Serie 7

Aufgabe 7.1

Im Juypter Notebook wurde das zweiseitige Vertrauensintervall hergeleitet. Wie sieht das einseitige Vertrauensintervall (nach oben und nach unten) aus? Gehen Sie wie im Juypter Notebook vor.

Aufgabe 7.2

Aufgrund langjähriger Untersuchungen ist bekannt, dass der Bleigehalt X von gewissen Bodenproben annähernd normal-verteilt ist

$$X \sim \mathcal{N}(\mu, \sigma^2)$$

a) Es wurden in 10 Bodenproben der Bleigehalt X gemessen. Dabei wurde ein Mittelwert von $\overline{x}_{10} = 31$ ppb erhalten. Die Standardabweichung sei bekannt und beträgt 6 ppb.

   Geben Sie das zweiseitige 99 % Vertrauensintervall für den Mittelwert an.

b) Wie viele Beobachtungen sind nötig, um die Breite des in Teilaufgabe a) bestimmten zweiseitigen Vertrauensintervalls auf die Hälfte zu reduzieren?

   Wie viele (unabhängige) Bestimmungen des Bleigehalts müssen geplant werden, falls der Bleigehalt mit einer Stichprobe „1 ppb genau" bestimmt werden soll, d. h., wenn die Breite des 99 % des Konfidenzintervalls nicht grösser als 1 ppb sein soll?

c) Normalerweise ist die Standardabweichung σ unbekannt. Um welchen Faktor verändert sich die Breite des zweiseitigen Vertrauensintervalls in Teilaufgabe a), wenn man die Standardabweichung aus den Daten geschätzt hat, also $\hat{\sigma} = 6$?

Aufgabe 7.3

Im National Bureau of Standards (USA) wurden regelmässig Wägungen des 10-Gramm - Standardgewichtstücks durchgeführt. Bei 9 Wägungen erhielt man als durchschnittliche Differenz -403 Mikrogramm vom 10 Gramm-Sollgewicht und eine Standardabweichung von 3.127 Mikrogramm für eine einzelne Wägung.

a) Geben Sie das exakte, zweiseitige 95 %-Vertrauensintervall für die wahre Differenz an, unter der Annahme, dass die Messfehler normal-verteilt sind.

b) Könnte die wahre Differenz $-400.0\,\mu g$ betragen? Entscheiden Sie aufgrund des Resultats in Aufgabe a). (Kurze Begründung)

Aufgabe 7.4

Eine Brücke soll aufgrund des höheren Verkehrsaufkommens renoviert werden. Im Bau wurden damals Schrauben mit einer mittleren Festigkeit von $500\,\mathrm{N/mm^2}$ benutzt. Da dies für nicht mehr sicher genug gehalten wird, sollen diese nun durch Schrauben mit einer mittleren Festigkeit von mehr als $500\,\mathrm{N/mm^2}$ ersetzt werden.

Um diesen Anforderungen gerecht zu werden, hat der alte Schraubenlieferant ein neues Verfahren entwickelt. Zur Baustelle werden allerdings unbeschriftete Schrauben geliefert, aus denen nicht sofort hervorgeht, ob es sich um die alten 500er oder um die neuen verbesserten Schrauben handelt.

Vor dem Verbau will der leitende Ingenieur zuerst sicherstellen, dass es sich um die besseren Schrauben handelt. Um dies herauszufinden, werden einige der Schrauben vermessen und ein statistischer Test durchgeführt. Je nach Ergebnis sollen die Schrauben verbaut oder zurückgeschickt werden.

	1	2	3	4	5
Schraubenfestigkeit (X_i)	520	512	499	524	505

Für den empirischen Mittelwert und empirische Varianz ergeben sich bei obiger Stichprobe $\overline{x}_5 = 512$ und $s_x^2 = 106.5$.

Wir modellieren die Daten mit einer Normalverteilung, d. h. X_i i.i.d. $\sim \mathcal{N}(\mu, \sigma^2)$.

a) Stellen Sie die geeigneten Null- und Alternativhypothesen auf und begründen Sie Ihre Wahl.

b) Sie führen nun einen einseitigen t-Test zum Niveau $\alpha = 0.05$ durch (unabhängig von Ihrer obigen Antwort). Stellen Sie die Teststatistik T auf und berechnen Sie deren Wert. Geben Sie die Verteilung der Teststatistik T unter H_0 und den Verwerfungsbereich des Tests an. Was ist der Testentscheid?

c) Berechnen Sie ein (zweiseitiges) 95 %-Vertrauensintervall für μ.

d) Wie würde das entsprechende Vertrauensintervall von c) aussehen, wenn wir die Streuung als bekannt voraussetzen würden (mit dem gleichen Wert wie der beobachtete)?

e) Betrachten Sie (unabhängig von dem oben aufgeführten Beispiel) einen einseitigen t-Test von $H_0 : \mu = 0$ gegen $H_A : \mu > 0$ zum Niveau 0.05.

