

Meddelelse 5

Forelæsningerne i sidste uge

(Onsdag den 18. september) Resumé af testene i en normalfordelt observationsrække. Gennemgang af konfidensintervaller for middelværdi og varians i en normalfordelt observationsrække.

(Fredag den 20. september) Afsnit 3.1 afsluttes med en gennemgang af Afsnit 3.1.4 *Likelihood metoden*. Jeg nåede dog kun at omtale tilfælde I. *Variansen kendt*: $\sigma^2 = \sigma_0^2$. Derefter begyndte jeg på Afsnit 3.2 *Flere observationsrækker* og nåede til øverst på side 3.33.

Forelæsningerne i denne uge

(Onsdag den 25. september) Forsatte gennemgangen af Afsnit 3.2.1 *To observationsrækker* og nåede til nederst side 3.36 i gennemgangen af *to observationsrækker med varainshomogenitet*.

(Fredag den 27. september) Jeg afslutter gennemgangen af Afsnit 3.2.1 *To observationsrækker* og når måske at begynde på Afsnit 3.2.2 *Flere observationsrækker*.

Forelæsningerne i næste uge

(Onsdag den 2. oktober og fredag den 4. oktober) Afslutter Afsnit 3.2.2 *k observationsrækker*. Næste punkt er Afsnit 3.3 *Lineær regression*.

Øvelserne i denne uge

(Uge 39: 23. september-29. september) Omfatter som minimum opgaverne 3.2, 3.3 og 3.4. Hvis der er mere tid kan den bruges i EDB lokalet, for eksempel med at regne opgaverne 3.2 og 3.3 ved hjælp af SAS.

Øvelserne i de to næste uger

(Uge 40: 30. september-6. oktober) Opgave 3.5, 3.6 (afleveringsopgaven) regnes færdig, 3.7, 3.8 samt opgaven sidst på denne meddelelse. (Bemærk at programmet er ændret i forhold til programmet for uge 40 på meddelelse 4.)

(Uge 41: 7. oktober-13. oktober) Opgave 3.9, 3.17 (men spring over spørgsmål 3) og besvar spørgsmål 4) alene med metoden fra Afsnit 3.2.2) 3.19, 3.13.

Afleveringsopgaver

(Uge 40: 30. september-6. oktober) Opgave 3.6 1), 2) og 3).

(Uge 41) Følgende opgave:

I tabellen nedenfor findes konditallene for to grupper fodboldspillere i samme klub. Den ene gruppe spiller i serie 6 og den anden spiller i serie 1.

	<i>kondital</i>									
<i>serie 6</i>	56.0	61.8	64.1	64.9	65.2	66.3	66.9	68.9	70.6	70.8
	72.4	73.1	73.9	74.1	74.1	75.3	78.7			
<i>serie 1</i>	68.9	75.2	74.3	72.9	72.0	63.9	76.3	76.3	75.4	66.0
	68.4	64.1	73.1	64.9	73.4	76.2	79.4	69.4	79.8	72.1

Analyser data med henblik på om der er forskel på konditallene for de to grupper.

Et SAS program, der indholder data, kan ses i en browser på adressen

<http://home.imf.au.dk/statbib/atkskurser/stat1/meddelelser/aktuelle/afl41.sas>

(Uge 42) Ingen afleveringsopgave. Efterårsferie!!!

Opgave til øvelserne i uge 40

Institut for Idræt ved Københavns Universitet arrangerer hvert år et atletikstævne for de 1. års studerende. Atletiklærer Mikkel Sørensen har stillet nogle af resultaterne for årene 1998 og 1999 til rådighed. Det er af interesse at vide om resultaterne for de to år er forskellige eller ej. Resultaterne af pigernes længdespring er vist i tabellen nedenfor.

<i>år</i>	<i>længde i m</i>									
1998	3.72	3.65	3.90	3.74	3.32	4.22	3.58	4.56	3.65	2.99
	3.91	3.65	3.65	3.88	3.65	4.20				
1999	4.32	3.79	3.53	3.54	4.27	3.75	4.21	3.66	4.58	3.73
	5.18	3.00	3.91	3.52						

Standardberegninger findes i tabellen nedenfor.

	<i>n</i>	<i>S</i>	<i>USS</i>
1998	16	60.27	229.0119
1999	14	54.99	219.8343

STAT1, TØ nr. 4

opg. 3.2

obs.	antal	kumulerede antal	ss (til fraktildiag.)
-0,2	1	1	$1/2 \cdot 9 \approx 0,056$
0,1	1	2	$2+1/2 \cdot 9 \approx 0,167$
0,2	2	4	$2+1/2 \cdot 9 \approx 0,333$
0,3	2	6	$4+6/2 \cdot 9 \approx 0,556$
0,6	2	8	$6+8/2 \cdot 9 \approx 0,778$
0,7	1	9	$8+9/2 \cdot 9 \approx 0,944$

1)

differeansen kan jvf. vedlagte fraktildiagram antages at være normalfordelt, dvs. vi har flg. model for diff.:

$$H_0: X_i \sim N(\mu, \sigma^2)$$

2)

Umiddelbart kunne man mene, at vi skulle betragte forsøget som to obs. rækker, men så ville vi smide information om de direkte sammenlign. af målemetoderne ræk (målinger på samme klippestykker med to forsk. metoder).
 Altså er det (med formål at sammenligne metoder) smartest at betragte forsøget som en obs. række hvor obs. er diff

Om metoden er ens svar til at teste

$$H_0: \mu = 0$$

teststatistikken er (skal vurderes i en $t(9-1)$ fordeling) (se s. 3.29)

$$t(x) = \frac{\bar{x}_1 - 0}{\sqrt{s^2/n}} = \frac{S/n}{\sqrt{(USS - S^2/n)/(n \cdot (n-1))}} \approx 3,277$$

$$E(x) = 2(1 - F_{t(8)}(3,277)) \leq 2(1 - 0,99) = 0,02 < 0,05$$

Derfor hypotesen forkastes: de to metoder giver ikke samme resultat

3)

Jvf s. 3.30 er et konfidensint. (95%) for diff.

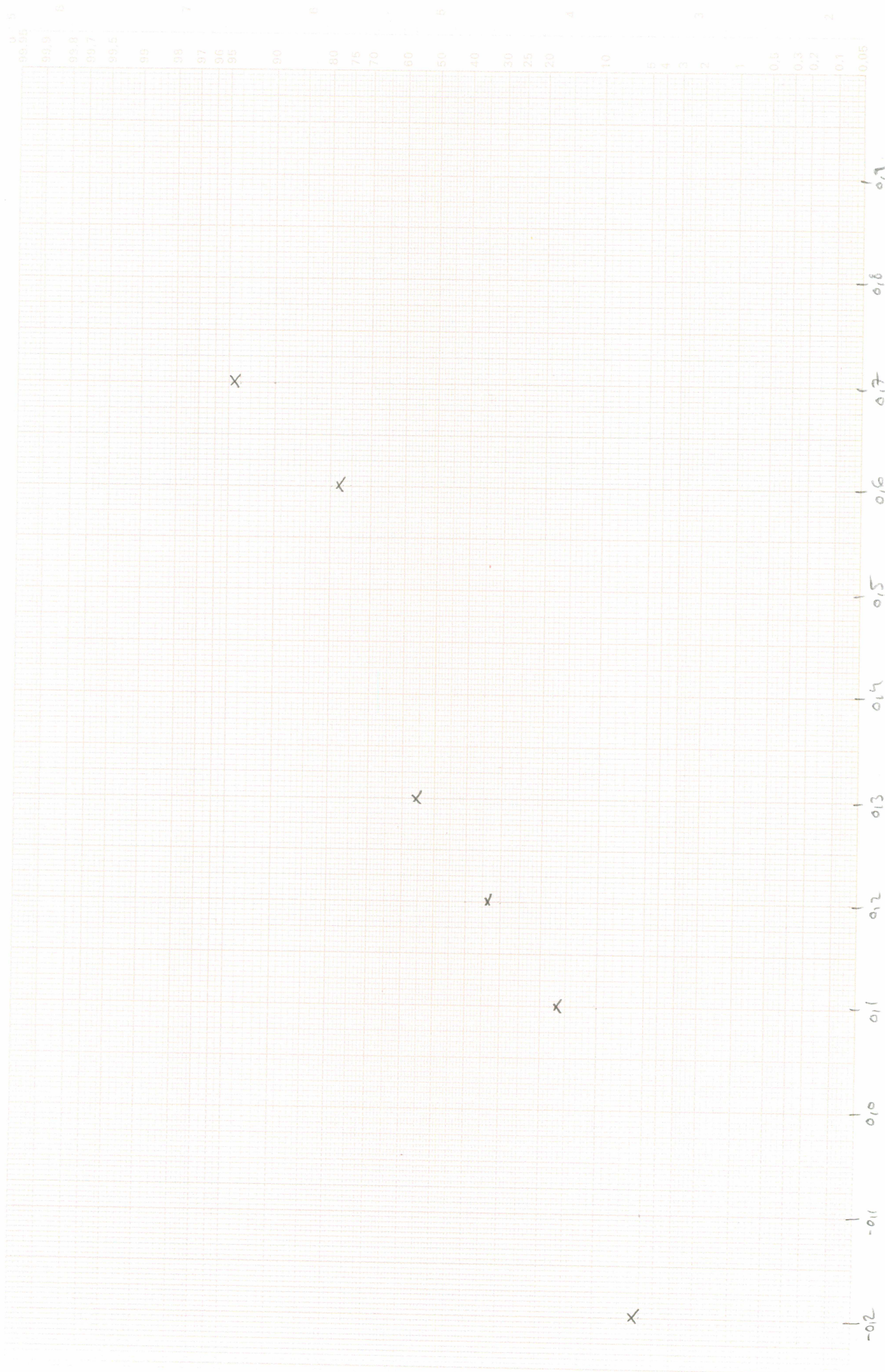
$$[\bar{x}_1 - \sqrt{s^2/n} t_{0,975}(n-1) ; \bar{x}_1 + \sqrt{s^2/n} t_{0,975}(n-1)]$$

$$= [0,0922 ; 0,5300]$$

$$\text{idet } \bar{x}_1 = S/n, \quad s^2 = \frac{USS - S^2/n}{n-1}, \quad t_{0,975}(9-1) = 2,306$$



OPG 3.2



opgave 3.3

Da variansen er et mål for målesikkerhed er det naturligt at se på forsøget som 2 delforsøg med hver 10 obs. rækker med hver 2 obs.

De i alt 20 obs. rækker kan antages at være normalfordelt (De store tals lov). Det er nyttigt at tegne prøvernes resultat i hver enkelt obs. række op mod hinanden (f.eks. p_{v-2} i måling 3 op mod p_{v-2} i måling 4)

Dette er gjort i vedlagte SAS udskrift, og der er en pæn lin. sammenhæng, hvilket ikke strider imod, at vi kan antage flg. grundmodel:

$$H_0: X_{ijk} \sim N(\mu_{ij}, \sigma_i^2) \quad i=1,2 \text{ (delforsøg)}, j=1, \dots, 10 \text{ (prøver)}, k=1,2$$

Da der ikke er mere end 2 obs. i hver obs. række kan vi ikke benytte Bartlett's test for ens varians, og vi vil derfor blot antage ens varians indenfor hvert delforsøg

Hvis der er en forskel på målesikkerheden i de 2 delforsøg bør flg. hypotese kunne afkræftes:

$$H_0: \sigma_1^2 = \sigma_2^2$$

Juf. s. 3.55 har vi flg. estimater

$$\sigma_i^2 \leftarrow s_i^2 = \frac{\sum SSD_{(i)}}{\sum f_{(i)}} = \frac{\sum (USS_{(i)} - \frac{S_{i.}^2}{2})}{10 \cdot (2-1)}$$

$$= \frac{1}{10} (0,0032 + 0,00405 + 0,00405 + 0,0018 + 0,0002 + 0,0018 + 0,00405 + 0,005 + 0,0032 + 0,0005) = 0,00314$$

Tilsvarende fås (af læsning i SAS udskrift i linien
ERROR under Mean Square) at $S_2^2 = 0,014465$

og jf. s. 3.56: $S_i^2 \sim \sigma_i^2 \chi^2(10)/10$

Under nulhypotesen er $F = \frac{S_2^2}{S_1^2} \sim F(10, 10)$

$F = \frac{0,014465}{0,00314} \approx 4,607$, denne størrelse bruger vi som teststatistik

Dermed fås at teststørrelsen $E(X)$ for H_0 er

$$E(X) = 2 (1 - F_{F(10,10)}(4,607)) < 2 (1 - 0,975) = 0,05$$

Altså forkastes hypotesen, og vi kan derfor konkludere
at der er forskel på målesikkerheden i de
to delforsøg, og at målesikkerheden i det
1. delforsøg er lavere end i det 2. delforsøg.

```
OPTIONS PAGENO=1 NODATE PS=35;
TITLE1 'OPGAVE 3.3';
```

```
DATA dobbelt;
INPUT e11 e12 e21 e22 nr;
      diff1=e11-e12;
      diff2=e21-e22;
```

```
DATALINES;
5.25 5.17 4.79 4.80 1
6.10 6.19 5.61 5.55 2
5.70 5.61 5.33 5.03 3
5.76 5.67 5.86 6.11 4
5.00 4.94 5.42 5.32 5
6.04 6.02 4.91 5.04 6
5.49 5.55 5.83 6.02 7
5.33 5.42 5.85 5.77 8
4.84 4.94 5.25 5.39 9
5.89 5.81 5.01 5.22 10
RUN;
```

```
PROC PLOT DATA=dobbelt;
      PLOT e11*e12='*';
```

```
PROC PLOT DATA=dobbelt;
      PLOT e21*e22='*';
```

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

```
TITLE2 'eksperiment 1';
%normgraf(DATA=dobbelt,var=diff1);
```

```
PROC MEANS DATA=dobbelt n mean stderr t prt USS;
      VAR diff1;
```

```
TITLE2 'eksperiment 2';
%normgraf(DATA= dobbelt, var=diff2);
```

```
PROC MEANS DATA=dobbelt n mean stderr t prt USS;
      VAR diff2;
```

```
RUN;
```

```
/*-----*/
/* De to gentagne maalinge paa samme proeve kan betragtes som
   een observationsraekke. Saadan opfattet bestaar hvert af de
   to eksperimenter af ti observationsraekker med to observationer
   i hver, og hvor man kun interesserer sig for variansen.
   For at kunne finde estimat for variansen i den model, skal
   man have data paa en form, hvor een variabel (her x) indeholder
   alle maalinge, og to andre variable holder styr paa hvilken proeve
   en maaling er fra (her proeve) og hvilket eksperiment en proeve er
   fra (her forsoeg).
*

```



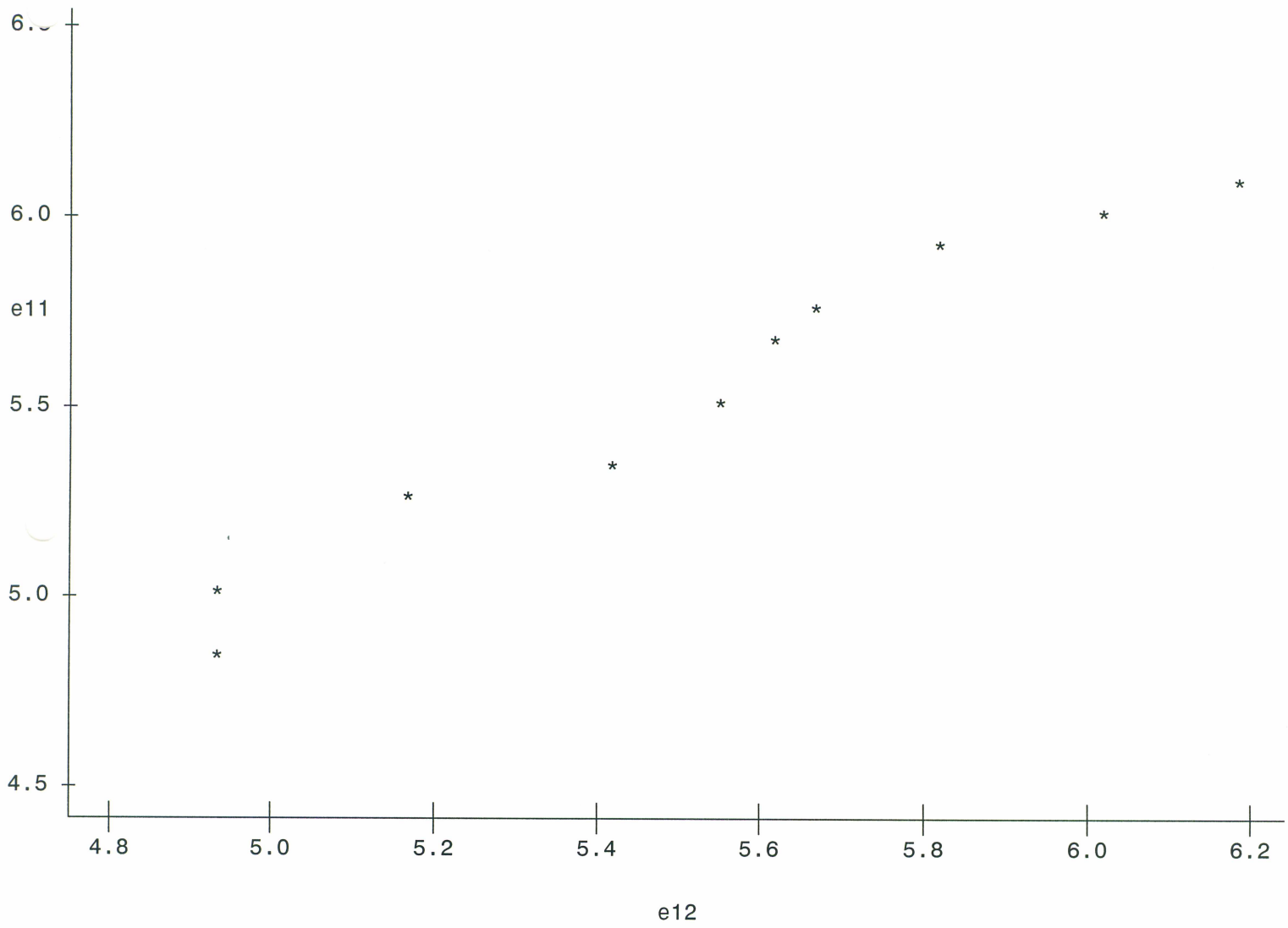
```
DATA lang;  
SET dobbelt;  
  x=e11; proeve=nr; forsoeg=1; output;  
  x=e12; proeve=nr; forsoeg=1; output;  
  x=e21; proeve=nr; forsoeg=2; output;  
  x=e22; proeve=nr; forsoeg=2; output;  
DROP e11 e12 e21 e22 diff1 diff2;  
RUN;
```

```
PROC SORT DATA=lang;  
BY forsoeg;  
RUN;
```

```
TITLE2 '10 observationsraekker med to observationer i hver';  
TITLE3 ' i hvert forsoeg';  
PROC GLM DATA=lang;  
CLASS proeve;  
MODEL x=proeve/SS1;  
BY forsoeg;  
RUN;
```

```
/* Oprydning */  
TITLE;
```

