Einfaktorielle Varianzanalyse mit festen Effekten

1. WARUM?

Varianzanalysen werden durchgeführt, um Hypothesen zu testen, d.h. ob sich ein oder mehrere Faktoren signifikant auf einen Messwert auswirken.
Bisher: ___________
Problem I: Faktor hat mehr als 2 Stufen
Bisher: ___________
Folgeproblem: Inflation des α Fehlers
= je mehr Paarvergleiche umso höher wird die Wahrscheinlichkeit einen α Fehler zu begehen und zwar exponential mit der Anzahl der Vergleich m

$$p(\text{Fehler}) = 1 - (1- \alpha)^m$$

z.B. für Faktor Konsonant aus der Klausur:
Anzahl der Paarvergleich: 15 R Befehl choose(6, 2)
$p= 1-(1-0.05)^{15}= 0.54$
d.h. die Wahrscheinlich bei 15 Paarvergleichen einen α Fehler zu begehen ist 54%

Lösung 1: **Bonferroni**-Korrektur = das α-Niveau für jeden Einzeltest wird soweit herabgesetzt, dass das Gesamtniveau nur noch 0.05 beträgt (α/Anzahl der Tests).
Lösung 2: Mehrstufige Faktoren können getestet werden ohne Korrektur mittels der Varianzanalyse
Problem II: mehrere Faktoren könnten sich auf eine abhängige Variable auswirken (z.B. Geschlecht und Akzent auf Grundfrequenz).

2. VORAUSSETZUNGEN

1) Mindestens Intervallskalenniveau und Normalverteilung innerhalb der Stichprobe bei der abhängigen Variablen
2) Mindestens 20 Elemente pro Stichprobe (Gruppe, Zelle)
3) Ähnlich stark besetzte Gruppen (gleiches N)
4) Varianzhomogenität der abhängigen Variablen zwischen den einzelnen Stichproben (s. Bartlett-Test)

3. BERECHNUNG

Beispiel Kieferhöhe während der Konsonanten /s, ʃ, t, d, n, l/ von einem Sprecher (AW) mit k=9 Messungen pro Faktorstufe, p=6 Faktorstufen und insgesamt N = 9*6 = 54 Messungen .

	Spalten	**i Faktorstufen**					
Zeilen	JC AW	S	$	T	D	N	L
j Messwerte	1	1.259	1.318	1.665	1.081	0.283	-1.133
	2	1.339	1.206	1.735	0.804	-0.351	-1.533
	3	1.291	0.909	1.569	0.374	-0.498	-1.846
	4	1.174	1.040	1.342	0.298	-0.066	-1.287
	5	1.178	1.004	1.127	0.274	-0.840	-1.284
	6	1.128	1.052	1.381	0.261	-0.335	-1.730
	7	1.298	1.129	1.469	0.963	0.154	-1.206
	8	1.325	0.827	1.495	0.404	-0.444	-1.900
	9	1.260	1.041	1.530	0.428	-1.500	-1.375

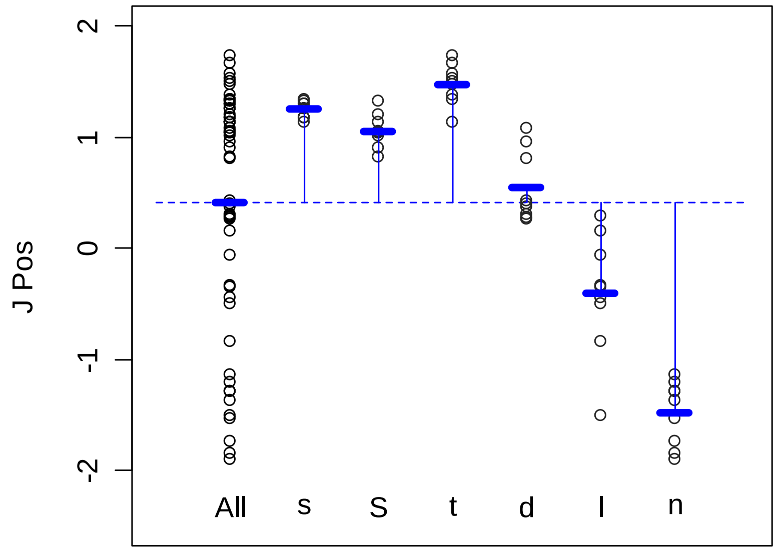
Messwert x_{ij}

= Vergleich der Varianzen der einzelnen Faktorstufen mit der Gesamtvarianz. Ist die Varianz der einzelnen Faktorstufen wesentlich größer als die zufällige Gesamtvarianz in den Daten, dann hat der Faktor einen signifikanten Einfluss.

