

Meddelelse 16

Forelæsningerne i uge 50

(Onsdag den 11. december) Preben Blæsild talte om det parrede t -test.

(Fredag den 13. december) Afslutning af Kapitel 5 og derefter Kapitel 8 om *Todimensionale normalfordelte data*.

Forelæsningerne i denne uge

(Onsdag den 18. december) Afsnit 8.2 *Estimation* og Afsnit 8.3 *Test for hypotesen $\rho = 0$* .

(Fredag den 20. december) Forelæsningen dækkede Afsnit 8.4-6 og 8.9. Eksempel 8.2 blev brugt til at illustrere anvendelsen af teknikkerne fra Afsnit 8.3 og 8.4: Test af $\rho = 0$, konfidensinterval for korrelationskoefficient, test af identitet af to korrelationskoefficienter, fælles estimat for korrelationskoefficient baseret på $k = 2$ uafhængige stikprøver, konfidensinterval for korrelationskoefficient baseret på $k = 2$ uafhængige stikprøver.

Gamle eksamensopgaver

Tre tidligere eksamenssæt i Statistik α kan findes på adressen

<http://home.imf.au.dk/statbib/atsskurser/statalfa/eksamen>

Der er også relevante opgaver i eksamenssættene i Statistik 1, som findes på adressen

<http://home.imf.au.dk/statbib/atsskurser/stat1/eksamen>

Meddelelse fra Gunnar Hellmund

Forelæsningerne er tilgængelige på adressen

<http://home.imf.au.dk/webcast/>

PENSUM FOR STATISTIK α PÅ NÆSTE SIDE!

Skriftlig eksamen i Statistik α

Eksamen består af én skriftlig prøve. Denne foregår fredag den 31. januar klokken 9.00 til 13.00 på Trøjborg, lokale 139, Niels Juelsgade 84.

Mødetiden er 8.40 og årskort skal medbringes.

Alle sædvanlige hjælpemidler er tilladt, det vil sige man kan medbringe: Statistik Bind 1 og Bind 2, Statistical Tables, meddelelser, egne notater og tidligere regnede opgaver. Indlysende undtagelser er udstyr der larmer (skrivemaskiner o.l.) og mobiltelefoner.

Man *skal* medbringe

- årskort
- skrivetøj
- Statistical Tables
- en lommeregner med et nyt/nyopladet batteri,

Man *bør* medbringe

- Statistik Bind 1 og Bind 2
- meddelelser.

Papir (både til kladde og indskrivning) udleveres af eksamensvagterne.

Den afleverede besvarelse bør være let læseligt skrevet, gerne med blyant. Siderne skal være fortløbende nummeret. På omslaget skal angives navn, årskortnummer, fag og antal afleverede sider. Opgavernes numre bør tydeligt fremgå, så lærer og censor ikke skal lede efter besvarelsen af et bestemt spørgsmål.

Resultatet offentliggøres den 21. februar 2003.

Pensum:

Preben Blæsild og Jørgen Granfeldt: Statistik, Bind 1 og Bind 2, August 2001:

Kapitel 2 - 4

Kapitel 5 fraregnet afsnit 5.1

Kapitel 6 - 7

Kapitel 8 fraregnet følgende: Afsnit 8.7 og 8.8, side 8.20¹⁵-8.21⁹, samt Anneks til Kapitel 8.

Appendiks A1

Opgave 1

I den biologiske forskning har der i de senere år været en betydelig interesse for at udvikle metoder til at beskrive inhomogeniteten af punktmønstre. Disse punktmønstre kan eksempelvis repræsentere udbredelsen af en planteart i et bestemt område.

Følgende undersøgelse blev lavet for at sammenligne to nye metoder til at måle inhomogeniteten. I undersøgelsen indgik 10 punktmønstre med varierende inhomogenitet. For hvert af de 10 punktmønstre blev inhomogeniteten bestemt ved de to metoder.

Punktmønster	1. metode	2. metode	differens
1	9.59	9.56	0.03
2	11.23	10.83	0.40
3	10.66	10.41	0.25
4	11.72	11.90	-0.18
5	10.83	11.04	-0.21
6	9.82	10.36	-0.54
7	11.66	11.89	-0.23
8	11.70	11.81	-0.11
9	10.50	10.36	0.14
10	10.02	10.30	-0.28

I slutningen af opgaven er angivet nogle beregnede størrelser, som kan bruges ved løsningen af opgaven.

- 1) Undersøg, om differensen mellem de to metoders målinger er normalfordelt.
- 2) Undersøg, om der er systematisk forskel på de to metoders målinger.
- 3) Angiv estimat og 95%-konfidensinterval for middelværdien og variansen af differensen mellem de to metoders målinger.

Standardberegninger for differenserne

n	S	USS
10	-0.73	0.7545

STAT1, TD nr. 15

STAT x, S2002 opg. 1

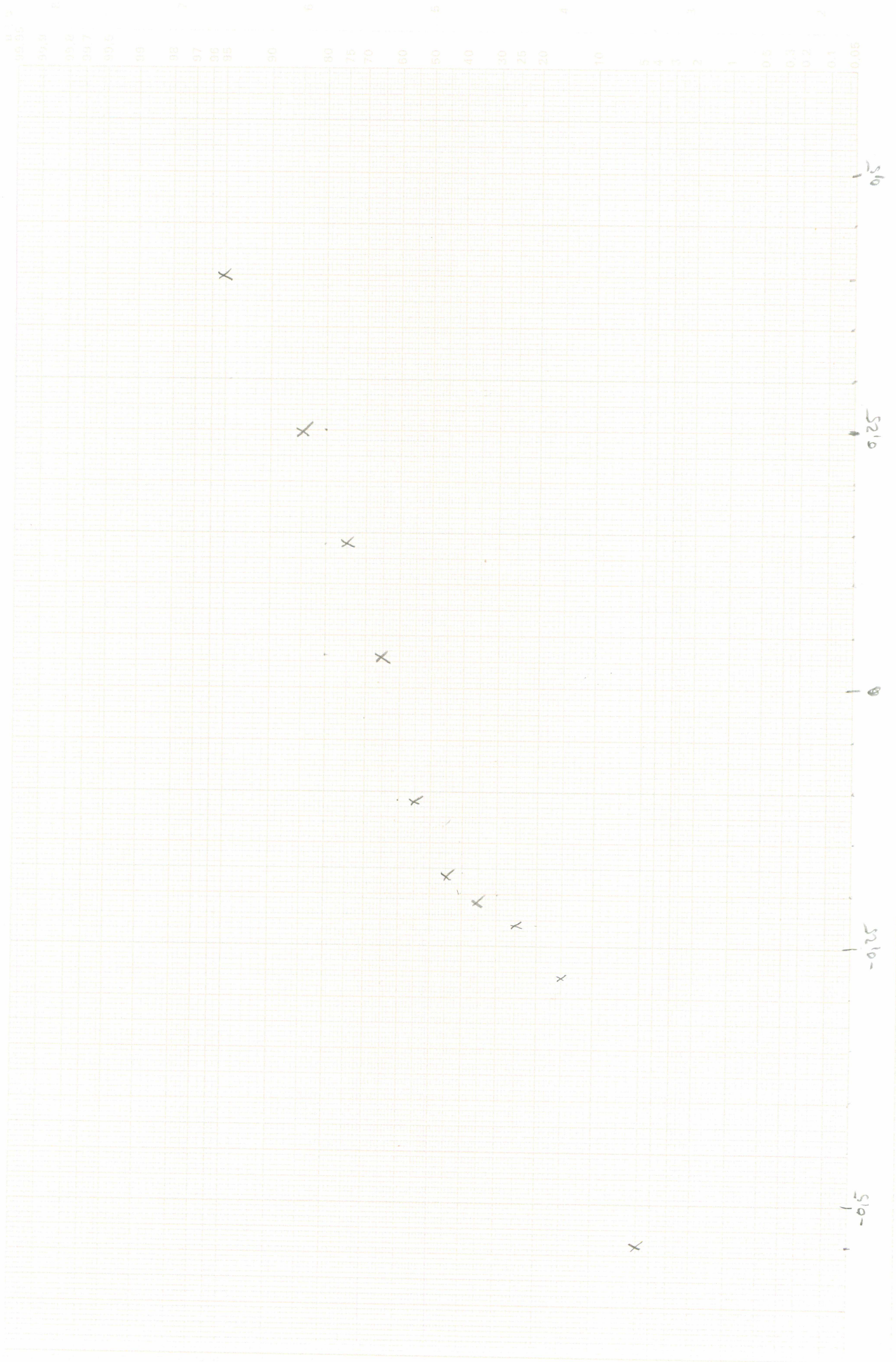
1)

Jf. vedlagte faktildiagram kan vi antage modellen $M_0: x_{i1} - x_{i2} \sim N(\delta, \sigma^2) \quad i=1, \dots, 10$

Til faktildiagrammet benyttede vi

obs.	ak. antal	ss.
-0,54	1	0,05
-0,28	2	0,15
-0,23	3	0,25
-0,21	4	0,35
-0,18	5	0,45
-0,11	6	0,55
0,03	7	0,65
0,14	8	0,75
0,25	9	0,85
0,40	10	0,95

Statix, S2002 opg. 1



2)

Vi kan benytte det pærede t-test når vi vil undersøge om der er systematisk forskel på metode 1 og metode 2, det svarer til at teste

$$H_0: \delta = 0$$

$$t = \frac{s/\sqrt{n}}{\sqrt{(USS - s^2/n)/(n-1) \cdot n}} = -0,827$$

$$E(X) = 2(1 - F_{t(10-1)}(|t|)) \geq 2(1 - 0,8) = 0,4 > 0,05$$

Dvs. vi kan ikke forkaste H_0 og vi dermed konkludere at der ikke er signifikant forskel på de to metodes målinger.

3)

$$\hat{\delta} = \frac{s}{n} = -0,073, \quad s^2 = (USS - s^2/n)/(n-1) = 0,0779$$

95% konf. int. j-f. $(3,18) + (3,12)$

$$\delta \in [-0,272; 0,127] \quad \sigma^2 \in [0,0369; 0,260]$$

Opgave 3.2

Er allerede regnet på den 'hævede' måde.

opgave 8.4

1) Mænd udeladt

198 -													
196 -													
194 -													
192 -													
190 -													
188 -													
186 -													
184 -													
182 -													
180 -						K							
178 -										2K			
176 -													
174 -						K					K		
172 -							K				K		
170 -						K		K					
168 -						K		K	K	K	K	K	
166 -						K				K	K		
164 -	K					2K	K	K					
162 -						K		2K					
160 -	K									K			
158 -	K									K			
156 -										K			
154 -	K												
152 -										K			

47 50 53 56 59 62 65 68 71 74 77 80 83 86

2)

Estimate for mend: (see s.8.5)

$$\hat{\mu}_1 = \sum h_i / 80 = 180,9 \quad \hat{\mu}_2 = \sum v_i / 80 = 72,4$$

$$\sigma_1^2 = \frac{1}{80} SSD_h = \frac{1}{80} (USS_h - s_h^2/80) = 45,69$$

$$\sigma_2^2 = \frac{1}{80} (USS_v - s_v^2/80) = 42,79 \quad n=80$$

$$\hat{\rho} = \frac{SP_{hv}}{\sqrt{SSD_h SSD_v}} = \frac{SP_{hv} - \frac{s_v s_h}{n}}{\sqrt{(USS_h - s_h^2/n)(USS_v - s_v^2/n)}} = 0,3435$$

Estimate for kinder:

$$\sum h_i = 5639 \quad \sum h_i^2 = 935132$$

$$\sum v_i = 1973 \quad \sum v_i^2 = 115897 \quad n=34$$

$$\sum h_i v_i = 328018$$

$$\hat{\mu}_1 = \sum h_i / 34 = 165,71 \quad \hat{\mu}_2 = \sum v_i / 34 = 58,03$$

$$\hat{\sigma}_1^2 = \frac{1}{34} (USS_h - s_h^2/34) = 45,44$$

$$\hat{\sigma}_2^2 = \frac{1}{34} (USS_v - s_v^2/34) = 41,32$$

$$\hat{\rho} = \frac{SP_{hv} - \frac{s_v s_h}{34}}{\sqrt{(USS_h - s_h^2/34)(USS_v - s_v^2/34)}} = 0,7332$$

$\bar{z}_k$ betegner Fishus Z mht. kvinder

$\bar{z}_m$ betegner Fishus Z mht. mændene

Under hypotesen $H_0: p_k = p_m$ gælder

jf. (8.9) at (approximativt):

$$z^* = \frac{\bar{z}_k - \bar{z}_m}{\sqrt{\frac{1}{n_k-3} + \frac{1}{n_m-3}}} \sim N(0,1) \quad (8.9)$$

Dvs.

$$E(X) = 2(1 - \Phi(|z^*|))$$

$$\bar{z}_k = \frac{1}{2} \ln \frac{1 + \hat{p}_k}{1 - \hat{p}_k}, \quad \bar{z}_m = \frac{1}{2} \ln \frac{1 + \hat{p}_m}{1 - \hat{p}_m} \quad (8.7)$$

$$z^* = 2,7152$$

Altså:

$$E(X) \leq 2(1 - 0,9966) = 0,0068 < 0,05$$

Dvs. hypotesen forkastes, korrelationen mellem højde og vægt er ikke den samme for de to køn.

Meddelelse 15

Forelæsningerne i sidste uge

(Onsdag den 4. december) Handlede om resten af Afsnit 7.4.2 om *Den multiplikative poisson-model* og begyndelsen af Afsnit 7.4.1 om *Poissonmodellen med proportionale parametre*, hvor jeg nåede til nederst side 7.14.

(Fredag den 6. december) Kapitel 7 blev afsluttet, og derefter begyndte jeg på Kapitel 5 *Lidt om styrke af forsøg og forsøgsplanlægning*, hvor jeg nåede til midt på side 5.8.

Forelæsningerne i denne uge

(Onsdag den 11. december) Preben Blæsild talte om de parrede t -test. (Fredag den 13. december) Afslutning af Kapitel 5 og derefter Kapitel 8 om *Todimensionale normalfordelte data*.

