Einführung Inferenz metrische Werte

Karsten Lübke

Random Walk

Beim Glücksspiel ist es offensichtlich, aber auch an vielen, vielen anderen Stellen im Leben begegnen wir dem *Zufall*. Daten, Beobachtungen sind häufig Realisationen von sogenannten Zufallsvariablen. Das sind Variablen, deren Werte vom Zufall (und damit auch seinen Modellen und Gesetzen) abhängen. So werden Aktienkurse und -renditen häufig als Random Walk aufgefasst und modelliert - häufig unter der *Annahme* einer Normalverteilung.[1]

Zur Analyse wird wieder das Paket mosaic verwendet:

```
require(mosaic)
```

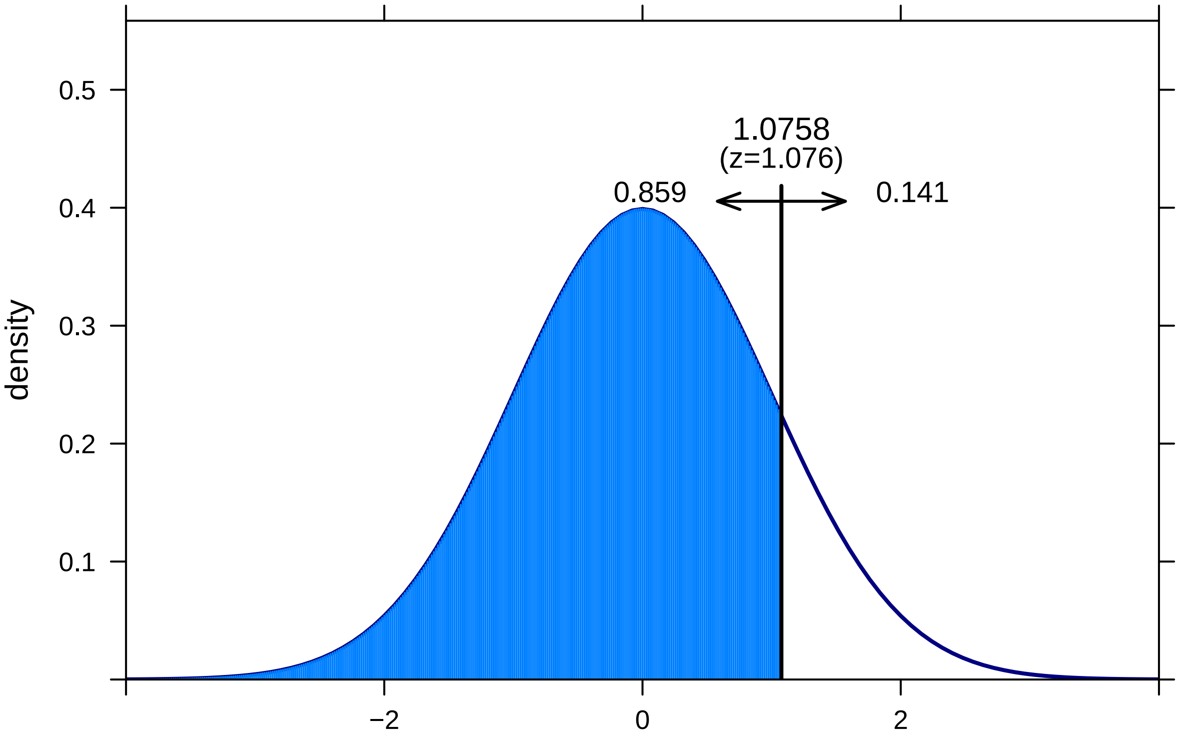
t-Test für eine Stichprobe

An den `n=251` Handelstagen des Jahres 2015 lag der arithmetische Mittelwert der (logarithmierten) Rendite der Facebook Aktie bei `0.11`, bei einer Standardabweichung (hier auch Volatilität genannt) von `1.62`. Zeigen diese Daten, dass die Rendite der Aktie nicht zufällig größer als Null ist, können wir aus dem Daten also auf die Alternativhypothese $H_A : \mu > 0$ schließen? Dabei gehen wir von eine unabhängigen, identischen Normalverteilung der Beobachtungen aus.

```
mFB <- 0.11 # Mittelwert
sdFB <- 1.62 # Standardabweichung
nFB <- 251 # Anzahl Beobachtungen
seFB <- sdFB / sqrt(nFB) # Standardfehler
z0 <- (mFB - 0) / seFB # z-Wert
xpnorm(z0, lower.tail = FALSE) # p-Wert
```

```
##
## If X ~ N(0, 1), then
##
##  P(X <= 1.075758) = P(Z <= 1.075758) = 0.8589822
##  P(X >  1.075758) = P(Z >  1.075758) = 0.1410178
```

[1] Sowohl die Annahme einer Normalverteilung, als auch die Annahme, dass die Renditen unabhängig voneinander sind (d. h. keine *Autokorrelation* vorliegt) und einer identischen Verteilung folgen (hier gleiche Varianz) sind in der Praxis kritisch zu hinterfragen.

