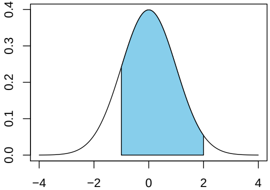
Hypothesentests

Statistik I

Sommersemester 2009


0.0
0.1
0.2
0.3
0.4
−4
−2
0
2
4

Zum Nachlesen

- Agresti/Finlay: Kapitel 6+7
- Gehring/Weins: Kapitel 12
- (Schumann: Kapitel 7.3)

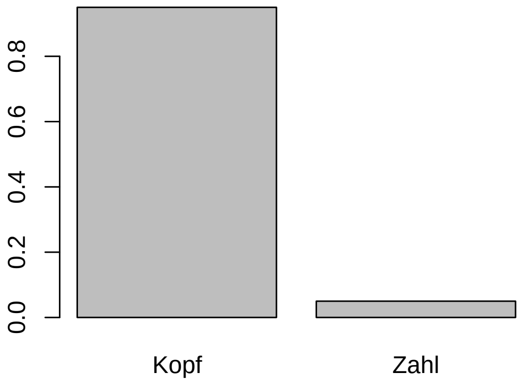
Gefälschte Münze $\longrightarrow$ Normalverteilung?

- Zentraler Grenzwertsatz:
  1. Komplexe Zufallsvariablen (z. B. Stichprobenmittelwerte),
  2. Die additiv aus vielen (> 30) einfachen Zufallsvariablen zusammengesetzt sind
  3. Verteilen sich über eine große Zahl von Wiederholungen normal
  4. Unabhängig davon, wie die einfachen Zufallsvariablen verteilt sind

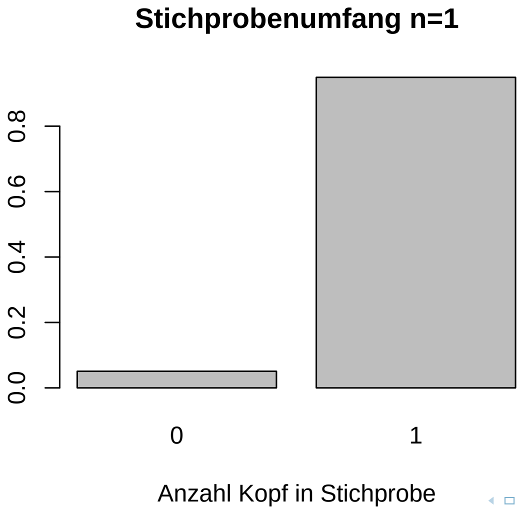
Gefälschte Münze $\longrightarrow$ Normalverteilung?

- Zentraler Grenzwertsatz:
  1. Komplexe Zufallsvariablen (z. B. Stichprobenmittelwerte),
  2. Die additiv aus vielen (> 30) einfachen Zufallsvariablen zusammengesetzt sind
  3. Verteilen sich über eine große Zahl von Wiederholungen normal
  4. Unabhängig davon, wie die einfachen Zufallsvariablen verteilt sind
- Eine Serie von Münzen ist gefälscht
- D. h. $p_{Kopf} = 0.95$
- „Grundgesamtheit" $=$ theoretische Verteilung
- „Stichprobenziehung" mit Umfang $n \longrightarrow n$ Münzen gleichzeitig werfen, Anzahl Kopf auszählen

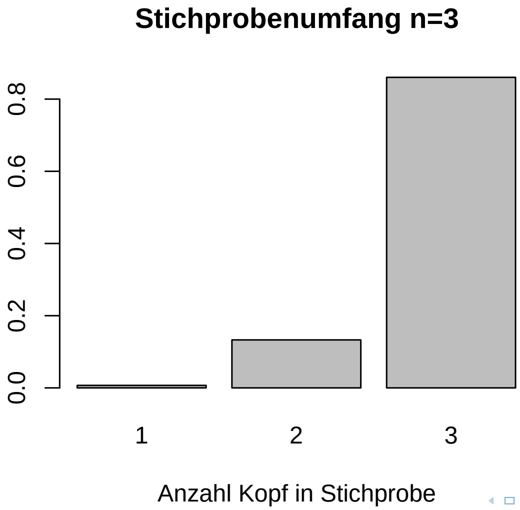
Theoretische Verteilung (GG)



