

Hypothesentest t-Test für abhängige Stichproben

Code ▼

1) Grundlegende Konzepte: Was ist ein t-Test für abhängige Stichproben?

Von “abhängigen Stichproben” wird gesprochen, wenn der Messwert und ein bestimmter anderer Messwert sich gegenseitig beeinflussen. Folgende Szenarien eignen sich für eine verbundene Stichprobe:

Messwiederholung: Die Messwerte stammen von der gleichen Person z. B. Messzeitpunkt #1 verglichen mit Messzeitpunkt #2.

Natürliche Paare: Die Messwerte stammen von verschiedenen Personen, die aber zusammen gehören: Ehefrau – Ehemann, Psychologe – Patient oder Zwillinge.

Matching: Die Messwerte stammen ebenfalls von verschiedenen Personen, die einander zugeordnet wurden. Aufgrund eines vergleichbaren Werts (Drittvariablen) werden Matching-Paare gebildet.

Wichtig: Es werden genau zwei Gruppen verglichen.

Die Zentrale Frage: Unterscheiden sich die Mittelwerte von zwei abhängigen Stichproben?

Um zu prüfen, ob die Mittelwertsunterschiede statistisch signifikant sind, muss die zugehörige Teststatistik berechnet werden. Die Verteilung der Teststatistik t folgt einer theoretischen t -Verteilung, deren Form sich in Abhängigkeit der Freiheitsgrade unterscheidet. Die dem Test zu Grunde liegende t -Verteilung gibt dem Test den Namen t -Test.

Mit Hilfe der Effektstärke wird zum Schluss die praktische Bedeutsamkeit des ermittelten Effekts beurteilt.

2) Aufgabenstellung - Beschreibung

Bei einer Gruppe von 25 teilnehmenden erwachsenen Geschwisterpaaren (Brüdern), also natürlichen Paaren, die aus 25 zufällig ausgewählten Familien stammen, wurden die Länge und Breite des Kopfes gemessen. Für die Studie wurden die Kopflängen verglichen. Alle Messangaben erfolgten in mm. Es soll untersucht, ob es einen Unterschied in der Kopflänge Mittelwert zwischen Brüdern gibt.

3) Hypothese

H_0 : Es gibt keinen Unterschied in der Länge des Kopfes von Brüdern.

$$M_{b1} = M_{b2}$$

H_1 : Es gibt einen Unterschied in der Länge des Kopfes von Brüdern.

$$M_{b1} \neq M_{b2}$$

Variable 1 = Kopflänge (mm) AV

Variable 2 = Brudergruppe(bruder_1/bruder_2) UV

4) Voraussetzungen

Der verbundene t-Test vergleicht zwei Gruppen. Die Abhängigkeit muss vorher gegeben sein. In unserer Studie liegen zwei verbundene Stichproben oder Gruppen vor, aber die verschiedenen Messwertpaare sind voneinander unabhängig: - Die beiden Gruppen (Bruder 1 und Bruder 2) sind verbundenen durch die Verwandtschaft (Brüder) - Die Paare innerhalb der Gruppen sind jedoch unabhängig, weil die Paare (Bruder 1 und Bruder 2) aus verschiedenen, zufällig gewählten Familien stammen.

Die abhängige Variable ist mind. intervallskaliert -> Ja, Längenmaße der Kopflänge sind metrisch.

Unsere unabhängige Variable muss kategorial sein, daher nominalskaliert, mit zwei Ausprägungen. Die beiden Ausprägungen beziehen sich auf die beiden Gruppen, die wir vergleichen, und sind in diesem Fall jeweils 'Bruder 1' und 'Bruder 2'.

Die Unterschiede zwischen den verbundenen Testwerten sind in der Grundgesamtheit normalverteilt (bei Stichproben > 30 sind Verletzungen unproblematisch). Siehe Deskriptive Statistiken und Korrelation: Die Unterschiede zwischen den verbundenen Testwerten sind in der Stichprobe normalverteilt.

