein Signifikanztest hier: Binomialtest

$$P(X_i = 1) = \theta$$
$$P(X_i = 0) = 1 - \theta$$

1. Modell definieren

$$X_1, X_2, \ldots, X_{100} \sim \text{Bernoulli}(\theta)$$

2. Null-/Gegenhypothese aufstellen

$$H_0 : \theta = 0.23, \quad H_1 : \theta \neq 0.23$$

Ein Signifikanztest hier: Binomialtest

$$P(X_i = 1) = \theta$$
$$P(X_i = 0) = 1 - \theta$$

1. Modell definieren

$$X_1, X_2, \ldots, X_{100} \sim \text{Bernoulli}(\theta)$$

2. Null-/Gegenhypothese aufstellen

$$H_0 : \theta = 0.23, \quad H_1 : \theta \neq 0.23$$

3. Teststatistik formulieren

$$T = \sum_{i=1}^{100} X_i$$

[hier einfache Interpretation: die Anzahl der Hits unter den 100 Versuchen]

Ein Signifikanztest hier: Binomialtest

$$P(X_i = 1) = \theta$$
$$P(X_i = 0) = 1 - \theta$$

1. Modell definieren $X_1, X_2, \ldots, X_{100} \sim \text{Bernoulli}(\theta)$

2. Null-/Gegenhypothese aufstellen $H_0 : \theta = 0.23, \quad H_1 : \theta \neq 0.23$

3. Teststatistik formulieren $T = \sum_{i=1}^{100} X_i$

[hier einfache Interpretation: die Anzahl der Hits unter den 100 Versuchen]

4. Verteilung der Teststatistik bestimmen
(unter H_0, Modellannahmen) $T \overset{H_0}{\sim} \text{Binom}(100, \theta)$

Ein Signifikanztest hier: Binomialtest

$$P(X_i = 1) = \theta$$
$$P(X_i = 0) = 1 - \theta$$

1. Modell definieren $X_1, X_2, \ldots, X_{100} \sim \text{Bernoulli}(\theta)$

2. Null-/Gegenhypothese aufstellen $H_0 : \theta = 0.23, \quad H_1 : \theta \neq 0.23$

3. Teststatistik formulieren $T = \sum_{i=1}^{100} X_i$

[hier einfache Interpretation: die Anzahl der Hits unter den 100 Versuchen]

4. Verteilung der Teststatistik bestimmen

(unter H_0, Modellannahmen) $T \overset{H_0}{\sim} \text{Binom}(100, \theta)$ $f(t) = \binom{100}{t}\theta^t(1-\theta)^{100-t}$

Ein Signifikanztest hier: Binomialtest

$$P(X_i = 1) = \theta$$
$$P(X_i = 0) = 1 - \theta$$

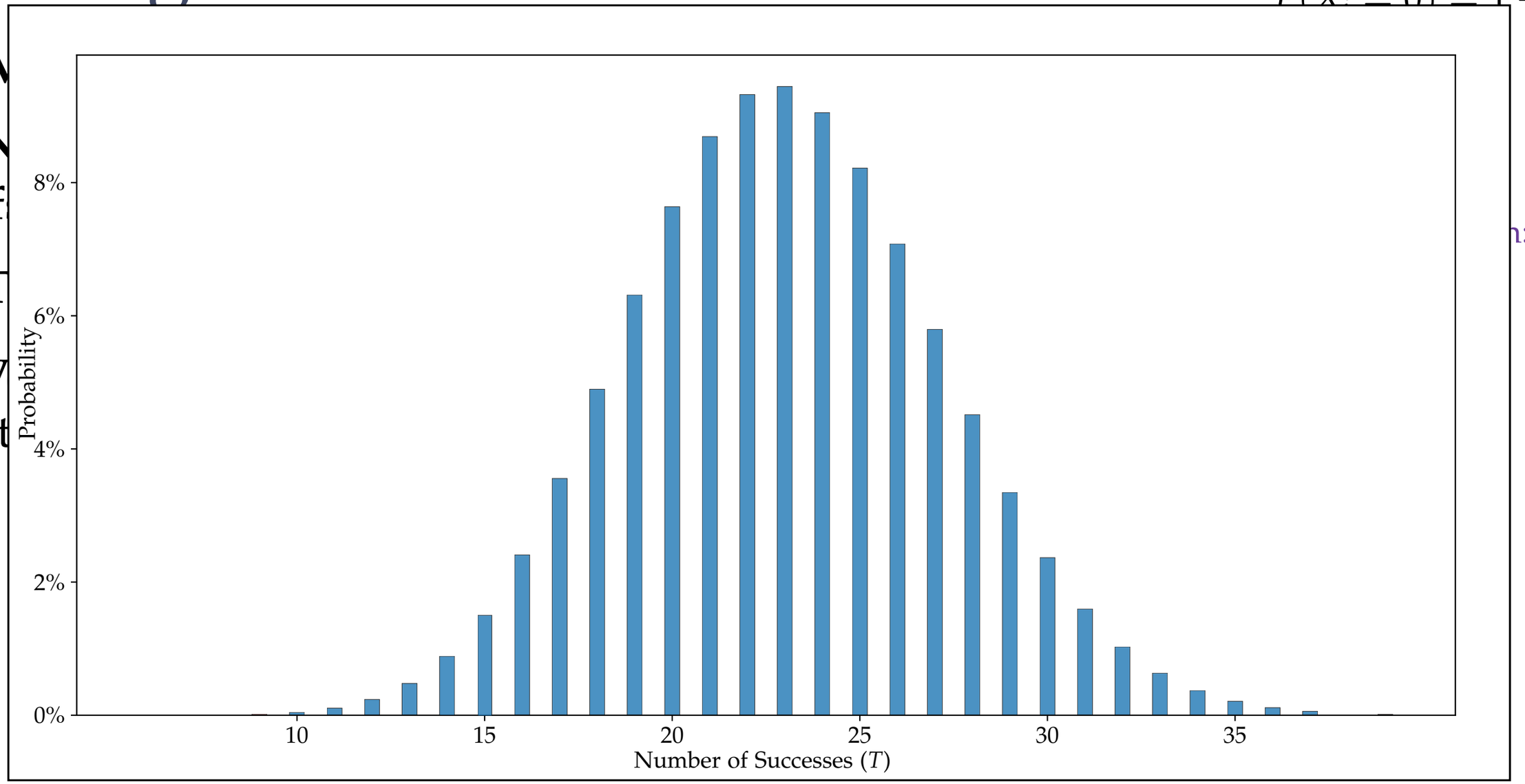
1. N

2. N
auf

3. T

4. V
(unt

n:



Ein Signifikanztest hier: Binomialtest

$$P(X_i = 1) = \theta$$
$$P(X_i = 0) = 1 - \theta$$

1. Modell definieren $X_1, X_2, \ldots, X_{100} \sim \text{Bernoulli}(\theta)$

2. Null-/Gegenhypothese aufstellen $H_0 : \theta = 0.23, \quad H_1 : \theta \neq 0.23$

3. Teststatistik formulieren $T = \sum_{i=1}^{100} X_i$

[hier einfache Interpretation: die Anzahl der Hits unter den 100 Versuchen]

4. Verteilung der Teststatistik bestimmen

(unter H_0, Modellannahmen) $T \overset{H_0}{\sim} \text{Binom}(100, \theta)$ $f(t) = \binom{100}{t}\theta^t(1-\theta)^{100-t}$

Ein Signifikanztest hier: Binomialtest

$$P(X_i = 1) = \theta$$
$$P(X_i = 0) = 1 - \theta$$

1. Modell definieren $X_1, X_2, \ldots, X_{100} \sim \text{Bernoulli}(\theta)$

2. Null-/Gegenhypothese aufstellen $H_0 : \theta = 0.23, \quad H_1 : \theta \neq 0.23$

3. Teststatistik formulieren $T = \sum_{i=1}^{100} X_i$

[hier einfache Interpretation: die Anzahl der Hits unter den 100 Versuchen]

4. Verteilung der Teststatistik bestimmen
(unter H_0, Modellannahmen) $T \overset{H_0}{\sim} \text{Binom}(100, \theta)$ $f(t) = \binom{100}{t}\theta^t(1-\theta)^{100-t}$

5. Daten erfassen und Prüfwert ausrechnen $x_1, \ldots, x_{100} \in \{0, 1\}$ Realisierungen von $X_1, \ldots, X_{100}$; $t_{\text{obs}} = \sum_{i=1}^{100} x_i \quad = 14$

Ein Signifikanztest hier: Binomialtest

$$P(X_i = 1) = \theta$$
$$P(X_i = 0) = 1 - \theta$$

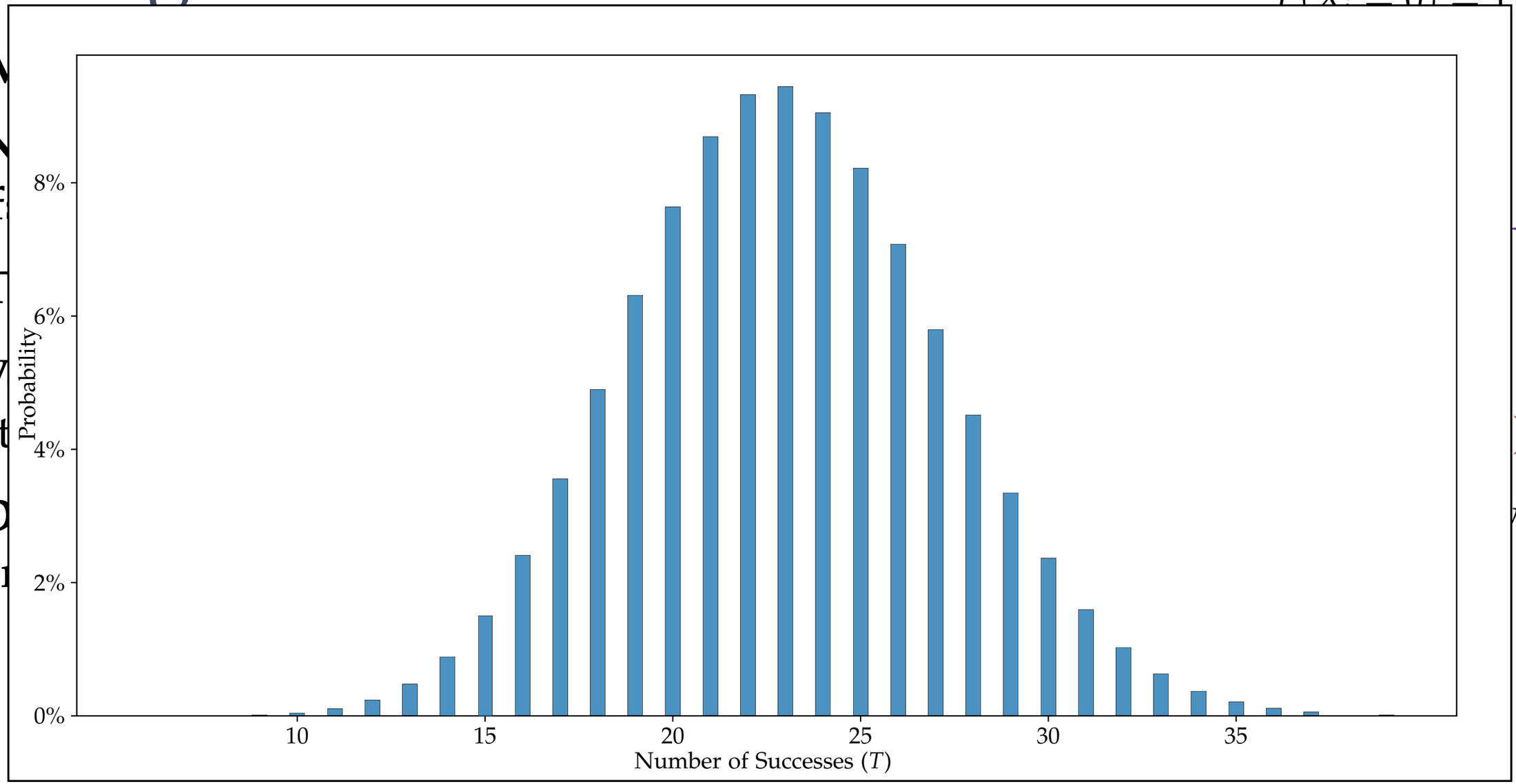
1. N

2. N
auf

3. T

4. V
(unt

5. D
aus



n:

$)^{100-t}$

von

Ein Signifikanztest hier: Binomialtest

$P(X_i = 1) = \theta$
$P(X_i = 0) = 1 - \theta$

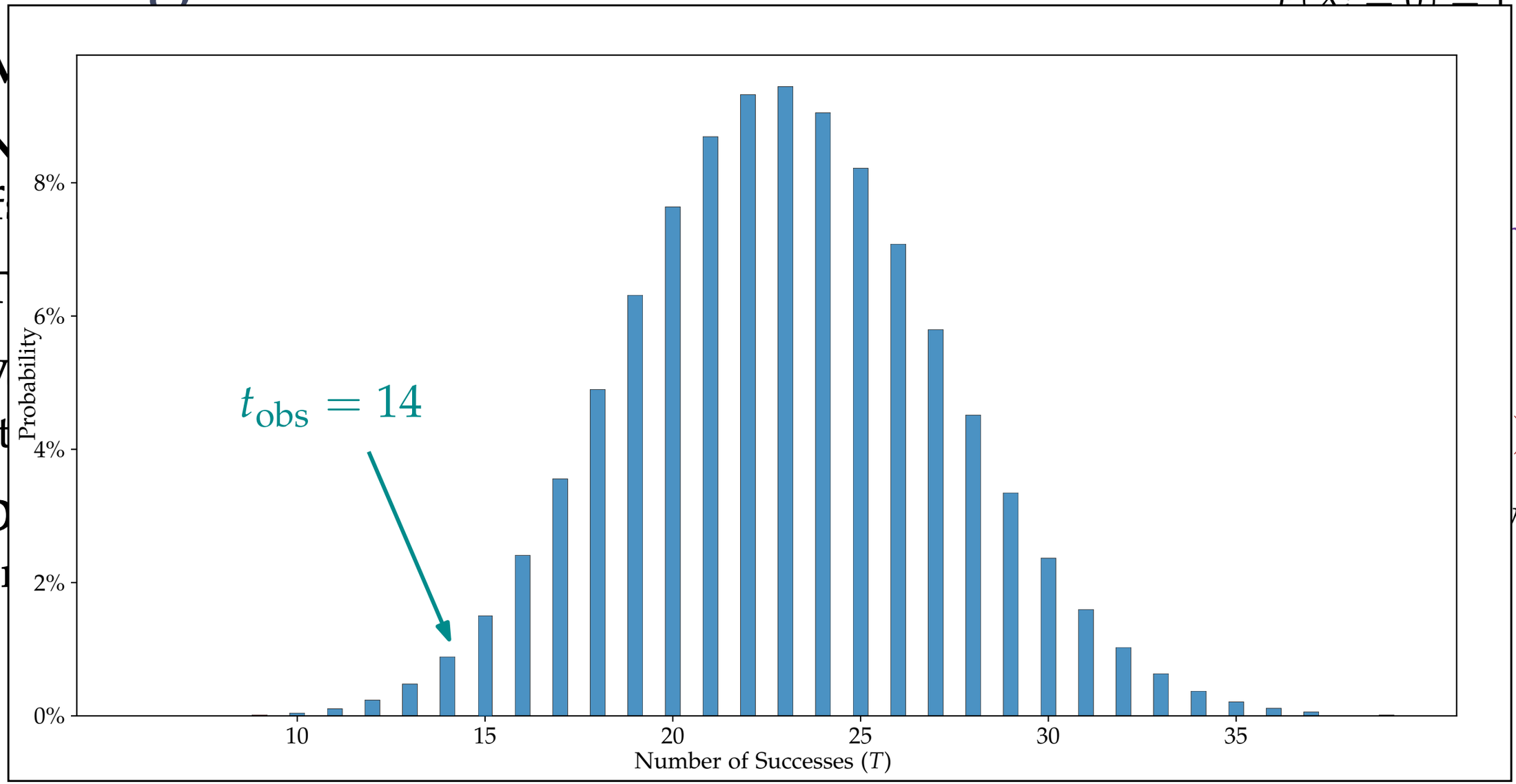
1. N

2. N
auf

3. T

4. V
(unt

5. D
aus



n:

$)^{100-t}$

von

Ein Signifikanztest hier: Binomialtest

$$P(X_i = 1) = \theta$$
$$P(X_i = 0) = 1 - \theta$$

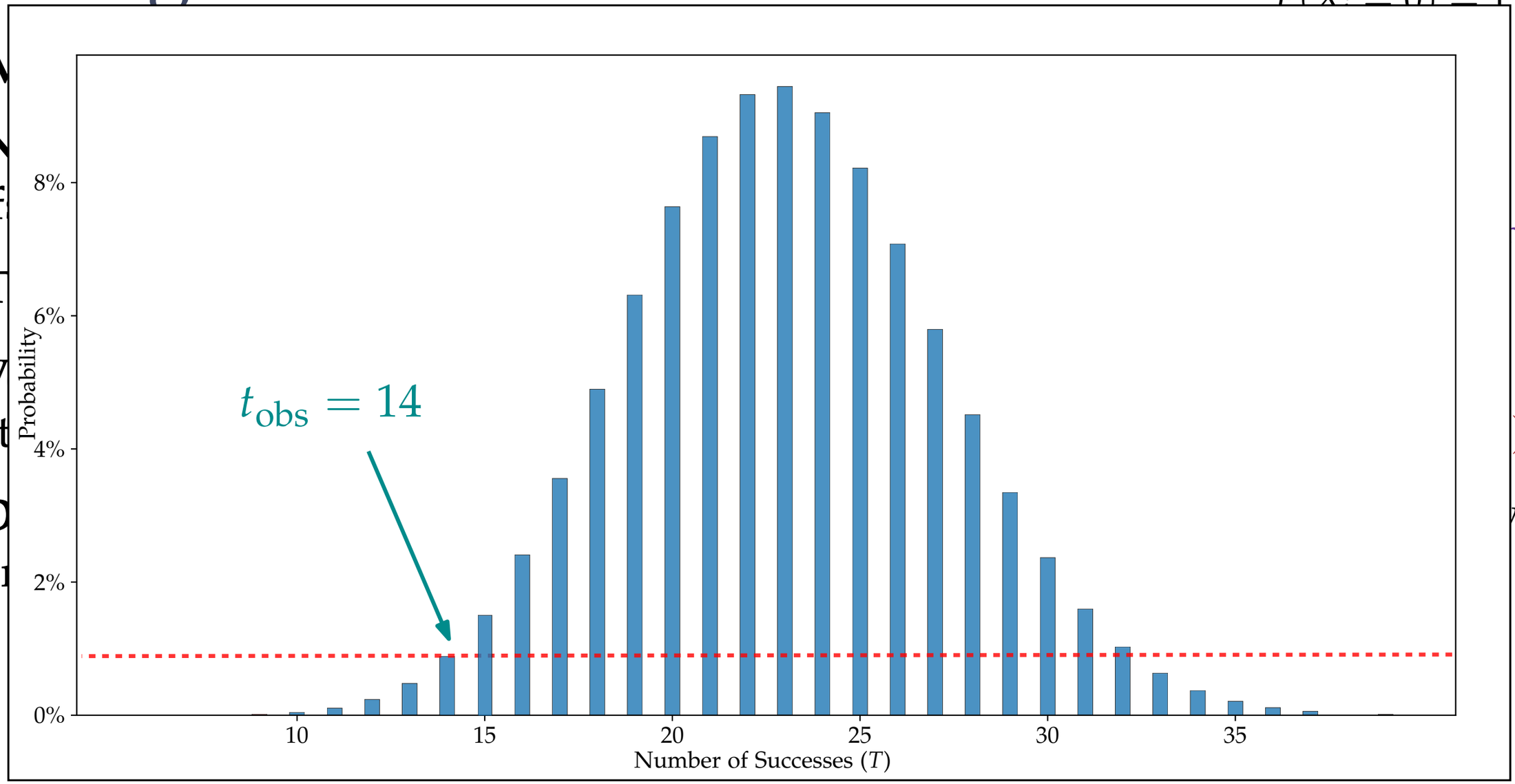
1. N

2. N
auf

3. T

4. V
(unt

5. D
aus



n:

$\theta^{t}(1-\theta)^{100-t}$

von

Ein Signifikanztest hier: Binomialtest

$P(X_i = 1) = \theta$
$P(X_i = 0) = 1 - \theta$

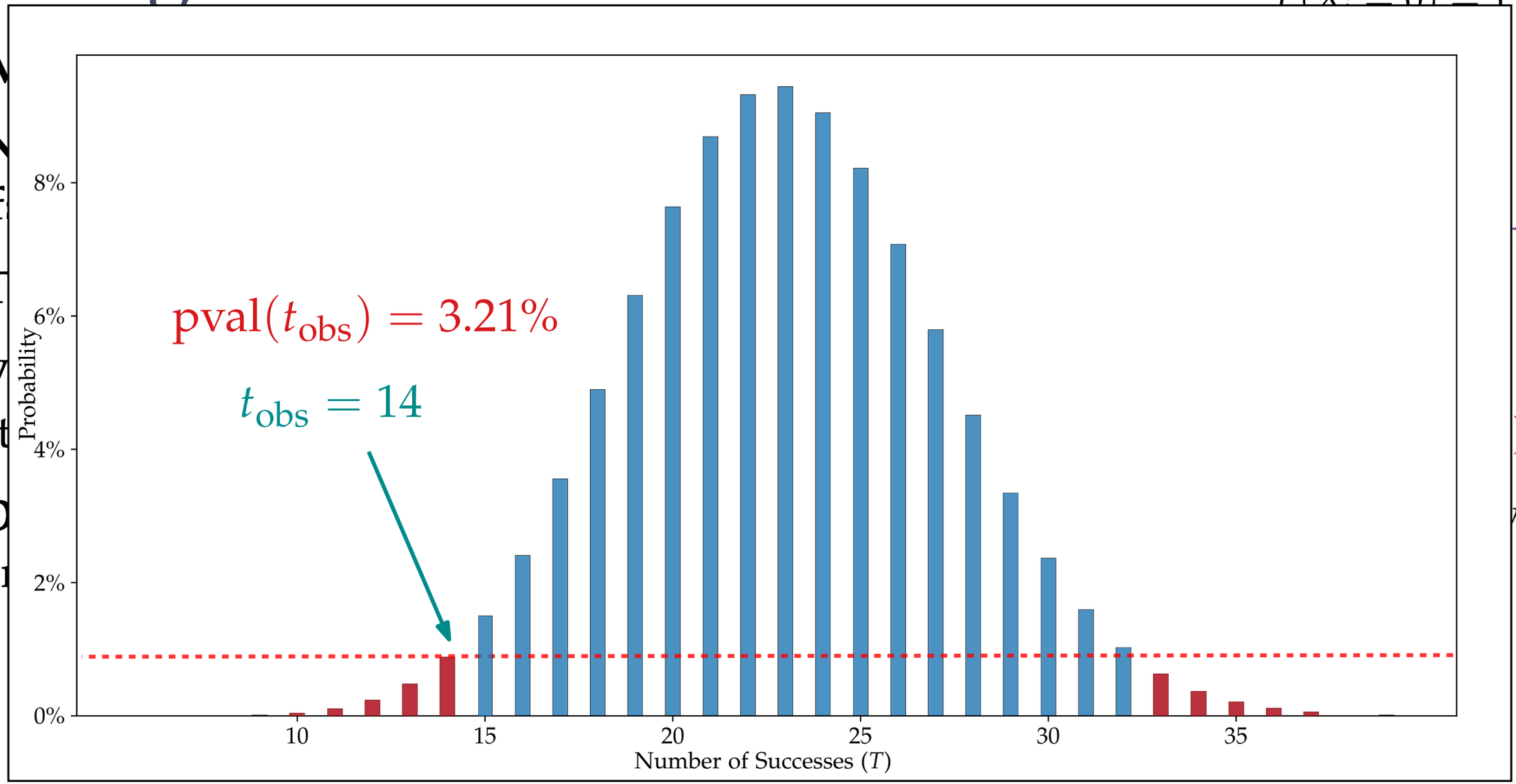
1. N

2. N
auf

3. T

4. V
(unt

5. D
aus



Ein Signifikanztest hier: Binomialtest

$$P(X_i = 1) = \theta$$
$$P(X_i = 0) = 1 - \theta$$

1. Modell definieren $X_1, X_2, \ldots, X_{100} \sim \text{Bernoulli}(\theta)$

2. Null-/Gegenhypothese aufstellen $H_0 : \theta = 0.23, \quad H_1 : \theta \neq 0.23$

3. Teststatistik formulieren $T = \sum_{i=1}^{100} X_i$

[hier einfache Interpretation: die Anzahl der Hits unter den 100 Versuchen]

4. Verteilung der Teststatistik bestimmen
(unter H_0, Modellannahmen) $T \overset{H_0}{\sim} \text{Binom}(100, \theta)$ $f(t) = \binom{100}{t}\theta^t(1-\theta)^{100-t}$

5. Daten erfassen und Prüfwert ausrechnen $x_1, \ldots, x_{100} \in \{0, 1\}$ Realisierungen von $X_1, \ldots, X_{100}$; $t_{\text{obs}} = \sum_{i=1}^{100} x_i \quad = 14$

Ein Signifikanztest hier: Binomialtest

$$P(X_i = 1) = \theta$$
$$P(X_i = 0) = 1 - \theta$$

1. Modell definieren $X_1, X_2, \ldots, X_{100} \sim \text{Bernoulli}(\theta)$

2. Null-/Gegenhypothese aufstellen $H_0 : \theta = 0.23, \quad H_1 : \theta \neq 0.23$

3. Teststatistik formulieren $T = \sum_{i=1}^{100} X_i$

[hier einfache Interpretation: die Anzahl der Hits unter den 100 Versuchen]

4. Verteilung der Teststatistik bestimmen
(unter H_0, Modellannahmen) $T \overset{H_0}{\sim} \text{Binom}(100, \theta)$ $f(t) = \binom{100}{t}\theta^t(1-\theta)^{100-t}$

5. Daten erfassen und Prüfwert ausrechnen $x_1, \ldots, x_{100} \in \{0, 1\}$ Realisierungen von $X_1, \ldots, X_{100}$; $t_{\text{obs}} = \sum_{i=1}^{100} x_i \quad = 14$

6. P-Value ausrechnen

$$\text{pval}(t_{\text{obs}}) = \sum_{t:\, P(T=t) \leq P(T=t_{\text{obs}})} P(T = t)$$

WK ein Wert mind. so extrem wie beobachtet zu sehen

Ein Signifikanztest hier: Binomialtest

$$P(X_i = 1) = \theta$$
$$P(X_i = 0) = 1 - \theta$$

1. Modell definieren

$$X_1, X_2, \ldots, X_{100} \sim \text{Bernoulli}(\theta)$$

2. Null-/Gegenhypothese aufstellen

$$H_0 : \theta = 0.23, \quad H_1 : \theta \neq 0.23$$

3. Teststatistik formulieren

$$T = \sum_{i=1}^{100} X_i$$

[hier einfache Interpretation: die Anzahl der Hits unter den 100 Versuchen]

4. Verteilung der Teststatistik bestimmen
(unter H_0, Modellannahmen)

$$T \overset{H_0}{\sim} \text{Binom}(100, \theta) \quad f(t) = \binom{100}{t}\theta^t(1-\theta)^{100-t}$$

5. Daten erfassen und Prüfwert ausrechnen

$$x_1, \ldots, x_{100} \in \{0, 1\} \text{ Realisierungen von } X_1, \ldots, X_{100}; \; t_{\text{obs}} = \sum_{i=1}^{100} x_i = 14$$

6. P-Value ausrechnen

$$\text{pval}(t_{\text{obs}}) = \sum_{t:\, P(T=t) \leq P(T=t_{\text{obs}})} P(T = t)$$

WK ein Wert mind. so extrem wie beobachtet zu sehen

7. P-Wert mit Signifikanzniveau vergleichen, entscheiden

$$\text{pval}(t_{\text{obs}}) \leq 5\% \Rightarrow H_0 \text{ verwerfen!}$$
$$\text{pval}(t_{\text{obs}}) > 5\% \Rightarrow H_0 \text{ nicht verwerfen!}$$

Ein Signifikanztest hier: Binomialtest

$$P(X_i = 1) = \theta$$
$$P(X_i = 0) = 1 - \theta$$

1. Modell definieren $X_1, X_2, \ldots, X_{100} \sim \text{Bernoulli}(\theta)$

2. Null-/Gegenhypothese aufstellen $H_0 : \theta = 0.23, \quad H_1 : \theta \neq 0.23$

3. Teststatistik formulieren $T = \sum_{i=1}^{100} X_i$

[hier einfache Interpretation: die Anzahl der Hits unter den 100 Versuchen]

4. Verteilung der Teststatistik bestimmen
(unter H_0, Modellannahmen) $T \overset{H_0}{\sim} \text{Binom}(100, \theta)$ $f(t) = \binom{100}{t}\theta^t(1-\theta)^{100-t}$

5. Daten erfassen und Prüfwert ausrechnen $x_1, \ldots, x_{100} \in \{0, 1\}$ Realisierungen von $X_1, \ldots, X_{100}$; $t_{\text{obs}} = \sum_{i=1}^{100} x_i \quad = 14$

6. P-Value ausrechnen

$$\text{pval}(t_{\text{obs}}) = \sum_{t:\, P(T=t) \leq P(T=t_{\text{obs}})} P(T = t)$$

WK ein Wert mind. so extrem wie beobachtet zu sehen

Unser Fall: $\text{pval}(t_{\text{obs}}) = 3.21\% \Rightarrow H_0$ verwerfen

7. P-Wert mit Signifikanzniveau vergleichen, entscheiden

$\text{pval}(t_{\text{obs}}) \leq 5\% \Rightarrow H_0$ verwerfen!

$\text{pval}(t_{\text{obs}}) > 5\% \Rightarrow H_0$ nicht verwerfen!

Ein Signifikanztest hier: Binomialtest

$P(X_i = 1) = \theta$
$P(X_i = 0) = 1 - \theta$

1. Modell definieren $X_1, X_2, \ldots, X_{100} \sim \text{Bernoulli}(\theta)$

2. Null-/Gegenhypothese aufstellen $H_0 : \theta = 0.23, \quad H_1 : \theta \neq 0.23$

3. Teststatistik formulieren $T = \sum_{i=1}^{100} X_i$

[hier einfache Interpretation: die Anzahl der Hits unter den 100 Versuchen]

4. Verteilung der Teststatistik bestimmen
(unter H_0, Modellannahmen) $T \overset{H_0}{\sim} \text{Binom}(100, \theta)$

Typ-1-Fehlerrate?
H_0 verworfen obwohl H_0 gilt

5. Daten erfassen und Prüfwert ausrechnen $x_1, \ldots, x_{100} \in \{0, 1\}$ Realisierungen von $X_1, \ldots, X_{100}$; $t_{\text{obs}} = \sum_{i=1}^{100} x_i \quad = 14$

6. P-Value ausrechnen

$$\text{pval}(t_{\text{obs}}) = \sum_{t:\, P(T=t) \leq P(T=t_{\text{obs}})} P(T = t)$$

WK ein Wert mind. so extrem wie beobachtet zu sehen

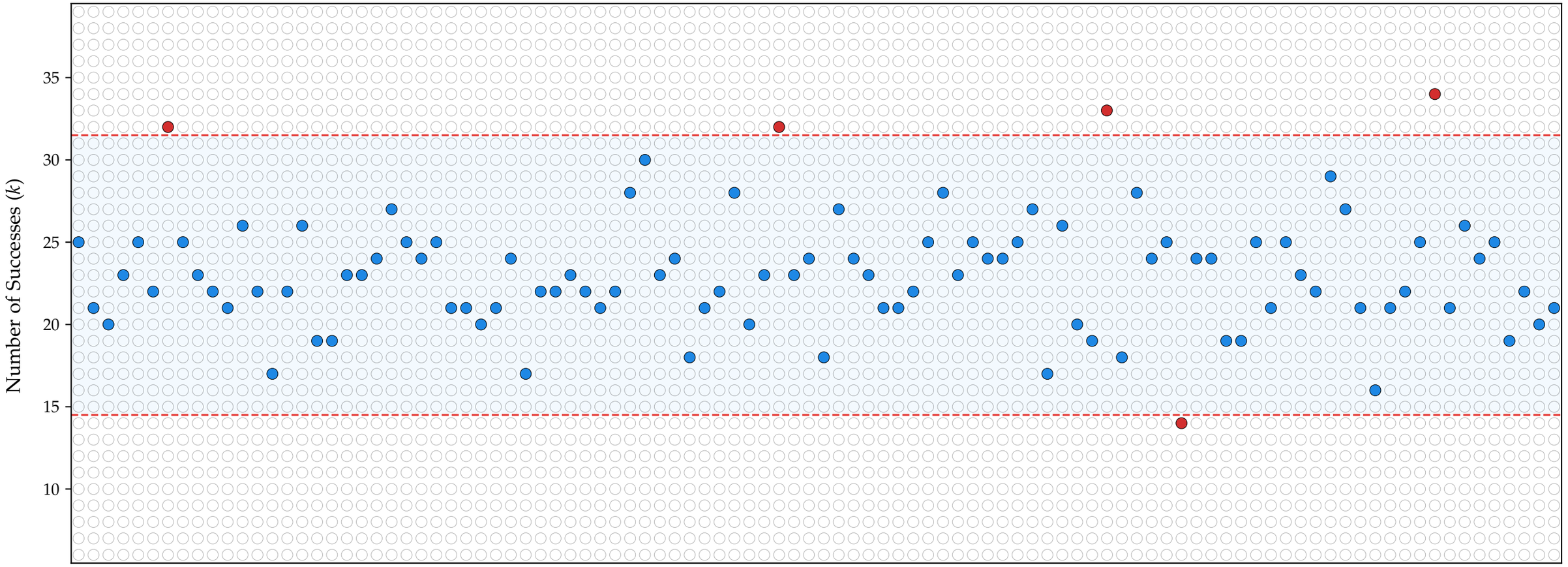
Unser Fall: $\text{pval}(t_{\text{obs}}) = 3.21\% \Rightarrow H_0$ verwerfen

7. P-Wert mit Signifikanzniveau vergleichen, entscheiden

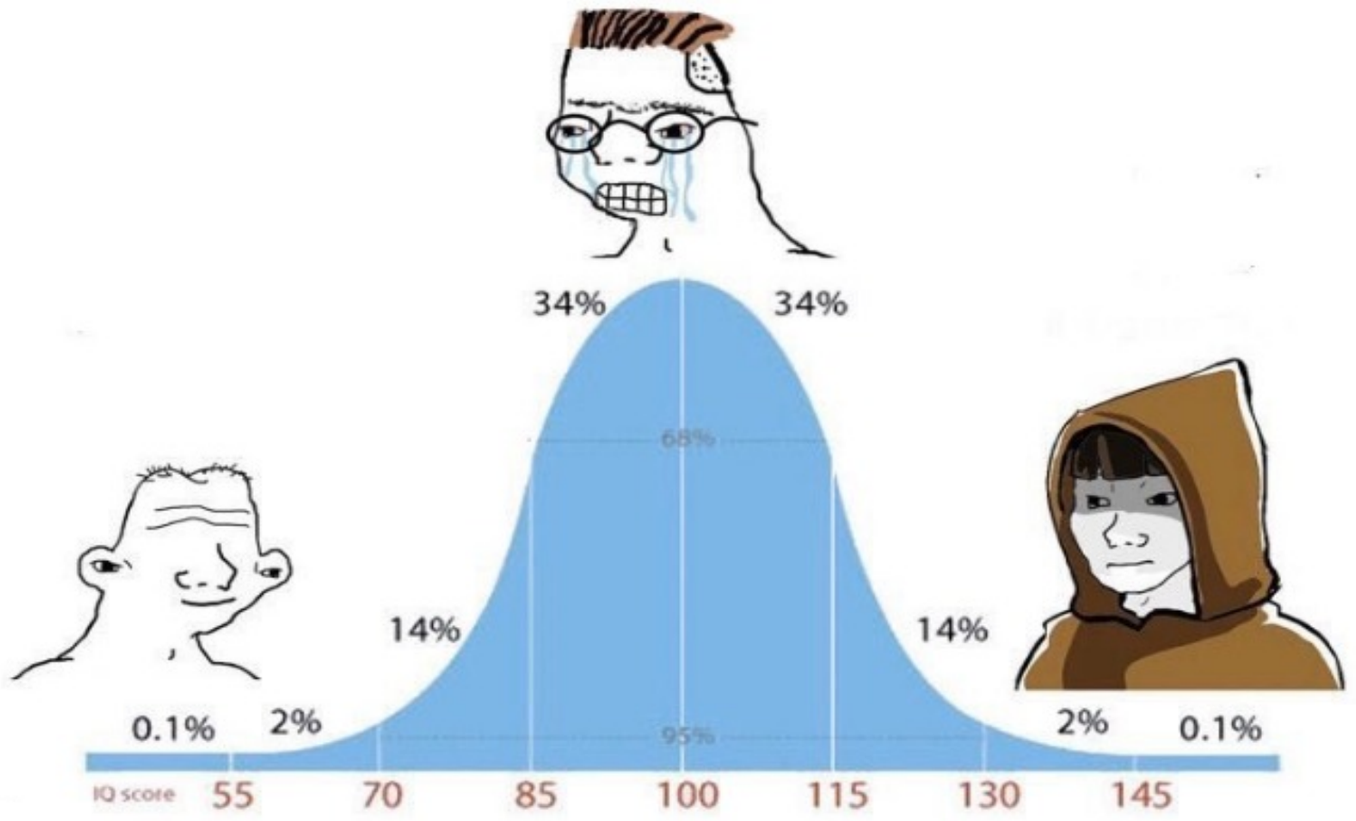
$\text{pval}(t_{\text{obs}}) \leq 5\% \Rightarrow H_0$ verwerfen!
$\text{pval}(t_{\text{obs}}) > 5\% \Rightarrow H_0$ nicht verwerfen!

