

Statistik 1

Seminar

Einheit 4

26.06.2025 | Janika Saretzki, MSc.

Termine

~~Einheit 1 02.05.25 14:45-16:15 Uhr A + B HS Audimax / P3~~

~~Einheit 2 15.05.25 13:05-15:30 Uhr A + B HS Audimax / P3~~

~~Einheit 3 12.06.25 13:05-15:30 Uhr A HS P5 005~~

~~Einheit 3 13.06.25 13:50-16:15 Uhr B HS P1 105~~

Einheit 4 26.06.25 13:20-15:45 Uhr A HS P5 005

Einheit 4 27.06.25 13:50-16:15 Uhr B HS P1 105

Einheit 5 17.07.25 13:20-15:45 Uhr A HS P5 005

Einheit 5 18.07.25 13:50-16:15 Uhr B HS P1 105

Termine

Einheit 1	02.05.25	14:45-16:15 Uhr	A + B	HS Audimax / P3
Einheit 2	15.05.25	13:05-15:30 Uhr	A + B	HS Audimax / P3
Einheit 3	12.06.25	13:05-15:30 Uhr	A	HS P5 005
Einheit 3	13.06.25	13:50-16:15 Uhr	B	HS P1 105
Einheit 4	26.06.25	13:20-15:45 Uhr	A	HS P5 005
Einheit 4	27.06.25	13:50-16:15 Uhr	B	HS P1 105
Einheit 5	17.07.25	13:20-15:45 Uhr	A	HS P5 005
Einheit 5	18.07.25	13:50-16:15 Uhr	B	HS P1 105
Tutorium	21.07.25	10:00-15:30 Uhr	A + B	HS Audimax / P3
Tutorium	22.07.25	10:00-15:30 Uhr	B + B	HS Audimax / P3

Recap

Boxplot

Ein Boxplot visualisiert die Verteilung zumindest ordinal skalierten Daten. Um einen Boxplot zu zeichnen, **werden zunächst Median, das erste (Q1) und das dritte Quartil (Q3) berechnet.** Diese drei Werte bilden die sogenannte Box. Der Interquartilsabstand (IQR) ergibt sich aus $Q3 - Q1$ und beschreibt die Streubreite der mittleren 50% der Daten. **An die Box schließen sich sogenannte Whisker an.** Sie reichen jeweils bis zum letzten Datenpunkt, der noch innerhalb des Bereichs von $Q1 - 1,5 \times IQR$ bzw. $Q3 + 1,5 \times IQR$ liegt. Werte, die außerhalb dieser Grenzen liegen, gelten als Ausreißer und werden im Plot als einzelne Punkte (oder Sternchen) dargestellt.

Ein Boxplot zeigt damit auf kompakte Weise die zentrale Tendenz, die Streuung und potenzielle Ausreißer - und eignet sich besonders gut, um mehrere Gruppen miteinander zu vergleichen.

Recap

Boxplot

Ein Dozent hat die Ergebnisse einer Statistik-Klausur in einem Kurs mit 12 Teilnehmenden erfasst. Die erreichten Punktzahlen liegen im Bereich von 0 bis 100. Der Median beträgt 75,0, das erste Quartil 72,5, das dritte Quartil 85,5

Daten (in Punkten): 58, 67, 71, 73, 74, 74, 76, 87, 80, 84, 88, 92

Aufgabe: Bestimmen Sie den Interquartilsabstand (IQR) und etwaige Ausreißer. Zeichnen Sie zudem einen Boxplot und beschriften dabei die Achsen sinnvoll.

Recap - Lösung

Die Daten sind bereits sortiert, d.h. in Form einer geordneten Urliste vorliegend. Wir kennen zudem den Median ($Mdn = 75,0$), das erste Quartil ($Q1 = 72,5$) und das dritte Quartil ($Q3 = 85,5$).

1. Berechnung des Interquartilsabstandes: $IQR = Q3 - Q1 = 85,5 - 72,5 = 13,0$

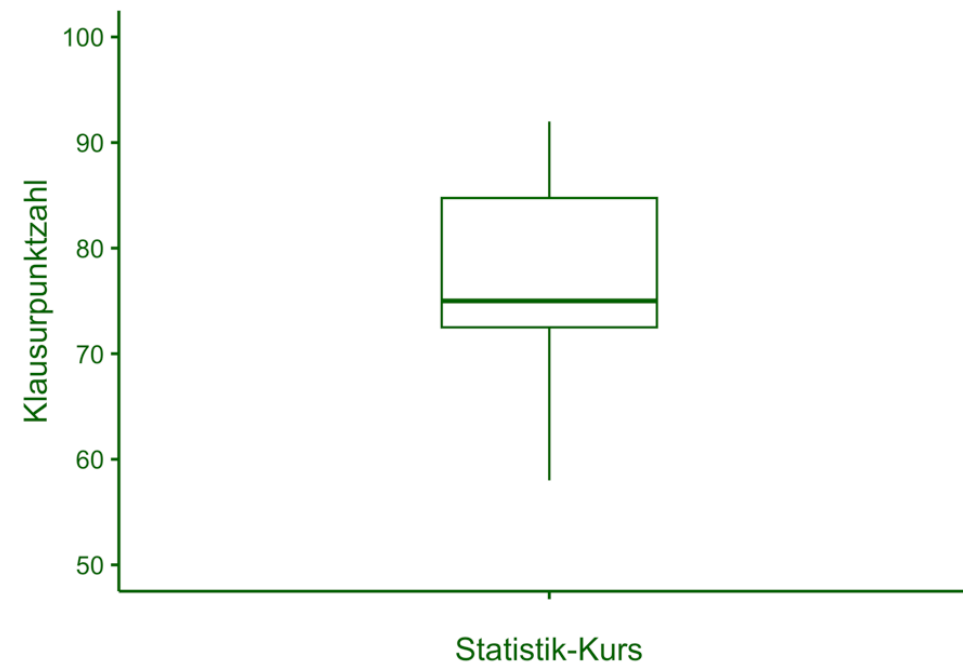
2. Berechnung der potenziellen Länge der Whisker:

- Unterer Whisker von $Q1$ bis $Q1 - 1,5 \times IQR$, d.h. von $72,5$ bis $72,5 - 1,5 \times 13,0 = 53,0$
- Oberer Whisker: von $Q3$ bis $Q3 + 1,5 \times IQR$, d.h. von $85,5$ bis $85,5 + 1,5 \times 13,0 = 105,0$

