

## Serie 8

### Aufgabe 8.1

Mit folgender Simulationsaufgabe sollen Sie mit dem Normalplot vertraut werden.

- a) Simulieren Sie  $n = 10, 20, 50$  und  $100$  (standard-) normalverteilte Zufallszahlen und betrachten Sie die jeweilige Folge mit einem Normalplot.

Wiederholen Sie diese Simulationen einige Male, bis Sie abschätzen können, wie weit zufällige Abweichungen von einer Geraden im Normalplot üblich sind.

*Python-Hinweise:*

```
import matplotlib.pyplot as plt
import numpy as np
import scipy.stats as st

x = st.norm.rvs(size=10)
st.probplot(x, plot=plt)
```

- b) *Langschwänzige Verteilung:* Simulieren Sie je  $n = 20$  und  $100$   $t$ -verteilte Zufallszahlen mit  $\nu = 20, 7$  und  $3$  Freiheitsgraden. Wiederholen Sie diese Simulationen einige Male, bis Sie abschätzen können, wie gross Abweichungen von einer Geraden im Normalplot üblich sind.

*Python-Hinweise:*

```
x = st.t.rvs(size=10, df=20)
```

- c) *Schiefe Verteilung:* Simulieren Sie je  $n = 20$  und  $100$  chiquadrat-verteilte Zufallszahlen mit  $\nu = 20$  und  $1$  Freiheitsgraden. Wiederholen Sie diese Simulationen einige Male, bis Sie abschätzen können, wie gross Abweichungen von einer Geraden im Normalplot üblich sind.

*Python-Hinweise:*

```
x = st.chi2.rvs(size=10, df=20)
```

### Aufgabe 8.2

Untenstehend finden Sie mehrere Beispiele für Vergleiche von 2 Stichproben. Beantworten Sie für jedes Beispiel *kurz* die folgenden Fragen:

- Handelt es sich um gepaarte oder um ungepaarte Stichproben? Begründen Sie!

- Ist der Test einseitig oder zweiseitig durchzuführen? Begründen Sie!
  - Wie lautet die Nullhypothese in Worten?
  - Wie lautet die Alternativhypothese in Worten?
- a) In einem Experiment sollte der Effekt von Zigarettenrauchen auf Blutplättchenanhäufungen untersucht werden. Dazu wurden 11 Probanden vor und nach dem Rauchen einer Zigarette Blutproben entnommen, und es wurde gemessen, wie stark sich die Blutplättchen anhäufeten. Es interessiert, ob sich Blutplättchen durch das Rauchen vermehrt anhäufen.
- b) Die nächsten Daten sind aus einer Studie von Charles Darwin über die Fremd- und Selbstbefruchtung. 15 Paare von Setzlingen mit demselben Alter, je einer durch Selbst- und einer durch Fremdbefruchtung produziert, wurden gezüchtet. Beide Teile je eines Paares hatten nahezu gleiche Bedingungen. Das Ziel bestand darin zu sehen, ob die fremdbefruchteten Pflanzen mehr Lebenskraft besitzen als die selbstbefruchteten (d.h., ob sie grösser werden). Es wurden die Höhen jeder Pflanze nach einer fixen Zeitspanne gemessen.
- c) Beeinflusst der Kalziumgehalt in der Nahrung den systolischen Blutdruck? Zur Überprüfung dieser Frage wurde einer Versuchsgruppe von 10 Männern während 12 Wochen ein Kalziumzusatz verabreicht. Einer Kontrollgruppe von 11 Männern gab man ein Placebopräparat.
- d) In einem Experiment wurde untersucht, ob Mäuse zwei Formen von Eisen ( $\text{Fe}^{2+}$  und  $\text{Fe}^{3+}$ ) unterschiedlich gut aufnehmen. Dazu wurden 36 Mäuse in zwei Gruppen zu je 18 unterteilt und die eine Gruppe mit  $\text{Fe}^{2+}$  und die andere mit  $\text{Fe}^{3+}$  „gefüttert“. Da das Eisen radioaktiv markiert war, konnte sowohl die Anfangskonzentration wie auch die Konzentration einige Zeit später gemessen werden. Daraus wurde für jede Maus der Anteil des aufgenommenen Eisens berechnet.

### Aufgabe 8.3

Zwei Tiefen-Messgeräte messen für die Tiefe einer Gesteins-Schicht an 9 verschiedenen Orten die folgenden Werte: Kennzahlen für die Differenz:  $\bar{d}_n$  beträgt  $-5.78$ , die

Messgerät A	120	265	157	187	219	288	156	205	163
Messgerät B	127	281	160	185	220	298	167	203	171
Differenz $d_i$	-7	-16	-3	2	-1	-10	-11	2	-8

Standardabweichung  $\sigma_D = 6.2$ .

Es wird vermutet, dass Gerät B systematisch grössere Werte misst. Bestätigen die Messwerte diese Vermutung oder ist eine zufällige Schwankung als Erklärung plausibel?

- Handelt es sich um verbundene (gepaarte) oder um unabhängige Stichproben?
- Führen Sie einen  $t$ -Test auf dem Niveau  $\alpha = 0.05$  durch. Formulieren Sie explizit: Modellannahmen, Nullhypothese, Alternative, Teststatistik, Verwerfungsbereich und Testergebnis.
- Sei  $Z$  die Zufallsvariable, die zählt, bei wie vielen der 9 Messungen Gerät A einen grösseren Wert misst, als Gerät B. Wie ist  $Z$  verteilt, wenn die Geräte bis auf Zufallsschwankungen das Gleiche messen?

### Aufgabe 8.4

Die Körpertemperatur von 10 Patienten wird zum Zeitpunkt der Verabreichung eines Medikaments ( $T_1$ ) und 2 Stunden später ( $T_2$ ) gemessen. Es soll geprüft werden, ob dieses Medikament eine fiebersenkende Wirkung hat.

Patient-Nr.	1	2	3	4	5	6	7	8	9	10
Temp. 1 in °C	39.1	39.3	38.9	40.6	39.5	38.4	38.6	39.0	38.6	39.2
Temp. 2 in °C	38.1	38.3	38.8	37.8	38.2	37.3	37.6	37.8	37.4	38.1

- Handelt es sich um einen gepaarten oder ungepaarten Test? Begründen Sie Ihre Antwort.
- Handelt es sich um einen ein- oder zweiseitigen Test? Begründen Sie Ihre Antwort.
- Formulieren Sie Null- und Alternativhypothese.
- Wir nehmen an, die Daten seien normalverteilt. Welchen Test wählen Sie? Führen Sie den Test mit **R** auf Signifikanzniveau 5 % durch.
- Wenn wir nicht davon ausgehen können, dass die Daten normalverteilt sind, welchen Test wählen Sie? Führen Sie diesen auf Signifikanzniveau 5 % durch.
- Erklären Sie den Unterschied der  $p$ -Werte in Teilaufgaben d) und e).

männlich $x_i$	120	107	110	116	114	111	113	117	114	112
weiblich $y_j$	110	111	107	108	110	105	107	106	111	111

### Aufgabe 8.5

In der folgenden Tabelle sind die Kieferlängen von 10 männlichen und 10 weiblichen Goldschakalen eingetragen: Einige Kennzahlen:  $\bar{x}_n = 113.4$ ,  $\bar{y}_n = 108.6$ ,  $\hat{\sigma}_x^2 = 13.82$ ,  $\hat{\sigma}_y^2 = 5.16$ . Es stellt sich nun die Frage, ob es einen Unterschied zwischen der Kieferlänge von Männchen und Weibchen gibt.

- Handelt es sich um gepaarte oder ungepaarte Stichproben? Begründen Sie Ihre Antwort.
- Formulieren Sie Nullhypothese und Alternativhypothese.
- Führen Sie den t-Test nun noch mit Hilfe von **Python** durch. Geben Sie den resultierenden  $p$ -Wert sowie den daraus folgenden Testentscheid an.

```
import pandas as pd
import scipy.stats as st

jackals = pd.read_table(r"*/jackals.txt", sep = " ")

jackals

st.ttest_ind(jackals["M"], jackals["W"], equal_var=False)
```

Für \* muss Ihr Pfad stehen, wo sich die Datei befindet.

