Applied Statistics for Data Science

Serie 10

Aufgabe 10.1

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 anhäuften. 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 (Fe^{2+} und Fe^{3+}) unterschiedlich gut aufnehmen. Dazu wurden 36 Mäuse in zwei Gruppen zu je 18 unterteilt und die eine Gruppe mit Fe^{2+} und die andere mit 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 10.2

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?

a) Handelt es sich um verbundene (gepaarte) oder um unabhängige Stichproben?

b) Führen Sie einen t-Test auf dem Niveau $\alpha = 0.05$ durch. Formulieren Sie explizit: Modellannahmen, Nullhypothese, Alternative, und Testergebnis.

Aufgabe 10.3

In der folgenden Tabelle sind die Kieferlängen von 10 männlichen und 10 weiblichen Goldschakalen eingetragen: Einige Kennzahlen: $\bar{x}_n = 113.4$, $\bar{y}_n = 108.6$, $\widehat{\sigma}_x^2 = 13.82$,

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

$\widehat{\sigma}_y^2 = 5.16$

a) Handelt es sich um gepaarte oder ungepaarte Stichproben? Begründen Sie Ihre Antwort.

b) Formulieren Sie Nullhypothese und Alternativhypothese.

c) Führen Sie den t-Test nun noch mit Hilfe von **R** durch. Geben Sie den resultierenden p-Wert sowie den daraus folgenden Testentscheid an.

```r
# Datensatz anschauen
jackals <- read.table(file = "../../../Themen/Statistik_Messdaten/Uebungen_de/Daten/jackals.txt",
    header = TRUE)  # Datensatz  einlesen
jackals
# t-Test durchfuehren
t.test(jackals[, "M"], jackals[, "W"])
```

d) Führen Sie mit Hilfe von **R** einen Wilcoxon-Test durch. Geben Sie wiederum p-Wert und Testentscheid an.

```
# Wilcoxon-Test durchfuehren
wilcox.test(jackals[, "M"], jackals[, "W"], )
```

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

Aufgabe 10.4

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?

a) Handelt es sich um einen gepaarten oder ungepaarten Test? Begründen Sie Ihre Antwort.

b) Handelt es sich um einen ein- oder zweiseitigen Test? Begründen Sie Ihre Antwort.

c) Formulieren Sie Null- und Alternativhypothese.

d) Berechnen Sie die Mittelwerte der beiden Testreihen. Welche Vermutung haben Sie?

e) Welchen Test würden Sie hier wählen: t-Test oder Wilcoxontest? Begründen Sie Ihre Antwort.

f) Führen Sie den entsprechenden Test mit **R** durch. Interpretieren Sie p-Wert. Bei welcher Hotdog-Art ist der Kaloriengehalt grösser?

Aufgabe 10.5

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 all, *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 gemessenen. Aufgrund dieser Studie war eine Aussage einer vielgelesenen Schweizer Gratiszeitung, dass in Zürich viel mehr Drogen konsumiert werden als anderswo.

In der nachfolgenden Tabellle sind für die Städte Zürich und Basel die an den untersuchten Tagen ausgeschiedenen Mengen MDMA augeführt - die Werte finden Sie in der Datei *mdma.txt* im Verzeichnis **Daten** auf dem Desktop. Die Angaben sind in mg pro 1000 Einwohner pro Tag. Nehmen Sie an, dass die täglichen Differenzen D_i zwi-

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

schen den pro tausend Einwohner ausgeschiedenen Mengen von MDMA im Abwasser von Zürich und Basel unabhängig voneinander normalverteilt mit Erwartungswert μ_D und Standardabweichung σ_D sind.

a) Schätzen Sie aus den Daten den Mittelwert und die Standardabweichung der Differenzen, d.h., $\widehat{\mu}_D$ und $\widehat{\sigma}_D$.

b) Handelt es sich um gepaarte oder ungepaarte Stichproben? Begründen Sie Ihre Antwort.

c) Formulieren Sie die Nullhypothese und die Alternativhypothese, wenn Sie die Aussage der besagten Gratiszeitung überprüfen möchten.

d) Führen Sie einen statistischen Test mit Hilfe von **R** auf dem Signifikanzniveau 5 % durch, unter der Annahme, dass die Daten normalverteilt sind. Wie lautet die Teststatistik und wie ist diese unter der Nullhypothese verteilt?

e) Geben Sie das (einseitiges) 95 %-Vertrauensintervall für die Differenzen D_i an (mit Hilfe von **R**). Wie interpretieren Sie dieses Vertrauensintervall?

f) Führen Sie nun einen statistischen Test mit Hilfe von **R** auf dem Signifikanzniveau 5 % durch, unter der Annahme, dass die Daten nicht normalverteilt sind.

