Übung 12

-

Testverfahren
Musterlösung

Aktuelle Version: 30. August 2022

Hinweise:

- Übungen sind mit Vorteil alleine zu lösen.
- Benutzen Sie die Musterlösungen nur zur Korrektur.
- Die Übungen sind wichtige Vorbereitungen für die Prüfung. Lösen sie die Übungen sorgfältig und stellen Sie die Lösungswege übersichtlich dar.
- (Ergänzte) Vorlesungsunterlagen und Fachbücher helfen beim Lösen von Übungen und bringen gleichzeitig eine erweiterte Ansicht auf die Problemstellung.
- Wenn Sie die Übungen nicht verstehen, fragen Sie!

Übung 1. *Fragen*

1. Welche Typen von statistischen Tests gibt es?

   (a) *Parametertests*: Es wird anhand einer Stichprobe eine Hypothese zu einem Parameter (Mittelwert, Varianz, etc.) geprüft.

   (b) *Verteilungstests*: Es wird anhand einer Stichprobe eine Hypothese zur Verteilung einer Zufallsvariable geprüft.

   (c) *Unabhängigkeitstest*: Es wird anhand einer Stichprobe eine Hypothese zur Unabhängigkeit von zwei Zufallsvariablen geprüft.

2. Welche zwei Arten von Fehler gibt es?

   (a) *Fehler erster Art*: fehlerhaftes Ablehnen einer Hypothese.

   (b) *Fehler erster Art*: fehlerhaftes Annehmen einer Hypothese.

3. Erklären Sie den Begriff Signifikanz!

   Signifikanz bezieht sich allgemein auf die Aussagekraft der Daten.

   Die *statistische Signifikanz* gibt Auskunft darüber, ob Stichprobendaten so stark von einer Annahme (Nullhypothese) abweichen, dass die Annahme verworfen wird.

   Neben der statistischen Signifikanz findet man im Bereich der Medizin und Psychologie auch noch die *klinische Signifikanz*. Diese sagt aus, ob Studienergebnisse von klinischer Bedeutsamkeit sind.

4. Was ist der Unterschied von abhängigen und unabhängigen Stichproben?

   Unabhängige Stichproben setzen sich aus voneinander unabhängigen Personen, Messungen, Maschinen, Verfahren etc. zusammen. Im Gegensatz dazu handelt es sich bei abhängigen Stichproben um Datenpaare oder Datengruppen, die zusammengehören und keine statistisch voneinander unabhängigen Messungen darstellen, z.B. bei Messwiederholungen.

Übung 2. *Parametertest*

Die mittlere Lebensdauer einer Stichprobe von 100 Glühbirnen, die von einer Firma hergestellt wurden, wurde mit 1570 h und die Standardabweichung mit 120 h berechnet. Man teste die Hypothese, dass die mittlere Lebensdauer $\mu = 1600h$ ist bei einer Irrtumswahrscheinlichkeit von 0.05 respektive 0.01.

Wir habe also die folgenden Hypothesen:

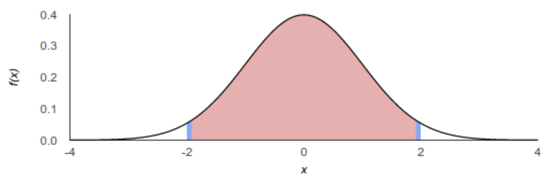
H_0: Die mittlere Lebensdauer ist $\mu \neq 1600h$.
H_1: Die mittlere Lebensdauer ist $\mu = 1600h$.

Es wurden $n = 100 > 30$ Glühbirnen entnommen, so dass wir eine Normalverteilung der Stichprobenmittelwerte annehmen dürfen. Diese Verteilung hat eine Standardabweichung von:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{120}{10} = 12$$

Aus der Tabelle ermitteln wir jeweils den z-Wert für ein beidseitiges Konfidenzintervall:

$$z_{(0.975)} = 1.96$$
$$z_{(0.995)} = 2.58$$



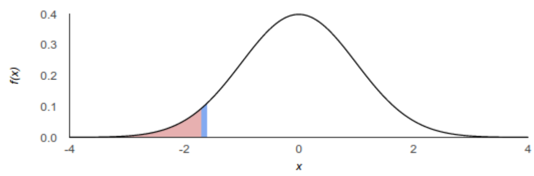
Wir können nun die Konfidenzintervalle berechnen:

$$\bar{x}_{95\%} = \mu \pm z_{(0.975)}\sigma_{\bar{x}} = 1600 \pm 1.96 \cdot 12 = [1576, 1624]$$
$$\bar{x}_{99\%} = \mu \pm z_{(0.995)}\sigma_{\bar{x}} = 1600 \pm 2.58 \cdot 12 = [1569, 1631]$$

