Faculty of Health Sciences

Basal Statistik

Sammenligning af grupper, Variansanalyse

Lene Theil Skovgaard

7. september 2020

Sammenligning af grupper

- Sammenligning af to grupper: T-test
- Dimensionering af undersøgelser
- Sammenligning af flere end to grupper: Ensidet variansanalyse
- Tosidet variansanalyse
- Appendix med kode

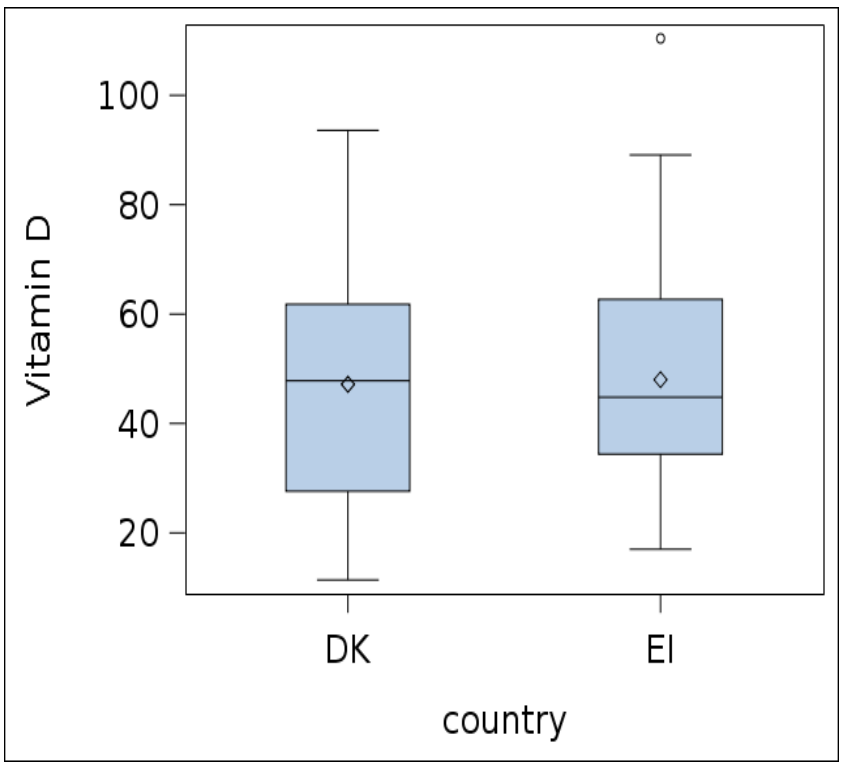
E-mail: `ltsk@sund.ku.dk`

$*$: Siden er lidt teknisk

$\triangle$: Siden gennemgås (nok) ikke

Vitamin D eksemplet

Er der forskel på vitamin D status for kvinder i Danmark og Irland?
Hvis der er en forskel på 5 nmol/l, vil det være af interesse.



Indlæsning, se s. 98

```
proc sgplot data=women;
    where country in (1,4);
    vbox vitd / category=country;
run;
```

Praktisk håndtering af data

Der er tale om 94 datalinier, en for hver kvinde,
men **to variable** for hver kvinde:

- Land (DK, EI), repræsenteret ved `country` (1,4)
- Vitamin D status, `vitd` (Serum 25(OH)D, nmol/l)

Summary statistics, opdelt efter land

```
proc means maxdec=2 data=women; where country in (1,4);
class country;
var vitd;
run;
```

country	N Obs	N	Mean	Std Dev	Minimum	Maximum
DK	53	53	47.17	22.78	11.40	93.60
EI	41	41	48.01	20.22	17.00	110.40

Model for uparret sammenligning

Antagelser:

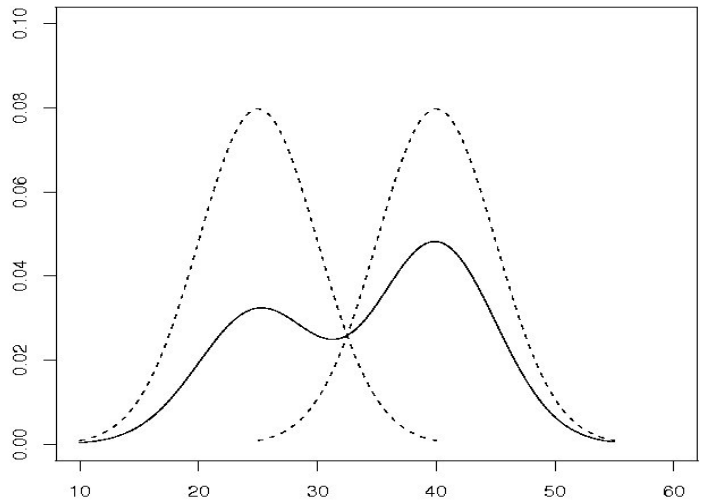
- Alle observationerne er uafhængige
  – kun 1 måling pr. person
  – personerne har ikke noget med hinanden at gøre
- Der er samme spredning(varians) i de to grupper
  – bør checkes/sandsynliggøres
- Observationerne følger en normalfordeling i hver gruppe, med hver deres middelværdi, μ_1 hhv. μ_2

– og det er disse 2 middelværdier (μ_1 og μ_2),
vi gerne vil sammenligne

Normalfordelingsmodel for to grupper

Bemærk:

Selv hvis hver gruppe er *eksakt normalfordelt*:



er det "totalt set" slet ikke en normalfordeling!!

men en blanding af to

Uparret t-test i praksis

til sammenligning af to gennemsnit

```
proc ttest data=women;
     where country in (1,4);
class country;
var vitd;
run;
```

where-sætningen udvælger de to lande, vi vil se på

Typisk output fra et uparret t-test

The TTEST Procedure

Variable: vitd (Vitamin D)

country	N	Mean	Std Dev	Std Err	Minimum	Maximum
DK	53	47.1660	22.7829	3.1295	11.4000	93.6000
EI	41	48.0073	20.2221	3.1582	17.0000	110.4
Diff (1-2)		-0.8413	21.7067	4.5147		

country	Method	Mean	95% CL Mean		Std Dev
DK		47.1660	40.8863	53.4458	22.7829
EI		48.0073	41.6244	54.3902	20.2221
Diff (1-2)	Pooled	-0.8413	-9.8078	8.1253	21.7067
Diff (1-2)	Satterthwaite	-0.8413	-9.6739	7.9913	

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	92	-0.19	0.8526
Satterthwaite	Unequal	90.213	-0.19	0.8503

Equality of Variances

Method	Num DF	Den DF	F Value	Pr > F
Folded F	52	40	1.27	0.4357

Kommentarer til output

- Først nogle summary statistics for hvert land
- Derefter konfidensintervaller (CI) for de to middelværdi-estimater,
  samt for deres differens
  i 2 forskellige udgaver, afhængig af, om spredningerne(varianserne) kan antages at være ens eller ej.
- Herefter 2 forskellige udgaver af T-testet, igen afhængig af, om spredningerne kan antages at være ens eller ej.
  Under alle omstændigheder er $P = 0.85$, dvs. vi kan ikke afvise, at middelværdierne er ens.
- Til sidst et test for ens varianser (spredninger), som ikke forkastes, idet P=0.44

∗Hvad er det, der udregnes?

Estimat for forskel i middelværdier:

$$\hat{\mu}_1 - \hat{\mu}_2 = \bar{Y}_1 - \bar{Y}_2 = 48.01 - 47.17 = 0.84 \quad \text{nmol/l}$$

med tilhørende usikkerhed

$$St.Err.(\bar{Y}_1 - \bar{Y}_2) = \text{pooled SD}\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} = 4.51$$

og teststørrelse

$$T = \frac{\bar{Y}_1 - \bar{Y}_2}{\text{St.Err.}(\bar{Y}_1 - \bar{Y}_2)} = -\frac{0.8413}{4.5147} = -0.19$$

som under H_0 er t-fordelt med 92 frihedsgrader

Hvad betyder teststørrelsens fordeling? - under H_0

Vi forestiller os mange ens undersøgelser af stikprøver på 94 kvinder fra samme land (svarende til H_0: ingen landeforskel):

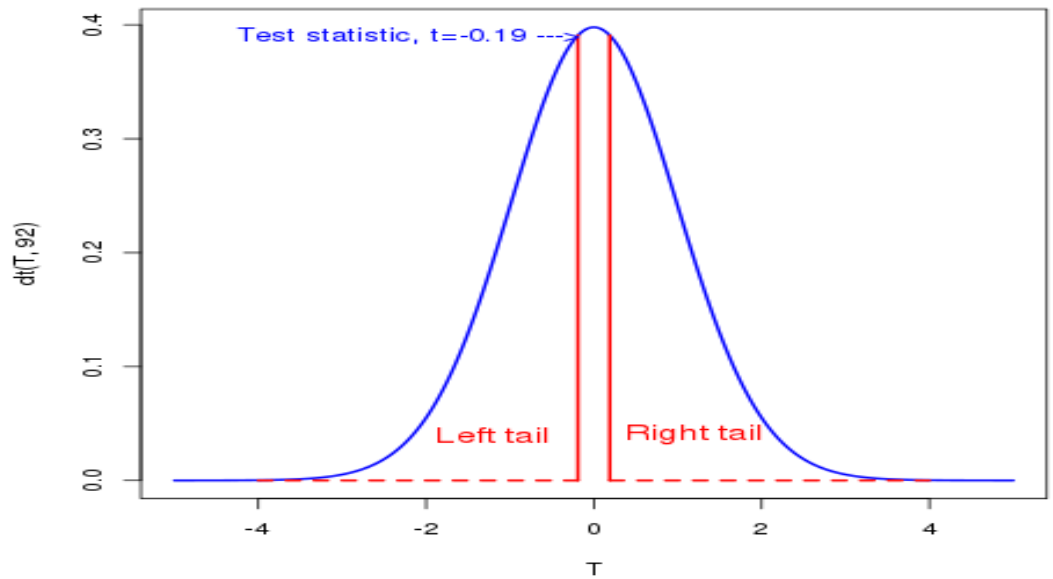
1. Fordel tilfældigt 53 i en gruppe, 41 i en anden, $\Longrightarrow t_1$
2. Fordel tilfældigt 53 i en gruppe, 41 i en anden, $\Longrightarrow t_2$
3. Fordel tilfældigt 53 i en gruppe, 41 i en anden, $\Longrightarrow t_3$
   osv. osv.

Fordeling af t'erne? ... kan udregnes til t(92)...

Vores faktiske T sammenlignes nu med denne fordeling,
Passer den pænt?

