Lehrstuhl fuer Statistik, Oekonometrie und empirische Wirtschaftsforschung, Universitaet Tuebingen

Varianzanalyse

Dr. S. Prohl

6. Mai 2007/ SoSe 2007

Varianzanalyse

- Problemstellung und Typen der Varianzanalyse
- Einfaktorielle Varianzanalyse:
  - Modellannahmen
  - Analyse der Abweichungsquadrate
  - Pruefung der statistischen Unabhaengigkeit
- Zweifaktorielle Varianzanalyse:
  - Analyse der Abweichungsquadrate
  - Pruefung der statistischen Unabhaengigkeit

Das einfache Modell

$$y_{gk} = \mu + \alpha_g + e_{gk}.$$

- wobei: g -Faktorstufe der unabhaengigen Variablen (g=1,..,G), k - Beobachtungswert innerhalb einer Faktorstufe (k=1,..,K)

Modellannahmen:

- A.1 $\sum_g K\alpha_g = 0$.
- A.2 $\sum_g \alpha_g = 0$.
- A.3 $e_{gk} \sim N(0, {\sigma_e}^2)$, i.i.d.
- A.4 $E(y_{gk}) = \mu$.
- A.5 $Var(y_{gk}) = {\sigma_e}^2$.
- A.6 $Cov(y_{gk}, y_{ij}) = 0$, wobei $g \neq i$ und $k \neq j$.

Analogie zur multiplen linearen Regression

Ziele der Varianzanalyse

Ausgangspunkt: das einfache Modell

$$y_{gk} = \mu + \alpha_g + e_{gk}\,.$$

Ziele:

- Aufteilung der Gesamtstreuung in einen:
  - erklaerten Anteil
  - nicht-erklaerten Anteil

Einfache Varianzanalyse

- Die Summe der quadrierten Abweichungen (SS_{total}, Sum of Squares)
  $\sum_{g=1}^{G}\sum_{k=1}^{K}(y_{gk}-\bar{y})^2 =$
  $\sum_{g=1}^{G}K(\bar{y_g}-\bar{y})^2+\sum_{g=1}^{G}\sum_{k=1}^{K}(y_{gk}-\bar{y_g})^2.$
  - wobei $SS_{total} = SS_{between} + SS_{within}$
- Die mittlere quadratische (Gesamt-)Abweichung (MS):
  $\sum_{g=1}^{G}\sum_{k=1}^{K}(y_{gk}-\bar{y})^2/(GK-1) =$
  $\sum_{g=1}^{G}K(\bar{y_g}-\bar{y})^2/(G-1)+\sum_{g=1}^{G}\sum_{k=1}^{K}(y_{gk}-\bar{y_g})^2/G(K-1).$
  - wobei $MS_{total} = MS_{between} + MS_{within}$.

Pruefung der statistischen Unabhaengigkeit

- Das Ausgangsmodell:
  $y_{gk} = \mu + \alpha_g + e_{gk}$.
- Aus der Beziehung:
  $MS_{between} = SS_{between}/(G-1)$.
- Varianzanalyse erlaubt einen F-Test:
  $F_{emp} = MS_{between}/MS_{within}$.
  mit $H_0 : \alpha_1 = \alpha_2 = ... = \alpha_g = 0$.
- F-Statistik folgt unter H_0 der F-Verteilung mit (G-1) und (G(K-1)) Freiheitsgraden.
- Falls $F_{emp} > F_{theor} \Rightarrow H_0$ wird verworfen

Zweifaktorielle Varianzanalyse

Das Ausgangsmodell:
$y_{ghk} = \mu + \alpha_g + \beta_h + (\alpha\beta)_{gh} + e_{ghk}$.

- wobei alle Annahmen der einfaktoriellen Varianzanalyse gelten.

Ziele der Varianzanalyse –> Aufteilung der Gesamtstreuung in einen:

- erklaerten Anteil
- nicht-erklaerten Anteil

Zweifaktorielle Varianzanalyse

Schaetzung der Parameter:

- Gesamtmittelwert (μ): $\bar{y}$
- Wirkung von Faktor A (α_g): $(\bar{y}_g - \bar{y})$
- Wirkung von Faktor B (β_h): $(\bar{y}_h - \bar{y})$
- Interaktionseffekt ($\alpha\beta_{gh}$): $(\bar{y}_{gh} - \hat{y}_{gh})$

Zweifaktorielle Varianzanalyse

- Gesamtstreuungszerlegung:
  $SS_{total} = SS_{between} + SS_{within}$
- Streuung zwischen den Gruppen:
  $SS_{between} = SS_A + SS_B + SS_{AB}$
- Mittlere quadratische Abweichung:
- fuer Haupteffekt A: $MS_A = SS_A/(G-1)$
- fuer Haupteffekt B: $MS_B = SS_B/(H-1)$
- fuer Interaktionsterm AB: $MS_{AB} = SS_{AB}/(G-1)(H-1)$
- innerhalb der Gruppen: $MS_{within} = SS_{within}/GH(K-1)$

Test fuer die einzelnen Hypothesen

- Haupteffekt A:
  $F_{A,emp} = MS_A / MS_{within}$
  - krit. Wert mit (G-1) Zaehler- und (DH(K-1)) Nennerfreiheitsgraden.
- Haupteffekt B
  $F_{B,emp} = MS_B / MS_{within}$
  - krit. Wert mit (H-1) Zaehler- und (GH(K-1)) Nennerfreiheitsgraden.
- Interaktionseffekt
  $F_{AB,emp} = MS_{AB} / MS_{within}$
  - krit. Wert mit (G-1)(H-1) Zaehler- und (GH(K-1)) Nennerfreiheitsgraden.

Varianzanalyse: Kontrollfragen

- Definieren Sie die Modellannahmen der einfaktoriellen Varianzanalyse.
- Erklaeren Sie die Zerlegung der Gesamtstreuung im Falle der einfaktoriellen Varianzanalyse.
- Was testet der F-Test mit $H_0 : \alpha_1 = \alpha_2 = ... = \alpha_g = 0$?
- Wie sieht die Zerlegung der Gesamtstreuung im Falle der zweifaktoriellen Varianzanalyse aus?
- Formulieren Sie H_O des F-Testes fuer Haupteffekt A, und den Interaktionsterm AB.