Datensatz:

Hide

```
path = "C:/Data_Science_Projects/Repository_2/Aufgabe5/koepfebrueder.csv"
koepfebrueder <- read.csv(path)
View(koepfebrueder)
```

5) Deskriptive Statistiken und Korrelation

Hide

```
head(koepfebrueder)
```

	X <int>	I1 <int>	b1 <int>	I2 <int>	b2 <int>
1	1	191	155	179	145
2	2	195	149	201	152
3	3	181	148	185	149
4	4	183	153	188	149
5	5	176	144	171	142
6	6	208	157	192	152
6 rows					

Der zu analysierende Datensatz enthält neben einer Paar-Probandennummer (ID) die vier Messungen der Kopflänge und -Breite von jeweils zwei Brüdern. Daraus wurde für die Auswertung jeweils die Differenz der Kopflänge von Bruder 1 und Bruder 2 gebildet. Die Spalten für die Kopfbreite werden nicht betrachtet.

Es wird die Längendifferenz gebildet und der Tabelle hinzugefügt, die Spalten für die Kopfbreite werden der besseren Übersicht wegen gelöscht:

Hide

```
koepfebrueder <- koepfebrueder[c(1,2,4)]
head(koepfebrueder)
```

	X <int>	I1 <int>	I2 <int>
1	1	191	179
2	2	195	201
3	3	181	185
4	4	183	188
5	5	176	171
6	6	208	192

6 rows

Hide

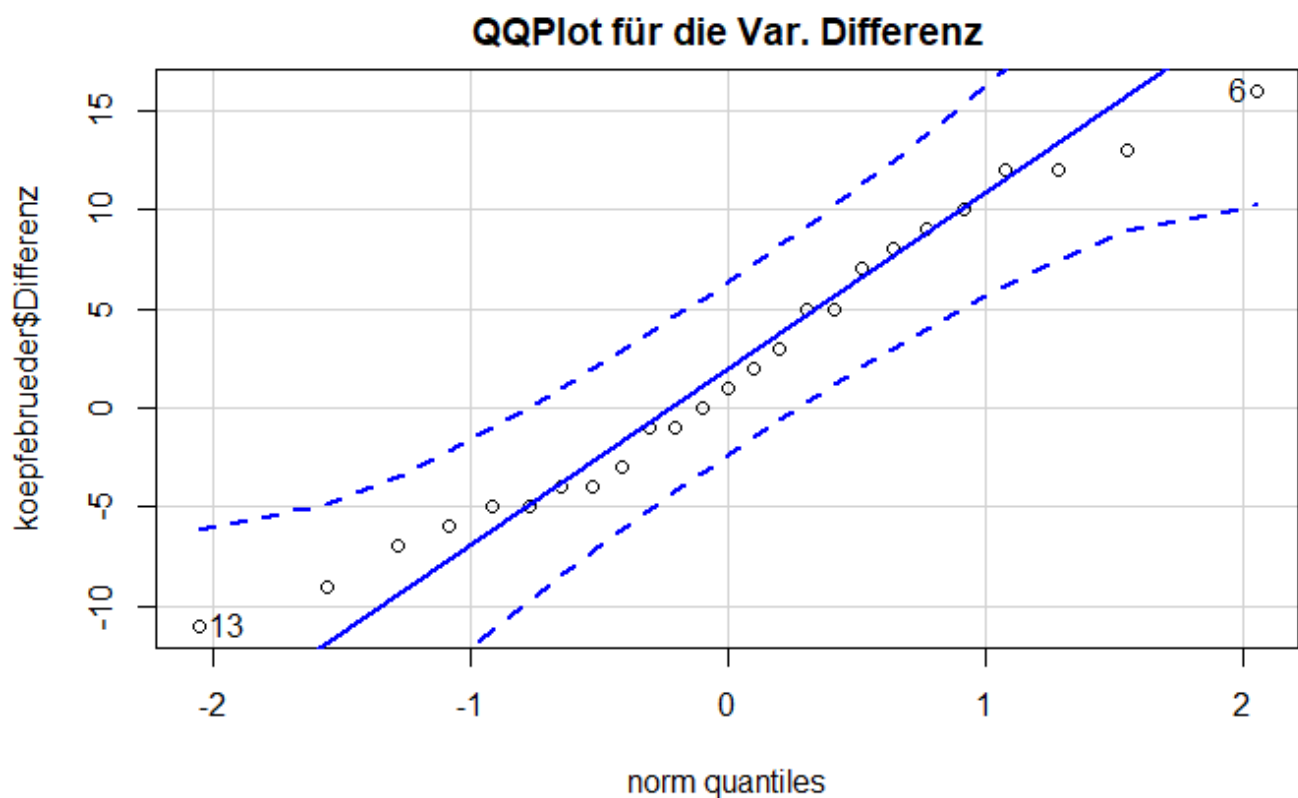
```
# Differenz-Spalte erstellen
zwischen <- koepfebrueder$I1 - koepfebrueder$I2
koepfebrueder <- cbind(koepfebrueder, "Differenz" = zwischen)
View(koepfebrueder)
```