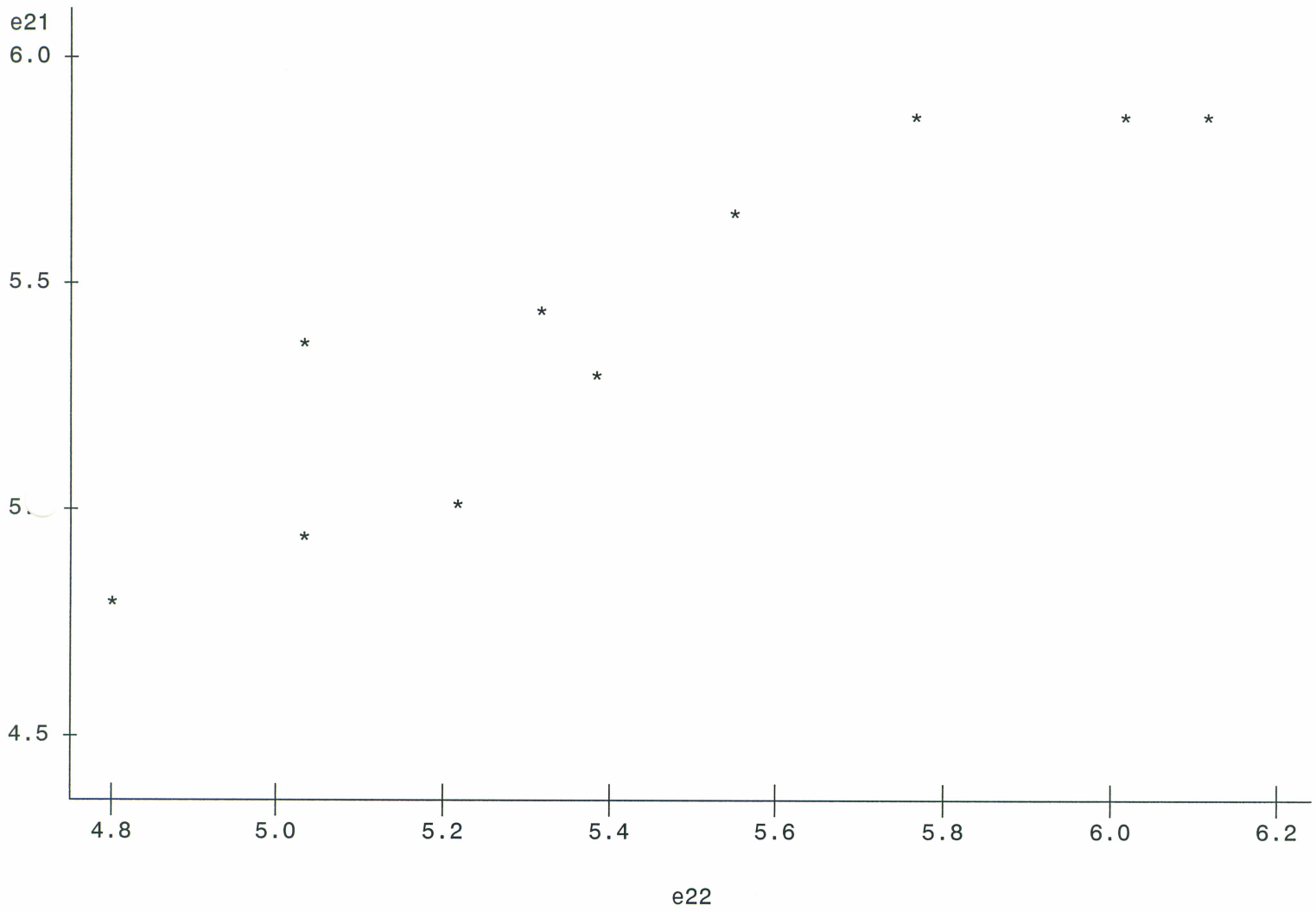
Plot of $e_{11} \cdot e_{12}$. Symbol used is '*'.



OPGAVE 3.3
eksperiment 1

2

Plot of $e_{21} \cdot e_{22}$. Symbol used is '*'.



OPGAVE 3.3
eksperiment 2

The MEANS Procedure

Analysis Variable : diff1

N	Mean	Std Error	t Value	Pr > t	USS
10	0.0080000	0.0262805	0.30	0.7677	0.0628000

OPGAVE 3.3
eksperiment 2

6

The MEANS Procedure

Analysis Variable : diff2

N	Mean	Std Error	t Value	Pr > t	USS
10	-0.0390000	0.0551855	-0.71	0.4976	0.2893000

10 observationsraekker med to observationer i hver
i hvert forsoeg

----- forsoeg=1 -----

The GLM Procedure

Dependent Variable: x

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	9	3.25968000	0.36218667	115.35	<.0001
Error	10	0.03140000	0.00314000		
Corrected Total	19	3.29108000			

R-Square	Coeff Var	Root MSE	x Mean
0.990459	1.012206	0.056036	5.536000

Source	DF	Type I SS	Mean Square	F Value	Pr > F
proeve	9	3.25968000	0.36218667	115.35	<.0001

10 observationsraekker med to observationer i hver
i hvert forsoeg

----- forsoeg=2 -----

The GLM Procedure

Dependent Variable: x

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	9	3.00324500	0.33369389	23.07	<.0001
Error	10	0.14465000	0.01446500		
Corrected Total	19	3.14789500			

R-Square	Coeff Var	Root MSE	x Mean
0.954049	2.224966	0.120271	5.405500

Source	DF	Type I SS	Mean Square	F Value	Pr > F
proeve	9	3.00324500	0.33369389	23.07	<.0001

opg. 3.4

$$x_i \sim \text{po}(\lambda) \quad i=1, \dots, n$$

1)

$$L(\lambda, \underline{x}) = \prod_{i=1}^n e^{-\lambda} \frac{\lambda^{x_i}}{x_i!} = e^{-\lambda n} \lambda^{\sum x_i} \prod_{i=1}^n \frac{1}{x_i!}$$

$$\ell(\lambda, \underline{x}) = -\lambda n + \sum x_i \cdot \ln(\lambda) - \sum \ln(x_i!)$$

$$0 = \frac{\partial \ell}{\partial \lambda} = -n + \sum_{i=1}^n \frac{x_i}{\lambda} = -n + \frac{1}{\lambda} \sum_{i=1}^n x_i$$

2)

$$\text{Påstand: } \hat{\lambda} = \bar{x}.$$

Basis:

$$0 = -n + \frac{1}{\lambda} \sum_{i=1}^n x_i \Rightarrow \lambda = \bar{x}.$$

$$\text{Idet } \lambda < \bar{x}: \frac{\partial \ell}{\partial \lambda}(\lambda) < 0$$

$$\lambda > \bar{x}: \frac{\partial \ell}{\partial \lambda}(\lambda) > 0$$

Se vi at ℓ har maksimum i $\lambda = \bar{x}$.

□

3)

Som vist i 2) er maksimum likelihood estimatet for $\hat{\lambda}$ givet ved $\bar{x}$. (forudsat,

$\exists i: x_i \neq 0$)*, dvs. ss. for at $\hat{\lambda}$ er veldef. er

$$P\left(\bigcap_{i=1}^n \{x_i \neq 0\}\right) = 1 - P\left(\bigcap_{i=1}^n \{x_i = 0\}\right)$$

$$= 1 - \prod_{i=1}^n e^{-\lambda} = 1 - e^{-n\lambda} \quad * \lambda > 0$$

4)

$$H_0: \lambda = \lambda_0$$

$$Q(x) = \frac{L(\lambda_0, x)}{L(\hat{\lambda}, x)} = \frac{e^{-\lambda_0 n} \lambda_0^{\sum x_i} \prod_{i=1}^n \frac{1}{x_i!}}{e^{-\hat{\lambda} n} \hat{\lambda}^{\sum x_i} \prod_{i=1}^n \frac{1}{x_i!}}$$

$$= e^{-\lambda_0 n + x} (\lambda_0 n / x)^x$$

5)

Pfand: $x \rightarrow e^{-\lambda_0 n + x} (\lambda_0 n / x)^x$ er
 strengt aft. for $x \geq n \lambda_0$
 strengt vok. for $x \leq n \lambda_0$

Beweis:

$$\ln(e^{-\lambda_0 n + x} (\lambda_0 n / x)^x) = -\lambda_0 n + x + x \ln(\lambda_0 n / x)$$

$$D_x \ln(\dots) = 1 + \ln(\lambda_0 n / x) - 1$$

$$\begin{cases} < 0 & , x \geq \lambda_0 n \\ > 0 & , x \leq \lambda_0 n \end{cases}$$

□

6)

$$n=5, \lambda_0=3, x_0=20$$

De x , som er mere eller ligeså kritiske for H_0

som $x = 20$ er jvf. 5) $x \geq 20$ og

$$\text{idet } e^{-15+10} (15/10)^{10} < e^{-15+20} (15/20)^{20} < e^{-15+11} (15/11)^{11}$$

og $x \leq 10$, altså alt i alt de x for hvilket
der gælder, at $x \leq 10$ eller $x \geq 20$

Under nulhypotesen gælder (jvf. teori om uafh. po-f.d.)

$$x \sim \text{po}(5,3)$$

Dermed har vi

$$E(x) = 1 - e^{-15} \sum_{k=11}^{19} \frac{15^k}{k!} \approx 0,2432 > 0,05$$

Dvs. vi forkaster ikke hypotesen ($\lambda_0 = 3$)

STAT1, afleveringsopgave nr. 2, Gunna Hellebrand

opg. 3.6

1)

$$L(x, n, \pi) = \binom{n}{x} \pi^x (1-\pi)^{n-x}$$

$$\ell(x, n, \pi) = \ln \binom{n}{x} + x \cdot \ln \pi + (n-x) \ln(1-\pi)$$

$$0 = \frac{\partial \ell}{\partial \pi} = \frac{x}{\pi} - \frac{n-x}{1-\pi}$$

2)

En løsn. til likelihood-lign., $\hat{\pi}$ opfylder $\frac{x}{\hat{\pi}} = \frac{n-x}{1-\hat{\pi}}$
dvs $\hat{\pi} = \frac{x}{n}$

idet $\frac{\partial \ell}{\partial \pi}(\pi) \leq 0$, $\pi \leq \frac{x}{n}$, $\frac{\partial \ell}{\partial \pi}(\pi) \geq 0$, $\pi \geq \frac{x}{n}$

se vi, at $\hat{\pi}$ er maksimum likelihood est., $\hat{\pi}$ giver kun
et udelst. estimat, hviss $\hat{\pi} = \frac{x}{n} \in]0, 1[$, dvs.

med ss $= 1 - P(X=0) - P(X=n) = 1 - (1-\pi)^n - \pi^n$

3)

$$H_0: \pi = \pi_0 = 1/2$$

$$Q(x) = \frac{L(\pi_0)}{L(\hat{\pi})} = \frac{\binom{n}{x} \left(\frac{1}{2}\right)^x (1-\frac{1}{2})^{n-x}}{\binom{n}{x} \left(\frac{x}{n}\right)^x (1-\frac{x}{n})^{n-x}} = \frac{(1/2)^n}{(x/n)^x (1-x/n)^{n-x}}$$

Meddelelse 4

Forelæsningerne i sidste uge

(Onsdag den 11. september) Gennemgangen af Kapitel 2 sluttede med Afsnit 2.5 *Transformation*. I forbindelse hermed blev proceduren `normgraf` omtalt.

Anneks til Kapitel 2 *Beregninger i SAS* vil blive omtalt ved en senere lejlighed. Men i Annekset beskrives, hvordan de grafiske metoder, som indføres i Kapitel 2 kan udføres i SAS. Programmer til alle eksempler i Kapitel 2 kan hentes på nettet, se til sidst på Meddelelse 3.

(Fredag den 13. september) Vi begynder på Kapitel 3 *Normalfordelte data* og begynder forfra, men ser også i Supplement til Kapitel 3, hvor vigtige resultater om normalfordelingen og relaterede fordelinger er samlet.

Forelæsningerne i denne uge

(Onsdag den 18. september) Resumé af testene i en normalfordelt observationsrække. Gennemgang af konfidensintervaller for middelværdi og varians i en normalfordelt observationsrække.

(Fredag den 20. september) Afsnit 3.1 afsluttes og jeg begynder på Afsnit 3.2 *Flere observationsrækker*. Jeg forventer at komme til side 3.33.

Forelæsningerne i næste uge

(Onsdag den 25. september og fredag den 27. september) Jeg afslutter gennemgangen af Afsnit 3.2.1 *To observationsrækker* og begynder på Afsnit 3.2.2 *Flere observationsrækker*.

Øvelserne i denne uge

(Uge 38: 16.-22. september) Opgave 3.1 (resten af opgaven) samt opgaver i tilknytning til Supplement til Kapitel 3: 3.21, 3.22, 3.23, 3.24. Hvis der er mere tid kan resten af øvelsestiden bruges i EDB lokalet.

Øvelserne i de to næste uger

(Uge 39: 23. september-29. september) Omfatter som minimum opgaverne 3.2, 3.3 og 3.4. Hvis der er mere tid kan den bruges i EDB lokalet, for eksempel med at regne opgaverne 3.2 og 3.3 ved hjælp af SAS.

(Uge 40: 30. september-6. oktober) Opgave 3.5, 3.6 (afleveringsopgaven) regnes færdig, samt opgaven sidst på denne meddelelse.

Afleveringsopgaver

(Uge 39: 23. september-29. september) Følgende opgave:

15 laborantstuderende er blevet bedt om - uafhængigt af hinanden - at bestemme koncentrationen af laktat i den samme blodprøve med en kendt koncentration. De studerendes målinger er i mg/l

12	10	9	14	2	8	6	9
86.5	81.0	79.4	90.9	73.8	80.2	79.9	80.3
71.9	78.6	83.1	79.6	80.0	87.0	92.7	
1	3	11	5	7	13	15	

og den kendte koncentration er 80.0 mg/l. Analyser data med henblik på at belyse om de studerende måler nøjagtigt og at beskrive måleusikkerheden.

(Uge 40: 30. september-6. oktober) Opgave 3.6 1), 2) og 3).

(Uge 40) Opgave 3.6 1), 2) og 3).

(Uge 41) En opgave i to observationsrækker, som vil komme på næste meddelelse.

Opgave til øvelserne i uge 40

Institut for Idræt ved Københavns Universitet arrangerer hvert år et atletikstævne for de 1. års studerende. Atletiklærer Mikkel Sørensen har stillet nogle af resultaterne for årene 1998 og 1999 til rådighed. Det er af interesse at vide om resultaterne for de tre år er forskellige eller ej. Resultaterne af pigernes længdespring er vist i tabellen nedenfor.

år	længde i m									
1998	3.72	3.65	3.90	3.74	3.32	4.22	3.58	4.56	3.65	2.99
	3.91	3.65	3.65	3.88	3.65	4.20				
1999	4.32	3.79	3.53	3.54	4.27	3.75	4.21	3.66	4.58	3.73
	5.18	3.00	3.91	3.52						

Standardberegninger findes i tabellen nedenfor.

	n	S	USS
1998	16	60.27	229.0119
1999	14	54.99	219.8343

STAT1, TØ nr. 3

opg. 3.21

X kont. s.v. med tæthed f_X

h er en strengt monoton og diff. reel fkt.

$$Y := h(X)$$

Juf. ssz er Y en kont. s.v.