Obwohl die beobachteten n Datenpunkte einen empirischen Mittelwert grösser Null haben, ergeben die Berechnungen, dass die Nullhypothese nicht verworfen wird.

Entscheiden Sie, ob folgende Aussagen wahr oder falsch sind:

i. Man verwirft H_0 für kein Niveau $\alpha < 0.05$.

ii. Es gibt ein Niveau $\alpha < 1$, bei dem man H_0 verwirft.

iii. Der p-Wert ist strikt kleiner als 0.5.

iv. Führt man statt eines einseitigen einen zweiseitigen Test zum Niveau 0.05 durch, verwirft man H_0 nicht.

v. Wenn man die Daten immer öfter kopiert (d. h., man betrachtet jeden Datenpunkt k-Mal, so dass man insgesamt $k \cdot n$ Datenpunkte erhält), verwirft man H_0 für ein grosses k beim Niveau 0.05.

Kurzlösungen vereinzelter Aufgaben

A 7.2:

a) $[26.1, 35.9]$

b) 959

c) $[24.8, 37.2]$

A 7.3:

a) $[-405.4, -400.6]$

Musterlösungen zu Serie 7

Lösung 7.2

a) Wir bezeichnen mit X_i den Bleigehalt in der i-ten Bodenprobe. Es gilt

$$X_i \sim \mathcal{N}(\mu, \sigma^2)$$

μ ist der wahre Mittelwert der Verteilung, wobei wir diesen Wert in der Praxis natürlich nicht kennen. $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ ist auch normalverteilt, allerdings mit Standardabweichung $\sigma/\sqrt{n}$, dem sogenannten *Standardfehler*

$$\overline{X}_n \sim \mathcal{N}(\mu, \sigma^2/n)$$

Nun betrachten wir in dieser Aufgabe $\overline{X}_{10}$, den Mittelwert von 10 Stichproben. $\overline{X}_{10}$ ist also normalverteilt mit Varianz $\sigma^2/n = \frac{36}{10} = 3.6$ und μ unbekannt. Der Fall, dass σ bekannt ist, μ aber nicht, ist natürlich nicht realistisch. Wir werden in einer späteren Teilaufgabe den Fall sehen, wo sowohl σ wie μ unbekannt sind.

Mit **Python**: (zu **R**)

```
from scipy.stats import norm, t
import numpy as np
norm.interval(alpha=0.99, loc=31, scale=6/np.sqrt(10))

## (26.112707522188142, 35.887292477811854)
```

$$[26.1, 35.9]$$

Zu 99 % liegt der wahre Mittelwert μ in diesem Bereich. Das Intervall hat eine Breite von 10.5, was relative breit ist. Dies hat damit zu tun, dass die Anzahl Messungen klein ist.

Je mehr Messungen wir haben, umso kleiner wird das Vertrauensintervall, umso sicherer sind wir, wo sich der Mittelwert mit einer (in diesem Fall) 99 %-er Sicherheit ist.

b) Die Breite des Vertrauensintervalles ist:

```
from scipy.stats import norm
import numpy as np

p_10 = norm.ppf(q=[0.005, 0.995], loc=31, scale=6/np.sqrt(10))

p_10[1] - p_10[0]
```

```
## 9.774584955623713
```

Wir versuchen nun das Vertrauensintervall zu halbieren, in dem wird die Anzahl Beobachtungen erhöhen.

```
n = 20
p_n = norm.ppf(q=[0.005, 0.995], loc=31, scale=6/np.sqrt(n))

p_n[1] - p_n[0]
```

```
## 6.911675305405538
```

Wir können nun n solange erhöhen, bis wir die Hälfte erreicht haben. Dies ist natürlich nicht sehr sinnvoll von Hand zu machen und machen dies mit einer Schlaufe.

```
n = 0

while True:
    n = n+1
    p_n = norm.ppf(q=[0.005, 0.995], loc=31, scale=6/np.sqrt(n))
    if (p_n[1] - p_n[0]) <= (p_10[1] - p_10[0])/2:
        break

n
```

```
## 40
```

Wir sehen, dass wir die Anzahl Versuche vervierfachen müssen, um das Vertrauensintervall zu halbieren. Dies wird auch das $\sqrt{n}$-Gesetz genannt.

