

# Die Varianzanalyse

Jonathan Harrington

```
library(ggplot2)
```

```
library(ez)
```

```
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"))
```

## 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**.

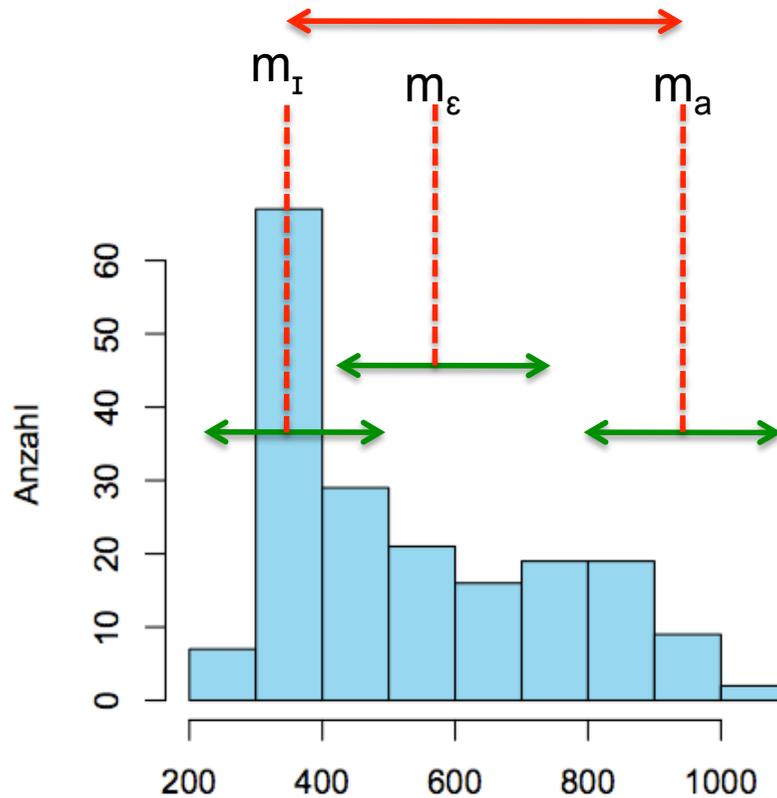
z.B. F1 von drei Vokalkategorien, /I,ε,a/.

**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

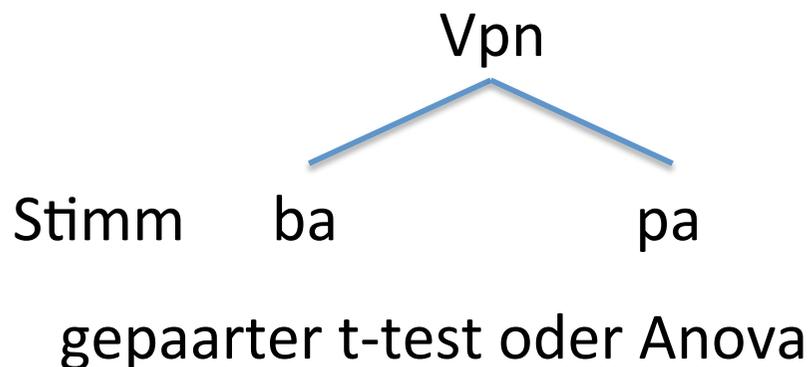
- für Analysen innerhalb derselben Person
- eine Messung pro Vpn. pro Stufe

/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

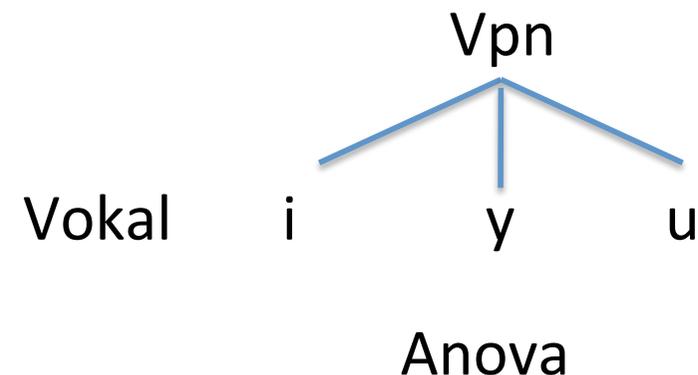


/i, y, u/ wurden von denselben Sprechern produziert.  
Unterscheiden sich /i, y, u/ in F2?

