Varianzanalyse

Idee

$\boxed{\text{Gesamtvariabilität der Beobachtungen}} = \boxed{\text{Variationsursache 1}} + \boxed{\text{Variationsursache 2}} + \boxed{\text{Variationsursache 3}} + \dots$

Begriffe

`Faktor:` diskrete, erklärende Variable
`Levels:` Werte, die der Faktor annimmt
`Ein-Weg-Varianzanalyse:` Einfluss eines Faktors wird untersucht
`Zwei-Weg-Varianzanalyse:` Einfluss von zwei Faktoren wird untersucht
`Treatment:` Faktorkombination
`Plot, experimentelle Einheit:` kleinste Einheit, die einem Treatment zugeteilt werden kann.

Ein-Weg-Varianzanalyse

Modell:

$$y_{ij} = \mu + A_i + \epsilon_{ij}, \quad \text{wobei } \sum A_i = 0$$

y_{ij} ist die Messung für Plot j mit Treatment i, $j = 1, \dots, J; i = 1, \dots, I$,
μ ist das Gesamtmittel,
A_i ist der Effekt des i-ten Levels von Faktor A, bzw. die Abweichung der Gruppe i vom Gesamtmittel und
ϵ_{ij} ist ein zufälliger „Fehler“ oder Rest, über den folgendes vorausgesetzt wird:

1) $E(\epsilon_{ij}) = 0$ für alle i und j
2) die ϵ_{ij} sind unabhängig und haben alle die gleiche Varianz σ^2
3) die ϵ_{ij} sind normalverteilt

Varianzanalyse-Grundgleichung

$$\underbrace{\sum_i \sum_j (y_{ij} - y_{..})^2}_{\text{Gesamtvariabilität}} = \underbrace{\sum_i \sum_j (y_{i.} - y_{..})^2}_{\text{Variabilität zwischen den Gruppen}} + \underbrace{\sum_i \sum_j (y_{ij} - yi.)^2}_{\text{Variabilität innerhalb der Gruppen}}$$

total sum of squares	=	treatment sum of squares	+	residual sum of squares
SS_{tot}	=	SS_{treat}	+	SS_{res}
N–1	=	I–1	+	N–I
df_{tot}	=	df_{treat}	+	df_{res}

Total mean square: $MS_{tot} = SS_{tot}/(N-1)$
Residual mean square: $MS_{res} = SS_{res}/I(J-1) = \hat{\sigma}^2, \quad E(MS_{res}) = \sigma^2$
Treatment mean square: $MS_{treat} = SS_{treat}/(I-1), \quad E(MS_{treat}) = \sigma^2 + \sum JA_i^2/(I-1)$

Anova-Tabelle

Source	SS	df	MS	F	P-Value
Treatment	$\sum_i \sum_j (y_{i.} - y_{..})^2$	I–1	MS_{treat}	$MS_{treat}/MSres$	$P_{H_0}(F > F^*)$
Residuals	$\sum_i \sum_j (y_{ij} - yi.)^2$	N–I	MS_{res}		
Total	$\sum_i \sum_j (y_{ij} - y_{..})^2$	N–1			

Tests und Parameterschätzungen

`F-Test`

$H_0:$ alle $A_i = 0$, $\quad H_A:$ mindestens ein $A_i \neq 0$

Wenn die ϵ_{ij} normalverteilt sind, dann ist $F = MS_{treat}/MS_{res}$ unter H_0 F-verteilt mit $I-1$ und $N-I$ Freiheitsgraden. Verwerfe H_0, falls $F > F_{95\%,I-1,N-I}$.

`Effekt Modell:` $\quad y_{ij} = \mu + A_i + \epsilon_{ij}, \qquad \sum A_i = 0$

Schätzungen: $\hat{\mu} = y_{..}$ $\quad \widehat{\mu + A_i} = y_{i.}$ $\quad \hat{A}_i = y_{i.} - y_{..}$

Voraussage: $\hat{y}_{ij} = \hat{\mu} + \hat{A}_i = yi.$ $\quad$ Residuum: $\hat{\epsilon}_{ij} = y_{ij} - y_{i.}$

`Mean Modell:` $\quad y_{ij} = \mu_i + \epsilon_{ij} \qquad \hat{\mu_i} = y_{i.}$

`Effekt Modell mit anderer Nebenbedingung`: $y_{ij} = \mu + A_i + \epsilon_{ij}$, $A_1 = 0$

$\hat{\mu} = y_{1.}$, $\quad \hat{A}_i = y_{i.} - y_{1.}$

`Treatmentvergleiche`

Treatmentunterschied $A_i - A_{i'}$ wird geschätzt durch $y_{i.} - y_{i'.}$, mit Standardfehler $\sqrt{\sigma^2(1/J + 1/J)} = \sqrt{2\sigma^2/J}$, geschätzt durch $\sqrt{2MS_{res}/J}$.

