

Die Varianzanalyse

Jonathan Harrington

Anova und t-test

t-test oder ANOVA (Analysis of Variance = Varianzanalyse)

Ein Faktor mit 2 Stufen

- Hat Geschlecht einen Einfluss auf die Dauer?

ANOVA

Ein Faktor mit mehr als 2 Stufen; oder mehr als ein Faktor

- 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?

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

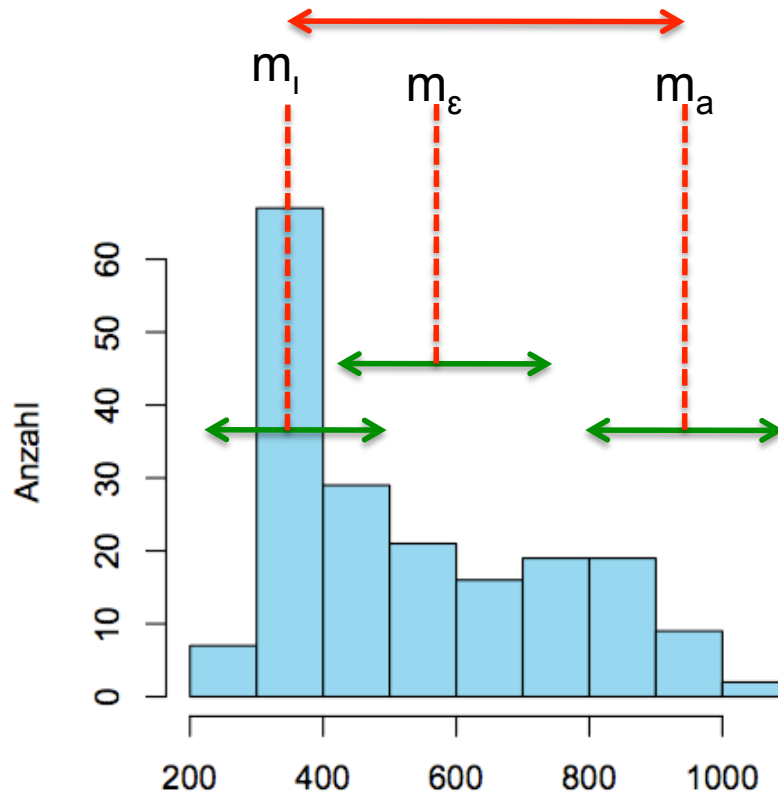
z.B. F1 von drei Vokalkategorien, /I,ε,a/.

innerhalb: Es gibt eine **randomisierte Variation von F1** innerhalb jeder Stufe (F1 von /I/ variiert, F1 von /ε/ variiert, F1 von /a/ variiert).

zwischen: F1 variiert, weil es eine **systematische Variation** zwischen den Verteilungen der Vokalkategorien gibt: die Werte von /I/, /ε/, und /a/ liegen in ganz unterschiedlichen F1-Bereichen, und je unterschiedlicher sie sind, **umso größer wird diese Varianz im Verhältnis zu der willkürlichen, randomisierten Varianz innerhalb der Stufen sein.**

Was ist die Varianzanalyse?

F1-Verteilung, drei Vokale



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

Ist F signifikant größer als 1?

Within-subject Faktor

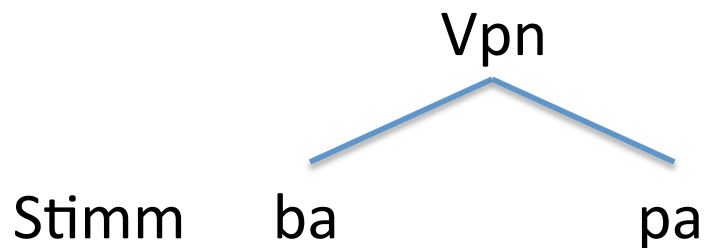
- für Analysen innerhalb derselben Person
- eine Messung pro Vpn. pro Stufe

