Faculty of Health Sciences

Basal Statistik

Sammenligning af grupper, Variansanalyse i R

Lene Theil Skovgaard

17. februar 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

Home pages:
`http://publicifsv.sund.ku.dk/~sr/BasicStatistics`
E-mail: `ltsk@sund.ku.dk`

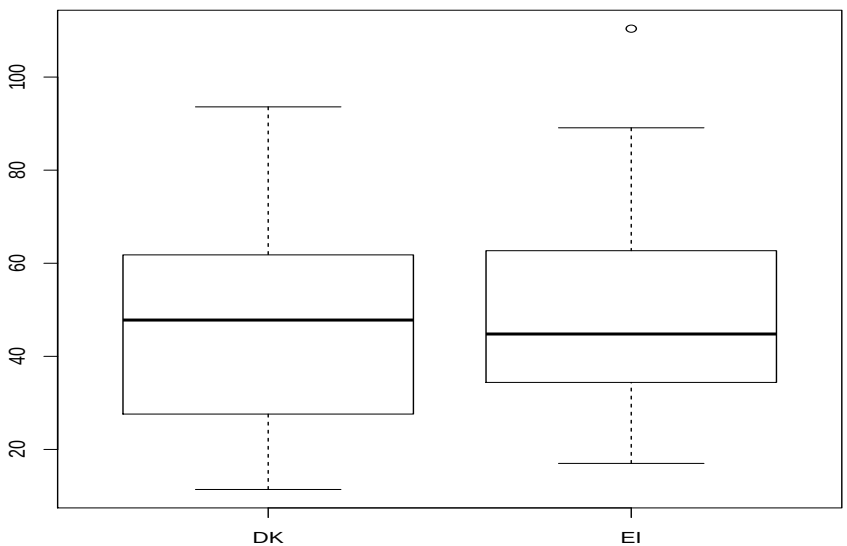
$*$: Siden er lidt teknisk
$\triangle$: Siden gennemgås (nok) ikke

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.

Data framen `vitd` indlæses, se s. 99

```
vitd2 <- subset(vitd, category==2 & country %in% c("DK","EI"))
boxplot(VitaminD~factor(country), data=vitd2)
```



Praktisk håndtering af data

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

- Country (DK, EI)
- Vitamin D status, kaldet `VitaminD` eller bare `VitD` (Serum 25(OH)D, nmol/l)

Summary statistics, opdelt efter land

```
install.packages("doBy")
library(doBy)

vitd2$VitD = vitd2$VitaminD

summaryBy(VitD ~ country, data = vitd2,
  FUN = function(x) { c(length=length(x), mean = mean(x,na.rm=T),
    min = min(x,na.rm=T),max = max(x,na.rm=T),sd = sd(x,na.rm=T))})
```

```
  country VitD.length VitD.mean VitD.min VitD.max  VitD.sd
1      DK          53  47.16604     11.4     93.6 22.78292
2      EI          41  48.00732     17.0    110.4 20.22212
```

Model for uparret sammenligning

Antagelser:

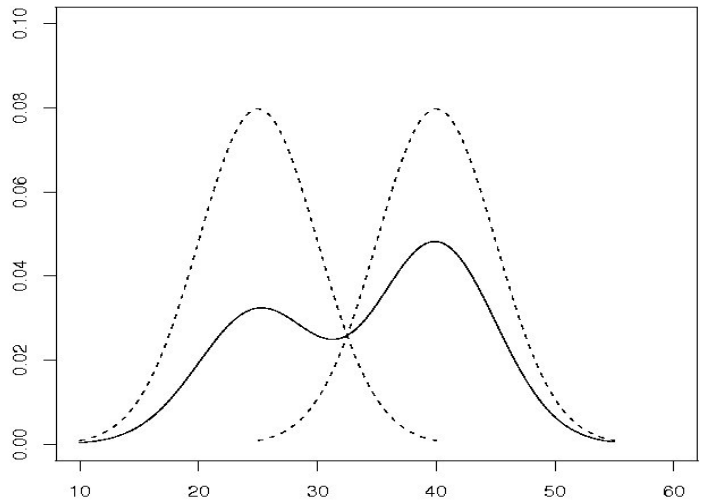
- Alle observationerne er uafhængige
  – kun 1 måling pr. person
  – personerne har ikke noget med hinanden at gøre
- Der er samme spredning(varians) i de to grupper
  – bør checkes/sandsynliggøres
- Observationerne følger en normalfordeling i hver gruppe, med hver deres middelværdi, μ_1 hhv. μ_2

– 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

```
vitd2 <- subset(vitd1, country %in% c("DK","EI"))

t.test(VitaminD ~ country, data=vitd2)
```

giver et T-test *uden* antagelse om ens spredninger (Welch test)

```
Welch Two Sample t-test

data:  vitd2$VitaminD by vitd2$country
t = -0.18922, df = 90.213, p-value = 0.8503
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -9.673908  7.991349
sample estimates:
mean in group DK mean in group EI
        47.16604         48.00732
```

Ens eller uens spredninger?

