

Meddelelse 29

Forelæsningerne i sidste uge

(Onsdag den 30. april): Afsluttede omtalen af beregninger i SAS i tilknytning til variansanalyse og begyndte på Kapitel 12 om den negative binomialfordeling.

(Fredag den 2. maj): Fortsætter gennemgangen af Kapitel 12 om den negative binomialfordeling.

Forelæsningerne i denne uge

(Onsdag den 7. maj): Den negative binomialfordeling. Maksimum likelihood estimation og beregninger i SAS. Begyndte på en gennemgang asymptotisk likelihood teori, som baggrund for at finde konfidensintervaller for parametrene i den negative binomialfordeling.

(Fredag den 9. maj): Asymptotisk likelihood teori.

Forelæsningerne i næste uge

(Onsdag den 14. maj): Dette er sidste forelæsning. Afrunding og moralsk oprustning.

(Fredag den 16. maj): Store bededag.

Øvelserne i de næste uger

(Uge 20: 12.-18. maj): (Næstsidste øvelsesgang) Opgave 3 i eksamenssættet (i Statistik 1) fra sommeren 2001, opgave 3 (Statistik 1) fra sommeren 1999, opgave 4 (Statistik 1) fra sommeren 2002.

(Uge 21: 19.-25. maj): (Sidste øvelsesgang) 12.2, 12.7, 12.8

IKKE Udleveret materiale

Pensum og eksamensspørgsmål kommer i næste uge.

STAT1, TØ m. 27

opgave 10.17

b_{ik} , p_{ik} , e_{ijl} er uafh.

1)

$$\bar{X}_{i..} \sim N(\mu, \sigma^2)$$

$$\mu = \sum_j \alpha_i + \beta_j / 24 = \alpha_i + \frac{1}{4} \sum_j \beta_j = \alpha_i + \bar{\beta}.$$

$$\sigma^2_{\text{uafh.}} = \frac{1}{24^2} (4^2 (6\sigma_B^2) + 4^2 (6\sigma_P^2) + 24\sigma_I^2)$$

$$= \frac{1}{24} (4\sigma_B^2 + 4\sigma_P^2 + \sigma_I^2)$$

Juf. (10.43) s. 130 gælder, at

$$\lambda_P = \sigma_I^2 + 4\sigma_P^2$$

$$\lambda_B = \sigma_I^2 + 4\sigma_P^2 + 12\sigma_B^2$$

$$\frac{1}{3}\lambda_B + \frac{2}{3}\lambda_P = \sigma_I^2 + 4\sigma_P^2 + 4\sigma_B^2$$

2)

$$\hat{\lambda}_1, \hat{\lambda}_2 \text{ uafh. } \hat{\lambda}_1 \sim \lambda_1 \chi^2(f_1)/f_1, \hat{\lambda}_2 \sim \lambda_2 \chi^2(f_2)/f_2$$

$$\hat{\lambda} := s\hat{\lambda}_1 + t\hat{\lambda}_2, \quad s, t > 0$$

$\hat{\lambda}$ er approx. $\lambda \chi^2(\bar{f})/\bar{f}$ fordelt, hvor

$\lambda \chi^2(\bar{f})/\bar{f}$ har samme middelværdi og varians som $\hat{\lambda}$

Jvf. kap 3.4.2 i Bind 1 er

middelværdien af en $\sigma^2 X(f)/f$ fordeling

$$\frac{\sigma^2}{f} f = \sigma^2 \quad \text{og} \quad \text{variansen} \quad \left(\frac{\sigma^2}{f}\right)^2 2f = \frac{2\sigma^4}{f}$$

Dvs. pga. uafh.

$$E[\hat{\lambda}] = s\lambda_1 + t\lambda_2 \quad \text{Var}(\hat{\lambda}) = 2 \left(s^2 \frac{\lambda_1^2}{f_1} + t^2 \frac{\lambda_2^2}{f_2} \right)$$

$$\text{Dvs.} \quad \bar{f} = \frac{2 E[\hat{\lambda}]^2}{\text{Var}(\hat{\lambda})} = \frac{2 (s\lambda_1 + t\lambda_2)^2}{2 \left(s^2 \frac{\lambda_1^2}{f_1} + t^2 \frac{\lambda_2^2}{f_2} \right)} = \frac{(s\lambda_1 + t\lambda_2)^2}{\frac{(s\lambda_1)^2}{f_1} + \frac{(t\lambda_2)^2}{f_2}}$$

3)

Frihedsgradene for den op. fnd. til $\frac{1}{3}\hat{\lambda}_B + \frac{2}{3}\hat{\lambda}_P$ er

$$\left(\frac{1}{3} \hat{\lambda}_B + \frac{2}{3} \hat{\lambda}_P \right)^2 / \left(\left(\frac{1}{3} \hat{\lambda}_B \right)^2 / 5 + \left(\frac{2}{3} \hat{\lambda}_P \right)^2 / 10 \right) = 8,86898 \approx 9$$

Antag $\bar{X} \sim N(\alpha_i + \beta_i, \frac{1}{24} (4\sigma_B^2 + 4\sigma_P^2 + \sigma_I^2))$ uafh. af $\bar{X}_{i..}$

$$(\bar{X}_{i..} - \bar{X}_{i..}) / \sqrt{\frac{1}{24} \left(\frac{1}{3} \hat{\lambda}_B + \frac{2}{3} \hat{\lambda}_P \right)} \sim t(\bar{f}) \text{ jvf.}$$

afsnit 3.4.3 i bind 1

Dvs. et 95% konf. int. for $\alpha_i + \beta_i$ er

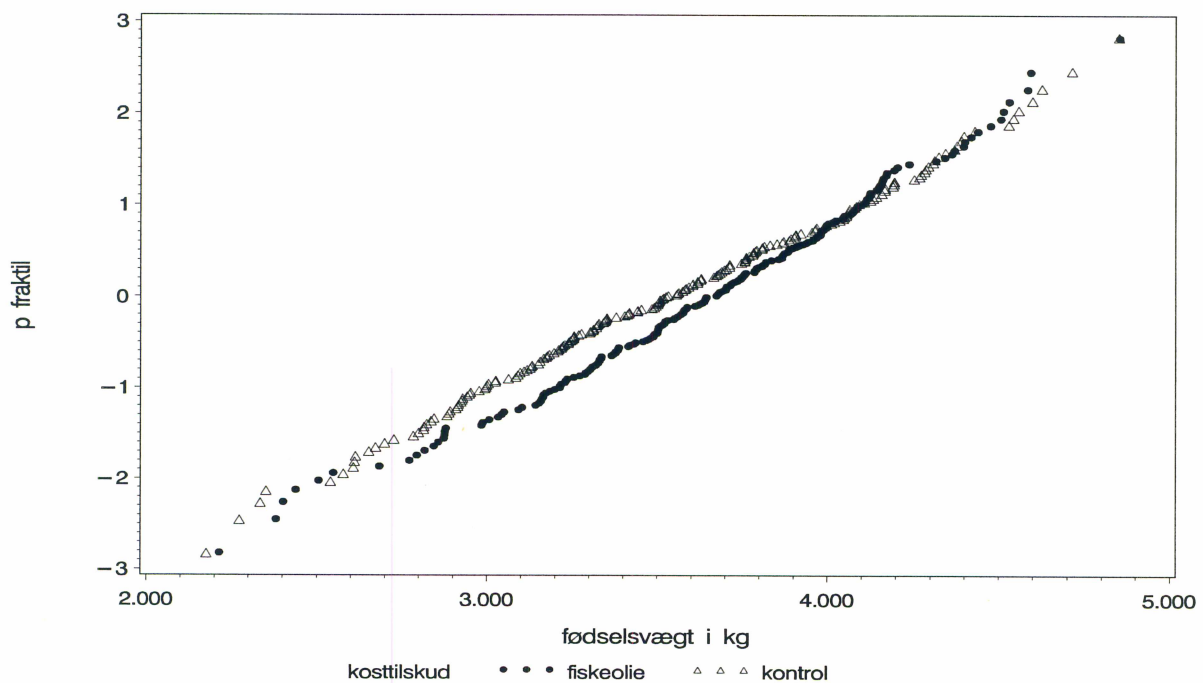
$$\left[-t_{0,975}(\bar{f}) \sqrt{\frac{1}{24} \left(\frac{1}{3} \hat{\lambda}_B + \frac{2}{3} \hat{\lambda}_P \right)} + \bar{X}_{i..}, t_{0,975}(\bar{f}) \sqrt{\frac{1}{24} \left(\frac{1}{3} \hat{\lambda}_B + \frac{2}{3} \hat{\lambda}_P \right)} + \bar{X}_{i..} \right]$$

Opgave 1

Man har iagttaget lange svangerskaber og høje fødselsvægte på Færøerne, og det har ført til en antagelse om at højt indtag af omega 3 fedtsyrer, som man får ved at spise fed fisk, kan forlænge svangerskabet. I et kontrolleret forsøg er raske kvinder i svangerskabets uge 30 blevet randomiseret til ét af to kosttilskud gennem resten af svangerskabet. Kosttilskuddet var et dagligt indtag af enten fiskeolie eller olivenolie. Gruppen af kvinder, som fik fiskeolie, vil vi kalde *fiskeolie* og den anden gruppe for *kontrol*. Problemstillingen i opgaven er ægte, men data er konstruerede.

210 kvinder kom i fiskeoliegruppen og 215 kvinder kom i kontrolgruppen.

I Figur 1 er vist et fraktildiagram for normalfordeling for fødselsvægten for de to grupper og i Tabel 1 er gengivet nogle beregninger, som kan benyttes i besvarelsen.



Figur 1: Fraktildiagrammer for normalfordeling for fødselsvægten i de to grupper.

Beregningsskema:								
i	ni	Si	USSi	Si ² /ni	SSDi	fi	Var	Mean
fiskeolie	210	764.151	2825.348699	2780.603575	44.745124	209	0.214092	3.6388
kontrol	215	760.128	2745.829530	2687.416634	58.412896	214	0.272957	3.5355
	425	1524.279	5571.178229	5468.020209	103.158020	423	0.243872	

Tabel 1: Standardberegninger på fødselsvægten i de to grupper.

1° Giver data anledning til at tro, at de to kosttilskud har forskellig virkning på fødselsvægten?

I Tabel 2 er svangerskabets længde opgjort i uger i de to grupper.

Table of gruppe by uge

gruppe	uge					
Frequency	37	38	39	40	41	Total
fiskeolie	6	11	30	104	59	210
kontrol	5	24	41	111	34	215
Total	11	35	71	215	93	425

Tabel 2: De observerede graviditetslængder i de to grupper opgjort i uger.

2° Sammenlign svangerskabets længde i de to grupper.

I det følgende vil vi nærmere beskrive sammenhængen mellem fødselsvægten og svangerskabets længde, og til det formål er der en lang række tegninger og tabeller, som må benyttes i besvarelsen. I Figur 2 og 3 er for henholdsvis kontrolgruppen og fiskeoliegruppen vist fraktildiagrammer for normalfordeling for fødselsvægten opdelt efter svangerskabets længde, og i Tabel 3 er vist standardberegninger for kontrolgruppen. Endelig er der i Figur 4 vist regressionen af fødselsvægten på graviditetslængden for de to grupper.

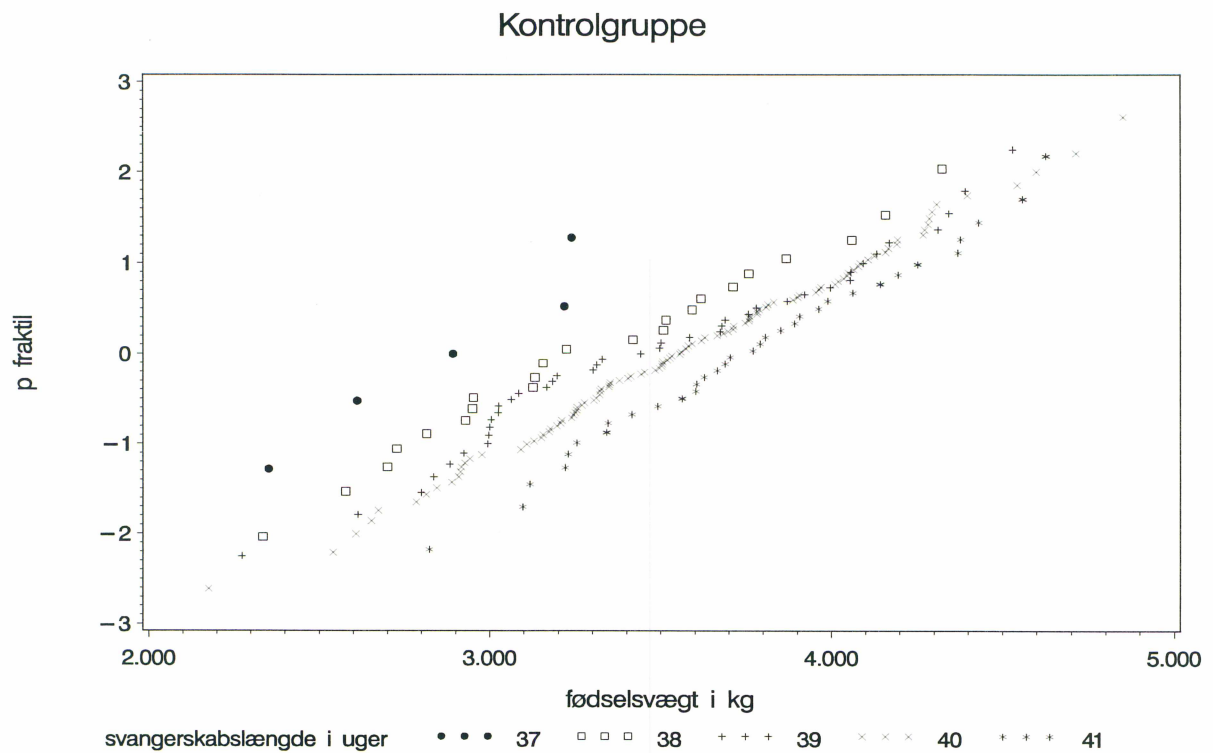
Beregningsskema: Kontrolgruppen

i	ni	Si	USSi	Si2/ni	SSDi	fi	Var	Mean
37	5	14.312	41.559534	40.966669	0.592865	4	0.148216	2.862
38	24	79.312	268.316548	262.099723	6.216825	23	0.270297	3.305
39	41	142.487	507.175651	495.184029	11.991622	40	0.299791	3.475
40	111	396.233	1441.898533	1414.419732	27.478801	110	0.249807	3.570
41	34	127.784	486.879264	480.257372	6.621892	33	0.200663	3.758
	215	760.128	2745.829530	2692.927525	52.902005	210	0.251914	

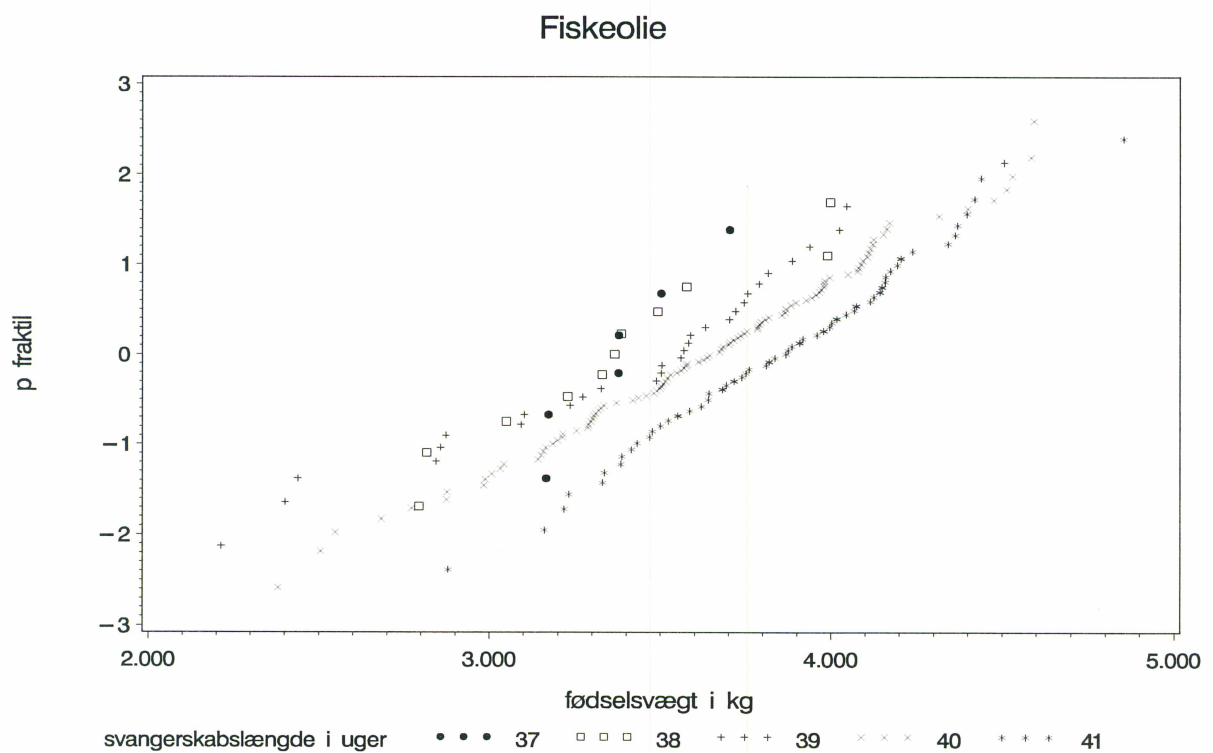
Tabel 3: Standardberegninger på fødselsvægten i kontrolgruppen opdelt efter svangerskabets længde.

