Vejledende besvarelse af hjemmeopgave, forår 2018

Udleveret 12. februar, afleveres senest ved øvelserne i uge 10 (6.-9.marts)

I forbindelse med reagensglasbehandling blev 100 par randomiseret til to forskellige former for hormonstimulation. Herefter blev der udtaget et eller flere æg samt (i princippet) to ægblærevæsker, en fra hver æggestok.

Kvaliteten af de udtagne æg blev vurderet, og et antal hormoner blev målt i ægblærevæskerne, heriblandt Testosteron (nmol/l) og InhibinB (ng/ml).

For 4 kvinder lykkedes det ikke at opnå nogle æg eller ægblærer, og for visse af de 96 resterende var det (af forskellige årsager) ikke muligt at foretage alle hormonbestemmelserne.

På hjemmesiden
`http://staff.pubhealth.ku.dk/~lts/basal18_1/hjemmeopgave.html`
ligger oplysninger på 96 kvinder, med betegnelserne:

id: Løbenummer for kvinden

treat: Stimulationsmetoden (A eller B)

kvalitet: 1 for god kvalitet, 0 for mindre god

testo: Målinger af Testosteron, hhv. *left* og *right*

inhibin: Målinger af InhibinB, hhv. *left* og *right*

Inden vi går i gang med spørgsmålene, skal vi lige have læst data ind:

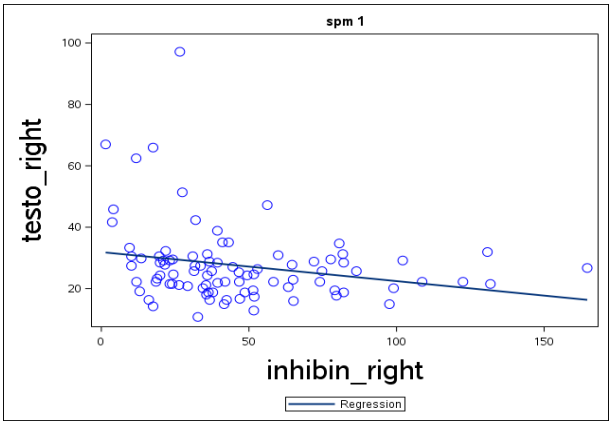
```
FILENAME hjopg URL
"http://publicifsv.sund.ku.dk/~lts/basal18_1/hjemmeopgave/hjemmeopgave.txt";

data a1;
infile hjopg firstobs=2;
input idnr treat$ kvalitet testo_left testo_right inhibin_left inhibin_right;
run;
```

I de første 3 spørgsmål fokuserer vi udelukkende på **prediktion af testosteron målinger på højre side**, og **ignorerer**, at der er benyttet to forskellige stimulationsmetoder.

1. *Lav en figur til at illustrere sammenhængen mellem testosteron og inhibin, begge målt på højre side. Udfør på denne baggrund en lineær regressionsanalyse (***på passende skala***), med testosteron (højre side) som outcome og inhibin (højre side) som forklarende variabel. Angiv estimater for intercept og hældning med konfidensintervaller samt spredning omkring regressionslinien. Giv også, så vidt muligt, fortolkninger af disse.*

   Her er det naturligt at starte med et scatterplot, hvor outcome (testosteron på højre side) sættes på Y-aksen, medens kovariaten (inhibin på højre side) sættes på X-aksen. Nedenfor er tilføjet regressionslinier som støtte for vurderingen af rimeligheden af forudsætningerne for at foretage regressionsanalysen.

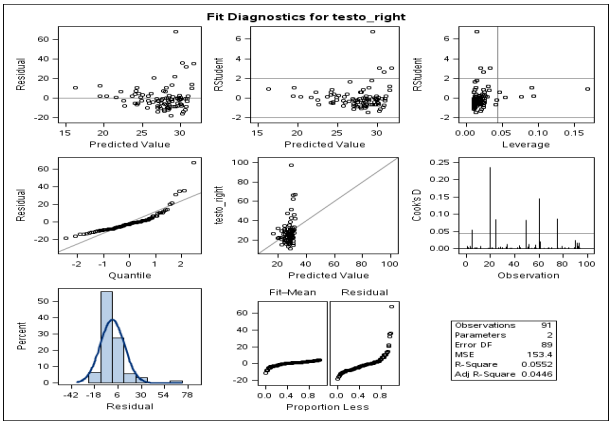
```
title 'spm 1';
proc sgplot data=a1;
reg Y=testo_right X=inhibin_right /
    markerattrs=(color=blue size=0.3cm);
  xaxis labelattrs=(size=0.8cm);
  yaxis labelattrs=(size=0.8cm);
run;
```



Denne figur giver ikke noget klart lineært indtryk, og residualerne fra en regressionsanalyse på disse utransformerede data ville have flere problemer:

- Residualerne har ikke samme spredning for alle niveuer af outcome, men har derimod en tendens til at være større (numerisk, dvs. både store negative og store positive værdier) for de høje (forventede) værdier af outcome.

  Dette illustreres også ved det trompetagtige udseende af plottet af residualer mod forventede værdier for denne regressionsanalyse på utransformeret skala



- Residualerne er ikke normalfordelte, men har en hale mod de store værdier. Dette illustreres bedst ved fraktildiagrammet i figuren ovenfor, der klart har "hængekøjefacon"

I lyset af disse problemer, vil vi logaritme-transformere vores outcome, og vi holder os til 10-tals logaritmen. Der er ikke noget i de ovenstående problemer, der indikerer, at vi skal transformere vores kovariat, men alligevel kan det rent fortolkningsmæssigt være rimeligt at transformere kovariaten også, da der i begge tilfælde er tale om koncentrationer. Vi vil derfor tranformere begge med logaritmen i de efterfølgende analyser.

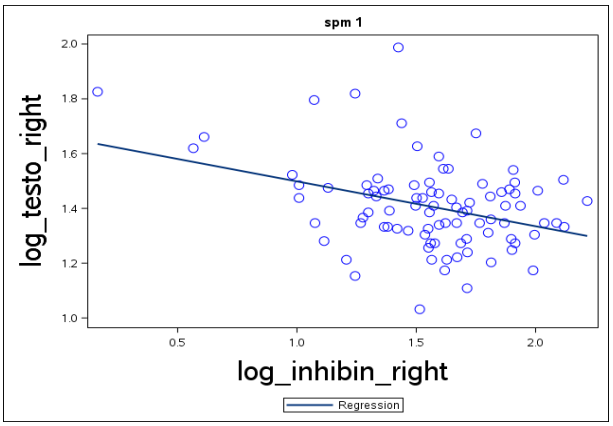
Vi definerer således (inden det første `run;`) alle de logaritmerede værdier (vi får brug for dem efterfølgende):

```
log_testo_left=log10(testo_left);
log_testo_right=log10(testo_right);
```

```
log_inhibin_left=log10(inhibin_left);
log_inhibin_right=log10(inhibin_right);
```

Vi kan nu gentage figuren fra før, denne gang på logaritmisk skala:

```
proc sgplot data=a1;
reg Y=log_testo_right X=log_inhibin_right /
    markerattrs=(color=blue size=0.3cm);
  xaxis labelattrs=(size=0.8cm);
  yaxis labelattrs=(size=0.8cm);
run;
```



Denne figur viser ingen særlige problemer med henblik på en regressionsanalyse, men vi ser nærmere på det ved modelkontrollen nedenfor. Der er dog nogle observationer svarende til de lave inhibinmålinger, der godt kan gå hen og blive indflydelsesrige. Dette kunne have været en grund til *ikke* at logaritmere inhibin....

