

Opgavebesvarelse, brain weight

(Matthews & Farewell: Using and Understanding Medical Statistics, 2nd. ed.)

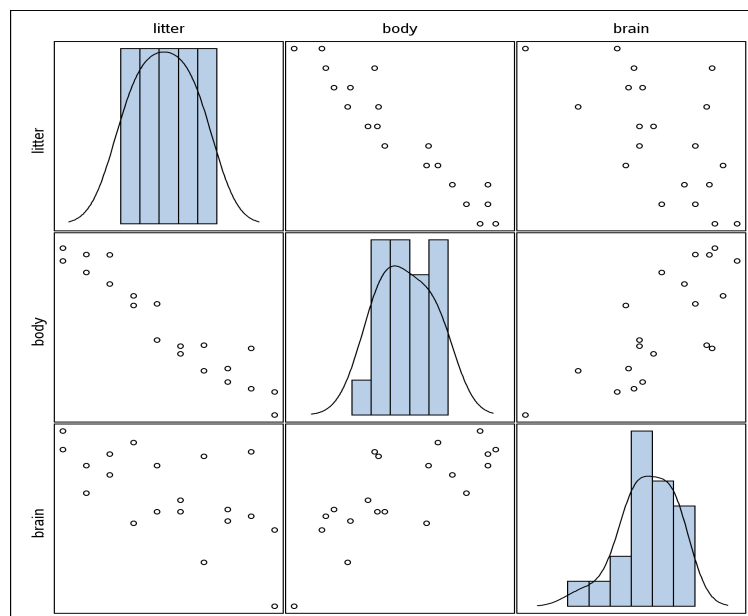
Spørgsmål 1

Data er indlagt på 'brain.txt' og variabelnavnene er `litter`, `body` og `brain`, som anført i første linie. Vi indlæser ved at skrive

```
data mus;  
infile 'C:\Basal\brain.txt' firstobs=2;  
input litter body brain;  
run;
```

Vi har nu 20 observationer, hver med 3 variable, og nedenfor ses de tre plottet parvist mod hinanden. Dette plot kan kun laves med en passende opgraderet version af SAS, ved at skrive:

```
ods graphics on;  
proc sgscatter data=mus;  
    matrix _numeric_ / diagonal=(kernel histogram);  
ods graphics off;
```



Spørgsmål 2

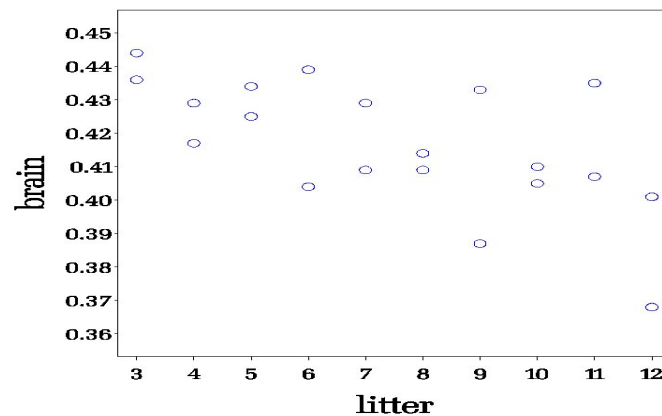
Vi ønsker at udtale os om hjernevægtens afhængighed af kuld størrelsen.

Ovenfor ses (bl.a.) et scatterplot af hjernevægten mod kuld størrelsen. Hvis vi skal se det separat, skriver vi

```
proc gplot data=mus;  
  plot brain*litter;  
run;
```

eller for at få en lidt pænere figur:

```
proc gplot data=mus;  
  plot brain*litter  
  / haxis=axis1 vaxis=axis2 frame;  
axis1 offset=(3,3) label=(H=3) value=(H=2) minor=NONE;  
axis2 offset=(1,1) value=(H=2) minor=NONE  
      label=(A=90 R=0 H=3);  
symbol1 v=circle c=blue i=none h=2 r=1;  
run;
```



På tegningen spores en negativ afhængighed af kuld størrelse, således at store kuld fører til mindre hjerner.