Forelæsningerne i næste uge

Kapitel 8 om *Todimensionale normalfordelte data*.

Øvelserne i denne uge

(Uge 50: 9. december - 15. december) Eksamensopgave 3 i Statistik 1 Sommeren 2001, hele eksamenssættet fra eksamen i Statistik α Vinteren 2001-2002.

Øvelserne i næste uge

(Uge 51: 16. december - 22. december) Eksamensopgave 1 i Statistik α Sommeren 2002 (tag også stilling til om der er spørgsmål, som ikke er stillet, men som burde besvares), opgave 3.2 (set i lyset af Kapitel 5), Opgave 8.4, samt eventuelle manglende opgaver.

Afleveringsopgaver (Uge 51) Ingen afleveringsopgave.

Udleveret materiale

Intet.

Gamle eksamensopgaver

Tre tidligere eksamenssæt i Statistik α kan findes på adressen

<http://home.imf.au.dk/statbib/atsskurser/statalfa/eksamen>

Der er også relevante opgaver i eksamenssættene i Statistik 1, som findes på adressen

<http://home.imf.au.dk/statbib/atsskurser/stat1/eksamen>

Meddelelse fra Gunnar Hellmund

Forelæsningerne bliver tilgængelige i næste uge på

<http://home.imf.au.dk/webcast/>

Opgave 3

I evalueringsrapporten fra 1998 *Evaluering af de de videregående matematik-, fysik- og kemi-uddannelser* var der en tabel med oplysning om 1. prioritetsansøgningerne på de matematisk-fysisk-kemiske uddannelser i årene 1990 til 1998. Oplysningerne for Aarhus Universitet er gengivet nedenfor i Tabel 6. Ifølge rapporten skulle tabellen illustrere et hovedproblem, nemlig den vigende søgning til uddannelserne.

	1990	1991	1992	1993	1994	1995	1996	1997	1998	I alt
Aarhus Universitet	445	474	436	416	404	323	293	296	287	3374

Tabel 6: Førsteprioritetsansøgninger på de matematisk-fysisk-kemiske uddannelser på Aarhus Universitet.

For at undersøge om udsvingene i antallet af 1. prioritetsansøgninger kan skyldes tilfældigheder, forsøger vi os med en model. Lad X_i betegne antallet af 1. prioritetsansøgninger til år i , og antag at

$$M_0: X_i \sim po(\lambda_i), \quad i = 1990, \dots, 1998,$$

samt at antallet af ansøgninger de forskellige år er uafhængige.

1° Undersøg om data strider mod modellen

$$M_\infty: X_i \sim po(\lambda), \quad i = 1990, \dots, 1998,$$

og kommenter sammenhængen med påstanden om den vigende søgning til uddannelserne.

Nærmere overvejelser viser at data i Tabel 6 ikke umiddelbart er sammenlignelige, fordi antallene fra årene 1990-1994 både er ansøgere med A og med B niveau i matematik, mens antallene fra årene 1995-1998 kun er ansøgere med A niveau. I Tabel 7 er antallet af ansøgere med A niveau for årene 1991-1998, idet det ikke har været muligt at skaffe oplysningerne for 1990. Desuden er der en kolonne med vægtet årgang, som er gennemsnittet af antallet af 19, 20 og 21 årige det pågældende år, som et groft udtryk for størrelsen af den population, som uddannelserne rekrutterer fra.

Lad Y_i betegne antallet af 1. prioritetsansøgninger med A niveau i matematik til år i , og lad m_i betegne Vægtet årgang i år i .

	<i>A niveau</i>	<i>Vægtet årgang</i>
1991	336	77100
1992	351	77200
1993	314	76100
1994	328	75300
1995	323	73300
1996	293	70000
1997	296	66500
1998	287	64100
<i>I alt</i>	2528	579600

Tabel 7: Førsteprioritetsansøgninger med A niveau i matematik på de matematisk-fysisk-kemiske uddannelser på Aarhus Universitet (*A niveau*), og gennemsnittet af antallet af 19, 20 og 21 årige det pågældende år (*Vægtet årgang*).

$Y.$	$\sum_i m_i$	$\sum_i Y_i \ln(Y_i)$	$\sum_i Y_i \ln(m_i)$	$\hat{\alpha} Y.$	$\hat{\beta} \sum_i Y_i \ln(m_i)$
2528	579600	14556.1932	28294.4264	-11186.5824	25741.6009

Tabel 8: Beregnede størrelser.

Vi skal i det følgende benytte grundmodellen at

$$M_0: Y_i \sim po(\lambda_i), \quad i = 1991, \dots, 1998,$$

samt at Y_i 'erne er uafhængige.

I besvarelsen af de følgende spørgsmål må de beregnede størrelser i Tabel 8 benyttes.

2° Test om grundmodellen kan reduceres til

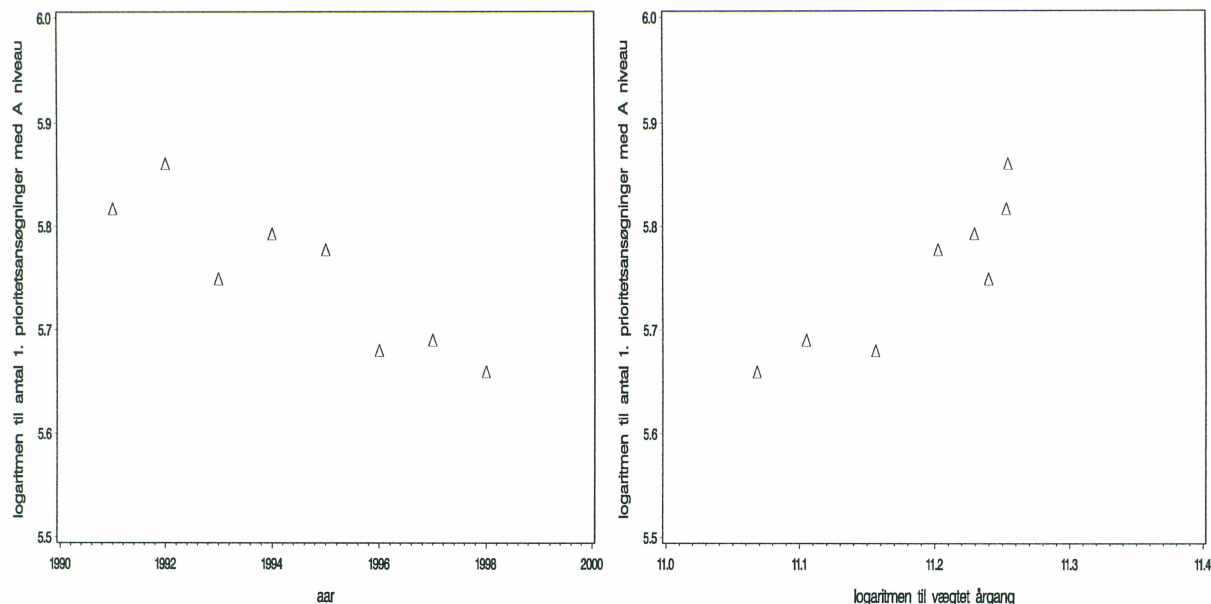
$$M_\infty: Y_i \sim po(\lambda), \quad i = 1991, \dots, 1998,$$

og giv en vurdering af om det er en rimelig beskrivelse af data, jævnfør Figur 5.

Vi vender nu tilbage til grundmodellen M_0 og vil på baggrund af den undersøge modellen

$$M_1: Y_i \sim po(e^{\alpha + \beta t_i}), \quad i = 1991, \dots, 1998,$$

hvor α og β er ukendte parametre og t_i er en forklarende variabel. Både årstallet i og logaritmen til *Vægtet årgang*, $\ln m_i$, vil kunne komme på tale som forklarende variable. Men kun logaritmen til *Vægtet årgang* vil blive inddraget i det følgende.



Figur 5: Logaritmen til antal ansøgere med A niveau mod årstallet og mod logaritmen til Vægtet årgang.

3° Opskriv likelihoodfunktionen og gør rede for at likelihood ligningerne for α og β er

$$\begin{aligned}\sum_i e^{\alpha + \beta t_i} &= \sum_i Y_i \\ \sum_i t_i e^{\alpha + \beta t_i} &= \sum_i t_i Y_i\end{aligned}$$

I det følgende benyttes logaritmen til Vægtet årgang som forklarende variabel, det vil sige $t_i = \ln m_i$. I dette tilfælde er løsningen til likelihood ligningerne i spørgsmål 3°

$$\hat{\alpha} = -4.4251 \quad \hat{\beta} = 0.9098,$$

og disse løsninger maksimerer likelihood funktionen.

4° Gør rede for at data ikke strider mod modellen M_1 med den forklarende variabel $t_i = \ln m_i$.

5° Test om data strider mod en reduktion af modellen fra M_1 til M_∞ .

I modellen M_1 med den forklarende variabel $t_i = \ln m_i$ betragtes hypotesen $H_\beta: \beta = 1$ eller med andre ord modellen

$$M_2: Y_i \sim \text{po}(e^{\alpha + t_i}) = \text{po}(e^\alpha m_i), \quad i = 1991, \dots, 1998,$$

6° Find maksimum likelihood estimatet for α i modellen M_2 og gør rede for at data ikke strider mod reduktionen til modellen M_2 .

STAT1, TØ nr. 14

S2001, Stat1 opgave 3

1)

$$H_0: X_i \sim \text{po}(\lambda)$$

$$\begin{aligned} -2 \ln Q(x) &= 2 \sum_i x_i \ln(x_i/e_i) = 2 \sum_i x_i \ln(10 \cdot x_i / x_{\cdot}) \\ &\quad \begin{matrix} (7,30) \\ 5,404 \\ 82 \end{matrix} \quad \begin{matrix} (7,29) \\ 5,404 \\ 82 \end{matrix} \\ &= 830,74 \end{aligned}$$

$$E(x) = 1 - F_{\chi^2(q)}(-2 \ln Q(x)) < 0,05 \quad (\text{da } e_i \geq 5)$$

$\begin{matrix} (7,31) \\ 5,404 \\ 82 \end{matrix}$

Dvs. vi forkaster hypotesen.

Umiddelbart ser det altså ud som om der er rigende segning (men dette dækker måske over demografikens udsving, eller fejl i materialet)

2)

$$H_0: \Xi_i \sim \text{po}(\lambda)$$

$$-2 \ln Q(x) = 2 \sum_i y_i \ln(y_i / e_i) = 2 \left(\sum_i y_i \ln(9 y_i / y_{\cdot}) \right)$$

$\begin{matrix} (7.31) \\ 5.404 \\ 82 \end{matrix}$
 $\begin{matrix} (7.29) \\ 5.704 \\ 82 \end{matrix}$

$$= 606,86$$

$$E(x) = 1 - F_{\chi^2(z)}(-2 \ln Q(x)) < 0,05 \quad (\text{da } e_i \geq 5)$$

$\begin{matrix} (7.31) \\ 5.404 \\ 82 \end{matrix}$

Dvs. vi forkaster hypotesen, hvilket også er rimeligt set i forhold til fig. 5:

Der tyder på at være en lin. sammenhæng mellem $\log_2$ ti antal ansejere og årgang, hvis H_0 kunne accepteres ville data fordelte sig omkring en vandret linie.

Den anden graf i fig. 5 peger på at et stigende antal ansejere kan forklares ved mindre årgange.

3)

Under M_1 :

$$L(\alpha, \beta) = \prod_{i=1}^8 e^{-e^{\alpha+\beta t_i}} \frac{(e^{\alpha+\beta t_i})^{y_i}}{y_i!}$$

$$\ell(\alpha, \beta) = -\sum_{i=1}^8 e^{\alpha+\beta t_i} + \sum_{i=1}^8 y_i (\alpha+\beta t_i) - \sum_{i=1}^8 \ln(y_i!)$$

$$\frac{\partial \ell}{\partial \alpha}(\alpha, \beta) = -\sum_{i=1}^8 e^{\alpha+\beta t_i} + \sum_{i=1}^8 y_i$$

$$\frac{\partial \ell}{\partial \beta}(\alpha, \beta) = -\sum_{i=1}^8 t_i e^{\alpha+\beta t_i} + \sum_{i=1}^8 y_i t_i$$

$$\frac{\partial \ell}{\partial \alpha} = \frac{\partial \ell}{\partial \beta} = 0 \iff \sum_{i=1}^8 y_i = \sum_{i=1}^8 e^{\alpha+\beta t_i} \wedge \sum_{i=1}^8 y_i t_i = \sum_{i=1}^8 t_i e^{\alpha+\beta t_i}$$

4)

De forventede antal under M_1 ($e^{\hat{\alpha}+\hat{\beta}t_i}$) er ≥ 5 (klart)

$$-2\ln Q(x) = 2(\hat{\ell}_0 - \hat{\ell}_1)$$

$$= 2(-y_0 + \sum_i y_i \ln y_i - \sum_i \ln(y_i!)) - (-\sum_i e^{\hat{\alpha}+\hat{\beta}t_i} + \sum_i y_i (\hat{\alpha}+\hat{\beta}t_i) - \sum_i \ln(y_i!))$$

$$= 2(\sum_i y_i \ln y_i - \hat{\beta} \sum_i y_i \ln m_i - \hat{\alpha} \sum_i y_i) = 2,3494$$

3)

$$E(x) = 1 - F_{\chi^2}(8-2) (-2\ln Q(x)) \geq 1 - 0,2 = 0,8 > 0,05$$

Dvs. vi kan ikke forkaste hypotesen om at reducere til M_1 ,
altså stude data ikke mod modellen M_1 .

5)

Under M_0 : $L(\lambda) = e^{-8\lambda} \lambda^{y \cdot} \prod_i (y_i!)^{-1}$

$\hat{\lambda} = \frac{y \cdot}{8} = \bar{y}$ jvf. setn. 7.1

$-2 \ln Q(x) = 2(\hat{\lambda} - \hat{\lambda}_0)$

$= 2(-\sum_i y_i + \sum_i y_i (\hat{\alpha} + \beta t_i) - \sum_i \ln(y_i!))$
 $- (-y \cdot + y \cdot \ln(\bar{y}) - \sum_i \ln(y_i!))$

$= 2(\hat{\alpha} \sum_i y_i + \beta \sum_i y_i \ln m_i - y \cdot \ln(\bar{y})) = 9,004$

$\mathcal{E}(x) = 1 - F_{\chi^2}(2-1)(-2 \ln Q(x)) \leq 1 - 0,995 = 0,005 < 0,05$

Dvs. vi kan ikke reducere til modellen M_0 fra M_1 .

