

Basal statistik

Logaritmer, Repetition, Kovariansanalyse, Interaktion. SPSS

Lene Theil Skovgaard

9. marts 2020

1 / 91



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 / 91



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:

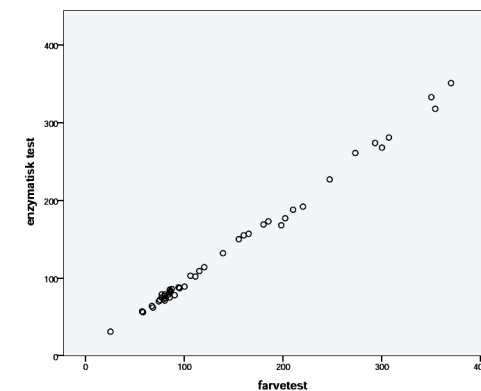
Enzymatisk test, mere specifikt for glucose.

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 / 91



Scatter plot af de to metoder



Det ser jo pænt ud...., eller?

4 / 91



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 / 91



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)

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
dif	46	9,8913	9,70276	1,43059

One-Sample Test

Test Value = 0					
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference Lower Upper
dif	6,914	45	,000	9,89130	7,0099 12,7727

Med en P-værdi ≈ 0.000 er der **stærk indikation** af forskel på de to metoder

6 / 91



Limits of agreement

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

$$9.89 \pm 2 \times 9.70 = (-9.51, 29.29)$$

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

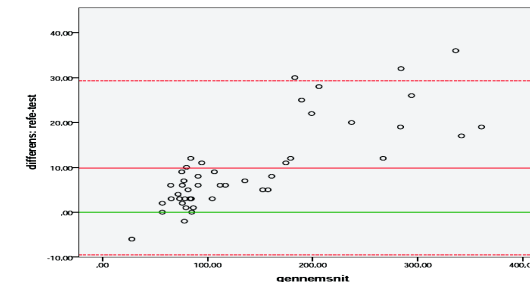
- ▶ differenserne stiger med niveauet (repræsenteret ved gennemsnittet)
- ▶ variationen stiger også med niveauet

7 / 91



Bland-Altman plot

Plot af differenser mod gennemsnit (af de to målinger på samme person), se mere s. 83



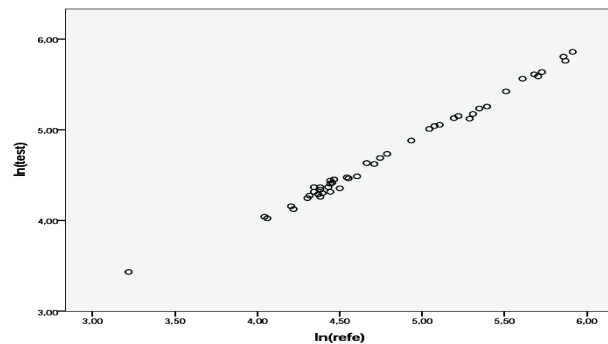
Store afvigelser ved høje målinger, dvs. **relative afvigelser**
Så er det smart at se på **logaritmer**

8 / 91



Scatter plot

efter logaritmetransformation (her “den naturlige”)



9 / 91



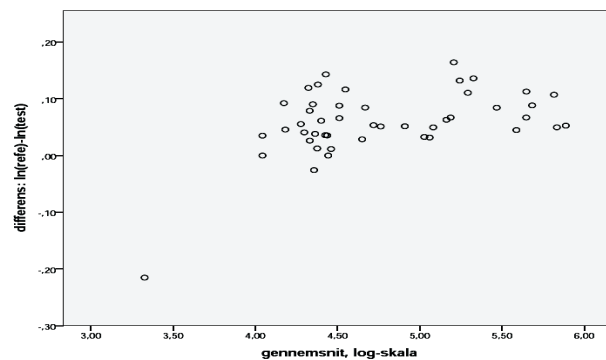
Bemærk

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

10 / 91



Bland-Altman plot for logaritmer



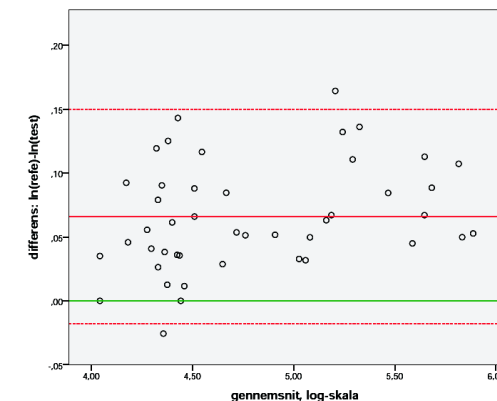
Der er en tydelig *outlier* (den mindste observation)

11 / 91



Vi udelader en outlier....

og laver igen et Bland-Altman plot



som bliver acceptabelt....

