

Beregningskemaer til variansanalyse

Peter Stoltze*

29. april 2010

Ensidet variansanalyse

I Tabel 1 på næste side er vist et beregningskema, der kan fungere som skabelon ved ensidet variansanalyse. Med X_{ij} menes den j 'te observation i den i 'te gruppe. Generelt er der n observationer i hver gruppe, og der er a grupper. Skemaet benytter *punktum-streg* notation. Erstattes et indeks (fx i) med et punktum betyder det, at der er summeret over dette indeks. Tilsvarende betyder en streg at der er delt med antallet af observationer for at få et gennemsnit.

$$X_{i.} = \sum_{j=1}^n X_{ij}$$

$$\bar{X}_{i.} = \frac{1}{n} \sum_{j=1}^n X_{ij}$$

$$X_{.j} = \sum_{i=1}^a X_{ij}$$

$$\bar{X}_{.j} = \frac{1}{a} \sum_{i=1}^a X_{ij}$$

$$X_{..} = \sum_{i=1}^a \sum_{j=1}^n X_{ij}$$

$$\bar{X}_{..} = \frac{1}{an} \sum_{i=1}^a \sum_{j=1}^n X_{ij}$$

Bemærk at *punktummet* i et operatorhierarki kommer før potensfunktioner, således at

$$X_{i.}^2 = \left(\sum_{j=1}^n X_{ij} \right)^2$$

Signifikanssandsynligheden p bestemmes ved at sammenligne F_{obs} med en F -fordeling med $a-1$ frihedsgrader i tælleren og $a(n-1)$ frihedsgrader i nævneren:

$$p = P(F \geq F_{\text{obs}})$$

Meget små værdier af p tyder på, at der er forskel mellem i hvert fald to af de a grupper.

*Email: stat@peterstoltze.dk

Variations-årsag	Kvadratsum SAK	Frihedsgrader DF	Middelvadratsum MS	Forventet MS EMS	F-værdi F_{obs}
Mellem grupper	$\text{SAK}_a = n \sum_{i=1}^a (\bar{X}_i - \bar{X}_{..})^2$ $= \frac{1}{n} \sum_{i=1}^a X_i^2 - \frac{1}{an} X_{..}^2$	$a - 1$	$\text{MS}_a = \frac{\text{SAK}_a}{a - 1}$	$\sigma^2 + n \frac{1}{a - 1} \sum_{i=1}^a \alpha_i^2$	$F_{\text{obs}} = \frac{\text{MS}_a}{\text{MS}_e}$
Indenfor grupper	$\text{SAK}_e = \sum_{i=1}^a \sum_{j=1}^n (X_{ij} - \bar{X}_i)^2$ $= \sum_{i=1}^a \sum_{j=1}^n X_{ij}^2 - \frac{1}{n} \sum_{i=1}^a X_i^2$	$a(n - 1)$	$\text{MS}_e = \frac{\text{SAK}_e}{a(n - 1)}$	σ^2	
Total	$\text{SAK}_{\text{tot}} = \sum_{i=1}^a \sum_{j=1}^n (X_{ij} - \bar{X}_{..})^2$ $= \sum_{i=1}^a \sum_{j=1}^n X_{ij}^2 - \frac{1}{an} X_{..}^2$	$an - 1$			

Tabel 1: Ensidet variansanalyse med lige mange observationer i hver gruppe.

Tosidet variansanalyse uden gentagelser

Et forsøg med to faktorer men ingen gentagelser kalder man ofte for et *blokforsøg*. Når der ikke er nogen gentagelser kan man ikke bestemme nogen vekselvirkning. Et eksempel på blokke kunne være skoleklasse — der kan være en effekt af den specifikke klasse, uden at den i sig selv er interessant.

Den statistiske model er

$$X_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij} ,$$

hvor α_i er virkningen af faktor A på niveau i , $i = 1, \dots, a$, og β_j er virkningen af faktor B på niveau j , $j = 1, \dots, b$. Vi antager at restleddene er uafhængige og stammer fra normalfordelinger med samme varians σ^2 og middelværdi nul, $\{\varepsilon_{ij}\} \sim N(0, \sigma^2)$. Desuden er $\sum_{i=1}^a \alpha_i = \sum_{j=1}^b \beta_j = 0$.

Variansanalysekemaet er vist i Tabel 2.