100 zufällige Durchführungen mit Teststatistik

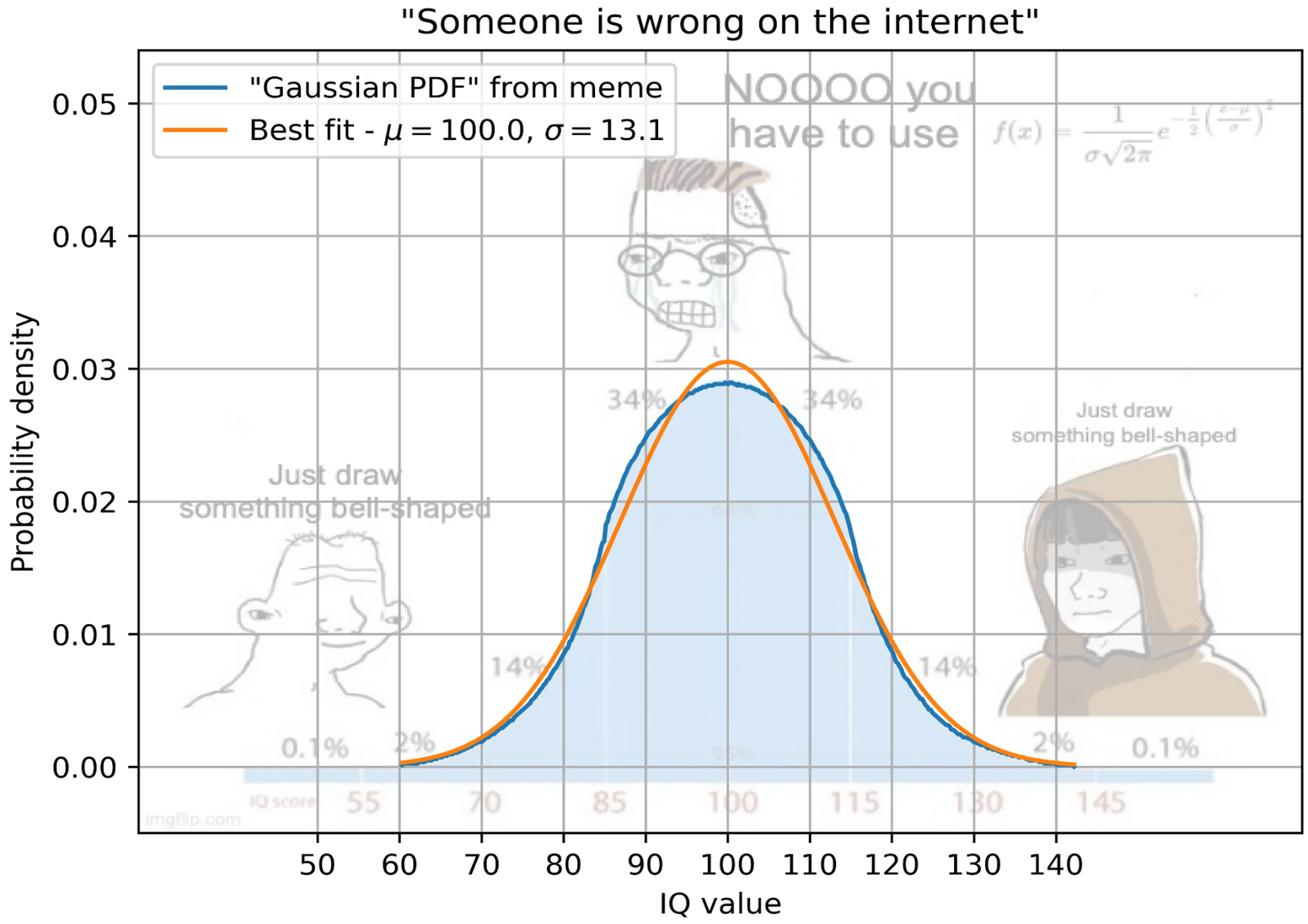
H_0 wahr: $\theta = .23$ rote Punkte mit pval $\leq 5\%$



Normalverteilung



Normalverteilung



Normalverteilung

Parametrisiert durch *Lage* μ und *Skala* σ

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right)$$

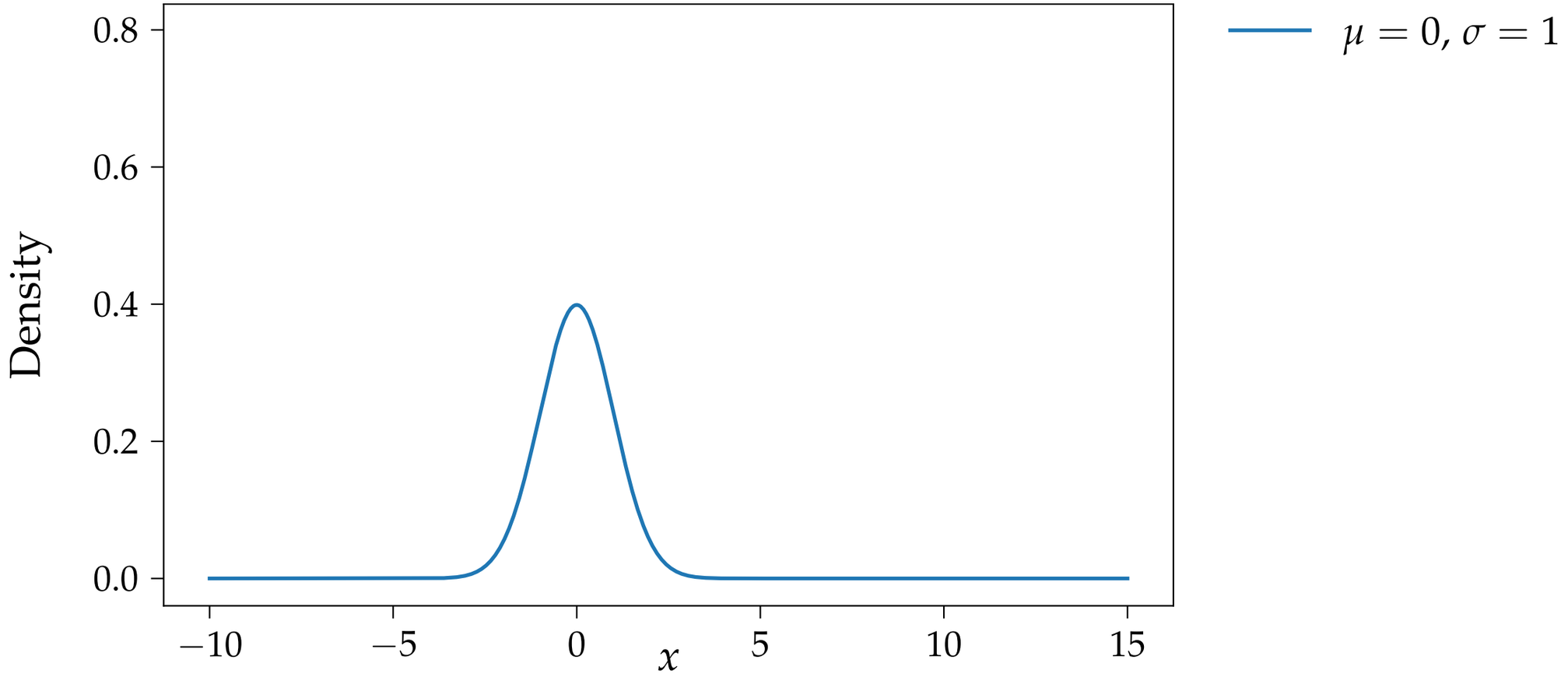
$$\mathrm{E}[X] = \mu$$
$$\mathrm{Var}[X] = \sigma^2$$

Normalverteilung

Parametrisiert durch *Lage* μ und *Skala* σ

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right)$$

$$\mathrm{E}[X] = \mu$$
$$\mathrm{Var}[X] = \sigma^2$$

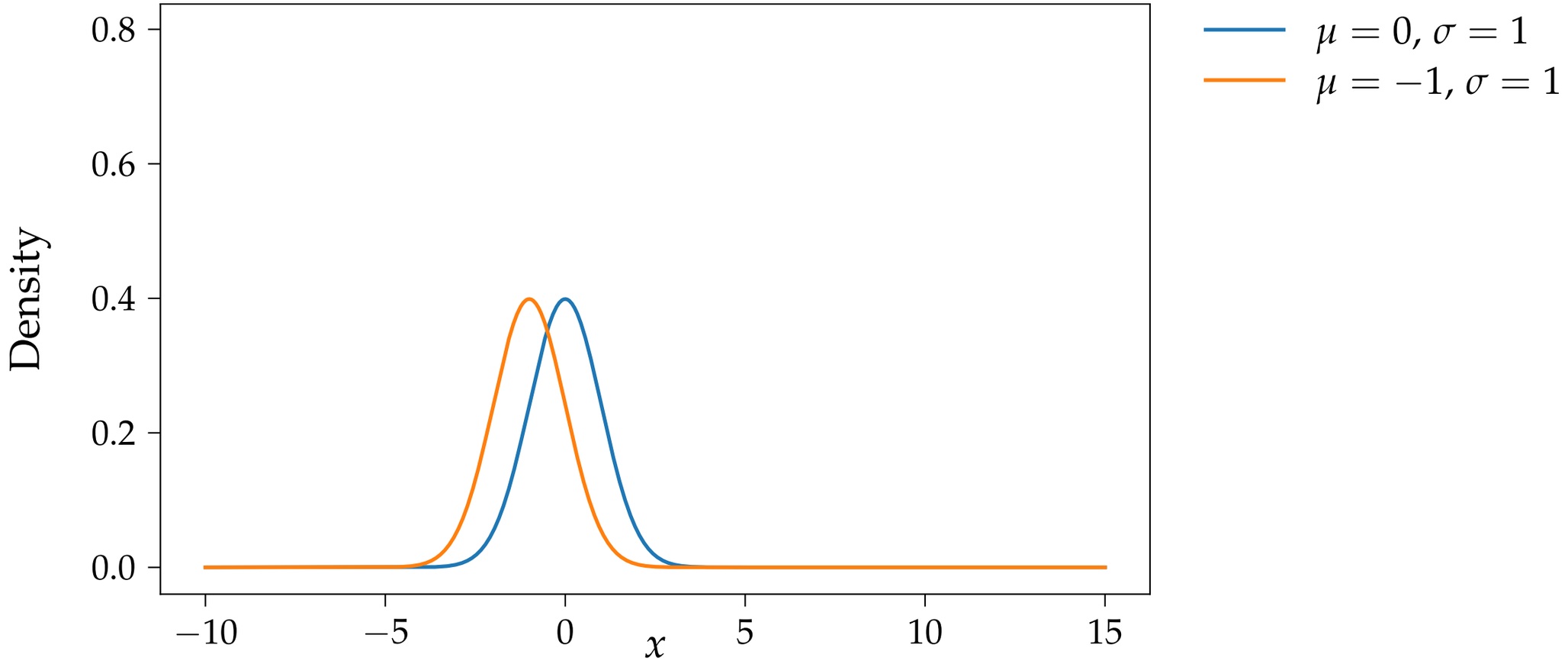


Normalverteilung

Parametrisiert durch *Lage* μ und *Skala* σ

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right)$$

$$\mathrm{E}[X] = \mu$$
$$\mathrm{Var}[X] = \sigma^2$$

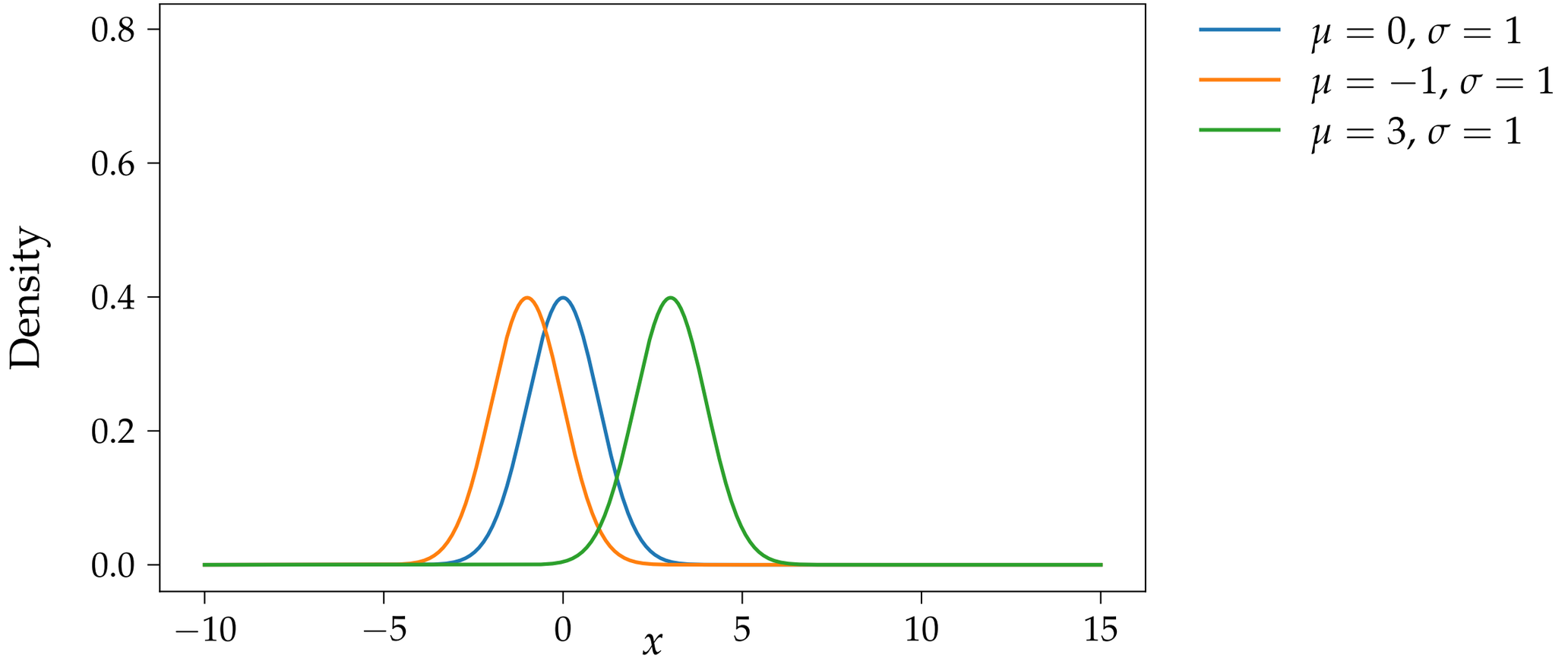


Normalverteilung

Parametrisiert durch *Lage* μ und *Skala* σ

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right)$$

$$\mathrm{E}[X] = \mu$$
$$\mathrm{Var}[X] = \sigma^2$$

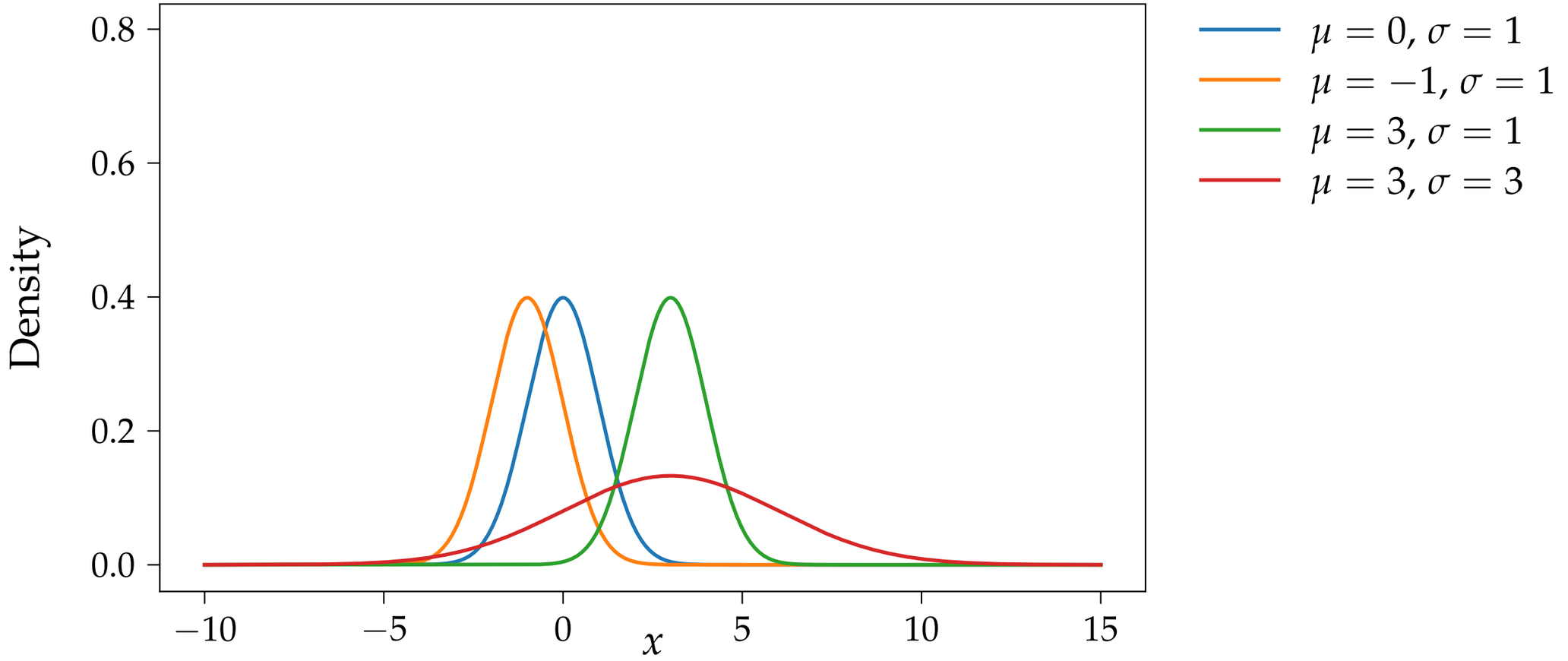


Normalverteilung

Parametrisiert durch *Lage* μ und *Skala* σ

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right)$$

$$\mathrm{E}[X] = \mu$$
$$\mathrm{Var}[X] = \sigma^2$$

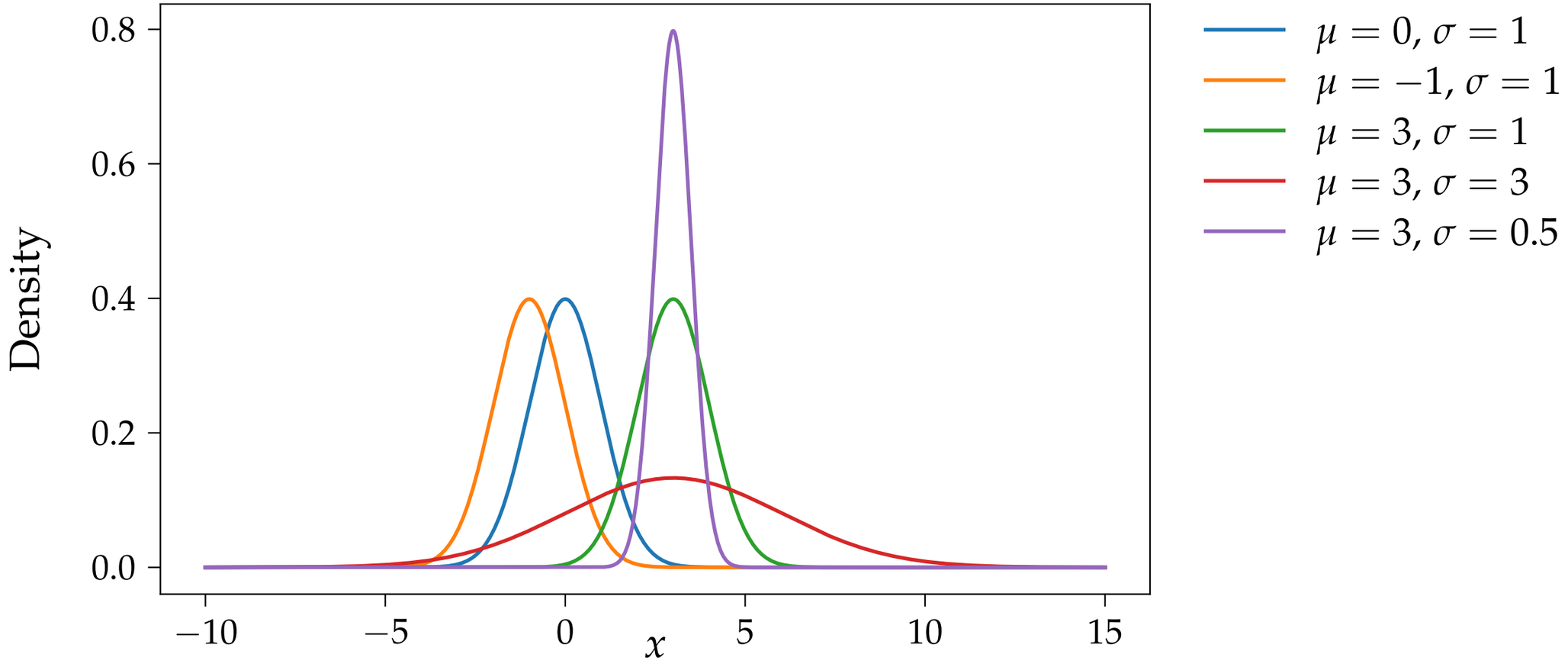


Normalverteilung

Parametrisiert durch *Lage* μ und *Skala* σ

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right)$$

$$\mathrm{E}[X] = \mu$$
$$\mathrm{Var}[X] = \sigma^2$$



Normalverteilung

Parametrisiert durch *Lage* μ und *Skala* σ

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right)$$

$$\mathrm{E}[X] = \mu$$
$$\mathrm{Var}[X] = \sigma^2$$

Szenario:

Vergleichen Laufzeit von Algorithmus A mit Algorithmus B.

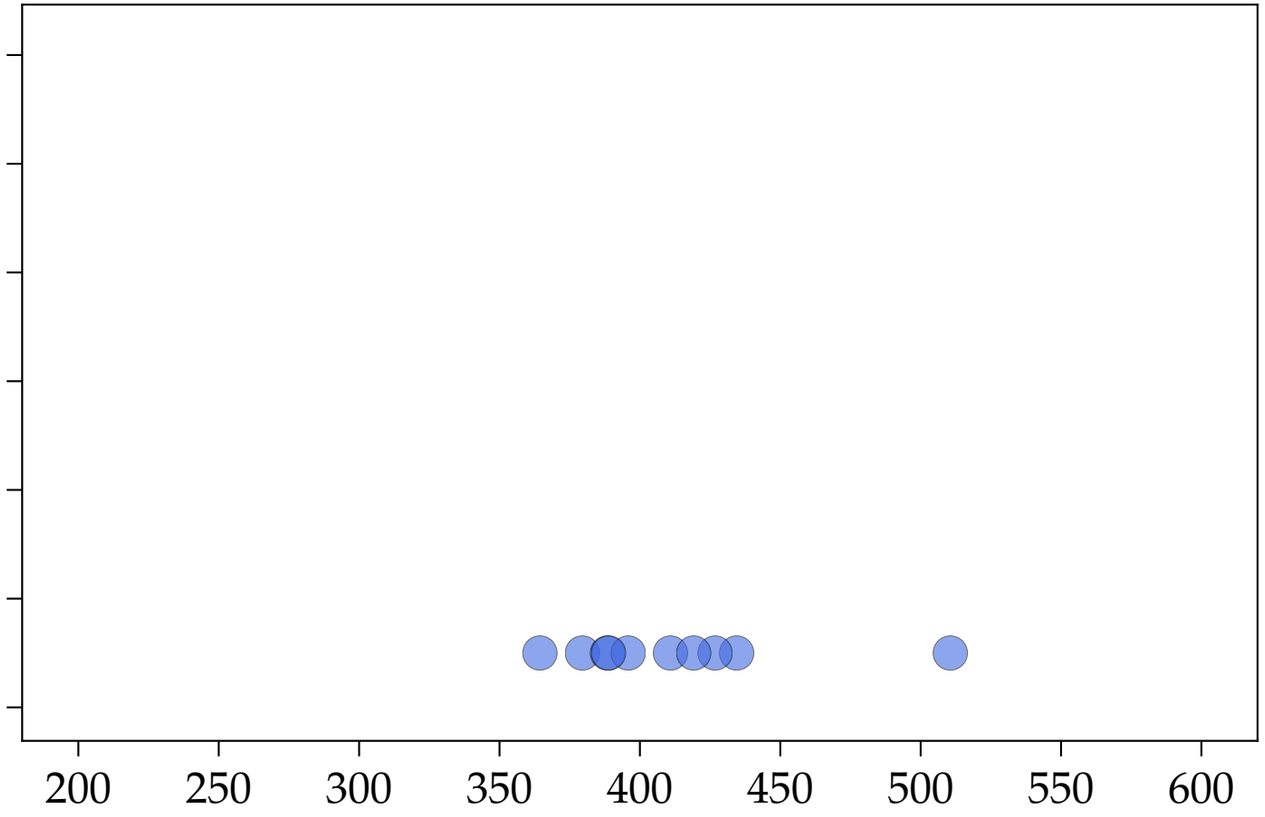
Modellieren Verteilung als Normalverteilung.

Normalverteilung

Parametrisiert durch *Lage* μ und *Skala* σ

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right)$$

$$\mathrm{E}[X] = \mu$$
$$\mathrm{Var}[X] = \sigma^2$$



Szenario:

Vergleichen Laufzeit von Algorithmus A mit Algorithmus B.

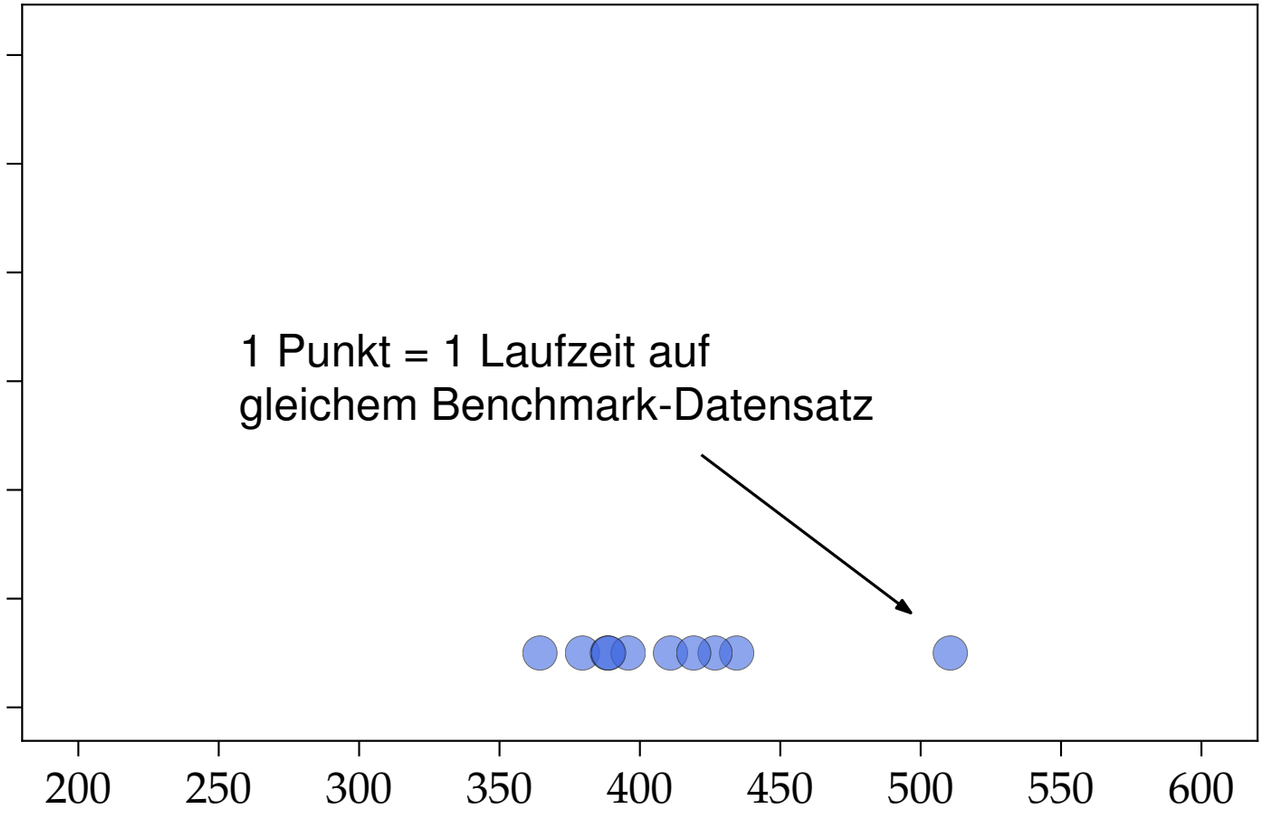
Modellieren Verteilung als Normalverteilung.

Normalverteilung

Parametrisiert durch *Lage* μ und *Skala* σ

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right)$$

$$\mathrm{E}[X] = \mu$$
$$\mathrm{Var}[X] = \sigma^2$$



Szenario:

Vergleichen Laufzeit von Algorithmus A mit Algorithmus B.

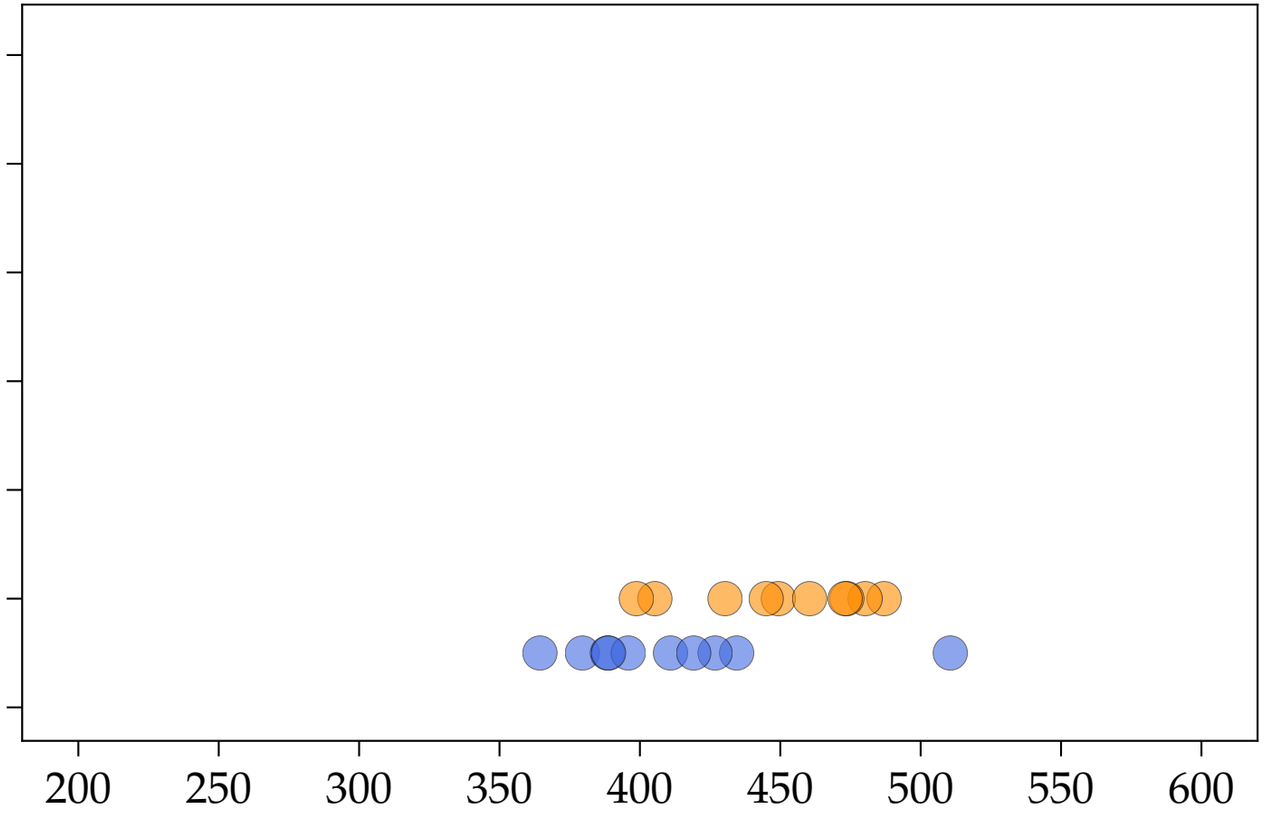
Modellieren Verteilung als Normalverteilung.

Normalverteilung

Parametrisiert durch *Lage* μ und *Skala* σ

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right)$$

$$\mathrm{E}[X] = \mu$$
$$\mathrm{Var}[X] = \sigma^2$$



Szenario:

Vergleichen Laufzeit von Algorithmus A mit Algorithmus B.

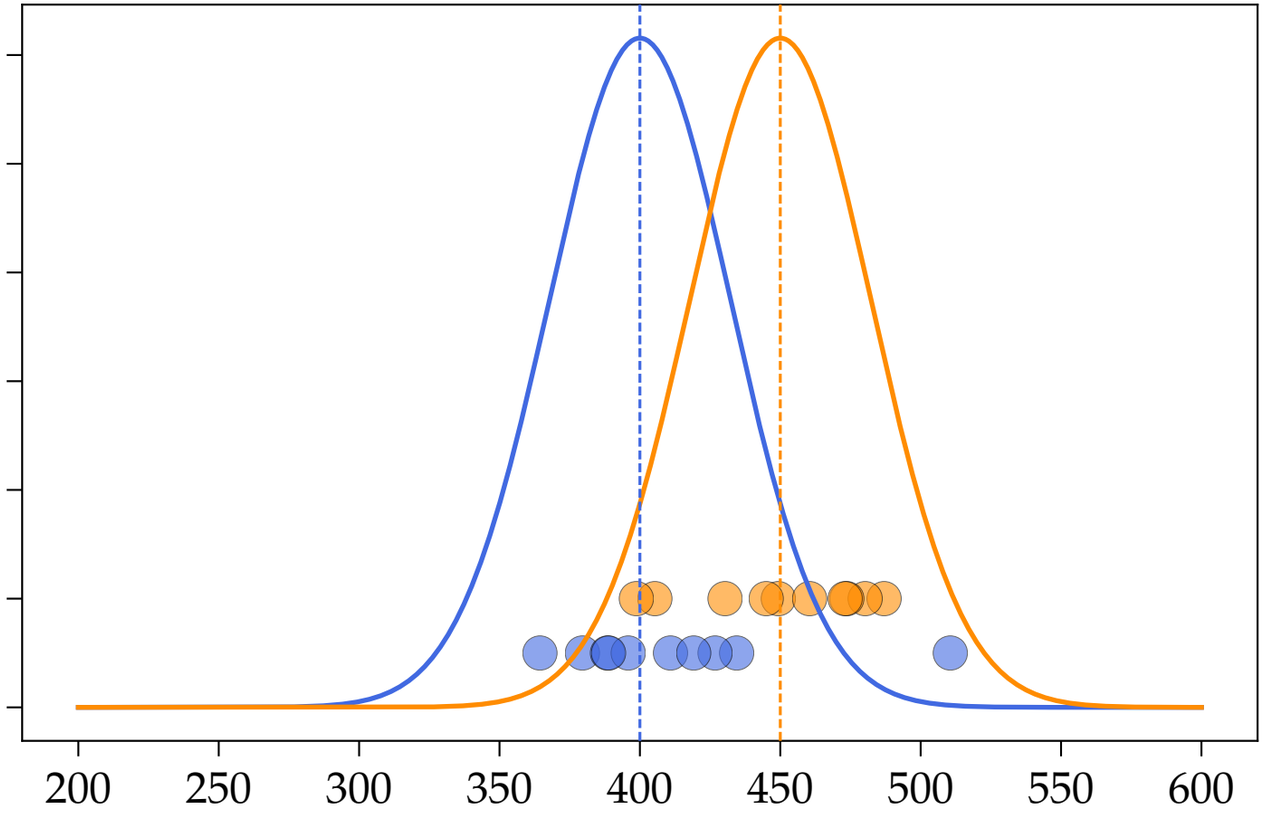
Modellieren Verteilung als Normalverteilung.

Normalverteilung

Parametrisiert durch *Lage* μ und *Skala* σ

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right)$$

$$\mathrm{E}[X] = \mu$$
$$\mathrm{Var}[X] = \sigma^2$$



Szenario:

Vergleichen Laufzeit von Algorithmus A mit Algorithmus B.

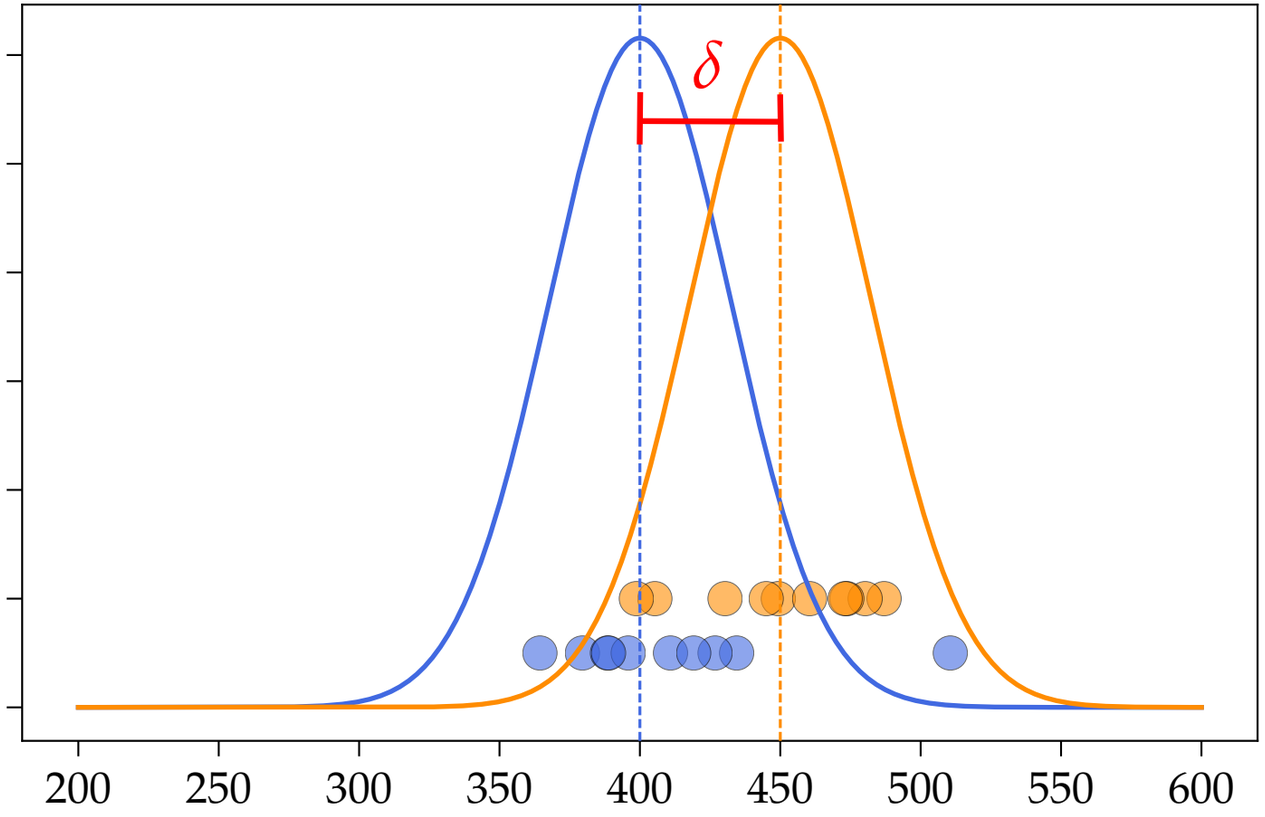
Modellieren Verteilung als Normalverteilung.

Normalverteilung

Parametrisiert durch *Lage* μ und *Skala* σ

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right)$$

$$\mathrm{E}[X] = \mu$$
$$\mathrm{Var}[X] = \sigma^2$$



Szenario:

Vergleichen Laufzeit von Algorithmus A mit Algorithmus B.

Modellieren Verteilung als Normalverteilung.

$H_0 : \delta = 0; \quad H_1 : \delta \neq 0$

Warum Normal…?

Warum Normal. . . ?

Warum Normal. . . ?

- Triffst du oft (empirisch) in der Natur, ist einfach zum Rechnen, das machen alle so, . . .

Warum Normal…?

- Triffst du oft (empirisch) in der Natur, ist einfach zum Rechnen, das machen alle so, …
- Die einzige Fehlerverteilung, welche den arithmetischen Mittelwert (als ML-Schätzer) rationalisiert, ist $\mathcal{N}$

Warum Normal…?

- Triffst du oft (empirisch) in der Natur, ist einfach zum Rechnen, das machen alle so, …
- Die einzige Fehlerverteilung, welche den arithmetischen Mittelwert (als ML-Schätzer) rationalisiert, ist $\mathcal{N}$
- Modellhaft bestehen Fehler aus vielen kleinen zufälligen „Mini-Fehlern“ $\pm\epsilon$. $\rightarrow$ Summe der Mini-Fehler konvergiert zu $\mathcal{N}$

Warum Normal…?