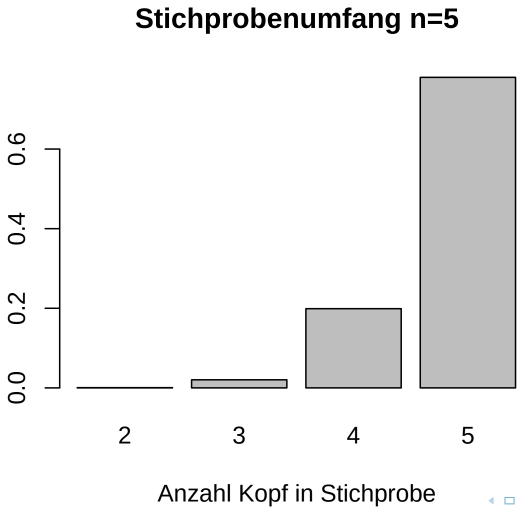
10 000 Wiederholungen



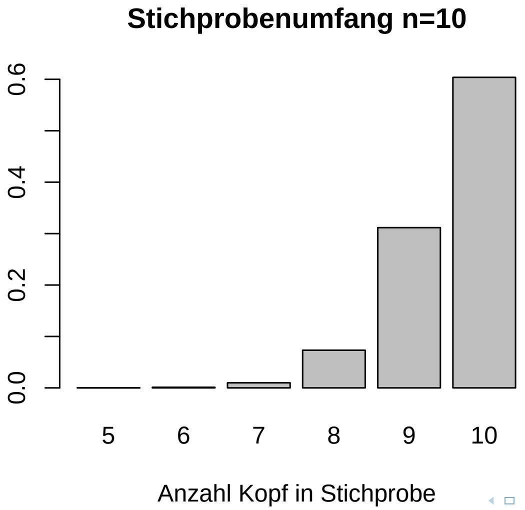
10 000 Wiederholungen



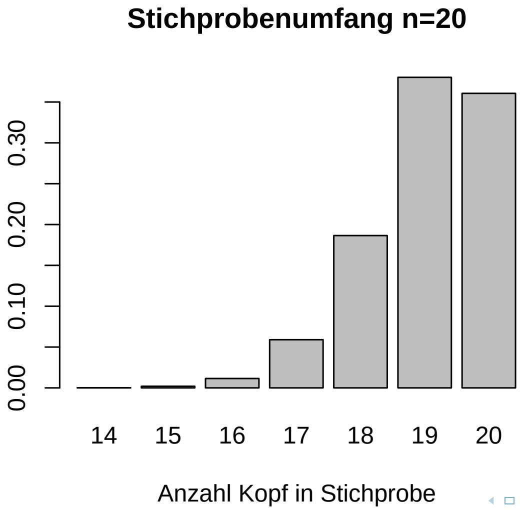
10 000 Wiederholungen



10 000 Wiederholungen



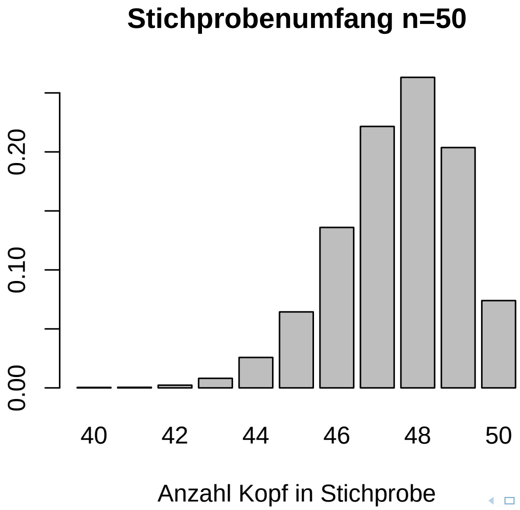
10 000 Wiederholungen



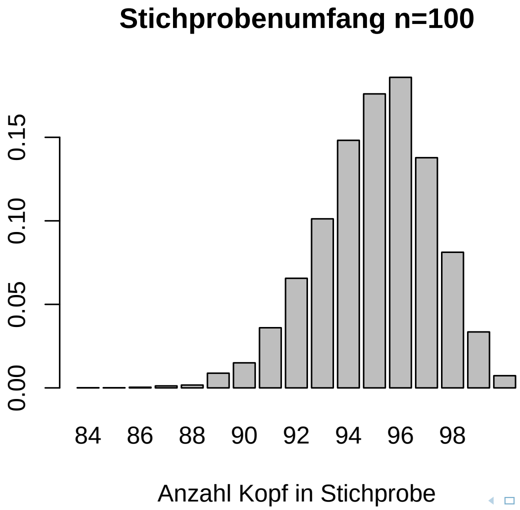
Was passiert hier?