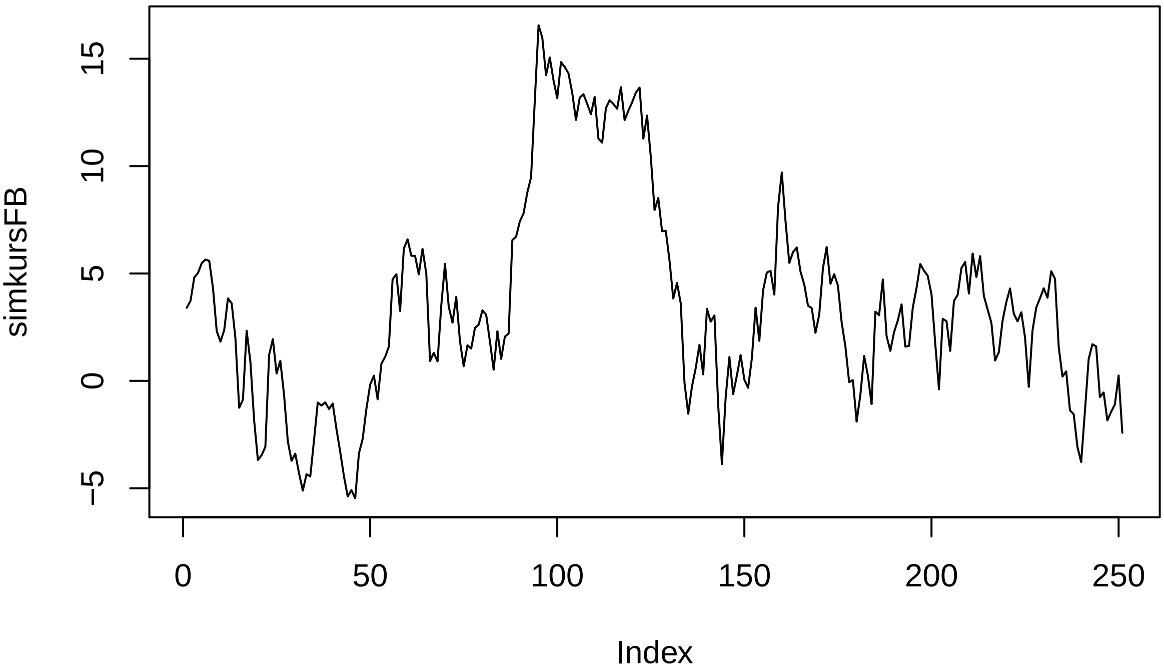



```
## [1] 0.1410178
```

Mit einem p-Wert von 0.141 kann die Nullhypothese, der Lageparameter der Rendite ist kleiner gleich 0 $H_0 : \mu \leq 0$ *nicht* verworfen werden.

Das zeigen auch Simulationen: Der Befehl `rnorm` erzeugt Zufallszahlen aus einer Normalverteilung:

```
set.seed(1896) # Zufallszahlengenerator setzen
simFB <- rnorm(n = nFB, mean = mFB, sd = sdFB) # Renditen simulieren
simkursFB <- cumsum(simFB) # Kursverlauf aus Renditen berechnen
plot(simkursFB, type="l") # Kurs plotten
```



Sieht doch *fast* realistisch aus, oder?

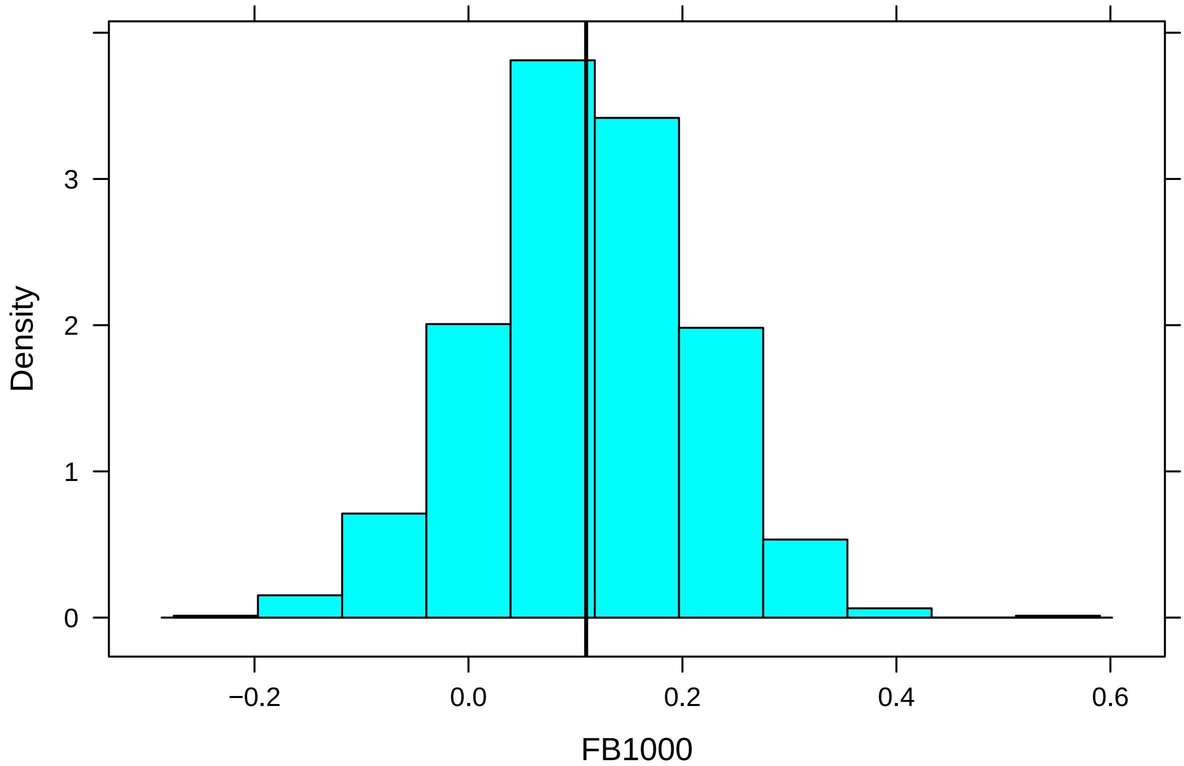
```r
mean(simFB)
```

```
## [1] -0.009637873
```

In der Simulation, für die wir die (Populations-) Parameter durch ihre (Stichproben) Punktschätzer ersetzt haben, haben wir einen Mittelwert der Rendite von -0.01 – und nicht wie (zufällig) beobachtet von 0.11.

