

Kursus 02403

Introduktion til Statistik

Forelæsning 12: Variansanalyse

Jan Kloppenborg Møller

DTU Informatik

Bygning 305 - rum 219

Danmarks Tekniske Universitet

2800 Lyngby – Danmark

e-mail: jkm@imm.dtu.dk

Oversigt

- 1 Envejs variansanalyse
 - Eksempel, model, hypotese
 - Beregning - variationsopspaltning
 - Hypotesetest
 - Variansanalysetabel
 - Grafisk præsentation
 - Test, F-fordeling
 - Eksempel 1
 - Eksempel 2
- 2 Tovejs variansanalyse
 - Model og opspaltning af varians
 - Variansanalysetabel
 - Eksempel 2
- 3 R (R note 11)

Oversigt

- 1 Envejs variansanalyse
 - Eksempel, model, hypotese
 - Beregning - variationsopspaltning
 - Hypotesetest
 - Variansanalysetabel
 - Grafisk præsentation
 - Test, F-fordeling
 - Eksempel 1
 - Eksempel 2
- 2 Tovejs variansanalyse
 - Model og opspaltning af varians
 - Variansanalysetabel
 - Eksempel 2
- 3 R (R note 11)

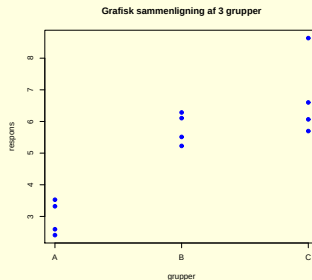
Envejs variansanalyse - eksempel

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.

Envejs variansanalyse - eksempel



Envejs variansanalyse, model

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

hvor det antages, at

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

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

Envejs variansanalyse, hypotese

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

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

dvs vi kan specificere hypotesen:

$$H_0 : \alpha_i = 0 \quad \text{for alle } i$$

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

Envejs variansanalyse, opspaltning

Svarende til modellen

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

kan den totale variation i data opspaltes:

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

'Envejs' 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

Formler for kvadratafgivelsessummer

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

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

hvor

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

Theorem 12.1

$$\underbrace{\sum_{i=1}^k \sum_{j=1}^{n_i} (y_{ij} - \bar{y})^2}_{\text{SST}} = \underbrace{\sum_{i=1}^k \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2}_{\text{SSE}} + \underbrace{\sum_{i=1}^k n_i (\bar{y}_i - \bar{y})^2}_{\text{SS(Tr)}}$$

hvor $\bar{y}_i$ er gruppegennemsnit, $\bar{y}$ er et estimat for μ og $\bar{y}_i - \bar{y}$ er estimer for α_i

Envejs variansanalyse, F-test

Vi har:

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

fås teststørrelsen, F :

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

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

Ensidig 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:

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

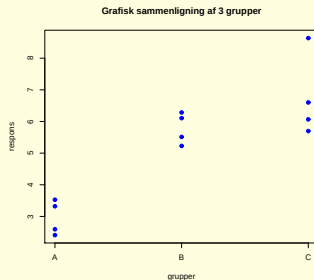
Variansanalysetabel

Variationskilde	Frihedsgrader	Kvadratafgvig. sum	Teststørrelse F
Behandlig	$k - 1$	$SS(Tr)$	$\frac{SS(Tr)/(k-1)}{SSE/(N-k)}$
Residual	$N-k$	SSE	
Total	$N-1$	SST	

Grafisk præsentation

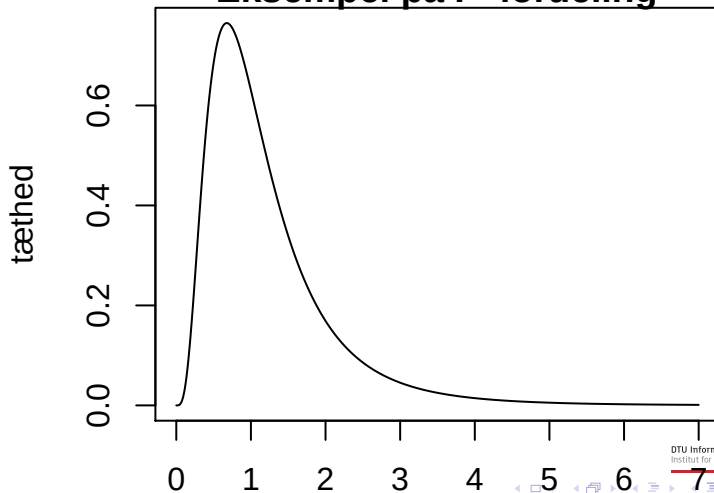
- Det er nyttigt at lave en grafisk præsentation af data inden man foretager variansanalysen.
- Er der en forskel mellem grupper?
- Er variansen ens indenfor grupper?
- Kan data overhovedet antages at være normalfordelt?