- Führen Sie mit Hilfe von **Python** einen Wilcoxon-Test durch. Geben Sie wiederum  $p$ -Wert und Testentscheid an.

```
st.mannwhitneyu(jackals["M"], jackals["W"], alternative="two-sided")
```

- Falls die Resultate der beiden Tests unterschiedlich ausgefallen wären, welchem würden Sie eher vertrauen? Weshalb?

### Aufgabe 8.6

Im Jahr 2013 wurden im Rahmen einer internationalen Zusammenarbeit unter der Leitung der EAWAG in Dübendorf Konzentrationen von illegalen Substanzen im Abwasser von 42 europäischen Städten während einer Woche untersucht (Ort C. et al, *Spatial differences and temporal changes in illicit drug use in Europe quantified by wastewater analysis*, Addiction 2014 Aug).

Dabei wurden an 7 aufeinanderfolgenden Tagen (6.-12. März) neben anderen Substanzen die medianen Konzentrationen von Ecstasy (MDMA) im Abwasser gemessen. Aufgrund dieser Studie war eine Aussage einer vielgelesenen Schweizer Gratiszeitung, dass in Zürich viel mehr Drogen konsumiert werden als anderswo.

In der nachfolgenden Tabelle sind für die Städte Zürich und Basel die an den untersuchten Tagen ausgeschiedenen Mengen MDMA aufgeführt. Die Angaben sind in mg pro 1000 Einwohner pro Tag.

Wochentage	Mi	Do	Fr	Sa	So	Mo	Di
Zürich	16.3	12.7	14.0	53.3	117	62.6	27.6
Basel	10.4	8.91	11.7	29.9	46.3	25.0	29.4

- Schätzen Sie aus den Daten den Mittelwert und die Standardabweichung der Differenzen, d.h.,  $\hat{\mu}_D$  und  $\hat{\sigma}_D$ .
- Handelt es sich um gepaarte oder ungepaarte Stichproben? Begründen Sie Ihre Antwort.
- Formulieren Sie die Nullhypothese und die Alternativhypothese, wenn Sie die Aussage der besagten Gratiszeitung überprüfen möchten.
- Führen Sie einen statistischen Test mit Hilfe von **Python** auf dem Signifikanzniveau 5 % durch, unter der Annahme, dass die Daten normalverteilt sind.
- Führen Sie nun einen statistischen Test mit Hilfe von **Python** auf dem Signifikanzniveau 5 % durch, unter der Annahme, dass die Daten nicht normalverteilt sind.

### Aufgabe 8.7

Ein U.S. Magazin, Consumer Reports, führte eine Untersuchung des Kalorien- und Salzgehaltes von verschiedenen Hotdog-Marken durch. Es gab drei verschiedene Typen von Hotdogs: Rind, „Fleisch“ (Rind, Schwein, Geflügel gemischt) und Geflügel.

Die Resultate unten führen den Kaloriengehalt verschiedener Marken von Rind- und Geflügel-Hotdogs auf.

Rinds-Hotdog

186, 181, 176, 149, 184, 190, 158, 139, 175, 148, 152, 111, 141, 153, 190, 157, 131, 149, 135, 132

Geflügel-Hotdog:

129, 132, 102, 106, 94, 102, 87, 99, 170, 113, 135, 142, 86, 143, 152, 146, 144

Haben die beiden Hotdog-Arten verschiedenen Kaloriengehalt?

- Handelt es sich um einen gepaarten oder ungepaarten Test? Begründen Sie Ihre Antwort.
- Handelt es sich um einen ein- oder zweiseitigen Test? Begründen Sie Ihre Antwort.
- Formulieren Sie Null- und Alternativhypothese.
- Berechnen Sie die Mittelwerte der beiden Testreihen. Welche Vermutung haben Sie?
- Welchen Test würden Sie hier wählen:  $t$ -Test oder Wilcoxon-Test? Begründen Sie Ihre Antwort.
- Führen Sie den entsprechenden Test mit **R** durch. Interpretieren Sie  $p$ -Wert. Bei welcher Hotdog-Art ist der Kaloriengehalt grösser?

### Aufgabe 8.8

Wir haben aus eigener Erfahrung das Gefühl, dass bei Ehepaaren der Mann eher älter ist als die Frau. Nun wollen wir statistisch untersuchen, ob dem so ist.

Im Datensatz **mannfrau.csv** sind die Werte von Körpergröße und Alter bei Mann und Frau von 170 britischen Ehepaaren aufgeführt. Die Körpergröße ist in cm und das Alter in Jahren angegeben.

*Hinweis:*

```
import pandas as pd
mf <- pd.read_csv(r".../mannfrau.csv")
```

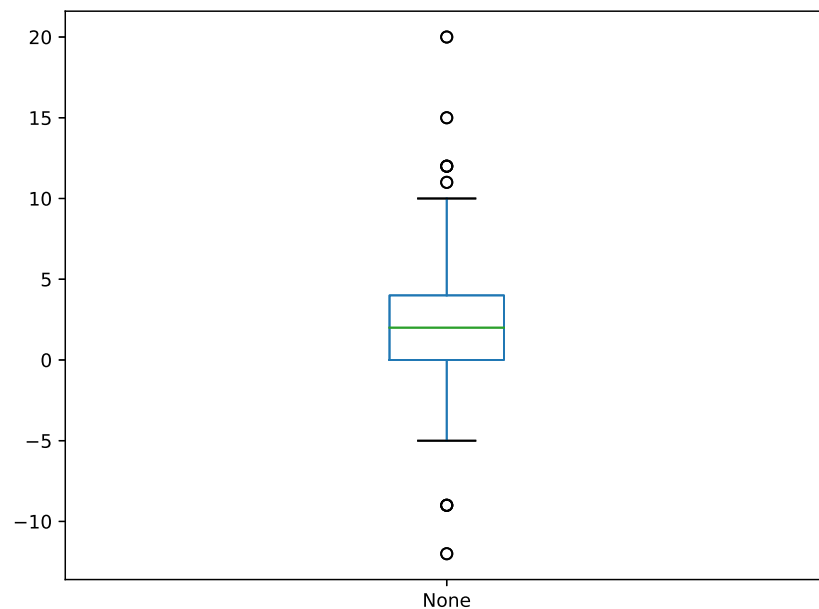
```
import pandas as pd
mf = pd.read_csv("../.../Themen/Deskriptive_Statistik/Uebungen_de/Daten/mannf")
mf.head()
```

##	alter.mann	groesse.mann	alter.frau	groesse.frau
## 0	49	180	43	159
## 1	25	184	28	156
## 2	40	165	30	162
## 3	52	177	57	154
## 4	58	161	52	142

Die ... stehen für den Pfad, wo die Datei abgespeichert wurde.

Wir zeichnen den Boxplot für die Altersdifferenz der Ehepaare.

```
diff = mf["alter.mann"] - mf["alter.frau"]
diff.plot(kind="box")
```



Bei rund 50 % der Ehepaare liegt der Altersunterschied zwischen 0 und etwa 5 Jahren. Bei rund 25 % der Ehepaaren ist die Ehefrau älter als der Ehemann. Unsere Vermutung scheint zu stimmen, aber die Frage ist dann noch, ob der Unterschied statistisch signifikant ist.

- a) Wir wollen unsere Vermutung, dass die Männer bei Ehepaaren eher älter sind als die Frauen, ebenfalls mit einem Hypothesentest untersuchen.
  - i) Wählen Sie einen gepaarten oder ungepaarten Test?
  - ii) Machen Sie einen ein- oder zweiseitigen Test?
  - iii) Wir gehen von Normalverteilung der Daten aus. Führen Sie diesen Hypothesentest auf Signifikanzniveau von 5 % durch.  
  
Formulieren Sie die Null- und Alternativhypothese und führen Sie den Test mit Testentscheid auf einem Signifikanzniveau von 5 % durch.  
  