Spørgsmål 3

Ved hjælp af en simpel lineær regressionsanalyse af hjernevægt, med kuld-
størrelse som forklarende variabel, estimerer vi denne relation:

```
proc reg data=mus;
  model brain=litter / clb;
run;
```

og finder

The REG Procedure
Dependent Variable: brain

Analysis of Variance

| Source | DF | Sum of Squares | Mean Square | F Value | Pr > F |
|-----------------|----|-------------------|----------------|---------|--------|
| Model | 1 | 0.00268 | 0.00268 | 11.33 | 0.0034 |
| Error | 18 | 0.00427 | 0.00023698 | | |
| Corrected Total | 19 | 0.00695 | | | |

| | | | |
|----------------|---------|----------|--------|
| Root MSE | 0.01539 | R-Square | 0.3862 |
| Dependent Mean | 0.41675 | Adj R-Sq | 0.3521 |
| Coeff Var | 3.69383 | | |

Parameter Estimates

| Variable | DF | Parameter Estimate | Standard Error | t Value | Pr > t |
|-----------|----|-----------------------|-------------------|---------|---------|
| Intercept | 1 | 0.44700 | 0.00962 | 46.44 | <.0001 |
| litter | 1 | -0.00403 | 0.00120 | -3.37 | 0.0034 |

| Variable | DF | 95% Confidence Limits | |
|-----------|----|-----------------------|----------|
| Intercept | 1 | 0.42678 | 0.46722 |
| litter | 1 | -0.00655 | -0.00152 |

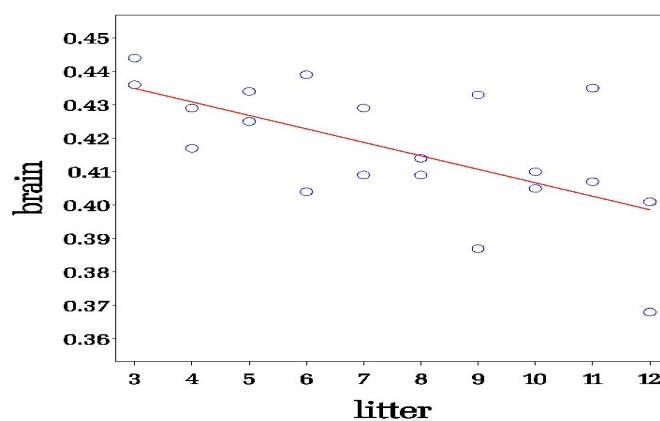
Vi ser (svarende til tegningen), at hældningskoefficienten er negativ

$$\hat{\beta}_1 = -0.0040(0.0012)$$

Dette betyder, at vi for hver ekstra mus i kullet forventer en hjernevægt på gennemsnitligt $0.004g$ mindre for hver enkelt mus.

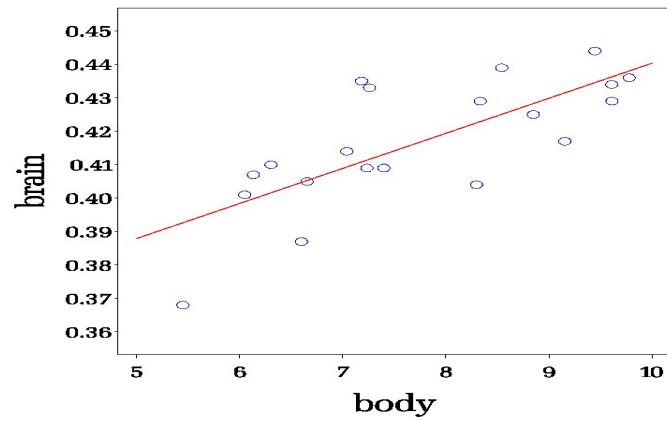
Vi indlægger regressionslinien på tegningen ved i symbol-sætningen i `proc gplot` at skrive `i=r1` i stedet for `i=none`, samt tilføjer ekstra options for farve, bredde og linietype:

```
symbol1 v=circle c=blue i=r1 h=2 l=1 ci=red r=1;
```



Spørgsmål 4

Nu ser vi tilsvarende på hjernevægtens afhængighed af kropsvægten, hvor regressionsanalysen giver nedenstående resultater.



The REG Procedure
Dependent Variable: brain

Analysis of Variance

| Source | DF | Sum of Squares | Mean Square | F Value | Pr > F |
|-----------------|----|----------------|-------------|---------|--------|
| Model | 1 | 0.00387 | 0.00387 | 22.62 | 0.0002 |
| Error | 18 | 0.00308 | 0.00017109 | | |
| Corrected Total | 19 | 0.00695 | | | |

| | | | |
|----------------|---------|----------|--------|
| Root MSE | 0.01308 | R-Square | 0.5569 |
| Dependent Mean | 0.41675 | Adj R-Sq | 0.5323 |
| Coeff Var | 3.13861 | | |

