Testen von Hypothesen

Marcus Hudec

Hypothesen über die Grundgesamtheit

- Aufgabenstellung der Testtheorie
- Hypothesen (Annahmen, Vermutungen oder Behauptungen) über die unbekannte Grundgesamtheit anhand einer Stichprobe als richtig oder falsch einzustufen.
- <u>Arten von Hypothesen</u>
  - Hypothesen über Parameter (z.B.: Merkmalsanteil; Mittelwert)
  - Hypothesen über die Unabhängigkeit von Merkmalen
  - Hypothesen über Form der Verteilung (z.B.: Normalverteilung)
  - Hypothesen über die Güte der Anpassung von Modellen (z.B.: passt eine lineare Trendgerade)

John Arbuthnot (1667 -1735)

- Leibarzt von Queen Anne
- Wissenschaftler und Satiriker (Freund von Jonathan Swift)
- Auszählung von Geburtsregistern von 82 Jahrgängen (1629 – 1710)
- Anzahl Knabengeburten > Anzahl Mädchengeburten
- Ergebnis: 82 zu 0
- Schlussfolgerung: „Das kann kein Zufall sein“
- Die Hypothese, dass Knaben- und Mädchengeburten gleich wahrscheinlich seien war damit bereits 1710 widerlegt.

Wo sind objektive Grenzen?

- Was wäre gewesen, wenn das Ergebnis nicht 82:0 sondern etwa 60:22, oder gar 50: 32 gelautet hätte?
- Ab welchem empirischen Ergebnis hätte sich Arbuthnot getraut, in seinen wissenschaftlichen Aufzeichnungen zu notieren?

„Das kann kein Zufall sein“

Testverfahren

- Ein statistischer Test (Signifikanztest) ist ein Verfahren, das es erlaubt - auf der Basis einer empirischen Stichprobe - sich zwischen zwei konkurrierenden wissenschaftlichen Hypothesen mit einer gewissen, kontrollierten Irrtumswahrscheinlichkeit zu entscheiden.

Knaben- vs. Mädchengeburten

- Nullhypothese: Knaben- und Mädchengeburten sind gleich wahrscheinlich Prob(K)=Prob(M)=0,5
- Alternativhypothese: Knaben- und Mädchengeburten sind nicht gleich wahrscheinlich Prob(K)<>Prob(M)

  Anmerkung:

  Es handelt sich hier um eine sog. zweiseitige Fragestellung (two-sided test), da keine spezifische Richtung einer Abweichung von der Nullhypothese festgelegt wird.

Prinzipielles Vorgehen der Testtheorie

1. Unter der Annahme, dass die Nullhypothese stimmt, bestimmen wir die Wahrscheinlichkeit der Stichprobe.
2. Wenn diese Wahrscheinlichkeit kleiner als eine vorher gewählte Irrtumswahrscheinlichkeit ist, können wir die Nullhypothese mit eben dieser Wahrscheinlichkeit verwerfen.

Prob(Knaben)=	0,5
Prob(Mädchen)=	0,5
Prob(Mehr Knabengeburten in einem Jahr)=	0,5
Prob(Mehr Mädchengeburten in einem Jahr)=	0,5
Anzahl der Jahre	82
Wahrscheinlichkeit, dass in allen 82 Jahren Knabengeburten überwiegen = $0{,}5^{82}$	0,00000000000000000000000002068

Beziehung zu Konfidenzintervall

- N=10.000 Haushalte einer Kleinstadt
- Null-Hypothese:
  25% der Haushalte verfügen über mehr als ein KFZ.
- Eine Stichprobe unter n=100 Haushalten hat ergeben, dass 30 der befragten Haushalte über mehr als ein Auto verfügen.
- Ist damit unsere Hypothese widerlegt?
- Einen Indikator für die Beantwortung liefert das Konfidenzintervall, welches wir bei einem Wert von α=0,05 mit [21%;39%] ermittelt haben. Offensichtlich werden wir unsere Hypothese, wenn wir mit 95% Sicherheit argumentieren wollen, nicht verwerfen, da das Intervall den Wert von 25% inkludiert.
- Umgekehrt: hätten wir die Null-Hypothese gehabt, dass 20% der Haushalte über mehr als ein KFZ verfügen, könnten wir diese wohl verwerfen, da der Wert von 20% nicht mehr im Intervall liegt.
- Aber Achtung: Konfidenzintervall ist immer symmetrisch –es gibt aber auch einseitige Hypothesen

Arten von Hypothesen

- Nullhypothese (Ausgangshypothese) H_0
- Alternativhypothese H_1 oder H_A
- Wichtigste Form von Hypothesen:
  - einseitige Hypothesen

    $H_0: \theta < \theta_0 \qquad H_1: \theta \geq \theta_0$
  - zweiseitige Hypothesen

    $H_0: \theta = \theta_0 \qquad H_1: \theta \neq \theta_0$

Hinweise:

Einseitige Hypothesen setzen voraus, dass die Richtung der Abweichung bekannt ist.
Zweiseitige Hypothesen entsprechen mathematisch einem Konfidenzintervall.

Testverfahren

- Ein statistischer Test (Signifikanztest) ist ein Verfahren, das es erlaubt auf der Basis einer Stichprobe entweder die Nullhypothese H_0 abzulehnen (auch zurückweisen oder verwerfen) und sich für die Alternativhypothese H_1 zu entscheiden oder aber die H_0 beizubehalten (anzunehmen).
- Das Annehmen oder Ablehnen von H_0 kann entweder die richtige oder die falsche Entscheidung sein.
- Insgesamt gibt es 4 Konstellationen

Entscheidungsproblem

		Ergebnis des Test	
		H_0 verworfen statistisch signifikant	H_0 beibehalten statistisch nicht signifikant
Wahrer Zustand	H_A	O.K.	*β-Fehler*
	H_0	*α-Fehler*	O.K.

Fehlerarten

- Die Irrtumswahrscheinlichkeit α (Fehler 1.Art) ist die Wahrscheinlichkeit, mit der die Nullhypothese fälschlicherweise abgelehnt wird.
  Entscheidung für die Alternativhypothese, obwohl die Nullhypothese richtig ist.
- Die Fehlerwahrscheinlichkeit β (Fehler 2.Art) ist die Wahrscheinlichkeit, mit der die Alternativhypothese nicht als richtig erkannt wird.
  Entscheidung für die Nullhypothese, obwohl die Alternativhypothese richtig ist.

Neyman-Pearson Prinzip

- Da die simultane Minimierung beider Fehlerarten scheitert, erfolgt eine asymmetrische Behandlung
- Der α-Fehler (Fehler 1.Art) wird a priori festgehalten (meist 0,01 oder 0,05)
- <u>Die interessierende Hypothese wird als Alternativhypothese</u> festgelegt
- ==> für signifikante Ergebnisse [Entscheidung für Alternativhypothese] kennen wir dann die Irrtumswahrscheinlichkeit, die wir a priori festsetzen
- ==> nicht signifikante Ergebnisse [Entscheidung für Nullhypothese] implizieren nicht, dass die Nullhypothese tatsächlich gilt (Fehler 2.Art)

Einstichprobentest für den Anteilswert (1)

- Fragestellung:
  Wir wollen zeigen, dass der Anteil (π) einer bestimmten Minderheit X in einer Population größer als 10% ist.
- H_0: π =0,10 (bzw. $\pi \leq 0{,}10$) versus
- H_1: $\pi > 0{,}10$
- Irrtumswahrscheinlichkeit α=0,05 bzw.
- Signifikanzniveau 1-α=0,95

Hinweis:
Für die Berechnung der Verteilung der Teststatistik genügt es den kritischen Fall π =0,10 zu betrachten.

Einstichprobentest für den Anteilswert (2)

- Bestimmung einer sinnvollen Prüfgröße
- Idee: standardisierte Abweichung des empirisch beobachteten Anteils vom hypothetisch unterstellten Anteil
- Bestimmung der Verteilung dieser Prüfgröße unter der Annahme, dass die Nullhypothese gültig ist

Einstichprobentest für den Anteilswert (3)

- $p \sim N(\pi, \pi(1-\pi)/n)$
- Falls H_0 zutrifft: $p \sim N(\pi_0, \pi_0(1-\pi_0)/n)$
  π_0...Grenzwert=0,10

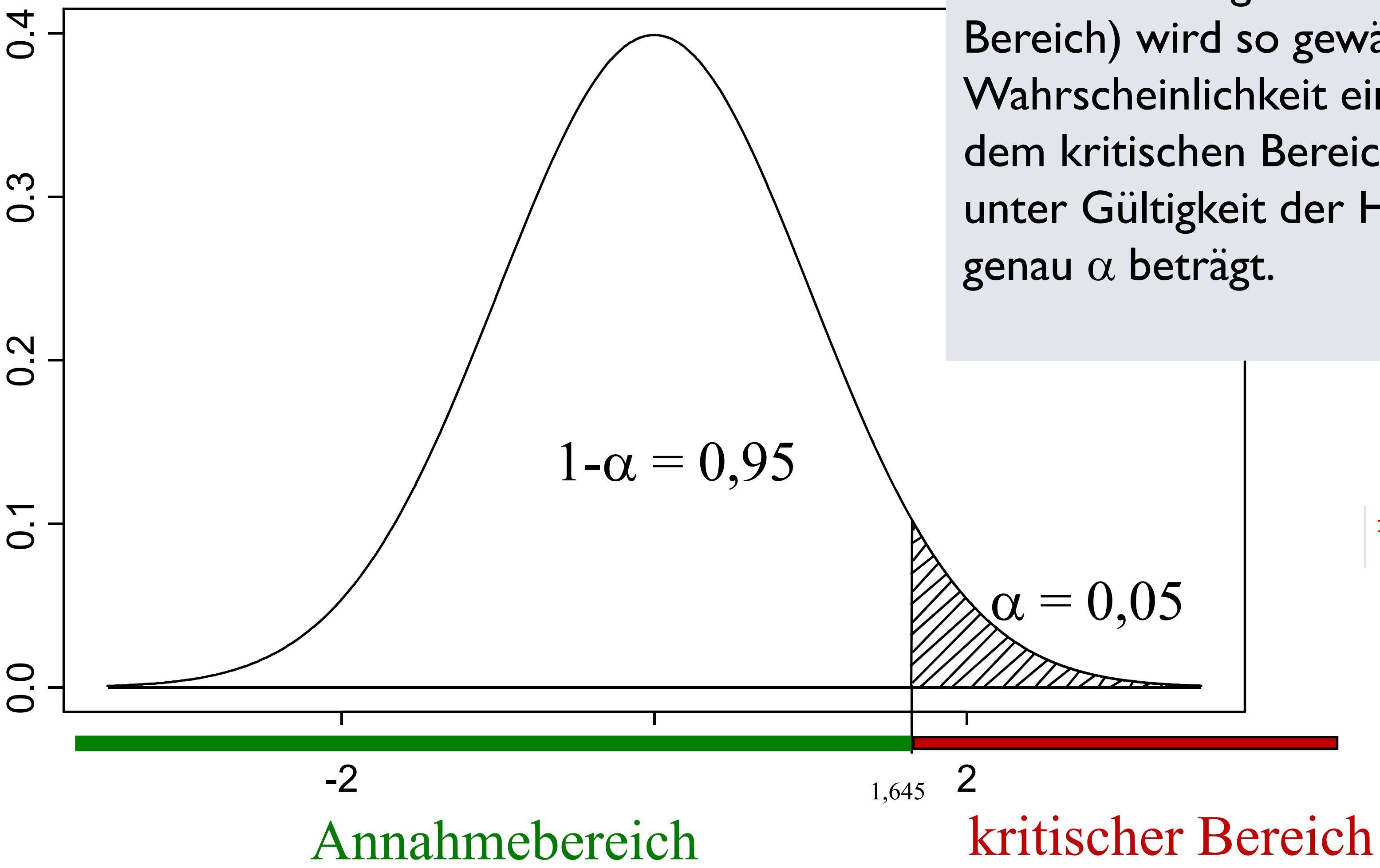
- Prüfgröße

$$Z = \frac{p - \pi_0}{\sigma_p} = \frac{p - \pi_0}{\sqrt{\frac{\pi_0(1-\pi_0)}{n}}}$$

- Die Prüfgröße Z ist bei Gültigkeit der Nullhypothese für große Stichproben und kleinem Auswahlsatz approximativ standardnormal verteilt

Der kritische Bereich

Der Ablehnungsbereich (Kritischer Bereich) wird so gewählt, dass die Wahrscheinlichkeit einen Wert aus dem kritischen Bereich zu erhalten unter Gültigkeit der H_0 höchstens genau α beträgt.




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

Kritischer Wert

- In unserem Beispiel ist der kritische Wert 1,645
- Liegt die standardisierte Abweichung des Stichprobenwertes vom zu testenden Wert der Nullhypothese unter diesem Wert, fällt die Beobachtung in den Annahmebereich und wir entscheiden uns für die Nullhypothese H_0.
- Liegt die Beobachtung oberhalb dieses Wertes, d.h. sie liegt im kritischen Bereich (Rückweisungs-, Ablehnungsbereich), so entscheiden wir uns für die Alternativhypothese H_1. Man spricht dann auch von einem signifikanten Ergebnis.

Beispiel:

- Stichprobe: n=100; x=15 ➔ p=0,15

$$Z = \frac{p - \pi_0}{\sqrt{\frac{\pi_0(1-\pi_0)}{n}}} = \frac{0{,}15 - 0{,}10}{\sqrt{\frac{0{,}1 \cdot 0{,}9}{100}}} = \frac{0{,}05}{0{,}03} = 1{,}66$$

Das bedeutet wir entscheiden uns (knapp aber doch) für H_1. Wir können die Nullhypothese bei einem Signifikanzniveau von 0,95 verwerfen.