- Triffst du oft (empirisch) in der Natur, ist einfach zum Rechnen, das machen alle so, …
- Die einzige Fehlerverteilung, welche den arithmetischen Mittelwert (als ML-Schätzer) rationalisiert, ist $\mathcal{N}$
- Modellhaft bestehen Fehler aus vielen kleinen zufälligen „Mini-Fehlern“ $\pm\epsilon$. $\rightarrow$ Summe der Mini-Fehler konvergiert zu $\mathcal{N}$

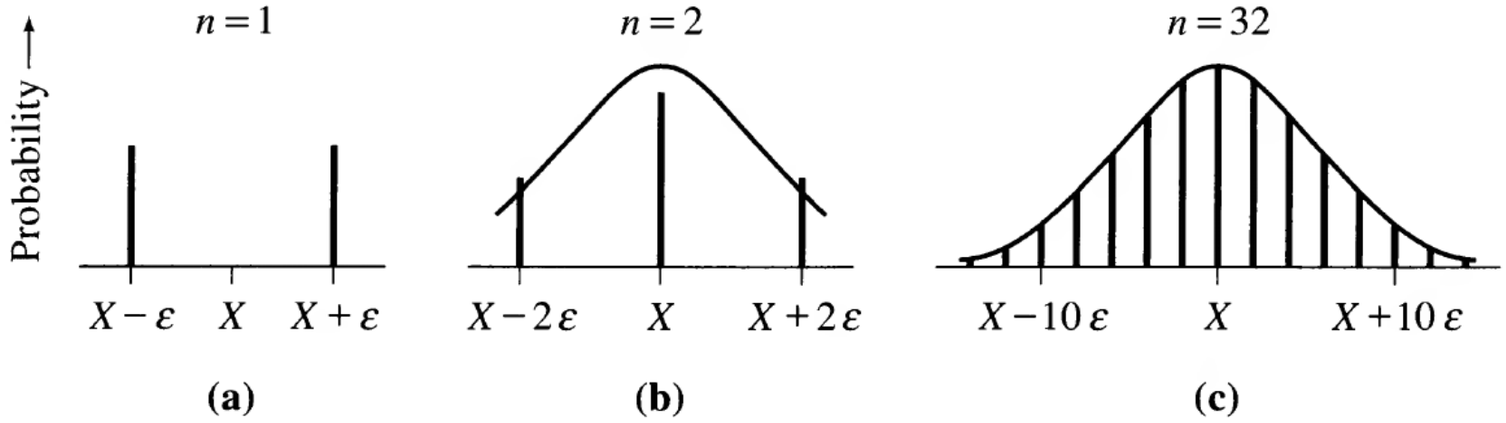


Figure 10.5. Distribution of measurements subject to n random errors of magnitude ε, for $n = 1$, 2, and 32. The continuous curves superimposed on (b) and (c) are Gaussians with the same center and width. (The vertical scales differ in the three graphs.)

Warum Normal…?

- Triffst du oft (empirisch) in der Natur, ist einfach zum Rechnen, das machen alle so, …
- Die einzige Fehlerverteilung, welche den arithmetischen Mittelwert (als ML-Schätzer) rationalisiert, ist $\mathcal{N}$
- Modellhaft bestehen Fehler aus vielen kleinen zufälligen „Mini-Fehlern“ $\pm\epsilon$. $\rightarrow$ Summe der Mini-Fehler konvergiert zu $\mathcal{N}$
- **Central Limit Theorem**: Für identisch & unabhängige ZV $X_1, \ldots, X_n$ mit $\mathrm{E}[X_i] = 0$, $\mathrm{Var}[X_i] = \sigma^2$ ist

$$\frac{X_1 + X_2 + \cdots + X_n}{n} \rightarrow \mathcal{N}\left(0, \frac{\sigma^2}{n}\right)$$

Warum Normal…?

- Triffst du oft (empirisch) in der Natur, ist einfach zum Rechnen, das machen alle so, …
- Die einzige Fehlerverteilung, welche den arithmetischen Mittelwert (als ML-Schätzer) rationalisiert, ist $\mathcal{N}$
- Modellhaft bestehen Fehler aus vielen kleinen zufälligen „Mini-Fehlern“ $\pm\epsilon$. $\rightarrow$ Summe der Mini-Fehler konvergiert zu $\mathcal{N}$
- **Central Limit Theorem**: Für ~~identisch~~ & unabhängige ZV $X_1, \ldots, X_n$ mit $\mathrm{E}[X_i] = 0$, ~~$\mathrm{Var}[X_i] = \sigma^2$~~ ist $\mathrm{Var}[X_i] = \sigma_i^2$

$$\frac{X_1 + X_2 + \cdots + X_n}{n} \rightarrow \mathcal{N}\left(0, \frac{\sigma^2}{n}\right)$$

Warum Normal…?

- Triffst du oft (empirisch) in der Natur, ist einfach zum Rechnen, das machen alle so, …
- Die einzige Fehlerverteilung, welche den arithmetischen Mittelwert (als ML-Schätzer) rationalisiert, ist $\mathcal{N}$
- Modellhaft bestehen Fehler aus vielen kleinen zufälligen „Mini-Fehlern“ $\pm\epsilon$. $\rightarrow$ Summe der Mini-Fehler konvergiert zu $\mathcal{N}$
- **Central Limit Theorem**: Für ~~identisch~~ & unabhängige ZV $X_1, \ldots, X_n$ mit $\mathrm{E}[X_i] = 0$, ~~$\mathrm{Var}[X_i] = \sigma^2$~~ ist $\mathrm{Var}[X_i] = \sigma_i^2$

$$\frac{X_1 + X_2 + \cdots + X_n}{n} \rightarrow \mathcal{N}\left(0, \frac{\sigma^2}{n}\right)$$

wobei

$$\frac{\sum_i^n \mathrm{E}[|X_i|^3]}{(\sum_i^n \sigma_i^2)^{3/2}} \rightarrow 0$$

Warum Normal…?

- Triffst du oft (empirisch) in der Natur, ist einfach zum Rechnen, das machen alle so, …
- Die einzige Fehlerverteilung, welche den arithmetischen Mittelwert (als ML-Schätzer) rationalisiert, ist $\mathcal{N}$
- Modellhaft bestehen Fehler aus vielen kleinen zufälligen „Mini-Fehlern“ $\pm\epsilon$. $\rightarrow$ Summe der Mini-Fehler konvergiert zu $\mathcal{N}$
- **Central Limit Theorem**: Für ~~identisch~~ & unabhängige ZV $X_1, \ldots, X_n$ mit $\mathrm{E}[X_i] = 0$, ~~$\mathrm{Var}[X_i] = \sigma^2$~~ ist $\mathrm{Var}[X_i] = \sigma_i^2$

$$\frac{X_1 + X_2 + \cdots + X_n}{n} \rightarrow \mathcal{N}\left(0, \frac{\sigma^2}{n}\right) \quad \sum_i \sigma_i^2 / n^2$$

wobei

$$\frac{\sum_i^n \mathrm{E}[|X_i|^3]}{(\sum_i^n \sigma_i^2)^{3/2}} \rightarrow 0$$

Warum Normal…?

- Triffst du oft (empirisch) in der Natur, ist einfach zum Rechnen, das machen alle so, …
- Die einzige Fehlerverteilung, welche den arithmetischen Mittelwert (als ML-Schätzer) rationalisiert, ist $\mathcal{N}$
- Modellhaft bestehen Fehler aus vielen kleinen zufälligen „Mini-Fehlern“ $\pm\epsilon$. $\rightarrow$ Summe der Mini-Fehler konvergiert zu $\mathcal{N}$
- **Central Limit Theorem**: Für ~~identisch~~ & unabhängige ZV $X_1, \ldots, X_n$ mit $\mathrm{E}[X_i] = 0$, ~~$\mathrm{Var}[X_i] = \sigma^2$~~ ist $\mathrm{Var}[X_i] = \sigma_i^2$

$$\frac{X_1 + X_2 + \cdots + X_n}{n} \rightarrow \mathcal{N}\left(0, \frac{\sigma^2}{n}\right)$$

$\sum_i \sigma_i^2 / n^2$

wobei

$$\frac{\sum_i^n \mathrm{E}[|X_i|^3]}{(\sum_i^n \sigma_i^2)^{3/2}} \rightarrow 0$$

- **MaxEnt:** Für feste μ, σ ist $X \sim \mathcal{N}(\mu, \sigma^2)$ die einzige Wahl mit $\mathrm{E}[X] = \mu$ und $\mathrm{Var}[X] = \sigma^2$ und X mit maximaler Entropie.

Zweistichproben-t-Test

Zweistichproben-t-Test

1. Modell definieren

$$X_1, \ldots, X_n \sim \mathcal{N}(\mu, \sigma^2)$$
$$Y_1, \ldots, Y_m \sim \mathcal{N}(\mu + \delta, \sigma^2)$$

alle paarweise unabhängig

Zweistichproben-t-Test

1. Modell definieren

$X_1, \ldots, X_n \sim \mathcal{N}(\mu, \sigma^2)$ alle paarweise
$Y_1, \ldots, Y_m \sim \mathcal{N}(\mu + \delta, \sigma^2)$ unabhängig

2. Null-/Gegenhypothese aufstellen

$H_0 : \delta = \delta_0, \quad H_1 : \delta \neq \delta_0$ [hier $\delta_0 = 0$]

3. Teststatistik formulieren

$$T = \sqrt{\frac{nm(n+m-2)}{n+m}} \frac{\bar{\mathbf{X}} - \bar{\mathbf{Y}} - \delta_0}{\sqrt{\|\mathbf{X} - \bar{\mathbf{X}}\|^2 + \|\mathbf{Y} - \bar{\mathbf{Y}}\|^2}}$$

4. Verteilung der Teststatistik bestimmen

(unter H_0, Modellannahmen)

$T \overset{H_0}{\sim} \text{StudentT}(n+m-2)$

Zweistichproben-t-Test

1. Modell definieren

$$X_1, \ldots, X_n \sim \mathcal{N}(\mu, \sigma^2)$$ alle paarweise
$$Y_1, \ldots, Y_m \sim \mathcal{N}(\mu + \delta, \sigma^2)$$ unabhängig

2. Null-/Gegenhypothese aufstellen

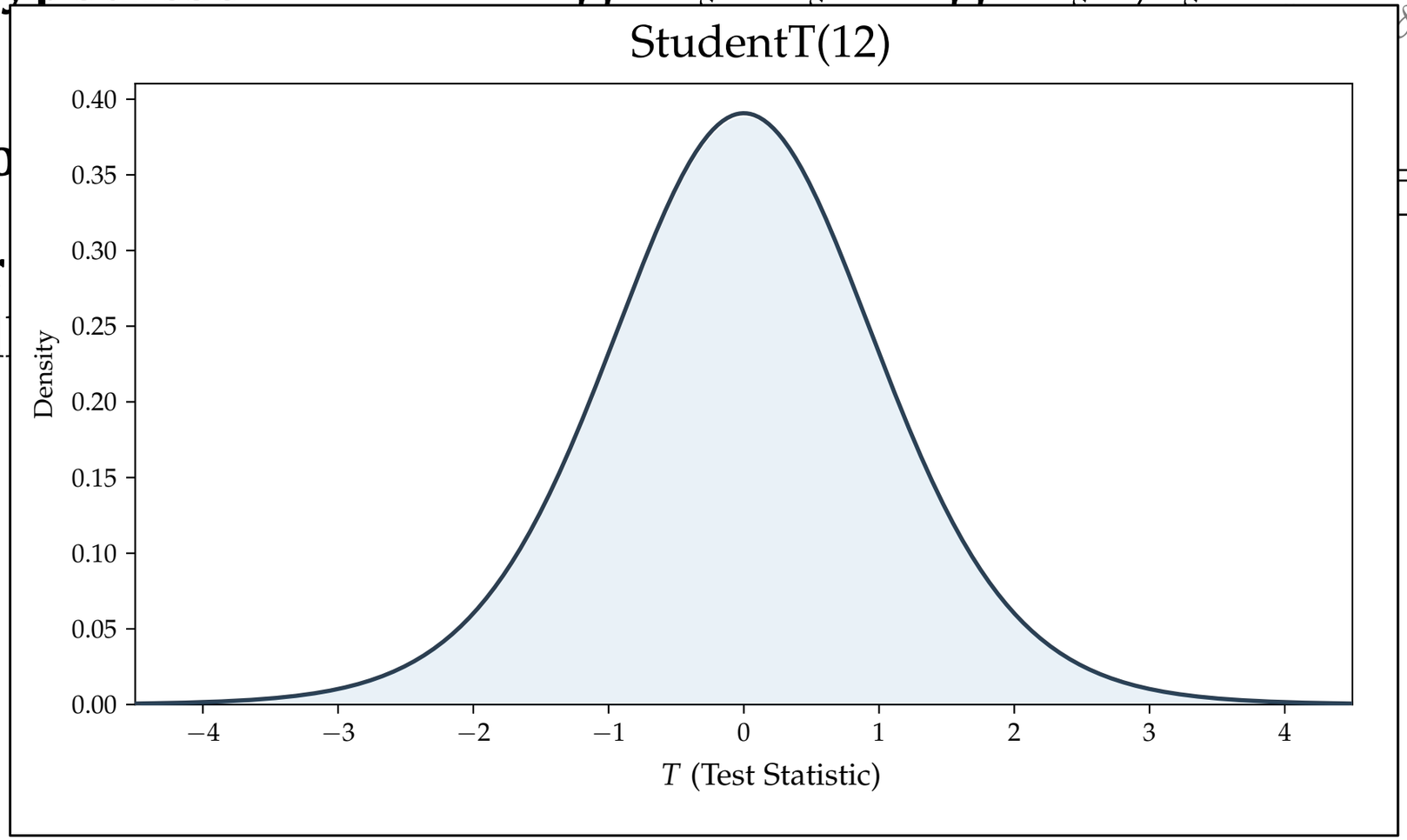
$\delta_0 = 0]$

3. Teststatistik fo

$$\frac{\quad - \delta_0}{\quad + \|\mathbf{Y} - \bar{\mathbf{Y}}\|^2}$$

4. Verteilung der

(unter H_0, Model



Zweistichproben-t-Test

1. Modell definieren

$X_1, \ldots, X_n \sim \mathcal{N}(\mu, \sigma^2)$ alle paarweise
$Y_1, \ldots, Y_m \sim \mathcal{N}(\mu + \delta, \sigma^2)$ unabhängig

2. Null-/Gegenhypothese aufstellen

$H_0 : \delta = \delta_0, \quad H_1 : \delta \neq \delta_0$ [hier $\delta_0 = 0$]

3. Teststatistik formulieren

$$T = \sqrt{\frac{nm(n+m-2)}{n+m}} \frac{\bar{\mathbf{X}} - \bar{\mathbf{Y}} - \delta_0}{\sqrt{\|\mathbf{X} - \bar{\mathbf{X}}\|^2 + \|\mathbf{Y} - \bar{\mathbf{Y}}\|^2}}$$

4. Verteilung der Teststatistik bestimmen

(unter H_0, Modellannahmen)

$T \overset{H_0}{\sim} \text{StudentT}(n+m-2)$

Zweistichproben-t-Test

1. Modell definieren

$$X_1, \dots, X_n \sim \mathcal{N}(\mu, \sigma^2) \quad \text{alle paarweise}$$
$$Y_1, \dots, Y_m \sim \mathcal{N}(\mu + \delta, \sigma^2) \quad \text{unabhängig}$$

2. Null-/Gegenhypothese aufstellen

$$H_0 : \delta = \delta_0, \quad H_1 : \delta \neq \delta_0 \quad [\text{hier } \delta_0 = 0]$$

3. Teststatistik formulieren

$$T = \sqrt{\frac{nm(n+m-2)}{n+m}} \frac{\bar{\mathbf{X}} - \bar{\mathbf{Y}} - \delta_0}{\sqrt{\|\mathbf{X} - \bar{\mathbf{X}}\|^2 + \|\mathbf{Y} - \bar{\mathbf{Y}}\|^2}}$$

4. Verteilung der Teststatistik bestimmen
(unter H_0, Modellannahmen)

$$T \overset{H_0}{\sim} \text{StudentT}(n+m-2)$$

5. Daten erfassen und Prüfwert ausrechnen

$$t_{\text{obs}} = \sqrt{\frac{nm(n+m-2)}{n+m}} \frac{\bar{\mathbf{x}} - \bar{\mathbf{y}} - \delta_0}{\sqrt{\|\mathbf{x} - \bar{\mathbf{x}}\|^2 + \|\mathbf{y} - \bar{\mathbf{y}}\|^2}}$$

6. P-Wert ausrechnen

$$\text{pval}(t_{\text{obs}}) = \int_{\{t:\, f(t) \leq f(t_{\text{obs}})\}} f(t)\, \mathrm{d}t \quad = P(|T| > |t_{\text{obs}}|)$$

Zweistichproben-t-Test

1. Modell definieren

$$X_1, \dots, X_n \sim \mathcal{N}(\mu, \sigma^2) \quad \text{alle paarweise}$$
$$Y_1, \dots, Y_m \sim \mathcal{N}(\mu + \delta, \sigma^2) \quad \text{unabhängig}$$

2. Null-/Gegenhypothese aufstellen

$$H_0 : \delta = \delta_0, \quad H_1 : \delta \neq \delta_0 \quad [\text{hier } \delta_0 = 0]$$

3. Teststatistik formulieren

$$T = \sqrt{\frac{nm(n+m-2)}{n+m}} \frac{\bar{\mathbf{X}} - \bar{\mathbf{Y}} - \delta_0}{\sqrt{\|\mathbf{X} - \bar{\mathbf{X}}\|^2 + \|\mathbf{Y} - \bar{\mathbf{Y}}\|^2}}$$

4. Verteilung der Teststatistik bestimmen
(unter H_0, Modellannahmen)

$$T \overset{H_0}{\sim} \text{StudentT}(n+m-2)$$

5. Daten erfassen und Prüfwert ausrechnen

$$t_{\text{obs}} = \sqrt{\frac{nm(n+m-2)}{n+m}} \frac{\bar{\mathbf{x}} - \bar{\mathbf{y}} - \delta_0}{\sqrt{\|\mathbf{x} - \bar{\mathbf{x}}\|^2 + \|\mathbf{y} - \bar{\mathbf{y}}\|^2}}$$

6. P-Wert ausrechnen

$$\text{pval}(t_{\text{obs}}) = \int_{\{t:\, f(t) \leq f(t_{\text{obs}})\}} f(t)\, \mathrm{d}t \quad = P(|T| > |t_{\text{obs}}|)$$

7. P-Wert mit Signifikanzniveau vergleichen, entscheiden

$$\text{pval}(t_{\text{obs}}) \leq 5\% \Rightarrow H_0 \text{ verwerfen!}$$
$$\text{pval}(t_{\text{obs}}) > 5\% \Rightarrow H_0 \text{ nicht verwerfen!}$$

Zweistichproben-t-Test

1. Modell definieren

$X_1, \ldots, X_n \sim \mathcal{N}(\mu, \sigma^2)$ alle paarweise
$Y_1, \ldots, Y_m \sim \mathcal{N}(\mu + \delta, \sigma^2)$ unabhängig

2. Null-/Gegenhypothese aufstellen

$\delta_0 = 0]$

3. Teststatistik fo

$- \delta_0$

$+ \|\mathbf{Y} - \bar{\mathbf{Y}}\|^2$

4. Verteilung der
(unter H_0, Model

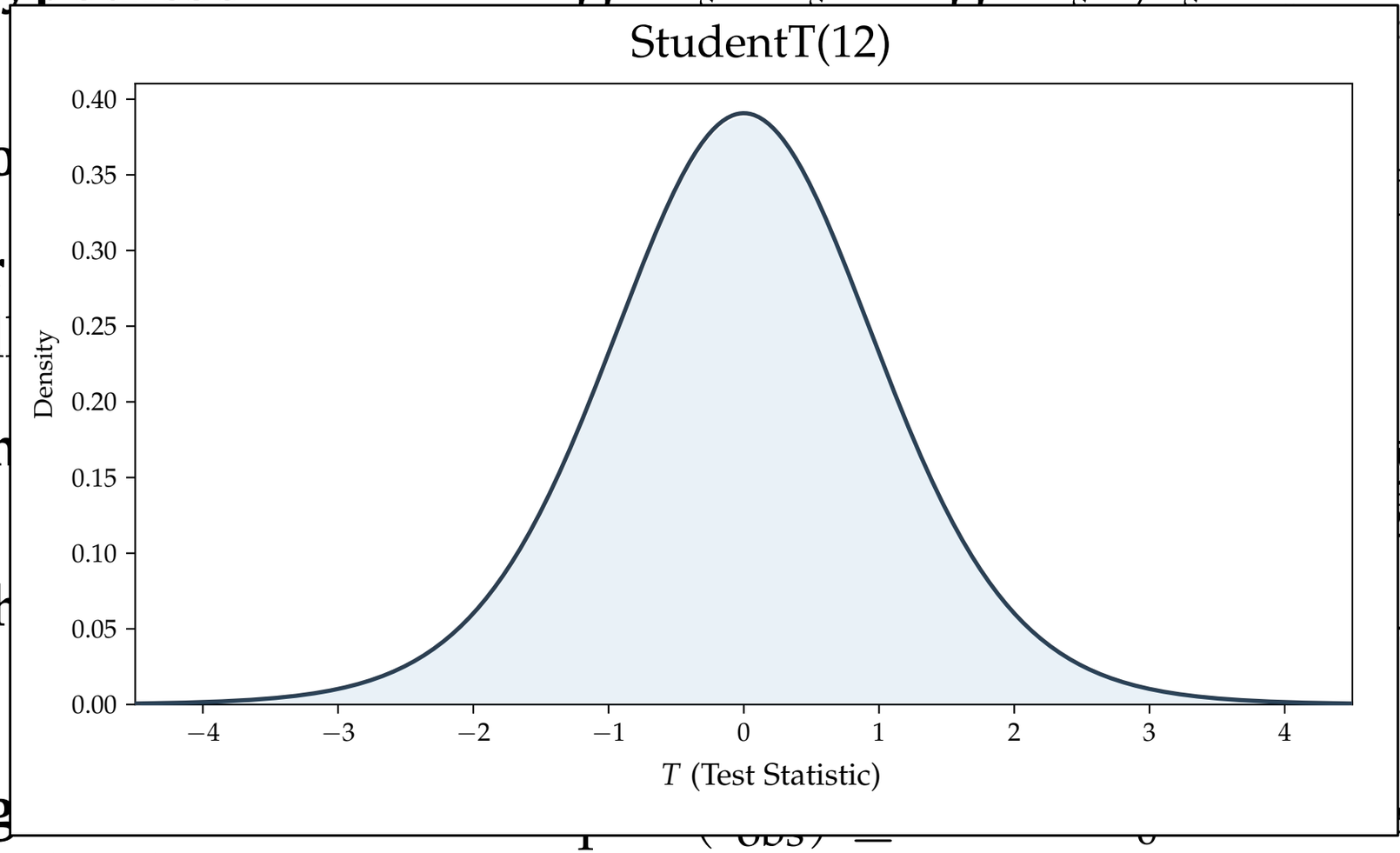
5. Daten erfassen
ausrechnen

$\bar{\mathbf{y}} - \delta_0$

$+ \|\mathbf{y} - \bar{\mathbf{y}}\|^2$

6. P-Wert ausrech

$'| > |t_{\text{obs}}|)$



7. P-Wert mit Sig n!
vergleichen, entscheiden

$\text{pval}(t_{\text{obs}}) > 5\% \Rightarrow H_0$ nicht verwerfen!

Zweistichproben-t-Test

1. Modell definieren

$X_1, \ldots, X_n \sim \mathcal{N}(\mu, \sigma^2)$ alle paarweise
$Y_1, \ldots, Y_m \sim \mathcal{N}(\mu + \delta, \sigma^2)$ unabhängig

2. Null-/Gegenhypothese aufstellen

$\delta_0 = 0]$

3. Teststatistik fo

$- \delta_0$

$+ \|\mathbf{Y} - \bar{\mathbf{Y}}\|^2$

4. Verteilung der
(unter H_0, Model

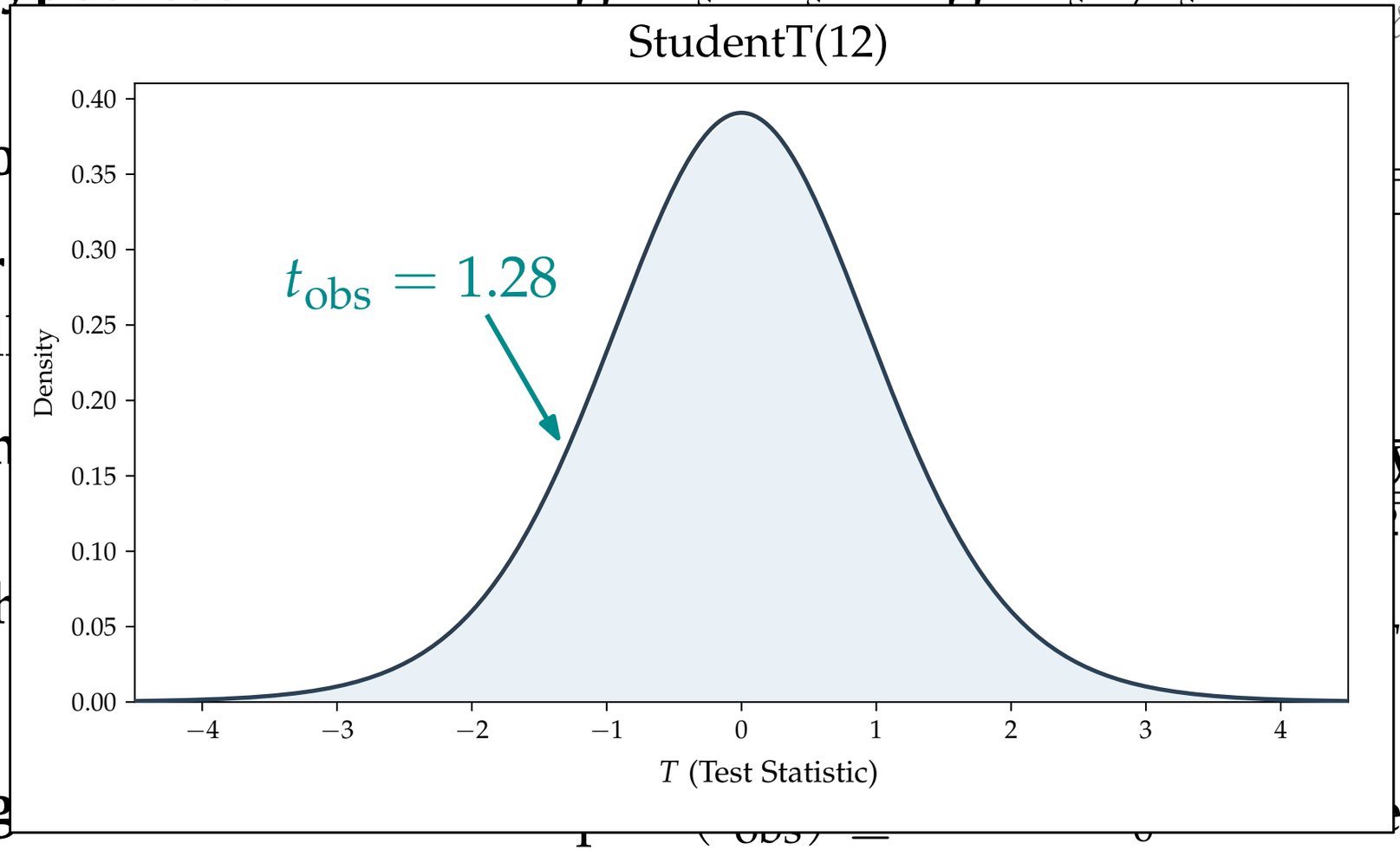
5. Daten erfassen
ausrechnen

$\bar{\mathbf{y}} - \delta_0$

$+ \|\mathbf{y} - \bar{\mathbf{y}}\|^2$

6. P-Wert ausrech

$'| > |t_{\text{obs}}|)$



7. P-Wert mit Sig
vergleichen, entscheiden

n!

$\text{pval}(t_{\text{obs}}) > 5\% \Rightarrow H_0$ nicht verwerfen!

Zweistichproben-t-Test

1. Modell definieren

$X_1, \ldots, X_n \sim \mathcal{N}(\mu, \sigma^2)$ alle paarweise
$Y_1, \ldots, Y_m \sim \mathcal{N}(\mu + \delta, \sigma^2)$ unabhängig

2. Null-/Gegenhypothese aufstellen

$\delta_0 = 0]$

3. Teststatistik fo

$$- \delta_0$$
$$+ \|\mathbf{Y} - \bar{\mathbf{Y}}\|^2$$

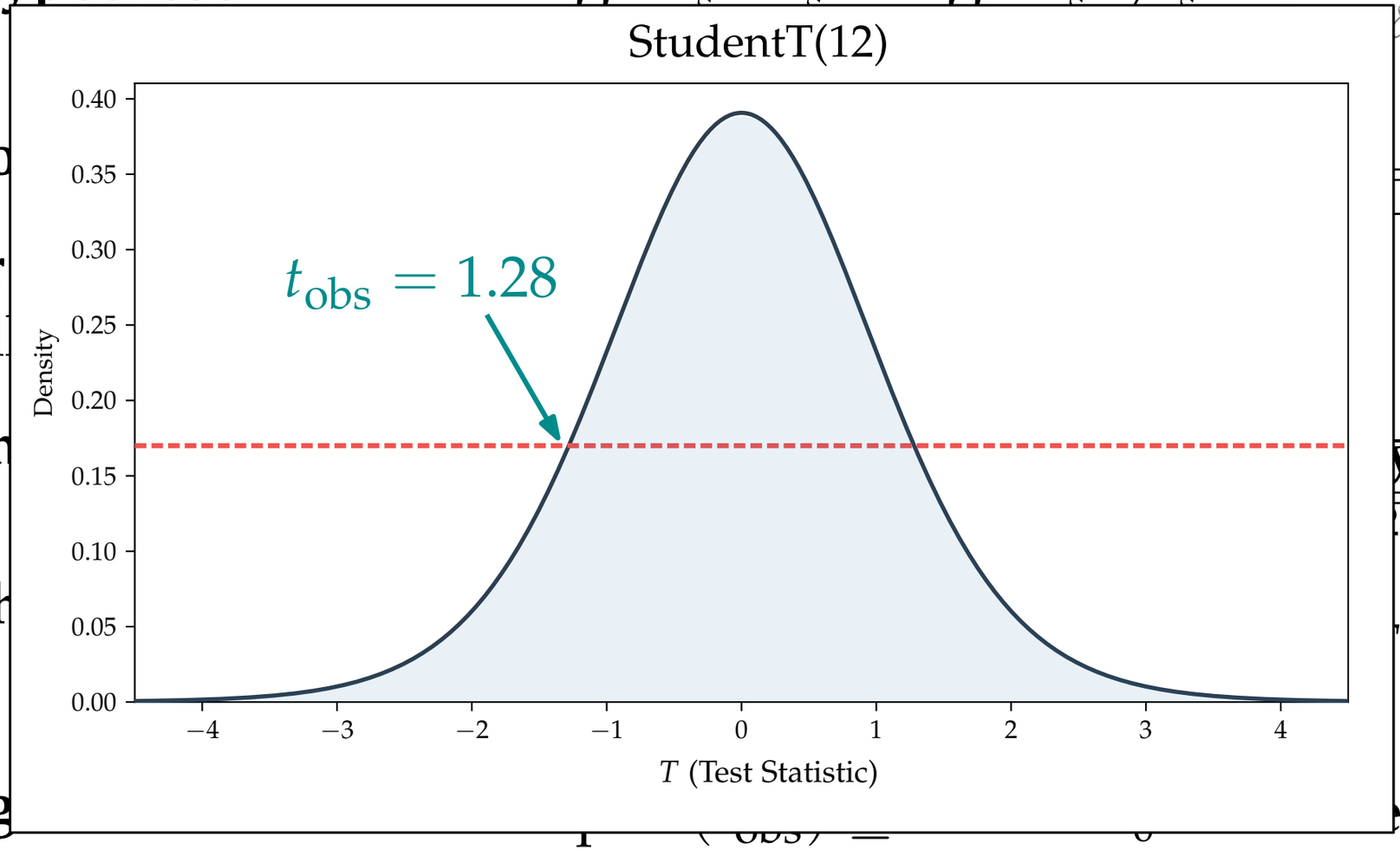
4. Verteilung der
(unter H_0, Model

5. Daten erfassen ausrechnen

$$\bar{\mathbf{y}} - \delta_0$$
$$+ \|\mathbf{y} - \bar{\mathbf{y}}\|^2$$

6. P-Wert ausrech

$$'| > |t_{\text{obs}}|)$$



7. P-Wert mit Sig vergleichen, entscheiden

$$\text{pval}(t_{\text{obs}}) > 5\% \Rightarrow H_0 \text{ nicht verwerfen!}$$

Zweistichproben-t-Test

1. Modell definieren

$X_1, \dots, X_n \sim \mathcal{N}(\mu, \sigma^2)$ alle paarweise
$Y_1, \dots, Y_m \sim \mathcal{N}(\mu + \delta, \sigma^2)$ unabhängig

2. Null-/Gegenhypothese aufstellen

$\delta_0 = 0]$

3. Teststatistik fo

$- \delta_0$

$+ \|\mathbf{Y} - \bar{\mathbf{Y}}\|^2$

4. Verteilung der
(unter H_0, Model

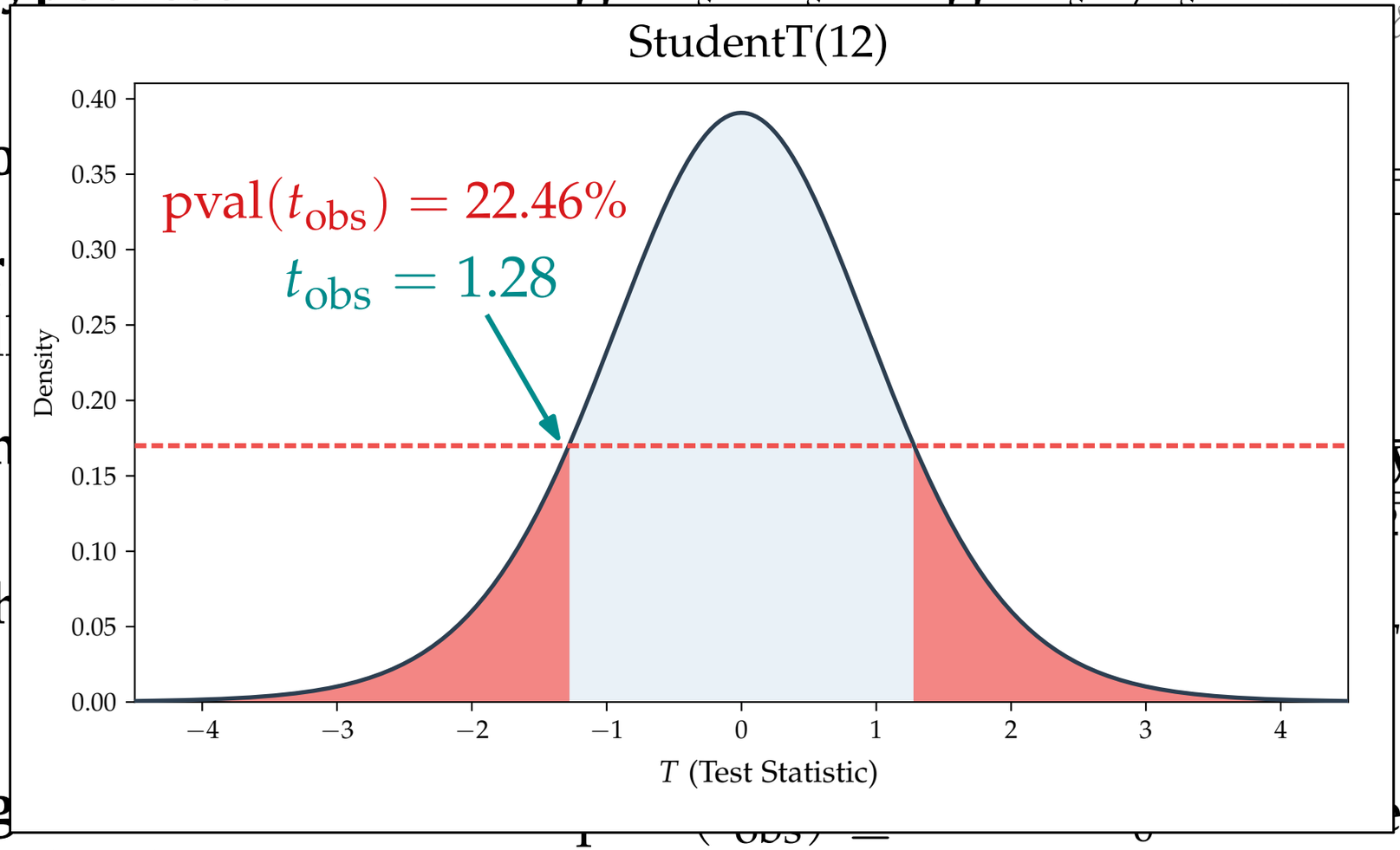
5. Daten erfassen ausrechnen

$\bar{\mathbf{y}} - \delta_0$

$+ \|\mathbf{y} - \bar{\mathbf{y}}\|^2$

6. P-Wert ausrech

$'| > |t_{\text{obs}}|)$



7. P-Wert mit Sig

n!

vergleichen, entscheiden

$\text{pval}(t_{\text{obs}}) > 5\% \Rightarrow H_0$ nicht verwerfen!

Zweistichproben-t-Test

1. Modell definieren

$$X_1, \dots, X_n \sim \mathcal{N}(\mu, \sigma^2) \quad \text{alle paarweise}$$
$$Y_1, \dots, Y_m \sim \mathcal{N}(\mu + \delta, \sigma^2) \quad \text{unabhängig}$$

2. Null-/Gegenhypothese aufstellen

$$H_0 : \delta = \delta_0, \quad H_1 : \delta \neq \delta_0 \quad [\text{hier } \delta_0 = 0]$$

3. Teststatistik formulieren

$$T = \sqrt{\frac{nm(n+m-2)}{n+m}} \frac{\bar{\mathbf{X}} - \bar{\mathbf{Y}} - \delta_0}{\sqrt{\|\mathbf{X} - \bar{\mathbf{X}}\|^2 + \|\mathbf{Y} - \bar{\mathbf{Y}}\|^2}}$$

4. Verteilung der Teststatistik bestimmen
(unter H_0, Modellannahmen)

$$T \overset{H_0}{\sim} \text{StudentT}(n+m-2)$$

5. Daten erfassen und Prüfwert ausrechnen

$$t_{\text{obs}} = \sqrt{\frac{nm(n+m-2)}{n+m}} \frac{\bar{\mathbf{x}} - \bar{\mathbf{y}} - \delta_0}{\sqrt{\|\mathbf{x} - \bar{\mathbf{x}}\|^2 + \|\mathbf{y} - \bar{\mathbf{y}}\|^2}}$$

6. P-Wert ausrechnen

$$\text{pval}(t_{\text{obs}}) = \int_{\{t:\, f(t) \leq f(t_{\text{obs}})\}} f(t)\, \mathrm{d}t \quad = P(|T| > |t_{\text{obs}}|)$$

7. P-Wert mit Signifikanzniveau vergleichen, entscheiden

$$\text{pval}(t_{\text{obs}}) \leq 5\% \Rightarrow H_0 \text{ verwerfen!}$$
$$\text{pval}(t_{\text{obs}}) > 5\% \Rightarrow H_0 \text{ nicht verwerfen!}$$

Software

Software

```
import numpy as np
import statsmodels.stats.api as sms

g1 = [20, 22, 24, 20, 22, 18, 19]
g2 = [25, 23, 20, 19, 22, 24, 22]

avg1 = np.mean(g1)
avg2 = np.mean(g2)

# descriptive statistics
print("g1 has mean {:.3f}".format(avg1))
print("g2 has mean {:.3f}".format(avg2))
print("difference {:.3f}".format(avg2 - avg1))
```

```
g1 has mean 20.714
g2 has mean 22.143
  difference 1.429
```

```
cm = sms.CompareMeans.from_data(g2, g1)

print(cm.summary())
```

```
                          Test for equality of means
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
subset #1      1.4286      1.116      1.280      0.225      -1.002       3.860
==============================================================================
```

Software

```
import numpy as np
import statsmodels.stats.api as sms

g1 = [20, 22, 24, 20, 22, 18, 19]
g2 = [25, 23, 20, 19, 22, 24, 22]

avg1 = np.mean(g1)
avg2 = np.mean(g2)

# descriptive statistics
print("g1 has mean {:.3f}".format(avg1))
print("g2 has mean {:.3f}".format(avg2))
print("difference {:.3f}".format(avg2 - avg1))
```

```
g1 has mean 20.714
g2 has mean 22.143
  difference 1.429
```

```
cm = sms.CompareMeans.from_data(g2, g1)

print(cm.summary())
```

p-value 22.5% $>$ 5% $\rightarrow$ H_0 nicht ablehnen :(

```
                          Test for equality of means
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
subset #1      1.4286      1.116      1.280      0.225      -1.002       3.860
==============================================================================
```

Normale Praxis

Viele Tests setzen Normalverteilung voraus.

Q: Woher weiß ich dass meine Daten normalverteilt sind?