3. Ausreißer identifizierten: Beobachtete Werte reichen von 58 bis 92, d.h. kein Wert $< 53,0$ und kein Wert $> 105,0$.

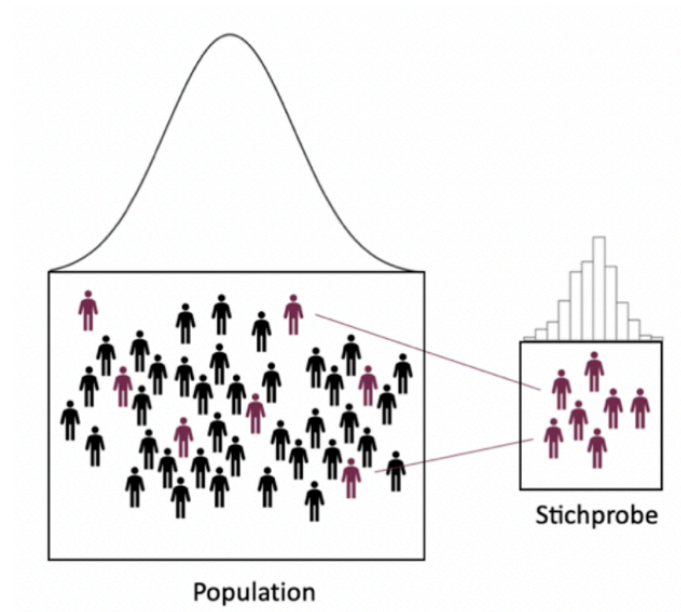
Es sind daher keine Ausreißer vorhanden!

Recap - Lösung



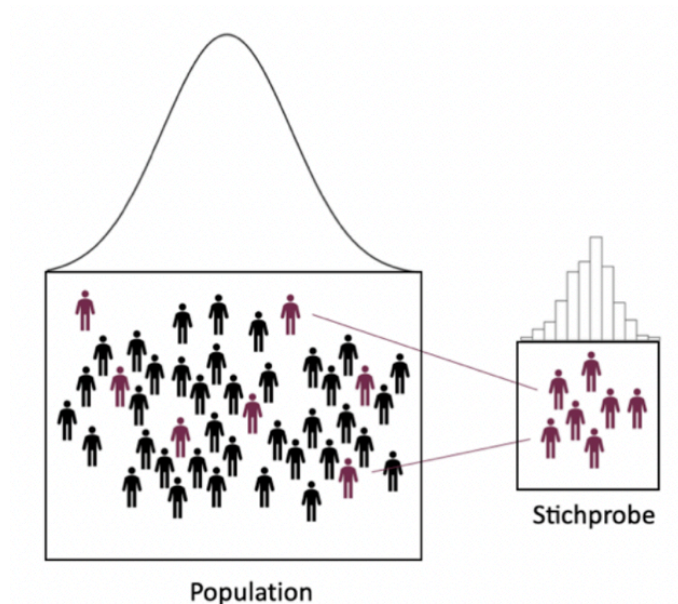
Wiederholung

Logik des Schließens von Stichprobe auf Population



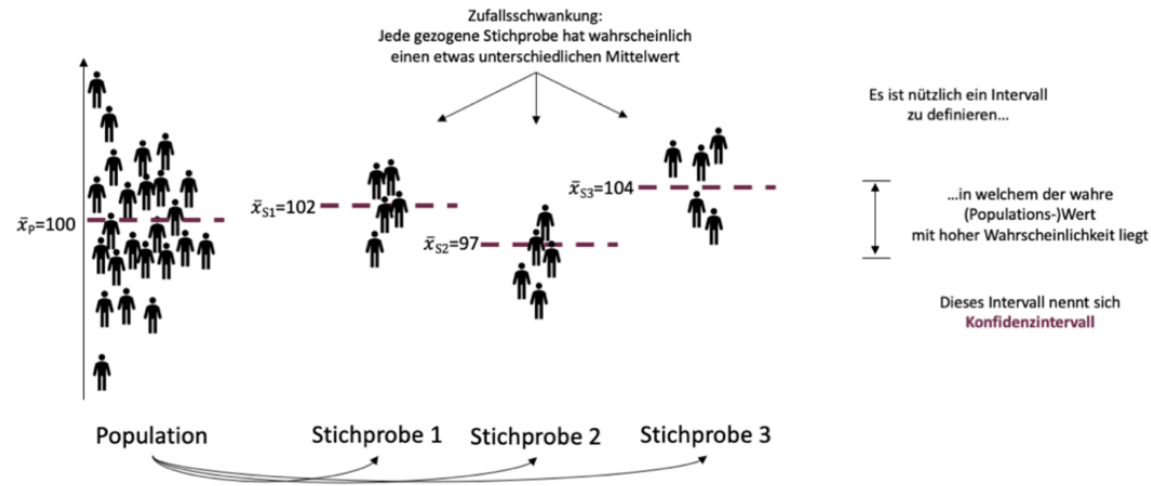
Wiederholung

Logik des Schließens von Stichprobe auf Population



Schätzwert = Wert, von dem wir auf Basis unserer Stichprobe davon ausgehen, dass er "möglichst nahe" an dem wahren Parameterwert bzw. dem wahren Wert der Maßzahl in der Population liegt (**\bar{x} und s als Schätzwerte für die Population**)