Führen Sie den Testentscheid noch mit dem Vertrauensintervall durch.
  - iv) Wenn Sie nicht von Normalverteilung ausgehen, welchen Test wählen Sie? Führen Sie diesen Test durch.
- b) Wir untersuchen noch die Größenunterschiede zwischen Frauen und Männer. Im Allgemeinen sind Männer grösser als Frauen. In England sind gemäss Wikipedia die Männer durchschnittlich 13 cm grösser als die Frauen.
  - i) Welcher Test ist hier angebracht (ein- oder zweiseitig, gepaart oder ungepaart)?
  - ii) Formulieren Sie die Null- und Alternativhypothese.
  - iii) Wird die Aussage durch unseren Datensatz auf einem Signifikanzniveau von 5 % statistisch signifikant widerlegt? Führen Sie den Test mit Testentscheid durch. Wir nehmen an, dass die Körpergrößen normalverteilt sind.  
  
Führen Sie den Testentscheid noch mit dem Vertrauensintervall durch.



## Kurzlösungen einzelner Aufgaben

**A 8.3:** a) gepaart

b)  $K = (-\infty, -1.86]$

**A 8.5:**

a) ungepaart

c)  $p$ -Wert: 0.0034

b)  $H_0 : \mu_1 = \mu_2$   
 $H_A : \mu_1 \neq \mu_2$

d)  $p$ -Wert: 0.0048

# Musterlösungen zu Serie 8

## Lösung 8.1

a) (zu R)

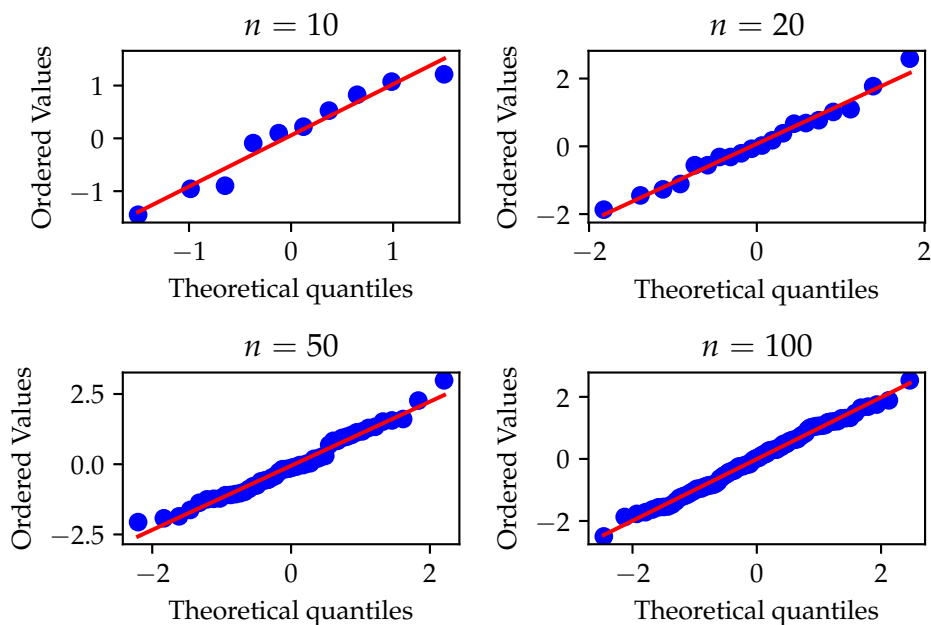
```
plt.subplot(2,2,1)
x = st.norm.rvs(size=10)
st.probplot(x,plot=plt)
plt.title("n=10")

plt.subplot(2,2,2)
x = st.norm.rvs(size=20)
st.probplot(x,plot=plt)
plt.title("n=20")

plt.subplot(2,2,3)
x = st.norm.rvs(size=50)
st.probplot(x,plot=plt)
plt.title("n=50")

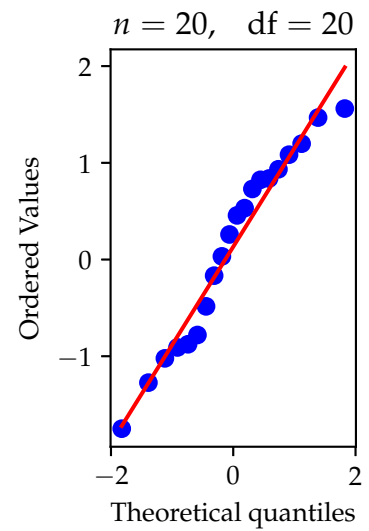
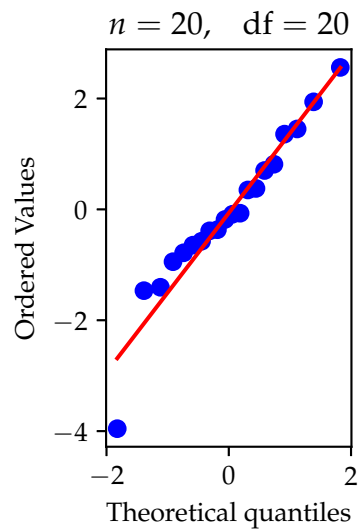
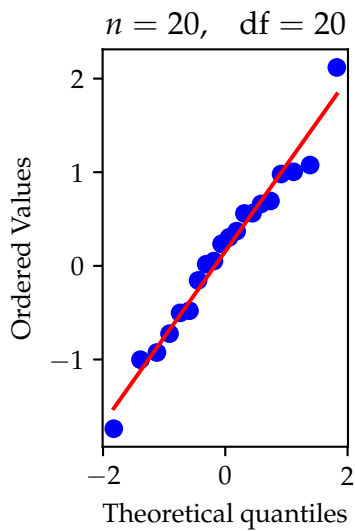
plt.subplot(2,2,4)
x = st.norm.rvs(size=100)
st.probplot(x,plot=plt)
plt.title("n=100")

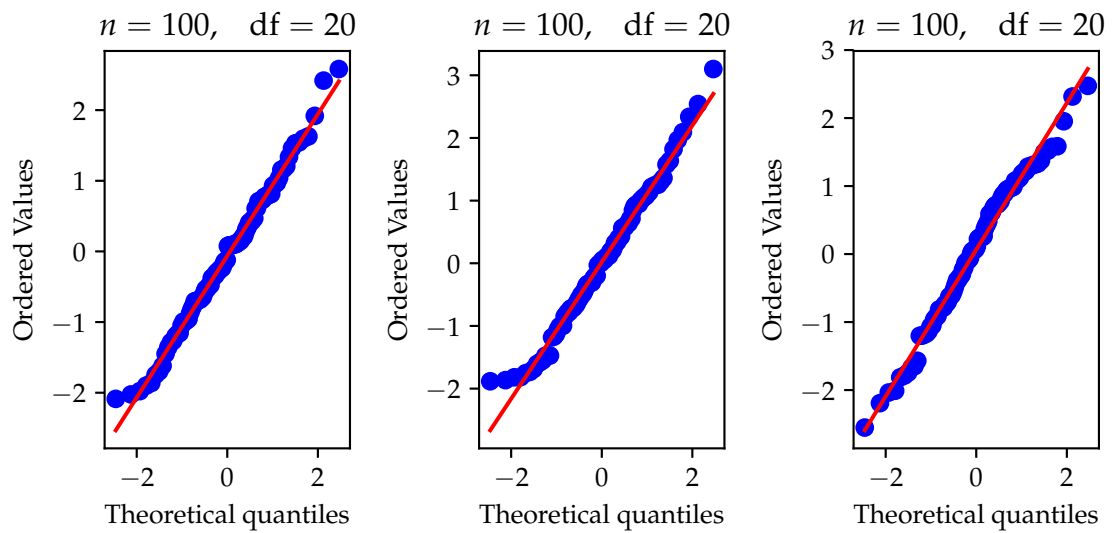
plt.tight_layout()
plt.show()
```



b) (zu R)

```
for i in range(1,4):  
    plt.subplot(1,3,i)  
    x = st.t.rvs(size=20, df=20)  
    st.probplot(x,plot=plt)  
    plt.title("n=20, df=20")  
  
plt.tight_layout()  
plt.show()  
  
for i in range(1,4):  
    plt.subplot(1,3,i)  
    x = st.t.rvs(size=100, df=20)  
    st.probplot(x,plot=plt)  
    plt.title("n=100, df=20")  
  
plt.tight_layout()  
plt.show()
```