Auf der Basis unseres Stichprobenergebnisses können wir bei Vorgabe einer Irrtumswahrscheinlichkeit von 5%, die Hypothese, dass der Anteil der Minderheit kleiner gleich 10% ist, zurückweisen bzw. die Hypothese, dass der Anteil größer als 10% ist annehmen.

Alternative Angabe des kritischen Bereichs

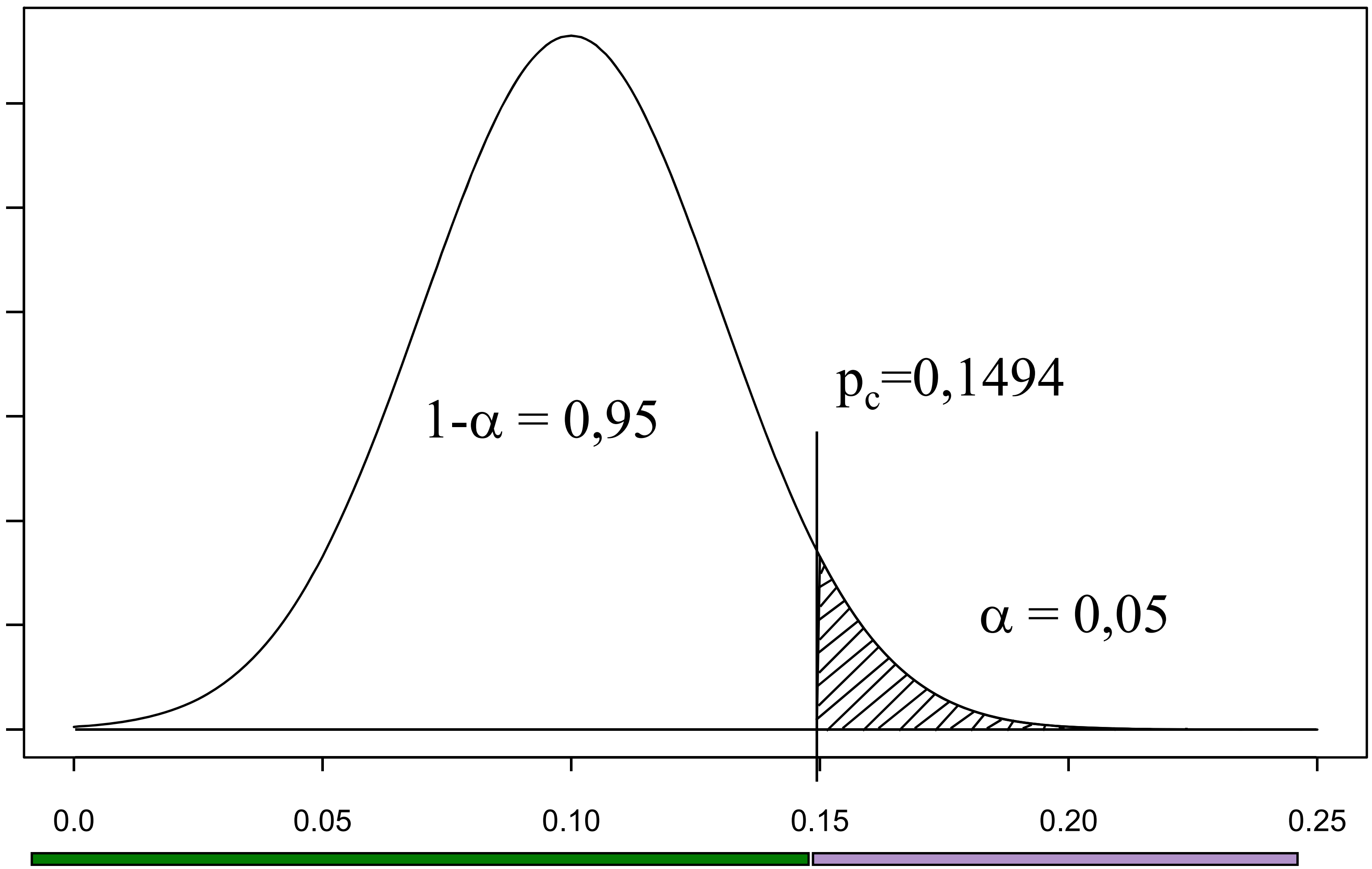
- In manchen Anwendungen (z.B. Qualitätssicherung in der Industrie) will man den kritischen Bereich in der eigentlich untersuchten Dimension ausdrücken, dazu errechnet man durch Umformung der Teststatistik die sog. Annahmekennzahl.
- In unserem Beispiel sucht man jenen Wert für die Anzahl der deklarierten Mitglieder einer Minderheit in einer **Stichprobe von n=100**, ab welchem die Nullhypothese $\pi \leq 0{,}10$ abzulehnen ist?

$$z_c = \frac{p_c - \pi_0}{\sqrt{\frac{\pi_0(1-\pi_0)}{n}}} \Rightarrow p_c = \pi_0 + z_c \cdot \sqrt{\frac{\pi_0(1-\pi_0)}{n}}$$

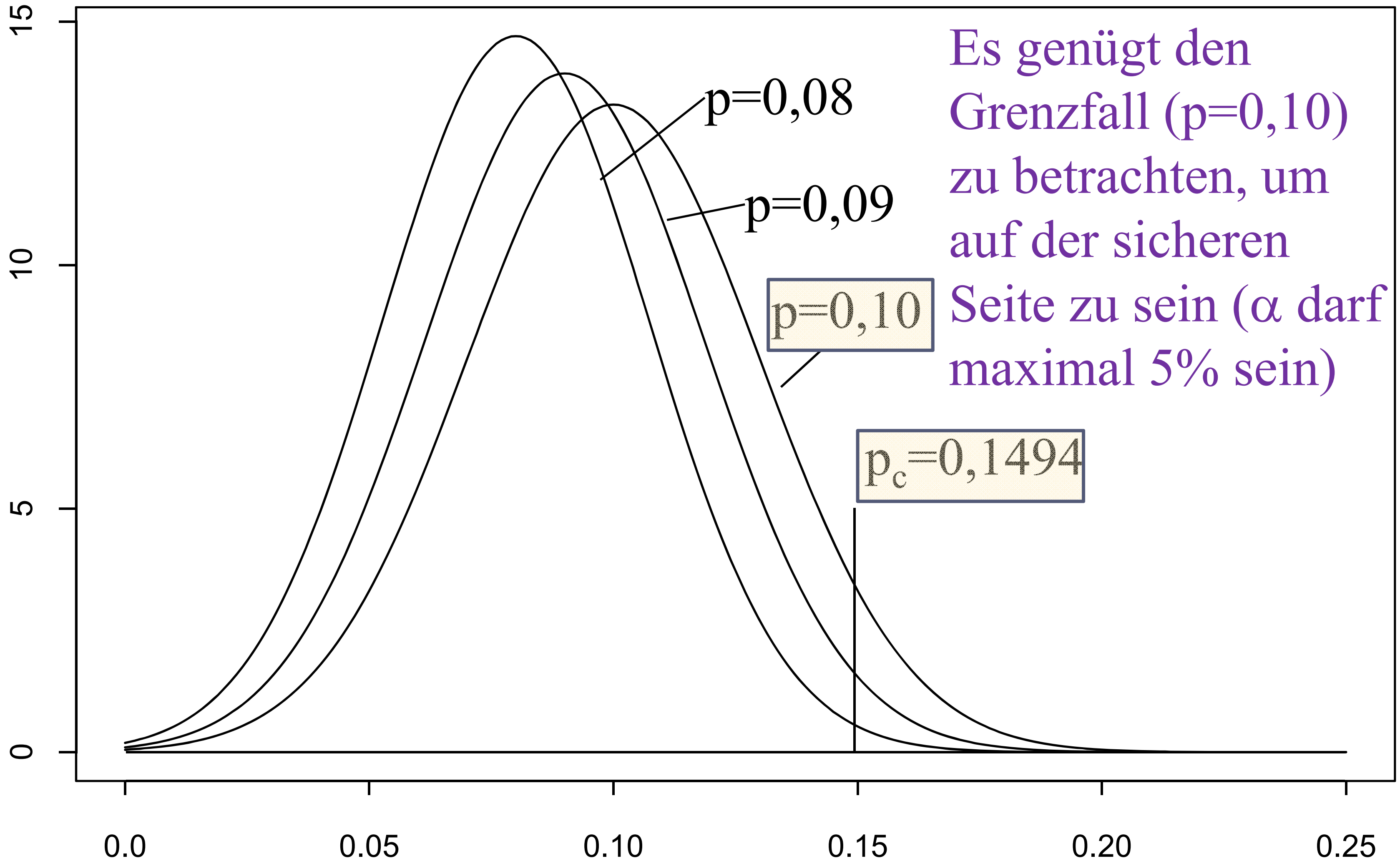
Beispiel

- p_c=0,10+1,645*0,03=0,1494
- Bei einer Stichprobe von 100 lehnen wir die H_0 dann ab einer Anzahl von 15 oder mehr deklarierten Mitgliedern der Minderheit in der Stichprobe mit einer Irrtumswahrscheinlichkeit von α=0,05 ab.
- Hat man weniger als 15 (also maximal 14) deklarierte Mitglieder empirisch erhoben, so kann die Nullhypothese nicht verworfen werden.
- Hätte man etwa x=13 beobachtet, so ergibt sich für die Testgröße z = (0,13-0,10)/0,03=1, was zur Beibehaltung der H_0 geführt hätte.

Kritischer Bereich ausgedrückt als Anteilswert



Zusammengesetzte Nullhypothese



Beispiel:

- Der Stimmenanteil einer Partei betrug bei der letzten Wahl 40%. Eine Umfrage unter n=600 Wahlberechtigten soll zeigen, ob sich der Anteil der Partei verändert hat.
- H_0: $\pi = 0{,}40$
- H_1: $\pi \neq 0{,}40$ (zweiseitige Alternative)
- Irrtumswahrscheinlichkeit $\alpha = 0{,}05$
- Signifikanzniveau $1-\alpha = 0{,}95$
- Prüfgröße:

$$Z = \frac{p - \pi_0}{\sigma_p} = \frac{p - \pi_0}{\sqrt{\frac{\pi_0(1-\pi_0)}{n}}}$$

Beispiel:

- Kritischer Bereich:
  $[-\infty,-1{,}96] \cup [1{,}96,+\infty]$
- Annahmebereich: $[-1{,}96,+1{,}96]$
- Alternative Angabe des Kritischen Bereichs entspricht dem Schwankungsintervall bei Gültigkeit von H_0:

$$\sigma_p = \sqrt{\frac{\pi_0(1-\pi_0)}{n}} = \sqrt{\frac{0{,}4 \cdot 0{,}6}{600}} = \sqrt{0{,}0004} = 0{,}02$$

$$p_c = \pi_0 + z_c \cdot \sqrt{\frac{\pi_0(1-\pi_0)}{n}} = 0{,}4 \pm 1{,}96 \cdot 0{,}02 = 0{,}4 \pm 0{,}0392$$

[0,3608;0,4392] Annahmebereich in Anteilen

[217;263] Annahmebereich in Anzahl

Beispiel:

- Durchführen der Befragung ergibt:
- n=600 x=261 p=0,435 43,5%
- Teststatistik:

$$Z = \frac{p - \pi_0}{\sigma_p} = \frac{p - \pi_0}{\sqrt{\frac{\pi_0(1-\pi_0)}{n}}} = \frac{0{,}435 - 0{,}4}{0{,}02} = 1{,}75$$

Entscheidung:

Beibehaltung der Nullhypothese

Aufgrund des Stichprobenergebnisses kann, die Nullhypothese π=0,40 mit einer Irrtumswahrscheinlichkeit von α=0,05 nicht abgelehnt werden.