- „Kopf" sehr viel wahrscheinlicher ($\frac{0.95}{0.05} = 19$) als „Zahl"
- Wahrscheinlichkeit, in einer „Stichprobe" vom Umfang n ausschließlich „Kopf" zu erhalten: 0.95^n
- Warum? $\rightarrow$ einzelne Münzen sind unabhängig voneinander (konditionale = marginale Wahrscheinlichkeit) $\rightarrow$ Einzelwahrscheinlichkeiten multiplizieren
  - $n = 2$ $\quad 0.95 \times 0.95 \approx 0.90$
  - $n = 4$ $\quad 0.95 \times 0.95 \times 0.95 \times 0.95 \approx 0.81$
  - $n = 10$ $\quad 0.95^{10} \approx 0.6$
  - $n = 20$ $\quad 0.95^{20} \approx 0.36$
- Für große n werden Werte im mittleren Bereich wahrscheinlicher als ausschließlich „Kopf"

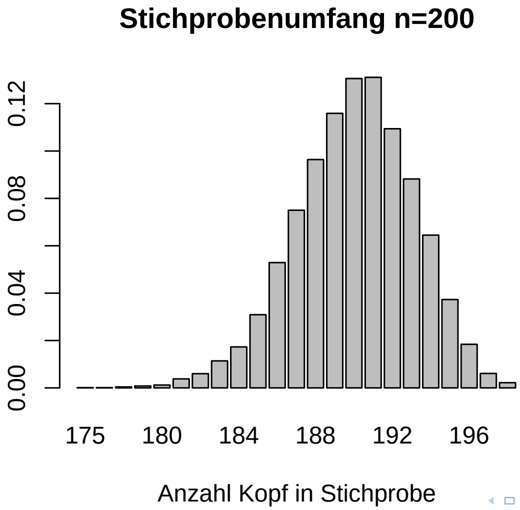
10 000 Wiederholungen



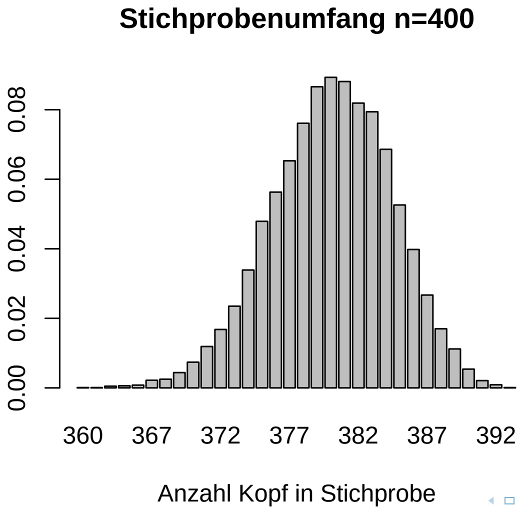
10 000 Wiederholungen



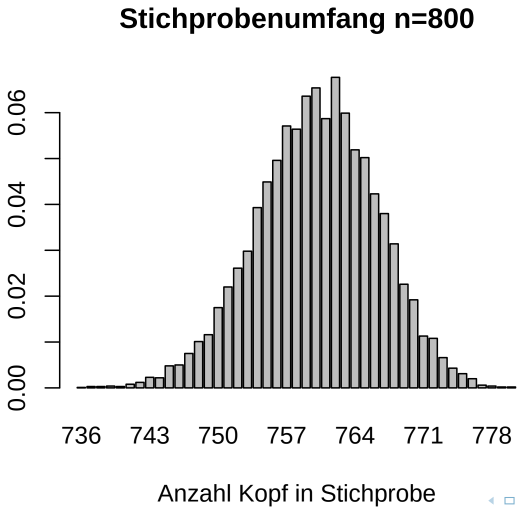
10 000 Wiederholungen



10 000 Wiederholungen



10 000 Wiederholungen



(Binomialformel)