Ensidig variansanalyse

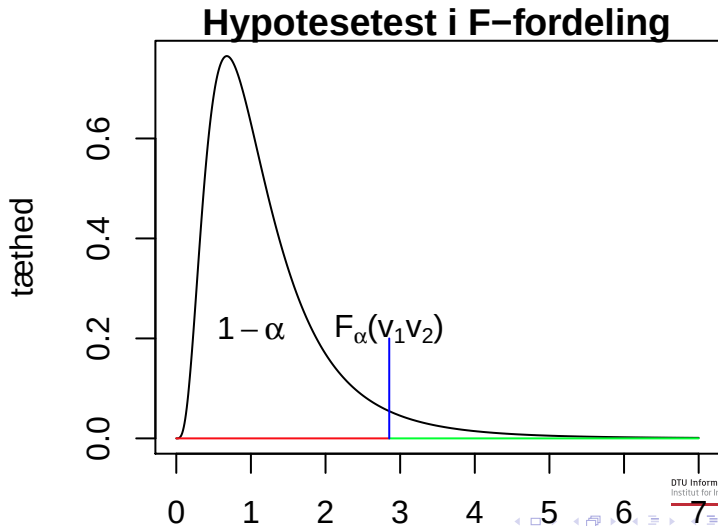


F-fordeling

Eksempel på F-fordeling



F-fordeling

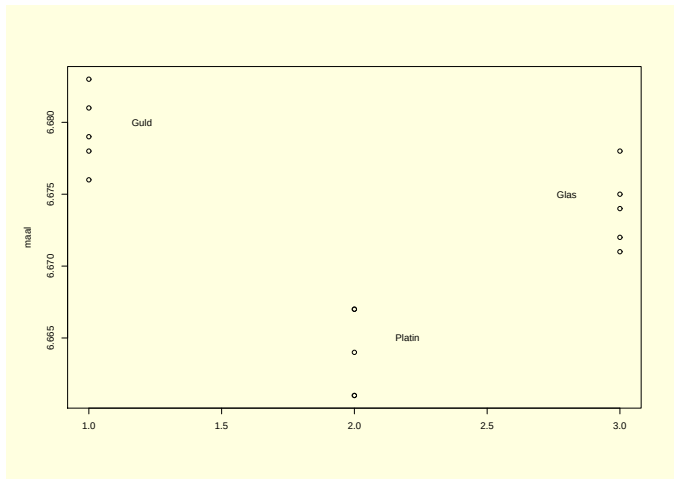


Eksempel 1

Ved målinger af gravitationskonstanten, G , benyttede Heyl (1930) kugler af tre forskellige materialer og fik bl.a. følgende observationer af G (enhed $10^{-11}Nm^2kg^{-2}$)

Materiale	Måling				
Guld	6.683	6.681	6.676	6.678	6.679
Platin	6.661	6.661	6.667	6.667	6.664
Glas	6.678	6.671	6.675	6.672	6.674

Eksempel 1



Eksempel I

Opstil en statistisk model for forsøget, og test om der er forskel på bestemmelse af gravitationskonstanten for de 3 materialer. Anvend signifikansniveau $\alpha = 5\%$

Var. kilde	SSQ	df	MS	F teststørrelse
Materiale	$6.1053 \cdot 10^{-4}$			
Residual	$9.5200 \cdot 10^{-5}$			
Total	$7.0573 \cdot 10^{-4}$			

Eksempel 2

Et høreapparat skal tilpasses individuelt. En måde at validere et apparat på er at afspille en liste (eller serie) ord ved lavt volumen og bede patienten om at gentage ordene. I et studie ville man sammenligne 4 forskellige lister, der skulle have samme sværhedsgrad og samme antal ord

Test	1	2	3	4	5	6	7	8	...	24
list I	28	24	32	30	34	30	36	32	...	40
list II	20	16	38	20	34	30	30	28	...	44
list III	24	32	20	14	32	22	20	26	...	34
list IV	26	24	22	18	24	30	22	28	...	42

Eksempel 2

Opstil en statistisk model for forsøget. Redegør for hvordan man kan teste om der er forskel i sværhedsgrad for de 4 lister. Burde undersøgelsen laves anderledes for at sikre valide resultater?