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

Abhängige Variable: VOT

Within-Faktor: Stimmhaftigkeit

Ein Wert für /ba/ ein Wert für /pa/ pro Vpn



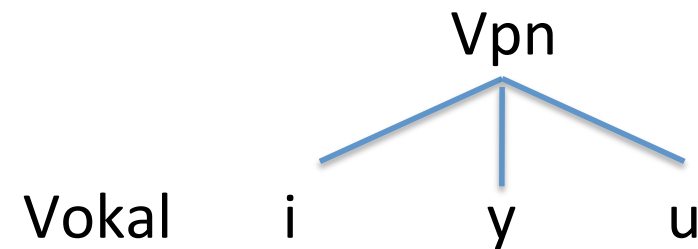
gepaarter t-test oder Anova

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

Abhängige Variable: F2

Within-Faktor: Vokal

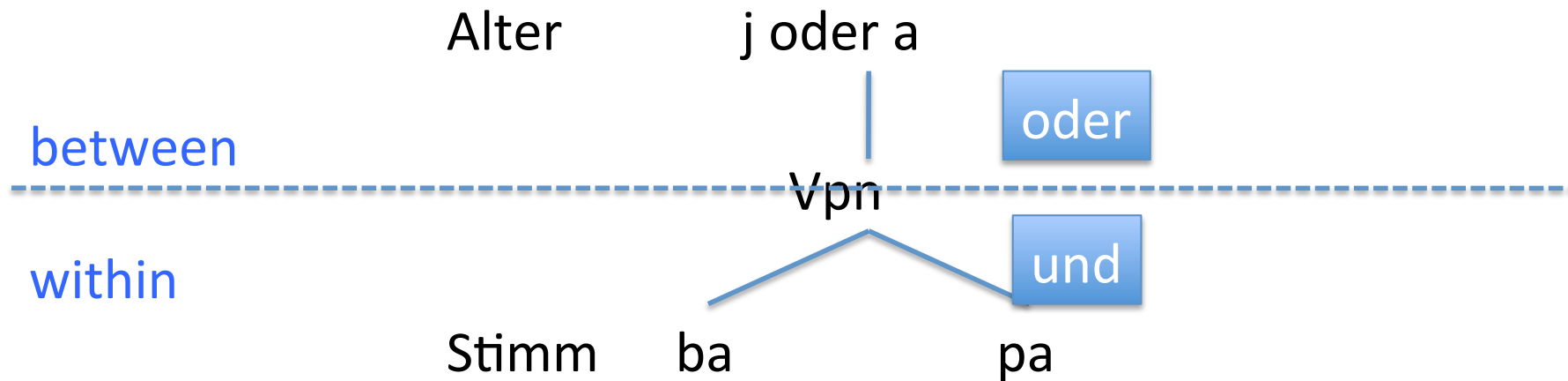
Ein Wert für /i/ ein Wert für /y/,
ein Wert für /u/ pro Vpn



Anova

Between-subjects Faktor

beschreibt meistens eine Eigenschaft der Vpn. Z.B.
Muttersprache (englisch *oder* deutsch *oder* französisch),
Geschlecht (m *oder* w), Alter (jung *oder* alt) usw.



Within- and between-subjects factors

Die Kieferposition wurde in 3 Vokalen /i, e, a/ und jeweils zu 2 Sprechtempi (langsam, schnell) gemessen. Die Messungen ($3 \times 2 = 6$ 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