Die Gesamtvarianz lässt sich demnach zerlegen in folgende Quadratsummen (SS)

$SS_{total} = SS_{treatment} - SS_{error}$
$SS_{treatment}$= Varianz, die sich aus den Faktorstufen ergibt (auch $SS_{between}$)
SS_{error} = Varianz, die sich aus mehreren Messungen ergeben (auch SS_{within})

Wichtig: Zusammenhang Quadratsummen – Varianz: SS=var*df



rep(0, nrow(pos_aw))

$SS_{treatment}$ = Summe der blauen senkrechten Linien SS_{error} = Abweichungen pro Faktorstufen SS_{total} = Abweichung aller Daten (vgl. All) Modellgleichung: $x_{ij} = \mu + \tau_i + \varepsilon_{ij}$

Schritte:

1. Berechnung Faktorstufenvarianzen $SS_{treatment}$

Summe der Abweichungsquadrate innerhalb der Faktorstufen = $SS_{treatment}$
(fettgedrucktes x bedeutet im Folgenden Mittelwert, i bezieht sich auf die Faktorstufen und j auf die Messungen).

	S	$	T	D	N	L	**x..**
$\mathbf{x_{i.}}$	1.25	1.06	1.48	0.54	-0.40	-1.48	0.41
$\mathbf{x_{i.}-x_{..}}$	0.84	0.65	1.07	0.13	-0.81	-1.89	
$(\mathbf{x_{i.}-x_{..}})^2$	0.71	0.42	1.14	0.02	0.65	3.6	Sum 6.5

$$SS_{treatment} = r\sum(\bar{x}_{i.} - \bar{x}_{..})^2 \quad \text{Sum of squares - treatment}$$

$$MS_{treatment} = \frac{SS_{treatment}}{df_{treatment}} \quad \text{Mean Square (variance) - treatment}$$

$SS_{treatment}$ = Sum * Anz.Messwerte = 6.5*9 = 58.5
$MS_{treatment} = SS_{treatment}/df_{treatment}$ = 58.5 / 5

2. Berechnung Fehlervarianz SS_{error}

= Varianz, die durch die Abweichungen vom Faktormittelwert bei z.B. mehrfachen Wiederholungen entstehen („weil die Versuchsperson nicht exakt immer das Gleiche gemacht hat“).

$$SS_{error} = \sum (x_{ij} - \bar{x}_{i.})^2$$

SS error = sum(tapply(pos_awJC, pos_awcons, var))*8 = 4.25

MS error = SS error /df = 4.25/48

3. Berechnung Gesamtvarianz SS_{total}

$$SS_{tot} = \sum (x_{ij} - \bar{x}_{..})^2$$

SS tot = var(pos_aw$JC) * (9*6-1) = 62.79

4. Berechnung F Wert

Zugrundeliegende Modellgleichung

$$\boxed{x_{ij} = \mu + \tau_i + \varepsilon_{ij}}$$

Jeder Messwert x_{ij} setzt sich zusammen aus dem Gesamtmittelwert, dem Einfluss des Faktors τ_i und zufälliger Variation ε_j, die nicht auf den Faktor zurückzuführen ist.

Ob der Faktor nun wichtiger ist als die Fehlervarianz, läßt sich durch den F-Wert schätzen

$$F = MS_{treatment}/MS_{error}$$

mit $df_{treatment}$ = Faktorstufen -1 (p-1) und df_{error} = Gesamt – Faktorstufen (N – p)

Nullhypothese:

H0: $\tau_s = \tau_S = \tau_t = \tau_d = \tau_l = \tau_n = 0$

4. INTERPRETATION

Befehle in R:

```
anova(lm(pos_aw$JC ~ pos_aw$cons))
```

oder

```
summary(aov(pos_aw$JC ~ pos_aw$cons))
```

Ergebnis aus R mit

- □□□□□□□□□□□ entspricht Werten für *treatment*
- □□□□□□□□□ entspricht den Werten für *error*
- SS sind in Spalte □□□ □□ und MS in Spalte □□□□ □□
- Da $MS_{treatment}$ wesentlich größer ist als MS_{error} , wird der F-Wert ziemlich groß (132.18).
- Ein großer F-Wert ist meistens auch signifikant (siehe Tabellen in Statistikbüchern).
- Ergebnis (wie es in wissenschaftlichen Artikeln, Magisterarbeiten und Statistik-klausuren berichtet werden sollte): der Konsonant hat einen signifikanten Einfluss **($F(5,48)=132.18$, $p<0.001$)** auf die Kieferposition während des Konsonanten.