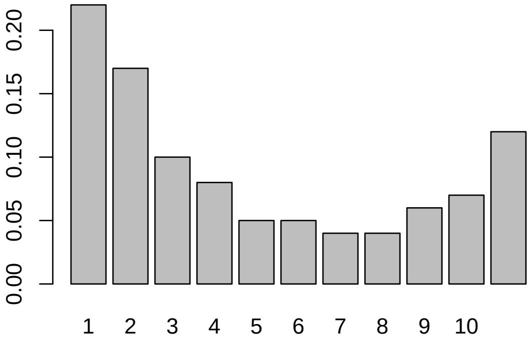
- Exakte Verteilung = Binomialverteilung (für Zusammenwirken binärer Zufallsvariablen)
- Z. B. Wahrscheinlichkeit, bei einer Stichprobengröße von $n = 800$ 760 mal Kopf zu erhalten

$$\begin{aligned} \binom{800}{760} &\times 0.95^{760} \times 0.05^{40} = \\ \frac{800 \times 799 \times \cdots \times 761}{760 \times 759 \times \cdots \times 1} &\times 0.95^{760} \times 0.05^{40} \\ &\approx 0.065 \end{aligned}$$

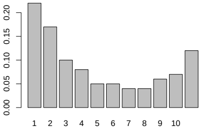
Ziehung aus hypothetischer Grundgesamtheit

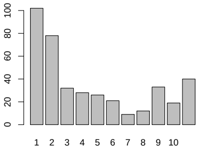
- Links-Rechts-Wert in hypothetischer Grundgesamtheit
- Werte zwischen 1 (links) und 11 (rechts)
- Verteilung linkssteil/rechtssschief
- Mittelwert 4.89
- Varianz 13.12

Verteilung in GG

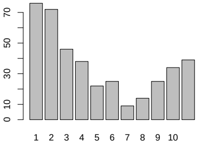


GG und drei Stichproben (n=400)

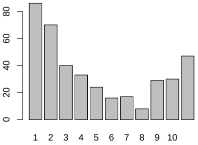




$\bar{x} = 4.52$



$\bar{x} = 4.85$



$\bar{x} = 4.89$

Grenzwertsatz reloaded

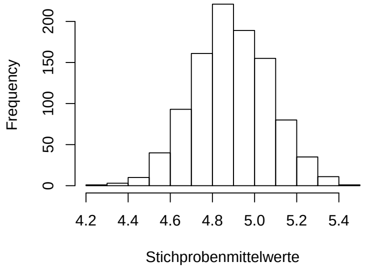
- Was erwarten wir für die Verteilung der Mittelwerte aus vielen (z. B. 1 000) Stichproben?

Grenzwertsatz reloaded

- Was erwarten wir für die Verteilung der Mittelwerte aus vielen (z. B. 1 000) Stichproben?
- Normalverteilung
- Um den wahren Mittelwert von 4.89
- Mit einem theoretischen Standardfehler (Standardabweichung) von:

$$\hat{\sigma}_{\bar{x}} = \sqrt{\frac{13.12}{400}} = 0.181$$

Verteilung Stichprobenmittelwerte



- Mittelwert der Stichprobenmittelwerte: 4.89
- Standardabweichung der Stichprobenmittelwerte: 0.182

Wie kommt man vom Grenzwertsatz zum Konfidenzintervall? I

- Normalverteilung der Stichprobenmittelwerte um wahren Mittelwert in der GG
- Bekannte Eigenschaften der Normalverteilung
  - Symmetrisch
  - 90% der Fläche (=Wahrscheinlichkeit) ± 1.64 Standardabweichungen vom Mittelwert
  - 95% der Fläche (=Wahrscheinlichkeit) ± 1.96 Standardabweichungen vom Mittelwert
  - 99% der Fläche (=Wahrscheinlichkeit) ± 2.58 Standardabweichungen vom Mittelwert
  - Weitere Integrale aus Tabelle ablesbar

Wie kommt man vom Grenzwertsatz zum Konfidenzintervall? II

- 95% aller *Stichproben* nicht mehr als 1.96 Standardabweichungen (=Standardfehler) vom wahren Mittelwert entfernt
- Standardfehler: $\hat{\sigma}_{\bar{x}} = \sqrt{\frac{\sigma^2}{n}}$ (schätzen)
- In 95% der Stichproben schließt Intervall von 1.96 Standardfehlern um *Stichprobenmittelwert* wahren Mittelwert in GG mit ein
- Wenn Varianz in GG nicht bekannt, Schätzung $\rightarrow$ t-Verteilung (mehr extreme Werte)
- Für große Fallzahlen geht t-Verteilung in Normalverteilung über

Was heißt Hypothesenprüfung?