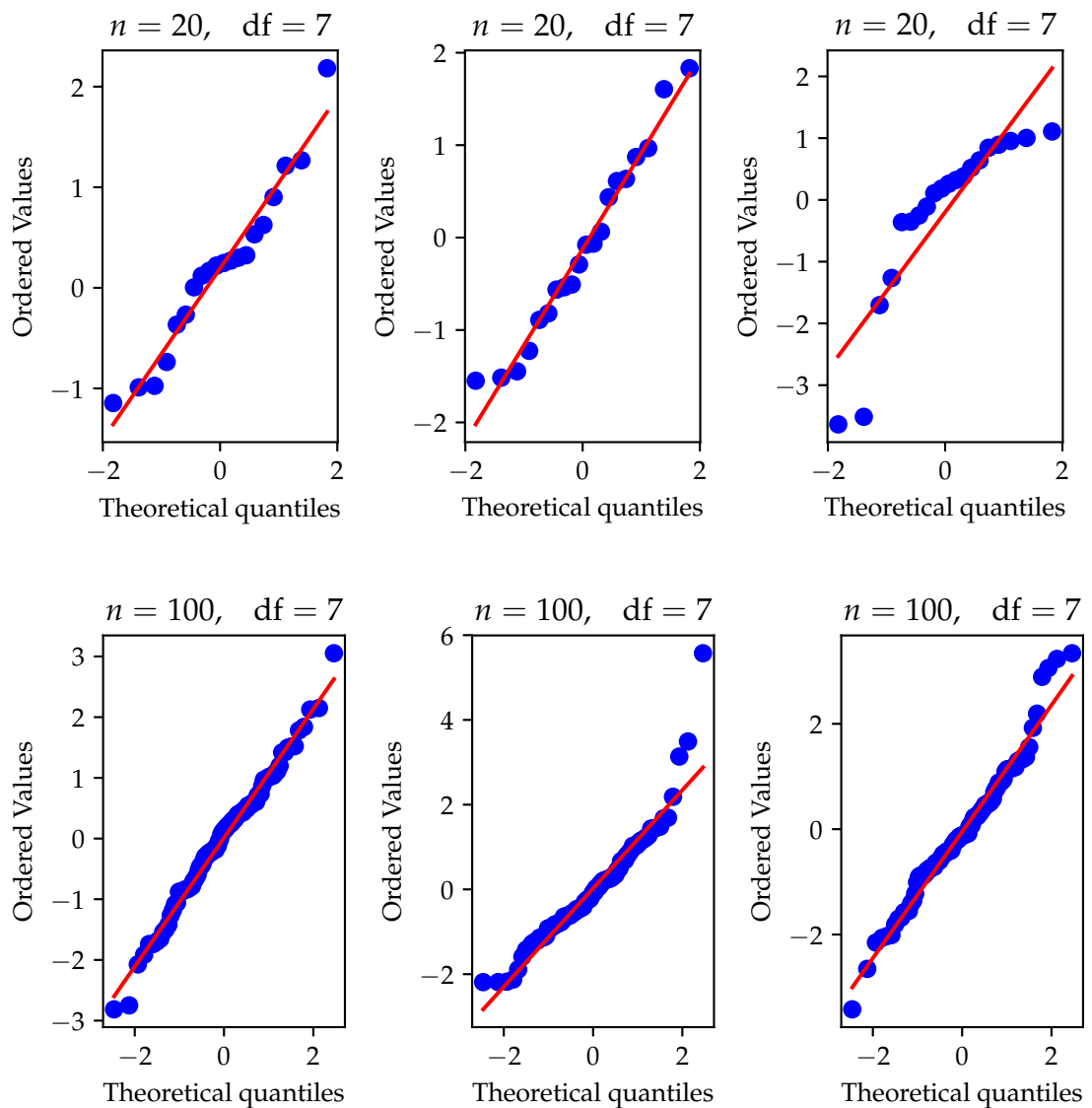
(zu R)

```
for i in range(1,4):
    plt.subplot(1,3,i)
    x = st.t.rvs(size=20, df=7)
    st.probplot(x,plot=plt)
    plt.title("n=20, df=7")

plt.tight_layout()
plt.show()

for i in range(1,4):
    plt.subplot(1,3,i)
    x = st.t.rvs(size=100, df=7)
    st.probplot(x,plot=plt)
    plt.title("n=100, df=7")

plt.tight_layout()
plt.show()
```



(zu R)

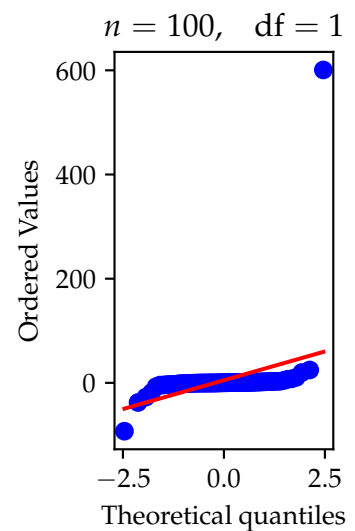
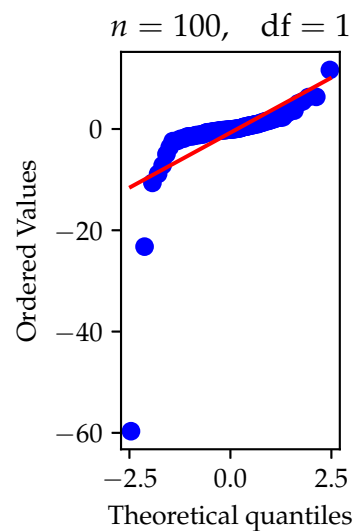
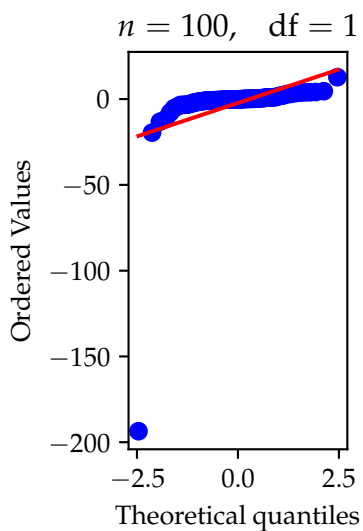
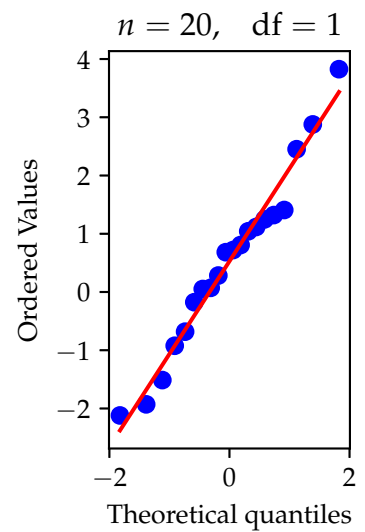
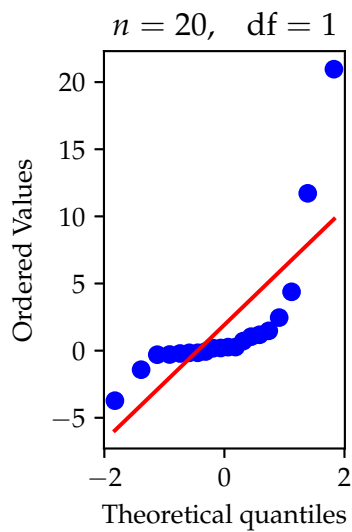
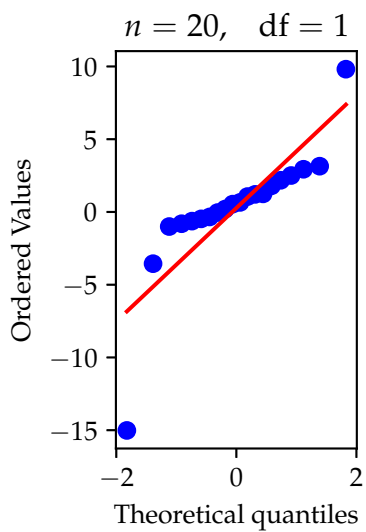
```
for i in range(1,4):
    plt.subplot(1,3,i)
    x = st.t.rvs(size=20, df=1)
    st.probplot(x,plot=plt)
    plt.title("n=20,df=1")

plt.tight_layout()
plt.show()
```

```
plt.figure(figsize=(6,3))

for i in range(1,4):
    plt.subplot(1,3,i)
    x = st.t.rvs(size=100, df=1)
    st.probplot(x,plot=plt)
    plt.title("n=100,df=1")

plt.tight_layout()
plt.show()
```