Vi laver dernæst regressionsanalysen:

```
proc glm plots=(Diagnosticspanel residuals(smooth)) data=a1;
    model log_testo_right = log_inhibin_right / solution clparm;
run;
```

og finder outputtet (beskåret):

```
The GLM Procedure

Number of Observations Read          96
Number of Observations Used          91

Dependent Variable: log_testo_right

                                   Sum of
Source                     DF       Squares   Mean Square  F Value  Pr > F
Model                       1    0.28786547    0.28786547    13.05  0.0005
Error                      89    1.96290261    0.02205509
Corrected Total            90    2.25076807

R-Square     Coeff Var      Root MSE    log_testo_right Mean
0.127897      10.55150      0.148510                1.407473

                                        Standard
Parameter                  Estimate           Error    t Value    Pr > |t|
Intercept              1.662266113      0.07222342      23.02      <.0001
log_inhibin_right     -0.163753077      0.04532617      -3.61      0.0005

Parameter               95% Confidence Limits
Intercept              1.518759708  1.805772518
log_inhibin_right     -0.253815219 -0.073690936
```

Estimatet for interceptet ses at være 1.662 med tilhørende konfidensinterval $(1.519, 1.806)$, men dette er rimeligt meningsløst, idet det refererer til en inhibinværdi på 1 (fordi logaritmen til 1 er 0), og vi vil derfor ikke bruge krudt på at tilbagetransformere det.

Estimatet for hældningen er -0.164, med et konfidensinterval på $(-0.254, -0.074)$, hvilket viser den negative association mellem de to hormoner.

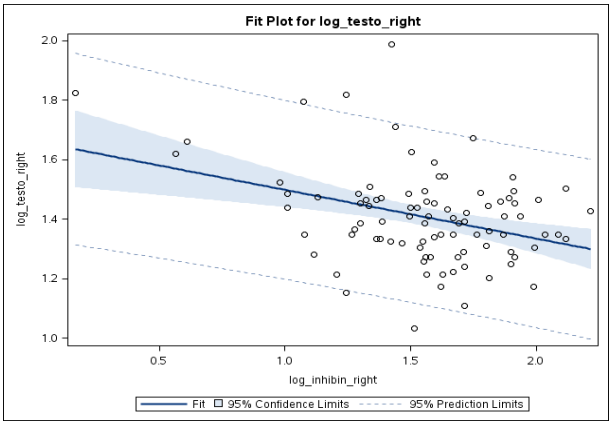
Når man tilbagetransformerer hældningen, får man effekten af at 10-doble inhibinværdien, $10^{-0.164} = 0.69$, altså at testosteron falder med 31% eller til 69% af hvad den var før. Konfidensintervallet for dette er $(10^{-0.254}, 10^{-0.074}) = (0.56, 0.84)$, svarende til et fald i testosteron på mellem 16 og 44%.

Interessant er det, at når begge variable (outcome såvel som kovariat) logaritmetransformeres, så får man samme hældning uanset hvilken logaritme, man benytter (blot man selvfølgelig benytter samme logaritme på begge). Det betyder, at man også direkte kan se effekten af

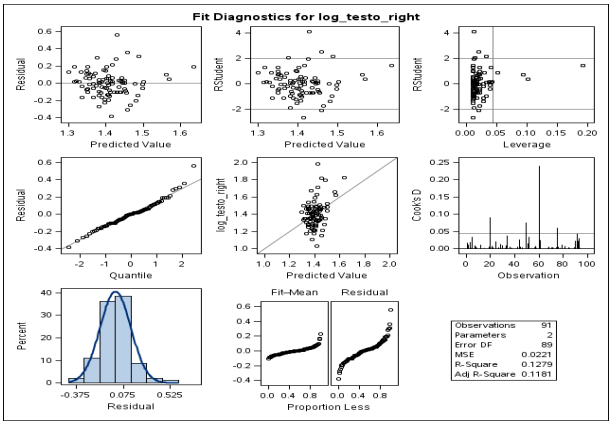
en fordobling af inhibinværdien, nemlig en faktor $2^{-0.164} = 0.89$, på testosteron-værdien, altså et fald på ca. 11%. Konfidensintervallet for dette er $(2^{-0.254}, 2^{-0.074}) = (0.84, 0.95)$, svarende til et fald i testosteron på mellem 5 og 16%.

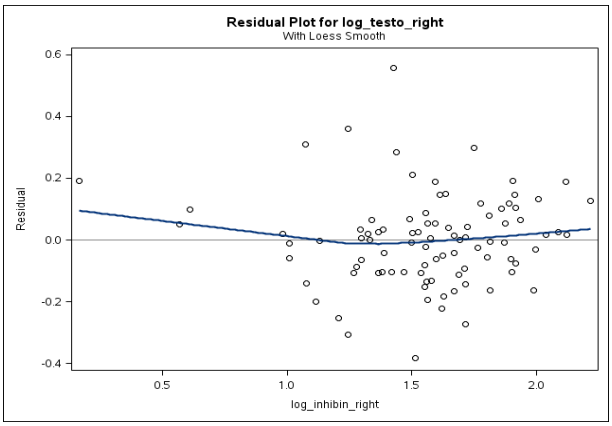
Estimatet for spredningen omkring regressionslinien (`Root MSE`) er 0.149, men det er ikke så let at fortolke. Det benyttes jo til at konstruere prediktionsintervaller, idet disse er givet ud fra linien ved at lægge $\pm 2 \times 0.149 = \pm 0.298$ til. Tilbagetransformerer vi, fås $10^{0.298} = 1.99$ og $10^{-0.298} = 0.504$. Det kan fortolkes som tilbagetransformerede prediktionsgrænser, der siger, at de individuelle testosteron-værdier kan være op til enten det dobbelte eller ned til halvdelen af, hvad prediktionen angiver.

På $\log_{10}$-skala ser vores grænser således ud



Vi skal også lige med et par residualplots (svarende til dem ovenfor for den utransformerede outcome variable) sikre os, at vores model er rimelig:





Det ses, at forudsætningerne ser fornuftige ud, idet residualerne har nogenlunde samme spredning uanset niveauet af outcome variablen, og at de er nogenlunde normalfordelte. Samme konklusion fås i øvrigt fra modelkontrol i tilfælde af, at kovariaten ikke er logaritmetransformeret.

Der ses mindst en observation med ret stor Cook-værdi. Vi kan se nærmere på denne ved at gemme Cook-værdierne i det nye datasæt `check` og bagefter udskrive dem med de største værdier:

```
proc glm plots=(Diagnosticspanel residuals(smooth)) data=a1;
    model log_testo_right = log_inhibin_right / solution clparm;
output out=check cookd=cook;
run;

proc print data=check; where cook>0.05;
run;
```

Obs	idnr	treat	kvalitet	testo_ left	testo_ right	inhibin_ left	inhibin_ right	log_testo_ left
20	20	A	1	62.46	97.16	27.80	26.60	1.79560
49	49	B	1	62.46	62.46	7.49	11.80	1.79560
60	60	B	0	53.79	66.97	24.70	1.46	1.73070
75	75	A	1	48.58	65.93	5.00	17.52	1.68646

Obs	log_testo_ right	dif_log_ testo	log_inhibin_ left	log_inhibin_ right	dif_log_ inhibin	cook
20	1.98749	0.19189	1.44404	1.42488	-0.01916	0.09133
49	1.79560	0.00000	0.87448	1.07188	0.19740	0.07587
60	1.82588	0.09518	1.39270	0.16435	-1.22834	0.24088
75	1.81908	0.13263	0.69897	1.24353	0.54456	0.06160

Den rigtig store Cook-værdi (0.24) ses at tilhøre individ nr. 60, som har en dårlig æg-kvalitet og en meget lille værdi af inhibin på højre side, faktisk den mindste i datamaterialet. Det *kunne* jo være en fejl....