Der gemessene Wert liegt mit einer Fehlerwahrscheinlichkeit von 5% nicht im Konfidenzintervall, wir lehnen die Hypothese zugunsten der Nullhypothese ab: Der Mittelwert μ beträgt nicht 1600h. Der gemessene Wert liegt mit einer Fehlerwahrscheinlichkeit von 1% jedoch im Konfidenzintervall, wir akzeptiere die Hypothese und verwerfen die Nullhypothese: Der Mittelwert μ beträgt 1600h.

Alternativ könnten wir auch einseitige Konfidenzintervalle berechnen. Aus der Tabelle ermitteln wir dazu die folgenden z-Werte und Intervallgrenzen:

$$z_{(0.95)} = 1.64$$
$$z_{(0.99)} = 2.32$$
$$\bar{x}_{u,95\%} = \mu - z_{(0.95)}\sigma_{\bar{x}} = 1600 - 1.64 \cdot 12 = 1580$$
$$\bar{x}_{u,99\%} = \mu - z_{(0.99)}\sigma_{\bar{x}} = 1600 - 2.32 \cdot 12 = 1572$$



Hier liegt nun der gemessene Wert für beide Fehlerwahrscheinlichkeiten ausserhalb des Konfidenzintervalls, wir lehnen die Hypothese zugunsten der Nullhypothese ab: Der Mittelwert μ beträgt nicht 1600h.

Übung 3. *Anteilswerte*
Bei einem Experiment über aussersinnliche Wahrnehmung verlangt man von einer Person, die in einem Raum sitzt, die Farbe (rot oder blau) einer Karte anzugeben, die von einer zweiten Person in einem anderen Raum aus einem Stapel gut gemischten Karten gezogen wurde. Der Testperson ist unbekannt, wie viele rote und blaue Karten in dem Spiel sind.

1. 32 Karten von 50 Karten werden richtig erkannt. Bestimmen Sie, ob dieses Ergebnis mit einer Irrtumswahrscheinlichkeit von 0.05 respektive 0.01 signifikant ist.

   Das Raten der Karten folgt einer Binomial-Verteilung mit $p_0 = 0.5$. Da $n \cdot p \cdot q = 50 \cdot 0.5 \cdot 0.5 = 12.5 > 9$ ist, können wir das Ergebnis als normalverteilt annehmen. Die Standardabweichung ist:

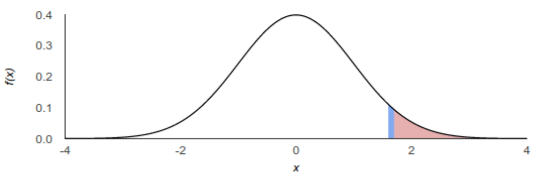
$$\sigma_{\bar{x}} = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.5 \cdot 0.5}{50}} = 0.071$$

   Wir stellen die folgenden Hypothesen auf:

   H_0: Das Ergebnis von $p = \frac{32}{50} = 0.64$ ist zufällig, d.h. der Hellseher hat geraten.
   H_1: Das Ergebnis von $p = \frac{32}{50} = 0.64$ ist überzufällig, d.h. der Hellseher ist hellsichtig.

   Aus der Tabelle ermitteln wir jeweils den z-Wert für ein einseitiges Konfidenzintervall:

$$z_{(0.95)} = 1.64$$
$$z_{(0.99)} = 2.32$$



   Wir können nun die oberen Konfidenzgrenzen berechnen:

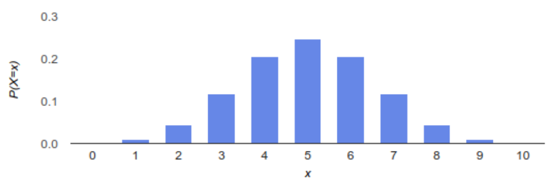
$$p_{o,95\%} = p + z_{(0.95)}\sigma_{\bar{x}} = 0.5 + 1.64 \cdot 0.071 = 0.62$$
$$p_{o,99\%} = p + z_{(0.99)}\sigma_{\bar{x}} = 0.5 + 2.32 \cdot 0.071 = 0.66$$

   Einerseits ist bei 5% Fehlerwahrscheinlichkeit $p > p_{o,95\%}$, wir nehmen die Hypothese an und verwerfen die Nullhypothese, d.h. das Ergebnis ist überzufällig, der Hellseher ist hellsichtig mit einer Fehlerwahrscheinlichkeit von 5%.