Multiple Vergleiche

Ein `Kontrast` C ist eine Linearkombination von Effekten: $C = \sum_{i=1}^{I} \lambda_i A_i$ mit $\sum \lambda_i = 0$ und wird durch $\hat{C} = \sum \lambda_i y_{i.}$ geschätzt.

Zwei Kontraste $C_1 = \sum \lambda_i A_i$ und $C_2 = \sum \lambda'_i A_i$ heissen `orthogonal`, wenn $\sum \lambda_i \lambda'_i = 0$. Die entsprechenden Schätzungen sind dann unkorreliert. Es gibt I–1 orthogonale Kontraste.

n geplante, orthog. Kontraste ($n \leq I-1$)	`Bonferroni` (-Holm) Signifikanzniveau α/n
alle Paarvergleiche	`Tukey`: krit. Werte für die Verteilung von $\max \lvert y_{i.} - y_{i'.} \rvert$
komplexe nichtorthogonale oder komplexe ungeplante Vergleiche	`Scheffé`: krit. Wert $\sqrt{(I-1)F_{I-1,N-I,95\%}}$

Vollständiges Blockdesign

Jedes Treatment kommt in jedem Block gleich oft vor.

Modell: $y_{ij} = \mu + A_i + b_j + \epsilon_{ij}$, $\quad b_j$: Effekt des Blocks j.

`Fixed-Effects Model / Modell I`	$\sum A_i = 0,\ \sum b_j = 0,\ \epsilon_{ij} \sim \mathcal{N}(0, \sigma^2)$
`Mixed Model / Modell III`	$\sum A_i = 0,\ b_j \sim \mathcal{N}(0, \sigma_b^2),\ \epsilon_{ij} \sim \mathcal{N}(0, \sigma_e^2)$, alle b_j und ϵ_{ij} unabhängig
`Random-Effects Model / Modell II`	alle Faktoren haben zufällige Effekte

Anova-Tabelle:

Source	SS	df	MS	F
Blocks	...	$J-1$	...	
Treatments	...	$I-1$	...	...
Residual	...	$(I-1)(J-1)$	...	
Total	...	$N-1$		

Multi–Faktor–Experimente

2–Weg–Varianzanalyse

Modell:

$$y_{ijk} = \mu + A_i + B_j + (AB)_{ij} + \epsilon_{ijk}, \quad i = 1, \ldots, I; j = 1, \ldots, J; k = 1, \ldots, K.$$

mit den Nebenbedingungen: $\sum A_i = 0$, $\sum B_j = 0$, $\sum_i (AB)_{ij} = 0$, $\sum_j (AB)_{ij} = 0$.

y_{ijk} ist die k-te Beobachtung mit Faktor A auf Level i und Faktor B auf Level j, μ ist das Gesamtmittel, A_i ist der Haupteffekt des i-ten Levels des Faktors A, B_j ist der Haupteffekt des j-ten Levels des Faktors B, $(AB)_{ij}$ ist die Interaktion des i-ten Levels von A mit dem j-ten Level von B und ϵ_{ijk} ist der zufällige Rest mit $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I})$.

Zerlegung der Gesamtvariabilität: $SS_{tot} = SS_A + SS_B + SS_{AB} + SS_{res}$

$SS_{tot} = \sum\sum\sum(y_{ijk} - y_{...})^2 \quad SS_A = \sum\sum\sum(y_{i..} - y_{...})^2 \quad SS_B = \sum\sum\sum(y_{.j.} - y_{...})^2$
$SS_{AB} = \sum\sum\sum(y_{ij.} - y_{i..} - y_{.j.} + y_{...})^2 \quad SS_{res} = $ ≪Differenz≫

Ein Faktor mit I Levels hat $I - 1$, eine Interaktion zwischen zwei Faktoren mit I und J Levels hat $(I-1)(J-1)$ Freiheitsgrade.

Anova-Tabelle

Source	SS	df	MS	F	P-Wert
A		$I-1$		MS_A/MS_{res}	
B		$J-1$		MS_B/MS_{res}	
AB		$(I-1)(J-1)$		MS_{AB}/MS_{res}	
Residual		≪Differenz≫			
Total		$IJK-1$			

Modell mit zufälligen Effekten (Modell II)

$$y_{ijk} = \mu + a_i + b_j + \epsilon_{ijk}, \qquad i = 1, \ldots, I; j = 1, \ldots, J; k = 1, \ldots, K.$$

μ ist das Gesamtmittel,
a_i ist der zufällige Effekt von Faktor a, $a_i \sim \mathcal{N}(0, \sigma_a^2)$,
b_j ist der zufällige Effekt von Faktor b, $b_j \sim \mathcal{N}(0, \sigma_b^2)$,
ϵ_{ijk} ist der zufällige Rest, $\epsilon_i \sim \mathcal{N}(0, \sigma_e^2)$, die a_i, b_j und ϵ_{ijk} sind alle unabhängig.

Schätzungen:
$\hat{\sigma}_e^2 = MS_{res}$
$\hat{\sigma}_a^2 = (MS_a - MS_{res})/JK$
$\hat{\sigma}_b^2 = (MS_b - MS_{res})/IK$