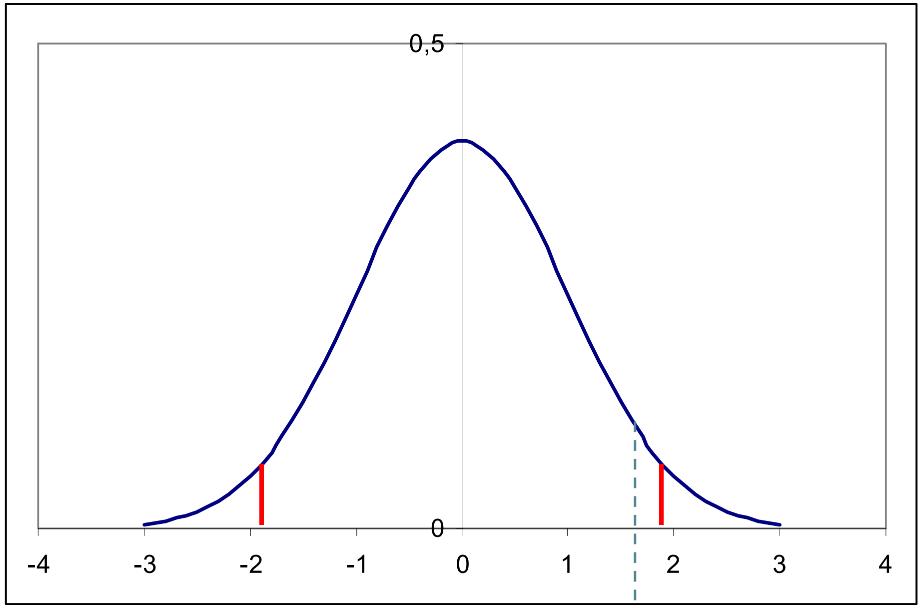
Einseitige Formulierung der Fragestellung

- Wir wollen mit 95% Wahrscheinlichkeit „nachweisen", dass der Stimmenanteil gewachsen ist:
- H_0: $\pi \leq 0{,}40$
- H_1: $\pi > 0{,}40$ (einseitige Alternative)
- Irrtumswahrscheinlichkeit $\alpha = 0{,}05$
- Konfidenzniveau $1-\alpha = 0{,}95$
- Prüfgröße bleibt unverändert
- Einseitiger Annahmebereich: $[-\infty, +1{,}645]$
- bzw. Rückweisungsbereich in Anteilen ausgedrückt
  0,40+1,645*0,02==> [43,29%; 100%]
- Im Beispiel: ==> signifikantes Ergebnis bei α=0,05

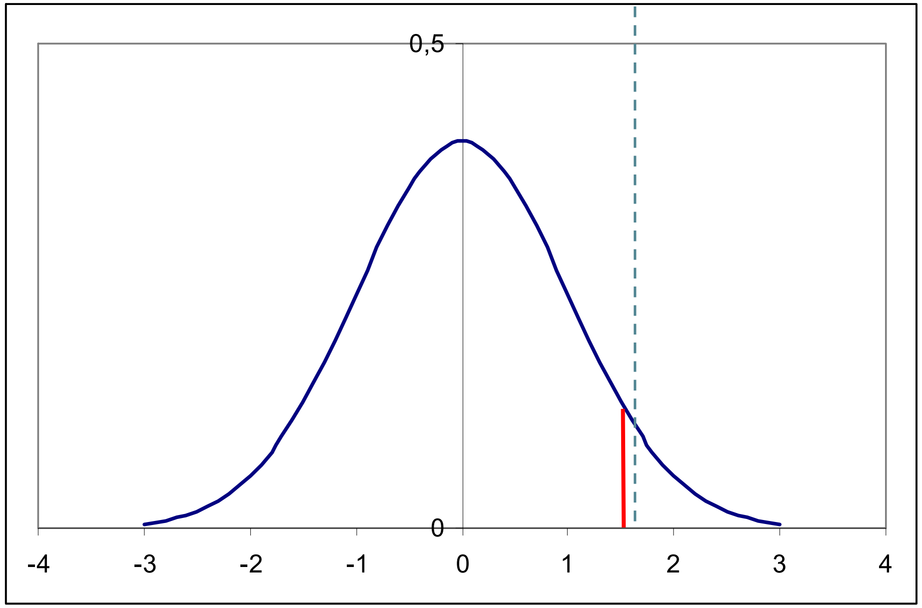
Beachte die Wahl von H_0 und H_1

Einseitige versus Zweiseitige Hypothesen

- Offensichtlich sind einseitige Fragestellungen trennschärfer als zweiseitige (d.h. wir erzielen ceteris paribus leichter ein signifikantes Ergebnis)
- Bei einseitiger Hypothese geht die Richtung der vermuteten Abweichung in die Testprozedur als zusätzliche Annahme ein
- Festlegen der Hypothesen muss immer vor der Durchführung der Untersuchung festgelegt werden
- <u>Eine nachträgliche datengesteuerte Spezifikation verfälscht das Signifikanzniveau</u>

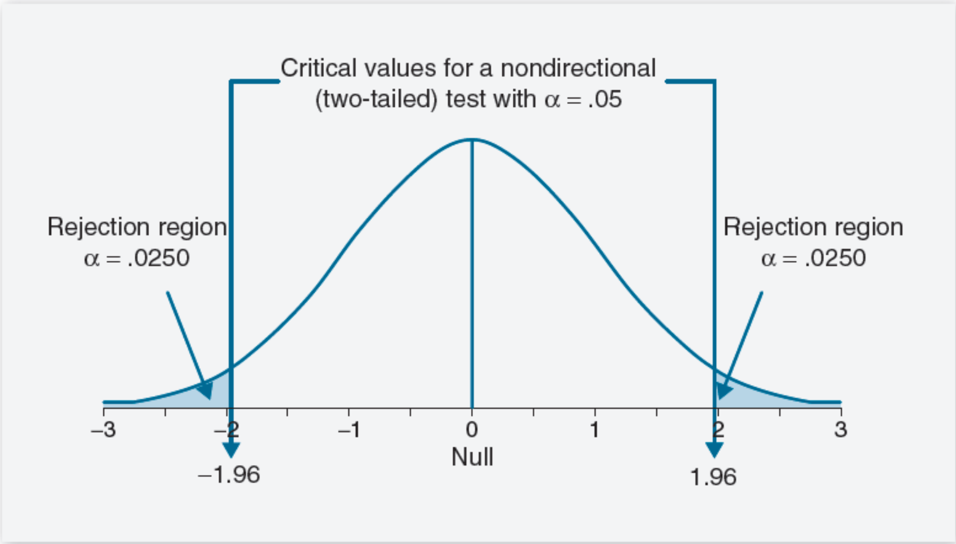


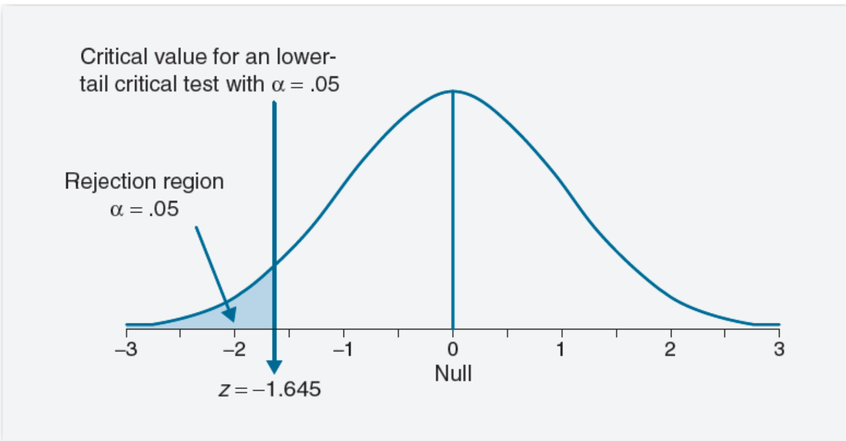
Oben:
Zweiseitiger Test liefert kein signifikantes Ergebnis

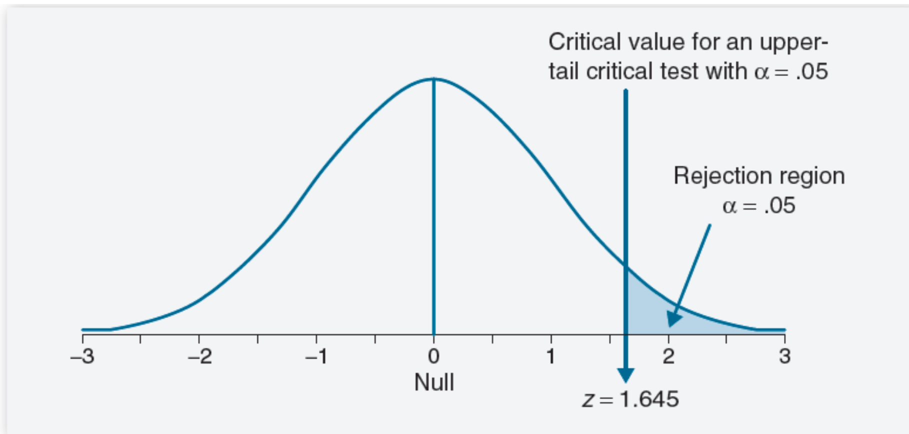


Unten:
Einseitiger Test liefert signifikantes Ergebnis

Gerichtete und ungerichtete (zweiseitige) Hypothesen







Streng konfirmatorische Vorgangsweise

Vor der Datenerhebung

- Aufstellen von Null- und Alternativhypothese
- Festlegen des Signifikanzniveaus
- Festlegen einer Prüfgröße (Stichprobenfunktion)
- Bestimmen der Testverteilung bei Gültigkeit der Nullhypothese
- Ermittlung des kritischen Bereichs (Ablehnungsbereich)

Nach

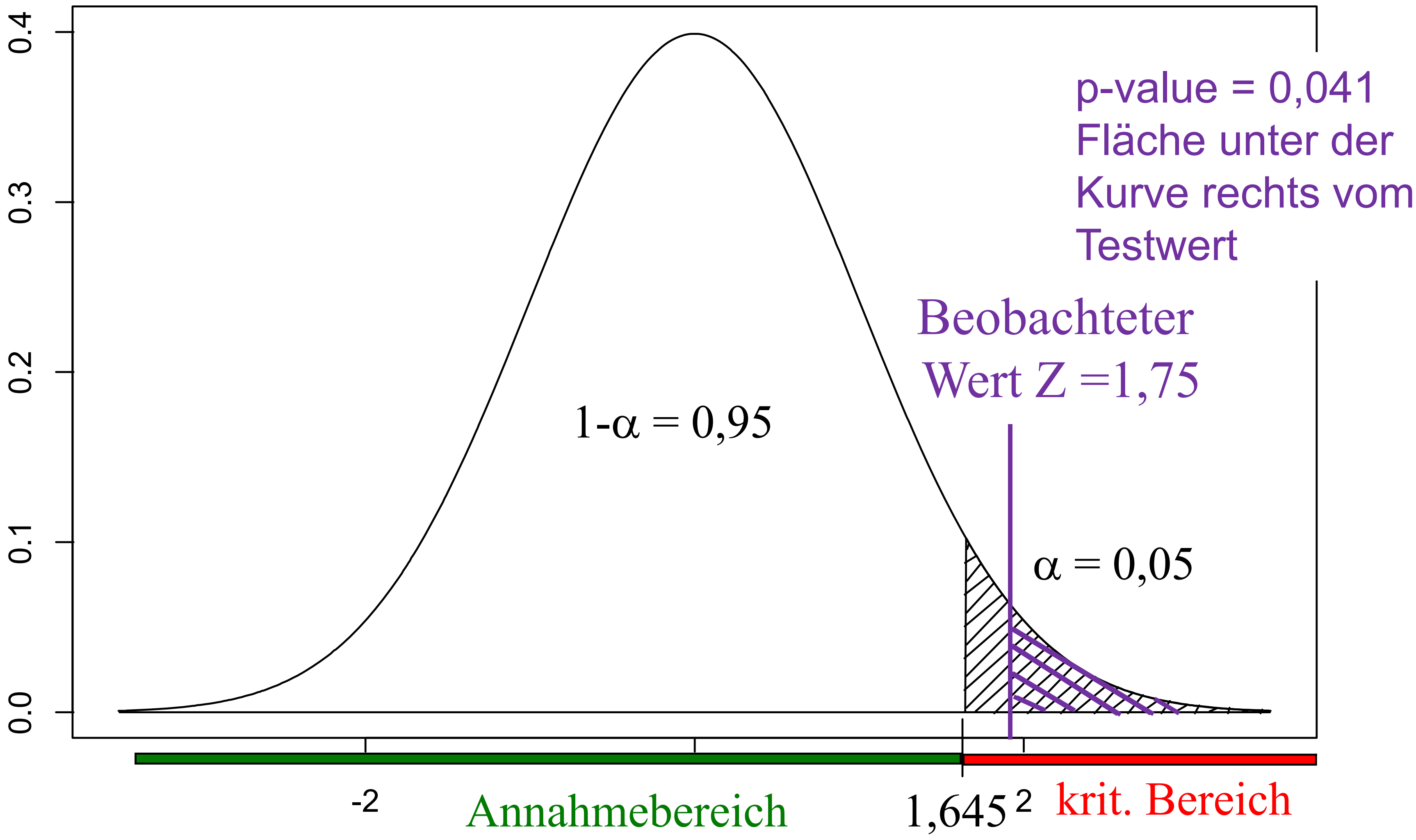
- Berechnung der Prüfgröße
- Entscheidung & Interpretation

Empirisches Signifikanzniveau (p-value)

- Das empirische Signifikanzniveau ist jene Wahrscheinlichkeit α^*, mit der die Prüfgröße einen Wert T annimmt, der unter H_0 genauso oder sogar noch unplausibler ist als der beobachtete Wert für die Teststatistik t^*.
- Wert der Teststatistik im Beispiel: $z=1{,}75$
- Bei einseitiger Fragestellung:
  $P(Z>1{,}75) = ?$ ==> $1-\Phi(1{,}75)=1-0{,}959=0{,}041$
- Häufige Schreibweise für das empirische Signifikanzniveau: p-value=0,041

Die Nullhypothese ist bei vorgegebenen α immer dann abzulehnen, wenn der p-value kleiner als das gewählte α ist.

Der kritische Bereich



Empirisches Signifikanzniveau

- Der p-value ist eine deskriptive Zusatzinformation, die angibt, wie weit liegt die Teststatistik im Ablehnungs-bereich bzw. wie weit ist sie vom kritischen Grenzwert entfernt.
- Der p-value kann für eine objektive Entscheidungsfindung nicht die a priori Wahl des Signifikanzniveaus ersetzen.
- Die datengesteuerte Wahl des Signifikanzniveaus verändert den Fehler 2.Art

Beispiel: Wirksamkeit von Medikamenten

- Eine Pharmafirma hat ein neues Medikament entwickelt, von dem vermutet wird, dass es die Heilungschance bei einer bestimmten Krankheit von 20% (Erfolgschance bei Standardmedikation) erhöht.
- Man plant eine Studie mit n=100 Patienten.
- Ab welchem Ergebnis, sprechen wir von einer signifikanten Verbesserung?
- Ab welchem Ergebnis, würden Sie die Entscheidung treffen, dass das neue Medikament besser als die Standardmedikation ist?