Check af ens spredninger:

```
> var.test(VitaminD ~ country, data=vitd2)

F test to compare two variances

data:  VitaminD by country
F = 1.2693, num df = 52, denom df = 40, p-value = 0.4357
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.6952686 2.2646896
sample estimates:
ratio of variances
          1.269303
```

T-test *med* antagelse om ens spredninger:

```
> t.test(VitaminD ~ country, data=vitd2, var.equal=TRUE)

Two Sample t-test

data:  VitaminD by country
t = -0.18634, df = 92, p-value = 0.8526
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -9.807835  8.125276
```

Kommentarer til output

- Først udføres et Welch test *uden* antagelse om ens spredninger.
- Derefter checkes om vi kan tillade os at antage ens spredninger. Dette gøres her med et F-test (der er flere muligheder, som vi senere skal se), og vi finder P=0.44, som siger, at der ikke er noget, der taler imod ens spredninger.
- Endelig T-testet under *antagelse* om ens spredninger, specificeret ved option `var.equal=T`.

Under alle omstændigheder er $P = 0.85$, dvs.
vi kan ikke afvise, at middelværdierne er ens.

∗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):

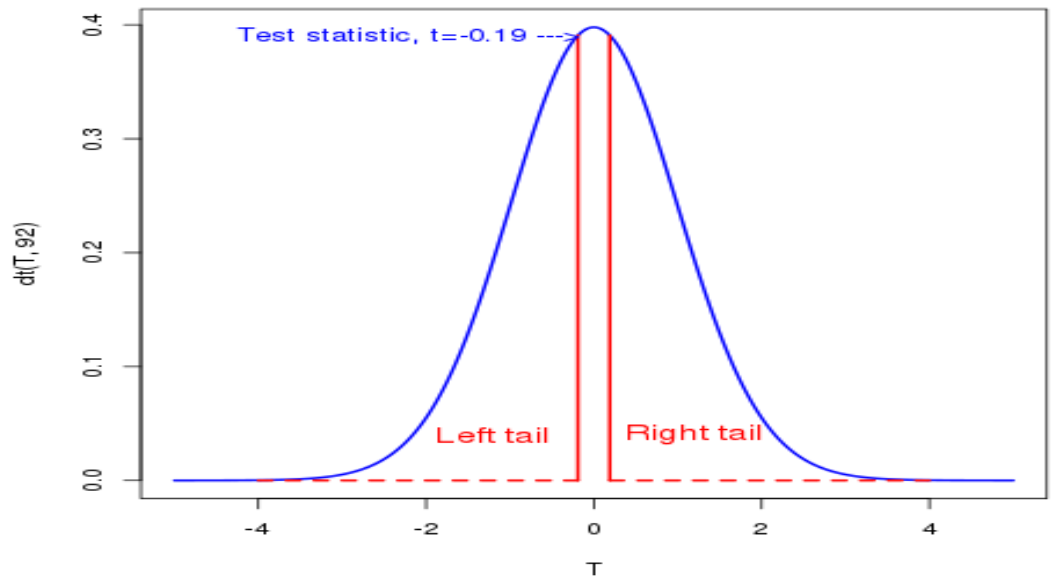
1. Fordel tilfældigt 53 i en gruppe, 41 i en anden, $\Longrightarrow t_1$
2. Fordel tilfældigt 53 i en gruppe, 41 i en anden, $\Longrightarrow t_2$
3. Fordel tilfældigt 53 i en gruppe, 41 i en anden, $\Longrightarrow t_3$
   osv. osv.

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.