Für Hypothesentests und Konfidenzintervalle reicht natürlich nicht eine Simulation:

```r
FB1000 <- do(1000) *  mean( rnorm(n = nFB, mean = mFB, sd = sdFB) )
histogram(~FB1000, v=mFB)
```



Das *simulierte* 95% Konfidenzintervall geht also von

```r
sort(FB1000$mean)[floor(1000*0.025)]
```

```
## [1] -0.08230827
```

bis

```r
sort(FB1000$mean)[ceiling(1000*0.975)]
```

```
## [1] 0.3073781
```

und enthält damit die 0. Von den 1000 Simulationen haben

```r
sum(FB1000$mean <= 0)
```

```
## [1] 140
```

einen Mittelwert in der Jahresrendite unter 0.

Übung:

1. Berechnen (nicht simulieren!) Sie das 95% Konfidenzintervall unter der Annahme einer Normalverteilung.
2. Wie ändern sich Standardfehler und p-Werte, wenn die gleichen Punktschätzer nicht auf der Basis von einem sondern von 3 Jahren ermittelt wurden.

t-Test für eine abhängige Stichprobe

Der B3 Datensatz *Heilemann, U. and Münch, H.J. (1996): West German Business Cycles 1963-1994: A Multivariate Discriminant Analysis. CIRET–Conference in Singapore, CIRET–Studien 50.* enthält Quartalsweise Konjunkturdaten aus (West-)Deutschland.

Er kann von https://goo.gl/0YCEHf heruntergeladen werden:

```r
download.file("https://goo.gl/0YCEHf", destfile = "B3.csv")
```

Anschließend können die Daten in R eingelesen werden:

```r
B3 <- read.csv2("B3.csv")
str(B3) # Datenstruktur
```

```
## 'data.frame':    157 obs. of  14 variables:
##  $ PHASEN  : int  2 2 3 3 3 3 3 3 3 3 ...
##  $ BSP91JW : num  10.53 10.6 9.21 5.17 4.93 ...
##  $ CP91JW  : num  9.31 12.66 6.55 7.87 8.6 ...
##  $ DEFRATE : num  0.05 0.06 0.05 0.05 0.04 0.04 0.04 0.03 0.03 0 ...
##  $ EWAJW   : num  5.7 5.2 4.8 3.3 2.1 3.2 2.5 2.7 3 0.3 ...
##  $ EXIMRATE: num  3.08 1.96 2.82 3.74 4.16 2.9 3.65 4.57 4.37 2.89 ...
##  $ GM1JW   : num  11.15 11.03 10.04 8.33 7.69 ...
##  $ IAU91JW : num  23.56 12.72 11.52 0.85 -2.08 ...
##  $ IB91JW  : num  14.69 24.95 14.9 7.55 3.23 ...
##  $ LSTKJW  : num  3 2.36 3.39 5.3 6.91 1.03 3.73 6.2 4.12 7.94 ...
##  $ PBSPJW  : num  2.89 2.59 3.01 3.03 3.46 1.95 3.18 3.98 3.29 5.63 ...
##  $ PCPJW   : num  1.91 2.2 3.09 2.08 1.48 1.65 1.47 3.29 3.59 4.19 ...
##  $ ZINSK   : num  6.27 4.6 6.19 6.71 7.1 4.96 5.21 4.83 4.5 3.83 ...
##  $ ZINSLR  : num  3.21 3.54 3.22 3.37 3.14 4.95 3.82 3.09 3.91 1.47 ...
```

```r
head(B3); tail(B3)
```

```
##   PHASEN BSP91JW CP91JW DEFRATE EWAJW EXIMRATE GM1JW IAU91JW IB91JW LSTKJW
## 1      2   10.53   9.31    0.05   5.7     3.08 11.15   23.56  14.69   3.00
## 2      2   10.60  12.66    0.06   5.2     1.96 11.03   12.72  24.95   2.36
## 3      3    9.21   6.55    0.05   4.8     2.82 10.04   11.52  14.90   3.39
## 4      3    5.17   7.87    0.05   3.3     3.74  8.33    0.85   7.55   5.30
## 5      3    4.93   8.60    0.04   2.1     4.16  7.69   -2.08   3.23   6.91
## 6      3    8.39   5.62    0.04   3.2     2.90  6.62   -3.76  14.58   1.03
##   PBSPJW PCPJW ZINSK ZINSLR
## 1   2.89  1.91  6.27   3.21
## 2   2.59  2.20  4.60   3.54
## 3   3.01  3.09  6.19   3.22
## 4   3.03  2.08  6.71   3.37
## 5   3.46  1.48  7.10   3.14
## 6   1.95  1.65  4.96   4.95
```

```
##     PHASEN BSP91JW CP91JW DEFRATE EWAJW EXIMRATE GM1JW IAU91JW IB91JW
## 152      3   -1.27   1.29   -4.87 -1.97     6.03  9.79  -18.29   1.73
## 153      3   -2.13  -0.57   -2.98 -2.05     7.59  0.72  -15.82  -3.23
## 154      3    1.39   2.33   -2.86 -1.84     7.49 11.33  -10.59   4.62
## 155      4    1.63   0.64    1.20 -1.58     7.75 11.38   -4.90   3.62
## 156      1    1.40   0.57   -3.56 -1.34     5.58  9.53   -0.76   2.19
## 157      1    1.83  -0.08   -2.22 -0.93     7.50 15.20    2.75   6.12
##     LSTKJW PBSPJW PCPJW ZINSK ZINSLR
## 152   1.08   2.73  2.98  6.83   3.55
## 153   1.67   2.67  3.31  6.35   3.05
## 154  -0.12   2.66  2.94  5.88   3.17
```

```
## 155  -1.81   1.77  2.58  5.29   4.82
## 156  -1.54   1.85  2.60  5.01   5.27
## 157  -0.92   1.79  2.49  5.28   5.62
```