- Zweites wichtiges Teilgebiet der Inferenzstatistik
- Wie lassen sich
  1. Mit Stichproben
  2. Annahmen/Hypothesen über eine (normalerweise unbekannte) Grundgesamtheit überprüfen?
- Wichtige Testverfahren u. a.
  - Mittelwertunterschiede bei bekannter Streuung (z-Test)
  - Mittelwertunterschiede für unabhängige/abhängige Stichproben (t-Test)
  - χ^2-Test auf Unabhängigkeit
- Test „Nullhypothese“ vs. „Alternativhypothese“
- Zahlreiche weitere Testverfahren mit identischer Logik

Was ist eine „Alternativhypothese“?

- Idealerweise neue, inhaltlich interessante, theoretische begründete Hypothese über Verhältnisse in GG
  - Gerichtet oder ungerichtet
  - Spezifisch oder unspezifisch
- Beispiele
  - „Frauen verdienen im Mittel schlechter als Männer“

Was ist eine „Alternativhypothese“?

- Idealerweise neue, inhaltlich interessante, theoretische begründete Hypothese über Verhältnisse in GG
  - Gerichtet oder ungerichtet
  - Spezifisch oder unspezifisch
- Beispiele
  - „Frauen verdienen im Mittel schlechter als Männer“ (gerichtet, unspezifisch)

Was ist eine „Alternativhypothese“?

- Idealerweise neue, inhaltlich interessante, theoretische begründete Hypothese über Verhältnisse in GG
  - Gerichtet oder ungerichtet
  - Spezifisch oder unspezifisch
- Beispiele
  - „Frauen verdienen im Mittel schlechter als Männer“ (gerichtet, unspezifisch)
  - „Es besteht ein Zusammenhang zwischen rechtsextremen Einstellungen und Wahl der NPD“

Was ist eine „Alternativhypothese“?

- Idealerweise neue, inhaltlich interessante, theoretische begründete Hypothese über Verhältnisse in GG
  - Gerichtet oder ungerichtet
  - Spezifisch oder unspezifisch
- Beispiele
  - „Frauen verdienen im Mittel schlechter als Männer“ (gerichtet, unspezifisch)
  - „Es besteht ein Zusammenhang zwischen rechtsextremen Einstellungen und Wahl der NPD“ (ungerichtet, unspezifisch)

Was ist eine „Alternativhypothese“?

- Idealerweise neue, inhaltlich interessante, theoretische begründete Hypothese über Verhältnisse in GG
  - Gerichtet oder ungerichtet
  - Spezifisch oder unspezifisch
- Beispiele
  - „Frauen verdienen im Mittel schlechter als Männer“ (gerichtet, unspezifisch)
  - „Es besteht ein Zusammenhang zwischen rechtsextremen Einstellungen und Wahl der NPD“ (ungerichtet, unspezifisch)
- Alternativhypothese wird mit $\mathbf{H}_A$ bezeichnet

Was ist eine „Nullhypothese"?

- **$\mathbf{H}_0$** Gegenstück zur Alternativhypothese, negiert diese
- „Frauen verdienen genauso viel oder mehr als Männer"
- „Zwischen rechtsextremen Einstellungen und der Wahl der NPD besteht kein Zusammenhang"
- Idealerweise steht H_0 für den etablierten Forschungsstand; H_0 als Grundannahme
- H_0 und H_A schließen sich gegenseitig aus und beschreiben Möglichkeitsraum vollständig, d. h.
- In der GG gilt entweder H_0 oder (exklusiv) H_A

Was ist das „Entscheidungsdilemma“ beim Hypothesentest?

		In der *GG* gilt ...	
		H_A	H_0
Entscheidung auf Grundlage Stichprobe	H_A	richtig	α-Fehler
	H_0	β-Fehler	richtig

$\beta \neq 1 - \alpha$!

Was ist das „Entscheidungsdilemma“ beim Hypothesentest?

		In der *GG* gilt …	
		H_A	H_0
Entscheidung auf Grundlage Stichprobe	H_A	richtig	α-Fehler
	H_0	β-Fehler	richtig

$\beta \neq 1 - \alpha$!

Was ist die „Irrtumswahrscheinlichkeit“?

- Wahrscheinlichkeit, einen α-Fehler (Fehler 1. Ordnung) zu begehen
- Wahrscheinlichkeit für β-Fehler (vs.) „Teststärke“ (Power) bleibt meistens unberücksichtigt
- Verfahren inhärent konservativ
- Irrtumswahrscheinlichkeit α wird vorab auf kleinen Wert festgelegt (1%, **5%**, 10%)

Was ist die „Irrtumswahrscheinlichkeit“?