Andererseits ist bei 1% Fehlerwahrscheinlichkeit $p < p_{o,99\%}$, wir verwerfen in diesem Fall die Hypothese zugunsten der Nullhypothese, d.h. das Ergebnis ist zufällig, der Hellseher hat geraten.

2. 7 Karten von 10 Karten werden richtig erkannt. Bestimmen Sie, mit welcher Irrtumswahrscheinlichkeit dieses Ergebnis signifikant ist.

   Da nun $n \cdot p \cdot q = 10 \cdot 0.5 \cdot 0.5 = 2.5 < 9$ ist, müssen wir mit der Binomialverteilung rechnen. Aus der Tabelle ermitteln wir den Wert $P(X \leq 7) = 0.9453$, damit ist die Irrtumswahrscheinlichkeit $P(X > 7) = 1 - P(X \leq 7) = 0.055$.



Übung 4. *Differenztests mit abhängigen Stichproben*
Ein Marktforschungsinstitut untersucht, ob sich die die Fernsehgewohnheiten infolge der Corona-Pandemie verändert und hat folgende Resultate ermittelt:

Proband	Vorher	Nachher
1	1.1	1.2
2	2.3	1.9
3	4.4	4.5
4	2.2	3.1
5	0.9	9.8
6	1.2	1.2
7	4.0	2.9
8	2.3	4.4
9	7.7	2.8
10	5.5	5.2

Überprüfen Sie, ob sich die beiden Stichproben signifikant ($\alpha = 0.05$) unterscheiden!

Wir machen einen zweiseitigen Test für abhängige Stichproben und berechnen das maximale Intervall. Da $n < 30$ ist, nehmen wir die t-Verteilung mit $\nu = n - 1 = 9$.

i	x_v	x_n	d	$(d-\bar{d})^2$
1	1.1	1.2	-0.1	0.19
2	2.3	1.9	0.4	0.88
3	4.4	4.5	-0.1	0.19
4	2.2	3.1	-0.9	0.13
5	0.9	9.8	-8.9	69.89
6	1.2	1.2	0	0.29
7	4	2.9	1.1	2.69
8	2.3	4.4	-2.1	2.43
9	7.7	2.8	4.9	29.59
10	5.5	5.2	0.3	0.71
Σ	31.6	37	-5.4	107.0

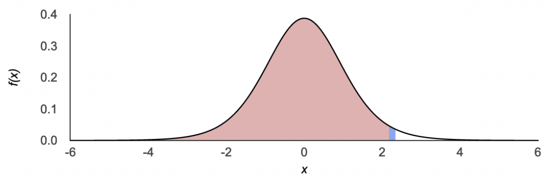
$$\bar{d} = \frac{1}{n}\sum d_i = \frac{-5.4}{10} = -0.54$$

$$s^2 = \frac{1}{n-1}\sum (d-\bar{d})^2 = \frac{107.0}{9} = 11.89$$

$$s = \sqrt{s} = \sqrt{11.89} = 3.45$$

$$t_{9,0.975} = 2.23$$

$$c = \bar{d} \pm t_{9,0.975}\frac{s}{\sqrt{n}} = -0.54 \pm 2.23\frac{3.45}{\sqrt{10}} = [-2.97, 1.89]$$



Die die Differenz $\bar{d}$ innerhalb des Intervalls liegt, können die Unterschiede auch durch Stichprobenfehler erklärt werden. Die Fernsehgewohnheiten haben sich nicht signifikant geändert.

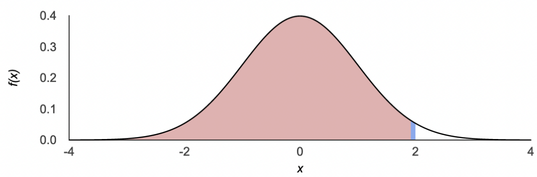
Übung 5. *Differenztests mit unabhängigen Stichproben*
Ein Marktforschungsinstitut untersucht, ob sich die West- und Deutschschweizer in ihren Fernsehgewohnheiten unterscheiden und hat folgende Resultate ermittelt:

- 800 Westschweizer sehen durchschnittlich 2 h fern bei einer Standardabweichung von 1 h.
- 600 Deutschschweizer sehen durchschnittlich 1.5 h fern bei einer Standardabweichung von 0.5 h.