Fortolkning af P-værdi

t-fordelingen (Student fordelingen) med 92 frihedsgrader:



- Teststørrelsen -0.19 ses at ligge meget centralt i fordelingen
- Arealet af området med "værre teststørrelser" kaldes halesandsynligheden
- – og det er også P-værdien, her 0.85

Konklusion

- Der ser ikke ud til at være forskel på vitamin D status i de to lande
  - Vi fandt nemlig en teststørrelse, der passer pænt med dem, vi ville finde, hvis vi havde valgt kvinder fra *samme* land, altså hvor forskellene *udelukkende* var tilfældige
- Men kan vi nu være *sikre på*, at der ikke er nogen forskel?
  - Nej, konfidensintervallet siger, at forskellen mellem de to lande med 95% sandsynlighed ligger mellem 8.13 i Danmarks favør og 9.81 i Irlands favør.
- Vi kan altså ikke udelukke en forskel på 5 nmol/l, som var det, vi ønskede at finde ud af....
- Vi skal måske prøve en større undersøgelse...

Signifikansbegrebet

Statistisk signifikans afhænger af:

- sand forskel
- antal observationer
- den *tilfældige variation*, dvs. den biologiske variation
- signifikansniveau

"Videnskabelig" signifikans afhænger af:

- størrelsen af den påviste forskel

Tænkt eksempel

To aktive behandlinger: A og B, vs. Placebo: P

Resultater fra to trials:

1. trial: A signifikant bedre end P (n=100)
2. trial: B *ikke* signifikant bedre end P (n=50)

Konklusion:
A er bedre end B ???

Nej, ikke nødvendigvis.

Hvis der ikke er signifikans

kan det skyldes

- At der ikke *er* en forskel
- At forskellen er så lille, at den er vanskelig at opdage
- At variationen er så stor, at en evt. forskel drukner
- At materialet er for lille til at kunne påvise nogensomhelst forskel af interesse.

Kan vi så konkludere, at der ikke er forskel?
Nej!!, ikke nødvendigvis

Se på konfidensintervallet for forskellen

Vurdering af konfidensintervaller

Konfidensintervaller for hver behandling for sig

- siger noget om den mulige effekt af *denne* behandling
- kan kun *somme tider* benyttes til at vurdere forskel på behandlinger
  - Hvis konfidensintervallerne *ikke* overlapper, er der signifikant forskel
  - Hvis estimatet for den ene behandlingseffekt er indeholdt i konfidensintervallet for den anden, så er der *ikke* signifikant forskel
  - At konfidensintervallerne blot overlapper, kan ikke benyttes til at argumentere for *ingen signifikans*,

Se på konfidensintervallet for forskellen

Risiko for fejlkonklusioner

Signifikansniveauet α (sædvanligvis 0.05) angiver den risiko, vi er villige til at løbe for at *forkaste en sand nulhypotese*,
også betegnet som **fejl af type I**.

	accept	forkast
H_0 sand	1-α	α fejl af type I
H_0 falsk	β fejl af type II	1-β styrke

1-β kaldes **styrken**, den angiver
sandsynligheden for at forkaste en *falsk* hypotese.

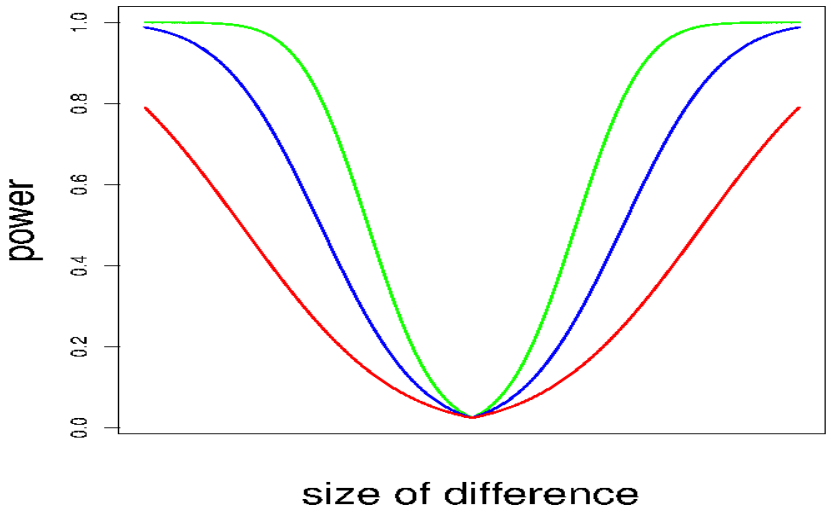
Styrke

Men hvad betyder "H_0 falsk"? Hvor store forskelle er der?

På figuren er n lille, n større, n størst

Styrkefunktion:

'Hvis forskellen er xx, hvad er så styrken, dvs. sandsynligheden for at opdage denne forskel – på 5% niveau'?



Styrken er en funktion af forskellen - og af antallet af observationer

Vigtigt

- Styrken udregnes for at **dimensionere** en undersøgelse
- Når resultaterne er i hus, præsenteres i stedet **konfidensintervaller**
- Post-hoc styrkebetragtninger giver kun mening, hvis man skal i gang med en ny undersøgelse

  - som f.eks. for vitamin D, fordi resultatet var inkonklusivt

Dimensionering af undersøgelser

Hvor mange patienter skal vi medtage?

Dette afhænger naturligvis af datas forventede beskaffenhed, samt af, hvad man ønsker at opnå:

- Hvilken forskel i respons er vi interesserede i at opdage?
  Fastsæt MIREDIF (mindste relevante differens)
  – her skal der tænkes - ikke regnes
- Med hvilken sandsynlighed (styrke = power)?
- På hvilket signifikansniveau?
- Hvor stor er spredningen (den biologiske variation)?

Hvordan skaffer man de nødvendige oplysninger?

- Klinisk relevant forskel (MIREDIF)
  Dette er noget, man fastsætter ud fra teoretiske/praktiske overvejelser om, hvilken forskel, der skønnes at være stor nok til at være vigtig.
  Det er altså ikke noget, man skal *regne* sig frem til!
  Her var vi interesseret i at kunne påvise forskellen,
  hvis den oversteg 5 nmol/l
- Styrke: bør være stor, mindst 80%
- Signifikansniveau: Sædvanligvis 5%
  - I tilfælde af mange sammenligninger, eller hvis det kan have fatale konsekvenser at forkaste en sand hypotese, bør det sættes lavere, f.eks. 1%
- Spredning:
  Dette er det sværeste, se næste side

Fornuftigt gæt på spredning

kan være ganske vanskeligt og kræver sædvanligvis et pilot-studie.

Her har vi yderligere oplysninger fra T-testet,
nemlig usikkerheden på spredningsestimatet
(ikke medtaget s. 8, men se kode s. 7):

```
The TTEST Procedure
Variable:  vitd  (Vitamin D)

country       Method              95% CL Std Dev
DK                                19.1229  28.1887
EI                                16.6026  25.8743
Diff (1-2)    Pooled              18.9725  25.3688
Diff (1-2)    Satterthwaite
```

For at være på den sikre side, bør vi vælge et spredningsskøn på 25 eller 28, hvorimod 20-22 let kan vise sig at være for lavt

Dimensionering i praksis

Der findes mange web-sider til at gøre dette
men i SAS benyttes

```
proc power;
    twosamplemeans test=diff
    meandiff=5
    stddev=20,28
    npergroup=.
    power=0.8,0.9;
run;
```

Bemærk, at man kan foretage adskillige dimensioneringer på samme tid

Output fra dimensionering

The POWER Procedure
Two-sample t Test for Mean Difference

Fixed Scenario Elements

Distribution	Normal
Method	Exact
Mean Difference	5
Alpha	0.05

Computed N Per Group

Index	Std Dev	Nominal Power	Actual Power	N Per Group
1	20	0.8	0.801	253
2	20	0.9	0.901	338
3	28	0.8	0.801	494
4	28	0.9	0.900	660

Vi skal altså op på ca. 500 personer fra hvert land for at kunne detektere en forskel af den relevante størrelse.

Vigtigheden af antagelserne for uparret sammenligning

- Uafhængighed: meget vigtig
  Hvis enkelte målinger (en lille procentdel) viser sig at stamme fra samme person eller nært beslægtede individer, gør det næppe nogen stor skade, men *hvis designet er parret*, eller der konsekvent er flere målinger på hvert individ, kan det have dramatiske konsekvenser
  Kodeordet her er gentagne målinger = repeated measurements
- Ens spredninger: relativt vigtig, specielt hvis grupperne ikke har nogenlunde samme størrelse
- Normalfordelingen: ikke så vigtig, specielt hvis grupperne er store, eller har nogenlunde samme størrelse
  (afvigelse mindre end en faktor 1.5)

Hvis antagelserne (slet) ikke holder

- Uafhængighed:
  Brug metoder fra "repeated measurements"
- Ens spredninger:
  - Brug test og konfidensintervaller, der er angivet med "Sattertwaite" (Welch test)
  - Transformer outcome variabel
- Normalfordeling:
  - Transformer outcome variabel
  - Lav et ikke-parametrisk test (non-parametrisk)

△Nonparametrisk uparret sammenligning

- Mann-Whitney test
  tester om sandsynligheden for, at den ene gruppe resulterer i større værdier end den anden, er 0.5

  eller om medianerne er ens
  (hvis der kun er tale om en forskydning)

  Her giver Mann-Whitney P=0.91, men intet konfidensinterval (kode næste side)
- Permutationstest
  en ide, der kan benyttes i mange sammenhænge, men som fører for vidt her, da det involverer simulationer.....

△Nonparametrisk uparret test i praksis

Mann-Whitney test, også kaldet Wilcoxon two-sample test, eller mere generelt, for flere end to grupper: Kruskal-Wallis test (approksimation for $n > 25$)

```
proc npar1way wilcoxon data=women;
   where country in (1,4);
class country;
*    exact hl;
var vitd;
run;
```

For små samples kan sætningen `"exact hl;"` give et *eksakt test*, men her ville det tage frygtelig lang tid

△Output fra Mann-Whitney test

```
The NPAR1WAY Procedure
             Wilcoxon Scores (Rank Sums) for Variable vitd
                   Classified by Variable country

                          Sum of      Expected       Std Dev          Mean
country         N         Scores      Under H0      Under H0         Score
DK             53       2502.50       2517.50    131.157521     47.216981
EI             41       1962.50       1947.50    131.157521     47.865854
                  Average scores were used for ties.

   Wilcoxon Two-Sample Test
Statistic              1962.5000

Normal Approximation
Z                         0.1106
One-Sided Pr >  Z         0.4560
Two-Sided Pr > |Z|        0.9120

t Approximation
One-Sided Pr >  Z         0.4561
Two-Sided Pr > |Z|        0.9122

Z includes a continuity correction of 0.5.

      Kruskal-Wallis Test
Chi-Square                0.0131
DF                             1
Pr > Chi-Square           0.9089
```

Jeg plejer at bruge P-værdien fra Z-størrelsen, altså her P=0.9120.

Andre bruger Kruskal-Wallis.....

Husk at skelne parret fra uparret

Som regel gør det ingen synderlig forskel *i P-værdi*
om man benytter parametriske eller non-parametriske metoder.

Men det er vigtigt at respektere sit design!