Wir müssen aus b) einfach die Abbruchbedingung ändern:

```
n = 0

while True:
    n = n+1
    p_n = norm.ppf(q=[0.005, 0.995], loc=31, scale=6/np.sqrt(n))
    if (p_n[1] - p_n[0]) < 1:
        break

n
```

```
## 956
```

Wir brauchen also $n = 956$ Messungen, damit das 95 Vertrauensintervall eine Breite von 1 ppm haben, also relativ sicher sind, wo sich der wahre Mittelwert befindet.

c) Da σ_x unbekannt, verwenden wir eine t-Verteilung mit 9 Freiheitsgraden.

Mit **Python**: (zu **R**)

```
from scipy.stats import norm, t
import numpy as np
t.interval(alpha=0.99, df=9, loc=31, scale=6/np.sqrt(10))

## (24.833870595963425, 37.166129404036575)
```

Für $\widehat{\sigma} = 6$, $n = 10$ und $\overline{x}_{10} = 31$ ergibt sich das Vertrauensintervall

$$[24.8, 37.2]$$

Lösung 7.3

a) $\left[-403 \pm t_{9-1;97.5\,\%} \cdot \frac{3.127}{\sqrt{9}}\right] = [-403 \pm 2.31 \cdot 1.042] = [-405.4, -400.6]$

Mit **Python**: (zu **R**)

```
from scipy.stats import norm, t
import numpy as np

t.interval(alpha=0.95, df=8, loc=-403, scale=3.127/np.sqrt(9))

## (-405.40362497674977, -400.59637502325023)

print(t.interval(alpha=0.95, df=8, loc=-403, scale=3.127/np.sqrt(9)))

## (-405.40362497674977, -400.59637502325023)
```

b) Da -400.0 nicht im 95 %-Vertrauensintervall liegt, würde die Nullhypothese $H_0 : \mu = -400.0$ zu Gunsten der Alternative $H_A : \mu \neq -400.0$ auf dem 5 %-Signifikanzniveau verworfen werden.

Die Beobachtungen und die Hypothese $H_0 : \mu = -400.0$ passen also nicht gut zusammen und daher ist die wahre Differenz wohl nicht -400.0.

Lösung 7.4

a) Wir wählen

$$H_0 : \quad \mu = 500$$

als Nullhypothese und

$$H_A : \quad \mu > 500$$

als Alternativhypothese.

Man möchte auf jeden Fall den Fehler „Es handelt sich um 500er Schrauben, doch wir denken, es seien die neuen verbesserten Schrauben." vermeiden, da dies im schlimmsten Fall zum Brückeneinsturz führen kann. Daher wird dies der Fehler 1. Art.

Die Geschichte schliesst $\mu < 500$ als Teil der Alternative aus. Die Alternative $H_A : \mu > 500$ führt zu einer grösseren Macht als $H_A : \mu \neq 500$ (eines der beiden „Argumente" reicht).

b) Mit **Python** (zu **R**)

```
import numpy as np
from scipy.stats import norm, t
from pandas import Series

x = Series([520, 512, 499, 524, 505])

x.mean()

## 512.0

x.var()

## 106.5

t.ppf(q=0.95, df=x.size-1, loc=500, scale=x.std()/np.sqrt(x.size))


## 509.8388828578604

print(x.mean())

## 512.0

print(x.var())

## 106.5

print(t.ppf(q=0.95, df=x.size-1, loc=500, scale=x.std()/np.sqrt(x.size)))

## 509.8388828578604
```

Der Verwerfungsbereich ist also

$$I = [509.8, \infty)$$

Der Wert 512 für den Mittelwert liegt im Verwerfungsbereich, somit wird die Nullhypothese verworfen. Die Schrauben sind also statistisch signifikant stärker.

Berechnung von Hand:

Die Teststatistik berechnet sich als

$$T = \sqrt{5}\frac{\overline{X}_n - 500}{\hat{\sigma}}$$

wobei $\overline{x}_5 = \frac{1}{5}\sum_{i=1}^{5} x_i = 512$ der empirische Mittelwert und

$$\hat{\sigma}^2 = \frac{1}{4}\sum_{i=1}^{5}(x_i - \bar{x}_5)^2 = 106.5$$

die empirische Varianz ist.

