

Basal statistik

Logaritmer, Kovariansanalyse, Interaktion, i R

Lene Theil Skovgaard

9. marts 2020

1 / 93



Logaritmer og kovariansanalyse

- ▶ Parret sammenligning af målemetoder, med logaritmer
- ▶ Kovariansanalyse: Confounding og mediering
- ▶ Sammenligning af hældninger
Interaktion - igen
- ▶ Kovariansanalyse:
Eksempel om baseline og follow-up

Home pages:

<http://publicifsv.sund.ku.dk/~sr/BasicStatistics>

E-mail: ltsk@sund.ku.dk

*: Siden er lidt teknisk

2 / 93



Sammenligning af målemetoder

(jvf. eksemplet MF vs. SV fra første forelæsning)

To forskellige metoder til bestemmelse af glucosekoncentration.

Ref: R.G. Miller et.al. (eds): Biostatistics Casebook. Wiley, 1980

REFE:

Farvetest, der kan 'forurenes' af urinsyre

TEST:

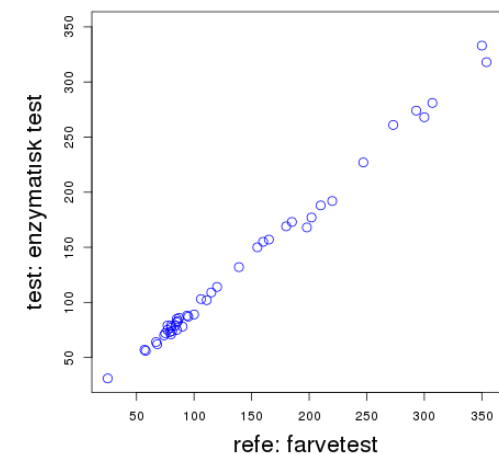
Enzymatisk test, mere specifikt for glucose.

nr.	REFE	TEST
1	155	150
2	160	155
3	180	169
·	·	·
·	·	·
·	·	·
44	94	88
45	111	102
46	210	188
$\bar{X}$	144.1	134.2
SD	91.0	83.2

3 / 93



Scatter plot af de to metoder



Det ser jo pænt ud...., eller?

4 / 93



Formål med undersøgelsen

kunne være:

- ▶ Kan vi erstatte en dårlig metode med en bedre metode?
- ▶ Kan vi erstatte en dyr metode med en billigere metode?
- ▶ Kan vi bruge de to metoder i flæng?

Under alle omstændigheder:

Hvor store forskelle kan vi forvente at finde mellem de to metoder

Og svaret er **ikke**:

Korrelationen er 0.99776.....

5 / 93



Parret sammenligning

Det er naturligt at se på differenser (dif=refe-test)

- ▶ Test om middelværdien kan være 0:
Systematisk forskel? (denne side)
- ▶ Kvantificer individuelle differenser:
Limits of agreement (næste side)

```
> t.test(refetest$dif)[c("p.value", "conf.int")]
$p.value
[1] 1.366351e-08

$conf.int
[1] 7.009941 12.772667
```

Med en P-værdi < 0.0001 er der **stærk indikation** af forskel på de to metoder

6 / 93



Limits of agreement

På basis af en normalfordelingsantagelse på differenserne finder vi referenceintervallet (normalområdet):

```
> mean(refetest$dif) + c(-2,2)*sd(refetest$dif)
[1] -9.514208 29.296817
```

med fortolkningen:

“Når vi måler med begge metoder på samme person, vil differensen typisk ligge i intervallet (-9.5, 29.3)”

På tegningen ses, at dette er en **dårlig** beskrivelse, idet

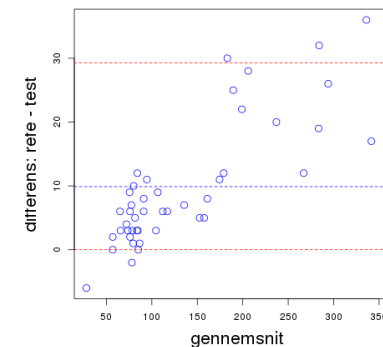
- ▶ differenserne stiger med niveauet (repræsenteret ved gennemsnittet)
- ▶ variationen stiger også med niveauet

7 / 93



Bland-Altman plot

Plot af differenser mod gennemsnit (af de to målinger på samme person), se kode s. 82



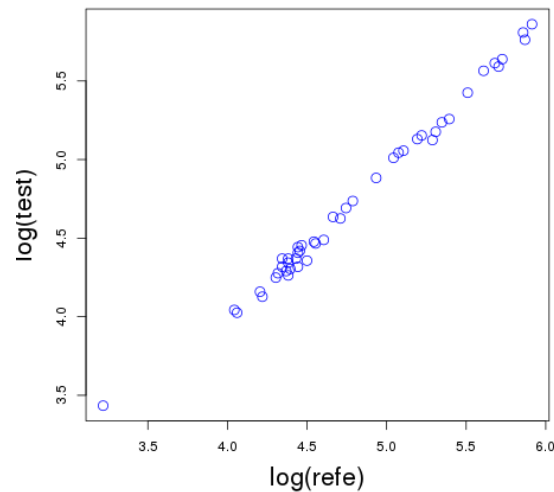
Store afvigelser ved høje målinger, dvs. **relative afvigelser**
Så er det smart at se på **logaritmer**

8 / 93



Scatter plot

efter **logaritmetransformation** (her “den naturlige”)



9 / 93



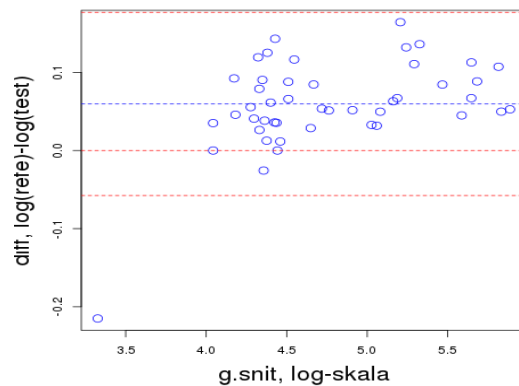
Bemærk

- ▶ Det er de **oprindelige målinger**, der skal logaritmetransformeres, **ikke** differenserne!
- ▶ Det er ligegyldigt, hvilken logaritmfunktion, der vælges (der er proportionalitet mellem alle logaritmer)
- ▶ Efter logaritmering gentages proceduren med differenser og konstruktion af limits of agreement

10 / 93



Bland-Altman plot **for logaritmer**



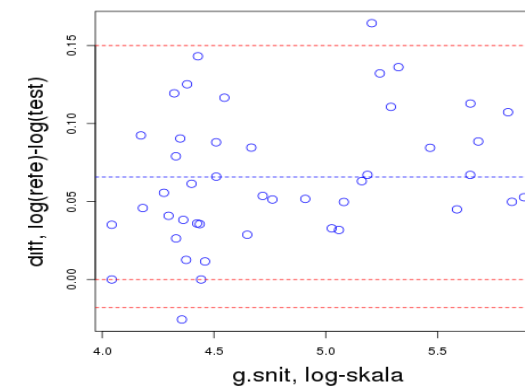
Der er en tydelig *outlier* (den mindste observation, nr. 31)

11 / 93



Vi udelader en outlier....nr. 31

og laver igen et **Bland-Altman plot**



som bliver acceptabelt....

12 / 93



En slags konklusion

```
> c(mean(refetest$ldif[-31]),sd(refetest$ldif[-31]))
[1] 0.06572945 0.04195475

> mean(refetest$ldif[-31]) + c(-2,2)*sd(refetest$ldif[-31])
[1] -0.01818005 0.14963895
```

Limits of agreement **på logaritmisk skala** er altså:

$$0.066 \pm 2 \times 0.042 = (-0.018, 0.150)$$

Det betyder, at der i 95% af tilfældene vil gælde

$$-0.018 < \log(\text{REFE}) - \log(\text{TEST}) = \log\left(\frac{\text{REFE}}{\text{TEST}}\right) < 0.150$$

Men hvad kan vi bruge det til?