12 / 91



En slags konklusion

Summary statistics for differensen $\ln(\text{refe}) - \ln(\text{test})$:

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
differens: ln(refe)-ln(test)	45	,0657	,04195	,00625

One-Sample Test

Test Value = 0

	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Mean Difference	95% Confidence Interval of the Mean Difference
					Lower	Upper
differens: ln(refe)-ln(test)	10,510	44	,000	,06573	,0531	,0783

13 / 91



En slags konklusion, fortsat

Limits of agreement på logaritmisk skala:

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

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?

14 / 91



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

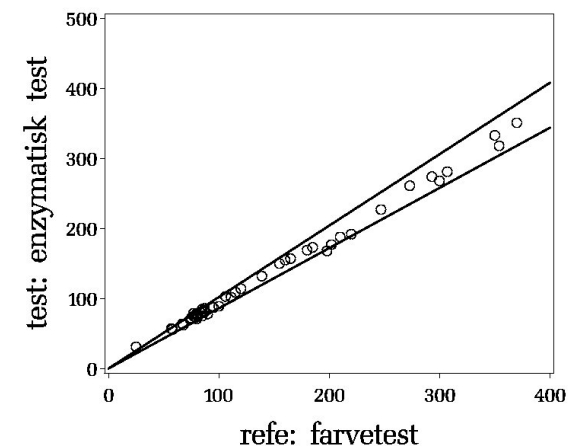
14% under og **2% over** REFE.

15 / 91



Limits of agreement

tilbagetransformeret til oprindelig skala



16 / 91



Regninger direkte på ratio-skala

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

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
ratio	45	,97	1,18	1,0689	,04512
Valid N (listwise)	45				

svarende til **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

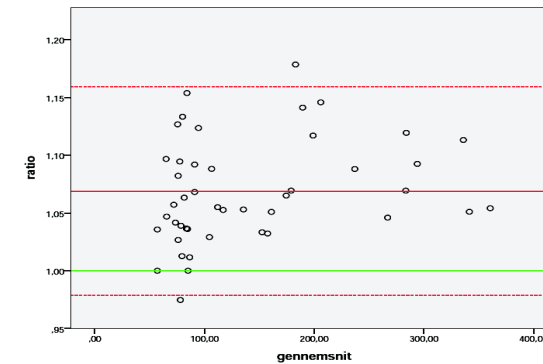
Dette er ikke altid tilfældet!!

17 / 91



Limits of agreement på ratio-skala

hvor er den med linier på??



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

18 / 91



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 **GLM**:

Begge typer kovariater i samme model

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

19 / 91



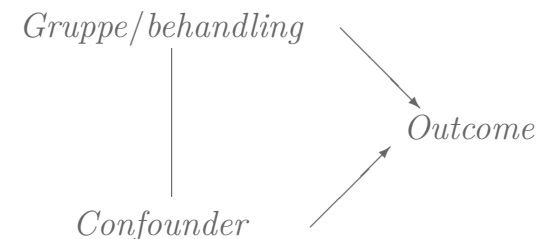
Sammenligning af to grupper

- 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"**.



20 / 91



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:

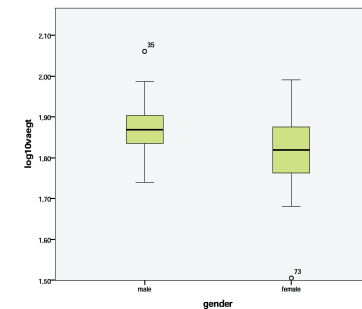
To forskellige videnskabelige spørgsmål

21 / 91



Vægtfordeling for mænd og kvinder

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

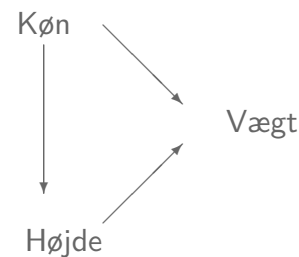
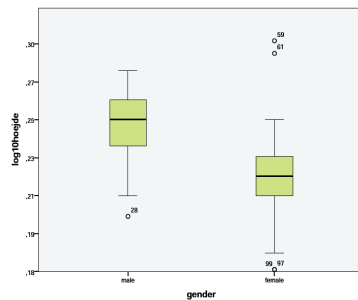


Et T-test giver $P=0.0002$, og en estimeret "fordel" til mænd på 14.5% (6.8-22.8%), se mere s. 84

22 / 91



Kan højden forklare forskellen på mænd og kvinder?



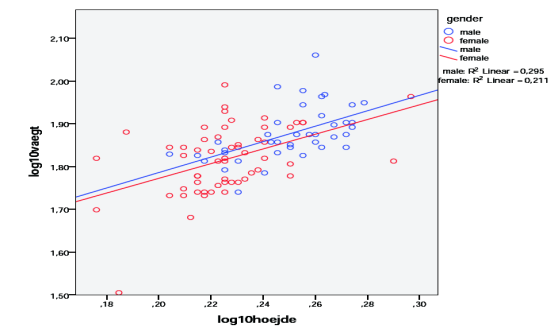
- ▶ Der er i hvert fald forskel på mænds og kvinders højde
- ▶ Men er der tillige en *ægte* kønseffekt?