Hier interessieren besonders die (Veränderung) der Investitionen in Ausrüstungsgüter (`IAUJW91`) und in Bauten (`IB91JW`). Die deskriptiven Kennzahlen zeigen,

```
favstats( ~ IAU91JW, data=B3)
```

```
##     min    Q1 median  Q3   max     mean       sd   n missing
##  -19.95 -1.25    5.3 9.1 27.25 3.992675 8.864805 157       0
```

```
favstats( ~ IB91JW, data=B3)
```

```
##     min    Q1 median   Q3   max     mean       sd   n missing
##  -21.59 -1.16    2.6 5.55 40.25 2.565096 7.481063 157       0
```

dass im betrachteten Zeitraum die Investitionen in Ausrüstungsgüter mit im arithmetischen Mittelwert von 3.99 im Mittel stärker gestiegen sind als die in Bauten mit 2.57. Da die Investitionen sicherlich in Zusammenhang mit der gesamten konjunkturellen Entwicklung stehen, ist davon auszugehen, dass es sich hier um vom jeweiligen Zeitpunkt abhängige Beobachtungen handelt. Daher wird hier die Differenz der Werte betrachtet: `IB91JW - IAU91JW`. Der R Befehl für einen t-Test lautet `t.test`:

```
t.test (~ (IB91JW - IAU91JW), data=B3)
```

```
##
##  One Sample t-test
##
## data:  B3$(IB91JW - IAU91JW)
## t = -1.9612, df = 156, p-value = 0.05164
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  -2.86544149  0.01028226
## sample estimates:
## mean of x
##  -1.42758
```

Der (umfangreichen) Ausgabe kann man neben den z- bzw. t-Wert (-1.96) unter der Nullhypothese der Gleichheit des Lageparameters

$$H_0 : \mu_{\text{IB91JW-IAU91JW}} = 0$$

insbesondere den p-Wert (0.0516) und das Konfidenzintervall (-2.87, 0.01) entnehmen. Zum Signifikanzn-vieau von 5% wird die Nullhypothese also gerade so *nicht* abgelehnt, da der p-Wert über 5% liegt.

Übung:

3. Testen Sie, ob es einen nicht zufälligen mittleren Lageunterschied zwischen der Veränderung des Preisindex des Bruttosozialproduktes `PBSPJW` und dem des privaten Verbrauchs `PCPJW` gibt.

---

t-Test für eine unabhängige Stichprobe

Untersuchen wir, ob sich makroökonomische Kennzahlen im Auf- und Abschwung unterscheiden. Zunächst stellen wir fest, dass die eigentlich kategorielle Variable `PHASEN` hier numerisch kodiert wurde, was aber schnell verwirren würde.

```
typeof(B3$PHASEN)
```

```
## [1] "integer"
```

Typänderung zu `factor` geht einfach:

```r
B3$PHASEN <- as.factor(B3$PHASEN)
```

Wenn wir die einzelnen `levels` des Faktors als numerische Werte verwenden wollen würde man den Befehl `as.numeric()` verwenden. Aber sicherheitshalber vorher über `levels()` gucken, ob die Reihenfolge auch stimmt.

Um die Interpretation zu erleichtern können wir hier einfach die Faktorstufe umbenennen.

```r
levels(B3$PHASEN) <- c("Aufschwung", "Oberer Wendepunkt",
                        "Abschwung", "Unterer Wendepunkt")
```

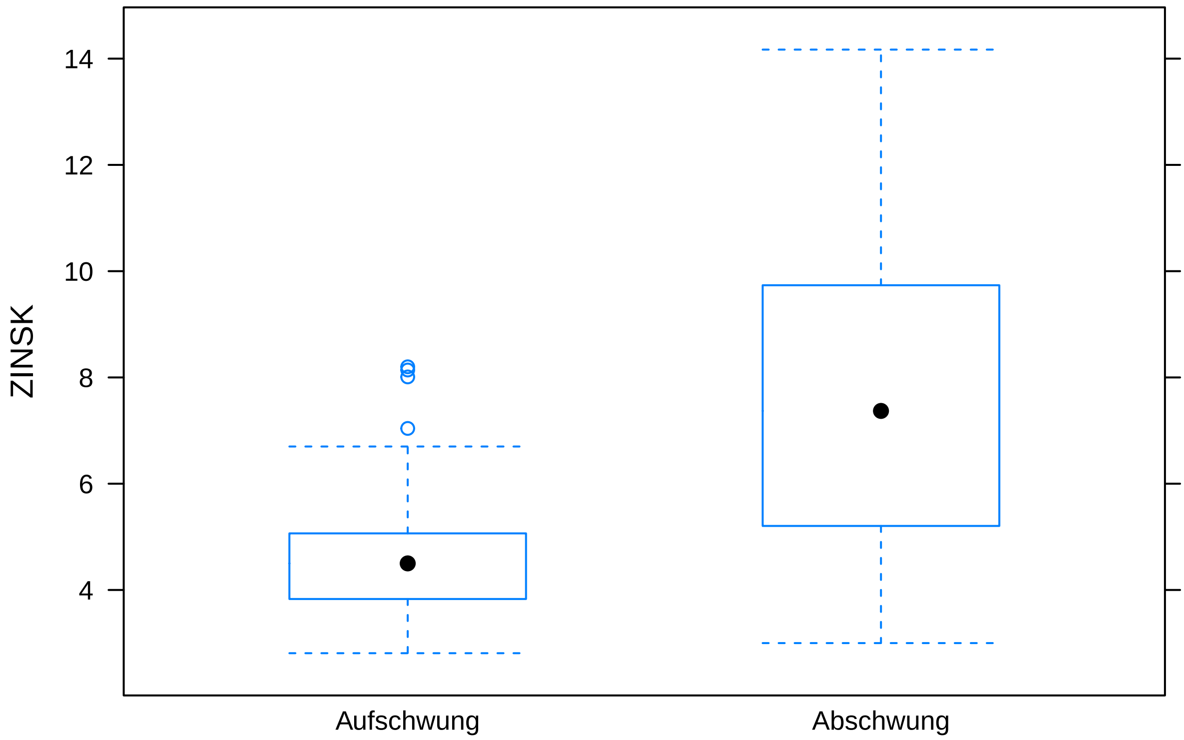
Jetzt ist keine Verwechselung von kategoriellen und metrischen Variablen mehr möglich.