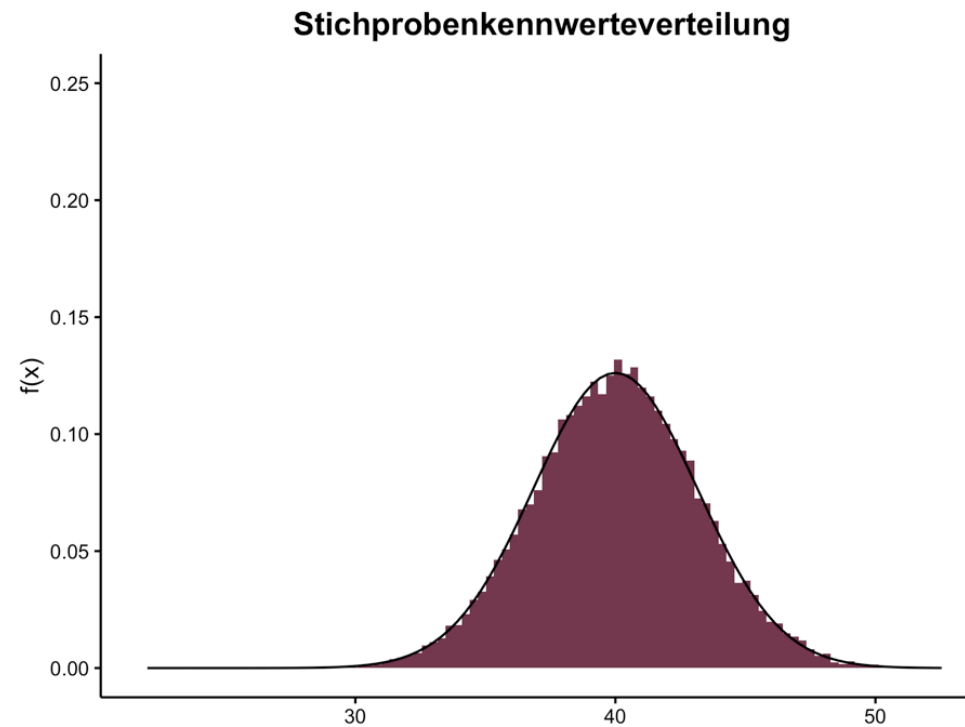
Wiederholung



Die Genauigkeit der Schätzung hängt von der Streuung der Stichprobenkennwerteverteilung ab. Diese Streuung ist die Standardabweichung der Schätzfunktion und wird **Standardfehler** genannt.

$$\sigma_{\bar{x}} = \sqrt{\frac{\sigma_x^2}{n}} = \frac{\sigma_x}{\sqrt{n}}$$

Wiederholung



Computersimulation einer Stichprobenkennwerteverteilung mit der Stichprobengröße $n = 10$

Theoretische Kennwerte: $\mu = 40,00$, $\sigma = 3,16$, $n = 10$; Empirische Kennwerte: $\bar{x} = 4,12$, $\sigma = 3,21$, $n = 10$, $a = 25.000$

Wiederholung

Ein **Konfidenzintervall** ist ein auf Stichprobendaten basierendes Intervall, das bei wiederholter Ziehung in einem festgelegten Anteil der Fälle (z.B. 95%) den wahren Wert des Populationsparameters enthalten würde.

KI für μ einer $N(\mu, \sigma^2)$ Variable bei bekannter σ^2

$$\mu_{1,2} = \bar{x} \pm z_{1-(\alpha/2)} \frac{\sigma}{\sqrt{n}}$$

KI für μ einer $N(\mu, \sigma^2)$ Variable bei unbekannter σ^2

$$\mu_{1,2} = \bar{x} \pm t_{1-(\alpha/2)} \frac{\hat{\sigma}}{\sqrt{n}}, \text{ df} = n - 1$$

Fläche*															
df	0,55	0,60	0,65	0,70	0,75	0,80	0,85	0,90	0,95	0,975	0,990	0,995	0,9995	$r_{0,05}$	$r_{0,01}$
1	0,158	0,325	0,510	0,727	1,000	1,376	1,963	3,078	6,314	12,706	31,821	63,657	636,619	0,997	1,000
2	0,142	0,289	0,445	0,617	0,816	1,061	1,386	1,886	2,920	4,303	6,965	9,925	31,598	0,950	0,990
3	0,137	0,277	0,424	0,584	0,765	0,978	1,250	1,638	2,353	3,182	4,541	5,841	12,941	0,878	0,959
4	0,134	0,271	0,414	0,569	0,741	0,941	1,190	1,533	2,132	2,776	3,747	4,604	8,610	0,811	0,917
5	0,132	0,267	0,408	0,559	0,727	0,920	1,156	1,476	2,015	2,571	3,365	4,032	6,859	0,754	0,874
6	0,131	0,265	0,404	0,553	0,718	0,906	1,134	1,440	1,943	2,447	3,143	3,707	5,959	0,707	0,834
7	0,130	0,263	0,402	0,549	0,711	0,896	1,119	1,415	1,895	2,365	2,998	3,499	5,405	0,666	0,798
8	0,130	0,262	0,399	0,546	0,706	0,889	1,108	1,397	1,860	2,306	2,896	3,355	5,041	0,632	0,765
9	0,129	0,261	0,398	0,543	0,703	0,883	1,100	1,383	1,833	2,262	2,821	3,250	4,781	0,602	0,735
10	0,129	0,260	0,397	0,542	0,700	0,879	1,093	1,372	1,812	2,228	2,764	3,169	4,587	0,576	0,708
11	0,129	0,260	0,396	0,540	0,697	0,876	1,088	1,363	1,796	2,201	2,718	3,106	4,437	0,553	0,684
12	0,128	0,259	0,395	0,539	0,695	0,873	1,083	1,356	1,782	2,179	2,681	3,055	4,318	0,532	0,661
13	0,128	0,259	0,394	0,538	0,694	0,870	1,079	1,350	1,771	2,160	2,650	3,012	4,221	0,514	0,641
14	0,128	0,258	0,393	0,537	0,692	0,868	1,076	1,345	1,761	2,145	2,624	2,977	4,140	0,497	0,623
15	0,128	0,258	0,393	0,536	0,691	0,866	1,074	1,341	1,753	2,131	2,602	2,947	4,073	0,482	0,606
16	0,128	0,258	0,392	0,535	0,690	0,865	1,071	1,337	1,746	2,120	2,583	2,921	4,015	0,468	0,590
17	0,128	0,257	0,392	0,534	0,689	0,863	1,069	1,333	1,740	2,110	2,567	2,898	3,965	0,456	0,575
18	0,127	0,257	0,392	0,534	0,688	0,862	1,067	1,330	1,734	2,101	2,552	2,878	3,922	0,444	0,561
19	0,127	0,257	0,391	0,533	0,688	0,861	1,066	1,328	1,729	2,093	2,539	2,861	3,883	0,433	0,549
20	0,127	0,257	0,391	0,533	0,687	0,860	1,064	1,325	1,725	2,086	2,528	2,845	3,850	0,423	0,537
21	0,127	0,257	0,391	0,532	0,686	0,859	1,063	1,323	1,721	2,080	2,518	2,831	3,819	0,413	0,526
22	0,127	0,256	0,390	0,532	0,686	0,858	1,061	1,321	1,717	2,074	2,508	2,819	3,792	0,404	0,515
23	0,127	0,256	0,390	0,532	0,685	0,858	1,060	1,319	1,714	2,069	2,500	2,807	3,767	0,396	0,505
24	0,127	0,256	0,390	0,531	0,685	0,857	1,059	1,318	1,711	2,064	2,492	2,797	3,745	0,388	0,496
25	0,127	0,256	0,390	0,531	0,684	0,856	1,058	1,316	1,708	2,060	2,485	2,787	3,725	0,381	0,487
26	0,127	0,256	0,390	0,531	0,684	0,856	1,058	1,315	1,706	2,056	2,479	2,779	3,707	0,374	0,478
27	0,127	0,256	0,389	0,531	0,684	0,855	1,057	1,314	1,703	2,052	2,473	2,771	3,690	0,367	0,470
28	0,127	0,256	0,389	0,530	0,683	0,855	1,056	1,313	1,701	2,048	2,467	2,763	3,674	0,361	0,463
29	0,127	0,256	0,389	0,530	0,683	0,854	1,055	1,311	1,699	2,045	2,462	2,756	3,659	0,355	0,456
30	0,127	0,256	0,389	0,530	0,683	0,854	1,055	1,310	1,697	2,042	2,457	2,750	3,646	0,349	0,449
40	0,126	0,255	0,388	0,529	0,681	0,851	1,050	1,303	1,684	2,021	2,423	2,704	3,551	0,304	0,393
60	0,126	0,254	0,387	0,527	0,679	0,848	1,046	1,296	1,671	2,000	2,390	2,660	3,460	0,250	0,325
120	0,126	0,254	0,386	0,526	0,677	0,845	1,041	1,289	1,658	1,980	2,358	2,617	3,373	0,178	0,232
z	0,126	0,253	0,385	0,524	0,674	0,842	1,036	1,282	1,645	1,960	2,326	2,576	3,291		