Wiederholung ANOVA

1. Warum heißt dieses Verfahren Varianzanalyse?
2. Erläutere die folgenden Modellgleichungen
   a) $x_{ij}= \mu+\tau_i$
   b) $x_{ij}= \mu+\tau_i+\varepsilon_{ij}$
3. Was bedeutet: $SS_{total} = SS_{treatment} + SS_{error}$
4. Was bedeutet folgende Tabelle

   [illegible]

5. Wie sollten die Ergebnisse dokumentiert werden?

Exkurs: Hypothesen und Prüfstatistik

Nullhypothese: es existiert kein Unterschied zwischen zwei Mittelwerten (z.B. Stichprobenmittelwert und Populationsmittelwert, oder Mittelwert und einem angenommenen Mittelwert, oder zwischen 2 Stichprobenmittelwerten) $H_0: \mu_1=\mu_2$

Alternativhypothese: Mittelwerte unterscheiden sich.
ungerichtete Alternativhypothese: es gibt einen Unterschied $H_1: \mu_1\neq\mu_2$
gerichtete Alternativhypothese gibt eine Richtung an (< oder >) $H_1: \mu_1<\mu_2$

α-Niveau

- Die Nullhypothese wird verworfen, wenn der empirisch ermittelte Kennwert außerhalb des Konfidenzintervalls liegt.
- Abhängig von $t_{\alpha,df}$
- Je kleiner α ist, desto größer muss der Mittelwertunterschied sein, um *signifikant* zu sein.
- Signifikant ≈ statisch relevant
- α-Niveau legt die Wahrscheinlichkeit fest, mit der die Nullhypothese abgelehnt wurde.
- Irrtumswahrscheinlichkeit bzw. Restrisiko für eine Fehlentscheidung gegen eine gültige Nullhypothese
- Umgangssprachlich ausgedrückt: wir haben blöderweise eine Stichprobe gezogen, die an den seitlichen Rändern der theoretischen Verteilungskurve aller Stichprobenmittelwerte liegt.
- α-Fehler, Fehler erster Art, *Type I error*

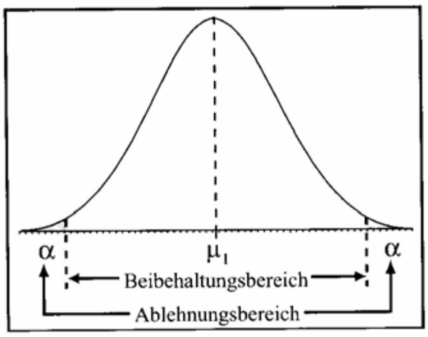


Abbildung 7.5: Das α-Niveau bei zweiseitiger Testung

Prüfung der Signifikanz hängt vom α-Niveau ab:

α=0.1	marginal signifikant	.
α=0.05	signifikant	*
α=0.01	hoch signifikant	**
α=0.001	höchst signifikant	***

Sind die Konsequenzen einer fälschlichen Ablehnung der Nullhypothese sehr gravierend, so setzt man das α-Niveau auf einen kleineren Wert (1% oder 1 Promille).

Testen von Hypothesen: zwei Mittelwerte, x_1 und x_2, sollen miteinander verglichen werden. Wir wollen feststellen, ob sie aus der gleichen Population stammen (= ____-Hypothese) oder aus verschiedenen (=____ - Hypothese). Bei einem α-Niveau von 5 % ist die Wahrscheinlichkeit, dass x_1 und x_2, wenn sie außerhalb des Beibehaltungsbereichs liegen, trotzdem aus der gleichen Population stammen, gleich 5%.

Bei einem **beidseitigen Test** entsprechen die beiden Ränder jeweils α/2. Der Beibehaltungsbereich ist 1-α.

FRAGE: Wie groß sind die α-Bereiche bei einem beidseitigen Test mit einem Beibehaltungsbereich von

- 95%
- 99%
- 99.9%

Bei einem **einseitigen Test** wissen wir aus der Literatur, dass einer der beiden Mittelwerte größer (kleiner) sein sollte als der andere, d.h. wir nehmen eine Richtung an.
Vorteil: der t-Test wird schon bei einem geringeren Mittelwertsunterschied signifikant.

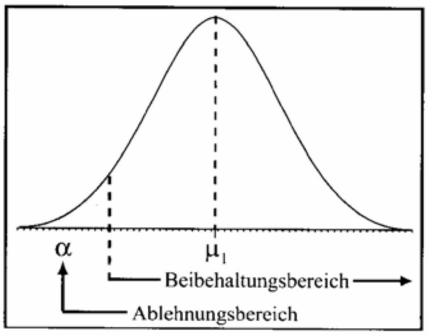


Abbildung 7.4: Das α-Niveau bei einseitiger Testung

β-Fehler

= Beibehaltung der falschen Nullhypothese bei gültiger Alternativhypothese
= Fehler zweiter Art, Type II error

Fehler 1. Art: Ablehnung einer gültigen Nullhypothese Fehler 2. Art: Beibehaltung der falschen Nullhypothese

		Realität	
		H0 ist wahr	H0 ist falsch
Entscheidung	akzeptiere H0	korrekt (es brennt nicht, kein Alarm)	Fehler 2. Art (es brennt, aber kein Alarm)
	lehne H0 ab	Fehler 1. Art (es brennt nicht, aber Alarm)	korrekt (es brennt und Alarm)

1-β: **Teststärke** (*test power*) ist die Wahrscheinlichkeit, dass ein in der Population vorhandener Unterschied bei statistischer Testung aufgedeckt wird.

