

# Die Varianzanalyse

Jonathan Harrington

04. Juni 2019

```
library(ggplot2)
library(dplyr)
library(ez)

source(file.path(pfadu, "phoc.txt"))
source(file.path(pfadu, "ph.step.R"))

blang = read.table(file.path(pfadu, "blang.txt"))
v.df = read.table(file.path(pfadu, "vokal.txt"))
dg = read.table(file.path(pfadu, "dg.txt"))
ssb = read.table(file.path(pfadu, "ssb.txt"))
```

## Vorbemerkung zur sogenannten “wissenschaftlichen Notation” von Zahlen

Wenn ein Effekt eines Faktors sehr groß ist, wird der p-Wert sehr klein, z.B. etwas wie 0.0000000147. Ganz generell ist eine so kleine Zahl sehr schwer zu lesen (und dasselbe gilt auch für sehr große Zahlen). Daher verwendet R die sogenannte “wissenschaftliche Notation” für sehr große und sehr kleine Zahlen. Die wissenschaftliche Notation beschreibt jede Zahl als Kommazahl  $m$ , die größer als 0, aber kleiner als 10 ist, multipliziert mit  $10^e$ , also

$$m * 10^e$$

So ist z.B. die Zahl eine Milliarde, wenn Sie sie in der R-Konsole eintippen (also als 1 mit 9 Nullen), wie folgt dargestellt,

```
1000000000
```

```
## [1] 1e+09
```

also als  $1 * 10^9$ .

Ganz genauso ist es mit einer sehr kleinen Zahl, wie z.B. einem Milliardstel,

```
0.000000001
```

```
## [1] 1e-09
```

das also als  $1 * 10^{-9}$  dargestellt wird (denn es ist ja auch eine 1 an der neunten Nachkommastelle).

Normalerweise stellt R nur Zahlen in dieser Schreibweise dar, die wirklich groß bzw. klein (und damit unleserlich) sind:

```
0.01
## [1] 0.01
0.001
## [1] 0.001
0.0001
## [1] 1e-04
0.00001
## [1] 1e-05
#usw. ...
```

Einige Packages (wie das heute zu verwendende ez) stellen aber in den Ergebnistabellen alle p-Werte in dieser Notation dar, so dass aus 0.18 ein 1.8e-01 wird.

Damit man trotzdem auf einen Blick erkennt, was signifikant ist und was nicht, also ob dieser Wert von  $p$  kleiner oder größer als 0.05 ist, wird in ez (und anderen Packages) zusätzlich ein \* angegeben, wenn  $p < 0.05$ .

## ANOVA und t-Test

### t-Test oder ANOVA (“Analysis of Variance” = Varianzanalyse)

Ein t-Test oder eine ANOVA sind dann äquivalent, wenn man nur **einen** Faktor hat, der nur **zwei Stufen** hat, z.B.

- Hat Geschlecht einen Einfluss auf die Dauer?

### ANOVA

Hat man hingegen einen **Faktor mit mehr als 2 Stufen**, oder **mehr als einen Faktor**, so kann man keinen t-Test mehr anwenden, und braucht stattdessen eine Varianzanalyse. Beispiele wären:

- Es gibt 3 Altersgruppen, *jung*, *mittel*, *alt*. Hat die Altersgruppe einen Einfluss auf die Dauer? (= ein Faktor mit 3 Stufen)
- Haben Geschlecht und Dialekt einen Einfluss auf die Dauer? (= 2 Faktoren)

### Was ist die Varianzanalyse?

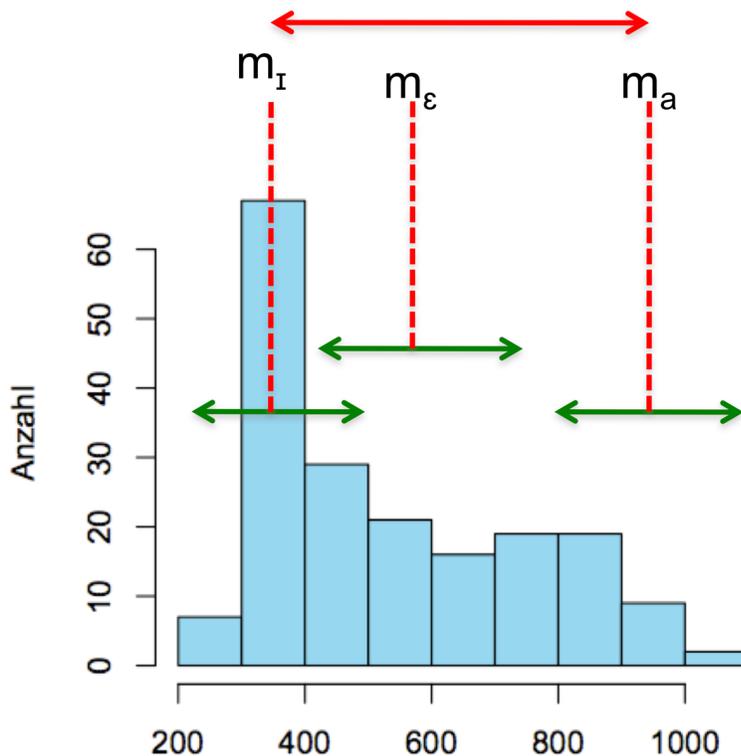
Die *Varianz* ist ein Streuungsmaß, das sich berechnet als Summe der quadrierten Abweichungen aller Einzelwerte einer Verteilung vom arithmetischen Mittel derselben

Verteilung geteilt durch die Gesamtzahl der Werte. Wir kennen dies Maß indirekt bereits: die uns schon bekannte Stichprobenstandardabweichung ( $sd()$ ) ist nämlich die Quadratwurzel der Varianz.

Mit der Varianzanalyse wird (durch einen sogenannten  $F$ -Test) ein Verhältnis zwischen zwei Varianzen berechnet: derjenigen **innerhalb von Stufen** und derjenigen **zwischen Stufen**.

Beispiel:  $F_1$  von drei Vokalkategorien, /i,ε,a/

## F<sub>1</sub>-Verteilung, drei Vokale



$$F = \frac{\text{Varianz zwischen den Stufen}}{\text{Varianz innerhalb der Stufen}}$$

### Innerhalb der Stufen

Es gibt eine **randomisierte Variation von F<sub>1</sub>** *innerhalb* jeder Stufe (F<sub>1</sub> von /i/ variiert, F<sub>1</sub> von /ε/ variiert, F<sub>1</sub> von /a/ variiert).

