

Eine Einführung in R: Varianzanalyse

Bernd Klaus, Verena Zuber

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

13. Januar 2009

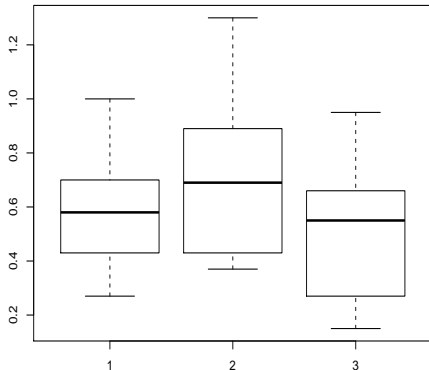
I. Varianzanalyse: Theorie

II. Varianzanalyse: Praxis

I. Varianzanalyse: Theorie

Beispiel: "toyicar"

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 $n_1 + \dots + n_k = n$ 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 $j = 1, \dots, k$ Gruppen und $i = 1, \dots, n_j$ 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
- ▶ $H_0: \mu_1 = \dots = \mu_k$

Streuungszerlegung

ANOVA: *AN*alysis Of *VA*riances

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

Für die Streuungszerlegung werden folgende Größen berechnet:

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

Der F -Test

- ▶ $H_0: \mu_1 = \dots = \mu_k$
- ▶ Aus der Streuungszersetzung 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 leicht 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

II. Varianzanalyse: Praxis

Beispiel: *toycar*-Daten

- ▶ Berechnung des linearen Modells `lm.car`:
`lm.car <- lm(distance ~ car)`
- ▶ R-Befehl zur Varianzanalyse: `anova(lm.car)`
- ▶ Output:
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			

⇒ Nur der Effekt von **SCR** ist signifikant von 0 verschieden

Beispiel - Schätzung der Effektgrößen / Koeffizienten

- ▶ 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