Påstand: $f_Y(y) = |h'(h^{-1}(y))|^{-1} f_X(h^{-1}(y))$, $y \in \text{supp } f_Y$

Beris:

Dette er en simpel konsekvens af (3.94)

(Se desuden SEG noterne s. 86, hvor sætn. er beviset)

□

opg. 3.22

1)

Påstand: $\gamma(x, \alpha, \lambda)$ i (3.72) er en tæthedsfkt. $\alpha > 0, \lambda > 0$ og γ er strengt aft. for $\alpha \leq 1$ og unimodal for $\alpha > 1$ med maksimum i $x = \frac{\alpha-1}{\lambda}$

Beris:

$$\begin{aligned} \int_{\mathbb{R}} \gamma(x, \alpha, \lambda) dx &= \int_0^{\infty} \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} dx = \frac{\lambda^\alpha}{\Gamma(\alpha)} \int_0^{\infty} x^{\alpha-1} e^{-\lambda x} dx \\ &= \frac{\lambda^\alpha}{\Gamma(\alpha)} \int_0^{\infty} (z/\lambda)^{\alpha-1} e^{-z} \frac{dz}{\lambda} = \frac{\lambda^\alpha}{\Gamma(\alpha)} \lambda^{-\alpha} \int_0^{\infty} z^{\alpha-1} e^{-z} dz = 1 \end{aligned}$$

$$\alpha = 1:$$

γ er klart strengt aft.

$$\alpha < 1:$$

$$\begin{aligned}\gamma'(x) &= \frac{\lambda^\alpha}{\Gamma(\alpha)} (\alpha-1) x^{\alpha-2} e^{-\lambda x} + \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} (-\lambda) e^{-\lambda x} \\ &= \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-2} e^{-\lambda x} (\alpha-1-\lambda x) < 0, \text{ dus.}\end{aligned}$$

γ er strengt aft.

$$\alpha > 1:$$

$$\gamma'(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-2} e^{-\lambda x} (\alpha-1-\lambda x)$$

$$\alpha-1-\lambda x = 0 \iff x = \frac{\alpha-1}{\lambda}$$

$$\alpha-1-\lambda x < 0 \iff x > \frac{\alpha-1}{\lambda}$$

$$\alpha-1-\lambda x > 0 \iff x < \frac{\alpha-1}{\lambda}$$

Dvs.

γ er voksende på $(0, \frac{\alpha-1}{\lambda})$ har maks. i $\frac{\alpha-1}{\lambda}$ og aftager på $(\frac{\alpha-1}{\lambda}, \infty)$

□

2)

Påstand: $X \sim \Gamma(\alpha, \lambda)$, $c > 0 \Rightarrow cX \sim \Gamma(\alpha, \lambda/c)$

Beris:

Juf. 3.21 er en metode for cX :

$$f_{cX}(t) = |c|^{-1} f_X(\frac{1}{c}t)$$

$$= \frac{1}{c} \frac{\lambda^\alpha}{\Gamma(\alpha)} (\frac{1}{c}t)^{\alpha-1} e^{-\lambda(\frac{1}{c}t)} = \frac{(\lambda/c)^\alpha}{\Gamma(\alpha)} t^{\alpha-1} e^{-(\lambda/c)t}$$

dvs. $cX \sim \Gamma(\alpha, \lambda/c)$

□

3)

Påstand:

$X_1 \sim \Gamma(\alpha_1, \lambda) \perp X_2 \sim \Gamma(\alpha_2, \lambda) \Rightarrow X_1 + X_2 \sim \Gamma(\alpha_1 + \alpha_2, \lambda)$

Beris:

For $X \sim \Gamma(\alpha, \lambda)$: $\varphi_X(t) = \left(\frac{\lambda}{\lambda - it}\right)^\alpha$, $t \in \mathbb{R}$

Juf. Ginnert & Storzaker Bed. 2001 s. 188

Af. 4.18 i Hoffmann følger pga. uafh. at

$$\varphi_{X_1+X_2}(t) = \varphi_{X_1}(t) \varphi_{X_2}(t) = \left(\frac{\lambda}{\lambda - it}\right)^{\alpha_1 + \alpha_2}, t \in \mathbb{R}$$

og dermed (4.18 i Hoffmann) at $X_1 + X_2 \sim \Gamma(\alpha_1 + \alpha_2, \lambda)$

□

4)

Påstand: $X \sim \Gamma(\alpha, \lambda) \Rightarrow E[X] = \alpha/\lambda$, $\text{Var} X = \alpha/\lambda^2$

Beris:

$$\begin{aligned} E[|X|^n] &= \int_0^\infty x^n \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} dx \\ &= \frac{\lambda^\alpha}{\Gamma(\alpha)} \int_0^\infty x^{n+\alpha-1} e^{-\lambda x} dx = \frac{\lambda^\alpha}{\Gamma(\alpha)} \frac{\Gamma(n+\alpha)}{\lambda^{n+\alpha}} = \frac{\alpha(\alpha+1) \cdots (\alpha+n-1)}{\lambda^n} < \infty \end{aligned}$$

Dus:

$$E[X] = E[|X|] = \alpha/\lambda$$

$$\text{Var}(X) = E[X^2] - E[X]^2 = \frac{\alpha(\alpha+1)}{\lambda^2} - \frac{\alpha^2}{\lambda^2} = \alpha/\lambda^2$$

□

5)

Påstand: $\sigma^2 \chi^2(f) \sim \Gamma(f/2, 1/(2\sigma^2))$, $\sigma^2 \chi^2(f)/f \sim \Gamma(f/2, f/2\sigma^2)$

Beris:

$\chi^2(f) \sim \Gamma(f/2, 1/2)$ pr. def.

Resultatet er dermed trivielt ud fra 2)

□

6)

Påstand: $\chi^2(2) \sim E(2)$

Beris:

Pr. def. gælder, at $\chi^2(2) \sim \Gamma(1, 1/2)$, dus

$$f_{\chi^2(2)}(x) = \frac{1/2}{\Gamma(1)} x^{1-1} e^{-x/2} = 1/2 e^{-x/2}, x > 0 \quad \text{og ellers}$$

Dus $\chi^2(2) \sim E(2)$

□

opg. 3.23

Påstand: $U \sim N(0,1) \Rightarrow U^2 \sim \chi^2(1)$

Beris:

Lad $x \in \mathbb{R}$ være vilk.

$$F_{U^2}(x) := P(U^2 \leq x)$$

$$x \leq 0: F_{U^2}(x) = 0$$

$$x > 0: F_{U^2}(x) = P(-\sqrt{x} \leq U \leq \sqrt{x}) = \int_{-\sqrt{x}}^{\sqrt{x}} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt$$

$$= \sqrt{\frac{2}{\pi}} \int_0^{\sqrt{x}} e^{-t^2/2} dt = \sqrt{\frac{2}{\pi}} \int_0^x \frac{1}{2\sqrt{s}} e^{-s/2} ds$$

$$= \int_0^x \frac{(1/2)^{1/2}}{\Gamma(1/2)} s^{1/2-1} e^{-s \cdot (1/2)} ds, \text{ dvs. } U^2 \text{ er a.k.}$$

se evt.
Lindström
9.2.7

mhk. λ_1 med tæthed $s \mapsto \frac{(1/2)^{1/2}}{\Gamma(1/2)} s^{1/2-1} e^{-(1/2)s} \cdot 1_{\mathbb{R}_+}(s)$
og dermed $U^2 \sim \Gamma(1/2, 1/2) \sim \chi^2(1)$

□

3.24

Find tætheden for $t(f)$ -fordelingen:

Dette er vist i afl. opg. nr. 15 i SS2

□

STAT 1, afleveringsopgave nr. 1, Emma Hellebrand.

obs	kom. antal	ss til brug i fraktildiagram
71,9	1	3,33%
73,8	2	10 %
78,6	3	16,67%
79,4	4	23,33%
79,6	5	30 %
- 79,9	6	36,67%
80,0	7	43,33%
- 80,2	8	50%
80,3	9	56,67%
81,0	10	63,33%
83,1	11	70 %
86,5	12	76,67%
87,0	13	83,33%
90,9	14	90 %
92,7	15	96,67%

USS = 100472,63
S = 1224,9

Juf. vedlagte fraktildiagram kan vi antage, at data er normalfordelt, dvs. vores model er

$$M_0: X_i \sim N(\mu, \sigma^2)$$

$$s^2 = \frac{1}{n-1} (USS - s^2/n) \quad , \text{ da } f = n-1 = 14 \text{ er}$$

(3.21)
(3.22)

et 95% konfidensint. for σ^2 (j.f. 3.14)

$$\left[\frac{f_{s^2}}{\chi^2_{0,975}(f)}, \frac{f_{s^2}}{\chi^2_{0,025}(f)} \right]$$

$$= [17,14 ; 79,45] , s^2 = 31,95 , \mu = 81,66$$

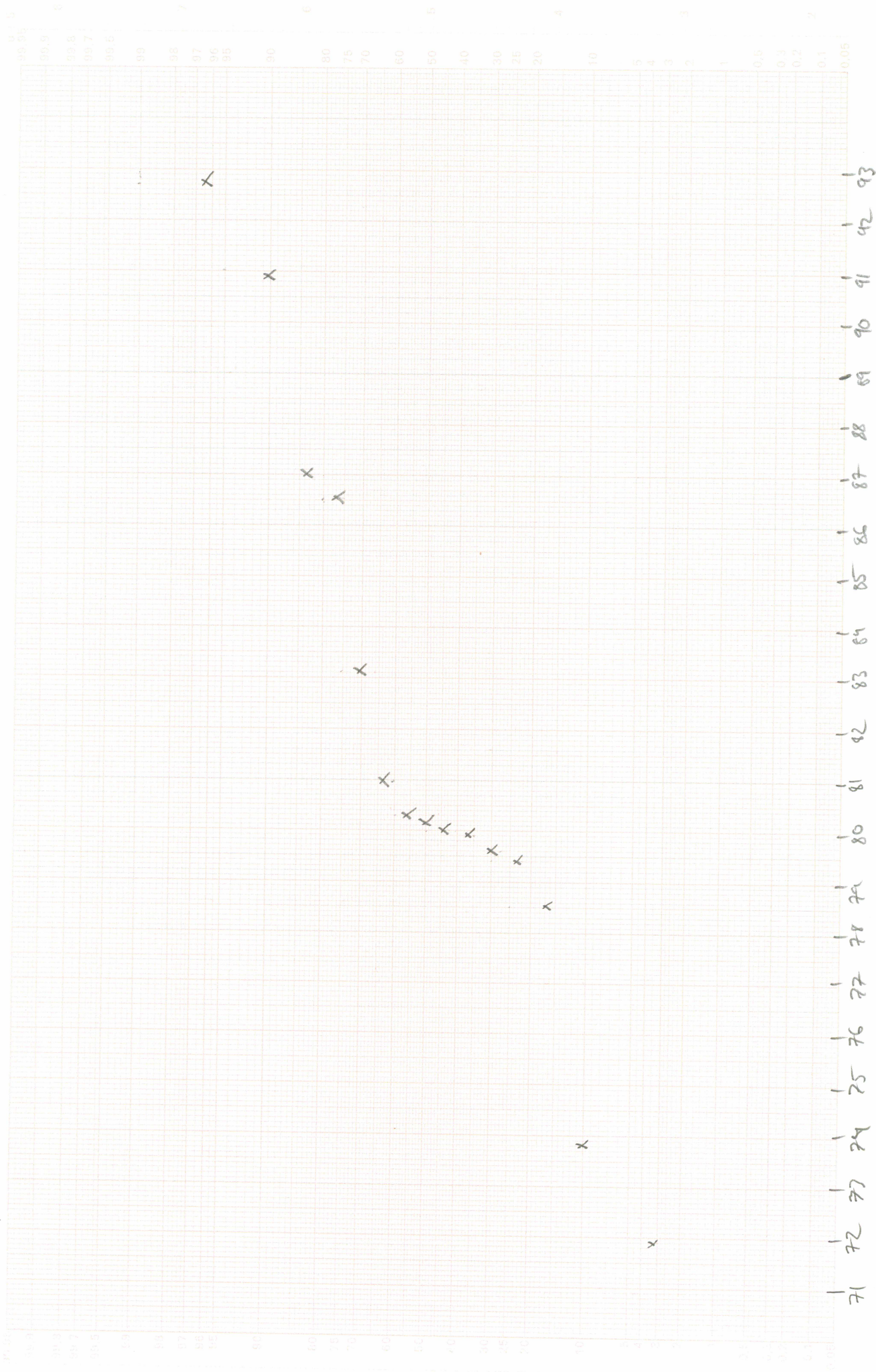
Da konfidensintervallet er meget bredt, kan vi konkludere at målingerne ikke er ret nøjagtige



Nr. 2110

1954 - 1955

Afl. nr. 1



Meddelelse 3

Forelæsningerne i sidste uge

(Onsdag den 4. september) Frank Dabelstein fortalte i 25 minutter om EDB systemet på Institut for Matematiske Fag. Efter nogle indledende bemærkninger gik jeg direkte i gang med Kapitel 2 om *Indledende undersøgelser*.

(Fredag den 6. september) Jeg fortsatte med Kapitel 2 og nåede ind i Afsnit 2.4 *Fraktildiagrammer for normalfordelingen*, hvor jeg som det sidste lige nåede at omtale Tabel 2.3: Beregnings-skema for fraktildiagrammer for et *ugrupperet* datasæt.

Forelæsningerne i denne uge

(Onsdag den 11. september) Gennemgangen af Kapitel 2 sluttede med Afsnit 2.5 *Transformation*. I forbindelse hermed blev proceduren `normgraf` omtalt.

Anneks til Kapitel 2 *Beregninger i SAS* vil blive omtalt ved en senere lejlighed. Men i Annekset beskrives, hvordan de grafiske metoder, som indføres i Kapitel 2 kan udføres i SAS. Programmer til alle eksempler i Kapitel 2 kan hentes på nettet, se til sidst på denne meddelelse.

(Fredag den 13. september) Vi begynder på Kapitel 3 *Normalfordelte data* og begynder forfra, men ser også i Supplement til Kapitel 3, hvor vigtige resultater om normalfordelingen og relaterede fordelinger er samlet.

Forelæsningerne i næste uge

(Onsdag den 18. og fredag den 29. september) Onsdag er Afsnit 3.1 ved at være gennemgået og vi fortsætter med Afsnit 3.2.

Øvelserne i denne uge

(Uge 37: 9.-15. september) Anden øvelsesgang brugte vi den første halvdel af øvelsestiden til at regne opgaver og fortsatte med introduktionen til SAS i den anden halvdel.

Programmet omfatter: Opgaverne 2.1, 2.2 og 3.1(kun spørgsmål (1)).

I EDB øvelserne fortsætter man hvor man slap med opgaverne bag i Introduktion til SAS. Opgave 2.2 er tænkt som et oplæg til diskussion i forbindelse med EDB øvelserne.

Øvelserne i de to næste uger

(Uge 38: 16.-22. september) Opgave 3.1 (resten af opgaven) samt opgaver i tilknytning til Supplement til Kapitel 3: 3.21, 3.22, 3.23, 3.24. Hvis der er mere tid kan resten af øvelsestiden bruges i EDB lokalet.

(Uge 39: 23. september-29. september) Omfatter som minimum opgaverne 3.2, 3.3 og 3.4. Hvis der er mere tid kan den bruges i EDB lokalet, for eksempel med at regne opgaverne 3.2 og 3.3 ved hjælp af SAS.

Afleveringsopgave

Jeg vil prøve at stille afleveringsopgaver til de fleste uger. Afleveringsopgaver skal ikke forveksles med obligatoriske opgaver. Afleveringsopgaver er et tilbud, som det meget anbefales at benytte. Jeg stiller opgaverne til en bestemt uge, og det betyder, at man på holdene skal aftale et afleveringstidspunkt, så instruktoren kan nå at rette besvarelsenerne, så de kan leveres tilbage og diskuteres på øvelserne i den pågældende uge.

(Uge 39: 23. september-29. september) Følgende opgave:

15 laborantstuderende er blevet bedt om - uafhængigt af hinanden - at bestemme koncentrationen af laktat i den samme blodprøve med en kendt koncentration. De studerendes målinger er i mg/l

86.5	81.0	79.4	90.9	73.8	80.2	79.9	80.3
71.9	78.6	83.1	79.6	80.0	87.0	92.7	

og den kendte koncentration er 80.0 mg/l. Analyser data med henblik på at belyse om de studerende måler nøjagtigt og at beskrive måleusikkerheden.

(Uge 40: 30. september-6. oktober) Opgave 3.6 1), 2) og 3).

Valg af browser

netscape er fuldt tilstrækkelig til øvelserne. Den åbner meget hurtigt og den tvinger ikke nettet i knæ, hvis et helt øvelseshold åbner den på én gang.

Skal man ud i verden, for eksempel til SAS Institutes hjemmeside <http://www.sas.com/> får man brug for en nyere browser. Brug da kommandoen `communicator` på Unix og enten `mozilla` eller `netscape` på Linux.

Rettettelse til oplysning om filer på nettet på Meddelse 2

Rettelsen består i at

<http://www.imf.au.dk/~statbib>

skal erstattes af

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

og skyldes at systemfolkene har flyttet placeringen af hjemmesider. Bemærk at bogens oplysninger om placering af programmer til opgaver også skal rettes.