- Wahrscheinlichkeit, einen α-Fehler (Fehler 1. Ordnung) zu begehen
- Wahrscheinlichkeit für β-Fehler (vs.) „Teststärke“ (Power) bleibt meistens unberücksichtigt
- Verfahren inhärent konservativ
- Irrtumswahrscheinlichkeit α wird vorab auf kleinen Wert festgelegt (1%, **5%**, 10%)

Definition α

- Wahrscheinlichkeit, daß H_0 zu unrecht aufgegeben wird
- Obwohl sie in GG gilt

Was ist ein „Signifikanztest“?

- Alle Tests prüfen statistische Signifikanz
- Voraussetzung: In GG gilt tatsächlich H_0
- Berechnung einer Prüfgröße (Teststatistik) auf Basis der Stichprobe
- Wenn H_0 in GG gilt, ist Prüfgröße
  - Realisierung einer Zufallsvariablen
  - Mit bekannter Verteilung
- Verteilung bekannt $\rightarrow$ „kritische Werte“
- Werden nur in 1/5/10% aller Stichproben überschritten, wenn H_0 in GG $\rightarrow$ „signifikant“
- Unterscheidung zwischen gerichteten und ungerichteten Hypothesen

Statistische vs. praktisch-politische Signifikanz

- Statistische Signifikanz: Ergebnis sehr unwahrscheinlich (1/5/10%), wenn H_0 in GG gilt
- Deshalb H_0 verwerfen
- Sehr unwahrscheinlich, dies zu unrecht zu tun

Statistische vs. praktisch-politische Signifikanz

- Statistische Signifikanz: Ergebnis sehr unwahrscheinlich (1/5/10%), wenn H_0 in GG gilt
- Deshalb H_0 verwerfen
- Sehr unwahrscheinlich, dies zu unrecht zu tun
- Aber: Plausibel, daß in GG *exakt gar kein* Zusammenhang besteht?

Statistische vs. praktisch-politische Signifikanz

- Statistische Signifikanz: Ergebnis sehr unwahrscheinlich (1/5/10%), wenn H_0 in GG gilt
- Deshalb H_0 verwerfen
- Sehr unwahrscheinlich, dies zu unrecht zu tun
- Aber: Plausibel, daß in GG *exakt gar kein* Zusammenhang besteht?
- Je größer Stichprobe, desto eher werden triviale Abweichungen von H_0 statistisch signifikant

Statistische vs. praktisch-politische Signifikanz

- Statistische Signifikanz: Ergebnis sehr unwahrscheinlich (1/5/10%), wenn H_0 in GG gilt
- Deshalb H_0 verwerfen
- Sehr unwahrscheinlich, dies zu unrecht zu tun
- Aber: Plausibel, daß in GG *exakt gar kein* Zusammenhang besteht?
- Je größer Stichprobe, desto eher werden triviale Abweichungen von H_0 statistisch signifikant
- Statistische Signifikanz vs. inhaltliche *Bedeutsamkeit*

Statistische vs. praktisch-politische Signifikanz

- Statistische Signifikanz: Ergebnis sehr unwahrscheinlich (1/5/10%), wenn H_0 in GG gilt
- Deshalb H_0 verwerfen
- Sehr unwahrscheinlich, dies zu unrecht zu tun
- Aber: Plausibel, daß in GG *exakt gar kein* Zusammenhang besteht?
- Je größer Stichprobe, desto eher werden triviale Abweichungen von H_0 statistisch signifikant
- Statistische Signifikanz vs. inhaltliche *Bedeutsamkeit*
- „Insignificance of significance testing"

z-Test

- Einfacher Test
- Frage:
  - „Kann eine Stichprobe mit einem gewissen Mittelwert
  - Tatsächlich aus einer *bekannten* GG mit *bekanntem Mittelwert/bekannter Streuung* stammen?“
- Relativ selten angewendet, aber
- Modell für komplexere Testverfahren
- Für große Fallzahlen t-Test $\rightarrow$ z-Test

Beispiel: Bildungsforschung

- Land X unterrichtet seit 1982 Mathematik nach festem Plan
- 10. Klasse: standardisierter Leistungstest $\longrightarrow$ GG bekannt:
  - Mittelwert $\mu = 40$ Punkte
  - Standardabweichung $\sigma = 8$ Punkte