* Die Flächenanteile für negative t -Werte ergeben sich nach der Beziehung $P(-t_{df}) = 1 - P(t_{df})$

Wiederholung

Wie lese ich die Tabelle?

Erster Schritt

- Berechne den Wert (d.h. das benötigte Quantil) von $1 - \alpha/2$ (zweiseitiges Konfidenzintervall) bzw. $1 - \alpha$ (einseitiges Konfidenzintervall)
- Bestimme die Freiheitsgrade: $n - 1$

Zweiter Schritt

- In der ersten Spalte stehen die Freiheitsgrade (df)
- In der ersten Zeile stehen die Werte von $1 - \alpha/2$ (zweiseitiges Konfidenzintervall) bzw. $1 - \alpha$ (einseitiges Konfidenzintervall)
- **Die Zelle, in der sich beide schneiden, enthält den gesuchten t-Wert**

Übungsaufgabe 1

Eine Psychologin interessiert sich für das Stresslevel von Universitätsstudierenden. Sie verwendet eine Skala von 0 bis 100, wobei 0 für "kein Stress" und 100 für "extremen Stress" steht. Dazu erhebt sie eine Stichprobe von 30 Studierenden. Die gemessenen Stresswerte ergeben im Durchschnitt einen Wert von 75, mit einer Standardabweichung von 12. Sie möchte nun Rückschlüsse auf das Stresslevel aller Universitätsstudierenden ziehen.

a) Welche Werte aus dem Text entsprechen den folgenden statistischen Kennwerten:

$$\mu, \sigma, \bar{x}, s$$

b) Berechnen Sie den **Standardfehler des Mittelwerts**. Was sagt dieser aus, und wovon hängt er ab?

c) Geben Sie die Formel für das **95%-Konfidenzintervall** an und berechnen Sie es.

Übungsaufgabe 1 - Lösung

Eine Psychologin interessiert sich für das Stresslevel von Universitätsstudierenden. Sie verwendet eine Skala von 0 bis 100, wobei 0 für "kein Stress" und 100 für "extremen Stress" steht. Dazu erhebt sie eine Stichprobe von 30 Studierenden. Die gemessenen Stresswerte ergeben im Durchschnitt einen Wert von 75, mit einer Standardabweichung von 12. Sie möchte nun Rückschlüsse auf das Stresslevel aller Universitätsstudierenden ziehen.

a) Welche Werte aus dem Text entsprechen den folgenden statistischen Kennwerten:

$$\mu, \sigma, \bar{x}, s$$

- $\bar{x} = 75$ (durchschnittliches Stresslevel in der Stichprobe)
- $s = 12$ (Streuung des Stresslevels in der Stichprobe, Stichprobenstandardabweichung)
- μ (durchschnittliches Stresslevel in der Population): unbekannt
- σ (Streuung des Stresslevels in der Population, Populationsstandardabweichung): unbekannt

→ Stichproben- vs. Populationsebene

→ Wir verwenden \bar{x} und s als Schätzwerte für μ und σ

Übungsaufgabe 1 - Lösung

Eine Psychologin interessiert sich für das Stresslevel von Universitätsstudierenden. Sie verwendet eine Skala von 0 bis 100, wobei 0 für "kein Stress" und 100 für "extremen Stress" steht. Dazu erhebt sie eine Stichprobe von 30 Studierenden. Die gemessenen Stresswerte ergeben im Durchschnitt einen Wert von 75, mit einer Standardabweichung von 12. Sie möchte nun Rückschlüsse auf das Stresslevel aller Universitätsstudierenden ziehen.

b) Berechnen Sie den **Standardfehler des Mittelwerts**. Was sagt dieser aus, und wovon hängt er ab?

$$\text{Standardabweichung der Stichprobenmittelwerte: } \sigma_{\bar{x}} = \sqrt{\frac{\sigma_x^2}{n}} = \frac{\sigma_x}{\sqrt{n}}$$

Der Standardfehler gibt an, wie stark der Mittelwert aus der Stichprobe typischerweise vom wahren Populationsmittelwert abweicht – also wie präzise die Schätzung ist. Er wird durch a) die Streuung der Messwerte in der Population/Stichprobenkennwerteverteilung und b) die Stichprobengröße beeinflusst (siehe Formel).

Der Standardfehler ist umso kleiner, je a) kleiner die Streuung ist und b) größer die Stichprobe ist.

In der Praxis: Verwendung der empirischen Standardabweichung (= Stichprobenebene)

$$\sigma_{\bar{x}} = \frac{s}{\sqrt{n}} = \frac{12}{\sqrt{30}} = 2,19$$

Übungsaufgabe 1 - Lösung

Eine Psychologin interessiert sich für das Stresslevel von Universitätsstudierenden. Sie verwendet eine Skala von 0 bis 100, wobei 0 für "kein Stress" und 100 für "extremen Stress" steht. Dazu erhebt sie eine Stichprobe von 30 Studierenden. Die gemessenen Stresswerte ergeben im Durchschnitt einen Wert von 75, mit einer Standardabweichung von 12. Sie möchte nun Rückschlüsse auf das Stresslevel aller Universitätsstudierenden ziehen.

c) Geben Sie die Formel für das **95%-Konfidenzintervall** an und berechnen Sie es.

Die Populationsvarianz ist nicht bekannt, daher: Verwendung der t-Verteilung!

$$\mu_{1,2} = \bar{x} \pm t_{1-\alpha/2} \cdot \frac{\hat{\sigma}}{\sqrt{n}}$$

$$df = n - 1$$

Werte einsetzen:

$$t_{1-\alpha/2} = t_{1-0,05/2} = t_{0,975}$$

$$df = 30 - 1 = 29$$

... mit diesen Werten werfen wir nun einen Blick in die t-Tabelle!