13 / 93



Konklusion på brugbar facon

Vi kan **tilbagetransformere** med **anti-logaritmen** (her er det eksponentialfunktionen $\exp()$) og få

$$\exp(-0.018) = 0.982 < \frac{\text{REFE}}{\text{TEST}} < 1.162 = \exp(0.150) \quad \text{eller 'omvendt'}$$

$$0.861 < \frac{\text{TEST}}{\text{REFE}} < 1.018$$

Det betyder:

TEST ligger typisk mellem

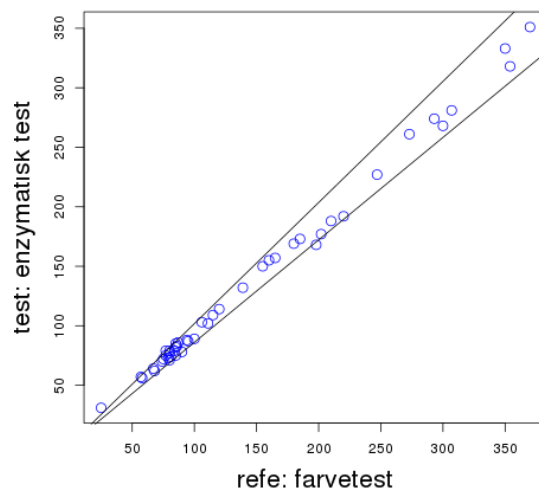
14% under og **2% over** REFE.

14 / 93



Limits of agreement

tilbagetransformeret til oprindelig skala



15 / 93



Regninger direkte på ratio-skala

Med definitionen $\text{ratio} = \frac{\text{refe}}{\text{test}}$ får vi:

```
> c(mean(refetest$ratio[-31]),sd(refetest$ratio[-31]))
[1] 1.06886072 0.04511841
> mean(refetest$ratio[-31]) + c(-2,2)*sd(refetest$ratio[-31])
[1] 0.9786239 1.1590975
```

dvs. **limits of agreement:**

$$1.069 \pm 2 \times 0.045 = (0.979, 1.159)$$

altså refe fra 2% under til 16% over test,

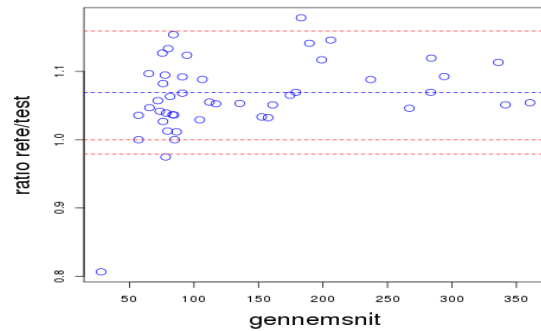
på 2 decimaler identisk med resultatet for logaritmerne

Dette er ikke altid tilfældet!!

16 / 93



Limits of agreement på ratio-skala



som jo er næsten identisk med figuren s. 11 bortset fra skalaen på akserne.

17 / 93



Terminologi

for **kvantitativt outcome**, f.eks. vitamin D

Regressionsanalyse: Kovariaterne er også **kvantitative**

- ▶ Simpel (lineær) regression: kun en enkelt kovariat
- ▶ Multipel (lineær) regression: to eller flere kovariater

Variansanalyse: Kovariaterne er **kategoriske** (grupper)

- ▶ Ensidedet variansanalyse: kun en enkelt kovariat
- ▶ Tosidedet variansanalyse: to kovariater

Generel lineær model: Begge typer kovariater i samme model

- ▶ **Kovariansanalyse (ANCOVA):**
Netop en kvantitativ og en kategorisk kovariat

18 / 93



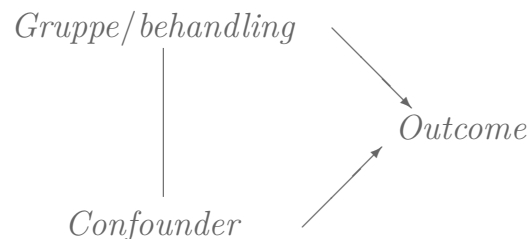
Sammenligning af to grupper

- som ikke er helt sammenlignelige, pga en **confounder**:

En variabel, som

- ▶ har en effekt på outcome
- ▶ er relateret til gruppen
(dvs. der er forskel på værdierne i de to grupper)

Herved **kan** opstå **"bias"**.



19 / 93



Eksempel: Vægt blandt mænd og kvinder

Mulig confounder: Højde

To forskellige sammenligninger:

- ▶ **T-test:**
Er der forskel på middel vægt blandt mænd og kvinder?
- ▶ **Kovariansanalyse:**
Er der forskel på middel vægt for mænd og kvinder, **med samme højde?**
Kønseffekten **korrigeres** for forskel i højde:

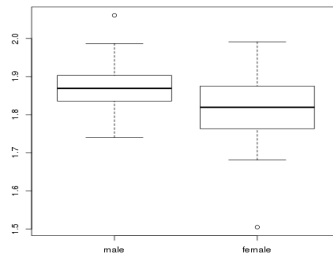
To forskellige videnskabelige spørgsmål

20 / 93



Vægtfordeling for mænd og kvinder

her for 100 tilfældigt udvalgte individer fra Sundby-materialet, på logaritmisk ($\log_{10}$ -skala):

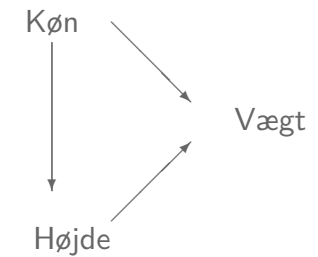
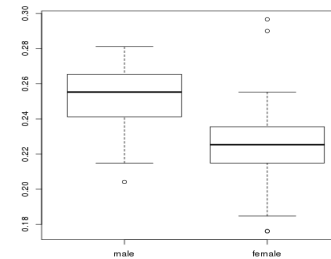


Et T-test giver $P=0.0002$, og en estimeret "fordel" til mænd på 14.5% (6.8-22.8%), se kode s. 83

21 / 93



Kan højden forklare forskellen på mænd og kvinder?



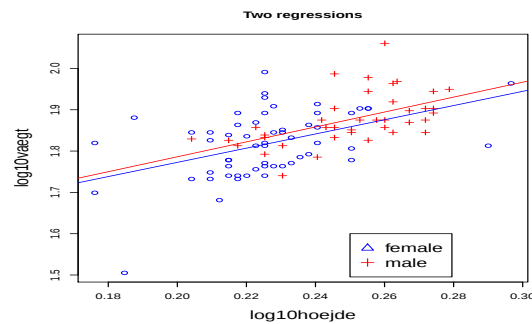
- Der er i hvert fald forskel på mænds og kvinders højde
- Men er der tillige en *ægte* kønseffekt?

22 / 93



Relation mellem højde og vægt

Igen på $\log_{10}$ -skala, og med indlagte regressionslinier for hvert køn, se kode s. 84



Der er et par meget lave kvinder, men ingen synderlige afvigelser fra linearitet...

23 / 93



Kovariansanalyse

Outcome: Vægt, logaritmeret ($\log_{10}$ vaegt)

Kovariater: 2 stk:

- Køn (gender)
- Højde, logaritmeret ($\log_{10}$ hoejde), med lineær effekt

```
ancova = lm(log10vaegt ~ log10hoejde + gender,
na.action=na.exclude, data=su)
```

24 / 93



Output fra kovariansanalyse

Den vigtigste del:

```
> summary(ancova)

Call:
lm(formula = log10vaegt ~ log10hoejde + gender,
    data = su, na.action = na.exclude)

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   1.43930    0.08237   17.475 < 2e-16 ***
log10hoejde   1.74854    0.32638    5.357 6.09e-07 ***
genderfemale -0.01716    0.01571   -1.093  0.277

> 10^(cbind(-confint(ancova)[3,1],
+          -ancova$coefficients[3], -confint(ancova)[3,2]))
              [,1]      [,2]      [,3]
genderfemale 1.117764 1.04031 0.9682231
```

Bemærk: Nu er kønseffekten helt forsvundet!