Normale Praxis

Viele Tests setzen Normalverteilung voraus.

Q: Woher weiß ich dass meine Daten normalverteilt sind?

A: Es gibt spezielle Tests dafür [Kolgomorov-Smirnov, . . .]

Normale Praxis

Viele Tests setzen Normalverteilung voraus.

Q: Woher weiß ich dass meine Daten normalverteilt sind?

A: Es gibt spezielle Tests dafür [Kolgomorov-Smirnov, . . .]

Aber die benutzt niemand. [Aus gutem Grund: Multi-stage Testing erschwert Kontrolle von α; In der Praxis werden alle Test auf Normalverteilung mit genügend Daten ablehnen.]

Normale Praxis

Viele Tests setzen Normalverteilung voraus.

Q: Woher weiß ich dass meine Daten normalverteilt sind?

A: Es gibt spezielle Tests dafür [Kolgomorov-Smirnov, . . .]

Aber die benutzt niemand. [Aus gutem Grund: Multi-stage Testing erschwert Kontrolle von α; In der Praxis werden alle Test auf Normalverteilung mit genügend Daten ablehnen.]

A: Plotte Histogramm und nutze Deine Augen

Normale Praxis

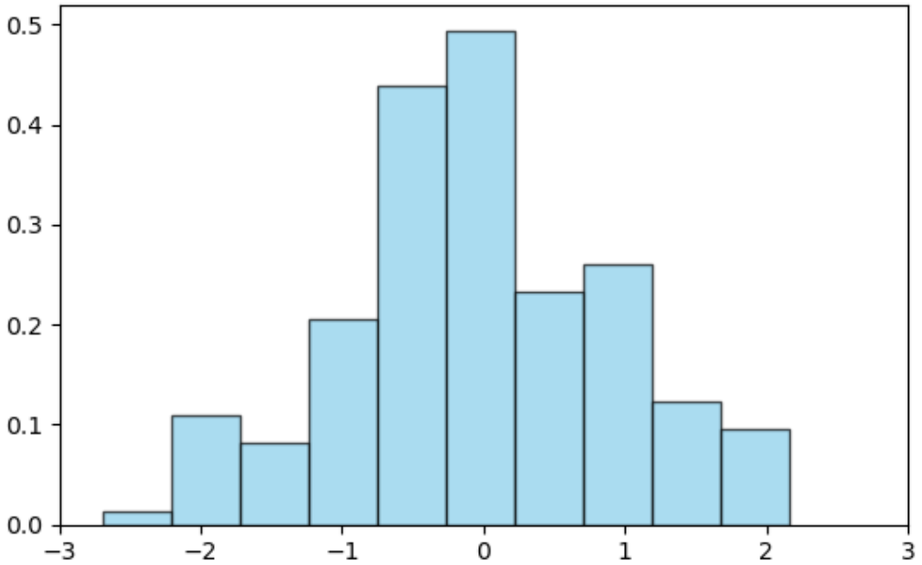
Viele Tests setzen Normalverteilung voraus.

Q: Woher weiß ich dass meine Daten normalverteilt sind?

A: Es gibt spezielle Tests dafür [Kolgomorov-Smirnov, . . .]

Aber die benutzt niemand. [Aus gutem Grund: Multi-stage Testing erschwert Kontrolle von α; In der Praxis werden alle Test auf Normalverteilung mit genügend Daten ablehnen.]

A: Plotte Histogramm und nutze Deine Augen



Normale Praxis

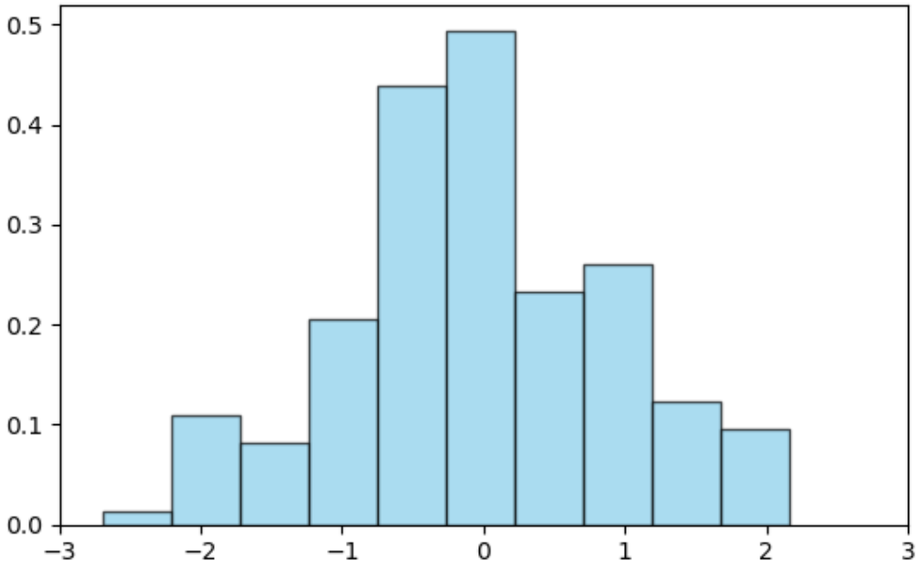
Viele Tests setzen Normalverteilung voraus.

Q: Woher weiß ich dass meine Daten normalverteilt sind?

A: Es gibt spezielle Tests dafür [Kolgomorov-Smirnov, . . .]

Aber die benutzt niemand. [Aus gutem Grund: Multi-stage Testing erschwert Kontrolle von α; In der Praxis werden alle Test auf Normalverteilung mit genügend Daten ablehnen.]

A: Plotte Histogramm und nutze Deine Augen



Normale Praxis

Viele Tests setzen Normalverteilung voraus.

Q: Woher weiß ich dass meine Daten normalverteilt sind?

A: Es gibt spezielle Tests dafür [Kolgomorov-Smirnov, . . .]

Aber die benutzt niemand. [Aus gutem Grund: Multi-stage Testing erschwert Kontrolle von α; In der Praxis werden alle Test auf Normalverteilung mit genügend Daten ablehnen.]

A: Plotte Histogramm und nutze Deine Augen

Normale Praxis

Viele Tests setzen Normalverteilung voraus.

Q: Woher weiß ich

A: Es gibt spezielle

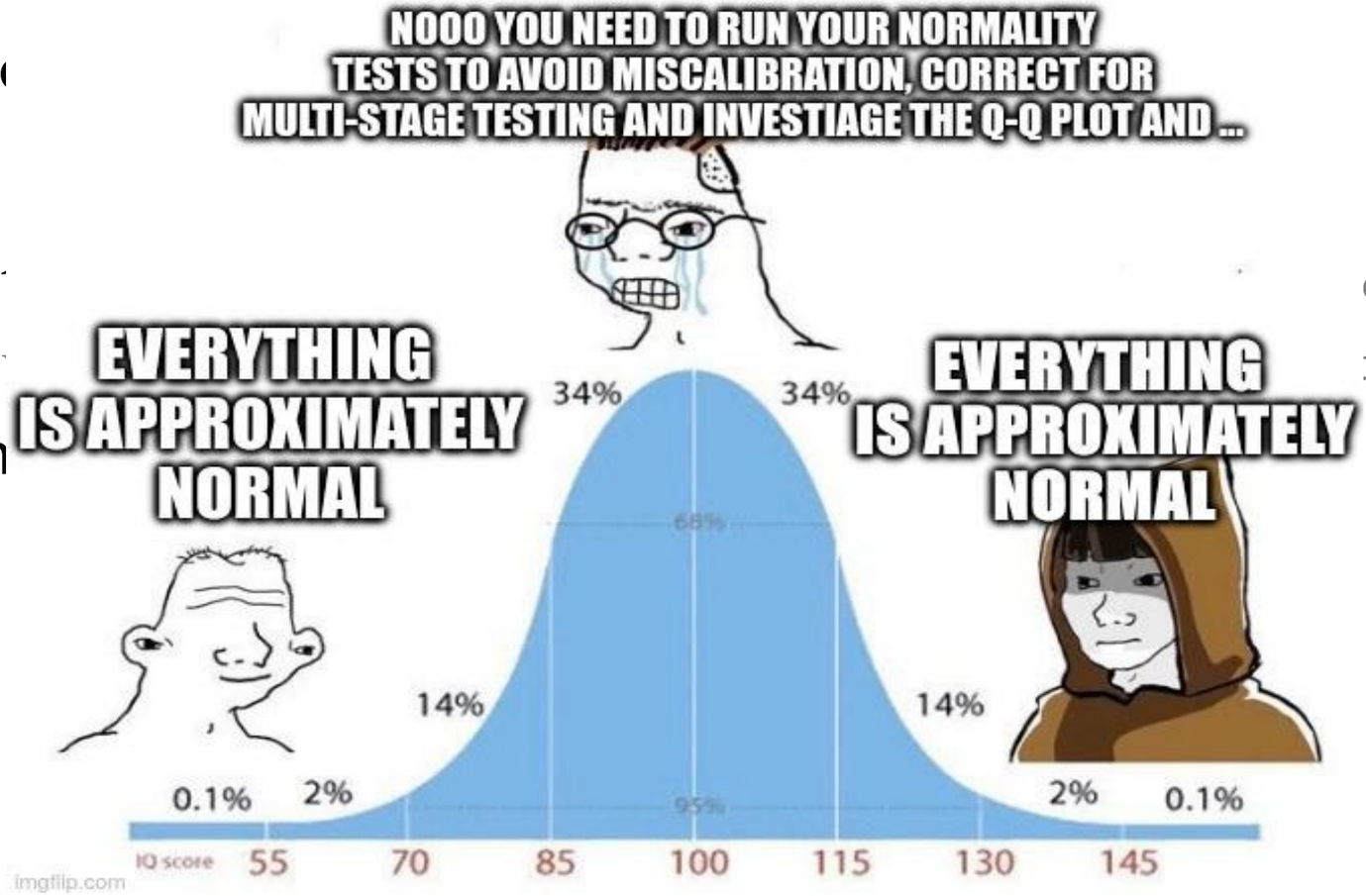
Aber die benutz

chwert Kontrolle

von α; In der Praxis

en ablehnen.]

A: Plotte Histogram



Konfidenzintervalle

Konfidenzintervalle

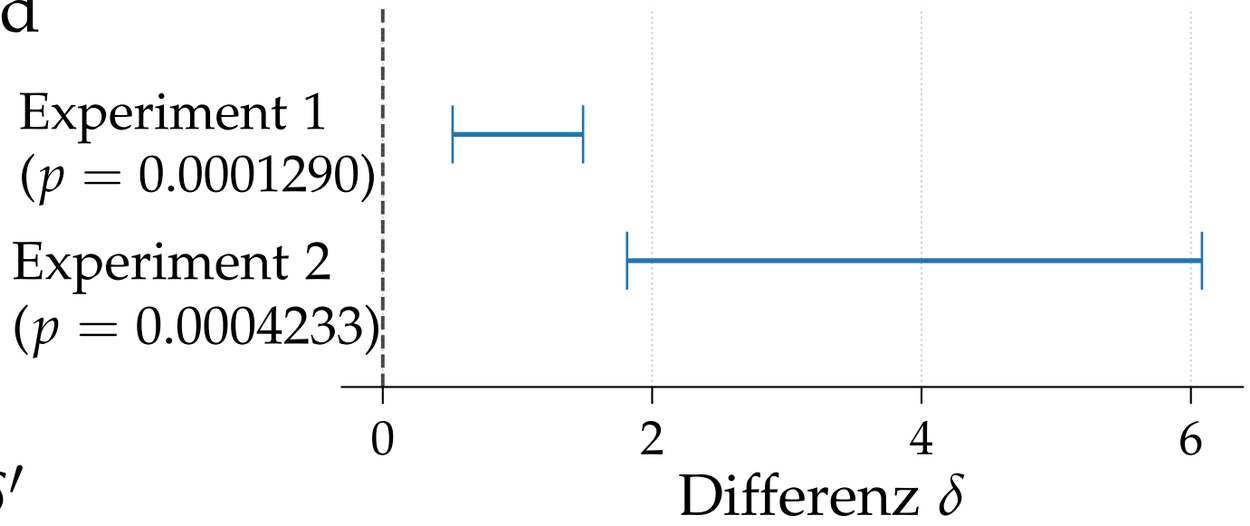
- Situation: Algorithmus A ist schneller als Algorithmus B mit $p = 0.0067$. (H_0 wird abgelehnt.)
- Frage: Heißt das
  - A ist *viel* schneller als B?
  - A ist *ziemlich sicher* schneller als B?

Konfidenzintervalle

- Situation: Algorithmus A ist schneller als Algorithmus B mit $p = 0.0067$. (H_0 wird abgelehnt.)
- Frage: Heißt das
  - A ist *viel* schneller als B?
  - A ist *ziemlich sicher* schneller als B?
- Konstruktion: Sammle alle Parameter δ' ein, für welche $H : \delta = \delta'$ *nicht* (mit $\alpha = 5\%$) verworfen wird.
- Das entstandene Intervall $[l(X), r(X)]$ überdeckt wahrscheinlich den gesuchten wahren Parameter: $P[l(X) \leq \delta \leq u(X)] = 95\%$
- H_0 ablehnen $\iff$ p-Wert $< 5\%$ $\iff$ $0 \notin [l(X), r(X)]$

Konfidenzintervalle

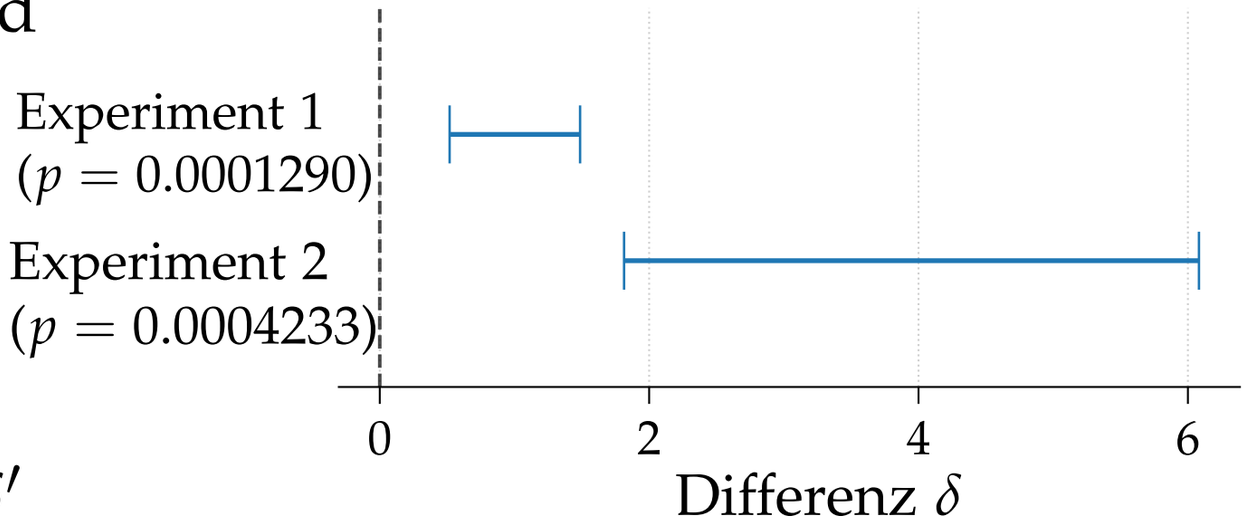
- Situation: Algorithmus A ist schneller als Algorithmus B mit $p = 0.0067$. (H_0 wird abgelehnt.)
- Frage: Heißt das
  - A ist *viel* schneller als B?
  - A ist *ziemlich sicher* schneller als B?



- Konstruktion: Sammle alle Parameter δ' ein, für welche $H : \delta = \delta'$ *nicht* (mit $\alpha = 5\%$) verworfen wird.
- Das entstandene Intervall $[l(X), r(X)]$ überdeckt wahrscheinlich den gesuchten wahren Parameter: $P[l(X) \leq \delta \leq u(X)] = 95\%$
- H_0 ablehnen $\iff$ p-Wert $< 5\%$ $\iff$ $0 \notin [l(X), r(X)]$

Konfidenzintervalle

- Situation: Algorithmus A ist schneller als Algorithmus B mit $p = 0.0067$. (H_0 wird abgelehnt.)
- Frage: Heißt das
  - A ist *viel* schneller als B?
  - A ist *ziemlich sicher* schneller als B?



- Konstruktion: Sammle alle Parameter δ' ein, für welche $H : \delta = \delta'$ *nicht* (mit $\alpha = 5\%$) verworfen wird.
- Das entstandene Intervall $[l(X), r(X)]$ überdeckt wahrscheinlich den gesuchten wahren Parameter: $P[l(X) \leq \delta \leq u(X)] = 95\%$

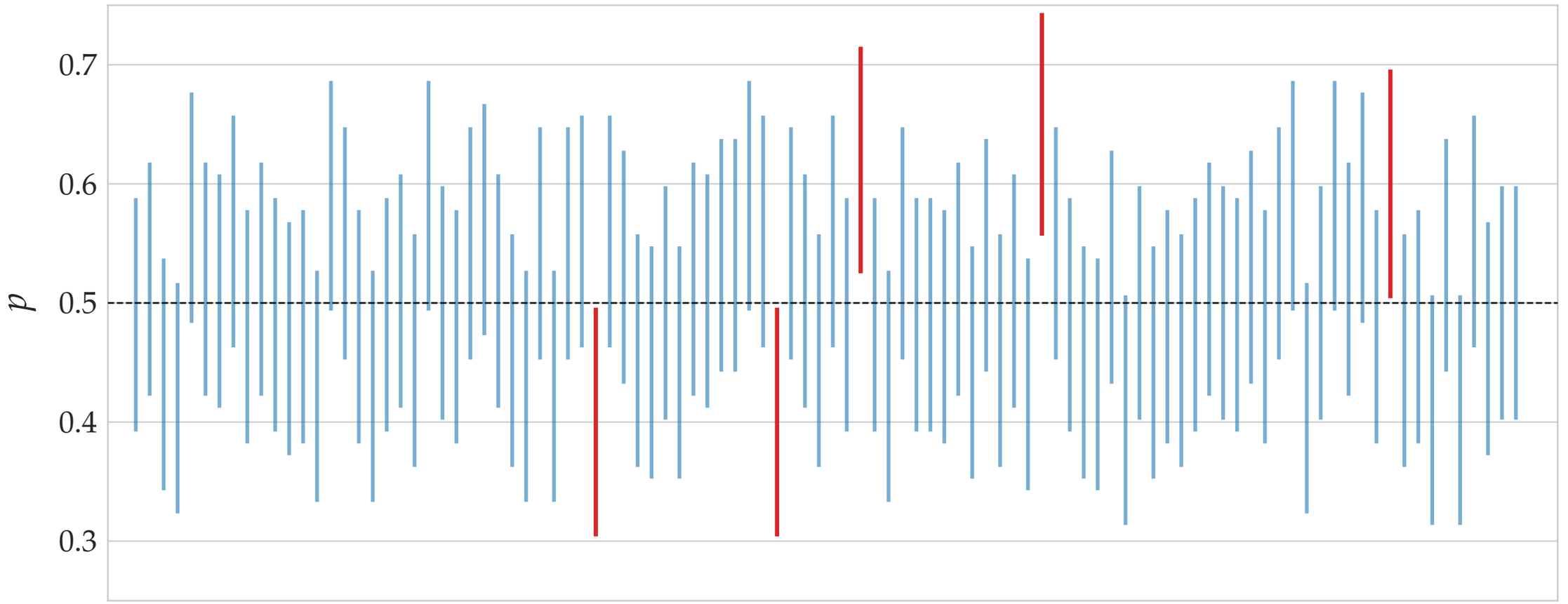
NICHT: δ liegt „wahrscheinlich“ in $[l(X), r(X)]$

- H_0 ablehnen $\iff$ p-Wert $< 5\%$ $\iff$ $0 \notin [l(X), r(X)]$

100 zufällige Durchführungen mit Konfidenzintervall

H_0 wahr: $p = 1/2$

$\rightarrow$ 5/100 Typ-1-Fehler



Praktische Hinweise

Praktische Hinweise

Praktische Hinweise

- Es existieren *nicht-parametrische Tests* die keine Voraussetzung an die Verteilung machen (aber höherer Typ-2-Fehler *(i.e. weniger Power)*!)

Praktische Hinweise

- Es existieren *nicht-parametrische Tests* die keine Voraussetzung an die Verteilung machen (aber höherer Typ-2-Fehler *(i.e. weniger Power)*!)
  z.B. Wilcoxon–Mann–Whitney test + Hodges–Lehmann estimate zur Abschätzung der Differenz des Medians von X vs Y unter einem *location shift model*

Praktische Hinweise

- Es existieren *nicht-parametrische Tests* die keine Voraussetzung an die Verteilung machen (aber höherer Typ-2-Fehler *(i.e. weniger Power)*!)
  z.B. Wilcoxon–Mann–Whitney test + Hodges–Lehmann estimate zur Abschätzung der Differenz des Medians von X vs Y unter einem *location shift model*
- Simuliert eure Tests mit RNGs!

Praktische Hinweise

- Es existieren *nicht-parametrische Tests* die keine Voraussetzung an die Verteilung machen (aber höherer Typ-2-Fehler *(i.e. weniger Power)*!)
  z.B. Wilcoxon–Mann–Whitney test + Hodges–Lehmann estimate zur Abschätzung der Differenz des Medians von X vs Y unter einem *location shift model*
- Simuliert eure Tests mit RNGs!
- Textbeispiel zum Reporting in Papers:

  We compared the runtime distributions of Algorithm A ($n_A = 50$) and Algorithm B ($n_B = 50$) across randomly generated problem instances. Median runtimes were 12.3s (IQR: 9.1–18.7) for Algorithm A and 15.8s (IQR: 11.4–22.3) for Algorithm B. A Wilcoxon–Mann–Whitney U test confirmed that this difference is statistically significant ($U = 847$, $p = 0.003$, two-sided). Under a location shift model, the Hodges–Lehmann estimator yields an estimated median difference of $\hat{\Delta} = 3.2$s (95% CI: 1.4–5.1) in favor of Algorithm A.

Drawbacks der NHST

Drawback: α-Fehler-Inflation bei vielen Tests

Table 1: Descriptive results, reported as M (SD) and p-values of the inferential statistics. Asterisks indicate significance levels as $^{*}p < .05$, $^{**}p < .01$, and $^{***}p < .001$. Square brackets indicate the values' range or unit.

	Neutral M (SD)	**Harassment** M (SD)	**Mann-Whitney U test** p-value
Sense of Embodiment (VEQ [1–7])			
Body Ownership	3.77 (1.29)	4.23 (1.33)	$p = 0.165$
Agency	5.66 (0.69)	5.92 (0.72)	$p = 0.203$
Change	3.08 (1.52)	3.84 (1.33)	$p = 0.048$ *
Self-Identification (VEQ+ [1-7])			
Self-Attribution	2.13 (1.15)	3.24 (0.99)	$p = 0.005$ **
Self-Similarity	2.80 (1.08)	3.28 (0.84)	$p = 0.225$
Self-Location	4.16 (0.98)	4.59 (1.33)	$p = 0.145$
Closeness (IOS Scale [1-7])			
	3.27 (1.31)	3.31 (1.78)	$p = 0.970$
Affect (PANAS-X [1-5])			
Positive Affect	2.63 (0.60)	2.67 (0.76)	$p = 0.804$
Negative Affect	1.30 (0.29)	1.38 (0.43)	$p = 0.817$
Joviality	2.71 (0.71)	2.69 (0.84)	$p = 0.963$
Self-Assurance	2.34 (0.64)	2.21 (0.89)	$p = 0.404$
Attentiveness	2.95 (0.62)	2.96 (0.77)	$p = 0.747$
Fear	1.34 (0.42)	1.37 (0.47)	$p = 1.000$
Sadness	1.32 (0.44)	1.40 (0.52)	$p = 0.742$
Guilt	1.29 (0.35)	1.29 (0.44)	$p = 0.712$
Hostile	1.17 (0.25)	1.31 (0.38)	$p = 0.143$
Shyness	1.61 (0.44)	1.67 (0.67)	$p = 0.926$
Fatigue	2.21 (0.91)	2.11 (0.74)	$p = 0.861$
Serenity	3.46 (0.77)	3.54 (0.87)	$p = 0.754$
Surprise	1.60 (0.66)	2.05 (0.81)	$p = 0.041$ *
State Self-Esteem (SSES [1-7])			
Social	3.24 (0.86)	3.34 (0.81)	$p = 0.633$
Performance	3.84 (0.71)	3.77 (0.82)	$p = 0.783$
Appearance	3.74 (0.80)	3.47 (0.75)	$p = 0.212$
Physiological Arousal			
RMSSD [ms]	27.29 (11.90)	30.52 (13.73)	$p = 0.402$
Movement Analysis			
Retreat Distance [m]	0.04 (0.04)	0.20 (0.12)	$p < 0.001$ ***
Control Measures and Manipulation Checks			
Similarity Rating [1 - 7]	3.12 (1.21)	3.38 (1.53)	$p = 0.448$
Spatial Presence [0 - 6]	3.78 (0.64)	3.93 (0.67)	$p = 0.348$
Humanness [-3 - 3]	-0.30 (1.33)	-0.21 (0.90)	$p = 0.927$
Impression [1 - 7]	4.58 (1.53)	2.43 (0.79)	$p < 0.001$ ***

Drawback: α-Fehler-Inflation bei vielen Tests

- 30 Tests wurden durchgeführt.
- 5 Tests sind signifikant ($\alpha = 5\,\%$).

Table 1: Descriptive results, reported as M (SD) and p-values of the inferential statistics. Asterisks indicate significance levels as $^{*}p < .05$, $^{**}p < .01$, and $^{***}p < .001$. Square brackets indicate the values' range or unit.

	Neutral M (SD)	**Harassment** M (SD)	**Mann-Whitney U test** p-value
Sense of Embodiment (VEQ [1–7])			
Body Ownership	3.77 (1.29)	4.23 (1.33)	$p = 0.165$
Agency	5.66 (0.69)	5.92 (0.72)	$p = 0.203$
Change	3.08 (1.52)	3.84 (1.33)	$p = 0.048$ *
Self-Identification (VEQ+ [1-7])			
Self-Attribution	2.13 (1.15)	3.24 (0.99)	$p = 0.005$ **
Self-Similarity	2.80 (1.08)	3.28 (0.84)	$p = 0.225$
Self-Location	4.16 (0.98)	4.59 (1.33)	$p = 0.145$
Closeness (IOS Scale [1-7])			
	3.27 (1.31)	3.31 (1.78)	$p = 0.970$
Affect (PANAS-X [1-5])			
Positive Affect	2.63 (0.60)	2.67 (0.76)	$p = 0.804$
Negative Affect	1.30 (0.29)	1.38 (0.43)	$p = 0.817$
Joviality	2.71 (0.71)	2.69 (0.84)	$p = 0.963$
Self-Assurance	2.34 (0.64)	2.21 (0.89)	$p = 0.404$
Attentiveness	2.95 (0.62)	2.96 (0.77)	$p = 0.747$
Fear	1.34 (0.42)	1.37 (0.47)	$p = 1.000$
Sadness	1.32 (0.44)	1.40 (0.52)	$p = 0.742$
Guilt	1.29 (0.35)	1.29 (0.44)	$p = 0.712$
Hostile	1.17 (0.25)	1.31 (0.38)	$p = 0.143$
Shyness	1.61 (0.44)	1.67 (0.67)	$p = 0.926$
Fatigue	2.21 (0.91)	2.11 (0.74)	$p = 0.861$
Serenity	3.46 (0.77)	3.54 (0.87)	$p = 0.754$
Surprise	1.60 (0.66)	2.05 (0.81)	$p = 0.041$ *
State Self-Esteem (SSES [1-7])			
Social	3.24 (0.86)	3.34 (0.81)	$p = 0.633$
Performance	3.84 (0.71)	3.77 (0.82)	$p = 0.783$
Appearance	3.74 (0.80)	3.47 (0.75)	$p = 0.212$
Physiological Arousal			
RMSSD [ms]	27.29 (11.90)	30.52 (13.73)	$p = 0.402$
Movement Analysis			
Retreat Distance [m]	0.04 (0.04)	0.20 (0.12)	$p < 0.001$ ***
Control Measures and Manipulation Checks			
Similarity Rating [1 - 7]	3.12 (1.21)	3.38 (1.53)	$p = 0.448$
Spatial Presence [0 - 6]	3.78 (0.64)	3.93 (0.67)	$p = 0.348$
Humanness [-3 - 3]	-0.30 (1.33)	-0.21 (0.90)	$p = 0.927$
Impression [1 - 7]	4.58 (1.53)	2.43 (0.79)	$p < 0.001$ ***

Drawback: α-Fehler-Inflation bei vielen Tests

- 30 Tests wurden durchgeführt.
- 5 Tests sind signifikant ($\alpha = 5\,\%$).
- Unter H_0 „gar kein Effekt“ werden $30 \cdot 5\,\% \approx 2$ Tests fehlerhaft ablehnen.

Table 1: Descriptive results, reported as M (SD) and p-values of the inferential statistics. Asterisks indicate significance levels as $^{*}p < .05$, $^{**}p < .01$, and $^{***}p < .001$. Square brackets indicate the values' range or unit.

	Neutral M (SD)	**Harassment** M (SD)	**Mann-Whitney U test** p-value
Sense of Embodiment (VEQ [1–7])			
Body Ownership	3.77 (1.29)	4.23 (1.33)	$p = 0.165$
Agency	5.66 (0.69)	5.92 (0.72)	$p = 0.203$
Change	3.08 (1.52)	3.84 (1.33)	$p = 0.048$ *
Self-Identification (VEQ+ [1-7])			
Self-Attribution	2.13 (1.15)	3.24 (0.99)	$p = 0.005$ **
Self-Similarity	2.80 (1.08)	3.28 (0.84)	$p = 0.225$
Self-Location	4.16 (0.98)	4.59 (1.33)	$p = 0.145$
Closeness (IOS Scale [1-7])			
	3.27 (1.31)	3.31 (1.78)	$p = 0.970$
Affect (PANAS-X [1-5])			
Positive Affect	2.63 (0.60)	2.67 (0.76)	$p = 0.804$
Negative Affect	1.30 (0.29)	1.38 (0.43)	$p = 0.817$
Joviality	2.71 (0.71)	2.69 (0.84)	$p = 0.963$
Self-Assurance	2.34 (0.64)	2.21 (0.89)	$p = 0.404$
Attentiveness	2.95 (0.62)	2.96 (0.77)	$p = 0.747$
Fear	1.34 (0.42)	1.37 (0.47)	$p = 1.000$
Sadness	1.32 (0.44)	1.40 (0.52)	$p = 0.742$
Guilt	1.29 (0.35)	1.29 (0.44)	$p = 0.712$
Hostile	1.17 (0.25)	1.31 (0.38)	$p = 0.143$
Shyness	1.61 (0.44)	1.67 (0.67)	$p = 0.926$
Fatigue	2.21 (0.91)	2.11 (0.74)	$p = 0.861$
Serenity	3.46 (0.77)	3.54 (0.87)	$p = 0.754$
Surprise	1.60 (0.66)	2.05 (0.81)	$p = 0.041$ *
State Self-Esteem (SSES [1-7])			
Social	3.24 (0.86)	3.34 (0.81)	$p = 0.633$
Performance	3.84 (0.71)	3.77 (0.82)	$p = 0.783$
Appearance	3.74 (0.80)	3.47 (0.75)	$p = 0.212$
Physiological Arousal			
RMSSD [ms]	27.29 (11.90)	30.52 (13.73)	$p = 0.402$
Movement Analysis			
Retreat Distance [m]	0.04 (0.04)	0.20 (0.12)	$p < 0.001$ ***
Control Measures and Manipulation Checks			
Similarity Rating [1 - 7]	3.12 (1.21)	3.38 (1.53)	$p = 0.448$
Spatial Presence [0 - 6]	3.78 (0.64)	3.93 (0.67)	$p = 0.348$
Humanness [-3 - 3]	-0.30 (1.33)	-0.21 (0.90)	$p = 0.927$
Impression [1 - 7]	4.58 (1.53)	2.43 (0.79)	$p < 0.001$ ***

Drawback: α-Fehler-Inflation bei vielen Tests

- 30 Tests wurden durchgeführt.
- 5 Tests sind signifikant ($\alpha = 5\,\%$).
- Unter H_0 „gar kein Effekt“ werden $30 \cdot 5\,\% \approx 2$ Tests fehlerhaft ablehnen.

Abhilfe: Bonferroni-Korrektur

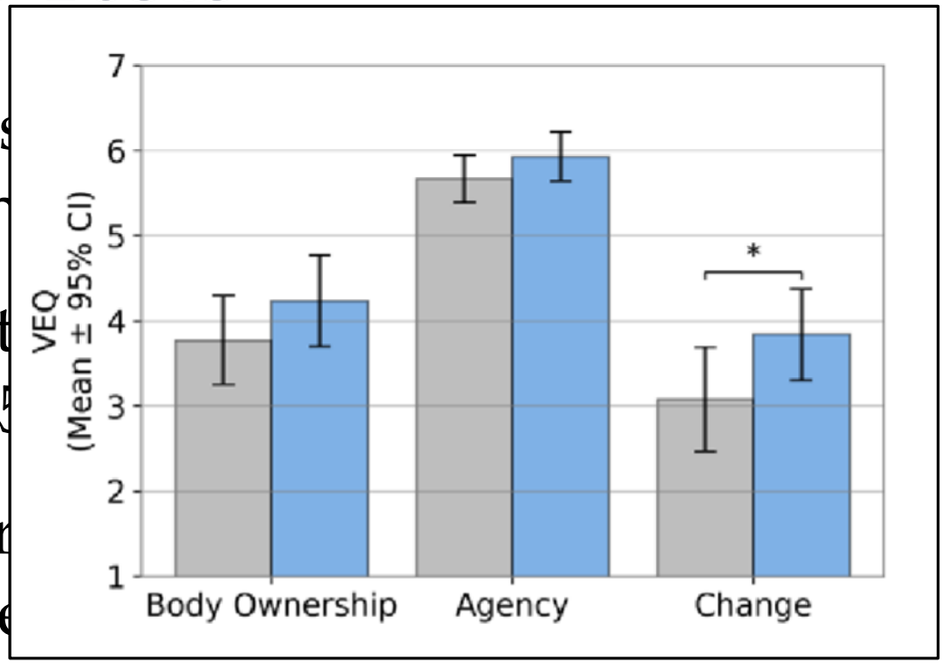
Effektives Signifikanzniveau auf $\alpha' = \alpha/30$ absenken

Table 1: Descriptive results, reported as M (SD) and p-values of the inferential statistics. Asterisks indicate significance levels as $^{*}p < .05$, $^{**}p < .01$, and $^{***}p < .001$. Square brackets indicate the values' range or unit.

	Neutral M (SD)	**Harassment** M (SD)	**Mann-Whitney U test** p-value
Sense of Embodiment (VEQ [1–7])			
Body Ownership	3.77 (1.29)	4.23 (1.33)	$p = 0.165$
Agency	5.66 (0.69)	5.92 (0.72)	$p = 0.203$
Change	3.08 (1.52)	3.84 (1.33)	$p = 0.048$ *
Self-Identification (VEQ+ [1-7])			
Self-Attribution	2.13 (1.15)	3.24 (0.99)	$p = 0.005$ **
Self-Similarity	2.80 (1.08)	3.28 (0.84)	$p = 0.225$
Self-Location	4.16 (0.98)	4.59 (1.33)	$p = 0.145$
Closeness (IOS Scale [1-7])			
	3.27 (1.31)	3.31 (1.78)	$p = 0.970$
Affect (PANAS-X [1-5])			
Positive Affect	2.63 (0.60)	2.67 (0.76)	$p = 0.804$
Negative Affect	1.30 (0.29)	1.38 (0.43)	$p = 0.817$
Joviality	2.71 (0.71)	2.69 (0.84)	$p = 0.963$
Self-Assurance	2.34 (0.64)	2.21 (0.89)	$p = 0.404$
Attentiveness	2.95 (0.62)	2.96 (0.77)	$p = 0.747$
Fear	1.34 (0.42)	1.37 (0.47)	$p = 1.000$
Sadness	1.32 (0.44)	1.40 (0.52)	$p = 0.742$
Guilt	1.29 (0.35)	1.29 (0.44)	$p = 0.712$
Hostile	1.17 (0.25)	1.31 (0.38)	$p = 0.143$
Shyness	1.61 (0.44)	1.67 (0.67)	$p = 0.926$
Fatigue	2.21 (0.91)	2.11 (0.74)	$p = 0.861$
Serenity	3.46 (0.77)	3.54 (0.87)	$p = 0.754$
Surprise	1.60 (0.66)	2.05 (0.81)	$p = 0.041$ *
State Self-Esteem (SSES [1-7])			
Social	3.24 (0.86)	3.34 (0.81)	$p = 0.633$
Performance	3.84 (0.71)	3.77 (0.82)	$p = 0.783$
Appearance	3.74 (0.80)	3.47 (0.75)	$p = 0.212$
Physiological Arousal			
RMSSD [ms]	27.29 (11.90)	30.52 (13.73)	$p = 0.402$
Movement Analysis			
Retreat Distance [m]	0.04 (0.04)	0.20 (0.12)	$p < 0.001$ ***
Control Measures and Manipulation Checks			
Similarity Rating [1 - 7]	3.12 (1.21)	3.38 (1.53)	$p = 0.448$
Spatial Presence [0 - 6]	3.78 (0.64)	3.93 (0.67)	$p = 0.348$
Humanness [-3 - 3]	-0.30 (1.33)	-0.21 (0.90)	$p = 0.927$
Impression [1 - 7]	4.58 (1.53)	2.43 (0.79)	$p < 0.001$ ***

Drawback: α-Fehler-Inflation bei vielen Tests

- 30 Tes
  durch
- 5 Test
  ($\alpha = 5$
- Unter
  werde
  fehlerhaft ablehnen.