Abhängige Variable: F2

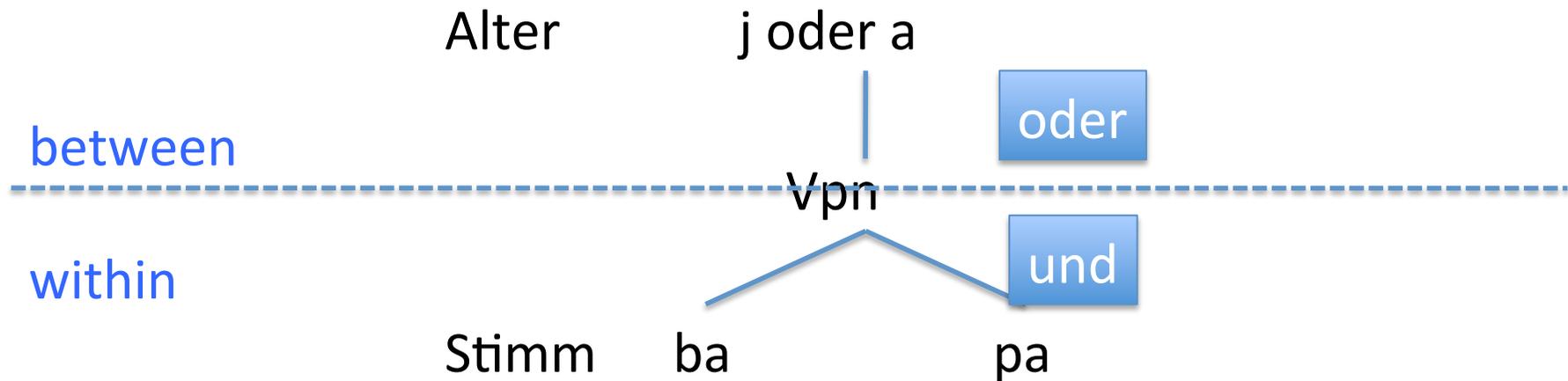
Within-Faktor: Vokal

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



## 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-Factor: gepaarter t-Test und Anova

`head(blang); dim(blang)`

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

Gepaarter t-Test oder within-subjects ANOVA

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

...

## Within-Factor: gepaarter t-Test und Anova

### 1. Differenz-Berechnung

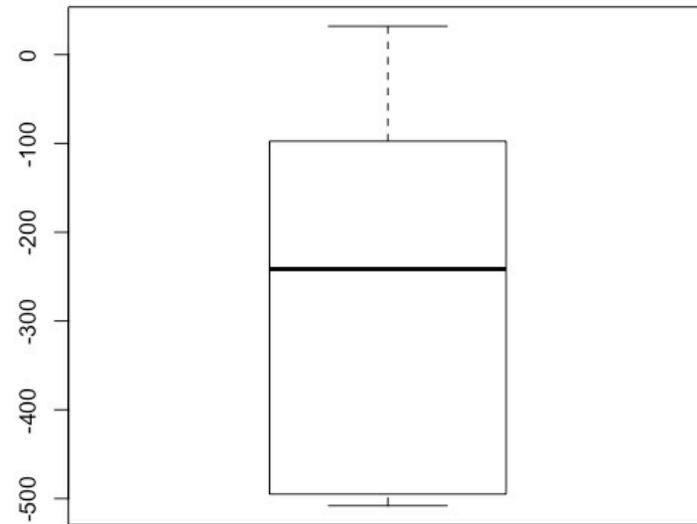
```
d = aggregate(F2 ~ Vpn, diff, data = blang)
```