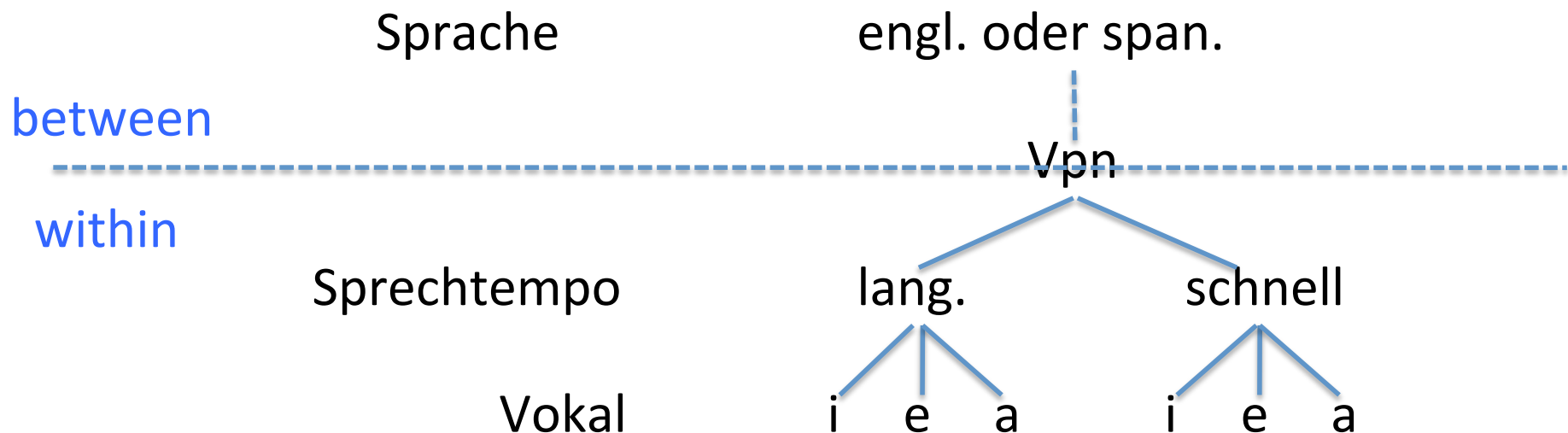
Sprache

Within

Sprechtempo, Vokal

Within- und between-subjects factors

Die Kieferposition wurde in 3 Vokalen /i, e, a/ und jeweils zu 2 Sprechtempi (langsam, schnell) gemessen. Die Messungen sind von 8 mit Muttersprache spanisch, 8 mit Muttersprache englisch aufgenommen worden.



(6 Werte pro Vpn)

Within-Factor: gepaarter t-Test und Anova

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

Abhängige Variable: F2

Within-Faktor: Betonung

Kurz-Format (für den t-Test)

```
b = read.table(file.path(pfadu, "bet.txt"))
```

Lang-Format (wird für den Anova benötigt)

```
blang = read.table(file.path(pfadu, "blang.txt"))
```

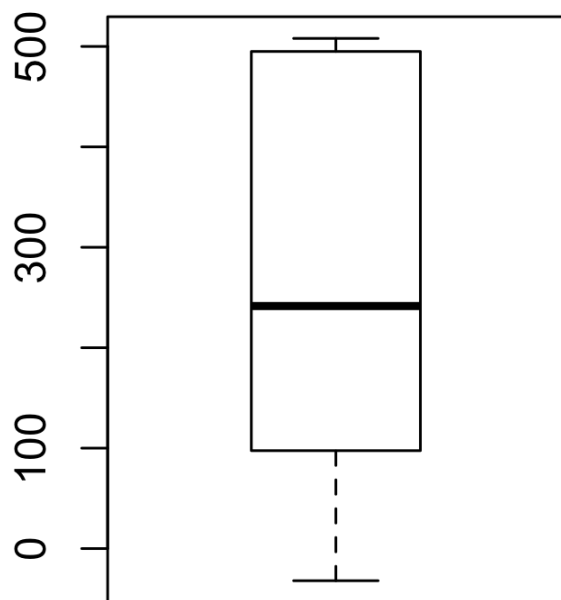
Within-factor: gepaarter t-Test und Anova

Gepaarter t-Test

Was ist die Wahrscheinlichkeit, dass der Mittelwert der Pro-Sprecher-Unterschiede (in F2 zwischen betontem und unbetontem /i/) 0 (Null) sein könnte?