Parameter Estimates

| Variable | DF | Parameter Estimate | Standard Error | t Value | Pr > t |
|-----------|----|--------------------|----------------|---------|---------|
| Intercept | 1 | 0.33557 | 0.01732 | 19.38 | <.0001 |
| body | 1 | 0.01048 | 0.00220 | 4.76 | 0.0002 |

| Variable | DF | 95% Confidence Limits | |
|-----------|----|-----------------------|---------|
| Intercept | 1 | 0.29918 | 0.37195 |

| | | | |
|------|---|---------|---------|
| body | 1 | 0.00585 | 0.01511 |
|------|---|---------|---------|

Vi ser heraf, at store mus har store hjerner, idet en musehjerne i gennemsnit er 0.010g tungere, når musen vejer 1g mere.

Spørgsmål 5

Vi undersøger nu, ved hjælp af korrelationstest, om der er signifikant korrelation mellem kuldstørrelse og kropsvægt.

Korrelationer udregnes ved at skrive

```
proc corr pearson spearman data=mus;
  var litter body;
run;
```

hvorved vi får

The CORR Procedure

2 Variables: litter body

Pearson Correlation Coefficients, N = 20

Prob > |r| under H0: Rho=0

| | | |
|--------|--------------------|--------------------|
| | litter | body |
| litter | 1.00000 | -0.95472 <.0001 |
| body | -0.95472 <.0001 | 1.00000 |

Spearman Correlation Coefficients, N = 20

Prob > |r| under H0: Rho=0

| | | |
|--------|--------------------|--------------------|
| | litter | body |
| litter | 1.00000 | -0.95095 <.0001 |
| body | -0.95095 <.0001 | 1.00000 |

Hvad enten vi benytter parametrisk eller nonparametrisk korrelation, er der helt klart en sammenhæng mellem kuldstorelse og kropsvægt.

Noget helt andet er så *fortolkningen* af korrelationskoefficienten. Vi kan i hvert fald ikke gå ud fra, at vores observationer passer med en todimensional normalfordeling, bl.a. fordi der er 2 af hver kuldstorelse.

Det ser ud som om denne **litter** er **valgt** på **systematisk** måde, og den faktiske størrelse af korrelationskoefficienten (hvad enten den er parametrisk eller ej) kan derfor ikke tillægges nogen fornuftig mening (idet korrelationskoefficientens størrelse som bekendt afhænger af samplingmetoden).

Spørgsmål 6

Vi undersøger nu, igen ved hjælp af en simpel lineær regressionsanalyse, om der er sammenhæng mellem kropsvægt og kuldstørrelse. Vi finder

The REG Procedure

Dependent Variable: body

Analysis of Variance

| Source | DF | Sum of Squares | Mean Square | F Value | Pr > F |
|-----------------|---------|----------------|-------------|---------|--------|
| Model | 1 | 32.12465 | 32.12465 | 185.38 | <.0001 |
| Error | 18 | 3.11931 | 0.17330 | | |
| Corrected Total | 19 | 35.24397 | | | |
| Root MSE | 0.41629 | R-Square | 0.9115 | | |
| Dependent Mean | 7.74710 | Adj R-Sq | 0.9066 | | |
| Coeff Var | 5.37346 | | | | |

Parameter Estimates

| Variable | DF | Parameter Estimate | Standard Error | t Value | Pr > t |
|-----------|----|-----------------------|----------------|---------|---------|
| Intercept | 1 | 11.05642 | 0.26027 | 42.48 | <.0001 |
| litter | 1 | -0.44124 | 0.03241 | -13.62 | <.0001 |
| Variable | DF | 95% Confidence Limits | | | |
| Intercept | 1 | 10.50960 | 11.60323 | | |
| litter | 1 | -0.50933 | -0.37316 | | |

Også for kropsvægten ses en negativ sammenhæng til kuldstørrelsen, idet hældningskoefficienten bliver signifikant negativ (i parentes bemærket er P-værdien præcis den samme, som vi fandt ovenfor i forbindelse med test af den parametriske korrelationskoefficient):

$$\hat{\beta} = -0.441(0.032)$$

Dette betyder, at vi for hver ekstra mus i kullet forventer en kropsvægt på gennemsnitligt 0.441g (knap et halvt gram) mindre for hver enkelt mus.