6)

Under M_2 : $L(\alpha) = \prod_i e^{-e^{\alpha+t_i}} \frac{(e^{\alpha+t_i})^{y_i}}{y_i!}$

$l(\alpha) = \sum_i -e^{\alpha+t_i} + \sum_i y_i (\alpha+t_i) - \sum_i \ln(y_i!)$

$\frac{\partial l}{\partial \alpha}(\alpha) = \sum_i -e^{\alpha+t_i} + \sum_i y_i \quad (*)$

$\frac{\partial l}{\partial \alpha}(\alpha) = 0 \iff \sum_i y_i = \sum_i e^{\alpha+t_i} = e^{\alpha} \sum_i e^{t_i} = e^{\alpha} \sum_i m_i$

$\iff \alpha = \ln(\sum_i y_i / \sum_i m_i) = \ln(y \cdot / m \cdot)$

Jvf. (*) og egenskaber ved exp ses at

$\hat{\alpha} = \ln(y \cdot / m \cdot) = -5,4349$ er et max.

$$-2 \ln Q(x) = 2 (\hat{\ell}_1 - \hat{\ell}_2)$$

$$= 2 (-\sum_i y_i + \sum_i y_i (\hat{\alpha} + \beta t_i) - \sum_i \ln(y_i!))$$

$$- (-\sum_i y_i + \sum_i y_i (\hat{\alpha} + \ln m_i) - \sum_i \ln(y_i!))$$

$$= 2 (\hat{\alpha} \sum_i y_i + \beta \sum_i y_i \ln m_i - \hat{\alpha} \sum_i y_i - \sum_i y_i \ln m_i)$$

$$= 0,0386$$

$$E(\chi) = 1 - F_{\chi^2(2-1)}(-2 \ln Q(x)) \geq 1 - 0,2 = 0,8 > 0,05$$

Dus, H_0 kan worden verworpen ten gunste van M_2

Opgave 1

Mucosa membranen danner afgrænsningen af muskelvævet ud mod mavesækken. I membranen findes et cellelag, der indeholder de såkaldte ECL-celler. Cellelaget kan underopdeles i to dellag, her kaldet A og B . For at undersøge fordelingen af ECL-cellerne er lavet en række vertikale vævssnit gennem membranen. For hvert vævssnit er antal ECL-celler bestemt i lag A og B . Data for 2 sådanne vævssnit er vist i tabellen nedenfor. Alle tællinger er foretaget i områder med lige stort volumen.

	lag A	lag B	Ialt
vævssnit 1	21	9	30
vævssnit 2	24	17	41
Ialt	45	26	71

Idet X_{ij} betegner antal ECL-celler i i 'te vævssnit og j 'te lag, kan det antages at $X_{ij} \sim po(\lambda_{ij})$, $i = 1, 2$, $j = A, B$, samt at alle stokastiske variable er uafhængige.

- 1) Vis, at den multiplikative Poisson model $\lambda_{ij} = \alpha_i \beta_j$ kan beskrive data.
- 2) Vis, at det kan antages indenfor hvert lag, at celleintensiteten er den samme i de to vævssnit, dvs. $\lambda_{ij} = \alpha \beta_j$.
- 3) Undersøg, om celleintensiteten er den samme i lag A og lag B .

Opgave 2

Thyroxin er et hormon, der dannes naturligt i kroppen. Stoffet er en nødvendig faktor for vækst, og mangel forventes at medføre dårlige vækstvilkår. For at belyse sidstnævnte påstand har man bestemt vægtforøgelsen i en 4 ugers periode for rotter inddelt tilfældigt i 3 grupper med henholdsvis 10, 7 og 10 rotter (anden gruppe indeholder færre rotter end de 2 øvrige på grund af et uheld i begyndelsen af eksperimentet). Den første gruppe er en kontrolgruppe, anden gruppe har fået thyroxin i drikkevandet og tredje gruppe har fået thiouracil i drikkevandet. Thiouracil er et præparat, der påstås at modvirke thyroxin. Resultaterne af eksperimentet er vist i tabellen nedenfor.

Kontrol	115, 117, 133, 115, 95, 107, 90, 91, 91, 112
Thyroxin	132, 84, 133, 118, 87, 88, 119
Thiouracil	68, 63, 80, 63, 89, 68, 52, 80, 61, 69

I slutningen af opgaven er angivet nogle beregnede størrelser, som kan bruges ved besvarelsen af opgaven. Det kan antages, at indenfor hver gruppe er vægtforøgelsen normalfordelt.

- 1) Vis, at det kan antages, at variansen af vægtforøgelsen er den samme i de 3 grupper.
- 2) Undersøg, om middelværdien af vægtforøgelsen i thyroxingruppen er den samme som i kontrolgruppen.
- 3) Undersøg, om tilsætning af thiouracil i drikkevandet påvirker væksten.
- 4) Angiv estimater og 95% konfidensintervaller for parametrene i slutmodellen.

Beregnete størrelser

i	n_i	S_i	USS_i	S_i^2/n_i	$SSD_{(i)}$	$f_{(i)}$	$s_{(i)}^2$	$\bar{x}_i$
1	10	1066	115508	113635.6	1872.4	9	208.0	106.6
2	7	761	85567	82731.6	2835.4	6	472.6	108.7
3	10	693	49093	48024.9	1068.1	9	118.7	69.3
Σ	27	2520	250168	244392.1	5775.9	24	240.7	93.3

Opgave 3

Lad $X_{ijk}, i = 1, \dots, r, j = 1, \dots, s, k = 1, \dots, t$, være diskrete stokastiske variable, der følger multinomialmodellen

$$M_0 : \mathbf{X} = \{X_{ijk}\} \sim m(n, \{\pi_{ijk}\}),$$

hvor $\pi_{ijk} > 0$ for alle i, j, k og $\sum_{i=1}^r \sum_{j=1}^s \sum_{k=1}^t \pi_{ijk} = 1$. Modellen M_0 har $rst - 1$ frie parametre. Vi er interesseret i at teste hypotesen, at sandsynlighederne er på formen $\pi_{ijk} = \alpha_{ik}\beta_{jk}$. Med henblik på at estimere under denne hypotese er det bekvemt at formulere hypotesen som

$$H_0 : \pi_{ijk} = \rho_{ik}\sigma_{jk}\tau_k,$$

Opgaven fortsættes

hvor $\rho_{ik} > 0, \sigma_{jk} > 0, \tau_k > 0$ for alle i, j, k , og

$$\begin{aligned}\sum_{i=1}^r \rho_{ik} &= 1, \quad k = 1, \dots, t, \\ \sum_{j=1}^s \sigma_{jk} &= 1, \quad k = 1, \dots, t, \\ \sum_{k=1}^t \tau_k &= 1.\end{aligned}$$

Hypotesen H_0 har $(r + s - 1)t - 1$ frie parametre. Ovenstående skal ikke vises. De følgende spørgsmål skal besvares under antagelse af, at de observerede antal $x_{ijk} > 0$ for alle i, j, k .

1) Vis, at maximum likelihood estimatorerne af parametrene under H_0 er

$$\hat{\rho}_{ik} = \frac{x_{i \cdot k}}{x_{\cdot \cdot k}}, \quad \hat{\sigma}_{jk} = \frac{x_{\cdot jk}}{x_{\cdot \cdot k}}, \quad \hat{\tau}_k = \frac{x_{\cdot \cdot k}}{n},$$

hvor $x_{i \cdot k} = \sum_{j=1}^s x_{ijk}$, $x_{\cdot jk} = \sum_{i=1}^r x_{ijk}$ og $x_{\cdot \cdot k} = \sum_{i=1}^r \sum_{j=1}^s x_{ijk}$.

Vink. Omskriv likelihood funktionen

$$L(\{\rho_{ik}\}, \{\sigma_{jk}\}, \{\tau_k\}) = \frac{n!}{x_{111}! \cdots x_{rst}!} \prod_{i=1}^r \prod_{j=1}^s \prod_{k=1}^t (\rho_{ik} \sigma_{jk} \tau_k)^{x_{ijk}}$$

og anvend derefter Sætning 6.1 flere gange.

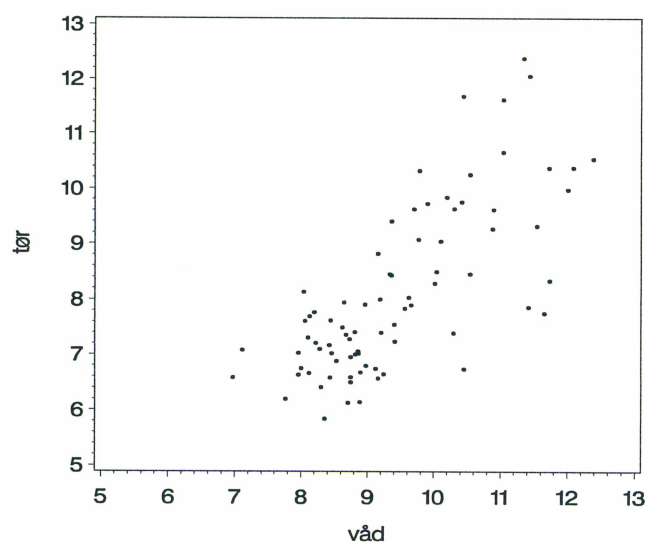
2) Vis, at $-2 \ln Q$ testet for H_0 er på formen

$$-2 \ln Q(\mathbf{x}) = 2 \sum_{i=1}^r \sum_{j=1}^s \sum_{k=1}^t x_{ijk} \ln \left(\frac{x_{ijk}}{x_{i \cdot k} x_{\cdot jk} / x_{\cdot \cdot k}} \right).$$

3) Argumenter for at $-2 \ln Q(\mathbf{x})$ skal vurderes i en $\chi^2((r-1)(s-1)t)$ -fordeling, hvis de forventede antal under H_0 alle er større end eller lig med 5.

Opgave 4

For at sammenligne 2 boremetoder, vådboring og tørboring, har man for hver af metoderne bestemt den tid, det tager at bore 5 feet i 80 forskellige dybder. Lad Y_{ij} betegne boretiden (i minutter) i i 'te dybde ved j 'te metode, $i = 1, \dots, 80$ og $j =$ våd, tør. I figuren nedenfor er sammenhørende boretider $(Y_{i1}, Y_{i2}), i = 1, \dots, 80$, afsat i et koordinatsystem. Tegningen tyder på at boretiden varierer med dybden (jordens beskaffenhed) samt at tørboring er hurtigere end vådboring. Sidstnævnte påstand skal undersøges nøjere i denne opgave.



Figur 1: Våd mod tør.

I forsøgsprotokollen er opgivet, at Y_{ij} er gennemsnit af n_{ij} målinger, hvor n_{ij} enten er 1, 2 eller 3. En naturlig model for boretiderne er derfor

$$M : Y_{ij} \sim N(\alpha_i + \beta_j, \frac{\sigma^2}{n_{ij}}), \quad i = 1, \dots, 80, j = \text{våd, tør},$$

hvor alle Y_{ij} 'erne er uafhængige.

1) Lad

$$t_i = \left(\frac{1}{n_{i1}} + \frac{1}{n_{i2}} \right)^{-1/2}$$

og lad $X_i = t_i(Y_{i1} - Y_{i2})$. Vis, at under modellen M for boretiderne gælder, at

$$X_i \sim N(\alpha + \beta t_i, \sigma^2), \quad i = 1, \dots, 80,$$

hvor $\alpha = 0$ og $\beta = \beta_1 - \beta_2$.

Nedenfor kaldes t_i 'erne for vægte, mens X_i 'erne kaldes for vægtede differenser. Ved besvarelsen af de følgende spørgsmål kan udskrifterne nedenfor fra de 2 små SAS programmer benyttes. SAS datasættet boring indeholder 2 variable: vægt, der indeholder vægtene, og vægtet_dif, der indeholder de vægtede differenser. Programstumpen vises før udskriften. NOINT angiver kørsel under regressionsmodellen med interceptet α lig med 0. En indledende

analyse viser, at en lineær regressionsmodel af vægtede differenser på vægtene

$$X_i \sim N(\alpha + \beta t_i, \sigma^2), \quad i = 1, \dots, 80,$$

kan accepteres.

- 2) Vis, at regressionsliniens intercept α kan antages at være 0.
- 3) Angiv estimat og 95% konfidensinterval for regressionsliniens hældning β .
- 4) Er der signifikant forskel på boretiderne ved de 2 boremetoder?

```
PROC GLM DATA=boring;
MODEL vaegtet_dif = vaegt/SS1;
RUN;
```

3

The GLM Procedure
 Number of observations 80

4

The GLM Procedure					
Dependent Variable: vaegtet_dif vaegtet differens					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	3.4392700	3.4392700	2.33	0.1306
Error	78	114.9072988	1.4731705		
Corrected Total	79	118.3465688			
The GLM Procedure					
R-Square	Coeff Var	Root MSE	vaegtet_dif Mean		
0.029061	78.02281	1.213742	1.555625		
Source	DF	Type I SS	Mean Square	F Value	Pr > F
vaegt	1	3.43926995	3.43926995	2.33	0.1306
Parameter	Estimate	Standard Error	t Value	Pr > t	
Intercept	-0.914885486	1.62257298	-0.56	0.5745	
vaegt	2.165396153	1.41719867	1.53	0.1306	


```
PROC GLM DATA=boring;
MODEL vaegtet_dif = vaegt /NOINT SS1;
RUN;
```

5

The GLM Procedure
 Number of observations 80

6

The GLM Procedure

Dependent Variable: vaegtet_dif vaegtet differens

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	196.5684431	196.5684431	134.59	<.0001
Error	79	115.3756569	1.4604514		
Uncorrected Total	80	311.9441000			

R-Square	Coeff Var	Root MSE	vaegtet_dif Mean
0.630140	77.68526	1.208491	1.555625

NOTE: No intercept term is used: R-square is not corrected for the mean.