### Zwischen den Stufen

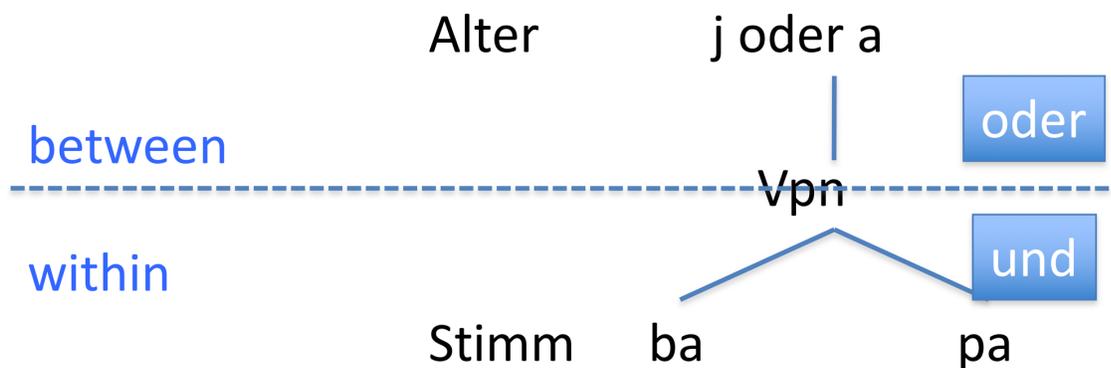
F<sub>1</sub> variiert, weil es eine **systematische Variation** zwischen den Verteilungen der Vokalkategorien gibt: die Werte von /i/, /ε/, und /a/ liegen in ganz unterschiedlichen F<sub>1</sub>-

Bereichen, und je unterschiedlicher sie sind, umso **größer wird diese Varianz im Verhältnis zu der willkürlichen, randomisierten Varianz innerhalb der Stufen sein:**

$$F = \frac{\text{systematische - Varianz - ZWISCHEN - den - Stufen}}{\text{randomisierte - Varianz - INNERHALB - der - Stufen}}$$

Der Wert von  $F$  hängt also nicht nur davon ab, wie weit die kategorienspezifischen Mittelwerte voneinander entfernt sind, sondern auch davon, wie weit die Werte pro Kategorie jeweils streuen. Die gleichen Mittelwerte könnten also mal zu einem signifikanten (wenn die Streuungen der Werte pro Kategorie relativ groß sind), und mal zu einem nicht-signifikanten Ergebnis (wenn die Streuungen der Werte pro Kategorie relativ klein sind) führen.

## Within-Subjects- vs. Between-Subjects-Faktoren



### Innersubjektfaktoren (within-subject factors)

Man spricht von Innersubjektfaktoren, wenn, wie beim gepaarten  $t$ -Test auch, Analysen innerhalb derselben Person vorgenommen wurden, so dass zumindest eine Messung pro Versuchsperson pro Stufe vorliegt:

#### Gepaarter $t$ -Test *oder* ANOVA mit Innersubjektfaktor

/ba, pa/ wurden von denselben Sprechern produziert. *Unterscheiden sich /ba, pa/ in VOT?*

- Abhängige Variable: **VOT**
- Within-Faktor: **Stimmhaftigkeit** (= /ba/ vs. /pa/)

Wir haben hier genau einen Wert für /ba/ und einen Wert für /pa/ pro Versuchsperson. Wir können also einen gepaarten  $t$ -Test durchführen, oder auch eine ANOVA mit dem Innersubjektfaktor **Stimmhaftigkeit**.

#### Kein gepaarter $t$ -Test möglich, aber ANOVA mit Innersubjektfaktor

/i, y, u/ wurden von denselben Sprechern produziert. *Unterscheiden sich /i, y, u/ in F2?*

- Abhängige Variable: **F2**
- Within-Faktor: **Vokal** (/i/ vs. /y/ vs. /u/)

Also ein Wert für /i/ ein Wert für /y/, ein Wert für /u/ pro Versuchsperson. Wegen der drei Stufen ist kein gepaarter *t*-Test möglich, aber eine *ANOVA* mit dem Innersubjektfaktor **Vokal**.

## Zwischensubjektfaktoren (between-subjects factors)

Ein Zwischensubjektfaktor beschreibt meistens eine Eigenschaft der Versuchsperson, z.B. Muttersprache (englisch oder deutsch oder französisch), Geschlecht (m oder w), Alter (jung oder alt) usw. .

In Analogie zum ungepaarten *t*-Test haben wir es also hier mit entweder-oder-Fällen zu tun; eine Versuchsperson ist z.B. *entweder* männlich *oder* weiblich.

Natürlich können Zwischensubjektfaktoren bei einer *ANOVA* aber auch mehr als zwei Stufen haben.

## Beispieldaten zur Unterscheidung zwischen Inner- und Zwischensubjektfaktoren

Die Kieferposition wurde in 3 **Vokalen** /i, e, a/ und jeweils zu 2 **Sprechtempi** (*langsam, schnell*) gemessen. Die Messungen (3 x 2 = 6 pro Vpn) sind von 16 Vpn erhoben worden, 8 mit **Muttersprache Spanisch**, 8 mit Muttersprache *Englisch*.

*Inwiefern haben die Faktoren Sprache, Sprechtempo, und Vokal einen Einfluss auf die Kieferposition?*

### Between

- Sprache

### Within

- Sprechtempo
- Vokal

## Within-Factor: gepaarter *t*-Test und *ANOVA*

```
head(blang); dim(blang)
```

```
##      F2 Vpn Betonung
## 1 2577 S1          B
## 2 2122 S2          B
## 3 2192 S3          B
## 4 2581 S4          B
## 5 2227 S5          B
## 6 2481 S6          B
## [1] 24 3
```

*12 Versuchspersonen produzierten /i/ in betonter und unbetonter Position. Hat Betonung einen Einfluss auf F2?*

Jede Stufe des unabhängigen within-Faktor wird von jeder Versuchsperson einmal belegt:

```
with(blang, table(Vpn, Betonung))
```

```
##      Betonung
## Vpn  B U
##  S1  1 1
##  S10 1 1
##  S11 1 1
##  S12 1 1
##  S2   1 1
##  S3   1 1
##  S4   1 1
##  S5   1 1
##  S6   1 1
##  S7   1 1
##  S8   1 1
##  S9   1 1
```

## Drei Schritte

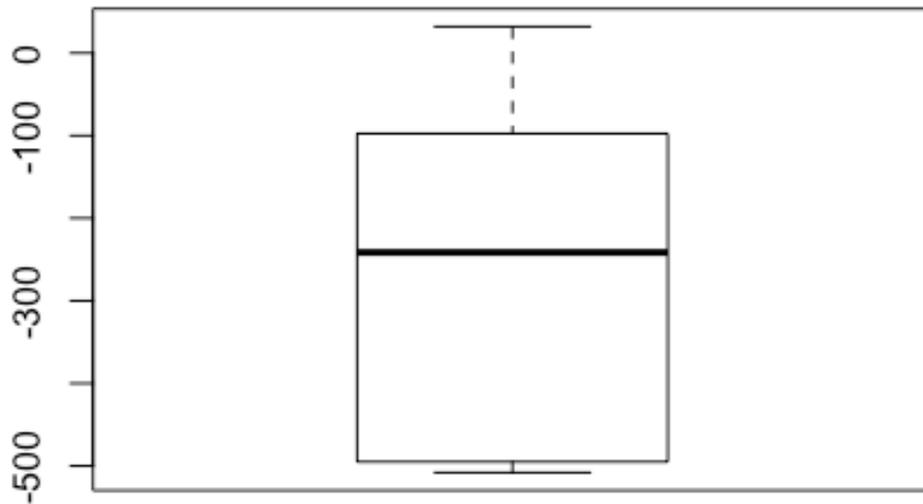
### 1. Differenz-Berechnung

```
d = blang%>%
  group_by(Vpn)%>%
  summarise(F2 = diff(F2))
d
```

```
## # A tibble: 12 x 2
##   Vpn      F2
##   <fct> <int>
## 1 S1      -495
## 2 S10     -497
## 3 S11       32
## 4 S12    -308
## 5 S2       -99
## 6 S3        15
## 7 S4     -480
## 8 S5       -96
## 9 S6     -508
## 10 S7     -146
## 11 S8     -495
## 12 S9     -175
```