SAS programmer til eksempler og opgaver ligger under brugernavnet `statbib`. Nemmeste adgang er via en browser. Adressen er

`http://home.imf.au.dk/statbib`

Specielt for øvelserne vedrørende introduktionen til SAS er adressen

`http://home.imf.au.dk/statbib/sasintro`

interessant, for her ligger alle eksemplerne, der skal køres igennem. Bemærk også `autoexec.sas` som ligger her, og indeholder fornuftige opsætninger til grafik.

Vi har lavet nogle SAS macroer til brug for kurset. De kan ses på adressen.

`http://home.imf.au.dk/statbib/sasmacros`

Det kan være en fordel at kunne se starten af en macro, hvis man ikke lige kan huske alle kodeordene. Var det nu `bygroup` eller `gruppe` man skulle skrive?

For Statistik 1 er adresserne til bogens eksempler og opgaver, samt meddelelserne

`http://home.imf.au.dk/statbib/atiskurser/stat1/eksempler`

(fordelt på kapitler)

`http://home.imf.au.dk/statbib/atiskurser/stat1/opgaver`

(ligeledes fordelt på kapitler) og

`http://home.imf.au.dk/statbib/atiskurser/stat1/meddelelser/aktuelle.`

For Statistik α er de tilsvarende adresser (det er nøjagtigt de samme filer som `stat1`'erne ser)

`http://home.imf.au.dk/statbib/atiskurser/statalfa/eksempler`

(fordelt på kapitler)

`http://home.imf.au.dk/statbib/atiskurser/statalfa/opgaver`

(ligeledes fordelt på kapitler) og

`http://home.imf.au.dk/statbib/atiskurser/statalfa/meddelelser/aktuelle.`

Det er bekvemt at kunne se eksemplerne i en browser og så kopiere relevante dele med musen over i SAS editoren.

STAT1 , TD nr. 2

opg. 2.1

$X_1, \dots, X_n$ uafh. s.v. $X_i \sim F, i=1, \dots, n$ (F f.d.f.k.t.)

Definer $F_n: \mathbb{R} \rightarrow [0, 1]$, $x \in \mathbb{R}: F_n(x) = \frac{\#\{i: X_i \leq x\}}{n} = \frac{1}{n} \sum_{i=1}^n 1_{(-\infty, x]}(X_i)$

1)

Påstand: $nF_n(x) \sim \text{bi}(n, F(x))$, $x \in \mathbb{R}$

Bewis:

$$\begin{aligned} P(nF_n(x) = k) &= P(\#\{i: X_i \leq x\} = k) \\ &= \binom{n}{k} P(X_1 \leq x)^k (1 - P(X_1 \leq x))^{n-k} = \binom{n}{k} F(x)^k (1 - F(x))^{n-k} \end{aligned}$$

for $k=1, \dots, n$, 0 ellers

□

2)

$$E[F_n(x)] = \frac{1}{n} E[nF_n(x)] = \frac{1}{n} n \cdot F(x) = F(x)$$

$$\text{Var}(F_n(x)) = \frac{1}{n^2} \text{Var}(nF_n(x)) = \frac{1}{n^2} n F(x)(1 - F(x)) = \frac{1}{n} F(x)(1 - F(x))$$

3)

Påstand: $F_n(x) \rightarrow F(x)$ i ss., dvs.

$$\forall \varepsilon > 0: \lim_{n \rightarrow \infty} P(|F_n(x) - F(x)| > \varepsilon) = 0$$

Bewis:

Lad $\varepsilon > 0$ være givet

$$P(|F_n(x) - F(x)| > \varepsilon) = P\left(\left|\frac{1}{n} \sum_{i=1}^n 1_{(-\infty, x]}(X_i) - F(x)\right| > \varepsilon\right) \xrightarrow{n \rightarrow \infty} 0$$

Juf. De store tals lov, idet X_i i.i.d. $\Rightarrow 1_{(-\infty, x]}(X_i)$ i.i.d.
og $E(1_{(-\infty, x]}(X_i)) = P(X_i \leq x) = F(x)$

□

opg. 2.2

→ se udskrift

opg. 3.1

1)

I numerisk rækkefølge er obs.:

15.934, 15.940, 15.943, 15.944, 15.947, 15.952, 15.953
15.957, 15.961

Dette giver flg. skema til brug for fraktildiagram

obs. , kum. antal , ss i %

1	1	$100/18 \approx 5,6\%$
2	2	$300/18 \approx 16,7\%$
3	3	$500/18 \approx 27,8\%$
4	4	$700/18 \approx 38,9\%$
5	5	$900/18 \approx 50\%$
6	6	$1100/18 \approx 61,1\%$
7	7	$1300/18 \approx 72,2\%$
8	8	$1500/18 \approx 83,3\%$
9	9	$1700/18 \approx 94,4\%$

Se vedlagte fraktildiagram

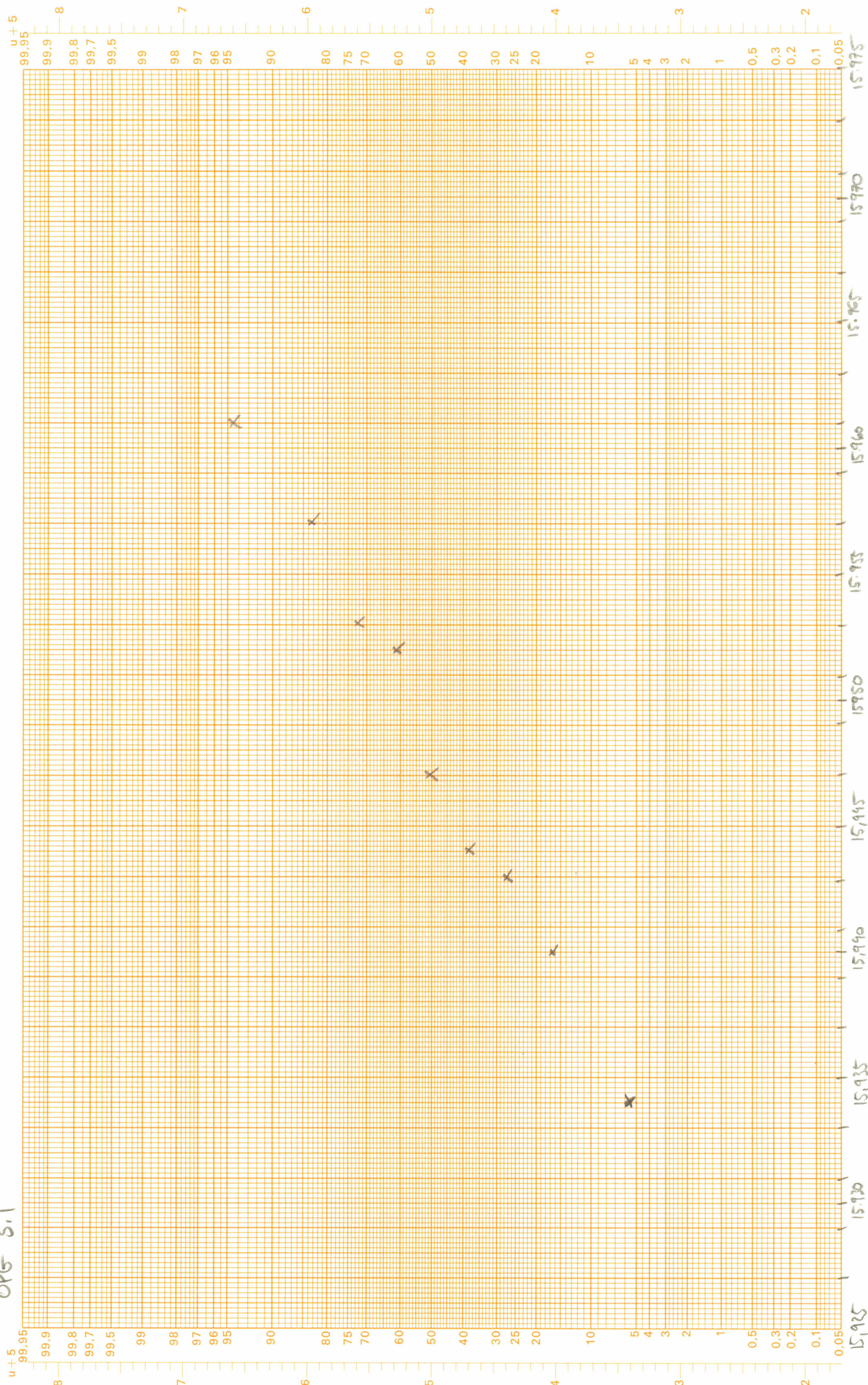
Fraktildiagrammet giver os ikke anledning til at tro på hypotesen om at data er normalfordelte, dvs. vores grundlæggende model er $H_0: X_i \sim N(\mu, \sigma^2)$



Nr. 2110

Lin. X Gauss

OPG 3.1



2)

$$H_0: \mu = 15.943$$

teststørrelsen som skal vurderes i en $t(9-1)$ -fordeling er

$$t(X) = \frac{\bar{S}_n - 15.943}{\sqrt{\frac{1}{n(n-1)}(U_{SS} - S^2/n)}} \approx 1.704$$

3.9
3.21
3.23

$$\text{Dus: } E(X) = 2(1 - F_{t(8)}(1.704)) \geq 2(1 - 0.9) = 0.2 > 0.05$$

Altså accepteres hypotesen og modellen er nu

$$H_1: x_i \sim N(15.943, \sigma^2)$$

Dus. at målingerne giver ikke anledning til at gribe ind i processen.

3)

Juf. (3.14) er et 95% konfidensinterval for σ^2 :

$$\left[\frac{f s^2}{\chi^2_{97.5}(f)}, \frac{f s^2}{\chi^2_{2.5}(f)} \right]_{3.21} = \left[\frac{(9-1) \frac{1}{9-1} (U_{SS} - \frac{S^2}{9})}{\chi^2_{97.5}(9-1)}, \frac{(9-1) \frac{1}{9-1} (U_{SS} - \frac{S^2}{9})}{\chi^2_{2.5}(9-1)} \right]$$

$$= \left[\frac{2289.017033 - \frac{143.531^2}{9}}{17.5}, \frac{2289.017033 - \frac{143.531^2}{9}}{2.18} \right] \approx [0.00879; 0.7059]$$

4)

Værdien 0.000074 ligger ikke i konfidensint. for σ^2 (se 3) og dermed kan variansen σ^2 ikke antages at være 0.000074.


```
OPTIONS NODATE PS=40 LS=72;
```

```
DATA hjerne;
```

```
INFILE 'H:\Documents and Settings\Ejer\Skrivebord\EKSAMENSOPGAVER\opgaver\kapitel_02\opgave_02.d
```

```
INPUT id sex $ weight age height N ln_N;
```

```
LABEL N='number of neurons/10**9'
```

```
HEIGHT='height in cm'
```

```
WEIGHT='weight in kg';
```

```
RUN;
```

```
/* Data fra Pakkenberg & Gundersen, J. Comp. Neurology, 1997 */
```

```
PROC PRINT DATA=hjerne (OBS=10) LABEL;
```

```
RUN;
```

```
/*
```

```
Med (OBS=10) udskrives kun observationer 1 til 10.
```

```
Med LABEL bestilles at labels skal bruges i stedet for  
variabelnavne i udskriften.
```

```
*/
```

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

Obs	id	sex	weight in kg	age	height in cm	number of neurons/ 10**9	ln_N
1	95	F	68	18	177	10.9521	2.39353
2	33	F	56	30	157	10.3261	2.33468
3	4	F	88	32	180	9.6636	2.26837
4	50	F	58	35	168	9.6400	2.26592
5	96	F	61	38	165	8.4899	2.13888
6	44	F	66	39	163	13.3586	2.59216
7	114	F	53	43	160	7.3590	1.99592
8	98	F	84	45	169	10.7028	2.37051
9	99	F	72	47	162	7.6997	2.04118
10	109	F	56	52	167	10.0783	2.31039

```

/*
Dette program - opgave_2b.sas - udfoerer noejagtigt det samme som
programmet opgave_2b.sas, med der er forsoegt at oeve indflydelse
paa udseendet af graferne og det er til en vis grad forklaret
hvordan det er opnaaet.
*/

GOPTIONS RESET=ALL; /* alle GOPTIONS saettes til deres default vaerdier */
GOPTIONS KEYMAP=ISO8859 DEVMAP=ISO8859;

/*
med denne GOPTIONS fortaeller man, at det er ISO8859 tegnsaettet
man bruger, og saa kan de danske karakterer æ,ø,å,Æ,Ø,Å udskrives.
I WINDOWS versionen skal 'ISO8859' erstattes af 'winansi'.
*/

GOPTIONS TARGETDEVICE=PS300A4
        HORIGIN=2 cm VORIGIN=2 cm
        ROTATE=landscape
        CBACK=white GUNIT=pct HTITLE=4 HTEXT=2.7
        FTEXT=swiss CTEXT=black CSYMBOL=black;
OPTIONS NODATE PS=40 LS=72;

/*
vigtigste af ovenstaaende GOPTIONS er
CSYMBOL=black, som tvinger SAS til at skelne mellem
formen paa plotsymbolerne i stedet for paa farven.
Det sidste er nemlig til ringe nytte, naar der bagefter
udskrives paa en sor-hvid printer.
GUNIT=pct som definerer at stoerrelsesangivelser (feks HTITLE=4)
er i procent af plotarealet */

DATA hjerne;
/* Data fra Pakkenberg & Gundersen, J. Comp. Neurology, 1997 */
INFILE 'H:\Documents and Settings\Ejer\Skrivebord\EKSAMENSOPGAVER\opgaver\kapitel_02\opgave_02.c
INPUT id sex $ weight age height N ln_N;
LABEL n='number of neurons/10**9'
        height='height in cm'
        weight='weight in kg'
        ln_N= 'ln(number of neurons/10**9)'
        age= 'age in years';

RUN;
/* Man kan knytte en label til sine variable. Denne label
bliver benyttet paa grafer og i procedurer, og derved kan
man have korte variabelnavne, som er lette at skrive og
alligevel faa laesbare grafer og udskrifter. */

PROC PRINT DATA=hjerne (OBS=10) LABEL;
RUN;
/*
Med (OBS=10) udskrives kun observationerne 1 til 10.
Med LABEL bestilles at labels skal bruges i stedet for
variabelnavne i udskriften.
*/

%INCLUDE 'H:\Documents and Settings\Ejer\Skrivebord\EKSAMENSOPGAVER\macros\normgraf.mac';

```



```

/*
  NORMGRAF, som laver fraktildiagrammer, har den ekstra
  facilitet, at man kan definere, hvilken 1.akse (dataakse),
  der skal bruges af NORMGRAF. Det sker ved at definere axis2
  inden kaldet af normgraf.

```

Naar data er opdelt i grupper laves desuden en tegning med fraktildiagrammerne for alle grupper. Desuden laver normgraf om paa stoerrelsen af plotsymboler.

I nedenstaaende eksempler paa brugen af NORMGRAF er valget af order=() - som kontrollerer akseinddelingen - inspireret ved at se paa resultatet af tidligere koersler.

```

*/

```

```

axis2 label=(h=3 pct) value=(h=2.7 pct) width=2.0
      order=(130 to 200 by 10);
%normgraf(data=hjerne, var=height, bygrp=sex,grafik=ja);
axis2; /* herved nulstilles axis2 definitionen */

```

```

axis2 label=(h=3 pct) value=(h=2.7 pct) width=2.0
      order=(20 to 160 by 20);
%normgraf(data=hjerne, var=weight, bygrp=sex,grafik=ja);
axis2;

```

```

axis2 label=(h=3 pct) value=(h=2.7 pct) width=2.0
      order=(7 to 16 by 1);
%normgraf(data=hjerne, var=N, bygrp=sex,grafik=ja);
axis2;

```

```

axis2 label=(h=3 pct) value=(h=2.7 pct) width=2.0
      order=(1.8 to 2.8 by .2);
%normgraf(data=hjerne, var=ln_N, bygrp=sex,grafik=ja);
axis2;

```

```

PROC GLOT DATA=hjerne;
  PLOT ln_N*sex;
RUN;

```

```

TITLE1 'Sammenhæng mellem antal neuroner og alder';
PROC GLOT DATA=hjerne;
  PLOT ln_N*age=sex;
RUN;

```

```

/* regressionslinjer indtegnet */
SYMBOL1 i=r1 v=dot l=1 h=2;
SYMBOL2 i=r1 v=plus l=3 h=2;
TITLE1 'Sammenhæng mellem antal neuroner og alder';

