Kursus 02403
Introduktion til Statistik

Klaus Kaae Andersen

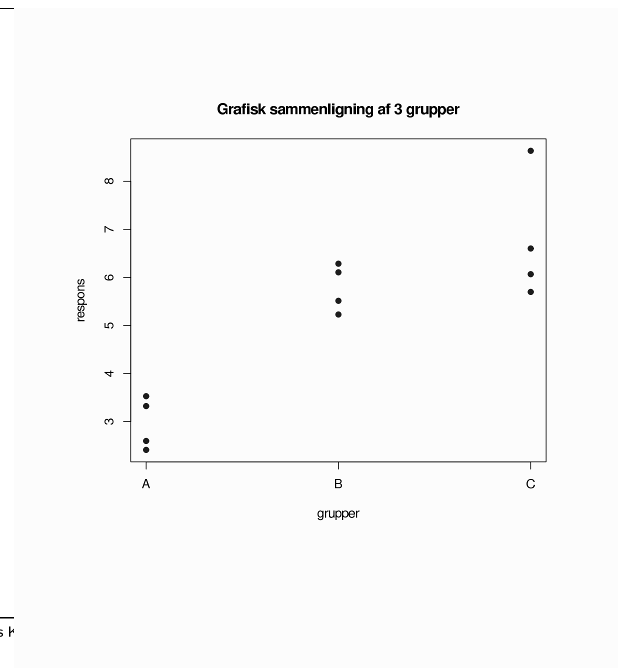
Informatics and Mathematical Modelling
Building 321 - room 011
Technical University of Denmark
2800 Lyngby – Denmark
e-mail: kka@imm.dtu.dk

Variansanalyse (kap 12)

- Ensidet variansanalyse
- Grafisk præsentation
- Tosidet variansanalyse

Gruppe A	Gruppe B	Gruppe C
2.8	5.5	5.8
3.6	6.3	8.3
3.4	6.1	6.9
2.3	5.7	6.1

- Er der forskel (i middel) på grupperne A, B og C?
- Variansanalyse (ANOVA) kan anvendes til analysen såfremt observationerne i hver gruppe kan antages at være normalfordelte.



En-sidet variansanalyse

- Vi betragter modellen

$$X_{ij} = \mu + \alpha_i + \epsilon_{ij}$$

hvor det antages, at

$$\epsilon_{ij} \sim N(0, \sigma^2)$$

- μ er gennemsnit for alle målinger
- α_i angiver niveau af 'gruppe' i

Ensidet variansanalyse

Vi vil nu sammenligne (flere end to) middelværdier $\mu + \alpha_i$ i modellen

$$X_{ij} = \mu + \alpha_i + \epsilon_{ij}, \quad \epsilon_{ij} \sim N(0, \sigma^2)$$

dvs vi kan specificere hypotesen:

$$H_0: \quad \alpha_i = \alpha_j \quad \text{for alle } i, j$$
$$H_1: \quad \alpha_i \neq \alpha_j \quad \text{for mindst et } i, j$$

Ensidet variansanalyse

Svarende til modellen

$$X_{ij} = \mu + \alpha_i + \epsilon_{ij}, \quad \epsilon_{ij} \sim N(0, \sigma^2)$$

kan den totale variation i data opspaltes:

$$SST = SS(Tr) + SSE$$

- 'Ensidet' hentyder til, at der kun er én faktor i forsøget, på i alt k nivauer
- Metoden kaldes variansanalyse, fordi testningen foregår ved at sammenligne varianser

Ensidet variansanalyse

Vi har:

$$SST = SS(Tr) + SSE$$

fås teststørrelsen, F:

$$F = \frac{SS(Tr)/(k-1)}{SSE/(N-k)}$$

hvor k er antal nivauer af faktoren, og N er antal observationer

Formler for kvadratafvigelsessummer

$$SST = \sum_{i=1}^{k}\sum_{j=1}^{n_i} y_{ij}^2 - C$$

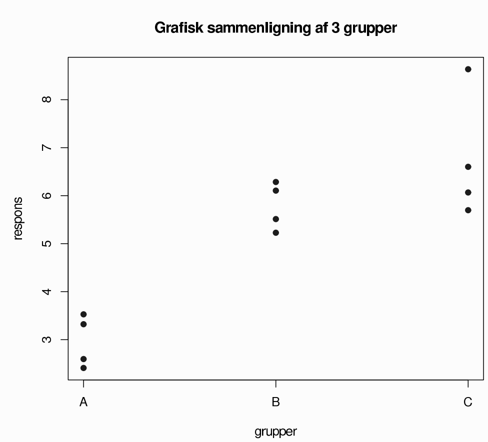
$$SS(Tr) = \sum_{i=1}^{k} \frac{T_i^2}{n_i} - C$$