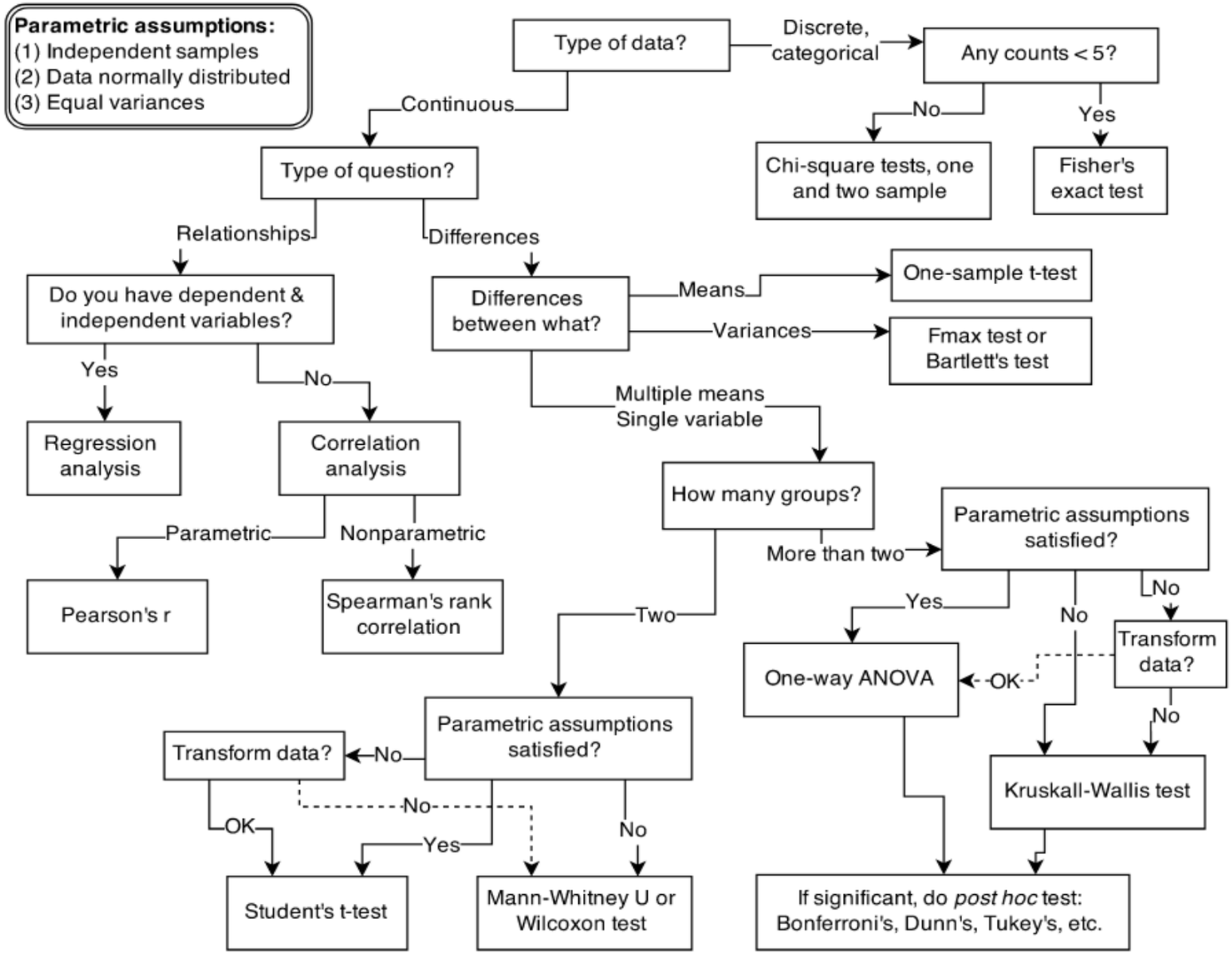
Abhilfe: Bonferroni-Korrektur

Effektives Signifikanzniveau auf $\alpha' = \alpha/30$ absenken

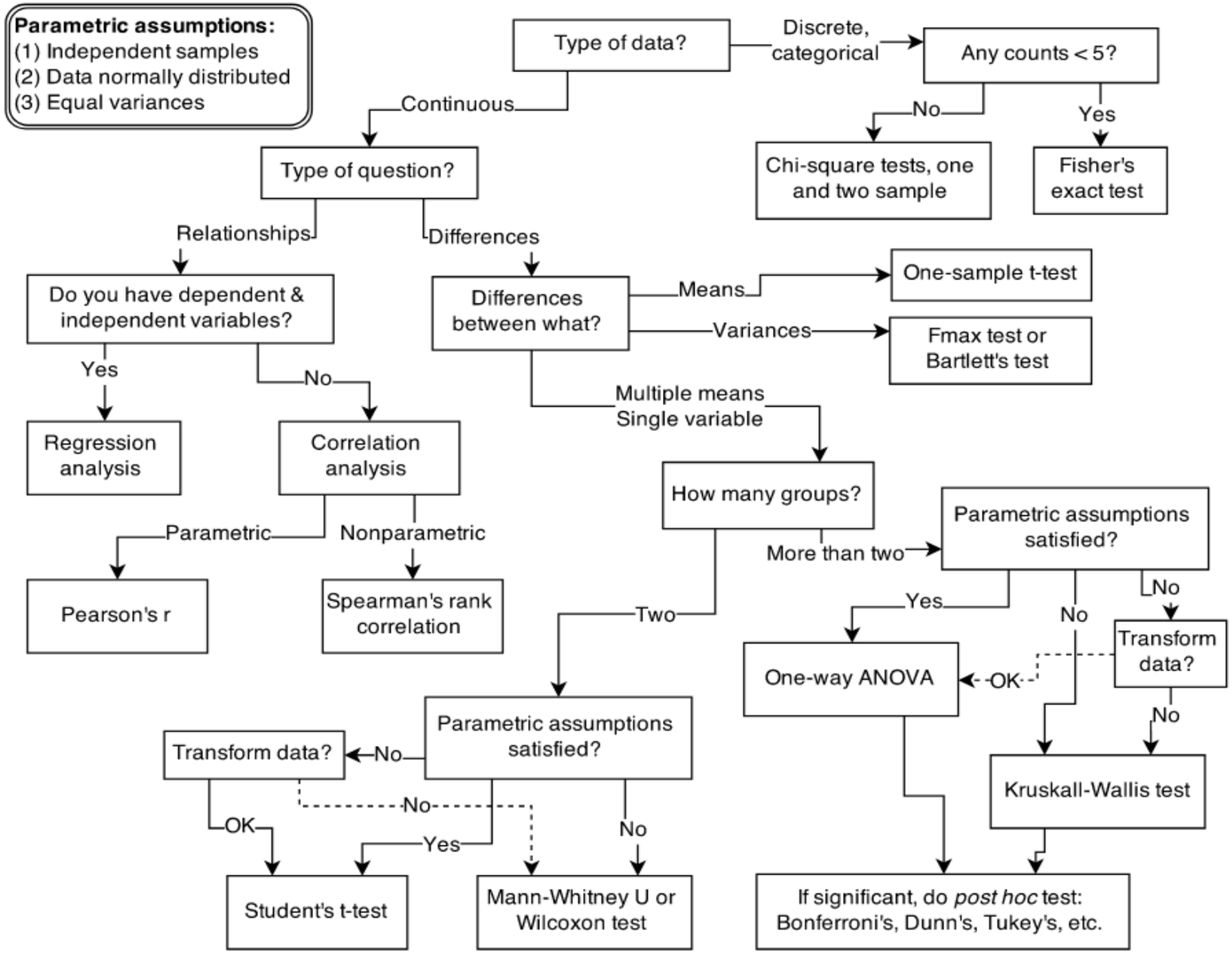
Table 1: Descriptive results, reported as M (SD) and p-values of the inferential statistics. Asterisks indicate significance levels as $^{*}p < .05$, $^{**}p < .01$, and $^{***}p < .001$. Square brackets indicate the values' range or unit.

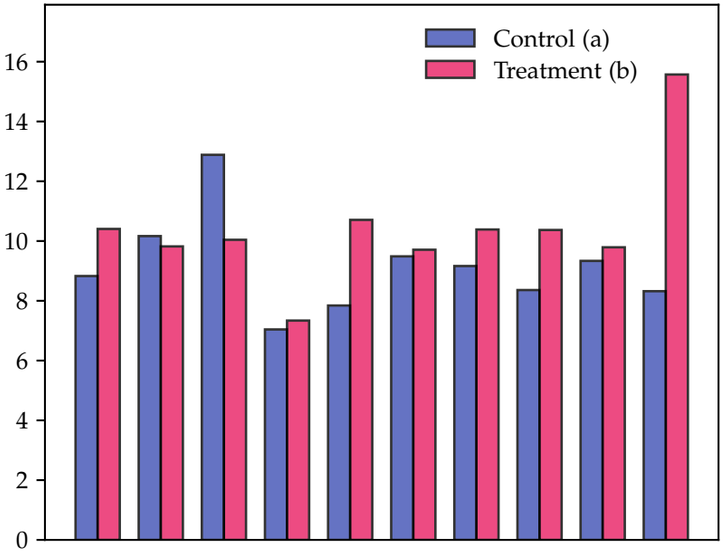
	Neutral M (SD)	**Harassment** M (SD)	**Mann-Whitney U test** p-value
Sense of Embodiment (VEQ [1–7])			
Body Ownership	3.77 (1.29)	4.23 (1.33)	$p = 0.165$
Agency	5.66 (0.69)	5.92 (0.72)	$p = 0.203$
Change	3.08 (1.52)	3.84 (1.33)	$p = 0.048$ *
Self-Identification (VEQ+ [1-7])			
Self-Attribution	2.13 (1.15)	3.24 (0.99)	$p = 0.005$ **
Self-Similarity	2.80 (1.08)	3.28 (0.84)	$p = 0.225$
Self-Location	4.16 (0.98)	4.59 (1.33)	$p = 0.145$
Closeness (IOS Scale [1-7])			
	3.27 (1.31)	3.31 (1.78)	$p = 0.970$
Affect (PANAS-X [1-5])			
Positive Affect	2.63 (0.60)	2.67 (0.76)	$p = 0.804$
Negative Affect	1.30 (0.29)	1.38 (0.43)	$p = 0.817$
Joviality	2.71 (0.71)	2.69 (0.84)	$p = 0.963$
Self-Assurance	2.34 (0.64)	2.21 (0.89)	$p = 0.404$
Attentiveness	2.95 (0.62)	2.96 (0.77)	$p = 0.747$
Fear	1.34 (0.42)	1.37 (0.47)	$p = 1.000$
Sadness	1.32 (0.44)	1.40 (0.52)	$p = 0.742$
Guilt	1.29 (0.35)	1.29 (0.44)	$p = 0.712$
Hostile	1.17 (0.25)	1.31 (0.38)	$p = 0.143$
Shyness	1.61 (0.44)	1.67 (0.67)	$p = 0.926$
Fatigue	2.21 (0.91)	2.11 (0.74)	$p = 0.861$
Serenity	3.46 (0.77)	3.54 (0.87)	$p = 0.754$
Surprise	1.60 (0.66)	2.05 (0.81)	$p = 0.041$ *
State Self-Esteem (SSES [1-7])			
Social	3.24 (0.86)	3.34 (0.81)	$p = 0.633$
Performance	3.84 (0.71)	3.77 (0.82)	$p = 0.783$
Appearance	3.74 (0.80)	3.47 (0.75)	$p = 0.212$
Physiological Arousal			
RMSSD [ms]	27.29 (11.90)	30.52 (13.73)	$p = 0.402$
Movement Analysis			
Retreat Distance [m]	0.04 (0.04)	0.20 (0.12)	$p < 0.001$ ***
Control Measures and Manipulation Checks			
Similarity Rating [1 - 7]	3.12 (1.21)	3.38 (1.53)	$p = 0.448$
Spatial Presence [0 - 6]	3.78 (0.64)	3.93 (0.67)	$p = 0.348$
Humanness [-3 - 3]	-0.30 (1.33)	-0.21 (0.90)	$p = 0.927$
Impression [1 - 7]	4.58 (1.53)	2.43 (0.79)	$p < 0.001$ ***

Drawback: Soo viele Tests

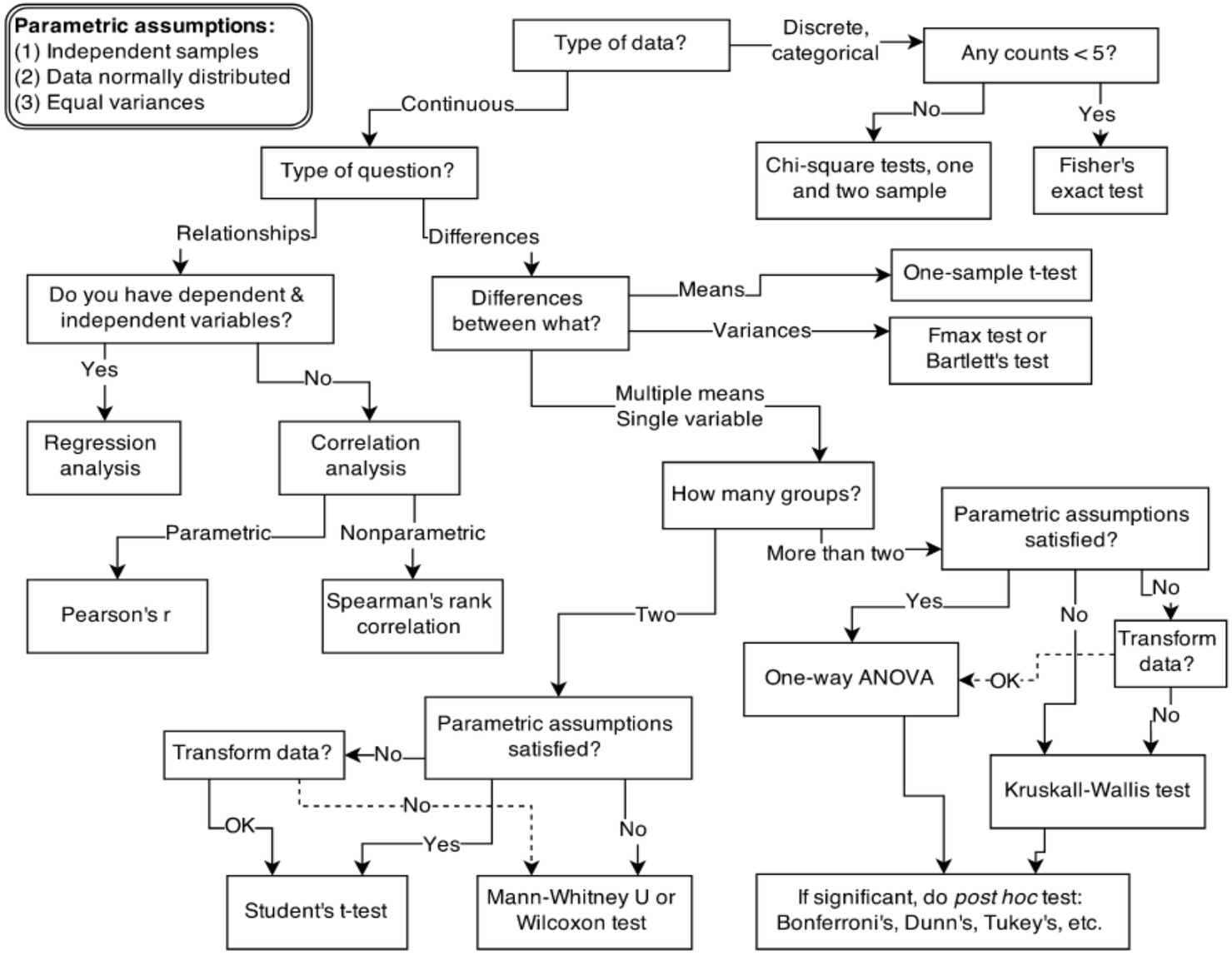


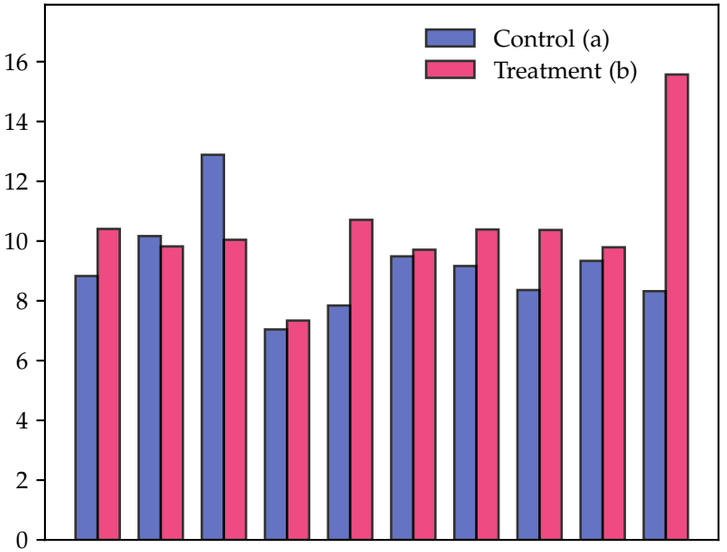
Drawback: Soo viele Tests





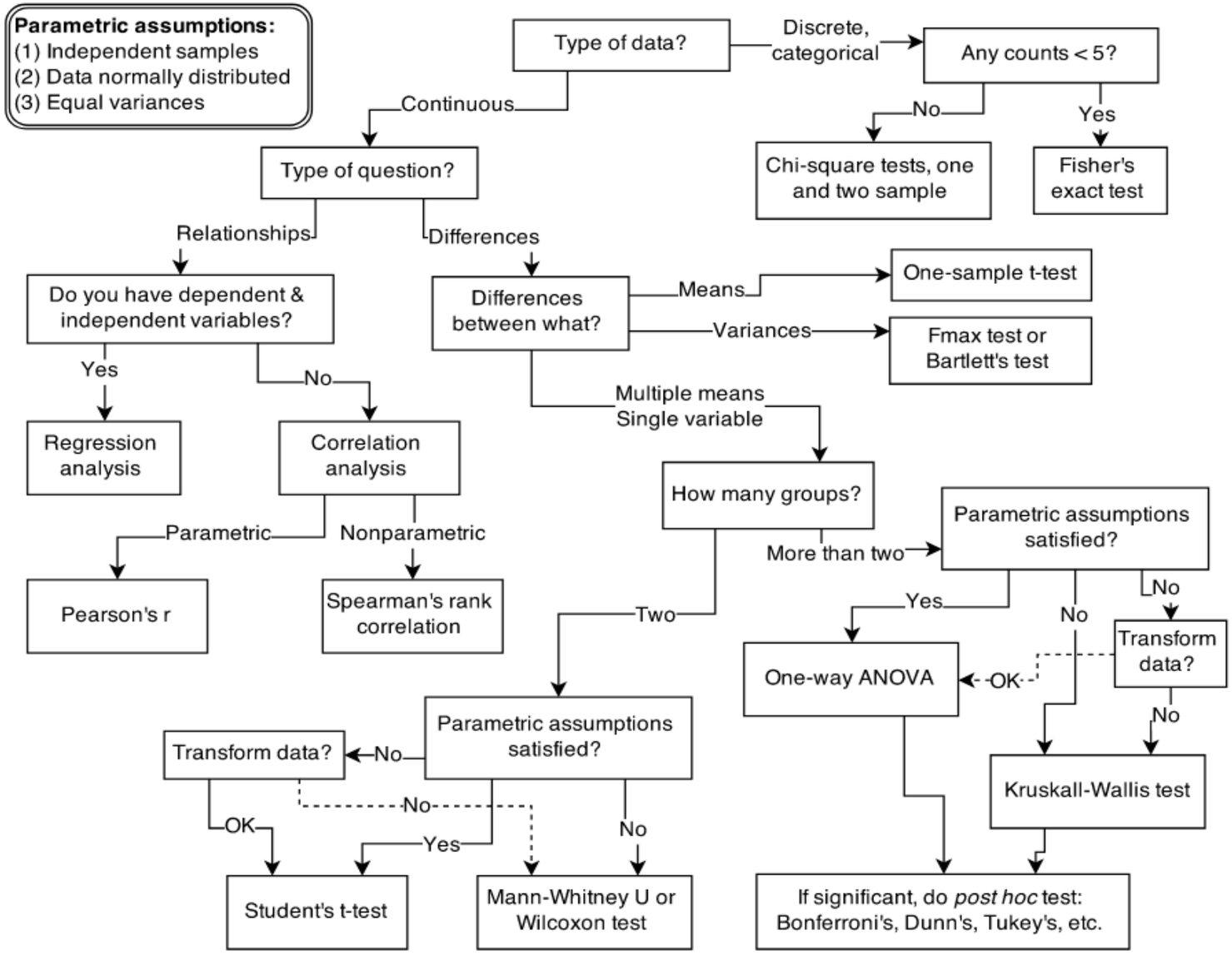
Drawback: Soo viele Tests

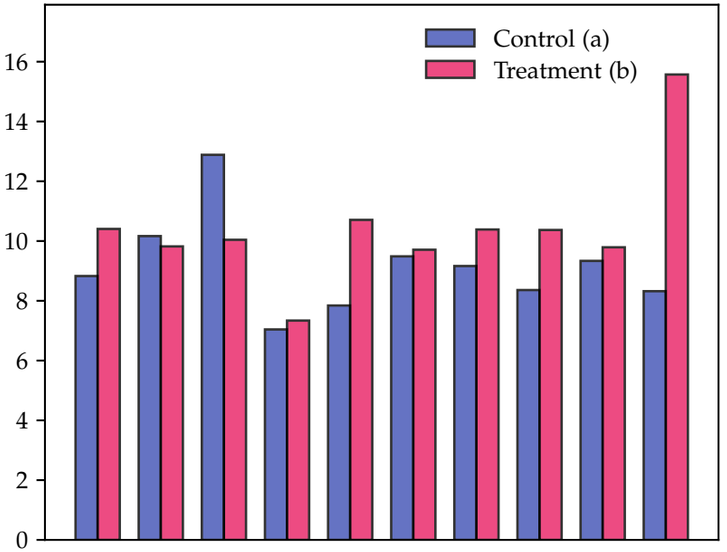




Paired t-test $p = 0.15765$

Drawback: Soo viele Tests

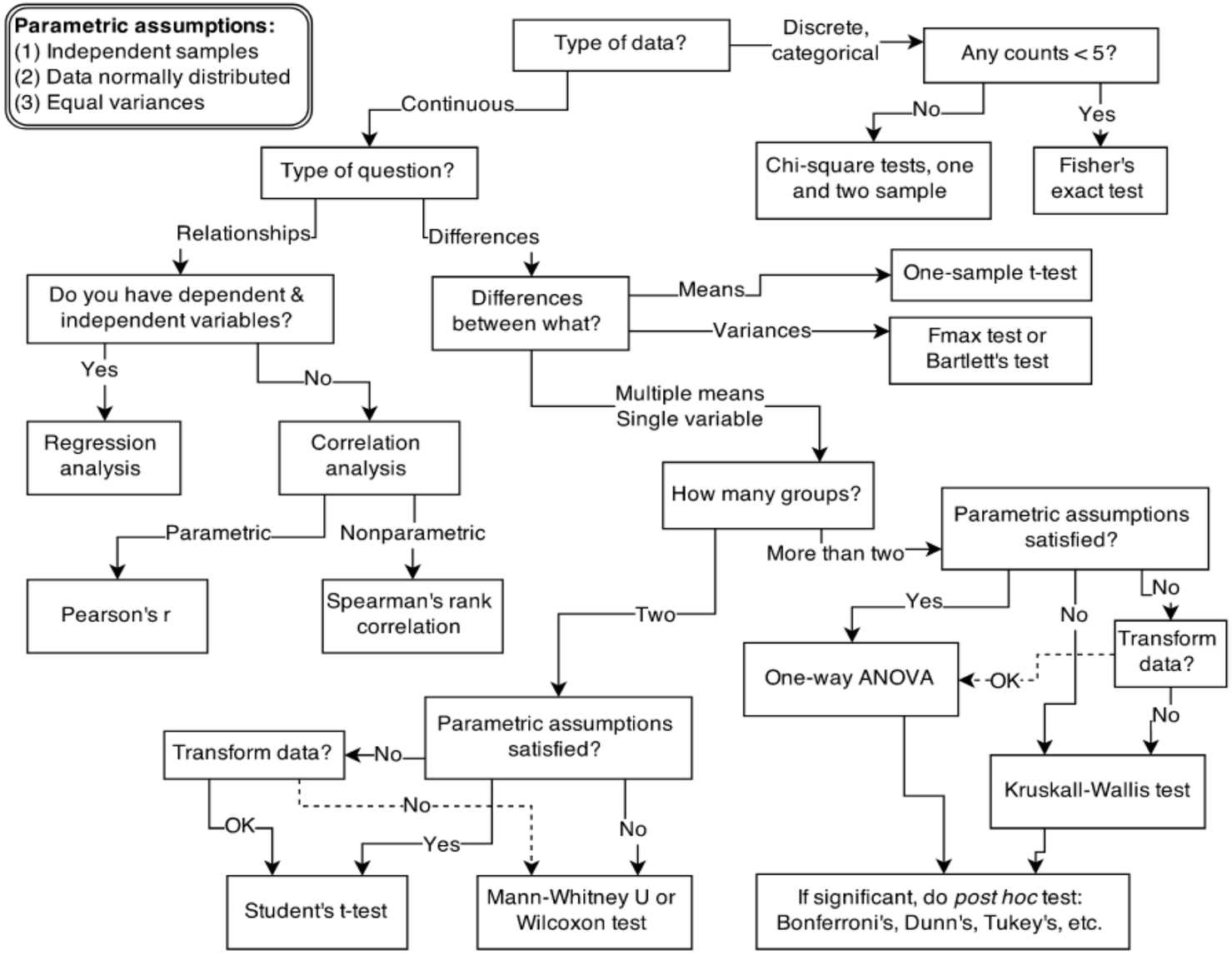


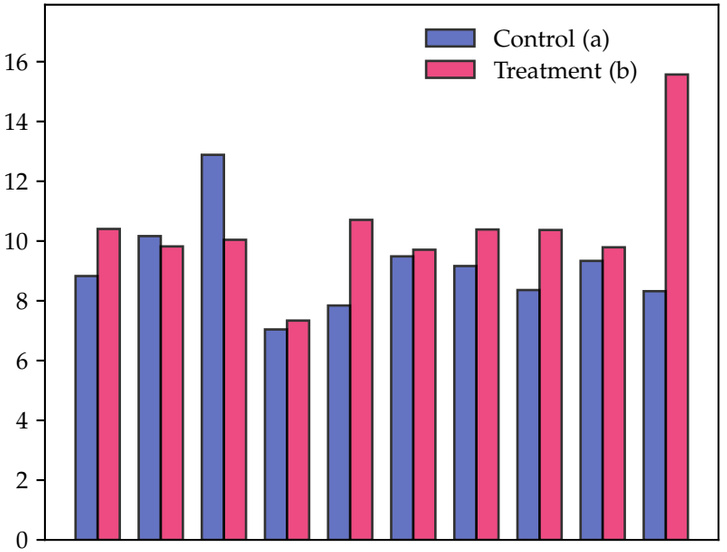


Paired t-test $p = 0.15765$

Welch's t-test $p = 0.13874$

Drawback: Soo viele Tests



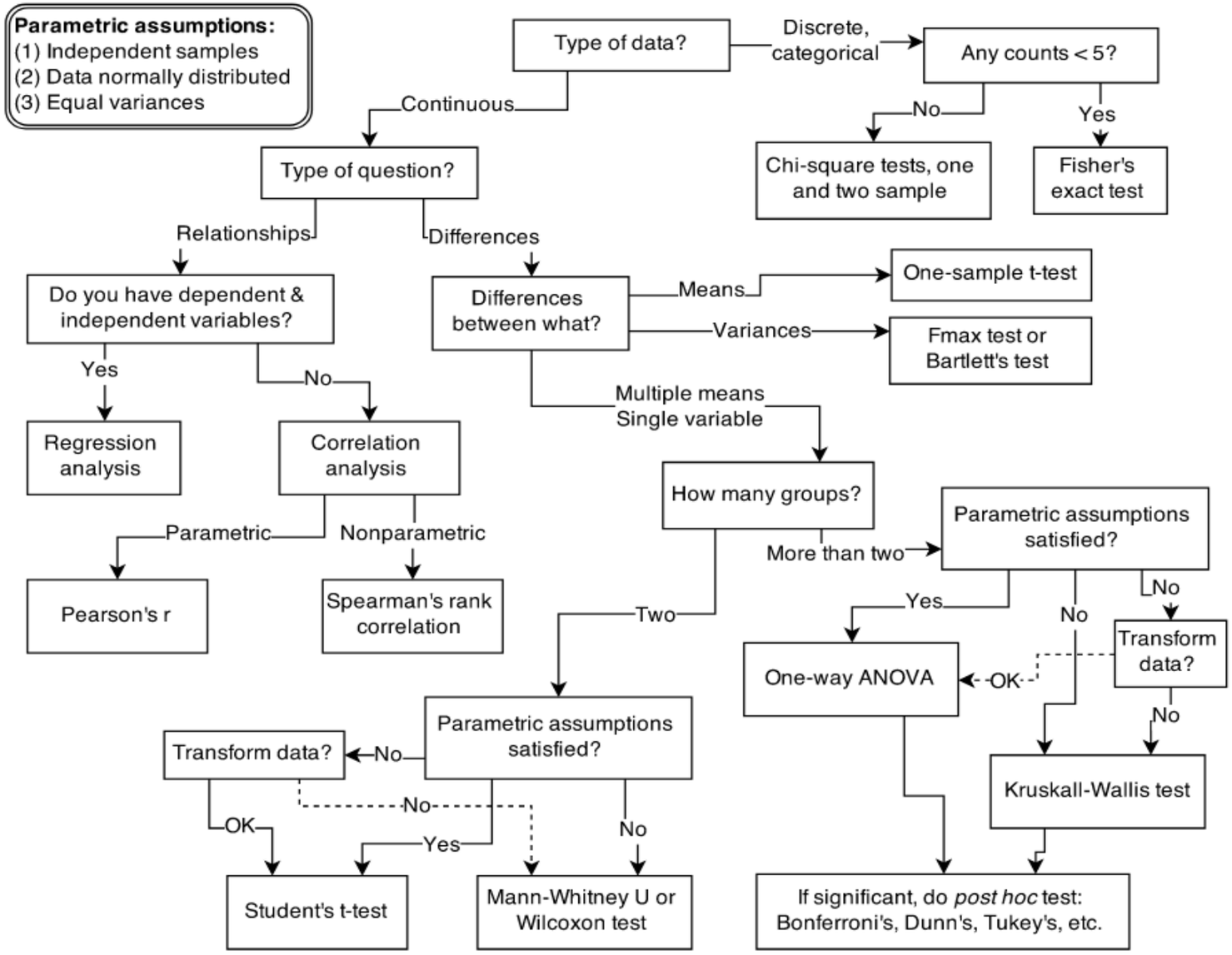


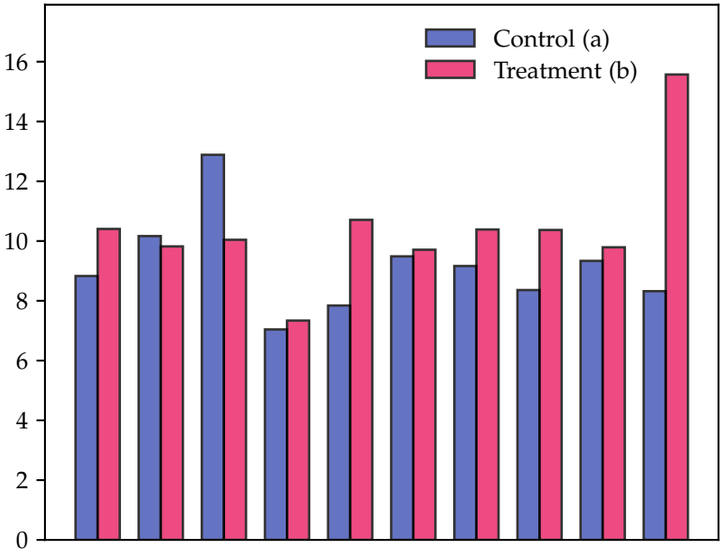
Paired t-test $p = 0.15765$

Welch's t-test $p = 0.13874$

Paired Sign Test $p = 0.10938$

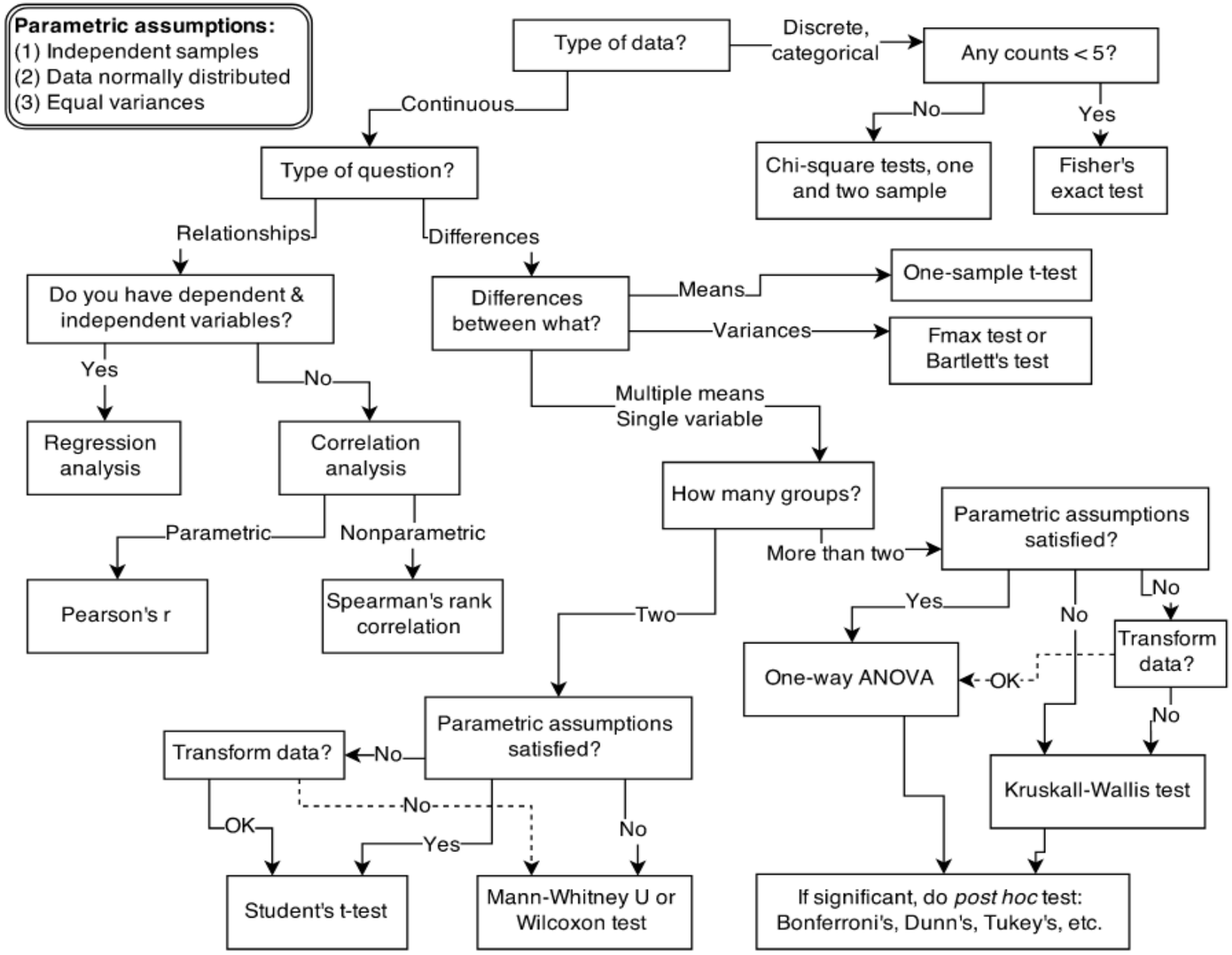
Drawback: Soo viele Tests

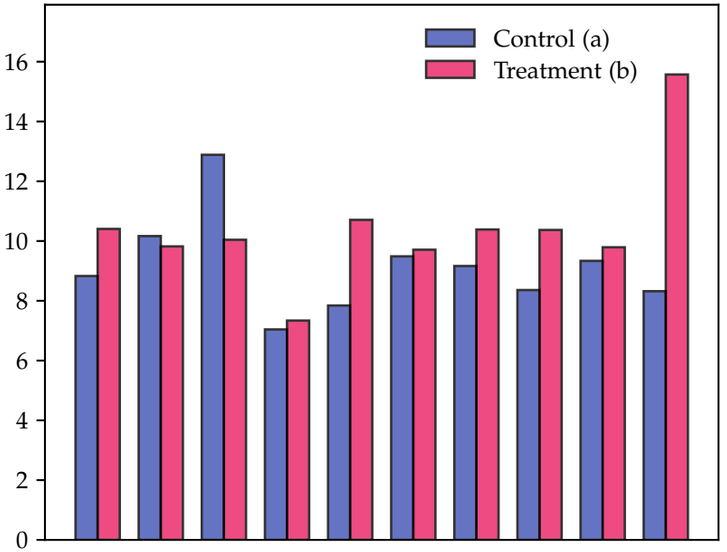




Paired t-test	$p = 0.15765$
Welch's t-test	$p = 0.13874$
Paired Sign Test	$p = 0.10938$
Mann-Whitney U	$p = \mathbf{0.03121}$

Drawback: Soo viele Tests

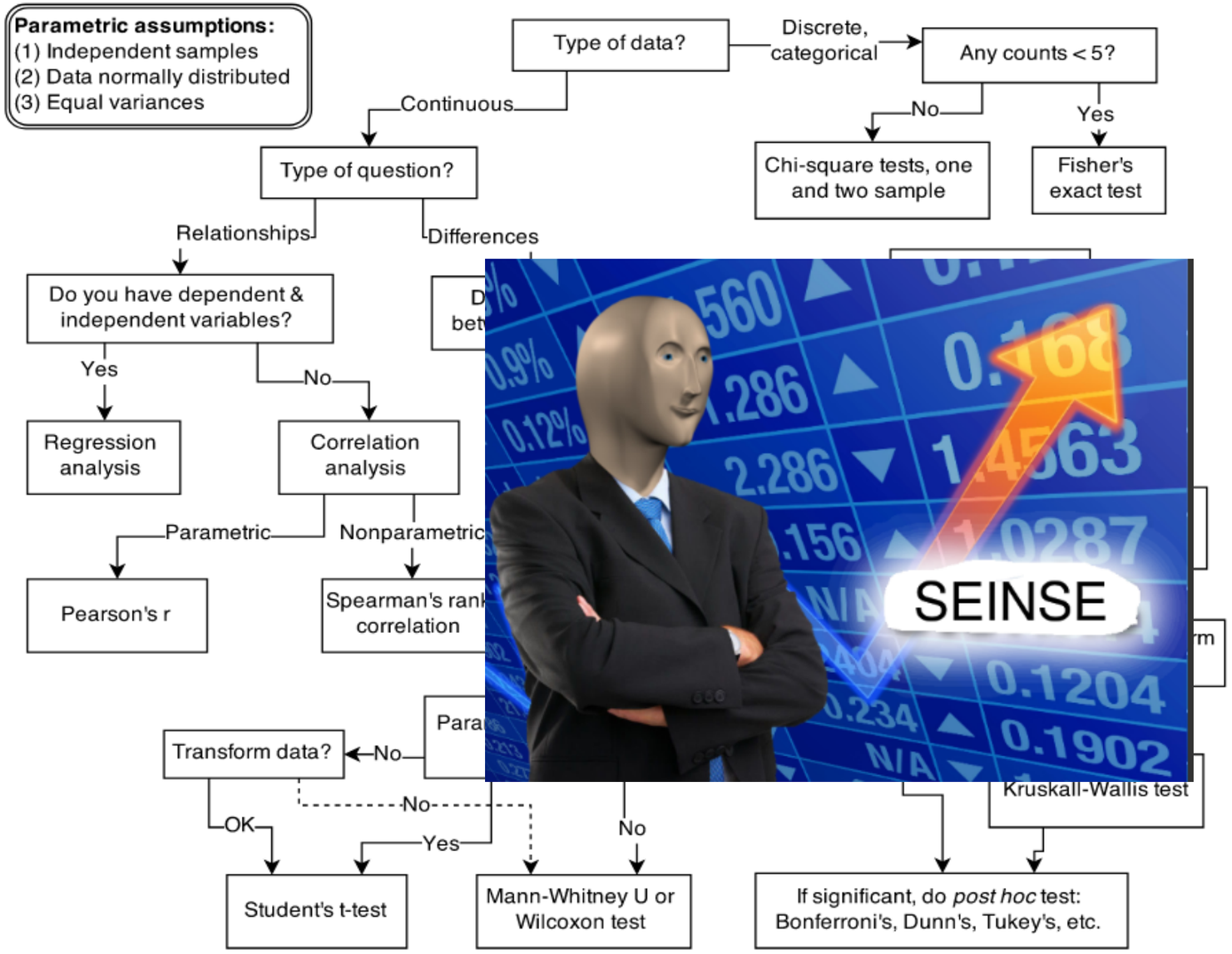


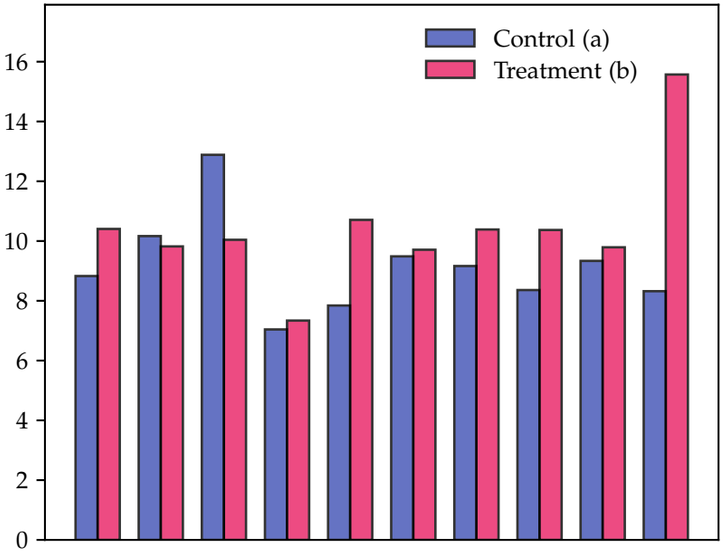


Paired t-test	$p = 0.15765$
Welch's t-test	$p = 0.13874$
Paired Sign Test	$p = 0.10938$
Mann-Whitney U	$p = \mathbf{0.03121}$

