

10. Vorlesung Statistik I

Einführung in statistische Hypothesentests



We are happy to share our materials openly:

The content of these Open Educational Resources by Lehrstuhl für Psychologische Methodenlehre und Diagnostik, Ludwig-Maximilians-Universität München is licensed under CC BY-SA 4.0. The CC Attribution-ShareAlike 4.0 International license means that you can reuse or transform the content of our materials for any purpose as long as you cite our original materials and share your derivatives under the same license.

Zwischengliederung

- Bislang:
 - Parameterschätzung
- Jetzt:
 - Statistische Hypothesentests

Statistische Hypothesentests

- Je nach Fragestellung, d.h. je nach Art der Aussage, die über einen Parameter getroffen werden soll, lassen sich inferenzstatistische Methoden zwei Gebieten zuordnen:
- **Parameterschätzung:**
 - Fragestellung: Welchen konkreten Wert hat ein Parameter?
 - Beispiele:
 - Welchen Wert hat π ?
 - Welche Werte haben μ und σ^2 ?
 - Resultate von Parameterschätzungen sind konkrete Zahlen oder Intervalle.
- **Statistische Hypothesentests:**
 - Fragestellung: Entspricht ein Parameter einem vorgegebenen Wert oder liegt er in einem vorgegebenen Bereich?
 - Beispiele:
 - Ist $\pi = 0.5$?
 - Ist $\mu = 100$?
 - Ist $\sigma^2 > 10$?
 - Das Resultat eines statistischen Hypothesentests ist eine Ja/Nein-Entscheidung.

Statistische Hypothesen

- Idealisiertes Vorgehen in der empirischen Psychologie:
 - Wir haben eine psychologische Theorie.
 - Aus dieser Theorie leiten wir eine empirisch überprüfbare Aussage ab, die wahr sein müsste, falls unsere Theorie wahr ist. Aussagen dieser Art werden auch **Hypothesen** genannt.
 - Wir überprüfen diese Hypothese empirisch.
- Zwei mögliche Ausgänge:
 - Wir stellen empirisch fest, dass die aus der Theorie abgeleitete Hypothese nicht wahr ist. In diesem Fall können wir logisch darauf schließen, dass unsere Theorie auch nicht wahr sein kann. Wir haben unsere Theorie **falsifiziert**.
 - Wir stellen empirisch fest, dass die aus der Theorie abgeleitete Hypothese wahr ist. In diesem Fall können wir **nicht** logisch darauf schließen, dass unsere Theorie auch wahr ist. Wir haben unsere Theorie lediglich **nicht falsifiziert**.

- Bemerkung:
 - Hypothesen können **verifiziert und falsifiziert** werden: Wir können empirisch feststellen, ob sie wahr oder falsch sind (zumindest prinzipiell).
 - Theorien können (in den meisten Fällen) **nur falsifiziert** werden. Wir können empirisch nur feststellen, ob sie falsch sind (mithilfe der Überprüfung der aus ihnen abgeleiteten Hypothesen), aber nicht, ob sie wahr sind.
 - Hypothesentests können auch benutzt werden, um direkt Aussagen über die Population zu überprüfen, ohne dass diese Aussagen aufwendig aus einer psychologischen Theorie abgeleitet wurden.

- Fast immer sind die empirisch überprüfbaren Hypothesen, die aus psychologischen Theorien abgeleitet werden können, Aussagen über deskriptivstatistische Maße in Populationen.
- Hypothesen dieser Art nennen wir **inhaltliche Hypothesen**.
- Beispiele für inhaltliche Hypothesen:
 - Die relative Häufigkeit der Depression in der Population ist größer als 30%:
 $h_D > 0.3$.
 - Die mittlere Reaktionszeit von depressiven Patient*innen auf einen Stimulus ist höher als die von gesunden Personen: $\bar{x}_{RT\text{ Depression}} > \bar{x}_{RT\text{ Normal}}$
- Inhaltliche Hypothesen können wir im Prinzip sowohl verifizieren als auch falsifizieren: Falls wir alle Personen in der jeweiligen Population untersuchen, können wir ohne Zweifel feststellen, ob z.B. die Hypothese $h_D > 0.3$ wahr oder nicht wahr ist.
- Aber: In den meisten Fällen können wir nicht alle Personen in der Population untersuchen.
- Was können wir in diesem Fall tun?

- Wir ziehen einfache Zufallsstichproben, formalisieren die zufällige Ziehung mithilfe der Wahrscheinlichkeitstheorie und greifen auf inferenzstatistische Verfahren zurück.
- Da unter bestimmten Voraussetzung die Parameter der Wahrscheinlichkeitsverteilungen der Zufallsvariablen X_1, X_2, \dots, X_n in unserer Stichprobe den uns interessierenden deskriptivstatistischen Maßzahlen in der Population entsprechen, können wir die inhaltlichen Hypothesen in Hypothesen über diese Parameter überführen.
- Hypothesen über Parameter von Wahrscheinlichkeitsverteilungen nennen wir **statistische Hypothesen**.
- Beispiel: Die inhaltliche Hypothese $h_D > 0.3$ ist aufgrund von $\pi = h_D$ im Fall einer einfachen Zufallsstichprobe äquivalent zur statistischen Hypothese $\pi > 0.3$ über den Parameter π der Bernoulli-Verteilung: Falls $\pi > 0.3$ wahr ist, ist auch $h_D > 0.3$ wahr. Falls $\pi > 0.3$ nicht wahr ist, ist auch $h_D > 0.3$ nicht wahr.

- Das heißt:
 - Falls wir eine einfache Zufallsstichprobe gezogen haben und mithilfe inferenzstatistischer Verfahren z.B. zu dem Schluss kommen, dass die statistische Hypothese $\pi > 0.3$ wahr ist, müssen wir wegen $\pi = h_D$ auch davon ausgehen, dass die inhaltliche Hypothese $h_D > 0.3$ wahr ist.
 - Falls wir eine einfache Zufallsstichprobe gezogen haben und mithilfe inferenzstatistischer Verfahren z.B. zu dem Schluss kommen, dass die statistische Hypothese $\pi > 0.3$ nicht wahr ist, müssen wir wegen $\pi = h_D$ auch davon ausgehen, dass die inhaltliche Hypothese $h_D > 0.3$ nicht wahr ist.

- Sowohl bei inhaltlichen als auch bei statistischen Hypothesen unterscheidet man zudem zwischen **Nullhypothesen** und **Alternativhypothesen**:
 - Die **Alternativhypothese** ist (fast immer) diejenige Hypothese, die aus der Theorie folgt, die wir empirisch überprüfen wollen.
 - Die **Nullhypothese** ist diejenige Hypothese, die sich aus der Negation der Alternativhypothese ergibt. Sie ist somit das „Gegenteil“ der Alternativhypothese.
- Notation:
 - Nullhypothesen werden mit H_0 bezeichnet.
 - Alternativhypothesen werden mit H_1 bezeichnet.

- Beispiel: Aus unserer Theorie könnte z.B. folgen, dass die relative Häufigkeit h_D der Depression in der Population größer als 0.3 sein sollte.
 - Inhaltliche Alternativhypothese: $H_1: h_D > 0.3$
 - Inhaltliche Nullhypothese: $H_0: h_D \leq 0.3$
 - Statistische Alternativhypothese: $H_1: \pi > 0.3$
 - Statistische Nullhypothese: $H_0: \pi \leq 0.3$
- Falls wir feststellen, dass die statistische und somit auch die inhaltliche Nullhypothese wahr ist bzw. die statistische und somit auch die inhaltliche Alternativhypothese falsch ist, haben wir unsere Theorie falsifiziert.