Beispiel: Bildungsforschung

- Land X unterrichtet seit 1982 Mathematik nach festem Plan
- 10. Klasse: standardisierter Leistungstest $\longrightarrow$ GG bekannt:
  - Mittelwert $\mu = 40$ Punkte
  - Standardabweichung $\sigma = 8$ Punkte
- Neue, angeblich bessere Unterrichtsmethode, Test an $n = 100$ zufällig ausgewählten Schülern
  - H_A: „Die neue Methode ist besser als die alte“
  - H_0: „Neue Methode genauso gut oder schlechter“
  - Irrtumswahrscheinlichkeit α wird auf 5% festgelegt
  - *Einseitiger Test*, da Hypothese gerichtet

Beispiel: Bildungsforschung

- Land X unterrichtet seit 1982 Mathematik nach festem Plan
- 10. Klasse: standardisierter Leistungstest $\longrightarrow$ GG bekannt:
  - Mittelwert $\mu = 40$ Punkte
  - Standardabweichung $\sigma = 8$ Punkte
- Neue, angeblich bessere Unterrichtsmethode, Test an $n = 100$ zufällig ausgewählten Schülern
  - H_A: „Die neue Methode ist besser als die alte"
  - H_0: „Neue Methode genauso gut oder schlechter"
  - Irrtumswahrscheinlichkeit α wird auf 5% festgelegt
  - *Einseitiger Test*, da Hypothese gerichtet
  - Mittelwert $\bar{x} = 42$ Punkte – Zufall?
  - Bzw. wie groß ist die Wahrscheinlichkeit, dieses oder ein höheres Ergebnis zu erhalten wenn Stichprobe aus alter GG (μ, σ) stammt?

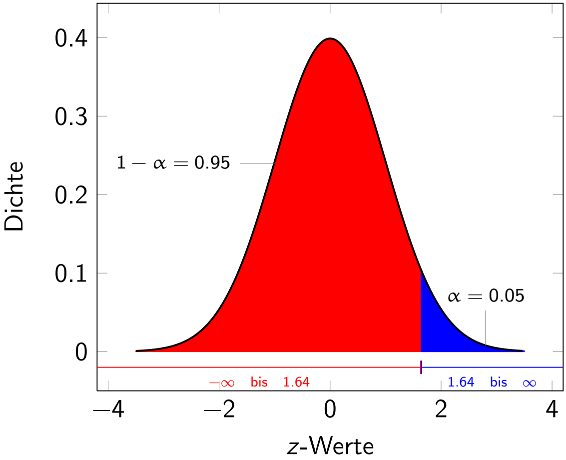
Vorgehensweise?

- *Richtung* der Abweichung von μ stimmt $\longrightarrow$ ok
- Wie wahrscheinlich ist *Größenordnung* der Abweichung?
- Hilfsannahme: Was wäre, wenn
  1. $\mu = 0$ und
  2. Wiederholte Stichprobenziehung und
  3. Stichprobenmittelwerte standard-normalverteilt mit $\sigma_{\bar{x}} = 1$?

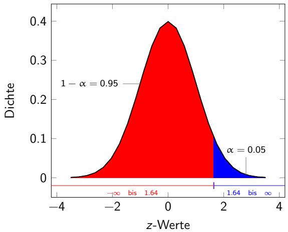
Vorgehensweise?

- *Richtung* der Abweichung von μ stimmt $\rightarrow$ ok
- Wie wahrscheinlich ist *Größenordnung* der Abweichung?
- Hilfsannahme: Was wäre, wenn
  1. $\mu = 0$ und
  2. Wiederholte Stichprobenziehung und
  3. Stichprobenmittelwerte standard-normalverteilt mit $\sigma_{\bar{x}} = 1$?
- In 95% der Stichproben $\bar{x} < 1.64$
- In 5% der Stichproben $\bar{x} \geqslant 1.64$

Standardnormalverteilung



Warum 1.64?



- *Einseitige* Hypothese $\rightarrow$ Ablehnungsbereich auf einer (positiver) Seite
- Welcher Wert trennt rechts 5% (α) der Standardnormalverteilung ab?