Aus unserem Datensatz ergibt sich $t = \frac{512-500}{\sqrt{106.5/5}} \approx 2.6$ T ist t-verteilt mit vier Freiheitsgraden. Der Verwerfungsbereich auf dem 5 % Niveau ist gegeben durch (zu **R**)

$$K = [2.132, \infty)$$

```
from scipy.stats import t

t.ppf(q=0.95, df=4)
## 2.13184678133629
print(t.ppf(q=0.95, df=4))
## 2.13184678133629
```

Die Nullhypothese wird demnach abgelehnt.

c) Das zweiseitige Vertrauensintervall ist gegeben durch (zu **R**)

$$\text{VI} := \left[\overline{x}_5 - \frac{\hat{\sigma}t_{n-1;1-\alpha/2}}{\sqrt{5}}; \overline{x}_5 + \frac{\hat{\sigma}t_{n-1;1-\alpha/2}}{\sqrt{5}}\right]$$

```
import numpy as np
from scipy.stats import norm, t
from pandas import Series

x = Series([520, 512, 499, 524, 505])

t.ppf(q=[0.025,0.975], df=x.size-1, loc=x.mean(),
scale=x.std()/np.sqrt(x.size))
## array([499.18617192, 524.81382808])
```

```
print(t.ppf(q=[0.025,0.975], df=x.size-1, loc=x.mean(), scale=x.std()/np.sq
)

## [499.18617192 524.81382808]
```

Für obige Daten ergibt sich $VI \approx [499.1882; 524.8118]$.

d) Setzt man die Streuung σ als bekannt voraus, so ist das zweiseitige Vertrauensintervall gegeben durch

$$\mathrm{VI} := \left[\hat{\mu} - \frac{\sigma z_{1-\alpha/2}}{\sqrt{5}}; \hat{\mu} + \frac{\sigma z_{1-\alpha/2}}{\sqrt{5}}\right]$$

wobei $z_{1-\alpha/2}$ das $1-\alpha$-Quantil einer Standard-Normalverteilung ist. (zu **R**)

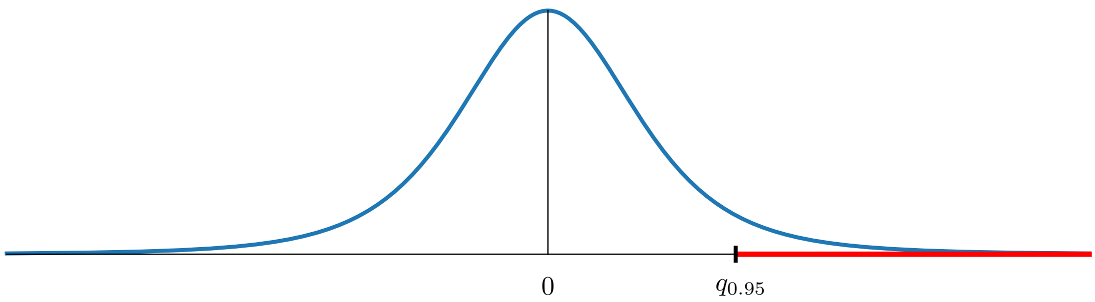
```
import numpy as np
from scipy.stats import norm, t
from pandas import Series

x = Series([520, 512, 499, 524, 505])

norm.ppf(q=[0.025,0.975], loc=x.mean(), scale=x.std()/np.sqrt(x.size))


## array([502.9543893, 521.0456107])

print(norm.ppf(q=[0.025,0.975], loc=x.mean(), scale=x.std()/np.sqrt(x.size)

## [502.9543893 521.0456107]
```

Für obige Daten ergibt sich dann $\mathrm{VI} \approx [502.9544; 521.0456]$.

e) Diese Aufgabe ist nicht ganz einfach und man muss den Text sehr genau lesen:

Betrachten Sie einen einseitigen t-Test von $H_0 : \mu = 0$ gegen $H_A : \mu > 0$ zum Niveau 0.05.

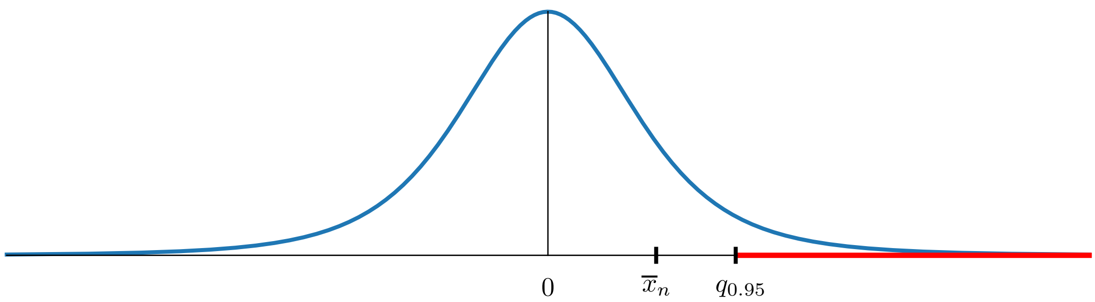
Wir zeichnen eine t-Verteilung mit $\mu = 0$ und $\sigma = 1$ und Freiheitsgrad (degree of freedom) 4 (Annahmen, die nichts zur Sache tun) und mit dem Verwerfungsbereich (Test nach oben).