```

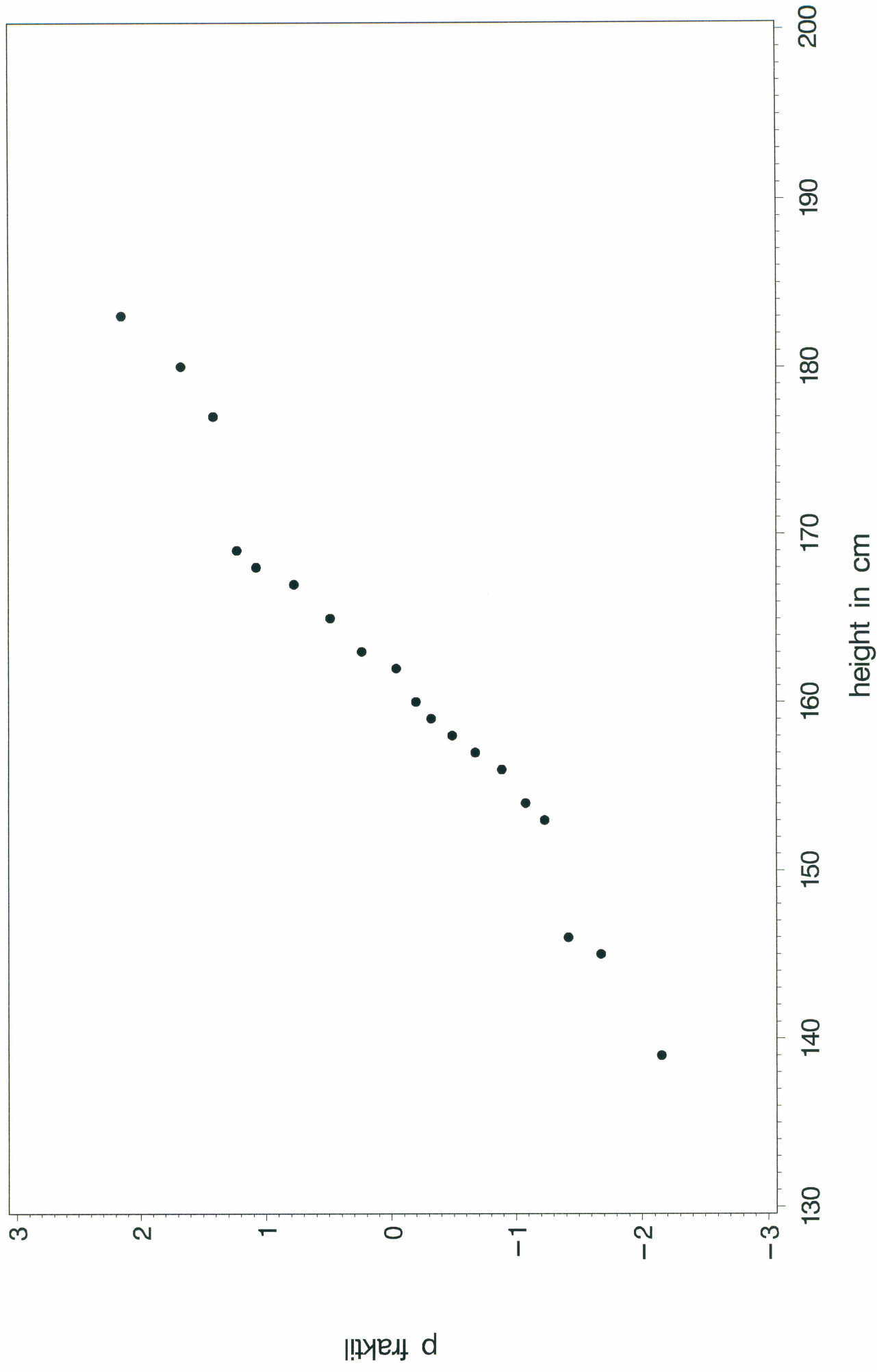
```
PROC Gplot DATA=hjerne;  
    PLOT ln_N*age=sex;  
RUN;
```

```
/* Hvis man vil rotere label paa andenaksen kan man gaa ind og  
kontrollere andenaksen (VAXIS). Foersteaksen hedder HAXIS og  
kan ogsaa kontrolleres
```

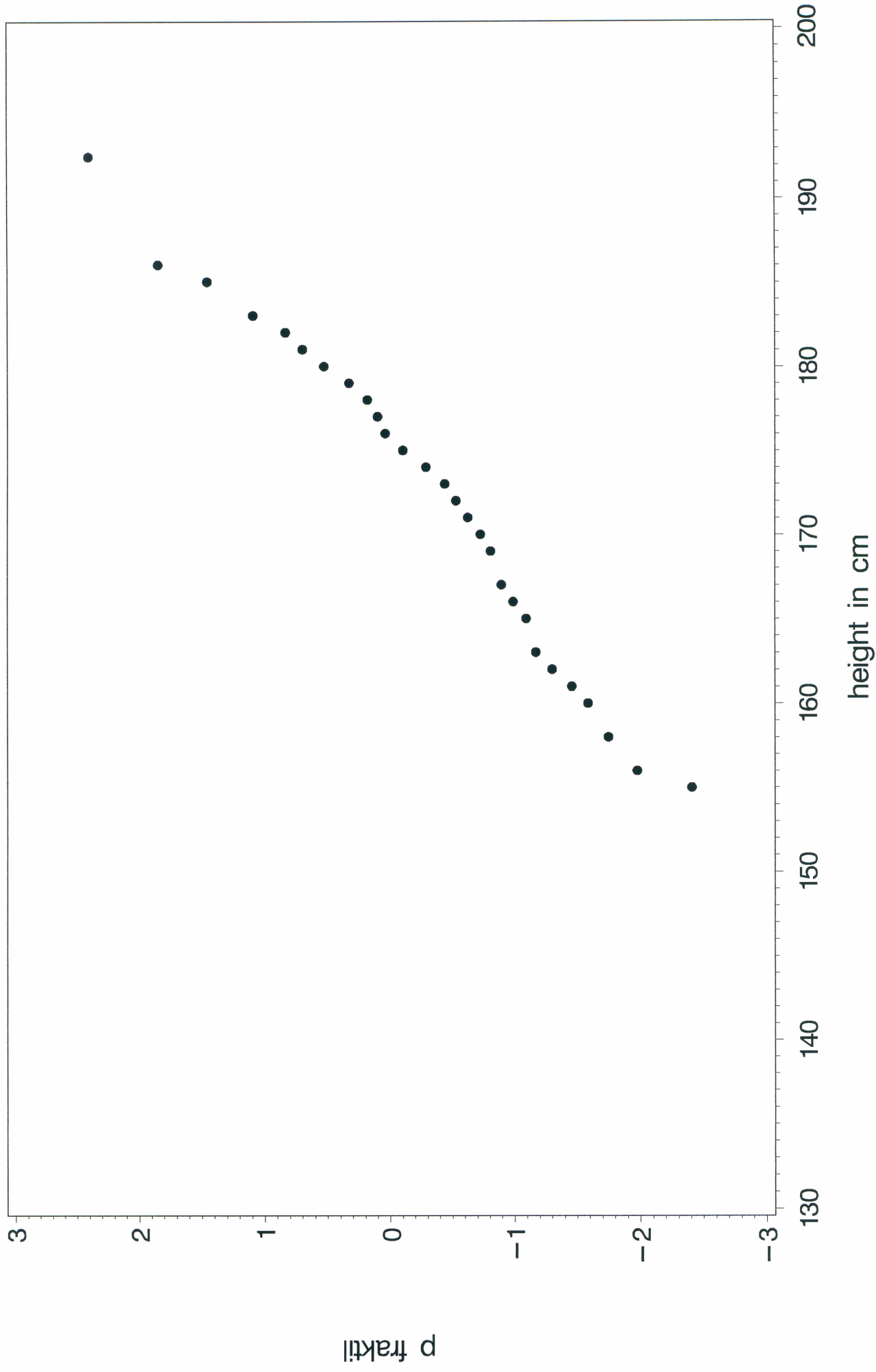
```
*/  
axis3 label=(angle=90);  
SYMBOL1 i=r1 v=dot l=1 h=2;  
SYMBOL2 i=r1 v=plus l=3 h=2;  
TITLE1 'Sammenhæng mellem antal neuroner og alder';  
PROC Gplot DATA=hjerne;  
    PLOT ln_N*age=sex/  
    VAXIS=axis3;
```

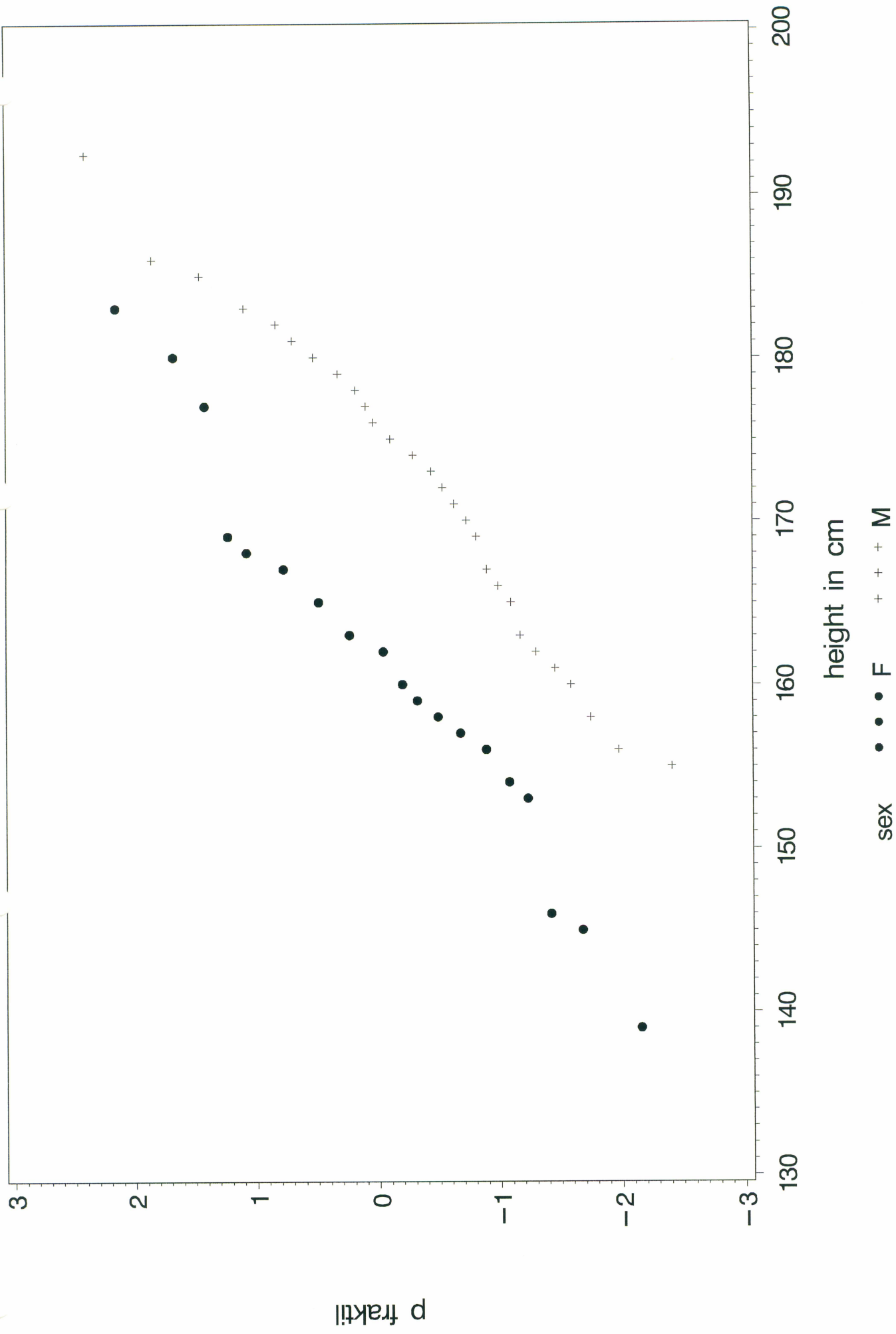
```
RUN;  
SYMBOL1;  
SYMBOL2;  
/* symboldefinitioner forbliver i kraft saa længe SAS  
koerer. Hvis man vil sikre sig mod overraskelser kan man  
genstille sine symboldefinitioner efter brug paa denne maade  
*/
```