`d = b$betont - b$unbetont`

`boxplot(d)`



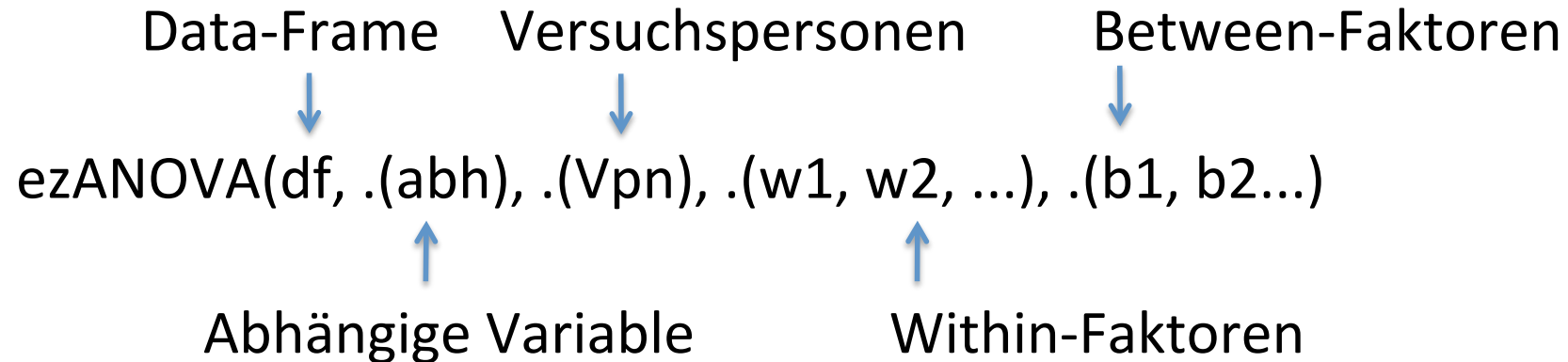
`t.test(d)`

$t = 4.3543$, $df = 11$, $p\text{-value} = 0.001147$

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

Within-factor: gepaarter t-Test und Anova

Lösung mit Anova



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

Effect	DFn	DFd	F	p	p<.05	ges2
Betonung	1	11	18.95986	0.001147148	*	0.4113659

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

Vergleich mit t-test

derselbe Wahrscheinlichkeitswert

der F-Wert ist der t-Wert hoch 2

`t.test(d)`

$t = 4.3543$, $df = 11$, $p\text{-value} = 0.001147$

Between-factor: t-test und Anova

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

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

Abhängige Variable: F2

Between-Faktor: Die Vpn. sind entweder Deutsch oder Englisch

```
table(v.df$Sprache)
```

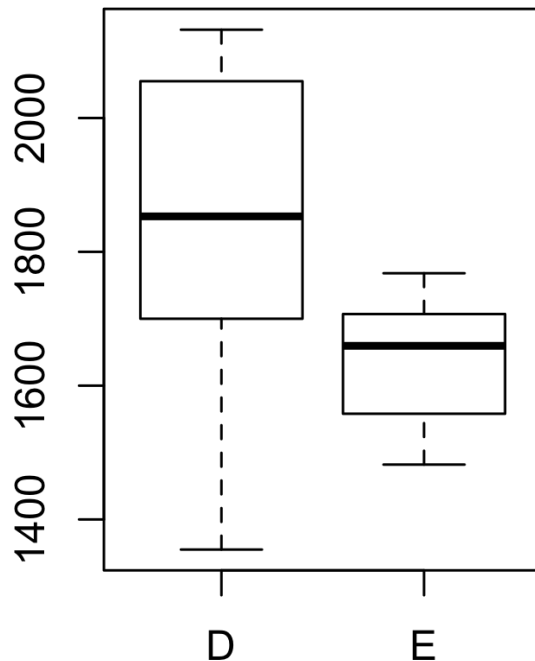
D	E
10	10

Between-factor: t-test und Anova

t-test

Was ist die Wahrscheinlichkeit, dass der Unterschied zwischen den Gruppen-Mittelwerten 0 (Null) sein könnte?

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



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

`t = 2.688, df = 11.806, p-value = 0.01999`

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

Between-factor: t-Test und Anova

Anova

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

```
$ANOVA
Effect  DFn  DFd      F      p  p<.05      ges1
Sprache    1   18  7.22526  0.01503014  *  0.2864296
```

Die Sprache hatte einen signifikanten Einfluss auf F2
($F[1,18] = 7.2, p < 0.05$).

Between-factor: t-Test und Anova

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

```
$`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
```

den man übrigens auch bekommt mit

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

Insofern bekommt man das gleiche Ergebnis mit einem t-test
unter dieser Annahme der gleichen Varianzen:

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

$t = 2.688$, $df = 18$, $p\text{-value} = 0.01503$

der F-Wert ist der t-Wert hoch 2

derselbe Wahrscheinlichkeitswert

```
$ANOVA  
Effect  DFn  DFd      F      p  p<.05      ges1  
Sprache    1   18 7.22526 0.01503014      * 0.2864296
```

Zwei Faktoren

F2 Daten, 60 Sprecher, 30 m, 30 w, drei Dialekte: es gibt einen Wert pro Sprecher.

```
dg = read.table(file.path(pfadu, "dg.txt"))
```

```
head(dg)
```

```
names(dg)          "F2"          "Region" "Gen"      "Vpn"
```

```
table(dg$Gen, dg$Region)
```

	A	B	C
m	10	10	10
w	10	10	10

between/within?

Gender between

Region between

Zwei Faktoren

Hat Gender einen Einfluss auf F2?