Obwohl die beobachteten n Datenpunkte einen empirischen Mittelwert grösser Null haben, ergeben die Berechnungen, dass die Nullhypothese nicht verworfen wird.

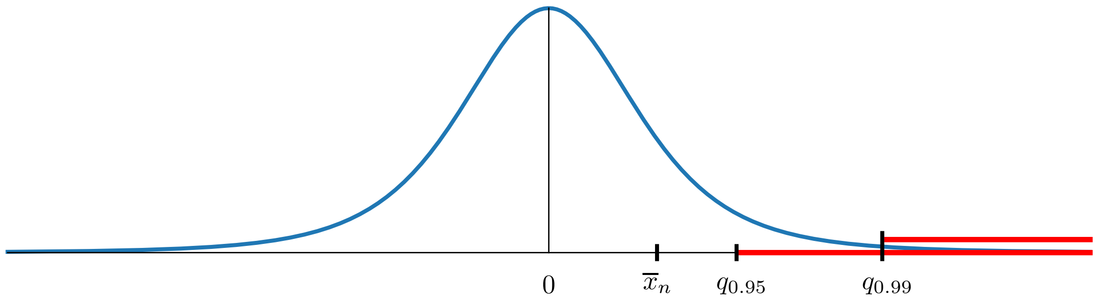
Was heisst das?

- Zunächst ist der Mittelwert $\overline{x}_n > 0$, er liegt rechts von 0 auf der x-Achse.
- H_0 wird nicht verworfen, dass heisst $\overline{x}_n$ liegt nicht im Verwerfungsbereich.
- $\overline{x}_n$ liegt also irgendwo zwischen dem 0 und dem Verwerfungsbereich.



i. *Man verwirft H_0 für kein Niveau $\alpha < 0.05$.*

   Was heisst das? Unser Signifikanzniveau ist $\alpha = 0.05$. Nun wählen wir ein Signifkanzniveau α^*, dass kleiner ist als 0.05, zum Beispiel $\alpha^* = 0.01$. Der Verwerfungsbereich wird kleiner wandert nach rechts. Das heisst $\overline{x}_n$ liegt immer noch nicht im Verwerfungsbereich (siehe Graphik unten).



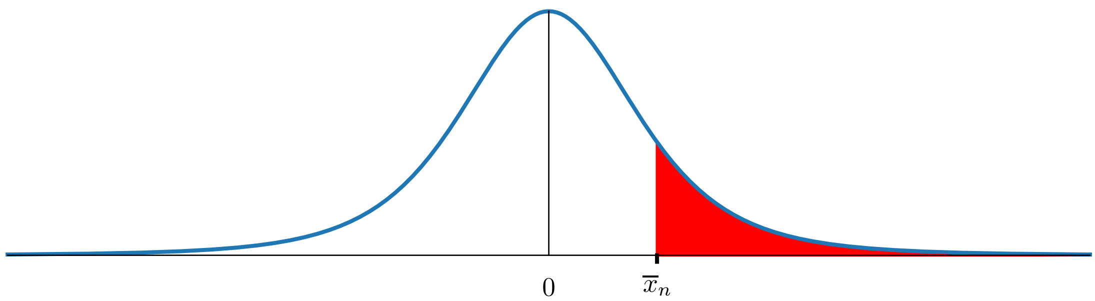
Also wird für alle $\alpha < 0.05$ die Nullhypothese nicht verworfen, da $\overline{x}_n$ nie im Verwerfungsbereich liegen kann. Also ist die Aussage richtig.

ii. *Es gibt ein Niveau $\alpha < 1$, bei dem man H_0 verwirft.*

Dies ist die umgekehrte Fragestellung von vorher: $\alpha = 0.05$ und jetzt *vergrössern* wir α, dann beginnt der Verwerfungsbereich nach links zu wandern und ab irgendeinem α liegt $\overline{x}_n$ im Verwerfungsbereich und die Nullhypothese wird verworfen. Also ist die Aussage richtig.