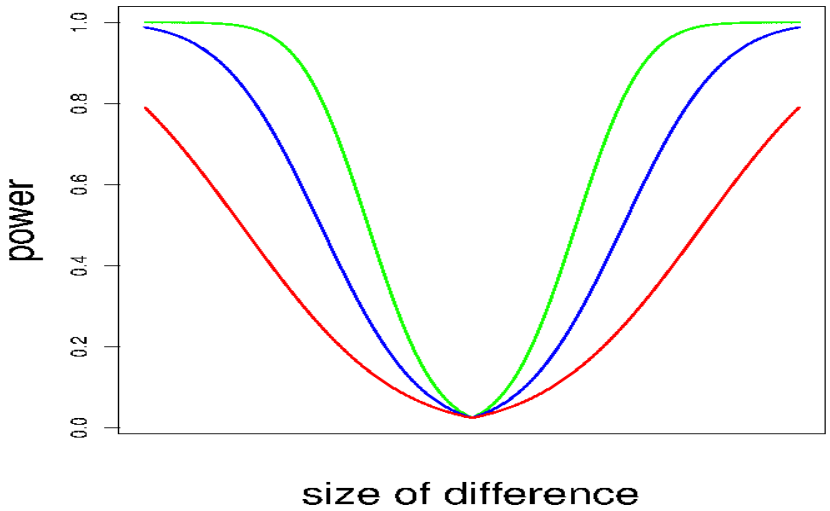
Styrke

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 benytter vi pakken `TeachingDemos` med funktionen `sigma.test`, som kan give os konfidensgrænser på de spredningsestimater, vi allerede har fra vores data:

```
install.packages("TeachingDemos")
library(TeachingDemos)

> sqrt(sigma.test(vitd2$VitaminD[vitd2$country=="DK"], sigma = 1)$conf.int)
[1] 19.12291 28.18872
attr(,"conf.level")
[1] 0.95
> sqrt(sigma.test(vitd2$VitaminD[vitd2$country=="EI"], sigma = 1)$conf.int)
[1] 16.60262 25.87426
attr(,"conf.level")
[1] 0.95
```

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 R kan man benytte

```
power.t.test(n = NULL, delta = 5,
             sd = 20, sig.level = 0.05,
             power = 0.8,
             type = c("two.sample"),
             alternative = c("two.sided"))
```

Tilsyneladende kan man kun udføre en enkelt dimensionering ad gangen, så vi sammenfatter 4 forskellige af disse på næste side.

Output fra dimensionering

```
p1=power.t.test(n = NULL, delta = 5, sd = 20, power = 0.8)$n
p2=power.t.test(n = NULL, delta = 5, sd = 28, power = 0.8)$n
p3=power.t.test(n = NULL, delta = 5, sd = 20, power = 0.9)$n
p4=power.t.test(n = NULL, delta = 5, sd = 28, power = 0.9)$n
antal=c(p1,p2,p3,p4)

delta=rep(5,4)
sd=rep(c(20,28),2)
power=c(rep(0.8,2),rep(0.9,2))
```

```
> cbind(delta, sd, power, antal)
     delta sd power    antal
[1,]     5 20   0.8 252.1281
[2,]     5 28   0.8 493.2439
[3,]     5 20   0.9 337.2008
[4,]     5 28   0.9 659.9874
```

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, fremkommet med 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 (se s. 29-30)
- 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

```
wilcox.test(VitaminD ~ country, data=vitd2)
```

For små samples vil man med fordel kunne anvende
den *eksakte* udgave:

```
install.packages("exactRankTests")
library(exactRankTests)

wilcox.exact(VitaminD ~ country, exact=T, data=vitd2)
```

△Output fra Mann-Whitney test

Approksimativt:

```
> wilcox.test(VitaminD ~ country, data=vitd2)

Wilcoxon rank sum test with continuity correction

data:  VitaminD by country
W = 1071.5, p-value = 0.912
alternative hypothesis: true location shift is not equal to 0
```

Eksakt:

```
> wilcox.exact(VitaminD ~ country, exact=T, data=vitd2)

Exact Wilcoxon rank sum test

data:  VitaminD by country
W = 1071.5, p-value = 0.9109
alternative hypothesis: true mu is not equal to 0
```

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:

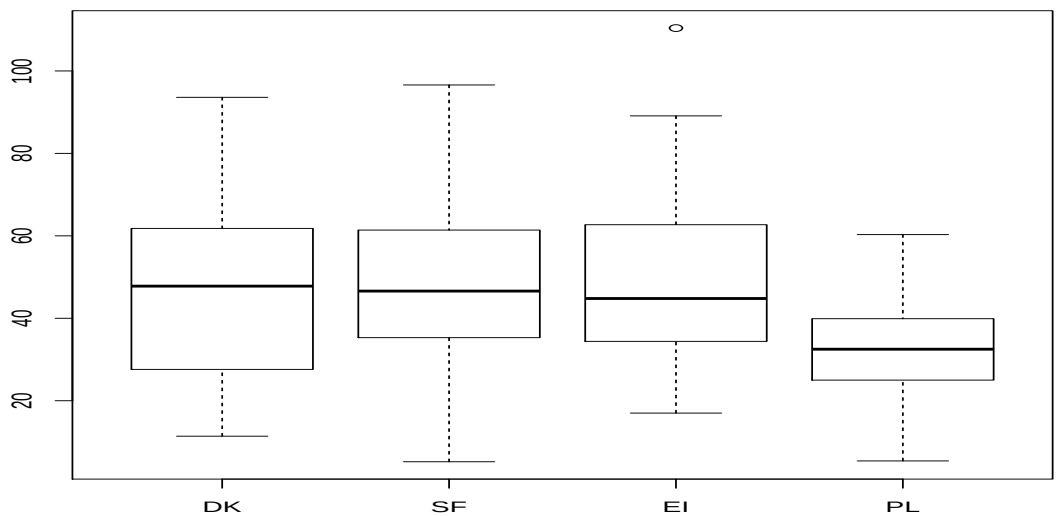
(-12.71 cm^3, 13.19 cm^3)