Prüfung der Normalverteilung mittels QQ-Plot

Hide

```
qqPlot(koepfebrueder$Differenz, main = "QQPlot für die Var. Differenz")
```

```
[1] 6 13
```

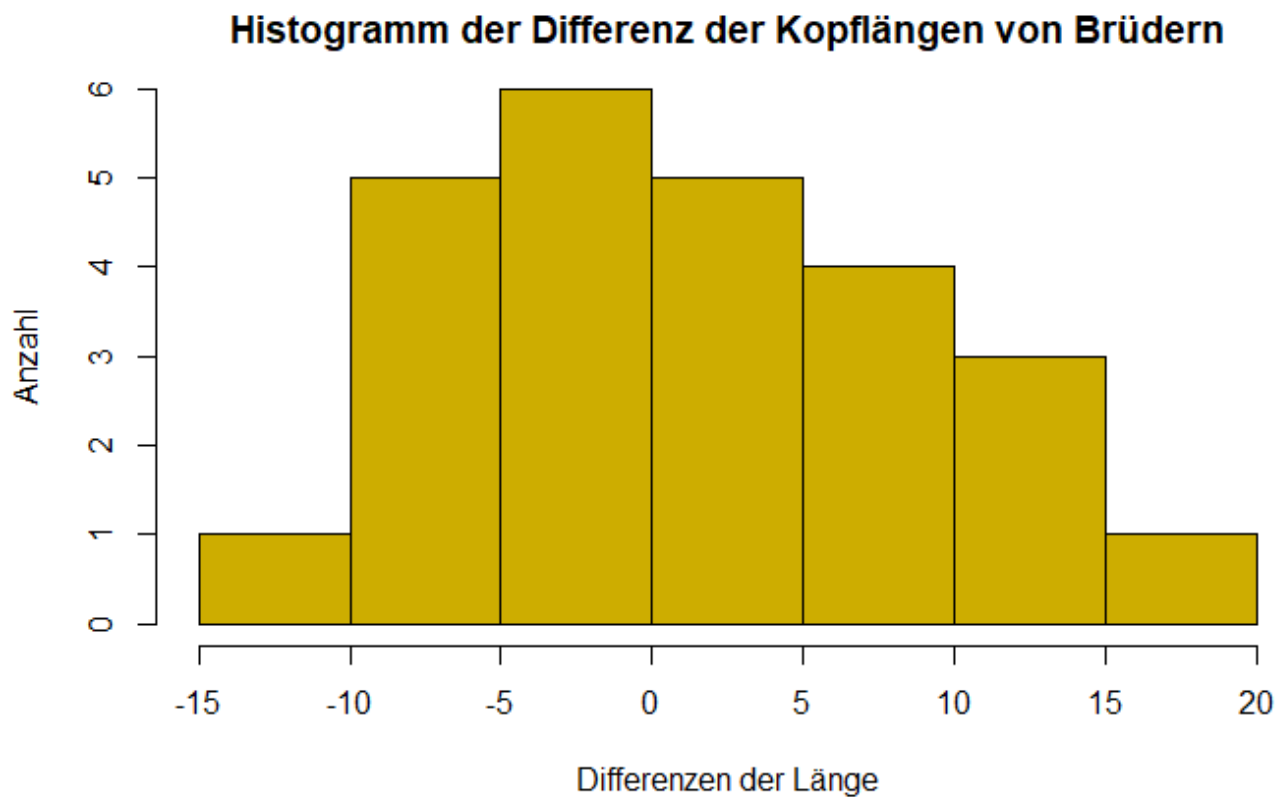


Der QQ-Plot zeigt eine Normalverteilung der Differenzen der Kopflängen an, da sich die meisten Datenpunkte nahe der durchgezogenen Linie befinden.