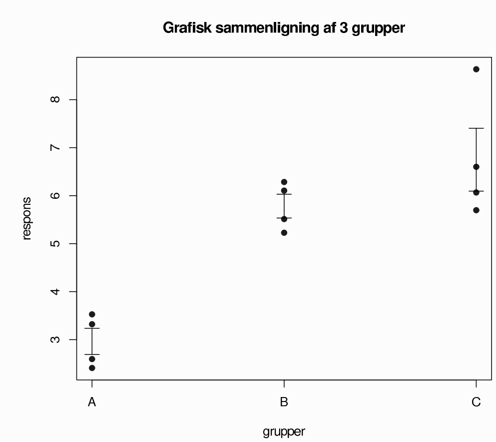
hvor

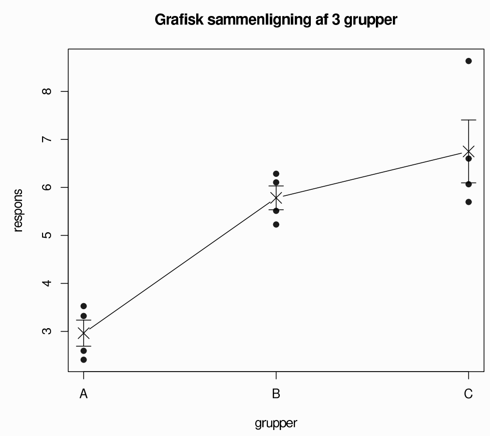
$$C = \frac{T_\cdot^2}{N} \quad T_i = \sum_{j=1}^{n_i} y_{ij} \quad T_\cdot = \sum_{i=1}^{k} T_i$$

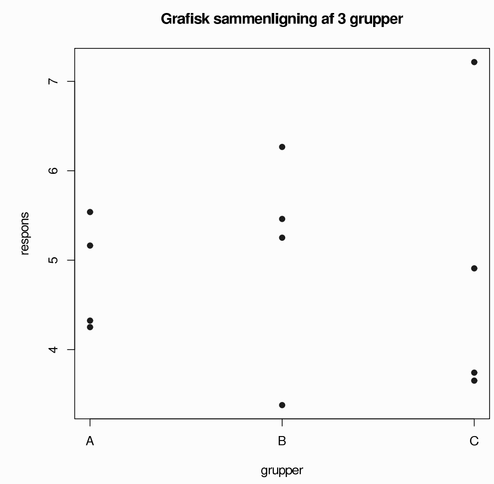
Variansanalysetabel

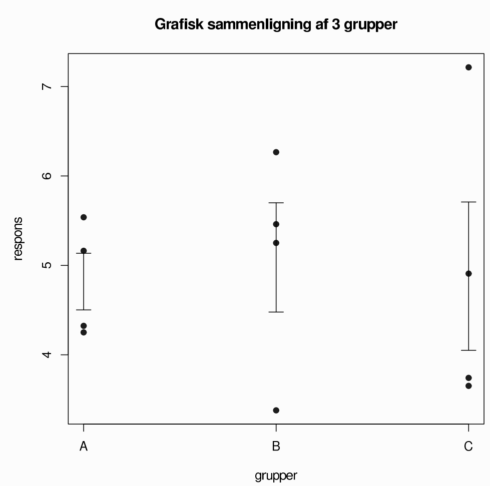
Variations-kilde	Friheds-grader	Kvadratafvig. sum	Test-størrelse F
Behandlig	$k-1$	$SS(Tr)$	$\frac{SS(Tr)/(k-1)}{SSE/(N-k)}$
Residual	N-k	SSE	
Total	N-1	SST	