### 2. Boxplot

```
boxplot(d$F2)
```



### 3. Statistischer Test

#### Gepaarter $t$ -Test

```
t.test(d$F2)
```

```
##
## One Sample t-test
##
## data:  d$F2
## t = -4.3543, df = 11, p-value = 0.001147
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  -407.9837 -134.0163
## sample estimates:
## mean of x
##      -271
```

**F2 wird signifikant von Betonung beeinflusst ( $t[11] = 4.4$ ,  $p < 0.01$ ).**

**Oder: Lösung mit *ANOVA* (`ezANOVA()` aus dem package `ez`, angewendet auf den ursprünglichen `data.frame`)**

```
ezANOVA(blang, .(F2), .(Vpn), .(Betonung))
```

```
## $ANOVA
##      Effect DFn DFd      F      p p<.05      ges
## 2 Betonung   1  11 18.95986 0.001147148 * 0.4113659
```

**F2 wird signifikant von Betonung beeinflusst ( $F[1,11] = 19.0$ ,  $p < 0.01$ ).**

## Generelle Syntax des Befehls `ezANOVA()`

```
ezANOVA(DATAFRAME,
        .(Abhängige numerische Variable),
        .(Versuchspersonen),
        .( WITHIN-FAKTOR(EN) ),
        between = .( BETWEEN-FAKTOR(EN) ))
```

## Vergleich *ANOVA* und *t*-Test-Ergebnisse

### Derselbe Wahrscheinlichkeitswert:

```
t.test(d$F2)$p.value
```

```
## [1] 0.001147148
```

```
ezANOVA(blang, .(F2), .(Vpn), .(Betonung))$ANOVA$p
```

```
## [1] 0.001147148
```

### Zusammenhang zwischen den Prüfstatistikwerten: $F = t^2$

```
t.test(d$F2)$statistic
```

```
##      t
## -4.354292
```

```
t.test(d$F2)$statistic^2
```

```
##      t
## 18.95986
```

```
ezANOVA(blang, .(F2), .(Vpn), .(Betonung))$ANOVA$F
```

```
## [1] 18.95986
```

## Between-factor: Two-sample *t*-Test und *ANOVA*

```
head(v.df); dim(v.df)
```

```
##      F2 Sprache Vpn
## 1 2009      D  S1
## 2 1700      D  S2
## 3 1678      D  S3
## 4 2070      D  S4
## 5 1832      D  S5
## 6 2132      D  S6

## [1] 20  3
```

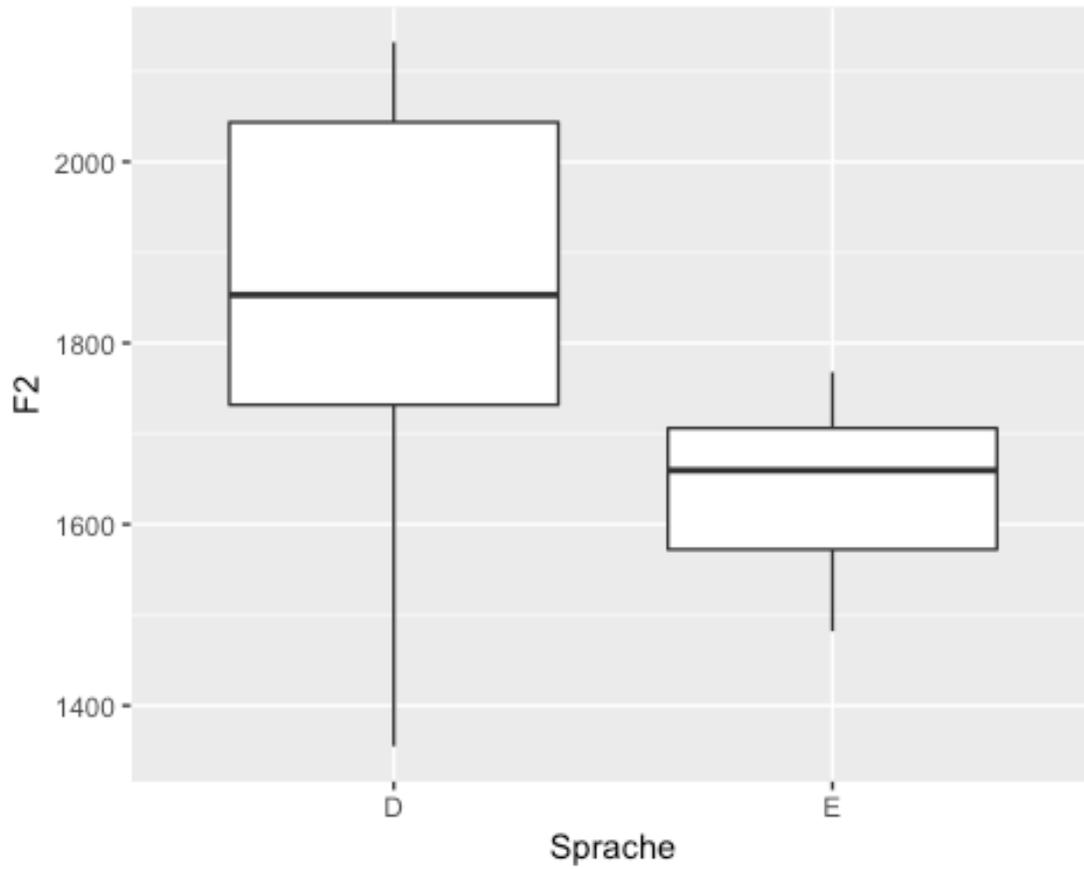
## Unterscheiden sich Deutsche und Engländer in F2 von /e/?