Übungsaufgabe 1 - Lösung

Eine Psychologin interessiert sich für das Stresslevel von Universitätsstudierenden. Sie verwendet eine Skala von 0 bis 100, wobei 0 für "kein Stress" und 100 für "extremen Stress" steht. Dazu erhebt sie eine Stichprobe von 30 Studierenden. Die gemessenen Stresswerte ergeben im Durchschnitt einen Wert von 75, mit einer Standardabweichung von 12. Sie möchte nun Rückschlüsse auf das Stresslevel aller Universitätsstudierenden ziehen.

c) Geben Sie die Formel für das **95%-Konfidenzintervall** an und berechnen Sie es.

$$\mu_{1,2} = 75 \pm 2,045 \cdot \frac{12}{\sqrt{30}}$$

$$\mu_1 = 70,52$$

$$\mu_2 = 79,48$$

Auf Basis unserer Stichprobe ergibt sich ein 95%-Konfidenzintervall von 70,52 bis 79,48. Dieses Intervall wurde so berechnet, dass es mit 95%iger Konfidenz den wahren Mittelwert der Population überdeckt.

Wenn wir viele Stichproben ziehen und jeweils ein 95%-Konfidenzintervall berechnen, dann würden etwa 95% dieser Intervalle den wahren Populationsmittelwert einschließen.

Übungsaufgabe 2

Im Rahmen eines psychologischen Experiments wird die Rundenzeit einer Ratte in einem Parcours untersucht. In einer Trainingssession absolviert die Ratte 54 Runden. Die durchschnittliche Rundenzeit beträgt 17 Minuten und 30 Sekunden. Es wird angenommen, dass die Rundenzeiten normalverteilt sind, mit einer (Populations-)Varianz von $2,5 \text{ Minuten}^2$.

Wie lautet das 90%-Konfidenzintervall für den Mittelwert der Rundenzeit?

Übungsaufgabe 2 - Lösung

Im Rahmen eines psychologischen Experiments wird die Rundenzeit einer Ratte in einem Parcours untersucht. In einer Trainingssession absolviert die Ratte 54 Runden. Die durchschnittliche Rundenzeit beträgt 17 Minuten und 30 Sekunden. Es wird angenommen, dass die Rundenzeiten normalverteilt sind, mit einer (Populations-)Varianz von 2,5 Minuten².

Wie lautet das 90%-Konfidenzintervall für den Mittelwert der Rundenzeit?

Gegeben ist:

$$n = 54$$

$$\bar{x} = 17,5 \text{ min}$$

$$\text{Populationsvarianz: } \sigma^2 = 2,5 \Rightarrow \sigma = \sqrt{2,5} = 1,58 \text{ min}$$

$$\text{Konfidenzniveau: } 90\% \Rightarrow \alpha = 0,10$$

Die Populationsvarianz ist bekannt, daher: Verwendung der z-Verteilung!

$$z_{1-\alpha/2} = z_{1-0,10/2} = z_{0,95}$$

$$\mu_{1,2} = \bar{x} \pm z_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

Übungsaufgabe 2 - Lösung

Im Rahmen eines psychologischen Experiments wird die Rundenzeit einer Ratte in einem Parcours untersucht. In einer Trainingssession absolviert die Ratte 54 Runden. Die durchschnittliche Rundenzeit beträgt 17 Minuten und 30 Sekunden. Es wird angenommen, dass die Rundenzeiten normalverteilt sind, mit einer (Populations-)Varianz von 2,5 Minuten².

Wie lautet das 90%-Konfidenzintervall für den Mittelwert der Rundenzeit?

Beachte: Beim Nachschlagen in der t-Tabelle sucht man den kritischen t-Wert anhand der Freiheitsgrade (df) und dem gewünschten Konfidenzniveau. In der z-Tabelle hingegen sucht man die Fläche (bzw. Wahrscheinlichkeit) unter der Normalverteilung und liest dazu den passenden z-Wert ab.

→ Fläche: 0,9495 (bzw. 0,9505)

→ z-Wert: 1,64 oder 1,65 (beides wäre richtig)

$$\mu_{1,2} = 17,5 \pm 1,64 \cdot \frac{1,58}{\sqrt{54}}$$

$$\mu_1 = 17,15$$

$$\mu_2 = 17,85$$

Auf Basis der Stichprobe ergibt sich ein 90%-Konfidenzintervall von 17,15 bis 17,85. Das bedeutet: Wir gehen mit 90% Konfidenz davon aus, dass der wahre Mittelwert μ der Population innerhalb dieses Intervalls liegt.

Wiederholung

Zur Durchführung eines Hypothesentests sind folgende Schritte notwendig:

1. Formulierung der Hypothesen:

H_0 : Kein Effekt H_1 : Es liegt ein Effekt vor

2. Auswahl der zugrunde liegenden Verteilung

3. Festlegung von Annahme- und Ablehnungsbereich: Bestimmung des kritischen Werts

4. Einordnung des Beobachtungswerts in die Verteilung

- **Binomialverteilung**: Wahrscheinlichkeiten
- **z-Verteilung**: Mittelwerte bei bekanntem σ
- **t-Verteilung**: Mittelwerte bei unbekanntem σ
- **F-Verteilung**: Varianzvergleiche
- **Chi-Quadrat-Verteilung**: Häufigkeiten / Proportionen

5. Vergleich von Teststatistik und kritischem Wert

6. Entscheidung: Test signifikant oder nicht signifikant

Wiederholung

Ein-Stichproben Test für Mittelwerte

1. Hypothesenpaar

Hypothese über μ einer normalverteilten Variable, wobei die Populationsvarianz bekannt ist

Statistische Hypothesenpaare:

$H_0: \mu = \mu_0,$	$H_1: \mu \neq \mu_0$	(ungerichtet)	→ zweiseitig
$H_0: \mu \leq \mu_0,$	$H_1: \mu > \mu_0$	(gerichtet)	→ einseitig (rechtsseitig)
$H_0: \mu \geq \mu_0,$	$H_1: \mu < \mu_0$	(gerichtet)	→ einseitig (linksseitig)

