Datenanalyse – Sitzung 5

Signifikanztest

Institut für Kommunikationswissenschaft und Medienforschung
Ludwig-Maximilians-Universität München

Ablauf der Sitzung

1. Wiederholung: Statistische Hypothesen
2. Übungsblatt: Aufgabe 1 (Hausaufgabe)
3. Wiederholung: Signifikanztest und Signifikanzniveau
4. Übungsblatt: Aufgabe 2 (Hausaufgabe)
5. Wiederholung: Fehler erster und zweiter Art
6. Übungsblatt: Aufgabe 3
7. Einfacher t-Test für den Mittelwert in R
8. Übungsblatt: Aufgabe 4 (mit R)

WIEDERHOLUNG:

STATISTISCHE HYPOTHESEN

Hypothesen

Klassifizierung von Hypothesen

- Unterschieds- *vs.* Zusammenhangshypothesen
- Gerichtete (einseitige) *vs.* ungerichtete (zweiseitige) Hypothesen
- Spezifische *vs.* unspezifische Hypothesen

Statistische Hypothesen

- Die **Alternativhypothese** (H_1)
  üblicherweise die zu prüfende Vermutung der Forschenden
- Die **Nullhypothese** (H_0)
  formale Gegenhypothese zu H_1
  H_0 gilt bis sie widerlegt wird – wir testen gegen die Nullhypothese

➔ H_1 und H_0 sind somit zueinander komplementäre Aussagen

ÜBUNGSBLATT: AUFGABE 1 (HAUSAUFGABE)

Aufgabe 1

Entwickeln Sie jeweils eine Null- und Alternativhypothese für folgende Fragestellungen. Geben Sie jeweils eine gerichtete und eine ungerichtete Hypothese an.

a) Bei der Bundestagswahl 2017 gab es deutschlandweit eine Wahlbeteiligung von 72%. Sie vermuten, dass sich die Wahlbeteiligung unter den Zuschauer*innen von „Maybrit Illner" hiervon unterscheidet.

b) Die Macher*innen der Sendung „ZDF Magazin Royale" geben an, dass seit Verleihung des Grimme-Preis 2023 jede dritte Person in Deutschland ihre Sendung zumindest dem Namen nach kennt. Sie vermuten, dass die tatsächliche Bekanntheit von der behaupteten abweicht.

Lösung: Aufgabe 1a

Bei der Bundestagswahl 2017 gab es deutschlandweit eine Wahlbeteiligung von 72%. Sie vermuten, dass sich die Wahlbeteiligung unter den Zuschauer*innen von „Maybrit Illner“ hiervon unterscheidet.

- Ungerichtete Hypothese

H0: Es gibt keinen Unterschied zwischen „Maybrit Illner“-Zuschauer*innen und der Gesamtbevölkerung hinsichtlich der Wahlbeteiligung ($p = .72$).
H1: Es gibt einen Unterschied zwischen „Maybrit Illner“-Zuschauer*innen und der Gesamtbevölkerung hinsichtlich der Wahlbeteiligung ($p \neq .72$).

- Gerichtete Hypothese

H0: Die Wahlbeteiligung unter „Maybrit Illner“-Zuschauer*innen ist gleich hoch oder geringer als in der Gesamtbevölkerung ($p \leq 0{,}72$).
H1: Die Wahlbeteiligung unter „Maybrit Illner“-Zuschauer*innen ist höher als in der Gesamtbevölkerung ($p > 0{,}72$).

Lösung: Aufgabe 1b

Die Macher*innen der Sendung „ZDF Magazin Royale“ geben an, dass seit Verleihung des Grimme-Preis 2023 jede dritte Person in Deutschland ihre Sendung zumindest vom Namen her kennt. Sie vermuten, dass die tatsächliche Bekanntheit von der behaupteten abweicht.

- Ungerichtete Hypothese

H_0: Die Bekanntheit von „ZDF Magazin Royale“ entspricht 33% $(p = 0{,}33)$.
H_1: Die Bekanntheit von „ZDF Magazin Royale“ entspricht nicht 33% $(p \neq 0{,}33)$.

- Gerichtete Hypothese

H_0: Die Bekanntheit von „ZDF Magazin Royale“ ist mindestens 33% $(p \geq 0{,}33)$.
H_1: Die Bekanntheit von „ZDF Magazin Royale“ ist kleiner als 33% $(p < 0{,}33)$.