Hypothesen

- Nullhypothese: Die Wirksamkeit des neuen Medikamentes ist höchstens gleich gut, wie die der bisherigen Standardmedikation:
  Prob(Response) <= 0,2
- Alternativhypothese: Die Wirksamkeit des neuen Medikamentes ist besser, als die der bisherigen Standardmedikation:

  Prob(Response) > 0,2

  Anmerkung:

  einseitige Fragestellung (one-sided test)

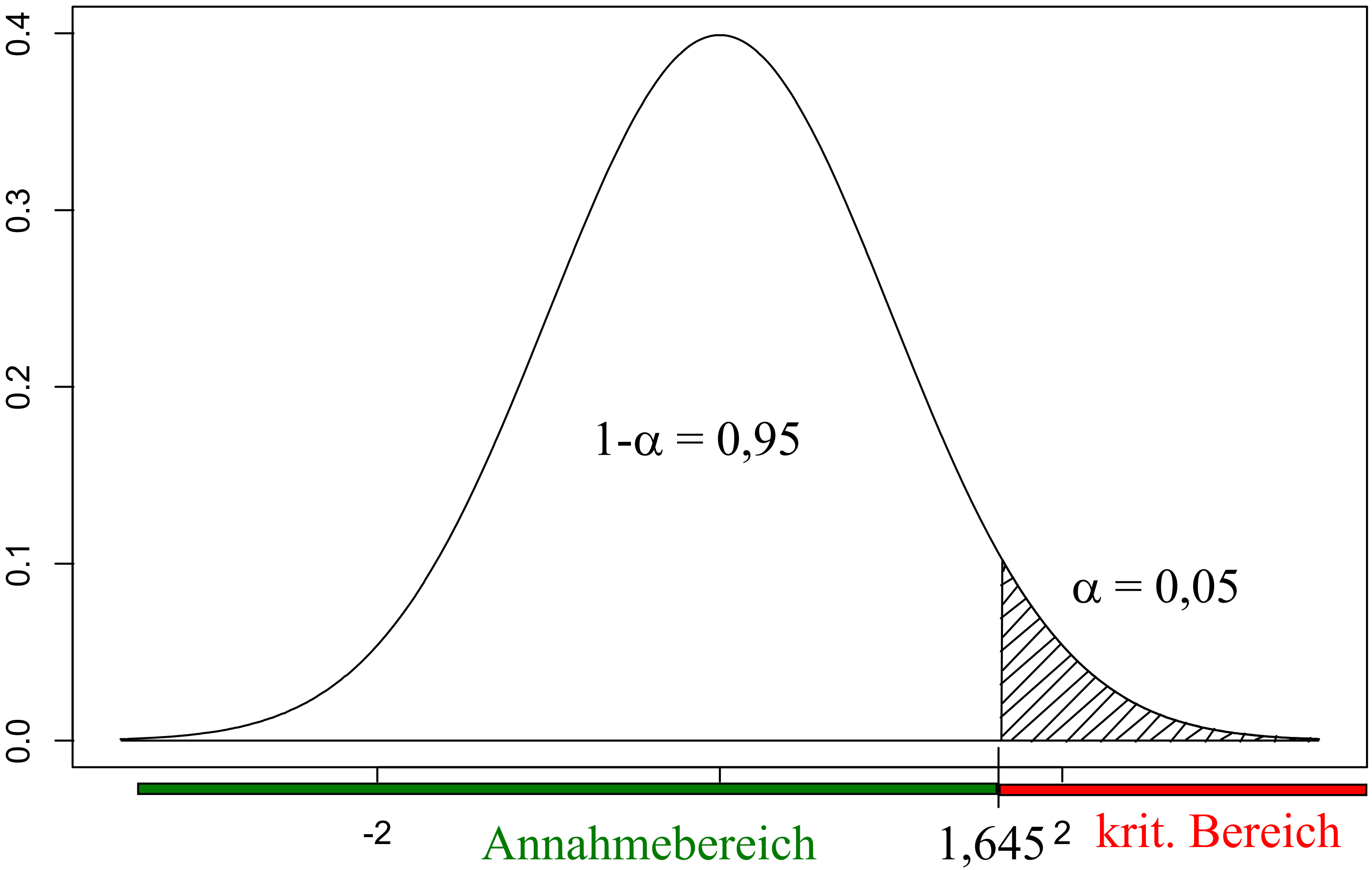
Einstichprobentest für den Anteilswert bei großem n

- Berechnen einer standardisierten Prüfgröße

$$Z = \frac{p - \pi_0}{\sqrt{\frac{\pi_0(1-\pi_0)}{n}}}$$

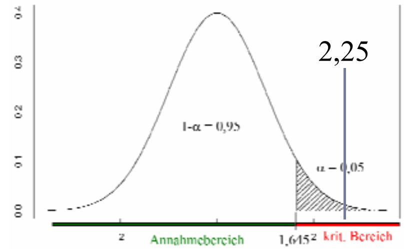
- p ist der tatsächlich beobachtete Anteil, π_0 ist der zu testende Anteil, n ist die Fallzahl
- Z ist **unter Gültigkeit der Nullhypothese standardnormalverteilt** und kann als Normierung der Responserate interpretiert werden

Der kritische Bereich



Berechnung der Testgröße

n= 100 π= 0,2

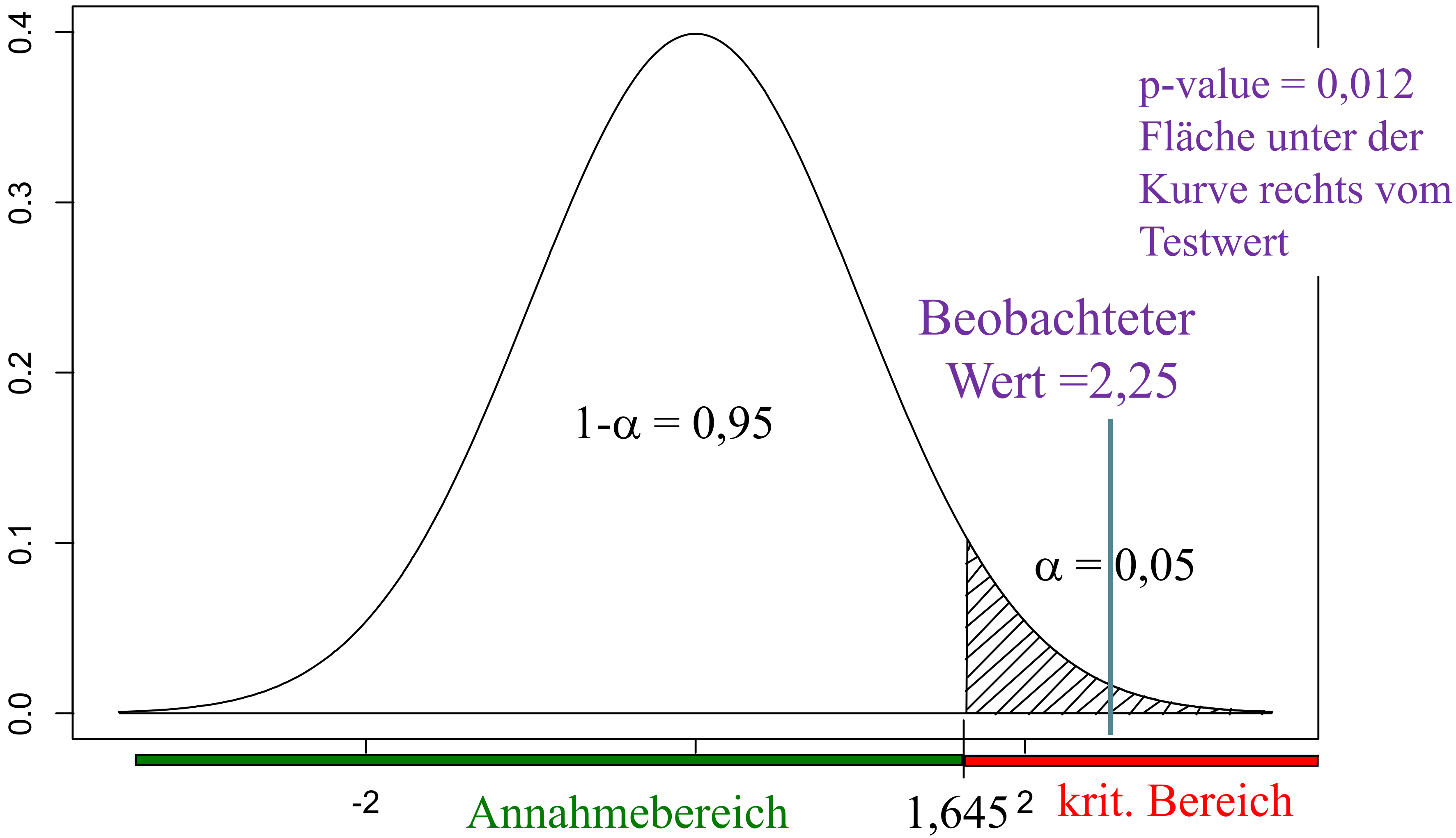


Anzahl der Respondenten= 29
Response-Rate in der Studie p= 0,29

Standardabweichung= 0,04
Differenz= 0,09
Teststatistik= 2,25
p-value= 0,012224

Aufgrund der durchgeführten Studie können wir bei einer Irrtumswahrscheinlichkeit von α=0,05 die Nullhypothese, dass die Wirksamkeit kleiner gleich 20% ist zurückweisen. Der p-value beträgt 0,0122.

Der kritische Bereich



Macht des statistischen Tests (power)

		Ergebnis der Studie	
		H_0 verworfen statistisch signifikant	H_0 beibehalten statistisch nicht signifikant
Realer Treatment Effekt	ja	O.K. *Macht*	*β-Fehler*
	nein	*α-Fehler*	O.K.

Fehlerarten

- Die beiden Fehler sind natürlich nicht unabhängig voneinander
- Verringert man ceteris paribus den Fehler 1.Art, so erhöht man gleichzeitig den Fehler 2.Art
- Für alle Parameterwerte θ aus dem Bereich der Alternativhypothese, kann man den Fehler 2.Art bestimmen.
- Dieser Fehler 2.Art gibt die Wahrscheinlichkeit an, dass das Testverfahren einen real existenten Zustand aus H_1 (Parameterwerte θ aus dem Bereich der Alternativhypothese) nicht erkennt.

Operationscharakteristik und Gütefunktion

- Fasst man den Fehler 2.Art als Funktion von θ auf

  $\beta(\theta)$: Operationscharakteristik (OC)

- Alternativ können wir auch das komplementäre Ergebnis betrachten.

  $1-\beta(\theta)$: Gütefunktion, Power oder Macht des Tests

- Die Gütefunktion gibt die Wahrscheinlichkeit an, dass das Testverfahren einen real existenten Zustand aus H_1 auch tatsächlich erkennt

Macht eines statistischen Tests (power)

- Wahrscheinlichkeit einen real bestehenden Unterschied durch den statistischen Test auch tatsächlich zu entdecken.
- Im Beispiel: Wie groß ist die Chance mit einer empirischen Studie die verbesserte Wirksamkeit nachzuweisen, wenn π tatsächlich einen Wert größer als 0,20 hat.
- Die Macht des statistischen Tests hängt ab von:
  - α-Fehler (Fehler 1. Art) (positiv)
  - Stärke des Effekts/Unterschieds (positiv)
  - Variabilität (negativ)
  - Fallzahl (positiv)

Beispiel:

- Der Stimmenanteil einer Partei betrug bei der letzten Wahl 40%. Eine Umfrage unter n=600 Wahlberechtigten soll zeigen, ob sich der Anteil der Partei verändert hat.
- H_0: $\pi \leq 0{,}40$
- H_1: $\pi > 0{,}40$ (einseitige Alternative)
- Irrtumswahrscheinlichkeit $\alpha = 0{,}05$
- Signifikanzniveau $1\text{-}\alpha = 0{,}95$
- Prüfgröße:

$$Z = \frac{p - \pi_0}{\sigma_p} = \frac{p - \pi_0}{\sqrt{\frac{\pi_0(1 - \pi_0)}{n}}}$$

Berechnung der Gütefunktion

- Einseitige Fragestellung
- H_0: $\pi \leq 0{,}40$
- H_1: $\pi > 0{,}40$ (einseitige Alternative)
- Ergebnis war für α=0,05:
- Rückweisungsbereich für Z :[1,645; +∞]

  bzw.
- Rückweisungsbereich in Anteilen ausgedrückt

  0,40+1,645*0,02==> [43,29%; 100%]

