UNIVERSITÄT LEIPZIG

Eine Einführung in R: Varianzanalyse

Bernd Klaus, Verena Zuber

Institut für Medizinische Informatik, Statistik und Epidemiologie (IMISE),

Universität Leipzig

13. Dezember 2012

1. Varianzanalyse: Theorie

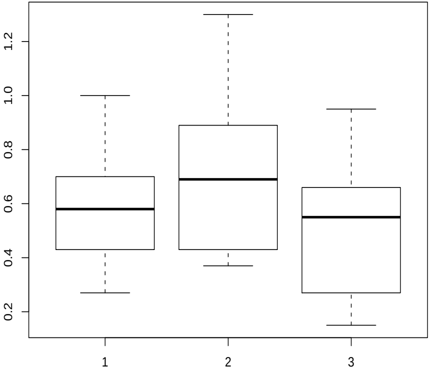
2. Varianzanalyse: Praxis

3. Gruppenvergleiche und multiples Testen

Varianzanalyse: Theorie

Beispiel: "*toycar*"

Fragestellung: Fahren die drei Autotypen unterschiedlich weit?



Oder wie untersucht man die Nullhypothese: $H_0 : \mu_1 = \mu_2 = \mu_3$?

Varianzanalyse

Daten: Gegeben ist eine metrische (normalverteilte) **Zielgröße** Y und mindestens $(p \leq 1)$ **Faktorstufen**, die jeweils mehrere Gruppen $(k \leq 2)$ umfassen.
Insgesamt sind $n_1 + ... + n_k = n$ Beobachtungen gegeben

- $p = 1$: Einfaktorielle Varianzanalyse
- $p = 1$ und $k = 2$: t-Test
- $p > 1$: Mehrfaktorielle Varianzanalyse

Frage: Unterscheiden sich die Erwartungswert der metrischen Zufallsvariable in den Gruppen?

Oder: Ist die Varianz zwischen den Gruppen größer als in den Gruppen?

Das Modell der einfaktoriellen Varianzanalyse

$p = 1$

- Spezialfall $k = 2$: t-Test
- Das Modell für $j = 1, .., k$ Gruppen und $i = 1, ..., n_j$ Beobachtungen in Gruppe j:

$$Y_{ji} = \mu_j + \epsilon_{ji}$$

- Voraussetzungen:
  1. $\epsilon_{ji} \sim N(0, \sigma)$
  2. ϵ_{ji} **ist normalverteilt mit Erwartungswert 0**
  3. **identischer Varianz** σ^2
- H_0: $\mu_1 = ... = \mu_k$

Streuungszerlegung

ANOVA: *ANalysis Of VAriances*

$$SQT = \sum_{i=1}^{n}(y_i - \bar{y})^2 = \underbrace{\sum_{j=1}^{k}\sum_{i=1}^{n_j}(y_{ij} - \bar{y}_j)^2}_{SQR} + \underbrace{\sum_{j=1}^{k} n_j(\bar{y}_j - \bar{y})^2}_{SQE}$$

Für die Streuungszerlegung werden folgende Größen berechnet:

- SQT: Sum of Squares Total, die Gesamtstreuung ($Var(Y)$)
- SQR: Sum of Squares Residuals, Streuung in den Gruppen
- SQE: Sum of Squares Explained, Streuung zwischen den Gruppen

Der F-Test

- H_0: $\mu_1 = ... = \mu_k$
- Aus der Streuungszerlegung wird verwendet:

Streuung	df	Mittlerer Quadr. Fehler
zwischen den Gruppen	k-1	SQE/(k-1)
in den Gruppen	n-k	SQR/(n-k)

- Die Prüfgröße F berechnet sich aus:

$$F = MQE/MQR = \frac{SQE}{k-1} / \frac{SQR}{n-k}$$

- F ist F-verteilt mit $(k-1, n-k)$ Freiheitsgraden

Mehrfaktorielle Varianzanalyse $p > 1$

- Natürlich können mehrere Faktoren und Wechselwirkungen zwischen Faktoren berücksichtigt werden
- Die Formeldarstellung kann dabei sehr schnell sehr kompliziert werden
- Wichtig in der Praxis ist dabei, dass jede der einzelnen Unterkategorien eine ausreichende Stichprobengröße besitzt
- Es gibt F-Tests für alle Faktoren und deren Wechselwirkungen

Varianzanalyse: Praxis

Beispiel: *toycar*-Daten

- Berechnung des linearen Modells `lm.car`:
  `lm.car <- lm(distance ~ car)`
- Output des `summary`-Befehl :
  `Call: lm(formula = distance ~ car)`
  `Coefficients:`