Eks: Målemetoderne MF og SV (fra forelæsningen sidste uge):

Parret T-test:

$t = 0.16, \qquad f = 20$

$P = 0.88$

Sikkerhedsinterval:

(-2.93 cm^3, 3.41 cm^3)

Uparret T-test (galt):

$t = 0.04, \qquad f = 40$

$P = 0.97$

Sikkerhedsinterval:

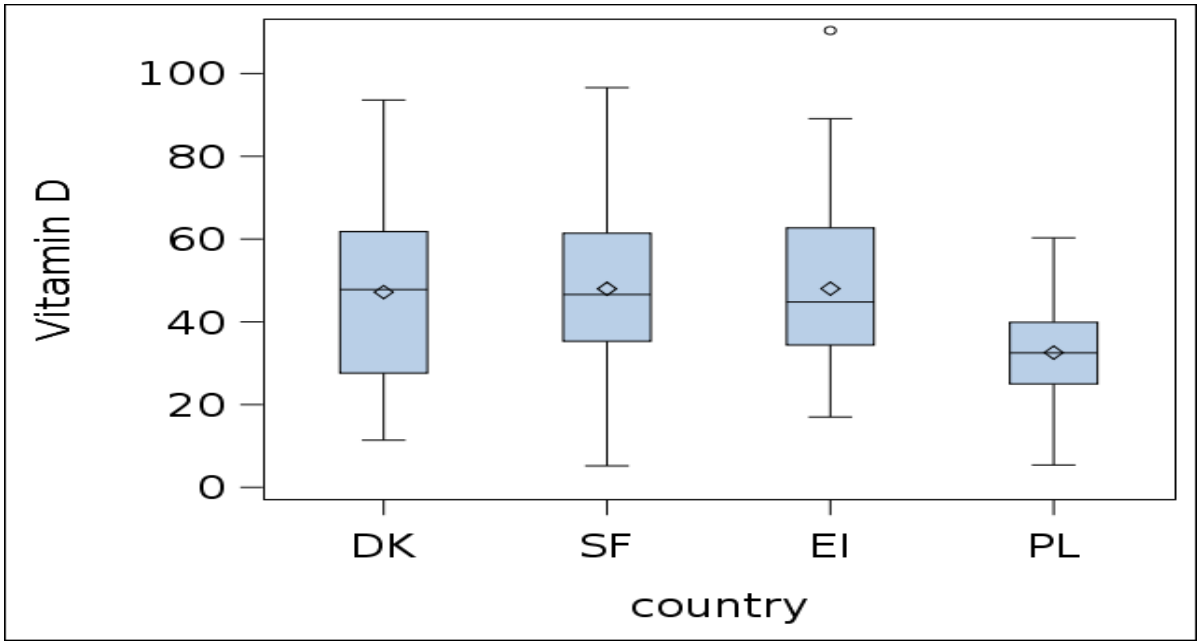
(-12.71 cm^3, 13.19 cm^3)

T-test kontra non-parametrisk alternativ

- T-test giver pr. automatik et konfidensinterval for forskellen på middelværdierne
- Man skal sno sig - og have stor tålmodighed - for at få et konfidensinterval baseret på et non-parametrisk test
- T-testet er lidt stærkere, dvs. man kan nøjes med lidt færre observationer
  - men det er jo fordi man lægger en *antagelse* ind i stedet for...
- Man skal ikke være så bange for normalfordelingsantagelsen, for det er i virkeligheden kun gennemsnittene, der behøver at være pænt normalfordelte, og det er de sædvanligvis, når man har mange observationer i hver gruppe
- Det er kun, hvis man skal udtale sig om enkeltindivider, at man skal være forsigtig med normalfordelingsantagelsen, altså ved prediktioner og diagnostik.

Vitamin D i alle 4 lande

Kode som s. 3, bare uden `where`-sætning



Polen synes at ligge lavere, både i niveau og spredning.

Sammenligning af alle 4 lande

- Vi har set, at Danmark og Irland ikke adskiller sig signifikant fra hinanden
- Er det simpelthen sådan, at alle landene er mere eller mindre identisk mht vitamin D status?
- Man kunne sammenligne alle landene parvis, men det er farligt pga risikoen for massesignifikans (kommer senere...)

I stedet kan man se på hypotesen om ens middelværdier for alle lande under et:

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4 \quad (= \mu)$$

Det kaldes ensidet variansanalyse eller one-way anova

Ensidet variansanalyse, ANOVA

- ensidet: fordi der kun er *et* inddelingskriterium, f.eks. som her `country`
- variansanalyse: fordi man sammenligner variansen *mellem grupper* med variansen *indenfor grupper*

Summary statistics for de 4 grupper:

		Analysis Variable : vitd Vitamin D			
country	N Obs	N	Mean	Std Dev	Variance
DK	53	53	47.1660377	22.7829216	519.0615167
SF	54	54	47.9907407	18.7247114	350.6148183
EI	41	41	48.0073171	20.2221214	408.9341951
PL	65	65	32.5615385	12.4644832	155.3633413

Poolet varians (vægtet gennemsnit): $343.897 = 18.54^2$
Varians mellem de 4 gennemsnit: $3457.997 = 58.80^2$

Antagelser for ensidet ANOVA

- Alle observationer er uafhængige
  (personerne går ikke igen flere gange, er ikke tvillinger o.l.)
- Der er samme spredning
  (samme varians, dvs. biologisk variation) i alle grupper
- Inden for hver gruppe er observationerne normalfordelt

Disse antagelser bør checkes *efter* estimationen,
(fordi man benytter størrelser fra selve analysen)
men *før* fortolkningen.

Ensidet ANOVA i praksis

Man kan med fordel undlade at benytte proceduren ANOVA, og i stedet bruge den mere generelle GLM, her med ekstra options, som kommenteres efterfølgende:

```
proc glm data=women;
class country;
model vitd=country / solution clparm;
means country / hovtest welch;
run;
```

Bemærkninger til GLM-kode

Options:

- `solution` betyder, at man gerne vil have printet estimater
- `clparm` betyder, at der skal være konfidensgrænser på disse
- `means`-sætningen giver
  - `hovtest`:
    homogeneity **o**f **v**ariance test, dvs. test af ens spredninger i de 4 grupper (lande),
    her i form af Levenes test, se s. 45
  - `welch`:
    Welch test, som bruges, når spredningerne afhænger af land,
    se s. 49

Output fra GLM (ensidet ANOVA)

Den typiske start på outputtet fra koden s. 37
(den mindre brugbare del):

```
The GLM Procedure
Dependent Variable: vitd   Vitamin D

                                 Sum of
Source                     DF       Squares   Mean Square  F Value  Pr > F
Model                       3   10373.99129    3457.99710    10.06  <.0001
Error                     209   71874.40589     343.89668
Corrected Total           212   82248.39718
```

Dette er en såkaldt *variansanalysetabel*,
som her giver testet for ens middelværdier,
men som generelt ikke kan bruges til så forfærdeligt meget.

Output, fortsat

Nu den mere brugbare del:

R-Square	Coeff Var	Root MSE	vitd Mean
0.126130	43.04626	18.54445	43.08028

Source	DF	Type III SS	Mean Square	F Value	Pr > F
country	3	10373.99129	3457.99710	10.06	<.0001

Parameter		Estimate		Standard Error	t Value	Pr > |t|
Intercept		47.99074074	B	2.52358020	19.02	<.0001
country	DK	-0.82470300	B	3.58567617	-0.23	0.8183
country	EI	0.01657633	B	3.84137747	0.00	0.9966
country	PL	-15.42920228	B	3.41455343	-4.52	<.0001
country	SF	0.00000000	B	.	.	.

Parameter		95% Confidence Limits	
Intercept		43.01580657	52.96567491
country	DK	-7.89343136	6.24402535
country	EI	-7.55623632	7.58938899
country	PL	-22.16058279	-8.69782177
country	SF	.	.

med fortolkning på de følgende sider

Bemærkninger til output, I

Variansanalysetabel s. 39

- Spredningen $\sigma \sim s$, `Root MSE`:
  Den poolede variation for de 4 grupper=lande (within groups)
- F Value:
  Teststørrelsen for test af ens middelværdier i de 4 grupper, med tilhørende P-værdi. Her er $P < 0.0001$, dvs. de 4 middelværdier kan *ikke* antages at være ens.

Vi forkaster nulhypotesen om ens middelværdier, hvis variationen *mellem grupper* er for stor
i forhold til variationen *indenfor grupper*,
for så tyder det på en reel forskel mellem grupperne.

Bemærkninger til output, II

Estimater fra s. 40:

- `Intercept` svarer til niveauet for referencegruppen (sidste gruppe, alfabetisk eller numerisk), dvs. Finland (SF)
- Estimatet ud for f.eks. `country DK` er forskellen i niveau mellem DK og SF (referencegruppen)

Bemærk:
Ved omkodning af grupper kan man få vilkårlige forskelle frem. Dette er årsagen til, at man risikerer at få en NOTE om noget med en singulær matrix.... og den er altså *ikke* farlig

Modelantagelse 1: Uafhængighed

Dette er noget, man skal vide

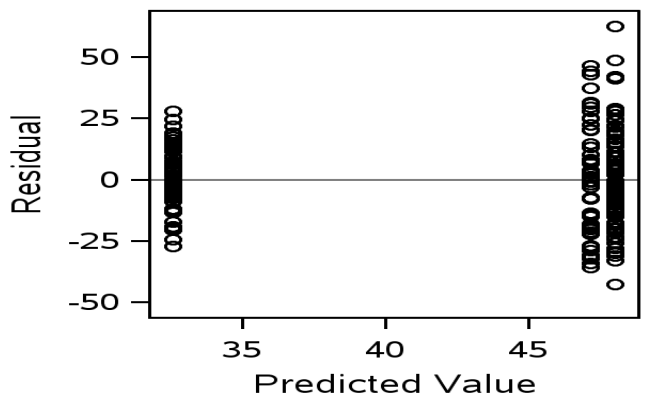
- ingen tvillinger, søskende etc.
- kun *en* observation for hver person
  (ellers hører det hjemme under emnet
  "Korrelerede målinger", kursets sidste emne)

Hvis observationerne er korrelerede (afhængige af hinanden), kan man få ganske betydelige fejl i sin analyse,
hovedsagelig i form af forkerte standard errors
og dermed forkerte konfidensintervaller og P-værdier,
og altså i sidste konsekvens forkerte konklusioner.

Modelantagelse 2: Identiske spredninger i grupperne

Kaldes som regel varianshomogenitet, og checkes ud fra

- Scatter plot eller Box plot af selve observationerne
- Test af hypotese om ens varianser
  (sædvanligvis Levenes test, se næste side)
- Residualer tegnet op mod predikterede
  (=forventede=fittede) værdier, skal være *jævnt*



Dette er en del
af den automatiske
grafik, se. s. 47

Levenes test for identiske spredninger

Brug option `hovtest` i `means`-sætningen s. 37
Vi har allerede set, at Polen måske har mindre spredning end de andre tre lande - men det *kunne* jo være en tilfældighed:

Levene's Test for Homogeneity of vitd Variance
ANOVA of Squared Deviations from Group Means

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
country	3	3934223	1311408	5.84	0.0008
Error	209	46968190	224728		

Ved sammenligning af de 4 variansestimater fås en P-værdi på P=0.0008, og altså kraftig signifikans!