Jede Stufe des unabhängigen between-Faktors wird einmal pro Versuchsperson belegt (between-Faktor: Die Versuchspersonen sind entweder Deutsch oder Englisch):

```
with(v.df, table(Vpn, Sprache))
```

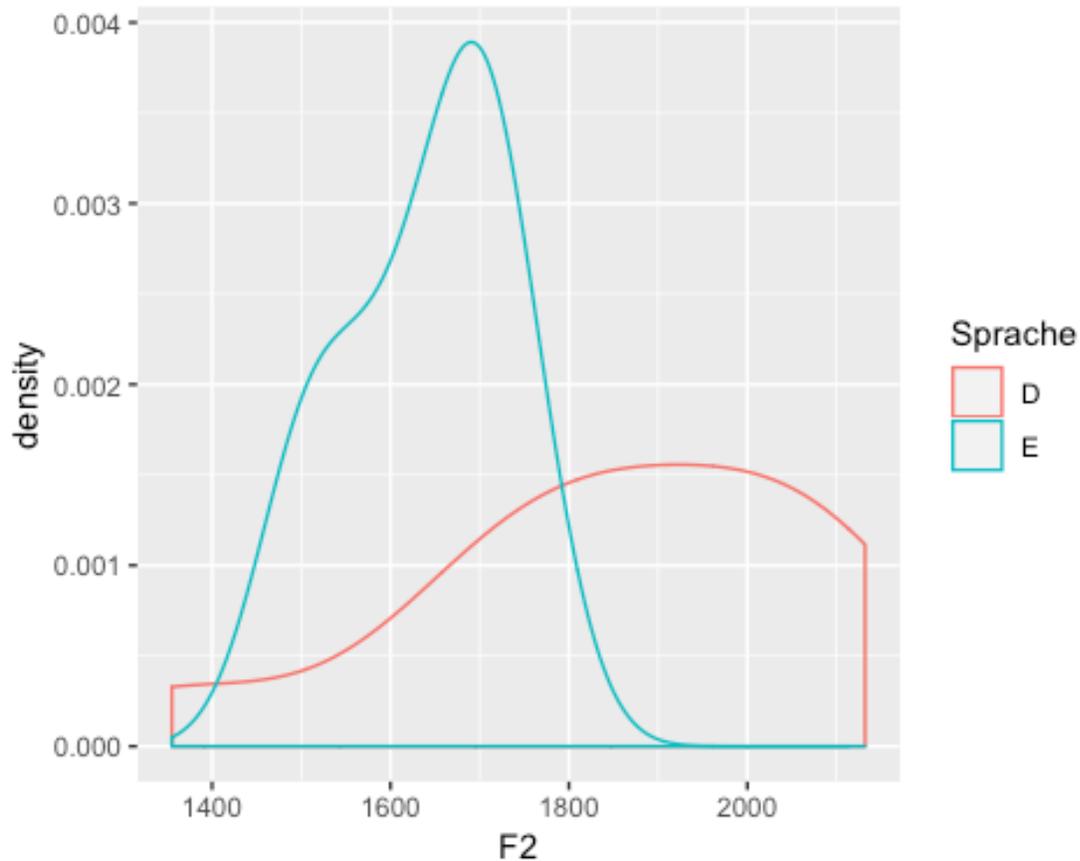
```
##      Sprache
## Vpn   D E
##  S1   1 0
##  S10  1 0
##  S11  0 1
##  S12  0 1
##  S13  0 1
##  S14  0 1
##  S15  0 1
##  S16  0 1
##  S17  0 1
##  S18  0 1
##  S19  0 1
##   S2   1 0
##  S20  0 1
##   S3   1 0
##   S4   1 0
##   S5   1 0
##   S6   1 0
##   S7   1 0
##   S8   1 0
##   S9   1 0
```

```
ggplot(v.df) +  
  aes(y = F2, x = Sprache) +  
  geom_boxplot()
```



Oder:

```
ggplot(v.df) +  
  aes(x = F2, col = Sprache) +  
  geom_density()
```



t-Test:

```
t.test(F2 ~ Sprache, data = v.df)

##
## Welch Two Sample t-test
##
## data:  F2 by Sprache
## t = 2.688, df = 11.806, p-value = 0.01999
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  40.29511 388.50489
## sample estimates:
## mean in group D mean in group E
##      1853.2      1638.8
```

**F2 wird signifikant von der Sprache beeinflusst ( $t[11.8] = 2.7, p < 0.05$ ).**

Oder:

ANOVA mit between-subjects-factor:

```
ezANOVA(v.df, .(F2), .(Vpn), between = .(Sprache))

## $ANOVA
##   Effect DFn DFd    F      p p<.05    ges
```

```
## 1 Sprache 1 18 7.22526 0.01503014 * 0.2864296
##
## $`Levene's Test for Homogeneity of Variance`
## DFn DFd SSn SSd F p p<.05
## 1 1 18 48807.2 213558.1 4.113773 0.05759797
```

**F2 wird signifikant von der Sprache beeinflusst ( $F[1,18] = 7.2$ ,  $p < 0.05$ ).**

Eine Anova mit between-Faktor wird unter der Annahme durchgeführt, dass sich die Varianzen der Stufen nicht unterscheiden. Daher wird der Levene-Test (wenn  $p > 0.05$ , dann ist die Anova berechtigt) automatisch durchgeführt.

Insofern bekommt man das gleiche Ergebnis mit einem  $t$ -test nur unter der Annahme, dass sich die Varianzen in den Stufen nicht unterscheiden:

```
t.test(F2 ~ Sprache, data = v.df, var.equal=T)

##
## Two Sample t-test
##
## data: F2 by Sprache
## t = 2.688, df = 18, p-value = 0.01503
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## 46.82548 381.97452
## sample estimates:
## mean in group D mean in group E
## 1853.2 1638.8

t.test(F2 ~ Sprache, data = v.df, var.equal=T)$statistic

## t
## 2.687984

t.test(F2 ~ Sprache, data = v.df, var.equal=T)$statistic^2

## t
## 7.22526

ezANOVA(v.df, .(F2), .(Vpn), between = .(Sprache))$ANOVA$F

## [1] 7.22526
```

## Zwei Faktoren

```
head(dg); names(dg)
```

```
## F2 Region Gen Vpn
## 1 1806 A m S1
## 2 1644 A m S2
## 3 1512 A m S3
## 4 1767 A m S4
```

```
## 5 1812      A   m   S5
## 6 1990      A   m   S6

## [1] "F2"      "Region" "Gen"    "Vpn"
```

### Inwiefern wird F2 vom Dialekt *und* vom Geschlecht beeinflusst?

```
with(dg, table(Vpn, interaction(Region, Gen)))
```

```
##
## Vpn   A.m B.m C.m A.w B.w C.w
## S1    1  0  0  0  0  0
## S10   1  0  0  0  0  0
## S11   0  1  0  0  0  0
## S12   0  1  0  0  0  0
## S13   0  1  0  0  0  0
## S14   0  1  0  0  0  0
## S15   0  1  0  0  0  0
## S16   0  1  0  0  0  0
## S17   0  1  0  0  0  0
## S18   0  1  0  0  0  0
## S19   0  1  0  0  0  0
## S2    1  0  0  0  0  0
## S20   0  1  0  0  0  0
## S21   0  0  1  0  0  0
## S22   0  0  1  0  0  0
## S23   0  0  1  0  0  0
## S24   0  0  1  0  0  0
## S25   0  0  1  0  0  0
## S26   0  0  1  0  0  0
## S27   0  0  1  0  0  0
## S28   0  0  1  0  0  0
## S29   0  0  1  0  0  0
## S3    1  0  0  0  0  0
## S30   0  0  1  0  0  0
## S31   0  0  0  1  0  0
## S32   0  0  0  1  0  0
## S33   0  0  0  1  0  0
## S34   0  0  0  1  0  0
## S35   0  0  0  1  0  0
## S36   0  0  0  1  0  0
## S37   0  0  0  1  0  0
## S38   0  0  0  1  0  0
## S39   0  0  0  1  0  0
## S4    1  0  0  0  0  0
## S40   0  0  0  1  0  0
## S41   0  0  0  0  1  0
## S42   0  0  0  0  1  0
## S43   0  0  0  0  1  0
## S44   0  0  0  0  1  0
## S45   0  0  0  0  1  0
## S46   0  0  0  0  1  0
```

```
## S47 0 0 0 0 1 0
## S48 0 0 0 0 1 0
## S49 0 0 0 0 1 0
## S5 1 0 0 0 0 0
## S50 0 0 0 0 1 0
## S51 0 0 0 0 0 1
## S52 0 0 0 0 0 1
## S53 0 0 0 0 0 1
## S54 0 0 0 0 0 1
## S55 0 0 0 0 0 1
## S56 0 0 0 0 0 1
## S57 0 0 0 0 0 1
## S58 0 0 0 0 0 1
## S59 0 0 0 0 0 1
## S6 1 0 0 0 0 0
## S60 0 0 0 0 0 1
## S7 1 0 0 0 0 0
## S8 1 0 0 0 0 0
## S9 1 0 0 0 0 0
```