Wir beobachten, dass je grösser  $n$ , desto klarer ist die Langschwänzigkeit sichtbar.

c) (zu R)

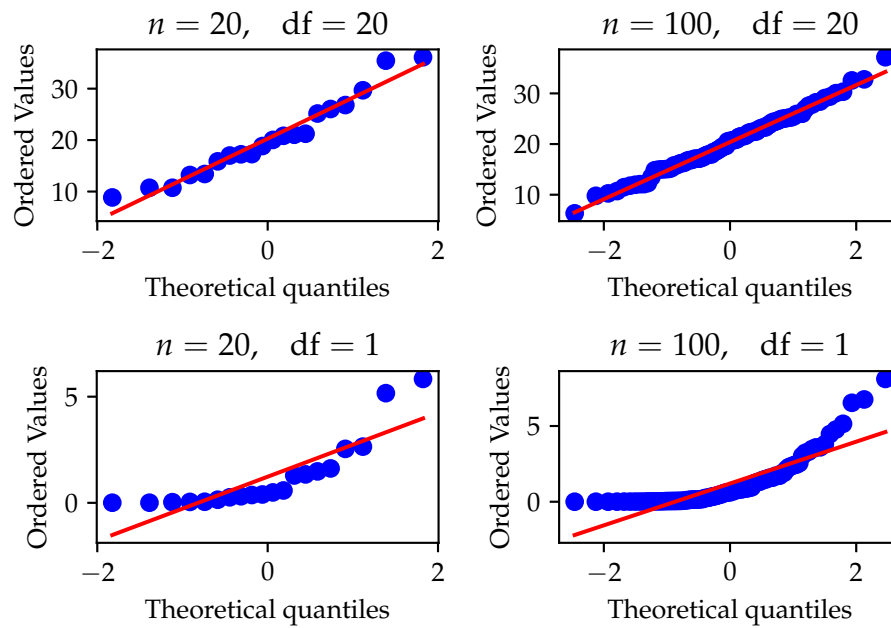
```
plt.subplot(2,2,1)
x = st.chi2.rvs(size=20, df=20)
st.probplot(x, plot=plt)
plt.title("n=20, df=20")

plt.subplot(2,2,2)
x = st.chi2.rvs(size=100, df=20)
st.probplot(x, plot=plt)
plt.title("n=100, df=20")

plt.subplot(2,2,3)
x = st.chi2.rvs(size=20, df=1)
st.probplot(x, plot=plt)
plt.title("$n=20, df=1")

plt.subplot(2,2,4)
x = st.chi2.rvs(size=100, df=1)
st.probplot(x, plot=plt)
plt.title("n=100, df=1")

plt.tight_layout()
plt.show()
```



Bei einem Freiheitsgrad ist die Rechtsschiefe klar sichtbar. Bei 20 Freiheitsgraden sind bloss noch Hinweise auf eine Rechtsschiefe. (zu R)

```
plt.subplot(2,2,1)
x = st.chi2.rvs(size=20, df=7)
st.probplot(x,plot=plt)
plt.title("n=20, df=7")

plt.subplot(2,2,2)
x = st.chi2.rvs(size=100, df=7)
st.probplot(x,plot=plt)
plt.title("n=100, df=7")

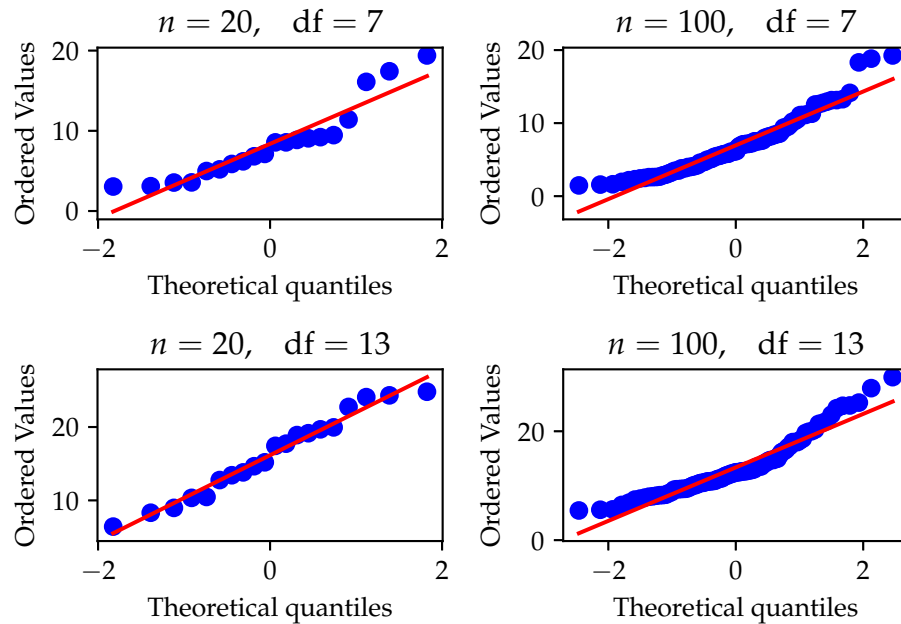
plt.subplot(2,2,3)
x = st.chi2.rvs(size=20, df=13)
st.probplot(x,plot=plt)
plt.title("n=20, df=13")

plt.subplot(2,2,4)
x = st.chi2.rvs(size=100, df=13)
st.probplot(x,plot=plt)
plt.title("n=100, df=13")

plt.tight_layout()
```



```
plt.show()
```



Bei 7 Freiheitsgraden und  $n = 100$  ist die Rechtsschiefe klar sichtbar. Ansonsten ist kaum erkennbar, dass die Verteilung rechtsschief ist.

## Lösung 8.2

- a) *Gepaarte Stichprobe*: Zu jeder Blutplättchenmenge vor dem Rauchen gehört die Blutplättchenmenge derselben Person nach dem Rauchen.

*Einseitiger Test*: Wir wollen nicht wissen, ob sich die Blutplättchenmenge *verändert* hat, sondern ob sie sich *erhöht* hat.

$H_0$ : Rauchen hat keinen Einfluss auf die Anhäufung der Blutplättchen:  $\mu_R = \mu_{NR}$

$H_A$ : Durch Rauchen erhöht sich die Anhäufung der Blutplättchen:  $\mu_R > \mu_{NR}$

- b) *Gepaarte Stichprobe*: Zu jeder Höhe eines selbstbefruchteten Setzlinge gehört die Höhe des fremdbefruchteten „Partners“.

*Einseitiger Test*: Wir wollen nicht wissen, ob sich die Höhen *unterscheiden*, sondern ob die fremdbefruchteten Setzlinge *größer* werden als die selbstbefruchteten.

$H_0$ : Die Höhen unterscheiden sich nicht:  $\mu_f = \mu_s$

$H_A$ : Fremdbefruchtete Setzlinge werden größer als selbstbefruchtete:  $\mu_f > \mu_s$

- c) *Ungepaarte Stichprobe*: Ungleiche Anzahl in den Gruppen. Zu einem Blutdruck aus der Versuchsgruppe gehört nicht ein spezifischer aus der Kontrollgruppe.

*Zweiseitiger Test*: Wir wollen nur wissen, ob das Kalzium einen Einfluss hat auf den Blutdruck, *egal* ob nach oben oder unten.