WIEDERHOLUNG:

SIGNIFIKANZTEST UND SIGNIFIKANZNIVEAU

Signifikanztest und Signifikanzniveau

Das Prinzip des Signifikanztests

- Annahme: Zufälliger Auswahlfehler
- Berechnung: In welchem Bereich sollten die Beobachtungen mehrerer gleichartiger empirischer Studien höchstwahrscheinlich (1 - α) liegen, wenn die Nullhypothese zutrifft?
- Test: Liegt der empirisch beobachtete Wert innerhalb/außerhalb dieses Bereichs?

Das Signifikanzniveau α

- α = Anteil der Testergebnisse, für die wir bereit sind, H_0 zurückzuweisen, obwohl sie zutreffend ist.
- α = Von Forschenden festgelegte Wahrscheinlichkeit (Risiko) für einen Fehler 1. Art (Alpha-Fehler)
- Meist werden die Signifikanzniveaus 5% ($\alpha = 0{,}05$) und 1% ($\alpha = 0{,}01$) verwendet.

z-Test für Anteilswerte

Approximation durch Normalverteilung

bei großen Stichproben (wenn gilt)

$N \cdot p_0 \cdot (1 - p_0) > 9$ [p_0 = erwartete Vergleichswahrscheinlichkeit]

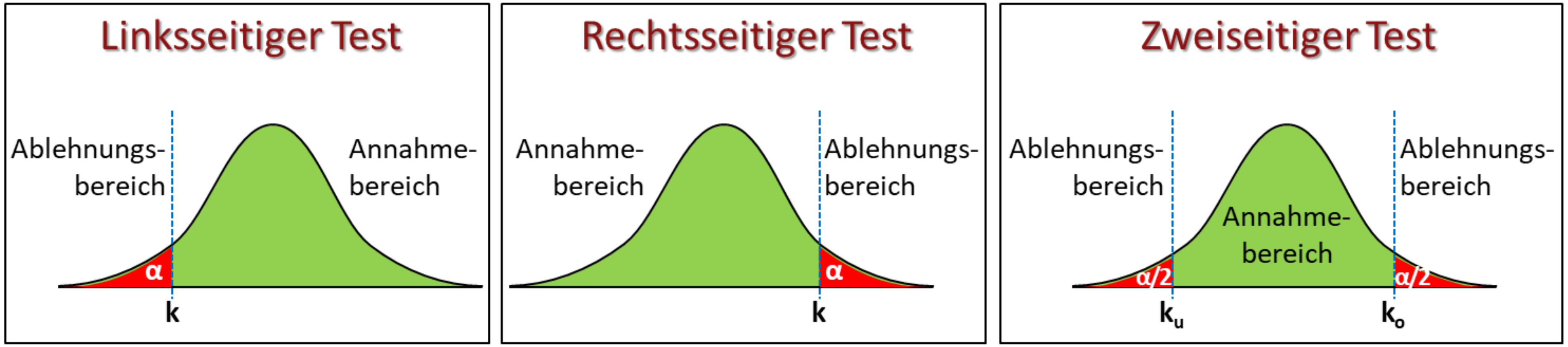
Bestimmung der kritischen Werte:

- Linksseitig: $k = N \cdot p_0 - z_{1-\alpha} \cdot \sqrt{N \cdot p_0 \cdot (1 - p_0)}$
- Rechtsseitig: $k = N \cdot p_0 + z_{1-\alpha} \cdot \sqrt{N \cdot p_0 \cdot (1 - p_0)}$
- Zweiseitig: $k_u = N \cdot p_0 - z_{1-\frac{\alpha}{2}} \cdot \sqrt{N \cdot p_0 \cdot (1 - p_0)}$

  $k_o = N \cdot p_0 + z_{1-\frac{\alpha}{2}} \cdot \sqrt{N \cdot p_0 \cdot (1 - p_0)}$