Zunächst wird der Datensatz, der auch die konjunkturellen Wendepunkte enthält, nur auf Auf- und Abschwung eingeschränkt.

```r
B3AufAb <- subset(B3, PHASEN %in% c("Aufschwung", "Abschwung"))
B3AufAb <- droplevels(B3AufAb)
```

In der politischen Diskussion werden immer niedrige Zinsen gefordert. Gucken wir mal wie die Zinsen in den Konjunkturphasen in der Vergangenheit (1955-1994) verteilt waren:

```r
bwplot(ZINSK ~ PHASEN, data=B3AufAb)
```



Anscheinend waren die Zinsen in Zeiten des Aufschwungs niedriger.

Was sagen die deskriptiven Kennzahlen:

```r
favstats(ZINSK ~ PHASEN, data=B3AufAb)
```

```
##       PHASEN  min    Q1 median    Q3   max     mean       sd  n missing
## 1 Aufschwung 2.81 3.830   4.50 5.065  8.20 4.715085 1.209989 59       0
## 2  Abschwung 3.00 5.205   7.37 9.735 14.17 7.682553 3.020254 47       0
```

Alle Lagemaße für die Zinskosten sind in der Aufschwungphase niedriger.

Der t-Test der Zinskosten für

$$H_0 : \mu_{\text{Aufschwung}} = \mu_{\text{Abschwung}} \Leftrightarrow \mu_{\text{Aufschwung}} - \mu_{\text{Abschwung}} = 0$$

hat dann den gleichen Aufbau des Syntax wie `bwplot` oder `favstats`: Teste die Zinskosten in Abhängigkeit der Konjunkturphase:

```
t.test(ZINSK ~ PHASEN, data=B3AufAb)
```

```
##
##  Welch Two Sample t-test
##
## data:  ZINSK by PHASEN
## t = -6.3426, df = 57.766, p-value = 3.743e-08
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -3.904085 -2.030852
## sample estimates:
## mean in group Aufschwung  mean in group Abschwung
##                 4.715085                 7.682553
```

Der kleine p-Wert von 3.7430775×10^{-8} zeigt, dass die Nullhypothese der Gleichheit der Lageparameter verworfen werden kann. Wir können der Funktion auch eine spezielle Alternativhypothese übergeben:

```
t.test(ZINSK ~ PHASEN, data=B3AufAb, alternative = "less")
```

```
##
##  Welch Two Sample t-test
##
## data:  ZINSK by PHASEN
## t = -6.3426, df = 57.766, p-value = 1.872e-08
## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:
##       -Inf -2.185354
## sample estimates:
## mean in group Aufschwung  mean in group Abschwung
##                 4.715085                 7.682553
```

Jetzt haben wir die Nullhypothese "Das Lagemaß für die Zinskosten ist im Aufschwung *nicht* kleiner als im Abschwung" gegen die Alternativhypothese (Forschungshypothese) "Das Lagemaß für die Zinskosten ist im Aufschwung kleiner als im Abschwung" getestet:

$$H_0 : \mu_{\text{Aufschwung}} \geq \mu_{\text{Abschwung}} \quad vs. \quad H_A : \mu_{\text{Aufschwung}} < \mu_{\text{Abschwung}}$$

bzw.

$$H_0 : \mu_{\text{Aufschwung}} - \mu_{\text{Abschwung}} \geq 0 \quad vs. \quad H_A : \mu_{\text{Aufschwung}} - \mu_{\text{Abschwung}} < 0$$

Übung:

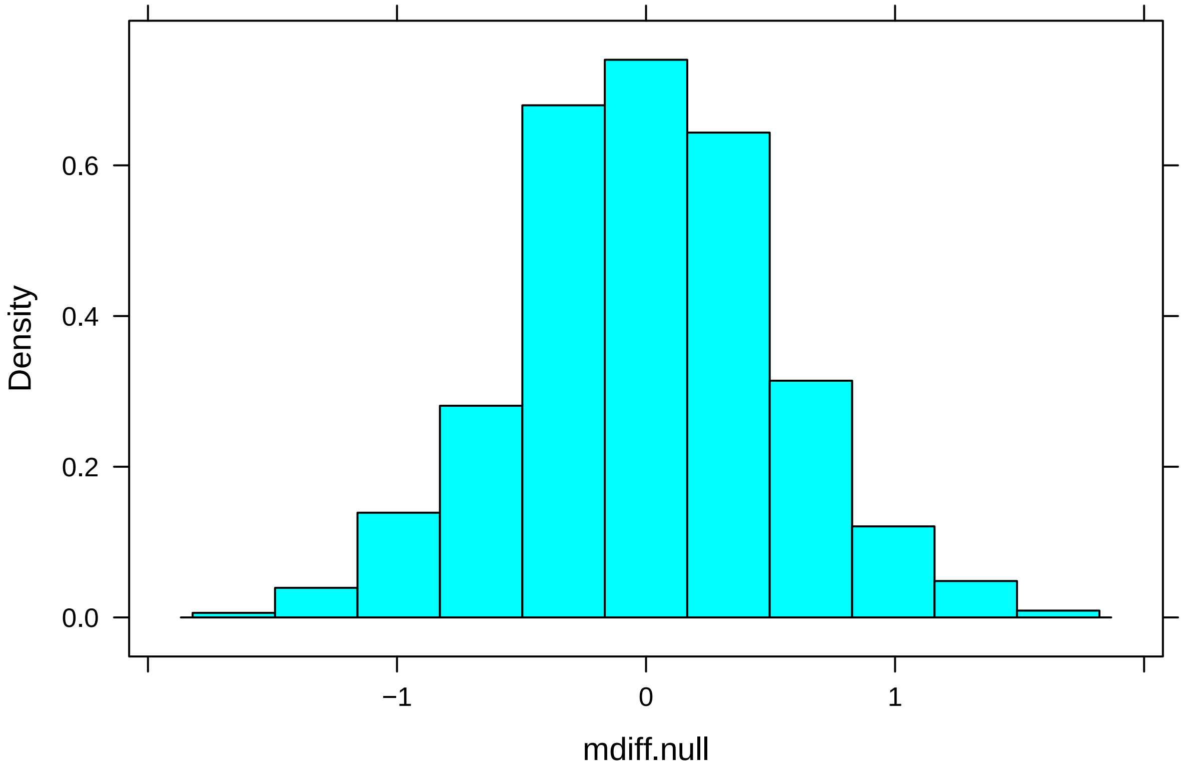
4. Untersuchen Sie, ob sich die mittlere Entwicklung des privaten Verbrauchs `CP91JW` zwischen den Konjunkturphasen unterscheidet.