23 / 91



Relation mellem højde og vægt

Igen på $\log_{10}$ -skala, og med indlagte regressionslinier for hvert køn, se mere s. 85



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

24 / 91



Kovariansanalyse i praksis

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

Kovariater: 2 stk:

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

Her benyttes Analyze/General Linear Model/Univariate, hvor vi sætter $\log_{10}\text{vægt}$ over i Dependent Variable, gender i Fixed Factor(s) og $\log_{10}\text{højde}$ i Covariate(s)

For at undgå at få interaktionen med i modellen, klikkes nu på Model, hvorefter man vælger Custom, markerer begge kovariater, skifter fra Interaction til Main Effects og klikker på pilen og Continue.

Husk også i Options at vælge Parameter Estimates

25 / 91



Output fra kovariansanalyse

Den vigtigste del:

Parameter Estimates

Dependent Variable: $\log_{10}\text{vægt}$

Parameter	B	Std. Error	t	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Intercept	1.422	.074	19.154	.000	1.275	1.570
[gender=1]	.017	.016	1.093	.277	-.014	.048
[gender=2]	0 ^a	.	.	.	.	.
$\log_{10}\text{højde}$	1.749	.326	5.357	.000	1.100	2.397

a. This parameter is set to zero because it is redundant.

Bemærk: Nu er kønseffekten helt forsvundet!

Højden kunne altså godt forklare den....måske

26 / 91



Fortolkning af kovariansanalysen

Vi fitter **to parallelle linier** (se s. 29) men både outcome og den kvantitative kovariat er logaritmeret....

Fortolkning af kønseffekt:

På $\log_{10}$ -skala ligger mænds vægt 0.017 højere end kvinders. Når vi tilbagetransformerer dette, får vi $10^{0.017} = 1.040$, svarende til, at mænd vejer ca. 4% mere end kvinder **med samme højde**.

Tilsvarende tilbagetransformerer konfidensintervallerne: $(10^{-0.014}, 10^{0.048}) = (0.968, 1.118)$, 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).

27 / 91



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 $1.1^{1.749} = 1.181$ på vægten, altså en forøgelse af vægten med 18.1%.

Tilsvarende tilbagetransformerer konfidensintervallerne:

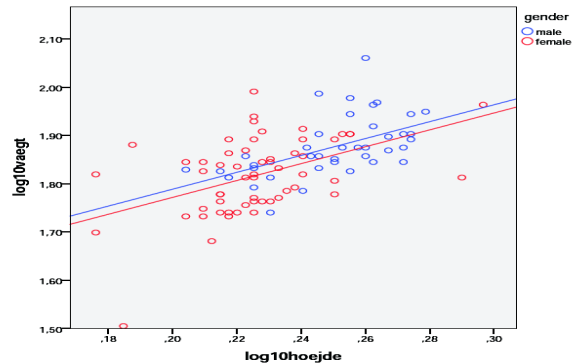
$(1.1^{1.100}, 1.1^{2.397}) = (1.111, 1.257)$, dvs. mellem 11.1% og 25.7% mere i vægt.

28 / 91



Ancova-fit (parallelle linier)

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



29 / 91



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.

30 / 91



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

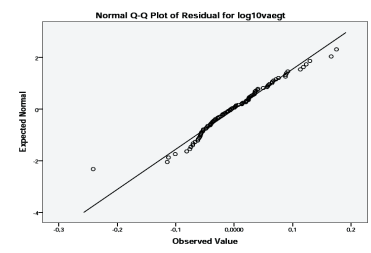
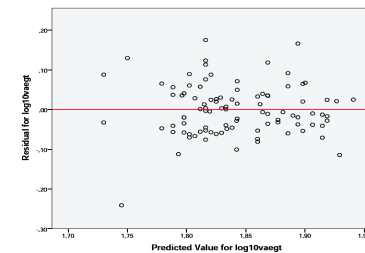
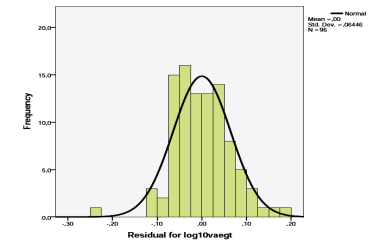
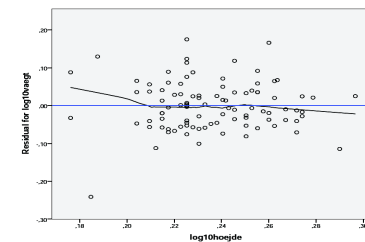
Højde er en **mellemkommende** (intermediate) variabel
 for effekten af køn på vægt

31 / 91



Husk modelkontrol:

Se mere s. 88



32 / 91



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.