β-Fehler ist abhängig von

- α-Niveau: je höher das vorher festgelegte α-Niveau, desto kleiner wird die Wahrscheinlichkeit für einen β-Fehler
- Einseitige vs. zweiseitige Testung: höhere Wahrscheinlichkeit für einen Fehler 2. Art bei zweiseitiger Testung
- Streuung des Merkmals: je einheitlicher sich die Stichprobenteilnehmer bezüglich eines Merkmals verhalten, umso geringer die Streuung. Je kleiner die Streuung umso kleiner ist auch der

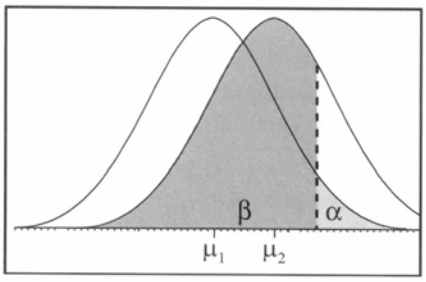


Abbildung 7.8: α- und β-Fehler, Teil 3

Standardfehler. Je kleiner der Standardfehler umso eher erhält man ein signifikantes Ergebnis.

- Stichprobenumfang: je größer die Stichprobe, umso kleiner der Standardfehler
- Mittelwertsunterschied: je größer der Unterschied zwischen zwei Stichproben (oder Faktorstufen) umso eher ein signifikantes Ergebnis
- β ist kleiner für abhängige als für unabhängige Stichproben
- Skalenniveau: je höher das Skalenniveau, desto kleiner β

R Befehle

Formelschreibweise in R: *AbhängigeVar ~ Faktor*

(entspricht sprachlich: Variable wird durch Faktor beschrieben bzw. hängt von dem Faktor ab)

t.test mit paired=T (t Test für abhängige Stichproben)

---

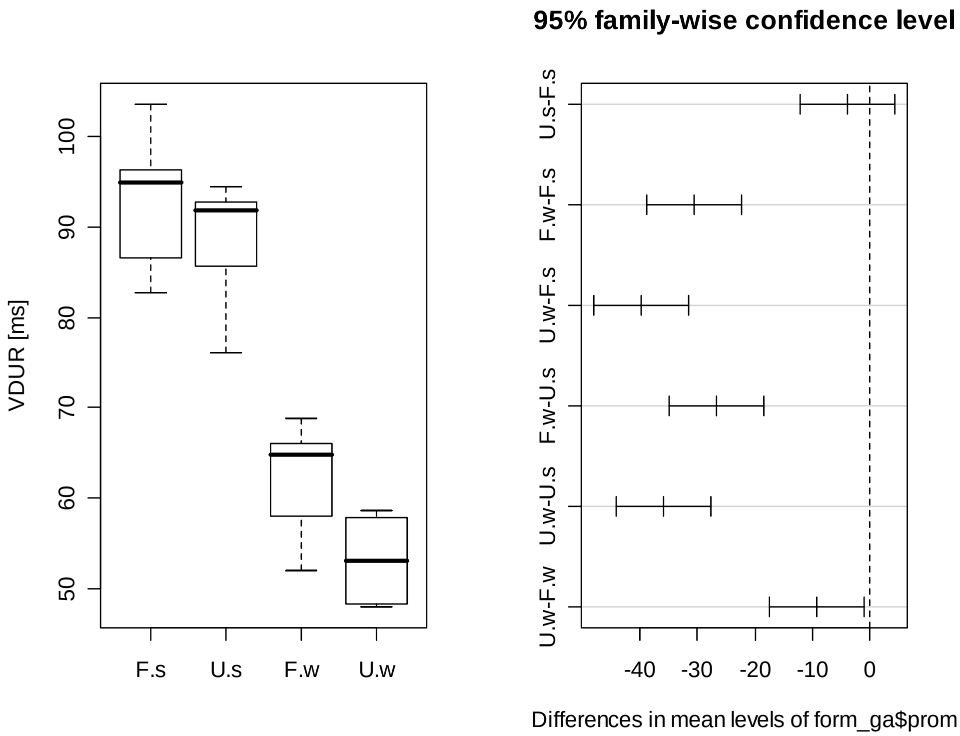
4. POST HOC TESTS

- Ziel: Welche Mittelwerte unterscheiden sich signifikant bei mehrstufigen Faktoren?
- **Nachdem** mittels einer ANOVA ein signifikanter Effekt festgestellt wurde, können so genannte *Post hoc Tests* durchgeführt werden.
- Es wird keine Unabhängigkeit der Stichproben gefordert.
- Automatische Anpassung des α-Niveaus

Tests:

- Sehr gebräuchlich: *Scheffé* Test (sehr konservativ)
- Pairwise.t.test mit Bonferroni Anpassung
- Auch in R implementiert *Tukey HSD* („honestly significant difference“)

