

Die Bestimmung der Zentrale Tendenz mittels t-Tests

Katharina Maubach

Table of contents

1	t-Test für unabhängige Stichproben	3
1.1	Datenmanagement	3
1.2	Deskriptives der abhängigen Variablen	4
1.3	Voraussetzungsprüfung	5
1.3.1	Skalierung der UV und AV	5
1.3.2	Normalverteilung der AV	5
1.3.3	Varianzhomogenität	6
1.4	Durchführung des t-Tests	6
1.5	Interpretation des Outputs	7
1.6	Nicht-parametrischer Test: Mann-Whitney-U & Wilcoxon-Rangsummen-Test	8
2	t-Test für verbundene Stichproben	8
2.1	Durchführung des t-tests für verbundene Stichproben	8
2.2	Nicht-parametrischer Test: Wilcoxon-Vorzeichen-Rangtest	8
3	t-Test für Einstichproben	9
3.1	Data Management	9
3.2	Voraussetzungsprüfung	9
3.3	Durchführung des t-tests für Einstichproben	10
3.4	Nicht-parametrischer Test: one sample Wilcoxon Test	11

Wir nutzen den t-test um Mittelwertunterschiede zwischen zwei Gruppen zu analysieren.

Video

<https://nc.uni-bremen.de/index.php/s/T7brcJ73tLGpCk7/download/%2311%20t-Test.mp4>

Wir unterscheiden grundsätzlich in zwei verschiedene Arten von t-Test:



Figure 1: Bild generiert von Midjourney

- den t-Test für unabhängige Stichproben (two-paired-test) sowie
- den t-Test für verbundene Stichproben (paired test)

t-test für unabhängige Stichproben

Mit dem t-test für unabhängige Stichproben überprüfen wir, ob sich die Mittelwerte in zwei gemessenen Stichproben unterscheiden. Die Stichproben sind dann unabhängig, wenn sie aus zwei unterschiedlichen Gruppen stammen und daher sich nicht gegenseitig beeinflussen können. Ein solches Beispiel liegt zum Beispiel dann vor, wenn wir schauen ob das Alter (dichotom kodiert in alt vs jung) einen Einfluss auf das Einkommen hat. In diesem Fall kann eine Person zum Untersuchungszeitpunkt nicht gleichzeitig alt und jung sein, sondern fällt in einer der beiden Gruppen. Ein anderes typisches Anwendungsbeispiel ist der Vergleich von zwei Gruppen im Rahmen eines Experimentaldesign. Hier kann ein Proband ebenfalls nur einer der Untersuchungsgruppe zugeordnet sein.

t-test für verbundene Stichproben

Der t-test für verbundene Stichproben wiederum misst, inwiefern sich Mittelwerte bei denselben Personen oder Fällen verändern. Wenn wir bei unserem Beispiel mit dem Einfluss des Alters auf das Einkommen bleiben, so könnten wir dieselben Personen zu zwei Zeitpunkten befragen: einmal wenn sie jung sind und dann zu einem späteren Zeitpunkt noch einmal. In diesem Fall sind unsere Stichproben nicht unabhängig, da die Messwerte von denselben Personen stammen und ggf. der frühere Messzeitpunkt den späteren beeinflussen kann.

Der Einstichprobentest

Zusätzlich gibt es noch den Einstichprobentest oder auch one-sample-test. Hier vergleichen wir nicht zwei Gruppenmittelwerte miteinander, sondern den Mittelwert einer Gruppe mit einem bereits bekannten feststehenden Wert. Wir können so beispielsweise überprüfen, ob das Einkommen in einer Stichprobe dem deutschlandweiten Mittelwert entspricht, sofern dieser uns vorab bekannt ist.

1 t-Test für unabhängige Stichproben

Wir schauen uns zunächst den t-Test für unabhängige Stichproben an.

1.1 Datenmanagement

Wir beginnen mit dem Laden der notwendigen Pakete.

```
if(!require("pacman")) {install.packages("pacman");library(pacman)}  
p_load(tidyverse, ggplot2, haven, mosaic, knitr, effectsize, car, broom) ①  
theme_set(theme_classic()) ②
```

- ① Mittels p_load laden wir alle benötigten Pakete gleichzeitig.
- ② Wir legen allgemein den Hintergrund theme_classic fest.