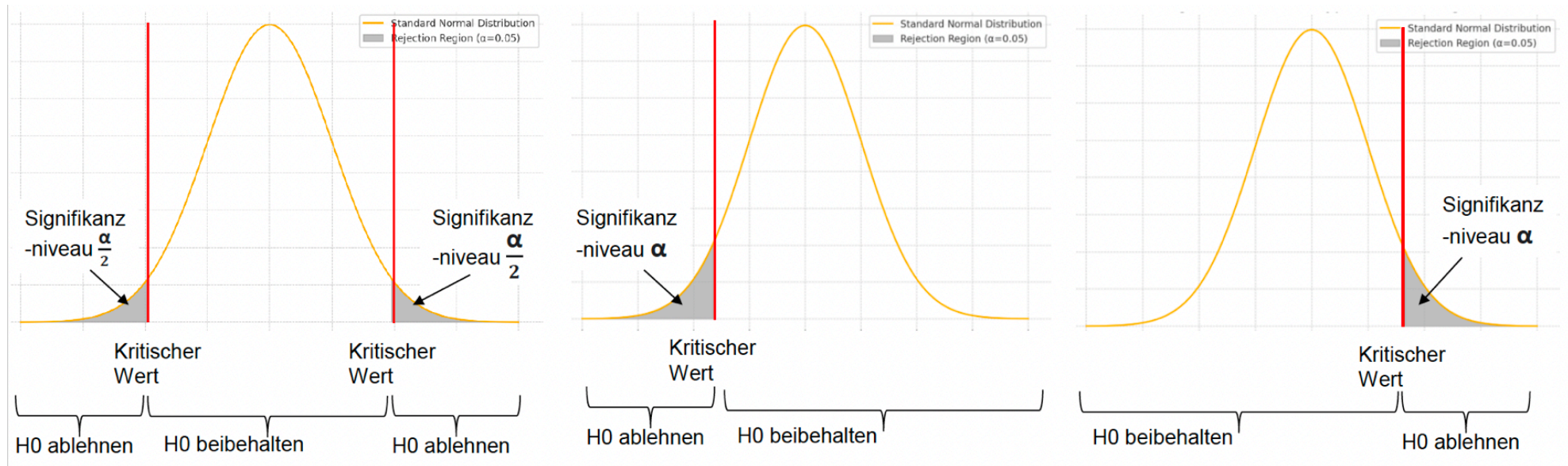
μ = Populationsmittelwert; μ_0 = hypothetischer Populationsparameter

2. Bestimmung einer zugrundeliegenden Verteilung

- Wenn wir an Mittelwerten interessiert sind UND die Populationsvarianz bekannt ist: **z-Verteilung**
- Wenn wir an Mittelwerten interessiert sind und die Populationsvarianz NICHT bekannt ist: **t-Verteilung**

Wiederholung

3. Annahme- und Ablehnungsbereich der Nullhypothese



Der kritische Wert wird in der Verteilungstabelle abgelesen:

Einseitiger Test: $z_{\text{kritisch}} = 1 - \alpha$

Zweiseitiger Test: $z_{\text{kritisch}} = 1 - \frac{\alpha}{2}$

Übungsaufgabe 3

Ein Psychologe möchte untersuchen, ob Kinder mit jüngeren Geschwistern im Durchschnitt eine höhere Intelligenzleistung aufweisen als die Population aller Kinder. In einer Stichprobe von $N = 225$ Kindern mit jüngeren Geschwistern wird ein Mittelwert von $\bar{x} = 101,8$ beobachtet. Der Intelligenzquotient (IQ) unter Kindern gilt als normalverteilt mit einem Erwartungswert von $\mu = 100$ und einer Standardabweichung von $\sigma = 15$.

- a) Handelt es sich bei der Fragestellung um eine gerichtete oder ungerichtete Hypothese? Formulieren Sie die inhaltliche und statistische Null- und Alternativhypothese entsprechend.
- b) Angenommen, der Forscher hat das Signifikanzniveau vorab auf 5% festgelegt - zu welcher Entscheidung hinsichtlich der Nullhypothese kommt er?
- c) Wie würde das Testergebnis ausfallen, wenn stattdessen ein ungerichteter Test auf dem 5%-Niveau durchgeführt würde?

Übungsaufgabe 3 - Lösung

Ein Psychologe möchte untersuchen, ob Kinder mit jüngeren Geschwistern im Durchschnitt eine höhere Intelligenzleistung aufweisen als die Population aller Kinder. In einer Stichprobe von $N = 225$ Kindern mit jüngeren Geschwistern wird ein Mittelwert von $\bar{x} = 101,8$ beobachtet. Der Intelligenzquotient (IQ) unter Kindern gilt als normalverteilt mit einem Erwartungswert von $\mu = 100$ und einer Standardabweichung von $\sigma = 15$.

a) Handelt es sich bei der Fragestellung um eine gerichtete oder ungerichtete Hypothese? Formulieren Sie die inhaltliche und statistische Null- und Alternativhypothese entsprechend.

Es handelt sich um eine gerichtete Hypothese, da ein Richtungseffekt („höher“) angenommen wird.

Inhaltliche Hypothesen:

- Nullhypothese (H_0): Es gibt keinen höheren IQ bei Kindern mit jüngeren Geschwistern im Vergleich zur Gesamtpopulation.
- Alternativhypothese (H_1): Kinder mit jüngeren Geschwistern haben einen höheren IQ als der Populationsmittelwert.

Statistische Hypothesen:

$$H_0 : \mu \leq 100$$

$$H_1 : \mu > 100$$

Übungsaufgabe 3 - Lösung

Ein Psychologe möchte untersuchen, ob Kinder mit jüngeren Geschwistern im Durchschnitt eine höhere Intelligenzleistung aufweisen als die Population aller Kinder. In einer Stichprobe von $N = 225$ Kindern mit jüngeren Geschwistern wird ein Mittelwert von $\bar{x} = 101,8$ beobachtet. Der Intelligenzquotient (IQ) unter Kindern gilt als normalverteilt mit einem Erwartungswert von $\mu = 100$ und einer Standardabweichung von $\sigma = 15$.

b) Angenommen, der Forscher hat das Signifikanzniveau vorab auf 5% festgelegt - zu welcher Entscheidung hinsichtlich der Nullhypothese kommt er?