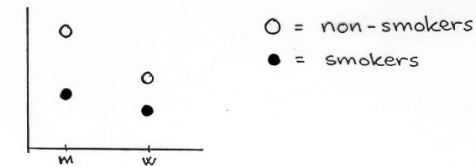
33 / 91



Vekselvirkning (**interaktion**)

Tænkt eksempel:

- ▶ To inddelingskriterier: køn og rygestatus
- ▶ Outcome: FEV₁



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

34 / 91



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

35 / 91



Eksempel: Rygnings effekt på fødselsvægt

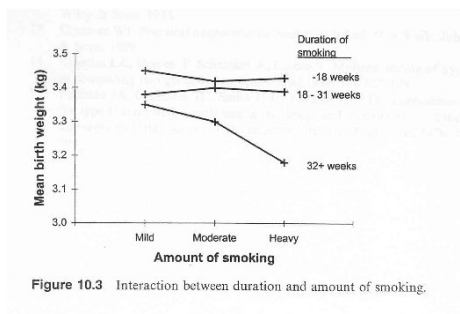
Table 10.1 Average Birth Weight of Children Born to Women with Different Amount and Duration of Smoking^a

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.

36 / 91





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**....

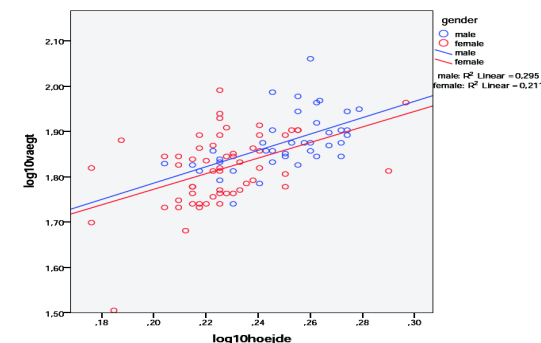
37 / 91



38 / 91

Model med interaktion

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



Model med interaktion i praksis

Testvenlig fremgangsmåde:

Start som på s. 25, men i Model markeres derefter begge samtidig, og der skiftes til Interaction og klikkes på pilen, så interaktionen kommer med over i modellen.

Mere estimat-venlig fremgangsmåde:

Udføres som ovenfor, men i Model gøres følgende

- ▶ sæt log10højde nederst (giver de to hældninger separat)
- ▶ fjern evt. fluebenet ved Include Intercept in model (giver de to intercepter separat)

39 / 91



Output fra testvenlig model

Se s. 39, øverste mulighed:

Parameter Estimates						
Dependent Variable: log10vægt						
Parameter	B	Std. Error	t	Sig.	95% Confidence Interval	
Intercept	1,429	,092	15,537	,000	1,246	1,611
[gender=1]	-,004	,168	-,022	,982	-,337	,330
[gender=2]	0 ^a	.	.	.	.	.
log10højde	1,719	,405	4,242	,000	,914	2,524
[gender=1] * log10højde	-,086	,691	-,125	,901	-1,285	1,458
[gender=2] * log10højde	0 ^a	.	.	.	.	.

a. This parameter is set to zero because it is redundant.

Interaktionsparameteren angiver
forskellen på hældninger i de to grupper



40 / 91

Fortolkning af interaktionsmodel

Vi har fittet to *vilkårlige linier* (se s. 38) 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. 38 varierer en smule

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

41 / 91



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. 28, idet koefficienten til $\log_{10}\text{hoejde}$ angiver effekten for referencegruppen for gender (altså mænd), medens koefficienten svarende til $\log_{10}\text{hoejde}*\text{gender female}$ angiver *forskellen* mellem effekten for kvinder og effekten for mænd.

Vi omregner disse estimater på s. 43, men se dog også estimationsvenlig fremgangsmåde s. 39 (nederst) og output fra denne s. 44

42 / 91



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 mere nederst s. 39 og output s. 44

43 / 91



Output fra estimationsvenlig fremgangsmåde

Se s. 39 (nederst)

Parameter Estimates

Dependent Variable: $\log_{10}\text{vægt}$

Parameter	B	Std. Error	t	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
[gender=1]	1,425	,140	10,152	,000	1,146	1,704
[gender=2]	1,429	,092	15,537	,000	1,246	1,611
[gender=1] * $\log_{10}\text{hoejde}$	1,805	,559	3,228	,002	,694	2,916
[gender=2] * $\log_{10}\text{hoejde}$	1,719	,405	4,242	,000	,914	2,524
$\log_{10}\text{hoejde}$	0 ^a	.	.	.	.	.

a. This parameter is set to zero because it is redundant.

44 / 91



* Fortolkning af interaktionsmodel, fortsat

Vi har nu direkte estimater for de to linier (se s. 43-44) og skal tilbagetransformere til relationen

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

for hvert køn for sig. Vi finder:

► **For mænd:**

$$\text{vægt} = 10^{1.4251} \times \text{højde}^{1.8051} = 26.61 \times \text{højde}^{1.8051}$$

► **For kvinder:**

$$\text{vægt} = 10^{1.4288} \times \text{højde}^{1.7189} = 26.84 \times \text{højde}^{1.7189}$$

Disse *kurver* (ikke linier) er vist på s. 46.

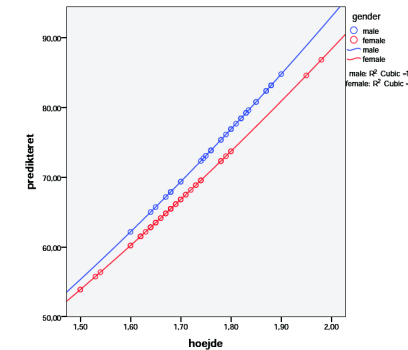
Bemærk dog, at de ikke afviger ret meget fra linier....