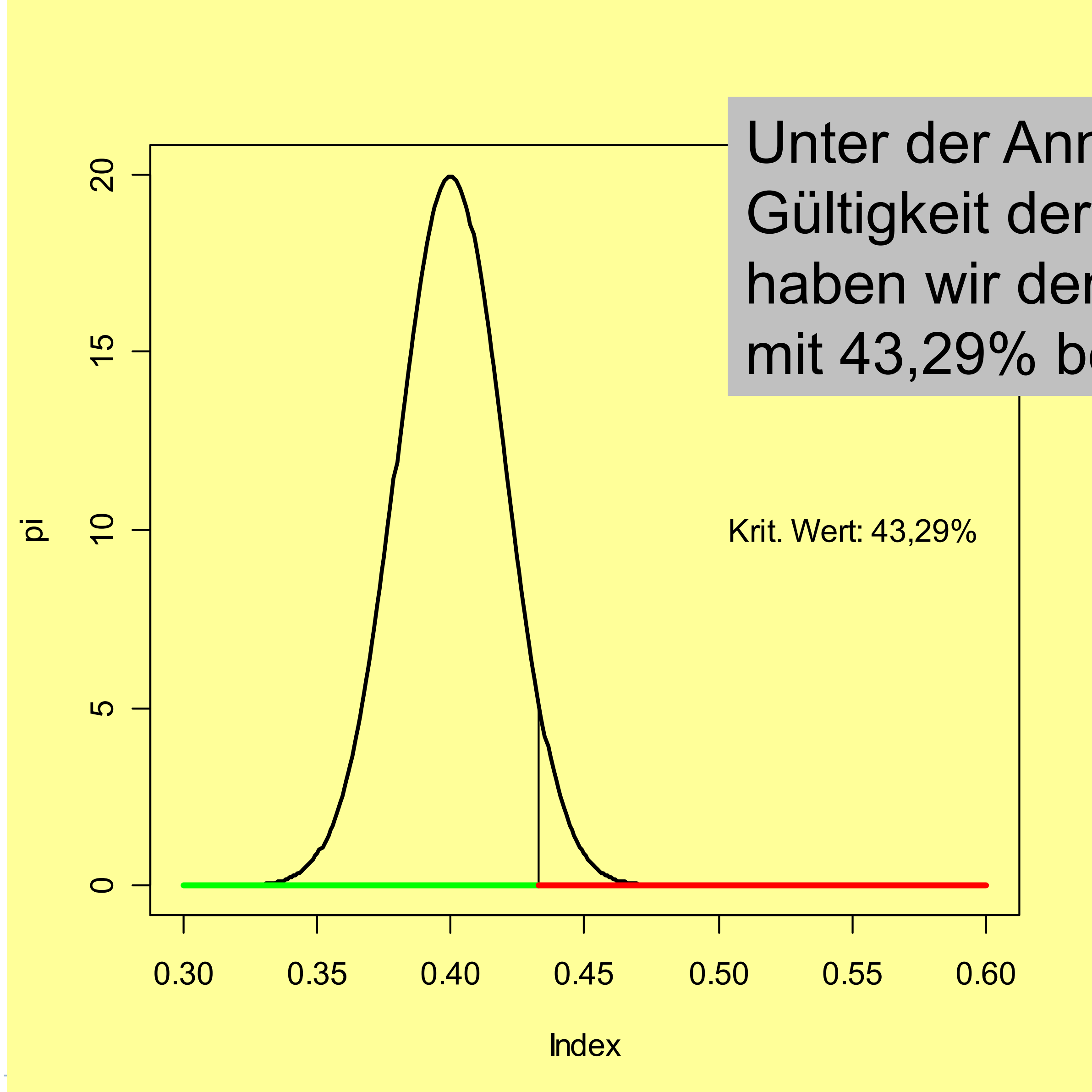
Berechnung der Gütefunktion

- Wie groß ist die Chance eine real existierende Abweichung von der Nullhypothese mit diesem Testverfahren auch tatsächlich zu entdecken?
- Konkret: Wie groß ist die Wahrscheinlichkeit, dass unsere Teststatistik bei n=600 und α=0,05 ein signifikantes Ergebnis liefert, wenn der wahre Stimmenanteil derzeit 0,425 (0,45 oder 0,5) beträgt?
- Antwort: Berechnen der Gütefunktion
- Dazu nehmen wir jetzt verschiedene Szenarien (aus dem Bereich der Alternativhypothese) für die Realität an und schauen, wie sich unsere Entscheidungsregel auswirkt.

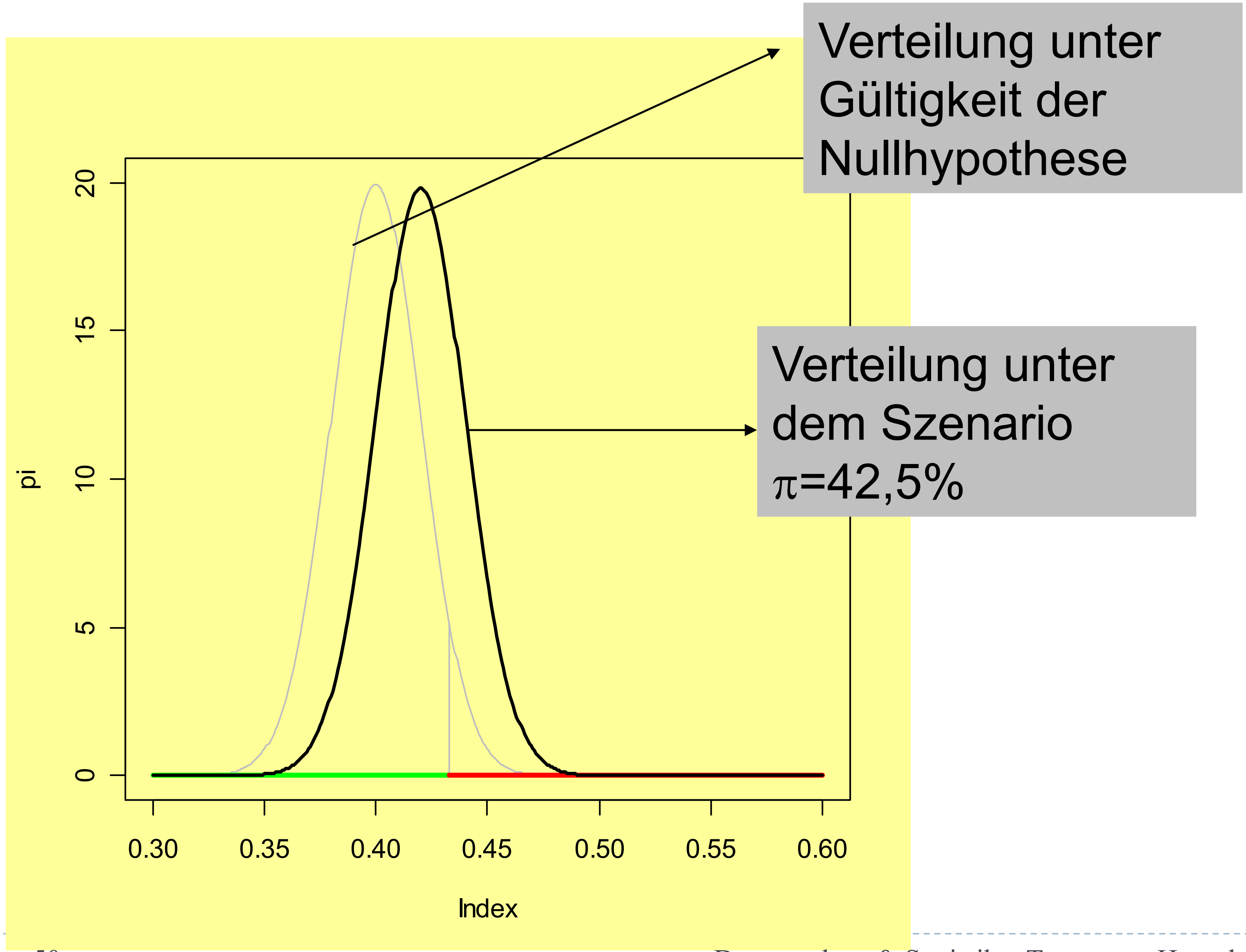
Berechnung der Gütefunktion

- Angenommen der wahre Anteil ist 42,5%.
- Wie groß ist die Wahrscheinlichkeit, dass unser Testverfahren, bei diesem Zustand der Grundgesamtheit ein richtiges (H_A) Ergebnis liefert?
- $P(z>1{,}645|\pi=0{,}425)$ bzw. $P(p>43{,}29|\pi=0{,}425)$

$$Power(0,425) = 1 - \Phi\left(\frac{0,4329 - 0,425}{\sqrt{\frac{0,425 \cdot (1-0,425)}{600}}} \right) = 1 - \Phi(0,391) = 0,35$$

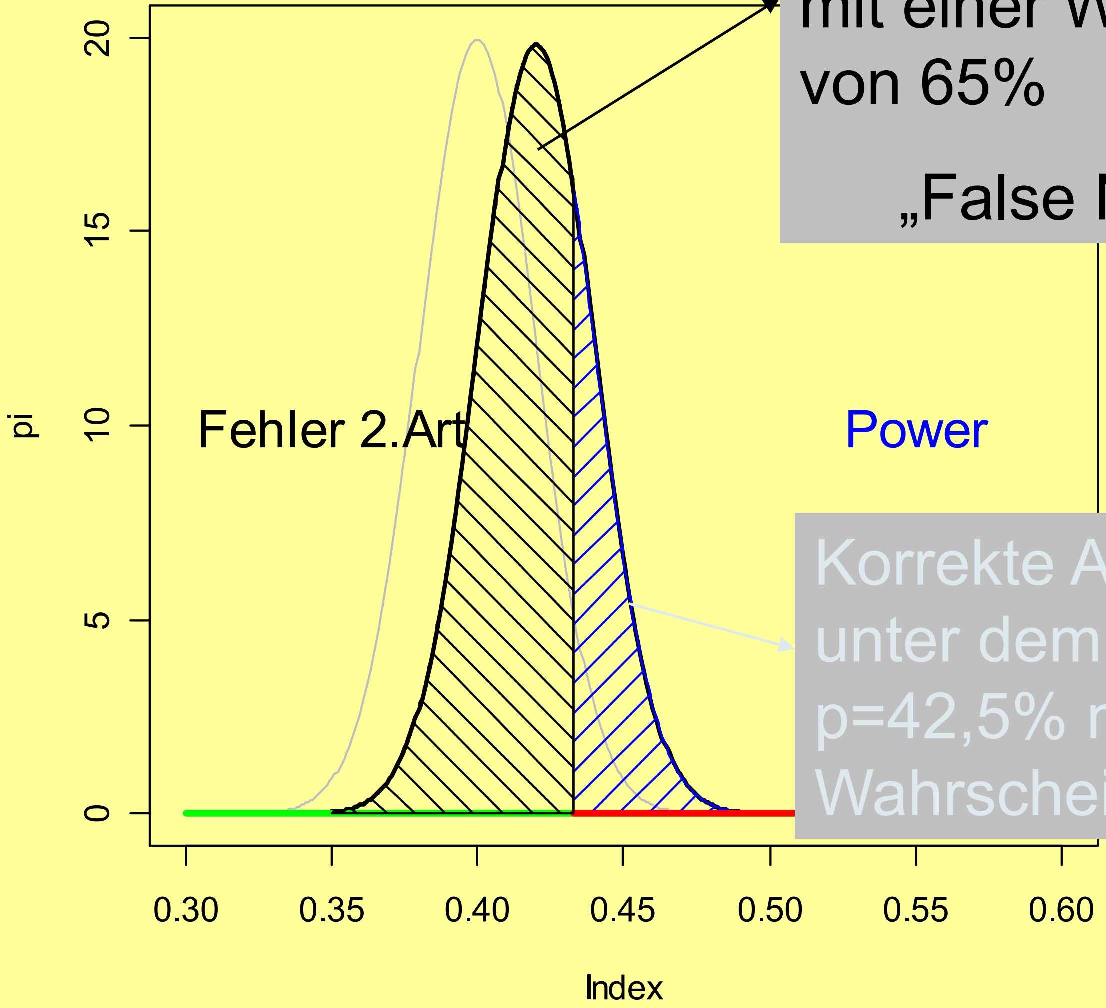


Unter der Annahmen der Gültigkeit der Nullhypothese haben wir den kritischen Wert mit 43,29% bestimmt.