Prüfung der Normalverteilung mittels Histogramm

Hide

```
hist(koepfebrueder$Differenz, main = "Histogramm der Differenz der Kopflängen von Brüdern", xlab = "Differenzen der Länge", ylab = "Anzahl", col = "gold3")
```

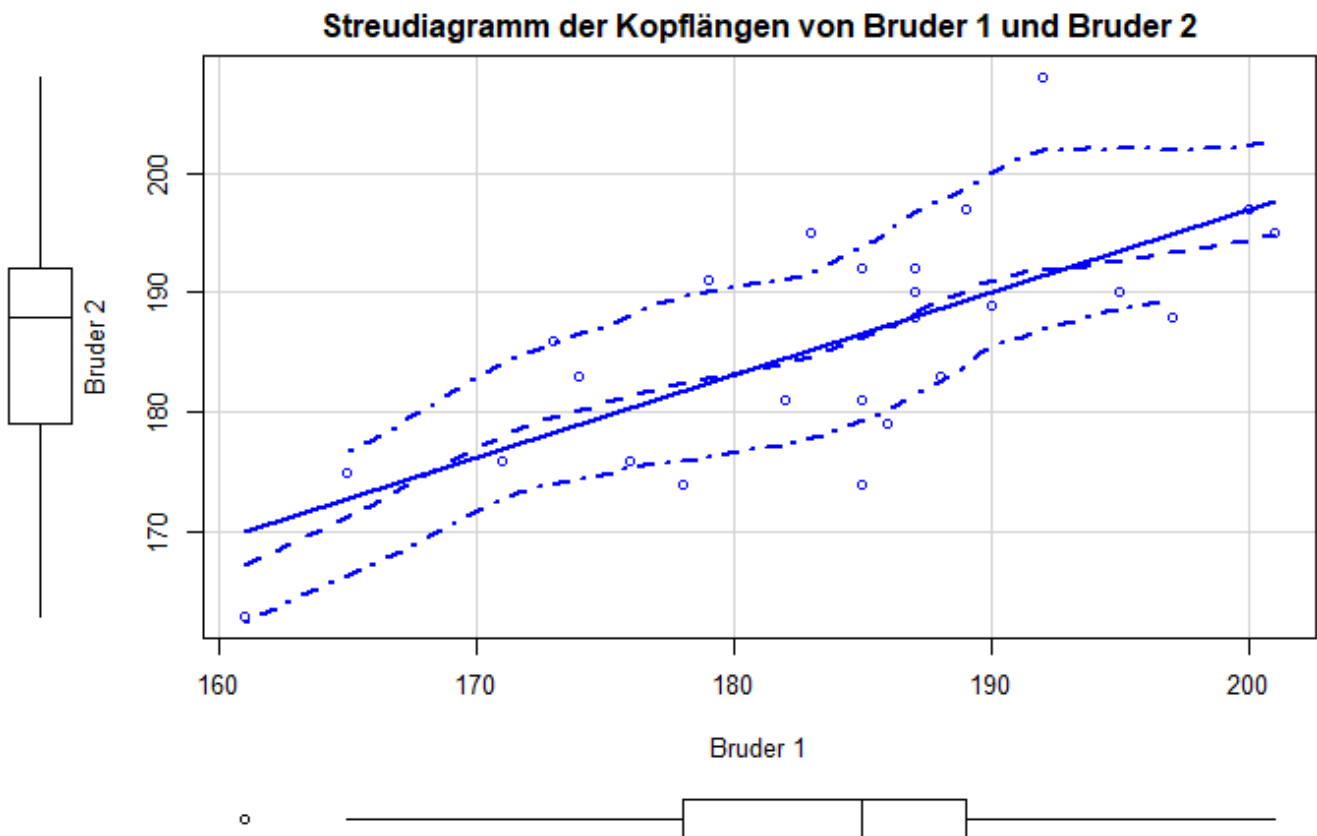


Die Differenzen der Kopflängen der Brüder zeigen, wie im Histogramm zu sehen ist, annähernd eine Normalverteilung.