3° Vis, at det kan antages, at der for kontrolgruppen er samme varians på fødselsvægten i de fem observationsrækker.

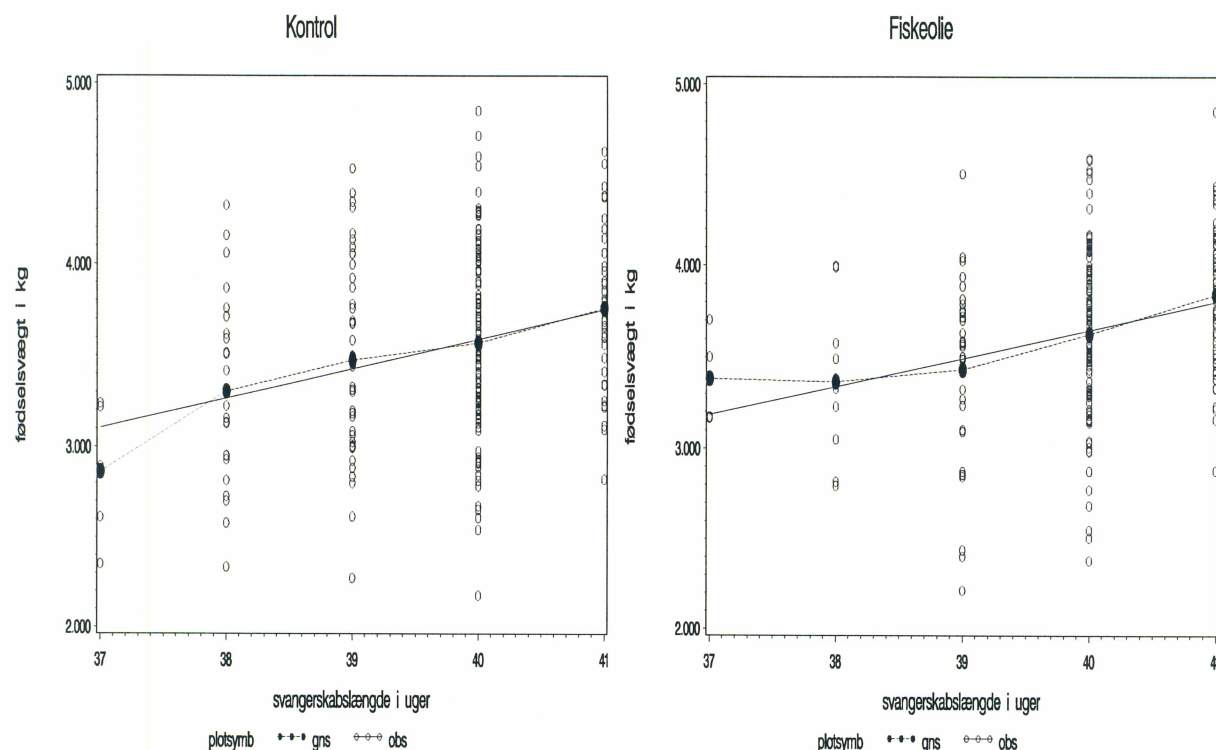
Det kan tilsvarende vises, at det også for fiskeoliegruppen kan antages at der er varianshomo-genitet mellem de fem observationsrækker defineret ved samme længde af svangerskabet.



Figur 2: Fraktildiagrammer for normalfordeling for fødselsvægt i kontrolgruppen opdelt efter svangerskabets længde.



Figur 3: Fraktildiagrammer for normalfordeling for fødselsvægt i fiskeoliegruppen opdelt efter svangerskabets længde.



Figur 4: Fødselsvægt tegnet op mod svangerskabets længde med både estimeret regressionslinje og gennemsnit i hver gruppe indtegnet.

Til brug for besvarelsen af de næste spørgsmål er der en række SAS udskrifter på de følgende sider. Beregningerne er baseret på datasættet `vaegt`, som indeholder de variable `fv`, som indeholder fødselsvægten, `uge`, som angiver svangerskabets længde og `gruppe`, som angiver kosttilskuddet. Inden hver udskrift er gengivet den programstump, der giver anledning til udskriften. Enkelte af udskrifterne er blevet let redigeret af pladshensyn.

- 4° Vis, at det for begge grupper kan antages, at der er lineær regression af fødselsvægt på svangerskabets længde.
- 5° Vis, at det kan antages at regressionslinierne for kontrolgruppen og fiskeoliegruppen har samme hældning.
- 6° Undersøg, om det kan antages at regressionslinierne for kontrolgruppen og fiskeoliegruppen er sammenfaldende. Angiv estimer og konfidensintervaller for alle parametre i slutmodellen.
- 7° Giv en sammenfatning af analysen i de foregående spørgsmål og en samlet konklusion.

```
PROC GLM DATA=vaegt;
CLASS uge;
BY gruppe;
MODEL fv=uge/SS1;
RUN;
```

----- kosttilskud=fiskeolie ----- 1

The GLM Procedure

Class Level Information

Class	Levels	Values
uge	5	37 38 39 40 41

Number of observations 210

----- kosttilskud=fiskeolie ----- 2

The GLM Procedure

Dependent Variable: fv foedselsvaegt i kg

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	4.91554024	1.22888506	6.32	<.0001
Error	205	39.82958352	0.19429065		
Corrected Total	209	44.74512376			

----- kosttilskud=kontrol ----- 3

The GLM Procedure

Class Level Information

Class	Levels	Values
uge	5	37 38 39 40 41

Number of observations 215

----- kosttilskud=kontrol ----- 4

The GLM Procedure

Dependent Variable: fv foedselsvaegt i kg

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	5.51089020	1.37772255	5.47	0.0003
Error	210	52.90200545	0.25191431		
Corrected Total	214	58.41289566			


```
PROC GLM DATA=vaegt;
BY gruppe;
MODEL fv=uge/SS1;
RUN;
```

----- kosttilskud=fiskeolie ----- 5

The GLM Procedure

Number of observations 210

----- kosttilskud=fiskeolie ----- 6

The GLM Procedure

Dependent Variable: fv foedselsvaegt i kg

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	4.42482847	4.42482847	22.83	<.0001
Error	208	40.32029529	0.19384757		
Corrected Total	209	44.74512376			

Parameter	Estimate	Standard Error	t Value	Pr > t
Intercept	-2.515624128	1.28851998	-1.95	0.0522
uge	0.154062709	0.03224627	4.78	<.0001

----- kosttilskud=kontrol ----- 7

The GLM Procedure

Number of observations 215

----- kosttilskud=kontrol ----- 8

The GLM Procedure

Dependent Variable: fv foedselsvaegt i kg

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	5.04490098	5.04490098	20.13	<.0001
Error	213	53.36799468	0.25055397		
Corrected Total	214	58.41289566			

Parameter	Estimate	Standard Error	t Value	Pr > t
Intercept	-2.875478142	1.42912801	-2.01	0.0455
uge	0.161589191	0.03601112	4.49	<.0001

```
PROC GLM DATA=vaegt;
CLASS gruppe;
MODEL fv= uge gruppe uge*gruppe/SS1 SOLUTION;
RUN;
```

5

```

The GLM Procedure

Class Level Information

Class          Levels      Values
gruppe          2      fiskeolie kontrol
```

Number of observations 425

6

```

The GLM Procedure

Dependent Variable: fv      foedselsvaegt i kg

Source          DF          Sum of Squares      Mean Square      F Value      Pr > F
Model              3          10.6041276          3.5347092          15.88      <.0001
Error            421          93.6882900          0.2225375
Corrected Total    424          104.2924176
```

```

R-Square      Coeff Var      Root MSE      fv Mean
0.101677      13.15304      0.471739      3.586539
```

```

Source          DF      Type I SS      Mean Square      F Value      Pr > F
uge              1      10.22164492      10.22164492      45.93      <.0001
gruppe          1       0.37710808       0.37710808       1.69      0.1937
uge*gruppe       1       0.00537464       0.00537464       0.02      0.8766
```

```

Parameter          Estimate      Standard Error      t Value      Pr > |t|
Intercept          -2.875478142 B      1.34685890      -2.13      0.0333
uge                 0.161589191 B      0.03393811       4.76      <.0001
gruppe fiskeolie    0.359854013 B      1.92874022       0.19      0.8521
gruppe kontrol      0.000000000 B      .               .         .
uge*gruppe fiskeolie -0.007526482 B      0.04843050      -0.16      0.8766
uge*gruppe kontrol  0.000000000 B      .               .         .
```

NOTE: The X'X matrix has been found to be singular, and a generalized inverse was used to solve the normal equations. Terms whose estimates are followed by the letter 'B' are not uniquely estimable.

```
PROC GLM DATA=vaegt;
CLASS gruppe;
MODEL fv= uge gruppe/SS1 SOLUTION;
RUN;
```

7

```

The GLM Procedure

Class Level Information

Class          Levels    Values
gruppe          2      fiskeolie kontrol
```

```
Number of observations      425
```

8

```

The GLM Procedure

Dependent Variable: fv      foedselsvaegt i kg

Source          DF          Sum of Squares    Mean Square    F Value    Pr > F
Model              2      10.5987530      5.2993765      23.87    <.0001
Error            422      93.6936646      0.2220229
Corrected Total   424      104.2924176

R-Square      Coeff Var      Root MSE      fv Mean
0.101625      13.13782      0.471193      3.586539

Source          DF      Type I SS      Mean Square    F Value    Pr > F
uge              1      10.22164492      10.22164492      46.04    <.0001
gruppe          1      0.37710808      0.37710808      1.70    0.1932

Parameter          Estimate          Standard Error    t Value    Pr > |t|
Intercept          -2.728842382 B      0.95999920      -2.84      0.0047
uge                 0.157893214      0.02418337      6.53      <.0001
gruppe fiskeolie    0.060198720 B      0.04619056      1.30      0.1932
gruppe kontrol      0.000000000 B      .              .              .
```

NOTE: The X'X matrix has been found to be singular, and a generalized inverse was used to solve the normal equations. Terms whose estimates are followed by the letter 'B' are not uniquely estimable.

```
PROC GLM DATA=vaegt;
CLASS gruppe;
MODEL fv= uge/SS1 SOLUTION;
RUN;
```

9

```

The GLM Procedure

Class Level Information

Class          Levels    Values
gruppe                2    fiskeolie kontrol

```

```

Number of observations    425

```

10

```

The GLM Procedure

Dependent Variable: fv    foedselsvaegt i kg

Source          DF          Sum of Squares    Mean Square    F Value    Pr > F
Model              1      10.2216449      10.2216449      45.96    <.0001
Error            423      94.0707727        0.2223895
Corrected Total   424     104.2924176

R-Square      Coeff Var      Root MSE      fv Mean
0.098009      13.14866      0.471582      3.586539

Source          DF      Type I SS      Mean Square    F Value    Pr > F
uge              1     10.22164492      10.22164492      45.96    <.0001

Parameter      Estimate      Standard Error    t Value    Pr > |t|
Intercept      -2.878562915      0.95388727      -3.02      0.0027
uge             0.162401338      0.02395446       6.78      <.0001

```


S2001 opgave 1

1)

Vi kan (jvf. fæltildingsdiagrammet i Fig. 1) antage, at
statmodellen er $M_0: X_{ij} \sim N(\mu_i, \sigma_i^2) \quad i = \text{Fisk, Kontrol} \quad j = 1, \dots, n_i$

$$H_{0\sigma}: \sigma_{\text{Fisk}}^2 = \sigma_{\text{Kontrol}}^2$$

$$s_{\text{Fisk}}^2 = \frac{1}{n_F - 1} (U_{SS_F} - \frac{S_F^2}{n_F}) = 0,214092$$

(3,23)
5,72

$$s_K^2 = \frac{1}{n_K - 1} (U_{SS_K} - \frac{S_K^2}{n_K}) = 0,272957 \quad \text{jvf. s. 106 B1 f\aa}s:$$

$$F = \frac{s_K^2}{s_F^2} = 1,27496$$

$$E(x) = 2(1 - F_{F(214, 209)}(F)) > 2(1 - 0,975) = 0,05$$

Dvs. vi kan reducere til

$$M_1: X_{ij} \sim N(\mu_i, \sigma^2)$$

$$H_{0\mu}: \mu_{\text{Fisk}} = \mu_{\text{Kontrol}}$$

$$s_1^2 = \frac{SSD_{(1)} + SSD_{(K)}}{f_F + f_K} = 0,243872 \quad \text{jvf. s. 106 B1 f\aa}s$$

(3,28)
5,91

$$t(x) = \frac{\bar{X}_F - \bar{X}_K}{\sqrt{s_1^2 (\frac{1}{n_F} + \frac{1}{n_K})}} = \frac{3,6388 - 3,5355}{\sqrt{0,243872 (\frac{1}{210} + \frac{1}{215})}} = 2,15602$$

$$E(x) = 2(1 - F_{t(423)}(|t(x)|)) < 2(1 - 0,975) = 0,05$$

Dvs. vi forkæster hypotesen og kan konkludere, at
fødselsvægten for fiskediegruppen er højere end i
kontrolgruppen

2)

Jvf. a) - d) er svangshabets længde i ugr multinomialfordelt

At svangshabetslængden er uafh. hvorvidt man tilhører fiskerie el. kontrolgruppen svarer til at teste for homogenitet af de to multinomialford.:

Jvf. (6.47) fis. de forventede antal:

e	37	38	39	40	41
F	5,43	17,29	35,08	106,29	45,95
K	5,56	17,71	35,92	108,76	47,05

$$-2 \ln Q = \underset{\substack{(6.49) \\ s. 352}}{13,72129}$$

$$E(\underline{x}) = 1 - F_{\chi^2(1,4)}(-2 \ln Q(\underline{x})) \leq 1 - 0,99 = 0,01 < 0,05$$

Dvs. svangshabetslængden afhænger af hvorvidt man tager fiskerie el. ej.