Spørgsmål 7

Vi har nu set, at der er signifikante sammenhænge mellem alle de tre målte størrelser:

- kropsvægten er negativt relateret til kuld størrelsen
- hjernevægten er positivt relateret til kropsvægten
- hjernevægten er negativt relateret til kuld størrelsen

Vi kan nu med rette spørge os selv, om den lavere hjernevægt blandt mus fra store kuld simpelthen er betinget af, at musene som sådan er mindre i store kuld og derfor også har mindre hjerner. For at undersøge denne påstand skulle man ideelt råde over data fra forskellige størrelser musekuld, hvor musene alle var lige tunge. Hvis hjernevægten her også kunne vises at falde med kuld størrelsen, kunne vi konstatere at kuld størrelse havde en *direkte* effekt på hjernevægten og ikke kun en effekt via kropsvægten.

Sådanne data har vi naturligvis ikke i praksis, men en multipel regressionsanalyse er præcis designet til at svare på dette spørgsmål, idet de enkelte effekter her netop fortolkes som effekten af den relevante kovariat for *fastholdt* værdi af alle de øvrige.

Vi foretager altså en multipel regressionsanalyse med hjernevægt som responsvariabel og såvel kuld størrelse som kropsvægt som forklarende variable:

```
proc reg data=mus;  
  model brain=body litter / clb;  
run;
```

hvorved vi får

The REG Procedure

Dependent Variable: brain

Analysis of Variance

| Source | DF | Sum of Squares | Mean Square | F Value | Pr > F |
|--------|----|----------------|-------------|---------|--------|
| Model | 2 | 0.00452 | 0.00226 | 15.81 | 0.0001 |
| Error | 17 | 0.00243 | 0.00014292 | | |

| | | | |
|-----------------|---------|----------|--------|
| Corrected Total | 19 | 0.00695 | |
| Root MSE | 0.01196 | R-Square | 0.6504 |
| Dependent Mean | 0.41675 | Adj R-Sq | 0.6093 |
| Coeff Var | 2.86864 | | |

Parameter Estimates

| Variable | DF | Parameter Estimate | Standard Error | t Value | Pr > t |
|-----------|----|--------------------|----------------|---------|---------|
| Intercept | 1 | 0.17877 | 0.07521 | 2.38 | 0.0295 |
| body | 1 | 0.02426 | 0.00677 | 3.58 | 0.0023 |
| litter | 1 | 0.00667 | 0.00313 | 2.13 | 0.0478 |

| Variable | DF | 95% Confidence Limits | |
|-----------|----|-----------------------|---------|
| Intercept | 1 | 0.02009 | 0.33746 |
| body | 1 | 0.00998 | 0.03854 |
| litter | 1 | 0.00007088 | 0.01327 |

Vi ser her, at begge kovariater er signifikante (kuldstørrelse er lige på kanten med $P=4.8\%$, medens kropsvægten er tydelig med $P=0.2\%$).

Spørgsmål 8

Vi sammenfatter resultaterne i tabellen:

| Respons: brain | koefficient til | | residual spredning | % forklaret variation |
|-------------------------------|---|---|-----------------------|--------------------------|
| Kovariater i model | litter $\hat{\beta}_1$ (se($\hat{\beta}_1$)) | body $\hat{\beta}_2$ (se($\hat{\beta}_2$)) | s | R^2 |
| kuldstørrelse alene | -0.004 (0.001) | – | 0.015 | 38.6 |
| kropsvægt alene | – | 0.010 (0.002) | 0.013 | 55.7 |
| kuldstørrelse og kropsvægt | 0.007 (0.003) | 0.024 (0.007) | 0.012 | 65.0 |

Bemærk her, at afhængigheden af kuldstørrelse skifter fortegn, idet den er negativ, når **body** ikke er med i modellen, men positiv, når **body** medtages. Fortolkningen af dette følger nedenfor.

Vi kan af dette konkludere

- at hjernevægten er positivt relateret til kropsvægten
 - Hvis mus A vejer $1g$ mere end mus B, forventer vi også at mus A's hjerne vejer i gennemsnit $0.010g$ mere end mus B's
(Bemærk, at vi i denne situation typisk har, at mus A kommer fra en mindre kuldstørrelse end mus B)
 - Hvis mus A og B kommer fra to kuld af **samme størrelse**, forventer vi dog en forskel i hjernevægt i gennemsnit på hele $0.024g$
- at hjernevægten er relateret til kuldstørrelsen, på følgende måde
 - Hvis mus C kommer fra en kuldstørrelse på en flere end mus D, da forventer vi, at mus C's hjerne vejer lidt mindre, i gennemsnit $0.004g$ mindre
(Bemærk, at vi i denne situation typisk har, at mus C tillige har en mindre kropsvægt end mus D)
 - Hvis mus C og D alligevel **vejer det samme**, da vil vi forvente, at mus C's hjerne vejer *mest*, nemlig i gennemsnit $0.007g$ mere end mus D's.