Einstichproben z-Test: Es wird geprüft, ob der Mittelwert einer Grundgesamtheit von einem bestimmten Wert abweicht, in diesem Fall ob der beobachtete Stichprobenmittelwert (signifikant) größer ist als der erwartete Populationsmittelwert. Wir berechnen also den empirischen z-Wert (die Teststatistik) auf Basis der Stichprobe und vergleichen diesen mit dem kritischen z-Wert, der dem gewählten Signifikanzniveau entspricht. Liegt der empirische z-Wert im Ablehnungsbereich, wird die Nullhypothese verworfen.

$$z_{\text{emp}} = \sqrt{n} \cdot \frac{\bar{x} - \mu_0}{\sigma}$$
$$z_{\text{emp}} = \sqrt{225} \cdot \frac{101,8 - 100}{15} = 1,8$$

Übungsaufgabe 3 - Lösung

Ein Psychologe möchte untersuchen, ob Kinder mit jüngeren Geschwistern im Durchschnitt eine höhere Intelligenzleistung aufweisen als die Population aller Kinder. In einer Stichprobe von $N = 225$ Kindern mit jüngeren Geschwistern wird ein Mittelwert von $\bar{x} = 101,8$ beobachtet. Der Intelligenzquotient (IQ) unter Kindern gilt als normalverteilt mit einem Erwartungswert von $\mu = 100$ und einer Standardabweichung von $\sigma = 15$.

b) Angenommen, der Forscher hat das Signifikanzniveau vorab auf 5% festgelegt - zu welcher Entscheidung hinsichtlich der Nullhypothese kommt er?

Bei einem Signifikanzniveau von 5% und rechtsseitiger Testung ist der kritische z-Wert 1,64, denn:

$$z_{\text{krit}} = z_{1-\alpha} = z_{1-0,05} = z_{0,95}$$

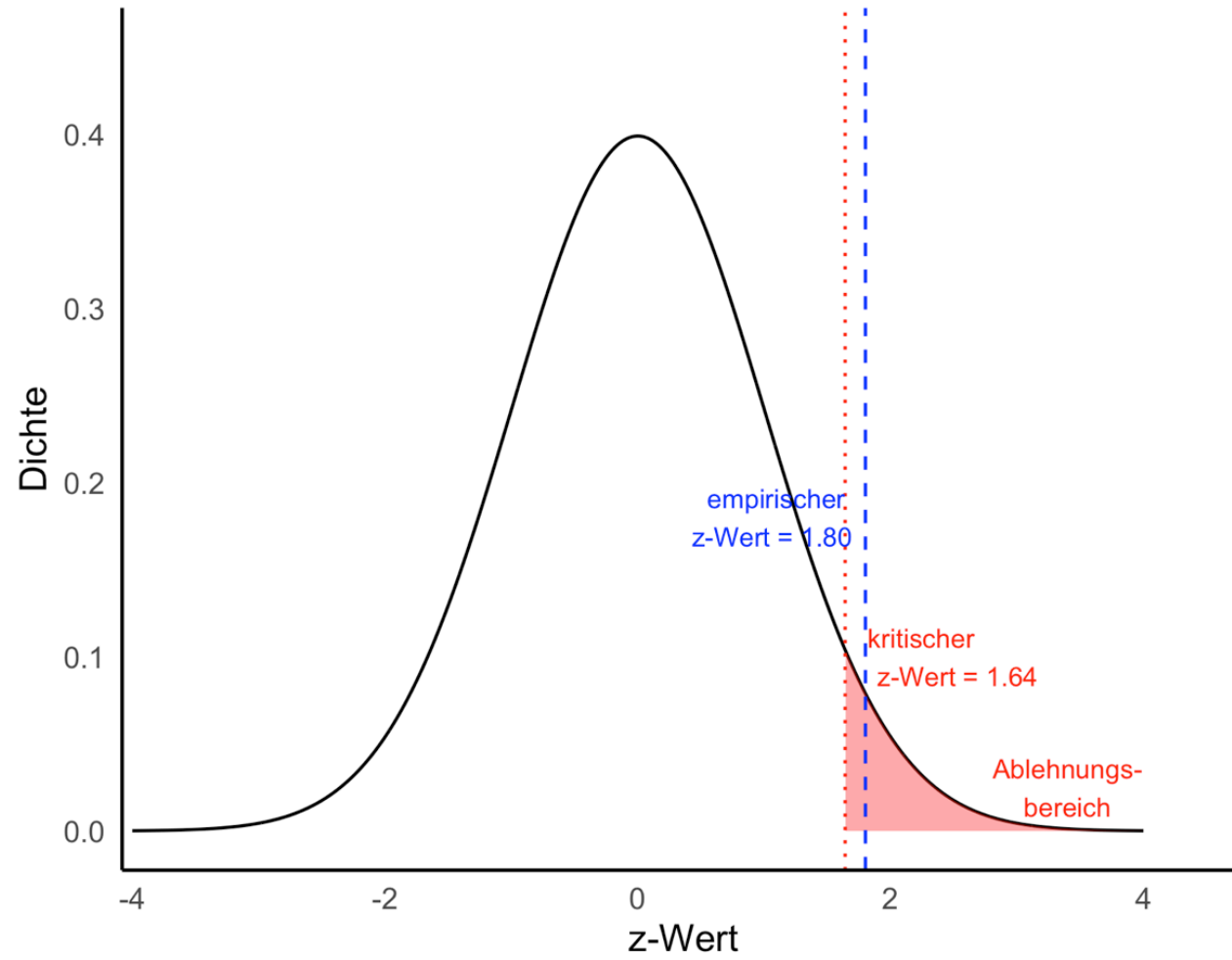
$$z_{0,95} = 1,64$$

$$1,8 > 1,64$$

Der empirische z-Wert liegt im Ablehnungsbereich der Nullhypothese – der kritische z-Wert markiert die Schwelle, ab der ein Ergebnis unter der Annahme der Nullhypothese als sehr unwahrscheinlich gilt. Ist in diesem Fall - entsprechend der Hypothese - die Bedingung ...

$$z_{\text{emp}} > z_{\text{krit}}$$

... erfüllt, betrachten wir den beobachteten Effekt als statistisch signifikant. Die Nullhypothese wird verworfen, da das Ergebnis nicht mehr mit bloßem Zufall erklärbar erscheint. Der Forscher würde daraus schließen, dass Kinder mit jüngeren Geschwistern im Mittel eine höhere Intelligenzleistung aufweisen als Kinder in der Gesamtpopulation.



Übungsaufgabe 3 - Lösung

Ein Psychologe möchte untersuchen, ob Kinder mit jüngeren Geschwistern im Durchschnitt eine höhere Intelligenzleistung aufweisen als die Population aller Kinder. In einer Stichprobe von $N = 225$ Kindern mit jüngeren Geschwistern wird ein Mittelwert von $\bar{x} = 101.8$ beobachtet. Der Intelligenzquotient (IQ) unter Kindern gilt als normalverteilt mit einem Erwartungswert von $\mu = 100$ und einer Standardabweichung von $\sigma = 15$.

c) Wie würde das Testergebnis ausfallen, wenn stattdessen ein ungerichteter Test auf dem 5%-Niveau durchgeführt würde?