Beispiel:



```
TukeyHSD(aov(form_ga$vdur ~ form_ga$prom))
```

[illegible]

```
par(mfcol=c(1,2))
boxplot(form_ga$vdur ~  form_ga$prom, ylab="VDUR [ms]")
plot(TukeyHSD(aov(form_ga$vdur ~ form_ga$prom)))
```

Mehrfaktorielle Varianzanalyse mit festen Effekten

- Ziel: Testen, ob mehrere Faktoren einen signifikanten Einfluss auf eine Variable haben, z.B. Geschlecht und Akzent auf f0
- Zweifaktorielles Design:

Stress / Accent	strong	weak
Focus	F.s	F.w
Unfocussed	U.s	U.w

Modellgleichung:

$$x_{ijk} = \mu + \alpha_i + \beta_j + \alpha_i\beta_j + \varepsilon_k$$

Haupteffekte:

Stress: α_1=strong
α_2=weak

Accent: β_1=focus
β_2=unfocussed

Stress / Accent	strong	weak	xbar
Focus	F.s	F.w	
Messung 1	103.62	66.250	77
Messung 2	96.72	52.001	
Messung 3	96.00	68.850	
Messung 4	82.78	65.679	
.k.			
Unfocussed	U.s	U.w	
Messung 1	92.100	48.03	71
Messung 2	94.406	49.31	
Messung 3	86.880	57.03	
Messung 4	91.510	58.57	
.k.			
xbar	91	58	μ 74

anova(lm(vdur ~ accent+stress, data=form_ga))

Interpretation:

- höchst signifikanter Effekt von Wortakzent (F(1,29)= 241.24, p<0.001) auf die Vokaldauer,
- hoch signifikanter Effekt von Satzakzent (F(1,29)= 9.45, p<0.01)) auf die Vokaldauer
- **beide Haupteffekte wirken sich signifikant auf die Vokaldauer aus.**

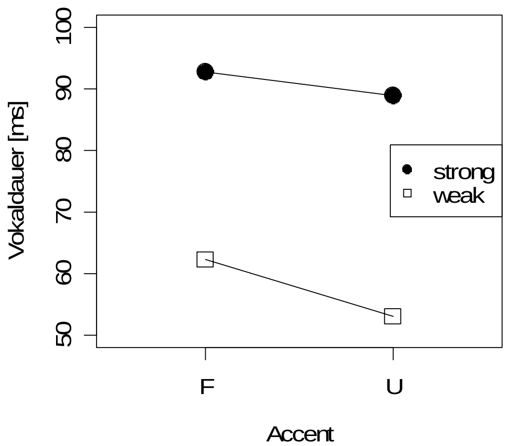
Interaktionen:

$$x_{ijk} = \mu + \alpha_i + \beta_j + \alpha_i\beta_j + \varepsilon_k$$

Interaktionen treten auf, wenn die Unterschiede zwischen den Faktorstufen eines Faktors nicht für alle Faktorstufen des zweiten Faktors gleich sind.

anova(lm(vdur ~ accent+stress+accent:stress, data=form_ga))
anova(lm(vdur ~ accent*stress, data=form_ga)) (Kurzform)

Zum Ausprobieren



Berechnung der Freiheitsgrade:

- Haupteffekte: p – 1=1, q-1=1
- Interaktion: (p-1)*(q-1)=1
- Fehler: pq(n-1) =2*2*7=28

- Deakzentuierung hat immer eine reduzierende Wirkung auf die Vokaldauer, unabhängig vom Wortkazent.
- Wortakzentuierte Vokale (strong) sind immer länger als wortunbetonte Vokale, unabhängig vom Satzakzent.

➔ Linien verlaufen ungefähr parallel
➔ Keine Interaktionen

Signifikante Interaktionen

```
anova(lm(cdur ~ accent+stress+accent:stress, data=form_ga))
```

[illegible]

- Deakzentuierung hat eine reduzierende Wirkung auf die Konsonantdauer für Konsonanten in starken Silben und den umgekehrten Effekt in schwachen Silben, unabhängig vom Wortkazent.

➔ Signifikante Interaktion

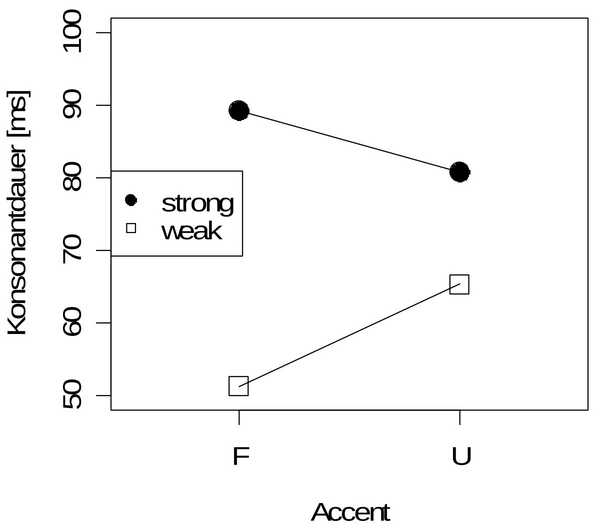
- Wortakzentuierte Konsonanten (strong) sind immer länger als wortunbetonte Konsonanten, unabhängig vom Satzakzent.