- Sowohl bei inhaltlichen als auch bei statistischen Hypothesen unterscheidet man außerdem zwischen ungerichteten und gerichteten Hypothesen.
- **Ungerichtete Hypothesen** treffen Aussagen darüber, ob eine deskriptivstatistische Maßzahl oder ein Parameter **gleich oder ungleich** einem bestimmten Wert ist.
 - Beispiele:
 - $\bar{x}_{Pop} = 100$
 - $\pi \neq 0.5$
 - $\mu_1 - \mu_2 = 0$
- **Gerichtete Hypothesen** treffen Aussagen darüber, ob eine deskriptivstatistische Maßzahl oder ein Parameter **kleiner, kleiner-gleich, größer oder größer-gleich** einem bestimmten Wert ist.
 - Beispiele:
 - $\bar{x}_{Pop} > 100$
 - $\pi \leq 0.5$
 - $\mu_1 - \mu_2 < 0$

Ungerichtete und gerichtete Hypothesen II

- Bei gerichteten Hypothesen spricht man manchmal von **linksgerichteten** (kleiner oder kleiner-gleich) und **rechtsgerichteten** Hypothesen (größer oder größer-gleich).
 - Beispiele linksgerichtet:
 - $\bar{x}_{Pop} < 100$
 - $\pi \leq 0.5$
 - $\mu_1 - \mu_2 < 0$
 - Beispiele rechtsgerichtet:
 - $\mu \geq 100$
 - $\bar{x}_{Pop} > 0.5$
 - $\mu_1 - \mu_2 > 0$

Zusammenfassung

- Wir unterscheiden zwischen
 - inhaltlichen Hypothesen und statistischen Hypothesen
 - Nullhypothesen und Alternativhypothesen
 - ungerichteten und gerichteten Hypothesen

- Unter statistischen Hypothesentests versteht man alle inferenzstatistischen Verfahren, die auf der Basis einer Stichprobe zu einer Entscheidung für eine statistische H_0 oder eine statistische H_1 führen.
- Da wir durch unsere Stichprobenziehung sicherstellen, dass die Parameter in unseren statistischen Hypothesen den deskriptivstatistischen Maßzahlen in der Population aus unseren inhaltlichen Hypothesen entsprechen, impliziert eine Entscheidung für eine statistische H_0 oder H_1 im Rahmen statistischer Hypothesentests immer auch eine Entscheidung für eine inhaltliche H_0 oder H_1 :
 - Falls wir uns z.B. für die statistische $H_0: \mu = 0$ entscheiden, entscheiden wir uns wegen $\mu = \bar{x}_{Pop}$ auch für die inhaltliche $H_0: \bar{x}_{Pop} = 0$.
 - Falls wir uns z.B. für die statistische $H_1: \mu \neq 0$ entscheiden, entscheiden wir uns wegen $\mu = \bar{x}_{Pop}$ auch für die inhaltliche $H_1: \bar{x}_{Pop} \neq 0$.

- Wie können wir auf der Basis einer einfachen Zufallsstichprobe zu einer sinnvollen Entscheidung entweder für die statistische H_0 oder für die statistische H_1 gelangen?
- Wir benötigen eine Entscheidungsregel dafür, in welchen Stichproben wir uns für die statistische H_0 und in welchen Stichproben wir uns für die statistische H_1 entscheiden sollen.
- Diese Entscheidungsregel wird je nach Parameter, je nach Art der Hypothese und je nach Art der Stichprobenziehung anders aussehen.
- **Um das allgemeine Prinzip statistischer Hypothesentests zu veranschaulichen, beginnen wir mit dem einfachsten Fall:**
 - Hypothesentests für ungerichtete statistische Hypothesen über μ

Hypothesentests für ungerichtete statistische Hypothesen über μ

- Aus einer psychologischen Theorie könnte folgen, dass der Mittelwert \bar{x}_{Pop} einer Variable in einer Population ungleich einem bestimmten Wert μ_0 sein sollte.
- μ_0 ist hierbei eine Zahl, z.B. $\mu_0 = 0$ oder $\mu_0 = 100$, deren genauer Wert sich aus der Theorie ergeben muss.
- Wir hätten also die folgenden ungerichteten inhaltlichen Hypothesen:

$$H_0: \bar{x}_{Pop} = \mu_0$$

$$H_1: \bar{x}_{Pop} \neq \mu_0$$

- Falls das Histogramm der Variable in der Population durch die Wahrscheinlichkeitsdichtefunktion der Normalverteilung approximiert werden kann und wir eine einfache Zufallsstichprobe mit Umfang n aus dieser Population ziehen, haben wir die folgende Ausgangssituation:

$$X_1, X_2, \dots, X_n \text{ mit } X_i \stackrel{\text{iid}}{\sim} N(\mu, \sigma^2)$$

- Da wir wissen, dass in diesem Fall $\bar{x}_{pop} = \mu$ ist, können wir die inhaltlichen Hypothesen in äquivalente statistische Hypothesen überführen:

$$H_0: \mu = \mu_0$$

$$H_1: \mu \neq \mu_0$$

- Wir wollen also überprüfen, ob μ einem vorgegebenen Wert μ_0 entspricht oder nicht.
- Wie können wir auf der Basis einer einfachen Zufallsstichprobe zu einer sinnvollen Entscheidung entweder für die statistische H_0 oder für die statistische H_1 gelangen?

- Wir wissen, dass wir als Schätzwert $\hat{\mu}_{Wert}$ für μ den Mittelwert \bar{x} in der Stichprobe verwenden können.
- **Idee:** Wir vergleichen diesen Schätzwert mit dem Wert μ_0 , dem μ unter der H_0 entsprechen sollte.
 - Falls der Schätzwert $\hat{\mu}_{Wert} = \bar{x}$ weit weg von μ_0 liegt, spricht dies eher dafür, dass die $H_1: \mu \neq \mu_0$ wahr ist.

Beispiel: $\hat{\mu}_{Wert} = \bar{x} = 100, \mu_0 = 0$

- Falls der Schätzwert $\hat{\mu}_{Wert} = \bar{x}$ nicht weit weg von μ_0 liegt, können wir eventuell davon ausgehen, dass die $H_0: \mu = \mu_0$ wahr ist.

Beispiel: $\hat{\mu}_{Wert} = \bar{x} = 0.01, \mu_0 = 0$

- **Aber:** Ab wann ist der Schätzwert $\hat{\mu}_{Wert}$ „weit weg“ bzw. „nicht weit weg“ von μ_0 ? Diese Einschätzung wäre ja auch von der Einheit von \bar{x} abhängig.

- Wir wissen, dass unser Schätzwert $\hat{\mu}_{Wert} = \bar{x}$ für μ die Realisation der Schätzfunktion

$$\hat{\mu} = \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

ist.

- Zusätzlich kennen wir die Wahrscheinlichkeitsverteilung von \bar{X} (VL 7, Folie 31):

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

- Falls die Nullhypothese $H_0: \mu = \mu_0$ tatsächlich wahr ist, gilt also:**

$$\bar{X} \sim N\left(\mu_0, \frac{\sigma^2}{n}\right)$$

Beispiel: Falls die $H_0: \mu = 0$ wahr ist, ist $\bar{X} \sim N\left(0, \frac{\sigma^2}{n}\right)$ verteilt.

- Das heißt: Wir könnten auf der Basis der Verteilungsfunktion F von

$$N\left(\mu_0, \frac{\sigma^2}{n}\right)$$

einen Bereich festlegen, bei dem es, **falls $H_0: \mu = \mu_0$ tatsächlich wahr ist, sehr unwahrscheinlich** wäre, dass sich die Schätzfunktion \bar{X} in einem Schätzwert \bar{x} realisiert, der in diesem Bereich liegt.

- „Weit weg“ von μ_0 hieße dann:

Sehr unwahrscheinlich, falls $H_0: \mu = \mu_0$ tatsächlich wahr sein sollte.