Den interessante konklusion er, at hjernevægten hos mus fra store kuld er **relativt større** set i forhold til kropsvægten end for mus fra små kuld. Den biologiske fortolkning af dette er, at den øgede konkurrence om næringen for store musekuld ikke går ud over hjernen i samme grad som den går ud over kropsvægten.

Opgavebesvarelse, biomasse

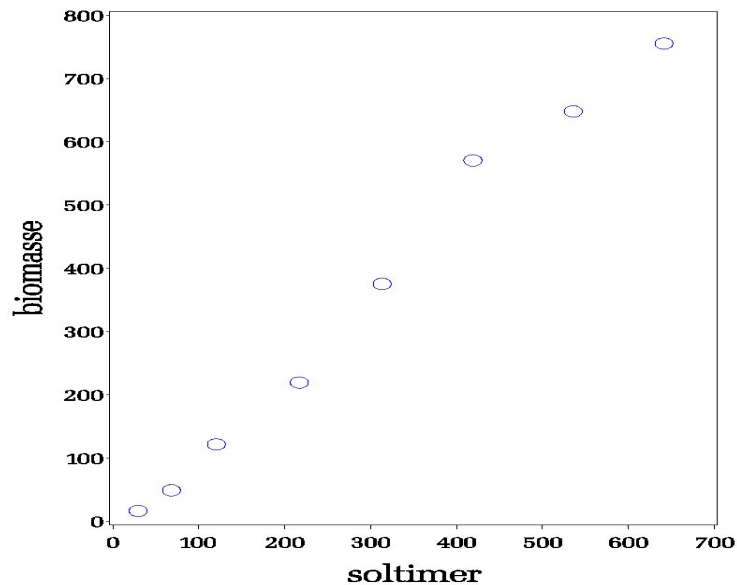
Spørgsmål 1.

Ser det ud til, at der en lineær sammenhæng mellem de to variable?

Data indtastes direkte i et SAS-program, og der tegnes:

```
data sol;
input soltimer biomasse;
cards;
29.7 16.6
68.4 49.1
120.7 121.7
217.2 219.6
313.5 375.5
419.1 570.8
535.9 648.2
641.5 755.6
;
run;

proc gplot data=sol;
  plot biomasse*soltimer
    / haxis=axis1 vaxis=axis2 frame;
axis1 value=(H=2) minor=NONE
      label=(H=3 'soltimer');
axis2 value=(H=2) minor=NONE
      label=(A=90 R=0 H=3 'biomasse');
symbol1 v=circle i=none c=BLUE h=3 w=2;
run;
```



Plottet (figur 1) ser jo rimeligt lineært ud. En nøjere granskning kan evt. foretages som modelkontrol efter fit af modellen. Bemærk, at en høj korrelation (specielt hvis det er en Spearman) ikke sikrer, at sammenhængen er lineær!

Under antagelse af en lineær regressionsmodel med normalfordelte fejl, ønskes følgende spørgsmål besvaret:

Spørgsmål 2.

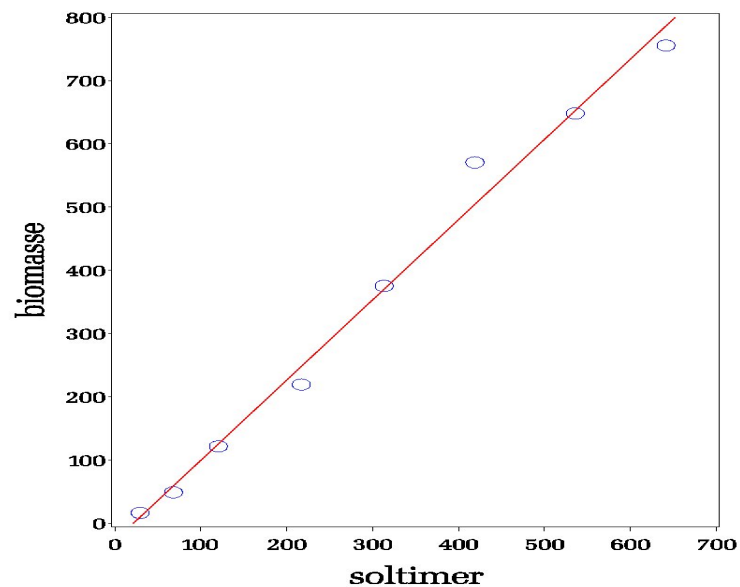
Giv et estimat for hældningen, med tilhørende 95% sikkerhedsinterval.