### Between

- Gender

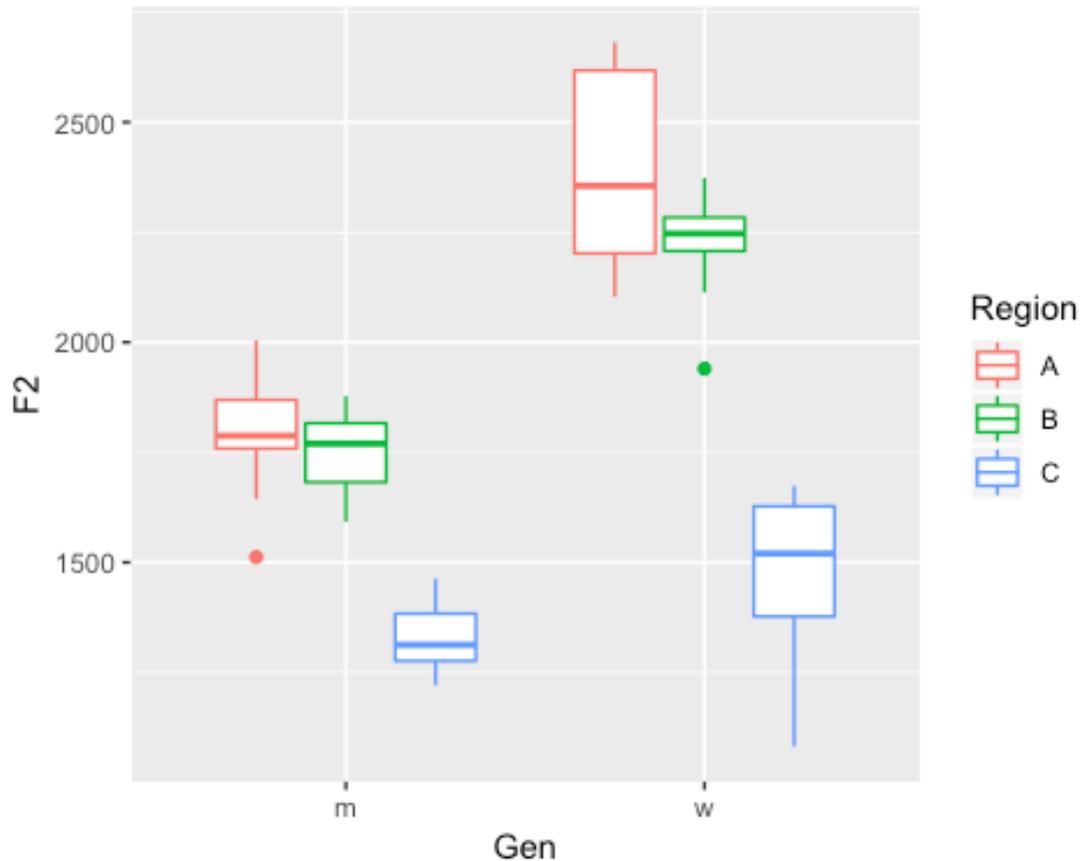
### Within

- Region

### Bei 2 Faktoren gibt es immer 3 Fragen:

- **Frage zu Faktor 1:** Hat *Gender* einen Einfluss auf F2?
- **Frage zu Faktor 2:** Hat *Region* einen Einfluss auf F2?
- **Frage zur Interaktion:** Gibt es eine *Interaktion* zwischen *Region* und *Gender*? = Ist der Unterschied zwischen männlich und weiblich derselbe in allen 3 Regionen?

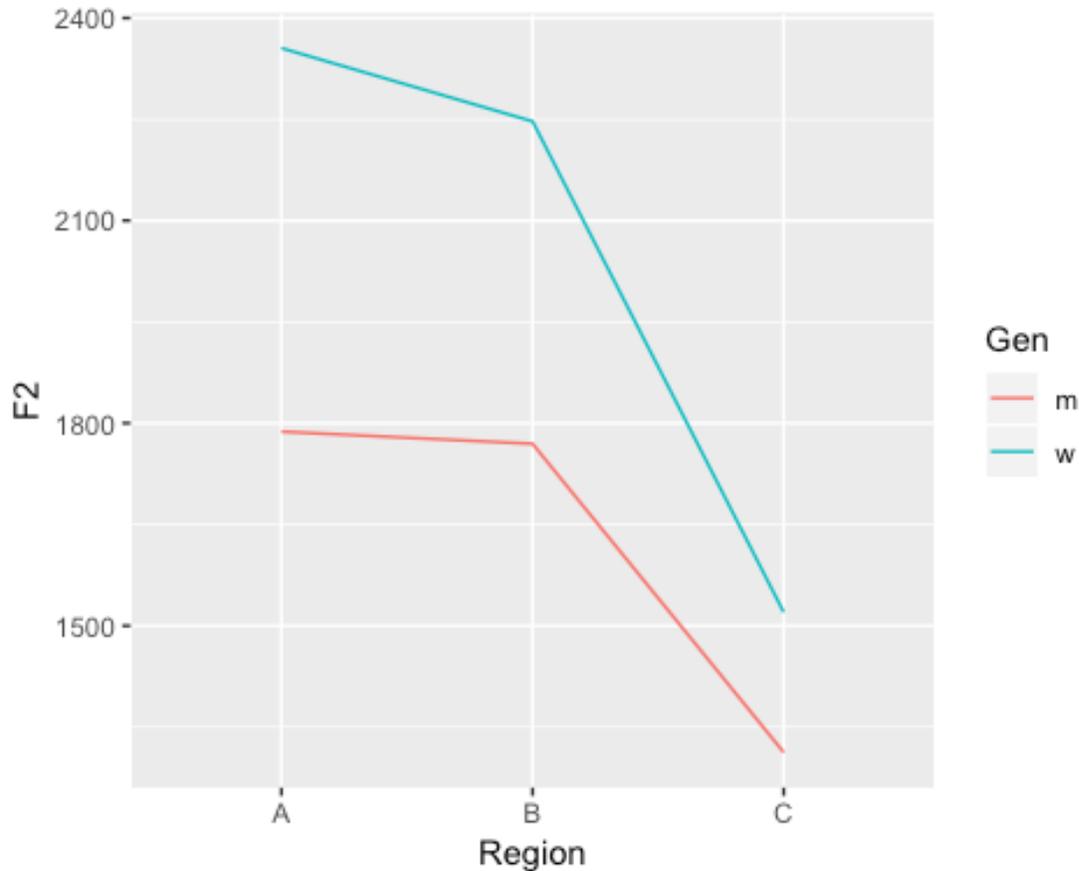
```
ggplot(dg) +
  aes(y = F2, x = Gen, colour = Region) +
  geom_boxplot()
```