Dette vil vi gerne gøre noget ved lige om lidt (s. 49).

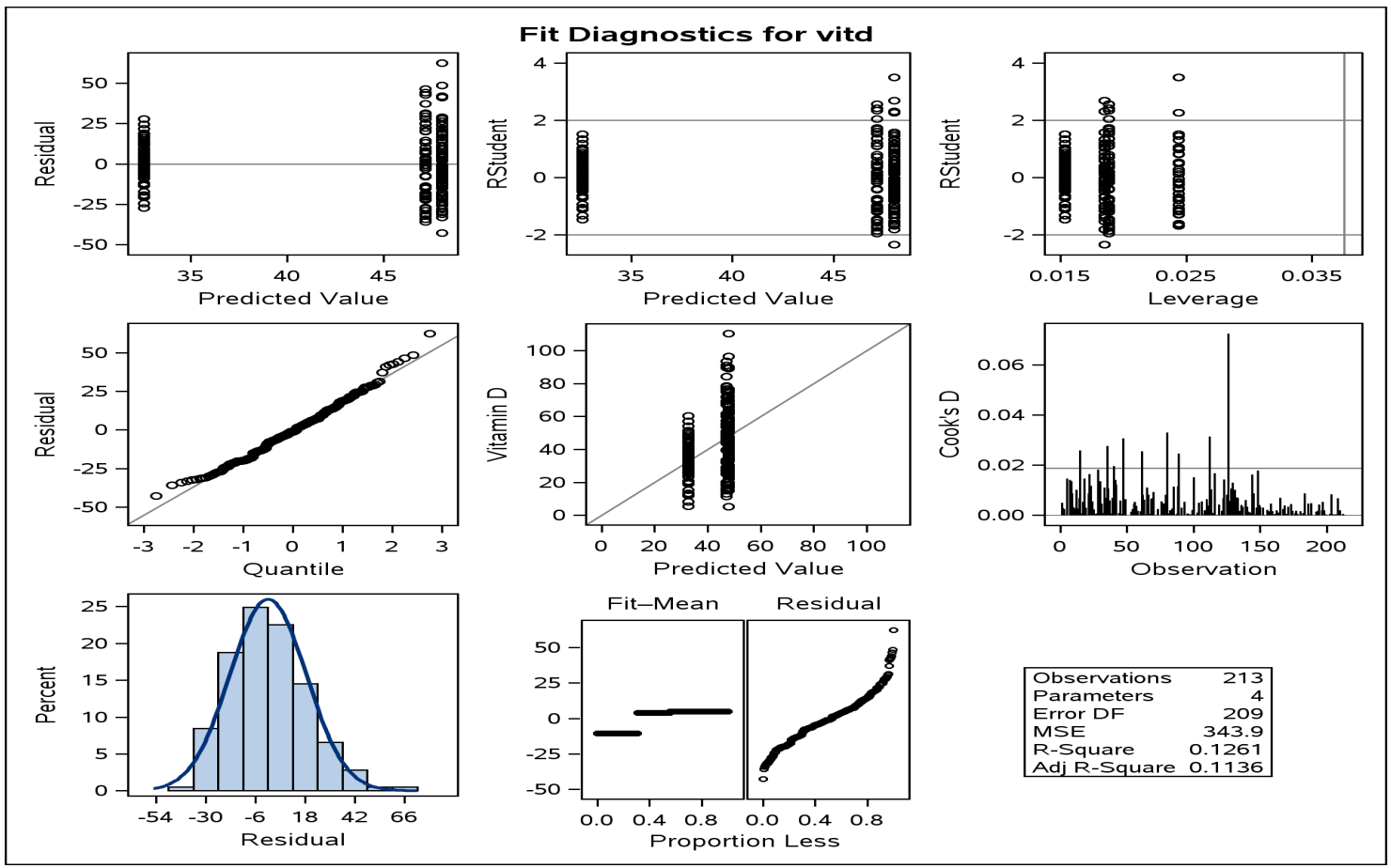
Modelantagelse 3: Normalfordelingsantagelsen

Det er antaget, at observationerne følger en normalfordeling *inden for hver gruppe.* Dette kan checkes:

- ved at tegne histogrammer eller fraktildiagrammer for hver gruppe (kun hvis man har rigtig mange observationer)
- ved at tegne histogram eller fraktildiagram for *residualerne* *="observation - fittet værdi"*
  som her blot er observation minus det relevante gruppegennemsnit
- Det er ikke særligt informativt med normalfordelingstest
  - Hvis man har mange observationer, bliver det stort set altid forkastet - uden at det betyder noget i praksis
  - Hvis man har få observationer, bliver det stort set altid godkendt - uden at man derved har påvist at der er tale om en normalfordeling

Modelkontrol: Diagnostics Panel

Automatisk plot i SAS, ellers brug `ods graphics on;`



Bemærkninger til Diagnostics Panel

Foreløbig beskæftiger vi os kun med første søjle (S1) på s. 47

- Figur (R1,S1): residualer mod predikterede værdier
  Har de samme spredning?
  Næh, den stiger vist lidt med den predikterede værdi
- Figur (R2-R3,S1): Fraktildiagram og histogram af residualerne: Ser de normalfordelte ud?
  Næsten, dog lidt "hængekøje"=skævhed=hale mod højre

Nogle SAS-versioner giver automatisk disse plots. Ellers kan man benytte ODS-systemet.

```
ods graphics on;
....
....
ods graphics off;
```

Hvad gør vi ved forskellen i spredninger?

- Er det slemt? Tja, ikke at dømme ud fra grafikken....
- Kan vi slippe for forudsætningen ligesom for T-testet?
  Ja: vi kan lave et `welch` test i stedet for
  (option `welch` i `means`-sætningen, se s. 37)

```
          Welch's ANOVA for vitd

Source              DF    F Value    Pr > F
country        3.0000      14.64    <.0001
Error           102.3
```

Vi kan altså godt føle os sikre på den fundne forskel
- men vi får ikke revideret vores sammenligninger landene imellem....

Konklusion...?

- Modellen er ikke helt rimelig
- F-test viser dog helt klart en forskel på middelværdien af vitamin D i de fire lande,

men var det i virkeligheden det, vi gerne ville vide?

Eller ville vi hellere vide, hvilke lande,
der adskiller sig fra hvilke andre?
– og hvor store forskellene er?

Multiple sammenligninger

Parvise t-test giver problemer med massesignifikans

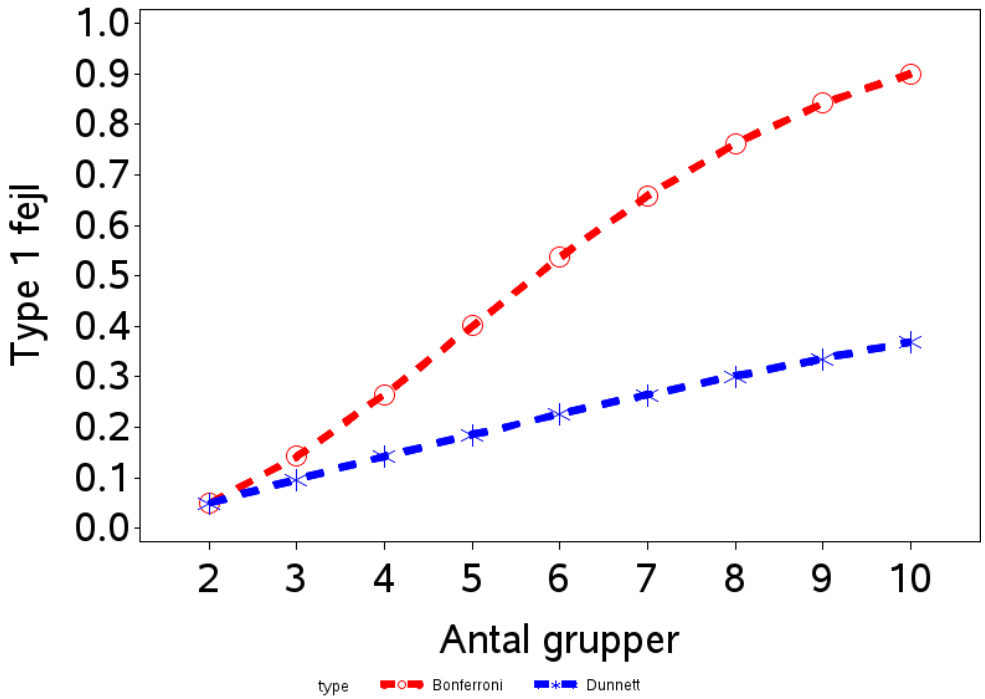
Hvis man sammenligner k grupper (lande) parvist, er der $m = k(k-1)/2$ mulige test, hver med signifikansniveau $\alpha = 0.05$.

Den *totale* risiko for at begå en type 1 fejl er derfor reelt væsentlig højere, men hvor høj?

Hvis testene var uafhængige af hinanden (det er de dog ikke), ville signifikansniveauet være: $1 - (1-\alpha)^m$,

f.eks. som her, for k=4: 0.26

Type 1 fejl ved *uafhængige* multiple sammenligninger



- Øverste graf: Alle grupper sammenlignes med alle andre
- Nederste graf:
  Alle grupper sammenlignes med en enkelt kontrolgruppe

Korrektion for multiple sammenligninger

- Bonferroni
  - benytter signifikansniveau $\frac{\alpha}{m}$
  - stærkt konservativ, dvs. for høje P-værdier (lav styrke)
- Sidak
  - benytter signifikansniveau $1-(1-\alpha)^{\frac{1}{m}} \approx \frac{\alpha}{m}$ for små m
  - lidt mindre konservativ, men stadig ret lav styrke
- Tukey eller Games-Howell
  - sidstnævnte i tilfælde af uens varianser (findes ikke i SAS)
  - giver større styrke
- Dunnett
  - korrigerer kun for test mod referencegruppe (typisk en kontrolgruppe eller 'tid 0')

Hvilken korrektion skal man vælge?

Dette er et meget vanskeligt spørgsmål, fordi:

- Der findes rigtig mange
- med hver deres fordele og ulemper

og hvilke (hvor mange) tests skal man korrigere for?

- dem i denne publikation?
- alle de, der vedrører dette projekt?
- hele min videnskabelige produktion?
- mine kollegers? ..?

Hvilken korrektion skal man vælge?, II

Jeg bruger oftest Tukey (eller Dunnett), fordi:

- Den sikrer lav type 1 fejls risiko
- Den tillader forskelle i gruppestørrelse

men den tillader ikke *vilkårlige sammenligninger*,
f.eks. Polen mod gruppen bestående af de 3 andre
(i så fald skal man bruge Scheffee)

Hvis det ikke drejer sig om en one-way anova, kan man altid pr. håndkraft benytte Bonferroni (eller Sidak).