Annahme- und Ablehnungsbereich

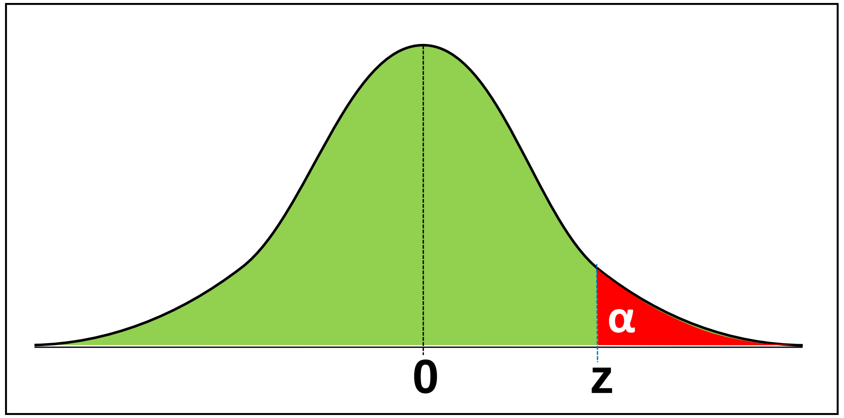
Annahmebereich	Ablehnungsbereich
Menge der Testergebnisse, die wir (noch) als mit der Nullhypothese H_0 vereinbar sehen	Menge der Testergebnisse, die wir als mit der Nullhypothese H_0 **unvereinbar** sehen



z-Test für Anteilswerte

Häufig verwendete α-Quantile der Normalverteilung:

$P(X \leq z)$	z
95,0%	1,64
97,5%	1,96
99,0%	2,33
99,5%	2,58



Übungsblatt: Aufgabe 2 (Hausaufgabe)

Der z-Test: Aufgabe 2 (Hausaufgabe)

Überprüfen Sie die folgenden Annahmen mittels eines Signifikanztests.

a) Bei der Bundestagswahl 2017 gab es deutschlandweit eine Wahlbeteiligung von 72%. Sie stellen die Hypothese auf, dass die Wahlbeteiligung unter „Maybrit Illner"-Zuschauer*innen höher ist als in der Gesamtbevölkerung. In einer Befragung unter 90 Zuschauer*innen gaben 75 Personen an, gewählt zu haben. Das Signifikanzniveau legen Sie auf 1% fest.

b) Die Macher*innen der Sendung „ZDF Magazin Royale" geben an, dass seit Verleihung des Grimme-Preis 2023 jede dritte Person in Deutschland ihre Sendung zumindest vom Namen her kennt. Sie stellen die Hypothese auf, dass die Bekanntheit der Sendung nicht 33% entspricht. Bei einer Quoten-Erhebung von 200 Personen aus der deutschen Wohnbevölkerung gaben 45 Personen an, dass sie die Sendung zumindest dem Namen nach kennen. Das Signifikanzniveau legen Sie auf 5% fest.

Der z-Test: Aufgabe 2a

Bei der Bundestagswahl 2017 gab es deutschlandweit eine Wahlbeteiligung von 72%. Sie stellen die Hypothese auf, dass die Wahlbeteiligung unter „Maybrit Illner"-Zuschauer*innen höher ist als in der Gesamtbevölkerung. In einer Befragung unter 90 Zuschauer*innen gaben 75 Personen an, gewählt zu haben. Das Signifikanzniveau legen Sie auf 1% fest. ($\alpha = 0{,}01$; $N = 90$)

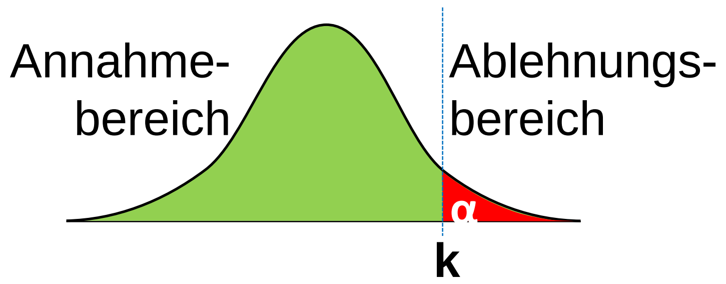
Testform: einseitig rechts

H₀: $\boldsymbol{p \leq 0,72}$ **H₁**: $\boldsymbol{p > 0,72}$

$N \cdot p_0 \cdot (1 - p_0) = 18{,}14 > 9$ → z-Test ist möglich

$$k = N \cdot p_0 + z_{1-\alpha} \cdot \sqrt{N \cdot p_0 \cdot (1 - p_0)}$$
$$k = 90 \cdot 0{,}72 + 2{,}33 \cdot \sqrt{90 \cdot 0{,}72 \cdot 0{,}28}$$
$$k = 64{,}8 + 9{,}92 = \mathbf{74,72}$$

Rechtsseitiger Test



Annahmebereich: **[0 ; 74]**
Ablehnungsbereich: **[75 ; 90]**
→ Die Hypothese (H_1) kann bestätigt werden, da 75 im Ablehnungsbereich von H_0 liegt.