Aufgabe 10.6

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:

```
mf <- read.csv(".../mannfrau.csv")
```

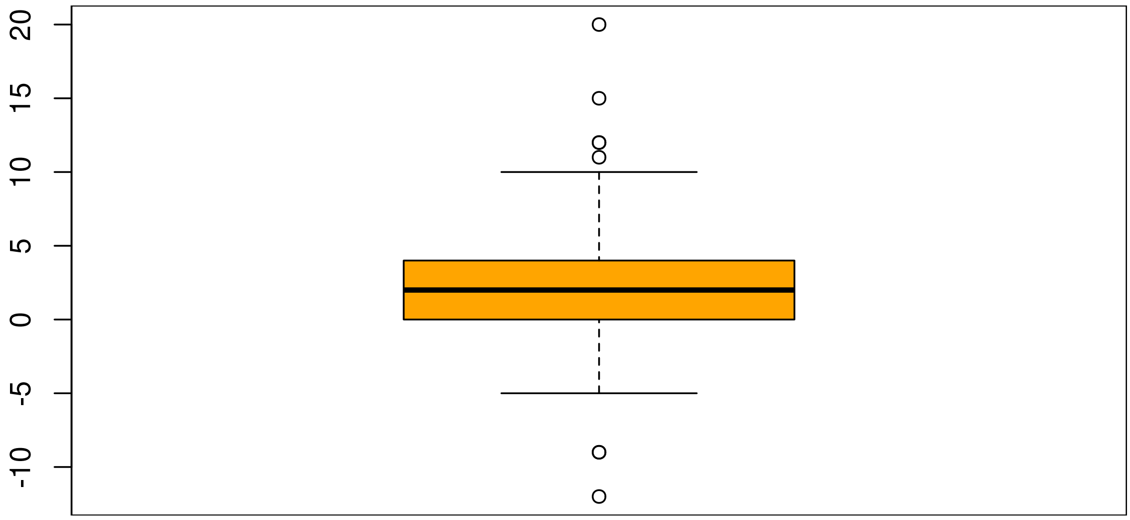
Die ... stehen für den Pfad, wo die Datei abgespeichert wurde.

Wir zeichnen den Boxplot für die Altersdifferenz der Ehepaare.

```
mf <- read.csv("../../../Themen/Deskriptive_Statistik/Uebungen_de/Daten/mannfrau.csv")

diff <- mf$alter.mann - mf$alter.frau

boxplot(diff, col = "orange")
```



Bemerkung **R**: Der Ausdruck

```
mf$alter.mann
```

ist gleichbedeutend mit

```
mf["alter.mann"]
```

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 zweisseitig, 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.

Aufgabe 10.7

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

a) Handelt es sich um einen gepaarten oder ungepaarten Test? Begründen Sie Ihre Antwort.

b) Handelt es sich um einen ein- oder zweiseitigen Test? Begründen Sie Ihre Antwort.

c) Formulieren Sie Null- und Alternativhypothese.

d) Wir nehmen an, die Daten seien normalverteilt. Welchen Test wählen Sie? Führen Sie den Test mit **R** auf Signifikanzniveau 5 % durch.

e) Wenn wir nicht davon ausgehen können, dass die Daten normalverteilt sind, welchen Test wählen Sie? Führen Sie diesen auf Signifikanzniveau 5 % durch.

f) Erklären Sie den Unterschied der p-Werte in Teilaufgaben d) und e).

Aufgabe 10.8

Betrachten Sie 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. Machen Sie ein vernünftige Skizze, wo Sie die relevanten Grössen einzeichnen.