- Konkret: Wir könnten zwei **kritische** Werte \bar{x}_{krit_links} und \bar{x}_{krit_rechts} bestimmen, so dass **unter der H_0**

$$P(\bar{X} \leq \bar{x}_{krit_links}) = F(\bar{x}_{krit_links}) = \frac{\alpha}{2}$$

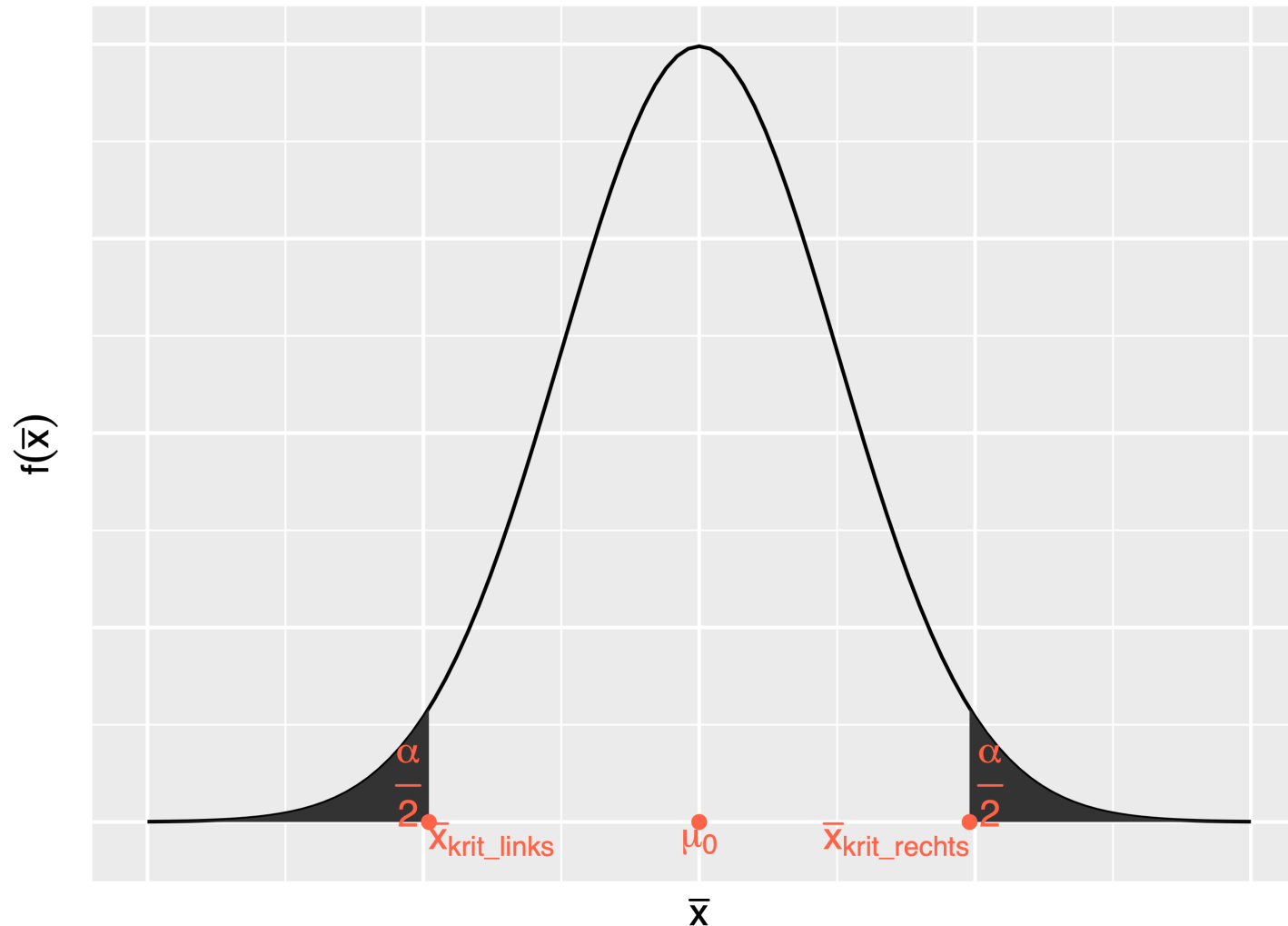
und

$$P(\bar{X} \geq \bar{x}_{krit_rechts}) = 1 - F(\bar{x}_{krit_rechts}) = \frac{\alpha}{2}$$

für eine vorgegebene (kleine) Wahrscheinlichkeit α ist.

- Graphische Veranschaulichung:

Wahrscheinlichkeitsdichtefunktion für \bar{X} falls $\mu = \mu_0$



- Der Bereich $K_{\bar{X}} =]-\infty, \bar{x}_{krit_links}] \cup [\bar{x}_{krit_rechts}, +\infty[$, also der Bereich links und rechts von den beiden kritischen Werten, wird **kritischer Bereich** genannt. Dies ist der Bereich, bei dem wir sagen würden, dass der Schätzwert “zu weit weg” von μ_0 wäre.
- Falls wir beispielsweise $\alpha = 0.005$ festlegen und auf der Basis von diesem α die kritischen Werte \bar{x}_{krit_links} , \bar{x}_{krit_rechts} und somit den kritischen Bereich bestimmen, bedeutet dies, dass **unter der H_0** die Wahrscheinlichkeit, dass sich die Schätzfunktion im kritischen Bereich realisiert, gleich 0.005 ist.

$$P(\bar{X} \in K_{\bar{X}}) = P(\bar{X} \leq \bar{x}_{krit_links}) + P(\bar{X} \geq \bar{x}_{krit_rechts}) = \frac{\alpha}{2} + \frac{\alpha}{2} = \alpha = 0.005$$

- Unsere Entscheidungsregel könnte dann wie folgt aussehen:
 - Falls der Schätzwert $\hat{\mu}_{wert} = \bar{x}$ für μ **im kritischen Bereich liegt**, also in einem Bereich, der unter der $H_0: \mu = \mu_0$ sehr unwahrscheinlich wäre, **entscheiden wir uns für die H_1** .
 - Falls der Schätzwert $\hat{\mu}_{wert} = \bar{x}$ für μ **nicht im kritischen Bereich liegt**, sondern in einem Bereich, der unter der $H_0: \mu = \mu_0$ relativ wahrscheinlich wäre, **entscheiden wir uns für die H_0** .

- Problem: Wir können die kritischen Werte \bar{x}_{krit_links} und \bar{x}_{krit_rechts} nur dann bestimmen, falls wir σ^2 kennen. Leider kennen wir σ^2 nicht.
- Aber: Wir kennen unabhängig von σ^2 die Wahrscheinlichkeitsverteilung von

$$T = \frac{\bar{X} - \mu}{\sqrt{\frac{S^2}{n}}}$$

- T folgt einer t -Verteilung mit Parameter $\nu = n - 1$ (siehe VL 7, Folie 56).
- Falls die $H_0: \mu = \mu_0$ tatsächlich wahr ist, folgt also die Zufallsvariable

$$T = \frac{\bar{X} - \mu_0}{\sqrt{\frac{S^2}{n}}}$$

einer t -Verteilung mit Parameter $\nu = n - 1$.

- Wir können daher den kritischen Bereich auf Basis dieser t -Verteilung bestimmen.

- Wir können auf der Basis der Verteilungsfunktion F der Zufallsvariable

$$T = \frac{\bar{X} - \mu_0}{\sqrt{\frac{S^2}{n}}}$$

einen Bereich festlegen, bei dem es, falls $H_0: \mu = \mu_0$ tatsächlich wahr ist, **sehr unwahrscheinlich** wäre, dass sich die Zufallsvariable in diesem realisiert.

- Konkret: Wir könnten zwei Werte t_{krit_links} und t_{krit_rechts} bestimmen, so dass **unter der H_0**