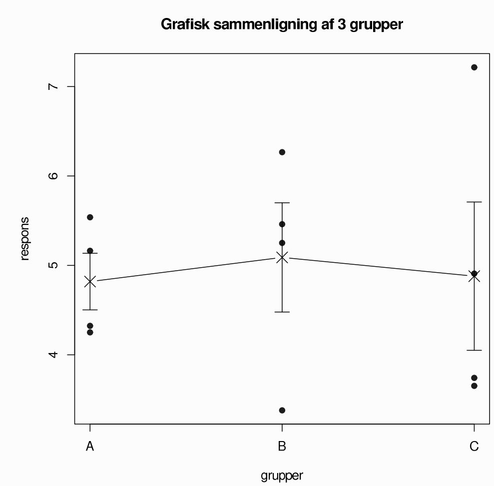





Grafisk sammenligning af 3 grupper
respons
3
4
5
6
7
8
A
B
C
grupper


Grafisk sammenligning af 3 grupper
respons
4
5
6
7
A
B
C
grupper


Grafisk sammenligning af 3 grupper
respons
4
5
6
7
A
B
C
grupper


Grafisk sammenligning af 3 grupper
respons
4
5
6
7
A
B
C
grupper

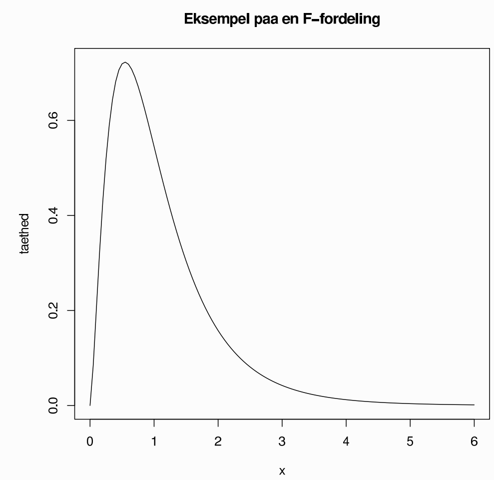
En-sidet variansanalyse: test

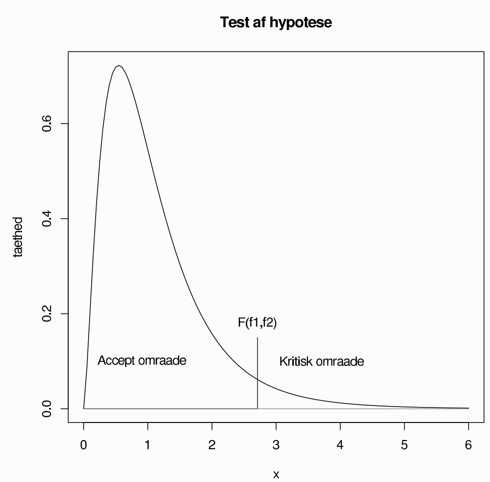
- Teststørrelsen F beregnes og signifikansniveau α vælges

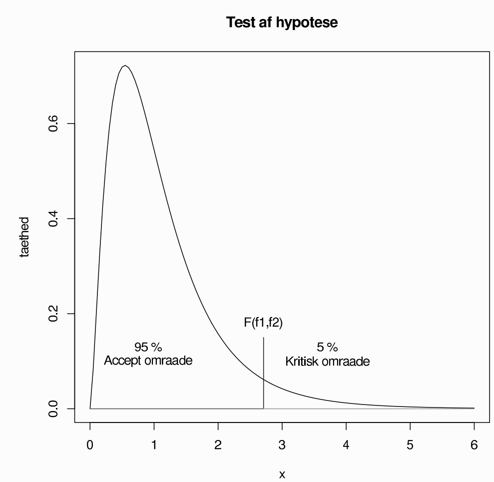
$$F = \frac{SS(Tr)/(k-1)}{SSE/(N-k)}$$

- Teststørrelsen sammenlignes med en fraktil i F fordelingen:

$$F \sim F_\alpha(k-1, N-k)$$







Grafisk præsentation

- Det er nyttigt at lave en grafisk præsentation af data inden man foretager variansanalysen.
  - ◇ Er der en forskel mellem grupper?
  - ◇ Er variansen ens indenfor grupper?
  - ◇ Kan data overhovedet antages at være normalfordelt?

Tosidet variansanalyse

- Vi antager nu, at vi har modellen

$$X_{ij} = \mu + \alpha_i + \beta_j + \epsilon_{ij}, \quad \epsilon_{ij} \sim N(0, \sigma^2)$$

dvs vi har to indelingskriterier, både α og β, hvor β også kan opfattes som en blok, hvorfor designet også kaldes et randomiseret blokforsøg

Tosidet variansanalyse

	A_1	A_2	A_3
B_1	x	x	x
B_2	x	x	x
B_3	x	x	x
B_4	x	x	x

Tosidet variansanalyse

- Svarende til modellen

$$X_{ij} = \mu + \alpha_i + \beta_j + \epsilon_{ij}, \quad \epsilon_{ij} \sim N(0, \sigma^2)$$

$$SST = SS(Tr) + SS(Bl) + SSE$$

Variansanalysetabel

Variations-kilde	Friheds-grader	Kvadrat-afvig.sum	Test-størrelse F
Behandlig	$a-1$	$SS(Tr)$	$\frac{SS(Tr)/(a-1)}{SSE/((a-1)(b-1))}$
Blokke	$b-1$	$SS(Bl)$	$\frac{SS(Bl)/(b-1)}{SSE/((a-1)(b-1))}$
Residual	$(a-1)(b-1)$	SSE	
Total	$N-1$	SST	

- Kritisk værdi for blokke: $F_\alpha(b-1,(a-1)(b-1))$
- Kritisk værdi for behandling: $F_\alpha(a-1,(a-1)(b-1))$

Variansanalyse (kap 12)

- Ensidet variansanalyse
- Grafisk præsentation
- Tosidet variansanalyse

PENSUM - SLUT