2. *Hvad er den forventede værdi af testosteron for en kvinde med en inhibin på 30? Og hvad er normalområdet for sådanne kvinder?*
Er det usædvanligt at se en kvinde med en testosteron-værdi på 80 og en inhibin-værdi på 30?

Vi tilføjer nu en `estimate`-sætning til vores regressionsanalyse, men da kovariaten inhibin også er logaritmetransformeret, skal vi lige huske at benytte den transformerede værdi i `estimate`-sætningen, altså `log10(30)=1.477`:

```
proc glm plots=(Diagnosticspanel residuals(smooth)) data=a1;
    model log_testo_right = log_inhibin_right / solution clparm;
estimate 'log-testo for inhibin 30' intercept 1 log_inhibin_right 1.477;
run;
```

hvorved vi får det ekstra output

Parameter	Estimate	Standard Error	t Value	Pr > |t|
log-testo for inhibin 30	1.42040282	0.01597409	88.92	<.0001

Parameter	95% Confidence Limits	
log-testo for inhibin 30	1.38866264	1.45214300

Den predikterede testosteron værdi er her $10^{1.4204} = 26.33$ nmol/l, og normalområdet på logaritmisk skala er $1.4204 \pm 2 \times 0.1485 = (1.1234, 1.7174)$. Da dette ikke indeholder $1.90 = log_{10}(80)$ er kombinationen usædvanlig. Vi kan også tilbagetransformere normalområdet til $10^{(1.1234,1.7174)} = (13.29, 52.17)$ nmol/l, som jo ikke indeholder 80.

Bemærk i øvrigt, at det i dette spørgsmål er **vigtigt** om normalfordelingen og den konstante varians (spredning) holder, idet vi udtaler os om enkeltobservationer.

3. *Har vi nogen glæde af at kende de to inhibin målinger, hvis vi også udnytter information om testosteron niveauet på venstre side som forklarende variabel?*

Ovenfor så vi, at højre sides inhibin målinger havde en signifikant evne til at prediktere testosteronmålingerne på højre side, omend vi kun forklarede ca. 13% af variationen. Det virker oplagt, at testosteron målingerne på venstre side skulle prediktere bedre, og vi afprøver nu dette ved at inddrage alle tre mulige kovariater i en regressionsmodel. Samtidig tilføjer vi en `contrast`-sætning (det har I ikke lært), som laver et samtidigt test af, om de to inhibinmålinger kan undværes:

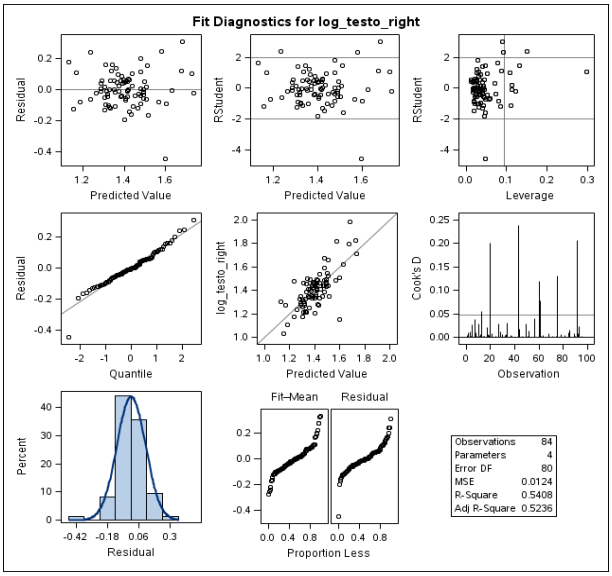
```
proc glm plots=(Diagnosticspanel residuals(smooth)) data=a1;
    model log_testo_right = log_testo_left
          log_inhibin_right log_inhibin_left / solution clparm;
    contrast "begge to" log_inhibin_left 1, log_inhibin_right 1;
output out=check r=residuals;
run;
```

og finder derved outputtet

```
The GLM Procedure

Number of Observations Read          96
Number of Observations Used          84

Dependent Variable: log_testo_right

                                 Sum of
Source                    DF       Squares   Mean Square  F Value  Pr > F
Model                      3    1.17152306    0.39050769    31.40  <.0001
Error                     80    0.99481498    0.01243519
Corrected Total           83    2.16633804
```

```
R-Square     Coeff Var      Root MSE    log_testo_right Mean
0.540785      7.929661      0.111513               1.406279

Source                    DF     Type I SS   Mean Square  F Value  Pr > F
log_testo_left             1    1.13578133    1.13578133    91.34  <.0001
log_inhibin_right          1    0.01791717    0.01791717     1.44  0.2335
log_inhibin_left           1    0.01782456    0.01782456     1.43  0.2347

Source                    DF   Type III SS   Mean Square  F Value  Pr > F
log_testo_left             1    0.85847579    0.85847579    69.04  <.0001
log_inhibin_right          1    0.03385518    0.03385518     2.72  0.1029
log_inhibin_left           1    0.01782456    0.01782456     1.43  0.2347

Contrast                  DF   Contrast SS   Mean Square  F Value  Pr > F
begge to                   2    0.03574173    0.01787087     1.44  0.2437

                                          Standard
Parameter                 Estimate           Error    t Value    Pr > |t|

Intercept           0.4770872892      0.15979744       2.99      0.0038
log_testo_left      0.6841443773      0.08233982       8.31      <.0001
log_inhibin_right   -.0729351423      0.04420286      -1.65      0.1029
log_inhibin_left    0.0556926887      0.04651735       1.20      0.2347

Parameter               95% Confidence Limits

Intercept           0.1590802406 0.7950943379
log_testo_left      0.5202829177 0.8480058369
log_inhibin_right   -.1609016450 0.0150313603
log_inhibin_left    -.0368797856 0.1482651630
```

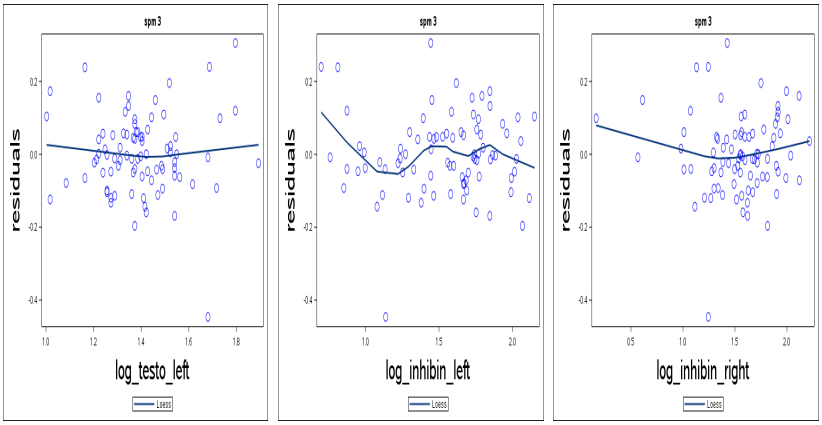
For en ordens skyld skal vi lige overbevise os selv om, at den multiple regressionsmodel overhovedet var rimelig, dvs. vi skal udføre de sædvanlige modelkontroltegninger. Det drejer sig om det sædvanlige `Diagnosticspanel`, inkluderende residualer plottet mod forventede værdier af outcome samt fraktildiagram over residualerne:



Endvidere vil vi gerne se plots af residualer mod hver af de tre kovariater, med udglattede kurver, for at vurdere lineariteten, men sådanne kommer desværre ikke automatisk med, når man er nået op på 3 kovariater. Vi har derfor i koden ovenfor tilføjet en `output`-sætning, så vi efterfølgende selv kan producere plottene, f.eks. således:

```
proc sgplot data=check;
loess Y=residuals X=log_inhibin_left /
    markerattrs=(color=blue size=0.3cm);
  xaxis labelattrs=(size=0.8cm);
  yaxis labelattrs=(size=0.8cm);
run;
```

hvorved vi får figurerne:



Disse er ikke alle helt fine i kanten, men heller ikke så slemme, at vi ikke kan fortsætte med vores forehavende.