„statistically significant“

Drawback: Soo viele Tests

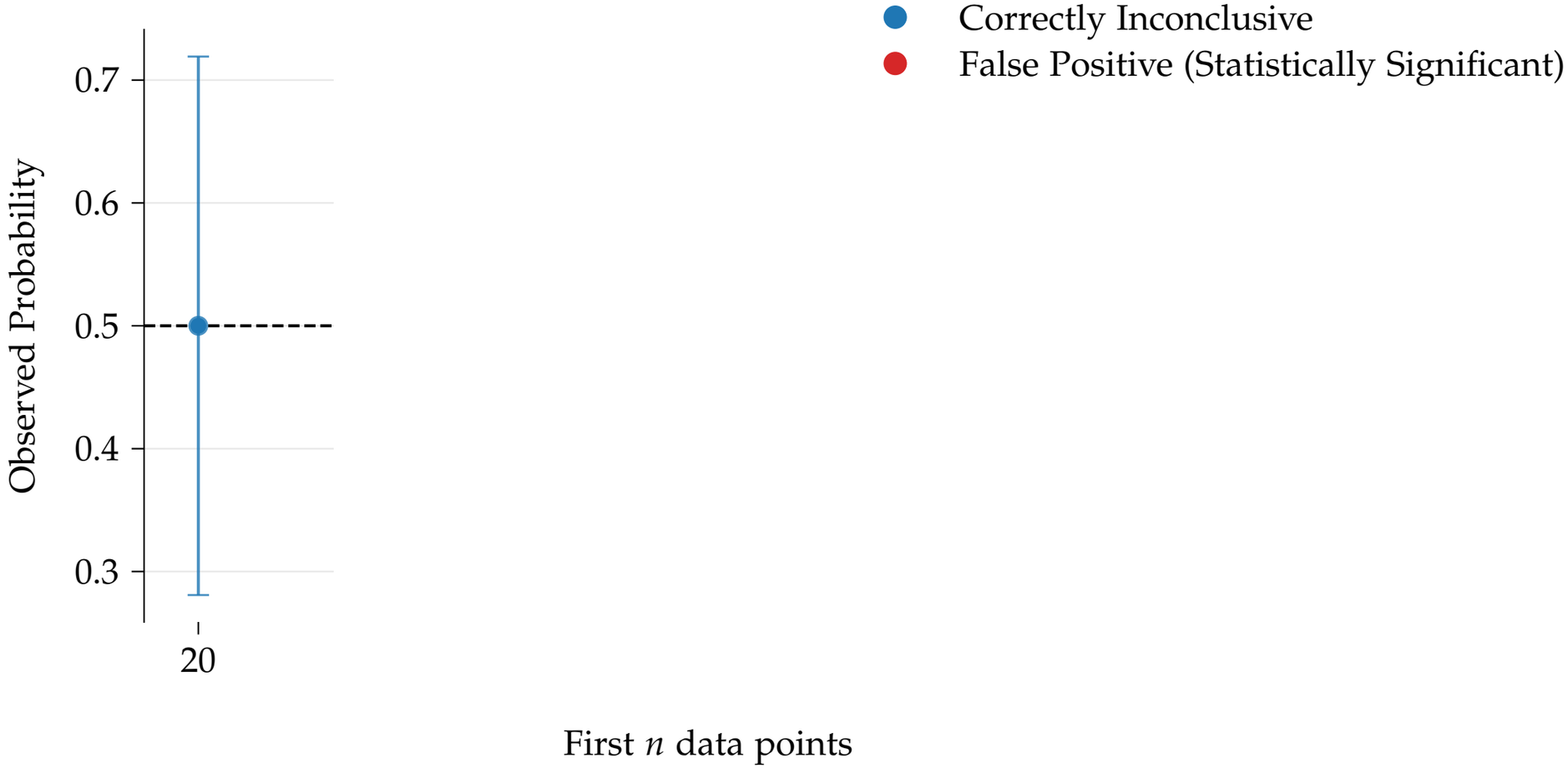




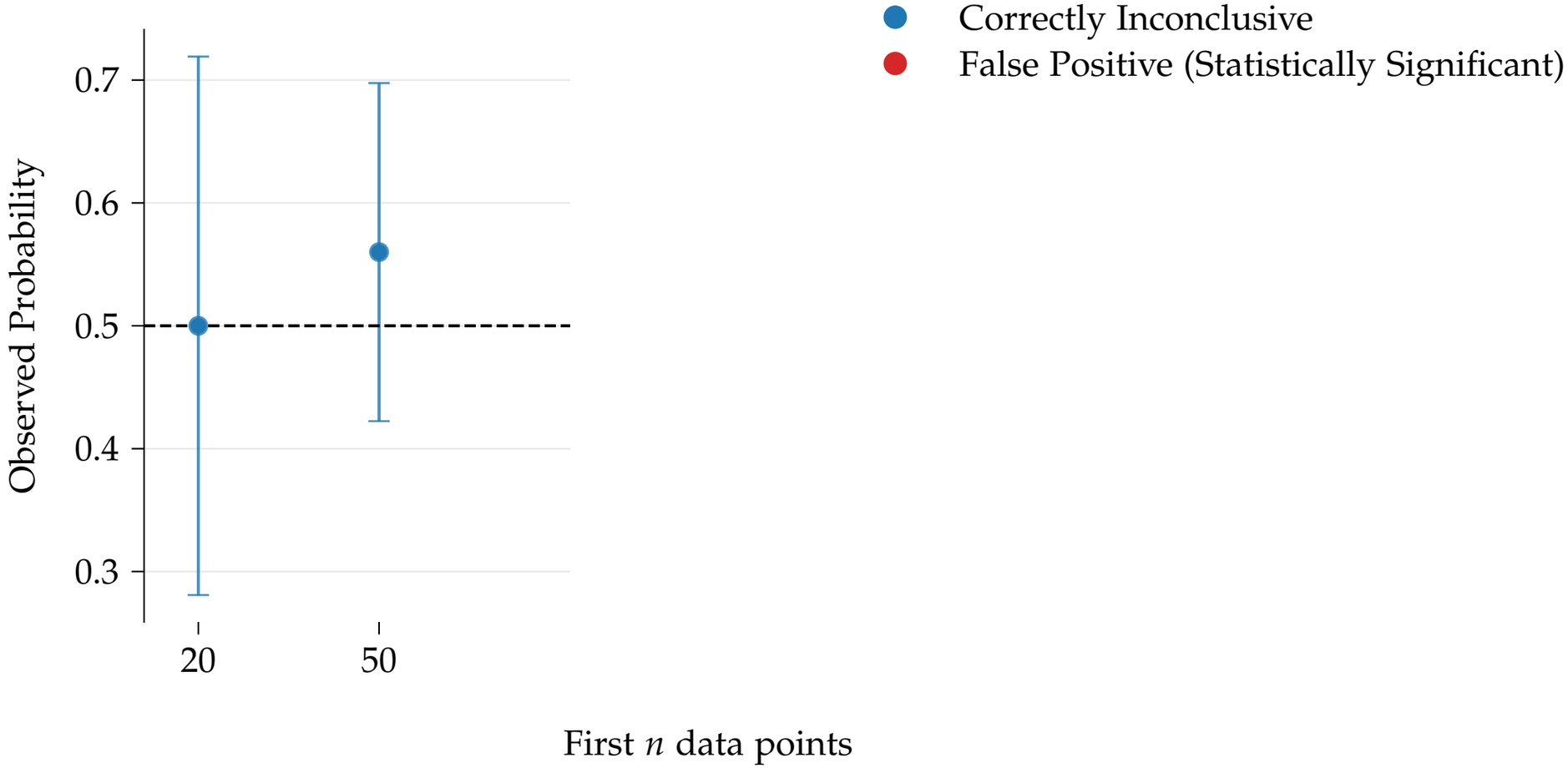
Paired t-test	$p = 0.15765$
Welch's t-test	$p = 0.13874$
Paired Sign Test	$p = 0.10938$
Mann-Whitney U	$p = \mathbf{0.03121}$

„statistically significant“

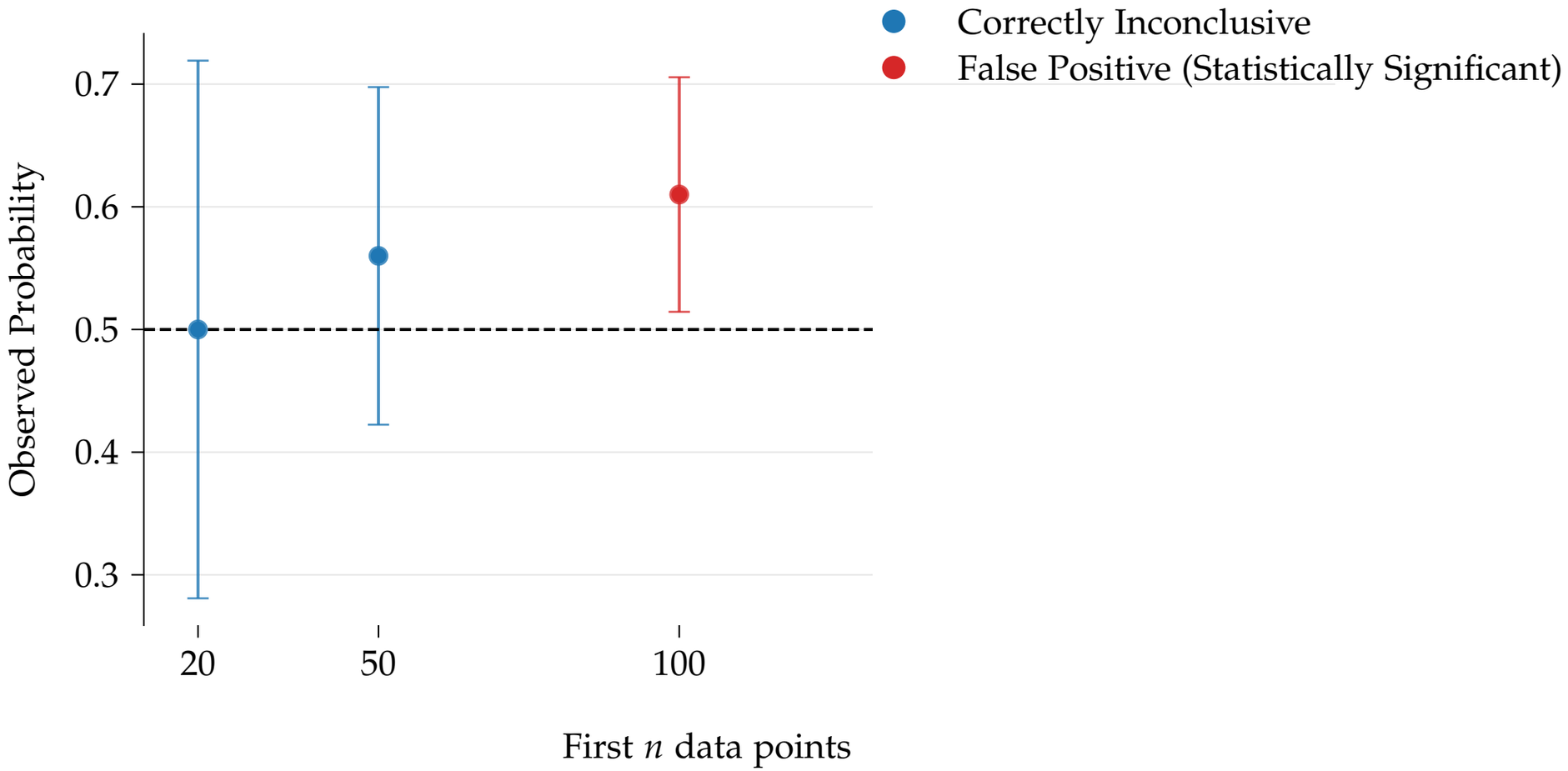
Drawback: Wiederholtes „peeking“ im A/B-Test H_0 wahr: $p = 1/2$



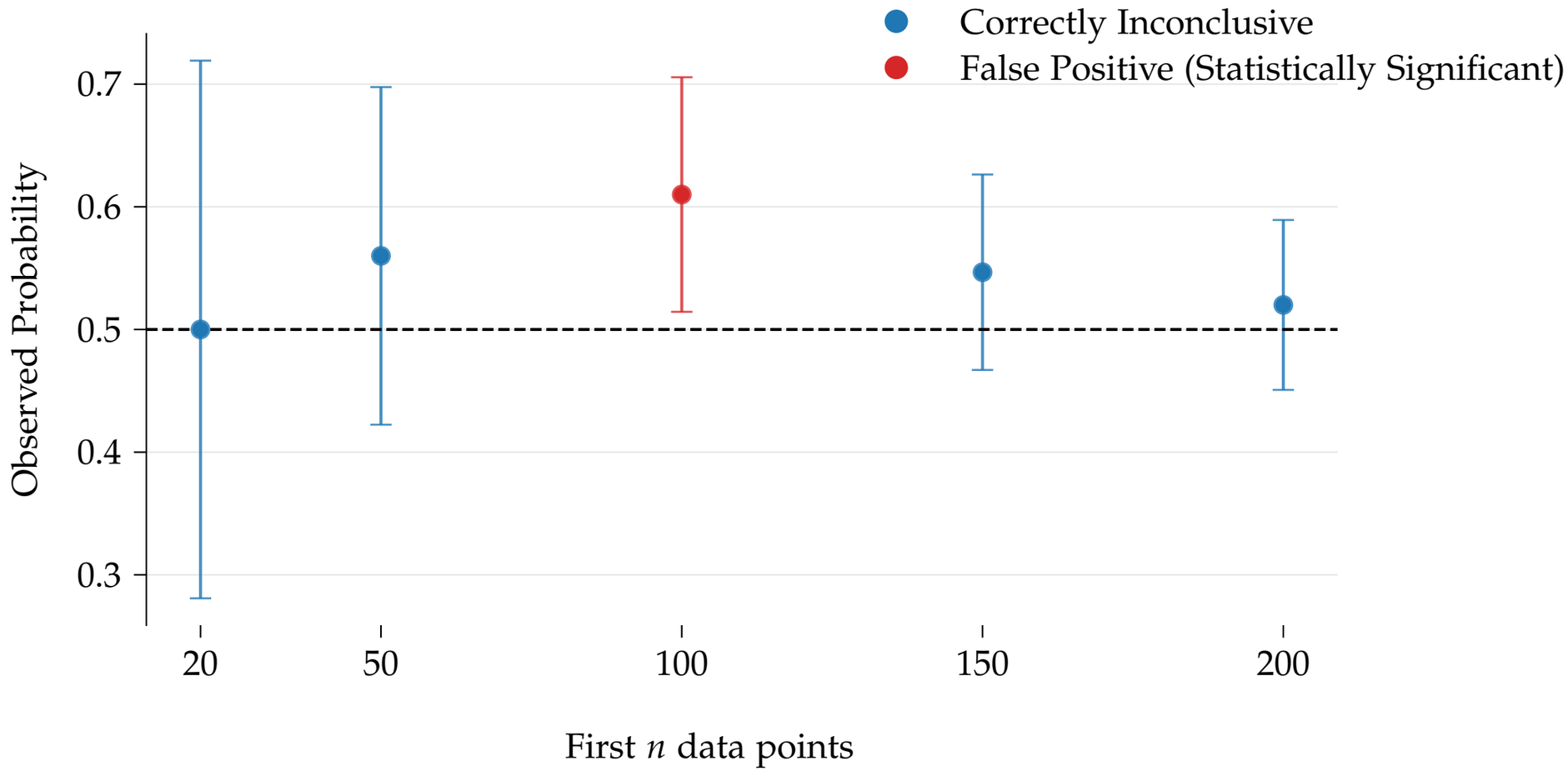
Drawback: Wiederholtes „peeking“ im A/B-Test H_0 wahr: $p = 1/2$



Drawback: Wiederholtes „peeking“ im A/B-Test H_0 wahr: $p = 1/2$



Drawback: Wiederholtes „peeking“ im A/B-Test H_0 wahr: $p = 1/2$



alles wird besser mit

Bayesian Statistics

Previously on OKInf…

Virgin Frequentistische Statistik

p-Werte unterscheiden nicht zwischen Effektstärke und Stichprobe

Hypothesentests geben keine Wahrscheinlichkeiten aus

Übereinfache Modelle

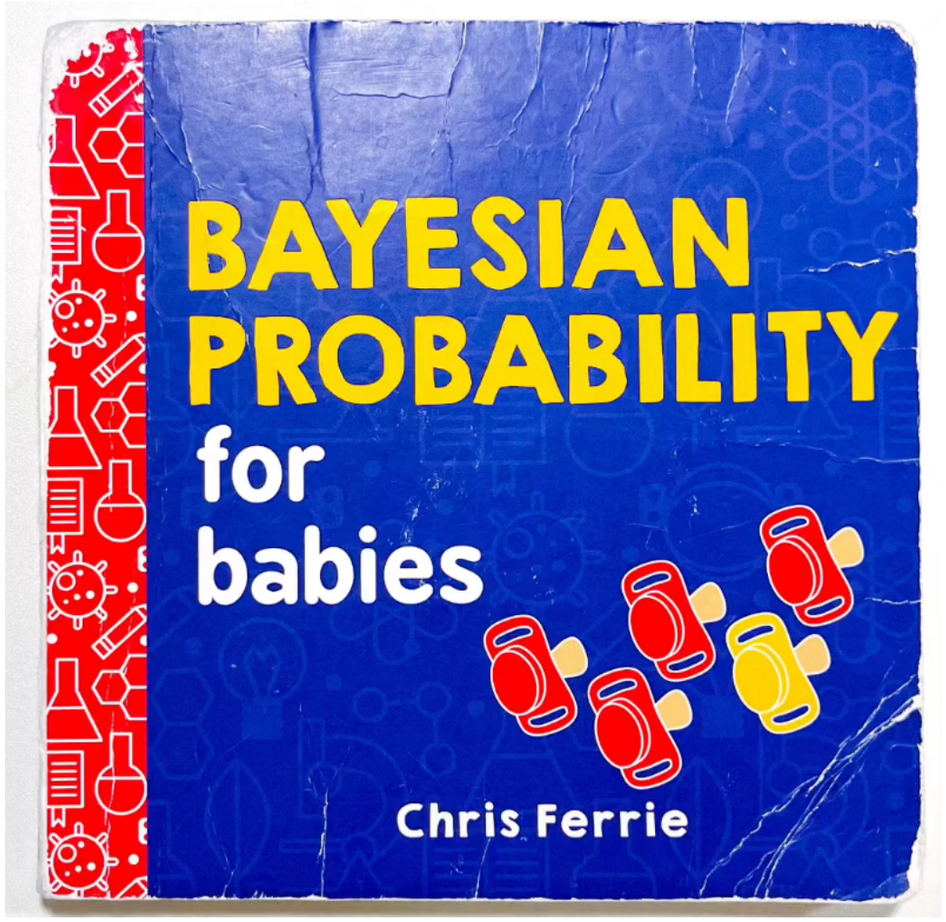
Missverständliche Konfidenzintervale

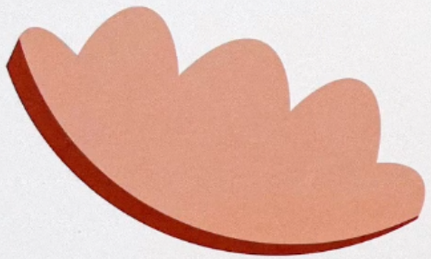
Chad Bayesians

erlauben komplexe Modelle für eine komplexe Welt

bekommen Wahrscheinlichkeitsverteil-ungen von Parametern

klare Interpretation von Ergebnissen


BAYESIAN
PROBABILITY
for
babies
Chris Ferrie

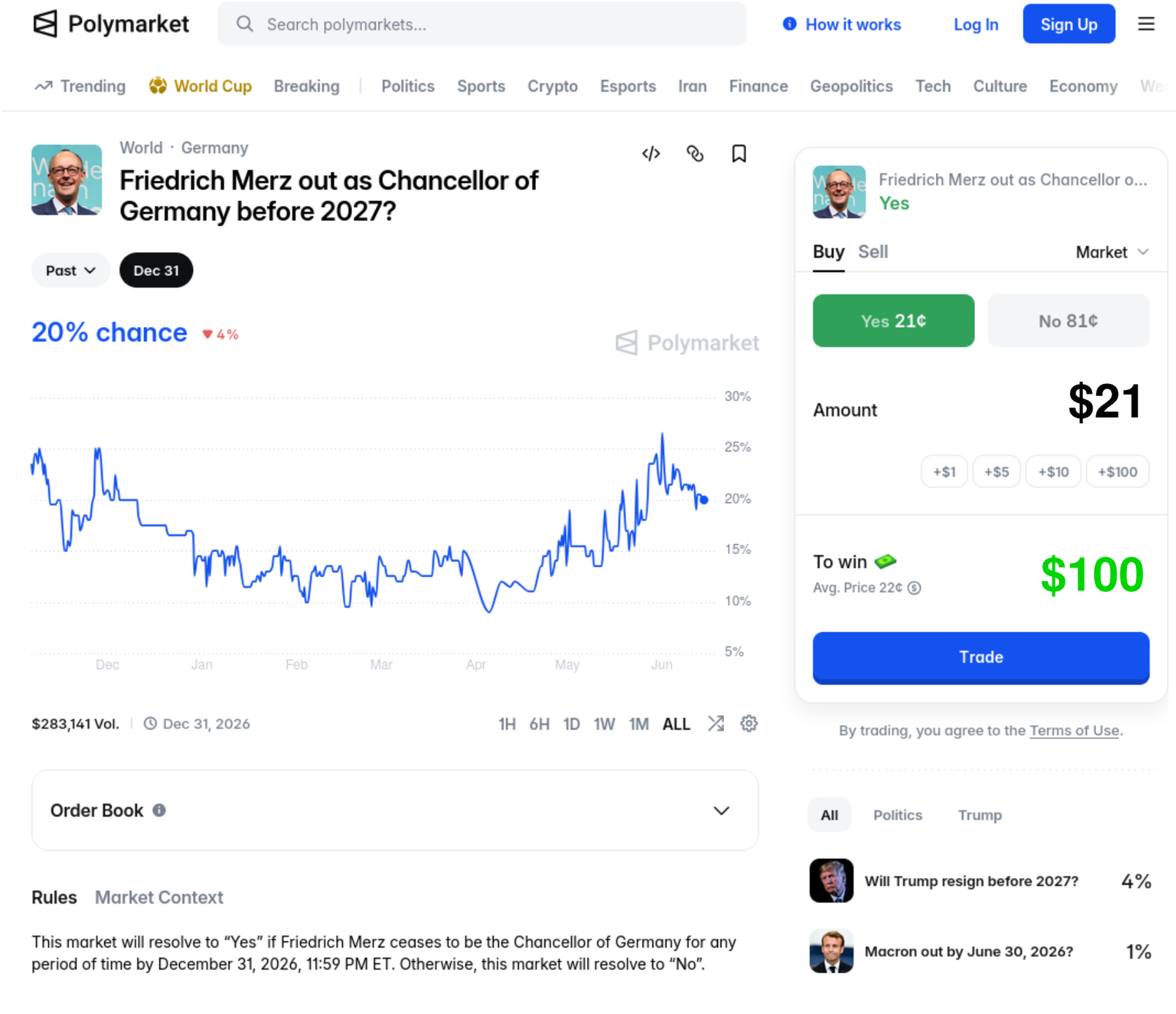


Take a bite. It has no candy.

Did it come from a candy cookie?

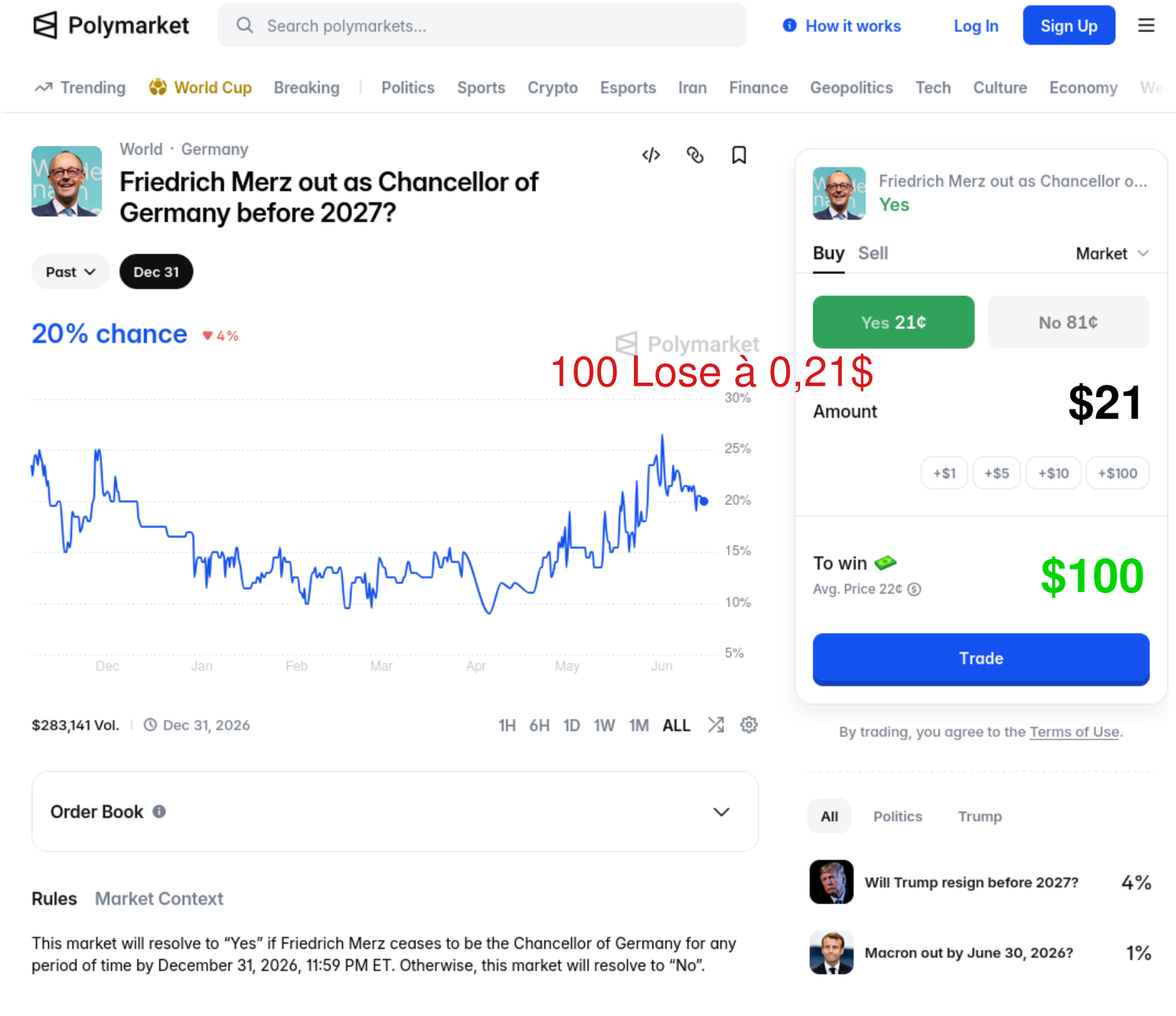
Bayesianische Statistik in Aktion

(dutch book argument)



Bayesianische Statistik in Aktion

(dutch book argument)



Operational subjective probabilities as wagering odds

Operational subjective probabilities as wagering odds

Spielregeln: [De Finetti; 1930s]

- Alice und Bob wetten auf (zukünftige) Ereignisse

Operational subjective probabilities as wagering odds

Spielregeln: [De Finetti; 1930s]

- Alice und Bob wetten auf (zukünftige) Ereignisse
- Für ein Los für das Ereignis A bekommt der Käufer $1\,€$ vom Verkäufer, falls A eintritt

Operational subjective probabilities as wagering odds

Spielregeln: [De Finetti; 1930s]

- Alice und Bob wetten auf (zukünftige) Ereignisse
- Für ein Los für das Ereignis A bekommt der Käufer $1\,€$ vom Verkäufer, falls A eintritt
- Alice *muss* A-Lose kaufen, falls Bob zum Preis $\leq p$ verkauft

Operational subjective probabilities as wagering odds

Spielregeln: [De Finetti; 1930s]

- Alice und Bob wetten auf (zukünftige) Ereignisse
- Für ein Los für das Ereignis A bekommt der Käufer $1\,€$ vom Verkäufer, falls A eintritt
- Alice *muss* A-Lose kaufen, falls Bob zum Preis $\leq p$ verkauft
- Alice *muss* A-Lose verkaufen, falls Bob zum Preis $\geq p$ kauft

Operational subjective probabilities as wagering odds

Spielregeln: [De Finetti; 1930s]

- Alice und Bob wetten auf (zukünftige) Ereignisse
- Für ein Los für das Ereignis A bekommt der Käufer $1\,€$ vom Verkäufer, falls A eintritt
- Alice *muss* A-Lose kaufen, falls Bob zum Preis $\leq p$ verkauft
- Alice *muss* A-Lose verkaufen, falls Bob zum Preis $\geq p$ kauft
- Alice setzt (je nach A) einen Preis p für ein Los fest

Operational subjective probabilities as wagering odds

Spielregeln: [De Finetti; 1930s]

- Alice und Bob wetten auf (zukünftige) Ereignisse
- Für ein Los für das Ereignis A bekommt der Käufer $1\,€$ vom Verkäufer, falls A eintritt
- Alice *muss* A-Lose kaufen, falls Bob zum Preis $\leq p$ verkauft
- Alice *muss* A-Lose verkaufen, falls Bob zum Preis $\geq p$ kauft
- Alice setzt (je nach A) einen Preis p für ein Los fest

Intuition: $p(A) \approx$ „WK für A“

Operational subjective probabilities as wagering odds

Spielregeln: [De Finetti; 1930s]

- Alice und Bob wetten auf (zukünftige) Ereignisse
- Für ein Los für das Ereignis A bekommt der Käufer $1\,€$ vom Verkäufer, falls A eintritt
- Alice *muss* A-Lose kaufen, falls Bob zum Preis $\leq p$ verkauft
- Alice *muss* A-Lose verkaufen, falls Bob zum Preis $\geq p$ kauft
- Alice setzt (je nach A) einen Preis p für ein Los fest

Intuition: $p(A) \approx$ „WK für A“

Satz: „Kohärent spielen“ $\implies$ nach den Axiomen der WK-Theorie spielen

Operational subjective probabilities as wagering odds

Spielregeln: [De Finetti; 1930s]

- Alice und Bob wetten auf (zukünftige) Ereignisse
- Für ein Los für das Ereignis A bekommt der Käufer $1\,€$ vom Verkäufer, falls A eintritt
- Alice *muss* A-Lose kaufen, falls Bob zum Preis $\leq p$ verkauft
- Alice *muss* A-Lose verkaufen, falls Bob zum Preis $\geq p$ kauft
- Alice setzt (je nach A) einen Preis p für ein Los fest

Intuition: $p(A) \approx$ „WK für A“

Beispiel:

Alice:

A. $p(\text{Wü wird exzellent}) = 0.7$

B. $p(\text{Wü wird nicht exzellent}) = 0.5$

Satz: „Kohärent spielen“ $\implies$ nach den Axiomen der WK-Theorie spielen

Operational subjective probabilities as wagering odds

Spielregeln: [De Finetti; 1930s]

- Alice und Bob wetten auf (zukünftige) Ereignisse
- Für ein Los für das Ereignis A bekommt der Käufer $1\,€$ vom Verkäufer, falls A eintritt
- Alice *muss* A-Lose kaufen, falls Bob zum Preis $\leq p$ verkauft
- Alice *muss* A-Lose verkaufen, falls Bob zum Preis $\geq p$ kauft
- Alice setzt (je nach A) einen Preis p für ein Los fest

Intuition: $p(A) \approx$ „WK für A“

Beispiel:

Alice:

A. $p(\text{Wü wird exzellent}) = 0.7$

B. $p(\text{Wü wird nicht exzellent}) = 0.5$

Bob: *infinite money glitch*

- Verkaufe A-Los und B-Los $(0.7 + 0.5)$ $\rightarrow$ Einnahmen 1.2

Satz: „Kohärent spielen“ $\implies$ nach den Axiomen der WK-Theorie spielen

Operational subjective probabilities as wagering odds

Spielregeln: [De Finetti; 1930s]

- Alice und Bob wetten auf (zukünftige) Ereignisse
- Für ein Los für das Ereignis A bekommt der Käufer $1\,€$ vom Verkäufer, falls A eintritt
- Alice *muss* A-Lose kaufen, falls Bob zum Preis $\leq p$ verkauft
- Alice *muss* A-Lose verkaufen, falls Bob zum Preis $\geq p$ kauft
- Alice setzt (je nach A) einen Preis p für ein Los fest

Intuition: $p(A) \approx$ „WK für A“

Beispiel:

Alice:

A. $p(\text{Wü wird exzellent}) = 0.7$

B. $p(\text{Wü wird nicht exzellent}) = 0.5$

Bob: *infinite money glitch*

- Verkaufe A-Los und B-Los $(0.7 + 0.5)$ $\rightarrow$ Einnahmen 1.2

	A	B	Gewinn
Wü wird exzellent	-1	0	0.2
Wü wird nicht exzellent	0	-1	0.2

Satz: „Kohärent spielen“ $\implies$ nach den Axiomen der WK-Theorie spielen

Bayes' Theorem in subjective beliefs

- Für ein Los für das Ereignis A bekommt der Käufer $1\,€$ vom Verkäufer, falls A eintritt
- Alice *muss* Lose kaufen, falls Bob zum Preis $\leq p$ verkauft
- Alice *muss* Lose verkaufen, falls Bob zum Preis $\geq p$ kauft

Bayes' Theorem in subjective beliefs

Neu: Kombi-Los $A \mid B$

- Falls A und B eintritt, bekommt der Käufer $1\,€$.
- Falls B nicht eintritt, bekommt der Käufer den Kaufpreis zurück [als ob das Los nie gekauft würde].

- Für ein Los für das Ereignis A bekommt der Käufer $1\,€$ vom Verkäufer, falls A eintritt
- Alice *muss* Lose kaufen, falls Bob zum Preis $\leq p$ verkauft
- Alice *muss* Lose verkaufen, falls Bob zum Preis $\geq p$ kauft

Bayes' Theorem in subjective beliefs

Neu: Kombi-Los $A \mid B$

- Falls A und B eintritt, bekommt der Käufer $1\,€$.
- Falls B nicht eintritt, bekommt der Käufer den Kaufpreis zurück [als ob das Los nie gekauft würde].

Intuition: $p(A|B) \approx$ „WK für A *falls B eintritt*“

- Für ein Los für das Ereignis A bekommt der Käufer $1\,€$ vom Verkäufer, falls A eintritt
- Alice *muss* Lose kaufen, falls Bob zum Preis $\leq p$ verkauft
- Alice *muss* Lose verkaufen, falls Bob zum Preis $\geq p$ kauft

Bayes' Theorem in subjective beliefs

Szenario: Kaum jemand ist in der Vorlesung. War gestern Party?

Neu: Kombi-Los $A \mid B$

- Falls A und B eintritt, bekommt der Käufer $1\,€$.
- Falls B nicht eintritt, bekommt der Käufer den Kaufpreis zurück [als ob das Los nie gekauft würde].

Intuition: $p(A|B) \approx$ „WK für A *falls B eintritt*“

- Für ein Los für das Ereignis A bekommt der Käufer $1\,€$ vom Verkäufer, falls A eintritt
- Alice *muss* Lose kaufen, falls Bob zum Preis $\leq p$ verkauft
- Alice *muss* Lose verkaufen, falls Bob zum Preis $\geq p$ kauft

Bayes' Theorem in subjective beliefs

Neu: Kombi-Los $A \mid B$

- Falls A und B eintritt, bekommt der Käufer $1\,€$.
- Falls B nicht eintritt, bekommt der Käufer den Kaufpreis zurück [als ob das Los nie gekauft würde].

Intuition: $p(A|B) \approx$ „WK für A *falls B eintritt*“

- Für ein Los für das Ereignis A bekommt der Käufer $1\,€$ vom Verkäufer, falls A eintritt
- Alice *muss* Lose kaufen, falls Bob zum Preis $\leq p$ verkauft
- Alice *muss* Lose verkaufen, falls Bob zum Preis $\geq p$ kauft

Szenario: Kaum jemand ist in der Vorlesung. War gestern Party?

$$p(\text{Party}) = 0.1$$
$$p(\text{Hörsaal leer} \mid \text{Party}) = 0.9$$
$$p(\text{Hörsaal leer} \mid \text{keine Party}) = 0.3$$

Bayes' Theorem in subjective beliefs

Neu: Kombi-Los $A \mid B$

- Falls A und B eintritt, bekommt der Käufer $1\,€$.
- Falls B nicht eintritt, bekommt der Käufer den Kaufpreis zurück [als ob das Los nie gekauft würde].

Intuition: $p(A|B) \approx$ „WK für A *falls B eintritt*“

- Für ein Los für das Ereignis A bekommt der Käufer $1\,€$ vom Verkäufer, falls A eintritt
- Alice *muss* Lose kaufen, falls Bob zum Preis $\leq p$ verkauft
- Alice *muss* Lose verkaufen, falls Bob zum Preis $\geq p$ kauft

Szenario: Kaum jemand ist in der Vorlesung. War gestern Party?

$$p(\text{Party}) = 0.1$$
$$p(\text{Hörsaal leer} \mid \text{Party}) = 0.9$$
$$p(\text{Hörsaal leer} \mid \text{keine Party}) = 0.3$$

$$p(\text{Party} \mid \text{Hörsaal leer}) = ???$$

Bayes' Theorem in subjective beliefs

Neu: Kombi-Los $A \mid B$

- Falls A und B eintritt, bekommt der Käufer $1\,€$.
- Falls B nicht eintritt, bekommt der Käufer den Kaufpreis zurück [als ob das Los nie gekauft würde].

Intuition: $p(A|B) \approx$ „WK für A *falls B eintritt*“

- Für ein Los für das Ereignis A bekommt der Käufer $1\,€$ vom Verkäufer, falls A eintritt
- Alice *muss* Lose kaufen, falls Bob zum Preis $\leq p$ verkauft
- Alice *muss* Lose verkaufen, falls Bob zum Preis $\geq p$ kauft

Szenario: Kaum jemand ist in der Vorlesung. War gestern Party?

$$p(\text{Party}) = 0.1$$
$$p(\text{Hörsaal leer} \mid \text{Party}) = 0.9$$
$$p(\text{Hörsaal leer} \mid \text{keine Party}) = 0.3$$

$$p(\text{Party} \mid \text{Hörsaal leer}) = ???$$

Bayes Theorem

$$P(A \mid B) = \frac{P(B \mid A) \cdot P(A)}{P(B)}$$

Bayes' Theorem in subjective beliefs

Neu: Kombi-Los $A \mid B$

- Falls A und B eintritt, bekommt der Käufer $1\,€$.
- Falls B nicht eintritt, bekommt der Käufer den Kaufpreis zurück [als ob das Los nie gekauft würde].

Intuition: $p(A|B) \approx$ „WK für A *falls B eintritt*“

- Für ein Los für das Ereignis A bekommt der Käufer $1\,€$ vom Verkäufer, falls A eintritt
- Alice *muss* Lose kaufen, falls Bob zum Preis $\leq p$ verkauft
- Alice *muss* Lose verkaufen, falls Bob zum Preis $\geq p$ kauft

Szenario: Kaum jemand ist in der Vorlesung. War gestern Party?

$$p(\text{Party}) = 0.1$$
$$p(\text{Hörsaal leer} \mid \text{Party}) = 0.9$$
$$p(\text{Hörsaal leer} \mid \text{keine Party}) = 0.3$$

$$p(\text{Party} \mid \text{Hörsaal leer}) = ???$$

Bayes Theorem

$$P(A \mid B) = \frac{P(B \mid A) \cdot P(A)}{P(B)}$$

$$P(B \mid A) \cdot P(A) + P(B \mid \neg A) \cdot P(\neg A)$$

Wahrscheinlichkeit Bayesianisch

Definition: Wahrscheinlichkeit ist ein Maß
für die Gewissheit
der (persönlichen, rationalen) **Einschätzung**
eines Sachverhalts.