Der z-Test: Aufgabe 2b

Die Macher*innen der Sendung „ZDF Magazin Royale“ geben an, dass seit Verleihung des Grimme-Preis 2023 jede dritte Person in Deutschland ihre Sendung zumindest vom Namen her kennt. Sie stellen die Hypothese auf, dass die Bekanntheit der Sendung nicht 33% entspricht. Bei einer Quoten-Erhebung von 200 Personen aus der deutschen Wohnbevölkerung gaben 45 Personen an, dass sie die Sendung zumindest dem Namen nach kennen. Das Signifikanzniveau legen Sie auf 5% fest. ($\alpha = 0{,}05$; $N = 200$)

H$_0$: $\boldsymbol{p} = 0{,}33$ **H$_1$:** $\boldsymbol{p} \neq 0{,}33$ Testform: zweiseitig

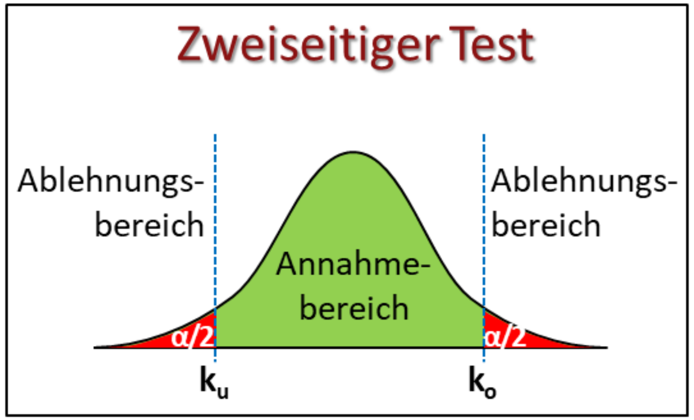
$N \cdot p_0 \cdot (1 - p_0) = 44{,}22 > 9$ → z-Test ist möglich

$$k_u = N \cdot p_0 - z_{1-\frac{\alpha}{2}} \cdot \sqrt{N \cdot p_0 \cdot (1 - p_0)}$$

$$k_u = 200 \cdot 0{,}33 - 1{,}96 \cdot \sqrt{44{,}2} = \mathbf{52{,}97}$$

$$k_o = N \cdot p_0 + z_{1-\frac{\alpha}{2}} \cdot \sqrt{N \cdot p_0 \cdot (1 - p_0)}$$

$$k_o = 200 \cdot 0{,}33 + 1{,}96 \cdot \sqrt{44{,}2} = \mathbf{79{,}03}$$



Annahmebereich: **[53 ; 79]**

Ablehnungsbereich: **[0 ; 52] + [80 ; 200]**

→ Die Hypothese (H$_1$) kann bestätigt werden, da 45 im Ablehnungsbereich von H$_0$ liegt.