Bei einem Signifikanzniveau von 5% und zweiseitiger Testung ergibt sich folgender kritischer z-Wert:

$$z_{\text{krit}} = z_{1-\frac{\alpha}{2}} = z_{1-0,025} = z_{0,975} \approx \pm 1,96$$

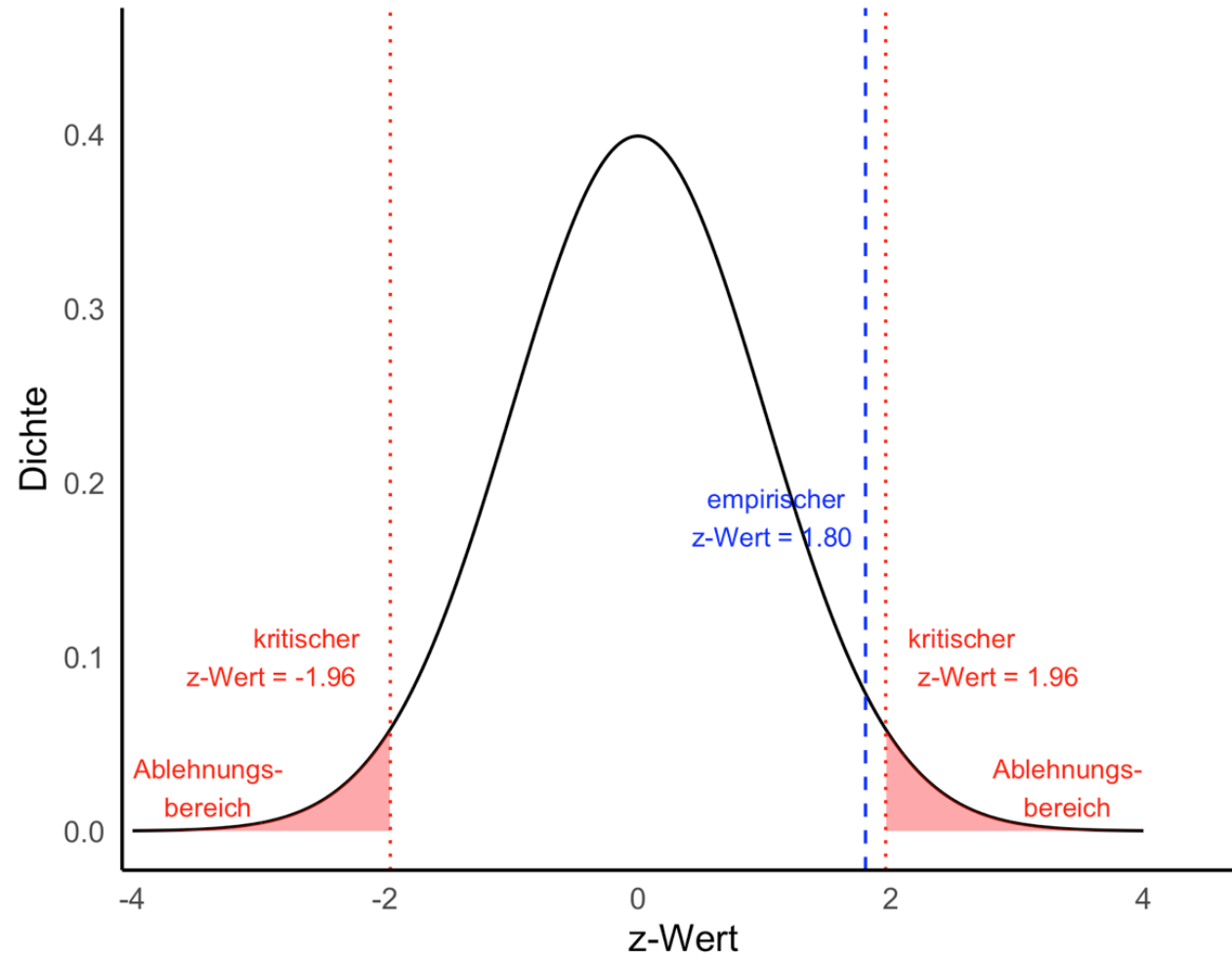
... das bedeutet: Jeweils 2,5% Irrtumswahrscheinlichkeit liegen in den beiden äußeren Bereichen der Normalverteilung.

Erinnerung:

$$z_{\text{emp}} = 1,8$$

$$1,8 < 1,96$$

Würde es sich um einen zweiseitigen Test handeln, läge der empirische Wert nicht im Ablehnungsbereich. Die Nullhypothese würde beibehalten, da das Ergebnis nicht signifikant ist. Der Forscher würde also keinen belegbaren Unterschied zur Gesamtpopulation feststellen.



z-Tabelle

z-Wert*	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,00	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,10	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,20	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,30	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,40	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,50	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,60	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,70	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,80	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,90	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,00	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,10	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,20	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,30	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,40	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,50	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,60	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,70	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,80	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1,90	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767
2,00	0,9772	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817
2,10	0,9821	0,9826	0,9830	0,9834	0,9838	0,9842	0,9846	0,9850	0,9854	0,9857
2,20	0,9861	0,9864	0,9868	0,9871	0,9875	0,9878	0,9881	0,9884	0,9887	0,9890
2,30	0,9893	0,9896	0,9898	0,9901	0,9904	0,9906	0,9909	0,9911	0,9913	0,9916
2,40	0,9918	0,9920	0,9922	0,9925	0,9927	0,9929	0,9931	0,9932	0,9934	0,9936
2,50	0,9938	0,9940	0,9941	0,9943	0,9945	0,9946	0,9948	0,9949	0,9951	0,9952
2,60	0,9953	0,9955	0,9956	0,9957	0,9959	0,9960	0,9961	0,9962	0,9963	0,9964
2,70	0,9965	0,9966	0,9967	0,9968	0,9969	0,9970	0,9971	0,9972	0,9973	0,9974
2,80	0,9974	0,9975	0,9976	0,9977	0,9977	0,9978	0,9979	0,9979	0,9980	0,9981
2,90	0,9981	0,9982	0,9982	0,9983	0,9984	0,9984	0,9985	0,9985	0,9986	0,9986
3,00	0,9987	0,9987	0,9987	0,9988	0,9988	0,9989	0,9989	0,9989	0,9990	0,9990

* Beispiel: $P(z \leq 1,52) = 0,936$. Der Flächenanteil 0,936 befindet sich in der Zeile für $z = 1,50$ und Spalte 0,02. Die Flächenanteile für negative z -Werte ergeben sich nach der Beziehung $P(-z) = 1 - P(z)$.

Möglichkeit der Visualisierung