**Interaktion:** *ist der Unterschied zwischen männlich und weiblich ähnlich in den 3 Regionen?*

Wenn ja, dann müsste der Abstand zwischen den *männlich-weiblich*-Medianen ähnlich sein (d.h. diese Linien müssten mehr oder weniger parallel zu einander sein):

```
# zunächst den Median pro Faktorstufenkombination ermitteln:
dg.m = dg %>%
  group_by(Gen, Region) %>%
  summarise(F2 = median(F2))
# diese Medianwerte dann plotten
ggplot(dg.m) +
  aes(y = F2, x = Region, group=Gen, colour = Gen) +
  geom_line()
```



```
ezANOVA(dg, .(F2), .(Vpn), between =.(Region, Gen))
```

```
## $ANOVA
##      Effect DFn DFd      F      p p<.05      ges
## 1      Region    2   54 119.63719 1.439560e-20 * 0.8158721
## 2         Gen    1   54 106.14696 2.353977e-14 * 0.6628097
## 3 Region:Gen    2   54  12.08336 4.602985e-05 * 0.3091690
##
## $`Levene's Test for Homogeneity of Variance`
##  DFn DFd      SSn      SSd      F      p p<.05
##  1    5   54 106131.6 502847.1 2.279462 0.05944605
```

**F2 wurde signifikant von der Region ( $F[2,54] = 119.6$ ,  $p < 0.001$ ) und von Geschlecht ( $F[1,54] = 106.1$ ,  $p < 0.001$ ) beeinflusst, und es gab eine signifikante Interaktion zwischen diesen Faktoren ( $F[2,54] = 12.1$ ,  $p < 0.001$ ).**

## Post-hoc *t*-Tests

Wenn eine Interaktion vorliegt, sollte durch *t*-Tests geprüft werden, ob sich alle Paare von Stufen-Kombinationen in der abhängigen Variable (hier F2) unterscheiden.

Die Anzahl dieser Tests berechnet sich nach:

- Region: 3 Stufen

- Geschlecht: 2 Stufen

=  $3 * 2 = 6$  Stufen.

Alle Paarungen davon:

```
factorial(6)/(factorial(4) * factorial (2))
```

```
## [1] 15
```

$6!/(4! \times 2!) = 15$  Testpaare!

Post-hoc-t-Tests lassen sich durchführen mittels der Funktion `phoc()`, die wir ganz am Anfang des Skriptes geladen hatten (mit `source(file.path(pfadu, "phoc.txt"))`):

```
phoc(dg, .(F2), .(Vpn), .(Region, Gen))
```

```
## $res
##           t           df      prob-adj
## A:m-B:m    0.8313356 15.22192 1.000000e+00
## A:m-C:m    8.7155048 13.98591 7.531888e-06
## A:m-A:w   -7.1586378 15.68960 3.814827e-05
## A:m-B:w   -7.0876370 17.28901 2.482025e-05
## A:m-C:w    4.1291502 16.66330 1.092264e-02
## B:m-C:m   10.6837180 17.65040 5.898958e-08
## B:m-A:w   -8.5319197 12.11771 2.708776e-05
## B:m-B:w   -9.8137671 16.97522 3.098841e-07
## B:m-C:w    3.9943383 12.84208 2.345945e-02
## C:m-A:w  -14.3108625 11.38030 1.881869e-07
## C:m-B:w  -19.4274325 15.79614 2.840451e-11
## C:m-C:w   -2.1074735 11.95523 8.530528e-01
## A:w-B:w    2.2029457 13.88744 6.749777e-01
## A:w-C:w    9.8529861 17.77397 1.896196e-07
## B:w-C:w   10.2391336 14.86067 5.992353e-07
##
## $name
## [1] "A:m-B:m" "A:m-C:m" "A:m-A:w" "A:m-B:w" "A:m-C:w" "B:m-C:m" "B:m-A:w"
## [8] "B:m-B:w" "B:m-C:w" "C:m-A:w" "C:m-B:w" "C:m-C:w" "A:w-B:w" "A:w-C:w"
## [15] "B:w-C:w"
##
## $paired
## [1] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [12] FALSE FALSE FALSE FALSE
##
## $bonf
## [1] 15
```

- `$res`: die Ergebnisse der t-tests
- `$name`: die Testpaare
- `$paired`: ob ein gepaarter oder ungepaarter t-test durchgeführt wurde
- `$bonf`: Anzahl der durchgeführten Tests (= Bonferroni-Faktor)

## Bonferroni-Korrektur

prob-adj in \$res ist der durch die sogenannte Bonferroni-Korrektur angepasste  $p$ -Wert.

Hierbei wird der Wahrscheinlichkeitswert der individuellen  $t$ -Tests mit der Anzahl der theoretisch möglichen Testkombinationen (= der Bonferroni-Faktor: 15 in diesem Fall) multipliziert. Sollte dies einen Wert  $> 1$  ergeben, wird der Wert einfach auf 1 (= 100%) gesetzt.

Der Grund liegt im Problem des multiplen Testens in derselben Stichprobe: Je mehr post-hoc Tests durchgeführt werden, um so wahrscheinlicher ist es, dass einer von den vielen Tests durch reinen Zufall signifikant sein wird.

Die Bonferroni-Korrektur ist eine Maßnahme dagegen.