$H_0$ : Kalzium hat keinen Einfluss auf den Blutdruck:  $\mu_{\text{Kalz}} = \mu_{\text{Kontr}}$

$H_A$ : Kalzium hat einen Einfluss auf den Blutdruck:  $\mu_{\text{Kalz}} \neq \mu_{\text{Kontr}}$

- d) *Ungepaarte Stichprobe*: Die Anzahlen in den beiden Gruppen brauchen nicht gleich zu sein. Zur Eisenmessung einer „Fe<sup>2+</sup>-Maus“ gehört nicht eine bestimmte Messung einer „Fe<sup>3+</sup>-Maus“.

*Zweiseitiger Test*: Wir wollen nur wissen, ob die Mäuse die verschiedenen Eisenformen *unterschiedlich* gut aufnehmen.

$H_0$ : Die Eisenaufnahme ist von der Form unabhängig:  $\mu_2 = \mu_3$

$H_A$ : Die Eisenaufnahme ist von der Form abhängig:  $\mu_2 \neq \mu_3$

### Lösung 8.3

- a) Es handelt sich um *gepaarte* Stichproben. Am gleichen Ort wird mit beiden Geräten gemessen.

- b) • *Modell*:

$$D_1, \dots, D_n \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(\mu, \sigma^2)$$

$\sigma$  wird durch  $\hat{\sigma}$  geschätzt.

- *Nullhypothese*:

$$H_0 : \mu_D = \mu_0 = 0$$

*Alternative*:

$$H_A : \mu_D < \mu_0$$

- *Teststatistik*:  $t$  Verteilung mit Freiheitsgrad 8  
• *Signifikanzniveau*:

$$\alpha = 5\%$$

- *Verwerfungsbereich*:

$$K = (-\infty, -3.843]$$

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

t.ppf(q=0.05, loc=0, scale=6.2/np.sqrt(9), df=8)
```

```
## -3.8430659442138757
print(t.ppf(q=0.05, loc=0, scale=6.2/np.sqrt(9), df=8))
## -3.8430659442138757
```

- *Testentscheid:*

Der Wert

$$\bar{d}_n = -5.78$$

liegt im Verwerfungsbereich, d.h. eine neue Eichung der Geräte ist angezeigt.

Mittel Standardisierung:

- *Modell:*

$$D_1, \dots, D_n \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(\mu, \sigma^2)$$

$\sigma$  wird durch  $\hat{\sigma}$  geschätzt.

- *Nullhypothese:*

$$H_0 : \mu_D = \mu_0 = 0$$

*Alternative:*

$$H_A : \mu_D < \mu_0$$

- *Teststatistik:*

$$T = \frac{\sqrt{n}(\bar{D}_n - \mu_0)}{\hat{\sigma}}$$

Verteilung der Teststatistik unter  $H_0$ :

$$T \sim t_{n-1}$$

- *Signifikanzniveau:*

$$\alpha = 5\%$$

- *Verwerfungsbereich:* (zu R)

$$K = (-\infty, t_{9-1,0.05}] = (-\infty, -1.86]$$

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

t.ppf(q=0.05, df=8)

## -1.8595480375228428
```

```
print(t.ppf(q=0.05, df=8))
## -1.8595480375228428
```

- *Testentscheid:*

$$t = \frac{\bar{d}_n - 0}{\hat{\sigma} / \sqrt{9}} = -2.8$$

Der Wert  $t$  der Teststatistik liegt im Verwerfungsbereich, d.h. eine neue Eichung der Geräte ist angezeigt.

- c)  $Z$  wäre binomialverteilt mit Parametern  $n = 9$  und  $\pi = 0.5$  unter der Nullhypothese  $\mu = 0$ . Darauf aufbauend kann man auch einen Vorzeichentest durchführen. Der Vorteil ist, dass man keine Normalverteilung mehr annehmen muss.

## Lösung 8.4

- a) Es ist ein gepaarter Test, da an einer Untersuchungseinheit (Patient) zwei Messungen durchgeführt wurden.
- b) Wir wollen die Wirksamkeit testen. Dazu berechnen wir den Durchschnitt der  $\mu_D$  der Unterschiede  $D_i$ . Damit die Wirksamkeit nachgewiesen werden kann, muss

$$\mu_D > 0$$

- c) Nullhypothese (keine Wirkung des Medikamentes)

$$H_0: \mu_D = 0$$

Alternativhypothese (Wirkung des Medikamentes)

$$H_A: \mu_D > 0$$

- d) **R-Output:**

```
import pandas as pd
import scipy.stats as st
t_1 = pd.Series([39.1, 39.3, 38.9, 40.6, 39.5, 38.4, 38.6, 39.0, 38.6,
39.2])
t_2 = pd.Series([38.1, 38.3, 38.8, 37.8, 38.2, 37.3, 37.6, 37.8, 37.4,
38.1])

st.ttest_rel(t_1, t_2)

## Ttest_relResult(statistic=5.656944533090247, pvalue=0.000310841377288432)
```

Da wir einen einseitigen Test haben, müssen wir den ausgegebenen  $p$ -Wert halbieren:

```
st.ttest_rel(t_1, t_2).pvalue / 2
## 0.00015542068864421603
```

Der  $p$ -Wert 0.000155 ist hier kleiner als 0.05. Somit ist die Differenz statistisch signifikant. Wir können also davon ausgehen, dass das Medikament fiebersenkend ist.

e) **R-Output:**

```
st.wilcoxon(t_1, t_2, alternative="greater")
## WilcoxonResult(statistic=55.0, pvalue=0.002446120450917513)
```

Der  $p$ -Wert 0.00245 ist hier kleiner als 0.05. Somit ist die Differenz statistisch signifikant. Wir können also davon ausgehen, dass das Medikament fiebersenkend ist.

f) Der  $p$ -Wert des Wilcoxon-Test ist grösser als der  $p$ -Wert des  $t$ -Testes.

Da Wilcoxon-Test von weniger ausgeht (keine Normalverteilung) als der  $t$ -Test, kommt eine zusätzliche Unsicherheit dazu. Die Nullhypothese wird weniger stark verworfen.

Allerdings suggeriert der  $t$ -Test hier, dass er „genauer“ ist. Dies ist auch so, gilt aber nur, sofern die Daten normalverteilt sind, was man oft apriori nicht weiss.

Darum ist der Wilcoxon-Test dem  $t$ -Test oft vorzuziehen.

## Lösung 8.5

a) Es handelt sich um ungepaarte Stichproben, da zu den einzelnen Männchen nicht jeweils ein bestimmtes Weibchen gehört. Die Anzahlen in den beiden Stichproben brauchen auch gar nicht gleich gross zu sein.

b) Wir führen folgende Bezeichnungen ein:

$X_i$ :  $i$ -ter Wert der Kieferlänge der männlichen Tiere,  $i = 1, \dots, n = 10$

$Y_j$ :  $j$ -ter Wert der Kieferlänge der weiblichen Tiere,  $j = 1, \dots, m = 10$

Nullhypothese  $H_0$ :  $X_i$  i.i.d.  $\mathcal{N}(\mu_1, \sigma^2)$ ,  $Y_j$  i.i.d.  $\mathcal{N}(\mu_2, \sigma^2)$  unabhängig,  $\mu_1 = \mu_2$

Alternative  $H_A$ :  $X_i \sim (\mu_1, \sigma^2)$ ,  $Y_j \sim \mathcal{N}(\mu_2, \sigma^2)$  mit  $\mu_1 \neq \mu_2$

c) Der **Python**-Output für den  $t$ -Test sieht folgendermassen aus: (zu **R**)

```
import pandas as pd
import scipy.stats as st
jackals = pd.read_table("../.../Themen/Statistik_Messdaten/Uebungen_de/Da
sep=" ")

st.ttest_ind(jackals["M"], jackals["W"], equal_var=False)

## Ttest_indResult(statistic=3.4843242131699643, pvalue=0.00335995243519250
```

Der  $p$ -Wert ist  $0.0034 < 0.05$ , also wird die Nullhypothese verworfen.