Korrelation

Hide

```
#library(car)
scatterplot(koepfebrueder$l1 ~ koepfebrueder$l2 , main = "Streudiagramm der Kopflängen von Bruder 1 und Bruder 2", xlab = "Bruder 1", ylab= "Bruder 2")
```



Die Verteilung der Kopflängen von Bruder 1 und Bruder 2 im Streudiagramm zeigt eine positive lineare Korrelation der Kopflängen.

Hide

```
test <- cor.test(koepfebrueder$l1, koepfebrueder$l2)
test
```

Pearson's product-moment correlation

```
data: koepfebrueder$l1 and koepfebrueder$l2
t = 4.8457, df = 23, p-value = 6.842e-05
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.4388744 0.8634043
sample estimates:
      cor
0.7107518
```

Die 'Kopflänge von Bruder 1' (l1) und die 'Kopflänge von Bruder 2' (l2) korrelieren signifikant positiv linear ($r = 0.7107518$, $p = 6.842e-05 < 0.05$, $n = 25$).

Bei natürlichen Paaren ist es möglich, dass die Daten der ersten und zweiten Gruppe, also eines Messwertpaares, miteinander korrelieren. Es ist plausibel, dass die Ergebnisse der Messungen der Paare sich ähnlich sind und dass innerhalb eines Messwertpaares eher geringere Unterschiede auftreten als zwischen den Paaren. Im R-Output wird daher eine Pearson Korrelation der beiden Messgruppen ausgegeben. Für das Beispiel ergibt sich eine starke Korrelation von $r = 0.7107518$ ($p = 6.842e-05 < 0.05$, $n = 25$), je länger der Kopf von Bruder 1, um so länger auch der Kopf von Bruder 2 und umgekehrt.

Deskriptive Statistiken

Hide

```
library(psych)
```

Attache Paket: `psych`

The following object is masked from `package:effsize`:

`cohen.d`

The following object is masked from `package:car`:

`logit`

The following objects are masked from `package:ggplot2`:

`%+%`, `alpha`

Hide

```
describe_koepfebrueder <- describe(koepfebrueder)
describe_koepfebrueder
```

	vars <dbl>	n <dbl>	mean <dbl>	sd <dbl>	median <dbl>	trimmed <dbl>	mad <dbl>	min <dbl>	max <dbl>
X	1	25	13.00	7.36	13	13.00	8.90	1	25
I1	2	25	185.72	9.76	188	185.76	10.38	163	208
I2	3	25	183.84	10.04	185	184.24	8.90	161	201
Differenz	4	25	1.88	7.54	1	1.81	8.90	-11	16

4 rows | 1-10 of 13 columns

Es zeigt sich, dass es einen leichten Mittelwertsunterschied zwischen den Längen der Köpfe von Brüdern gibt: I1-mean = 185.72 mm (SD = 9.76 mm, n = 25), I2-mean = 183.84 mm (SD = 10.04 mm, n = 25). Allerdings ist der Unterschied mit 1.88 mm (I1-mean > I2-mean) sehr klein und die Abweichung liegt innerhalb der Standardabweichungen. Die Mediane unterscheiden sich auch leicht, (I1-med = 188 mm > I2-med = 185 mm). Sie liegen nahe bei den Mittelwerten (innerhalb einer Standardabweichung). Somit dürften die Mittelwertsunterschiede nicht signifikant sein, was im Folgenden überprüft wird.

6) Ergebnisse des t-Tests für abhängige Stichproben

Hide

```
ttest_koepfebrueder<- t.test(koepfebrueder$I1, koepfebrueder$I2, alternative = "two.sided", p
aired = TRUE, conf.level = 0.95)
```

```
ttest_koepfebrueder
```

Paired t-test

```
data: koepfebrueder$I1 and koepfebrueder$I2
t = 1.2475, df = 24, p-value = 0.2242
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -1.230307  4.990307
sample estimates:
mean of the differences
      1.88
```

Hier wird der ungerichtete t-test betrachtet, weil wir nur wissen wollen, ob ein Unterschied besteht oder nicht (daher Unterschiedshypothese). Die Richtung ist daher in dieser Betrachtung nicht relevant.

t-kritisch

Hide

```
qt(0.975, df=24)
```

```
[1] 2.063899
```

Die Teststatistik beträgt $t(24) = 1.24754$ (t-empirisch (1.2475) ist kleiner als t-kritisch (2.063899)), und der zugehörige Signifikanzwert $p(\text{empirisch}) = 0.2242 > p(\text{kritisch}) = 0.05$. Damit ist der Unterschied nicht signifikant: Die Mittelwerte der beiden Messungen der Kopflängen von Bruder 1 und Bruder 2 (I1 und I2) unterscheiden sich somit nicht signifikant voneinander ($t(24) = 1.24754$, $p = 0.2242$, $n = 25$).