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

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

c) Der p-Wert ist strikt kleiner als 0.5.

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

e) 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.

Applied Statistics for Data Science

Musterlösungen zu Serie 10

Lösung 10.1

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 Bluplättchenmenge *verändert* hat, sondern ob sie sich *erhöht* hat.

H_0: Rauchen hat keinen Einfluss auf die Anhäufung der Blutplättchen. ($\mu_{\text{R}} = \mu_{\text{NR}}$)

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^{2+}-Maus" gehört nicht eine bestimmte Messung einer "Fe^{3+}-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 10.2

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

b)
- *Modell:* $D_1, \ldots, D_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mu, \sigma^2)$, σ wird durch $\widehat{\sigma}$ geschätzt.
- *Nullhypothese:*
$$H_0: \quad \mu_D = \mu_0 = 0$$
Alternative:
$$H_A: \quad \mu_D < \mu_0$$
- *Signifikanzniveau:*
$$\alpha = 5\,\%$$
- *Testentscheid:*

```
x <- c(120, 265, 157, 187, 219, 288, 156, 205, 163)
y <- c(127, 281, 160, 185, 220, 298, 167, 203, 171)

t.test(x, y, paired = T, alternative = "less")

##
##  Paired t-test
##
## data:  x and y
## t = -2.7955, df = 8, p-value = 0.01168
## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:
##      -Inf -1.93449
## sample estimates:
## mean of the differences
##               -5.777778
```

Der p-Wert ist kleiner als das Signifikanzniveau und somit wird die Nullhypothese verworfen. Das heisst, das Gerät B statistisch signifikant grössere Werte misst. Die Geräte müssen neu geeicht werden.

Lösung 10.3

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, $i = 1, \ldots, n = 10$

- Y_j: j-ter Wert der Kieferlänge der weiblichen Tiere, $j = 1, \ldots, m = 10$
- X_i i.i.d. $\mathcal{N}(\mu_1, \sigma^2)$, Y_j i.i.d. $\mathcal{N}(\mu_2, \sigma^2)$ unabhängig

Nullhypothese H_0:

$$H_0: \quad \mu_1 = \mu_2$$

Alternative H_A:

$$\mu_1 \neq \mu_2$$

u

c) Der **R**-Output für den t-Test sieht folgendermassen aus:

```
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
```

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

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

```
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 p-Wert ist $0.0048 < 0.05$, 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 10.4

a) 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.

b) Keine Präferenz apriori zwischen Geflügel- und Rindshotdog erkennbar, also zweiseitiger Test.

c) Da es sich um einen ungepaarten Test handelt, vergleichen wird die Mittelwerte μ_X und μ_Y.

Nullhypothese (kein Unterschied im Kaloriengehalt)

$$H_0 : \quad \mu_X = \mu_Y$$

Alternativhypothese (Unterschied im Kaloriengehalt)

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

d) **R**-Output:

```
x <- c(186, 181, 176, 149, 184, 190, 158, 139, 175, 148, 152, 111,
    141, 153, 190, 157, 131, 149, 135, 132)

y <- c(129, 132, 102, 106, 94, 102, 87, 99, 170, 113, 135, 142, 86,
    143, 152, 146, 144)

mean(x)

## [1] 156.85

mean(y)

## [1] 122.4706
```

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 Wilcoxontest.

f) **R**-Output:

```
wilcox.test(x, y, paired = FALSE)
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  x and y
## W = 285.5, p-value = 0.0004549
## alternative hypothesis: true location shift is not equal to 0
```

Der p-Wert ist 0.00046 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 10.5

a) Wir schätzen mit **R** den Mittelwert und die Standardabweichung folgendermassen:

```
mdma_zuerich <- c(16.3, 12.7, 14, 53.3, 117, 62.6, 27.6)
mdma_basel <- c(10.4, 8.91, 11.7, 29.9, 46.3, 25, 29.4)
d <- mdma_zuerich - mdma_basel
mean(d)

## [1] 20.27

sd(d)