	Estimate	Std.Error	t value	Pr(> | *t* |)	
Intercept	0.5911	0.0902	6.555	8.86 e-07	⋆⋆⋆
car2	0.1111	0.1275	0.871	0.392	
car3	-0.0822	0.1275	-0.645	0.525	

```
Multiple R-squared: 0.08797,
Adjusted R-squared: 0.01197
F-statistic: 1.158 on 2 and 24 DF, p-value: 0.3312
```

Beispiel: *toycar*-Daten

- R-Befehl zur Varianzanalyse: `anova(lm.car)`
- Output des anova-Befehl :

```
Analysis of Variance Table
Response: distance
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
car	2	0.16945	0.084726	1.1575	0.3312
Residuals	24	1.75673	0.073197		

Beispieldaten: "Taste"

Untersuchung von zwei verschiedenen Einflussfaktoren auf den Geschmack eines Nahrungsmittels:

- `SCORE`: *Geschmackspunktzahl*
- `LIQ`: *Flüssigkeitskomponente: hohe (1) oder niedrige (0) Konzentration*
- `SCR`: *Textur des Nahrungsmittels: rauh (0) oder fein (1))*

Beispiel für 2-faktorielle Varianzanalyse: *Taste*-Daten

- Berechnung des linearen Modells `taste`:
  `taste <- lm(SCORE ~ LIQ * SCR)`
- R-Befehl zur Varianzanalyse: `anova(taste)`
- Output:

```
Analysis of Variance Table
Response: SCORE
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
LIQ	1	1024.0	1024.0	2.6321	0.1306	
SCR	1	10609.0	10609.0	27.2696	0.0002	***
LIQ:SCR	1	420.2	420.2	1.0802	0.3191	
Residuals	12	4668.5	389.0			

$\Rightarrow$ Nur der Effekt von SCR ist signifikant von 0 verschieden

Beispiel - Schätzung der Effektgrößen

- Schätzer der Effektgrößen des Modells `taste`:
  `summary(taste)`
- Output wie im linearen Modell:
  `Coefficients:`

	Estimate	Std. Error	t value	Pr(>|t|)	
Intercept	41.75	9.862	4.233	0.0011	**
LIQ1	-5.75	13.947	-0.412	0.6874	
SCR1	61.75	13.947	4.427	0.0008	***
LIQ1:SCR1	-20.50	19.724	-1.039	0.3191	

Gruppenvergleiche und multiples Testen

Gruppenvergleiche

Die Nullhypothese der Varianzanalyse ist:

- H_0: $\mu_1 = ... = \mu_k$
- $\Rightarrow$ Ablehnen der Nullhypothese gibt noch keine Auskunft darüber welche Gruppen unterschiedlich sind!
- Es sind weitere Gruppenvergleiche ("post-hoc tests") nötig

Beispiel *Taste*

Die Varianzanalyse beim *taste* Datensatz hat **2 Faktoren** mit **2 Stufen**, also ergeben sich insgesamt **4 nicht disjunkte Gruppen**.

- $\Rightarrow$ Wo liegen Unterschiede zwischen den Gruppen?
- Rechnen von vier t-Tests nötig!
- Problem: Die Wahrscheinlichkeit die Nullhypothese bei $\alpha = 0.05$ für mindestens einen Test fälschlicherweise abzulehnen ist bei vier Tests: 0.185
- $\Rightarrow$ Korrektur für mehrfaches Testen angebracht

Vergleich der Gruppen in R

... mittels der Funktion:

```
pairwise.t.test(x, g, p.adjust.method =
p.adjust.methods, pool.sd = !paired, paired = FALSE,
alternative = c("two.sided", "less", "greater"), ...)
```

Parameter ähnlich dem normalen t-Test. Zusätzlich gibt es

- `g`: Faktor, der die Gruppe der Stichprobe angibt
- `p.adjust.method`: Zeichenkette, die eine Adjustierungsmethode angibt: **"holm"**, **"hochberg"**, **"hommel"**, **"bonferroni"**, **"BH"**, **"BY"**, **"fdr"**, **"none"**

Vergleich der Gruppen bei *Taste*

- Der Faktor `groups` beschreibt die Gruppezugehörigkeit
- Wir benutzen die FDR Korrektur:
  `pairwise.t.test(SCORE,groups, p.adjust =“fdr”)`

	LIQ = 1	LIQ=SCR=0	LIQ=SCR=1
LIQ=SCR=0	0.6970		
LIQ=SCR=1	0.0234	0.0394	
SCR = 1	0.0031	0.0041	0.3738

- Wie schon die ANOVA Tabelle nahelegt, hat die Textur offenbar einen sehr großen Einfluss auf den Geschmackswert, der Flüssigkeitsgehalt dagegen kaum!