## Auswahl relevanter post-hoc $t$ -Tests

Nicht alle  $t$ -Tests werden benötigt, sondern eher nur *Vergleiche zwischen Stufen* von **einem** Faktor, wenn die Stufen aller anderen Faktoren **konstant** sind. So interessiert es uns beispielsweise *nicht*, ob sich die Werte von Männern aus der Region A von den Werten der Frauen aus Region B unterscheiden oder nicht.

```
vok.ph = phoc(dg, .(F2), .(Vpn), .(Region, Gen))
```

- **Unterscheiden sich die Regionen innerhalb desselben Geschlechts? (Region variiert, Geschlecht ist konstant).**

```
phsel(vok.ph$res, 1)
```

```
##           t           df      prob.adj
## A:m-B:m  0.8313356 15.22192 1.000000e+00
## A:m-C:m  8.7155048 13.98591 7.531888e-06
## B:m-C:m 10.6837180 17.65040 5.898958e-08
## A:w-B:w  2.2029457 13.88744 6.749777e-01
## A:w-C:w  9.8529861 17.77397 1.896196e-07
## B:w-C:w 10.2391336 14.86067 5.992353e-07
```

- **Unterscheiden sich Männer und Frauen in derselben Region? (Geschlecht variiert, Region ist konstant).**

```
phsel(vok.ph$res, 2)
```

```
##           t           df      prob.adj
## A:m-A:w -7.158638 15.68960 3.814827e-05
## B:m-B:w -9.813767 16.97522 3.098841e-07
## C:m-C:w -2.107474 11.95523 8.530528e-01
```

- **Aber *nicht*, wenn beide Faktoren variieren.**

## Runden der adjustierten $p$ -Werte

Noch ersichtlicher werden die Signifikanz der relevanten Tests, wenn auf z.B. 3 Zahlen aufgerundet wird:

```
p1 = phsel(vok.ph$res, 1)
round(p1, 3)

##           t      df prob.adj
## A:m-B:m  0.831 15.222   1.000
## A:m-C:m  8.716 13.986   0.000
## B:m-C:m 10.684 17.650   0.000
## A:w-B:w  2.203 13.887   0.675
## A:w-C:w  9.853 17.774   0.000
## B:w-C:w 10.239 14.861   0.000
```

```
p2 = phsel(vok.ph$res, 2)
round(p2, 3)

##           t      df prob.adj
## A:m-A:w -7.159 15.690   0.000
## B:m-B:w -9.814 16.975   0.000
## C:m-C:w -2.107 11.955   0.853
```

**Post-hoc Bonferroni-adjusted  $t$ -Tests zeigten signifikante F2-Unterschiede zwischen A vs C ( $p < 0.001$ ) und zwischen B vs C ( $p < 0.001$ ), jedoch nicht zwischen A vs. B. F2 von Männern und Frauen unterschieden sich signifikant für die Regionen A ( $p < 0.001$ ) und B ( $p < 0.001$ ), jedoch nicht für C.**

## Zwei Bedingungen für die Durchführung der Varianzanalyse

### Wiederholungen in within-Stufen

Wenn es  $n$  within-Stufen gibt, dann müssen es  $n$  Werte pro Vpn sein, also pro Vpn ein Wert pro within-Stufe z.B:

*Englische und spanische Vpn produzierten /i, e, a/ zu 2 Sprechgeschwindigkeiten*

**Within:** *Vokal* (3 Stufen) und *Sprechgeschwindigkeit* (2 Stufen)

Daher:  $3 \times 2 = 6$  within-Werte pro Versuchsperson (ein Wert pro within-Stufe pro Vpn).

Jedoch haben die meisten phonetischen Untersuchungen mehrere Werte pro within-Stufe, da in diesen Untersuchungen häufig Wiederholungen eingesetzt werden. Als Beispiel: jede Vpn. erzeugte /i, e, a/ zu einer langsamen und schnellen Sprechgeschwindigkeit, und das jeweils 10 Mal.

Wiederholungen in derselben within-Stufe sind in einer ANOVA aber nicht zulässig und müssen gemittelt werden – damit wir pro Vpn. einen Wert pro within-Stufe haben (6 Mittelwerte pro Vpn. in diesem Beispiel).

```
summary(ssb)

##           F2           Alter           Wort           Vpn
## Min.      : 6.925    alt :179    swoop:119    arkn   : 30
## 1st Qu.:10.342    jung:180    used :120    frwa   : 30
```

```
## Median :12.649          who'd:120   gisa   : 30
## Mean   :12.108          jach   : 30
## 3rd Qu.:14.158          jeny   : 30
## Max.   :15.378          kapo   : 30
##                                     (Other):179
```

In einer Untersuchung zur /u/-Frontierung im Standardenglischen wurde von 12 Sprecherinnen (6 *alt*, 6 *jung*) F2 zum zeitlichen Mittelpunkt in drei verschiedenen /u/-Wörtern erhoben (*used*, *swoop*, *who'd*). Jedes Wort ist von jeder Vpn. 10 Mal erzeugt worden.

### Inwiefern wird F2 vom Alter und Wort beeinflusst?

Wieviele Werte pro Vpn. dürfen in der ANOVA vorkommen? 3!

Wieviele Werte darf es insgesamt in der ANOVA geben? 36!

```
dim(ssb)
## [1] 359  4

with(ssb, table(Vpn, interaction(Wort, Alter)))

##
## Vpn      swoop.alt used.alt who'd.alt swoop.jung used.jung who'd.jung
## arkn         10      10      10         0         0         0
## elwi          9      10      10         0         0         0
## frwa         10      10      10         0         0         0
## gisa         10      10      10         0         0         0
## jach          0         0         0        10        10        10
## jeny          0         0         0        10        10        10
## kapo          0         0         0        10        10        10
## mapr         10      10      10         0         0         0
## nata         10      10      10         0         0         0
## rohi          0         0         0        10        10        10
## rusy          0         0         0        10        10        10
## shle          0         0         0        10        10        10
```

Über die Wort-Wiederholungen pro Versuchsperson mit `group_by()` mitteln ("Alter" steht mit in `group_by()`, um die Altersinformation nicht zu verlieren):

```
ssbm = ssb %>%
  group_by(Wort,Alter,Vpn) %>%
  summarise(F2 = mean(F2))
dim(ssbm); head(ssbm)

## [1] 36  4

## # A tibble: 6 x 4
## # Groups:   Wort, Alter [1]
## Wort Alter Vpn      F2
## <fct> <fct> <fct> <dbl>
## 1 swoop alt   arkn  10.5
```