Undersøg, om hældningen kan antages at være 1.

Vi skal foretage en sædvanlig lineær regression med **biomasse** (**Y**) som respons og **soltimer** (**X**) som forklarende variabel:

$$Y_i = \alpha + \beta X_i + \varepsilon_i$$

men først tegner vi lige en linie ind på plottet ved at skrive `i=r1` i symbolsætningen, samt tilføje `l=1 ci=red w=3` for at få en fuldt optrukket linie (andre værdier af `l=` giver forskellige stiplinger) og bred (`w=3`) rød linie.:



```
proc gplot data=sol;
  plot biomasse*soltimer
  /  haxis=axis1 vaxis=axis2 frame;
axis1 offset=(3,3) label=(H=3) value=(H=2) minor=NONE;
axis2 offset=(1,1) value=(H=2) minor=NONE
      label=(A=90 R=0 H=3);
symbol1 v=circle c=blue i=r1 h=2 l=1 ci=red w=3 r=1;
run;

proc reg data=sol;
model biomasse=soltimer / clb;
run;
```

Vi får outputtet:

Dependent Variable: biomasse

Analysis of Variance

| Source | DF | Sum of Squares | Mean Square | F Value | Pr > F |
|-----------------|----|----------------|-------------|---------|--------|
| Model | 1 | 567033 | 567033 | 531.90 | <.0001 |
| Error | 6 | 6396.28000 | 1066.04667 | | |
| Corrected Total | 7 | 573429 | | | |

| | | | |
|----------------|-----------|----------|--------|
| Root MSE | 32.65037 | R-Square | 0.9888 |
| Dependent Mean | 344.63750 | Adj R-Sq | 0.9870 |
| Coeff Var | 9.47383 | | |

Parameter Estimates

| Variable | DF | Parameter Estimate | Standard Error | t Value | Pr > t |
|-----------|----|--------------------|----------------|---------|---------|
| Intercept | 1 | -27.56895 | 19.84220 | -1.39 | 0.2141 |
| soltimer | 1 | 1.26925 | 0.05503 | 23.06 | <.0001 |

Parameter Estimates

| Variable | DF | 95% Confidence Limits | |
|-----------|----|-----------------------|----------|
| Intercept | 1 | -76.12106 | 20.98316 |
| soltimer | 1 | 1.13458 | 1.40391 |

Vi ser af ovenstående output, at effekten af solskin er stærkt signifikant, idet et test af hældning = 0 giver $T=23.06$, svarende til en P-værdi, der er mindre end 0.0001. Samme P-værdi får man ved at anvende F-testet, idet $F=23.06^2=531.90 \sim F(1,6)$. Bemærk, at der i outputtet **ikke** findes noget test for linearitet!

Vi er imidlertid ikke blot interesserede i at påvise en effekt af solskin, vi vil *kvantificere* denne i form af et konfidensinterval for hældningen, som jo er tilvæksten i biomasse forårsaget af en enkelt solskinstime. Estimatet med tilhørende spredning (standard error) er

$$\hat{\beta} = 1.2692(0.0550)$$

og et 95% konfidensinterval fås som estimat \pm ca. 2 gange spredningen. Nu er de 'ca. 2' jo egentlig en t-fraktil, som for store datamaterialer nærmer sig 1.96. Her har vi kun 8 observationer, og dermed 6 frihedsgrader til estimation af variansen (se også ovenstående output), og t-fraktilen er derfor en del større end 2. Vi kan slå den op i bogen s. 521 og finder værdien 2.447, og vi kan nu udregne konfidensintervallet til:

$$1.2692 \pm 2.447 \times 0.055 = (1.13, 1.40)$$

sådan ca. det samme som vi fik i outputtet.

Vi kan altså sige, at intervallet (1.13,1.40) med 95% sandsynlighed indeholder den sande tilvækst i biomasse efter en enkelt solskinstime.

Da værdien 1 ikke er indeholdt i dette interval, kan vi med det samme sige, at på et 5% signifikansniveau kan vi ikke acceptere hypotesen om at hældningen er 1 (vi forkaster hypotesen $H_0 : \beta = 1$ på et 5% niveau).

Vi kan også direkte udregne et test for $H_0 : \beta = 1$ ved at omskrive hypotesen til $H_0 : \beta - 1 = 0$ og udregne teststørrelsen:

$$\frac{1.2692 - 1}{0.055} = 4.89$$

som helt klart er for stor, svarende til at vi allerede ved, at hypotesen skal forkastes. P-værdien kan slås op i t-tabellen (med sølle 6 frihedsgrader), hvilket giver $0.001 < P < 0.01$.