### 2. Boxplot

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

### 3. gepaarter t-Test

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

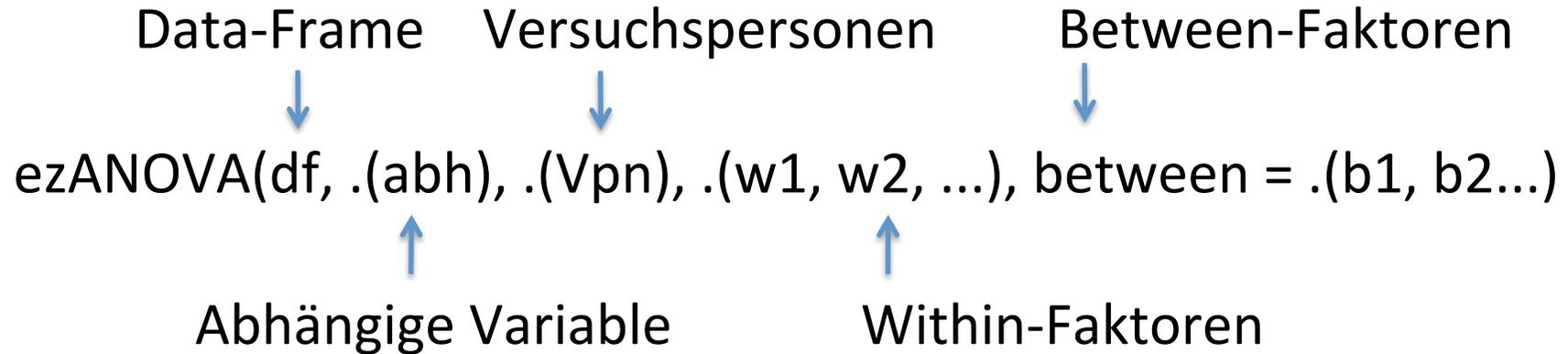


$t = 4.3543$ ,  $df = 11$ ,  $p\text{-value} = 0.001147$

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-value = 0.001147

## Between-factor: t-test und Anova

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

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

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

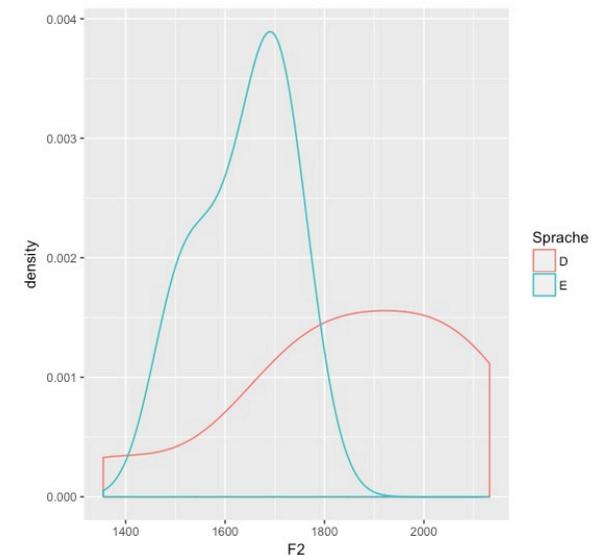
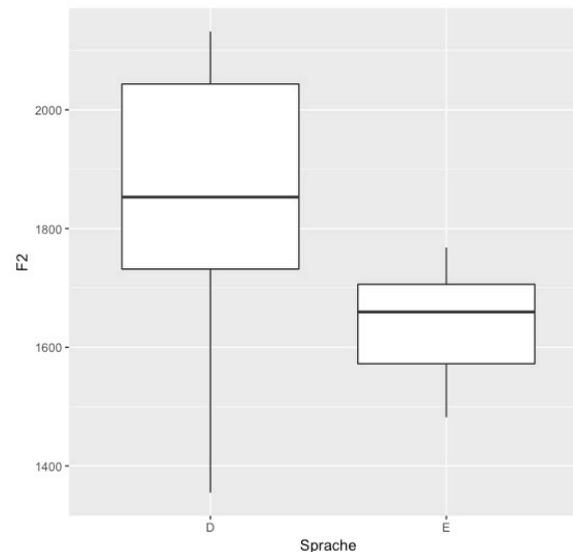
`with(v.df, table(Vpn, Sprache))`

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

## Between-factor: t-test und Anova

Boxplot, densityplot

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



t-Test

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

$t = 2.688$ ,  $df = 11.806$ ,  $p\text{-value} = 0.01999$

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

Ein Anova mit between-Faktor 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
```

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

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

```
t = 2.688, df = 18, p-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
```