

```

## Anteilstest
# Bei einem Anteilstest wird getestet, ob es plausibel ist, dass der
# Anteil von Objekten/Personen/... mit einer bestimmten Eigenschaft
# in der Gesamtheit einen bestimmten, vorgegebenen Wert hat (zweiseitiger
# Test) oder mindestens bzw. höchstens einen bestimmten, vorgegebenen
# Wert hat (einseitiger Test).
# Zur Durchführung eines Anteilstests in R muss eine Variable als
# zweiseitiger Faktor definiert sein; zudem muss die uns interessierende
# Eigenschaft der erste der beiden Werte der Variable sein. Als Beispiel
# wollen wir aufgrund des Datensatzes "daten" mit 10 Beobachtungen und
# der Variable "haarfarbe" mit den Werten "schwarz", "braun", "blond",
# "rot" in der angegebenen Reihenfolge auf dem Signifikanzniveau  $\alpha=0.05$ 
# testen, ob es plausibel ist, dass der Anteil von Personen mit brauner
# Haarfarbe in der Gesamtheit 60% beträgt (Nullhypothese).
# Dazu erzeugen wir mit Hilfe von "haarfarbe" zuerst eine neue Indikator-
# Variable "braunInd", welche angibt, ob jemand braune Haare hat:
daten$braunInd <- ifelse(
  daten$haarfarbe == "braun",
  1, # falls haarfarbe == "braun"
  2 # sonst
)
daten$braunInd <- factor( # "braunInd" als Faktor definieren
  daten$braunInd,
  labels = c("braun", "nicht braun")
)
table( # Kontrolle der Variable "braunInd": Kreuztabelle mit "haarfarbe"
  daten$braunInd,
  daten$haarfarbe
)

# Der Anteilstest kann aber nicht direkt auf der Variable "braunInd"
# durchgeführt werden. Stattdessen wird zuerst eine Tabelle erzeugt (und
# gespeichert),
# welche die absoluten Häufigkeiten der Werte von "braunInd" zählt. Mit
# dieser Tabelle führen wir dann zuerst einen (exakten) Binomialtest durch:
braunTab <- table( # Tabelle erzeugen (wird von R nicht angezeigt)
  daten$braunInd
)
binom.test( # Binomialtest durchführen
  braunTab,
  p = 0.6, # Anteil, auf welchen getestet werden soll
  alternative = "two.sided", # zweiseitiger Test
  conf.level = 0.95 # Konfidenzniveau  $\backslash(1-\alpha\backslash)$  für das Konf. intervall
  des Anteils
)
remove(braunTab) # aufräumen

# Der Binomialtest liefert für kleine Stichproben exakte p-Werte, kann
# aber für grosse Stichproben sehr rechenintensiv und zeitaufwendig werden
# (allerdings ist das mit den heutigen Rechnern üblicherweise kein grosses
# Problem mehr). Für sehr grosse Stichproben ist deshalb der
# Anteilstest prop.test(), basierend auf einer Näherung durch eine
# Normalverteilung, sinnvoll:

braunTab <- table( # Tabelle erzeugen (wird von R nicht angezeigt)
  daten$braunInd
)

```

```

prop.test( # Anteilstest mit Näherung durch Normalverteilung durchführen
braunTab,
p = 0.6, # Anteil, auf welchen getestet werden soll
alternative = "two.sided", # zweiseitiger Test
conf.level = 0.95 # Konfidenzniveau  $\alpha$  für das Konf.intervall
des Anteils
)
remove(braunTab) # aufräumen

## Einstichproben-t Test
# Wollen wir beispielsweise aufgrund der Variable "groesse" (Körpergrösse
in cm) im Datensatz "daten" auf dem Signifikanzniveau  $\alpha=0.1$ 
# testen, ob in der Gesamtheit der Mittelwert der Körpergrösse grösser als
170 cm ist (d.h. die Nullhypothese wäre "Mittelwert der
# Körpergrösse ist höchstens 170 cm"), geben wir das Folgende ein:
t.test(
  daten$groesse,
  alternative='greater', # rechtsseitiger Test
  mu=170, # postulierter Mittelwert  $\mu_0$  in der Gesamtheit
  conf.level=.9 # Konfidenzniveau  $\alpha$  für das Konf.intervall
des Mittelwerts
)

## Ungepaarter Zweistichproben-t-Test
# Wollen wir beispielsweise aufgrund der Variablen "geschlecht" (Faktor; 1
= "männlich", 2 = "weiblich") und "groesse" (Körpergrösse in cm)
# im Datensatz "daten" auf dem Signifikanzniveau  $\alpha=0.05$  testen, ob sich in
der Gesamtheit die Mittelwerte der Körpergrösse zwischen den
# beiden Geschlechtern unterscheiden (d.h. Nullhypothese wäre, dass sich
die Mittelwerte in der Gesamtheit nicht voneinander
# unterscheiden), geben wir Folgendes ein:
t.test(
  groesse~geschlecht, # Formeleingabe, zu interpretieren als "groesse in
  Abhängigkeit von geschlecht"
  alternative="two.sided", # zweiseitiger Test
  conf.level=.95, # Konfidenzniveau  $\alpha$  für das Konf.intervall
  der Differenz der Mittelwerte
  data=daten # damit R weiss, wo es die Variablen suchen soll
)

##Chi-Quadrat-Test
chisq.test(data.xls$Geschlecht, data.xls$Sportnote)

```