Anschließend laden wir die Daten aus dem Allbus:

```
daten = haven::read_dta("Datensatz/Allbus_2021.dta")
```

Um mit dem Datensatz zu arbeiten benötigen wir einige grundlegende Schritte des Datenmanagements (für ausführliche Erklärungen siehe [hier](#)). Für unseren t-test für unabhängige Stichproben möchten wir uns anschauen, wie sich der Wohnort der Befragten auf ihr Vertrauen in die Bundesregierung auswirkt. Wir nutzen dazu die folgenden Variablen:

Variable	Beschreibung	Ausprägungen
eastwest	Erhebungsgebiet West-Ost	1= Alte Bundesländer (Westen) 2 = Neue Bundesländer (Osten)

Variable	Beschreibung	Ausprägungen
pt12	Vertrauen in die Bundesregierung	-42 = Datenfehler -11 = TNZ Split -9 = Keine Angabe 1 = Gar kein Vertrauen ... 7 = Großes Vertrauen

Innerhalb unseres Datenmanagements schließen wir fehlerhafte und fehlende Werte der Variablen pt12 aus und benennen diese anschließend um:

```
daten <- daten %>%
  filter(., pt12 > 0) %>% #<1>
  rename(., trustreg = pt12) %>% #<2>
  mutate(trustreg = as.numeric(trustreg), #<3>
         eastwest = haven::as_factor(eastwest)) #<3>
```

- ① Wir schließen fehlende Werte (siehe Tabelle oben aus). Da diese alle negativ sind, können wir einfach alle Werte kleiner als 0 ausschließen.
- ② Hier benennen wir die Variablen anders um unsere weitere Arbeit zu vereinfachen.
- ③ Wir bringen unsere Daten in das korrekte Format. Für den t-Test muss die abhängige Variable als Vektor vorliegen und die unabhängige Variable als Faktor.

1.2 Deskriptives der abhängigen Variablen

Als ersten Schritt schauen wir uns die deskriptiven Statistiken unserer abhängigen Variablen Vertrauen in die Regierung für die beiden Faktorstufen unserer unabhängigen Variablen an.

```
des_stat <- (favstats(trustreg ~ eastwest, data = daten))
rownames(des_stat) <- c("Alte Bundesländer", "Neue Bundesländer")
kable(des_stat[,2:10], digits = 2)
```

	min	Q1	median	Q3	max	mean	sd	n	missing
Alte Bundesländer	1	3	4	5	7	4.15	1.58	2335	0
Neue Bundesländer	1	3	4	5	7	3.86	1.71	1183	0

Anhand dieser deskriptiven Daten sehen wir, dass das Vertrauen in die Bundesregierung in den alten Bundesländern im Mittelwert mit 4,15 stärker ausgeprägt ist als in den neuen

Bundesländern (Mittelwert von 3,86. Um zu prüfen, ob dieser Unterschied des Mittelwertes statistisch signifikant ist, berechnen wir in einem zweiten Schritt die Teststatistik des t-tests. Zunächst müssen wir jedoch prüfen, ob unsere Daten für die Berechnung geeignet sind. Dazu führen wir eine Voraussetzungsprüfung durch.

1.3 Voraussetzungsprüfung

Bei dem t-test handelt es sich um ein statistisches Verfahren. Um die Güte der Ergebnisse sicherzustellen müssen dabei einige Voraussetzungen beachtet werden. Dies sind im Überblick:

- die korrekte Skalierung der UV (=Faktor) und AV (=numerisch)
- die Normalverteilung der abhängigen Variablen
- die Varianzhomogenität

1.3.1 Skalierung der UV und AV

Die abhängige Variable muss als numerischer Vektor vorliegen, die unabhängige Variable wiederum als Faktor (mit zwei Ausprägungen). Die korrekte Skalierung der Variablen haben wir bereits innerhalb des Data Managements sichergestellt, können uns aber zusätzlich mit dem Befehl `class` noch einmal einen Überblick über das Skalenniveau unserer Variablen verschaffen.

```
daten %>%  
  select(eastwest, trustreg) %>%  
  lapply(class)
```

```
$eastwest  
[1] "factor"
```

```
$trustreg  
[1] "numeric"
```

1.3.2 Normalverteilung der AV

```
daten %>%  
  group_by(eastwest) %>%  
  do(tidy(shapiro.test(. $trustreg)))
```

```
# A tibble: 2 x 4
# Groups:   eastwest [2]
  eastwest      statistic p.value method
  <fct>          <dbl>   <dbl> <chr>
1 ALTE BUNDESLAENDER 0.937 2.69e-30 Shapiro-Wilk normality test
2 NEUE BUNDESLAENDER 0.930 5.06e-23 Shapiro-Wilk normality test
```

1.3.3 Varianzhomogenität

```
daten %>%
  drop_na(eastwest, trustreg) %>%
  mutate(eastwest = as.factor(eastwest)) %>%
  leveneTest(trustreg~eastwest, data = .)
```

Levene's Test for Homogeneity of Variance (center = median)

```
      Df F value    Pr(>F)
group   1  9.5219 0.002046 **
      3516
```