sex=F

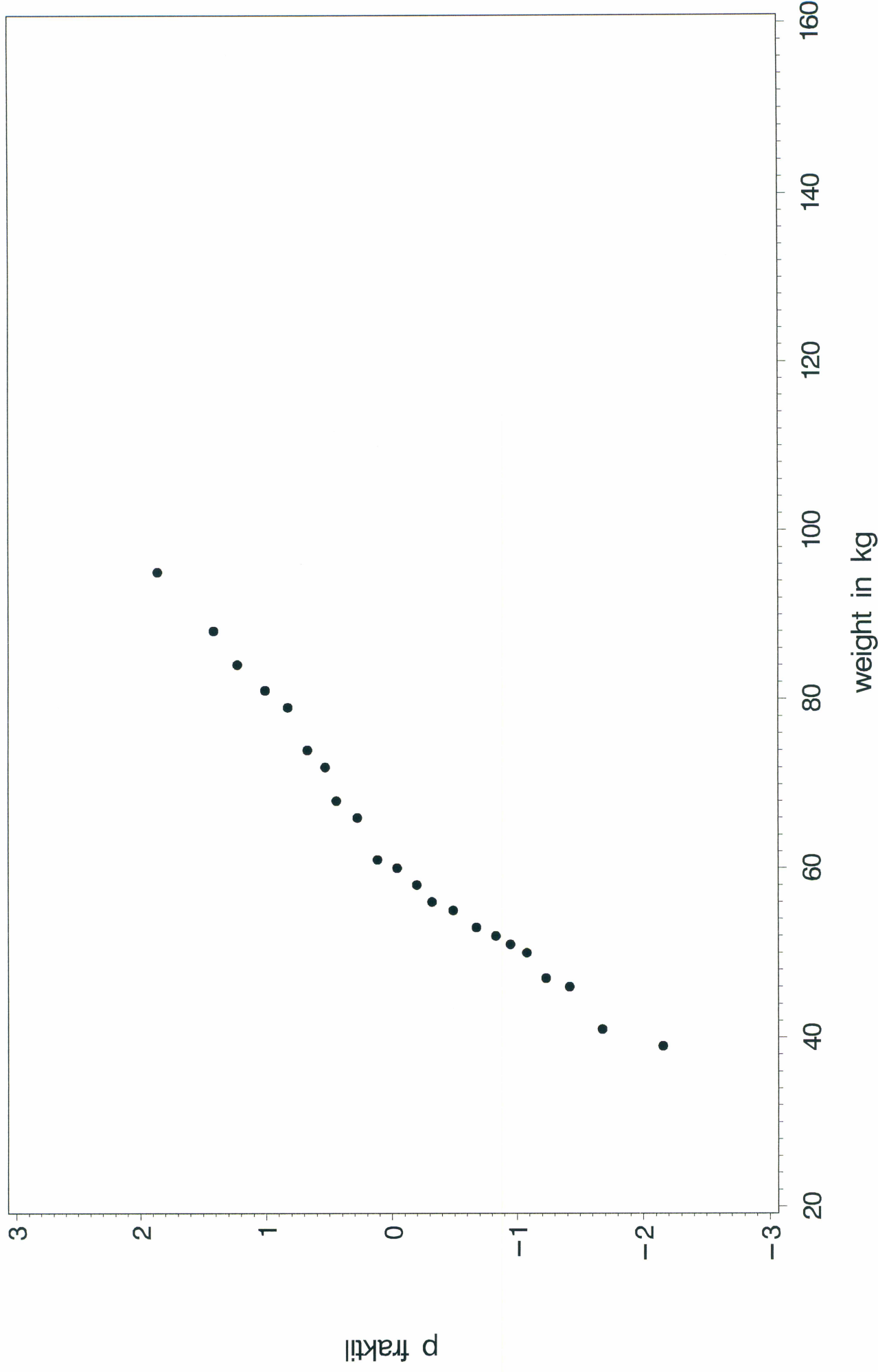


sex = M

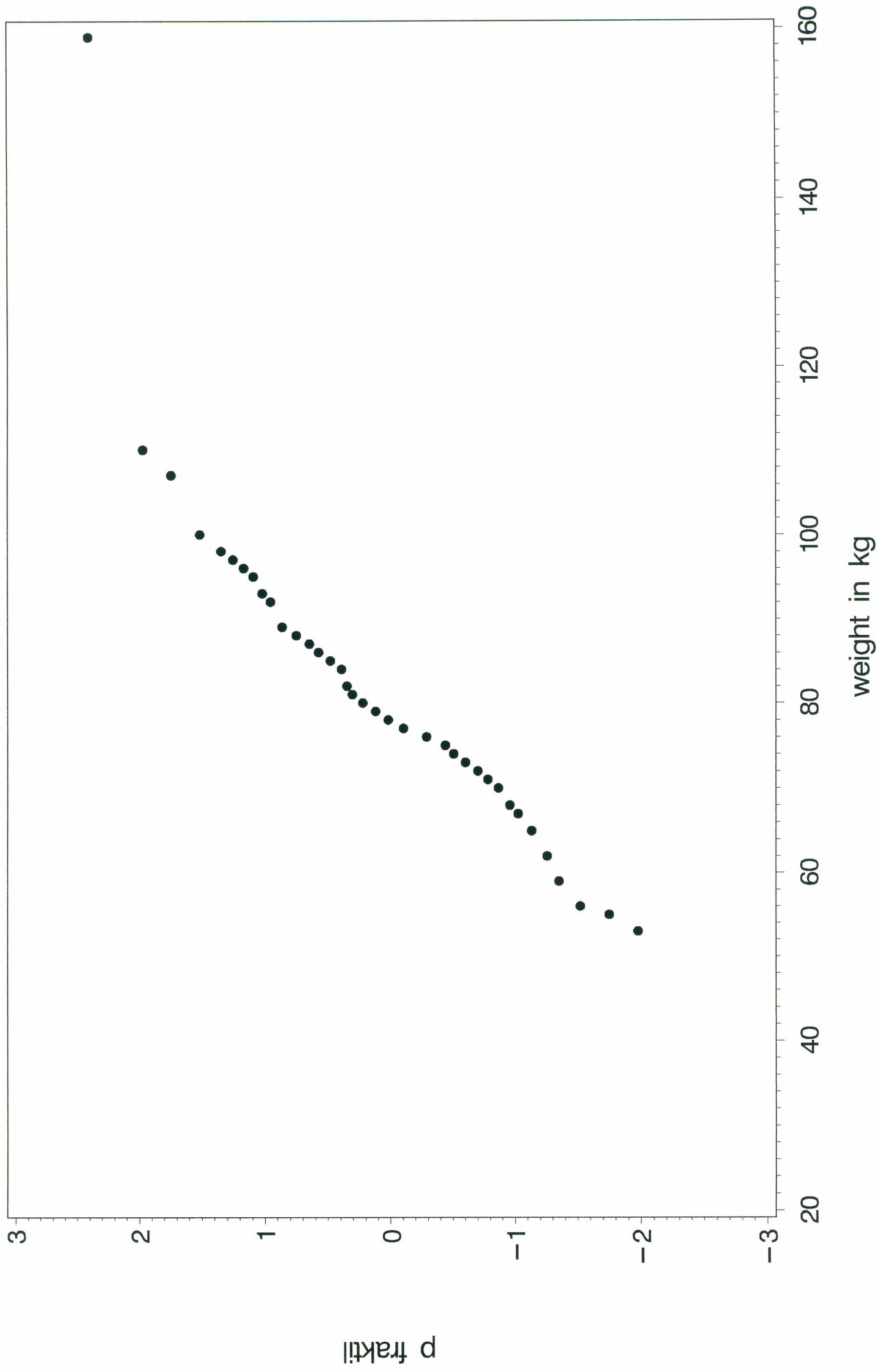


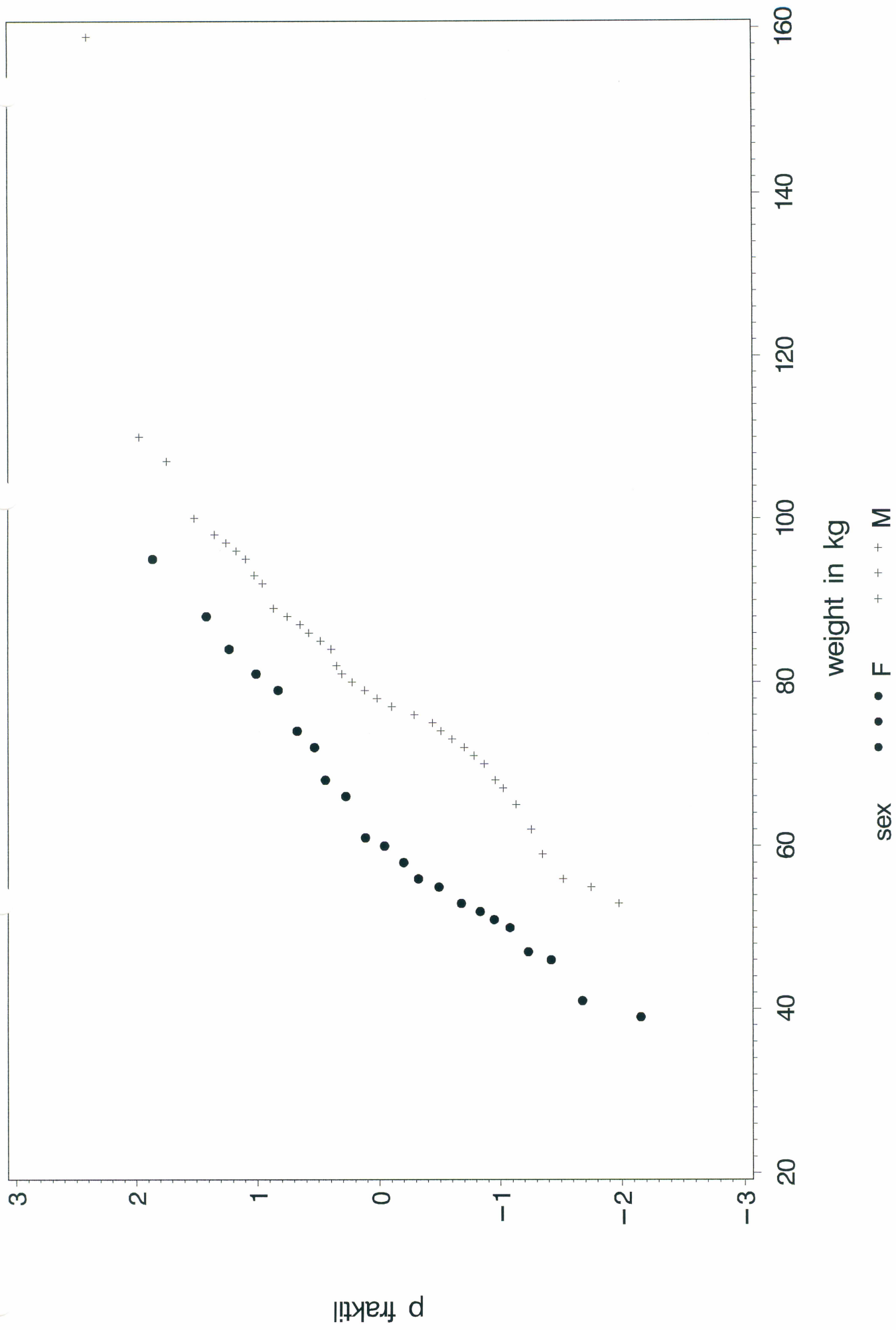


sex = F

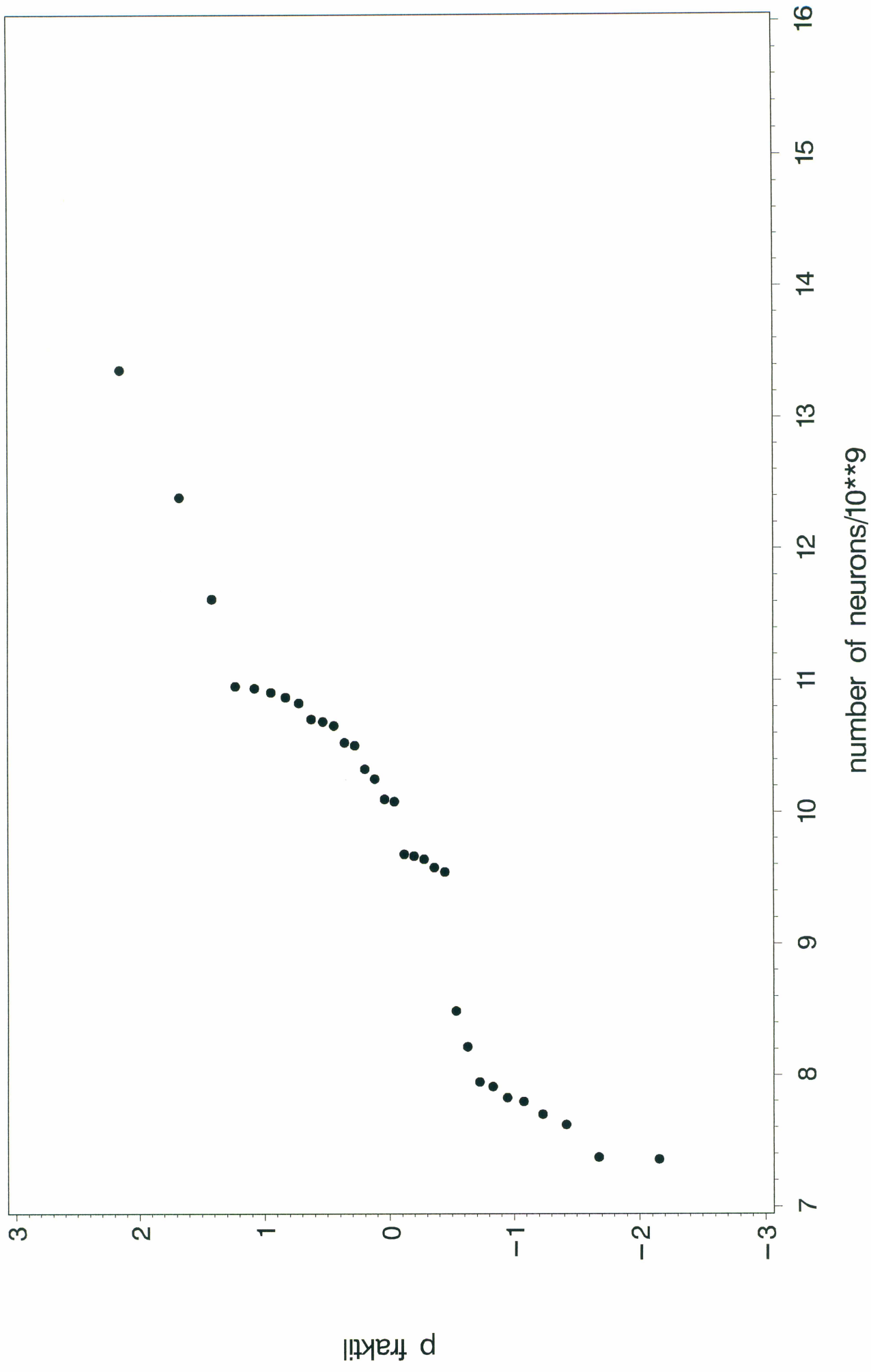


sex = M

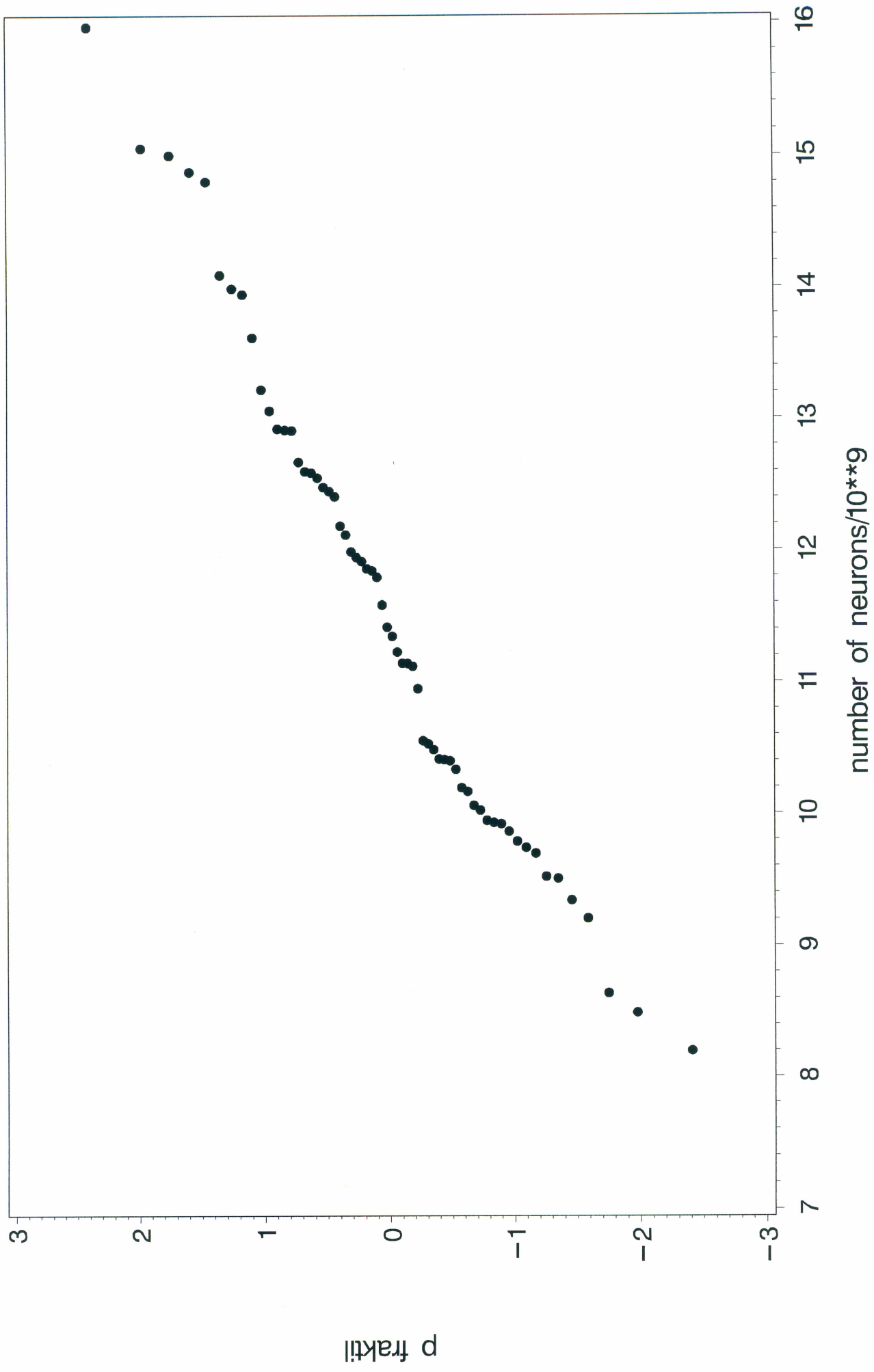


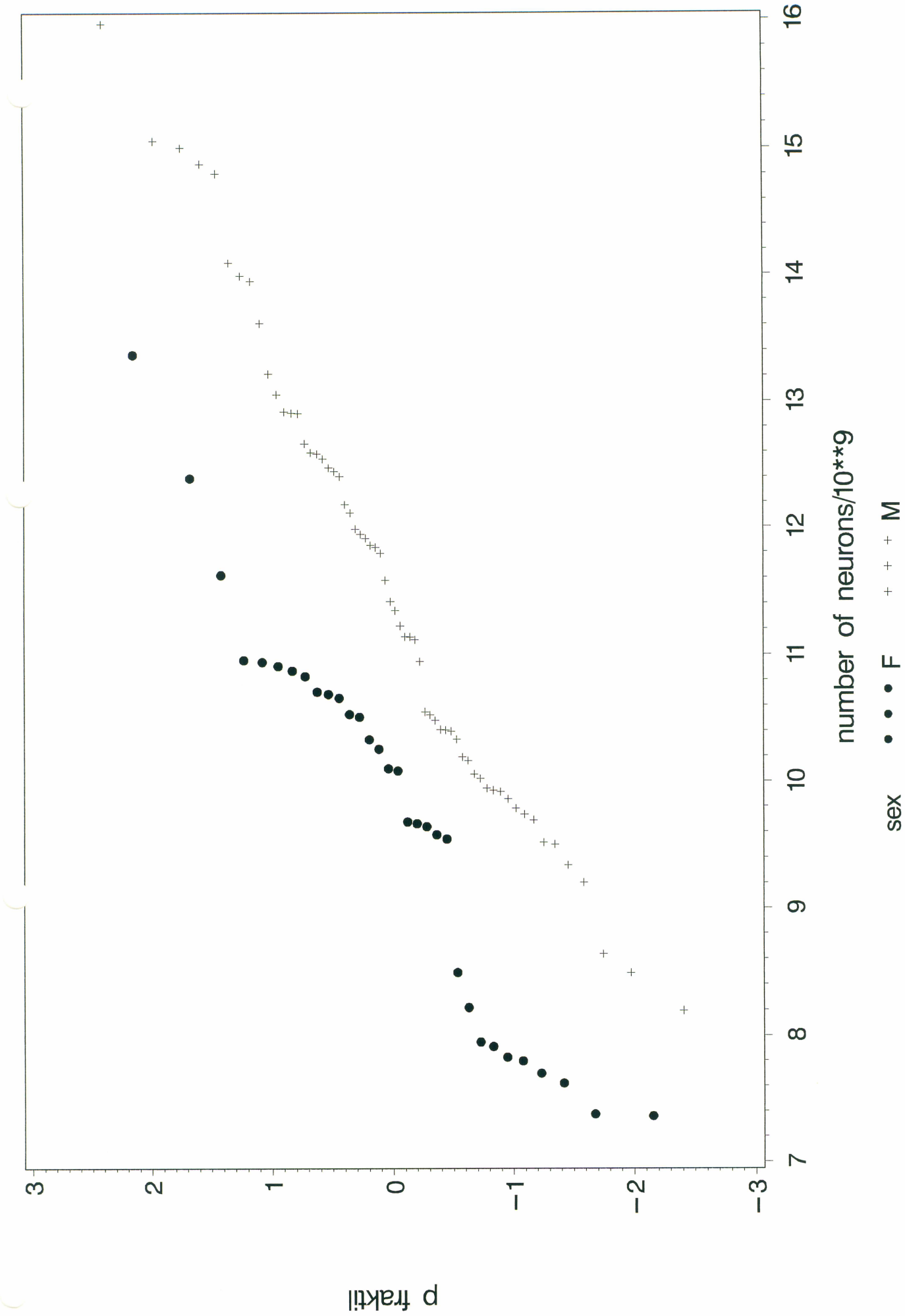


sex = F

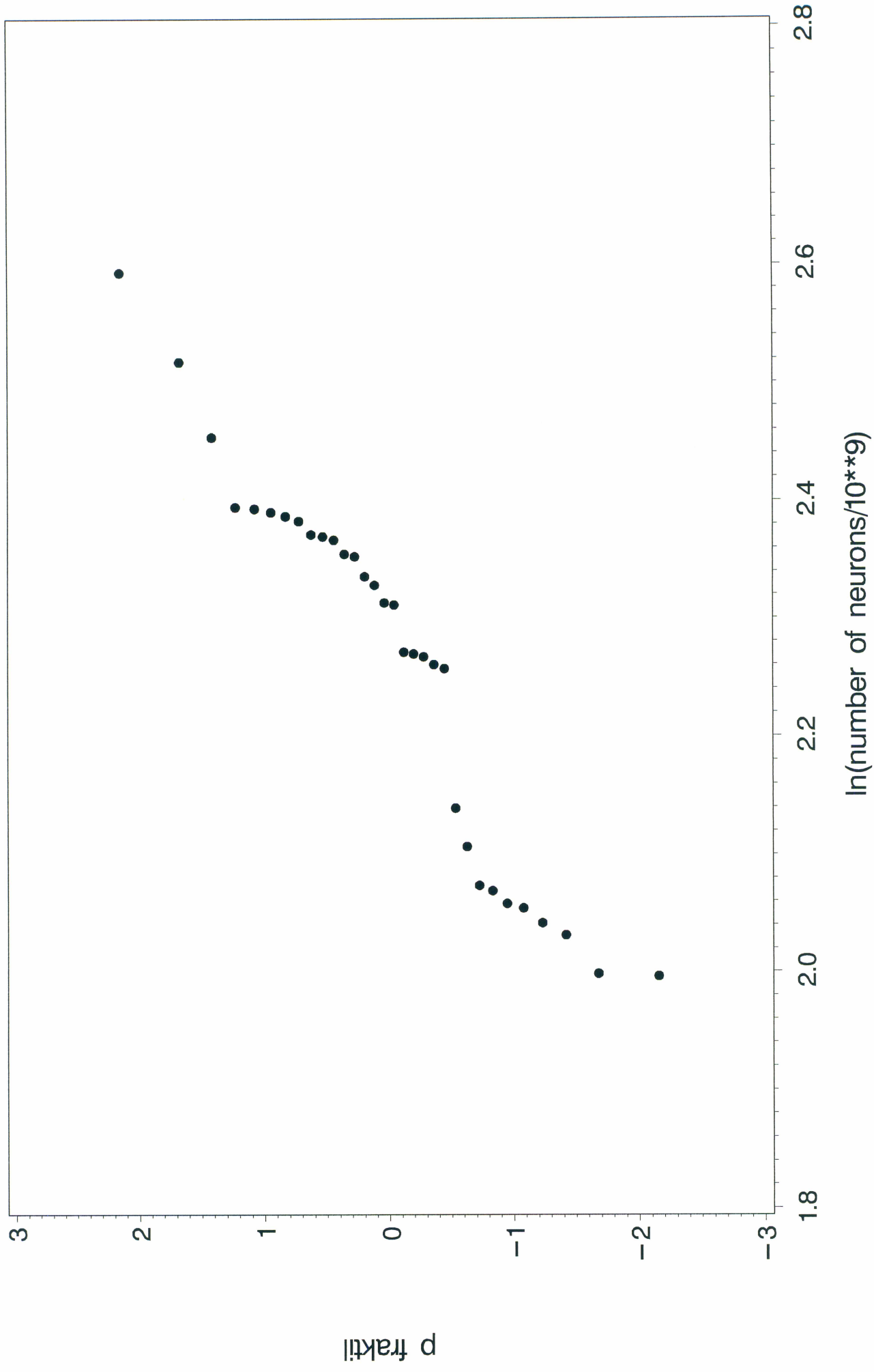


sex = M

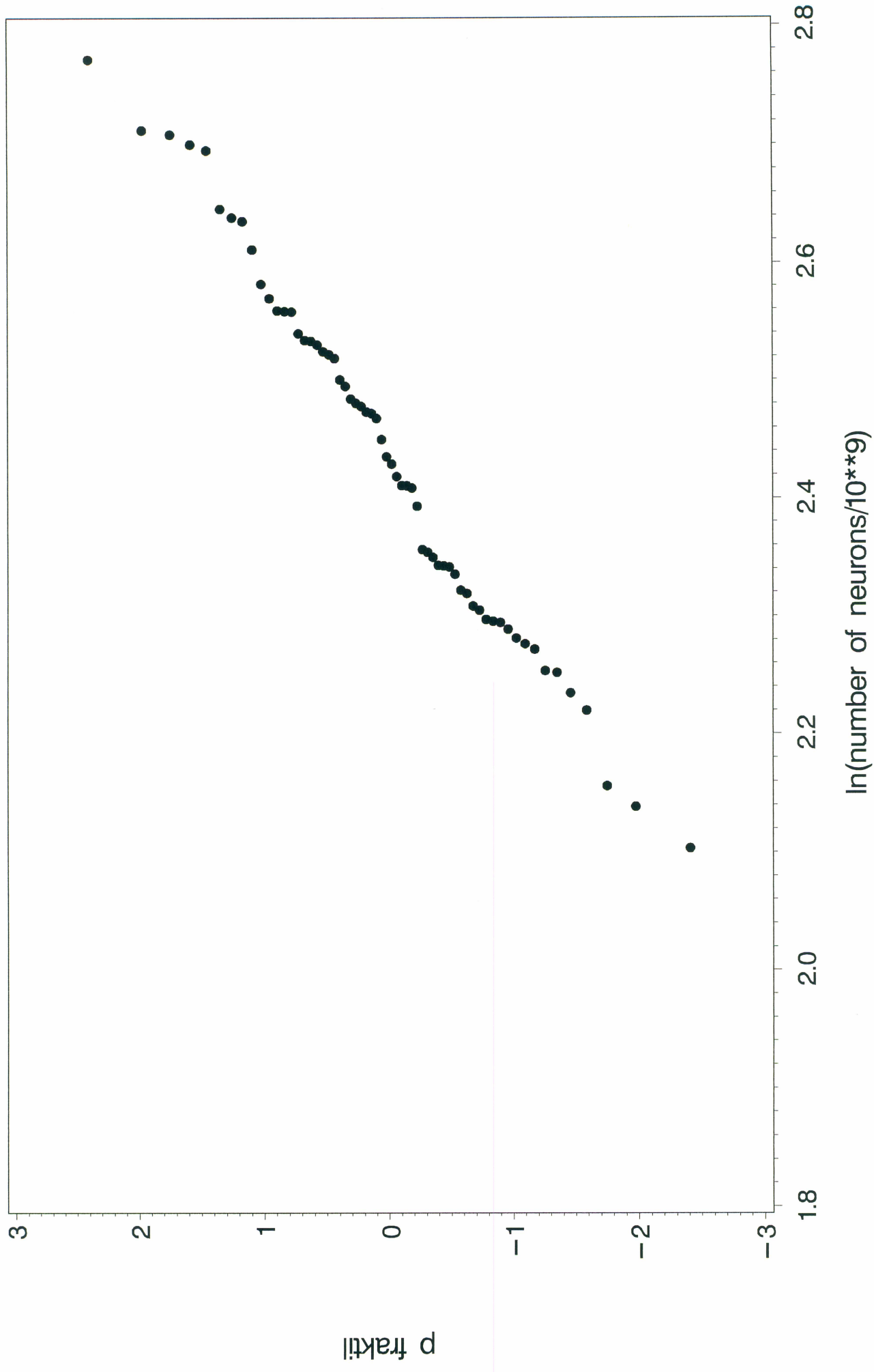




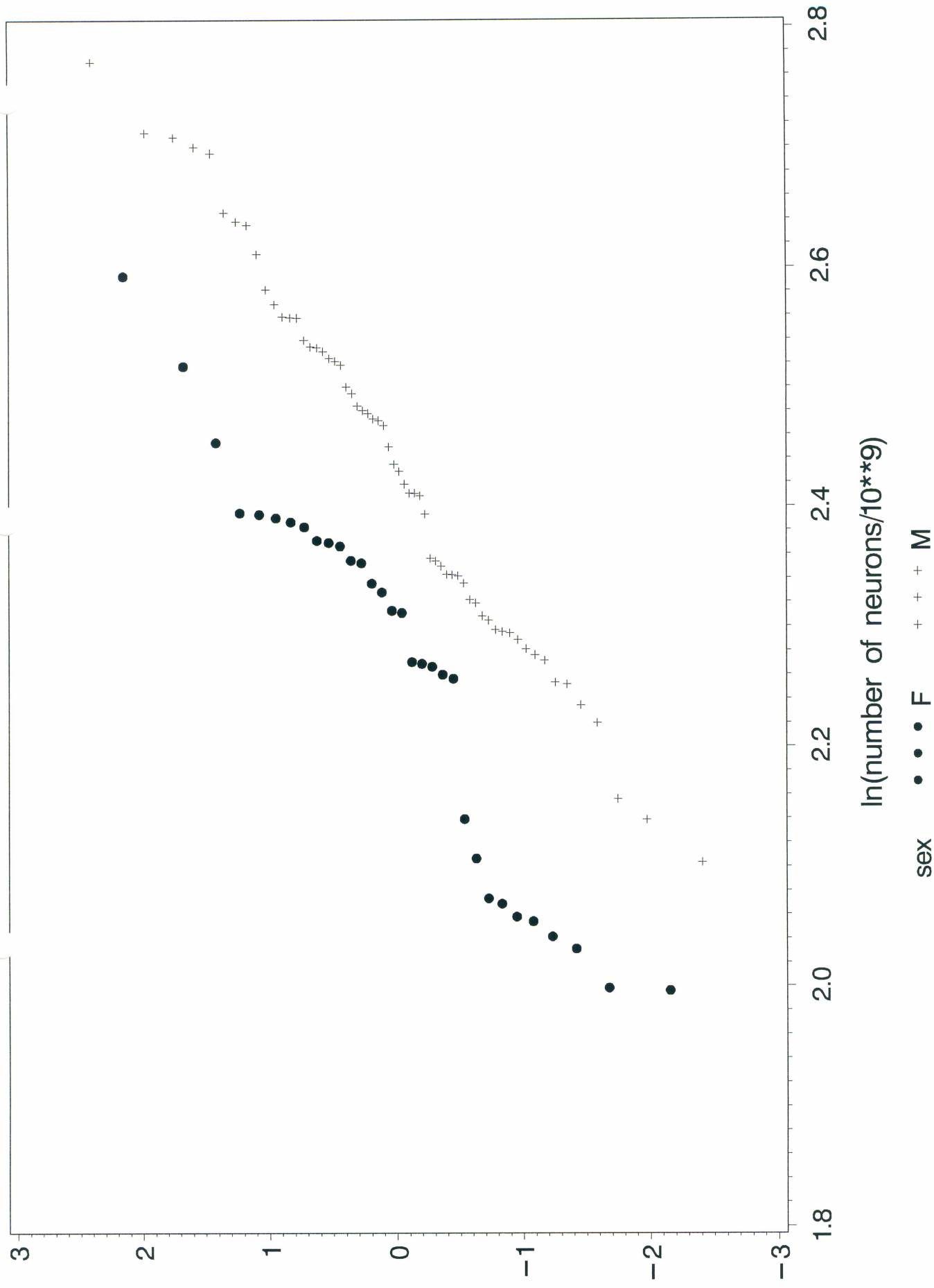
sex = F

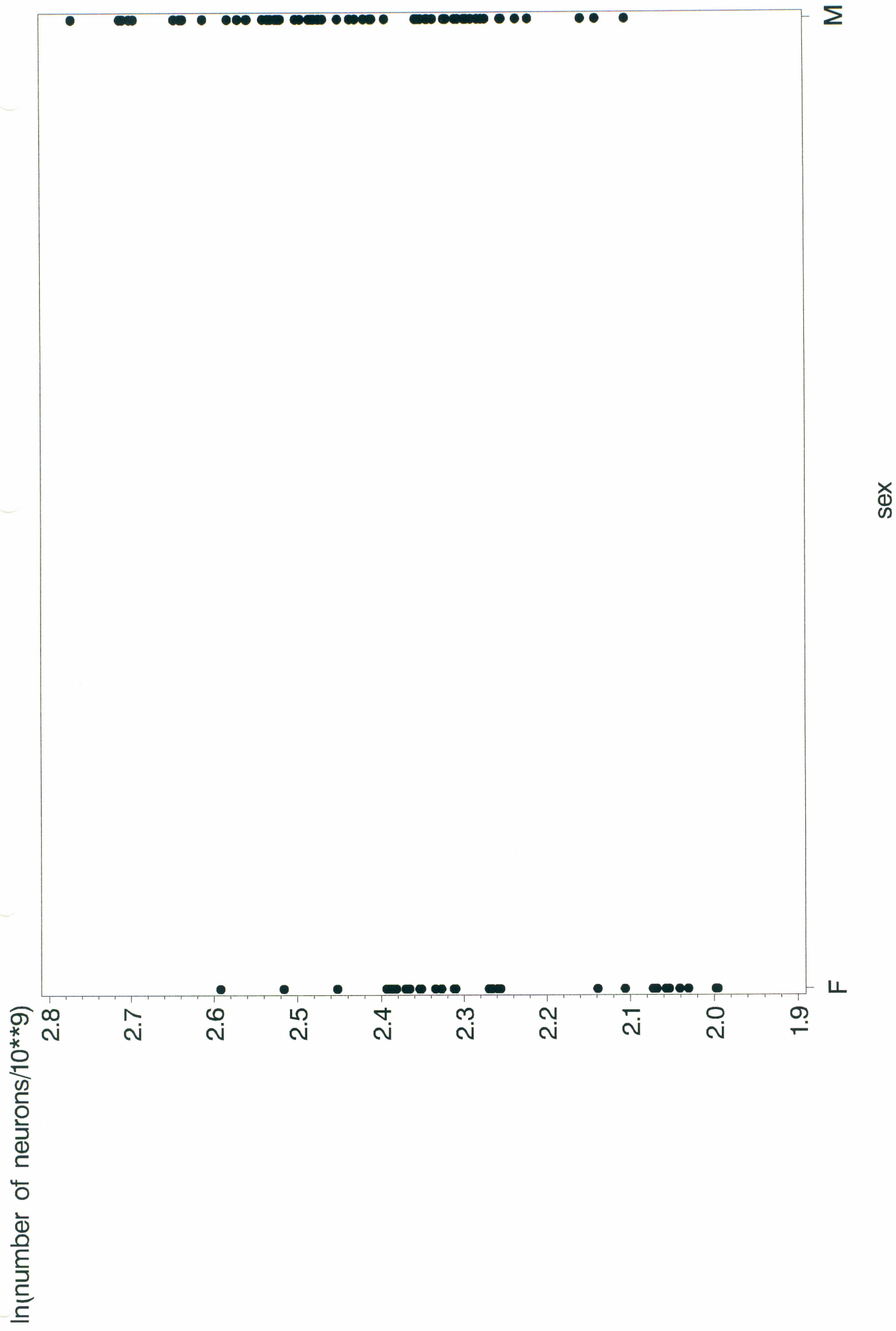


sex = M

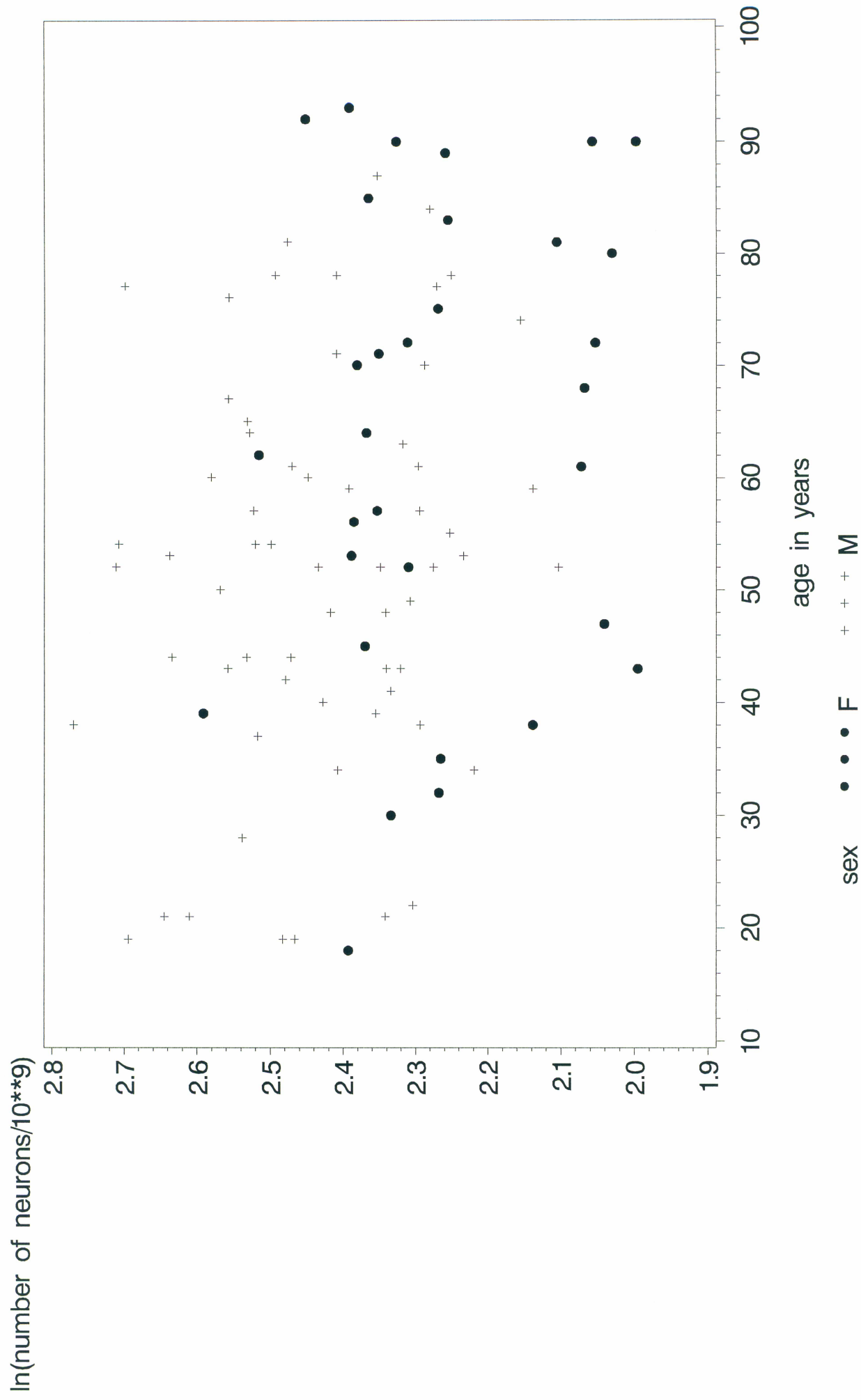


p fraktíl

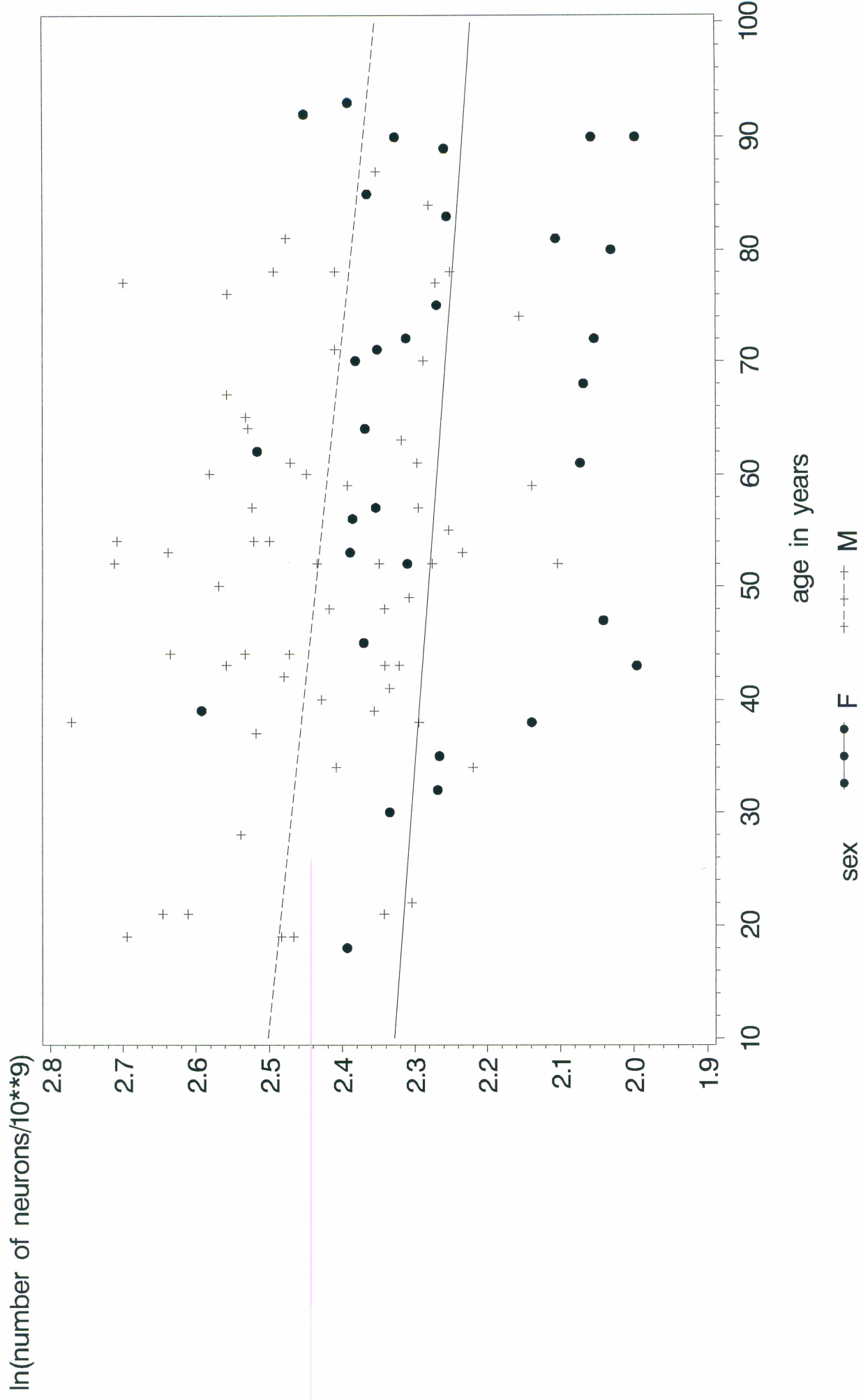




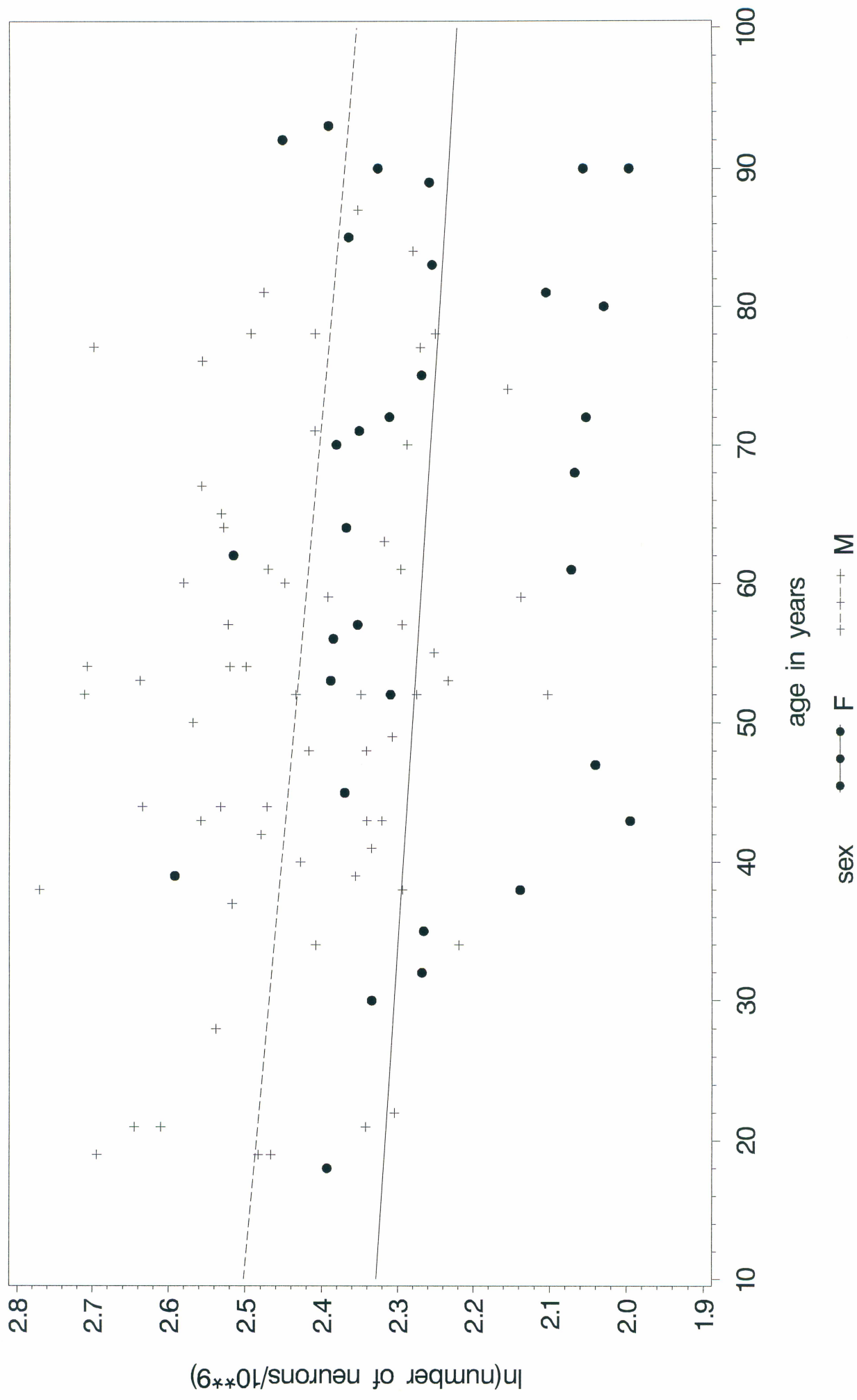
Sammenheng mellom antal neuroner og alder



Sammenheng mellom antal neuroner og alder



Sammenheng mellom antal neuroner og alder



Meddelelse 2

Forelæsningerne i denne uge (36: 2.9.-8.9)

(Onsdag den 4. september) Frank Dabelstein fortalte i 25 minutter om EDB systemet på Institut for Matematiske Fag. Efter nogle indledende bemærkninger gik jeg direkte i gang med Kapitel 2 om *Indledende undersøgelser*.

(Fredag den 6. september) Jeg fortsætter med Kapitel 2 og forventer at nå ind i Afsnit 2.4 *Fraktildiagrammer for normalfordelingen*.

Forelæsningerne i næste uge (37: 9.9.-15.9)

(Onsdag den 11. september) Det er planen at afslutte gennemgangen af Kapitel 2 ved at gennemgå Afsnit 2.5 *Transformation*. Hvis der er mere tid starter gennemgangen af Kapitel 3.

(Fredag den 13. september) Gennemgangen af Kapitel 3 begynder i Supplement til Kapitel 3, hvor vi nævner vigtige resultater om normalfordelingen og afledte fordelinger.

Anneks til Kapitel 2 *Beregninger i SAS* vil blive omtalt ved en senere lejlighed. Men i Annekset beskrives, hvordan de grafiske metoder, som indføres i Kapitel 2 kan udføres i SAS. Programmer til alle eksempler i Kapitel 2 kan hentes på nettet, se nedenfor på denne meddelelse.

Øvelserne i denne uge

(Uge 36: 2.-8. september) Første øvelsesgang bruger vi til en introduktion til EDB systemet, herunder oprettelse af brugernavne, og en introduktion til programpakken SAS.

Øvelserne i de to næste uger

(Uge 37: 9.-15. september) Anden øvelsesgang bruger vi den første halvdel af øvelsestiden til at regne opgaver og fortsætter med introduktionen til SAS i den anden halvdel. Holdene mødes altså i øvelseslokalerne og bevæger sig til EDB-lokalet senere.

Programmet omfatter: Opgaverne 2.1, 2.2 og 3.1 (kun spørgsmål (1)).

I EDB øvelserne fortsætter hvor man slap med opgaverne bag i Introduktion til SAS. Opgave 2.2 er tænkt som et oplæg til diskussion i forbindelse med EDB øvelserne.