Verteilung unter Gültigkeit der Nullhypothese
Verteilung unter dem Szenario π=42,5%
pi
Index
0
5
10
15
20
0.30
0.35
0.40
0.45
0.50
0.55
0.60

Fälschliche Annahme von H_0 unter dem Szenario p=42,5% mit einer Wahrscheinlichkeit von 65%

„False Negative Result“

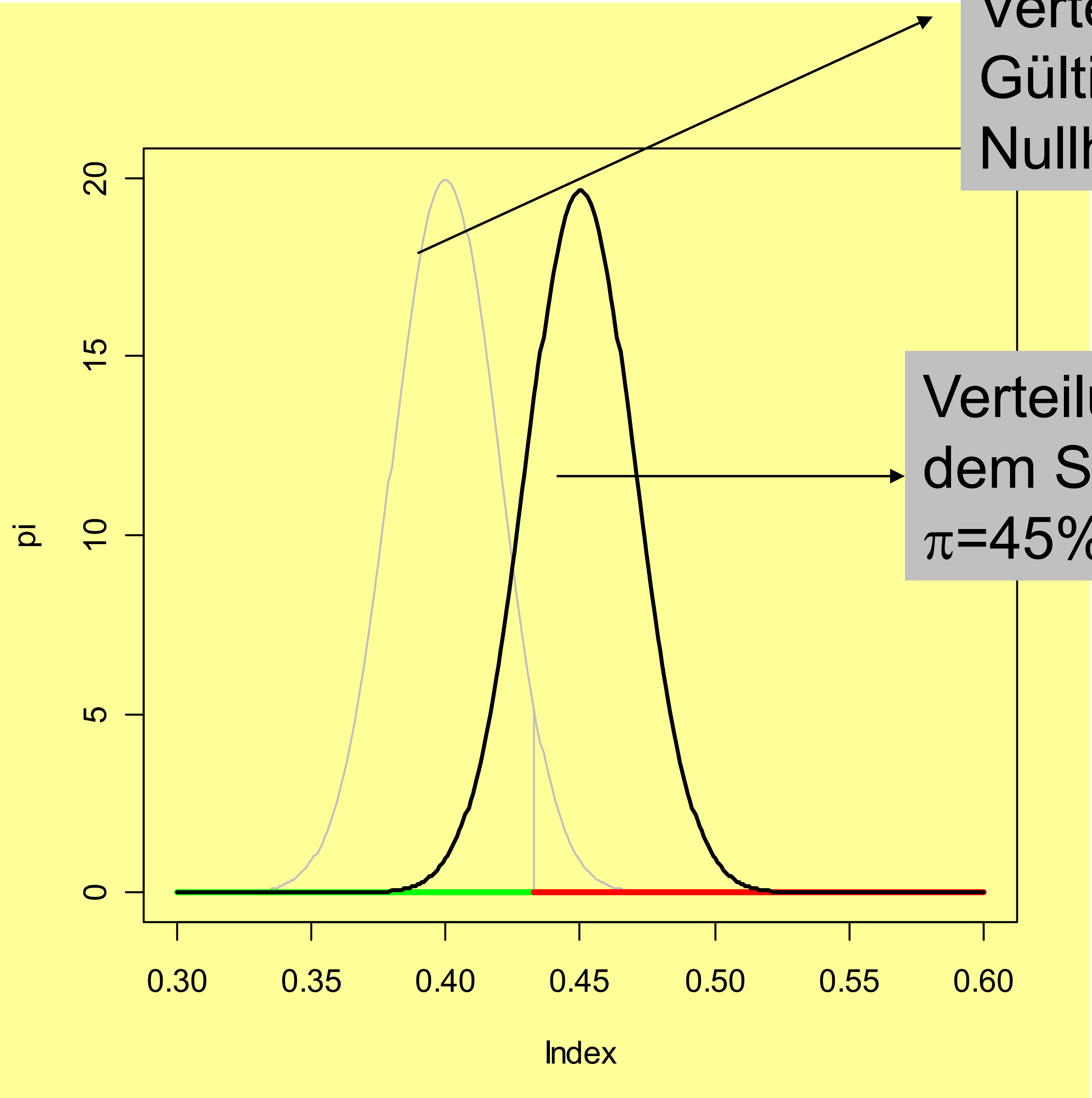


Korrekte Annahme von H_A unter dem Szenario p=42,5% mit einer Wahrscheinlichkeit von 35%

Berechnung der Gütefunktion

- Angenommen der wahre Anteil ist 45%.
- Wie groß ist die Wahrscheinlichkeit, dass unser Testverfahren, bei diesem Zustand der Grundgesamtheit ein richtiges (H_A) Ergebnis liefert?
- $P(z>1{,}645|\pi=0{,}45)$ bzw. $P(p>43{,}29|\pi=0{,}45)$

$$Power(0,45) = 1 - \Phi\left(\frac{0,4329 - 0,45}{\sqrt{\frac{0,45 \cdot (1-0,45)}{600}}} \right) = 1 - \Phi(-0,842) = 0,80$$

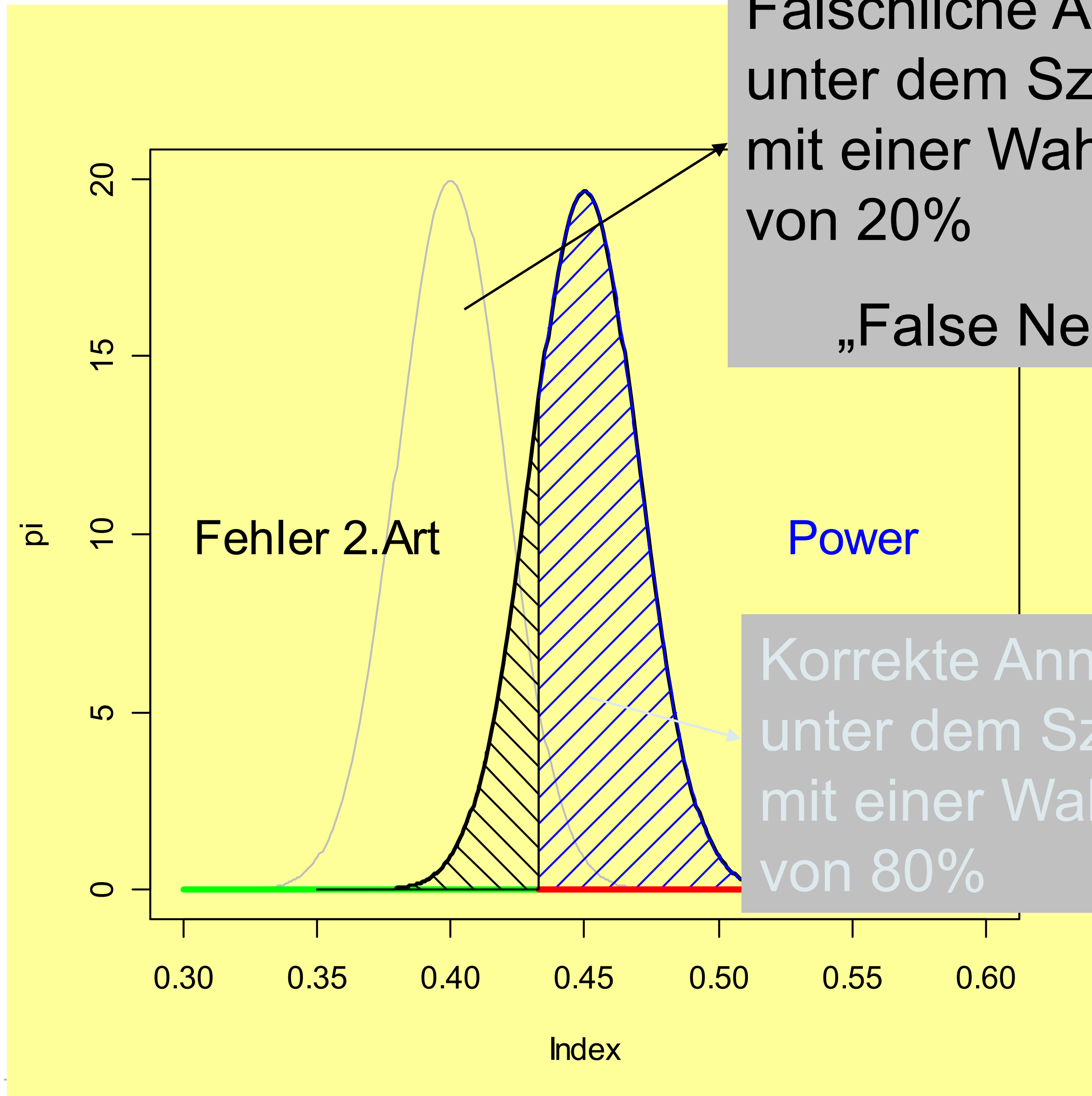


Verteilung unter Gültigkeit der Nullhypothese

Verteilung unter dem Szenario π=45%

Fälschliche Annahme von H_0 unter dem Szenario p=45% mit einer Wahrscheinlichkeit von 20%

„False Negative Result“

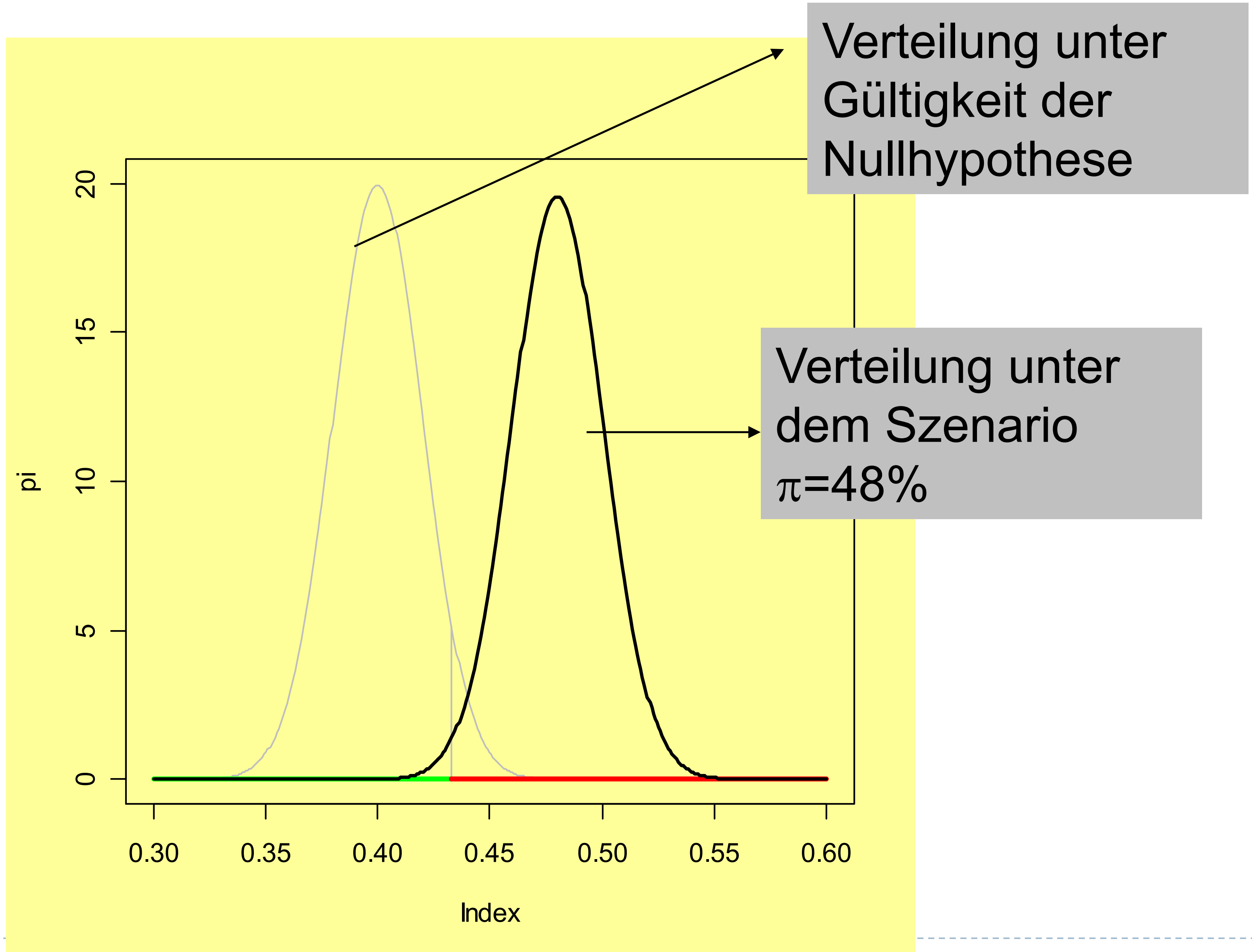


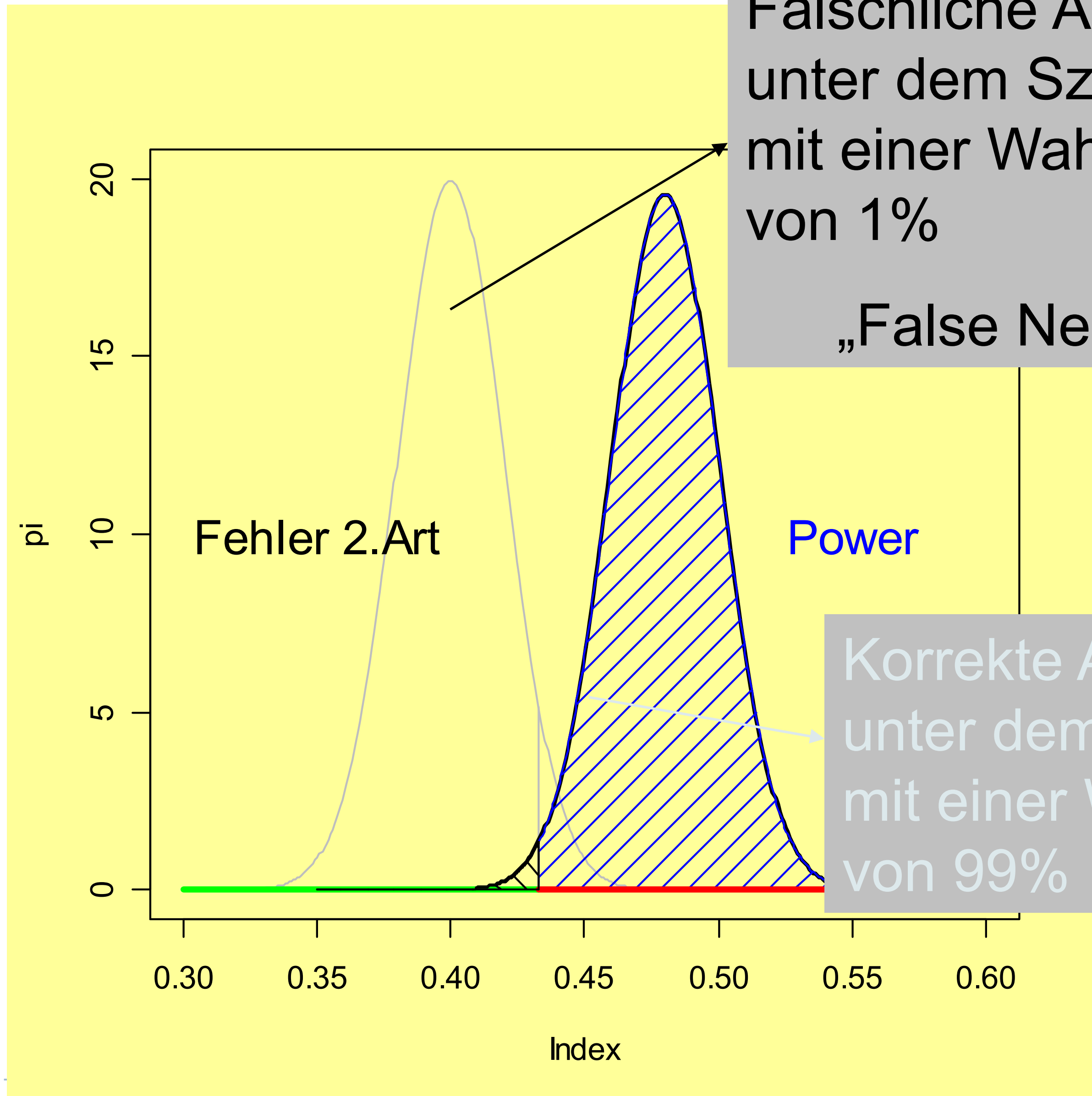
Korrekte Annahme von H_A unter dem Szenario p=45% mit einer Wahrscheinlichkeit von 80%