Gibt es bei den Fernsehgewohnheiten signifikante Unterschiede bei einem Signifikanzniveau von 5%?

Wir machen einen zweiseitigen Test für unabhängige Stichproben und berechnen das maximale Intervall. Da $n > 30$ ist, können wir auf die Normalverteilung zurückgreifen.

$$z_{(0.975)} = 1.96$$
$$c = \pm z_{(0.975)} \sqrt{\frac{\sigma_w^2}{n_w} + \frac{\sigma_d^2}{n_d}}$$
$$= \pm 1.96 \sqrt{\frac{2^2}{800} + \frac{0.5^2}{600}} = \pm 0.07$$



Da die Mittelwertsdifferenz $\bar{d} = \bar{x}_w - \bar{x}_d = 2 - 1.5 = 0.5$ ausserhalb des Intervalls liegt können die Unterschiede nicht durch Stichprobenfehler erklärt werden. Die Fernsehgewohnheiten unterscheiden sich damit signifikant.

Übung 6. *Verteilungstest*
Sie haben Verkaufszahlen von vier Produkten in zwei verschiedenen Warenhäusern:

Produkt	Warenhaus 1	Warenhaus 2
A	560	70
B	680	120
C	640	110
D	700	100

Folgen die Verkaufszahlen des zweiten Warenhauses der Verteilung des ersten Warenhauses (bei einer Irrtumswahrscheinlichkeit von 0.05)?

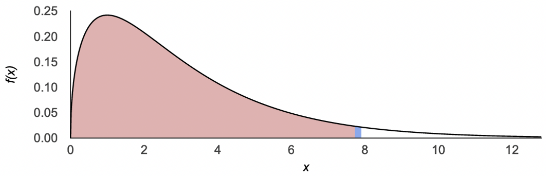
Wir machen einen χ^2-Test, um die empirische und die theoretische Häufigkeit zu vergleichen und stellen dazu die Hypothesen auf:

H_0: Die empirische Häufigkeit entspricht der theoretischen Häufigkeit.
H_1: Die empirische Häufigkeit entspricht nicht der theoretischen Häufigkeit.

Um die beiden Häufigkeiten vergleichen zu können, müssen beide Häufigkeiten in der Summe gleich sein. Wir skalieren deshalb die erste Häufigkeit und erhalten damit die theoretische Häufigkeit:

Produkt	h	h_t
A	560	$560 \cdot 400/2580 = 87$
B	680	$680 \cdot 400/2580 = 105$
C	640	$640 \cdot 400/2580 = 99$
D	700	$700 \cdot 400/2580 = 109$
	2580	400

Aus der Anzahl der Produkte folgt $\nu = 4 - 1 = 3$ Freiheitsgrade. Aus der Tabelle ermitteln wir $\chi^2_{(0.95,3)} = 7.81$.



Der Testwert berechnet sich mit:

$$\begin{aligned}\chi^2 &= \sum \frac{h_{b,i} - h_{t,i}}{h_{t_i}} \\ &= \frac{(70-87)^2}{87} + \frac{(120-105)^2}{105} + \frac{(110-99)^2}{99} + \frac{(100-109)^2}{109} = 7.43\end{aligned}$$

Da nun $\chi^2 < \chi^2_{(0.95,5)}$, müssen wir die Hypothese zugunsten der Nullhypothese verwefen. D.h. die Verteilungen unterscheiden sich nicht signifikant.

Zusatzaufgaben

Übung 7. *Anteilswerte*
Man stelle eine Entscheidungsregel zum Testen der Hypothese auf, dass eine Münze echt ist (gleich oft Kopf wie Zahl), wenn man eine Stichprobe von 46 Würfen der Münze nimmt und eine Irrtumswahrscheinlichkeit von 0.05 resp. 0.01 verwendet wird.

Der Münzwurf folgt einer Binomial-Verteilung mit $p_0 = 0.5$. Da $n \cdot p \cdot q = 46 \cdot 0.5 \cdot 0.5 = 11.5 > 9$ ist, können wir das Ergebnis als normalverteilt annehmen. Die Standardabweichung ist:

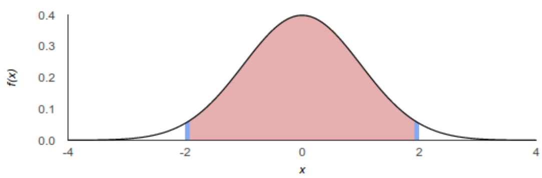
$$\sigma = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.5 \cdot 0.5}{46}} = 0.074$$

Wir stellen die folgenden Hypothesen auf:

H_0: Kopf und Zahl werden unterschiedlich oft gewürfelt, d.h. die Münze ist nicht echt.
H_1: Kopf und Zahl werden gleich oft gewürfelt, d.h. die Münze ist echt.