- d) Der **Python**-Output für den Wilcoxon-Test sieht folgendermassen aus: (zu R)

```
import pandas as pd
import scipy.stats as st
jackals = pd.read_table("../.../Themen/Statistik_Messdaten/Uebungen_de/Da
sep=" ")

st.mannwhitneyu(jackals["M"], jackals["W"], alternative="two-sided")

## MannwhitneyuResult(statistic=87.5, pvalue=0.004845462884722891)
```

Der  $P$ -Wert ist somit  $0.0049 < 0.05$ , also wird auch bei diesem Test die Nullhypothese verworfen.

- e) Das Resultat des Wilcoxon-Tests ist vertrauenswürdiger, da er im Gegensatz zum  $t$ -Test nicht annimmt, dass die Daten normal-verteilt sind und wir diese Voraussetzung in keiner Weise überprüft haben. Allerdings ist die stark unterschiedliche Standardabweichung in den zwei Gruppen problematisch für beide Tests.

## Lösung 8.6

- a) Wir schätzen mit **Python** den Mittelwert und die Standardabweichung folgendermassen: (zu R)

```
import numpy as np
from pandas import Series

mdma_zuerich = Series([16.3, 12.7, 14.0, 53.3, 117, 62.6, 27.6])
mdma_basel = Series([10.4, 8.91, 11.7, 29.9, 46.3, 25.0, 29.4])

d = mdma_zuerich - mdma_basel

d.mean()

## 20.27
```

```
d.std()
## 26.272304175056032
```

Auf den ersten Blick scheint die Sachlage klar : in Zürich wird mehr MDMA konsumiert. Wir stellen dann allerdings fest, dass die Differenzen eine ziemlich grosse Streuung aufweisen.

- b) Die Städte Zürich und Basel können als unterschiedliche Versuchseinheiten aufgefasst werden. Wir fassen die Stichproben also als ungepaart auf.
- c) Die Nullhypothese lautet, dass es keinen Unterschied zwischen den beiden Städten in Bezug auf die ausgeschiedene Menge an MDMA gibt, also  $\mu_D = \mu_{\text{Zürich}} - \mu_{\text{Basel}} = 0$ . Die Alternativhypothese entspricht der Behauptung der Gratiszeitung, nämlich dass in Zürich mehr Drogen konsumiert werden und damit mehr MDMA ausgeschieden wird, also  $\mu_D > \mu_0 = 0$ .
- d) Da es sich um ungepaarte Stichproben handelt, gilt: (zu R)

```
import numpy as np
from pandas import Series
import scipy.stats as st

mdma_zuerich = Series([16.3, 12.7, 14.0, 53.3, 117, 62.6, 27.6])
mdma_basel = Series([10.4, 8.91, 11.7, 29.9, 46.3, 25.0, 29.4])

st.ttest_ind(mdma_zuerich,mdma_basel, equal_var=False)

## Ttest_indResult(statistic=1.3273296255450922, pvalue=0.2232566584677775)
```

In diesem Fall ist der zweiseitige  $p$ -Wert  $pvalue = 0.22$ , d.h., der einseitige  $p$ -Wert ist grösser als 0.05, womit die Nullhypothese beibehalten wird.

Würden wir die Daten als gepaarte Stichproben auffassen, dann haben wir für  $\bar{D}_n \sim \mathcal{N}\left(\mu_D, \frac{\sigma_D^2}{n}\right)$  (Zentraler Grenzwertsatz). Dann ist die Teststatistik:

$$T = \frac{\bar{D}_n - \mu_0}{\hat{\sigma}_D / \sqrt{n}} = \frac{\sqrt{n}\bar{D}_n}{\hat{\sigma}_D}$$

$$t = 2.0413$$

Verteilung von  $T$  unter  $H_0$ :  $T \sim t_{n-1} = t_6$ . Mit Python ergibt sich : (zu R)

```
import numpy as np
from pandas import Series
import scipy.stats as st
```

```
mdma_zuerich = Series([16.3, 12.7, 14.0, 53.3, 117, 62.6, 27.6])
mdma_basel = Series([10.4, 8.91, 11.7, 29.9, 46.3, 25.0, 29.4])

st.ttest_rel(mdma_zuerich,mdma_basel)

## Ttest_relResult(statistic=2.041289516059924, pvalue=0.08728840175305683)
```

Da wir hier einen einseitigen Test haben und **Python** den *P*-Wert für eine zweiseitige Alternative berechnet, müssen wir den Wert für **pvalue** halbieren und somit ist der *P*-Wert 0.04364. Er ist (knapp) kleiner als  $\alpha = 0.05$  und somit wird auf dem 5% Signifikanzniveau die Nullhypothese verworfen. Falls die Daten also als gepaarte Stichproben aufgefasst würden, lautet die Schlussfolgerung, dass signifikant mehr Ecstasy in Zürich als in Basel konsumiert wird.

e) (zu R)

```
import numpy as np
from pandas import Series
import scipy.stats as st

mdma_zuerich = Series([16.3, 12.7, 14.0, 53.3, 117, 62.6, 27.6])
mdma_basel = Series([10.4, 8.91, 11.7, 29.9, 46.3, 25.0, 29.4])

st.mannwhitneyu(mdma_zuerich, mdma_basel, alternative='greater')

## MannwhitneyuResult(statistic=34.0, pvalue=0.12507650397969117)
```

Für den Mann-Whitney-U-Test kann in **Python** die Richtung der Alternativhypothese definiert werden, und zwar mit **alternative='...'**. Wir können folglich den Wert von **pvalue** direkt für den Testentscheid herbeiziehen: Der (einseitige) *P*-Wert beträgt in diesem Fall 0.13 und ist somit grösser als das Signifikanzniveau 5%. Wir behalten die Nullhypothese also bei. Offenbar hängt der Testentscheid stark von der Verteilungsannahme der Differenzen ab.

## Lösung 8.7

- Die Messwerte des einen Datensatzes können wir nicht eindeutig den Werten der anderen Messwerte zu ordnen. Also handelt es um einen ungepaarten Test. Zudem sind die Messreihen verschieden lang.
- Keine Präferenz apriori zwischen Geflügel- und Rindshotdog erkennbar, also zweiseitiger Test.
- Da es sich um einen ungepaarten Test handelt, verglichen wird die Mittelwerte  $\mu_X$  und  $\mu_Y$ .



Nullhypothese (kein Unterschied im Kaloriengehalt)

$$H_0: \mu_X = \mu_Y$$

Alternativhypothese (Unterschied im Kaloriengehalt)

$$H_A: \mu_X \neq \mu_Y$$

d) R-Output:

```
import pandas as pd
import scipy.stats as st
x = pd.Series([186, 181, 176, 149, 184, 190, 158, 139, 175, 148, 152,
               111, 141, 153, 190, 157, 131, 149, 135, 132])
y = pd.Series([129, 132, 102, 106, 94, 102, 87, 99, 170, 113, 135,
               142,
               86, 143, 152, 146, 144])

x.mean()
## 156.85

y.mean()
## 122.47058823529412
```

Der Kaloriengehalt der Rinds-Hotdogs scheint doch um einiges höher als bei den Geflügel-Hotdogs zu sein. Die Nullhypothese dürfte verworfen werden. Trotzdem machen wir einen zweiseitigen Test, da dieser Ausgang nicht von vorneherein klar war.

Man *könnte* argumentieren, dass man schon „weiss“, das Rind fetthaltiger als Geflügel ist und dann wäre ein einseitiger Test angebracht.

e) Da keine Angabe vorhanden ist, ob die Daten normalverteilt sind, wählen wir sicherheitshalber den Wilcoxonstest.

f) R-Output:

```
st.mannwhitneyu(x, y)
## MannwhitneyUResult(statistic=54.5, pvalue=0.0002274490378782059)
```

Der  $p$ -Wert ist 0.00023 und damit weit unter dem Signifikanzniveau von 0.05. Somit wird die Nullhypothese, dass die beiden Hotdog-Arten den gleichen Kaloriengehalt haben *verworfen*. Somit hat es im Rinds-Hotdog statistisch signifikant mehr Kalorien als im Geflügel-Hotdog.