➔ Signifikanter Haupteffekt stress

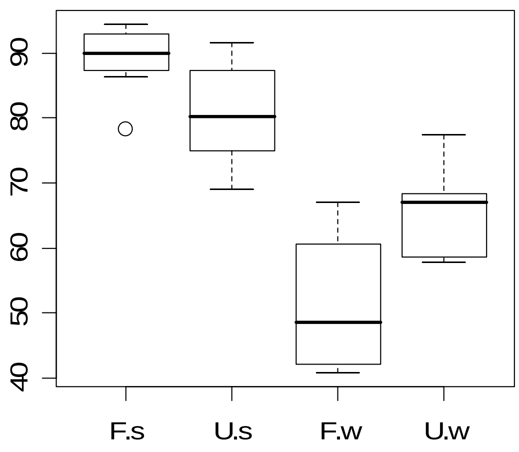
- Durch die signifiante Interaktion streut die Konsonantdauer so stark, dass der Akzent keinen signifikanten Haupteffekt hat

Kennzeichen der Interaktion

➔ **Linien verlaufen nicht parallel**



Aufgabe:
Untersuche mittels zweifaktorieller Varianzanalyse, ob die Intensität (rms) von Wortakzent (stress) und Satzakzent (accent) beeinflusst werden.
Nimm hierzu auch Abbildungen und Post hoc Tests zuhilfe.

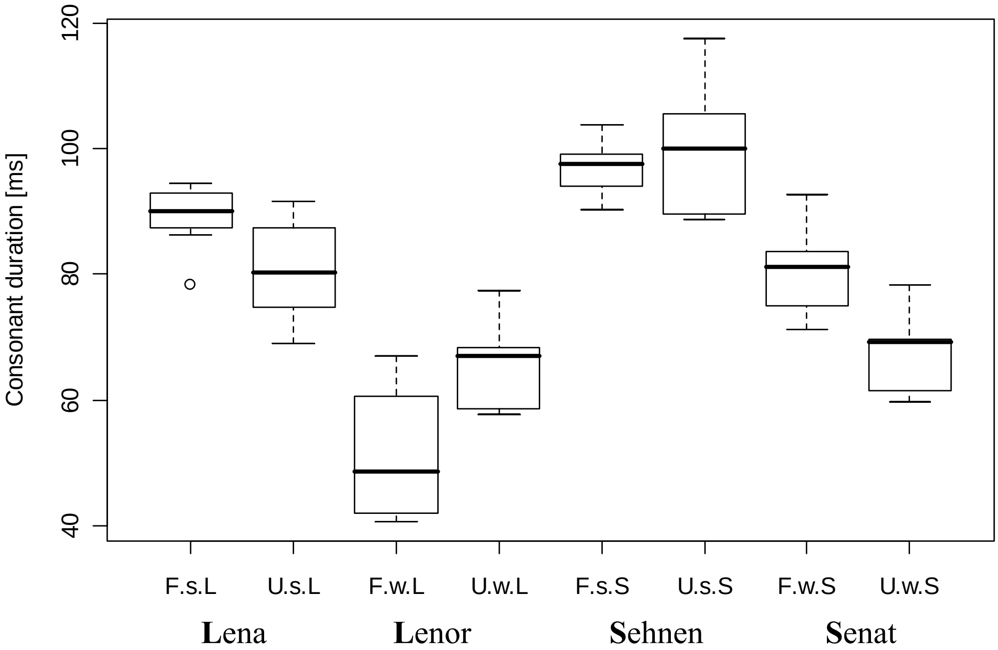


Zusammenfassung Vorgehensweise:

1. Prüfen, ob die Voraussetzungen für eine Varianzanalyse gegeben sind
2. Berechnen einer vollständigen Varianzanalyse

```
anova(lm(cdur ~ accent*stress*cons, data=formls))
```

[illegible]



Weitere Vorgehensweise je nach Ergebnis:

I. Mehrstufiger Haupteffekt ist signifikant ➔ Post hoc Test

II. Interaktion ist signifikant ➔ 2 Möglichkeiten

a) „Aufdröseln“, d.h. ANOVAs getrennt für die Faktorstufen eines Faktors, z.B. **cons**

```
anova(lm(cdur ~ accent*stress, subset=cons=="L"))
```