Højden kunne altså godt forklare den....måske

25 / 93



Fortolkning af kovariansanalysen

Vi fitter **to parallelle linier** (se s. 28) men både outcome og den kvantitative kovariat er logaritmeret....

Fortolkning af kønseffekt:

På $\log_{10}$ -skala ligger mænds vægt 0.01716 højere end kvinders. Når vi tilbagetransformerer dette, får vi $10^{0.01716} = 1.040$, svarende til, at mænd vejer ca. 4% mere end kvinder **med samme højde**.

Tilsvarende tilbagetransformerer konfidensintervallerne:

$(10^{-0.0140245}, 10^{0.04835}) = (0.968, 1.118)$, dvs. det kan tænkes, at mænds vægt er hele 11.8% højere end kvinders **med samme højde**, men det kan også tænkes, at den faktisk er 3.2% lavere (svarende til at gange med faktoren 0.968).

26 / 93



Fortolkning af kovariansanalysen, fortsat

Fortolkning af effekt af højde:

Højden på log-skala antages at have en lineær effekt på vægten på log-skala. Det betyder, at vi skal fortolke den tilbagetransformerede koefficient som **effekten af en 10-dobling af højden**, hvilket jo ikke er særligt interessant....

Når både outcome og kovariat er logaritmeret **med samme logaritme**, er det imidlertid ligegyldigt, hvad grundtallet for logaritmen er, da koefficienten altid vil blive den samme.

Vi kan derfor blot sige, at en 10% forøgelse af højden (svarende til faktoren 1.1) bevirker en faktor $1.1^{1.74854} = 1.181$ på vægten, altså en forøgelse af vægten med 18.1%.

Tilsvarende tilbagetransformerer konfidensintervallerne:

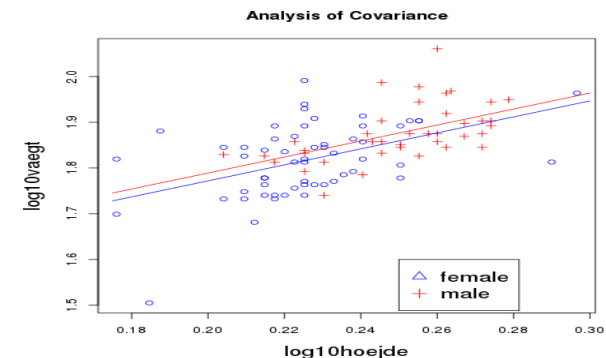
$(1.1^{1.10043}, 1.1^{2.39666}) = (1.111, 1.257)$, dvs. mellem 11.1% og 25.7% mere i vægt.

27 / 93



Ancova-fit (parallelle linier)

Bemærk, at de parallelle linier er næsten sammenfaldende, svarende til den ringe kønseffekt.



Se kode s. 86

28 / 93



Estimation af kønseffekt

i 2 forskellige modeller, nemlig **med** og **uden** højde som kovariat

Outcome er $\log_{10}$ vægt:

Kovariater	Mænd vs. kvinder ratio (CI)	P-værdi
kun kønnet	1.14 (1.07, 1.23)	0.0002
køn og højde	1.04 (0.97, 1.12)	0.28

Den observerede forskel i ($\log_{10}$) vægt mellem mænd og kvinder **kan** altså evt. tilskrives højdeforskellen mellem kønnene.

29 / 93



Er der så en kønseffekt eller ej?

Ja for mænd og kvinder har jo ikke samme middel vægt **men** forskellen i niveau kan forklares ud fra den højdeforskel, der generelt er mellem mænd og kvinder **måske** men vi kan ikke *påvise* forskel mellem en mand og en kvinde med samme højde

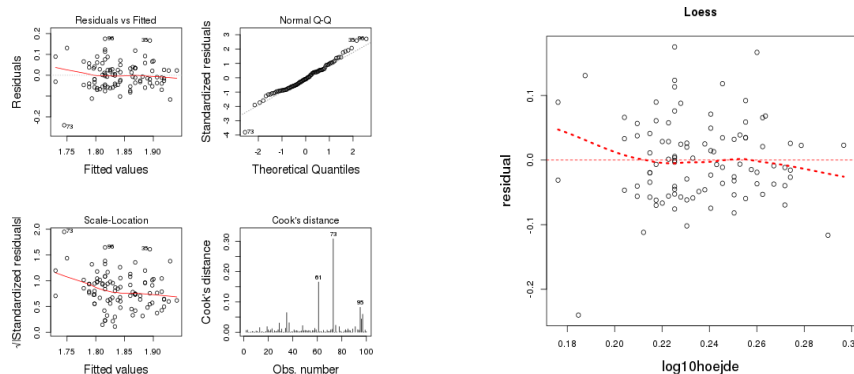
Højde er en **mellemkommende** (intermediate) variabel for effekten af køn på vægt

30 / 93



Husk modelkontrol:

Se kode s. 87



31 / 93



Additivitet – eller interaktion?

Kovariansanalyse: dækker sædvanligvis over situationen med to **parallelle** linier, altså med identiske hældninger, også kaldet **additivitet**. **Her:** Effekten af højde på vægt antages at være *den samme* for mænd og kvinder

Mere generel model tillader

interaktion=vekselvirkning=effektmodifikation, altså *forskellige* hældninger. Det betyder:

- Effekten af højde afhænger af kønnet
- Forskellen på kønnene afhænger af højden

I tilfælde af interaktion kan man **ikke** udtale sig om **én** generel effekt af hverken højde eller køn.

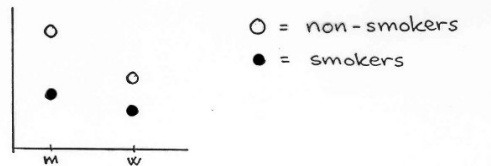
32 / 93



Vekselvirkning (interaktion)

Tænkt eksempel:

- To inddelingskriterier: køn og rygestatus
- Outcome: FEV₁



- Effekten af rygning **afhænger af** køn
- Forskellen på kønnene **afhænger af** rygestatus

33 / 93

Mulige forklaringer

- **biologisk kønsforskel** på effekt af rygning
 - holder vist ikke i praksis, men eksemplet er jo også blot 'tænkt'
- måske ryger kvinderne ikke helt så meget
 - antal pakkeår confounder for køn
- måske virker rygningen som en **relativ** (%-vis) nedsættelse af FEV₁
 - kunne undersøges ved en longitudinel undersøgelse

34 / 93

Eksempel: Rygnings effekt på fødselsvægt

Table 10.1 Average Birth Weight of Children Born to Women with Different Amount and Duration of Smoking^a

Duration of smoking in pregnancy	Amount of smoking			All
	Mild	Moderate	Heavy	
–18 weeks	3.45 (n = 15)	3.42 (n = 12)	3.43 (n = 7)	3.44 (n = 34)
18–31 weeks	3.38 (n = 8)	3.40 (n = 10)	3.39 (n = 6)	3.39 (n = 24)
32+ weeks	3.35 (n = 5)	3.30 (n = 3)	3.18 (n = 9)	3.25 (n = 17)
All	3.41 (n = 28)	3.40 (n = 25)	3.32 (n = 22)	3.38 (n = 75)

^a Entries are average birth weight in kg.

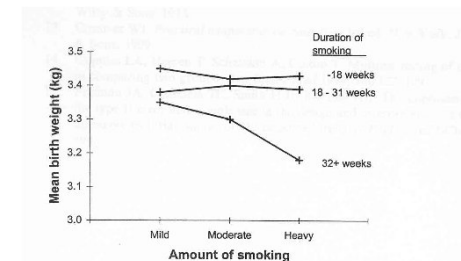


Figure 10.3 Interaction between duration and amount of smoking.