Man kan også gøre det med SAS ved at indføje en `test`-sætning:

```
proc reg data=sol;
  model biomasse=soltimer / clb;
  test soltimer=1;
run;
```

som vil give det ekstra output

Test 1 Results for Dependent Variable biomasse

| Source | DF | Mean Square | F Value | Pr > F |
|-------------|----|----------------|---------|--------|
| Numerator | 1 | 25516 | 23.94 | 0.0027 |
| Denominator | 6 | 1066.04667 | | |

Nu er der jo egentlig heller ikke rigtigt nogen grund til, at β skulle være 1, så vi burde slet ikke have opstillet sådan en hypotese, bare fordi $\hat{\beta}$ så ud til at være tæt på 1!

Spørgsmål 3.

Undersøg om interceptet kan antages at være 0. Hvad bliver hældningsestimatet under denne hypotese? Hvad sker der med spredningsestimatet for hældningen ved overgang fra modellen med intercept til modellen uden intercept (dvs. intercept=0)?

Fra outputtet svarende til den ovenfor udførte lineære regressionsanalyse finder vi estimatet for afskæringen (interceptet) med tilhørende spredning (standard error) til

$$\hat{\alpha} = -27.5690(19.8422)$$

T-test størrelsen for test af hypotesen $H_0 : \alpha = 0$ står også i output som -1.39, med en tilhørende P-værdi på 0.21, svarende til, at vi ikke kan forkaste denne hypotese (bemærk at vi ikke hermed har *bevist*, at $\alpha = 0$, vi har blot ikke her evidens for det modsatte).

Hvis vi antager, at interceptet er 0, skal vi reestimere hældningen ved at foretage en lineær regressionsanalyse gennem (0,0). Modellen hedder nu

$$Y_i = \beta X_i + \varepsilon_i$$

og den kan fittes ved at benytte option `noint` i `model`-sætningen:

```
proc reg data=sol;
model biomasse=soltimer / noint clb;
run;
```

Output bliver:

Dependent Variable: biomasse

NOTE: No intercept in model. R-Square is redefined.

Analysis of Variance

| Source | DF | Sum of Squares | Mean Square | F Value | Pr > F |
|-------------------|----|----------------|-------------|---------|--------|
| Model | 1 | 1515175 | 1515175 | 1254.54 | <.0001 |
| Error | 7 | 8454.24118 | 1207.74874 | | |
| Uncorrected Total | 8 | 1523629 | | | |

| | | | |
|----------------|-----------|----------|--------|
| Root MSE | 34.75268 | R-Square | 0.9945 |
| Dependent Mean | 344.63750 | Adj R-Sq | 0.9937 |
| Coeff Var | 10.08384 | | |

Parameter Estimates

| Variable | DF | Parameter Estimate | Standard Error | t Value | Pr > t |
|----------|----|--------------------|----------------|---------|---------|
| soltimer | 1 | 1.20705 | 0.03408 | 35.42 | <.0001 |

Vore nye hældningsestimat, med tilhørende spredning (standard error) bliver:

$$\hat{\beta} = 1.2071(0.0341)$$

medens vi i modellen **med intercept** fik

$$\hat{\beta} = 1.2692(0.0550)$$

Vi bemærker, at vi ved at tvinge interceptet til at være 0 (større end det oprindelige estimat, som jo var negativt) har fået et mindre hældningsestimat. Dette sker p.g.a. den negative korrelation mellem disse estimater. Vi bemærker endvidere, at spredningen på estimatet er faldet betragteligt (vi vinder generelt præcision ved at smide insignifikante effekter væk, specielt hvis disse er korrelerede med de interessante effekter). Populært kan man sige at vi nu arbejder i en model med *større viden*, hvilket naturligvis øger vores sikkerhed.

Spørgsmål 4.

Bestem et 95% sikkerhedsinterval for den estimerede biomasse produktion når det kumulerede antal solskinstimer når op på 200, for modellen *med* hhv. *uden* intercept. Forklar forskellen.