[illegible]

anova(lm(cdur ~ accent*stress, subset=cons=="S"))

```
□□□□□□□□□ □□□□
                □□ □□□ □□ □□□□ □□ □ □□□□□     □□□□□□
□□□□□□           □  □□□□□   □□□□□  □□□□□□  □□□□□□□□ □
□□□□□□           □ □□□□□□  □□□□□□ □□□□□□□ □□□□□□□□□ □□□
□□□□□□□□□□□□□    □  □□□□□   □□□□□  □□□□□□  □□□□□□□□ □□
□□□□□□□□□       □□ □□□□□□     □□□□
```

b) Erstellen eines mehrstufigen Faktors, z.B. promcons aus stress, accent und cons
promcons=paste(accent, stress, cons, sep=".")
pairwise.t.test(cdur, promcons, p.adj="bonf")
➔ Nachteil: sehr viele t-Tests, d.h. kaum noch Signifikanzen

Aufgabe: Mehrfaktorielle ANOVA
Untersuche anhand der Datenbasis *formls.Rdata* für einen Sprecher wie die Vokaldauer (*vdur*) durch die Faktoren *accent*, *stress* und *cons* beeinflusst werden. Welche Folgeanalysen zur Ermittlung der einzelnen Faktoreinflüsse müssen durchgeführt werden, wenn die experimentelle Fragestellung sich nur auf die Prosodie bezieht? Wie können die Ergebnisse interpretiert werden? Nimm hierzu auch Abbildungen und Post hoc Tests zu Hilfe.

Mehrfaktorielle Varianzanalyse mit Messwiederholungen

- ANOVA bisher: nur für eine Versuchsperson, unterschiedliche Bedingungen (Faktorstufen), mehrere Wiederholungen pro Bedingung
- Ziel: Generalisierung für die gesamte Population anhand einer Stichprobe
- Annahme bei **ANOVA mit festen Faktoren**: die Stichproben sind unabhängig voneinander. Das heißt, für jede Messung (auch Faktorstufe) wurde ein neuer Sprecher verwendet
- Messwiederholungen bedeutet, dass ein Sprecher unter verschiedenen Bedingungen (z.B. beim Sprechen in verschiedenen Lautstärken) gemessen wurde = ***within subject factor***.
  Analog zu ______________________
- Unterschiedliches Verhalten der einzelnen Versuchspersonen wird dabei als Zufallseffekt angesehen.
- Vergleich von Sprechergruppen (z.B. Stotterer vs. Nicht-Stotterer, bzw. Deutsche vs. Koreaner) = ***between subjects factor***
- Strukturgleichung: $y_{ij} = \mu + \alpha_j + \pi_i + \varepsilon_{ij}$ mit
  α_j fester Effekt der Faktorstufe j,
  π_i Zufallseffekt der Versuchsperson i
  ε_{ij} Fehler

Datenbasis bei einfaktoriellem Design

	Stress	Strong	Weak
Sprecher		F0	F0
BD		s	w
	Messung 1	126.507	106.072
	Messung 2	132.109	106.348
	Messung 3	147.076	112.089
	Messung 4	125.079	107.023
		…	…
	xbar	**137**	**108**
BP		s	w
	Messung 1	125.687	111.952
	Messung 2	113.411	103.334
	Messung 3	128.160	105.855
	Messung 4	119.850	107.437
		…	…
	xbar	**129**	**102**
...			
Weitere Sprecher			
…			

- Errechnen der **Zellenmittelwerte**, da sonst jede Messung wie eine neue Versuchsperson behandelt wird
- Zellenmittelwert = Mittelwerte für jeden Sprecher und jede Bedingung über die Wieder-holungen

Zellenmittelwerte in R:

```
lab=paste(formNF$vp, substring(formNF$lab,5,5),"")
f0=as.vector(tapply(formNF$f0, lab, mean))
labnew=names(tapply(formNF$f0, lab, mean))
```

Erzeugen einer Matrize mit den Faktoren Sprecher und Stress sowie der Variablen f0:

____

____

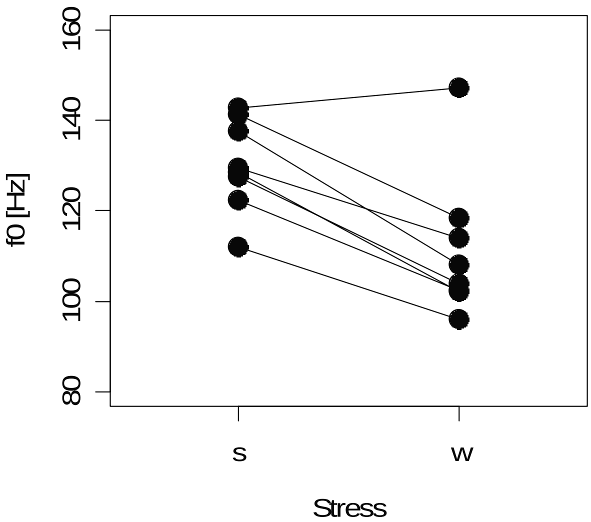
____

Erstellen des Modells in R

Falsch:

```
summary(aov(f0~stress), data=mat))
            □□  □□□ □□ □□□□ □□ □ □□□□□  □□□□□□
□□□□□□       □ □□□□□□□ □□□□□□□  □□□□□□ □□□□□□□ □
□□□□□□□□□   □□ □□□□□□□  □□□□□□
□□□
□□□□□□□ □□□□□□  □ □□□□□ □□□□□ □□□□ □□□□ □□□ □□□□ □□□ □□□ □ □ □
```

Warum falsch????