Vi vender derfor tilbage til det output, vi fandt fra vores analyse, og her bemærker vi, at vi i hvert fald ikke har meget glæde af at kende inhibinværdien på venstre side, så denne udelader vi, hvorved vi får

```
proc glm plots=(Diagnosticspanel residuals(smooth)) data=a1;
    model log_testo_right = log_inhibin_right
                            log_testo_left / solution clparm;
run;
```

```
The GLM Procedure

Number of Observations Read          96
Number of Observations Used          85

R-Square     Coeff Var      Root MSE    log_testo_right Mean
0.531280      7.914363      0.111284               1.406107

                                        Standard
Parameter                 Estimate           Error    t Value    Pr > |t|
Intercept            0.5689774965      0.14066835       4.04      0.0001
log_inhibin_right    -.0445568519      0.03753222      -1.19      0.2386
log_testo_left       0.6487671859      0.07712057       8.41      <.0001

Parameter               95% Confidence Limits
Intercept            0.2891433391 0.8488116540
log_inhibin_right    -.1192203922 0.0301066884
log_testo_left       0.4953497983 0.8021845734
```

Vi ser, at der heller ikke ligger signifikant information i at kende inhibin værdien på højre side.

Faktisk kunne vi allerede have set dette af det oprindelige output (fra modellen med 3 kovariater), idet `Type I SS`-testene kan *trevles op fra neden* og dermed med det samme vise, at `log_inhibin_right` ikke ville blive signifikant, når vi udelod `log_inhibin_left`.

Ligeledes kunne vi af output fra `contrast`-sætningen se, at det fælles test for udeladelse af de to inhibin-kovariater gav P=0.24.

Bemærk, at man ikke bør sammenligne R^2-værdierne for modeller, der ikke har det samme antal kovariater, da denne altid vil være mindre, når der er flere kovariater med. Så er residalspredningen (`Root MSE`) et bedre mål, og her ser vi, at modellen med udelukkende `log_testo_left` som forklarende variabel giver næsten helt den samme residualspredning, som hvis en eller begge inhibin-målinger er med i modellen.

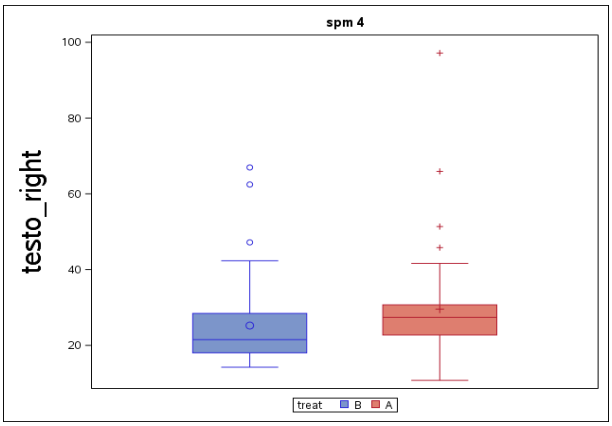
```
proc glm data=a1;
    model log_testosteron_left = log_testosteron_right
          / solution clparm;
run;


The GLM Procedure

Number of Observations Read          96
Number of Observations Used          85

R-Square     Coeff Var      Root MSE    log_testo_right Mean
0.523224      7.933855      0.111558               1.406107

                                    Standard
Parameter             Estimate          Error    t Value    Pr > |t|
Intercept          0.4520662275     0.10069320       4.49      <.0001
log_testo_left     0.6833105476     0.07159666       9.54      <.0001

Parameter            95% Confidence Limits
Intercept          0.2517915039 0.6523409512
log_testo_left     0.5409076775 0.8257134177
```

Nu kommer et par spørgsmål, der fokuserer på
forskellen på stimulationsmetoderne:

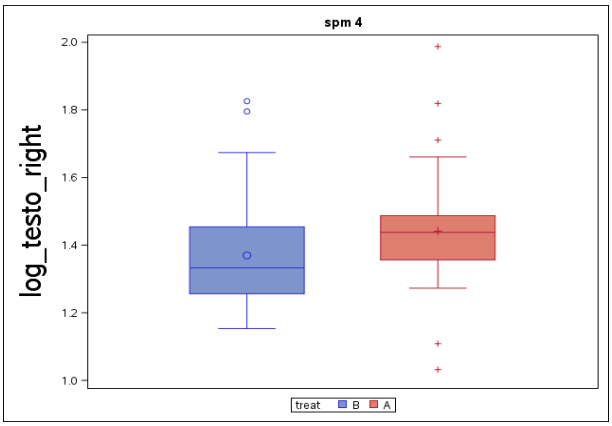
4. *Lav en figur til illustration af forskellene på de to behandlingsgrupper forsåvidt angår testosteron på højre side. Udfør på baggrund af denne tegning en sammenligning af de to grupper* **på passende skala** *og kvantificer forskellen med konfidensinterval.*

Vi starter med et simpelt Boxplot

```
proc sgplot data=a1;
vbox testo_right / group=treat;
  yaxis labelattrs=(size=0.8cm);
run;
```



Man bemærker her (igen), at fordelingen i begge grupper er en anelse skæv, med en hale mod de høje værdier. Da vi er på vej mod en sammenligning af grupperne med et T-test, forsøger vi igen at transformere med logaritmen for at opnå symmetriske fordelinger (helst normalfordelinger). Her er anvendt 10-tals logaritmer, men det spiller ingen rolle.



Ud fra denne figur ser det fornuftigt ud at foretage et uparret T-test,

idet

- Observationerne antages uafhængige (forskellige kvinder)
- Der synes at være nogenlunde samme variation i begge grupper
- Fordelingerne ser rimeligt symmetriske/normalfordelte ud

Man kan selvfølgelig supplere med fraktildiagrammer, men det skønnes ikke at være nødvendigt her, ikke mindst fordi fordelingsantagelsen ikke er voldsomt kritisk, hvis vi blot skal sammenligne middelværdier og ikke udtale os om enkeltindivider.