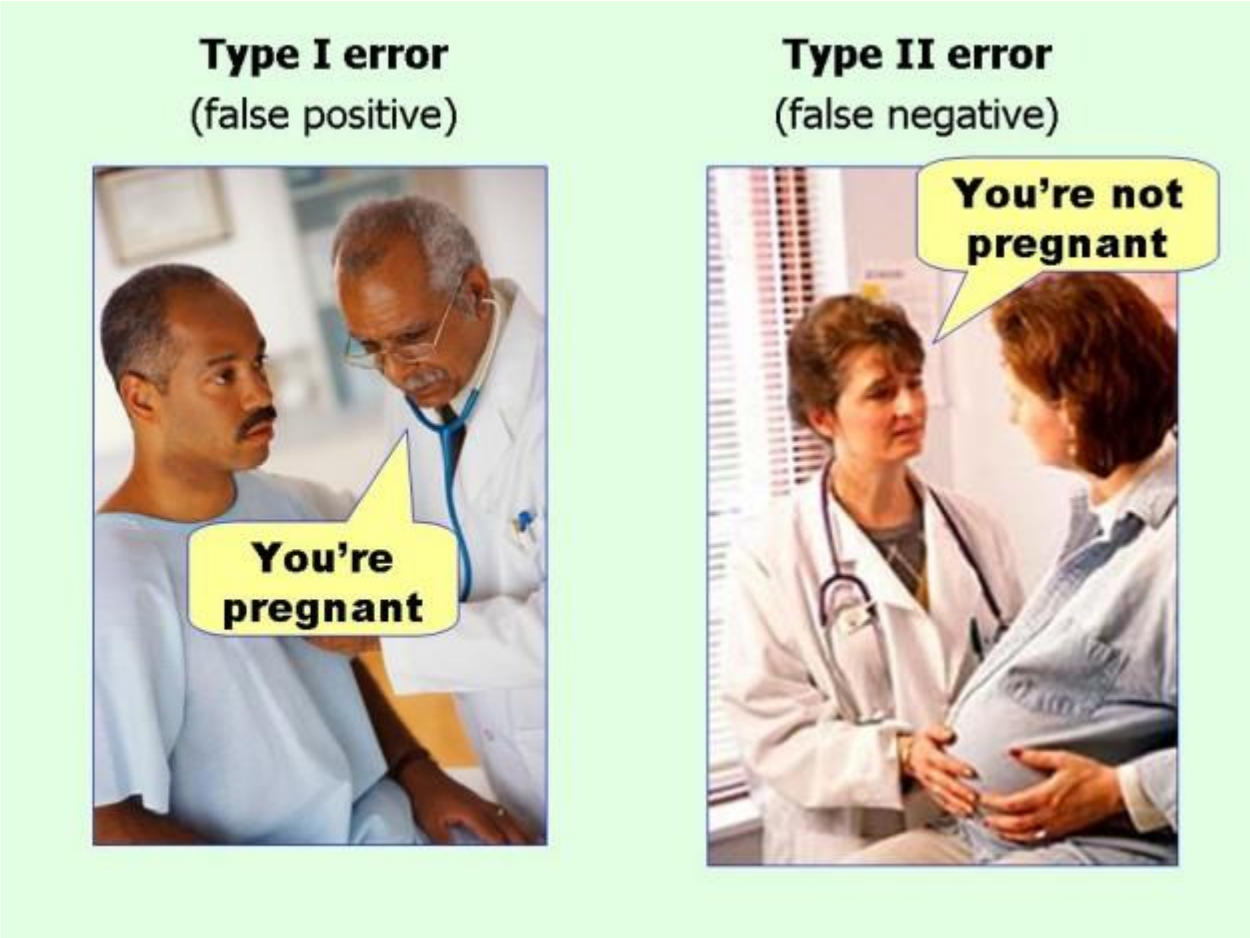
WIEDERHOLUNG:

FEHLER ERSTER UND ZWEITER ART

Fehler 1. und 2. Art

		In der Grundgesamtheit gilt (tatsächlich)	
		tatsächlich H_0	**tatsächlich H_1**
Entscheidung anhand Testgröße	**Entscheidung H_0**	Richtige Entscheidung (1-α)	**β-Fehler** Fehler 2. Art
	α **Entscheidung H_1**	**α-Fehler** Fehler 1. Art Irrtumswahrscheinlichkeit	Richtige Entscheidung Statistische Power (1-β)

Fehler 1. und 2. Art



Achtung: HARKing vermeiden

Quelle: Cockburn, A., Dragicevic, P., Besançon, L., & Gutwin, C. (2020). Threats of a replication crisis in empirical computer science. *Communications of the ACM*, 63(8), 70–79.

HARKing
„Hypothesizing After the Results are Known“

Stellen sie <u>immer erst Hypothesen</u> auf und testen sie diese im Anschluss statistisch.

Sonst wächst der tatsächliche α-Fehler massiv an.