Auch hier können wir, ohne eine Verteilungsannahme zu verwenden permutieren.

```
mdiff <- diff(mean(ZINSK ~ PHASEN, data=B3AufAb))
mdiff
```

```
## Abschwung
##  2.967468
```

```
mdiff.null <- do(1000) * diff(mean(ZINSK ~ sample(PHASEN), data=B3AufAb))
histogram( ~ mdiff.null)
```



Unter der Nullhypothese der Gleichheit der Lagemaße kommt eine gleich große oder größere Differenz also

```
sum(mdiff.null >= mdiff)
```

```
## [1] 0
```

mal bei 1000 Permutationen vor!

Da die statistische *Signifikanz* vom Standardfehler abhängt, welcher wiederum vom Stichprobenumfang abhängt wurde von Cohen ein Maß für die *Effektstärke*, **Cohen's d** vorgeschlagen:

$$d = \frac{\bar{x}_1 - \bar{x}_2}{sd_{\text{pool}}}$$

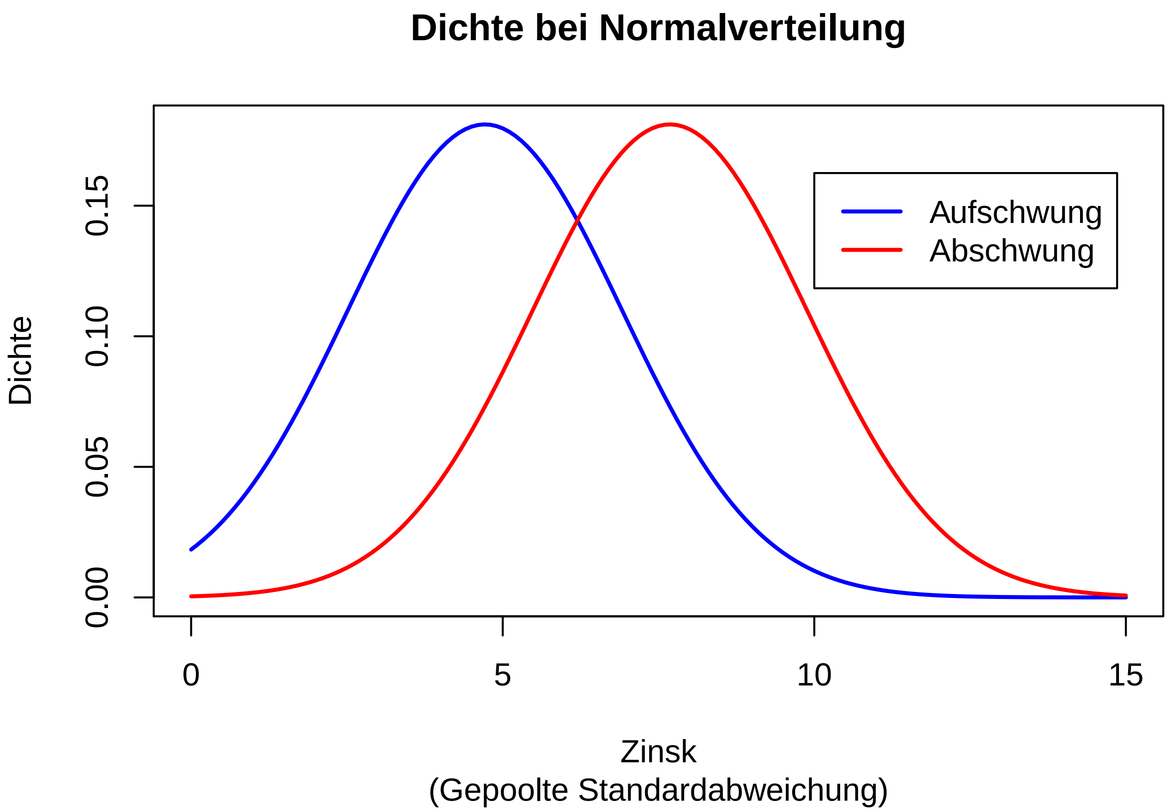
mit

$$sd_{\text{pool}} = \sqrt{\frac{1}{n_1 + n_2 - 2}\left((n_1 - 1)sd_1^2 + (n_2 - 1)sd_2^2\right)}$$

```
# Kennzahlen 1. Stichprobe
m1 <- mean(B3$ZINSK[B3$PHASEN=="Aufschwung"])
sd1 <- sd(B3$ZINSK[B3$PHASEN=="Aufschwung"])
n1 <- length(B3$ZINSK[B3$PHASEN=="Aufschwung"])
# Kennzahlen 2. Stichprobe
m2 <- mean(B3$ZINSK[B3$PHASEN=="Abschwung"])
sd2 <- sd(B3$ZINSK[B3$PHASEN=="Abschwung"])
n2 <- length(B3$ZINSK[B3$PHASEN=="Abschwung"])
# Gepoolte Standardabweichung
sdpool <- sqrt( ((n1-1)*sd1^2 + (n2-1)*sd2^2) / (n1+n2-2))
# Cohen's d
d <- (m1-m2)/sdpool
d
```

```
## [1] -1.347291
```