3)

$$H_0: \sigma_{\text{B1K}}^2 = \dots = \sigma_{\text{G1K}}^2$$

Jvf. s. 128 i B1 fis.:

$$-2 \ln Q(\underline{x}) = f_1 \ln s_1^2 - \sum_i f_{1i} \ln(s_{1i}^2) = 1,97190$$

$$C = 1 + \frac{1}{3(5-1)} \left(\sum \frac{1}{f_{1i}} - \frac{1}{f_1} \right) = 1,02943$$

$$B_a = -2 \ln Q(\underline{x}) / C = 1,91553$$

$$E(X) = 1 - F_{\chi^2(4)}(B_1) \geq 1 - 0,3 = 0,7 > 0,05$$

Dus, vi forkaster ikke hypotesen om ens varians

4)

Testet for lin. regression er (jvf. s. 211 i B1)

Fisherlic:

$$F = \frac{((40,32029529 - 39,82958352) / (202 - 205))}{0,19429065} = 0,84189$$

$$E(X) = 1 - F_{F(3,205)}(F) \geq 1 - 0,7 = 0,3 > 0,05$$

Kontrol:

$$F = \frac{((53,36799468 - 52,90200545) / (213 - 210))}{0,25191431} = 0,4660$$

$$E(X) = 1 - F_{F(3,210)}(F) \geq 1 - 0,5 = 0,5 > 0,05$$

Dus, vi kan antage at der er lin. reg. i begge grupper.

5)

For at kunne behandle hypotesen om fælles hældning som en lin. normal model, må vi først teste om der er ens varians i de to grupper

Ud. s. 106 fås

$$F = \frac{0,25055397}{0,19384757} = 1,29253$$

$$E(X) = 2(1 - F_{F(213,208)}(F)) > 2(1 - 0,975) = 0,05$$

Dus, der er ens varians i de to reg. modeller

Jvf. s. 211 i B1 er testet for samme hældning

$$F = \frac{((93,6936646 - 93,6882900) / (422 - 421))}{0,2225375} = 0,024151$$

$$E(x) = 1 - F_{F(1,421)}(F) = F_{F(421,1)}(1/F) \geq 0,7 > 0,05$$

Dvs. vi kan antage at reg. lin. har samme hældning. (Jvf SAS $\rightarrow E(x) = 0,8766$)

6)

Testet for sammenfaldende reg. lin. $\leftarrow$ (jvf. s. 211 i B1)

$$F = \frac{((94,0707727 - 93,6936646) / (423 - 422))}{0,2220229} = 1,69851$$

$$E(x) = 1 - F_{F(1,422)}(F) \geq 1 - 0,9 = 0,1 > 0,05$$

Dvs. vi kan reducere til modellen om sammenfaldende regressionslinier. (Jvf SAS $\rightarrow E(x) = 0,1932$)

Estimation i slutmodellen $\leftarrow$ (jvf SAS)

$$\text{varians} = \hat{\sigma}^2 = 0,2223895$$

$$\text{afskæring} = \hat{\alpha} = -2,878562915$$

$$\text{hældning} = \hat{\beta} = 0,162401338$$

95% Konf. int. (jvf SAS + bn. s. 146-147 + s. 160n) :

$$\sigma^2 : \left[\frac{0,2223895}{\chi^2_{0,95}(423)/423} ; \frac{0,2223895}{\chi^2_{0,025}(423)/423} \right] = [0,19522 ; 0,25569]$$

$$\chi^2_{0,95}(423)/423 = 1,13919, \chi^2_{0,025}(423)/423 = 0,86975$$

(uhjælp interpolation i tabellene)

$$\alpha: [-2,878562915 \pm 0,95388727 \cdot 1,966] = [-4,754; -1,003]$$

$$\beta: [0,162401338 \pm 0,02395446 \cdot 1,966] = [0,115; 0,209]$$

$$t_{0,975}(423) = 1,966$$

(Jvf. SAS udskriften kan modellen ikke undersøges ydeligere)

7)

Phd. 1) (2) peger på forskelle i de to grupper
men analyse v.h.j. a lin. normale modeller 5) + 6)

peger på at der ikke er signifikante forskelle.

Jvf. testet for ens varians i 5) kan vi sige at
med flere forsøgspersoner (frihedsgrad) ville hypotesen ($\sigma_F^2 = \sigma_K$)
måske blive forkastet. En varianskomponentmodel
ville måske forklare forsøget bedre (at det
netop er forskelle i varians som kan være centrale
funktion af forholdsdiagrammet i fig. 1).

Sammenfattende kan vi sige, at vi ikke
kan afvise (på grundlag af det givne materiale)
at der ingen forskel er på de to grupper
- et mere omfattende studie (med en
bedre forsøgstilrettelæggelse?) kan måske påvise
forskelle i variationen.

Opgave 2

I et studie af børns vækst indgik 27 børn, 11 piger og 16 drenge, som 8, 10, 12 og 14 år gamle fik målt afstanden i mm fra centret af *glandula pituitaria* (hypofysen) til den *pterygomaxillære fissur* (en revne lige bag næsehulen). Data er gengivet i nedenstående tabel.

Person	Køn	Alder			
		8 år	10 år	12 år	14 år
1	pige	21.0	20.0	21.5	23.0
2	pige	21.0	21.5	24.0	25.5
3	pige	20.5	24.0	24.5	26.0
4	pige	23.5	24.5	25.0	26.5
5	pige	21.5	23.0	22.5	23.5
6	pige	20.0	21.0	21.0	22.5
7	pige	21.5	22.5	23.0	25.0
8	pige	23.0	23.0	23.5	24.0
9	pige	20.0	21.0	22.0	21.5
10	pige	16.5	19.0	19.0	19.5
11	pige	24.5	25.0	28.0	28.0
12	dreng	26.0	25.0	29.0	31.0
13	dreng	21.5	22.5	23.0	26.5
14	dreng	23.0	22.5	24.0	27.5
15	dreng	25.5	27.5	26.5	27.0
16	dreng	20.0	23.5	22.5	26.0
17	dreng	24.5	25.5	27.0	28.5
18	dreng	22.0	22.0	24.5	26.5
19	dreng	24.0	21.5	24.5	25.5
20	dreng	23.0	20.5	31.0	26.0
21	dreng	27.5	28.0	31.0	31.5
22	dreng	23.0	23.0	23.5	25.0
23	dreng	21.5	23.5	24.0	28.0
24	dreng	17.0	24.5	26.0	29.5
25	dreng	22.5	25.5	25.5	26.0
26	dreng	23.0	24.5	26.0	30.0
27	dreng	22.0	21.5	23.5	25.0

Nogle relevante summer er gengivet i Tabel 4.

Betragt designet $\mathcal{D}$, der består af faktorerne,

$$\mathcal{D} = \{I, \text{Person}, \text{Alder} \times \text{Køn}, \text{Alder}, \text{Køn}, O\}.$$

<i>Køn</i>	<i>Alder</i>				<i>I alt</i>
	<i>8 år</i>	<i>10 år</i>	<i>12 år</i>	<i>14 år</i>	
<i>pige</i>	233.0	244.5	254.0	265.0	996.5
<i>dreng</i>	366.0	381.0	411.5	439.5	1598.0
<i>I alt</i>	599.0	625.5	665.5	704.5	2594.5

Tabel 4: Summer.

<i>Faktor</i>	<i>Kort navn</i>	<i>Frihedsgrader</i>	<i>SSD</i>
<i>Køn</i>	<i>K</i>	1	140.465
<i>Person</i>	<i>P</i>	25	377.915
<i>Alder</i>	<i>A</i>	3	237.192
<i>Alder</i> $\times$ <i>Køn</i>	<i>A</i> $\times$ <i>K</i>	3	13.993
<i>I</i>	<i>I</i>	75	148.128
<i>I alt</i>		107	917.693

Tabel 5: Variansanalysetabel.

1° Vis, at $\mathcal{D}$ er et ortogonalt design.

2° Tegn et faktor struktur diagram og angiv frihedsgrader for kvadratsummerne.

I Tabel 5 er angivet variansanalysetabellen for designet $\mathcal{D}$.

Betragt modellen, hvor mængden af tilfældige faktorer er $\mathcal{B} = \{I, Person\}$ og mængden af systematiske faktorer er $\mathcal{T} = \{Alder \times Køn, Alder, Køn, O\}$.

3° Undersøg, om det kan antages at varianskomponenten σ_{Person}^2 er 0 og angiv estimer for varianskomponenterne i slutmodellen.

4° Vis, at modellen for middelværdien kan reduceres til $\mathcal{T}_1 = \{Alder, Køn, O\}$ og undersøg om modellen kan reduceres yderligere.

5° Gør rede for at kontrasten for *Køn* er estimerbar i modellen $\mathcal{T}_1 + \mathcal{B}$ og angiv et 95% konfidensinterval for kontrasten for *Køn*.

S2001 opgave 2

1)

Jvf. afsnit 10.7.2 s. 106 i B3 er

$A \times K$, A , K parvis ortogonale (jeg bruger selvfølgelig forkortede forkortelser samt at $P \perp A$ (jvf. tabellen)

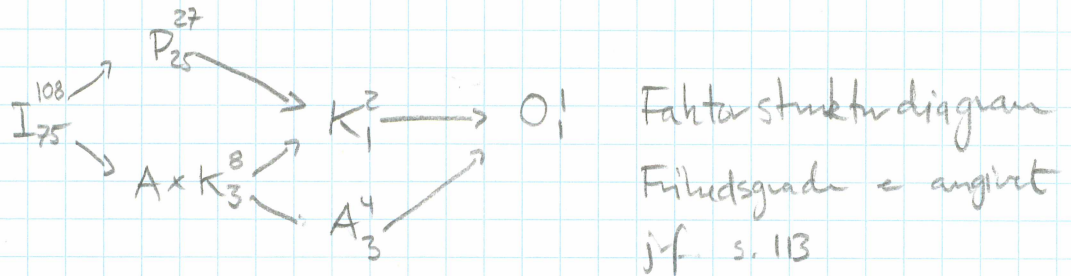
Af sætn. 10.9 følger, at $P \perp K$ og $P \perp A \times K$

(jvf. tabellen) (Det er ligeledes jvf. sætn. 10.9

klart at hh. I , O er ortogonale m. de andre faktorer)

Da D klart er lukket under minimumsdannelse og $I \in D$ er D jvf. def. 10.10 s. 108 et ortogonalt design.

2)



3)

$$H_0: \sigma_P^2 = 0$$

$$F = \frac{\hat{\lambda}_P}{\hat{\lambda}_I} = \frac{SSD_P / d_P}{SSD_I / d_I} = \frac{377,915 / 25}{148,128 / 75} = 7,65382$$

$$1 - F_F(25, 75) (F) \leq 1 - 0,9995 = 0,0005$$

Dus faktoren P er signifikant ; dvs. $\sigma_P^2 > 0$

Jvf. (10.43) s. 130 ~

$$\hat{\sigma}_I^2 = \hat{\lambda}_I, \quad 4\hat{\sigma}_P^2 + \hat{\sigma}_I^2 = \hat{\lambda}_P$$

og jvf. (10.46) s. 133 ~ $\hat{\lambda}_I = \text{SSD}_I / d_I$, $\hat{\lambda}_P = \text{SSD}_P / d_P$

Dvs. $\hat{\sigma}_I^2 = \text{SSD}_I / d_I = 1,975$

$$\hat{\sigma}_P^2 = (\hat{\lambda}_P - \hat{\lambda}_I) / 4 = (\text{SSD}_P / d_P - \text{SSD}_I / d_I) / 4 = 3,285$$

4)

Testet for reduktion til $T_1 + [B]$ er jvf. s. 131-135 ~

$$F = \frac{\text{SSD}_{A \times K} / d_{A \times K}}{\text{SSD}_I / d_I} = 2,36164$$

$$E(x) = 1 - F_{F(3,75)}(F) > 1 - 0,95 = 0,05$$

Dvs. vi kan reducere til $T_1 + [B]$

ANOVA-tabellen for $T_1 + [B]$ er som

for blot er $d_I' = d_I + d_{A \times K}$, $\text{SSD}_I' = \text{SSD}_I + \text{SSD}_{A \times K}$
jvf. (10.26) + (10.24) s. 113.

Det er klart, blot ved at se på ANOVA-tabellen at vi ikke kan reducere modellen ydeligere.

Jvf. s. 134-135 ~

$$F_K = \frac{\text{SSD}_K / d_K}{\text{SSD}_P / d_P} = 9,292, \quad F_A = \frac{\text{SSD}_A / d_A}{\text{SSD}_I' / d_I'} = 57,059$$

$$E_K(x), E_A(x) < 0,05$$

Altså kan modellen ikke reduceres ydeligere.

5)

At $K=0$ og jvf. sætn. 10.17 er kontrasten
for K dermed estimbar

$$\text{Da } t_{0,975}(25) = 2,060, \quad \hat{\lambda}_D = \frac{SSD_D/d_D}{s. 133} = 15,117$$

$$f_D = 25, \quad n_{\text{pige}} = 44, \quad n_{\text{dring}} = 64$$

$$\hat{\alpha}_{\text{dring}}^K = 1598 / (16 \cdot 4) = 24,969$$

$$\hat{\alpha}_{\text{pige}}^K = 996,5 / (11 \cdot 4) = 22,648$$

er et 95% konf. int. for kontrast $\alpha_{\text{dring}}^K - \alpha_{\text{pige}}^K$

$$[-0,752; 3,890], \quad \text{jvf. (10.51) s. 138}$$

Meddelelse 28

Forelæsningerne i sidste uge

(Onsdag den 23. april): Afsluttede gennemgangen af teorien om variansanalyse.

(Fredag den 25. april): Beregninger i SAS i tilknytning til variansanalyse. Næste ikke afsnittet *Tredje del. Modelsætningen og ANOVA tabellen.*

Forelæsningerne i denne uge

(Onsdag den 30. april): Afsluttede omtalen af beregninger i SAS i tilknytning til variansanalyse og begyndte på Kapitel 12 om den negative binomialfordeling.

(Fredag den 2. maj): Fortsætter gennemgangen af Kapitel 12 om den negative binomialfordeling.

Forelæsningerne i næste uge

Onsdag den 7. maj og fredag den 9. maj.

Øvelserne i de næste uger

(Uge 19: 5.-11. maj): 10.17, samt opgave 1 og 2 i eksamenssættet (i Statistik 1) fra sommeren 2001.

(Uge 20: 12.-18. maj): (Næstsidste øvelsesgang) Opgave 3 i eksamenssættet (i Statistik 1) fra sommeren 2001, opgave 3 (Statistik 1) fra sommeren 1999, opgave 4 (Statistik 1) fra sommeren 2002.

(Uge 21: 19.-25. maj): (Sidste øvelsesgang) Opgaver stilles senere.