45 / 91



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...)



46 / 91



Bemærk

Da der absolut ikke kunne spores nogen interaktion i dette eksempel, går vi tilbage til den additive model, *kovariansanalysen*, fra s. 25-26. 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. 30).

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

Sådan et eksempel skal vi se nu

47 / 91



Nyt eksempel: Hormoner hos P-pille brugere

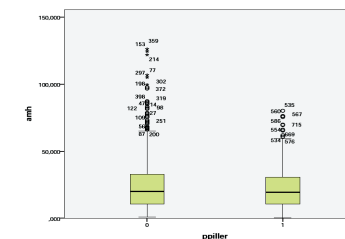
Anti Müllersk Hormon (**AMH**) er et hormon, som dannes i de små tidlige ægblærer (follikler) i æggestokkene.

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



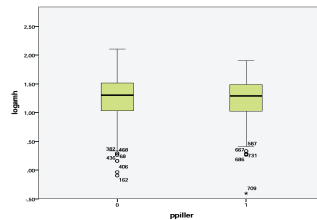
48 / 91



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 logaritmetransformerede ($\log_{10}$) AMH-værdier



Heller ikke på denne skala er der perfekt symmetri

49 / 91



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).



50 / 91

Sammenligning af grupper - på logaritmisk skala

Uparret T-test giver først lidt summary statistics og test for identiske varianser:

Group Statistics

	ppiller	N	Mean	Std. Deviation	Std. Error Mean
logamh	0	504	1,2646	,36547	,01628
	1	228	1,2394	,36255	,02401

Independent Samples Test

			Levene's Test for Equality of Variances		t-test for Equality of Means	
			F	Sig.	t	df
logamh	Equal variances assumed		,035	,851	,866	730
	Equal variances not assumed				,869	441,560

og derefter selve testet, se næste side:

51 / 91



Sammenligning af grupper - på logaritmisk skala, II

Independent Samples Test					
		t-test for Equality of Means			
		Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence ... Lower
logamh	Equal variances assumed	,387	,02520	,02910	-,03193
	Equal variances not assumed	,386	,02520	,02901	-,03182

Independent Samples Test			
		t-test for Equality of Means	
		95% Confidence Interval of the ...	
		Upper	
logamh	Equal variances assumed	,08232	
	Equal variances not assumed	,08221	

52 / 91



Kommentarer til output

- Der er ikke tegn på forskellige varianser i de to grupper ($P=0.851$ for Levenes test)
- Det er derfor ligegyldigt, hvilket T-test, vi vælger at se på (de giver i øvrigt også næsten præcis det samme)
- P-værdien er 0.39, altså **ingen evidens for forskel** i AMH blandt brugere og ikke-brugere af P-piller

53 / 91



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

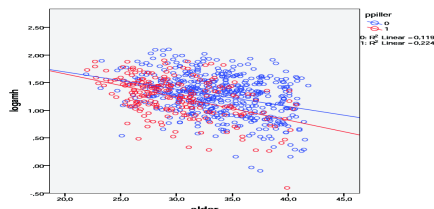
Kan der være en confounder? Alder?

54 / 91



Er alder en confounder?

- Har den en effekt på hormon-niveauet?**
Det skulle man tro, ud fra vores viden om, hvad hormonet afspejler, se figuren nedenfor
- Er der forskellig alder i de to grupper?**
Formentlig er P-pille brugerne yngre end ikke-P-pille brugerne, se tabellen nedenfor



Descriptive Statistics

ppiller		N	Minimum	Maximum	Mean	Std. Deviation
0	alder	504	21,9	41,8	33,724	4,0227
	Valid N (listwise)	504				
1	alder	228	22,0	40,7	30,016	4,0996
	Valid N (listwise)	228				

Alder er en confounder!

55 / 91



Kovariansanalyse - justering for aldersforskel

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

Her benyttes Analyze/General Linear Model/Univariate, se tilsvarende analyse for det forrige eksempel, s. 25

Parameter Estimates

Dependent Variable: logamh

Parameter	B	Std. Error	t	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Intercept	2,280	,095	23,983	,000	2,094	2,467
[ppiller=0]	,154	,029	5,267	,000	,096	,211
[ppiller=1]	0 ^a	.	.	.	.	.
alder	-,035	,003	-11,263	,000	-,041	-,029

a. This parameter is set to zero because it is redundant.

56 / 91



Kommentarer til kovariansanalyse

Kvinder, der tager P-piller estimeres til et 0.154 lavere niveau end kvinder, der ikke tager P-piller, med CI=(0.096, 0.211).