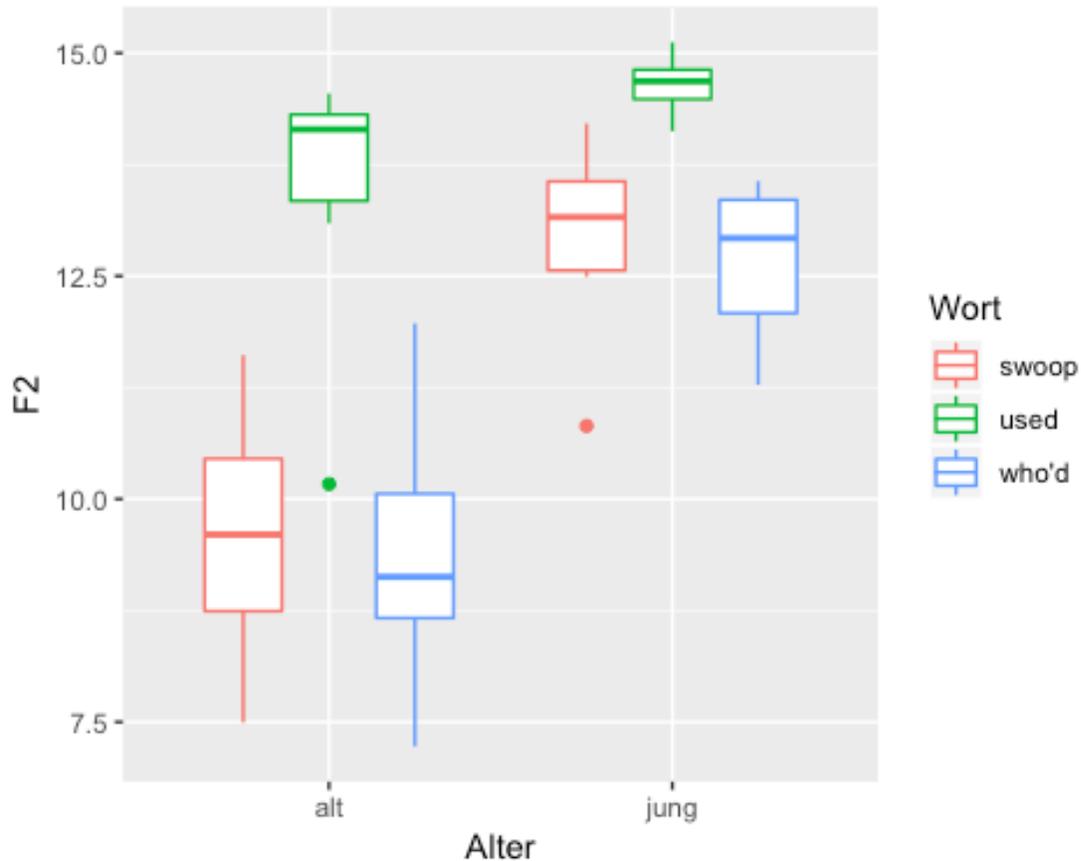
```
## 2 swoop alt elwi 8.66
## 3 swoop alt frwa 7.50
## 4 swoop alt gisa 11.6
## 5 swoop alt mapr 10.2
## 6 swoop alt nata 8.98
```

```
with(ssbm, table(Vpn, interaction(Wort, Alter)))
```

```
##
## Vpn swoop.alt used.alt who'd.alt swoop.jung used.jung who'd.jung
## arkn 1 1 1 0 0 0
## elwi 1 1 1 0 0 0
## frwa 1 1 1 0 0 0
## gisa 1 1 1 0 0 0
## jach 0 0 0 1 1 1
## jeny 0 0 0 1 1 1
## kapo 0 0 0 1 1 1
## mapr 1 1 1 0 0 0
## nata 1 1 1 0 0 0
## rohi 0 0 0 1 1 1
## rusy 0 0 0 1 1 1
## shle 0 0 0 1 1 1
```

Erst jetzt, mit den gemittelten Daten, sollte man eine Abbildung erstellen:

```
ggplot(ssbm) +
  aes(y = F2, x = Alter, colour = Wort) +
  geom_boxplot()
```



```
ezANOVA(ssbm,
        .(F2),
        .(Vpn),
        .(Wort),
        between = .(Alter))
```

```
## $ANOVA
##      Effect DFn DFd      F      p p<.05      ges
## 2      Alter   1  10 14.876957 3.175409e-03 * 0.5519903
## 3       Wort   2  20 78.505534 3.390750e-10 * 0.5742513
## 4 Alter:Wort   2  20  9.890888 1.031474e-03 * 0.1452519
##
## $`Mauchly's Test for Sphericity`
##      Effect      W      p p<.05
## 3       Wort 0.5423826 0.06373468
## 4 Alter:Wort 0.5423826 0.06373468
##
## $`Sphericity Corrections`
##      Effect      GGe      p[GG] p[GG]<.05      HFe      p[HF]
## 3       Wort 0.6860511 1.340736e-07 * 0.7587667 3.342362e-08
## 4 Alter:Wort 0.6860511 4.370590e-03 * 0.7587667 3.120999e-03
##      p[HF]<.05
## 3      *
## 4      *
```

## Sphericity-Korrektur

*Sphericity* ist die Annahme, dass die Unterschiede zwischen den Stufen eines within-Faktors **dieselbe Varianz haben**.

Wenn Sphericity nicht gegeben ist, werden **die Wahrscheinlichkeiten durch Änderungen in den Freiheitsgraden nach oben** gesetzt.

Dieses Problem kommt *nur dann* vor, wenn **ein within-Faktor mehr als 2 Stufen** hat.

**Man soll grundsätzlich immer für Sphericity korrigieren**, wenn Sphericity Corrections in der Ausgabe von `ezANOVA()` erscheint.

Die betroffenen Freiheitsgrade werden mit

- dem *Greenhouse-Geisser-Epsilon* (GGe) multipliziert, wenn es unter 0.75 liegt,
- sonst (also wenn das GGe > 0.75) mit dem *Huynh-Feldt-Epsilon* (HFe):
- sollte in diesem letzten Fall das HFe > 1 sein, dann einfach die ursprünglichen Freiheitsgrade nehmen d.h. keine Korrektur einsetzen.

Für

```
ezANOVA(ssbm,
        .(F2),
        .(Vpn),
        .(Wort),
        between = .(Alter))

## $ANOVA
##      Effect DFn DFd      F      p p<.05      ges
## 2      Alter   1  10 14.876957 3.175409e-03 * 0.5519903
## 3       Wort   2  20 78.505534 3.390750e-10 * 0.5742513
## 4 Alter:Wort   2  20  9.890888 1.031474e-03 * 0.1452519
##
## $`Mauchly's Test for Sphericity`
##      Effect      W      p p<.05
## 3       Wort 0.5423826 0.06373468
## 4 Alter:Wort 0.5423826 0.06373468
##
## $`Sphericity Corrections`
##      Effect      GGe      p[GG] p[GG]<.05      HFe      p[HF]
## 3       Wort 0.6860511 1.340736e-07 * 0.7587667 3.342362e-08
## 4 Alter:Wort 0.6860511 4.370590e-03 * 0.7587667 3.120999e-03
##      p[HF]<.05
## 3          *
## 4          *
```

Wort:  $F[2,20] \rightarrow F[2 * 0.6860511, 20 * 0.6860511] = F[1.4, 13.7]$

Alter × Wort Interaktion:  $F[2,20] \rightarrow F[1.4, 13.7]$

Die neuen damit verbundenen Wahrscheinlichkeiten sind  $p[GG]$  (wenn mit  $GGe$  multipliziert wurde), sonst  $p[HF]$ .

Das sind die Wahrscheinlichkeiten mit den korrigierten Freiheitsgraden (vergleiche mit  $p[GG]$  für Wort und die Alter:Wort-Interaktion):

```
1 - pf(78.505534, 2 * 0.6860511, 20 * 0.6860511)
```

```
## [1] 1.340735e-07
```

```
1 - pf(9.8908882, 2 * 0.6860511, 20 * 0.6860511)
```

```
## [1] 0.004370589
```

**Alter ( $F[1,10] = 14.9$ ,  $p < 0.001$ ), Wort ( $F[1.4, 13.7] = 78.5$ ,  $p < 0.001$ ) sowie die Interaktion von Wort und Alter ( $F[1.4, 13.7] = 9.9$ ,  $p < 0.01$ ) hatten einen signifikanten Einfluss auf  $F2$ .**