Berechnung der Gütefunktion

- Angenommen der wahre Anteil ist 48%.
- Wie groß ist die Wahrscheinlichkeit, dass unser Testverfahren, bei diesem Zustand der Grundgesamtheit ein richtiges (H_A) Ergebnis liefert?
- $P(z>1{,}645|\pi=0{,}48)$ bzw. $P(p>43{,}29|\pi=0{,}48)$

$$Power(0,48) = 1 - \Phi\left(\frac{0,4329 - 0,48}{\sqrt{\frac{0,48\cdot(1-0,48)}{600}}}\right) = 1 - \Phi(-2,31) = 0,99$$


Verteilung unter Gültigkeit der Nullhypothese
Verteilung unter dem Szenario π=48%
pi
0
5
10
15
20
0.30
0.35
0.40
0.45
0.50
0.55
0.60
Index



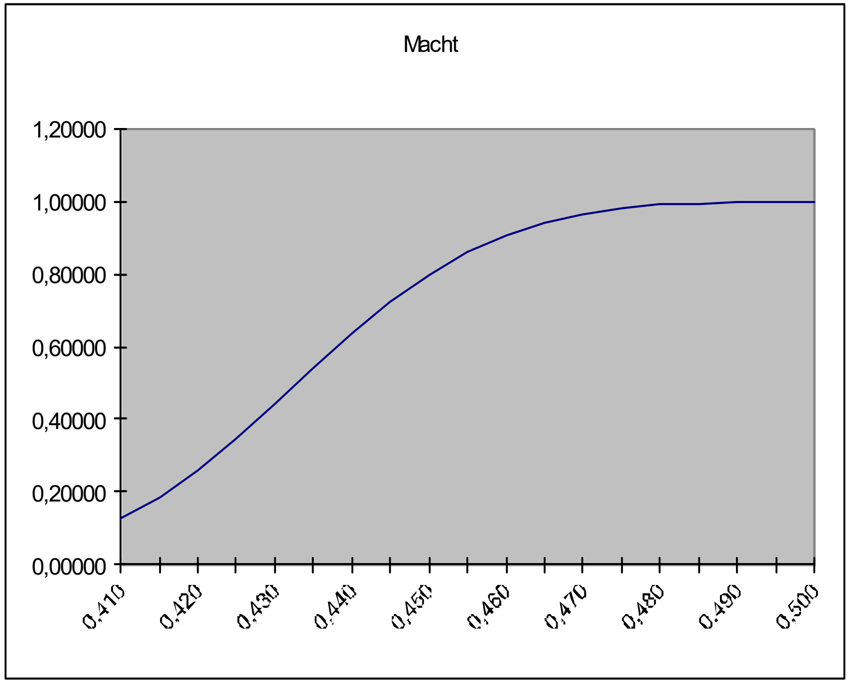
Fälschliche Annahme von H_0 unter dem Szenario p=48% mit einer Wahrscheinlichkeit von 1%

„False Negative Result“

Korrekte Annahme von H_A unter dem Szenario p=48% mit einer Wahrscheinlichkeit von 99%

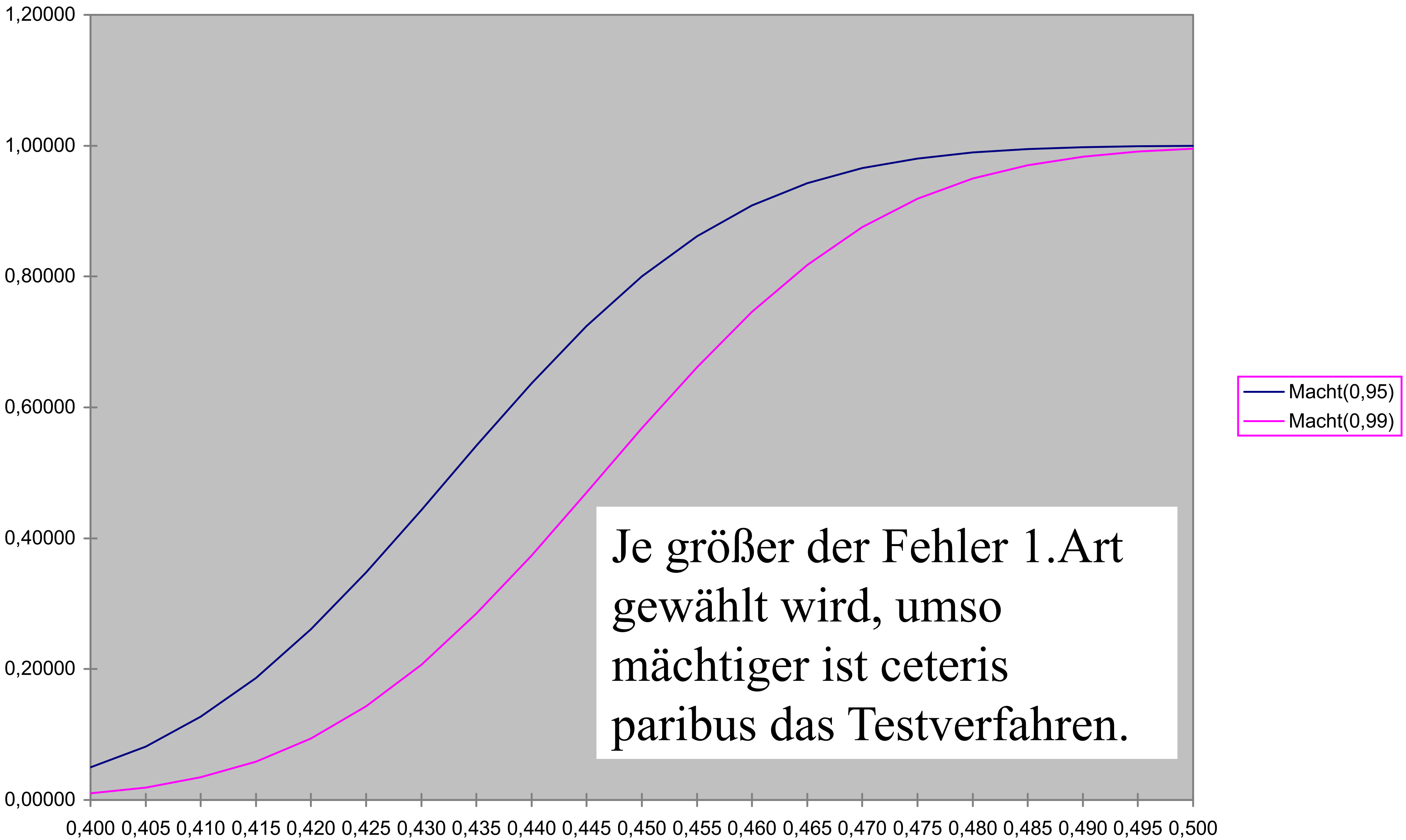
Macht des Tests im Beispiel

π	σ_p	z	Macht
0,410	0,02008	1,14049	0,12704
0,415	0,02012	0,88987	0,18677
0,420	0,02015	0,64022	0,26102
0,425	0,02018	0,39145	0,34773
0,430	0,02021	0,14348	0,44295
0,435	0,02024	–0,10376	0,54132
0,440	0,02026	–0,35036	0,63697
0,445	0,02029	–0,59640	0,72454
0,450	0,02031	–0,84195	0,80009
0,455	0,02033	–1,08709	0,86150
0,495	0,02041	–3,04242	0,99883
0,500	0,02041	–3,28722	0,99949

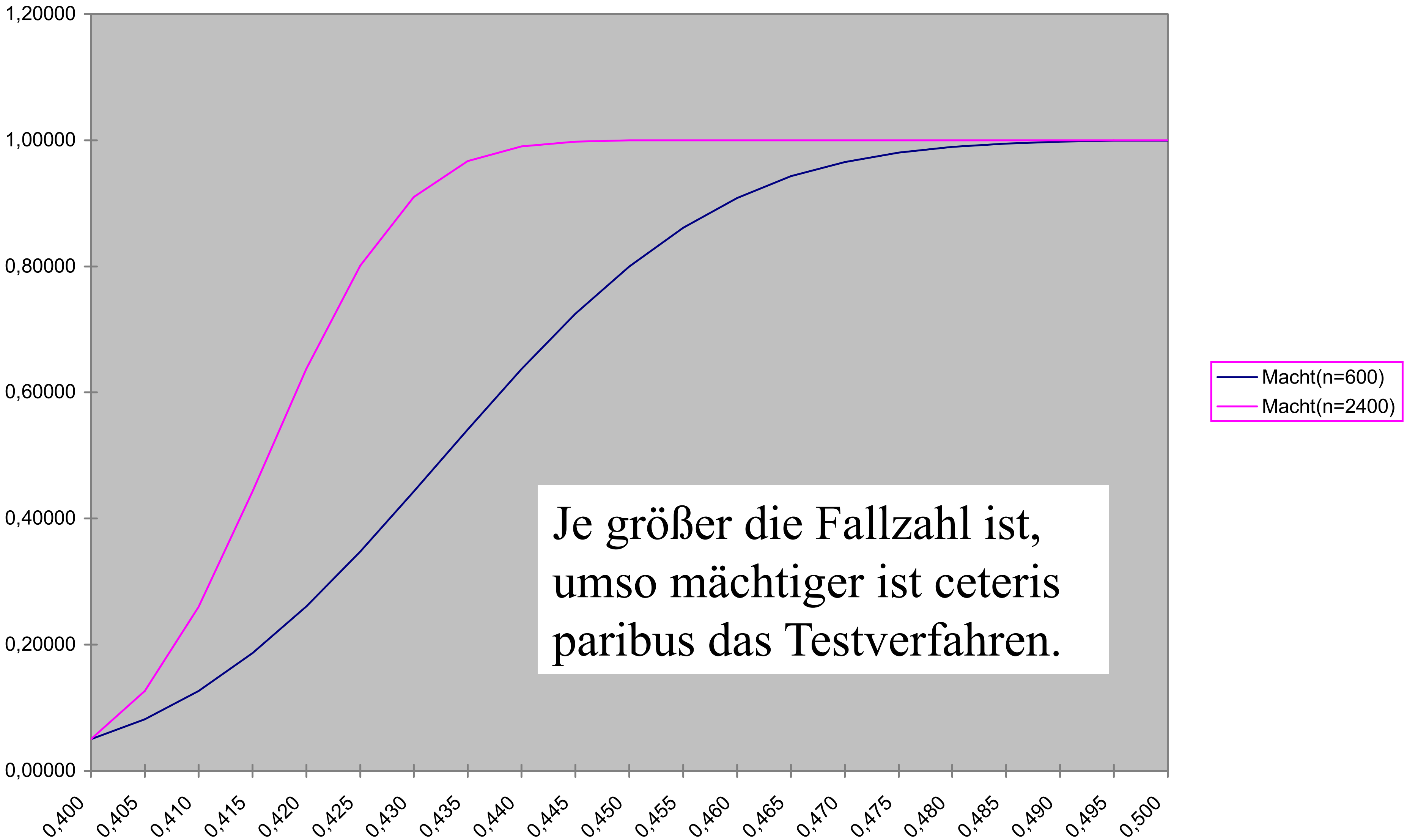


Je stärker der wahre Wert von der Nullhypothese abweicht, desto größer ist ceteris paribus, die Chance diesen Unterschied auch zu entdecken

Macht für verschiedene Werte von α



Macht für unterschiedliche n



Formulierung der Hypothesen

- Die Annahme von H_0 ist in der Regel eine wesentlich unsicherere Entscheidung als deren Ablehnung.
- Interpretation:

  H_0 kann aufgrund der Stichprobe beim gewählten Signifikanzniveau nicht verworfen werden
- Bei der Formulierung einseitiger Alternativen, muss daher jene Hypothese, die man beweisen möchte bzw. jene Hypothese, die mit schwerwiegenderen Konsequenzen verbunden ist als Alternativhypothese formuliert werden.