Richtig:

```
summary(aov(f0~stress + Error(sp/stress), data=mat))
□□□□□□ □□
          □□  □□□ □□ □□□□ □□ □ □□□□□ □□□□□□
□□□□□□□□□  □ □□□□□□□  □□□□□□

□□□□□□ □□□□□□□□□
          □□  □□□ □□ □□□□ □□ □ □□□□□   □□□□□□
□□□□□□     □ □□□□□□□ □□□□□□□  □□□□□□ □□□□□□□□ □□
□□□□□□□□□  □  □□□□□□   □□□□□
```

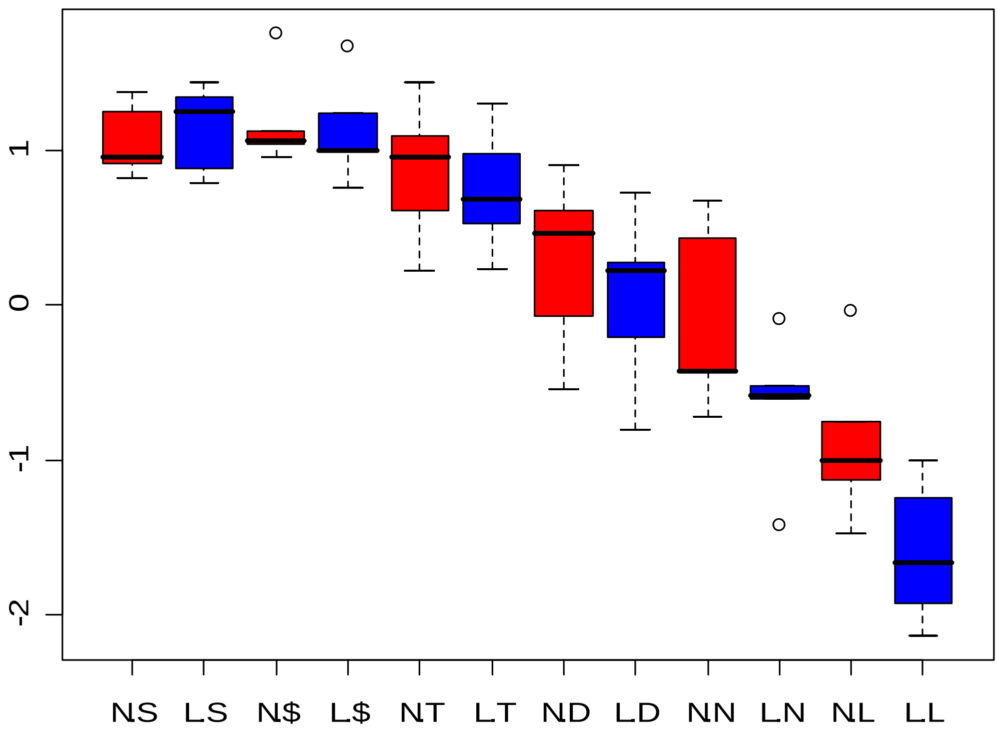
□□□□□□□□□□□□□□□□□ bedeutet, dass Änderungen in der Grundfrequenz aufgrund des Faktors *stress* immer innerhalb des Subjekts betrachtet werden sollte.

Aufgabe ANOVA mit Messwiederholungen

Untersuche anhand der Datenbasis *formants.Rdata* für die Teilmenge der akzentuierten (*accent*=F) und wortbetonten Daten (*stress* =s), ob die Lautstärke einen signifikanten Einfluss auf die Variable *f0* hat. Erzeuge hierfür eine Matrix mit den Zellenmittelwerten. Wie können die Ergebnisse interpretiert werden? Nimm hierzu auch Abbildungen und Post hoc Tests zu Hilfe.

Mehrfaktorielles Design

```
summary(aov( JC ~ cons * loudness + Error(subj/(cons*loudness)), data=pos))
```



Zwischensubjekt-Faktoren

Daten *stutt*: Zwischensubjektfaktor NS = non-stutterer, ST = persons who stutter

R Befehl für ANOVA mit Messwiederholung für den Innersubjektfaktor *loud* und den Zwischensubjektfaktor *group*

```
summary(aov(f0~loud * group + Error(sp/loud), data=stutt))
```

Wichtig: Zwischensubjektfaktor steht nicht im Error Term aber der Innersubjektfaktor