De nævnte eksamensopgaver ligger på nettet, men helt i tråd med det høje serviceniveau på Stat1 er de også lagt frem på Stat1 hylden.

Temperatur i grader Celsius										
Materiale type	-5			20			45			Sum
1	130	74		64	20		70	82		1238
	155	180	(709)	105	110	(299)	20	58	(230)	
	170									
2	155	193		136	122		25	70		1517
	164	131	(840)	106	115	(479)	58	45	(198)	
	197									
3	148	120		174	120		96	104		1688
	178	160	(763)	150	139	(583)	82	60	(342)	
	157									
Sum	2312			1361			770			4443

Tabel 1: Levetider for 39 batterier. Summen af levetiderne for samme kombination af materiale type og temperatur er angivet i parentes.

Opgave 2

Under udviklingen af et batteri, som skulle bruges under meget varierende temperaturer, interesserede man sig for, om materialevalget til en bestemt komponent af batteriet ville påvirke levetiden. Man havde tre typer af materiale at vælge imellem, og 13 batterier med hvert materialevalg blev afprøvet ved temperaturerne -5°C , 20°C og 45°C . Batteriernes levetid i timer blev registreret. Data fra forsøget er gengivet i Tabel 1. For hver kombination af materialevalg og temperatur er summen af levetiderne gengivet i Tabel 1 i parentes.

Faktoren defineret ved de tre niveauer af materiale type betegnes med M og faktoren defineret ved de tre niveauer af temperatur betegnes med T .

- 1° Undersøg, om designet $\mathcal{D} = \{I, M, T, T * M, 0\}$ er et ortogonalt design, tegn faktor struktur diagrammet og angiv frihedsgraderne.

Vi vender os nu mod analysen af forsøget, hvor vi vil benytte en lineær normal model med $\mathcal{T} = \{T * M\}$ og $\mathcal{B} = \{I\}$.

- 2° Undersøg om det er rimeligt at antage, at variansen på levetiden ikke afhænger af materiale type eller temperatur. Beregningerne i Tabel 2 må gerne benyttes.

Beregningsskema:								
i	ni	Si	USSi	Si2/ni	SSDi	fi	Var	Mean
1	5	709.0000	107701.000	100536.200	7164.80	4	1791.2	141.80
2	5	840.0000	144140.000	141120.000	3020.00	4	755.0	168.00
3	5	763.0000	118237.000	116433.800	1803.20	4	450.8	152.60
4	4	299.0000	27621.000	22350.250	5270.75	3	1756.92	74.75
5	4	479.0000	57841.000	57360.250	480.75	3	160.25	119.75
6	4	583.0000	86497.000	84972.250	1524.75	3	508.25	145.75
7	4	230.0000	15388.000	13225.000	2163.00	3	721.0	57.50
8	4	198.0000	10914.000	9801.000	1113.00	3	371.0	49.50
9	4	342.0000	30356.000	29241.000	1115.00	3	371.67	85.50

39	4443.0000	598695.000	575039.750	23655.25	30	788.50	=====	

Tabel 2: Standardberegninger for data i Tabel 1.

I det følgende regner vi i den lineære normale model med $\mathcal{T} = \{T * M\}$ og $\mathcal{B} = \{I\}$. I besvarelsen af spørgsmålene må SAS udskrifterne benyttes. For at gøre identifikationen af den model, der hører til en udskrift så sikker som muligt, er det SAS program, der giver udskriften, gengivet inden udskriften.

- 3° Angiv *ANOVA* tabellen for designet, og undersøg, om der er vekselvirkning mellem temperatur og materiale type og om temperatur og om materiale type har indflydelse på batteriernes levetid.
- 4° Angiv 95% konfidensintervaller dels for forskellen i den forventede levetid ved temperaturen -5 og ved temperaturen 45 °C og dels for forskellen i den forventede levetid ved materiale type 1 og materiale type 3 i slutmodellen fra spørgsmål 3°.

Alternativt kunne man have hæftet sig ved, at temperaturen er en kontinuert variabel og forsøgt sig med lineær regression for at beskrive levetidens afhængighed af temperaturen.

- 5° Undersøg, om man kan antage, at levetiden afhænger lineært af temperaturen, og om det kan antages, at der er samme hældning for de tre materialetyper.

Én person har brugt variansanalyse og er nået til additivitetsmodellen som slutmodel, mens en anden person har brugt lineær regression og er nået til modellen med fælles hældning som slutmodel.

- 6° Beskriv relationen mellem de to personers slutmodeller, og test reduktionen fra den groveste til den fineste af de to modeller.


```
TITLE1 'Model M1';
PROC GLM DATA=batteri;
CLASS m t;
MODEL levetid=m t m*t/SS1 SOLUTION;
RUN;
```

giver følgende let reviderede udskrift:

Model M1						2
General Linear Models Procedure						
Class Level Information						
Class	Levels	Values				
M	3	1	2	3		
T	3	-5	20	45		
Model M1						3
General Linear Models Procedure						
Dependent Variable: LEVETID						
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F	
Model	8	68879.519	8609.940	10.92	0.0001	
Error	30	23655.250	788.508			
Corrected Total	38	92534.769				
Source	DF	Type I SS	Mean Square	F Value	Pr > F	
M	2	7938.000	3969.000	5.03	0.0130	
T	2	53964.453	26982.226	34.22	0.0001	
M*T	4	6977.067	1744.267	2.21	0.0915	
Parameter	Estimate	T for H0: Parameter=0	Pr > T	Std Error of Estimate		
INTERCEPT	85.50000000 B	6.09	0.0001	14.04019527		
M	1 -28.00000000 B	-1.41	0.1688	19.85583457		
	2 -36.00000000 B	-1.81	0.0798	19.85583457		
	3 0.00000000 B	.	.	.		
T	-5 67.10000000 B	3.56	0.0013	18.83689863		
	20 60.25000000 B	3.03	0.0049	19.85583457		
	45 0.00000000 B	.	.	.		
M*T	1 -5 17.20000000 B	0.65	0.5234	26.63939752		
	1 20 -43.00000000 B	-1.53	0.1362	28.08039055		
	1 45 0.00000000 B	.	.	.		
	2 -5 51.40000000 B	1.93	0.0632	26.63939752		
	2 20 10.00000000 B	0.36	0.7242	28.08039055		
	2 45 0.00000000 B	.	.	.		
	3 -5 0.00000000 B	.	.	.		
	3 20 0.00000000 B	.	.	.		
	3 45 0.00000000 B	.	.	.		

Opgaven fortsættes

```

TITLE1 'Model M2';
PROC GLM DATA=batteri;
CLASS m t;
MODEL levetid=m t/SS1 SOLUTION;
RUN;

```

giver følgende udskrift:

Model M2

4

General Linear Models Procedure
Class Level Information

Class	Levels	Values
M	3	1 2 3
T	3	-5 20 45

Number of observations in data set = 39

Model M2

5

General Linear Models Procedure

Dependent Variable: LEVETID

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	61902.453	15475.613	17.18	0.0001
Error	34	30632.317	900.950		
Corrected Total	38	92534.769			
R-Square		C.V.	Root MSE	LEVETID Mean	
0.668964		26.34746	30.016	113.92	

Source	DF	Type I SS	Mean Square	F Value	Pr > F
M	2	7938.000	3969.000	4.41	0.0199
T	2	53964.453	26982.226	29.95	0.0001

Parameter	Estimate	T for H0: Parameter=0	Pr > T	Std Error of Estimate
INTERCEPT	80.08974359 B	7.27	0.0001	11.01280149
M 1	-34.61538462 B	-2.94	0.0059	11.77318002
2	-13.15384615 B	-1.12	0.2717	11.77318002
3	0.00000000 B	.	.	.
T -5	89.96666667 B	7.74	0.0001	11.62508381
20	49.25000000 B	4.02	0.0003	12.25391427
45	0.00000000 B	.	.	.

```
TITLE1 'M3';
PROC GLM DATA=batteri;
CLASS m;
MODEL levetid=m t m*t/SS1 SOLUTION;
RUN;
```

giver følgende udskrift:

M3

6

General Linear Models Procedure
 Class Level Information

Class	Levels	Values
M	3	1 2 3

Number of observations in data set = 39

M3

7

General Linear Models Procedure

Dependent Variable: LEVETID

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	64868.761	12973.752	15.48	0.0001
Error	33	27666.009	838.364		
Corrected Total	38	92534.769			

R-Square	C.V.	Root MSE	LEVETID Mean
0.701020	25.41585	28.955	113.92

Source	DF	Type I SS	Mean Square	F Value	Pr > F
M	2	7938.000	3969.000	4.73	0.0156
T	1	53813.795	53813.795	64.19	0.0001
T*M	2	3116.966	1558.483	1.86	0.1718

Parameter	Estimate	T for H0: Parameter=0	Pr > T	Std Error of Estimate
INTERCEPT	153.4396552 B	14.40	0.0001	10.65892878
M 1	-27.1103448 B	-1.80	0.0813	15.07400164
M 2	5.8206897 B	0.39	0.7019	15.07400164
M 3	0.0000000 B	.	.	.
T	-1.3051724 B	-3.37	0.0019	0.38772067
T*M 1	-0.4151724 B	-0.76	0.4543	0.54831983
T*M 2	-1.0496552 B	-1.91	0.0643	0.54831983
T*M 3	0.0000000 B	.	.	.


```
TITLE1 'M4';
PROC GLM DATA=batteri;
CLASS m;
MODEL levetid=m t /SS1 SOLUTION;
RUN;
```

giver følgende udskrift:

M4 22

General Linear Models Procedure
 Class Level Information

Class	Levels	Values
M	3	1 2 3

Number of observations in data set = 39

M4 23

General Linear Models Procedure

Dependent Variable: LEVETID

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	61751.795	20583.932	23.40	0.0001
Error	35	30782.974	879.514		
Corrected Total	38	92534.769			

R-Square	C.V.	Root MSE	LEVETID Mean
0.667336	26.03212	29.657	113.92

Source	DF	Type I SS	Mean Square	F Value	Pr > F
M	2	7938.000	3969.000	4.51	0.0180
T	1	53813.795	53813.795	61.19	0.0001

Parameter	Estimate	T for H0: Parameter=0	Pr > T	Std Error of Estimate
INTERCEPT	162.2661804 B	17.62	0.0001	9.21048363
M 1	-34.6153846 B	-2.98	0.0053	11.63227306
2	-13.1538462 B	-1.13	0.2658	11.63227306
3	0.0000000 B	.	.	.
T	-1.7934483	-7.82	0.0001	0.22927849

STAT 1, TØ nr. 26

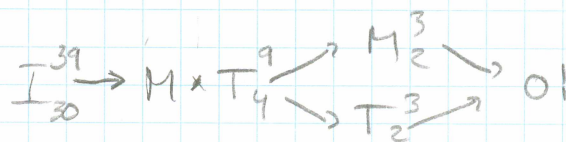
opg. 3 VOI/02 $\rightarrow$ se TØ nr. 25

opgave 2 S1998

1)

D er lukket under minimumsdaemelse, og $I \in D$

Af sætn. 10.9 følger at faktorerne er parvis ortogonale, jvf. def. 10.10 s. 108 er D altså et ortogonalt design.



2)

$$H: \sigma_{mt}^2 = \sigma^2 \quad m \in M, t \in T$$

$$-2 \ln Q(X) = \underbrace{f_1}_{s.113} \ln s_1^2 - \sum_{k=1}^9 \underbrace{f_{1k}}_{i \text{ bind 1}} \ln s_{1k}^2 = 7,60882$$

$$C = \underbrace{1}_{s.113} + \frac{1}{3(9-1)} \left(\sum_{k=1}^9 \frac{1}{f_{1k}} - \frac{1}{f_1} \right) = 1,11319$$

$$Ba = \underbrace{-\frac{2 \ln Q}{C}}_{s.113 \text{ bind 1}} = 6,83572$$

$$E(X) = 1 - F_{\chi^2(8)}(Ba) \geq 1 - 0,5 = 0,5 > 0,05$$

Der. at det er rimeligt at antage hypotesen H

3)

ANOVA:

Faktor	d	SSD
M	2	7938
T	2	53964,453
M×T	4	6977,067
I	30	23655,25

Jf. udskriften er testss for hypotese om ingen vekselvirkning $0,0915 > 0,05$ (SAS s.3)
 Dus. vi kan reducere til additivitetsmodellen
 Vi kan ikke reducere ydeligere (testss for hh. at fjerne M og T er jf SAS s.5 $0,0199$ og $0,0001$) Dus. der er ingen vekselvirkning, men hvilken M el. T kan testes væk.

4)

Konfidensintervallerne for de to forskelle er jf. bm. s.86 i bilag 1 ($\text{Mean} \pm \text{Stdev} \cdot t_{0,975}(f)$) hvor $f=34$ i dette tilfælde ($\alpha=0,05$)
 $[66,326^\circ; 113,607^\circ]$, $[-58,557; -10,673]$

($t_{0,975}(34) = 2,0336$ v. interpolation)

5)

At teste for lin. regression svarer til at teste hvorvidt vi kan gå fra modellen m. reduktion til modellen for regression
Jf. SAS (s. 3 + s. 7 i udskriften) fås

$$F = \frac{27666,009 - 23655,250}{33 - 30} = 1,696$$

$\frac{782,508}{782,508}$

s. 209
 i bilag
 (4.12)

$$E(X) = 1 - F_{F(3,30)}(F) \geq 1 - 0,1 = 0,9 > 0,05$$

Dvs. vi kan reducere til modellen om to regressionslinier

Jf. SAS (s. 7 + s. 23 i udskriften) samt (4.12) s. 209 i bilag 1 fås at testet for reduktion til modellen m. parallelle regressionslinier er

$$F = \frac{30782,974 - 27666,009}{35 - 33} = 1,859$$

$\frac{838,364}{838,364}$

$$E(X) = 1 - F_{F(2,33)}(F) \geq 1 - 0,1 = 0,9 > 0,05$$

Vi forkaster ikke hypotesen om reduktion til modellen m. parallelle regressionslinier

6)

Middelværdinummeret for modellen m. to regressionslinier m. samme hældning er indeholdt i middelværdinummeret for additivitetsmodellen - vi kan derfor teste for reduktion fra additivitetsm. til regressionsm. (jvf. SAT s. 5 + s. 23)

$$F = \frac{\frac{30782,974 - 30632,317}{35 - 34}}{900,950} = 0,167 \quad ((4,12) \text{ s. } 209 \text{ i bind 1})$$

$$E(X) = 1 - F_{F(1,34)}(F) \geq 1 - 0,5 = 0,5 > 0,05^*$$

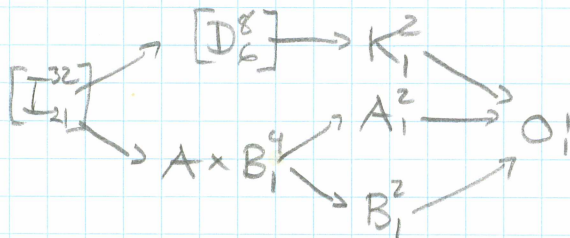
Dvs. vi kan reducere til regressionsmodellen (m. linier m. samme hældning)

*
Man kunne finde et mere præcist udtryk v.h.j.a. bn s. 100 i bind 1

opgave 10.13

Faktorene er I, D dag, K køn, A, B, O

Faktorstrukturdiagrammet er



ANOVA-tabellen bliver

Faktor	d	SSD
K	1	10,92781250
D	6	36,33187500
A	1	11,88281250
B	1	10,46531250
AxB	1	5,52781250
I	21	35,2965625

Juf. s. 134-135 er testet for 'ingen udskejelser, m. A og B'

$$F = \frac{SSD_{A \times B} / d_{A \times B}}{SSD_I / d_I} = 3,289$$

$$E(K) = 1 - F_{F(1,21)}(F) > 1 - 0,95 = 0,05$$

Dvs. modellen reduceres til

$$A + B + K + [D + I]$$

I den reducerede model er testss. for at
 fene hh. K, B, A hh. 0,239; 0,0267; 0,0191,
 dvs. vi kan ikke umiddelbart teste flere systematiske
 faktorer vkh.