$$P(T \leq t_{krit_links}) = F(t_{krit_links}) = \frac{\alpha}{2}$$

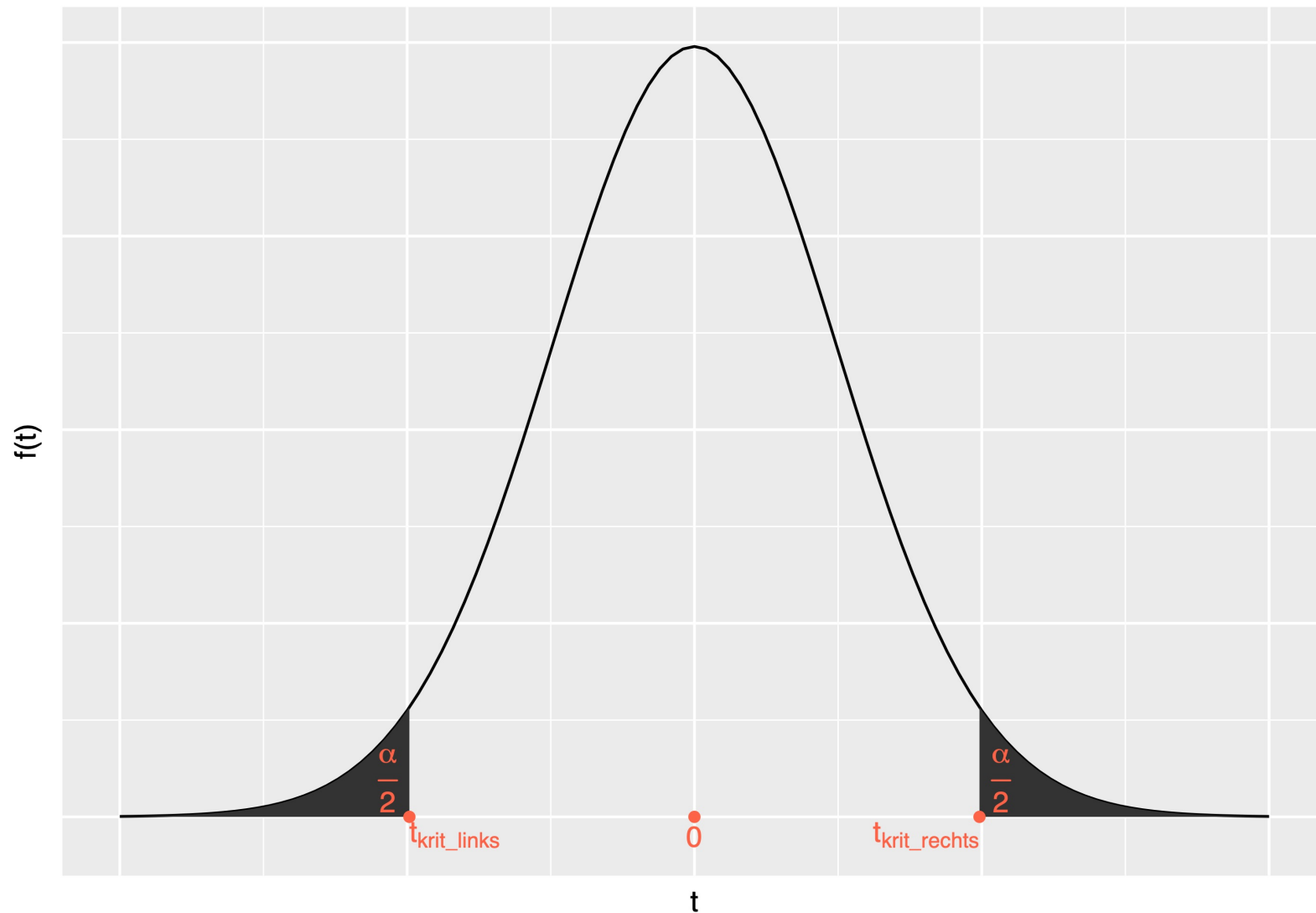
und

$$P(T \geq t_{krit_rechts}) = 1 - F(t_{krit_rechts}) = \frac{\alpha}{2}$$

für eine vorgegebene (kleine) Wahrscheinlichkeit α ist.

Problem und Lösung III

- Graphische Veranschaulichung:



- Der kritische Bereich ist in diesem Fall $K_T =]-\infty, t_{krit_links}] \cup [t_{krit_rechts}, +\infty[$, also der Bereich links und rechts von den beiden kritischen Werten.
- Falls wir beispielsweise ein $\alpha = 0.005$ festlegen und auf der Basis von diesem α die kritischen Werte t_{krit_links} und t_{krit_rechts} somit den kritischen Bereich K_T bestimmen, bedeutet dies, dass **unter der H_0** die Wahrscheinlichkeit dafür, dass sich die Zufallsvariable T im kritischen Bereich realisiert, gleich 0.005 ist.

$$P(T \in K_T) = P(T \leq t_{krit_links}) + P(T \geq t_{krit_rechts}) = \frac{\alpha}{2} + \frac{\alpha}{2} = \alpha = 0.005$$

- Bemerkung: Aufgrund der Symmetrie der t-Verteilung ist $t_{krit_rechts} = -t_{krit_links}$.

- Unsere Entscheidungsregel sieht dann wie folgt aus:
 - Falls die Realisation t der Zufallsvariable T **im kritischen Bereich liegt**, also in dem Bereich, der unter der $H_0: \mu = \mu_0$ sehr unwahrscheinlich wäre, **entscheiden wir uns für die H_1** .
 - Falls die Realisation t der Zufallsvariable T **nicht im kritischen Bereich liegt**, sondern in einem Bereich, der unter der $H_0: \mu = \mu_0$ relativ wahrscheinlich wäre, **entscheiden wir uns für die H_0** .
- Bemerkung: Falls wir uns für die H_1 entscheiden, sagen wir auch, dass das Ergebnis des statistischen Hypothesentests **signifikant** ist.

- Eine Zufallsvariable wird **Teststatistik** genannt, falls wir auf der Basis ihrer Realisation unsere Testentscheidung treffen.
- Im Fall von ungerichteten Hypothesen für μ ist also

$$T = \frac{\bar{X} - \mu_0}{\sqrt{\frac{S^2}{n}}}$$

die Teststatistik.

- Anforderungen an Teststatistiken:
 - Wir können ihre Realisation aus den Stichprobendaten berechnen.
 - Wir kennen ihre Wahrscheinlichkeitsverteilung unter der H_0 und können daher für eine vorgegebene Wahrscheinlichkeit α den kritischen Bereich bestimmen.
 - Der durch die Teststatistik definierte Hypothesentest erfüllt bestimmte Gütekriterien, die wir später noch besprechen werden.

Definition Signifikanzniveau

- Die vorgegebene Wahrscheinlichkeit α wird **Signifikanzniveau** genannt.
- Sie entspricht der Wahrscheinlichkeit dafür, dass sich die Teststatistik im kritischen Bereich realisiert, falls die H_0 wahr ist, also der Wahrscheinlichkeit dafür, dass wir uns fälschlicherweise für die H_1 entscheiden, obwohl die H_0 wahr ist.
- Das Signifikanzniveau α sollte daher auf jeden Fall niedrig sein. Welchen Wert wir genau für α wählen, werden wir später noch diskutieren.

Einstichproben t-Test für ungerichtete Hypothesen über μ

- Der besprochene statistische Hypothesentest mit der Teststatistik

$$T = \frac{\bar{X} - \mu_0}{\sqrt{\frac{S^2}{n}}} \stackrel{H_0}{\sim} t(n-1)$$

und dem kritischen Bereich

$$K_T =]-\infty, t_{krit_links}] \cup [t_{krit_rechts}, +\infty[$$

wird **Einstichproben t-Test für ungerichtete Hypothesen über μ** genannt.

Beispiel

Beispiel I

- Aus unserer Theorie folgt, dass der durchschnittliche IQ \bar{x}_{IQ_Pop} in der Population der Studierenden ungleich 100 sein sollte.
- In diesem Fall ist $\mu_0 = 100$ und die inhaltlichen Hypothesen lauten:

$$H_0: \bar{x}_{IQ_Pop} = 100$$

$$H_1: \bar{x}_{IQ_Pop} \neq 100$$

- Wir nehmen an, dass das Histogramm des IQ in der Population der Studierenden durch die Wahrscheinlichkeitsdichtefunktion einer Normalverteilung approximiert werden kann und ziehen eine einfache Zufallsstichprobe der Größe $n = 500$.
- In diesem Fall ist $\bar{x}_{IQ_Pop} = \mu$ und wir erhalten die folgenden statistischen Hypothesen:

$$H_0: \mu = 100$$

$$H_1: \mu \neq 100$$

Beispiel II

- Unsere Teststatistik ist

$$T = \frac{\bar{X} - \mu_0}{\sqrt{\frac{S^2}{n}}} = \frac{\bar{X} - 100}{\sqrt{\frac{S^2}{500}}}$$

und wir wissen, dass diese Teststatistik unter der H_0 einer t-Verteilung mit

$$\nu = n - 1 = 500 - 1 = 499$$

folgt.