Interaktion/vekselvirkning mellem mængden og varigheden af rygningen

- Der er effekt af mængden, men **kun** hvis man har røget længe.
- Der er effekt af varigheden, og denne effekt **øges** med mængden.

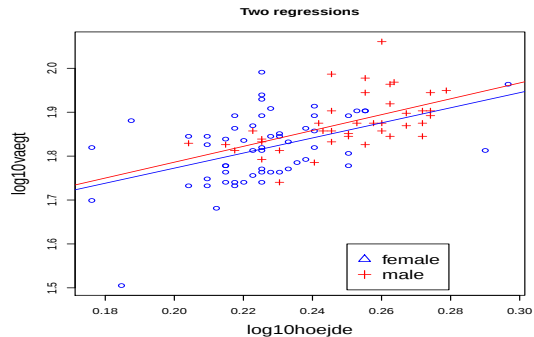
Effekten af mængden **afhænger af**....
og effekten af varigheden **afhænger af**....

35 / 93

36 / 93

Model med interaktion

Nu er linierne ikke mere *tvunget* til at være parallelle, men de er det næsten alligevel...



Kode til figuren s. 84

37 / 93



Model med interaktion i praksis

Testvenlig kode:

```
interaktion = lm(log10vægt ~ log10højde + gender +
  gender:log10højde,
  na.action=na.exclude, data=su)
```

Mere estimat-venlig kode:

```
interaktion1 = lm(log10vægt ~ gender +
  gender:log10højde -1,
  na.action=na.exclude, data=su)
```

38 / 93



Output fra testvenlig model

Kode s. 38, foroven

```
> summary(interaktion)

Call:
lm(formula = log10vægt ~ log10højde + gender + gender:log10højde,
    data = su, na.action = na.exclude)

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   1.425146   0.140381  10.152 < 2e-16 ***
log10højde    1.805082   0.559213   3.228 0.00173 **
genderfemale  -0.003698   0.167823   0.022 0.98247
log10højde:genderfemale -0.086223   0.690582  -0.125 0.90091

> confint(interaktion)
              2.5 %    97.5 %
(Intercept)  1.1463374  1.7039541
log10højde   0.6944371  2.9157269
genderfemale -0.3296129  0.3370087
log10højde:genderfemale -1.4577795  1.2853334
```

Interaktionsparameteren angiver
forskellen på hældninger i de to grupper

39 / 93



Fortolkning af interaktionsmodel

Vi har fittet to *vilkårlige linier* (se s. 37) men stadig med
logaritmeret outcome og kovariat.....

Fortolkning af kønseffekt:

Da linierne ikke er parallelle, er der ikke mere noget, der hedder
kønseffekt **en**, da afstanden mellem linierne på s. 37 varierer en
smule

Det angivne estimat for gender angiver forskellen på mænd og
kvinder med $\log_{10}\text{højde}=0$, dvs. en højde på 1 meter, altså
overhovedet ikke interessant.

40 / 93



Fortolkning af interaktionsmodel, fortsat

Fortolkning af effekt af højde:

Nu er der *to forskellige effekter af højde*, svarende til de to køn. De fortolkes ligesom s. 27, idet koefficienten til $\log_{10}\text{hoejde}$ angiver effekten for referencegruppen for gender (altså mænd), medens koefficienten svarende til $\log_{10}\text{hoejde}:\text{genderfemale}$ angiver *forskellen* mellem effekten for kvinder og effekten for mænd.

Vi omregner disse estimater på s. 42, men se estimationsvenlig kode s. 38 (nederst) og output fra denne s. 43

41 / 93



Omregning til de to linier:

Linie for mænd:

$$\log_{10}(\text{vægt}) = 1.4251 + 1.8051 \times \log_{10}(\text{højde})$$

Linie for kvinder:

$$\begin{aligned} \log_{10}(\text{vægt}) &= 1.4251 + 0.0037 + (1.8051 - 0.0862) \times \log_{10}(\text{højde}) \\ &= 1.4288 + 1.7189 \times \log_{10}(\text{højde}) \end{aligned}$$

Men så får vi ikke nogen konfidensgrænser....
så man bør benytte programmet til udregningerne,
se kode nederst s. 38

42 / 93



Output fra estimationsvenlig kode

Kode s. 38 (nederst)

```
> summary(interaktion1)

Call:
lm(formula = log10vaegt ~ gender + gender:log10hoejde - 1, data = su,
    na.action = na.exclude)

Residuals:
    Min       1Q   Median       3Q      Max
-0.241152 -0.045355 -0.005657  0.035528  0.175107

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
gendermale      1.42515    0.14038  10.152 < 2e-16 ***
genderfemale    1.42884    0.09197  15.537 < 2e-16 ***
gendermale:log10hoejde 1.80508    0.55921   3.228 0.00173 **
genderfemale:log10hoejde 1.71886    0.40520   4.242 5.27e-05 ***

> confint(interaktion1)
              2.5 %    97.5 %
gendermale    1.1463374  1.703954
genderfemale  1.2461910  1.611496
gendermale:log10hoejde 0.6944371  2.915727
genderfemale:log10hoejde 0.9141012  2.523617
```

43 / 93



* Fortolkning af interaktionsmodel, fortsat

Vi har nu direkte estimater for de to linier (se s. 42-43) og skal tilbagetransformere til relationen

$$\text{vægt} = 10^\alpha \times \text{højde}^\beta$$

for hvert køn for sig. Vi finder:

- ▶ **For mænd:**
 $\text{vægt} = 10^{1.4251} \times \text{højde}^{1.8051} = 26.61 \text{højde}^{1.8051}$
- ▶ **For kvinder:**
 $\text{vægt} = 10^{1.4288} \times \text{højde}^{1.7189} = 26.84 \text{højde}^{1.7189}$

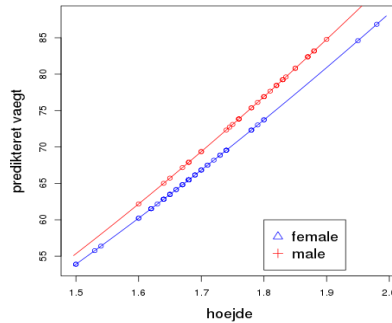
Disse *kurver* (**ikke linier**) er vist på s. 45.
Bemærk dog, at de ikke afviger ret meget fra linier....

44 / 93



Estimerede relationer i interaktionsmodellen

Da der er tale om linier på dobbeltlogaritmisk skala, er der ved tilbagetransformation tale om krumme kurver (dog ikke særligt krumme...). Se kode s. 88



45 / 93



Bemærk

Da der absolut ikke kunne spores nogen interaktion i dette eksempel, går vi tilbage til den additive model, *kovariansanalysen*, fra s. 24. I **dette** eksempel så vi

- Den observerede forskel i ($\log_{10}$) vægt mellem mænd og kvinder **kunne godt** tilskrives højdeforskellen mellem kønnene.

Der **kan** dog stadig være en kønsforskel **op til ca. 12% øget vægt hos mænd** i forhold til en kvinde på samme højde (se s. 29).

- Det *omvendte* kan også forekomme, altså at **effekter dukker op, når der kontrolleres for andre**

Sådan et eksempel skal vi se nu

46 / 93



Nyt eksempel: Hormoner hos P-pille brugere

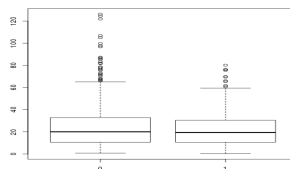
Anti Müllersk Hormon (**AMH**) er et hormon, som dannes i de små tidlige ægblærer (follikler) i æggestokkene.

Videnskabeligt spørgsmål:

Er AMH-niveauet nedsat hos P-pille brugere?

Og i givet fald hvor meget?

Der foreligger en undersøgelse af 732 kvinder, heraf 228, der tager P-piller



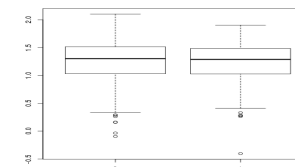
47 / 93



Valg af skala

- Boxplottet på forrige side viser klart en tendens til hale mod høje værdier
- Man får derfor en ide om, at logaritmer vil være påkrævet
- På den anden side er der mange observationer i hver gruppe, så det er muligvis ikke strengt nødvendigt....