## Lösung 8.8

- a) i) Es ist ein gepaarter Test. Zu jeder Untersuchungseinheit (Ehepaar) gibt es zwei zugehörige Messungen (Alter Mann, Alter Frau).
- ii) Wir sind uns nicht sicher ob, die Ehemänner wirklich älter sind als die Ehefrauen. Also machen wir einen zweiseitigen Test.

Allerdings könnten aufgrund des Boxplots oben auch für einen einseitigen Test argumentieren.

- iii) Nullhypothese

$$H_0: \mu_D = 0$$

Alternativhypothese:

$$H_0: \mu_D \neq 0$$

```
import scipy.stats as st
diff = mf["alter.mann"] - mf["alter.frau"]
st.ttest_1samp(diff, popmean=0).pvalue
## 2.473721344120317e-11
```

oder

```
st.ttest_rel(mf["alter.mann"], mf["alter.frau"]).pvalue
## 2.473721344120317e-11
```

Der  $p$ -Wert ist bei weitem unter dem Signifikanzniveau von 5 % und somit wird die Nullhypothese verworfen. Die Ehemänner sind statistisch signifikant älter als ihre Ehefrauen.

Das Vertrauensintervall ist

$$(1.61, 2.85)$$

```
import numpy as np
st.t.interval(alpha=0.95, loc=diff.mean(), scale=diff.std()/np.sqrt(diff.size), df=diff.size-1)
## (1.6182861331431675, 2.8523021021509503)
```

Mit 95 % liegt der wahre Mittelwert in diesem Bereich. Die Nullhypothese war  $\mu_D = 0$ , also kein Altersunterschied. Dieser Wert liegt *nicht* im Vertrauensintervall und somit wird auch hier die Nullhypothese verworfen. Es gibt einen statistisch signifikanten Altersunterschied innerhalb der Ehepaare.

- iv) Wilcoxontest:

```
st.wilcoxon(diff, correction=True, alternative='two-sided').pvalue
## 3.977268697340033e-12
```

oder

```
st.wilcoxon(mf["alter.mann"], mf["alter.frau"], correction=True).pvalue
## 3.977268697340033e-12
```

Auch hier ist der  $p$ -Wert weit unter dem Signifikanzniveau von 5% und somit wird die Nullhypothese verworfen. Die Ehemänner sind statistisch signifikant älter als ihre Ehefrauen.

- b) i) Hier werden nur die Durchschnittsgrößen der Männer und Frauen verglichen, somit ist es ein ungepaarter Test. Da wir nicht wissen, ob die Abweichung nach oben oder nach unten ist, machen wir wieder einen zweiseitigen Test.
- ii) Sei  $\mu_F$  der Durchschnitt der Körpergröße der Frauen und  $\mu_M$  der Durchschnitt der Körpergrößen der Männer. Dann lautet die Nullhypothese

$$H_0: \mu_F = \mu_M - 13$$

und die Alternativhypothese

$$H_A: \mu_F \neq \mu_M - 13$$

- iii) Wir nehmen Normalverteilung der Körpergrößen an:

```
st.ttest_ind(mf["groesse.mann"]-13, mf["groesse.frau"], equal_var=False).pvalue
## 0.5272102605793969
```

oder

Der  $p$ -Wert ist weit grösser als das Signifikanzniveau und somit wird die Nullhypothese nicht verworfen. Die Daten widerlegen den Größenunterschied von 13 cm statistisch signifikant *nicht*.

Das Vertrauensintervall ist

$$(-1.81, 0.93)$$

```
import statsmodels.stats.api as sms
X1 = mf["groesse.mann"]
X2 = mf["groesse.frau"]
```

```
cm = sms.CompareMeans(sms.DescrStatsW(X1 - 13), sms.DescrStatsW(X2))
cm.tconfint_diff(usevar='unequal')

## (-1.8122809664878108, 0.9299280253113469)
```

Mit 95 % liegt der wahre Mittelwert in diesem Bereich. Die Nullhypothese war  $\mu_F = \mu_M - 13$ , also keine Abweichung vom Größenunterschied von 13 cm. Dieser Wert 0 liegt im Vertrauensintervall und somit wird auch hier die Nullhypothese *nicht* verworfen. Die Daten widerlegen den Größenunterschied von 13 cm statistisch signifikant *nicht*.

Gehen wir nicht von Normalverteilung aus, so wählen wir einen ungepaarten Wilcoxon Test (Mann-Whitney-U).

```
st.mannwhitneyu(mf["groesse.mann"]-13, mf["groesse.frau"], use_continuity=True, alternative="two-sided").pvalue

## 0.446086539266324
```

Auch hier wird die Nullhypothese klar nicht verworfen.

# R-Code

## Aufgabe ??

a) `pbinom(c(2,3), 12, p=0.5)`

```
## [1] 0.01928711 0.07299805
```

```
binom.test(7, 12, p=0.5, alternative="less")
```

```
##
## Exact binomial test
##
## data: 7 and 12
## number of successes = 7, number of trials =
## 12, p-value = 0.8062
## alternative hypothesis: true probability of success is less than 0.5
## 95 percent confidence interval:
## 0.0000000 0.8189752
## sample estimates:
## probability of success
## 0.5833333
```

b) `wilcox.test(c(71, 69, 67, 68, 73, 72, 71, 71, 68, 72, 69, 72),  
mu=70, alternative="less")`

```
##
## Wilcoxon signed rank test with continuity
## correction
##
## data: c(71, 69, 67, 68, 73, 72, 71, 71, 68, 72, 69, 72)
## V = 44.5, p-value = 0.6838
## alternative hypothesis: true location is less than 70
```

## Aufgabe 8.3

b) (zu Python)

```
qt(0.05, df=8)
```

```
## [1] -1.859548
```

## Aufgabe 8.5

a)

b) (zu Python)

```
jackals <- read.table(file="./Daten/jackals.txt",header=TRUE)
t.test(jackals[, "M"], jackals[, "W"])

##
## Welch Two Sample t-test
##
## data: jackals[, "M"] and jackals[, "W"]
## t = 3.4843, df = 14.894, p-value = 0.00336
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## 1.861895 7.738105
## sample estimates:
## mean of x mean of y
## 113.4 108.6
```

c) (zu Python)

```
wilcox.test(jackals[, "M"], jackals[, "W"])

##
## Wilcoxon rank sum test with continuity
## correction
##
## data: jackals[, "M"] and jackals[, "W"]
## W = 87.5, p-value = 0.004845
## alternative hypothesis: true location shift is not equal to 0
```

Der Wert ist hier leicht grösser als bei Python. Der Grund liegt an *Stetigkeitskorrekturen*, die jeweils leicht anders gehandhabt werden. Die Grössenordnung bleibt allerdings gleich.

## Aufgabe 8.6

a) (zu Python)

```
mdma_zuerich <- c(16.3, 12.7, 14.0, 53.3, 117, 62.6, 27.6)
mdma_basel <- c(10.4, 8.91, 11.7, 29.9, 46.3, 25.0, 29.4)
d <- mdma_zuerich - mdma_basel
mean(d)

## [1] 20.27

sd(d)

## [1] 26.2723
```

d) (zu Python)

```
t.test(mdma_zuerich, mdma_basel, alternative="greater",
       paired=TRUE)

##
## Paired t-test
##
## data: mdma_zuerich and mdma_basel
## t = 2.0413, df = 6, p-value = 0.04364
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  0.9742245      Inf
## sample estimates:
## mean of the differences
##                20.27
```

(zu Python)

```
t.test(mdma_zuerich, mdma_basel, alternative="greater",
       paired=FALSE)

##
## Welch Two Sample t-test
##
## data: mdma_zuerich and mdma_basel
## t = 1.3273, df = 7.5245, p-value = 0.1116
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
## -8.362064      Inf
## sample estimates:
## mean of x mean of y
##  43.35714  23.08714
```

f) (zu Python)

```
wilcox.test(mdma_zuerich, mdma_basel,
            alternative="greater", paired=FALSE)

##
## Wilcoxon rank sum test
##
## data: mdma_zuerich and mdma_basel
## W = 34, p-value = 0.1297
## alternative hypothesis: true location shift is greater than 0
```