Juf. s. 135m - 136 i Bnd 3 er testes for $\sigma_D^2 = 0$

$$F = \frac{\hat{\lambda}_D}{\hat{\lambda}_I} = \frac{\text{SSD}_D / d_D}{\text{SSD}_I / d_I} = \frac{\text{SSD}_D / d_D}{\text{SSD}_I + \text{SSD}_{A \times B} / (d_I + d_{A \times B})}$$

$\begin{matrix} (10,46) \\ s. 133 \end{matrix} \quad \begin{matrix} (10,24) \\ (10,26) \\ s. 113 \end{matrix}$

$$= 3,263$$

$$1 - F_{F(6,22)}(F) < 1 - 0,975 = 0,025 \quad (\text{dvs. } D \text{ er signifikant})$$

Dvs. slutmodellen bliver

$$A + B + K + [D + I]$$

opgave 10.15 $\rightarrow$ se TB m. 25

opgave 10.14

$\mathcal{I}_1, \dots, \mathcal{I}_k, U_{11}, \dots, U_{kn_k}$ uafh.

$$\mathcal{I}_i \sim N(0, \omega^2) \quad i=1, \dots, k$$

$$U_{ij} \sim N(0, \sigma^2) \quad i=1, \dots, k \quad j=1, \dots, n_i$$

$$X_{ij} = \mu + \mathcal{I}_i + U_{ij} \quad i=1, \dots, k \quad j=1, \dots, n_i$$

$$N = n.$$

1)

Pastand:

$$\begin{aligned} \ell(\mu, \sigma^2, \omega^2) = & -\frac{1}{2}N \log 2\pi - \frac{1}{2}(N-k) \log \sigma^2 - \frac{1}{2} \sum_i \log \tau_i^2 \\ & - \frac{SSD_i}{2\sigma^2} - \sum_i \frac{n_i (\bar{X}_i - \mu)^2}{2\tau_i^2} \end{aligned}$$

$$\tau_i^2 := \sigma^2 + (n_i - 1)\omega^2, SSD_i := \sum_j (X_{ij} - \bar{X}_i)^2$$

Beris:

$$\text{Cov}(X_{i_1 j_1}, X_{i_2 j_2}) = \begin{cases} 0, & i_1 \neq i_2 \\ \omega^2, & i_1 = i_2, j_1 \neq j_2 \\ \omega^2 + \sigma^2, & i_1 = i_2, j_1 = j_2 \end{cases} \quad (*)$$

J-f. sætn. 9.17 (B) er $(X_{11}, \dots, X_{kn_k}) \sim N_N(\underline{\mu}, \underline{\Sigma})$

hvor $\underline{\Sigma}$ er kovariansmatricen bestemt ved (*)

Tuf. opgave 10.2 og transformationsseth. (for tetheder)

gælder for $(Y_1, \dots, Y_n)' = A(X_1, \dots, X_n)'$

A ortogonal, $a_{ij} = \frac{1}{\sqrt{n_i}}$, $j=1, \dots, n_i$

$$L(\mu, \sigma^2, \omega^2) = \prod_{i=1}^k \frac{1}{\sqrt{2\pi} \tau_i^2} e^{-\frac{1}{2} (\sqrt{n_i} \mu - Y_{i1})^2 / \tau_i^2} \left(\prod_{i=1}^k \frac{1}{\sqrt{2\pi \sigma^2}} e^{-\frac{1}{2} Y_{ij}^2 / \sigma^2} \right)$$

$$= (2\pi)^{\frac{N}{2}} (\sigma^2)^{\frac{N-k}{2}} \prod_{i=1}^k (\tau_i^2)^{\frac{1}{2}} e^{-\frac{1}{2} \sum_{i=1}^k (\sqrt{n_i} \mu - Y_{i1})^2 / \tau_i^2} e^{-\frac{1}{2\sigma^2} \sum_{i=1}^k \sum_{j=2}^{n_i} Y_{ij}^2}$$

$$= (2\pi)^{\frac{N}{2}} (\sigma^2)^{\frac{N-k}{2}} \prod_{i=1}^k (\tau_i^2)^{\frac{1}{2}} e^{-\frac{1}{2} \sum_{i=1}^k n_i (\bar{Y}_{i1} / \sqrt{n_i} - \mu)^2 / \tau_i^2} e^{-\frac{1}{2\sigma^2} \sum_{i=1}^k \sum_{j=2}^{n_i} Y_{ij}^2}$$

$$= (2\pi)^{\frac{N}{2}} (\sigma^2)^{\frac{N-k}{2}} \prod_{i=1}^k (\tau_i^2)^{\frac{1}{2}} e^{-\frac{1}{2} \sum_{i=1}^k n_i (\bar{X}_{i1} - \mu)^2 / \tau_i^2} e^{-\frac{1}{2\sigma^2} \sum_{i=1}^k \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_{i1})^2}$$

$$\text{idet } Y_{i1} = \frac{1}{\sqrt{n_i}} X_{i1}, \quad \sum_{i=1}^k \sum_{j=2}^{n_i} Y_{ij}^2 = - \sum_{i=1}^k Y_{i1}^2 + \sum_{i=1}^k \sum_{j=1}^{n_i} Y_{ij}^2$$

$$= - \sum_{i=1}^k \frac{X_{i1}^2}{n_i} + \sum_{i=1}^k \sum_{j=1}^{n_i} X_{ij}^2 = \sum_{i=1}^k \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_{i1})^2$$

($\sum_{j=1}^{n_i} X_{ij} = \sum_{j=1}^{n_i} X_{ij}$)

Ved at tage logaritmen på begge sider følger det ensbede $\square$

Påstand:

$$l_\mu = \frac{\partial \ell(\mu, \sigma^2, \omega^2)}{\partial \mu} = \sum_{i=1}^k \frac{n_i (\bar{X}_{i1} - \mu)}{\tau_i^2}$$

$$l_{\sigma^2} = \frac{\partial \ell(\mu, \sigma^2, \omega^2)}{\partial \sigma^2} = - \frac{(N-k)}{2\sigma^2} - \frac{1}{2} \sum_{i=1}^k \frac{1}{\tau_i^2} + \frac{SSD_1}{2\sigma^4} + \sum_{i=1}^k \frac{n_i (\bar{X}_{i1} - \mu)^2}{2\tau_i^4}$$

$$l_{\omega^2} = \frac{\partial \ell(\mu, \sigma^2, \omega^2)}{\partial \omega^2} = - \frac{1}{2} \sum_{i=1}^k \frac{n_i}{\tau_i^2} + \sum_{i=1}^k \frac{n_i^2 (\bar{X}_{i1} - \mu)^2}{2\tau_i^4}$$

Beweis: Folg direkte af 1) idet $\frac{d}{dx} \frac{1}{x} = -\frac{1}{x^2}$

$$\frac{d}{dx} \log x = \frac{1}{x}$$

$\square$

3)

Antag $(\hat{\mu}, \hat{\sigma}^2, \hat{\omega}^2)$ er en ksn. til ligh. i 2)

$$\begin{aligned} L_{\mu} \text{ giver } \hat{\mu} \cdot \sum_i \frac{n_i}{t_{i,2}} &= \sum_i \frac{n_i \bar{x}_{i,1}}{t_{i,2}} \Rightarrow \hat{\mu} = \frac{\sum_i \frac{n_i \bar{x}_{i,1}}{t_{i,2}}}{\sum_i \frac{n_i}{t_{i,2}}} \\ &= \frac{\sum_i \frac{n_i \bar{x}_{i,1}}{n_i \omega^2 + \sigma^2}}{\sum_i \frac{n_i}{n_i \omega^2 + \sigma^2}} \end{aligned}$$

Vi ser, at for $n_1 = \dots = n_k = n$: $\hat{\mu} = \frac{\frac{x_{..}}{n \omega^2 + \sigma^2}}{\frac{k \cdot n}{n \omega^2 + \sigma^2}} = \bar{x}_{..}$

4)

Af 2) følger $L_{\sigma^2} = 0 \wedge L_{\omega^2} = 0 \iff$

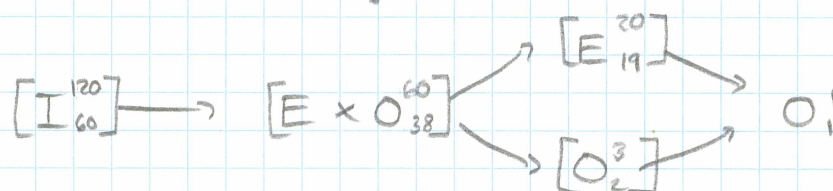
$$-\frac{(N-k)}{\hat{\sigma}^2} + \frac{SSD_1}{2\hat{\sigma}^4} + \sum_i \frac{n_i (\bar{x}_{i,1} - \hat{\mu})^2}{(\hat{\sigma}^2 + n_i \hat{\omega}^2)} - \sum_i \frac{1}{(\hat{\sigma}^2 + n_i \hat{\omega}^2)} = 0$$

$$\wedge \sum_i \frac{n_i^2 (\bar{x}_{i,1} - \hat{\mu})^2}{(\hat{\sigma}^2 + n_i \hat{\omega}^2)} = \sum_i \frac{n_i}{\hat{\sigma}^2 + n_i \hat{\omega}^2}$$

opgave 10.19

Faktorerne er E, emner; O, operator

Faktorstrukturdiagrammet er



ANOVA -tabellen:

Faktor	d	SSD
O	2	2,616667
E	19	1185,425000
ExO	38	27,050000
I	60	59,500000

Juf. s. 135m-136 i Bind 3 får følgende test for hypotesen $\sigma_{ExO}^2 = 0$:

$$F = \frac{\hat{\lambda}_{ExO}}{\hat{\lambda}_I} = \frac{SSD_{ExO} / d_{ExO}}{SSD_I / d_I} = 0,71782$$

(10,46)
5,83

$1 - F_{F(38,60)}(F) > 0,5$ dvs. vi kan reducere til

$$[E + O + I]$$

Juf. (10.24) + (10.25) s. 113 fås ANOVA-tabellen
for den nye model ($SSD_I = SSD_E + SSD_{Eco}$, $d_I = d_E + d_{Eco}$)

Faktor	d	SSD
O	2	2,616667
E	19	1185,425000
I	98	86,550000

Juf. SAS udskriften s. 7 og formelen s. 133 + 135 - 136
ses at alle de resterende faktorer er signifikante.

Meddelelse 27

Forelæsningerne i sidste uge før påske

(Onsdag den 9. april): Fortsatte fra side 132 og nåede frem midt på side 134 i *Test af hypoteser i middelværdien*. Brugte desuden en del tid på at bruge Eksempel 10.6 som illustration.

(Fredag den 11. april): Nåede til midt i beviset for Korollar 10.27.

Forelæsningerne i denne uge

(Onsdag den 23. april): Afsluttede gennemgangen af teorien om variansanalyse.

(Fredag den 25. april): Beregninger i SAS i tilknytning til variansanalyse.

Forelæsningerne i næste uge

Onsdag den 30. april og fredag den 2. maj.

Øvelserne i de næste uger

(Uge 18: 28. april - 4. maj.): opgave 3 fra vintereksamen i Stat 1 2001/2002 (stillet på Meddelelse 26) (Mortens hold), opgave 2 fra sommeren 1998 (opgaven kan findes på nettet, men bliver også lagt frem på Stat 1 hylden), 10.13, 10.15 (Jans hold), 10.14, 10.19.

(Uge 19: 5.-11. maj): Ikke endeligt fastlagt, men kunne omfatte 10.17, samt opgaver fra Stat 1 eksamenssættet fra sommeren 2001.

Sidste obligatoriske opgave

Den sidste obligatoriske opgave er stillet til aflevering den 12. maj. Det er muligt at lave individuel aftale om en anden afleveringsdato ved henvendelse til Jan eller Morten.

Data og SASprogrammer

Husk data og SASprogrammer på adressen:

http://home.imf.au.dk/statbib/atkurser/stat1/opgaver/kapitel_10

Bind 3 på nettet

Opdaterede pdf og ps versioner af bind 3 findes på adressen:

<http://home.imf.au.dk/statbib/atkurser/stat1/noter/bind3/>

STAT1, TØ nr. 25

opgave 10.12

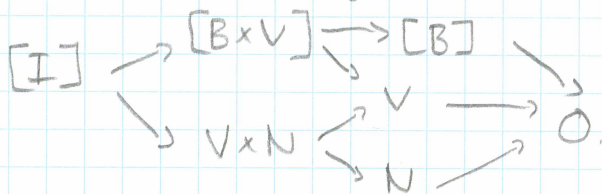
I subplots, P plots, B blokke, V søtte, N kvælstof

Vi ser, at $P \sim B \times V$

Designet er $D := \{I, B \times V, B, V, V \times N, N, O\}$

D er et ortogonalt design jvf. def. 10.10 og
bm. s. 106n i 10.3.2

Faktorstrukturdiagrammet er



ANOVA-tabelen for D er

Faktor	SSD	d
B	15875,278	5
V	1786,361	2
BxV	6013,306	10
N	20020,5	3
VxN	321,75	6
I	7968,75	45

H_0 : Vi kan fjerne $V \times N$ (der er ingen interaktion mellem V og N)

$$F = \frac{SSD_{V \times N} / d_{V \times N}}{SSD_I / d_I} = 0,30282$$

$$E(X) = 1 - F_{F(6,45)}(F) \geq 1 - 0,5 = 0,5 > 0,05$$

J.f. s. 134-135m

Dvs. vi reducerer til

$$V + N + [B + B \times V + I]$$

ANOVA-tabellen for den reducerede model er

som før blot er $SSD_I = SSD_{I_{gmi.}} + SSD_{V \times N}$,

$$d_I = d_{I_{gmi.}} + d_{V \times N} \quad \text{j.f. (10.24) + (10.26) s. 113}$$

H_0 : Vi kan fjerne V (dvs. der er ikke forskel på landesatir)

$$F = \frac{SSD_V / d_V}{SSD_{B \times V} / d_{B \times V}} = 1,48534$$

$$E(X) = 1 - F_{F(2,10)}(F) \geq 1 - 0,9 = 0,1 > 0,05$$

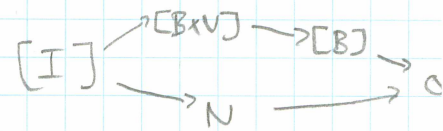
J.f. s. 134-135m

Dvs. vi reducerer til

$$N + [B + B \times V + I]$$

ANOVA-tabellen for den reducerede model j/f. (10.24)
+ (10.26) s. 113

Faktor	d	SSD
B	5	15875,2778
BxV	12	7799,6667
N	3	20020,5
I	51	8290,5



H_0 : Vi kan fjerne N (dvs. niveau kvælstof har ingen betydning)

$$F = \frac{SSD_N / d_N}{SSD_I / d_I} = 41,05283$$

$$P(X) = 1 - F_{F(3,51)}(F) \leq 1 - 0,9995 = 0,0005 < 0,05$$

Dvs. hypotesen forkastes og slutmodellen bliver

$$N + [B + BxV + I]$$

Source	Pr > F
Model	<.0001
Error	
Corrected Total	

R-Square	Coeff Var	Root MSE	udbytte Mean
0.846713	12.79887	13.30727	103.9722

Source	DF	Type I SS	Mean Square	F Value
blok	5	15875.27778	3175.05556	17.93
v	2	1786.36111	893.18056	5.04
blok*v	10	6013.30556	601.33056	3.40
n	3	20020.50000	6673.50000	37.69
v*	6	321.75000	53.62500	0.30