Einfacher t-Test für den Mittelwert

Einfacher t-Test für den Mittelwert

Prüft, ob der empirische Mittelwert $\overline{x}$ aus einer Grundgesamtheit stammt, deren Erwartungswert μ bekannt ist

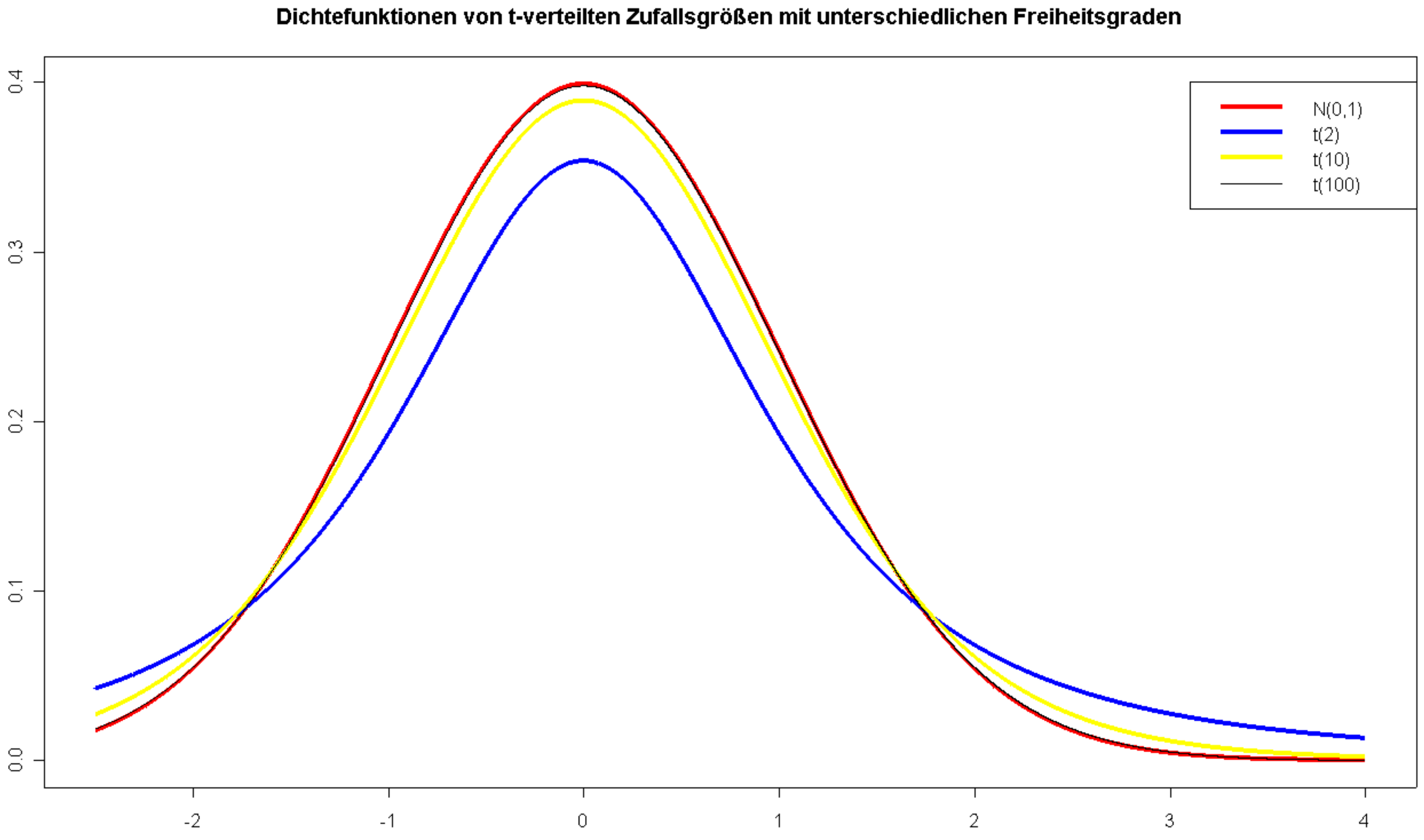
$$\mathbf{H_0}: \; \boldsymbol{\overline{x} = \mu}$$

$$\mathbf{H_1}: \; \boldsymbol{\overline{x} \neq \mu}$$

$$t = \sqrt{N} \cdot \left(\frac{\overline{x} - \mu}{s} \right)$$

- $df = N - 1$ Freiheitsgrade
- H_0 wird abgelehnt, wenn $|t| > t_{krit}$
- Einseitiger Test: t_{krit} = $(1 - \alpha)$-Quantil der t-Verteilung
- Zweiseitiger Test: t_{krit} = $\left(1 - \frac{\alpha}{2}\right)$-Quantil der t-Verteilung

t-Verteilung versus Normalverteilung (z-Werte)