Boxplot af logaritmetransformererede ($\log_{10}$) AMH-værdier



Heller ikke på denne skala er der perfekt symmetri

48 / 93



Valg af skala

er ikke altid oplagt....

Man kan lade det afhænge af, hvordan man helst vil præsentere sit resultat:

- ▶ som en forskel i enheden antal/ml, eller
- ▶ som en **relativ forskel**, i %

Hvis man vælger andre skalaer (kvadratrødder, kubikrødder etc.) kan parametrene i modellen ikke direkte fortolkes (forskellene kan ikke kvantificeres på en enkel måde).

49 / 93



Sammenligning af grupper - på logaritmisk skala

Uparret T-test giver:

```
> t.test(pill$logamh ~ pill$ppiller)

Welch Two Sample t-test

data:  pill$logamh by pill$ppiller
t = 0.86861, df = 441.56, p-value = 0.3855
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.03181567  0.08221097
sample estimates:
mean in group 0 mean in group 1
 1.264578      1.239381

> var.test(pill$logamh ~ pill$ppiller)

F test to compare two variances

data:  pill$logamh by pill$ppiller
F = 1.0162, num df = 503, denom df = 227, p-value = 0.8987
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.8099674 1.2628745
sample estimates:
ratio of variances
 1.016165
```

50 / 93



Kommentarer til output

- ▶ Der er ikke tegn på forskellige varianser i de to grupper (var.test: P=0.90)
- ▶ Det er derfor ligegyldigt, hvilket T-test, vi vælger at se på (de giver i øvrigt også næsten præcis det samme)
- ▶ R giver dog kun Welch-testet, med mindre man specificerer var.equal=TRUE
- ▶ P-værdien er 0.39, altså **ingen evidens for forskel** i AMH blandt brugere og ikke-brugere af P-piller
- ▶ R giver *ikke* selve differensen:
Ikke-P-piller vs. P-piller: $1.2646 - 1.2394 = 0.0252$

51 / 93



Kommentarer til output, II

Måske er det mere naturligt at udregne det *den anden vej*, altså P-pille brugere i forhold til ikke P-pille brugere, og så skal vi skifte fortegn:

$$10^{-0.0252} = 0.94, \quad CI = (10^{-0.0823}, 10^{0.0319}) = (0.83, 1.08)$$

Her læser vi, at P-pille brugere i gennemsnit har et ca. 6% reduceret niveau af AMH, men at konfidensintervallet strækker sig fra en reduktion på ca. 17% til et forøget niveau på omkring 8%.

Det lyder som en **ret betydelig potentiel forskel** i mine ører....
men altså ikke signifikant

Kan der være en confounder? Alder?

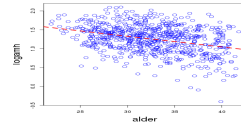
52 / 93



Er alder en confounder?

► Har den en effekt på hormon-niveauet?

Det skulle man tro, ud fra vores viden om, hvad hormonet afspejler



► Er der forskellig alder i de to grupper?

Formentlig er P-pille brugerne yngre end ikke-P-pille brugerne

```
> tapply(pill$alder, pill$ppiller, summary)
$'0'
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  21.90  30.90   33.70   33.72  36.90   41.80

$'1'
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  22.00  26.90   29.20   30.02  32.60   40.70
```

Alder er en confounder!

53 / 93



Kovariansanalyse - justering for aldersforskel

Nu sammenligner vi kvinder, der tager P-piller med kvinder, der *ikke* tager P-piller, men som har **samme alder**:

```
> ancova = lm(logamh ~ alder + factor(ppiller),
+   na.action=na.exclude, data=pill)
> summary(ancova)
```

```
Call:
lm(formula = logamh ~ alder + factor(ppiller),
    data = pill, na.action = na.exclude)
```

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   2.434239   0.104927  23.199 < 2e-16 ***
alder         -0.034684   0.003079 -11.263 < 2e-16 ***
factor(ppiller)1 -0.153784   0.029197  -5.267 1.83e-07 ***
```

```
> confint(ancova)
              2.5 %      97.5 %
(Intercept)  2.22824403  2.64023483
alder        -0.04072937 -0.02863812
factor(ppiller)1 -0.21110449 -0.09646268

> 10^(confint(ancova))
              2.5 %      97.5 %
(Intercept) 169.1391073 436.7519270
alder        0.9104805  0.9361854
factor(ppiller)1 0.6150289 0.8008244
```

54 / 93



Kommentarer til kovariansanalyse

Kvinder, der tager P-piller estimeres til et 0.1538 lavere niveau end kvinder, der ikke tager P-piller, med CI=(0.0965, 0.2111).

Når vi tilbagetransformerer til selve koncentrationsskalaen, giver det en estimeret ratio $10^{-0.1538} = 0.70$, med konfidensinterval $(10^{-0.2111}, 10^{-0.0965}) = (0.62, 0.80)$

Dette skal fortolkes som, at P-pille brugere ligger **30% lavere** end ikke-P-pille brugerne, med konfidensinterval fra 20% til 38% under.

Og forskellen er stærkt signifikant ($P < 0.0001$)

55 / 93



Kommentarer til de to analyser

Vi sammenfatter i en tabel:

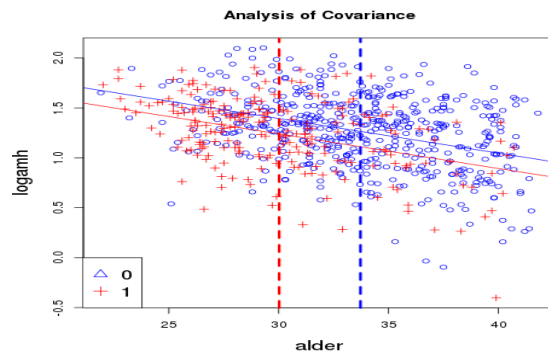
Model	Reduktion i AMH: P-pille brugere vs. ikke P-pille brugere	P-værdi
T-test (s. 50-52)	-6% (-17%, +8%)	0.39
Kovariansanalyse (s. 54-55)	-30% (-38%, -20%)	<0.0001

56 / 93



Kovariansanalyse modellen grafisk

med fortolkning næste side



Men kan der tænkes at være interaktion?

57 / 93

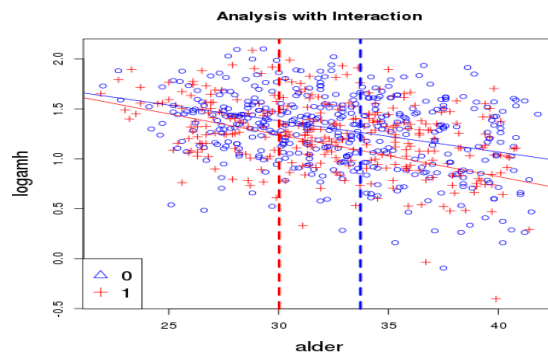


58 / 93



Figur med *frie* regressionslinier

svarende til *mulig interaktion*



Er de to linier ikke-parallele?

59 / 93



60 / 93



Fortolkning af figur s. 57

- Kovariansanalysen svarer til de to parallelle regressionslinier
- Den lodrette afstand imellem disse svarer til effekten af P-piller i kovariansanalysen s. 54-55
- De to lodrette linier er anbragt i aldersgennemsnittene:

Rød P-pille brugere

Blå Ikke-P-pille brugere

Disse linier ses at skære regressionslinierne i *nogenlunde* samme højde (dvs. værdi af $\log_{amh}$), og ca. svarende til det insignifikante T-test s. 50-52

Fortolkning af figur s. 59

- Interaktionsmodellen svarer til de to *ikke*-parallelle regressionslinier
- Den lodrette afstand imellem disse varierer nu med alderen, så man ikke mere kan tale om en *generel* effekt af P-piller
- De to lodrette linier skærer nu regressionslinierne i *præcis* de gennemsnitlige værdier af $\log_{amh}$, svarende til det insignifikante T-test s. 50-52

Er effekten af alder forskellig i de to grupper?

Regression for hver gruppe for sig (kode s. 91)