Når vi tilbagetransformerer til selve koncentrationsskalaen, giver det en estimeret ratio $10^{-0.154} = 0.70$, med konfidensinterval $(10^{-0.211}, 10^{-0.096}) = (0.62, 0.80)$

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 ($P = 0.000$)

57 / 91



Kommentarer til de to analyser

Vi sammenfatter i en tabel:

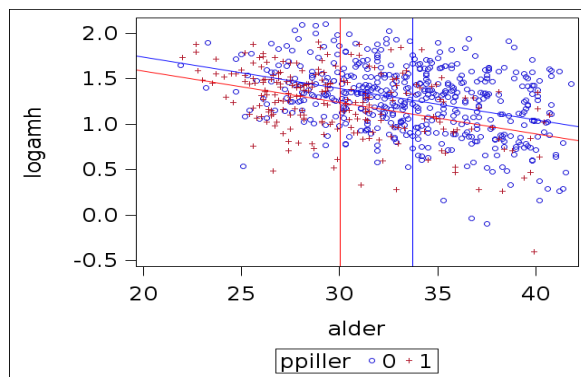
Model	Reduktion i AMH: P-pille brugere vs. ikke P-pille brugere	P-værdi
T-test (s. 51-54)	-6% (-17%, +8%)	0.39
Kovariansanalyse (s. 56-57)	-30% (-38%, -20%)	<0.0001

58 / 91



Kovariansanalyse modellen grafisk

med fortolkning næste side



Men kan der tænkes at være interaktion?

59 / 91



Fortolkning af figur s. 59

- Kovariansanalysen svarer til de to parallelle regressionslinier
- Den lodrette afstand imellem disse svarer til effekten af P-piller i kovariansanalysen s. 56-57
- De to lodrette linier er anbragt i aldersgennemsnittene:

Rød P-pille brugere

Blå Ikke-P-pille brugere

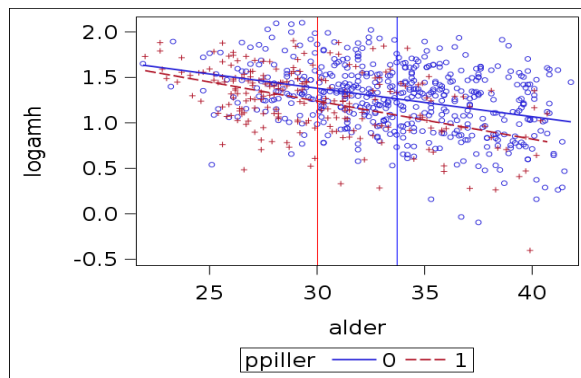
Disse linier ses at skære regressionslinierne i **nogenlunde** samme højde (dvs. værdi af logAMH), og ca. svarende til det insignifikante T-test s. 51-54

60 / 91



Figur med *frie* regressionslinier

svarende til *mulig interaktion*



Er de to linier ikke-parallelle?

61 / 91



Fortolkning af figur s. 61

- 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 logAMH, svarende til det insignifikante T-test s. 51-54

Er effekten af alder forskellig i de to grupper?

62 / 91



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

Coefficients ^a						
ppiller	Model		Unstandardized Coefficients		Standardized Coefficients	
			B	Std. Error	Beta	t
0	1	(Constant)	2,321	,129		17,952
		alder	-,031	,004	-,345	-8,228
1	1	(Constant)	2,496	,157		15,901
		alder	-,042	,005	-,473	-8,079

Coefficients ^a				
95,0% Confidence Interval for B				
ppiller	Model		Lower Bound	Upper Bound
0	1	(Constant)	2,067	2,575
		alder	-,039	-,024
1	1	(Constant)	2,187	2,805
		alder	-,052	-,032

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

63 / 91



Interaktion mellem P-piller og alder?

De to analyser s. 63 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	-0.155 (-0.196, -0.120)	0.70 (0.64, 0.76)
P-pille brugere	-0.208 (-0.260, -0.160)	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).

64 / 91



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, altså **inkluderende interaktion**

Se mere vedr. tidligere eksempel s. 89

Parameter Estimates

Dependent Variable: logamh

Parameter	B	Std. Error	t	Sig.	95% Confidence Interval	
Intercept	2,496	,165	15,131	,000	2,172	2,820
[ppiller=0]	-,175	,208	-,843	,400	-,584	,233
[ppiller=1]	0 ^a	.	.	.	.	.
alder	-,042	,005	-7,688	,000	-,053	-,031
[ppiller=0] * alder	,011	,007	1,598	,110	-,002	,024
[ppiller=1] * alder	0 ^a	.	.	.	.	.

a. This parameter is set to zero because it is redundant.

65 / 91



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 0.024, svarende til, at P-pille brugere på 5 år havde et ekstra tab på 24% (fordi $10^{-5*0.024} = 0.76$)

66 / 91



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, ved successivt at ændre Y-aksens placering ved at flytte nulpunktet, så effekten kvantificeres som forskellen i intercepter

67 / 91



Kvantificering af specifikke forskelle

f.eks. for alder 30 år, hvor vi definerer alder30=alder-30 og benytter denne i stedet for alder i regressionsmodellen fra s. 65.