(Uge 38: 16.-23. september) Opgaver i tilknytning til Supplement til Kapitel 3: 3.21, 3.22, 3.23, 3.24. Hvis der er mere tid kan resten af øvelsestiden bruges i EDB lokalet.

Udleveret:

Program for øvelserne i uge 36 (2.9-8.9) blev udleveret ved øvelserne.

Netbrowser og filer på nettet

SAS programmer til eksempler og opgaver ligger under brugernavnet `statbib`. Nemmeste adgang er via en browser. Vigtige adresser er

`http://home.imf.au.dk/statbib/sasintro`

Specielt for øvelserne vedrørende introduktionen til SAS, for her ligger alle eksemplerne, der skal køres igennem. Bemærk også `autoexec.sas` som ligger her, og indeholder fornuftige opsætninger til grafik.

Vi har lavet nogle SAS macroer til brug for kurset. De kan ses på adressen.

`http://home.imf.au.dk/statbib/sasmacros`

Det kan være en fordel at kunne se starten af en macro, hvis man ikke lige kan huske alle kodeordene. Var det nu `bygroup` eller `gruppe` man skulle skrive?

For Statistik 1 er adresserne til bogens eksempler og opgaver, samt meddelelserne

`http://home.imf.au.dk/statbib/atiskurser/stat1/eksempler`

(fordelt på kapitler)

`http://home.imf.au.dk/statbib/atiskurser/stat1/opgaver`

(ligeledes fordelt på kapitler) og

`http://home.imf.au.dk/statbib/atiskurser/stat1/meddelelser/aktuelle.`

For Statistik α er de tilsvarende adresser (det er nøjagtigt de samme filer som `stat1`'erne ser)

`http://home.imf.au.dk/statbib/atiskurser/statalfa/eksempler`

(fordelt på kapitler)

`http://home.imf.au.dk/statbib/atiskurser/statalfa/opgaver`

(ligeledes fordelt på kapitler) og

`http://home.imf.au.dk/statbib/atiskurser/statalfa/meddelelser/aktuelle.`

Det er bekvemt at kunne se eksemplerne i en browser og så kopiere relevante dele med musen over i SAS editoren.

Valg af browser

`netscape` er fuldt tilstrækkelig til øvelserne, og den åbner hurtigt.

Meddelelse 1

Tid og sted

Der er 3 forelæsninger hver uge på følgende tidspunkter og sted:

onsdag	11-12	Institut for Matematiske Fag, auditorium D1 (1-531-113)
fredag	12-14	Institut for Matematiske Fag, auditorium D1 (1-531-113)

Desuden er der 3 timers øvelser hver uge. Alle øvelserne finder sted i lokaler på Institut for Matematiske Fag eller på Fysisk Institut, bygning 520. Nedenfor er angivet tidspunkter for de enkelte hold samt lokale for de teoretiske øvelser. EDB-øvelserne foregår alle i D02 (1-531-015). Nøgle til EDB lokalerne kan mod depositum på 50 kr fås på Informationskontoret på første sal på Institut for Matematiske Fag.

Hold	Ugedag	Tid	Lokale	Computerrum	Instruktør
hold 1 Statistik α	torsdag	14-17	1-531-011	1-531-015	Jacob Jacobsen