Vi udfører så det uparrede T-test:

```
proc ttest data=a1;
    class treat;
    var log_testo_right;
run;
```

og får outputtet

```
The TTEST Procedure

Variable:  log_testo_right

treat          N       Mean    Std Dev    Std Err    Minimum    Maximum
A             48     1.4415     0.1512     0.0218     1.0318     1.9875
B             43     1.3695     0.1588     0.0242     1.1532     1.8259
Diff (1-2)           0.0719     0.1548     0.0325

treat         Method               Mean       95% CL Mean        Std Dev
A                                  1.4415      1.3976   1.4854      0.1512
B                                  1.3695      1.3207   1.4184      0.1588
Diff (1-2)    Pooled               0.0719     0.00734   0.1365      0.1548
Diff (1-2)    Satterthwaite        0.0719     0.00714   0.1367

Method           Variances        DF    t Value    Pr > |t|
Pooled           Equal            89       2.21      0.0295
Satterthwaite    Unequal      86.762       2.21      0.0300

              Equality of Variances

Method      Num DF    Den DF    F Value    Pr > F
Folded F        42        47       1.10    0.7384
```

Her ser vi

- Der indgår i alt 91 personer i vores sammenligning, idet nogle kvinder har manglende værdier for testosteron på højre side.
- Spredningerne i de to grupper er meget ens (0.1512 hhv. 0.1588), P-værdi for test af identitet af disse er 0.74.
- Middelværdierne i de to grupper er signifikant forskellige, med P=0.03.
- Den estimerede forskel på de to behandlinger er (på logaritmisk skala) 0.0719 (stimulation A mod stimulation B), med et konfidensinterval på (0.00734, 0.1365). Tilbagetransformeres dette, fås et ratio-estimat på $10^{0.0719} = 1.18$, altså at stimulation A giver 18% højere værdier end stimulation B, med et konfidensinterval på (1.017, 1.369), dvs. svarende til fra 1.7% over til hele 36.9% over.

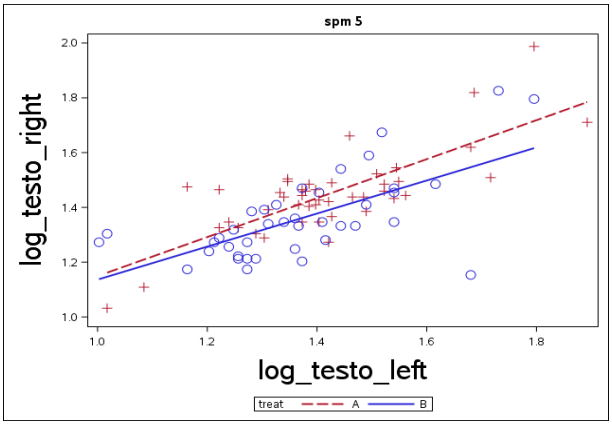
5. *Hvis to personer med forskellige stimulationsmetoder har opnået samme testosteron-værdi på venstre side, kan vi da forvente, at de også har nogenlunde samme værdi på højre side, eller kan afvigelserne komme op på f.eks. 50 nmol/l (eller 25%)?*

   Der er her lagt op til en kovariansanalyse, hvor vi ved sammenligning af `log_testo_right` for de to stimulationsmetoder korrigerer (betinger, justerer) for forskelle på venstre side, således at vi sammenligner værdier på højre side, *for fastholdt værdi* af værdi på venstre side.

   Først skal vi dog lige tegne:

```
ods graphics / imagename="log-testo-ancova";
proc sgplot data=a1;
reg Y=log_testo_right X=log_testo_left / group=treat
    markerattrs=(size=0.3cm);
  xaxis labelattrs=(size=0.8cm);
  yaxis labelattrs=(size=0.8cm);
run;
```

   hvorved vi får figuren:



Selv om disse regressionslinier ikke er helt parallelle, vil vi fortsætte med en kovariansanalyse, da vi ikke på forhånd havde formodning om en interaktion her. Vi medtager også den tilhørende modelkontrol

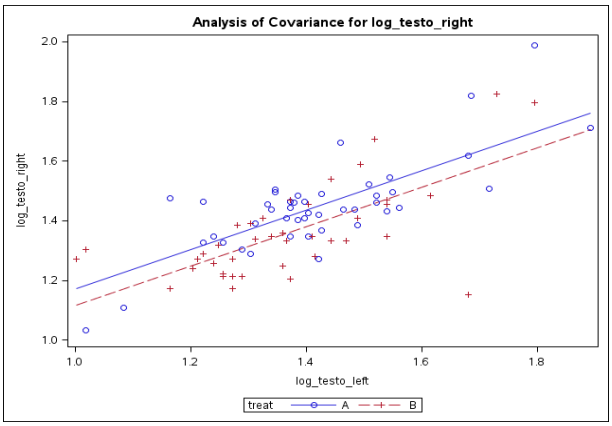
```
ods graphics / imagename="glm1-logtesto-treat";
proc glm plots=(Diagnosticspanel residuals(smooth)) data=a1;
    class treat;
    model log_testo_right = treat log_testo_left / solution clparm;
run;
```

og finder outputtet

```
The GLM Procedure

   Class Level Information

Class         Levels    Values
treat              2    A B

Number of Observations Read          96
Number of Observations Used          85

Dependent Variable: log_testo_right

                                 Sum of
Source                     DF       Squares   Mean Square  F Value  Pr > F
Model                       2    1.19791411    0.59895706    50.70  <.0001
Error                      82    0.96863569    0.01181263
Corrected Total            84    2.16654980

R-Square     Coeff Var      Root MSE    log_testo_right Mean
0.552913      7.729563      0.108686                1.406107
```