Herved fås outputtet:

Parameter Estimates

Dependent Variable: logamh

Parameter	B	Std. Error	t	Sig.	95% Confidence Interval	
Intercept	1,240	,022	55,671	,000	1,196	1,284
[ppiller=0]	,141	,030	4,670	,000	,082	,200
[ppiller=1]	0 ^a	.	.	.	.	.
alder30	-,042	,005	-7,688	,000	-,053	-,031
[ppiller=0] * alder30	,011	,007	1,598	,110	-,002	,024
[ppiller=1] * alder30	0 ^a	.	.	.	.	.

a. This parameter is set to zero because it is redundant.

og forskellen på 0.141 skal så tilbagetransformeres....

68 / 91



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**

69 / 91



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....

70 / 91



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

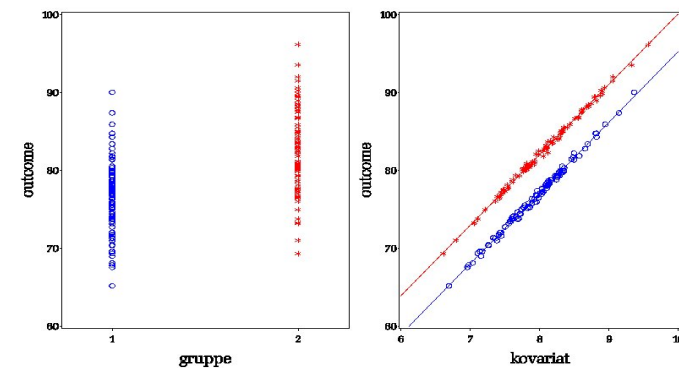
Men pas på med at gå for meget på **fisketur!!**

og husk at **fortolkningen afhænger af modellen**

71 / 91



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)

72 / 91



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

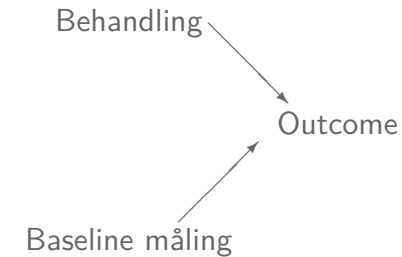
Høje scorere er gode

73 / 91



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?

74 / 91



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**

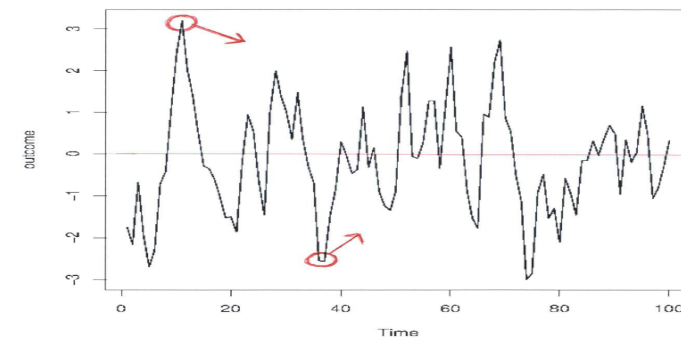
75 / 91



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

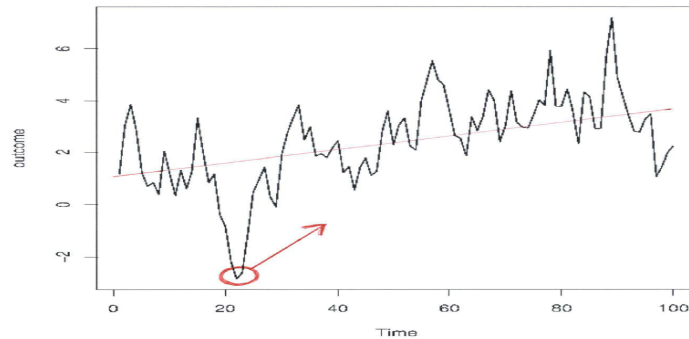
76 / 91



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.

77 / 91



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 variere hurtigt, f.eks.

- ▶ blodtryk
- ▶ hæmoglobin?
- ▶ smerter

men ikke

- ▶ højde
- ▶ knogletæthed

78 / 91



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

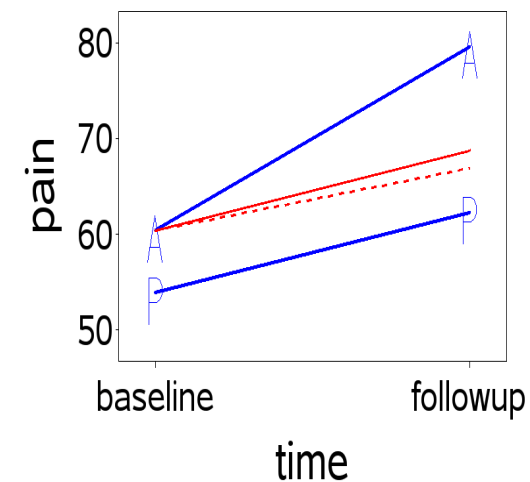
* resultater publiceret i Kleinhenz et.al. *Pain* 1999; 83:235-41.