Source	DF	Type I SS	Mean Square	F Value	Pr > F
vaegt	1	196.5684431	196.5684431	134.59	<.0001

Parameter	Estimate	Standard Error	t Value	Pr > t
vaegt	1.369110179	0.11801169	11.60	<.0001

STAT α , V2001/2002

opgave 1

1)

$$H_0: \lambda_{ij} = \alpha_i \beta_j$$

Under H_0 :

e	A	B
v_1	19,01	10,99
v_2	25,99	15,01

(7.45)

$$e_{ij} \geq 5$$

$$-2 \ln Q(x) = 2(\hat{\lambda}_0 - \hat{\lambda}_{H_0})$$

$$= 2 \sum x_{ij} \ln(x_{ij}/e_{ij}) = 0,995 \quad ((7.41) + (7.43))$$

$$\underline{E(x)} = 1 - F_{\chi^2_{(2-1)(2-1)}}(-2 \ln Q(x)) \stackrel{(7.51)}{\geq} 1 - 0,7 = 0,3 > 0,05$$

Dvs. hypotesen forkastes ikke og vi reducerer til

$$M_1: x_{ij} \sim \text{po}(\alpha_i \beta_j) \quad i=1,2 \quad j=1,2$$

2)

$$H_{01}: \alpha(B_j) = \alpha(B_j)$$

e	A	B
v1	22,5	13
v2	22,5	13

(7.49)

$$-2 \ln Q(x) = 2(\hat{\ell}_1 - \hat{\ell}_r)$$

$$= 2 \left(\sum_{i=1}^2 x_{i.} \ln(x_{i.}) - x_{..} \ln\left(\frac{x_{..}}{2}\right) \right) = 1,71$$

(7.57)

$$E(x) = 1 - F_{\chi^2}(2-1, -2 \ln Q(x)) \geq 1 - 0,9 = 0,1 > 0,05$$

(7.58)

Da hypotesen forkastes ikke og vi reducerer til

$$M_2^S: x_{ij} \sim \text{po}(\alpha(B_j))$$

3)

Hvis celleintensiteten er den samme i A som i B da vil hypotesen $H_{02}: \alpha(B_j) = \alpha(B_j)$ ikke blive forkastet.

e	A	B
v1	17,75	17,75
v2	17,75	17,75

$$e_{ij} \geq 5 \quad (7.51)$$

$$-2 \ln Q(x) = 2 \left(\sum_{j=1}^2 x_{\cdot j} \ln(x_{\cdot j}) - x_{\cdot \cdot} \ln\left(\frac{x_{\cdot \cdot}}{2}\right) \right)$$

$$= 5,15$$

$$E(x) = 1 - F_{\chi^2(2-1)}(-2 \ln Q(x)) < 1 - 0,975 = 0,025 < 0,05$$

Hypotesen forkastes.

Vi kan konkludere, at cellintensitet er forskellig i lag A og lag B.

opgave 2

$$M_0: x_{ij} \sim N(\mu_i, \sigma_i^2) \quad i=1,2,3 \quad j=1, \dots, n_i$$

1)

$$H_{0\sigma}: \sigma_1^2 = \sigma_2^2 = \sigma_3^2$$

Jvf. formelen s. 3.55:

$$-2 \ln Q(x) = 3,63$$

$$C = 1,0579$$

$$B_a = 3,43$$

$$\varepsilon(x) = 1 - F_{\chi^2(3-1)}(B_a) \geq 1 - 0,9 = 0,1 > 0,05$$

Vi kan derfor reducere til $M_1: x_{ij} \sim N(\mu_i', \sigma^2)$

2)

$$H_{0\mu_2} \quad \mu_1 = \mu_2$$

$$t(x) = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s_1^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} = -0,2747 \quad (3.28)$$

$$\underline{E(x)} = \underset{S. 3.35}{2(1 - F_{t(24)}(|t|))} \geq 2(1 - 0,7) = 0,6 > 0,05$$

Dus. μ_1 reduceren til

$$H_2: X_{ij} \sim N(\mu_i, \sigma^2), \mu_1 = \mu_2 = \mu$$

3)

Huis tils. af thi. ihku prøver udelukke må μ_1 acceptere

$$H_{0\mu_3}: \mu = \mu_3$$

$$t(x) = \frac{(x_1 + x_2) / (n_1 + n_2) - \bar{x}_3}{\sqrt{s_1^2 \left(\frac{1}{n_1 + n_2} + \frac{1}{n_3} \right)}} = 6,1735 \quad (3.29)$$

$$\underline{E(x)} = \underset{S. 3.35}{2(1 - F_{t(24)}(|t|))} \leq 2(1 - 0,9975) = 0,005 < 0,05$$

Hypotesen forkastes, dvs. thi. prøver udelukke

4)

Shitmodellen er M_2

$$\hat{\mu} = \frac{x_{1.} + x_{2.}}{n_1 + n_2} = 103,94 \quad \hat{\mu}_3 = 69,3$$

$$s^2 = 240,7$$

95% konfidensint. (jvf. 3.15 og 3.12)

$$\mu \in [96,17 ; 111,71] \quad \mu_3 \in [59,17 ; 79,43]$$

$$\sigma^2 \in [146,62 ; 465,87]$$

(Bem. usikkerheden på målingerne er stor)

opgave 3

1)

$$L(\underline{\mu}, \underline{\sigma}, \underline{\tau}) = \binom{n}{x_{11} \dots x_{rst}} \left(\prod_{i,k} p_{ik}^{x_{i,k}} \right) \left(\prod_{j,k} \sigma_{jk}^{x_{j,k}} \right) \left(\prod_k \tau_k^{x_{..k}} \right)$$

(1) (2) (3)

Da $p_{..k} = 1$, $\sigma_{..k} = 1$, $\tau_{..} = 1$ følger det enskede direkte af sætn. 6.1 når man ser på faktorerne (1), (2), (3) separat (og deler hver faktor i (1) og (2) i 2 faktorer).

2)

$$\begin{aligned} -2 \ln Q(x) &= 2 \sum_{ijk} x_{ijk} \ln (x_{ijk} / e_{ijk}) \\ &= 2 \sum_{ijk} x_{ijk} \ln (x_{ijk} / (x_{i..} x_{.j.} / x_{..k})) \end{aligned}$$

idet

$$e_{ijk} = \frac{1}{\prod_{ijk} n} = \frac{1}{p_{i..}} \frac{1}{\sigma_{.j.}} \frac{1}{\tau_{..k}} n = \frac{x_{i..}}{x_{..k}} \frac{x_{.j.}}{x_{..k}} \frac{x_{..k}}{n} \cdot n = \frac{x_{i..} x_{.j.}}{x_{..k}}$$

3)

Da H_0 er en glat hypotese (s. 6.52)

følger (jvf. s. 6.56) at $-2 \ln Q(x)$ testet skal

vurderes i en $\chi^2(f)$ forl., hvis $n \geq 5$, hvor

f = frihedsgrad i M_0 - frihedsgrad under H_0

$$= rst - 1 - ((r+s-1)t - 1) = (r-1)(s-1)t$$

opgave 4

1)

Antag M er vor model

Påstand: $\bar{X}_i = t_i (\bar{Y}_{i1} - \bar{Y}_{i2}) \sim N((\beta_1 - \beta_2) t_i, \sigma^2)$

Beweis:

Da $\bar{Y}_{i1} \perp \bar{Y}_{i2}$ følger (SSI regning)

$$t_i(\bar{Y}_{i1} - \bar{Y}_{i2}) \sim N(t_i((\alpha_i + \beta_1) - (\alpha_i + \beta_2)), t_i^2 \sigma^2 (\frac{1}{n_{i1}} + \frac{1}{n_{i2}}))$$

hvoraf det ønsede følger direkte. $\square$

2)

$$H_0: \alpha = 0$$

Ved at betragte modellen som lin. norm. model

fås:

$$F = \frac{\frac{115,3756569 - 114,9072981}{79 - 78}}{1,4731705} = 0,317925$$

$$E(x) = 1 - F_{F(79-78, 78)}(F) \geq 1 - 0,5 = 0,5 > 0,05$$

Hvilket stemmer overens med det alternative
t-test i SAS udskriften s. 4 linie 2 fra neden

Dvs. vi kan antage, at $\alpha=0$.

3)

Juf SAS udskriften s. 6 :

$$\hat{\beta} = 1,369110179$$

og et 95% konf.int. er givet med endepkt.:

$$\text{Estimate} \pm \text{Stand. error} \cdot t_{0,975}(79) \quad (s. 3.94)$$

$$\beta \in [1,1343 ; 1,6040]$$

4)

Hvis der ikke var signifikant forskel på de to boremetoder ville vi kunne godtgøre hypotesen

$H_{0\beta} : \beta = 1$, men denne hypotese må vi juf 3) klart forkaste, idet $0 \notin C_{95\%}(\beta)$

Meddelelse 14

Forelæsningerne i sidste uge

(Onsdag den 27. november) Jeg begyndte på gennemgangen af Kapitel 7 *Poissonfordelte data* og nåede til øverst side 7.7.

(Fredag den 29. november) Jeg fortsatte med Kapitel 7 *Poissonfordelte data* og gennemgik Afsnit 7.3 om *Én (Poissonfordelt) observationsrække*. I Afsnit 7.4 sprang jeg til Afsnit 7.4.2 om *Den multiplikative poissonmodel*, hvor jeg nåede til side 7.26.

Forelæsningerne i denne uge

(Onsdag den 4. december) Handlede om resten af Afsnit 7.4.2 om *Den multiplikative poissonmodel* og begyndelsen af Afsnit 7.4.1 om *Poissonmodellen med proportionale parametre*, hvor jeg nåede til nederst side 7.14.

(Fredag den 6. december) Kapitel 7 afsluttes, og derefter begynder jeg på Kapitel 5 *Lidt om styrke af forsøg og forsøgsplanlægning*.

Forelæsningerne i næste uge

(Onsdag den 11. december og fredag den 13. december) Kapitel 5 og derefter Kapitel 8 om *Todimensionale normalfordelte data*.

Øvelserne i denne uge

(Uge 49: 2. december - 8. december) 7.14, 7.16, 7.17 samt opgaverne 1 og 2, som er stillet på Meddelelse 12.

Øvelserne i de n næste uger

De to sidste uger bruges til repetition samtidig med at der regnes opgaver i tilknytning til kapitel 5 og kapitel 8.

(Uge 50: 9. december - 15. december) Eksamensopgave 3 i Statistik 1 Sommeren 2001, hele eksamenssættet fra eksamen i Statistik α Vinteren 2001-2002.

(Uge 51: 16. december - 22. december) Eksamensopgave 1 i Statistik α Sommeren 2002 (tag også stilling til om der er spørgsmål, som ikke er stillet, men som burde besvares), opgave 3.2 (set i lyset af Kapitel 5), Opgave 8.4, samt eventuelle manglende opgaver.

Afleveringsopgaver (Uge 50) Ingen afleveringsopgave.

(Uge 51) Ingen afleveringsopgave.

Af- og tilbagelevering af 3. obligatoriske opgave

For at få tid til eventuelle genafleveringer er afleveringstidspunktet sat til tirsdag den 10. december klokken 12. Med denne afleveringsfrist har instruktorerne lovet at have opgaverne rettet således at

Morten afleverer dem til øvelserne fredag den 13. december

Jacob afleverer dem til øvelserne torsdag den 12. december

Berit og Jørgen har opgaverne rettet til afhentning fredag den 13. december klokken 13.

Udleveret materiale

Intet.

Gamle eksamensopgaver

Tre tidligere eksamenssæt i Statistik α kan findes på adressen

<http://home.imf.au.dk/statbib/atsskurser/statalfa/eksamen>

Der er også relevante opgaver i eksamenssættene i Statistik 1, som findes på adressen

<http://home.imf.au.dk/statbib/atsskurser/stat1/eksamen>

STAT 1, TØ nr. 13

opgave 7.14

Det er naturligt (jvf. s. 7.6) at antage at antallet af biler som kører over for rødt og antallet af biler som ikke kører over for rødt er poissonfordelt, dvs. vor grundmodel er

$$H_0: X_{ij} \sim \text{po}(\lambda_{ij}) \quad i=1, \dots, 7 \quad j=1, 2 \text{ (1 rødt \& ikke-rødt)}$$

1)

Opgaven svarer til, at vi ikke forkaster hypotesen

$$H_0: \lambda_{ij} = \alpha_i \beta_j, \quad i=1, \dots, 6 \quad j=1, 2$$

e	rødt	ikke-rødt
m	37,34	1301,66
t	35,92	1252,08
o	35,50	1237,50
l	34,95	1218,05
f	36,68	1278,32
1	33,61	1171,39

$$(x_{.1} = 214 \quad x_{.2} = 7459, x_{..} = 7673)$$

Jvf. (7.45)

Vi ser at alle de forventede > 5 .