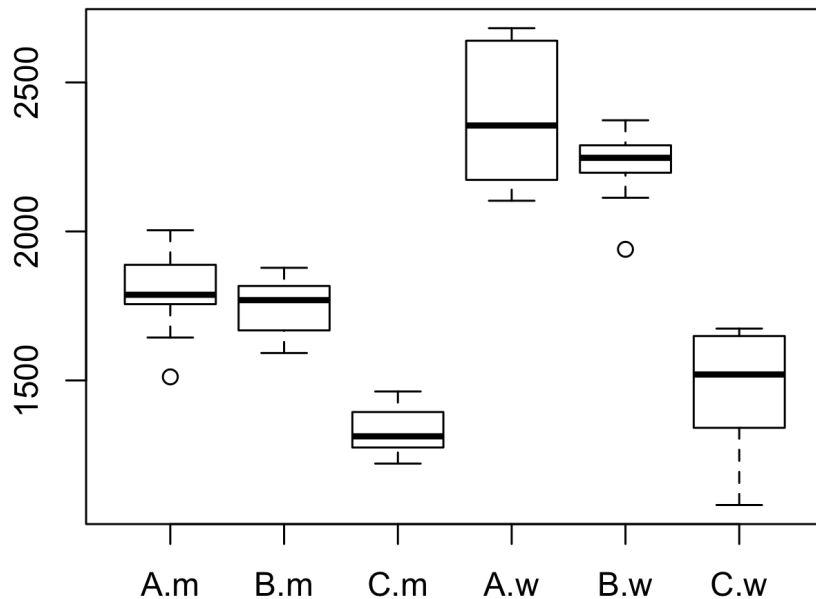
Hat Region einen Einfluss auf F2?

Gibt es eine Interaktion zwischen
Region und Gender? = Ist der
Unterschied zwischen männlich und
weiblich derselbe in allen 3 Regionen?

Zwei Faktoren

Boxplot Abbildung

`boxplot(F2 ~ Gen * Region, data = dg)`

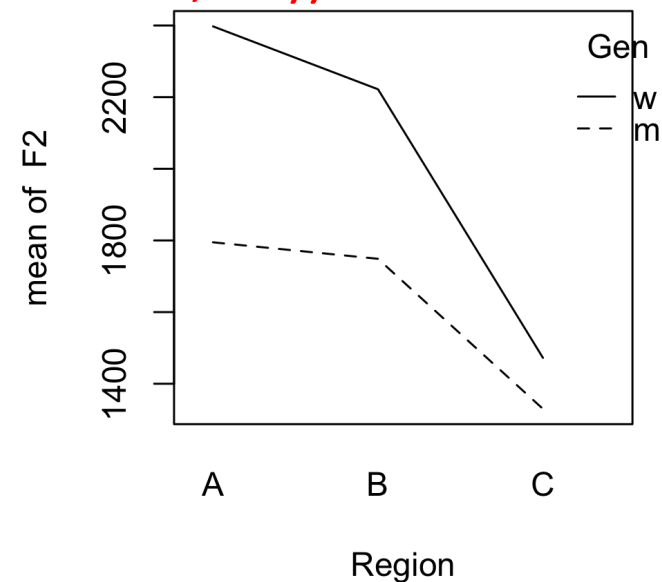


Hat Region einen Einfluss auf F2?

Hat Gender einen Einfluss auf F2?

Interaktion-Abbildung

`with(dg,
interaction.plot(Region,
Gen, F2))`



Gibt es eine Interaktion zwischen Region und Gender?