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

79 / 91



Udvikling i smerter, faktisk og forventeligt



80 / 91



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)

81 / 91



Udlægning af resultater, II

Differenser/ændringer

- ▶ Lav baseline værdi giver forventning om stor positiv ændring (**regression to the mean**)
- ▶ Derfor forventes placebogruppen at stige mest, og akupunkturgruppen 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**.

82 / 91



Bland-Altman plot

Slide 8

Benyt Graphs/Chart Builder og vælg Scatter (det simple længst til venstre), og dobbeltklik det op i det store felt. Sæt average over på X-aksen og dif over på Y-aksen.

For at lægge en vandret linie i 0 oveni, dobbeltklikker man på figuren, klikker på Properties-ikonet og vælger Add a reference line from Equation, hvorefter man i Custom Equation skriver $y=0$ og herefter Apply.

På tilsvarende vis kan linier svarende til andre specifikke værdier lægges ind.

Evt kan man vælge liniens farve og stipling under Lines: klik på farven og Apply/Close.

83 / 91



T-test på logaritmer

Slide 22

Først skal vi logaritmetransformere vægten. Dette gøres under Transform/Compute, hvor man som Target Variable sætter det nye variabelnavn, her $\log_{10}(\text{vægt})$, og i feltet Numeric Expression skriver $\text{LG10}(\text{vægt})$

T-testet udføres med Analyze/Compare Means/Independent Samples T-test, hvor vi sætter $\log_{10}(\text{vægt})$ over i Test Variable(s) og gender i Grouping Variable.

Under Define Groups vælges Group1 til male og Group2 til female.

84 / 91



Scatter plot, med regressionslinier

Slide 24

Først laves plottet ved at gå ind i Graph/Chart Builder, vælge Scatter plot (nr. 2 fra venstre), og i den fremkomne boks trækker man log10vægt over på Y-aksen, log10hoejde over på X-aksen og gender over i Set Color

For at tegne linierne dobbeltklikker man efterfølgende på grafen og klikker på ikonet Add Fit Line at Subgroups og derefter i Properties-boksen afkrydse Linear og klikke Apply.

Man kan endvidere vælge, om man vil have liniens ligning skrevet på linien eller ej (flueben ved Attach label to line kan fjernes)

Farven på linien kan vælges under Lines: klik på farven og Apply/Close

85 / 91



Kovariansanalyse - ANCOVA

Slide 25, 29, og 56

Her benyttes Analyze/General Linear Model/Univariate, hvor vi sætter log10vægt over i Dependent Variable, køn i Fixed Factor(s) og log10hoejde i Covariate(s)

For at undgå at få interaktionen med i modellen, klikkes nu på Model, hvorefter man vælger Custom, markerer begge kovariater, skifter fra Interaction til Main Effects og klikker på pilen og Continue.

Husk også i Options at vælge Parameter Estimates

86 / 91



ANCOVA-plot, med parallelle linier

Slide 29

Her må man først lave scatterplottet, med symbol for de enkelte grupper, se s. 85.

Herefter dobbeltklikker man på scatterplottet og vælger Add a reference line from Equation, hvorefter man skriver ligningen for den første linie, f.eks. svarende til referencen male: $y = 1.749x + 1.439$ og trykker Apply. Herefter dobbeltklikker man igen for at tilføje den anden linie, nu for female: $y = 1.749x + 1.422$.

Der er desværre ingen automatik til at lægge de fittede linier ind for denne model....

87 / 91



Modelkontrol i ANCOVA

Slide 32

I analysen fra s. 86 benyttes Options og under Display afkrydses Residual Plot, som dog giver nogle meget grimme figurer.

I stedet kan man gemme såvel predikterede værdier og residualer ved at benytte Save og vælge Predicted Values, samt Residuals: Studentized. De kommer til at hedde hhv. PRE_1 og SRE_1 i datasættet.

Herefter går man i Graph/Chart Builder/Scatter og vælger den ønskede graf.

Linier indlægges som beskrevet s. 83.

88 / 91



Model med interaktion

dsv 2 separate regressionslinier

Slide 38-40 og s. 65

Start som på s. 86, men i Model markeres begge, og der skiftes til Interaction og klikkes på pilen, så interaktionen kommer med over i modellen.

Ligesådan for eksemplet om P-piller, naturligvis med andre variable.

89 / 91



Omregning til to linier

Slide 43-44

Udføres som s. 89, men i Model gøres følgende

- ▶ fjern log10hoejde (behold de to øvrige), eller sæt hoejde nederst, efter interaktionen
- ▶ fjern fluebenet ved Include Intercept in model

90 / 91



Opdelte analyser

Slide 63

For at opdele analyserne, går vi først ind i Data/Split File og vælger Compare groups.

Herefter sættes ppiller over i groups based on og man klikker OK.

Derefter benyttes menuen Analyze/Regression/Linear, og i boksen sættes logamh som Dependent og alder som Independent(s)
(et uheldigt navn til forklarende variable...)

Man skal efterfølgende huske at gå ind i Statistics og afkrydse Parameter Estimates og Confidence intervals

91 / 91