Variations-årsag	Kvadratsum SAK	Frihedsgrader DF	Middelvadratsum MS	Forventet MS EMS	F-værdi F_{obs}
Rækker (behandling)	$SAK_a = b \sum_{i=1}^a (\bar{X}_i - \bar{X}_{..})^2$ $= \frac{1}{b} \sum_{i=1}^a X_i^2 - \frac{1}{ab} X_{..}^2$	$a - 1$	$MS_a = \frac{SAK_a}{a - 1}$	$\sigma^2 + \frac{b}{a - 1} \sum_{i=1}^a \alpha_i^2$	$F_a = \frac{MS_a}{MS_e}$
Søjler (blok)	$SAK_b = a \sum_{j=1}^b (\bar{X}_j - \bar{X}_{..})^2$ $= \frac{1}{a} \sum_{j=1}^b X_{.j}^2 - \frac{1}{ab} X_{..}^2$	$b - 1$	$MS_b = \frac{SAK_b}{b - 1}$	$\sigma^2 + \frac{a}{b - 1} \sum_{j=1}^b \beta_j^2$	$F_b = \frac{MS_b}{MS_e}$
Residual	$SAK_e = \sum_{i=1}^a \sum_{j=1}^b (X_{ij} - \bar{X}_i - \bar{X}_j + \bar{X}_{..})^2$ $= \sum_{i=1}^a \sum_{j=1}^b X_{ij}^2 - \frac{1}{b} \sum_{i=1}^a X_i^2 - \frac{1}{a} \sum_{j=1}^b X_{.j}^2 + \frac{1}{ab} X_{..}^2$	$(a - 1)(b - 1)$	$MS_e = \frac{SAK_e}{(a - 1)(b - 1)}$	σ^2	
Total	$SAK_{tot} = \sum_{i=1}^a \sum_{j=1}^b (X_{ij} - \bar{X}_{..})^2$ $= \sum_{i=1}^a \sum_{j=1}^b X_{ij}^2 - \frac{1}{ab} X_{..}^2$	$ab - 1$			

Tabel 2: Tosidet variansanalysetabel uden gentagelser.

Tosidet variansanalyse med gentagelser

Når der er gentagelser indenfor hvert niveau af to faktorer er det muligt at bestemme en *vekselvirkning* i mellem disse to faktorer. Fortolkningen af en signifikant vekselvirkning er, at effekten af den ene faktor afhænger af niveauet på den anden faktor. Der er ikke ren additivitet imellem de to faktorer.

Den statistiske model er

$$X_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_{ijk} ,$$

hvor α_i er virkningen af faktor A på niveau i , $i = 1, \dots, a$, β_j er virkningen af faktor B på niveau j , $j = 1, \dots, b$. Vekselvirkningen mellem de to faktorer, $A \times B$, betegnes γ_{ij} . Nogle gange bruges notationen $(\alpha\beta)_{ij}$. For hver kombination af de to faktorer er der n observationer således at $k = 1, \dots, n$. Vi antager at restleddene er uafhængige og stammer fra normalfordelinger med samme varians σ^2 og middelværdi nul, $\{\varepsilon_{ijk}\} \sim N(0, \sigma^2)$. Desuden er $\sum_{i=1}^a \alpha_i = \sum_{j=1}^b \beta_j = \sum_{i=1}^a \gamma_{ij} = \sum_{j=1}^b \gamma_{ij} = 0$.

Variansanalysekemaet er vist i Tabel 3.

Et forsøg med to krydsede faktorer kan naturligvis også laves som et blokforsøg. Den statistiske model er her

$$X_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \gamma_k + \varepsilon_{ijk} ,$$

hvor γ_k beskriver blokeffekten. Analysen af denne model gennemgås ikke her.

Variations-årsag	Kvadratsum SAK	Frihedsgrader DF	Middelkvadratsum MS	Forventet MS EMS	F-værdi F_{obs}
Rækker (Faktor A)	$SAK_A = bn \sum_{i=1}^a (\bar{X}_{i..} - \bar{X}...)^2$ $= \frac{1}{bn} \sum_{i=1}^a X_{i..}^2 - \frac{1}{abn} X...^2$	$a - 1$	$MS_A = \frac{SAK_A}{a - 1}$	$\sigma^2 + \frac{bn}{a - 1} \sum_{i=1}^a \alpha_i^2$	$F_A = \frac{MS_A}{MS_e}$
Søjler (Faktor B)	$SAK_B = an \sum_{j=1}^b (\bar{X}_{.j.} - \bar{X}...)^2$ $= \frac{1}{an} \sum_{j=1}^b X_{.j.}^2 - \frac{1}{abn} X...^2$	$b - 1$	$MS_B = \frac{SAK_B}{b - 1}$	$\sigma^2 + \frac{an}{b - 1} \sum_{j=1}^b \beta_j^2$	$F_B = \frac{MS_B}{MS_e}$
Vekselvirkning (A × B)	$SAK_{AB} = n \sum_{i=1}^a \sum_{j=1}^b (\bar{X}_{ij.} - \bar{X}_{i..} - \bar{X}_{.j.} + \bar{X}...)^2$ $= SAK_{tot} - SAK_A - SAK_B - SAK_{AB}$	$(a - 1)(b - 1)$	$MS_{AB} = \frac{SAK_{AB}}{(a - 1)(b - 1)}$	$\sigma^2 + \frac{n}{(a - 1)(b - 1)} \sum_{i=1}^a \sum_{j=1}^b \gamma_{ij}$	$F_{AB} = \frac{MS_{AB}}{MS_e}$
Residual	$SAK_e = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n (X_{ijk} - \bar{X}_{ij.})^2$ $= \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n X_{ijk}^2 - \frac{1}{n} \sum_{i=1}^a \sum_{j=1}^b X_{ij.}^2$	$ab(n - 1)$	$MS_e = \frac{SAK_e}{ab(n - 1)}$	σ^2	
Total	$SAK_{tot} = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n (X_{ijk} - \bar{X}...)^2$ $= \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n X_{ijk}^2 - \frac{1}{abn} X...^2$	$abn - 1$			

Tabel 3: Tosidet variansanalysetabel med vekselvirkning og lige mange observationer i hver gruppe.