Voraussetzungen für den einfachen t-Test

- Die Stichprobe entstammt einer normalverteilten Grundgesamtheit. Dafür ist nach dem zentralen Grenzwertsatz ein Stichprobenumfang größer 30 hinreichend.
- Die abhängige Variable ist (quasi-)metrisch skaliert.
- Der Erwartungswert μ der Grundgesamtheit ist bekannt.

Übungsblatt: Aufgabe 3

Aufgabe 3

Die Forscher*innen stellen die Hypothese auf, dass sich Schüler*innen im Internat der Benediktinerabtei Ettal bezüglich ihrer Internet-Nutzung von deutschen Durchschnittsjugendlichen unterscheiden. Als Referenzwert ziehen sie ein Ergebnis der aktuellen Shell-Jugendstudie heran, wonach Jugendliche in Deutschland durchschnittlich 18,4 Stunden online sind.

In ihrer Untersuchung im Benediktiner-Internat finden die Forscher*innen heraus, dass die befragten 60 Schüler*innen wöchentlich durchschnittlich 12,5 Stunden online sind ($s = 5{,}0$).

Kann die Hypothese mit diesem Ergebnis bei einem Signifikanzniveau von 1% angenommen werden?

Lösung: Aufgabe 3

Informationen aus der Aufgabenstellung

$$\mu = 18{,}4 \quad \bar{x} = 12{,}5 \quad s = 5{,}0 \quad N = 60 \quad \alpha = 0{,}01$$

Hypothesen H_0: $\bar{x} = \mu$ H_1: $\bar{x} \neq \mu$

$$t = \sqrt{N} \cdot \left(\frac{\bar{x} - \mu}{s}\right) = \sqrt{60} \cdot \left(\frac{12{,}5 - 18{,}4}{5{,}0}\right) = \mathbf{-9{,}14}$$

$df = N - 1 = 59$ → $t_{krit} = \mathbf{2{,}68}$ (zweiseitig)

$|t| = 9{,}14 > t_{krit}$ → **H_0 wird abgelehnt**

Ergebnis: Die Internatsschüler*innen unterscheiden hinsichtlich ihrer Internet-Nutzung signifikant von den deutschen Jugendlichen ($t = -9{,}14$; $df = 59$; $p < 0{,}01$).

Wichtige Take-Aways

- **Nullhypothese H_0**: Gegenhypothese zur zu prüfenden **Alternativhypothese H_1**
- **Signifikanztest** (übergeordneter Begriff)**:** Prüft, ob ein empirisch beobachteter Wert mit vorab festgelegter Irrtumswahrscheinlichkeit (**Signifikanzniveau α**) tatsächlich in der Grundgesamtheit vorliegt
- **z-Test:** Testet anhand der Häufigkeit eines Merkmals in der Stichprobe Hypothesen über dessen Häufigkeit in der Grundgesamtheit
- **Einfacher t-Test:** Prüft, ob sich der Mittelwert aus einer Stichprobe vom (vorgegebenen) Erwartungswert aus der Grundgesamtheit unterscheidet

Einfacher t-Test für den Mittelwert in R

Einfachten t-Test berechnen mittels `tidycomm`

Sie können den einfachen t-Test mit dem `tidycomm` Package berechnen.