```
> pill.0 <- subset(pill, ppiller==0,)
>
> reg.0 = lm(logamh ~ alder, na.action=na.exclude, data=pill.0)

> cbind(confint(reg.0)["alder",1],
+       reg.0$coefficients["alder"], confint(reg.0)["alder",2])
      [,1]      [,2]      [,3]
alder -0.03879627 -0.03131774 -0.02383922

-----

> pill.1 <- subset(pill, ppiller==1,)
> reg.1 = lm(logamh ~ alder, na.action=na.exclude, data=pill.1)
> cbind(confint(reg.1)["alder",1],
+       reg.1$coefficients["alder"], confint(reg.1)["alder",2])
      [,1]      [,2]      [,3]
alder -0.05207592 -0.04186509 -0.03165426
```

Konfidensintervallerne for alderseffekterne er noget overlappende, så vi kan ikke umiddelbart udtale os om, hvorvidt de er signifikant forskellige.

61 / 93



Interaktion mellem P-piller og alder?

De to analyser s. 61 viser, at effekten af 5 år (5 gange hældningen) kan opsummeres som i nedenstående tabel, der også tilbagetransformerer effekten til original skala:

Gruppe	Effekt af 5 år på $\log_{10}$ -skala	Tilbagetransformeret effekt
Ikke P-pille brugere	-0.1566 (-0.1940, -0.1192)	0.70 (0.64, 0.76)
P-pille brugere	-0.2093 (-0.2604, -0.1583)	0.62 (0.55, 0.69)

Fortolkningen er, at for P-pille brugere falder AMH-niveauet med 38% på 5 år (faktoren 0.62), medens der for ikke P-pille brugere kun er tale om et fald på 30% (faktoren 0.7).

62 / 93



Interaktion mellem P-piller og alder, II

Er der evidens for reelle forskelle i alderseffekten i de to grupper?

For at undersøge dette, skal vi først gøre os klart, at spørgsmålet går på, om de to hældninger ovenfor er signifikant forskellige, altså om alderseffekten afhænger af, om man er P-pille bruger eller ej.

```
> interaktion = lm(logamh ~ alder + ppiller + ppiller:alder,
+ na.action=na.exclude, data=pill)
> summary(interaktion)
```

```
Call:
lm(formula = logamh ~ alder + ppiller + ppiller:alder, data = pill,
    na.action = na.exclude)
```

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.320726    0.126612  18.329 < 2e-16 ***
alder        -0.031318    0.003728  -8.401 2.32e-16 ***
ppiller       0.175287    0.207947   0.843   0.40
alder:ppiller -0.010547    0.006599  -1.598   0.11
---
```

```
> confint(interaktion)
              2.5 %      97.5 %
(Intercept)  2.07215815  2.569293108
alder        -0.03863669 -0.023998801
ppiller       -0.23295989  0.583534066
alder:ppiller -0.02350310  0.002408406
```

63 / 93



Kommentarer til output vedr. interaktion

- ▶ Testet for ens hældninger aflæses på linien `alder:ppiller`, og P-værdien ses at være 0.11, altså ingen signifikans.
- I hvert fald ikke, hvis vi holder os til et 5% signifikansniveau.
- ▶ Vi kan således ikke i dette materiale påvise en forskel i alderseffekt for de to grupper, omend der synes at være tendens til en større effekt i P-pille gruppen.

Der *kunne* således være en forskel i hældninger på op imod $0.0105 + 2 \times 0.0066 = 0.0237$ svarende til, at P-pille brugere på 5 år havde et ekstra tab på 24% (fordi $10^{-5 \times 0.0237} = 0.76$)

64 / 93



Kvantificering i tilfælde af interaktion

Her finder vi ingen signifikant interaktion mellem alder og indtagelse af P-piller, men dette *udelukker* jo ikke, at der kan være interaktion.

Og hvordan kvantificerer man så effekten af P-piller i tilfælde af interaktion?

- ▶ Man kan kvantificere effekten for udvalgte aldre, f.eks. 25, 30, 35 og 40, eller for en lang række aldre ved at inkludere disse i et nyt datasæt, som man så predikterer for, se s. 66
- ▶ eller ved at ændre Y-aksens placering ved at flytte nulpunktet, så effekten kvantificeres som forskellen i intercepter

Selve niveauerne kan kvantificeres for en lang række aldre ved at prediktere fra et datasæt uden AMH-værdier, se s. 93

65 / 93



Kvantificering for mange forskellige aldre

```
> e.P25=confint(summary(glm(interaktion, linfct=rbind(c(0,0,1,25)))))
> e.P30=confint(summary(glm(interaktion, linfct=rbind(c(0,0,1,30)))))
> e.P35=confint(summary(glm(interaktion, linfct=rbind(c(0,0,1,35)))))
> e.P40=confint(summary(glm(interaktion, linfct=rbind(c(0,0,1,40)))))
```

som giver

```
> e.P25$confint
      Estimate      lwr      upr
1 -0.08839662 -0.1870355  0.01024224

> e.P30$confint
      Estimate      lwr      upr
1 -0.1411334 -0.2004643 -0.08180239

> e.P35$confint
      Estimate      lwr      upr
1 -0.1938701 -0.2693902 -0.11835

> e.P40$confint
      Estimate      lwr      upr
1 -0.2466068 -0.374196 -0.1190177
```

som så skal tilbagetransformeres....

66 / 93



Metoder til at undgå bias (her pga alder)

Matchning. Dvs. udvælge individer, således at de er nogenlunde ens med hensyn til de vigtige *forstyrrende* kovariater.

Husk at tage matchningsvariablen med som kovariat

Randomisering. Dvs. trække lod om behandling (gruppe)
NB: Dette kan naturligvis *kun* lade sig gøre, hvis grupperne er noget, man selv bestemmer over.
og det sikrer ikke helt identiske fordelinger i den enkelte undersøgelse

Korrektion Dvs. at medtage den (muligvis skævt fordelte) variabel som kovariat, **også somme tider i randomiserede undersøgelser**

67 / 93



Matching / randomisering / korrektion

▶ Matching:

- ▶ god metode, når den ene gruppe er vanskelig at få tag i
- ▶ kan være ret besværligt og kan reducere antallet af observationer
- ▶ skævvrider udtrykket for populationen som helhed

▶ Randomisering:

- ▶ vældig god metode, når den er mulig og sample size er stor
- ▶ vi kan ikke randomisere til P-piller ja/nej

▶ Korrektion:

- ▶ er altid mulig, men problematisk hvis der er for stor forskel på grupperne
- ▶ kan reducere residualvariationen (væsentligt)
- ▶ sample size bestemmer hvor meget, det er muligt at korrigere for....

68 / 93



Metoder til at øge styrken

- ▶ inkludere flere observationer/personer
- ▶ benytte fornuftige inklusions- eller eksklusionskriterier
- ▶ inddrage vigtige forklarende variable (kovariater)
også selv om de ikke er confoundere,
se eksempel på en sådan næste side
- ▶ udelade irrelevante kovariater

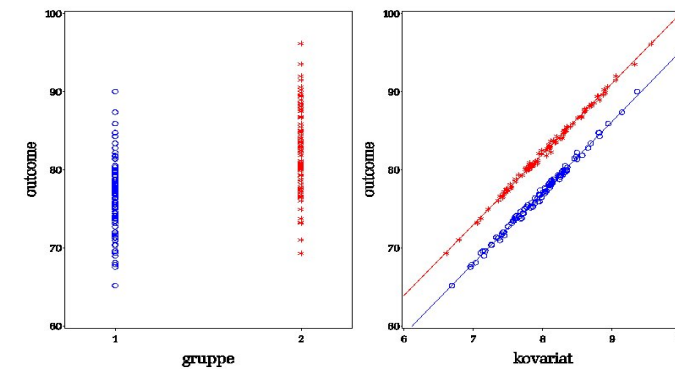
Men pas på med at gå for meget på **fisketur!!**

og husk at fortolkningen afhænger af modellen

69 / 93



Simuleret eksempel



Uden x i modellen: Ingen særlig forskel på grupperne...?