## [1] 26.2723
```

b) Betrachten wir den Versuchstag als Versuchseinheit, dann handelt es sich um gepaarte Stichproben. Es könnte allerdings auch argumentiert werden, dass die Städte als Versuchseinheiten unterschiedlich sind. In diesem Fall würde man die Stichproben als ungepaart auffassen.

c) Die Nullhypothese lautet, dass es keinen Unterschied zwischen den beiden Städten in Bezug auf die ausgeschiedene Menge an MDMA gibt, also $\mu_D = \mu_0 = 0$. 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)
```
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
```

Aus dem **R**-Output entnehmen wir, dass der P-Wert 0.04364 ist und somit kleiner als $\alpha = 0.05$. Auf dem 5 % Signifikanzniveau verwerfen wir also die Nullhypothese. Es wird also signifikant mehr Ecstasy in Zürich konsumiert. Fassen wir die Stichproben als ungepaart auf, so gilt:

```
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
```

Wir haben für $\overline{D}_n \sim \mathcal{N}(\mu_D, \frac{\sigma_D^2}{n})$ (Zentraler Grenzwertsatz). Dann ist die Teststatistik:

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

$$t = 2.0413$$

Verteilung von T unter H_0: $T \sim t_{n-1} = t_6$

e) Aus dem **R**-Output entnehmen wir, dass das 95 %-Vertrauensintervall für die Teststatistik T gegeben ist durch:

$$[0.9742245, \infty]$$

Mit 95 % Wahrscheinlichkeit liegt die wahre Differenz zwischen den ausgeschiedenen Mengen MDMA pro tausend Einwohner in Zürich und Basel im Intervall $[0.9742245, \infty]$. Daraus erkennen wir auch, dass 0 nicht im 95 %-Vertrauensintervall enthalten ist, so dass wir die Nullhypothese auf dem 5 %-Signifikanzniveau verwerfen können.

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

##
##  Wilcoxon signed rank test
##
## data:  mdma_zuerich and mdma_basel
## V = 27, p-value = 0.01563
```

```
## alternative hypothesis: true location shift is greater than 0
```

Auch in diesem Fall können wir die Nullhypothese verwerfen, da der p-Wert 0.01563 ist und somit kleiner als das Signifikanzniveau 5 %.

Lösung 10.6

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

$$H_0 : \quad \mu_D = 0$$

Alternativhypothese:

$$H_0 : \quad \mu_D \neq 0$$

```
t.test(diff)

##
##  One Sample t-test
##
## data:  diff
## t = 7.1518, df = 169, p-value = 2.474e-11
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  1.618286 2.852302
## sample estimates:
## mean of x
##  2.235294
```

oder

```
t.test(mf$alter.mann, mf$alter.frau, paired = TRUE)

##
##  Paired t-test
##
## data:  mf$alter.mann and mf$alter.frau
## t = 7.1518, df = 169, p-value = 2.474e-11
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  1.618286 2.852302
## sample estimates:
## mean of the differences
```

```
##                 2.235294
```

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

$$(1.61,\ 2.85)$$

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:

```
wilcox.test(diff)

##
##  Wilcoxon signed rank test with continuity correction
##
## data:  diff
## V = 9460, p-value = 3.977e-12
## alternative hypothesis: true location is not equal to 0
```

oder

```
wilcox.test(mf$alter.mann, mf$alter.frau, paired = TRUE)

##
##  Wilcoxon signed rank test with continuity correction
##
## data:  mf$alter.mann and mf$alter.frau
## V = 9460, p-value = 3.977e-12
## alternative hypothesis: true location shift is not equal to 0
```

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 μ_F der Durchschnitt der Körpergrösse der Frauen und μ_M der Durchschnitt der Körpergrössen der Männer. Dann lautet die Nullhypothese

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

und die Alternativhypothese

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

iii) Wir nehmen Normalverteilung der Körpergrössen an:

```
t.test(mf$groesse.mann - 13, mf$groesse.frau, paired = FALSE)

##
##  Welch Two Sample t-test
##
## data:  mf$groesse.mann - 13 and mf$groesse.frau
## t = -0.63293, df = 336.53, p-value = 0.5272
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -1.812281  0.929928
## sample estimates:
## mean of x mean of y
##  159.8471  160.2882
```

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

$$(-1.81, \ 0.92)$$

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).

```
wilcox.test(mf$groesse.mann - 13, mf$groesse.frau, paired = FALSE)

##
##  Wilcoxon rank sum test with continuity correction
##
## data:  mf$groesse.mann - 13 and mf$groesse.frau
## W = 13760, p-value = 0.4461
## alternative hypothesis: true location shift is not equal to 0
```

Auch hier wird die Nullhypothese klar nicht verworfen.