△Korrektion for massesignifikans i praksis

f.eks. Tukey-korrektion:
Benyt option `adjust=tukey` i `lsmeans`-sætningen:

```
proc glm data=women;
class country;
model vitd=country / solution clparm;
LSMEANS country / ADJUST=TUKEY pdiff cl;
run;
```


Tukey korrektion for sammenlingning af lande

Least Squares Means
Adjustment for Multiple Comparisons: Tukey-Kramer

country	vitd LSMEAN	LSMEAN Number
DK	47.1660377	1
EI	48.0073171	2
PL	32.5615385	3
SF	47.9907407	4

Least Squares Means for effect country
Pr > |t| for H0: LSMean(i)=LSMean(j)

Dependent Variable: vitd

i/j	1	2	3	4
1		0.9963	0.0002	0.9957
2	0.9963		0.0003	1.0000
3	0.0002	0.0003		<.0001
4	0.9957	1.0000	<.0001	

i	j	Difference Between Means	Simultaneous 95% Confidence Limits for LSMean(i)-LSMean(j)	
1	2	-0.841279	-10.830178	9.147620
1	3	14.604499	5.715969	23.493029
1	4	-0.824703	-10.110959	8.461553
2	3	15.445779	5.867491	25.024067
2	4	0.016576	-9.931900	9.965052
3	4	-15.429202	-24.272281	-6.586124

Kommentarer til Tukey korrektion

Selv om Tukey-korrektionen ikke er optimal pga de uens varianser, er P-værdierne så tydelige, at vi kan tillade os at konkludere, at:

- Land nr. 3 (Polen) adskiller sig signifikant fra de 3 øvrige.
- Herudover er der ingen forskelle, dvs. Danmark, Irland og Finland adskiller sig ikke parvist fra hinanden

Eksempelvis finder vi Finland vs. Polen:

- Sammenligning fra GLM (s. 40): 15.43 (8.70, 22.16)
- Tukey-korrigeret: 15.43 (6.59, 24.27)

Bemærk, at korrektionen gør CI bredere

ANOVA vs Multiple Sammenligninger (MS)

- Kan man risikere, at ANOVA er insignifikant, men at parvise tests findes signifikante?
  Ja, nemt, fordi ANOVA er et svagt test pga mange frihedsgrader
  Også efter Tukey-korrektion? Formentlig
- Kan man risikere at ANOVA er signifikant, uden at der er nogensomhelst parvise T-tests, der er signifikante?
  Formentlig *kun*, hvis de er Tukey-korrigerede...?
- Er vi overhovedet interesseret i ANOVA?
  – eller skal vi bare gå direkte til parvise sammenligninger?
  Det kunne vi godt, men de laves i praksis i tilslutning til ANOVA'en, såe....

Hvis antagelserne ikke holder

- Vægtet analyse (Welch's test, som vi så tidligere)
- Transformation (ofte logaritmer)
  kan afhjælpe såvel varians**in**homogenitet som dårlig normalfordelingstilpasning
- Non-parametrisk sammenligning
  - Kruskal-Wallis test
  - (Permutationstest)

Husk: Antagelserne er ikke altid lige vigtige,
vigtigst når man skal udtale sig om enkeltindivider

Non-parametrisk Kruskal-Wallis test

Udvidelse af Mann-Whitney testet til flere end 2 grupper (ingen ændring i selve koden, se s. 29)

```
   Kruskal-Wallis Test

Chi-Square          26.6819
DF                        3
Pr > Chi-Square      <.0001
```

Bemærk:

- Dette er et approksimativt test
- Man kan også få en eksakt vurdering af teststørrelsen (se s. 29), men pas på i tilfælde af store materialer (som f.eks. her)

  Det tager forfærdeligt lang tid - dagevis

Sammenligning af Finland og Polen

...som om vi kun havde disse to lande, dvs. et T-test:

```
proc ttest data=women; where country in (2,6);
class country;
var vitd;
run;
```

som giver os forskellen Finland vs. Polen:
15.43 (9.51, 21.35), $P < 0.0001$

Bemærk: Samme estimat som tidligere,
men smallere konfidensgrænser (se s. 40). Hvorfor?

Og hvorfor mon der er den forskel?
Kan de godt lide sol i Finland?

Solvaner i Finland og Polen

Vi definerer en variabel `sol` (se kode s. 110) som

- 0 Solhadere: Folk, der undgår solen (`sunexp=1`)
- 1 Solelskere: Folk, der godt kan lide solen (`sunexp=2,3`)

Der kunne være confounding mellem land og solvaner

En simpel optælling af solelskere (se s. 110) giver:

- Finland: 40 ud af 54, dvs. 74.1%
- Polen: 39 ud af 65, dvs. 60.0%

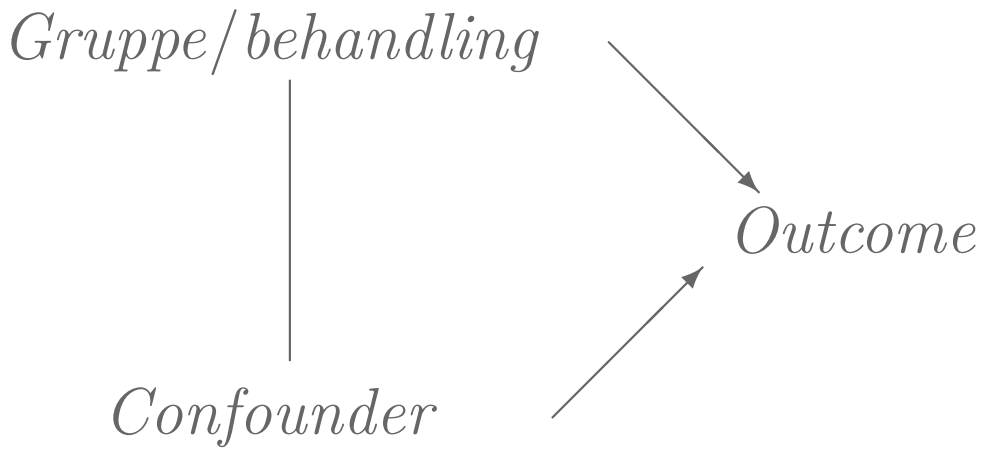
Kan denne forskel i sol-præferencer være forklaringen på forskellen i Vitamin-D?

Confoundere

En confounder er en variabel, som

- har en effekt på outcome
- er relateret til gruppen
  (dvs. der er forskel på værdierne i de to grupper)
- men er ikke en direkte konsekvens af gruppen

En sådan variabel kan give **"bias"** (meget mere om dette senere).



Mediator variabel

En mediator er en variabel, som

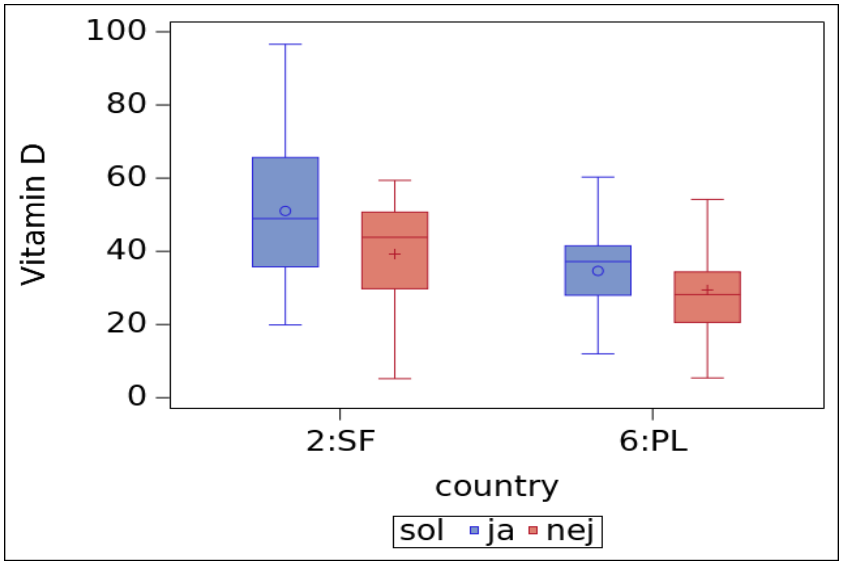
- har en effekt på outcome
- er relateret til gruppen
  som en direkte konsekvens af denne

Modeller med og uden en sådan variabel kan have helt forskellig fortolkning!

Er solvaner en mediator?

Har solvanerne overhovedet en betydning?

```
proc sgplot data=women;
  where country in (2,6);
vbox vitd /
     category=country group=sol;
run;
```



Sammenligning af blå og røde bokse tyder på en positiv effekt af sol - som ventet

Hvordan korrigerer vi for forskelle i solvaner?

Vi vil gerne sammenligne folk fra Finland og folk fra Polen,
som har samme forhold til solen,
uanset hvad dette forhold måtte være.

Vi siger, at vi fastholder værdien af solvaner,
når vi vurderer forskellen mellem landene.

Her antages altså *samme* forskel på landene uanset solvaner

eller

samme effekt af solvaner i de to lande

Tosidet variansanalyse: Additiv model

Tosidet, fordi der nu er to inddelingskriterier:

- Land: Finland, Polen
- Solvaner: Kan lide / kan ikke lide

Additiv betyder: Uden interaktion
(vedr. interaktion, se. s. 75ff)

Vi vil sammenligne folk fra Finland og Polen, der har samme præference for sol, dvs.

- 14 finner vs. 26 polakker, der *ikke* kan lide sol
- 40 finner vs. 39 polakker, der *godt* kan lide sol

og disse to forskelle *pooles* så til en *generel forskel på landene*.

To-sidet ANOVA i praksis

```
proc glm data=women;
  where country in (2,6);
class country sol;
model vitd=country sol / solution clparm;
run;
```

som giver outputtet

```
     Class Level Information
Class          Levels    Values
country             2    2:SF 6:PL
sol                 2    ja nej

Number of Observations Used         119

R-Square     Coeff Var      Root MSE     vitd Mean
0.242377      38.51354      15.23712      39.56303
```

med fortolkning s. 71

Output fra 2-sidet ANOVA af Vitamin D, fortsat

Source	DF	Type III SS	Mean Square	F Value	Pr > F
country	1	5920.804988	5920.804988	25.50	<.0001
sol	1	1594.131737	1594.131737	6.87	0.0100

Parameter		Estimate	Standard Error	t Value	Pr > |t|	95% Confidence Limits	
Intercept		27.86071 B	2.60580	10.69	<.0001	22.69960	33.02182
country	2:SF	14.32654 B	2.83696	5.05	<.0001	8.70757	19.94550
country	6:PL	0.00000 B	.	.	.	.	.
sol	ja	7.83471 B	2.98995	2.62	0.0100	1.91274	13.75668
sol	nej	0.00000 B	.	.	.	.	.

.