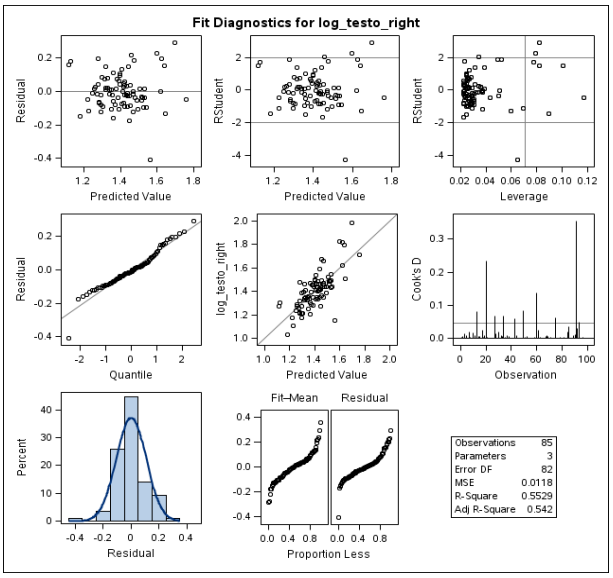
```
                                                Standard
Parameter                    Estimate               Error    t Value    Pr > |t|
Intercept                0.4545048528 B         0.09810597       4.63      <.0001
treat          A         0.0556541100 B         0.02384976       2.33      0.0221
treat          B         0.0000000000 B          .                .         .
log_testo_left           0.6604610414           0.07043701       9.38      <.0001

Parameter                  95% Confidence Limits
Intercept                0.2593408213 0.6496688843
treat          A         0.0082093388 0.1030988813
treat          B          .            .
log_testo_left           0.5203393899 0.8005826928

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

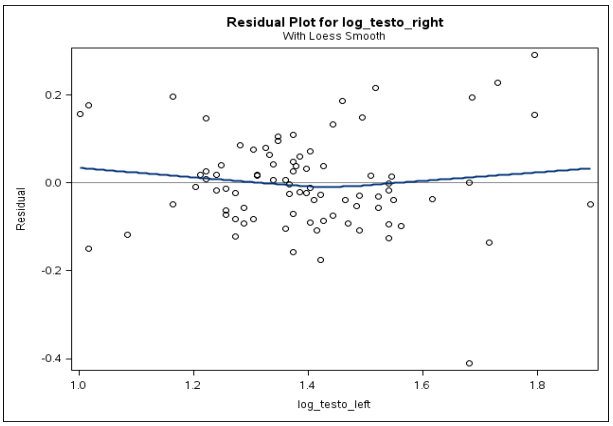
Den model, vi har fittet, svarer til to parallelle linier, som skitseret nedenfor:



og det ses af outputtet, at der stadig er signifikant forskel på de to stimulationsmetoder (P=0.022), omend forskellen er formindsket noget, fra 0.0719(0.0325) i spm. 4 til nu 0.0557(0.0238), på logaritmisk skala. Vi kvantificerer yderligere nedenfor.

Modelkontrol tegningerne bliver:





Vi skal nu kvantificere forskellen på de to stimulationsmetoder for `log_testo_right`, for *fastholdt* værdi af `log_testo_left`, altså forskellen 0.0557(0.0238) på logaritmisk skala, dvs. CI=(0.0082, 0.1031). Når vi tilbagetransformerer dette, får vi $10^{0.0557} = 1.137$, med CI=$(10^{0.0082}, 10^{0.1031}) = (1.019, 1.268)$. For fastholdt værdi af testosteron på venstre side, finder vi altså, at stimulationsmetode A stadig estimeres til at give 13.7% højere værdier på højre side end stimulationsmetode B, men det kan være helt op til 26.8% højere.

Da `testo_right`-værdierne ligger et sted mellem 10 og 60-70 stykker, kan forskellen **på middelværdierne** altså *ikke* komme helt op mod de 50 nmol/l, idet 26.8% af 60 kun er ca. 16.

Nu er der imidlertid spurgt til forskelle for to **personer**, og vi skal derfor over i noget med prediktionsintervaller i stedet for. Og det er ikke engang et helt almindeligt prediktionsinterval, fordi det drejer sig om en *differens* mellem to personer, ikke *en enkelt* person, som det plejer.

Vi indfører lige nogle lettere betegnelser, nemlig H_A, H_B, V_A, og V_B for `log_testo`, på hhv Højre og Venstre side, for stimulationsmetode A og B.

Vi har så her set på modellen med 2 parallelle regressionslinier:

$$H_A = \alpha_A + \beta V_A + \varepsilon_A$$
$$H_B = \alpha_B + \beta V_B + \varepsilon_B$$

og vi vil gerne finde prediktionsinterval for differensen $H_A - H_B$ under forudsætning af, at $V_A = V_B$. Men så er

$$H_A - H_B = \alpha_A - \alpha_B + \varepsilon_A - \varepsilon_B$$

som har middelværdi $\alpha_A - \alpha_B$, men varians $2 \times \sigma^2$.

I prediktionsintervallet skal vi således gange `Root MSE` med kvadratroden af 2 ($\sqrt{2}$), så vi finder:

$$0.0557 \pm 2 \times \sqrt{2} \times 0.1087 = (-0.2518, 0.3632)$$

og når vi tilbagetransformerer dette, får vi CI=$(10^{-0.2518}, 10^{0.3632}) = (0.560, 2.308)$, svarende til, at stimulationsmetode A kan være over dobbelt så høj som stimulationsmetode B, men også ned til ca. 44% lavere.

Her er således rigelig plads til en forskel på 50 nmol/l.

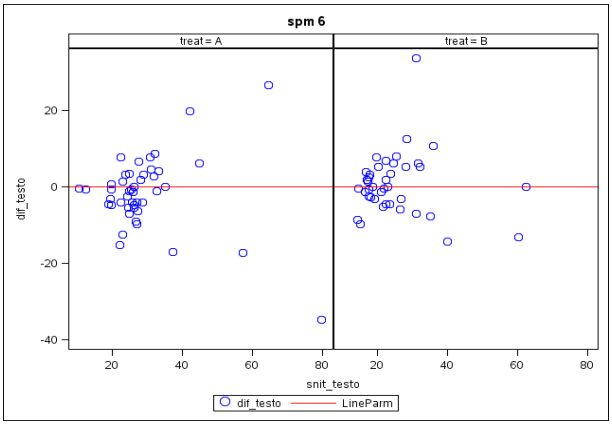
6. *Kvantificer middelværdien af sideforskellene for testosteron, for hver stimulationsmetode for sig.*

Vi går nu op inden det første `run;` og definerer nogle nye variable, nemlig differenser og gennemsnit for testosteron på log-skala. Husk, at man skal lave differenser af logaritmer, **aldrig** logaritmer af differenser:

```
dif_log_testo=log_testo_right-log_testo_left;
dif_testo=testo_left-testo_right;
snit_testo=(testo_left+testo_right)/2;
```

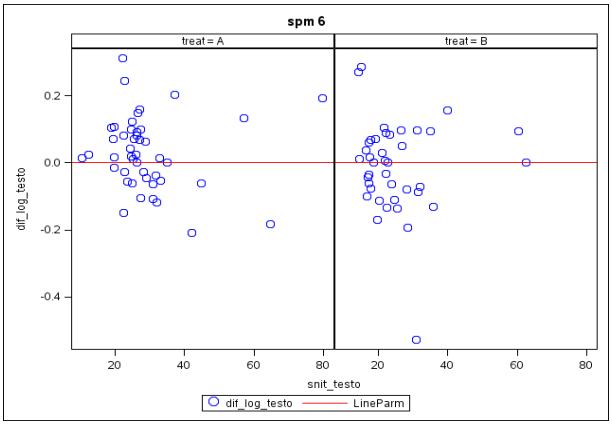
Når vi skal se på sideforskelle, skal vi (naturligvis) stadig arbejde på logaritmisk skala, som det fremgår af nedenstående Bland-Altman plots, der tydeligt viser trompetfacon:

```
proc sgpanel data=a1;
   panelby treat;
   scatter Y=dif_testo X=snit_testo /
   markerattrs=(color=blue size=0.3cm);
   lineparm x=10 y=0 slope=0 / lineattrs=(color=red);
run;
```



De tilsvarende Bland-Altman plots, for de logaritmerede værdier, bliver:

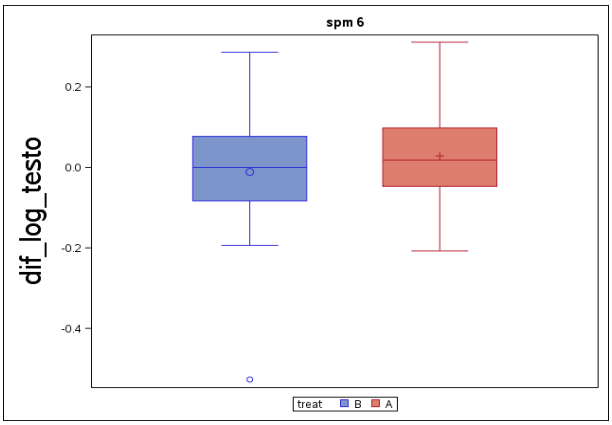
```
proc sgpanel data=a1;
   panelby treat;
   scatter Y=dif_log_testo X=snit_testo /
   markerattrs=(color=blue size=0.3cm);
   lineparm x=10 y=0 slope=0 / lineattrs=(color=red);
run;
```



som ser noget bedre ud, omend ikke fantastiske. Vi kvantificerer derfor differenserne på log-skala, både med et Box-plot:

```
proc sgplot data=a1;
vbox dif_log_testo / group=treat;
  yaxis labelattrs=(size=0.8cm);
run;
```

og med summary statistics:



som giver:

```
proc means N mean clm data=a1;
class treat;
var dif_log_testo;
run;
```

og får resultatet