Beispiel:

- Behauptung einer Pharmafirma:
  In 90% der Fälle bewirkt ein Medikament eine Erleichterung einer allergischen Reaktion (n=200)
- 1) Hypothesenformulierung durch „ARGE Kritische Medizin“
  $H_0: \pi \geq 0{,}9$ $\quad\quad$ $H_1: \pi < 0{,}9$
- 2) Hypothesenformulierung durch Pharmafirma

  $H_0: \pi \leq 0{,}9$ $\quad\quad$ $H_1: \pi > 0{,}9$

Beide wählen eine Irrtumswahrscheinlichkeit $\alpha=0{,}05$

1) $s_p=0{,}021$ $\quad\quad$ $z=1{,}645$ $\quad\quad$ $p_c=0{,}90-0{,}035=0{,}865$
   Falls in Stichprobe weniger als $x_c=173$ Annahme der Alternativhypothese (mit 95% - Wahrscheinlichkeit korrekt)
2) $s_p=0{,}021$ $\quad\quad$ $z=1{,}645$ $\quad\quad$ $p_c=0{,}90+0{,}035=0{,}935$
   Falls in Stichprobe mehr als $x_c=187$ Annahme der Alternativhypothese (mit 95% - Wahrscheinlichkeit korrekt)

Studienplanung

- Eine Pharmafirma hat ein neues Medikament entwickelt, von dem vermutet wird, dass es die Heilungschance bei einer bestimmten Krankheit von 20% (Erfolgschance bei Standardmedikation) auf etwa 30% erhöht.
- Man plant eine Studie mit n=30 (n=100) Patienten.
- Wie groß ist die Chance, den vermuteten Effekt (ca. 10%) mittels der Studie nachweisen zu können?
- Signifikanzniveau α=0,05 bzw. α=0,01

Beispiel Studienplanung (n=30)

$$n = 30 \quad \alpha = 0{,}05$$

$$H_0 : \pi_0 \leq 0{,}2 \quad H_1 : \pi_0 > 0{,}2$$

$$\hat{\sigma}_p = \sqrt{\frac{0{,}2 * 0{,}8}{29}} = 0{,}07$$

p_c, ist der kritische Wert, der erzielt werden, muss damit die H_0 abgelehnt wird

$$p_c = 0{,}2 + 1{,}645 * 0{,}07 = 0{,}322$$

$$Power(0{,}3) = P(p > p_c \mid \pi = 0{,}3) = P(p > 0{,}322 \mid \pi = 0{,}3) =$$

$$1 - \phi\left(\frac{0{,}322 - 0{,}3}{\sqrt{\frac{0{,}3 * 0{,}7}{29}}}\right) = 1 - \phi(0{,}26) = 0{,}40$$

Die Power(0,3) quantifiziert die Wahrscheinlichkeit, dass eine Behandlungsverbesserung von 20 auf 30% mit dieser Studie erkannt werden kann

Beispiel Studienplanung (n=100)

$$n = 100 \quad \alpha = 0{,}05$$

$$H_0 : \pi_0 \leq 0{,}2 \quad H_1 : \pi_0 > 0{,}2$$

$$\hat{\sigma}_p = \sqrt{\frac{0{,}2 * 0{,}8}{100}} = 0{,}04$$

$$p_c = 0{,}2 + 1{,}645 * 0{,}04 = 0{,}266$$

p_c, ist der kritische Wert, der erzielt werden, muss damit die H_0 abgelehnt wird

$$Power(0{,}3) = P(p > p_c \mid \pi = 0{,}3) = P(p > 0{,}266 \mid \pi = 0{,}3) =$$

$$1 - \phi\left(\frac{0{,}266 - 0{,}3}{\sqrt{\frac{0{,}3 * 0{,}7}{100}}}\right) = 1 - \phi(-0{,}75) = 0{,}77$$

Die Wahrscheinlichkeit, dass eine Behandlungsverbesserung von 20 auf 30% mit dieser Studie von 100 Patienten erkannt werden kann, beträgt 77%.

Beispiel Studienplanung (n=100; α=0,01)

$$n = 100 \quad \alpha = 0{,}01$$

$$H_0 : \pi_0 \leq 0{,}2 \quad H_1 : \pi_0 > 0{,}2$$

$$\hat{\sigma}_p = \sqrt{\frac{0{,}2 * 0{,}8}{100}} = 0{,}04$$

$$p_c = 0{,}2 + 2{,}33 * 0{,}04 = 0{,}293$$

$$\text{Power}(0{,}3) = P(p > p_c \mid \pi = 0{,}3) = P(p > 0{,}293 \mid \pi = 0{,}3) =$$

$$1 - \phi\left(\frac{0{,}293 - 0{,}3}{\sqrt{\frac{0{,}3 * 0{,}7}{100}}}\right) = 1 - \phi(-0{,}15) = 0{,}56$$

Die Power sinkt durch, die Verringerung des alpha-Fehlers (von 5% auf 1%) von 77% auf 56%.

Einstichprobentests für das arithmetische Mittel

- $H_0: \mu = \mu_0$ $H_1: \mu \neq \mu_0$ (zweiseitiger Test)
- bzw.
- $H_0: \mu \leq \mu_0$ $H_1: \mu > \mu_0$ (einseitiger Test)
- Unter Gültigkeit von H_0 ist die nachfolgende Teststatistik standardnormalverteilt. Sie kann als standardisierte Abweichung des empirischen Mittelwertes vom Wert der Nullhypothese

$$Z = \frac{\bar{x} - \mu_0}{\sigma_{\bar{x}}} = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$$

$$Z \sim N(0;1)$$

Einstichprobentests für das arithmetische Mittel

- Nur falls die Varianz der Grundgesamtheit bekannt ist kann die Teststatistik unmittelbar berechnet werden
- Beispiel: Produktionsprozess mit Sollwert: 500g
  Aus langjähriger Beobachtung weiß man, dass die Varianz $9g^2$ beträgt.
- Stichprobe: $n = 36 \quad \bar{x} = 501g$
- H_0: $\mu = \mu_0 = 500$ H_1: $\mu \neq 500$ (zweiseitiger Test)
- Unter H_0

$$Z = \frac{\bar{x} - \mu_0}{\sigma_{\bar{x}}} = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$$

$$Z \sim N(0;1)$$

Beispiel:

- α=0,05 z=1,96
- Kritischer Bereich (Ablehnungsbereich):

  [- ∞,-1,96] ∪ [1,96,+ ∞]
- Annahmebereich:

  [-1,96,+1,96]

$$\sigma_{\overline{x}} = 3/\sqrt{36} = 0,5$$

$$Z = \frac{501-500}{0,5} = 2$$

- Interpretation: Da der Mittelwert einer Stichprobe von 36 Verpackungseinheiten 501 beträgt, wird die Nullhypothese verworfen, und man kann mit einer 95%-igen Sicherheit davon ausgehen, dass die Norm in der Produktion nicht eingehalten wird.

Beispiel:

- Annahmebereich in der Originalskala:

$$\mu_0 \pm z_{1-\alpha/2} \cdot \sigma / \sqrt{n}$$

$$500 \pm 1,96 \cdot 0,5 = 500 \pm 0,98$$

- d.h. für unser obiges Beispiel ergibt sich ein Toleranzbereich von 499,02 bis 500,98 für das durchschnittliche Gewicht bei einer Losgröße von 36.

Einstichprobentests für das arithmetische Mittel

- In der Praxis wesentlich bedeutsamer ist der Fall der unbekannten Varianz der Grundgesamtheit
- Einsetzen der Schätzung für die Varianz auf Basis der Stichprobe ➔ t-Verteilung
- Unter H_0:

$$T = \frac{\bar{x} - \mu_0}{\hat{\sigma}_{\bar{x}}} = \frac{\bar{x} - \mu_0}{\hat{\sigma} / \sqrt{n}}$$

$$\hat{\sigma} = s = \sqrt{1/(n-1) \sum_{i=1}^{n} (x_i - \bar{x})^2}$$

$$T \sim t_{n-1}$$

Beispiel:

- Produktionsprozess mit Sollwert: 500g

  Stichprobe:

Nr	Gewicht	Nr	Gewicht	Nr	Gewicht
1	502,6	13	502,7	25	501,3
2	500,5	14	498,3	26	500,2
3	502,6	15	502,0	27	500,8
4	501,5	16	500,5	28	503,4
5	498,9	17	502,3	29	500,6
6	498,7	18	502,7	30	502,8
7	499,3	19	502,8	31	499,0
8	500,0	20	502,8	32	500,5
9	503,1	21	499,6	33	498,5
10	501,8	22	500,1	34	502,4
11	498,8	23	501,0	35	500,6
12	503,6	24	500,1	36	501,2

- $H_0: \mu = \mu_0 = 500$ $\qquad H_1: \mu \neq 500$
- $\alpha = 0{,}05$ $\qquad t_{0,975;35} = 2{,}03$
- Kritischer Bereich (Ablehnungsbereich):
- $[-\infty; -2{,}03] \cup [2{,}03; +\infty]$
- Annahmebereich: $[-2{,}03; +2{,}03]$

Beispiel:

Nr	Gewicht	Nr	Gewicht	Nr	Gewicht
1	502,6	13	502,7	25	501,3
2	500,5	14	498,3	26	500,2
3	502,6	15	502,0	27	500,8
4	501,5	16	500,5	28	503,4
5	498,9	17	502,3	29	500,6
6	498,7	18	502,7	30	502,8
7	499,3	19	502,8	31	499,0
8	500,0	20	502,8	32	500,5
9	503,1	21	499,6	33	498,5
10	501,8	22	500,1	34	502,4
11	498,8	23	501,0	35	500,6
12	503,6	24	500,1	36	501,2

Mittelwert	501,04
Varianz der Stichporbe	2,41
Standardabweichung	1,55
Standardfehler	0,26
Teststatistik	4,03
kritischer Wert bei α =0,05	2,03

$$\hat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

$$\hat{\sigma}/\sqrt{n}$$

$$T = \frac{501-500}{0,26} = 4,03$$

<u>Interpretation:</u>
Da der Mittelwert einer Stichprobe von 36 Verpackungseinheiten 501g beträgt, wird die Nullhypothese verworfen, und man kann mit einer 95%-igen Sicherheit davon ausgehen, dass die Norm in der Produktion nicht eingehalten wird.

Excel-Sheet

Stichprobenergebnisse

n=	36	
xq=	501,04	
s=	1,55	
s(xq)=	0,258	

Irrtumswahrscheinlichkeit

alpha=	0,05	
Freiheitsgrade=	35	
kritischer t-Wert=	2,0301	bei zweiseitigem Test
kritischer t-Wert=	1,6896	bei einseitigem Test

Hypothese über den Erwartungswert

H0:	500,00	

Wert der Teststatistik:	4,03	
p-value=	0,0001	bei einseitigem Test
p-value=	0,0003	bei zweiseitigem Test

Legende:

xq...arithmetisches Mittel (x-quer)
s...Standardabweichung der Stichprobenwerte
s(xq)…Standardfehler

Beispiel

- Canadian Survey of Labour and Income Dynamics
- Gesamt-Mittelwert für den Stundenlohn beträgt 15,55 $
- Im Rahmen einer Erhebung in einer Region wird eine Stichprobe von n=200 Angestellten befragt um festzustellen, ob sich der Lohn in dieser Region signifikant vom Gesamtmittel unterscheidet.

  H_0: Lohnniveau gleich H_A: Lohnniveau ungleich

  H_0: $\mu = \mu_0 = 15{,}55$ H_1: $\mu \neq 15{,}55$ (zweiseitiger Test)

  Mittelwert der Stichprobe: 14,86   Standardabweichung: 7,2

  Test-Statistik: -1,36   p-value (2-seitiger Test): 0,1769

Beispiel

Stichprobenergebnisse

n=	200	
xq=	14,86	
s=	7,2	
s(xq)=	0,509	

Irrtumswahrscheinlichkeit

alpha=	0,05	
Freiheitsgrade=	199	
kritischer t-Wert=	1,9720	bei zweiseitigem Test
kritischer t-Wert=	1,6525	bei einseitigem Test

Hypothese über den Erwartungswert

H0:	15,55	

Wert der Teststatistik:	-1,36	
p-value=	0,0884	bei einseitigem Test
p-value=	0,1769	bei zweiseitigem Test

Legende:

xq...arithmetisches Mittel (x-quer)
s...Standardabweichung der Stichprobenwerte
s(xq)…Standardfehler

Beispiel

- In einer empirischen Studie werden 10 Raucher befragt, wie viele Zigaretten sie täglich rauchen:
- Daten:
- 26, 34, 5, 20, 50, 44, 18, 39, 29, 19
- Überprüfen Sie die Hypothese, dass durchschnittlich mehr als 25 Zigaretten täglich geraucht werden mit einem $\alpha=0{,}01$.
- $H_0: \mu_0 \leq 25$ $H_1: \mu_0 > 25$

Berechnung

$$T = \frac{\overline{x} - \mu_0}{\hat{\sigma}_{\overline{x}}} = \frac{\overline{x} - \mu_0}{\hat{\sigma} / \sqrt{n}} \qquad \hat{\sigma} = s = \sqrt{1/(n-1)\left(\sum_{i=1}^{n} x_i^2 - n \cdot \overline{x}^2\right)}$$

$$\overline{x} = 28,4, \quad \sum x_i^2 = 9740 \quad s^2 = \frac{1}{9}\left(9740 - 10 \cdot 28,4^2\right) = 186,04$$

$$s = 13,64 \qquad T = \frac{28,4 - 25}{13,64 / \sqrt{10}} = 0,789$$

- Tabellenwert $t_{(0,99;9)}$=2,82 **einseitiger** Test
- offensichtlich kein signifikantes Ergebnis ➔ Die Nullhypothese wird beibehalten