Juf. (7.41) + (7.43) + (7.45) ses

$$-2 \ln Q(x) = \sum_{ij} x_{ij} \ln \left(\frac{x_{ij}}{e_{ij}} \right) = 3,91$$

$$E(x) = 1 - F_{\chi^2((2-1)(6-1))}(-2 \ln Q(x)) \geq 1 - 0,5 = 0,5 > 0,05$$

Dus. hypotesen forkastes ikke.

z)

Juf. (7.60) og s. 7.6 gælder

$$(\bar{X}_{\cdot 1}, \bar{X}_{\cdot 2} | \bar{X}_{\cdot \cdot} = x_{\cdot \cdot}) \sim m(x_{\cdot \cdot}, (p_1, p_2)^*)$$

Hvis ss. for at køre over fr. vødt er den samme om søndagen som på en hverdag^(*), da gælder ligeledes, under en trivial antagelse om, at $x_{7j} \sim \text{po}(\lambda_{7j})$, at $(\bar{X}_{71}, \bar{X}_{72} | \bar{X}_{7\cdot} = x_{7\cdot})$ (s. 7.6) $\sim m(x_{7\cdot}, (q_1, q_2)^*) \sim m(x_{7\cdot}, (p_1, p_2)^*)$

Dus. hypotesen (*) kan forkastes hvis flg. forkastes:

$$H_0: (p_1, p_2) = (q_1, q_2)$$

	rød	ikke-rød	Σ
hverdag	214	7459	7673
søndag	5	981	986
Σ	219	8440	8659

e	rød	ikke-rød
hver	194,06	7478,94
sen	24,94	961,06

$$-2 \ln Q(x) = 26,26 \quad (6.47)$$

$$E(x) = 1 - F_{\chi^2(1)}(-2 \ln Q(x))$$

$$< 1 - 0,9995 < 0,05$$

Dus. hypotesen forkastes.


```

TITLE1 'Roedkoersler';

OPTIONS PS=60 PAGENO=1;

DATA observa;
  I _JT dag $ overkrt $ tal;
  IF dag='soendag' THEN slagsdag='soendag';
                        ELSE slagsdag='hverdag';

CARDS;
mandag_  roedt      41
mandag_  groent     1298
tirsdag  roedt      26
tirsdag  groent     1262
onsdag_  roedt      39
onsdag_  groent     1234
torsdag  roedt      36
torsdag  groent     1217
fredag_  roedt      37
fredag_  groent     1278
loerdag  roedt      35
loerdag  groent     1170
s _rdag  roedt      5
soendag  groent     981
;

/* Foerst sammenlignes sandsynligheden for roedkoersel
   de forskellige dage. */

DATA hverdag;
SET observa;
IF dag='soendag' THEN delete;
RUN;

/* samme sandsynlighed paa alle hverdage */
PROC FREQ DATA=hverdag ORDER=data;
WEIGHT tal;
TABLES dag*overkrt / EXPECTED
                        CHISQ
                        NOROW
                        NOCOL
                        NOPERCENT;

RUN;

/* samme sandsynlighed paa hverdage og soendag */
PROC FREQ DATA=observa ORDER=data;
WEIGHT tal;
TABLES slagsdag*overkrt / EXPECTED
                        CHISQ
                        NOROW
                        NOCOL
                        NOPERCENT;

RUN;

/* sammenligning mellem alle dage i et skridt */
PROC FREQ DATA=observa ORDER=data;
WEIGHT tal;
TABLES dag*overkrt / EXPECTED
                        CHISQ

```

```
NOROW
NOCOL
NOPERCENT;
```

```
RUN;
```

```
/* Man kan ogsaa sammenligne den totale trafik,
   det spoerges der ikke om i opgaven, men
   det er rimeligt at bruge data til det
   i en multiplikativ poissonmodel */
```

```
/* samme trafik paa alle hverdage */
```

```
PROC FREQ DATA=hverdag ORDER=data;
```

```
WEIGHT tal;
```

```
TABLES dag /          EXPECTED
                      CHISQ
                      NOROW
                      NOCOL
                      NOPERCENT
                      OUT=traf_hvd;
```

```
RUN;
```

```
/* sammenligning mellem alle dage i et skridt */
```

```
PROC FREQ DATA=observa ORDER=data;
```

```
WEIGHT tal;
```

```
TABLES dag /          EXPECTED
                      CHISQ
                      NOROW
                      NOCOL
                      NOPERCENT
                      OUT=traf_al;
```

```
RUN;
```

```
%INCLUDE 'H:\Documents and Settings\Ejer\Skrivebord\EKSAMENSOPGAVER\sasmacros\poisson.mac';
```

```
DATA split;
```

```
ARRAY split {2} c1-c2;
```

```
DO I=1 TO 2;
```

```
    SET observa;
```

```
    split{i}=tal;
```

```
    trafik=c1+c2;
```

```
    roedt=c1;
```

```
    groent=c2;
```

```
END;
```

```
KEEP trafik roedt groent dag;
```

```
RUN;
```

```
PROC PRINT DATA=split;
```

```
RUN;
```

```
%poisson(data=split,obs=trafik);
```

```
%poisson(data=split(where=(dag NE 'soendag')),obs=trafik);
```

The FREQ Procedure

Table of dag by overkrt

dag		overkrt	
Frequency			
Expected	roedt	groent	Total
mandag_	41 37.345	1298 1301.7	1339
tirsdag	26 35.922	1262 1252.1	1288
onsdag_	39 35.504	1234 1237.5	1273
torsdag	36 34.946	1217 1218.1	1253
fredag_	37 36.675	1278 1278.3	1315
loerdag	35 33.607	1170 1171.4	1205
Total	214	7459	7673

Statistics for Table of dag by overkrt

Statistic	DF	Value	Prob
Chi-Square	5	3.6365	0.6028
Likelihood Ratio Chi-Square	5	3.9076	0.5628
Mantel-Haenszel Chi-Square	1	0.1189	0.7303
Phi Coefficient		0.0218	
Contingency Coefficient		0.0218	
Cramer's V		0.0218	

Sample Size = 7673

The FREQ Procedure

Table of slagsdag by overkrt

slagsdag		overkrt	
Frequency			
Expected	roedt	groent	Total
hverdag	214 194.06	7459 7478.9	7673
soendag	5 24.938	981 961.06	986
Total	219	8440	8659

Statistics for Table of slagsdag by overkrt

Statistic	DF	Value	Prob
Chi-Square	1	18.4551	<.0001
Likelihood Ratio Chi-Square	1	26.2512	<.0001
Continuity Adj. Chi-Square	1	17.5411	<.0001
Mantel-Haenszel Chi-Square	1	18.4530	<.0001
Phi Coefficient		0.0462	
Contingency Coefficient		0.0461	
Cramer's V		0.0462	

Fisher's Exact Test

Cell (1,1) Frequency (F)	214
Left-sided Pr <= F	1.0000
Right-sided Pr >= F	4.278E-07
Table Probability (P)	3.552E-07
Two-sided Pr <= P	7.703E-07

Sample Size = 8659

The FREQ Procedure

Table of dag by overkrt

dag		overkrt	
Frequency			
Expected	roedt	groent	Total
mandag_	41 33.865	1298 1305.1	1339
tirsdag	26 32.576	1262 1255.4	1288
onsdag_	39 32.196	1234 1240.8	1273
torsdag	36 31.69	1217 1221.3	1253
fredag_	37 33.258	1278 1281.7	1315
loerdag	35 30.476	1170 1174.5	1205
soendag	5 24.938	981 961.06	986
Total	219	8440	8659

Statistics for Table of dag by overkrt

Statistic	DF	Value	Prob
Chi-Square	6	22.4545	0.0010
Likelihood Ratio Chi-Square	6	30.1589	<.0001
Mantel-Haenszel Chi-Square	1	4.6909	0.0303
Phi Coefficient		0.0509	
Contingency Coefficient		0.0509	
Cramer's V		0.0509	

Sample Size = 8659

The FREQ Procedure

dag	Frequency	Cumulative Frequency
mandag_	1339	1339
tirsdag	1288	2627
onsdag_	1273	3900
torsdag	1253	5153
fredag_	1315	6468
loerdag	1205	7673

Chi-Square Test
for Equal Proportions

Chi-Square	8.7305
DF	5
Pr > ChiSq	0.1203

Sample Size = 7673

The FREQ Procedure

dag	Frequency	Cumulative Frequency
mandag_	1339	1339
tirsdag	1288	2627
onsdag_	1273	3900
torsdag	1253	5153
fredag_	1315	6468
loerdag	1205	7673
soendag	986	8659

Chi-Square Test
for Equal Proportions

Chi-Square	68.4446
DF	6
Pr > ChiSq	<.0001

Sample Size = 8659

Obs	dag	trafik	roedt	groent
1	mandag_	1339	41	1298
2	tirsdag	1288	26	1262
3	onsdag_	1273	39	1234
4	torsdag	1253	36	1217
5	fredag_	1315	37	1278
6	loerdag	1205	35	1170
7	soendag	986	5	981

Macro "POISSON"

Dataset: SPLIT
 Observationer: TRAFIK
 Vaegte: Alle vaegte er sat til 1
 Antal grupper: 7

gruppe	TRAFIK	forventet

1	1339	1237.00
2	1288	1237.00
3	1273	1237.00
4	1253	1237.00
5	1315	1237.00
6	1205	1237.00
7	986	1237.00

sum	8659	8659.00

Test for hypotesen om proportionale parametre:

testtype	stoerrelse	testsandsynlighed

-2lnQ	71.9339	0.0000
chi-i-anden	68.4446	0.0000

Estimater og 95% konfidensinterval for lambda
under hypotesen om proportionale parametre:

estimat	nedre graense	oevre graense

1237.00	1211.2179	1263.3309

Macro "POISSON"

Dataset: SPLIT(WHERE=(DAG NE 'SOENDAG'))
 Observationer: TRAFIK
 Vaegte: Alle vaegte er sat til 1
 Antal grupper: 6

gruppe	TRAFIK	forventet
1	1339	1278.83
2	1288	1278.83
3	1273	1278.83
4	1253	1278.83
5	1315	1278.83
6	1205	1278.83
sum	7673	7673.00

Test for hypotesen om proportionale parametre:

testtype	stoerrelse	testsandsynlighed
-2lnQ	8.7655	0.1188
chi-i-anden	8.7305	0.1203

Estimator og 95% konfidensinterval for lambda
 under hypotesen om proportionale parametre:

estimat	nedre graense	oevre graense
1278.83	1250.5371	1307.7698

opgave 7.16

1)

Det er naturligt (jvf. s. 7.6) at starte med flg. model:

$$H_0: x_{ij} \sim p(\lambda_{ij}) \quad i=1,2 \quad j=1,2,3$$

$$\text{Vi bm. } x_{1.} = 383, \quad x_{2.} = 441, \quad x_{.1} = 453$$

$$x_{.2} = 188, \quad x_{.3} = 183, \quad x_{..} = 824$$

Det er naturligt, at teste...

$$H_0: \lambda_{ij} = \alpha_i \beta_j$$

e	fag.	ufag.	lev.
fer.	210,56	87,38	85,06
eft.	242,44	100,62	97,94

$$\text{Bm } e_{ij} > 5$$

Jvf. (7.41), (7.43) og (7.45) fås

$$-2 \ln Q(x) = 2 \left(\sum_{ij} x_{ij} \ln(x_{ij}/e_{ij}) \right) = 1,19$$

$$P(x) = 1 - F_{\chi^2(2)}(-2 \ln Q(x)) \geq 1 - 0,5 = 0,5 > 0,05 \quad (7.54)$$

Dvs. vi kan reducere til

$$M_1: x_{ij} \sim p_0(\alpha_i \beta_j)$$

Det er nærliggende at teste

$$H_{05}: \alpha_i \beta_j = \alpha \beta_j$$

$$x_{\cdot 1}/2 = 226,5 \quad x_{\cdot 2}/2 = 94 \quad x_{\cdot 3}/2 = 91,5$$

$$(alle > 5)$$

$$-2 \ln Q(x) = 4,09 \quad (7.57)$$

$$E(x) = 1 - F_{\chi^2(1)}(-2 \ln Q(x)) \geq 1 - 0,9 = 0,1 > 0,05 \quad (7.57)$$

Dus, vi reducerer til

$$M_2^S: x_{ij} \sim \text{po}(\alpha \beta_j)$$

Vi teste (for en sikkerheds skyld, selvom det ligner sikker forkastelse)

$$H_{0r}: \alpha \beta_j = \alpha \beta$$

$$x_{1\cdot}/3 = 127,67 \quad x_{2\cdot}/3 = 147 \quad (> 5)$$

$$-2 \ln Q(x) = 672,29 \quad (7.55)$$

$$E(x) = 1 - F_{\chi^2(2)}(-2 \ln Q(x)) < 0,05 \quad (7.56)$$

En fornuftig model er altså M_2^S

2)

Hvis ulikheshyppigheden er forskellig i de 3 grupper
forhæstus

$$H_0: X_{.j} \sim \text{po}(m_j \lambda) \quad m_1 = 2413, m_2 = 988, m_3 = 539$$

$$\text{Vi ser } m_{.} = 3940$$

$$e_1 = x_{.1} \cdot \frac{m_1}{m_{.}} = 504,65$$

$$x_{.1} = 453$$

$$e_2 = 206,63$$

$$(7.29)$$

$$x_{.2} = 188$$

$$e_3 = 112,72$$

$$x_{.3} = 183$$

$$-2 \ln Q(x) = 44,01 \quad (7.30)$$

$$\varepsilon(x) = 1 - F_{\chi^2(2)}(-2 \ln Q(x)) < 0,05 \quad (7.31)$$

Dvs. ulikheshyppigheden er forskellig i de
3 faggrupper.