- Die kritischen Werte können wir auf Basis dieser t-Verteilung in R berechnen.
- Wir wählen ein Signifikanzniveau von $\alpha = 0.005$.
- t_{krit_links} ist derjenige Wert, für den $F(t_{krit_links}) = \frac{0.005}{2} = 0.0025$ ist:

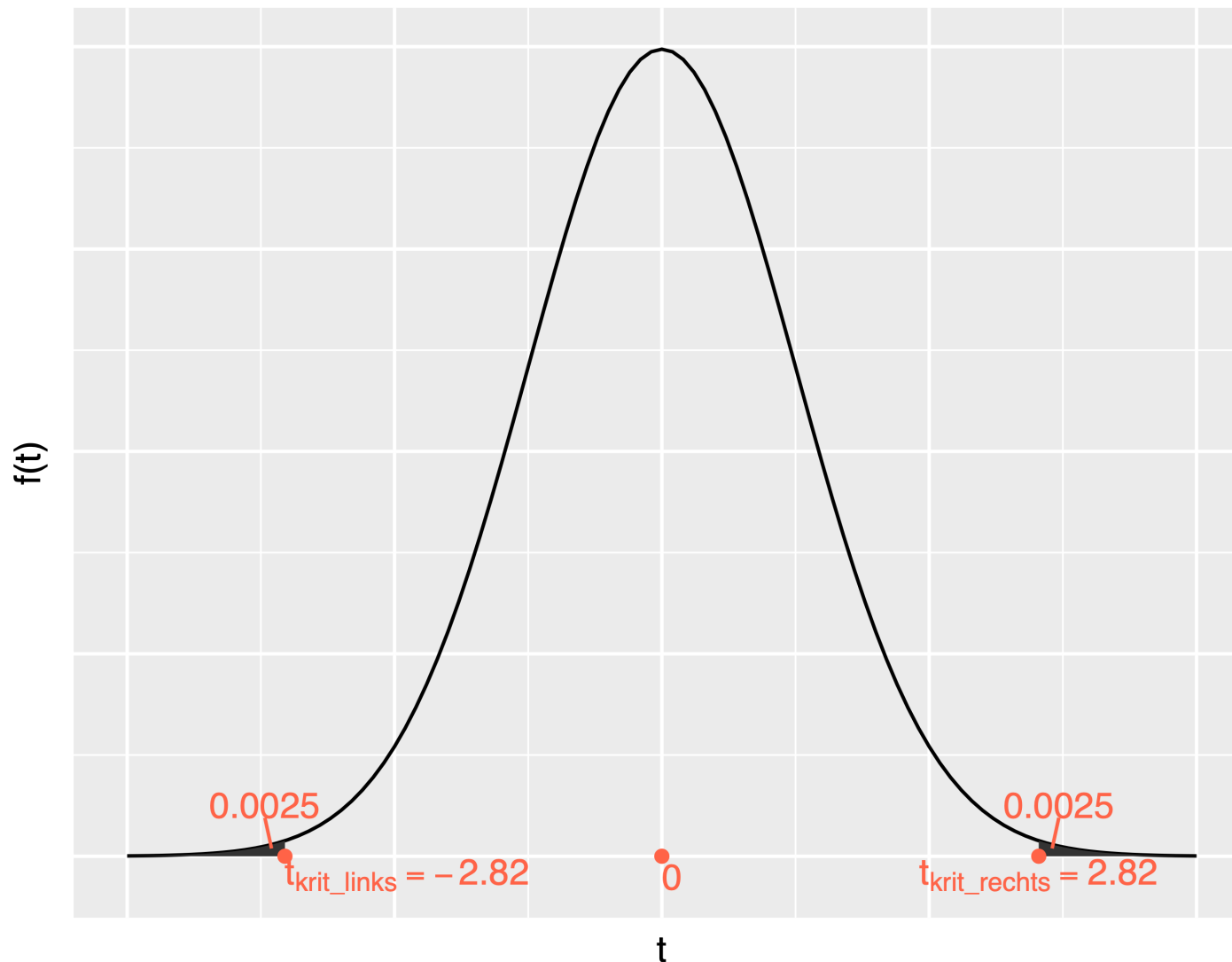
```
> qt(0.0025, 499)
[1] -2.819573
```

- Also ist $t_{krit_links} \approx -2.82$
- Wegen der Symmetrie der t-Verteilung ist $t_{krit_rechts} = -t_{krit_links} = 2.82$
- Der kritische Bereich ist damit

$$K_T =]-\infty, -2.82] \cup [2.82, \infty[$$

Beispiel IV

- Graphische Veranschaulichung:



- Um die Realisation

$$t = \frac{\bar{x} - 100}{\sqrt{\frac{s^2}{500}}}$$

der Teststatistik

$$T = \frac{\bar{X} - 100}{\sqrt{\frac{S^2}{500}}}$$

berechnen zu können, benötigen wir die Schätzwerte \bar{x} und s^2 .

Beispiel VI

- Beispielsweise könnte sich in unserer Stichprobe $\bar{x} = 110$ und $s^2 = 250$ ergeben.
- Die Realisation der Teststatistik wäre dann

$$t = \frac{\bar{x} - 100}{\sqrt{\frac{s^2}{500}}} = \frac{110 - 100}{\sqrt{\frac{250}{500}}} \approx 14.14$$

- Diese Realisation liegt im kritischen Bereich, da $t = 14.14 > 2.82 = t_{krit_rechts}$.
- Wir entscheiden uns daher für die statistische Alternativhypothese $\mu \neq 100$ und somit auch für die inhaltliche Alternativhypothese $\bar{x}_{IQ_Pop} \neq 100$.
- Dies bedeutet, dass wir uns dafür entscheiden, unsere Theorie nicht zu falsifizieren.
- Ist dies eine gute Entscheidung?

Gütekriterien für statistische Hypothesentests

Fehlentscheidungen

- Da wir unsere Entscheidung für die H_0 bzw. die H_1 nur auf der Basis einer zufälligen Stichprobe treffen, ist es immer möglich dass wir eine **Fehlentscheidung** treffen.
- **Dies gilt für jeden statistischen Hypothesentest!**

Arten von Fehlentscheidungen

- Folgende Fehlentscheidungen sind möglich:

		Testergebnis	
		<i>Entscheidung für H_0</i>	<i>Entscheidung für H_1</i>
Realität	H_0 wahr	Kein Fehler	Fehler 1. Art
	H_1 wahr	Fehler 2. Art	Kein Fehler

- Fehler 1. Art** (Synonym: **α -Fehler**):
 - Wir entscheiden uns für die Alternativhypothese, obwohl die Nullhypothese wahr ist.
- Fehler 2. Art** (Synonym: **β -Fehler**):
 - Wir entscheiden uns für die Nullhypothese, obwohl die Alternativhypothese wahr ist.

- Um Gütekriterien für statistische Hypothesen formulieren zu können, betrachten wir die **Wahrscheinlichkeiten** für diese beiden Fehler.
- Da wir unsere Testentscheidung auf der Basis der Realisation einer Zufallsvariablen treffen, können wir diese Wahrscheinlichkeiten bestimmen und kontrollieren.
- Wir können somit sicherstellen, dass wir auf lange Sicht – d.h. falls wir unendlich oft eine einfache Zufallsstichprobe ziehen – sehr häufig richtige Entscheidungen treffen würden.
- Ob eine konkrete Testentscheidung basierend auf einer konkreten Stichprobe richtig oder falsch ist, wissen wir aber nie.

- Die Wahrscheinlichkeit für einen Fehler 1. Art eines statistischen Hypothesentests entspricht der Wahrscheinlichkeit dafür, dass wir uns für die H_1 entscheiden, obwohl eigentlich die H_0 wahr ist.
- Da wir uns genau dann für die H_1 entscheiden, wenn die Realisation unserer Teststatistik im kritischen Bereich liegt, entspricht die Wahrscheinlichkeit für einen Fehler 1. Art dem Signifikanzniveau, da unter der H_0

$$P(\text{Entscheidung für } H_1) = P(T \in K_T) = \alpha$$

gilt.

- Daher sollte das Signifikanzniveau auf jeden Fall klein sein. Welchen Wert wir genau wählen, werden wir noch besprechen.
- Wir können die Wahrscheinlichkeit für einen Fehler 1. Art also durch die Wahl eines geeigneten kritischen Bereichs festlegen und somit kontrollieren.