Med x i modellen: Tydelig forskel på grupperne
(her den lodrette afstand mellem linierne)

70 / 93



Før-og-efter studie

Et typisk eksempel:

Vickers, A.J. & Altman, D.G.: Analysing controlled clinical trials with baseline and follow-up measurements.

British Medical Journal 2001; **323**: 1123-24.:

52 patienter med skuldersmerter **randomiseres** til enten

- ▶ Akupunktur (n=25)
- ▶ Placebo (n=27)

Smerte vurderes på en 100-trins skala
både før og efter behandling

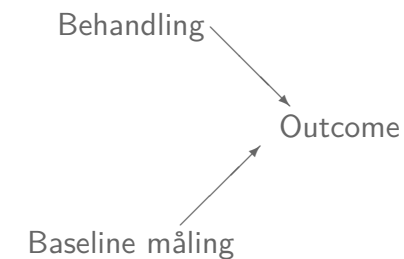
Høje scorer er gode

71 / 93



Baseline og Follow-up

I princippet:



Der er fokus på sammenligning af behandlingerne:

Er placebo ligeså godt som akupunktur?

Det lægger op til et simpelt T-test, **men**:

Hvad nu, hvis de to grupper ikke er *helt* identiske fra start,
selv om der er randomiseret?

72 / 93



Gentagne målinger over tid

– egentlig først emnet den sidste uge af kurset, men her er det simpelt, fordi vi kun har 2 målinger pr. individ, så:

Vi kan se på ændringerne, men er dette fornuftigt?

Ikke altid, pga **regression to the mean** eller “reversion” towards the mean, som kan oversættes til en tendens til at falde ind mod midten af en fordeling

Talrige eksempler fra dagligdagen:

- ▶ Efterfølgeren til en rigtig god film bliver sjældent helt så god
- ▶ Lungekræftdødeligheden i Fredericia er aftaget
- ▶ Placebo-effekten

Det handler om **seleksion**

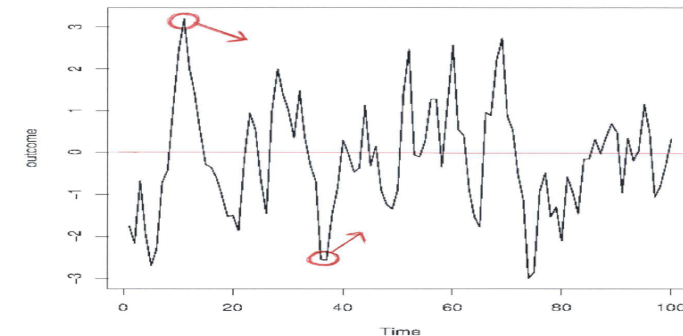
73 / 93



Hypotetisk situation

Stabilt outcome:

Hvordan udvikler folk sig, hvis de udvælges, når de udviser en ekstrem værdi?



- ▶ Placebo-behandling af folk med forhøjet blodtryk virker...
- ▶ Folk med for lav hæmoglobin har godt af at tage ekstra

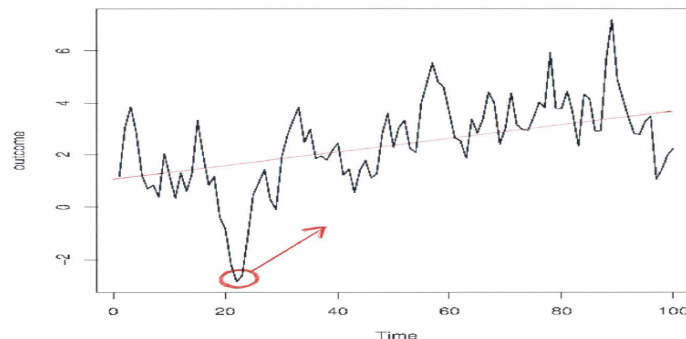
74 / 93



Hypotetisk BMD el.lign.

Outcome med positiv trend:

Hvordan udvikler folk sig, hvis de tilfældigvis udvælges, når de har et lavt niveau?



Den gruppe, der tilfældigvis ligger lavt fra starten, vil have tendens til den største stigning.

75 / 93



Hvor vigtigt er det?

Selektionseffektion er størst, når korrelationen mellem gentagne målinger (ρ) er *lille*,
men hvad betyder det?

Det betyder, at vi har at gøre med et outcome, der kan **variare hurtigt**, f.eks.

- ▶ blodtryk
- ▶ hæmoglobin?
- ▶ smerter

men ikke

- ▶ højde
- ▶ knogletæthed

76 / 93



Resultater vedr. skuldersmerter

Sammenligninger af de to grupper:

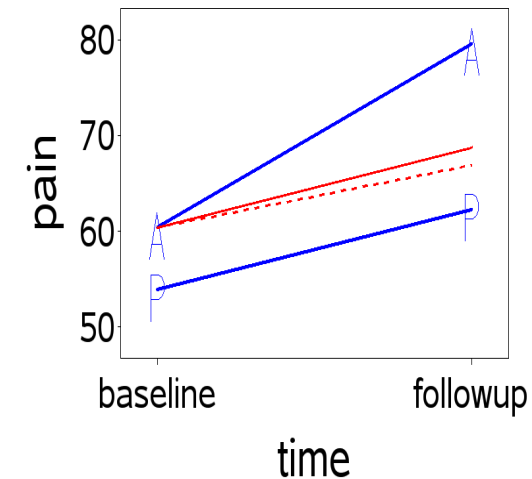
	Gennemsnitlig smertescore		Behandlinseffekt	
	placebo (n=27)	akupunktur (n=25)	differens (95% CI)	P-value
Baseline	53.9 (14.0)	60.4 (12.3)	6.5	0.09
Analyse:				
Follow-up	62.3 (17.9)	79.6 (17.1)	17.3 (7.5; 27.1)	0.0008
Ændringer*	8.4 (14.6)	19.2 (16.1)	10.8 (2.3; 19.4)	0.014
Ancova			12.7 (4.1; 21.3)	0.005

* resultater publiceret i Kleinhenz et.al. *Pain* 1999; 83:235-41.

De to baseline værdier er ikke signifikant forskellige,
men de er heller ikke helt ens....

77 / 93

Udvikling i smerter, faktisk og forventeligt



78 / 93

Udlægning af resultater, I

Baseline

- ▶ Akupunktur gruppen ligger lidt højere end placebo

Follow-up

- ▶ Vi vil forvente, at akupunktur gruppen stadig ligger det samme stykke højere end placebo gruppen efter behandlingen, selv hvis behandlingen ikke virker (**fuldt optrukket rød linie**).
- ▶ Det er derfor urimeligt blot at afgøre behandlingens effekt ved at sammenligne follow-up værdier (forskellen bliver for stor i dette tilfælde)

79 / 93

Udlægning af resultater, II

Differenser/ændringer

- ▶ Lav baseline værdi giver forventning om stor positiv ændring (**regression to the mean**)
- ▶ Derfor forventes placebo gruppen at stige mest, og akupunktur gruppen lidt mindre (**den stiplede røde linie**)
- ▶ og en direkte sammenligning af ændringer er derfor ikke rimelig (forskellen bliver for lille i dette tilfælde)

Ancova

- ▶ vurderer forskellen under hensyntagen til, at baseline forklarer *noget*, men *ikke al* forskellen ved follow-up: Vi sammenligner personer, der har fået forskellig behandling, men som **starter samme sted**.