iii. *Der p-Wert ist strikt kleiner als 0.5.*

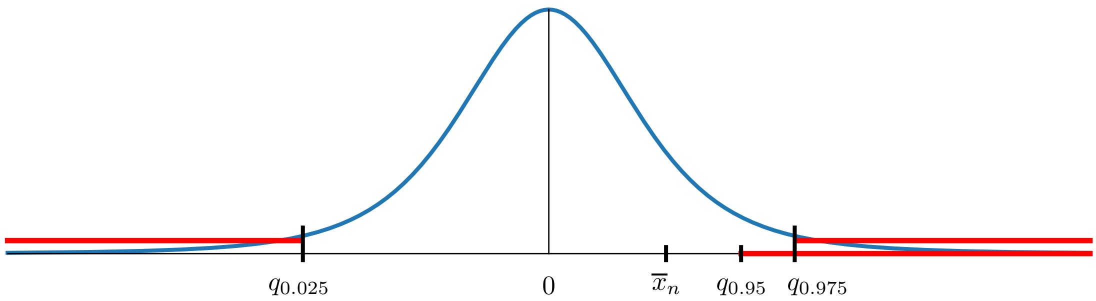
Nun was ist der pp-Wert? Das ist die Wahrscheinlichkeit einen bestimmten Wert (hier $\overline{x}_n$) oder einen extremeren zu beobachten in Richtung der Alternativhypothese. Wahrscheinlichkeiten können wir bei stetigen Verteilungen als Flächen darstellen und dies ist die Fläche unter der Kurve rechts von $\overline{x}_n$ (siehe Graphik unten).



Nun ist die Gesamtfläche unter der Kurve ist 1. Die Fläche rechts von 0 ist 0.5 und dann muss die rote Fläche (p-Wert) kleiner als 0.05 sein. Also ist die Aussage richtig.

iv. *Führt man statt eines einseitigen einen zweiseitigen Test zum Niveau 0.05 durch, verwirft man H_0 nicht.*

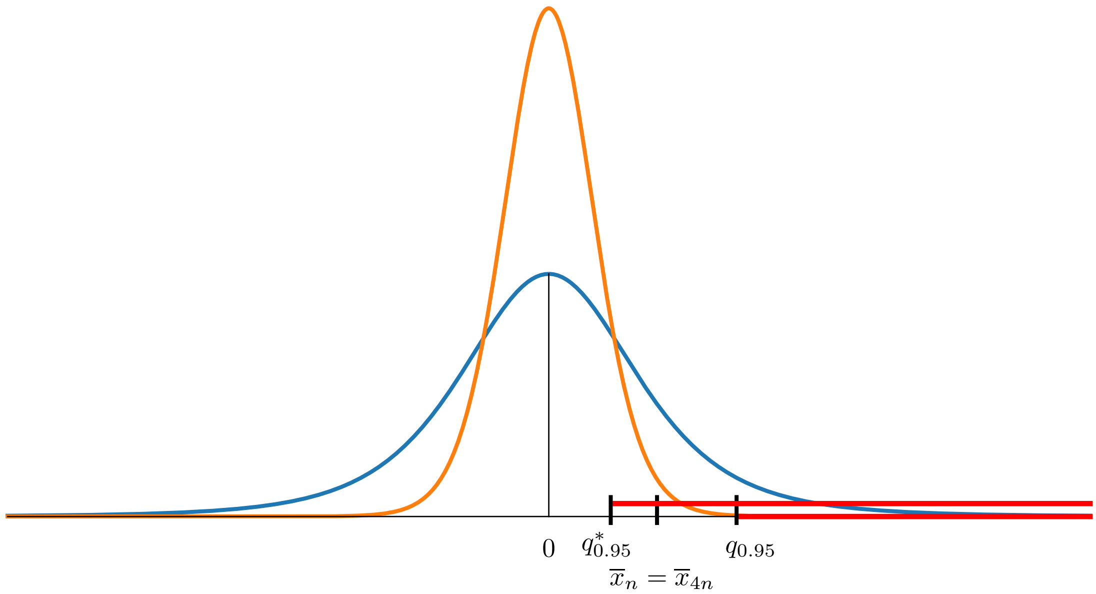
Ähnliche Frage wie weiter oben: Wechseln wir von einseitig zu zweiseitig, so verkleinert sich der Verwerfungsbereich auf der rechten Seite. Da $\overline{x}_n$ beim einseitigen Test schon nicht im Verwerfungsbereich liegt, dann kann er es beim zweiseitigen sowieso nicht. Also ist die Aussage richtig.



v. *Wenn man die Daten immer öfter kopiert (d. h., man betrachtet jeden Datenpunkt k-Mal, so dass man insgesamt $k \cdot n$ Datenpunkte erhält), verwirft man H_0 für ein grosses k beim Niveau* 0.05.

Wenn wir die Datenpunkte beispielsweise vervierfachen ($k = 4$), dann bleibt der Durchschnitt und die Standardabweichung nimmt für den Durchschnit aber ab mit $\sqrt{k} = 2$.