OPTIONS NODATE PAGENO=1 PS=65;

```
TITLE1 'Opgave 7.16';
DATA opg7_16;
INPUT tid$ gruppe$ antal;
CARDS;
  foer  fagl 203
  foer  ufagl 90
  foer  laerl 90
  efter fagl 250
  efter ufagl 98
  efter laerl 93
;
RUN;
```

```
TITLE2 'spoergsmaal 1';
PROC FREQ DATA=opg7_16 ORDER=data;
WEIGHT antal;
TABLES tid*gruppe/EXPECTED CHISQ NOROW NOCOL NOPERCENT;
RUN;
```

```
DATA opg7_16a;
INPUT skade ansatte;
CARDS;
453 2413
188 988
183 539
;
```

```
TITLE2 'spoergsmaal 2';
%INCLUDE 'H:\Documents and Settings\Ejer\Skrivebord\EKSAMENSOPGAVER\sasmacros\poisson.mac';
%poisson(data=opg7_16a,obs=skade,m=ansatte);
RUN;
```

Opgave 7.16
spoergsmaal 1

1

The FREQ Procedure

Table of tid by gruppe

tid		gruppe			
Frequency Expected		fagl	ufagl	laerl	Total
foer		203	90	90	383
		210.56	87.383	85.059	
efter		250	98	93	441
		242.44	100.62	97.941	
Total		453	188	183	824

Statistics for Table of tid by gruppe

Statistic	DF	Value	Prob
Chi-Square	2	1.1894	0.5517
Likelihood Ratio Chi-Square	2	1.1890	0.5518
Mantel-Haenszel Chi-Square	1	1.1453	0.2845
Phi Coefficient		0.0380	
Contingency Coefficient		0.0380	
Cramer's V		0.0380	

Sample Size = 824

Macro "POISSON"

Dataset: OPG7_16A
Observationer: SKADE
Vaegte: ANSATTE
Antal grupper: 3

gruppe	ANSATTE	SKADE	forventet
1	2413	453	504.65
2	988	188	206.63
3	539	183	112.72
sum	3940	824	824.00

Test for hypotesen om proportionale parametre:

testtype	stoerrelse	testsandsynlighed
-2lnQ	43.9981	0.0000
chi-i-anden	50.7761	0.0000

Estimer og 95% konfidensinterval for lambda
under hypotesen om proportionale parametre:

estimat	nedre graense	oevre graense
0.2091	0.1953	0.2239

opgave 7.17

1)

Hvis der ikke er sammenhæng mellem diag. alder og diag. per. forløb, vi ikke:

$$H_0: x_{ij} \sim \text{po}(\alpha_i \beta_j) \quad i=1,2 \quad j=1,2$$

e	35-44	45-54
35-44	7,72	11,28
45-54	5,28	7,72

(7.45)

$$-2 \ln Q(x) = 2 \sum x_{ij} \ln(x_{ij}/e_{ij}) = 7,63 \quad (7.41) + (7.43)$$

$$\xi(x) = 1 - F_{\chi^2(1)}(-2 \ln Q(x)) < 1 - 0,99 = 0,01 < 0,05 \quad (7.54)$$

Hypotesen forkastes, dvs. vi kan antage at der er sammenhæng mellem diagnose alder og periode.

2)

Hvis der ikke er forskel i sygdomsint. for de 35-44 i de to perioder
gælder

$$H_0: x_{i2} \sim \text{po}(m_i \lambda) \quad , \quad m_1 = 20 \quad m_2 = 5 \frac{1}{2} \quad m. = 25 \frac{1}{2}$$

e	35-44
70-89	15,15
90-96	3,85 < 5

Hvis de forventede antal alle var større/lige 5
ville vi benytte

$$-2 \ln Q(x) = 2 \sum_{i=1}^2 x_{i2} \ln \left(\frac{x_{i2}}{e_{i2}} \right) \quad (7.30)$$

som skal vurderes i en $\chi^2(1)$ -fide. (7.31)

$$\text{hvor} \quad e_{12} = \frac{x_{12} + x_{22}}{m_1 + m_2} m_1 \quad , \quad e_{22} = \frac{x_{12} + x_{22}}{m_1 + m_2} m_2$$

4)

$$H_0: \lambda_{ii} = \gamma_{ii} m_i, \lambda_{iz} = \gamma_{iz} m_i \quad i=1,2 \quad (\gamma_{ii}, \gamma_{iz} > 0 \quad i=1,2)$$

$$L(\gamma_{11}, \gamma_{21}, \gamma_{22}) = e^{-\gamma_{11} m_1} \frac{(\gamma_{11} m_1)^{x_{11}}}{x_{11}!} e^{-\gamma_{21} m_2} \frac{(\gamma_{21} m_2)^{x_{21}}}{x_{21}!} \\ \cdot e^{-\gamma_{22} m_1} \frac{(\gamma_{22} m_1)^{x_{12}}}{x_{12}!} e^{-\gamma_{22} m_2} \frac{(\gamma_{22} m_2)^{x_{22}}}{x_{22}!}$$

$$\text{Vier } \hat{\gamma}_{11} = \frac{x_{11}}{m_1}, \hat{\gamma}_{21} = \frac{x_{21}}{m_2} \quad (\text{siehe 7.1})$$

$$\hat{\gamma}_{22} = \frac{x_{12} + x_{22}}{m_1 + m_2}$$

$$-2 \ln Q(x) = 2(\hat{l}_0 - \hat{l}_a)$$

$$\stackrel{5.7.13}{=} 2 \left(-x_{..} + \sum_{ij} x_{ij} \cdot \ln(x_{ij}) - \sum_{ij} \ln(x_{ij}!) - \right. \\ \left. - (\hat{\gamma}_{11} m_1 + \hat{\gamma}_{21} m_2 + \hat{\gamma}_{22} (m_1 + m_2) + \sum_{ij} x_{ij} \ln(e_{ij}) - \sum_{ij} \ln(x_{ij}!)) \right)$$

$$= 2 \sum_{ij} x_{ij} \cdot \ln(x_{ij}/e_{ij}), \text{ id est}$$

$$e_{11} = \hat{\gamma}_{11} m_1 = x_{11}, \quad e_{21} = \hat{\gamma}_{21} m_2 = x_{21},$$

$$e_{12} = \hat{\gamma}_{22} m_1 = \frac{m_1}{m_1 + m_2} (x_{12} + x_{22}), \quad e_{22} = \hat{\gamma}_{22} m_2 = \frac{m_2}{m_1 + m_2} (x_{12} + x_{22})$$

D.h.

$$-2 \ln Q(x) = 2(x_{12} \ln(x_{12}/e_{12}) + x_{22} \ln(x_{22}/e_{22}))$$

sonn mis $e_{ij} \geq 5$ kann verwendet werden $\chi^2(4-3)$

für at finde $E(x) = 1 - F_{\chi^2(1)}(-2 \ln Q(x))$

$$H_{1b} : \lambda_{ij} = \lambda \alpha_i \beta_j m_i \quad \left(\begin{array}{l} \lambda > 0, \alpha_i m_i > 0, \alpha_1 m_1 + \alpha_2 m_2 = 1 \\ \beta_j > 0, \beta_1 + \beta_2 = 1, i=1,2, j=1,2 \end{array} \right) \quad (*)$$

$$\begin{aligned} L(\lambda, \alpha_1, \alpha_2, \beta_1, \beta_2) &= e^{-\lambda \alpha_1 \beta_1 m_1} \frac{(\lambda \alpha_1 \beta_1 m_1)^{x_{11}}}{x_{11}!} e^{-\lambda \alpha_1 \beta_2 m_1} \frac{(\lambda \alpha_1 \beta_2 m_1)^{x_{12}}}{x_{12}!} \\ &\quad e^{-\lambda \alpha_2 \beta_1 m_2} \frac{(\lambda \alpha_2 \beta_1 m_2)^{x_{21}}}{x_{21}!} e^{-\lambda \alpha_2 \beta_2 m_2} \frac{(\lambda \alpha_2 \beta_2 m_2)^{x_{22}}}{x_{22}!} \\ &= e^{-\lambda((\beta_1 + \beta_2)\alpha_1 m_1 + (\beta_1 + \beta_2)\alpha_2 m_2)} \lambda^{x_{..}} \\ &\quad \cdot \alpha_1^{x_{11} + x_{12}} \alpha_2^{x_{21} + x_{22}} \beta_1^{x_{11} + x_{21}} \beta_2^{x_{12} + x_{22}} m_1^{x_{11} + x_{12}} m_2^{x_{21} + x_{22}} \prod_{ij} \frac{1}{x_{ij}!} \\ &= e^{-\lambda} \lambda^{x_{..}} (\alpha_1 m_1)^{x_{1.}} (\alpha_2 m_2)^{x_{2.}} \beta_1^{x_{.1}} \beta_2^{x_{.2}} \prod_{ij} \frac{1}{x_{ij}!} \end{aligned}$$

Af betingelserne (*) og sætn. 7.1 + 6.1 fås

$$\hat{\lambda} = x_{..}, \quad \hat{\alpha}_1 m_1 = \frac{x_{1.}}{x_{..}}, \quad \hat{\alpha}_2 m_2 = \frac{x_{2.}}{x_{..}}$$

$$\hat{\beta}_1 = \frac{x_{.1}}{x_{..}}, \quad \hat{\beta}_2 = \frac{x_{.2}}{x_{..}}$$

$$\text{Dvs. } \hat{\alpha}_1 = \frac{x_{1.}}{x_{..}} \frac{1}{m_1}, \quad \hat{\alpha}_2 = \frac{x_{2.}}{x_{..}} \frac{1}{m_2}$$

$$-2 \ln Q(x) = 2 (\hat{\ell}_0 - \hat{\ell}_{1b})$$

$$= 2 \left(-x_{..} + \sum_{ij} x_{ij} \ln(x_{ij}) - \sum_{ij} \ln(x_{ij}!) - \right.$$

s. 7.13

$$\left(-x_{..} + x_{..} \ln x_{..} + x_{1.} \ln \left(\frac{x_{1.}}{x_{..}} \right) + x_{2.} \ln \left(\frac{x_{2.}}{x_{..}} \right) \right. \\ \left. + x_{.1} \ln \left(\frac{x_{.1}}{x_{..}} \right) + x_{.2} \ln \left(\frac{x_{.2}}{x_{..}} \right) - \sum_{ij} \ln(x_{ij}!) \right) \Bigg)$$

$$= 2 \left(x_{11} \ln \left(x_{11} \cdot \left(\frac{x_{1.}}{x_{..}} \right)^{-1} \left(\frac{x_{.1}}{x_{..}} \right)^{-1} \cdot x_{..}^{-1} \right) \right. \\ + x_{12} \ln \left(x_{12} \left(\frac{x_{1.}}{x_{..}} \right)^{-1} \left(\frac{x_{.2}}{x_{..}} \right)^{-1} x_{..}^{-1} \right) \\ + x_{21} \ln \left(x_{21} \left(\frac{x_{2.}}{x_{..}} \right)^{-1} \left(\frac{x_{.1}}{x_{..}} \right)^{-1} x_{..}^{-1} \right) \\ \left. + x_{22} \ln \left(x_{22} \left(\frac{x_{2.}}{x_{..}} \right)^{-1} \left(\frac{x_{.2}}{x_{..}} \right)^{-1} x_{..}^{-1} \right) \right)$$

$$= 2 \sum_{ij} x_{ij} \ln(x_{ij}/e_{ij})$$

$$e_{ij} = \hat{\lambda} \hat{\alpha}_i \hat{\beta}_j m_i = x_{..} \frac{x_{i.}}{x_{..}} \frac{x_{.j}}{x_{..}} = \frac{x_{i.} x_{.j}}{x_{..}}$$

Hvis $e_{ij} \geq 5$ vurderes dette i en $\chi^2(4-3)$ ford.
 således: $\varepsilon(x) = 1 - F_{\chi^2(1)}(-2 \ln Q(x))$

(4-3 diff. i antal frie parametre)

$$H_2: \lambda_{ij} = \delta_j m_i, \quad \delta_j > 0 \quad i=1,2 \quad j=1,2$$

$$L(\delta_1, \delta_2) = e^{-\delta_1 m_1} \frac{(\delta_1 m_1)^{x_{11}}}{x_{11}!} e^{-\delta_2 m_1} \frac{(\delta_2 m_1)^{x_{12}}}{x_{12}!} \\ \cdot e^{-\delta_1 m_2} \frac{(\delta_1 m_2)^{x_{21}}}{x_{21}!} e^{-\delta_2 m_2} \frac{(\delta_2 m_2)^{x_{22}}}{x_{22}!}$$

Juf. sætn. 7.1:

$$\hat{\delta}_1 = \frac{x_{11} + x_{21}}{m_1 + m_2}, \quad \hat{\delta}_2 = \frac{x_{12} + x_{22}}{m_1 + m_2}$$

$$e_{ij} = \hat{\delta}_j m_i = \frac{x_{.j}}{m_{.}} m_i$$

Hvis vi tester fra M_{1a} :

$$-2 \ln Q(x) = 2(\hat{\ell}_1 - \hat{\ell}_2)$$

$$= 2(-x_{..} + \sum_{ij} x_{ij} \ln(e_{ij}^1) - \sum_{ij} \ln(x_{ij}!)) \\ - (-x_{..} + \sum_{ij} x_{ij} \ln(e_{ij}^2) - \sum_{ij} \ln(x_{ij}!))$$

$$= 2 \sum_{ij} x_{ij} \ln(e_{ij}^1 / e_{ij}^2), \text{ som skal undersøges}$$

i en $\chi^2(3-2)$ fad:

$$\mathcal{E}(x) = 1 - F_{\chi^2(1)}(-2 \ln Q(x)), \text{ hvis } e_{ij}^2 \geq 5$$

Hvis vi tester fra M_{1b} :

$$-2 \ln Q(x) = 2(\hat{\ell}_{1b} - \hat{\ell}_2)$$

$$= 2 \sum_{ij} x_{ij} \ln(e_{ij}^{1b} / e_{ij}^2) \quad \text{som skal vurderes i en}$$

$$\chi^2(3-2) \text{ - fad: } \varepsilon(x) = 1 - F_{\chi^2(1)}(-2 \ln Q(x))$$

Testet i 1) svarer til at teste

$$H_{01}: \lambda_{ij} = \alpha_i \beta_j$$

$$-2 \ln Q(x) = 2 \sum_{ij} x_{ij} \ln(x_{ij} / e_{ij}) \quad e_{ij} = \frac{x_{i.} x_{.j}}{x_{..}}$$

som skal vurderes i en $\chi^2(4-3)$ - fad.

dvs. testet for H_{01} er ækvivalent med testet for H_{01b}

Testet i 2) svarer til at teste

$$H_{02}: \lambda_{iz} = \lambda_{m_i}$$

$$-2 \ln Q(x) = 2 \sum_{i=1}^2 x_{iz} \ln(x_{iz} / e_{iz}) \quad e_{iz} = \frac{x_{.z}}{m_{.}} m_i$$

som skal vurderes i en $\chi^2(4-3)$ - fad.

dvs. testet er ækvivalent med testet for H_{01a}

Opgave 1 på VS 12

1)

$$\bar{x}_n = \frac{1}{200} 122 = 0,61,$$

$$U_{SS} = 196,$$

$$s^2 = \frac{1}{200-1} (196 - 122^2/200) = 0,610955$$

$$t = \frac{s^2}{\bar{x}_n} = 1,001565 \in [0,81; 1,09] = [\chi^2_{0,025}(199)/199; \chi^2_{0,975}(199)/199]$$