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

```
boxplot(VitaminD~factor(country), data=vitd1)
```



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:

```
  country VitD.length VitD.mean  VitD.sd VitD.var
1      DK          53  47.16604 22.78292 519.0615
2      SF          54  47.99074 18.72471 350.6148
3      EI          41  48.00732 20.22212 408.9342
4      PL          65  32.56154 12.46448 155.3633
```

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 `aov`, og i stedet bruge den mere generelle `lm`, her suppleret med test for varianshomogenitet:

```
model1 = lm(VitaminD ~ relevel(country, ref="SF"),
         data=vitd1, na.action=na.exclude)

bartlett.test(VitaminD ~ country, data=vitd1)

install.packages("lawstat")
library(lawstat)

levene.test(vitd1$VitaminD, vitd1$country)
```

Bemærkninger til kode på forrige side

- I `model1` vælges `SF` som reference, for at få samme output som i SAS. I R vælges ellers pr. default det første land i alfabetet.
- Bartletts test undersøger, om der er signifikant forskel på varianserne for de involverede lande. Testet er ret sensibelt mht normalfordelingsantagelsen.
- Levenes test er også et test for varianshomogenitet, og det er mere robust overfor normalfordelingsantagelsen.
  Det kræver dog den ekstra pakke `lawstat`.

Output fra ensidet ANOVA

Variansanalysetabellen (den mindre brugbare del):

```
> anova(model1)
Analysis of Variance Table

Response: VitaminD
                              Df Sum Sq Mean Sq F value    Pr(>F)
relevel(country, ref = "SF")   3  10374  3458.0  10.055 3.227e-06 ***
Residuals                    209  71874   343.9
```

Denne giver testet for ens middelværdier ($P = 3.227 \times 10^{-6}$), men kan ellers ikke kan bruges til så forfærdeligt meget.

Output, fortsat

Nu den mere brugbare del:

```
> summary(model1)

Call:
lm(formula = VitaminD ~ relevel(country, ref = "SF"), data = vitd1,
    na.action = na.exclude)

Coefficients:                    Estimate Std. Error t value Pr(>|t|)
(Intercept)                      47.99074    2.52358  19.017  < 2e-16 ***
relevel(country, ref = "SF")DK  -0.82470    3.58568  -0.230    0.818
relevel(country, ref = "SF")EI   0.01658    3.84138   0.004    0.997
relevel(country, ref = "SF")PL -15.42920    3.41455  -4.519 1.04e-05 ***

Residual standard error: 18.54 on 209 degrees of freedom
Multiple R-squared:  0.1261, Adjusted R-squared:  0.1136
F-statistic: 10.06 on 3 and 209 DF,  p-value: 3.227e-06

> confint(model1)
                                    2.5 %    97.5 %
(Intercept)                      43.015807 52.965675
relevel(country, ref = "SF")DK  -7.893431  6.244025
relevel(country, ref = "SF")EI  -7.556236  7.589389
relevel(country, ref = "SF")PL -22.160583 -8.697822
```

med fortolkning på de følgende sider

Bemærkninger til output, I

- Variansanalysetabellen s. 39 giver F-testet for test af ens middelværdier i de 4 grupper, med tilhørende P-værdi. Her er $P = 3.227 \times 10^{-6}$, dvs. de 4 middelværdier kan *ikke* antages at være ens.
- Spredningen $\sigma \sim s$, `Residual standard error`:
  Den poolede variation for de 4 grupper=lande, ses s. 40 at være 18.54

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 (normalt første gruppe, alfabetisk eller numerisk), men her eksplicit valgt som Finland (SF)
- Estimatet ud for f.eks. `country DK` er forskellen i niveau mellem DK og SF (referencegruppen)

Modelantagelse 1: Uafhængighed

Dette er noget, man skal vide

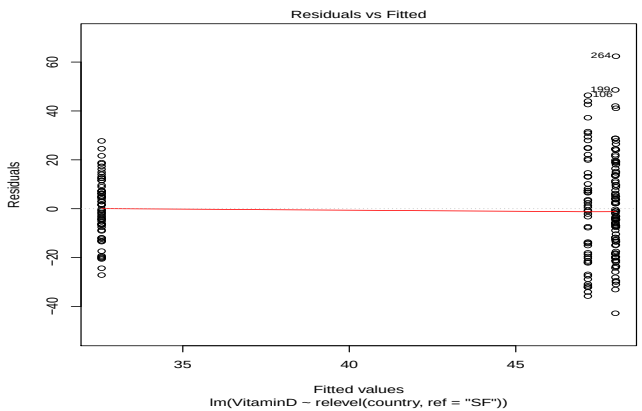
- ingen tvillinger, søskende etc.
- kun *en* observation for hver person
  (ellers hører det hjemme under emnet
  "Korrelerede målinger", kursets sidste emne)