Das heisst, $\overline{x}_n = \overline{x}_{nk}$, aber die Anzahl Beobachtungen nimmt zu. Das heisst die Kurve wird schmaler und der Verwerfungsbereich wächst gegen links und für ein genügend grosse k liegt dann $\overline{x}_n$ im Verwerfungsbereich und die Nullhypothese wird verworfen. Also stimmt die Aussage.


0
$q^*_{0.95}$
$q_{0.95}$
$\overline{x}_n = \overline{x}_{4n}$

R-Code

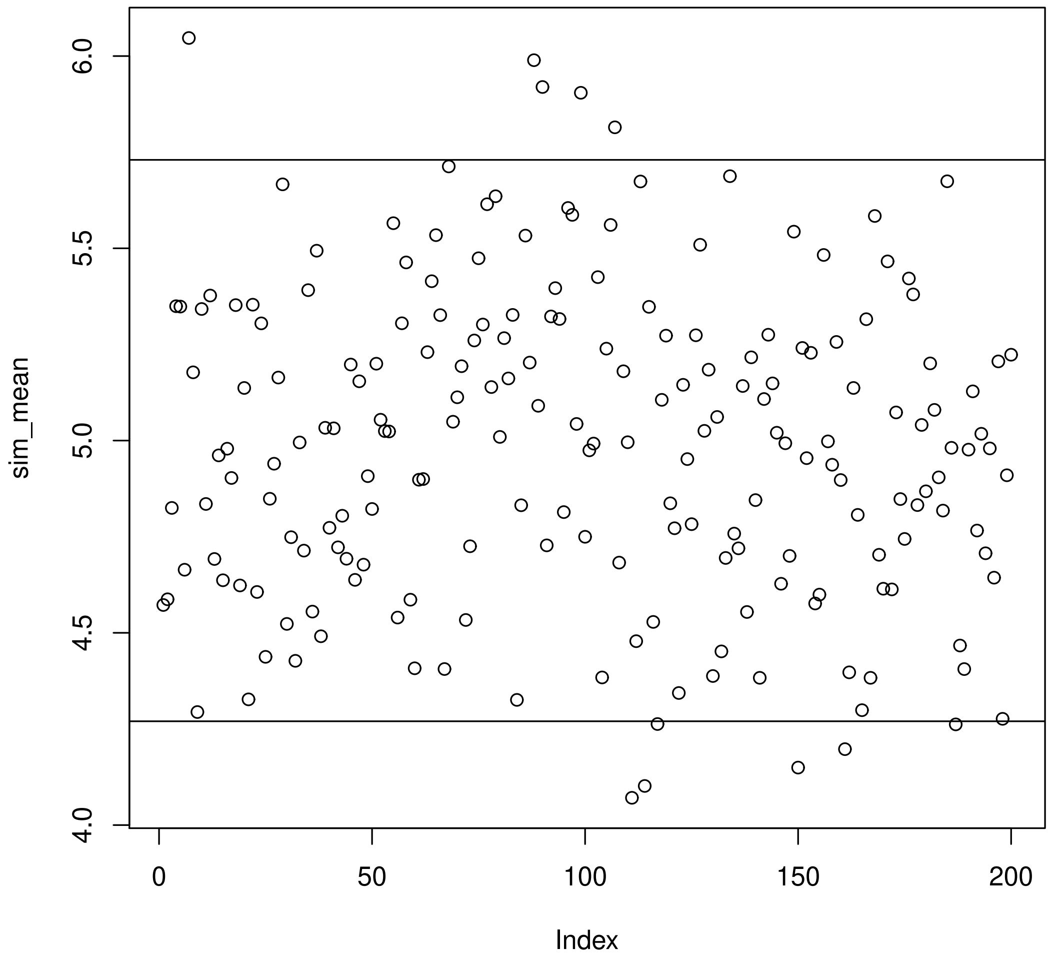
Aufgabe ??

a)

b)

c) (zu **Python**)

```
n <- 60
sim <- matrix(runif(n*200, min=0, max=10), ncol=n)
sim_mean <- apply(sim, 1, "mean")
plot(sim_mean)
abline(h=5.73)
abline(h=4.27)
```