Der. jvf 7.15 (Fiskeris dispersionsindeks) er data poissonfordelt

2)

$$\hat{\lambda} = \bar{x}_n = 0,61 \quad (7.12) \text{ og et 95\% h\u00e5uf. int er:}$$

$$[0,5109; 0,7283] \text{ jvf. (7.21)-(7.22)}$$

opgave 2 på US12

Hvis dødeligheden er større i vest end i øst
 må vi forudsætte flg. hypotese ($M_0: x_{ij} \sim \text{po}(\lambda_{ij})$)
 hvor vi har taget højde for forskellig dødelighed p.g.
 alder og forskellig sygdomsspredning over alders
 grupperne i vest og øst (total antal erstattes af antal
 overlevende):

$H_0:$

$$\lambda_{11} = \alpha_{11} \beta_{11}, \lambda_{12} = \alpha_{11} \beta_{12}, \lambda_{13} = \alpha_{12} \beta_{11}, \lambda_{14} = \alpha_{12} \beta_{12}$$

$$\lambda_{21} = \alpha_{21} \beta_{21}, \lambda_{22} = \alpha_{21} \beta_{22}, \lambda_{23} = \alpha_{22} \beta_{21}, \lambda_{24} = \alpha_{22} \beta_{22}$$

$$\lambda_{31} = \alpha_{31} \beta_{31}, \lambda_{32} = \alpha_{31} \beta_{32}, \lambda_{33} = \alpha_{32} \beta_{31}, \lambda_{34} = \alpha_{32} \beta_{32}$$

$$\beta_{i.} = 1 \quad i=1,2,3$$

$$L(\underline{\alpha}, \underline{\beta}) = e^{-\alpha_{11} \beta_{11}} \frac{(\alpha_{11} \beta_{11})^{x_{11}}}{x_{11}!} e^{-\alpha_{11} \beta_{12}} \frac{(\alpha_{11} \beta_{12})^{x_{12}}}{x_{12}!}$$

$$e^{-\alpha_{12} \beta_{11}} \frac{(\alpha_{12} \beta_{11})^{x_{13}}}{x_{13}!} e^{-\alpha_{12} \beta_{12}} \frac{(\alpha_{12} \beta_{12})^{x_{14}}}{x_{14}!} \dots$$

$$\hat{\alpha}_{11} = x_{11} + x_{12}, \hat{\alpha}_{12} = x_{13} + x_{14}$$

$$\hat{\alpha}_{21} = x_{21} + x_{22}, \hat{\alpha}_{22} = x_{23} + x_{24}$$

$$\hat{\alpha}_{31} = x_{31} + x_{32}, \hat{\alpha}_{32} = x_{33} + x_{34}$$

$$\hat{\beta}_{11} = 1 - \frac{x_{12} + x_{14}}{x_{1.}}$$

$$\hat{\beta}_{21} = 1 - \frac{x_{22} + x_{24}}{x_{2.}}$$

$$\hat{\beta}_{31} = 1 - \frac{x_{32} + x_{34}}{x_{3.}}$$

$$\hat{\beta}_{11} = 0,8732, \hat{\beta}_{21} = 0,8653, \hat{\beta}_{31} = 0,5940$$

$$(\hat{\beta}_{12} = 1 - \hat{\beta}_{11}, \hat{\beta}_{22} = 1 - \hat{\beta}_{21}, \hat{\beta}_{32} = 1 - \hat{\beta}_{31})$$

e	$\overset{v}{lw}$	$\overset{v}{død}$	$\overset{\sigma}{lw}$	$\overset{\sigma}{død}$
<15	94,31	13,69	270,69	39,31
15-39	228,44	35,56	163,54	25,46
>40	222,75	152,25	96,23	65,77

X_{ij}	$\overset{v}{lw}$	$\overset{v}{død}$	$\overset{\sigma}{lw}$	$\overset{\sigma}{død}$
<15	95	13	270	40
15-39	224	40	168	21
>40	218	157	101	61

$$-2 \ln Q(x) = 2(\hat{\ell}_0 - \hat{\ell}_{H_0}) = 2 \sum_{ij} x_{ij} \ln(x_{ij}/e_{ij}) = 2,4616$$

$$E(x) = 1 - F_{\chi^2(2-1)}(-2 \ln Q(x)) \geq 1 - 0,6 = 0,4 > 0,05$$

Dvs. den 'relative' dødelighed er ens i øst og vest

Meddelelse 13

Forelæsningerne i sidste uge

(Onsdag den 20. november) Kapitel 6 *Multinomialfordelte data*. Afsluttede uafhængighed af inddelingskriterier nåede frem til midt på side 6.29 i Afsnit 6.3 *Inferens i flere multinomialfordelinger*.

(Fredag den 22. november) Afsluttede Kapitel 6 *Multinomialfordelte data*.

Forelæsningerne i denne uge

(Onsdag den 27. november) Jeg begyndte på gennemgangen af Kapitel 7 *Poissonfordelte data* og nåede til øverst side 7.7.

(Fredag den 29. november) Jeg fortsætter med Kapitel 7 *Poissonfordelte data* og planlægger at dække Afsnit 7.4.2 *Den multiplikative poissonmodel*.

Forelæsningerne i næste uge

(Onsdag den 4. december og fredag den 6. december) Kapitel 7 afsluttes og derefter gennemgår jeg Kapitel 5 *Lidt om styrke af forsøg og forsøgsplanlægning*

Øvelserne i denne uge

(Uge 48: 25. november - 1. december) 6.11, 7.6, 7.8, opgave 1 og 2 på Meddelelse 11, samt eventuelle manglende opgaver fra tidligere øvelsesgange.

Øvelserne i de n næste uger

(Uge 49: 2. december - 8. december) 7.14, 7.16, 7.17 samt opgaverne 1 og 2, som er stillet på Meddelelse 12 og som gentages på denne meddelelse.

De to sidste uger bruges til repetition samtidig med at der regnes opgaver i tilknytning til kapitel 5 og kapitel 8.

(Uge 50: 9. december - 15. december) Eksamensopgave 3 i Statistik 1 Sommeren 2001, hele eksamenssættet fra eksamen i Statistik α Vinteren 2001-2002.

(Uge 51: 16. december - 22. december) Eksamensopgave 1 i Statistik α Sommeren 2002 (tag også stilling til om der er spørgsmål, som ikke er stillet, men som burde besvares), opgave 3.2 (set i lyset af Kapitel 5), Opgave 8.4, samt eventuelle manglende opgaver.

Afleveringsopgaver (Uge 48) Ingen afleveringsopgave.

(Uge 49) Eksamensopgave 2 i Statistik α Sommeren 2002. Opgavesættet for sommeren 2002 findes på adressen

<http://home.imf.au.dk/statbib/atsskurser/statalfa/eksamen/2002/>

Opgaven er også gengivet sidst på denne meddelelse.

(Uge 50) Ingen afleveringsopgave.

Udleveret materiale

Intet.

Gamle eksamensopgaver

Tre tidligere eksamenssæt i Statistik α kan findes på adressen

<http://home.imf.au.dk/statbib/atsskurser/statalfa/eksamen>

Der er også relevante opgaver i eksamenssættene i Statistik 1, som findes på adressen

<http://home.imf.au.dk/statbib/atsskurser/stat1/eksamen>

Opgave 1

Ved 10 prøjsiske kavalleriregimenter optalte man over 20 på hinanden følgende år (1875-94) antallet af soldater, der omkom som følge af hestespark. Resultatet er gengivet i Tabel 1.

<i>Antal dødsfald</i>	<i>Antal regimentår</i>
0	109
1	65
2	22
3	3
4	1

Tabel 1: Antal døde ved hestespark i 10 prøjsiske regimenter 1875-94.

- 1) Vis, at det kan antages at antallet af dødsfald per år er Poissonfordelt.
- 2) Estimer middelværdien og angiv et 95% konfidensinterval for middelværdien.

Opgave 2

Det foreliggende materiale stammer fra en amerikansk undersøgelse af en art plettyfus *Rocky Mountain Spotted Fever*, der spredes med blodmider som smittebærere.

<i>alder i år</i>	<i>“vestlig” type</i>		<i>“østlig” type</i>	
	<i>totalt antal tilfælde</i>	<i>tilfælde med dødelig udgang</i>	<i>totalt antal tilfælde</i>	<i>tilfælde med dødelig udgang</i>
<i>under 15</i>	108	13	310	40
<i>15-39</i>	264	40	189	21
<i>40 og derover</i>	375	157	162	61
	747	210	661	122

Tabel 2: Totalt antal tilfælde og tilfælde med dødelig udgang af *Rocky Mountain Spotted Fever* af henholdsvis “vestlig” og “østlig” type .

Blandt 747 tilfælde af sygdommen i den vestlige del af USA var der 210 dødsfald, mens der blandt 661 tilfælde i den østlige del af USA var 122 dødsfald. Umiddelbart kunne det derfor se ud som om der var en “vestlig” og en “østlig” type af sygdommer, hvor den “vestlige” var den alvorligste. Imidlertid viser mikrobiologiske undersøgelser ingen forskel mellem de “vestlige” og de “østlige” bakteriestammer.

I Tabel 2 er angivet det totale antal tilfælde og antallet af tilfælde med dødelig udgang i henholdsvis den vestlige og den østlige del af USA delt op i aldersgrupper.

- 1) Undersøg om der på grundlag af dette materiale kan antages at være forskel på den “vestlige” og den “østlige” type plettyfus.

Afleveringsopgave i uge 49

Data stammer fra en undersøgelse af sammenhængen mellem anvendelse af fysisk afstraffelse, politisk tilhørsforhold og oplevelse af fysisk afstraffelse i barndommen. Undersøgelsen er foretaget af Gallup Institutet.

Første del af undersøgelsen omfatter 162 kvinder, der alle har oplevet fysisk afstraffelse i barndommen. I tabellen nedenfor er de opdelt efter om de selv anvender fysisk afstraffelse (ja, nej) og deres politiske tilhørsforhold, der er opgjort som enten socialdemokratiet (*s*), partier, der anses for at være til venstre (*v*) eller til højre (*h*) for dette parti.

	<i>v</i>	<i>s</i>	<i>h</i>	Ialt
ja	12	27	58	97
nej	7	28	30	65
Ialt	19	55	88	162

- 1) Vis, at det kan antages, at der er uafhængighed mellem anvendelse af fysisk afstraffelse og politisk tilhørsforhold.
- 2) Angiv et 95% konfidensinterval for sandsynligheden for at anvende fysisk afstraffelse, når man selv har oplevet fysisk afstraffelse i barndommen.

I anden del af undersøgelsen undersøgte man de samme spørgsmål, men nu for ialt 75 kvinder, der *ikke* havde oplevet fysisk afstraffelse i barndommen. Blandt disse kvinder var der også uafhængighed mellem anvendelse af fysisk afstraffelse og politisk tilhørsforhold.

- 3) Blandt de 75 kvinder, der ikke havde oplevet fysisk afstraffelse i barndommen, anvendte 23 selv fysisk afstraffelse. Undersøg om sandsynligheden for at anvende fysisk afstraffelse afhænger af, om man selv har oplevet fysisk afstraffelse i barndommen.

opgave 6.11

→ se udskrift

Konklusion

Der er ikke nævnt mellem mellem hovedtema

tilføj 1 kriminel / ikke kriminel tilføj 2 kriminel (ikke-kriminel)

```
OPTIONS NODATE PAGENO=1 PS=65;
```

```
DATA opg6_11a;  
INPUT gruppe sex$ type$ antal;  
CARDS;
```

```
1 m mz 200  
1 m dz 420  
1 k mz 306  
1 k dz 552  
2 m mz 81  
2 m dz 154  
2 k mz 18  
2 k dz 38  
3 m mz 50  
3 m dz 37  
3 k mz 4  
3 k dz 3
```

```
;  
RUN;
```

```
PROC FREQ DATA=opg6_11a;  
BY DESCENDING sex;  
RUN;
```

```
TITLE1 'spoergsmaal 1';  
PROC FREQ DATA=opg6_11a ORDER=DATA;  
TABLES gruppe*type/EXPECTED CHISQ NOROW NOCOL NOPERCENT  
EXACT;  
WEIGHT antal;  
BY DESCENDING sex;  
RUN;
```

```
%INCLUDE 'H:\Documents and Settings\Ejer\Skrivebord\EKSAMENSOPGAVER\sasmacros\hw.mac';
```

```
TITLE1 'spoergsmaal 2';  
TITLE2 'dizygote maend';  
%hw(420,154,37);  
RUN;
```

```
TITLE2 'dizygote kvinder';  
%hw(552,38,3);  
RUN;
```

```
TITLE2 'spoergsmaal 3';  
DATA opg6_11b;  
INPUT maend$ kvinder$ antal;  
CARDS;  
- - 1219  
- + 24  
+ - 286  
+ + 16  
;  
RUN;
```

```
PROC FREQ DATA=opg6_11b ORDER=DATA;  
TABLES maend*kvinder/EXPECTED CHISQ NOROW NOCOL NOPERCENT  
EXACT;
```

WEIGHT antal;
RUN;

sex=m

The FREQ Procedure

Table of gruppe by type

Frequency		type	
Expected	mz	dz	Total
1	200 217.86	420 402.14	620
2	81 82.574	154 152.43	235
3	50 30.57	37 56.43	87
Total	331	611	942

Statistics for Table of gruppe by type

Statistic	DF	Value	Prob
Chi-Square	2	21.3421	<.0001
Likelihood Ratio Chi-Square	2	20.2990	<.0001
Mantel-Haenszel Chi-Square	1	15.0296	0.0001
Phi Coefficient		0.1505	
Contingency Coefficient		0.1488	
Cramer's V		0.1505	

Fisher's Exact Test

Table Probability (P)	2.321E-07
Pr <= P	3.711E-05

Sample Size = 942

----- sex=k -----

The FREQ Procedure

Table of gruppe by type

Frequency			
Expected	mz	dz	Total
1	306 305.56	552 552.44	858
2	18 19.944	38 36.056	56
3	4 2.4929	3 4.5071	7
Total	328	593	921

Statistics for Table of gruppe by type

Statistic	DF	Value	Prob
Chi-Square	2	1.7101	0.4253
Likelihood Ratio Chi-Square	2	1.6402	0.4404
Mantel-Haenszel Chi-Square	1	0.0635	0.8011
Phi Coefficient		0.0431	
Contingency Coefficient		0.0431	
Cramer's V		0.0431	

WARNING: 33% of the cells have expected counts less than 5. Chi-Square may not be a valid test.