hold 2 Statistik α	onsdag	8-11	1-531-011	1-531-015	Berit Lindum
hold 1 Statistik 1	mandag	10-13	1-531-011	1-531-015	Jørgen Vinsløv Hansen
hold 2 Statistik 1	fredag	9-12	1-531-011	1-531-015	Morten Fenger Grøn

Eksamen og obligatoriske opgaver

I løbet af kurset skal der afleveres 3 obligatoriske opgaver, som skal godkendes før man kan gå til eksamen. Kurset afsluttes med en 4-timers skriftlig eksamen, som bedømmes efter 13-skalaen.

Materiale

Lærebogen er Preben Blæsild og Jørgen Granfeldt: *Statistik, Bind 1 og 2*. Begge bind er til salg i Studiebogladen i vandrehallen.

Desuden skal vi benytte Preben Blæsild: *Statistical Tables* (ST). En kort introduktion til EDB-systemet og den programpakke, som bruges i kurset, kan ligeledes købes i Studiebogladen. Det er Jørgen Granfeldt: *Introduktion til EDB på Bio- og Geostatistik. Introduktion til SAS*.

Det forudsættes, at deltagerne i kurset har en lommeregner, som gerne må være programmerbar.

EDB-systemet

EDB-systemet bliver vedligeholdt af Frank Dabelstein og Michael K. Sørensen. Skulle der - mod forventning - opstå problemer, kan Frank og Michael træffes på henholdsvis lokal 3538 og 3516 eller på personsøger 3535 og nås med elektronisk post på adressen *staff@imf.au.dk*. Vi henstiller til, at telefonen kun bruges i nødsituationer (systemsammenbrud, printersammenbrud), mens spørgsmål om brug af systemet henvises til elektronisk post.

Forelæsningerne i denne uge (36: 2.9.-8.9)

(Onsdag den 4.9 og fredag den 6.9) Der vil blive givet en introduktion til vigtige statistiske begreber. Formålet er hurtigt at komme i gang med at regne enkle statistikopgaver både i hånden og i SAS. Gennemgangen følger i store træk Kapitel 2 og Kapitel 3.1 i noterne.

Øvelserne i de næste uger

(Uge 36: 2.-9. september) Første øvelsesgang bruger vi til en introduktion til EDB systemet, herunder oprettelse af brugernavne, og en introduktion til programpakken SAS. Man skal medbringe bogen *Introduktion til EDB på Bio- og Geostatistik. Introduktion til SAS*, og man må gerne have orienteret sig i den på forhånd. Man møder altså i det EDB lokale, der er reserveret til holdet.

(Uge 37: 10.-16. september) Anden øvelsesgang bruger vi den første halvdel af øvelsestiden til at regne opgaver og fortsætter med introduktionen til SAS i den anden halvdel. Holdene mødes altså i øvelseslokalerne og bevæger sig til EDB-lokalet senere.

Programmet omfatter: Opgaverne 2.1, 2.2 og 3.1.

Dog gemmes spørgsmål 3 og 4 i Opgave 3.1 til en senere øvelsesgang fordi vi ikke har gennemgået de nødvendige begreber til forelæsningerne.