Aus der Tabelle ermitteln wir jeweils den z-Wert für ein beidseitiges Konfidenzintervall:

$$z_{(0.975)} = 1.96$$
$$z_{(0.995)} = 2.58$$



Wir können nun die Konfidenzgrenzen berechnen:

$$p_{o,95\%} = p \pm z_{(0.975)}\sigma = 0.5 \pm 1.96 \cdot 0.074 = [0.3550.645]$$
$$p_{o,99\%} = p \pm z_{(0.995)}\sigma = 0.5 \pm 2.58 \cdot 0.074 = [0.3090.690]$$

Bei einem zweiseitigen Fehler ist die Münze ëcht", wenn das Laplace-Experiment mit einer Wahrscheinlichkeit im Konfidenzintervall zu liegen kommt, d.h. bei 46 Würfen muss die Anzahl Kopf (oder Zahl) innerhalb des Intervalls $k_{95\%} = [17, 29]$ resp. $k_{99\%} = [15, 31]$ liegen.

Übung 8. *Verteilungstest*
Die nachfolgende Tabelle zeigt die beobachteten Häufigkeiten bei 120 Würfelwürfen. Bei einer Irrtumswahrscheinlichkeit von 0.05 teste man die Hypothese, dass der Würfel echt ist.

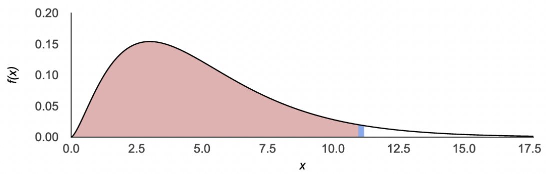
Augenzahl	i	1	2	3	4	5	6
Beobachtete Häufigkeit	h_b	25	17	15	23	24	16

Der Würfel ist dann echt, wenn jede Augenzahl gleich oft vorkommt, es muss also eine Gleichverteilung zwischen allen möglichen Ergebnissen vorliegen ($h_{t,i} = 20$). Wir haben hier also die Aufgabe, zwei Verteilungen zu überprüfen, ob diese gleich sind.

Wir machen einen χ^2-Test, um die empirische und die theoretische Häufigkeit zu vergleichen und stellen dazu die Hypothesen auf:

H_0: Die empirische Häufigkeit entspricht der theoretischen Häufigkeit.
H_1: Die empirische Häufigkeit entspricht nicht der theoretischen Häufigkeit.

Aus der Anzahl der Kategorien oder Klassen (Augenzahl 1, 2, 3, 4, 5, 6) folgt $\nu = 6 - 1 = 5$ Freiheitsgrade. Aus der Tabelle ermitteln wir $\chi^2_{(0.95,5)} = 11.1$.



Der Testwert berechnet sich mit:

$$\begin{aligned}\chi^2 &= \sum \frac{h_{b,i} - h_{t,i}}{h_{t_i}} \\ &= \frac{1}{20}\left(5^2 + 3^2 + 5^2 + 3^2 + 4^2 + 4^2\right) = 5\end{aligned}$$

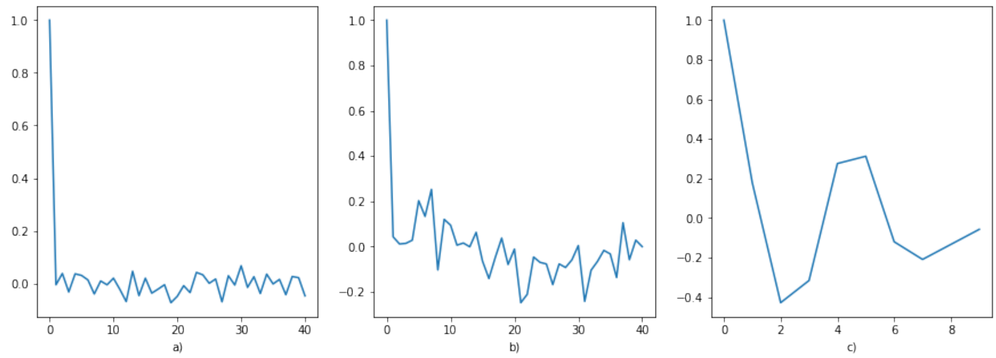
Da nun $\chi^2 < \chi^2_{(0.95,5)}$, müssen wir die Hypothese zugunsten der Nullhypothese verwefen. Wir nehmen also an, der Würfel ist echt.