```
The MEANS Procedure

                Analysis Variable : dif_log_testo

              N                                Lower 95%       Upper 95%
treat       Obs     N            Mean      CL for Mean      CL for Mean
-------------------------------------------------------------------------
A            51    45       0.0285724       -0.0040477       0.0611925
B            45    40      -0.0110996       -0.0546868       0.0324877
-------------------------------------------------------------------------
```

Den estimerede forskel (højre minus venstre) er (på logaritmisk skala) 0.0286 hhv -0.0111 for hhv stimulationsmetode A og B, dvs. sideforskellene går i forskellige retninger. Vi kan dog se af konfidensintervallerne, at ingen af dem er signifikant forskellige fra 0, dvs. der er ingen evidens for sideforskelle for nogen af stimulationsmetoderne.

For at forstå disse sideforskelle, skal vi tilbagetransformere til ratio'er, f.eks. for stimulationsmetode A: $10^{0.0285} = 1.07$, hvilket betyder, at værdierne på højre side i gennemsnit ligger ca. 7% over de tilsvarende i venstre side, med konfidensintervallet $(10^{-0.0040}, 10^{0.0612}) = (0.99, 1.15)$, altså fra mellem 1% under til 15% over.

Tilsvarende fås for stimulationsmetode B: $10^{-0.0111} = 0.97$, hvilket betyder, at værdierne på højre side i gennemsnit ligger ca. 3% *under* de tilsvarende i venstre side, med konfidensintervallet $10^{-0.0547}, 10^{0.0325} = (0.88, 1.08)$, altså fra mellem 12% under til 8% over.

Vi kan få P-værdier for testene af middelværdi 0 for disse sideforskelle ved at inkludere keyword `probt` i `proc means` ovenfor, men af pædagogiske årsager vises det i stedet ved at udføre tests for middelværdi 0, dvs. ved at lave parrede T-tests for de to sider, for hver stimulationsmetode for sig:

```
proc sort data=a1; by treat;
run;

proc ttest data=a1; by treat;
    paired log_testo_left*log_testo_right;
run;
```

hvilket giver

```
treat=A

The TTEST Procedure
Difference:  log_testo_left - log_testo_right

  N        Mean     Std Dev     Std Err     Minimum     Maximum
 45     -0.0286      0.1086      0.0162     -0.3113      0.2076

    Mean       95% CL Mean        Std Dev      95% CL Std Dev
 -0.0286     -0.0612  0.00405      0.1086      0.0899   0.1372

    DF    t Value    Pr > |t|
    44      -1.77      0.0845


treat=B

The TTEST Procedure
Difference:  log_testo_left - log_testo_right

  N        Mean     Std Dev     Std Err     Minimum     Maximum
 40      0.0111      0.1363      0.0215     -0.2864      0.5270

    Mean       95% CL Mean        Std Dev      95% CL Std Dev
  0.0111     -0.0325   0.0547      0.1363      0.1116   0.1750

    DF    t Value    Pr > |t|
    39       0.52      0.6094
```

Herved får vi P-værdierne 0.084 hhv. 0.61, og dermed ingen signifikante sideforskelle.

Ser det ud som om forskellen på stimulationsmetoderne er den samme på de to sider (for testosteron)?

Umiddelbart kunne det her se ud til, at der var lagt op til uparrede T-tests for hver af siderne, og da vi allerede i spm. 4 har set på højre side, supplerer vi her med venstre side:

```
proc ttest data=a1;
    class treat;
    var log_testo_left;
run;
```

som ville give os outputtet:

```
The TTEST Procedure

Variable:  log_testo_left

treat          N        Mean     Std Dev     Std Err     Minimum     Maximum
A             48      1.4199      0.1687      0.0244      1.0175      1.8925
B             42      1.3649      0.1681      0.0259      1.0026      1.7956
Diff (1-2)            0.0550      0.1684      0.0356

treat          Method                Mean        95% CL Mean         Std Dev
A                                    1.4199       1.3709   1.4689       0.1687
B                                    1.3649       1.3125   1.4173       0.1681
Diff (1-2)     Pooled                0.0550      -0.0157   0.1257       0.1684
Diff (1-2)     Satterthwaite         0.0550      -0.0157   0.1257

Method            Variances         DF     t Value     Pr > |t|
Pooled            Equal             88        1.55       0.1258
Satterthwaite     Unequal       86.505        1.55       0.1258

                Equality of Variances

Method       Num DF     Den DF     F Value     Pr > F
Folded F         47         41        1.01     0.9859
```

Sammenfattende kvantificerer vi forskellen på de to stimulationsmetoder, for hver side (for A vs. B), på $\log_{10}$-skala:

Venstre side: 0.0550 (-0.0157, 0.1257),
tilbagetransformeret til 1.135 (0.965, 1.336), altså 13.5% højere for A end for B, med CI fra 3.5% under til 33.6% over.

Højre side: 0.0719 (0.0073, 0.1365),
tilbagetransformeret til 1.180 (1.017, 1.369), altså 18.0% højere for A end for B, med CI fra 1.7% over til 36.9% over.

For højre side er der således signifikant forskel, medens der på venstre side *ikke* er. Dette er imidlertid *absolut* ikke nok til at sige, at der er forskel på effekterne på de to sider! Ydermere er der et lille problem: Det er ikke (helt) de samme personer for de to sider, idet manglende observationer får enkelte individer til at ryge ud af analyserne.....

Men hvordan finder vi så en P-værdi for testet af, om de to forskelle på stimulationsmetoderne er ens??

Vi benytter her igen de tidligere indførte betegnelser, nemlig H_A, H_B, V_A, og V_B for `log_testo`, på hhv Højre og Venstre side, for stimulationsmetode A og B.

Vi vil gerne undersøge, om forskellen på A og B er ens for de 2 sider, dvs. om

$$H_A - H_B = V_A - V_B$$

men ved at bytte lidt rundt på disse led, ses dette at være det samme som at undersøge om

$$H_A - V_A = H_B - V_B$$

altså om side-differenserne `dif_log_testo` er ens for de to stimulationsmetoder.

Vi laver derfor et uparret T-test til sammenligning af stimulation A og B for sideforskellene `dif_log_testo=log_testo_left-log_testo_right`:

```
proc ttest data=a1;
    class treat;
    var dif_log_testo;
run;
```

```
The TTEST Procedure

Variable:  dif_log_testo

treat          N       Mean    Std Dev    Std Err    Minimum    Maximum
A             45     0.0286     0.1086     0.0162    -0.2076     0.3113
B             40    -0.0111     0.1363     0.0215    -0.5270     0.2864
Diff (1-2)           0.0397     0.1224     0.0266

treat         Method               Mean       95% CL Mean        Std Dev
A                                  0.0286    -0.00405   0.0612      0.1086
B                                 -0.0111     -0.0547   0.0325      0.1363
Diff (1-2)    Pooled               0.0397     -0.0132   0.0926      0.1224
Diff (1-2)    Satterthwaite        0.0397     -0.0140   0.0934

Method           Variances        DF    t Value    Pr > |t|
Pooled           Equal            83       1.49      0.1396
Satterthwaite    Unequal      74.422       1.47      0.1452

              Equality of Variances

Method      Num DF    Den DF    F Value    Pr > F
Folded F        39        44       1.58    0.1444
```

Vi ser, at der ikke findes klare tegn på forskel her (P=0.14), så vi må konkludere, at det sagtens kan tænkes, at stimulationstyperne virker ens på de to sider.

Bemærk, at der i denne sidste analyse indgår endnu færre personer end i de tidligere, fordi nu kun personer med målte værdier for alle hormoner indgår.

I de to sidste spørgsmål ser vi på **ægkvaliteten**.

7. *Giv et estimat for sandsynligheden for høj ægkvalitet, for hver af de to stimulationsmetoder hver for sig. Ser de to estimater forskellige ud?*

Når vi skal udføre noget for hver stimulationstype for sig, starter vi med at sortere efter denne, hvorefter vi benytter eksakte binomialtests til at finde konfidensintervaller (selv om vi slet ikke er interesserede i selve testet):