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
  (Bartletts eller Levenes test, se side 37 og 45)
- Residualer tegnet op mod predikterede
  (=forventede=fittede) værdier, skal være *jævnt*



Dette er en del
af den halvautomatiske
grafik i R, se. s. 47

Levenes test for identiske spredninger

Vi har allerede set, at Polen måske har mindre spredning end de andre tre lande - men det *kunne* jo være en tilfældighed.
Her ser vi på Levenes test (se også side 37):

```
> levene.test(vitd1$VitaminD, vitd1$country)

modified robust Brown-Forsythe Levene-type test based on the absolute
deviations from the median

data:  vitd1$VitaminD
Test Statistic = 6.4484, p-value = 0.0003394
```

Ved sammenligning af de 4 variansestimater fås en P-værdi på P=0.0003, og altså kraftig signifikans!

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

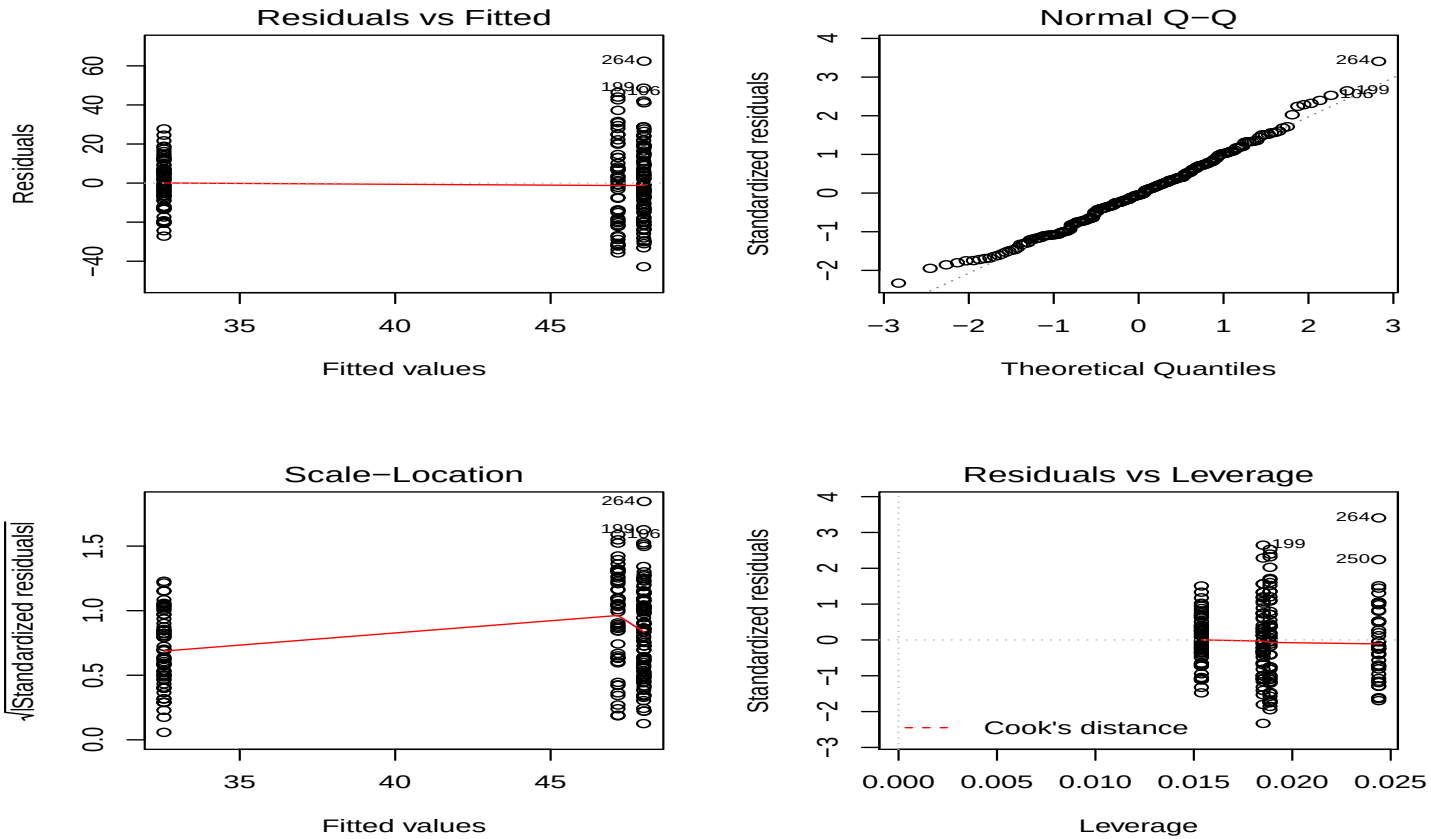
Modelantagelse 3: Normalfordelingsantagelsen