Dafür können Sie die folgende Funktion nutzen:

`t_test()`

Struktur:

```
data %>%
   t_test(variable, mu = X)
```

Einfacher *t*-Test in R

Beispielhypothese: Die Berufserfahrung der Journalist*innen aus dem WoJ-Datensatz unterscheidet sich vom Durchschnittswert der Grundgesamtheit aller Journalist*innen (24 Jahre, fiktiver Wert)

Schritt 1: Mithilfe der Funktion `tidycomm::describe` Voraussetzungen prüfen
Einsetzen: Variable (work_experience)

Befehl:

```
WoJ %>% describe(work_experience)
```

Ausgabe:

```
> WoJ %>% describe(work_experience)
# A tibble: 1 × 15
  Variable            N Missing     M    SD   Min   Q25   Mdn   Q75   Max Range CI_95_LL CI_95_UL Skewness Kurtosis
* <chr>           <int>   <int> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>    <dbl>    <dbl>    <dbl>    <dbl>
1 work_experience  1187      13  17.8  10.9     1     8    17    25    53    52     17.2     18.5    0.427     2.41
```

- ✓ Stichprobe > 30
- ✓ Abhängige Variable (work_experience) = metrisch
- ✓ (Erwartungswert μ der Grundgesamtheit gegeben)

Einfacher *t*-Test in R

Schritt 2: Mithilfe der Funktion `tidycomm::t_test()` den einfachten t-Test berechnen
Einsetzen: Variable (work_experience), Erwartungswert μ der Grundgesamtheit (24)

Befehl:

```
WoJ %>% t_test(work_experience , mu = 24)
```

Ausgabe:

```
> WoJ %>% t_test(work_experience , mu = 24)
# A tibble: 1 × 9
  Variable            M    SD CI_95_LL CI_95_UL    Mu     t    df        p
* <chr>           <dbl> <dbl>    <dbl>    <dbl> <dbl> <dbl> <dbl>    <dbl>
1 work_experience  17.8  10.9     17.2     18.5    24 -19.4  1186 7.38e-73
```

Tidycomm gibt immer den zweiseitigen p-Wert aus
(Bei Bedarf: einseitiger p-Wert = p / 2)

Ergebnis: Die Berufserfahrung der vorliegenden Stichprobe beträgt 17,8 Jahre (SD=10,9) und unterscheidet sich signifikant von der durchschnittlichen Berufserfahrung von 24 Jahren innerhalb der Grundgesamtheit an Journalist*innen (t = -19,4; df = 1186; p < ,001).

ÜBUNGSBLATT: AUFGABE 4 (MIT R)

Aufgabe 4

Laut der European Social Survey (ESS) 2018 liegt das Vertrauen von Europäer*innen in Politiker*innen auf einer Skala von 1 (gering) bis 5 (hoch) durchschnittlich bei einem Wert von 3,1.

Rufen Sie den Datensatz „WoJ“ in R auf.

Prüfen Sie die Hypothese, dass sich die Journalist*innen aus der Stichprobe hinsichtlich ihres Vertrauens in Politiker*innen vom Durchschnitt der europäischen Bevölkerung unterscheiden. Das Signifikanzniveau wird auf 5% festgelegt.

Lösung: Aufgabe 4

Schritt 1: Voraussetzungsprüfung

```
WoJ %>% describe(trust_politicians)
A tibble: 1 x 15
Variable              N Missing     M    SD   Min   Q25   Mdn   Q75   Max Range CI_95_LL CI_95_UL Skewness Kurtosis
<chr>             <int>   <int> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>    <dbl>    <dbl>    <dbl>    <dbl>
trust_politicians  1200       0  2.52 0.712     1     2     3     3     4     3     2.48     2.56   -0.320     2.78
```

Schritt 2: Einfacher t-Test

```
> WoJ %>% t_test(trust_politicians, mu = 3.1)
# A tibble: 1 x 9
  Variable              M    SD CI_95_LL CI_95_UL    Mu     t    df         p
* <chr>             <dbl> <dbl>    <dbl>    <dbl> <dbl> <dbl> <dbl>     <dbl>
1 trust_politicians  2.52 0.712     2.48     2.56   3.1 -28.2  1199 8.05e-135
```

Ergebnis: Das Vertrauen in Politiker*innen der vorliegenden Stichprobe auf einer Skala von 1 (gering) bis 5 (hoch) liegt bei 2,52 Punkten (SD = 0,71) und unterscheidet sich damit signifikant von dem durchschnittlichen Wert von 3,1 in der europäischen Bevölkerung (t = -28,2; df = 1199; p = < ,001).

Wichtige Take-Aways

- **Einfacher *t*-Test in R**: berechnen mit
  - `tidycomm::describe()` für die Voraussetzungsprüfung und
  - `tidycomm::t_test(, mu =)`

Vielen Dank für Ihre Aufmerksamkeit!