Cohen's d ist ein Maß der Überlappung der Verteilungen:



Häufig werden Werte

- |d|<=0.2 als kleine
- |d|<=0.5 als mittlere
- |d|>=0.8 als große Effekte

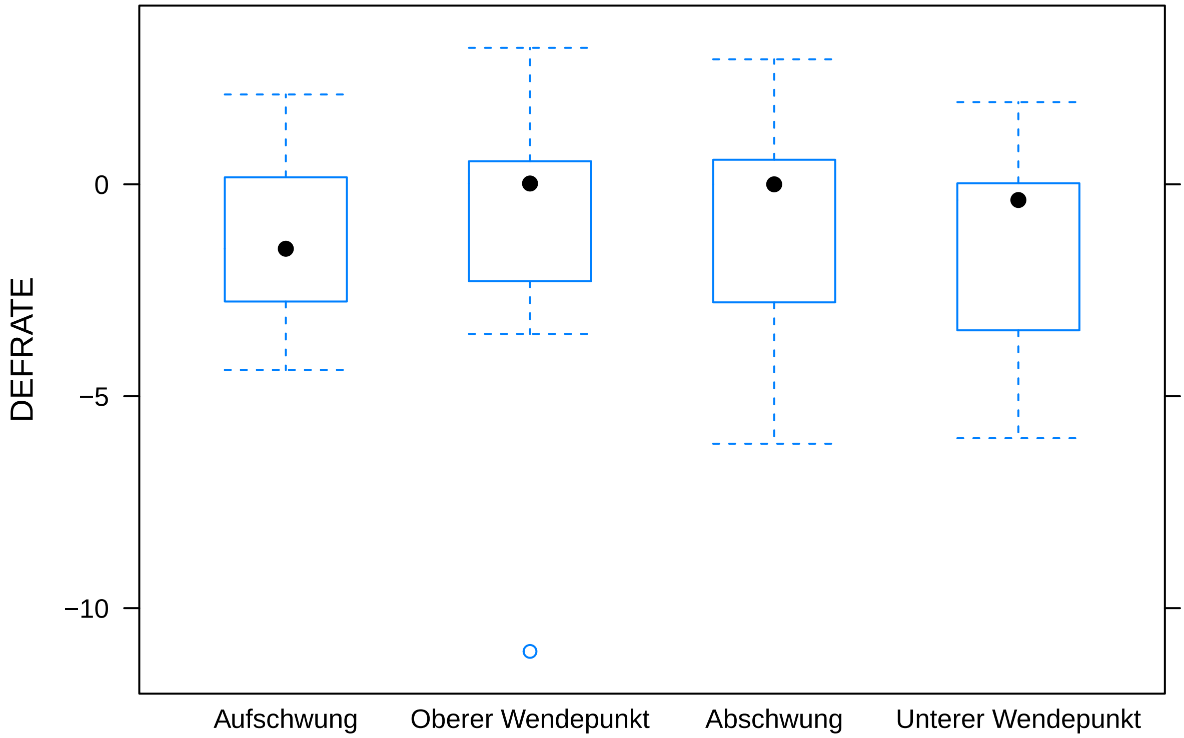
bezeichnet.

ANOVA

Bei mehr als zwei Gruppen funktionieren die Techniken des t-Tests nicht mehr. Um Lagemaßunterschiede zu testen wird anstelle der Mittelwerte die Streuung verglichen: Ist die Streuung zwischen den Gruppen groß im Vergleich zur Streuung innerhalb der Gruppen?

Unterscheidet sich der mittlere Anteil des Staatsdefizits `DEFRATE` nicht zufällig zwischen den Konjunkturphasen?

```
bwplot( DEFRATE ~ PHASEN, data=B3)
```




```
favstats( DEFRATE ~ PHASEN, data=B3)
```

```
##              PHASEN    min      Q1 median     Q3  max       mean       sd
## 1        Aufschwung  -4.38 -2.7650  -1.52 0.1650 2.12 -1.3394915 1.680638
## 2 Oberer Wendepunkt -11.02 -2.2475   0.02 0.4775 3.22 -0.8479167 2.836558
## 3         Abschwung  -6.12 -2.7850   0.00 0.5800 2.95 -0.8380851 2.287536
## 4 Unterer Wendepunkt  -5.99 -3.4450  -0.37 0.0250 1.94 -1.6548148 2.364026
##    n missing
## 1 59       0
## 2 24       0
## 3 47       0
## 4 27       0
```

Vielleicht, vielleicht nicht.