Wahrscheinlichkeit Bayesianisch

Definition: Wahrscheinlichkeit ist ein Maß
für die Gewissheit
der (persönlichen, rationalen) **Einschätzung**
eines Sachverhalts.

Bayesianisches Programm: Die Mechanismen der Umwelt sind unbekannt; wir können daher nur Einschätzungen abgeben – einmal gemessene experimentellen Daten sind dagegen fest.
Unsere Inferenzprozeduren sollen derart sein, dass sie aus den festen Daten möglichst rational unsere Einschätzungen verbessern.

Wahrscheinlichkeit Bayesianisch

Definition: Wahrscheinlichkeit ist ein Maß
für die Gewissheit
der (persönlichen, rationalen) **Einschätzung**
eines Sachverhalts.

Bayesianisches Programm: Die Mechanismen der Umwelt sind unbekannt; wir können daher nur Einschätzungen abgeben – einmal gemessene experimentellen Daten sind dagegen fest.
Unsere Inferenzprozeduren sollen derart sein, dass sie aus den festen Daten möglichst rational unsere Einschätzungen verbessern.

$$P(\text{Hypothese} \mid \text{Daten}) = \frac{P(\text{Daten} \mid \text{Hypothese}) \cdot P(\text{Hypothese})}{P(\text{Daten})}$$

Wahrscheinlichkeit Bayesianisch

Definition: Wahrscheinlichkeit ist ein Maß
für die Gewissheit
der (persönlichen, rationalen) **Einschätzung**
eines Sachverhalts.

Bayesianisches Programm: Die Mechanismen der Umwelt sind unbekannt; wir können daher nur Einschätzungen abgeben – einmal gemessene experimentellen Daten sind dagegen fest.
Unsere Inferenzprozeduren sollen derart sein, dass sie aus den festen Daten möglichst rational unsere Einschätzungen verbessern.

$$\underbrace{P(\text{Hypothese} \mid \text{Daten})}_{\text{Posterior}} = \frac{\overbrace{P(\text{Daten} \mid \text{Hypothese})}^{\text{Likelihood}} \cdot \overbrace{P(\text{Hypothese})}^{\text{Prior}}}{\underbrace{P(\text{Daten})}_{\text{(Model evidence)}}}$$

Wahrscheinlichkeit Bayesianisch

Definition: Wahrscheinlichkeit ist ein Maß
für die Gewissheit
der (persönlichen, rationalen) **Einschätzung**
e

Frequentistisch: Die Welt / der Parameter θ ist fest, aber meine Daten sind zufällig.

Ich frage: Wie wahrscheinlich sind diese Daten unter meiner Annahme?

Ba nismen der Umwelt sind unbekannt;
wi abgeben – einmal gemessene
ex est.
Ur art sein, dass sie aus den festen Daten
mö gen verbessern.

$$P(\text{Hypothese} \mid \text{Daten}) = \frac{P(\text{Daten} \mid \text{Hypothese}) \cdot P(\text{Hypothese})}{P(\text{Daten})}$$

Posterior — Likelihood — Prior — (Model evidence)

Wahrscheinlichkeit Bayesianisch

Definition: Wahrscheinlichkeit ist ein Maß
für die Gewissheit
der (persönlichen, rationalen) **Einschätzung**
e

Frequentistisch: Die Welt / der Parameter θ ist fest, aber meine Daten sind zufällig.

Ich frage: Wie wahrscheinlich sind diese Daten unter meiner Annahme?

Bayesianisch: Mein Wissen über die Welt (θ) ist unsicher. Aber die Daten sind fest (ich habe sie ja gemessen!).

Ich frage: Wie wahrscheinlich ist meine Hypothese, nachdem ich diese Daten gesehen habe?

Ba nism
wi abg
exp est.
Ur art s
mö gen

$$\underbrace{P(\text{Hypothese} \mid \text{Daten})}_{\text{Posterior}} = \frac{\overbrace{P(\text{Daten} \mid \text{Hypothese})}^{\text{Likelihood}} \cdot \overbrace{P(\text{Hypothese})}^{\text{Prior}}}{\underbrace{P(\text{Daten})}_{\text{(Model evidence)}}}$$

Bayesian Statistics

Idee: Modellieren Wissen über eine Eigenschaft einer Population als Zufallsvariable Θ. Aktualisieren unser Wissen aus den beobachteten Daten mittels Bayes' Theorem.

Bayesian Statistics

Idee: Modellieren Wissen über eine Eigenschaft einer Population als Zufallsvariable Θ. Aktualisieren unser Wissen aus den beobachteten Daten mittels Bayes' Theorem.

1. Prior aufstellen: $p(\theta)$ für $\theta \in \Theta$; subjektives Vorwissen

Bayesian Statistics

Idee: Modellieren Wissen über eine Eigenschaft einer Population als Zufallsvariable Θ. Aktualisieren unser Wissen aus den beobachteten Daten mittels Bayes' Theorem.

1. Prior aufstellen: $p(\theta)$ für $\theta \in \Theta$; subjektives Vorwissen

2. Datengenerierung modellieren: ordnet festem Parameter θ eine WK zu, gewisse Daten $\mathbf{X}$ zu beobachten; $p(\mathbf{X}|\theta)$

Bayesian Statistics

Idee: Modellieren Wissen über eine Eigenschaft einer Population als Zufallsvariable Θ. Aktualisieren unser Wissen aus den beobachteten Daten mittels Bayes' Theorem.

1. Prior aufstellen: $p(\theta)$ für $\theta \in \Theta$; subjektives Vorwissen

2. Datengenerierung modellieren: ordnet festem Parameter θ eine WK zu, gewisse Daten $\mathbf{X}$ zu beobachten; $p(\mathbf{X}|\theta)$

3. Posterior ausrechnen (algebraisch oder Simulation)

$$p(\theta|\mathbf{X}) = \frac{p(\mathbf{X}|\theta)p(\theta)}{p(\mathbf{X})} = \frac{p(\mathbf{X}|\theta)p(\theta)}{\int_\Theta p(\mathbf{X}|\theta)p(\theta)\,\mathrm{d}\theta} \propto p(\mathbf{X}|\theta)p(\theta)$$

Bayesian Statistics

Idee: Modellieren Wissen über eine Eigenschaft einer Population als Zufallsvariable Θ. Aktualisieren unser Wissen aus den beobachteten Daten mittels Bayes' Theorem.

1. Prior aufstellen: $p(\theta)$ für $\theta \in \Theta$; subjektives Vorwissen

2. Datengenerierung modellieren: ordnet festem Parameter θ eine WK zu, gewisse Daten $\mathbf{X}$ zu beobachten; $p(\mathbf{X}|\theta)$

3. Posterior ausrechnen (algebraisch oder Simulation)

$$p(\theta|\mathbf{X}) = \frac{p(\mathbf{X}|\theta)p(\theta)}{p(\mathbf{X})} = \frac{p(\mathbf{X}|\theta)p(\theta)}{\int_\Theta p(\mathbf{X}|\theta)p(\theta)\,\mathrm{d}\theta} \propto p(\mathbf{X}|\theta)p(\theta)$$

4. Hypothesen mit Posterior prüfen

Bayesian Statistics

Idee: Modellieren Wissen über eine Eigenschaft einer Population als Zufallsvariable Θ. Aktualisieren unser Wissen aus den beobachteten Daten mittels Bayes' Theorem.

1. Prior aufstellen: $p(\theta)$ für $\theta \in \Theta$; subjektives Vorwissen $\theta \sim \text{Uniform}(0, 1)$

2. Datengenerierung modellieren: ordnet festem Parameter θ eine WK zu, gewisse Daten $\mathbf{X}$ zu beobachten; $p(\mathbf{X}|\theta)$

3. Posterior ausrechnen (algebraisch oder Simulation)

$$p(\theta|\mathbf{X}) = \frac{p(\mathbf{X}|\theta)p(\theta)}{p(\mathbf{X})} = \frac{p(\mathbf{X}|\theta)p(\theta)}{\int_\Theta p(\mathbf{X}|\theta)p(\theta)\,\mathrm{d}\theta} \propto p(\mathbf{X}|\theta)p(\theta)$$

4. Hypothesen mit Posterior prüfen

Bayesian Statistics

Idee: Modellieren Wissen über eine Eigenschaft einer Population als Zufallsvariable Θ. Aktualisieren unser Wissen aus den beobachteten Daten mittels Bayes' Theorem.

1. Prior aufstellen: $p(\theta)$ für $\theta \in \Theta$; subjektives Vorwissen $\quad \theta \sim \text{Uniform}(0, 1)$

2. Datengenerierung modellieren: ordnet festem Parameter θ eine WK zu, gewisse Daten $\mathbf{X}$ zu beobachten; $p(\mathbf{X}|\theta)$ $\quad X \sim \text{Bernoulli}(\theta)$

3. Posterior ausrechnen (algebraisch oder Simulation)

$$p(\theta|\mathbf{X}) = \frac{p(\mathbf{X}|\theta)p(\theta)}{p(\mathbf{X})} = \frac{p(\mathbf{X}|\theta)p(\theta)}{\int_\Theta p(\mathbf{X}|\theta)p(\theta)\,\mathrm{d}\theta} \propto p(\mathbf{X}|\theta)p(\theta)$$

4. Hypothesen mit Posterior prüfen

Bayesian Statistics

Idee: Modellieren Wissen über eine Eigenschaft einer Population als Zufallsvariable Θ. Aktualisieren unser Wissen aus den beobachteten Daten mittels Bayes' Theorem.

1. Prior aufstellen: $p(\theta)$ für $\theta \in \Theta$; subjektives Vorwissen $\quad \theta \sim \text{Uniform}(0,1)$

2. Datengenerierung modellieren: ordnet festem Parameter θ eine WK zu, gewisse Daten $\mathbf{X}$ zu beobachten; $p(\mathbf{X}|\theta)$ $\quad X \sim \text{Bernoulli}(\theta)$

3. Posterior ausrechnen (algebraisch oder Simulation)

$$p(\theta|\mathbf{X}) = \frac{p(\mathbf{X}|\theta)p(\theta)}{p(\mathbf{X})} = \frac{p(\mathbf{X}|\theta)p(\theta)}{\int_\Theta p(\mathbf{X}|\theta)p(\theta)\,\mathrm{d}\theta} \propto p(\mathbf{X}|\theta)p(\theta)$$

4. Hypothesen mit Posterior prüfen

$$p(\theta \mid x_1, \ldots, x_n) = \frac{\left[\prod_i^n \theta^{x_i}(1-\theta)^{1-x_i}\right] \cdot 1}{\frac{(\sum_i^n x_i)! \cdot (n - \sum_i^n x_i)!}{(n+1)!}}$$

Bayesian Statistics

Idee: Modellieren Wissen über eine Eigenschaft einer Population als Zufallsvariable Θ. Aktualisieren unser Wissen aus den beobachteten Daten mittels Bayes' Th

1. Prior au

$\theta \sim \text{Uniform}(0,1)$

2. Dateng r θ

eine WK z

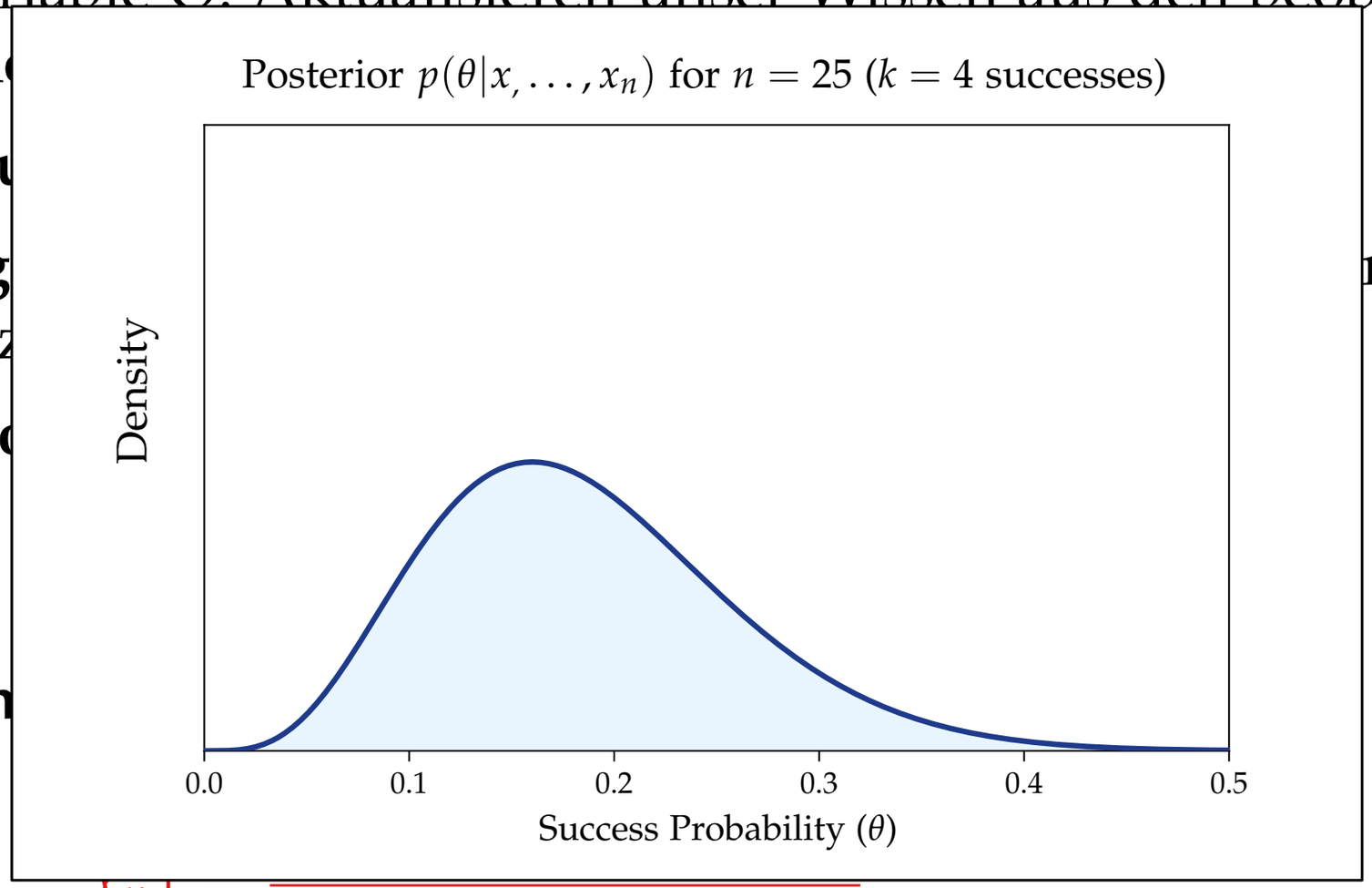
$X \sim \text{Bernoulli}(\theta)$

3. Posteri

$p(\theta|\mathbf{X}) =$

4. Hypoth

$$p(\theta \mid x_1, \ldots, x_n) = \frac{}{\frac{(\sum_i^n x_i)! \cdot (n - \sum_i^n x_i)!}{(n+1)!}}$$



Bayesian Statistics

Idee: Modellieren Wissen über eine Eigenschaft einer Population als Zufallsvariable Θ. Aktualisieren unser Wissen aus den beobachteten Daten mittels Bayes' Th

1. Prior au

$\theta \sim \text{Uniform}(0,1)$

2. Dateng r θ

eine WK z

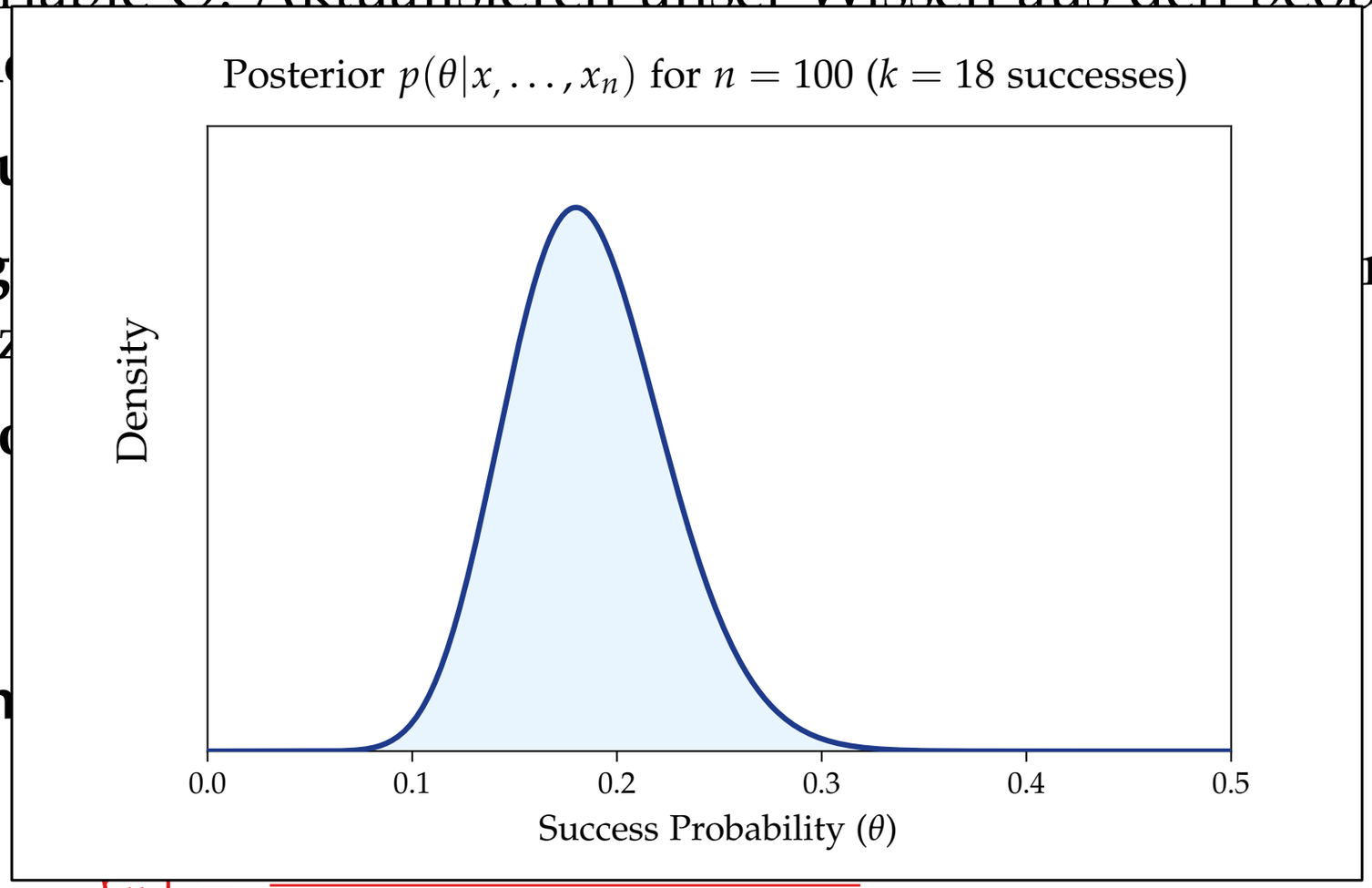
$X \sim \text{Bernoulli}(\theta)$

3. Posteri

$p(\theta|\mathbf{X}) =$

4. Hypoth

$$p(\theta \mid x_1, \ldots, x_n) = \frac{}{\frac{(\sum_i^n x_i)! \cdot (n - \sum_i^n x_i)!}{(n+1)!}}$$



Bayesian Statistics

Idee: Modellieren Wissen über eine Eigenschaft einer Population als Zufallsvariable Θ. Aktualisieren unser Wissen aus den beobachteten Daten mittels Bayes' Th

1. Prior au

$\theta \sim \text{Uniform}(0,1)$

2. Dateng r θ

eine WK z

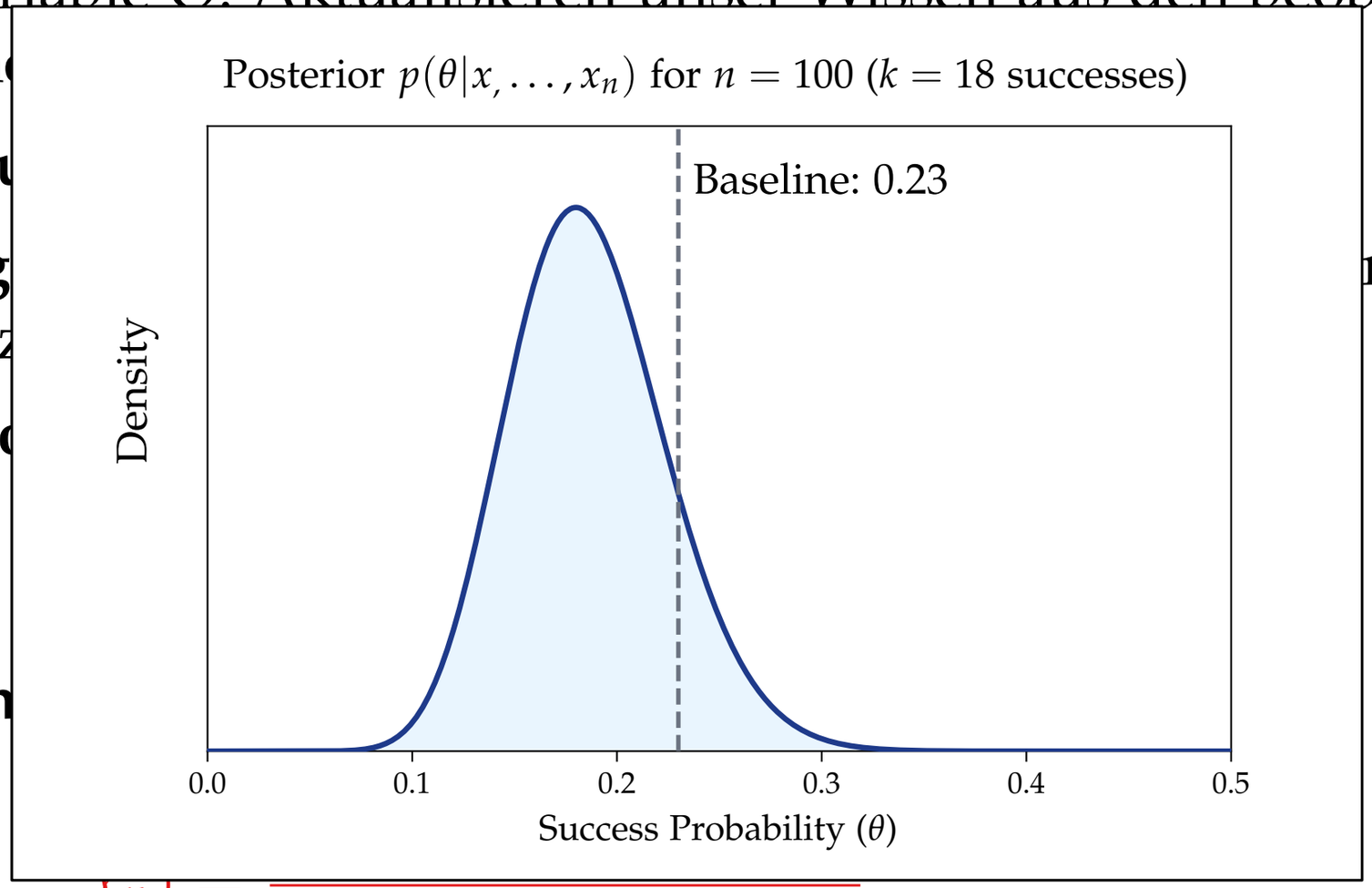
$X \sim \text{Bernoulli}(\theta)$

3. Posteri

$p(\theta|\mathbf{X}) =$

4. Hypoth

$$p(\theta \mid x_1, \ldots, x_n) = \frac{}{\frac{(\sum_i^n x_i)! \cdot (n - \sum_i^n x_i)!}{(n+1)!}}$$



Bayesian Statistics

Idee: Modellieren Wissen über eine Eigenschaft einer Population als Zufallsvariable Θ. Aktualisieren unser Wissen aus den beobachteten Daten mittels Bayes' Th

1. Prior au

$\theta \sim \text{Uniform}(0,1)$

2. Dateng r θ

eine WK z

$X \sim \text{Bernoulli}(\theta)$

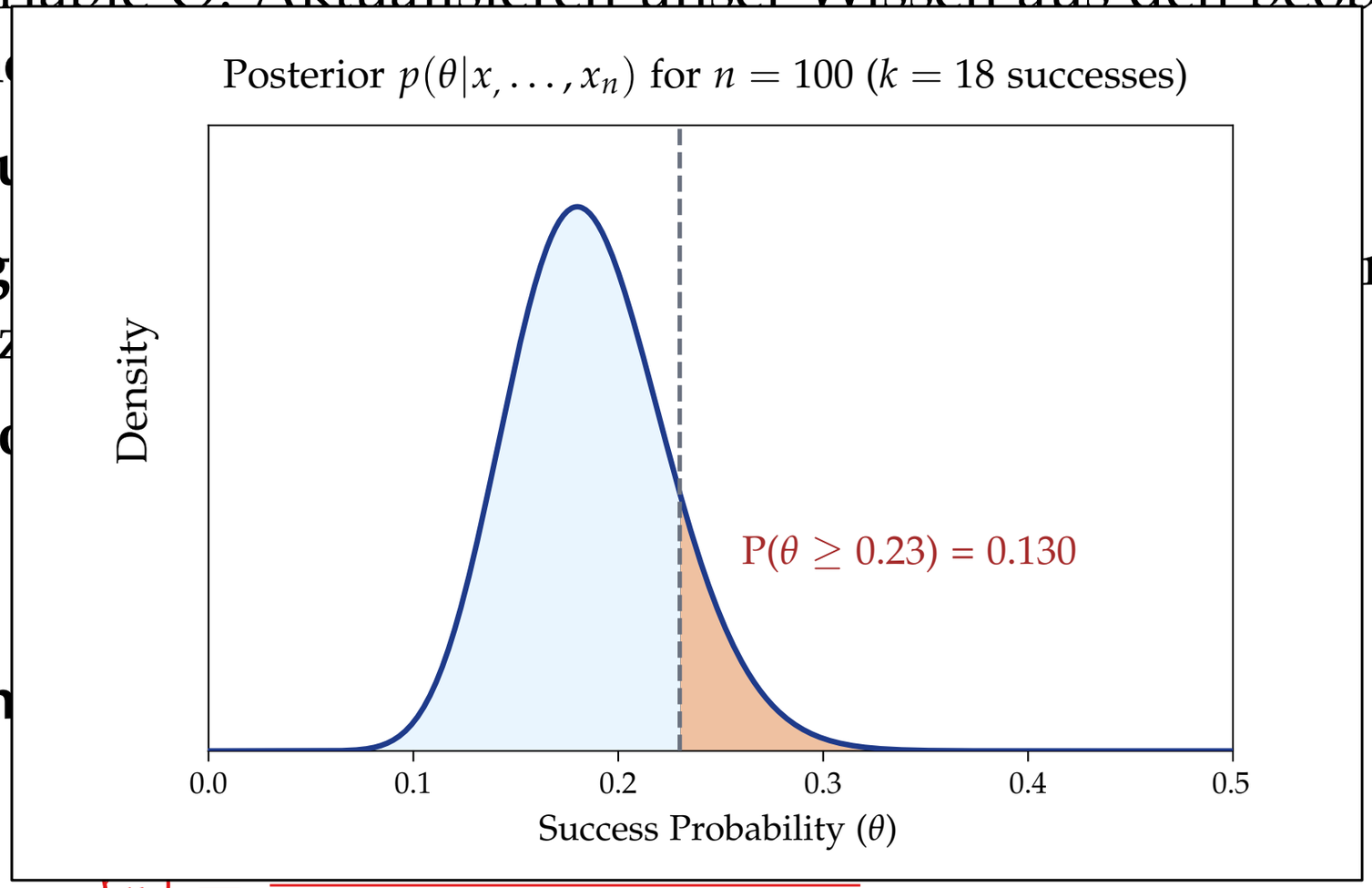
3. Posteri

$p(\theta|\mathbf{X}) =$

4. Hypoth

$p(\theta \mid x_1, \ldots, x_n) =$

$$\frac{\frac{(\sum_i^n x_i)! \cdot (n - \sum_i^n x_i)!}{(n+1)!}}{}$$



Bayesian Statistics

Idee: Modellieren Wissen über eine Eigenschaft einer Population als Zufallsvariable Θ. Aktualisieren unser Wissen aus den beobachteten Daten mittels Bayes' Th

1. Prior au

$\theta \sim \text{Uniform}(0,1)$

2. Dateng r θ

eine WK z

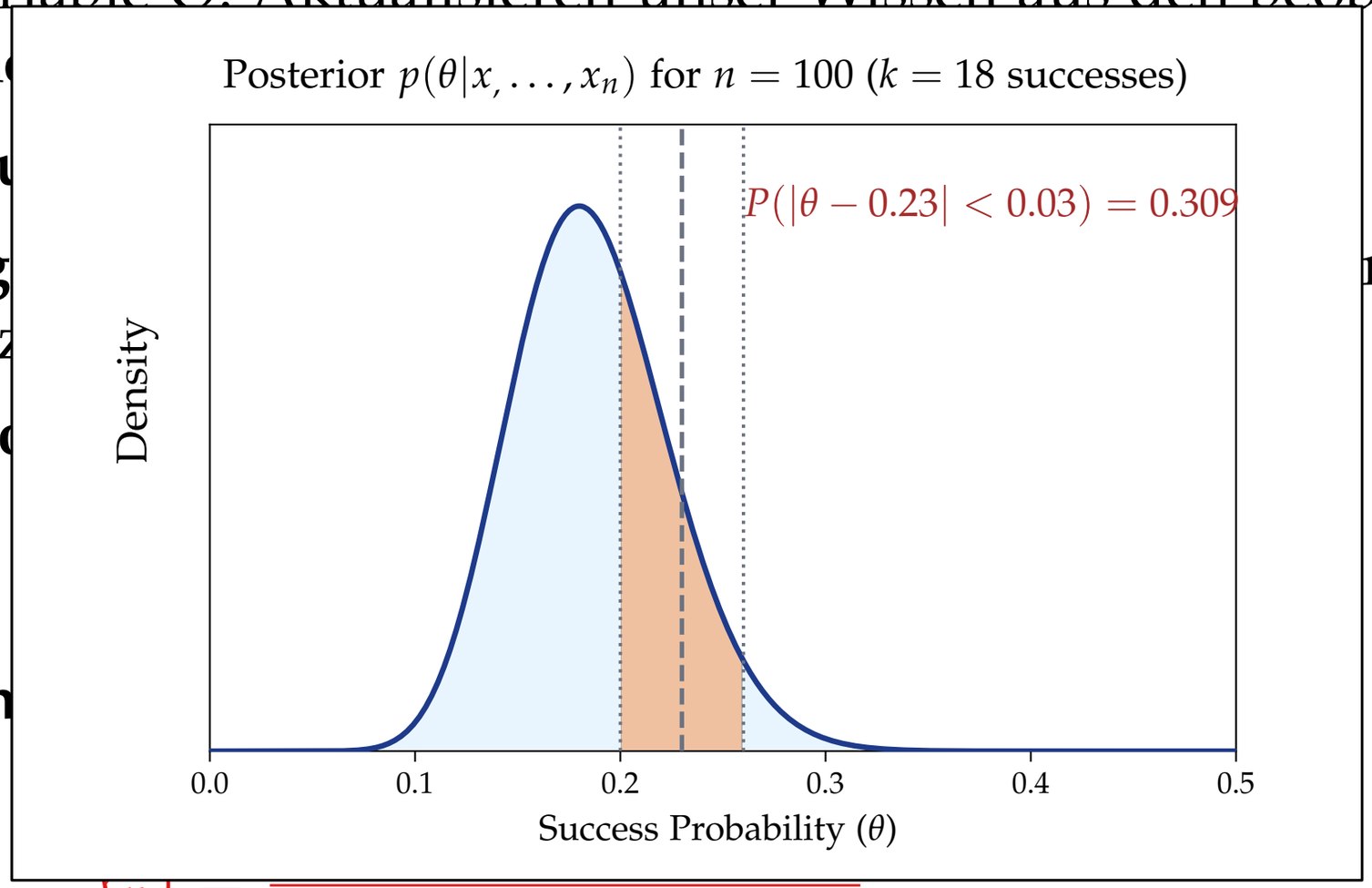
$X \sim \text{Bernoulli}(\theta)$

3. Posteri

$p(\theta|\mathbf{X}) =$

4. Hypoth

$$p(\theta \mid x_1, \ldots, x_n) = \frac{}{\frac{(\sum_i^n x_i)! \cdot (n - \sum_i^n x_i)!}{(n+1)!}}$$



Bayesian Statistics

Idee: Modellieren Wissen über eine Eigenschaft einer Population als Zufallsvariable Θ. Aktualisieren unser Wissen aus den beobachteten Daten mittels Bayes' Th

1. Prior au

$\theta \sim \text{Uniform}(0,1)$

2. Dateng r θ

eine WK z

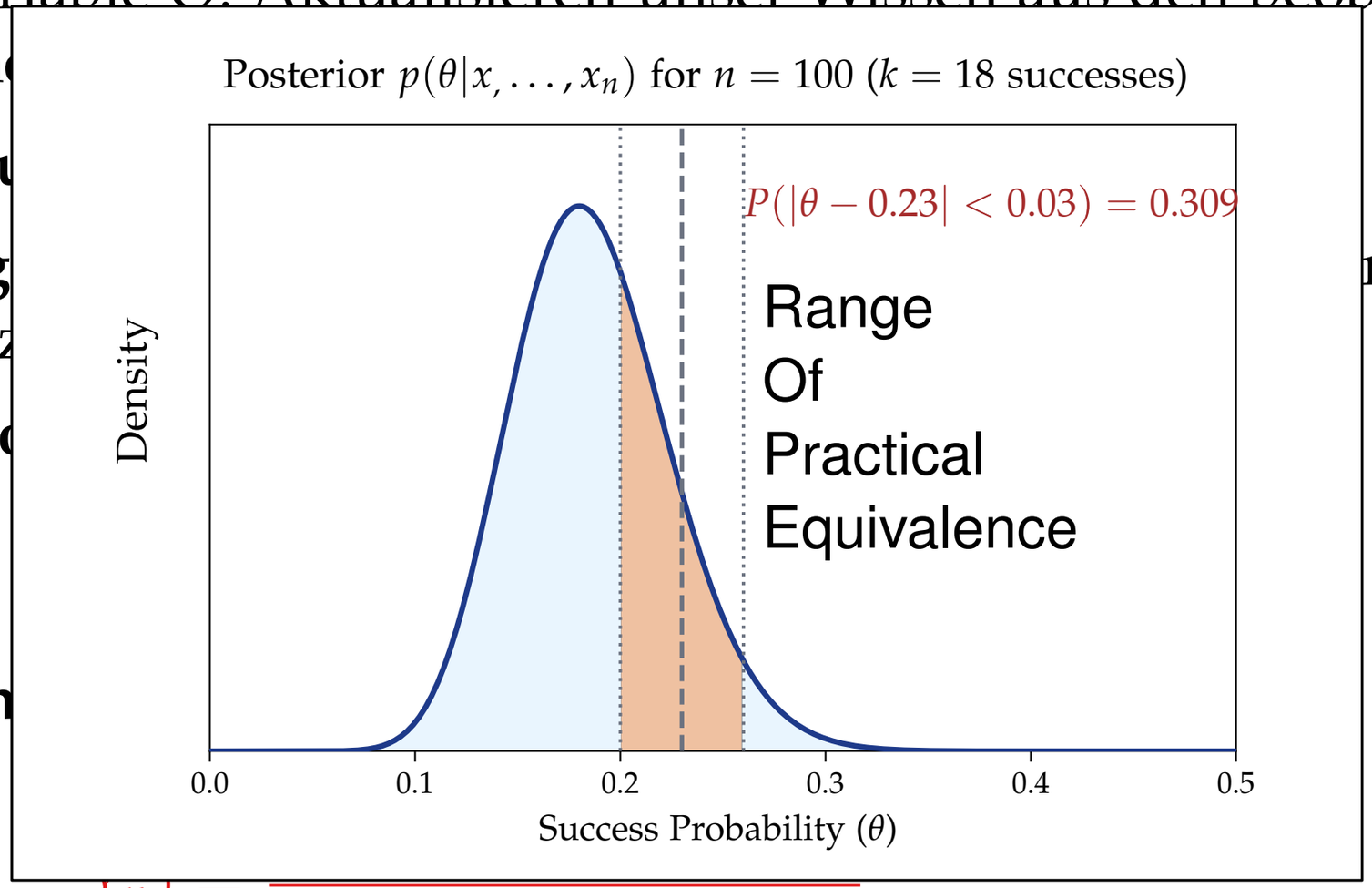
$X \sim \text{Bernoulli}(\theta)$

3. Posteri

$p(\theta|\mathbf{X}) =$

4. Hypoth

$p(\theta \mid x_1, \ldots, x_n) = \frac{}{\frac{(\sum_i^n x_i)! \cdot (n - \sum_i^n x_i)!}{(n+1)!}}$



Bayesian Statistics

Idee: Modellieren Wissen über eine Eigenschaft einer Population als Zufallsvariable Θ. Aktualisieren unser Wissen aus den beobachteten Daten mittels Bayes' Th

1. Prior au

$\theta \sim \text{Uniform}(0,1)$

2. Dateng r θ

eine WK z

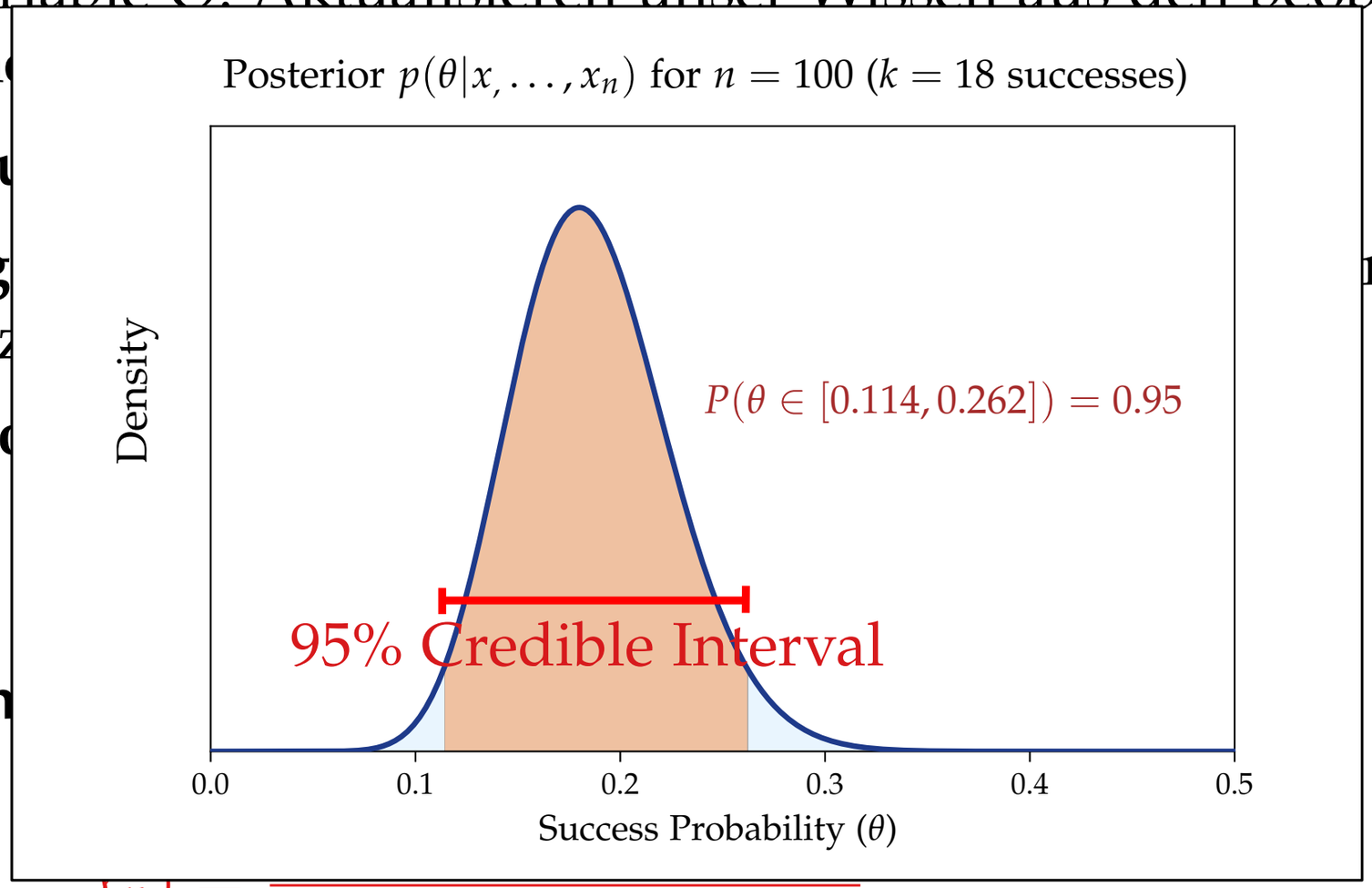
$X \sim \text{Bernoulli}(\theta)$

3. Posteri

$p(\theta|\mathbf{X}) =$

4. Hypoth

$p(\theta \mid x_1, \ldots, x_n) = \frac{}{\frac{(\sum_i^n x_i)! \cdot (n - \sum_i^n x_i)!}{(n+1)!}}$