80 / 93

APPENDIX

Programbidder svarende til diverse slides:

- ▶ Bland-Altman plot, s. 82
- ▶ T-test, s. 83
- ▶ Scatterplot. s. 84, 86, 88
- ▶ Kovariansanalyse, s. 85-87, 90
- ▶ Model med interaktion, s. 88-89, 92
- ▶ Opdelte analyser, s. 91

81 / 93



Bland-Altman plot

Slide 8

```
refetest <-
read.table("http://publicifsv.sund.ku.dk/~lts/basal/Rdata/refe_test.txt",
header=T,na.strings=c("."))
```

```
refetest$dif <- refetest$refe - refetest$test
refetest$snit <- (refetest$refe + refetest$test)/2
```

```
plot(refetest$snit, refetest$dif,
      ylab="differens: refe - test", xlab="gennemsnit",
      cex.lab=1.8, cex=1.5, col="blue")
abline(0,0,lty=2,col="red")
abline(9.89,0,lty=2,col="blue")
abline(29.29,0,lty=2,col="red")
```

82 / 93



T-test på logaritmer

Slide 21

```
> ttest=lm(su$log10vaegt ~ su$gender)
> confint(ttest)

                2.5 %      97.5 %
(Intercept)    1.85349462  1.90053311
su$genderfemale -0.08920598 -0.02837495
```

Nu skifter vi lige fortegn inden tilbagetransformationen:

```
> 10^(-confint(ttest))

                2.5 %      97.5 %
(Intercept)    0.01401217  0.01257381
su$genderfemale 1.22802153  1.06751736
```

83 / 93



Scatter plot, med regressionslinier

Slide 23 og 37

```
reg.M = lm(su$log10vaegt[su$gender=="male"] ~
           su$log10hoejde[su$gender=="male"])
reg.F = lm(su$log10vaegt[su$gender=="female"] ~
           su$log10hoejde[su$gender=="female"])
```

```
color = rep(NA, length=length(su$gender))
color[which(su$gender=="male")] = "red"
color[which(su$gender=="female")] = "blue"
```

```
mycolor = c("red", "blue")[su$gender]
mypch = c(3,1)[su$gender]
```

```
plot(su$log10hoejde, su$log10vaegt, main="Two regressions",
      ylab="log10vaegt", xlab="log10hoejde",
      pch=mypch, col=mycolor, cex.lab=1.5)
legend(0.25, 1.6, legend=c("female", "male"), pch=c(2,3),
      col=c("blue", "red"), cex=1.5)
abline(reg.M,col="red")
abline(reg.F,col="blue")
```

84 / 93



Kovariansanalyse - ANCOVA

Slide 24 og 25

```
ancova = lm(log10vaegt ~ log10hoejde + gender,
            na.action=na.exclude, data=su)
summary(ancova)
confint(ancova)
```

Herefter kan bruges

- ▶ summary giver estimer
- ▶ confint giver konfidensintervaller
- ▶ anova giver en variansanalysetabel

85 / 93



ANCOVA-plot, med parallelle linier

Slide 28

```
ny.F = data.frame(log10hoejde=seq(0.175,0.300,0.0025),gender="female")
pred.F = predict(ancova, ny.F)
ny.M = data.frame(log10hoejde=seq(0.175,0.300,0.0025),gender="male")
pred.M = predict(ancova, ny.M)

plot(su$log10hoejde, su$log10vaegt, main="Analysis of Covariance",
     ylab="log10vaegt", xlab="log10hoejde",
     pch=mypch, col=mycolor, cex.lab=1.5)
legend(0.25, 1.6, legend=c("female", "male"), pch=c(2,3),
      col=c("blue", "red"), cex=1.5)
lines(ny.F$log10hoejde, pred.F, type = "l", col="blue")
lines(ny.M$log10hoejde, pred.M, type = "l", col="red")
```

86 / 93



Modelkontrol i ANCOVA

Slide 31

```
par(mfrow=c(2,2))
plot(ancova,which=1, cex.lab=1.5)
plot(ancova,which=2, cex.lab=1.5)
plot(ancova,which=3, cex.lab=1.5)
plot(ancova,which=4, cex.lab=1.5)
dev.off()

scatter.smooth(su$log10hoejde, resid(ancova), main="Loess",
              ylab="residual", xlab="log10hoejde", cex.lab=1.5,
              lpars = list(col = "red", lwd = 3, lty = 3))
abline(0,0,col="red",lty=2)
```

Hvis man ønsker fuld kontrol over modelkontrollen, kan man arbejde videre med nye variable, såsom predict(ancova) og resid(ancova).

87 / 93



Model med to regressionslinier, tilbagetransformeret

altså **inkluderende interaktion**

Slide 38 og 45

```
interaktion = lm(log10vaegt ~ log10hoejde + gender + gender:log10hoejde,
                 na.action=na.exclude, data=su)
summary(interaktion)
confint(interaktion)

pred.F = 10^(predict(interaktion, ny.F))
pred.M = 10^(predict(interaktion, ny.M))

plot(10^(ny.F$log10hoejde), pred.F,
     ylab="predikteret vaegt",
     xlab="hoejde", cex.lab=1.5,type="n")
legend(1.8,60, legend=c("female", "male"), pch=c(2,3),
      col=c("blue", "red"), cex=1.5)
lines(10^(ny.F$log10hoejde), pred.F, type = "l", col="blue")
lines(10^(ny.M$log10hoejde), pred.M, type = "l", col="red")
points(su$hoejde[su$gender=="female"],
       10^(fitted(interaktion)[su$gender=="female"]),col="blue")
points(su$hoejde[su$gender=="male"],
       10^(fitted(interaktion)[su$gender=="male"]),col="red")
```

88 / 93



Omregning til to linier

Slide 42-43

Her indgår `log10hoejde` kun i leddet `gender:log10hoejde`, hvorved man får begge hældninger i output, og der skrives endvidere `-1` for at få begge intercepter.

```
interaktion1 = lm(log10vaegt ~ gender +
  gender:log10hoejde -1,
  na.action=na.exclude, data=su)
```

```
summary(interaktion1)
```

```
confint(interaktion1)
```

89 / 93



Kovariansanalyse

Slide 54

```
> ancova = lm(logamh ~ alder + factor(ppiller),
+   na.action=na.exclude, data=pill)
> summary(ancova)
```

Herefter kan bruges

- ▶ `summary` giver estimer
- ▶ `confint` giver konfidensintervaller
- ▶ `anova` giver en variansanalysetabel

90 / 93



Opdelte analyser

Slide 61

```
pill.0 <- subset(pill, ppiller==0,)
reg.0 = lm(logamh ~ alder,
  na.action=na.exclude, data=pill.0)

cbind(confint(reg.0)["alder",1],
  reg.0$coefficients["alder"],
  confint(reg.0)["alder",2])
```

```
pill.1 <- subset(pill, ppiller==1,)
reg.1 = lm(logamh ~ alder,
  na.action=na.exclude, data=pill.1)
```

```
cbind(confint(reg.1)["alder",1],
  reg.1$coefficients["alder"],
  confint(reg.1)["alder",2])
```

91 / 93



Test for interaktion

Slide 63

```
interaktion = lm(logamh ~ alder + factor(ppiller) +
  factor(ppiller):alder,
  na.action=na.exclude, data=pill)
summary(interaktion)
```

92 / 93



Prediktion i tilfælde af interaktion

Slide 65

Hvis vi vil prediktere niveauet for begge grupper for hver hele alder fra 20 til 40 (21 aldre), laver vi et nyt datasæt med 2×21 observationer, og predikterer outcome fra disse ud fra den anvendte model, her kaldet `interaktion`:

```
ny.ppillar=c(rep(0,21),rep(1,21))  
ny.alder=rep(20:40,2)  
ny = data.frame(ppillar=ny.ppillar, alder=ny.alder)  
  
pred = predict(interaktion, ny)  
cbind(ny.ppillar,ny.alder,pred)
```

Outcome `logamh` vil da blive predikteret for hver af disse personer, og man kan derefter fratrække for samme alder, men forskellig P-pille status og lave en figur.