Lösung 10.7

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 μ_D der Unterschiede D_i. Damit die Wirksamkeit nachgewiesen werden kann, muss
$$\mu_D > 0$$

c) Nullhypothese (keine Wirkung des Medikamentes)
$$H_0: \quad \mu_D = 0$$
Alternativhypothese (Wirkung des Medikamentes)
$$H_A: \quad \mu_D > 0$$

d) **R**-Output:

```
t.1 <- c(39.1, 39.3, 38.9, 40.6, 39.5, 38.4, 38.6, 39, 38.6, 39.2)
t.2 <- c(38.1, 38.3, 38.8, 37.8, 38.2, 37.3, 37.6, 37.8, 37.4, 38.1)

t.test(t.1, t.2, paired = T, alternative = "greater")

##
##  Paired t-test
##
## data:  t.1 and t.2
## t = 5.6569, df = 9, p-value = 0.0001554
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  0.7976252       Inf
## sample estimates:
## mean of the differences
##                    1.18
```

Der p-Wert 0.0001554 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:

```
wilcox.test(t.1, t.2, paired = T, alternative = "greater")

##
##  Wilcoxon signed rank test with continuity correction
##
## data:  t.1 and t.2
## V = 55, p-value = 0.002865
## alternative hypothesis: true location shift is greater than 0
```

Der p-Wert 0.002865 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 der 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.

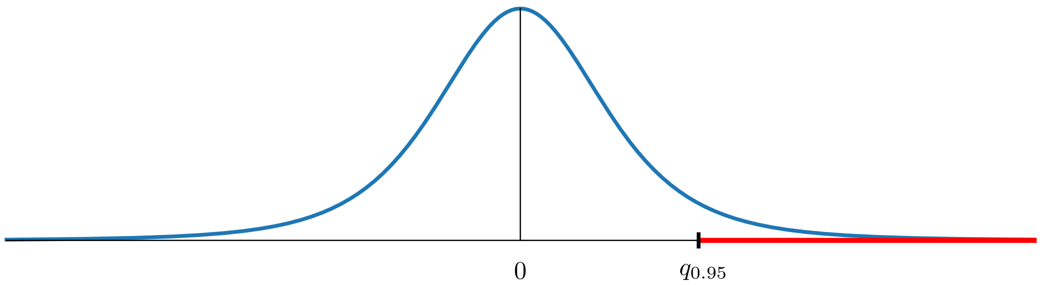
Darum ist der Wilcoxon-Test dem t-Test oft vorzuziehen.

Lösung 10.8

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.

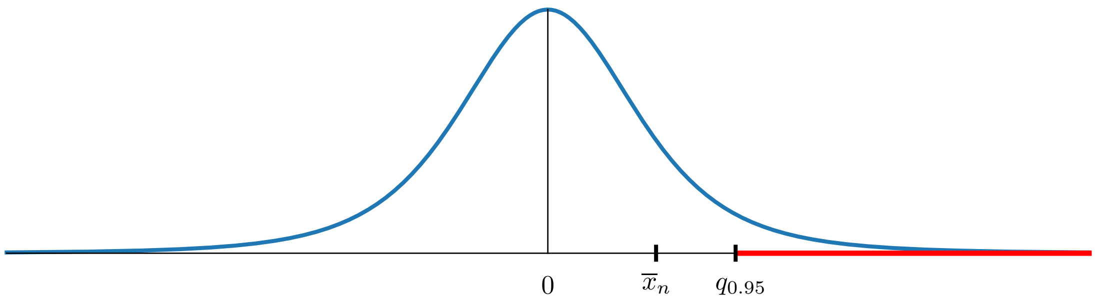
Wir zeichnen eine t-Verteilung mit $\mu = 0$ und $\sigma = 1$ und Freiheitsgrad (degree of freedom) 4 (Annahmen, die nichts zur Sache tun) und mit dem Verwerfungsbereich (Test nach oben).