Übung 9. *Autokorrelation*

Autokorrelation ist gegeben, wenn Beobachtungen in einer Zeitreihe x_n nicht unabhängig voneinander sind, d.h. wenn ein Teil einer Zeitreihe mit sich selbst zu einem anderen Zeitpunkt korreliert. Die Autokorrelation wird für ein gegebene Verschiebung $m = 0, 1, ...M$ folgendermassen berechnet:

$$\rho_m = \frac{1}{N-M} \sum_{n=1}^{N-M} (x_n - \bar{x})(x_{n+m} - \bar{x})$$

1. Welche der folgenden (normierten) Autokorrelationsfunktionen würden Sie am ehsten für eine Folge zufällig erzeugter Werte erwarten?



Für eine Folge zufällig generierte Zahlen erwarten wir, dass die Autokorrelation ausser bei $m = 0$ möglichst gegen Null geht: a).

2. $x_n = (2, 5, 6, 1, 4, 3)$

$$\begin{aligned}
N &= 6 \\
\bar{x} &= 3.5 \\
x_n - \bar{x} &= (-1.5, 1.5, 2.5, -2.5, 0.5, -0.5) \\
\rho_0 &= \frac{1}{4}(-1.5 \cdot -1.5 + 1.5 \cdot 1.5 + 2.5 \cdot 2.5 + -2.5 \cdot -2.5) = 4.25 \\
\rho_1 &= \frac{1}{4}(-1.5 \cdot 1.5 + 1.5 \cdot 2.5 + 2.5 \cdot -2.5 + -2.5 \cdot 0.5) = -1.5 \\
\rho_2 &= \frac{1}{4}(-1.5 \cdot 2.5 + 1.5 \cdot -2.5 + 2.5 \cdot 0.5 + -2.5 \cdot -0.5) = -1.25
\end{aligned}$$

Übung 10. *Parametertest*

In der letzten Prüfungssession einer Hochschule fanden $n = 30$ Prüfungen statt, der Durchschnitt lag bei $\bar{x} = 4.3$ Notenpunkten, die Standardabweichung bei $s = 1.5$. Testen Sie mit einer Irrtumswahrscheinlichkeit von $\alpha = 5\%$ die Hypothese, dass die Ergebnisse bei Prüfungen im Mittel $\mu = 4.5$ Notenpunkte sind.

Wir habe also die folgenden Hypothesen:

H_0: Die Notendurchschnitt ist $\mu \neq 4.5$.
H_1: Die Notendurchschnitt ist $\mu = 4.5$.

Die Standardabweichung der Population schätzen wir aus der Stichprobenvarianz:

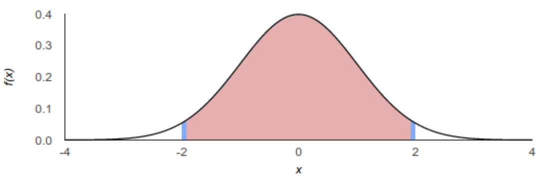
$$\sigma = \sqrt{\frac{n}{n-1}s^2} = \sqrt{\frac{30}{29}1.5^2} = 1.53$$

Mit $n = 30$ kann für die Verteilung der Mittelwerte entweder die t-Verteilung oder die Normalverteilung verwendet werden. Diese Verteilung hat eine Standardabweichung von:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{1.53}{\sqrt{30}} = 0.28$$

Aus der Tabelle ermitteln wir den z-Wert und das beidseitiges Konfidenzintervall:

$$\begin{aligned} z_{(0.975)} &= 1.96 \\ \bar{x}_{95\%} &= \mu \pm z_{(0.975)}\sigma_{\bar{x}} = 4.5 \pm 1.96 \cdot 0.28 = [3.95, 5.05] \end{aligned}$$



Der Durchschnitt der Prüfungssession $\bar{x}$ liegt mit einer Fehlerwahrscheinlichkeit von 5% im Konfidenzintervall, wir akzeptiere die Hypothese und verwerfen die Nullhypothese: Der Mittelwert beträgt $\mu = 4.5$.