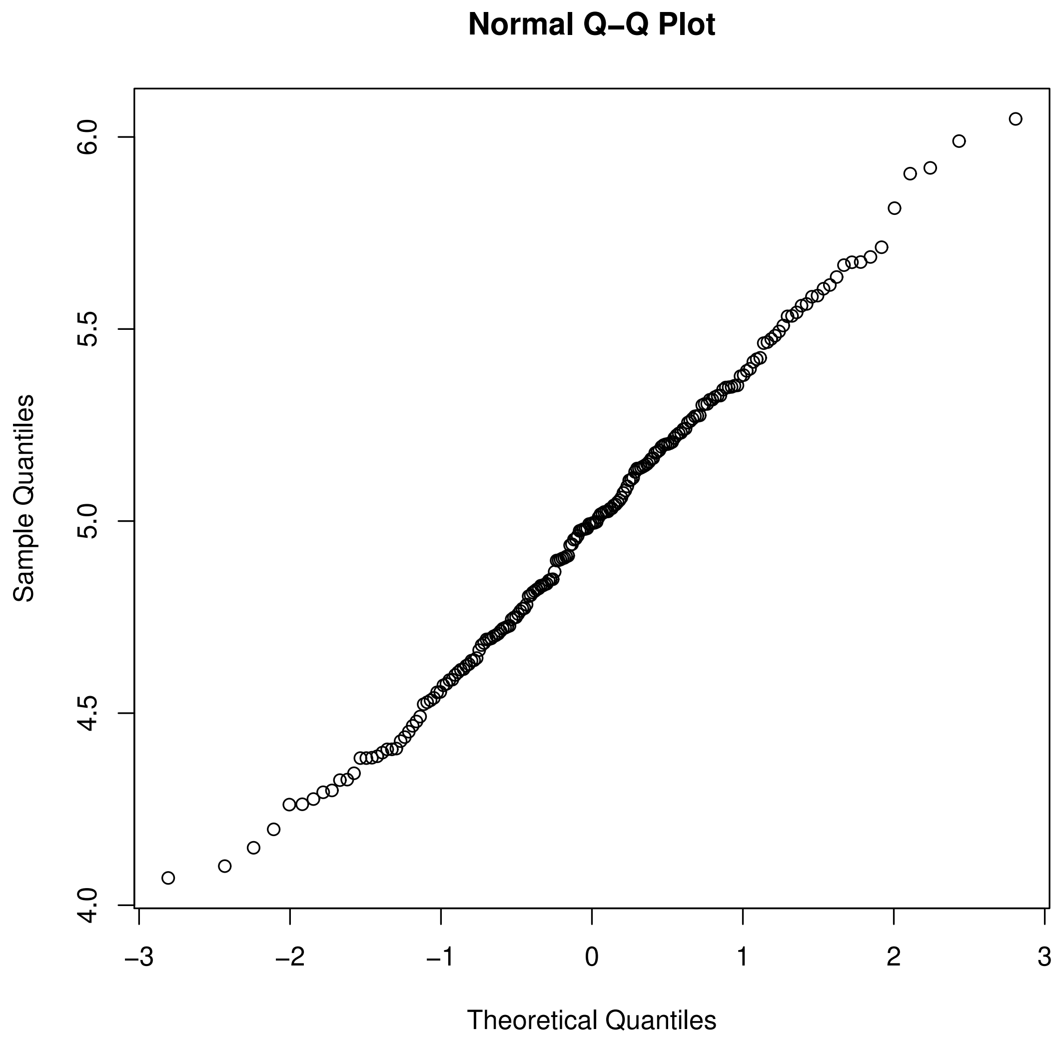



```
d <- sum(sim_mean>5.73) + sum(sim_mean<4.27)
```

In unserem Beispiel sind es 11.

Weiter bestätigt sich auch der Zentrale Grenzwertsatz: (zu **Python**)

```
qqnorm(sim_mean)
```



Aufgabe 7.2

a) (zu **Python**)

```
qnorm(c(0.005,0.995), 31, 6/sqrt(10))

## [1] 26.11271 35.88729
```

b)

c) (zu **Python**)

Diese Option gibt es bei **R** nicht.

(zu **Python**)

```r
qt(0.995, 9)

## [1] 3.249836
```

Aufgabe 7.3

a) (zu **Python**)

   Dieser Befehl existiert in **R** nicht.

Aufgabe 7.4

a)

b) (zu **Python**)

```r
x <- c(520, 512, 499, 524, 505)

mean(x)

## [1] 512

var(x)

## [1] 106.5

length(x)-1

## [1] 4

qt(0.95, df=length(x)-1)*(sd(x)/sqrt(length(x)))+500

## [1] 509.8389
```

(zu **Python**)

```r
qt(0.95, df=4)

## [1] 2.131847
```

c) (zu **Python**)

```r
x <- c(520, 512, 499, 524, 505)

qt(c(0.025,0.975), df=length(x)-1)*(sd(x)/sqrt(length(x)))+512

## [1] 499.1862 524.8138
```

d) (zu **Python**)

```
x <- c(520, 512, 499, 524, 505)

qnorm(c(0.025,0.975), mean=mean(x), sd=sd(x)/sqrt(length(x)))

## [1] 502.9544 521.0456
```