Source	Pr > F
blok	<.0001
v	0.0106
blok*v	0.0023
n	<.0001
v*n	0.9322

Source	DF	Type III SS	Mean Square	F Value
blok	5	15875.27778	3175.05556	17.93
v	2	1786.36111	893.18056	5.04
blok*v	10	6013.30556	601.33056	3.40
n	3	20020.50000	6673.50000	37.69
v*n	6	321.75000	53.62500	0.30

Source	Pr > F
blok	<.0001
v	0.0106
blok*v	0.0023
n	<.0001
v*n	0.9322

opgave 10.15

$$T + B + [B_0] \rightarrow T + [B] \quad (*)$$

$$B = \{B\} \cup B_0, \quad B_0 = \bigwedge_{\substack{B' \in \mathcal{B} \\ B \subseteq B'}} B'$$

$\hat{\lambda}_B, \hat{\lambda}_{B_0}$ er estimatorerne for de kanoniske varianskomponenter λ_B, λ_{B_0} i modellen $T + [B]$

Antag at $T + [B]$ opfylder (T&B) og B og B_0 opfylder (B1)-(B4) s. 124-125

Da gælder at F-testoren for reduktionen (*) er

$$F = \frac{\hat{\lambda}_B}{\hat{\lambda}_{B_0}}$$

Beweis:

Da B_0 er lukket under minimumsdannelse er B_0 element i B_0 (dvs. $B_0 \neq B$)

Jrf. (10.43) s. 130 er $\lambda_B = n_B \sigma_B^2 + \lambda_{B_0}$

Af bem. s. 133 følger, jf. antagelserne, at

$$\lambda_B = \frac{SSD_B}{d_B}, \quad \lambda_{B_0} = \frac{SSD_{B_0}}{d_{B_0}}$$

$$F = \frac{SSD_B / d_B}{SSD_{B_0} / d_{B_0}} = \frac{\hat{\lambda}_B}{\hat{\lambda}_{B_0}}, \quad \text{da } B \in \mathcal{D}_{B_0} \text{ (pr. def af } B_0, \mathcal{D}_{B_0})$$

(X) Jrf. s. 134-135 m

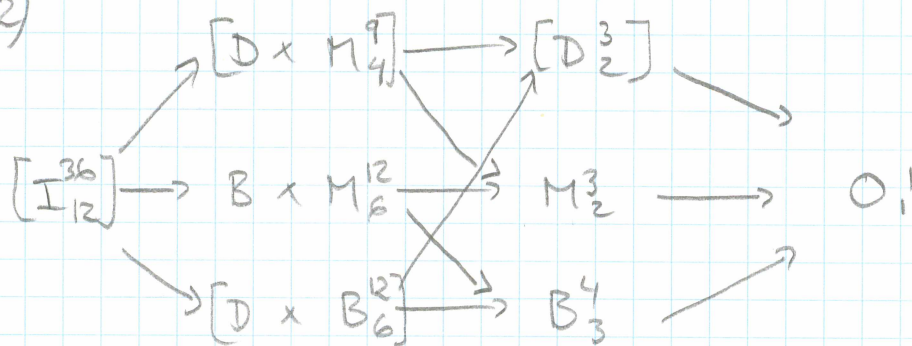
□

V01/02 opgave 3

1)

Dag x Metode x Blanding er balanceret derved følger jvf. bm. i 10.3.2 s. 106, at faktorerne i D er parvis ortogonale, da $I \in D$ og vi, ligeledes i flg. 10.3.2, har at D er lukket mht. minimumsdannelse er D et ortogonalt design jvf. def. 10.10 s. 108

2)



3)

Modellen $T + [B]$ er jvf. 1) et ortogonalt design og B opfylder betingelserne på s. 124-125

(bl.a. jvf. prop. 10.24) Dette opfyldes også af

$T + [B']$, $T + [B'']$, $T + [B_1]$, hvor

$B' = \{I, D \times M, D\}$, $B'' = \{I, D \times M\}$

H. : Reduktion til $T + [B']$

$$F = \frac{\hat{\lambda}_{D \times B}}{\hat{\lambda}_I} = \frac{\text{SSD}_{D \times B} / d_{D \times B}}{\text{SSD}_I / d_I} = 1,03110$$

$\begin{matrix} \text{s. 135} & & \text{s. 132} \\ -136m & & -133 \end{matrix}$

$$E(x) = 1 - F_{F(6,12)}(F) \geq 1 - 0,7 = 0,3 > 0,05$$

$\begin{matrix} \text{s. 136} & & \text{s. 132} \end{matrix}$

Dvs. vi reduserer til $T + [B']$

ANOVA-tabellen <v som før, dog sættes (jvf. (10.24)+(10.26) s. 113

$$d_I^1 = d_I + d_{D \times B}, \quad SSD_I^1 = SSD_I + SSD_{D \times B}$$

$$H: T + [B'] \rightarrow T + [B'']$$

$$F = \frac{SSD / d_D}{SSD_{D \times M} / d_{D \times M}} = 2,07987$$

s. 135-136
s. 132-133

$$E(X) = 1 - F_{F(2,4)}(F) \geq 1 - 0,09 = 0,170105$$

Dvs. vi reducerer til $T + [B'']$

Jvf. (10.24)+(10.26) s. 113 sættes

$$d_{D \times M}'' = d_{D \times M} + d_D, \quad SSD_{D \times M}'' = SSD_{D \times M} + SSD_D$$

$$H: T + [B''] \rightarrow T + [B_1]$$

$$F = \frac{SSD_{D \times M}'' / d_{D \times M}''}{SSD_I^1 / d_I^1} = \frac{(SSD_{D \times M} + SSD_D) / (d_{D \times M} + d_D)}{(SSD_I + SSD_{D \times B}) / (d_I + d_{D \times B})}$$

s. 135-136
s. 132-133

$$= 0,90237$$

$$E(X) = 1 - F_{F(6,18)}(F) \geq 1 - 0,5 = 0,5 > 0,05$$

Dvs. vi kan reducere til $T + [B_1] = T + [I]$

Jvf. (10.24) + (10.26) s. 113 sættes

$$d_I^1 = d_I^1 + d_{D \times M}'' = d_I + d_{D \times B} + d_{D \times M} + d_D$$

$$SSD_I^1 = SSD_I^1 + SSD_{D \times M}'' = SSD_I + SSD_{D \times B} + SSD_{D \times M} + SSD_D$$

4)

$$H: T + [I] \rightarrow T_1 + [I]$$

$$F = \frac{SSD_{B \times M} / d_{B \times M}}{SSD_I^1 / d_I^1} = 2,31781$$

S.134
-BSm

$$E(\underline{X}) = 1 - F_{F(6,24)}(F) > 1 - 0,95 = 0,05$$

S.134

Dus, vi kan redusere til slutmodellen $T_1 + [I]$

Meddelelse 26

Forelæsningerne i sidste uge

(Onsdag den 2. april): Afsnit 10.7 *Varianskomponentmodeller* frem til Sætning 10.23.

(Fredag den 4. april): Afsnit 10.7 *Varianskomponentmodeller* fra Sætning 10.23 og frem til midt på side 132. Desuden gennemgik jeg Eksempel 10.16 om split plot design, som blev oversprunget ved gennemgangen af de øvrige eksempler den 28. marts.

Forelæsningerne i denne uge

(Onsdag den 9. april): Fortsatte fra side 132 og nåede frem midt på side 134 i *Test af hypoteser i middelværdien*. Brugte desuden en del tid på at bruge Eksempel 10.6 som illustration.

(Fredag den 11. april): Afslutter gennemgangen af teorien om variansanalyse, og begynder så småt at omtale beregninger i SAS.

Forelæsningerne i „næste“ uge

Onsdag den 23. og fredag den 25. april efter påske fortsætter og afslutter jeg beregninger i SAS, og så er gennemgangen af variansanalyse afsluttet.

Øvelserne i de næste uger

(Uge 16-17: 14. april - 27. april): 10.4, 10.5, (fortsat) 10.12, og 10.15, samt nedenstående opgave, som er en opgave 3 fra vintereksamen 2001/2002.

Afleveringsopgaver

Data og SASprogrammer

Husk data og SASprogrammer på adressen:

http://home.imf.au.dk/statbib/atskurser/stat1/opgaver/kapitel_10

Opgave 3, Vinteren 2001/2002

For at undersøge betydningen af fire forskellige blandinger af farvepigment i en maling udførte man følgende forsøg. Først lavede man de fire blandinger af farvepigment, som blev blandet i samme maling, hvorved man fik fire forskellige malinger. Disse blev påført en flade med tre forskellige metoder: pensel, sprøjte og rulle. Hele denne procedure blev udført ialt tre gange på tre forskellige dage. Den målte respons er reflektansen af den tørrede maling. De målte

<i>Faktor</i>	<i>Kort navn</i>	<i>Frihedsgrader</i>	<i>SSD</i>
<i>Dag</i>	<i>D</i>	2	2.0417
<i>Blanding</i>	<i>B</i>	3	307.4789
<i>Metode</i>	<i>M</i>	2	222.0950
<i>Blanding</i> $\times$ <i>Metode</i>	<i>B</i> $\times$ <i>M</i>	6	10.0361
<i>Dag</i> $\times$ <i>Metode</i>	<i>D</i> $\times$ <i>M</i>	4	1.9633
<i>Dag</i> $\times$ <i>Blanding</i>	<i>D</i> $\times$ <i>B</i>	6	4.5294
<i>I</i>	<i>I</i>	12	8.7856
<i>I alt</i>		35	556.9300

Tabel 1: Variansanalysetabel.

responser er vist i nedenstående tabel, som samtidig giver definitionen af de tre basale faktorer i forsøget: *Dag*, *Metode* og *Blanding*.

<i>Dag</i>	<i>Metode</i>	<i>Blanding</i>			
		1	2	3	4
1	1	64.5	66.3	74.1	66.5
	2	68.3	69.5	73.8	70.0
	3	70.3	73.1	78.0	72.3
2	1	65.2	65.0	73.8	64.8
	2	69.2	70.3	74.5	68.3
	3	71.2	72.8	79.1	71.5
3	1	66.2	66.5	72.3	67.7
	2	69.0	69.0	75.4	68.6
	3	70.8	74.2	80.1	72.4

Betragt designet $\mathcal{D}$, der består af faktorerne,

$$\mathcal{D} = \{I, Blanding \times Metode, Dag \times Metode, Dag \times Blanding, Dag, Metode, Blanding, O\}.$$

- 1) Vis, at $\mathcal{D}$ er et ortogonalt design.
- 2) Tegn et faktorstrukturdiagram og angiv frihedsgrader for kvadratsummerne.

I Tabel 1 er angivet variansanalysetabellen for designet $\mathcal{D}$.

Betragt modellen, hvor *Dag* opfattes som en tilfældig faktor og mængden af tilfældige faktorer dermed er $\mathcal{B} = \{I, Dag \times Metode, Dag \times Blanding, Dag\}$ og mængden af systematiske faktorer er $\mathcal{T} = \{Blanding \times Metode, Metode, Blanding, O\}$.

- 3) Gør rede for at modellen kan reduceres til $\mathcal{T} + \mathcal{B}_1$ hvor $\mathcal{B}_1 = \{I\}$.
- 4) Gør dernæst rede for at modellen kan reduceres til $\mathcal{T}_1 + \mathcal{B}_1$ hvor $\mathcal{T}_1 = \{Metode, Blanding, O\}$.

STAT1, TD nr. 25

opgave 10.4

$$H: \eta = 0 \iff H: \tau^2 = \sigma^2 \quad (\text{se opg. 10.3})$$

1)

Påstand:

Likelihood ratio testeren for H er:

$$Q = \frac{(S_1^2)^{\frac{k(n-1)}{2}} \left(\frac{k-1}{k} S_2^2\right)^{\frac{k}{2}}}{\left(\frac{1}{kn} (k(n-1)S_1^2 + (k-1)S_2^2)\right)^{kn/2}}$$

Beris:

Under H : (jf. opg. 10.3 2))

$$L_H(\alpha, \tau^2) = \left(\frac{1}{2\pi\tau^2}\right)^{kn/2} e^{-\frac{1}{2\tau^2} \left(\sum_{i=1}^k (y_{i1} - \alpha)^2 + \sum_{i=2, j=2}^k y_{ij}^2\right)}$$

Vi ser (ved at indsette $\hat{\alpha}$) at

$$\hat{\tau}^2 = \frac{1}{kn} \left(\sum_{i=1}^k (y_{i1} - \bar{y}_{\cdot 1})^2 + \sum_{i=2, j=2}^k y_{ij}^2\right) = \frac{1}{kn} ((k-1)S_2^2 + k(n-1)S_1^2)$$

Af opg. 10.3 3) følger dermed

$$\begin{aligned} Q &= \frac{\sup_H L}{\sup L} = \frac{\left(\frac{1}{2\pi\hat{\tau}_H}\right)^{kn/2} e^{-\frac{1}{2\hat{\tau}_H^2} \left(\sum_{i=1}^k (\bar{y}_{i1} - \alpha)^2 + \sum_{i=2, j=2}^k \bar{y}_{ij}^2\right)}}{\left(\frac{1}{2\pi\hat{\tau}^2}\right)^{kn/2} \left(\frac{1}{2\pi\hat{\tau}^2}\right)^{\frac{k(n-1)}{2}} e^{-\frac{1}{2\hat{\tau}^2} \left(\sum_{i=1}^k (\bar{y}_{i1} - \alpha)^2 - \frac{1}{2\hat{\tau}^2} \sum_{i=2, j=2}^k \bar{y}_{ij}^2\right)}} \\ &= \frac{\left(\frac{1}{2\pi\hat{\tau}_H}\right)^{kn/2}}{\left(\frac{1}{2\pi\hat{\tau}^2}\right)^{kn/2} \left(\frac{1}{2\pi\hat{\tau}^2}\right)^{\frac{k(n-1)}{2}}} \frac{e^{-\frac{kn}{2}}}{e^{-\frac{k}{2} - \frac{k(n-1)}{2}}} = \frac{\left(\frac{k}{k-1} S_2^2\right)^{k/2} (S_1^2)^{\frac{k(n-1)}{2}}}{\left(\frac{1}{kn} ((k-1)S_2^2 + k(n-1)S_1^2)\right)^{kn/2}} \end{aligned}$$

↑
opg. 10.3 5)

□

2)

$$Q = \frac{(S_1^2)^{\frac{kn}{2}} \left(\frac{k}{k-1} \frac{S_2^2}{S_1^2} \right)^{k/2}}{(S_1^2)^{\frac{kn}{2}} \left(\frac{1}{kn} \left((k-1) \frac{S_2^2}{S_1^2} + k(n-1) \right) \right)^{kn/2}}$$

$$= \frac{\left(\frac{k}{k-1} F \right)^{k/2}}{F^{\frac{kn}{2}} \left(\frac{1}{kn} ((k-1)F + k(n-1)) \right)^{kn/2}} \quad (*)$$