Fortolkning af 2-sidet ANOVA

Effekter:

- Finland vs. Polen, for fastholdte solvaner:
  Finland estimeres til at ligge 14.33 nmol/l højere end Polen,
  med 95% CI: (8.71, 19.95)
  Dette afviger noget fra T-testet (s. 62), hvor vi fik estimatet 15.43, med 95% CI: (9.51, 21.35).
  Vi fik her indsnævret konfidensintervallet en anelse
  fordi vi fjernede noget af residualvariationen

- Folk, der kan lide sol har et 7.83 højere niveau end de fra samme land, som ikke kan lide sol,
  med 95% CI: (1.91, 13.76), P=0.01

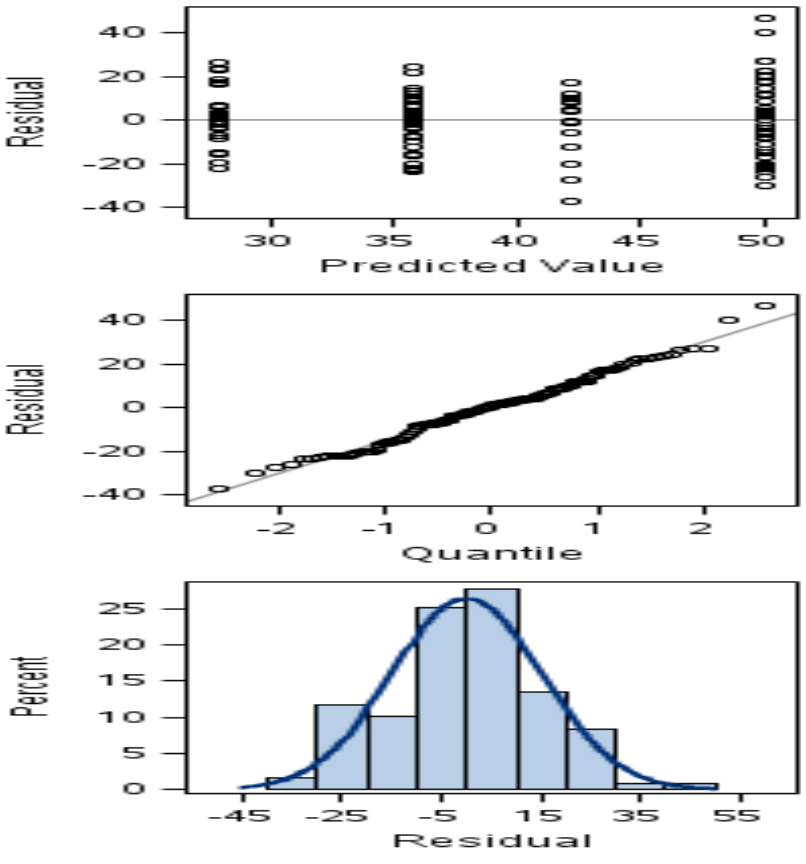
Vurdering af Finland vs. Polen

Model indeholder	estimat	CI
kun country (T-test)	15.43	(9.51, 21.35)
solvaner og land	14.33	(8.71, 19.95)
SF vs. PL, solelskere	16.40	(9.37, 23.43)
SF vs. PL, solhadere	9.82	(-0.47, 20.11)

Foreløbig har vi antaget, at forskellen på landene er den samme, uanset solvaner, dvs. at der ikke er nogen interaktion.

De to sidste rækker i tabellen antyder, at dette *måske ikke er rimeligt*, fordi forskellen på landene er noget større for solelskere end for solhadere.

Modelkontrolplots - kun venstre kolonne

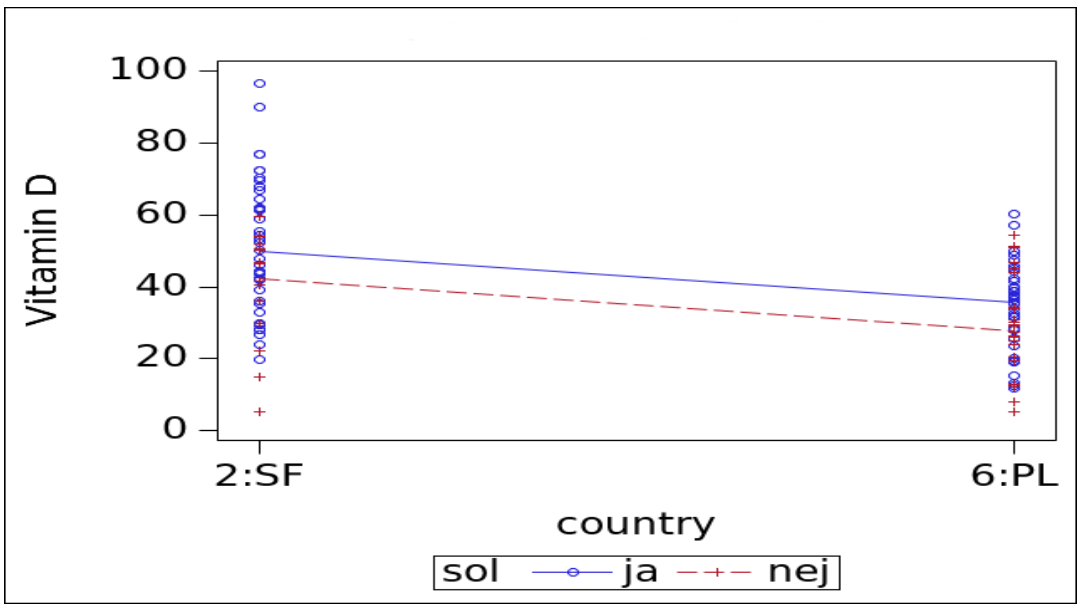


Intet specielt mønster her

Disse to figurer viser en udmærket tilpasning af normalfordelingen.

Den additive model - Predikterede værdier

Predikterede værdier svarende til samme sol-gruppe er forbundet, og hældningen viser derfor landeforskellen



Modellen antager samme forskel på lande, dvs. parallelle linier.

Mulig interaktion?

Hvad betyder interaktion mellem `country` og `sunexposure`?

- at forskellen på landene (Finland vs. Polen) afhænger af om man kan lide sol eller ej (altså kategorien af `sunexposure`)

eller *vendt den omvendte vej*:

- at effekten af sol (elskere vs. hadere) afhænger af, hvilket land, man bor i (altså kategorien af `country`)

Interaktion:
Effekten af den ene kovariat afhænger af, hvad den anden er.

Vurderinger af sol-effekt

Effekten af sol kunne tænkes at afhænge af landet
(breddegraden eller vejret)
Før antog vi, at effekten var den samme i begge lande
(additivitet=parallelle linier på plottet s. 74)

Vi opdeler nu efter land (helt separate T-tests):

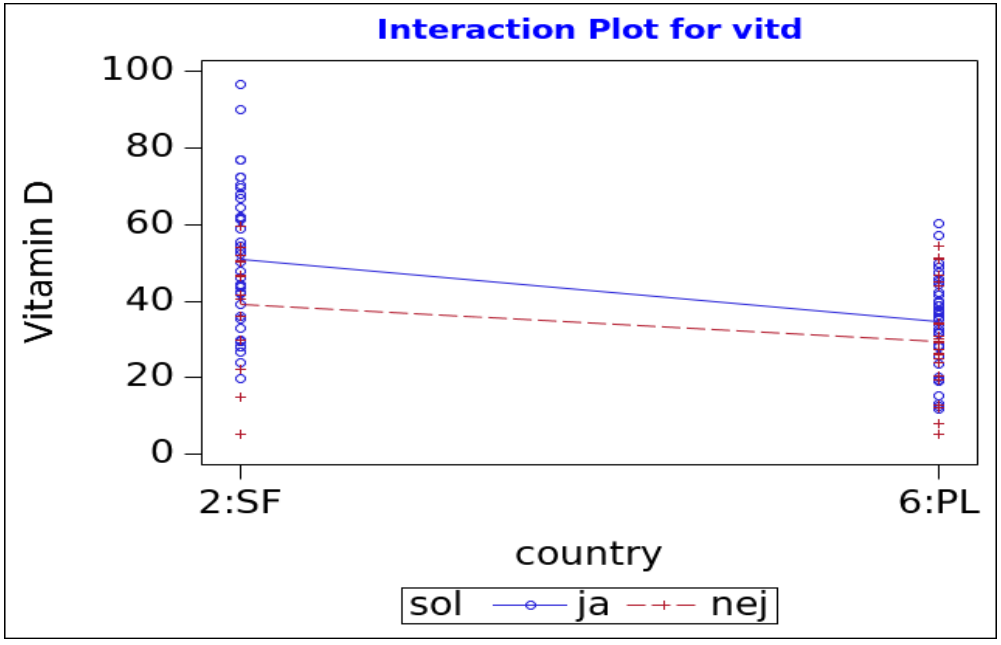
Vitamin D for solelskere vs. solhadere:

- i Finland:
  11.79 (5.64), 95% CI: (0.48, 23.10), P=0.04

- i Polen:
  5.21 (3.11), 95% CI: (-1.01, 11.42), P=0.10

Er de to vurderinger af solvanernes betydning forskellige?
I så fald siger vi, at der er interaktion

Modellen med interaktion



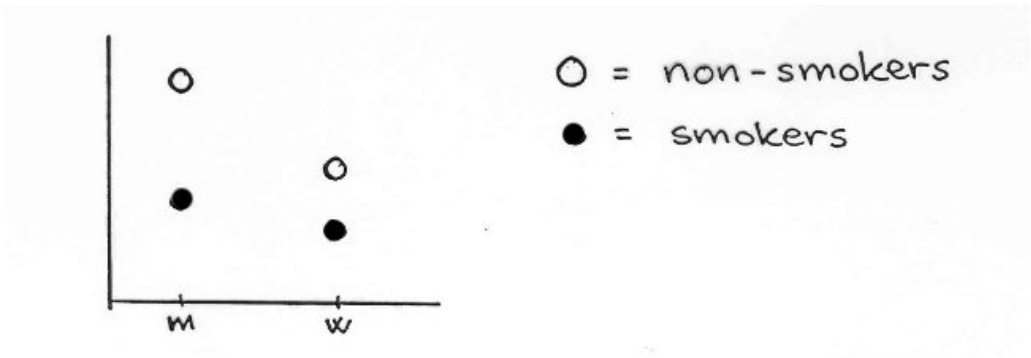
Er der forskel på sol-effekterne i de to lande?

Ikke så meget, ser det ud til....
for linierne er *næsten* parallelle (fortsættes s. 82)

Vekselvirkning = Interaktion

Tænkt eksempel:

- To inddelingskriterier: køn og rygestatus
- Outcome: FEV_1



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

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 FEV_1
  – kunne undersøges ved en longitudinel undersøgelse

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			
	Mild	Moderate	Heavy	All
–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.