```
proc sort data=a1; by treat;
run;

proc freq data=a1; by treat;
  tables kvalitet;
  exact binomial;
run;
```

Herved finder vi

```
treat=A

The FREQ Procedure
                                          Cumulative    Cumulative
kvalitet    Frequency     Percent     Frequency      Percent
-----------------------------------------------------------------
       0          21       41.18            21        41.18
       1          30       58.82            51       100.00

        Binomial Proportion
           kvalitet = 0

Proportion (P)               0.4118
ASE                          0.0689
95% Lower Conf Limit         0.2767
95% Upper Conf Limit         0.5468
```

```
Exact Conf Limits
95% Lower Conf Limit         0.2758
95% Upper Conf Limit         0.5583

Sample Size = 51



treat=B

The FREQ Procedure

                                        Cumulative    Cumulative
kvalitet    Frequency     Percent     Frequency      Percent
-----------------------------------------------------------------
       0          15       33.33            15        33.33
       1          30       66.67            45       100.00


        Binomial Proportion
           kvalitet = 0

Proportion (P)               0.3333
ASE                          0.0703
95% Lower Conf Limit         0.1956
95% Upper Conf Limit         0.4711

Exact Conf Limits
95% Lower Conf Limit         0.2000
95% Upper Conf Limit         0.4895

Sample Size = 45
```

De to estimater **for dårlig ægkvalitet!**, med tilhørende (eksakte) konfidensintervaller ses at være:

A: 0.41 (0.28, 0.55)

B: 0.33 (0.20, 0.49)

Hvis vi ønsker estimater for **god kvalitet** i stedet, skal vi bare trække fra 1:

A: 0.59 (0.45, 0.72)

B: 0.67 (0.51, 0.80)

De ser jo ikke særligt forskellige ud, og vi tester hypotesen om identitet nedenfor:

8. *Lav et test for om kvaliteten af de udtagne æg afhænger af stimulationsmetoden. Kvantificer forskellen, dels som en forskel på de to sandsynligheder fundet ovenfor og dels som en relativ risiko. Husk i begge tilfælde at supplere med et konfidensinterval. Kan vi udelukke, at der er dobbelt så stor sandsynlighed for at have god ægkvalitet i den ene gruppe i forhold til den anden?*

   Vi udvider vores analyse af to gange to tabellen med et par ekstra options:

   - `expected`: De forventede counts i hver celle, under hypotesen om "ingen forskel på behandlingerne". dem skal vi bruge til at vurdere om Chi-i-anden testet er tilladeligt at udføre
   - `chisq`: Chi-i-anden testet skal udskrives
   - `riskdiff`: Estimater for forskellen mellem sandsynlighederne for god ægkvalitet skal udskrives
   - `relrisk`: Estimat for den relative risiko skal udskrives

```
proc freq data=a1;
  tables treat*kvalitet / nopercent nocol expected chisq riskdiff relrisk;
  weight antal;
run;
```

som giver outputtet

```
The FREQ Procedure
Table of treat by kvalitet

treat     kvalitet

Frequency|
Expected |
Row Pct  |       0|       1|  Total
---------+--------+--------+
A        |     21 |     30 |     51
         | 19.125 | 31.875 |
         |  41.18 |  58.82 |
---------+--------+--------+
B        |     15 |     30 |     45
         | 16.875 | 28.125 |
         |  33.33 |  66.67 |
---------+--------+--------+
Total          36       60       96


Statistics for Table of treat by kvalitet
```

```
Statistic                     DF       Value      Prob
------------------------------------------------------
Chi-Square                     1      0.6275    0.4283
Likelihood Ratio Chi-Square    1      0.6294    0.4276
Continuity Adj. Chi-Square     1      0.3374    0.5613

       Fisher's Exact Test
---------------------------------
Two-sided Pr <= P           0.5272

Statistics for Table of treat by kvalitet

                        Column 2 Risk Estimates

                                     (Asymptotic) 95%       (Exact) 95%
                   Risk       ASE   Confidence Limits    Confidence Limits
-------------------------------------------------------------------------
Row 1           0.5882    0.0689    0.4532    0.7233    0.4417     0.7242
Row 2           0.6667    0.0703    0.5289    0.8044    0.5105     0.8000
Total           0.6250    0.0494    0.5282    0.7218    0.5203     0.7218
Difference     -0.0784    0.0984   -0.2713    0.1145

                     Difference is (Row 1 - Row 2)

                  Odds Ratio and Relative Risks

Statistic                        Value       95% Confidence Limits
-----------------------------------------------------------------
Odds Ratio                      1.4000        0.6082        3.2227
Relative Risk (Column 1)        1.2353        0.7289        2.0936
Relative Risk (Column 2)        0.8824        0.6479        1.2017

Sample Size = 96
```

Vi ser af såvel χ^2-test og Fishers eksakte test, at der ikke er signifikant forskel på ægkvaliteten for de to behandlinger (P=0.43 hhv. 0.53). De forventede antal er alle et pænt stykke over de krævede 5 (minimum 16.9), så det er ikke nødvendigt at lave Fishers eksakte test, da chi-i-anden approksimationen er god nok.

De estimerede sandsynligheder for god ægkvalitet ses under `Row Pct` i kolonne 2 (svarende til `kvalitet=1`). De er A: 58.82, og B: 66.67, og her får vi igen angivet de eksakte konfidensgrænser som fundet ovenfor.

Den estimerede forskel på de to sandsynligheder for god ægkvalitet (Column 2 risk) er 0.0784 (=0.6667-0.5882), altså ca. 8%*point*, med konfidensgrænser (-0.1145, 0.2713). Der kan altså med rimelighed være op til 27%-point større sandsynlighed for god ægkvalitet for treatment B.

Den relative "risiko" for god ægkvalitet for stimulationsmetode B vs. A udregnes nemt i hånden til

$$\frac{0.6667}{0.5882} = 1.13$$

men for at få konfidensgrænser på dette tal, må vi benytte SAS. Desværre er 1.13 dog ikke blandt de tal, SAS giver os, og det er fordi den angiver den relative risiko for treatment A vs. B i stedet. For kolonne 2 er denne angivet som 0.8824, og ved at invertere denne, fås

$$\frac{1}{0.8824} = 1.13$$

Tilsvarende kunne vi invertere det tilhørende konfidensinterval:

$$\left(\frac{1}{1.2017}, \frac{1}{0.0.6479}\right) = (0.83, 1.54)$$

Man kan derfor *ikke* sige, at sandsynligheden for god ægkvalitet kan være dobbelt så stor i den ene gruppe i forhold til den anden, men bemærk, at man *godt* kan sige, at den ene gruppe (gruppe B) kan have dobbelt så store *odds* for god ægkvalitet.

Vi kunne få SAS til at udregne disse inverteringer ved at bytte om på rækkefølgen af vores stimulationsformer, f.eks. ved at kalde dem `2:A` hhv. `1:B`:

```
if treat="A" then nytreat="2:A";
if treat="B" then nytreat="1:B";
```

I så fald ville vi (blandt meget andet) få outputtet

```
                Odds Ratio and Relative Risks

Statistic                        Value       95% Confidence Limits
-----------------------------------------------------------------
Odds Ratio                      0.7143        0.3103        1.6442
Relative Risk (Column 1)        0.8095        0.4776        1.3720
Relative Risk (Column 2)        1.1333        0.8322        1.5435

Sample Size = 96
```

som netop giver ovennævnte resultater.