Safari: Laufzeitvergleich à la T-Test

Modell/ Likelihood: $X_1, \ldots, X_n \sim \mathcal{N}(\mu_X, \sigma_X^2)$
$Y_1, \ldots, Y_n \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$

Relevanter Parameter: $\Delta = \mu_Y - \mu_X$

Priors: $\mu_X, \mu_Y \sim \text{Uniform}(0, 10^5)$;
$\sigma_X, \sigma_Y \sim \text{HalfCauchy}(10)$

Safari: Laufzeitvergleich à la T-Test

```
import numpy as np
import pymc as pm

group_a = [20, 22, 24, 20, 22, 18, 19]
group_b = [29, 23, 24, 19, 29, 24, 22]

with pm.Model() as two_sample_model:
    mu_a = pm.Uniform("mu_a", lower=0, upper=100)
    mu_b = pm.Uniform("mu_b", lower=0, upper=10_000)
    sigma_a = pm.HalfCauchy("sigma_a", beta=10)
    sigma_b = pm.HalfCauchy("sigma_b", beta=10)

    likelihood_a = pm.Normal("obs_a",
          mu=mu_a, sigma=sigma_a, observed=group_a)
    likelihood_b = pm.Normal("obs_b",
          mu=mu_b, sigma=sigma_b, observed=group_b)

    diff_means = pm.Deterministic("diff_means",
          mu_b - mu_a)
    trace = pm.sample()

print(az.summary(trace))
```

Modell/ Likelihood: $X_1, \ldots, X_n \sim \mathcal{N}(\mu_X, \sigma_X^2)$
$Y_1, \ldots, Y_n \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$

Relevanter Parameter: $\Delta = \mu_Y - \mu_X$

Priors: $\mu_X, \mu_Y \sim \text{Uniform}(0, 10^5)$;
$\sigma_X, \sigma_Y \sim \text{HalfCauchy}(10)$

Safari: Laufzeitvergleich à la T-Test

```
import numpy as np
import pymc as pm

group_a = [20, 22, 24, 20, 22, 18, 19]
group_b = [29, 23, 24, 19, 29, 24, 22]

with pm.Model() as two_sample_model:
    mu_a = pm.Uniform("mu_a", lower=0, upper=100)
    mu_b = pm.Uniform("mu_b", lower=0, upper=10_000)
    sigma_a = pm.HalfCauchy("sigma_a", beta=10)
    sigma_b = pm.HalfCauchy("sigma_b", beta=10)

    likelihood_a = pm.Normal("obs_a",
          mu=mu_a, sigma=sigma_a, observed=group_a)
    likelihood_b = pm.Normal("obs_b",
          mu=mu_b, sigma=sigma_b, observed=group_b)

    diff_means = pm.Deterministic("diff_means",
          mu_b - mu_a)
    trace = pm.sample()

print(az.summary(trace))
```

Modell/ Likelihood: $X_1, \ldots, X_n \sim \mathcal{N}(\mu_X, \sigma_X^2)$
$Y_1, \ldots, Y_n \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$

Relevanter Parameter: $\Delta = \mu_Y - \mu_X$

Priors: $\mu_X, \mu_Y \sim \text{Uniform}(0, 10^5)$;
$\sigma_X, \sigma_Y \sim \text{HalfCauchy}(10)$

Safari: Laufzeitvergleich à la T-Test

Modell/ Likelihood: $X_1, \ldots, X_n \sim \mathcal{N}(\mu_X, \sigma_X^2)$
$Y_1, \ldots, Y_n \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$

Relevanter Parameter: $\Delta = \mu_Y - \mu_X$

Priors: $\mu_X, \mu_Y \sim \text{Uniform}(0, 10^5)$;
$\sigma_X, \sigma_Y \sim \text{HalfCauchy}(10)$

```
import numpy as np
import pymc as pm

group_a = [20, 22, 24, 20, 22, 18, 19]
group_b = [29, 23, 24, 19, 29, 24, 22]

with pm.Model() as two_sample_model:
    mu_a = pm.Uniform("mu_a", lower=0, upper=100)
    mu_b = pm.Uniform("mu_b", lower=0, upper=10_000)
    sigma_a = pm.HalfCauchy("sigma_a", beta=10)
    sigma_b = pm.HalfCauchy("sigma_b", beta=10)

    likelihood_a = pm.Normal("obs_a",
          mu=mu_a, sigma=sigma_a, observed=group_a)
    likelihood_b = pm.Normal("obs_b",
          mu=mu_b, sigma=sigma_b, observed=group_b)

    diff_means = pm.Deterministic("diff_means",
          mu_b - mu_a)
    trace = pm.sample()

print(az.summary(trace))
```

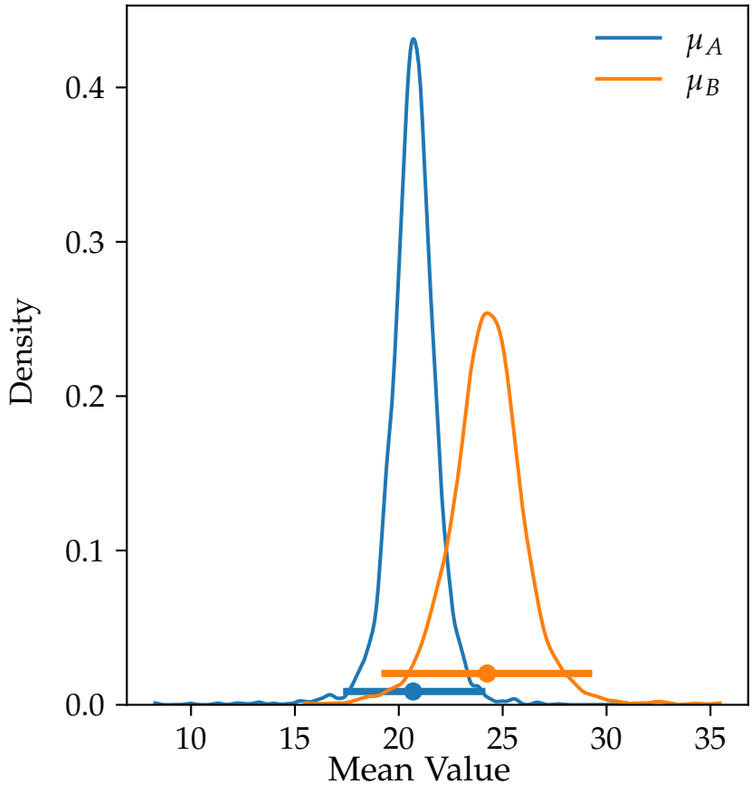
	mean	sd	eti89_lb	eti89_ub	ess_bul
mu_a	20.73	1.1	19	22	222
mu_b	24.28	1.88	21	27	235
sigma_a	2.68	1.2	1.5	4.6	196
sigma_b	4.71	1.9	2.7	8	203
diff_means	3.55	2.17	0.17	6.9	225

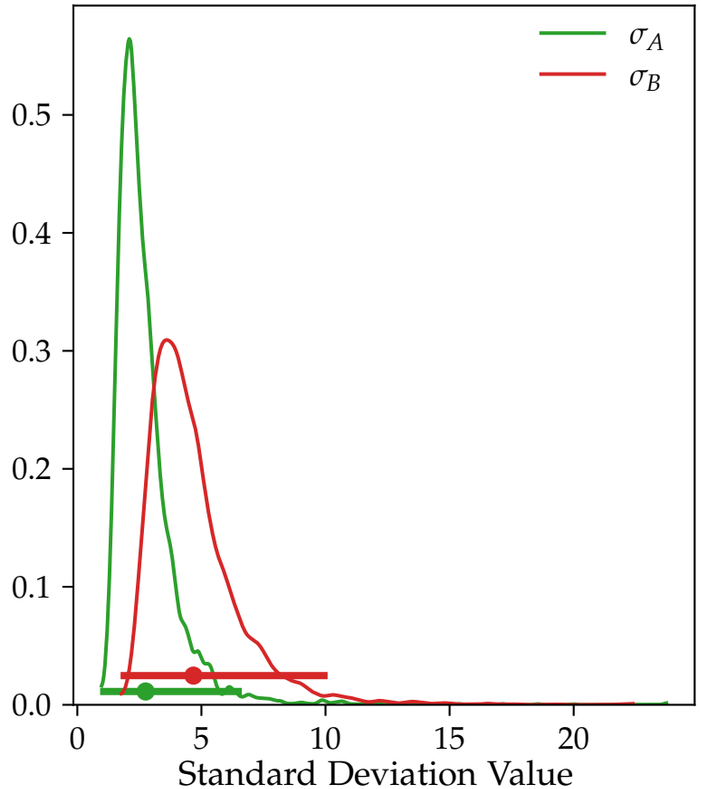
Safari: Laufzeitvergleich à la T-Test

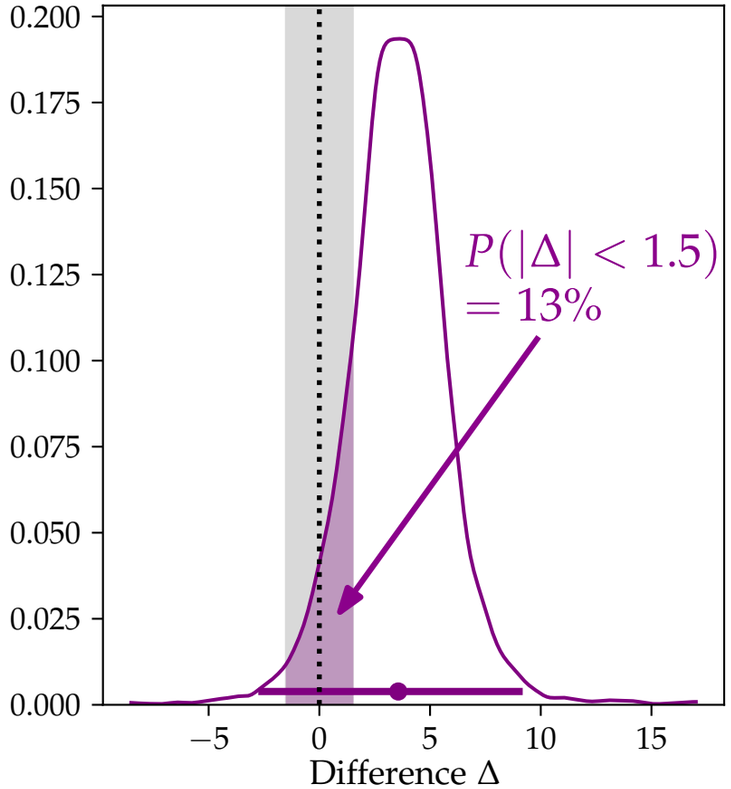
Modell/ Likelihood: $X_1, \ldots, X_n \sim \mathcal{N}(\mu_X, \sigma_X^2)$
$Y_1, \ldots, Y_n \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$

Relevanter Parameter: $\Delta = \mu_Y - \mu_X$

```
import numpy as np
import pymc as pm
```








```
    diff_means = pm.Deterministic("diff_means",
                                  mu_b - mu_a)
    trace = pm.sample()

print(az.summary(trace))
```

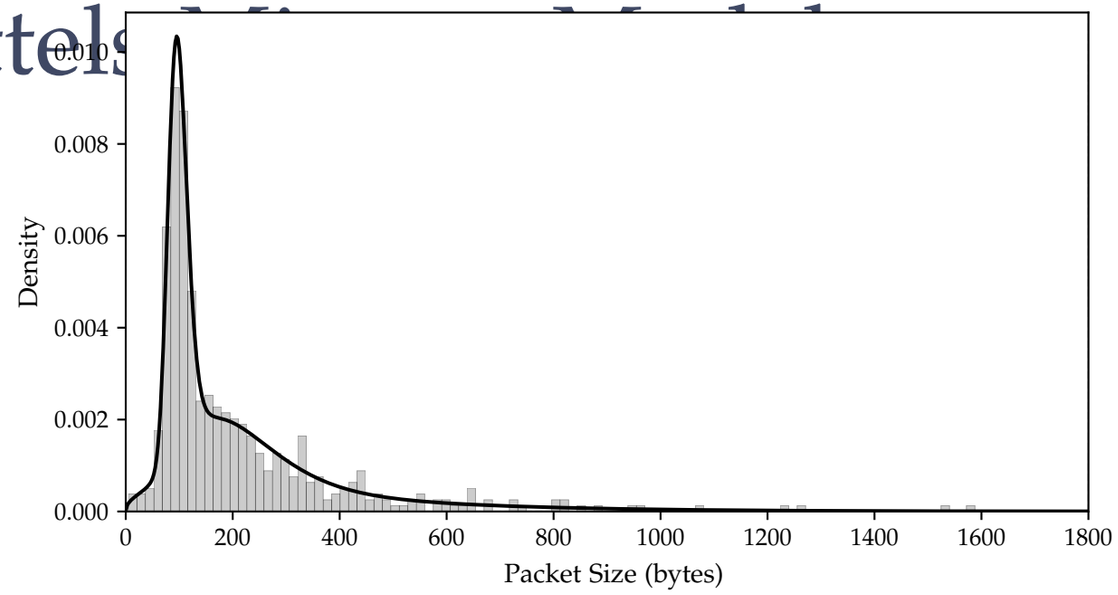
					ess_bul
mu_a	20.73	1.1	19	22	222
mu_b	24.28	1.88	21	27	235
sigma_a	2.68	1.2	1.5	4.6	196
sigma_b	4.71	1.9	2.7	8	203
diff_means	3.55	2.17	0.17	6.9	225

Safari: Traffic klassifizieren mittels Mixture Model

```
K = 3
with pm.Model() as mixture_model:
    # 1. Mixing weights
    w = pm.Dirichlet("w", a=np.ones(K), shape=K)

    # 2. Gamma Means (mu)
    mu = pm.LogNormal("mu",
        mu=np.log(100), sigma=2.0, shape=K,
        transform=pm.distributions.transforms.ordered,
        initval=np.log([50, 100, 200])
    )

    # 3. Scale (sigma)
    sigma = pm.HalfNormal("sigma", sigma=200, shape=K)

    # 4. Gamma Mixture
    y_obs = pm.Mixture("y_obs",
        w=w,
        comp_dists=pm.Gamma.dist(mu=mu, sigma=sigma),
        observed=y,
    )

    trace = pm.sample()
```

Safari: Traffic klassifizieren mittels

```
K = 3
with pm.Model() as mixture_model:
    # 1. Mixing weights
    w = pm.Dirichlet("w", a=np.ones(K), shape=K)

    # 2. Gamma Means (mu)
    mu = pm.LogNormal("mu",
        mu=np.log(100), sigma=2.0, shape=K,
        transform=pm.distributions.transforms.ordered,
        initval=np.log([50, 100, 200])
    )

    # 3. Scale (sigma)
    sigma = pm.HalfNormal("sigma", sigma=200, shape=K)

    # 4. Gamma Mixture
    y_obs = pm.Mixture("y_obs",
        w=w,
        comp_dists=pm.Gamma.dist(mu=mu, sigma=sigma),
        observed=y,
    )

    trace = pm.sample()
```



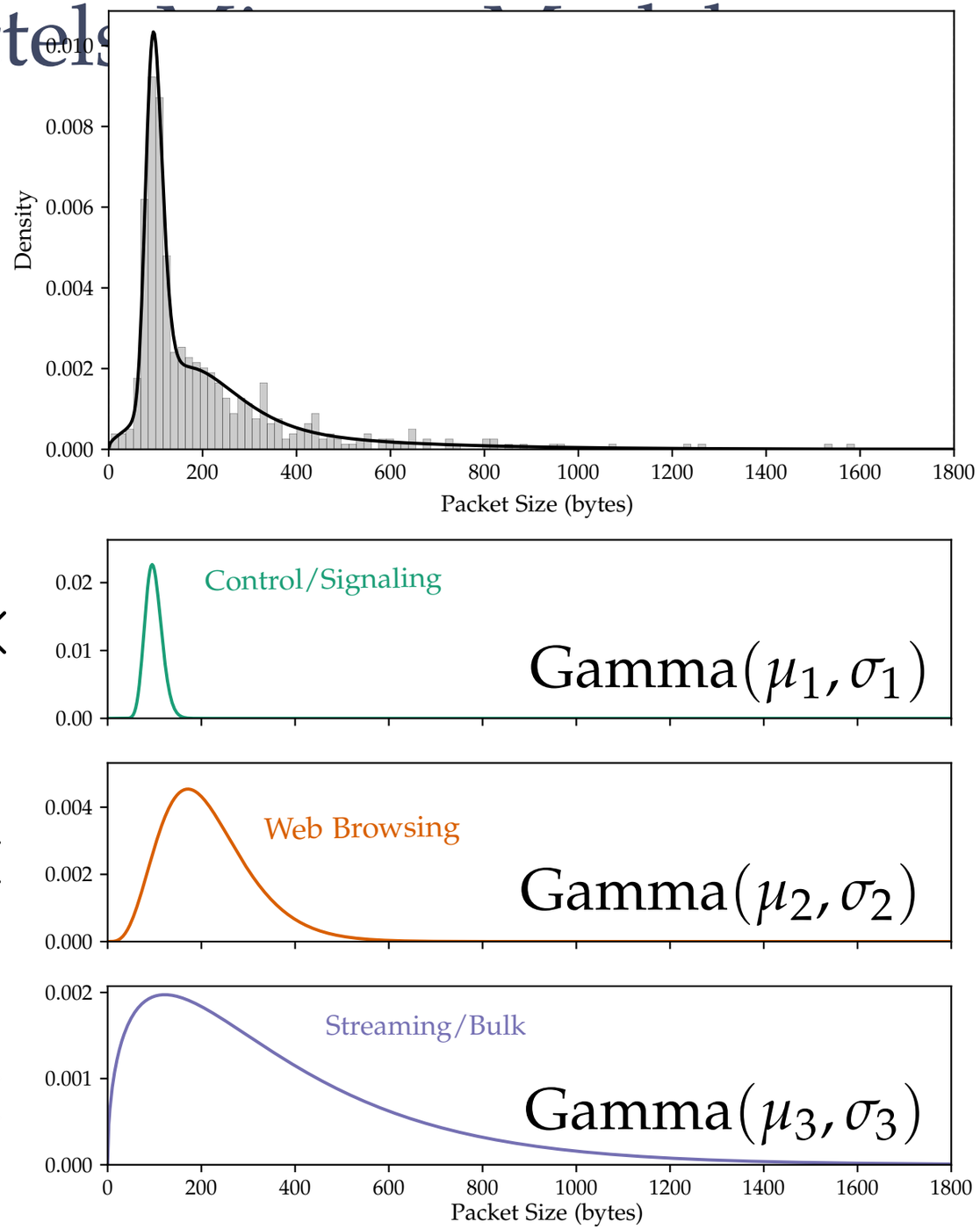
Safari: Traffic klassifizieren mittels

```
K = 3
with pm.Model() as mixture_model:
    # 1. Mixing weights
    w = pm.Dirichlet("w", a=np.ones(K), shape=K)

    # 2. Gamma Means (mu)
    mu = pm.LogNormal("mu",
        mu=np.log(100), sigma=2.0, shape=K,
        transform=pm.distributions.transforms.ordered,
        initval=np.log([50, 100, 200])
    )

    # 3. Scale (sigma)
    sigma = pm.HalfNormal("sigma", sigma=200, shape=K)

    # 4. Gamma Mixture
    y_obs = pm.Mixture("y_obs",
        w=w,
        comp_dists=pm.Gamma.dist(mu=mu, sigma=sigma),
        observed=y,
    )

    trace = pm.sample()
```

$$\pi_1 \times$$

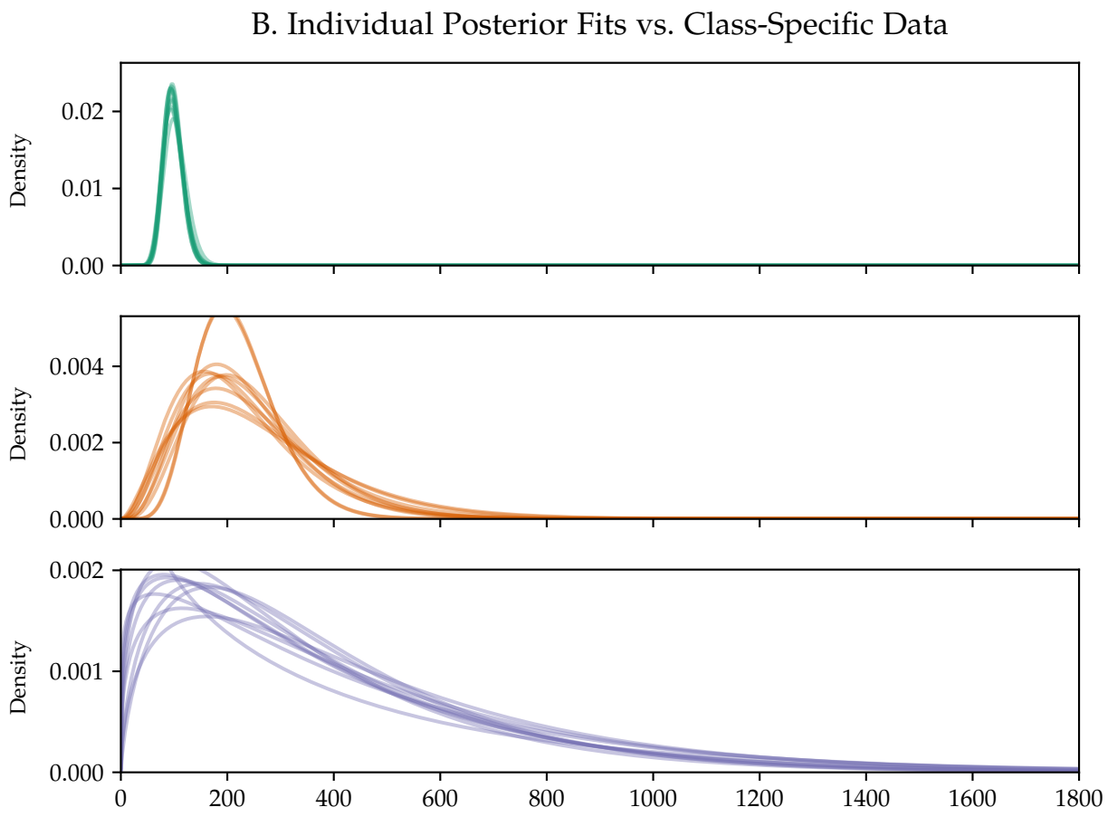
$$\pi_2 \times$$

$$\pi_3 \times$$



Safari: Traffic klassifizieren mittels Mixture Model

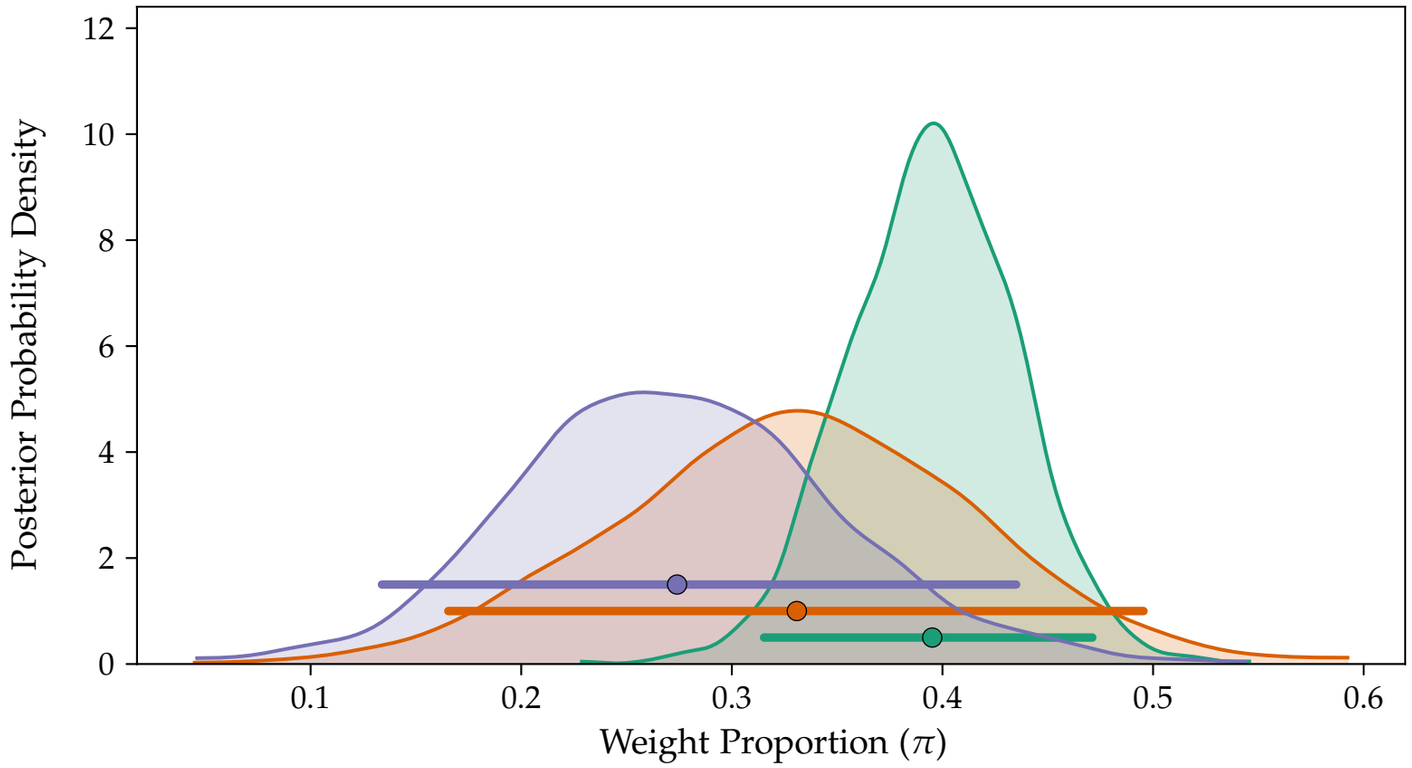
```
K = 3
with pm.Model() as mixture_model:
    # 1. Mixing weights
    w = pm.Dirichlet("w", a=np.ones(K), shape=K)
```

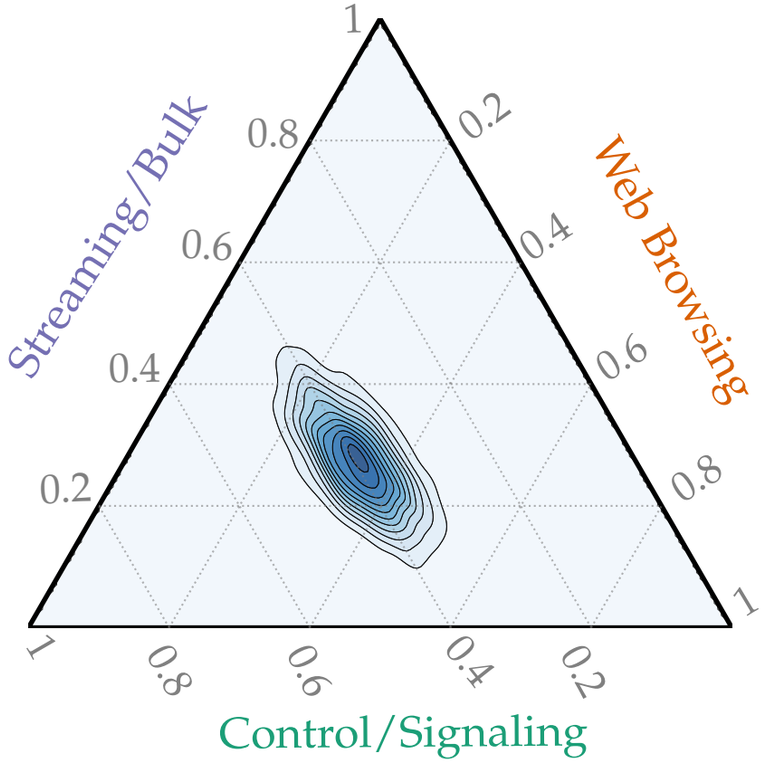



```
    trace = pm.sample()
```

Safari: Traffic klassifizieren mittels Mixture Model

```
K = 3
with pm.Model() as mixture_model:
    # 1. Mixing weights
    w = pm.Dirichlet("w", a=np.ones(K), shape=K)
```

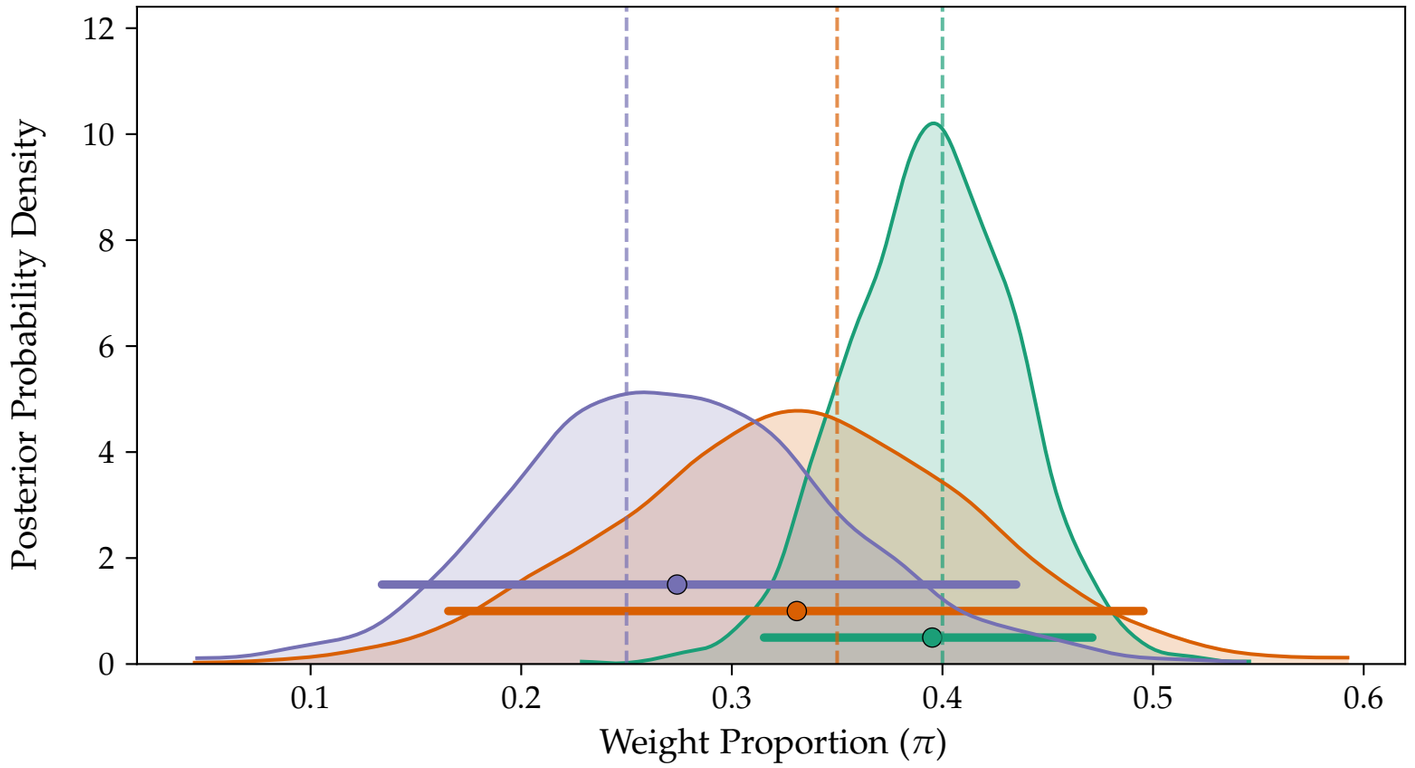


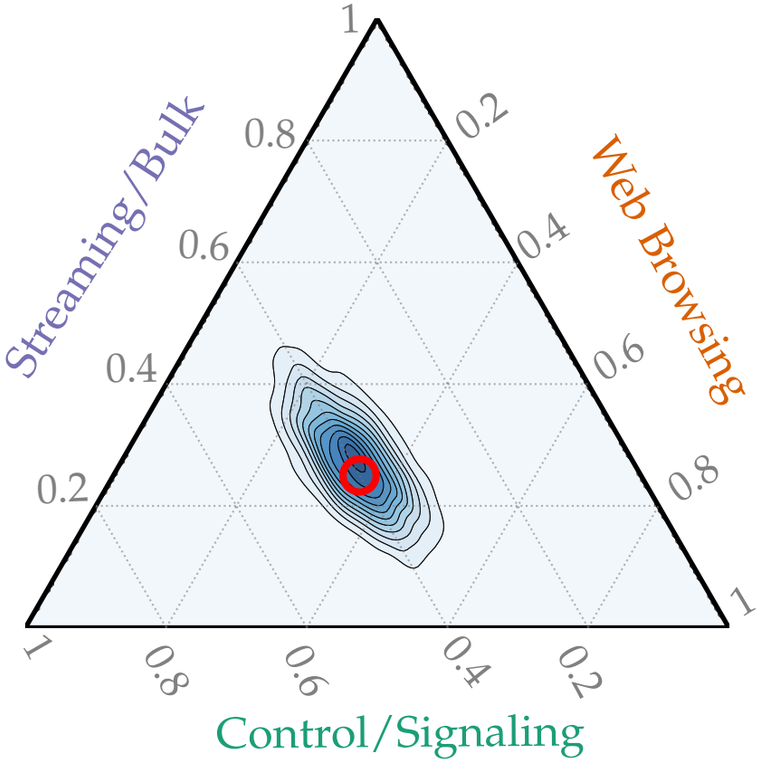



```
    trace = pm.sample()
```

Safari: Traffic klassifizieren mittels Mixture Model

```
K = 3
with pm.Model() as mixture_model:
    # 1. Mixing weights
    w = pm.Dirichlet("w", a=np.ones(K), shape=K)
```

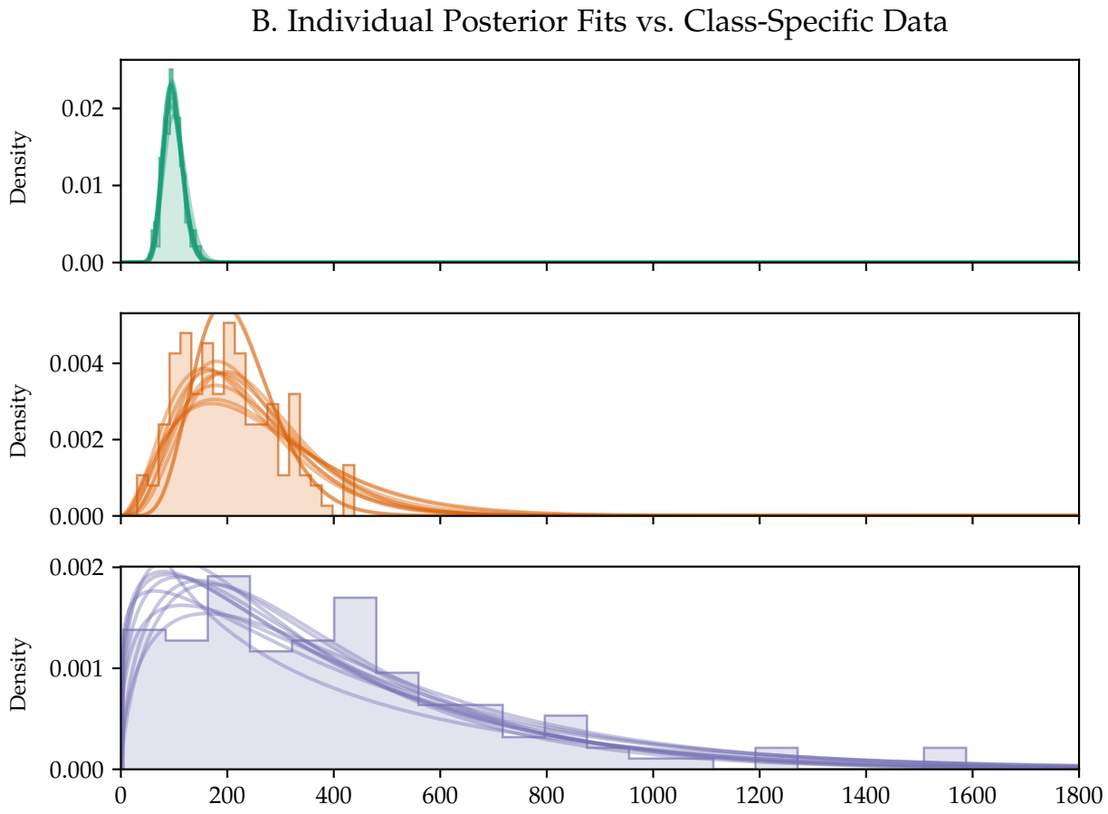


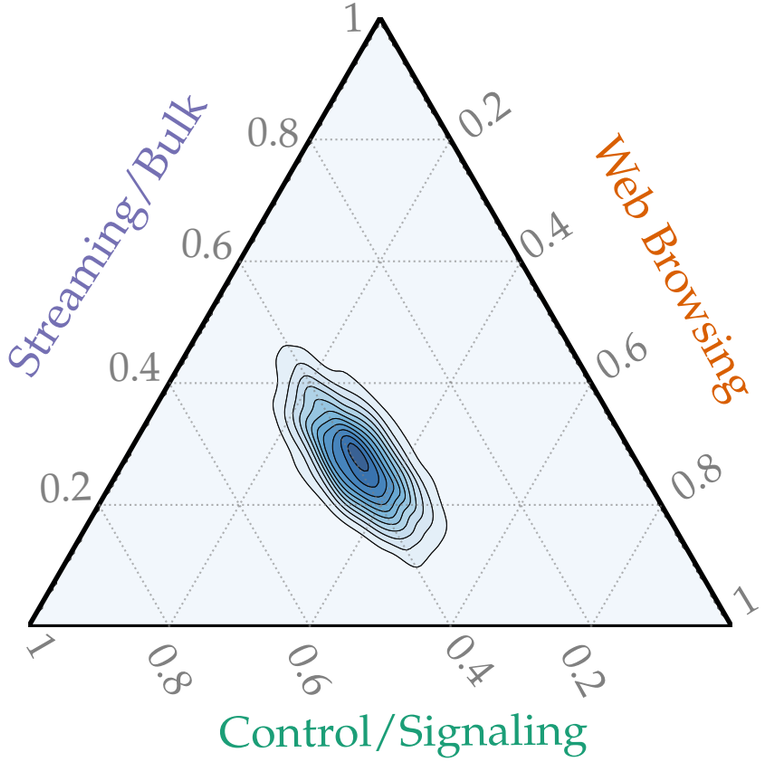



```
    trace = pm.sample()
```

Safari: Traffic klassifizieren mittels Mixture Model

```
K = 3
with pm.Model() as mixture_model:
    # 1. Mixing weights
    w = pm.Dirichlet("w", a=np.ones(K), shape=K)
```






```
trace = pm.sample()
```

Zum Weiterlesen/-schauen