Interaktion mellem mængden og varigheden af rygningen

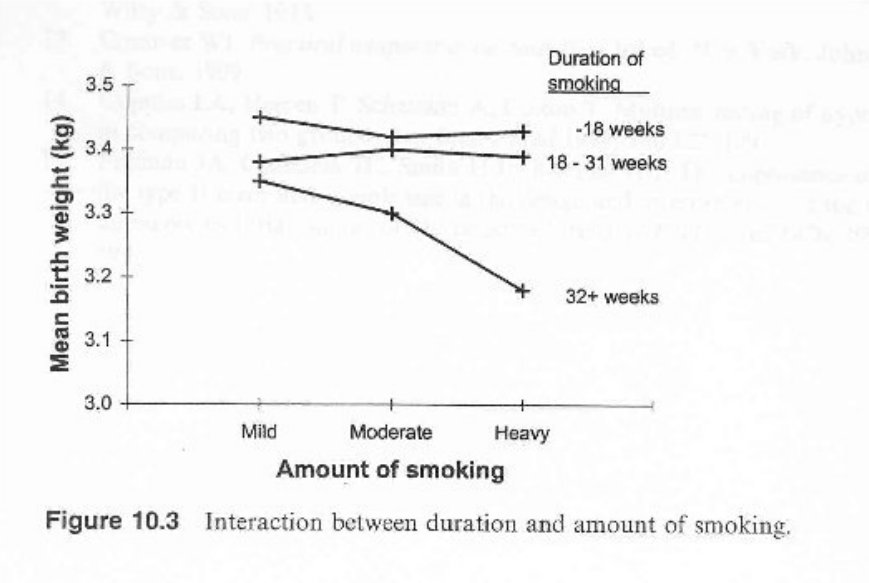


Figure 10.3 Interaction between duration and amount of smoking.

- 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**....

Interaktion mellem solvaner og land?

for Finland og Polen:

```
proc glm data=women;
  where country in (2,6);
class country sol;
model vitd=country sol country*sol / solution clparm;
run;
```

som giver outputtet

```
Dependent Variable: vitd   Vitamin D

R-Square     Coeff Var      Root MSE     vitd Mean
0.249976      38.48615      15.22628      39.56303

Source              DF     Type III SS     Mean Square    F Value    Pr > F
country              1     4283.432660     4283.432660      18.48    <.0001
sol                  1     1799.317754     1799.317754       7.76    0.0062
country*sol          1      270.136008      270.136008       1.17    0.2827
```

Output vedr. interaktion, fortsat

Parameter		Estimate	Standard Error	t Value	Pr > |t|	95% Confidence Limits	
Intercept		29.43846154 B	2.98612015	9.86	<.0001	23.52353223	35.35339085
country	2:SF	9.81868132 B	5.04746430	1.95	0.0542	-0.17937402	19.81673666
country	6:PL	0.00000000 B	.	.	.	.	.
sol	ja	5.20512821 B	3.85506453	1.35	0.1796	-2.43101269	12.84126911
sol	nej	0.00000000 B	.	.	.	.	.
country*sol	2:SF ja	6.58522894 B	6.10061461	1.08	0.2827	-5.49891449	18.66937237
country*sol	2:SF nej	0.00000000 B	.	.	.	.	.
country*sol	6:PL ja	0.00000000 B	.	.	.	.	.
country*sol	6:PL nej	0.00000000 B	.	.	.	.	.

Det ses her, at der ikke er signifikant interaktion (P=0.28)

Fortolkningen af estimaterne følger på de næste sider.

Fortolkning af estimater

Betydningen af de enkelte estimater, fra outputtet på forrige side:

- `Intercept`=29.44:
  Det estimerede niveau (her blot gennemsnittet) af vitamin D for referencegruppen, dvs. solhadere fra Polen.
- `country 2:SF`=9.82:
  Finlands forspring frem for Polen for sol-referencegruppen, dvs. for solhadere
- `sol ja`=5.21:
  Effekten af at kunne lide sol vs. at hade den, for country-referencegruppen, dvs. for polakker

Estimater, fortsat

- `country*sol 2:SF ja`=6.59:
  Den ekstra effekt af soldyrkning i Finland
  i forhold til i Polen,

  eller

  Den ekstra fordel af at være finne blandt soldykere,
  i forhold til blandt solhadere

  Den totale effekt af solen i Finland er således
  5.21+6.59=11.80, som vi også fandt før, se s. 76

  Denne ekstra effekt er ikke signifikant, men
  konfidensintervallet er (-5.50, 18.67), altså meget bredt, set i
  relation til effekternes størrelse, så vi kan faktisk ikke afgøre,
  om der er interaktion eller ej!!

Predikterede værdier - opsummeret

Referenceniveauerne er:
`country=6:PL, sol=nej`
(de sidste i den alfabetiske rækkefølge)

Denne gruppe har et forventet vitamin D niveau på `intercept=29.44`

For de andre niveauer skal der adderes et eller flere ekstra led, som angivet i skemaet.

solelsker?	country: Finland	country: Polen
ja	29.44	29.44
	+ 5.21	+ 5.21
	+ 9.82	
	+ 6.59	
	=51.05	=34.64
nej	29.44	29.44
	+ 9.82	
	=39.26	

Herved fremkommer de predikterede værdier,
(som her også blot er gennemsnittene)

Estimationsvenlig kode

Separate effekter af land (udelad country):

```
proc glm data=women;
  where country in (2,6);
class country sol;
model vitd=sol country*sol /
      solution clparm;
run;
```

Parameter		Estimate		Standard Error
Intercept		34.64358974	B	2.43815689
sol	0	-5.20512821	B	3.85506453
sol	1	0.00000000	B	.
country*sol	SF 0	9.81868132	B	5.04746430
country*sol	SF 1	16.40391026	B	3.42645631
country*sol	PL 0	0.00000000	B	.
country*sol	PL 1	0.00000000	B	.

Separate effekter af sol (udelad sol):

```
proc glm data=women;
  where country in (2,6);
class country sol;
model vitd=country country*sol /
      solution clparm;
run;
```

Parameter		Estimate		Standard Error
Intercept		34.64358974	B	2.43815689
country	SF	16.40391026	B	3.42645631
country	PL	0.00000000	B	.
country*sol	SF 0	-11.79035714	B	4.72821066
country*sol	SF 1	0.00000000	B	.
country*sol	PL 0	-5.20512821	B	3.85506453
country*sol	PL 1	0.00000000	B	.

Fokus på effekt af sol

Soldyrkere vs. solhadere, stadig kun Finland og Polen:

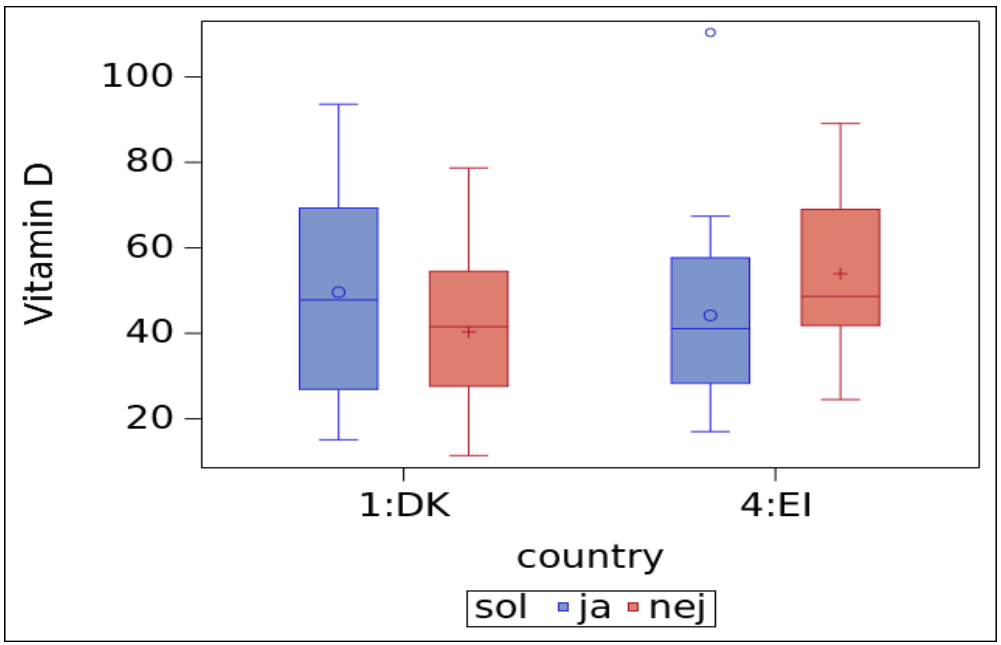
Model indeholder	estimat	CI
kun solvaner (T-test)	10.07	(3.63, 16.51)
solvaner og land	7.83	(1.91, 13.76)
solvaner, kun SF	11.79	(0.48, 23.10)
solvaner, kun PL	5.21	(-1.01, 11.42)

Confounding mellem land og sol giver forskellen i de to første linier.

De to sidste linier viser den insignifikante interaktion (P=0.28)

Sammenligning, nu af Danmark og Irland

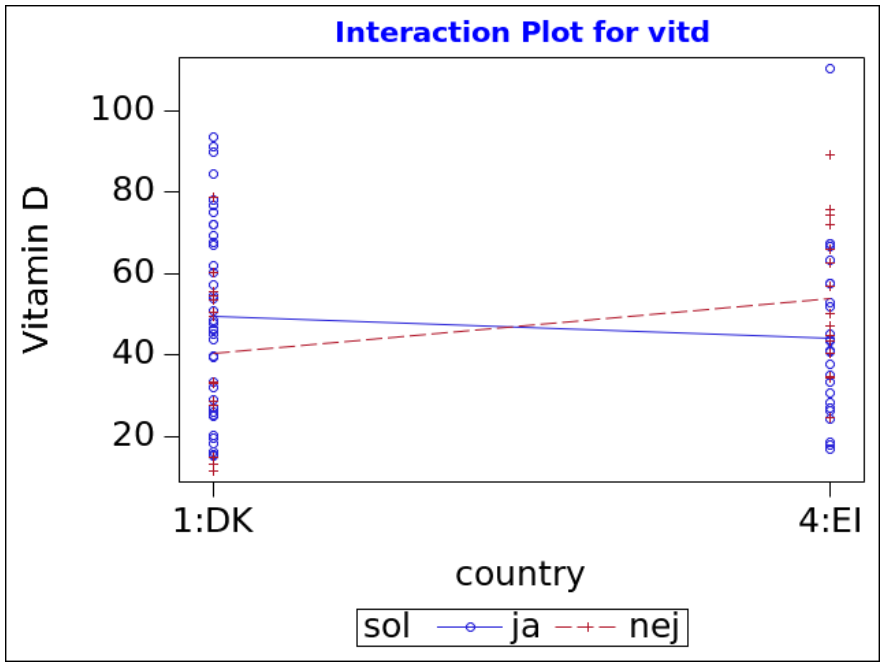
(ganske som Finland vs. Polen, s. 66)



Forskellen på blå og røde viser effekterne af sol
– og de ses at være modsatrettede!?