Modellen uden intercept er den letteste. Her er den estimerede biomasse ved 200 solskinstimer blot givet ved

$$200\hat{\beta} = 200 \times 1.2071 = 241.42$$

og usikkerheden er tilsvarende givet ved

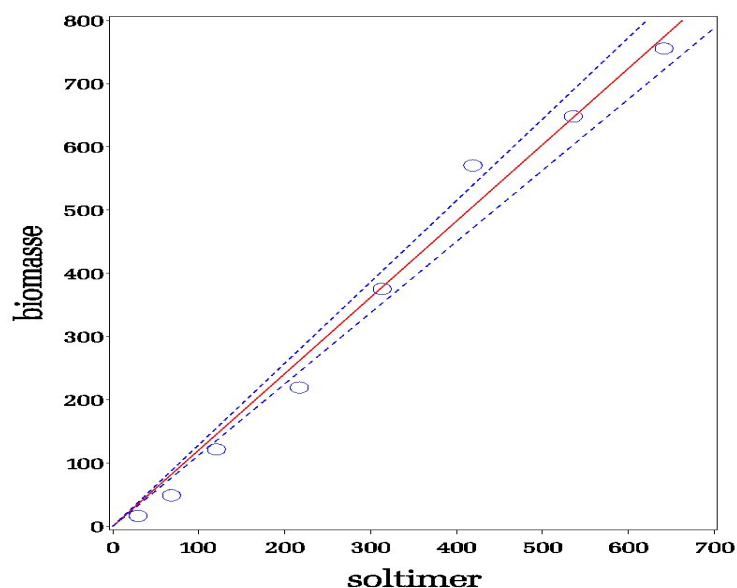
$$s.e.(200 \times \hat{\beta}) = 200 \times s.e.(\hat{\beta}) = 200 \times 0.0341 = 6.82$$

således at konfidensgrænserne bliver

$$200 \times (1.2071 \pm 2.365 \times 0.0341) = (225.29, 257.55)$$

Bemærk, at vi her anvender t-fraktilen svarende til 7 frihedsgrader (i stedet for som tidligere 6). Det er naturligvis fordi vi nu kun har en enkelt parameter i modellen.

Konfidensgrænserne kommer til at se således ud:



For modellen **med** intercept er det vanskeligt at udregne grænserne med håndkraft; her er det så vidt jeg kan se, lettest at gå over til at bruge `proc glm` og så tilføje en `estimate`-sætning:

```
proc glm data=sol;
  model biomasse=soltimer / clparm;
  estimate '200 soltimer' intercept 1 soltimer 200;
run;
```

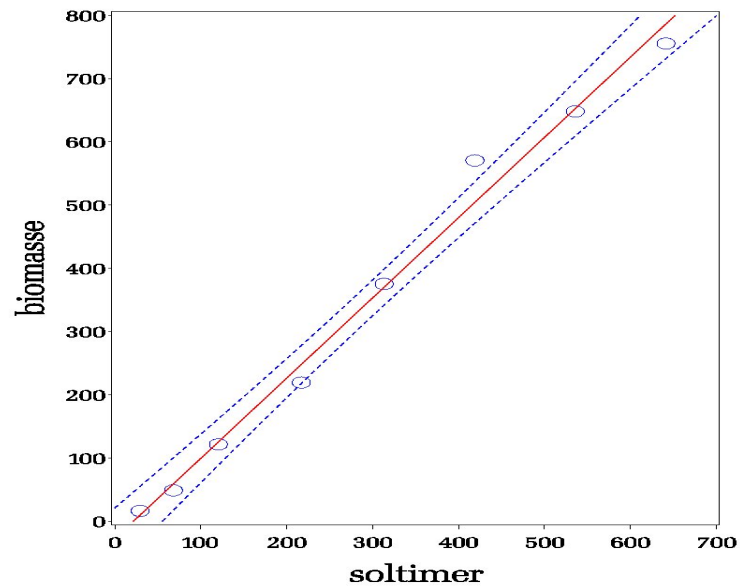
hvorved vi får det ekstra output

| Parameter | Estimate | Standard Error | t Value | Pr > t |
|--------------|------------|----------------|---------|---------|
| 200 soltimer | 226.280291 | 12.6329849 | 17.91 | <.0001 |

| Parameter | 95% Confidence Limits | |
|--------------|-----------------------|------------|
| 200 soltimer | 195.368490 | 257.192091 |

altså et estimat på 226.28 med konfidensintervallet (195.37,257.19).

Det tilhørende konfidensinterval fremgår af nedenstående figur:



Vi ser, at modellen *uden intercept* giver et noget højere predikteret udbytte ved 200 solskinstimer (241.42 mod 226.28 i modellen med intercept), svarende til, at vi stadig er 'i nærheden af 0', hvor linien jo er blevet løftet ved at smide interceptet ud. Konfidensgrænserne er tillige væsentligt smallere, igen på grund af den øgede præcision i en model med kun en parameter.