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

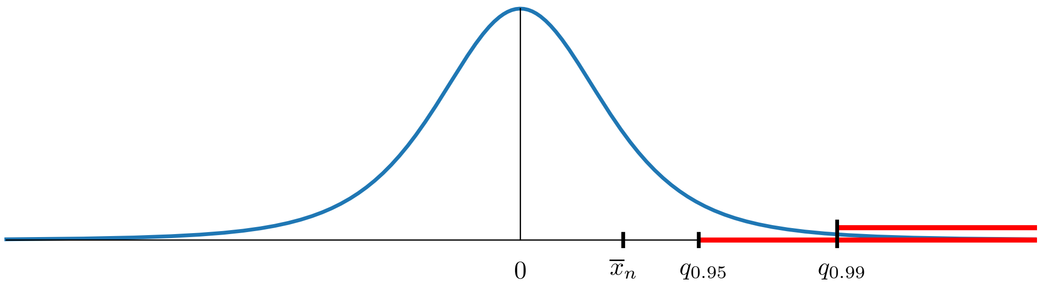
Was heisst das?

- 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.



a) *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).



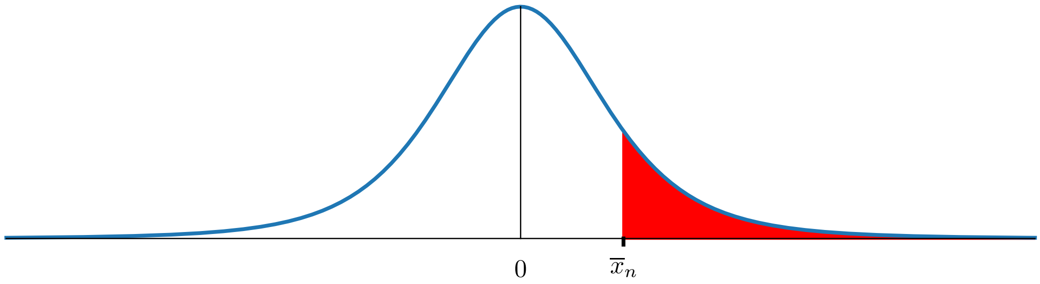
Also wird für alle $\alpha < 0.05$ die Nullhypothese nicht verworfen, da $\overline{x}_n$ nie im Verwerfungsbereich liegen kann. Also ist die Aussage richtig.

b) *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.

c) *Der p-Wert ist strikt kleiner als 0.5.*

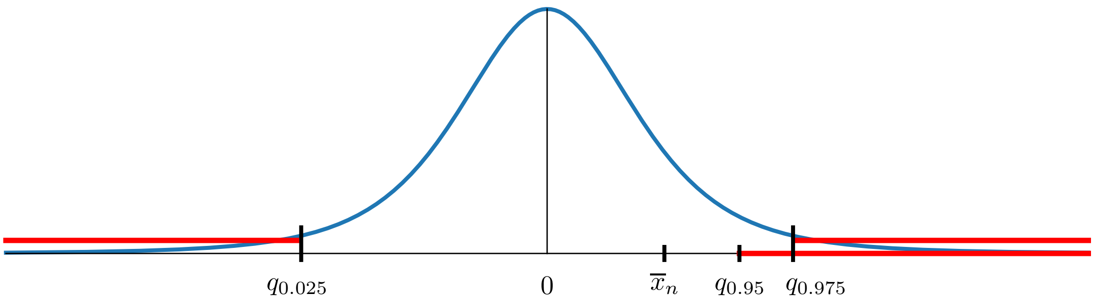
Nun was ist der pp-Wert? Das ist die Wahrscheinlichkeit einen bestimmten Wert (hier $\overline{x}_n$) oder einen extremeren zu beobachten in Richtung der Alternativhypothese. Wahrscheinlichkeiten können wir bei stetigen Verteilungen als Flächen darstellen und dies ist die Fläche unter der Kurve rechts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

d) *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.



e) *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.