Sammenligning af Danmark og Irland

Prediktion i interaktionsmodel (dvs. gennemsnit), ganske som s. 77



Interaktion mellem solvaner og land?

for Danmark og Irland:

```
     Class Level Information
Class         Levels    Values
country            2    1:DK 4:EI
sol                2    ja nej

Number of Observations Used          94

R-Square     Coeff Var      Root MSE     vitd Mean
0.042260      45.19359      21.48186      47.53298

Source                     DF     Type III SS     Mean Square    F Value    Pr > F
country                     1      337.137692      337.137692       0.73    0.3950
sol                         1        1.125877        1.125877       0.00    0.9607
country*sol                 1     1816.227997     1816.227997       3.94    0.0503

                                                      Standard
Parameter                        Estimate                 Error   t Value   Pr > |t|     95% Confidence Limits
Intercept                    53.95625000 B          5.37046507     10.05     <.0001    43.28688377  64.62561623
country     1:DK            -13.62053571 B          7.86155589     -1.73     0.0866   -29.23888864   1.99781721
country     4:EI              0.00000000 B                   .         .          .             .            .
sol         ja                -9.75625000 B          6.87755102     -1.42     0.1595   -23.41970551   3.90720551
sol         nej               0.00000000 B                   .         .          .             .            .
country*sol 1:DK ja           19.03848443 B          9.59663723      1.98     0.0503    -0.02691043  38.10387930
country*sol 1:DK nej           0.00000000 B                   .         .          .             .            .
country*sol 4:EI ja            0.00000000 B                   .         .          .             .            .
country*sol 4:EI nej           0.00000000 B                   .         .          .             .            .
```

Kommentarer til Danmark vs. Irland

Der er *næsten signifikant* interaktion (P=0.0503)

- men effekterne af sol er modsatrettede! - mystisk...
- Forskellen i sol-effekt kan være fra ca. 0 og helt op til en forskel på 38.1, svarende til, at danskere får en effekt på 38.1 nmol/l *mere* ud af at dyrke sol i forhold til Irland
- Det er godt nok en meget stor forskel.....
  Vi kan ikke konkludere på sikkert grundlag her,
  vi ved simpelthen for lidt

△Fokus på effekt af sol

- I `model`-sætningen *bytter vi om på* effekterne `sol` og `country*sol`. Det bevirker, at vi får
  - stadig får testet for interaktion (type III)
  - estimater for sol-effekten for hvert land
  - samlet effekt af `sol`, for alle lande på en gang (type I)
    (Se evt. forskellen fra output s. 82-83)
- Hvis man (som det er gjort her) også tilføjer `noint` i `model`-sætningen, fås estimater for middelværdien for hver af de to lande for `sol`-referencegruppen (`sol=1`, dvs. solelskere)

△Fokus på effekt af sol, output

Source	DF	Type I SS	Mean Square	F Value	Pr > F
country	2	212398.4633	106199.2317	230.13	<.0001
country*sol	2	1816.2377	908.1189	1.97	0.1457
sol	0	0.0000	.	.	.

Source	DF	Type III SS	Mean Square	F Value	Pr > F
country	1	337.137692	337.137692	0.73	0.3950
country*sol	1	1816.227997	1816.227997	3.94	0.0503
sol	1	1.125877	1.125877	0.00	0.9607

Parameter		Estimate		Standard Error	t Value	Pr > |t|
country	DK	49.61794872	B	3.43985063	14.42	<.0001
country	EI	44.20000000	B	4.29637206	10.29	<.0001
country*sol	DK 0	-9.28223443	B	6.69288713	-1.39	0.1689
country*sol	DK 1	0.00000000	B	.	.	.
country*sol	EI 0	9.75625000	B	6.87755102	1.42	0.1595
country*sol	EI 1	0.00000000	B	.	.	.
sol	0	0.00000000	B	.	.	.
sol	1	0.00000000	B	.	.	.

Konfidensgrænser udeladt af pladshensyn

Effekt af sol, alle 4 lande, output

The GLM Procedure

Class Level Information

Class	Levels	Values
country	4	DK EI SF PL
sol	2	1 0

Number of Observations Used 213

Dependent Variable: vitd Vitamin D

R-Square	Coeff Var	Root MSE	vitd Mean
0.170879	42.33673	18.23878	43.08028

Source	DF	Type I SS	Mean Square	F Value	Pr > F
country	3	10373.99129	3457.99710	10.40	<.0001
country*sol	4	3680.50546	920.12637	2.77	0.0286
sol	0	0.00000	.	.	.

Source	DF	Type III SS	Mean Square	F Value	Pr > F
country	3	8875.628260	2958.542753	8.89	<.0001
country*sol	3	2777.147251	925.715750	2.78	0.0420
sol	1	758.007243	758.007243	2.28	0.1327

Output, fortsat

Estimater for sol-effekt, for hvert land separat

Parameter		Estimate		Standard Error	t Value	Pr > |t|
Intercept		29.43846154	B	3.57691946	8.23	<.0001
country	DK	10.89725275	B	6.04609740	1.80	0.0730
country	EI	24.51778846	B	5.79527187	4.23	<.0001
country	SF	9.81868132	B	6.04609740	1.62	0.1059
country	PL	0.00000000	B	.	.	.
country*sol	DK 1	9.28223443	B	5.68247388	1.63	0.1039
country*sol	DK 0	0.00000000	B	.	.	.
country*sol	EI 1	-9.75625000	B	5.83925939	-1.67	0.0963
country*sol	EI 0	0.00000000	B	.	.	.
country*sol	SF 1	11.79035714	B	5.66367992	2.08	0.0386
country*sol	SF 0	0.00000000	B	.	.	.
country*sol	PL 1	5.20512821	B	4.61778316	1.13	0.2610
country*sol	PL 0	0.00000000	B	.	.	.
sol	1	0.00000000	B	.	.	.
sol	0	0.00000000	B	.	.	.

Konfidensgrænser udeladt af pladshensyn

APPENDIX

med SAS-programbidder svarende til nogle af slides

- Figurer: s. 99, 106, 111
- T-tests mv.: s. 101, 103
- Ensidet ANOVA: s. 104ff
- Tosidet ANOVA: s. 112-114

Indlæsning af Vitamin D data

Slide 3

```
PROC FORMAT;
    VALUE categoryf
        1 = "Girl" 2 = "Woman";
    VALUE sunf
        1 = "Avoid sun" 2 = "Sometimes in sun" 3 = "Prefer sun";
    VALUE countryf
        1 = "DK" 2 = "SF" 4 = "EI" 6 = "PL";
RUN;

DATA vitamind;
DATA a1;
    INFILE "http://staff.pubhealth.ku.dk/~lts/basal/data/VitaminD.txt"
           URL DLM=';' FIRSTOBS=2;
    INPUT country category vitd age bmi sunexp vitdintake;
    FORMAT category categoryf. ;
    FORMAT country countryf. ;
    FORMAT sunexp sunf. ;
    LABEL  vitd = "Vitamin D"
           sunexp="Sun exposure"
           bmi = "BMI";
RUN;

DATA women;
SET vitamind; WHERE category=2;

RUN;
```

Boxplots

Slide 3

```
proc sgplot data=women; where country in (1,4);
    vbox vitd / category=country;
run;
```

Slide 33

```
proc sgplot data=women;
  vbox vitd / category=country;
run;
```

Summary statistics

Slide 4

```
proc means data=women;
  where country in (1,4);
class country;
var vitd;
run;
```

where-sætningen udvælger de to lande, vi vil se på

Uparret T-test

Slide 7-8

```
proc ttest data=women;
     where country in (1,4);
class country;
var vitd;
run;
```

where-sætningen udvælger de to lande, vi vil se på

Dimensionering i SAS

Slide 24-25

```
proc power;
    twosamplemeans test=diff
    meandiff=5
    stddev=20,28
    npergroup=.
    power=0.8,0.9;
run;
```

Bemærk, at man kan foretage adskillige dimensioneringer på samme tid

Nonparametrisk uparret test i SAS

Slide 29-30

Mann-Whitney test eller Kruskal-Wallis test
(approksimation for $n > 25$)

```
proc npar1way wilcoxon data=women;
   where country in (1,4);
class country;
*    exact hl;
var vitd;
run;
```

For små samples kan sætningen `"exact hl;"` give et *eksakt test*, men her ville det tage frygtelig lang tid

Ensidet ANOVA i SAS

Slide 37ff

```
proc glm data=women;
class country;
model vitd=country / solution clparm;
run;
```

Levenes test for identiske spredninger

Slide 45

Benyt `hovtest` i `means`-sætningen:

```
proc glm data=women;
class country;
model vitd=country / solution clparm;
means country /hovtest;
run;
```

Modelkontrolplots for Vitamin D eksemplet

Slide 47

Med ODS-systemet og option plots=all:

```
ods graphics on;
proc glm plots=all data=women;
class country;
model vitd=country / solution clparm;
run;
ods graphics off;
```

Welch test - ANOVA for uens varianser

Slide 49

Option `welch` i `means`-sætningen:

```
proc glm data=women;
class country;
model vitd=country / solution clparm;
means country / welch;
run;
```

Tukey korrektion for vitamin D

Slide 56-57

Option `adjust=tukey` i `lsmeans`-sætningen:

```
proc glm data=women;
class country;
model vitd=country / solution clparm;
LSMEANS country / ADJUST=TUKEY pdiff cl;
run;
```

Non-parametrisk Kruskal-Wallis test

Slide 29

```
proc npar1way wilcoxon data=women;
class country;
var vitd;
run;
```

Bemærk: Man kan også få en eksakt vurdering af teststørrelsen ved at tilføje linien

```
 exact hl;
```

men pas på i tilfælde af store materialer

Optælling af solvaner

Slide 63

```
data women;
set women;

sol=(sunexp in (2,3));
run;

proc freq data=women;
tables country*sol / nopercent nocol;
run;
```

Box-plot, opdelt efter to kategorier

Slide 66

```
proc sgplot data=women;
  where country in (2,6);
vbox vitd / category=country group=sol;
run;
```

- category angiver X-aksen
- group angiver farven

Additiv tosidet ANOVA

dvs. uden interaktion

Slide 69-70

```
ods graphics on;
proc glm plots=all data=women;
   where country in (2,6);
class country sol;
model vitd=country sol / solution clparm;
run;
ods graphics off;
```

Tosidet ANOVA med interaktion

Slide 82, 83

```
ods graphics on;
proc glm plots=all data=women;
  where country in (2,6);
class country sol;
model vitd=country sol country*sol / solution clparm;
run;
ods graphics off;
```

Udregning af predikterede værdier

Slide 86

```
proc glm data=women; where country in (2,6);
class country sol;
model vitd=country sol country*sol / solution clparm p;
ods output PredictedValues=pred;
id country sol;
run;
```

Datasættet pred indeholder så:

- Observed
- Predicted
- Residual

Man kan også benytte en output-sætning
(mere om dette senere)