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

Halvautomatisk plot i R, se s. 107



Bemærkninger til Diagnostics Panel

Foreløbig beskæftiger vi os kun med de øverste plots på s. 47

- Figuren til venstre: residualer mod predikterede værdier
  Har de samme spredning?
  Næh, den stiger vist lidt med den predikterede værdi
- Figuren til højre: Fraktildiagram af residualerne: Ser de normalfordelte ud?
  Næsten, dog lidt "hængekøje"=skævhed=hale mod højre

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, se s. 108

```
> welch.test(VitaminD ~ country, data=vitd1)

  Welch's Heteroscedastic F Test
-------------------------------------------------------
  data : VitaminD and country

  statistic  : 14.64244
  num df     : 3
  denom df   : 102.3286
  p.value    : 5.270896e-08

  Result     : Difference is statistically significant.
-------------------------------------------------------
```

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

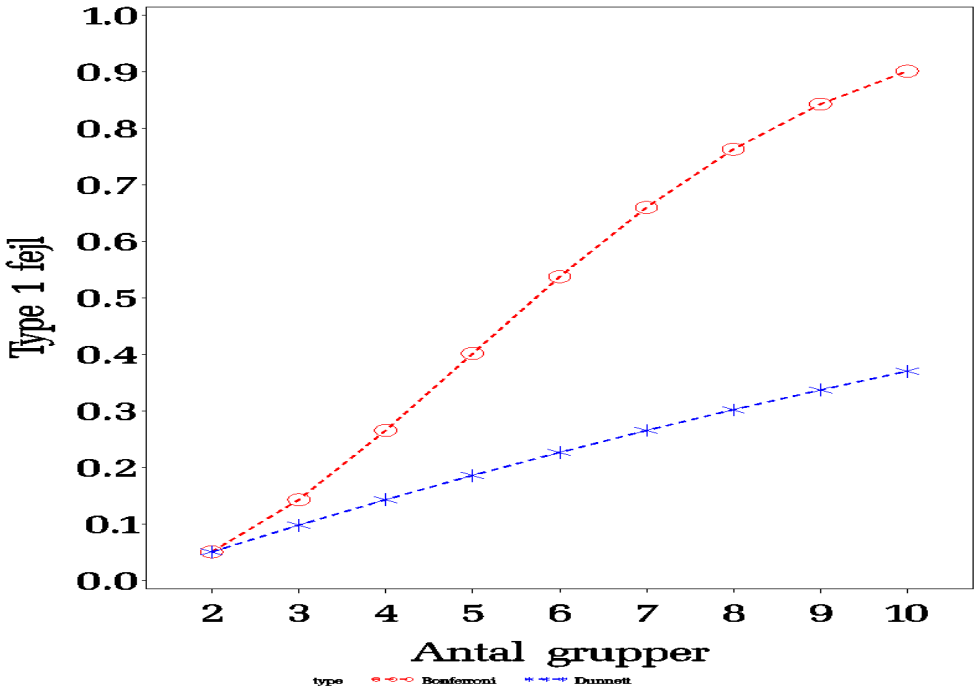
Hvis man sammenligner k grupper (lande) parvist, er der $m = k(k-1)/2$ mulige test, hver med signifikansniveau $\alpha = 0.05$.

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
  - 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:

Her er man nødt til at starte med aov-funktionen:

```
model2 = aov(VitaminD ~ relevel(country, ref="SF"),
         data=vitd1, na.action=na.exclude)

TukeyHSD(model2)
```

hvorefter man kan skrive

```
model2 = aov(VitaminD ~ relevel(country, ref="SF"),
         data=vitd1, na.action=na.exclude)

TukeyHSD(model2)
```

Tukey korrektion for sammenlingning af lande

```
> TukeyHSD(model2)
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = VitaminD ~ relevel(country, ref = "SF"),
     data = vitd1, na.action = na.exclude)

$`relevel(country, ref = "SF")`
               diff       lwr       upr     p adj
DK-SF  -0.82470300 -10.11096  8.461553 0.9957015
EI-SF   0.01657633  -9.93190  9.965052 1.0000000
PL-SF -15.42920228 -24.27228 -6.586124 0.0000612
EI-DK   0.84127934  -9.14762 10.830178 0.9963264
PL-DK -14.60449927 -23.49303 -5.715969 0.0001838
PL-EI -15.44577861 -25.02407 -5.867491 0.0002529
```

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 PL=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 `lm` (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

```
> kruskal.test(VitaminD ~ country, data=vitd1)

Kruskal-Wallis rank sum test

data:  VitaminD by country
Kruskal-Wallis chi-squared = 26.682, df = 3, p-value = 6.864e-06
```

Bemærk:

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

  Det kan tage forfærdeligt lang tid

Sammenligning af Finland og Polen

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

```
vitd3 <- subset(vitd1, country %in% c("SF","PL"))
t.test(VitaminD ~ country, data=vitd3)
```

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. 111) 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. 111) giver:

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

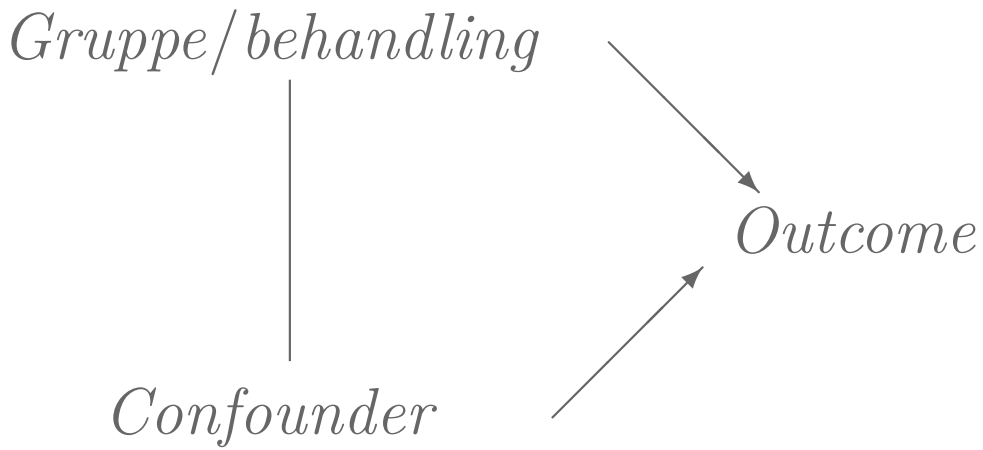
Kan denne forskel i sol-præferencer være forklaringen på forskellen i Vitamin-D?

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

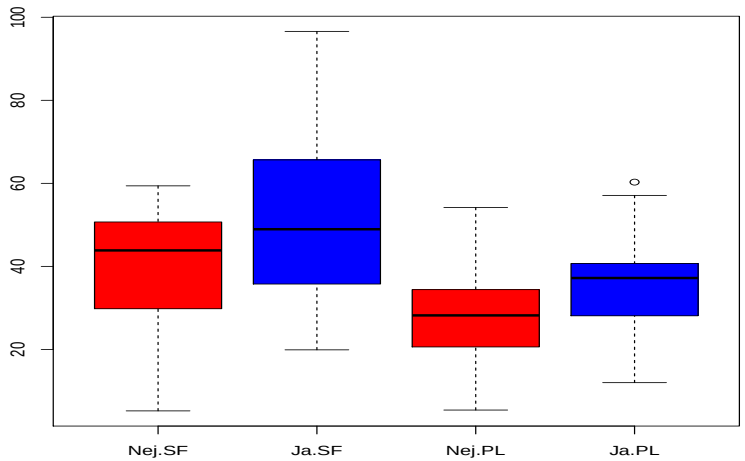
- har en effekt på outcome
- er relateret til gruppen
  som en direkte konsekvens af denne

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

Er solvaner en mediator?

Har solvanerne overhovedet en betydning?

Se kode s. 112



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

```
model2 = lm(VitaminD ~ relevel(country, ref="SF") + sol,
         data=vitd3, na.action=na.exclude)
```

som giver outputtet

```
> anova(model2)
Analysis of Variance Table

Response: VitaminD
                              Df  Sum Sq Mean Sq F value    Pr(>F)
relevel(country, ref = "SF")   1  7021.8  7021.8 30.2441 2.301e-07 ***
sol                            1  1594.1  1594.1  6.8662  0.009959 **
Residuals                    116 26931.7   232.2
---
```

med fortolkning s. 71

Output fra 2-sidet ANOVA af Vitamin D, fortsat

```
> summary(model2)

Call:
lm(formula = VitaminD ~ relevel(country, ref = "SF") + sol, data = vitd3,
    na.action = na.exclude)

Coefficients:
                              Estimate Std. Error t value Pr(>|t|)
(Intercept)                     42.187      3.034   13.90  < 2e-16 ***
relevel(country, ref = "SF")PL  -14.327      2.837   -5.05 1.66e-06 ***
solJa                             7.835      2.990    2.62  0.00996 **

Residual standard error: 15.24 on 116 degrees of freedom
Multiple R-squared:  0.2424, Adjusted R-squared:  0.2293
F-statistic: 18.56 on 2 and 116 DF,  p-value: 1.019e-07
```

```
> confint(model2)
                                    2.5 %    97.5 %
(Intercept)                      36.17819 48.196315
relevel(country, ref = "SF")PL -19.94550 -8.707574
solJa                            1.91274 13.756684
```