- Die Wahrscheinlichkeit für einen Fehler 2. Art eines statistischen Hypothesentests entspricht der Wahrscheinlichkeit dafür, dass wir uns für die H_0 entscheiden, obwohl eigentlich die H_1 wahr ist.
- Die Wahrscheinlichkeit für einen Fehler 2. Art eines statistischen Hypothesentests wird mit β bezeichnet.
- Häufig wird statt der Wahrscheinlichkeit für einen Fehler 2. Art auch die Wahrscheinlichkeit dafür betrachtet, dass wir uns für die H_1 entscheiden, falls die H_1 wahr ist.
- Diese Wahrscheinlichkeit wird **Power** (auch Teststärke) des statistischen Hypothesentests genannt und entspricht $1 - \beta$.
- Wie wir die Power eines statistischen Tests bestimmen und kontrollieren können, werden wir später noch besprechen.

- Wir können nun anhand dieser Fehlerwahrscheinlichkeiten Gütekriterien für statistische Hypothesentests formulieren:
- Ein statistischer Hypothesentest sollte gleichzeitig
 - eine geringe Wahrscheinlichkeit für einen Fehler 1. Art, also ein geringes Signifikanzniveau α aufweisen.
 - eine geringe Wahrscheinlichkeit für einen Fehler 2. Art, also eine hohe Power $1 - \beta$ aufweisen.

- Das Signifikanzniveau α eines statistischen Tests können wir direkt durch die Wahl eines geeigneten kritischen Bereichs festlegen.
- Die Power $1 - \beta$ eines statistischen Tests können wir nur indirekt durch die Größe unserer einfachen Zufallsstichprobe festlegen. Wir werden sehen, dass die Power umso **höher** ist, je **größer** die Stichprobe ist.
- **Sehr, sehr, sehr wichtig:**
 - Die Verwendung eines statistischen Hypothesentests ist nur dann sinnvoll, falls er **sowohl ein geringes Signifikanzniveau als auch eine hohe Power aufweist**.
 - Der Fehler 1. Art ist **nicht** „wichtiger“ als der Fehler 2. Art.
Ein statistischer Hypothesentest, der lediglich ein geringes Signifikanzniveau, aber keine hohe Power aufweist, ist kein guter Hypothesentest.
 - Wir werden später noch ausführlich darauf eingehen, warum dies so ist.

Hypothesentests für gerichtete statistische Hypothesen über μ

Gerichtete Hypothesen

- Zur Erinnerung: **Gerichtete Hypothesen** treffen Aussagen darüber, ob eine deskriptivstatistische Maßzahl oder ein Parameter **kleiner, kleiner-gleich, größer oder größer-gleich** einem bestimmten Wert ist.
- Wir unterscheiden zwischen **rechtsgerichteten** Hypothesen (größer oder größer-gleich) und **linksgerichteten** Hypothesen (kleiner oder kleiner-gleich).
- Je nachdem, ob wir eine linksgerichtete oder eine rechtsgerichtete Alternativhypothese vorliegen haben, ergeben sich unterschiedliche statistische Hypothesentests.

- Aus einer psychologischen Theorie könnte folgen, dass der Mittelwert \bar{x}_{Pop} einer Variable in einer Population **größer** als ein bestimmter Wert μ_0 sein sollte.
- μ_0 ist hierbei eine Zahl, z.B. $\mu_0 = 0$ oder $\mu_0 = 100$, deren genauer Wert sich aus der Theorie ergeben muss.
- Wir hätten also die folgenden gerichteten inhaltlichen Hypothesen:

$$H_0: \bar{x}_{Pop} \leq \mu_0$$

$$H_1: \bar{x}_{Pop} > \mu_0$$

- Falls das Histogramm der Variable in der Population durch die Wahrscheinlichkeitsdichtefunktion der Normalverteilung approximiert werden kann und wir eine einfache Zufallsstichprobe mit Umfang n aus dieser Population ziehen, haben wir genau wie bei ungerichteten Hypothesen die folgende Ausgangssituation:

$$X_1, X_2, \dots, X_n \text{ mit } X_i \stackrel{\text{iid}}{\sim} N(\mu, \sigma^2)$$

Rechtsgerichtete Alternativhypothesen II

- Da wir wissen, dass in diesem Fall $\bar{x}_{pop} = \mu$ ist, können wir die inhaltlichen Hypothesen in äquivalente statistische Hypothesen überführen:

$$H_0: \mu \leq \mu_0$$

$$H_1: \mu > \mu_0$$

- Wir wollen also überprüfen, ob μ **größer** als ein vorgegebener Wert μ_0 ist oder nicht.

Linksgerichtete Alternativhypothesen I

- Aus einer psychologischen Theorie könnte folgen, dass der Mittelwert \bar{x}_{Pop} einer Variable in einer Population **kleiner** als ein bestimmter Wert μ_0 sein sollte.
- μ_0 ist hierbei eine Zahl, z.B. $\mu_0 = 0$ oder $\mu_0 = 100$, deren genauer Wert sich aus der Theorie ergeben muss.
- Wir hätten also die folgenden gerichteten inhaltlichen Hypothesen:

$$H_0: \bar{x}_{Pop} \geq \mu_0$$

$$H_1: \bar{x}_{Pop} < \mu_0$$

- Falls das Histogramm der Variable in der Population durch die Wahrscheinlichkeitsdichtefunktion der Normalverteilung approximiert werden kann und wir eine einfache Zufallsstichprobe mit Umfang n aus dieser Population ziehen, haben wir genau wie bei ungerichteten und rechtsgerichteten Alternativhypothesen die folgende Ausgangssituation:

$$X_1, X_2, \dots, X_n \text{ mit } X_i \stackrel{\text{iid}}{\sim} N(\mu, \sigma^2)$$

Linksgerichtete Alternativhypothesen II

- Da wir wissen, dass in diesem Fall $\bar{x}_{pop} = \mu$ ist, können wir die inhaltlichen Hypothesen in äquivalente statistische Hypothesen überführen:

$$H_0: \mu \geq \mu_0$$

$$H_1: \mu < \mu_0$$

- Wir wollen also überprüfen, ob μ **kleiner** als ein vorgegebener Wert μ_0 ist oder nicht.

- Gerichtete statistische Alternativhypothesen über μ können entweder rechtsgerichtet oder linksgerichtet sein. Daraus ergeben sich unterschiedliche statistische Hypothesenpaare:

- rechtsgerichtete Alternativhypothese:

$$H_0: \mu \leq \mu_0$$

$$H_1: \mu > \mu_0$$

- linksgerichtete Alternativhypothese:

$$H_0: \mu \geq \mu_0$$

$$H_1: \mu < \mu_0$$

- Wir werden sehen, dass die statistischen Hypothesentests für beide Fälle die gleiche Teststatistik wie im ungerichteten Fall aufweisen. Die Lage des kritischen Bereichs wird jedoch jeweils unterschiedlich sein.

- Da sich die Ausgangssituation X_1, X_2, \dots, X_n mit $X_i \stackrel{\text{iid}}{\sim} N(\mu, \sigma^2)$ in beiden Fällen nicht von der Ausgangssituation bei ungerichteten Hypothesen unterscheidet, können wir als Schätzwert $\hat{\mu}_{\text{Wert}}$ für μ genau wie dort den Mittelwert \bar{x} in der Stichprobe verwenden.
- In beiden Fällen wollen wir diesen Schätzwert wieder mit dem vorgegebenen Wert μ_0 vergleichen und entscheiden, ob er so weit von diesem entfernt ist, dass wir uns für die H_1 entscheiden.