Hypotesen er kritisk for små verdier af Q

Af (*) ser vi, at Q er lille når F er lille

$$(*) = F^{k/2 - kn/2} \frac{\left(\frac{k}{k-1} \right)^{k/2}}{\left(\frac{1}{kn} ((k-1) + \frac{k(n-1)}{F}) \right)^{kn/2}} \quad (+)$$

Af (+) følger at Q er lille for F stor

idet $k/2 - kn/2 < 0$, $n > 1$

opgave 10.5

1)

$$\begin{aligned} \text{Cov}(X_{ij}, X_{rs}) &= E[(X_{ij} - \mu)(X_{rs} - \mu)] \\ &= E[(\bar{Y}_i + U_{ij})(\bar{Y}_r + U_{rs})] \\ &= E[\bar{Y}_i \bar{Y}_r] + E[U_{ij} \bar{Y}_r] + E[\bar{Y}_i U_{rs}] + E[U_{ij} U_{rs}] \\ &= \begin{cases} \sigma^2 + \omega^2, & i=r, j=s \\ E[\bar{Y}_i^2] = \omega^2, & i=r, j \neq s \\ 0, & i \neq r, j \neq s \end{cases} \end{aligned}$$

2)

Regningerne i opg. 10.2-3 bliver ikke ændret udt $\hat{\mu}$,
dvs. $\hat{\mu} = \bar{X}_{..}$

Jf. beregningerne i opg. 10.3 fås, at
 $(\hat{\sigma}^2, \hat{\tau}^2) = (s_1^2, s_2^{2'})$, $s_1^2 \leq s_2^{2'}$ (med betegnelse fra 10.3)

Hvis $s_2^{2'} < s_1^2$ (hvilket indikerer at $\eta < 0$) må
vi tolke dette som at data peger på $\eta = 0$,
dvs. $\tau^2 = \sigma^2$ og dermed bliver (j.f. udregn. i opg. 10.4)

$$\hat{\tau}^2 = \hat{\sigma}^2 = s_0^{2'}, \quad s_2^{2'} < s_1^2$$

3)

Jvf. opg. 10.4 er $H: \omega^2 = 0$ ækvivalent
med $H: \sigma^2 = \tau^2$

Jvf. opg. 10.4 og 2) fås

$$Q = \begin{cases} \frac{(s_1^2)^{\frac{k(n-1)}{2}} \left(\frac{k-1}{k} s_2^2 \right)^{k/2}}{s_2^{k/2}} & s_2^{21} \geq s_1^2 \\ 1 & s_2^{21} < s_1^2 \end{cases}$$

4)

Idet $Q \neq 1$: $s_2^{21} \geq s_1^2$, dvs. $F = \frac{s_2^{21}}{s_1^2}$, dvs. F
praktisk talt $(s_2^{21} \neq s_1^2)$ større end 1
Forhælder vi jvf. opg. 10.4 2) for store værdier
af F

Meddelelse 25

Forelæsningerne i sidste uge

(Onsdag den 26. marts): Jeg resumerede resultatet i hovedsætningen (Sætning 10.11), og afsluttede Afsnit 10.4. Fortsatte med konstruktion af ANOVA tabellen og nåede til og med Eksempel 10.13 om tosidet variansanalyse på side 114.

(Fredag den 28. marts): Fortsatte med eksemplerne på konstruktion af ANOVA tabeller i Afsnit 10.5, men sprang over Eksempel 10.16 om split-plot design. Gennemgik derefter hele Afsnit 10.6 om Lineære normale modeller.

Forelæsningerne i denne uge

(Onsdag den 2. april): Afsnit 10.7 *Varianskomponentmodeller* frem til Sætning 10.23.

(Fredag den 4. april): Afsnit 10.7 *Varianskomponentmodeller* fra Sætning 10.23 og frem til midt på side 132. Desuden gennemgik jeg Eksempel 10.16 om split plot design, som blev oversprunget ved gennemgangen af de øvrige eksempler den 28. marts.

Forelæsningerne i næste uge

(Onsdag den 9. april og fredag den 11. april): Fortsætter fra side 132.

Øvelserne i de næste uger

(Uge 15: 7. april - 13. april): 10.11 (fortsat), 10.3, 10.4, 10.5

(Uge 16-17: 14. april - 27. april): 10.12, 10.14, 10.15

Afleveringsopgaver

Data og SASprogrammer

Husk data og SASprogrammer på adressen:

http://home.imf.au.dk/statbib/atsskurser/stat1/opgaver/kapitel_10

STAT 1 TØ nr. 24

opgave 10.8

1)

Påstand:

$L_1, L_2 \subseteq \mathbb{R}^I$ lin. undrum

P_1, P_2 ortogonale projection på hh. L_1, L_2

L_1, L_2 er geometrisk ortogonale $\iff P_1 P_2 = P_2 P_1$

Beris:

Antag at L_1, L_2 er geometrisk ortogonale, dvs.

$$V_1 \perp V_2, \quad V = L_1 \cap L_2, \quad V_1 = (L_1 \cap L_2)^\perp \cap L_1 \\ V_2 = (L_1 \cap L_2)^\perp \cap L_2, \quad W = L_1^\perp \cap L_2^\perp$$

Vi ser, at $\mathbb{R}^I = V \oplus V_1 \oplus V_2 \oplus W$ (*)

Lad $x \in \mathbb{R}^I$ være vilk.

Jf. (*) $\exists!$ $x_v \in V, x_{v_1} \in V_1, x_{v_2} \in V_2, x_w \in W$ s

$$x = x_v + x_{v_1} + x_{v_2} + x_w \quad \text{og}$$

$$P_1 x = x_v + x_{v_1}, \quad P_2 x = x_v + x_{v_2}$$

$$P_2 P_1 x = x_v = P_1 P_2 x, \quad \text{dvs.}$$

$$P_1 P_2 = P_2 P_1$$

Antag omvendt, at $P_1 P_2 = P_2 P_1$ (+)

For x vilk.: $P_1 P_2 x \in L_1$, $P_2 P_1 x \in L_2$
 $\stackrel{(1)}{\implies} P_1 P_2 x = P_2 P_1 x \in L_1 \cap L_2$

Lad $x_1 \in V_1$, $x_2 \in V_2$ (Dvs. $P_1 P_2 x_2 \perp x_1$)

$$x_1^* x_2 = (P_1 x_1)^* (P_2 x_2) = x_1^* P_1^* P_2 x_2 = x_1^* (P_1 P_2 x_2) = 0$$

Altså $V_1 \perp V_2$

□

2)

Påstand:

Hvis L_1, L_2 er geometrisk ortogonale da
er $L_1, L_1^\perp, L_2, L_2^\perp$ parvis geometrisk
ortogonale

Beris:

Vi skal vise at hh. $L_1, L_2^\perp$; $L_2, L_1^\perp$; $L_1^\perp, L_2^\perp$
er geo. ort., det er nok at vise at $L_1, L_2^\perp$; $L_1^\perp, L_2^\perp$
er geo. ort.

Den ort. projection på $L_2^\perp$ er givet ved $(I - P_2)$

$$(I - P_2)P_1 = P_1 - P_2 P_1 = P_1 - P_1 P_2 = P_1(I - P_2)$$

Dvs. j.f. 1) er $L_1, L_2^\perp$ geo. ort. tilsvarende ses
at $L_1^\perp, L_2^\perp$ er geo. ort., idet

$$\begin{aligned} (I - P_1)(I - P_2) &= I - P_2 - P_1 - P_1 P_2 \\ &= I - P_2 - P_1 - P_2 P_1 = (I - P_2)(I - P_1) \end{aligned}$$

□

3)

Påstand:

$L_1, \dots, L_k$ parvis geo. ort. $P_1, \dots, P_k$ tilhørende
 ort. proj., da gælder $P = P_1 \cdot \dots \cdot P_k$ ort. proj.
 på $L_1 \cap \dots \cap L_k$

Beris:

Det er (pga. induction) nok at vise påstanden
 at $P_1 \cdot P_2$ er den ort. proj. på $L_1 \cap L_2$

Lad $x \in \mathbb{R}^F$ være vilk.

Jvf. udregningene i 1) er $P_1 P_2$ den ort. proj.
 på $L_1 \cap L_2$

□

4)

Faktorerne $F_1, \dots, F_k$ antages at være parvis ort.

Påstand: Ort. proj. på $L_{F_1} + \dots + L_{F_k}$ er givet ved
 $P = I - (I - P_{F_1})(I - P_{F_2}) \cdot \dots \cdot (I - P_{F_k})$

Beris:

$$L_{F_1} + \dots + L_{F_k} = (L_{F_1}^c \cap \dots \cap L_{F_k}^c)^c$$

Den ort. proj. på $L_{F_1}^c \cap \dots \cap L_{F_k}^c$ er jvf. 2) + 3)

$(I - P_{F_1}) \cdot \dots \cdot (I - P_{F_k})$ og dermed flg. det ønskede

□

opgave 10.9

$F: F_1; G: G_1$ betegner faktorer $F \subseteq F_1, G \subseteq G_1$

1)

Påstand: $F_1 \perp G_1$ og $F_1 \wedge G_1 = F \wedge G \implies F \perp G$

Beweis:

Betingelserne giver, at $L_F \subseteq L_{F_1}, L_G \subseteq L_{G_1}$

og $L_F \cap L_G = L_{F_1} \cap L_{G_1}$

samt
$$\begin{array}{ccc} L_{F_1} \cap (L_{F_1} \cap L_{G_1})^\perp & \perp & L_{G_1} \cap (L_{F_1} \cap L_{G_1})^\perp \\ \cup & & \cup \\ L_F \cap (L_F \cap L_G)^\perp & & L_G \cap (L_F \cap L_G)^\perp \end{array}$$

Dermed jvf. def. på geo. ort. s. 104 at $F \perp G$ □

2)

Påstand: Hvis $F_1 \times G_1$ er balanceret, så er $F \perp G$ og $F \wedge G = 0$

Beweis:

$F_1 \times G_1$ balanceret $\implies F_1 \wedge G_1 = 0$ jvf. s. 102

og jvf. sætn. 10.9 er $F_1 \perp G_1$ (idet F_1, G_1 er balanceret når $F_1 \times G_1$ er balanceret)

$F \wedge G \subseteq F_1 \wedge G_1$, dvs. $F_1 \wedge G_1 = 0$, dermed

følger det ønskede af 1) □

opgave 10.3

1)

Ud fra opg. 10.2 (spg. 2) har vi

$$\mu \in \mathbb{R}, \quad v^2 > 0, \quad -\frac{v^2}{n-1} < \mu < v^2 \quad \text{idet}$$

Σ_i er regulært normalfordelt.

idet $\alpha = \sqrt{n} \mu$, så vi, at $\alpha \in \mathbb{R}$

$$\text{Idet } (\tau^2, \sigma^2) \longrightarrow \left(\frac{\tau^2 - \sigma^2}{n}, \sigma^2 + \frac{\tau^2 - \sigma^2}{n} \right)$$

er en injektiv fkt. $\mathbb{R}_+^2 \longrightarrow \{(v^2, \mu) \mid v^2 > 0, -\frac{v^2}{n-1} < \mu < v^2\}$

og $(\mu, v^2) \longrightarrow (v^2 + (n-1)\mu, v^2 - \mu)$ er en fkt.

over i $\mathbb{R}_+^2$ så vi at variationsområdet for $(\alpha, \sigma^2, \sigma^2)$ er $\mathbb{R} \times \mathbb{R}_+ \times \mathbb{R}_+$

2)

$$L(\alpha, \tau^2, \sigma^2) \stackrel{\substack{\uparrow \\ \text{opg. 10.2}}}{=} \prod_{i=1}^k \frac{1}{\sqrt{2\pi\tau^2}} e^{-\frac{1}{2}(y_{i1}-\alpha)^2/\tau^2} \left(\prod_{j=2}^n \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{y_{ij}^2}{2\sigma^2}} \right)$$

$$= \left(\frac{1}{2\pi\tau^2} \right)^{k/2} \left(\frac{1}{2\pi\sigma^2} \right)^{k(n-1)/2} e^{-\frac{1}{2\tau^2} \sum_{i=1}^k (y_{i1}-\alpha)^2 - \frac{1}{2\sigma^2} \sum_{i=1}^k \sum_{j=2}^n y_{ij}^2}$$

3)

Juf. 2) :

$$L(\alpha, \tau^2, \sigma^2) = \left(\frac{1}{2\pi\tau^2} \right)^{k/2} e^{-\frac{1}{2\tau^2} \sum_{i=1}^k (y_{i1} - \alpha)^2} \\ \cdot \left(\frac{1}{2\pi\sigma^2} \right)^{k(n-1)/2} e^{-\frac{1}{2\sigma^2} \sum_{i=2, \dots, n} y_{ij}^2} = L_1(\alpha, \tau^2) \cdot L_2(\sigma^2)$$

Hvor L_1 er likelihood fkt. for en obs. vektor af $N(\alpha, \tau^2)$ fordelt obs. og L_2 er lik. fkt. for obs. fra $N(0, \sigma^2)$ fordelingen, da variationsområdet for $(\alpha, \tau^2, \sigma^2)$ er $\mathbb{R} \times \mathbb{R}_+ \times \mathbb{R}_+$ jf. 1) følger af teorien om 1. dim. normalfordelte s.v. at

$$\hat{\alpha} = \frac{1}{k} \mathbb{I}_{\cdot 1}$$

$$\hat{\tau}^2 = \frac{1}{k} \sum_{i=1}^k (\mathbb{I}_{i1} - \bar{\mathbb{I}}_{\cdot 1})^2$$

$$\hat{\sigma}^2 = \frac{1}{k(n-1)} \sum_{i=2, \dots, n} \mathbb{I}_{ij}^2$$

4)

Udfr. sætn. 4.1 er det klart, at $\hat{\alpha} \perp \hat{\tau}^2$ da $\mathbb{I}_{ij}$ 'ene er uafh. er $\hat{\sigma}^2$ uafh. af $\hat{\alpha}, \hat{\tau}^2$.
Dvs. $\hat{\alpha}, \hat{\tau}^2, \hat{\sigma}^2$ er uafh.