Hier ist schon ersichtlich, dass die H_0 angenommen werden muss. Für die Vollständigkeit der statistischen Untersuchungen wird im Folgenden noch die Effektstärke betrachtet.

7) Berechnung der Effektstärke

Cohen und Pearson

$$r = \left\| \sqrt{\frac{t^2}{t^2 + df}} \right\|$$

Hide

```
eff_fret <- abs(sqrt(ttest_koepfebrueder$statistic^2 / (ttest_koepfebrueder$statistic^2 + tte
st_koepfebrueder$parameter)))
```

```
sprintf("Effektstärke: %.4f", eff_fret)
```

```
[1] "Effektstärke: 0.2468"
```

Zur Beurteilung der Größe des Effektes dient die Einteilung von Cohen (1992):

Schwacher Effekt: $0.10 < ||r|| < 0.30$

Schwacher bis mittlerer Effekt: $0.30 = ||r||$

Mittlerer Effekt: $0.30 < ||r|| < 0.50$

Mittlerer bis starker Effekt: $0.50 = ||r||$

Starker Effekt: $0.50 < ||r||$

Die ermittelte Effektstärke von 0.25 entspricht einem schwachen Effekt, es liegt also ein schwacher Effekt vor, der wahrscheinlich dem Zufall geschuldet ist.

Alternative Berechnung mit Hedges \hat{g}

$$\text{Hedges } \hat{g} = \left| \frac{\hat{\mu}_1 - \hat{\mu}_2}{\hat{\sigma}_d} \right|$$

Hide

```
diff <- ttest_koepfebrueder$estimate  
  
sed <- sd(koepfebrueder$I1 - koepfebrueder$I2)  
  
g <- diff/sed  
  
sprintf("Effektstärke: %.4f",g)
```

```
[1] "Effektstärke: 0.2495"
```

Zur Beurteilung der Größe des Effektes:

Schwacher Effekt: $0.20 < ||g|| < 0.50$
Schwacher bis mittlerer Effekt: $0.50 = ||g||$
Mittlerer Effekt: $0.50 < ||g|| < 0.80$
Mittlerer bis starker Effekt: $0.80 = ||g||$
Starker Effekt: $0.80 < ||g||$

Dasselbe Ergebnis: Eine Effektstärke von 0.25 entspricht einem schwachen Effekt, es liegt also ein schwacher Effekt vor.

Kommentar zum Ergebnis der schwachen Effektstärke: Da die Mittelwertsunterschiede bereits als nicht signifikant ermittelt wurden, bietet die Berechnung der Effektstärke keinen Mehrwert, denn das Ergebnis ist eventuell zufällig. Das Ergebnis eines Schwachen Effekts kann nicht auf die Grundgesamtheit übertragen werden. Die H1 wird also trotzdem verworfen und die H0 angenommen.

8) Eine Aussage

Es zeigt sich, dass die Kopflängen von Brüdern sich statistisch nicht signifikant unterscheiden ($t(24) = 1.24754$, $p = 0.2242$, $n = 25$). Der schwache Mittelwertsunterschied der Gruppen 'Kopflänge von Bruder 1' (I1) und 'Kopflänge von Bruder 2' (I2) ist sehr klein (1.88 mm): $M(I1) = 185.72$ mm, $SD = 9.76$ mm, $n = 25$; $M(I2) = 183.84$ mm, $SD = 10.04$ mm, $n = 25$.

Die der Vollständigkeit halber ermittelte Effektstärke nach Cohen (1992) liegt bei $r = 0.25$ und entspricht damit einem schwachen Effekt, der jedoch eher dem Zufall geschuldet ist.

H1 wird dementsprechend verworfen, H0 wird angenommen: Die Kopflängen von Brüdern unterscheiden sich nicht.