Um eine Varianzanalyse mit

$$H_0 : \mu_1 = \mu_2 = \ldots = \mu_k$$

gegen

$$H_A : \text{Mindestens ein } \mu \text{ ist verschieden.}$$

durchzuführen kann in R u. a. der Befehl `aov` verwendet werden:

```
DEFaov <- aov(DEFRATE ~ PHASEN, data=B3)
summary(DEFaov)
```

```
##              Df Sum Sq Mean Sq F value Pr(>F)
## PHASEN        3   15.7   5.236    1.09  0.355
## Residuals   153  734.9   4.803
```

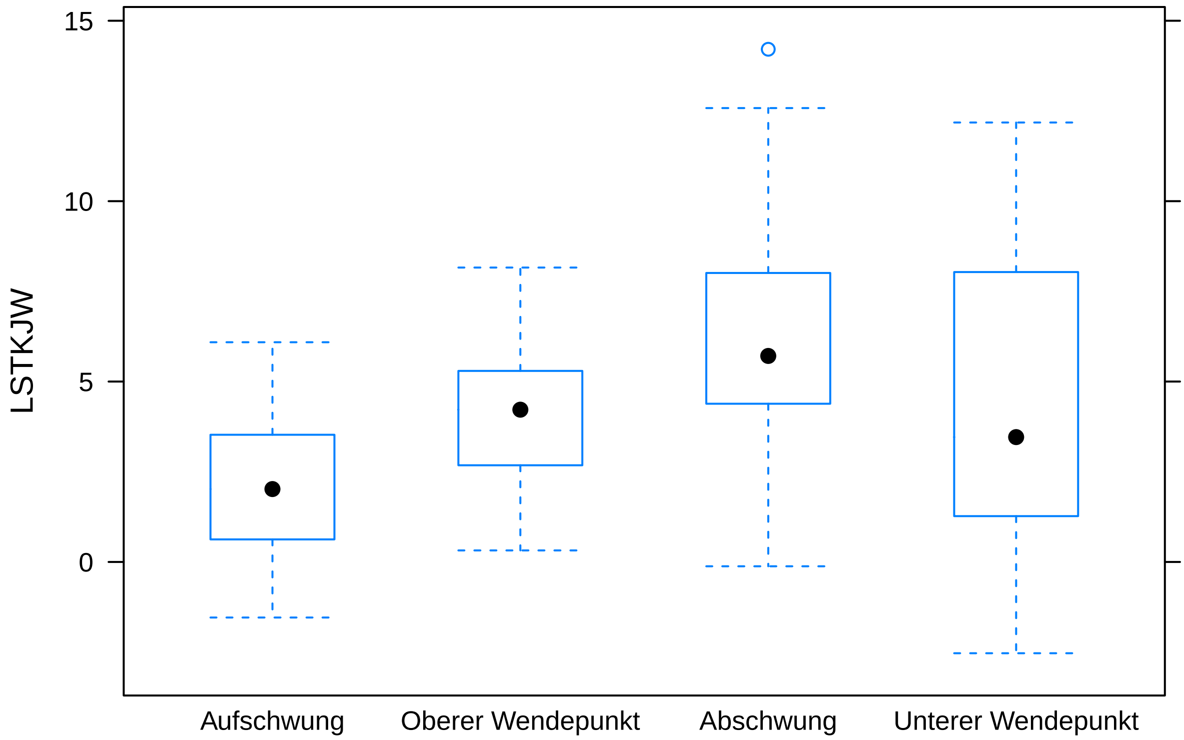
Der p-Wert des F-Tests der Nullhypothese

$$H_0 : \mu_{\text{Aufschwung}} = \mu_{\text{Oberer Wendepunt}} = \mu_{\text{Abschwung}} = \mu_{\text{Unterer Wendepunkt}}$$

der Gleichheit der Lage ist mit 0.3552 größer als 0.05, die Nullhypothese kann also für das Staatsdefizit nicht verworfen werden.

Unterscheidet sich das Lagemaß der Veränderung der Lohnstückkosten `LSTKJW` nicht zufällig?

```r
bwplot(LSTKJW ~ PHASEN, data=B3)
```




```r
favstats(LSTKJW ~ PHASEN, data=B3)
```

```
##              PHASEN   min    Q1 median     Q3   max     mean       sd  n
## 1        Aufschwung -1.54 0.625   2.02 3.5250  6.09 2.111017 1.837423 59
## 2  Oberer Wendepunkt  0.32 2.840   4.22 5.2225  8.16 4.195833 2.074516 24
## 3         Abschwung -0.12 4.385   5.71 8.0100 14.21 6.291064 3.122604 47
## 4 Unterer Wendepunkt -2.53 1.270   3.46 8.0350 12.18 4.249630 4.449861 27
##   missing
## 1       0
## 2       0
## 3       0
## 4       0
```

```r
LSTKaov <- aov(LSTKJW ~ PHASEN, data=B3)
summary(LSTKaov)
```

```
##              Df Sum Sq Mean Sq F value   Pr(>F)    
## PHASEN        3  459.5  153.15   18.62 2.37e-10 ***
## Residuals   153 1258.2    8.22                     
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Die Nullhypothese der Gleichheit wird hier also verworfen. Interessanterweise unterscheiden sich insbesondere die Lagemaße von Auf- und Abschwung, die beiden Wendepunkte liegen dazwischen.

Übung:

5. Gibt es nicht zufällige Lageunterschiede bei der Änderung der Erwerbstätigen `EWAJW` zwischen den Konjunkturphasen?

Übung: Teaching Rating

Dieser Datensatz analysiert u. a. den Zusammenhang zwischen Schönheit und Evaluierungsergebnis von Dozenten:

Hamermesh, D.S., and Parker, A. (2005). Beauty in the Classroom: Instructors' Pulchritude and Putative Pedagogical Productivity. Economics of Education Review, 24, 369–376.

Sie können ihn, sofern noch nicht geschehen, von https://goo.gl/6Y3KoK als `csv`herunterladen.

1. Ist das arithmetische Mittel der Evaluierung `eval` nicht zufällig größer als befriedigend (3)?
2. Gibt es einen nicht zufälligen Unterschied im Lagemaß der Evaluation `eval` zwischen mönnlichen und weiblichen Dozent/innen (`gender`)?

Versionshinweise:

- Datum erstellt: 2016-12-13
- R Version: 3.3.2
- `mosaic` Version: 0.14.4