Juf. sætn. 4.1 (og kap. 3.1) :

$$\hat{\alpha} \sim N\left(\alpha, \frac{\tau^2}{k}\right), \quad k\hat{\tau}^2 \sim \tau^2 \chi^2(k-1)$$

$$\hat{\sigma}^2 \sim \sigma^2 \chi^2(k(n-1)) / k(n-1)$$

5)

Juf. opg. 10.2.3):

$$\begin{aligned}\hat{\mu} &= \frac{1}{\sqrt{n}} \bar{X} = \frac{1}{\sqrt{n}} \bar{Y}_{..} = \frac{1}{\sqrt{n}} \frac{1}{k} \sum_{i=1}^k \sqrt{n} \frac{1}{n} \sum_{j=1}^n X_{ij} \\ &= \frac{1}{k \cdot n} \sum_{i,j} X_{ij} = \bar{X}_{..}\end{aligned}$$

$$\begin{aligned}\hat{\tau}^2 &= \frac{1}{k} \sum_{i=1}^k (\bar{Y}_{i.} - \bar{Y}_{..})^2 \\ &= \frac{1}{k} \sum_{i=1}^k (\sqrt{n} \bar{X}_{i.} - \sqrt{n} \bar{X}_{..})^2 = \frac{1}{k} \sum_{i=1}^k n (\bar{X}_{i.} - \bar{X}_{..})^2\end{aligned}$$

$$\hat{\sigma}^2 = \frac{1}{k(n-1)} \sum_{i=1}^k \sum_{j=2}^n Y_{ij}^2 = \frac{1}{k(n-1)} \sum_{i=1}^k \sum_{j=1}^n (X_{ij} - \bar{X}_{i.})^2$$

6)

Når vi transformerer 'tilbage' til estimatøren for ν^2, n fra (s^2, τ^2) svarer dette til transformationen

$$T(s_1^2, s_2^2) = \begin{pmatrix} s_1^2 + \frac{1}{n} (s_2^2 - s_1^2) \\ \frac{1}{n} (s_2^2 - s_1^2) \end{pmatrix} = \begin{pmatrix} 1 - \frac{1}{n} & \frac{1}{n} \\ -\frac{1}{n} & \frac{1}{n} \end{pmatrix} \begin{pmatrix} s_1^2 \\ s_2^2 \end{pmatrix}$$

$$\|T'\| = \left| \left(1 - \frac{1}{n}\right) \frac{1}{n} - -\frac{1}{n} \frac{1}{n} \right| = \frac{1}{n}$$

Lad $\tilde{\nu}^2, \tilde{n}$ betegne estimatorne for ν^2, n

Juf. transformationsbetragtningen er tætheden for

$(\tilde{\nu}^2, \tilde{n})$ koncentreret på $\{(\tilde{\nu}^2, \tilde{n}) \in \mathbb{R}^2 \mid \tilde{\nu}^2 > 0, \frac{\tilde{\nu}^2}{n-1} < \tilde{n} < \tilde{\nu}^2\}$ med tæthed

$$n \cdot f_{s_1^2}(\tilde{\nu}^2 - \tilde{n}) f_{s_2^2}(\tilde{\nu}^2 + (n-1)\tilde{n})$$

$$= n \frac{1}{\Gamma(k(n-1)/2)} \frac{1}{2}^{k(n-1)/2} (\tilde{y}^2 - \tilde{\mu})^{k(n-1)/2 - 1} e^{-\frac{1}{2}(\tilde{y}^2 - \tilde{\mu})}$$

$$\cdot \frac{1}{\Gamma((k-1)/2)} \frac{1}{2}^{(k-1)/2} (\tilde{y}^2 + (n-1)\tilde{\mu})^{(k-1)/2 - 1} e^{-\frac{1}{2}(\tilde{y}^2 + (n-1)\tilde{\mu})}$$

$$= \frac{n}{\Gamma(\frac{k(n-1)}{2})\Gamma(\frac{k-1}{2})} \left(\frac{1}{2}\right)^{(kn-1)/2} (\tilde{y}^2 - \tilde{\mu})^{k(n-1)/2 - 1} (\tilde{y}^2 + (n-1)\tilde{\mu})^{(k-1)/2 - 1}$$

$$\cdot e^{-\frac{1}{2}(2\tilde{y}^2 + (n-2)\tilde{\mu})}$$

Da $\hat{\mu}$ er uafh. af s_1^2, s_2^2 fås den simultane
tæthed for $(\hat{\mu}, \tilde{y}^2, \tilde{\mu})$ ved at gange tætheden
for $\hat{\mu}$ på ovenstående tæthed

$$\hat{\mu} \sim N\left(\mu, \frac{\sigma^2 - (n-1)\mu}{kn}\right)$$

Opgave 10.11

Der er 3 faktorer

R rækkevirk., S sejlvirk., G gødning

Da de jf. opg. formuleringen kan foretages en additiv struktur af R/S-virk. er vort Design:

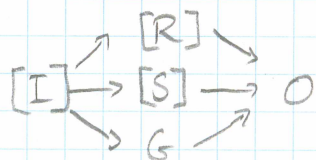
$$D := \{I, R, S, G, O\}$$

$I \times R \times G$ er balanceret (en obs. pr. niveau)

Jf. bn. s. 106 i 10.3.2 er alle faktorer parvis ortogonale, og $R1S \sim R1G \sim S1G \sim 0$, da $I \in D$

ser vi jf. def. 10.10 s. 108, at D er et ortogonalt design.

Faktorstrukturdiagram for D :



ANOVA-tabel:

Faktor	l. l	Frihedsg.	SSD
O	1	1	
R	6	5	54198,58
S	6	5	24467,25
G	6	5	248179,92
I	36	20	30541,00 (se udskrift)

Som det fremgår af faktorstrukturdyagrammet

er $B := \{I, R, S\}$ de tilfældige faktorer og

$T := \{G, O\}$ de systematiske faktorer

Sejvn B ikke opfylder betingelserne på s. 124-125, da

$RA S \sim 0 \notin B$, må vi gå ud fra, at opgaven ikke er

en nonsens, og at vi kan benytte metoderne i kap. 10

Hvis der ikke er forskel på de forskellige niveauer af G , kan denne faktor testes væk

H_0 : Der er ikke forskel på de forskellige niveauer af G

G tilhører stratum I

Jvf. s. 134-135 m. for

$$F = \frac{\sum_{G \in T, T_i} SSD_G / \sum_{G \in T, T_i} d_G}{\sum_{G \in D_{I \setminus T}} SSD_G / \sum_{G \in D_{I \setminus T}} d_G} = \frac{SSD_G / d_G}{SSD_I / d_I}$$

$$= \frac{248179,92 / 5}{30541,00 / 20} = 32,50449 \quad (T_i = \{O\})$$

$$E(X) = 1 - F_{F(5,20)}(F) \leq 1 - 0,9995 = 0,0005 < 0,05$$

Dvs. hypotesen forkastes og vi kan konkludere, at de forskellige niveauer af gødning af forskellig effekt.

H_0 : Der er ikke forskel på niveauet for NxF

NxF tilhører stratum I

Jvf. §134-135m fås

$$F = \frac{\sum_{G \in T|T_1} SSD_G / \sum_{G \in T|T_1} d_G}{\sum_{G \in D_I|T} SSD_G / \sum_{G \in D_I|T} d_G} = \frac{SSD_{NxF} / d_{NxF}}{SSD_I / d_I}$$
$$= \frac{6117,0556 / 2}{30541,0000 / 20} = 2,00290 \quad (T_1 = \{N, F, O\})$$

$$E(X) = 1 - F_{F(2,20)}(F) \geq 1 - 0,9 = 0,1 > 0,05$$

Dvs. vi kan ikke afvise, at der ikke er forskel på effekten af de forskellige niveauer af NxF

Vi reducerer til

$$D := \{I, R, S, N, F, O\}$$

$$B := \{I, R, S\}, \quad T := \{N, F, O\}$$

Dvs. vi har reduceret til en model uden vekselvirkning mellem N og F

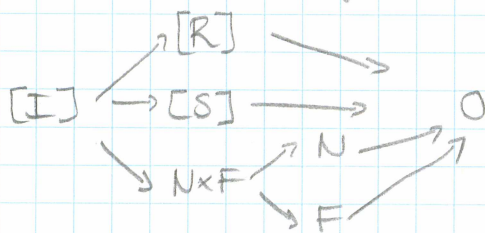
Vi ser på et andet design for eksperimentet, med faktorerne N kvælstof, F fosfat

$$D := \{I, R, S, N \times F, N, F, O\}$$

$$B := \{I, R, S\}, \quad T := \{N \times F, N, F, O\}$$

(som for ses, at D er et ortogonalt design, se også foregående bnr. om B)

Faktorstrukturdiagram:



ANOVA:

Faktor	d	SSD
O	1	
R	5	54198,5833
S	5	24467,25
N	1	77191,3611
F	2	164871,50
N x F	2	6117,0556
I	20	30541,00

En ANOVA tabel for den reducerede model:

Som før, men $d_I = d_{I_{\text{genl.}}} + d_{N \times F}$ og

$$SSD_I = SSD_{I_{\text{genl.}}} + SSD_{N \times F} \quad (\text{jf. (10.24) + (10.26) s. 113})$$

Vi kan, igen jf. s. 134-135, forsøge at teste hh. N og F væk

$$F_N = \frac{SSD_N / d_N}{SSD_I / d_I} = \frac{77191,3611 / 1}{(305411 + 6117,0536) / (20+2)} = 46,32570$$

$$E(x) = 1 - F_{F(1,22)}(F_N) \leq 1 - 0,9995 = 0,0005 < 0,05$$

$$F_F = \frac{SSD_F / d_F}{SSD_I / d_I} = \frac{164871,50 / 2}{(305411 + 6117,0536) / (20+2)} = 49,47307$$

$$E(x) = 1 - F_{F(2,22)}(F_F) \leq 1 - 0,9995 = 0,0005 < 0,05$$

Dus. slutmodellen er

$$N + F + [R + S + I]$$

Jvf. bkm. s. 136n og sætn. 10.17 er
kontrasterne estimerbare

```
OPTIONS PAGENO=1 PS=60 LS=72;
/* På filen opgave_10a.sas findes et program, som kun indlæser data */
```

```
DATA potatoes;
INJ yield dressing $ row $ col $ @@;
DATALINES;
633 E 1 1 527 B 1 2 652 F 1 3 390 A 1 4 504 C 1 5 416 D 1 6
489 B 2 1 475 C 2 2 415 D 2 3 488 E 2 4 571 F 2 5 282 A 2 6
384 A 3 1 481 E 3 2 483 C 3 3 422 B 3 4 334 D 3 5 646 F 3 6
620 F 4 1 448 D 4 2 505 E 4 3 439 C 4 4 323 A 4 5 384 B 4 6
452 D 5 1 432 A 5 2 411 B 5 3 617 F 5 4 594 E 5 5 466 C 5 6
500 C 6 1 505 F 6 2 259 A 6 3 366 D 6 4 326 B 6 5 420 E 6 6
RUN;
```

```
TITLE3 'Fuld model. Denne PROC GLM skal kun give ANOVA tabel';
PROC GLM DATA=potatoes;
  CLASS row col dressing;
  MODEL yield = row col dressing;
RUN; QUIT;
```

```
TITLE3 'RANDOM og TEST tilføjet';
PROC GLM DATA=potatoes;
  CLASS row col dressing;
  MODEL yield = row col dressing;
  RANDOM row col/TEST;
RUN; QUIT;
```

```
DATA potatoes;
SET potatoes;
IF dressing='A' THEN DO P=0; N=0; END;
IF dressing='B' THEN DO P=1; N=0; END;
IF dressing='C' THEN DO P=2; N=0; END;
IF dressing='D' THEN DO P=0; N=1; END;
IF dressing='E' THEN DO P=1; N=1; END;
IF dressing='F' THEN DO P=2; N=1; END;
```

```
QUIT;
PROC PRINT DATA=potatoes;
RUN;
```

```
TITLE3 'Betydning af P og N - fuld model';
PROC GLM DATA=potatoes;
  CLASS row col P N;
  MODEL yield = row col P N P*N;
  RANDOM row col/TEST;
RUN; QUIT;
```

```
TITLE3 'Betydning af P og N - dressing og P*N fjernet';
PROC GLM DATA=potatoes;
  CLASS row col P N;
  MODEL yield = row col P N ;
  RANDOM row col/TEST;
QUIT;
```

```
DATA potatoes;
SET potatoes;
P_cont=P;
RUN;
```



```
TITLE3 'Betydning af P og N - regressionsvirkning af P?';  
PROC GLM DATA=potatoes;  
  CLASS row col N P;  
  MODEL yield = row col N P_cont P;  
  RANDOM row col/TEST;  
RUN; QUIT;  
TITLE3 ' ';
```

```
TITLE3 'Betydning af P og N - slutmodel';  
PROC GLM DATA=potatoes;  
  CLASS row col N P;  
  MODEL yield = row col N P_cont;  
  RANDOM row col/TEST;  
RUN; QUIT;  
TITLE3 ' ';
```

Fuld model. Denne PROC GLM skal kun give ANOVA tabel

General Linear Models Procedure
Class Level Information

Class	Levels	Values
ROW	6	1 2 3 4 5 6
COL	6	1 2 3 4 5 6
DRESSING	6	A B C D E F

Number of observations in data set = 36

Fuld model. Denne PROC GLM skal kun give ANOVA tabel

General Linear Models Procedure

Dependent Variable: YIELD

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	15	326845.75	21789.72	14.27	0.0001
Error	20	<u>30541.00</u>	1527.05		
Corrected Total	35	357386.75			
R-Square		C.V.	Root MSE		YIELD Mean
	0.914544	8.444622	39.077		462.75

Source	DF	Type I SS	Mean Square	F Value	Pr > F
ROW	5	<u>54198.58</u>	10839.72	7.10	0.0006
COL	5	<u>24467.25</u>	4893.45	3.20	0.0276
DRESSING	5	<u>248179.92</u>	49635.98	32.50	0.0001

Source	DF	Type III SS	Mean Square	F Value	Pr > F
ROW	5	54198.58	10839.72	7.10	0.0006
COL	5	24467.25	4893.45	3.20	0.0276
DRESSING	5	248179.92	49635.98	32.50	0.0001

RANDOM og TEST tilføjet

General Linear Models Procedure
Class Level Information

Class	Levels	Values
ROW	6	1 2 3 4 5 6
COL	6	1 2 3 4 5 6
DRESSING	6	A B C D E F

Number of observations in data set = 36

The SAS System

4

RANDOM og TEST tilføjet

General Linear Models Procedure

Dependent Variable: YIELD

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	15	326845.75	21789.72	14.27	0.0001
Error	20	30541.00	1527.05		
Corrected Total	35	357386.75			

R-Square	C.V.	Root MSE	YIELD Mean
0.914544	8.444622	39.077	462.75

Source	DF	Type I SS	Mean Square	F Value	Pr > F
ROW	5	54198.58	10839.72	7.10	0.0006
COL	5	24467.25	4893.45	3.20	0.0276
DRESSING	5	248179.92	49635.98	32.50	0.0001

Source	DF	Type III SS	Mean Square	F Value	Pr > F
ROW	5	54198.58	10839.72	7.10	0.0006
COL	5	24467.25	4893.45	3.20	0.0276
DRESSING	5	248179.92	49635.98	32.50	0.0001

The SAS System

5

The GLM Procedure

Dependent Variable: yield

Source	DF	Sum of Squares	Mean Square	F Value
Model	15	326845.7500	21789.7167	14.27
Error	20	30541.0000	1527.0500	
Corrected Total	35	357386.7500		

Source	Pr > F
Model	<.0001
Error	
Corrected Total	

R-Square	Coeff Var	Root MSE	yield Mean
0.914544	8.444622	39.07749	462.7500

Source	DF	Type I SS	Mean Square	F Value
row	5	54198.5833	10839.7167	7.10
col	5	24467.2500	4893.4500	3.20
P	2	164871.5000	82435.7500	53.98
N	1	77191.3611	77191.3611	50.55
P*N	2	6117.0556	3058.5278	2.00

Source	Pr > F
row	0.0006
col	0.0276
P	<.0001
N	<.0001
P*N	0.1611

Source	DF	Type III SS	Mean Square	F Value
row	5	54198.5833	10839.7167	7.10
col	5	24467.2500	4893.4500	3.20
P	2	164871.5000	82435.7500	53.98

Source	Pr > F
--------	--------

row	0.0006
col	0.0276
P	<.0001