Eksempel 2

Det viser sig, at man har anvendt de samme 24 personer til hver af de fire lister, som vist i tabellen

Test/Person	1	2	3	4	5	6	7	8	...	24
list I	28	24	32	30	34	30	36	32	...	40
list II	20	16	38	20	34	30	30	28	...	44
list III	24	32	20	14	32	22	20	26	...	34
list IV	26	24	22	18	24	30	22	28	...	42

Oversigt

- 1 Envejs variansanalyse
 - Eksempel, model, hypotese
 - Beregning - variationsopspaltning
 - Hypotesetest
 - Variansanalysetabel
 - Grafisk præsentation
 - Test, F-fordeling
 - Eksempel 1
 - Eksempel 2
- 2 Tovejs variansanalyse
 - Model og opspaltning af varians
 - Variansanalysetabel
 - Eksempel 2
- 3 R (R note 11)

Tovejs variansanalyse

Vi antager nu, at vi har modellen

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

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

Tovejs 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

Tovejs variansanalyse, Model og opspaltning af varians

- Svarende til modellen

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

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

Variansanalysetabel

Variationskilde	Frihedsgrader	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: $F_{\alpha}(a - 1, (a - 1)(b - 1))$

Eksempel 2

Det viser sig, at man har anvendt de samme 24 personer til hver af de fire lister, som vist i tabellen

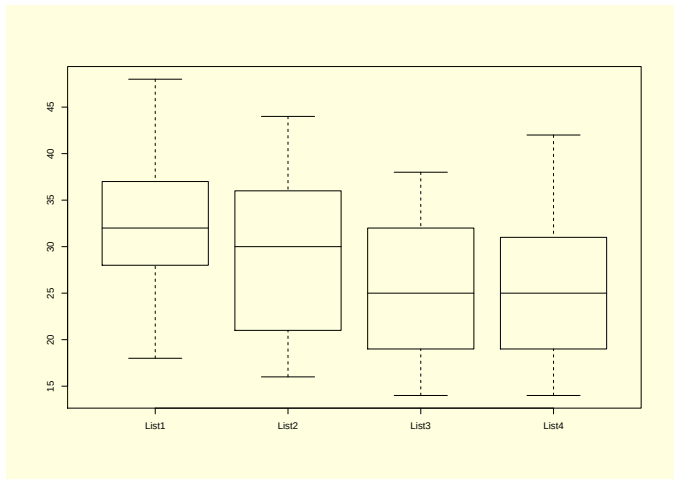
Test/Person	1	2	3	4	5	6	7	8	...	24
list I	28	24	32	30	34	30	36	32	...	40
list II	20	16	38	20	34	30	30	28	...	44
list III	24	32	20	14	32	22	20	26	...	34
list IV	26	24	22	18	24	30	22	28	...	42

Eksempel 2

Opstil en statistisk model for forsøget. Redegør for hvordan man kan teste om der er forskel i sværhedsgrad for de 4 lister.

Var. kilde	SSQ	df	MS	F teststørrelse
Liste	920.5			
Person	3231.6			
Residual	2506.5			
Total	6658.6			

Eksempel 2



Oversigt

- 1 Envejs variansanalyse
 - Eksempel, model, hypotese
 - Beregning - variationsopspaltning
 - Hypotesetest
 - Variansanalysetabel
 - Grafisk præsentation
 - Test, F-fordeling
 - Eksempel 1
 - Eksempel 2
- 2 Tovejs variansanalyse
 - Model og opspaltning af varians
 - Variansanalysetabel
 - Eksempel 2
- 3 R (R note 11)

R (R note 11)

```
> attach(C12tin)
> Lab <- factor(Lab)
> anova(lm(weight~Lab))
Analysis of Variance Table
```

Response: weight

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Lab	3	0.013006	0.0043354	2.8097	0.05038 .
Residuals	44	0.067892	0.0015430		

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R (R note 11)

```
> attach(example)
> treatm <- factor(treatm)
> block <- factor(block)
> anova(lm(y~treatm+block))
```

Analysis of Variance Table

Response: y

Terms added sequentially (first to last)

	Df	Sum of Sq	Mean Sq	F Value	Pr(F)
treatm	2	56	28.00000	3.230769	0.1116192
block	3	90	30.00000	3.461538	0.0913831
Residuals	6	52	8.66667		

Oversigt

- 1 Envejs variansanalyse
 - Eksempel, model, hypotese
 - Beregning - variationsopspaltning
 - Hypotesetest
 - Variansanalysetabel
 - Grafisk præsentation
 - Test, F-fordeling
 - Eksempel 1
 - Eksempel 2
- 2 Tovejs variansanalyse
 - Model og opspaltning af varians
 - Variansanalysetabel
 - Eksempel 2
- 3 R (R note 11)