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

1.4 Durchführung des t-Tests

```
daten %>%
  drop_na(eastwest, trustreg) %>%
  filter(.,trustreg >0) %>%
  t.test(trustreg ~ eastwest,
         alternative = "two.sided",
         var.equal = T,
         data = .)
```

Two Sample t-test

data: trustreg by eastwest

t = 4.9349, df = 3516, p-value = 8.391e-07

alternative hypothesis: true difference in means between group ALTE BUNDESLAENDER and group NEUE BUNDESLAENDER

95 percent confidence interval:

0.1726028 0.4001616

sample estimates:

mean in group ALTE BUNDESLAENDER	mean in group NEUE BUNDESLAENDER
4.147752	3.861369

Effektstärke cohens d berechnen

```
daten %>%  
  drop_na(eastwest, trustreg) %>%  
  cohens_d(trustreg~eastwest, data = .)
```

Cohen's d	95% CI
0.18	[0.11, 0.25]

- Estimated using pooled SD.

1.5 Interpretation des Outputs

Der Output des t-tests gibt uns die folgenden Informationen aus:

- data
- t
- df
- p-value
- 95% confidence intervall
- means in group 1 & 2

💡 Wie gebe ich die Ergebnisse korrekt an?

Die Ergebnisse des t-tests werden üblicherweise im Text angegeben. Für diese Angabe werden die folgenden Informationen benötigt:

- die Mittelwerte und Standardabweichung der einzelnen Vergleichsgruppen
- der df-Wert
- der F-Wert
- der p-Wert
- die Effektgröße

Das Format ist üblicherweise:

Beispiel: Personen mit einem sehr guten Gesundheitszustand haben durchschnittlich ein höheres Vertrauen in das Gesundheitswesen ($M = 5.13; SD = 1.45$) als Personen mit einem guten ($M = 4.99; SD = 1.32$), zufriedenstellenden ($M = 4.85; SD = 1.37$), weniger guten ($M = 4.82; SD = 1.48$) oder

schlechtem ($M = 4.71; SD = 1.55$) Gesundheitszustand. Der Gesundheitszustand hat dabei einen signifikanten Einfluss auf das Vertrauen in das Gesundheitswesen ($F(3,470) = 5.46, p < 0.001$). Die Effektstärke nach Cohen (1992) liegt bei $\alpha = 0.006$ und entspricht einem kleinen Effekt. Post-Hoc Paarvergleiche mit Tamhane's ergaben, dass sich der Mittelwert für die Personen mit sehr guten Gesundheitszustand signifikant von den Personen mit zufriedenstellendem ($p < 0.029$) und weniger gutem ($p < 0.0174$) Zustand unterscheidet. Die anderen Gesundheitsgruppen unterscheiden sich hingegen nicht signifikant voneinander.

1.6 Nicht-parametrischer Test: Mann-Whitney-U & Wilcoxon-Rangsummen-Test

Wilcoxon-Rangsummentest [Historie der Testverfahren](#)

2 t-Test für verbundene Stichproben

Voraussetzungsprüfung

2.1 Durchführung des t-tests für verbundene Stichproben

2.2 Nicht-parametrischer Test: Wilcoxon-Vorzeichen-Rangtest

```
pairwise.wilcox.test(daten$trustges, daten$gesund,  
                     p.adjust.method = "BH")
```

Pairwise comparisons using

data: daten\$trustges and daten\$gesund

<0 x 0 matrix>

P value adjustment method: BH

3 t-Test für Einstichproben

Beim t-Test für Einstichproben testen wir, ob die Werte unserer Stichprobe einem vorab festgelegten Wert entsprechen. So können wir beispielsweise mit Blick auf das Einkommen schauen, ob das monatliche Nettoeinkommen aus unserer Stichprobe dem Bundesdurchschnitt gemäß dem [statistischen Bundesamt](#) von 3813€ entspricht.

3.1 Data Management

Dazu benennen wir zunächst die entsprechende Variable aus den Allbus-Daten in `einkommen` um und filtern fehlende Werte hinaus.

```
daten_one <- daten %>%  
  filter(., di01a > 0) %>%  
  rename(., einkommen = di01a)
```

3.2 Voraussetzungsprüfung

Der Einstichprobentest setzt ähnlich wie der two-sided und paired test eine Normalverteilung der abhängigen Variablen voraus. Diese können wir erneut mit dem Shapiro-Wilk-Test überprüfen:

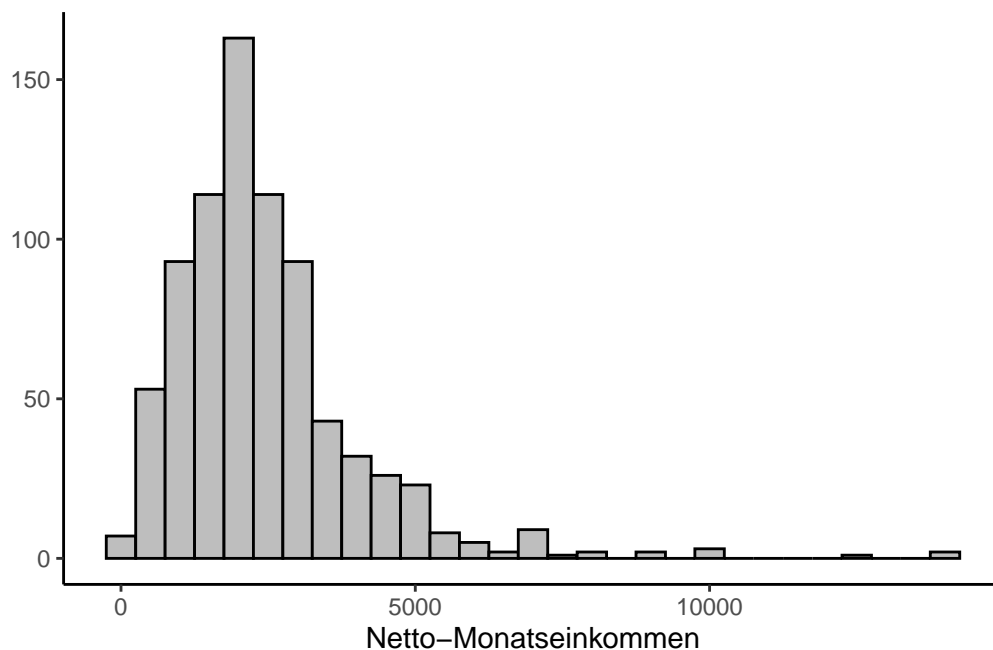
```
daten_one %>%  
  do(tidy(shapiro.test(.$einkommen)))
```

```
# A tibble: 1 x 3  
  statistic p.value method  
    <dbl>    <dbl> <chr>  
1      0.831 3.73e-28 Shapiro-Wilk normality test
```

Da unser Ergebnis signifikant ausfällt, müssen wir hier leider die Normalverteilung unserer AV ablehnen (wir testen ja auf signifikante Abweichung von der Normalverteilung, daher möchten wir bei diesem Test nicht ein signifikantes Ergebnis haben). Um weitere Gewissheit zu erlangen, prüfen wir die Normalverteilung zusätzlich optisch mit einem Histogramm:

```
ggplot(daten_one, aes(einkommen)) + #<1>  
  geom_histogram(aes(y = after_stat(count)), #<2>  
                 color = "black", fill = "grey",  
                 binwidth = 500) + #<3>  
  labs(x = "Netto-Monatseinkommen", #<4>  
       y = "")
```

- ① In R nutzen wir das Paket `ggplot2` von Wickham et al. um ein Histogramm auszugeben. Zunächst müssen wir hier das Paket `ggplot2` mit dem Befehl `ggplot()` aufrufen. Anschließend spezifizieren wir innerhalb der Klammer unseren Datensatz (hier `daten_one`) und unter `aes` unsere Variable (hier `einkommen`).
- ② Die Spezifizierungen innerhalb der Klammern unseres Histogramms geben an, dass dieses auf den Zahlen unseres Datensatzes beruhen soll (`aes(y = after_stat(count))`), wir die Außenumrandung schwarz `color = black` und die Füllfarbe grau wünschen (`fill = grey`). Diese Spezifikationen sind optional, sorgen jedoch für ein schöneres Aussehen unserer Grafik.
- ③ Mit `binwidth = 500` verweisen wir hier auf die Breite der Balken unseres Histogramms.
- ④ Für ein verschönertes Aussehen unseres Graphen nutzen wir den Befehl `labs` um zusätzlich die Achsen zu beschriften.



Die Grafik verstärkt die Einschätzung des Shapiro-Wilk-Tests, dass keine Normalverteilung vorliegt. In einem solchen Fall würden wir auf den nicht-parametrische one-sample Wilcoxon-Test ausweichen.

3.3 Durchführung des t-tests für Einstichproben

```
t.test(daten_one$einkommen , mu = 3813)
```

One Sample t-test

```
data: daten_one$einkommen
t = -23.889, df = 795, p-value < 2.2e-16
alternative hypothesis: true mean is not equal to 3813
95 percent confidence interval:
 2343.44 2566.61
sample estimates:
mean of x
 2455.025
```

3.4 Nicht-parametrischer Test: one sample Wilcoxon Test

```
wilcox.test(daten_one$einkommen, mu = 3813)
```

Wilcoxon signed rank test with continuity correction

```
data: daten_one$einkommen
V = 30680, p-value < 2.2e-16
alternative hypothesis: true location is not equal to 3813
```