Zwei Faktoren

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

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

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

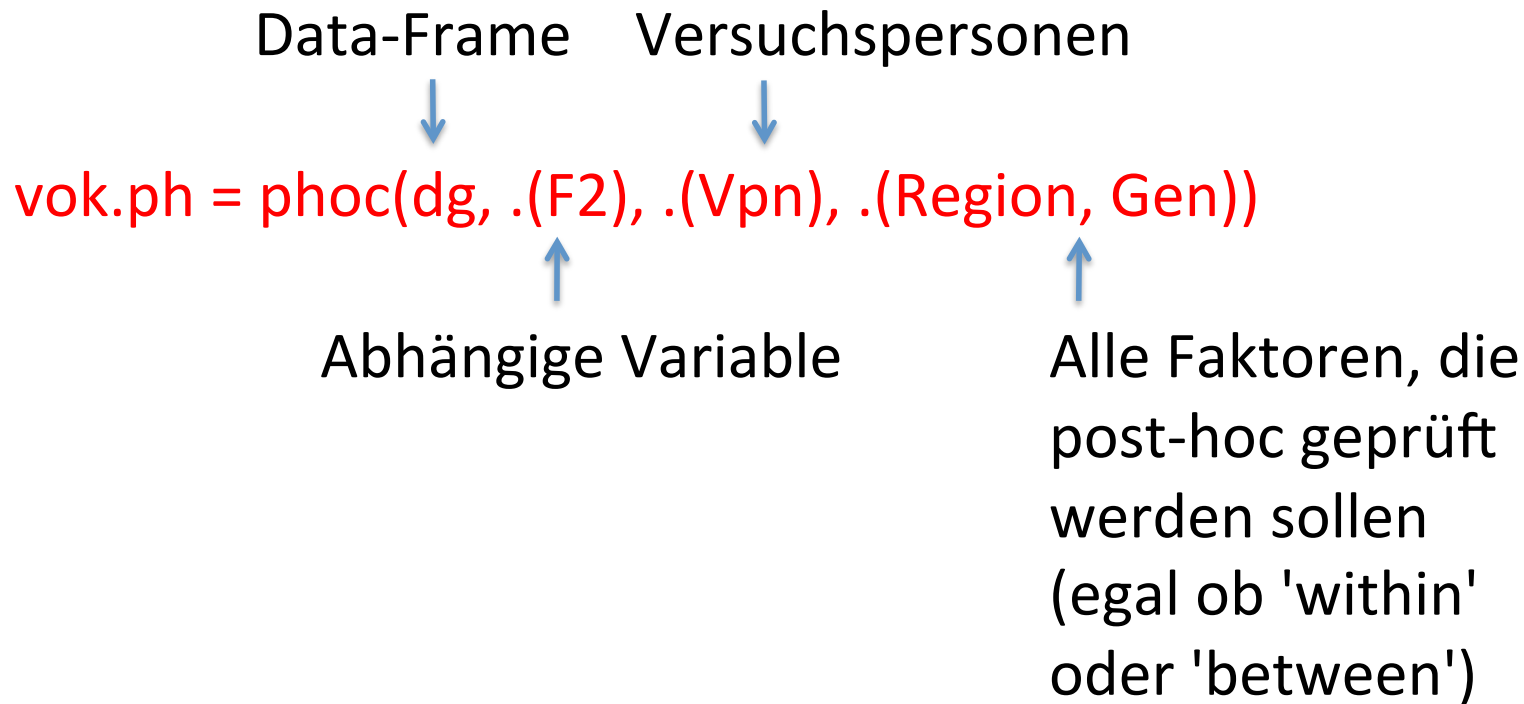
Wenn eine Interaktion vorliegt, sollte geprüft werden, ob sich die Stufen-Kombinationen in dem Parameter (hier F2) unterscheiden.

Dazu werden Bonferroni-korrigierte t-tests für alle Stufen-Kombination durchgeführt

Bonferroni-Korrektur: Der Wahrscheinlichkeitswert der individuellen t-tests wird mit der Anzahl der theoretisch möglichen Testkombinationen multipliziert.

Der Grund: Je mehr Tests wir post-hoc anwenden, um so wahrscheinlicher ist es, dass wir Signifikanzen per Zufall bekommen werden. Die Bonferroni-Korrektur ist eine Maßnahme dagegen

post-hoc t-Tests



`phoc()` und `phsel()` sind Funktionen, die von Jonathan Harrington geschrieben wurden. Siehe

<http://www.phonetik.uni-muenchen.de/~jmh/lehre/sem/ss11/statfort/posthoc.pdf> für ein Beispiel in dem `phoc()` und eine SPSS-Funktion dieselben Ergebnisse liefern

post-hoc t-tests

Wir benötigen aber nur Vergleiche zwischen Stufen von **einem** Faktor, wenn die Stufen aller anderen Faktoren **konstant** sind.

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

A vs B in Männern

A vs B in Frauen

A vs C in Männern

A vs C in Frauen

B vs C in Männern

B vs C in Frauen

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

m vs. w in A

m vs. w in B

m vs. w in C

Aber *nicht* wenn beide Faktoren variieren.

m-A vs. w-B, m-C vs w-A usw.

1. Unterscheiden sich die Regionen im selben Geschlecht (Region variiert, Geschlecht ist konstant)?

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

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

alles andere ist unwichtig

post-hoc t-tests

Die benötigten Faktoren können mit `phsel()` ausgesucht werden

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

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

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

oder

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

gibt die post-hoc Tests für
Region (mit Gender konstant)

gibt die post-hoc Tests für
Gender (mit Region konstant)

post-hoc t-tests

ersichtlicher wenn auf z.B. 3 Zahlen aufgerundet:

```
p1 = phsel(vok.ph$res)
```

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

```
round(p1, 3)
```

```
round(p2, 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

	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 **Regionen A** ($p < 0.001$) und **B** ($p < 0.001$), jedoch nicht für C.