Fisher's Exact Test

Table Probability (P)	0.0152
Pr <= P	0.3987

Sample Size = 921

Hardy-Weinberg 1

genotype	observeret	forventet	observeret- forventet
AA	420	404.270	15.7300
Aa	154	185.460	-31.4599
aa	37	21.270	15.7300

Hardy-Weinberg 2

n	p	antal A gener	-2lnQ	testss. -2lnQ	chi-i- anden	testss. chi-i- anden
611	0.8134	994	15.779	0.0001	17.581	0.0000

Hardy-Weinberg 1

genotype	observeret	forventet	observeret- forventet
AA	552	549.816	2.18381
Aa	38	42.368	-4.36762
aa	3	0.816	2.18381

Hardy-Weinberg 2

n	p	antal A gener	-2lnQ	testss. -2lnQ	chi-i- anden	testss. chi-i- anden
593	0.9629	1142	3.9179	0.0478	6.3020	0.0121

The FREQ Procedure

Table of maend by kvinder

maend		kvinder	
Frequency			
Expected	-	+	Total
-	1219 1210.8	24 32.181	1243
+	286 294.18	16 7.8188	302
Total	1505	40	1545

Statistics for Table of maend by kvinder

Statistic	DF	Value	Prob
Chi-Square	1	10.9232	0.0009
Likelihood Ratio Chi-Square	1	9.1190	0.0025
Continuity Adj. Chi-Square	1	9.6288	0.0019
Mantel-Haenszel Chi-Square	1	10.9161	0.0010
Phi Coefficient		0.0841	
Contingency Coefficient		0.0838	
Cramer's V		0.0841	

Fisher's Exact Test

Cell (1,1) Frequency (F)	1219
Left-sided Pr <= F	0.9994
Right-sided Pr >= F	0.0020
Table Probability (P)	0.0014
Two-sided Pr <= P	0.0022

Sample Size = 1545

opgave 7.6

$$X_1 \perp\!\!\!\perp X_2, \quad X_i \sim \text{po}(\lambda_i) \quad i=1,2$$

$$\text{Pøststand: } (X_1 | X_1 + X_2 = n) \sim \text{bi}(n, \frac{\lambda_1}{\lambda_1 + \lambda_2})$$

Bewis:

Lad $x \in \{0, 1, \dots, n\}$:

$$\begin{aligned} P(X_1 = x | X_1 + X_2 = n) &:= \frac{P(X_1 = x, X_1 + X_2 = n)}{P(X_1 + X_2 = n)} \\ &= \frac{P(X_1 = x) P(X_2 = n - x)}{P(X_1 + X_2 = n)} \\ &= \frac{e^{-\lambda_1} \frac{\lambda_1^x}{x!} e^{-\lambda_2} \frac{\lambda_2^{n-x}}{(n-x)!}}{e^{-\lambda_1 - \lambda_2} \frac{(\lambda_1 + \lambda_2)^n}{n!}} = \frac{n!}{(n-x)! x!} \left(\frac{\lambda_1}{\lambda_1 + \lambda_2} \right)^x \left(\frac{\lambda_2}{\lambda_1 + \lambda_2} \right)^{n-x} \end{aligned}$$

$X_1 + X_2 \sim \text{po}(\lambda_1 + \lambda_2)$
SSI

□

$$M_0: X_i \sim \text{po}(\lambda_i) \quad i=1,2$$

$$H_0: \lambda_1 = \lambda_2$$

Ved at sette $n=x$, i ovenstående får, at vi under hypotesen har

$$(X_1 | X_1 + X_2 = x) \sim b(x, 1/2)$$

Dvs. log-likelihood er $\ln\left(\binom{x}{x}\right) + x \ln(1/2) + (x-x) \ln(1)$
 $= \ln(x!) - \ln((x-x)!) - \ln(x!) - x \ln(2)$

Under M_0 er $(X_1 | X_1 + X_2 = x) \sim b(x, (\frac{\lambda_1}{\lambda_1 + \lambda_2}))$

dvs. log-likelihood er

$$\ln\left(\binom{x}{x}\right) + x \ln\left(\frac{\lambda_1}{\lambda_1 + \lambda_2}\right) + (x-x) \ln\left(\frac{\lambda_2}{\lambda_1 + \lambda_2}\right)$$

$$= \ln(x!) - \ln((x-x)!) - \ln(x!) + x \ln\left(\frac{\lambda_1}{\lambda_1 + \lambda_2}\right) + (x-x) \ln\left(\frac{\lambda_2}{\lambda_1 + \lambda_2}\right)$$

Maks. like. est. er jvf. sætn. 6.1 hh. $1/2$ og $\frac{x}{x}$,
 dvs. (jvf. kap. 1.7)

$$-2 \ln Q = 2 \left(x \ln 2 + x \ln\left(\frac{x}{x}\right) + (x-x) \ln\left(\frac{x-x}{x}\right) \right)$$

$$\hat{E}(x) = 1 - F_{\chi^2(1-\alpha)}(-2 \ln Q(x))$$

Artiklen indeholder 3 påstande

Antallet af ombodne i oht. 91 er signifikant
 større end i oht. 90 (1)

ombodne i mnd. 1-10 i 91 < # ombodne i mnd. 1-10 i 90 (2)

tilskadekomne i mnd. 1-10 i 91 < # tilskadekomne i mnd. 1-10 i 90 (3)

Ud fra betingelserne på s. 7.6 kan vi antage at

omkomme i okt. 90, 91, med 1-10 90, 91

tilskadekomme med. 1-10 i 90, 91 er po. fordelte og

at # ulykker i 90 $\parallel$ # ulykker i 91

Test for at (1) er falsk:

$$x. = 58 + 53 = 111, \text{ sæt } x = 58$$

$$-2 \ln Q(x) = 0,2253 \quad (\text{jvf. 1. del af opg.})$$

$$E(x) = 1 - F_{\chi^2(1-0)}(-2 \ln Q(x)) \geq 1 - 0,4 = 0,6 > 0,05$$

Der påstand om signifikant forskel er falsk.

(2):

$$x. = 974, \quad x = 472$$

$$-2 \ln Q(x) = 0,9242$$

$$E(x) = 1 - F_{\chi^2(1-0)}(-2 \ln Q(x)) \geq 1 - 0,7 = 0,3 > 0,05$$

Der er altså ikke en signifikant forskel

(3)

$$x. = 16869$$

$$x = 8007$$

$$-2 \ln Q(x) = 43,96$$

$$E(x) = 1 - F_{\chi^2(1-0)}(-2 \ln Q(x)) \leq 1 - 0,995 < 0,05$$

Der er en signifikant forskel mellem antal tilskadekomme i 90 og 91.

opgave 7.8

1)

$$\bar{x} = 0,93229$$

$$s^2 = \frac{1}{n-1} (U_{SS} - \frac{s^2}{n}) = 0,96937$$

$$t = \frac{s^2}{\bar{x}} = 0,96175$$

$$0,8577 = \chi_{0,025}^2 (576-1) / (576-1)$$

$$\leq t \leq 1,1189 = \chi_{0,975}^2 (576-1) / (576-1)$$

Dvs. (jvf. s. 7.8) at vi kan opstille modellen

$$M_0 : x_i \sim \sim po(\lambda)$$

2)

$\bar{x} = 0,93229$ er estimatet for λ og dermed for antal bær pr $1/4 \text{ km}^2$

Jvf. (7.21) + (7.22) er et 95% konfidensint.
idet $u_{0,975} = 1,960$

$$[0,8567 ; 1,0146]$$

c)

Et estimat for ss. for at ingen bombe
faldt i et givet 1/4 km² område ^(*) er

$$e^{-\hat{\lambda}} \frac{\hat{\lambda}^0}{0!} = e^{-\hat{\lambda}} = e^{-\bar{x}} = 0,3937$$

Ved at benytte konf.int. for $\hat{\lambda}$ fås et konf.int. for (*):

$$[0,3625; 0,4246]$$

Opgave på VS II

opgave 1

1) For at teste hypotesen om ens fordeling betragte vi

<u>e</u>	DD	DP	PP	(jvf. (6.46))
1956	334,37	335,49	283,14	
1969	259,63	260,57	219,86	

$$-2 \ln Q(x) = 2,523$$

$$E(x) = 1 - F_{\chi^2((2-1)(3-1))}(-2 \ln Q(x)) \geq 1 - 0,8 = 0,2 > 0,05$$

D.s. vi kan ikke afvise, at fordelingen på hen er identisk i 1956 og 1969.

2)

Lad $(p, r) \in]0, 1[{}^2$

$\pi_1 = \pi_1(p, r)$, $\pi_2 = \pi_2(p, r)$, $\pi_3 = \pi_3(p, r)$

$$\frac{2\pi_1 + \pi_2}{2} = \frac{2rp + 2(1-r)p^2 + 2(1-r)p(1-p)}{2}$$

$$= p$$

$$1 - \frac{2\pi_2}{(2\pi_1 + \pi_2)(2\pi_3 + \pi_2)} = 1 - \frac{\pi_2}{p(2\pi_3 + \pi_2)}$$

$$= 1 - \frac{2(1-r)p(1-p)}{p(2r(1-p) + 2(1-r)(1-p)^2 + 2(r-1)p(1-p))}$$

$$= 1 - \frac{2(1-r)p(1-p)}{p(1-p)(2r + 2(1-r)(1-p) + 2(r-1)p)}$$

$$= 1 - \frac{2(1-r)p(1-p)}{p(1-p)(2r + 2(1-r))} = 1 - (1-r) = r$$

Hvilket viser det ønskede

3)

$$\text{Vi ved } \hat{\pi}_1 = \frac{x_1}{n}, \hat{\pi}_2 = \frac{x_2}{n}, \hat{\pi}_3 = \frac{x_3}{n}$$

Indsættes dette i udtrykket i 2) for hh. p og r fås estimator:

$$\hat{p} = \frac{2\hat{\pi}_1 + \hat{\pi}_2}{2} = \frac{x_1}{n} + \frac{1}{2} \frac{x_2}{n} = \frac{2x_1 + x_2}{2n}$$

$$\hat{r} = 1 - \frac{2\hat{\pi}_2}{(2\hat{\pi}_1 + \hat{\pi}_2)(2\hat{\pi}_1 + \hat{\pi}_2)}$$

$$= 1 - \frac{2n \frac{x_2}{n}}{(2x_1 + x_2) \left(\frac{2x_1}{n} + \frac{x_2}{n} \right)}$$

$$= 1 - \frac{2nx_2}{(2x_1 + x_2)(2x_1 + x_2)}$$

Indsættes tallene fra 1) fås

$$\hat{p} = 0,526875 \quad \hat{r} = 0,293884$$

4)

Under antagelsen om at $p = 1/2$ bliver

$$\pi_1^{x_1} \pi_2^{x_2} \pi_3^{x_3} = \left(\frac{1}{4}r + \frac{1}{4}\right)^{x_1+x_3} \left(\frac{1}{2} - \frac{1}{2}r\right)^{x_2}$$

Juf. sætn. 6.1 har vi, at

$\left(\frac{1}{2}r + \frac{1}{2}\right)^{x_1+x_3} \left(\frac{1}{2} - \frac{1}{2}r\right)^{x_2}$ har maks. i

$$\frac{1}{2}r + \frac{1}{2} = \frac{x_1+x_3}{n}, \quad \frac{1}{2} - \frac{1}{2}r = \frac{x_2}{n}, \text{ dvs.}$$

$$r = \frac{x_1+x_3 - x_2}{n} \quad \text{hvilket også er maks.}$$

like, est., dvs.

$$\hat{r} = \frac{x_1+x_3 - x_2}{n} = 0,295924$$

Vort udgangspkt. er modellen

$$M_1: (X_1, X_2, X_3)^T \sim m(1693, (\pi_1(p, r), \pi_2(p, r), \pi_3(p, r)))$$

$$H_0: p = 1/2$$

$$-2 \ln Q(x) = 2 \sum_{i=1}^3 x_i \ln \left(\frac{\hat{\pi}_i^1}{\hat{\pi}_i''} \right) \quad (\text{se evt. s. 6.9})$$

Hvor $\hat{\pi}_i^1, \hat{\pi}_i''$ er maks. like, est. i hh. M_1 og under H_0

$$\pi_1^1 = \frac{x_1}{n}, \quad \pi_2^1 = \frac{x_2}{n}, \quad \pi_3^1 = \frac{x_3}{n}$$

$$\pi_1'' = \pi_3'' = \frac{1}{4}\hat{r} + \frac{1}{4}, \quad \pi_2'' = \frac{1}{2} - \frac{1}{2}\hat{r}$$

$$-2 \ln Q(x) = 7,5575$$

$$E(x) = 1 - F_{\chi^2(2-1)}(-2 \ln Q(x)) \leq 1 - 0,995 < 0,05$$

Altså fihaastus H_0 .

Opagave 2 (US 11)

1)

$$H_0: (X_1, \dots, X_{12})^* \sim m(59136, (\pi_1, \dots, \pi_{12})^*)$$

$$H_0: \pi_i = \frac{\text{antal dage i m\dd{a}neden}}{\text{antal dage i \dd{a}ret}}$$

\dd{A}ret var 1988 som havde 29 dage i feb.,
dvs. 366 i alt.

N\dd{a}r d\dd{e}df\dd{o}lte medregnes bliver f\dd{o}delstallene

$$\begin{array}{ccccccc} 4531, & 4498, & 5220, & 5079, & 5193, & 5012, & 5187 \\ 5312, & 5199, & 4738, & 4497, & 4620 & & (\varepsilon = 59136) \end{array}$$

og de forventede under hypotesen

$$5008,79 ; 4685,64 ; 5008,79 ; 4847,21$$

osv.

$$-2 \ln Q(x) = 218,74 \quad (\text{se 6.16})$$

$$\varepsilon(x) = 1 - F_{\chi^2(11-0)}(-2 \ln Q(x)) < 0,05$$