- Nachschauen in kompletter z-Tabelle (z. B. Gehring/Weins)
- Nachschauen in letzter Zeile der t-Tabelle
- Computer
- Gedächtnis

z-Tabelle

	.0	.1	.2	.3	.4	.5	.6	.7	.8	.9
0.0.	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1.	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2.	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3.	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4.	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5.	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6.	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7.	0.7580	0.7611	0.7642	0.7673	0.7703	0.7734	0.7764	0.7794	0.7823	0.7852
0.8.	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9.	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0.	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1.	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2.	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3.	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4.	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5.	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6.	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7.	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8.	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

t-Tabelle/Computer

	Fläche (1-α)									
df	0.65	0.7	0.75	0.8	0.85	0.9	0.95	0.975	0.99	0.995
1	0.510	0.727	1.000	1.376	1.963	3.078	6.314	12.706	31.821	63.656
2	0.445	0.617	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925
3	0.424	0.584	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841
4	0.414	0.569	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604
5	0.408	0.559	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032
6	0.404	0.553	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707
7	0.402	0.549	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499
8	0.399	0.546	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355
9	0.398	0.543	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
200	0.386	0.525	0.676	0.843	1.039	1.286	1.653	1.972	2.345	2.601
500	0.386	0.525	0.675	0.842	1.038	1.283	1.648	1.965	2.334	2.586
1000	0.385	0.525	0.675	0.842	1.037	1.282	1.646	1.962	2.330	2.581
z-Wert	0.385	0.524	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576

```
> qnorm(0.95)
[1] 1.644854
```

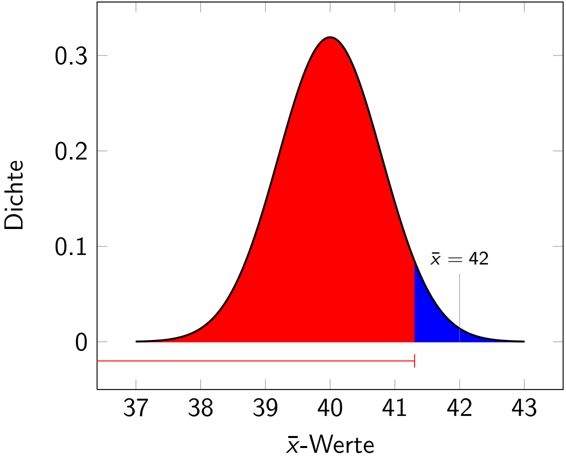
Bildungsforschung

- Wie weit ist Stichprobenmittelwert $\bar{x}$ von Mittelwert GG μ entfernt?
- Wie wahrscheinlich ist diese/größere Distanz wenn Stichprobe aus GG stammt?
- $\mu = 40 \neq 0 \rightarrow$ Differenz betrachten
- $\bar{x} - \mu = 42 - 40 = 2$
- Alle Normalverteilungen haben selbe Form (5% links von 1.64 etc.) $\rightarrow$ Differenz in Standardabweichungen umrechnen
- Standardabweichung (=Streuung) Stichprobenmittelwerte um μ = Standardfehler $\sigma_{\bar{x}}$

Standardfehler/Signifikanz Schulexperiment

- Standardfehler: $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{8}{\sqrt{100}} = 0.8$
- Differenz in Standardfehlern ausgedrückt: $\frac{2}{0.8} = 2.5 = z$
- Mehr als 1.64 $\rightarrow$ „signifikant auf 5%-Niveau“
- H_0 (Stichprobe kommt tatsächlich aus alter GG, d. h. neue Methode nicht besser) kann
- Mit geringer Irrtumswahrscheinlichkeit abgelehnt werden.

Stichprobenmittelwert innerhalb Stichprobenmittelwertverteilung



Wie groß ist die Irrtumswahrscheinlichkeit genau?

- z-Wert $=$ Quotient aus Mittelwertdifferenz (Effekt) und Standardfehler
- Standardnormalverteilte Teststatistik
- Wert von 2.5 in Tabellen/Computer suchen
- $\alpha = 0.006$

Wovon hängt die Irrtumswahrscheinlichkeit ab?

- Bei konstanten Verhältnissen in der GG (Mittelwert und Streuung) ...
- Von der Stärke des Effektes (hier Mittelwertdifferenz)
- Und vom Umfang der Stichprobe (über Standardfehler)
- Mit hinreichend großen Stichproben werden auch triviale Effekte „signifikant"
- Gängige Niveaus 1/5/10% sind *Konventionen*

Zusammenfassung

- Testverfahren ermöglichen es, mit Hilfe einer Stichprobe Hypothesen über eine GG zu überprüfen
- Wenn man ein gewisses Irrtumsrisiko in Kauf nimmt, das vorab festgelegt wird
- Logik: Testgröße berechnen (hier Mittelwertdifferenz/Standardfehler) und mit theoretischer Verteilung vergleichen
- Irrtumswahrscheinlichkeit kleiner als gewünschtes Niveau von $\alpha \rightarrow$ Alternativhypothese annehmen, sonst Nullhypothese beibehalten