- Im Gegensatz zu den statistischen Hypothesentests bei ungerichteten Hypothesen sprechen hier jedoch nicht alle großen Abweichungen des Schätzwerts von μ_0 für die H_1 , sondern nur Abweichungen in Richtung der H_1 :
 - Im Fall einer **rechtsgerichteten** H_1 würden wir uns nur für die H_1 entscheiden, falls der Schätzwert \bar{x} für μ deutlich **größer** als μ_0 ist.
 - Im Fall einer **linksgerichteten** H_1 würden wir uns nur für die H_1 entscheiden, falls der Schätzwert \bar{x} für μ deutlich **kleiner** als μ_0 ist.
- Beispiel: Im Falle einer rechtsgerichteten H_1 mit $\mu_0 = 10$, d.h.

$$H_0: \mu \leq 10$$

$$H_1: \mu > 10$$

würden wir uns bei einem Schätzwert $\hat{\mu}_{wert} = \bar{x} = 1000$ eher für die H_1 entscheiden, bei einem Schätzwert $\hat{\mu}_{wert} = \bar{x} = -1000$ jedoch eher für die H_0 .

- Genau wie beim Hypothesentest für ungerichtete Hypothesen können wir leider den kritischen Bereich nicht direkt für die Schätzfunktion \bar{X} bestimmen, da wir hierfür den unbekannten Parameter σ^2 kennen müssten.
- Wir verwenden daher wieder die folgende Transformation der Schätzfunktion als Teststatistik:

$$T = \frac{\bar{X} - \mu_0}{\sqrt{\frac{S^2}{n}}}$$

- Diese Teststatistik verwenden wir sowohl für rechtsgerichtete als auch für linksgerichtete Alternativhypothesen. Sie ist identisch mit der Teststatistik beim Hypothesentest für ungerichtete Hypothesen.

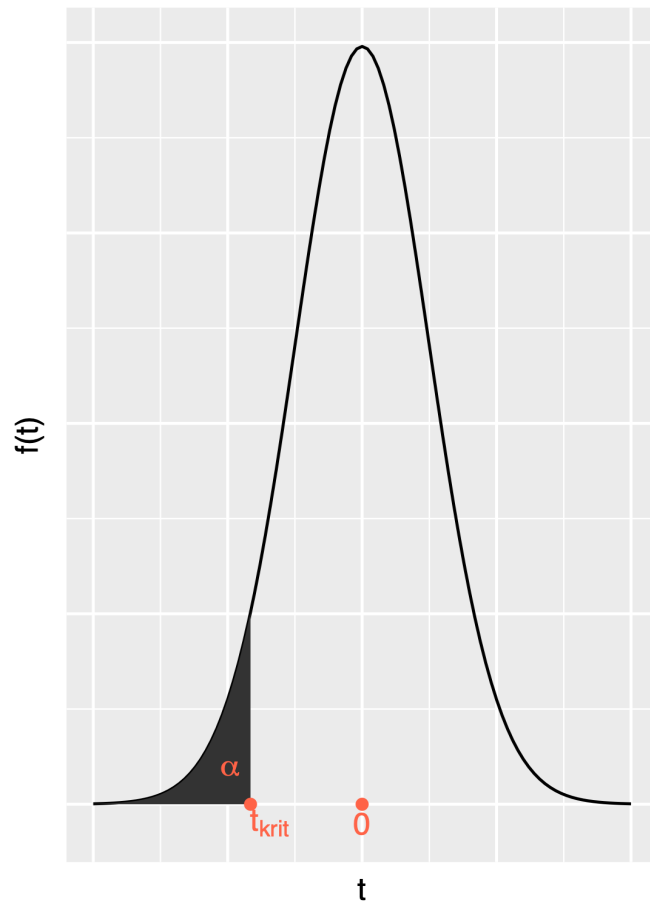
- Um den kritischen Bereich zu bestimmen, können wir bei gerichteten Hypothesen nicht einfach die Wahrscheinlichkeitsverteilung der Teststatistik unter der Voraussetzung, dass die H_0 wahr ist, betrachten.
- Es gibt unter der H_0 nämlich unendlich viele mögliche Parameterwerte und daher auch unendlich viele mögliche Wahrscheinlichkeitsverteilungen der Teststatistik unter der H_0 .
- Beispielsweise könnte μ unter der $H_0: \mu \leq 10$ alle Werte annehmen, die kleiner oder gleich 10 sind. Für jeden dieser Werte hätte die Teststatistik eine andere Wahrscheinlichkeitsverteilung.

- Wir können jedoch die Wahrscheinlichkeitsverteilung der Teststatistik unter der Voraussetzung betrachten, dass $\mu = \mu_0$ ist, also derjenige Parameterwert unter der H_0 der wahre Wert ist, der an der Grenze zur H_1 liegt.
- Wir wissen aus dem ersten Teil der Vorlesung, dass die Teststatistik

$$T = \frac{\bar{X} - \mu_0}{\sqrt{\frac{S^2}{n}}}$$

unter der Voraussetzung $\mu = \mu_0$ einer t-Verteilung mit $\nu = n - 1$ folgt.

- Auf der Basis dieser Wahrscheinlichkeitsverteilung berechnen wir nun sowohl für rechtsgerichtete als auch für linksgerichtete Alternativhypothesen den kritischen Bereich.



- Die Realisation

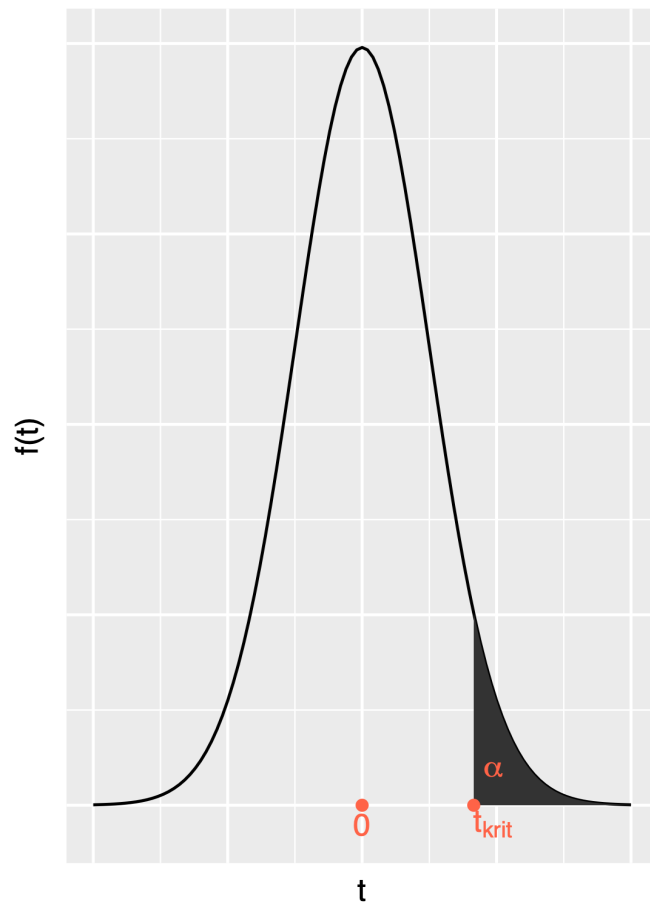
$$t = \frac{\bar{x} - \mu_0}{\sqrt{\frac{s^2}{n}}}$$

der Teststatistik T ist **negativ**, falls der Schätzwert \bar{x} für μ **kleiner** als μ_0 ist, also **links** von μ_0 und im Falle einer **linksseitigen Alternativhypothese** im Bereich der H_1 liegt.

- Bei **linksseitigen Alternativhypothesen** liegt der kritische Bereich somit sinnvollerweise auf der **linken Seite**:

$$K_T =]-\infty, t_{krit}]$$

- Der kritische Wert t_{krit} ist derjenige Wert, für den $P(T \leq t_{krit}) = F(t_{krit}) = \alpha$ ist, wobei P eine t-Verteilung mit $\nu = n - 1$ und F deren Verteilungsfunktion ist.



- Die Realisation

$$t = \frac{\bar{x} - \mu_0}{\sqrt{\frac{s^2}{n}}}$$

der Teststatistik T ist **positiv**, falls der Schätzwert \bar{x} für μ **größer** als μ_0 ist, also **rechts** von μ_0 und im Falle einer **rechtsseitigen Alternativhypothese** im Bereich der H_1 liegt.

- Bei **rechtsseitigen Alternativhypothesen** liegt der kritische Bereich somit sinnvollerweise auf der **rechten Seite**:

$$K_T = [t_{krit}, +\infty[$$

- Der kritische Wert t_{krit} ist derjenige Wert, für den $P(T \geq t_{krit}) = 1 - F(t_{krit}) = \alpha$ ist, wobei P eine t-Verteilung mit $\nu = n - 1$ und F deren Verteilungsfunktion ist.

- Die Entscheidungsregel ist wie bei allen statistischen Hypothesentests:
 - Falls die Realisation t der Teststatistik T **im kritischen Bereich liegt, entscheiden wir uns für die H_1 .**
 - Falls die Realisation t der Teststatistik T **nicht im kritischen Bereich liegt, entscheiden wir uns für die H_0 .**

- Da wir den kritischen Bereich nur auf der Basis der Wahrscheinlichkeitsverteilung der Teststatistik unter der Voraussetzung $\mu = \mu_0$ bestimmt haben, entspricht das Signifikanzniveau α **nicht mehr** der Wahrscheinlichkeit für einen Fehler 1. Art, sondern zunächst nur der Wahrscheinlichkeit für eine falsche Entscheidung für die H_1 , falls μ dem spezifischen Wert μ_0 entspricht.
- Man kann jedoch zeigen, dass die Wahrscheinlichkeit für einen Fehler 1. Art für alle anderen Parameterwerte unter der H_0 kleiner oder gleich α ist.
- Das **Signifikanzniveau α** entspricht also bei gerichteten Alternativhypothesen nicht mehr exakt der Wahrscheinlichkeit für einen Fehler 1. Art, sondern ist “nur“ die **maximale Wahrscheinlichkeit für einen Fehler 1. Art**.
- Das ist natürlich gut genug, da wir dann z.B. bei einem Signifikanzniveau von $\alpha = 0.005$ wissen, dass die Wahrscheinlichkeit für einen Fehler 1. Art im schlechtesten Fall immer noch gleich 0.005 ist.

- Der besprochene statistische Hypothesentest mit der Teststatistik

$$T = \frac{\bar{X} - \mu_0}{\sqrt{\frac{S^2}{n}}} \stackrel{H_0}{\sim} t(n-1)$$

und dem kritischen Bereich

$$K_T =]-\infty, t_{krit}]$$

oder

$$K_T = [t_{krit}, +\infty[$$

wird **Einstichproben t-Test für gerichtete Hypothesen über μ** genannt.

Bemerkungen II

- Das Vorgehen bei ungerichteten, rechtsgerichteten und linksgerichteten Alternativhypothesen bzgl. μ unterscheidet sich lediglich in der Lage und konkreten Berechnung des kritischen Bereichs. Die Teststatistiken und die Wahrscheinlichkeitsverteilung, auf der die Berechnung des kritischen Bereichs basiert, sind identisch.

Beispiel I

- Aus unserer Theorie folgt, dass der durchschnittliche IQ \bar{x}_{IQ_Pop} in der Population der Studierenden größer als 100 sein sollte.
- In diesem Fall ist $\mu_0 = 100$ und die inhaltlichen Hypothesen lauten:

$$H_0: \bar{x}_{IQ_Pop} \leq 100$$

$$H_1: \bar{x}_{IQ_Pop} > 100$$

- Wir nehmen an, dass das Histogramm des IQ in der Population der Studierenden durch die Wahrscheinlichkeitsdichtefunktion einer Normalverteilung approximiert werden kann und ziehen eine einfache Zufallsstichprobe der Größe $n = 500$.
- In diesem Fall ist $\bar{x}_{IQ_Pop} = \mu$ und wir erhalten die folgenden statistischen Hypothesen:

$$H_0: \mu \leq 100$$

$$H_1: \mu > 100$$

- Unsere Teststatistik ist

$$T = \frac{\bar{X} - \mu_0}{\sqrt{\frac{S^2}{n}}} = \frac{\bar{X} - 100}{\sqrt{\frac{S^2}{500}}}$$

und wir wissen, dass diese Teststatistik im Fall von $\mu = \mu_0 = 100$ einer t-Verteilung mit

$$\nu = n - 1 = 500 - 1 = 499$$

folgt.

Beispiel III

- Den kritischen Wert können wir auf Basis dieser t-Verteilung in R berechnen.
- Wir wählen ein Signifikanzniveau von $\alpha = 0.005$.
- Da wir eine rechtsseitige Alternativhypothese vorliegen haben, ist t_{krit} derjenige Wert, für den $1 - F(t_{krit}) = \alpha = 0.005$, also $F(t_{krit}) = 1 - 0.005 = 0.995$ ist:

```
> qt(0.995, 499)
[1] 2.585718
```

- Also ist $t_{krit} \approx 2.59$
- Der kritische Bereich ist aufgrund der rechtsseitigen Alternativhypothese

$$K_T = [2.59, \infty[$$

Beispiel IV

- Um die Realisation

$$t = \frac{\bar{x} - 100}{\sqrt{\frac{s^2}{500}}}$$

der Teststatistik

$$T = \frac{\bar{X} - 100}{\sqrt{\frac{S^2}{500}}}$$

berechnen zu können, benötigen wir die Schätzwerte \bar{x} und s^2 .

Beispiel V

- Beispielsweise könnte sich in unserer Stichprobe $\bar{x} = 110$ und $s^2 = 250$ ergeben.
- Die Realisation der Teststatistik wäre dann

$$t = \frac{\bar{x} - 100}{\sqrt{\frac{s^2}{500}}} = \frac{110 - 100}{\sqrt{\frac{250}{500}}} \approx 14.14$$

- Diese Realisation liegt im kritischen Bereich, da $t = 14.14 > 2.59 = t_{krit}$.
- Wir entscheiden uns daher für die statistische Alternativhypothese $\mu > 100$ und somit auch für die inhaltliche Alternativhypothese $\bar{x}_{IQ_Pop} > 100$.
- Dies bedeutet, dass wir uns dafür entscheiden, unsere Theorie nicht zu falsifizieren.

Allgemeines Vorgehen

Grundsätzliches Vorgehen beim Testen einer statistischen Hypothese

1. Ausgangspunkt: inhaltliche Nullhypothese und inhaltliche Alternativhypothese.
2. Aus diesen inhaltlichen Hypothesen leiten wir die statistische Nullhypothese und die statistische Alternativhypothese ab.
3. Wir wählen eine Teststatistik.
4. Wir legen ein Signifikanzniveau α und eine Power $1 - \beta$ fest.
5. Wir ziehen eine einfache Zufallsstichprobe, deren Größe von der von uns gewünschten Power $1 - \beta$ abhängt.
6. Wir bestimmen den kritischen Bereich, der von dem von uns gewünschten Signifikanzniveau α abhängt.
7. Wir berechnen die Realisation der Teststatistik in unserer Stichprobe.
8. Testentscheidung: Entscheidung für die H_0 oder die H_1 in Abhängigkeit der Realisation der Teststatistik und des kritischen Bereichs.

Zusammenfassung

- Inhaltliche Hypothesen beziehen sich für gewöhnlich auf deskriptivstatistische Maßzahlen in einer Population.
- Null- und Alternativhypothese schließen sich gegenseitig vollständig aus.
- Hypothesen können gerichtet (links oder rechts) oder ungerichtet sein.
- Statistische Hypothesen beziehen sich auf Parameter einer Verteilung, die als Modell einer interessierenden Variablen verwendet wird, und sind bei Ziehung einer einfachen Zufallsstichprobe äquivalent zu inhaltlichen Hypothesen.
- Die Schätzfunktion des Parameters kann in eine Teststatistik umgewandelt werden, die unter Annahme der H_0 einer bekannten Verteilung folgt.
- Mit Hilfe dieser Verteilung wird der kritische Bereich festgelegt, in dem sich die Teststatistik **wenn die H_0 gilt**, nur mit einer kleinen Wahrscheinlichkeit α realisieren sollte.
- Aus den Stichprobendaten wird die Realisation der Teststatistik berechnet und mit dem kritischen Bereich verglichen - liegt sie im Bereich, wird die H_0 verworfen.
- Damit entspricht α der (maximalen) Wahrscheinlichkeit für einen Fehler 1. Art, der Wahrscheinlichkeit, dass die H_0 gilt, jedoch verworfen wird.