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

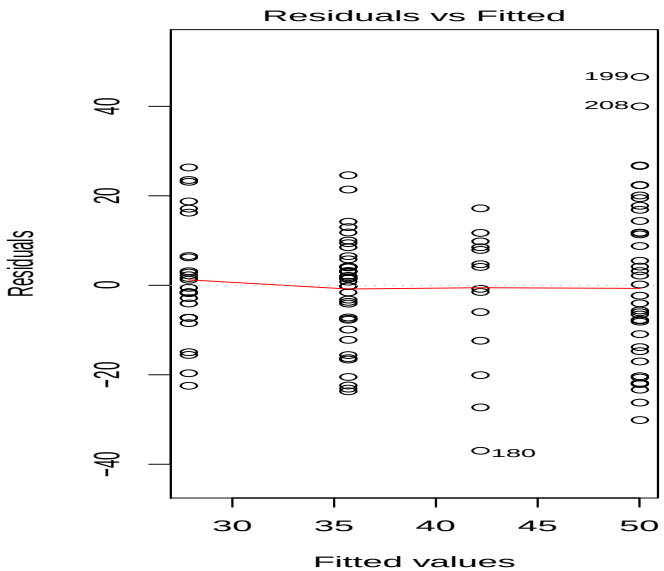
Model indeholder	estimat	CI
kun country (T-test)	15.43	(9.51, 21.35)
solvaner og land	14.33	(8.71, 19.95)
SF vs. PL, solelskere	16.40	(9.37, 23.43)
SF vs. PL, solhadere	9.82	(-0.47, 20.11)

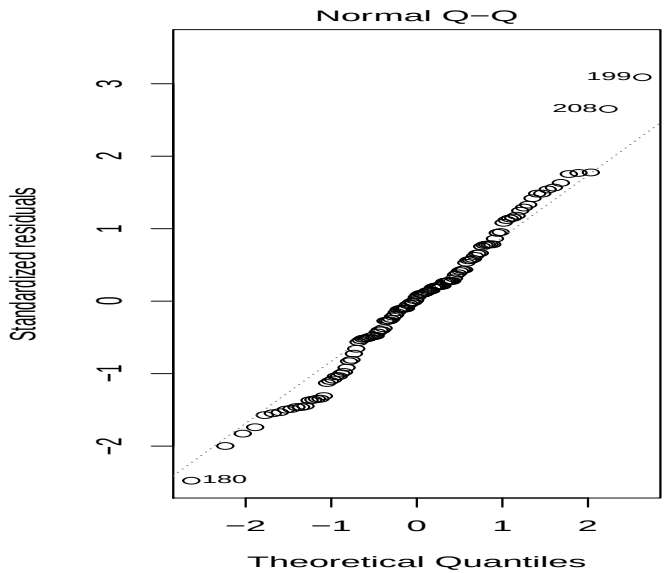
Foreløbig har vi antaget, at forskellen på landene er den samme, uanset solvaner, dvs. at der ikke er nogen interaktion.

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

Se kode s. 114



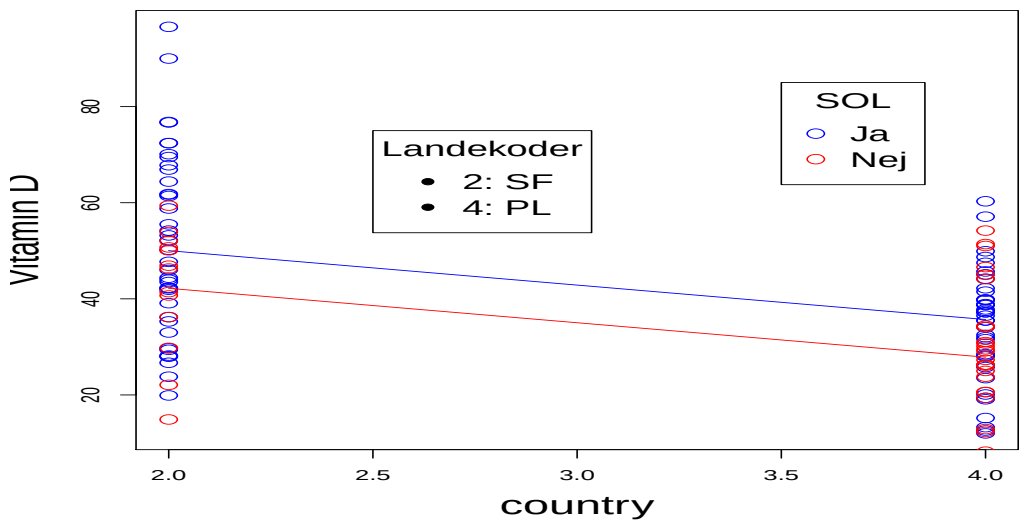


Intet specielt mønster her

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 (se kode s. 115)



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

Mulig interaktion?

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

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

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):

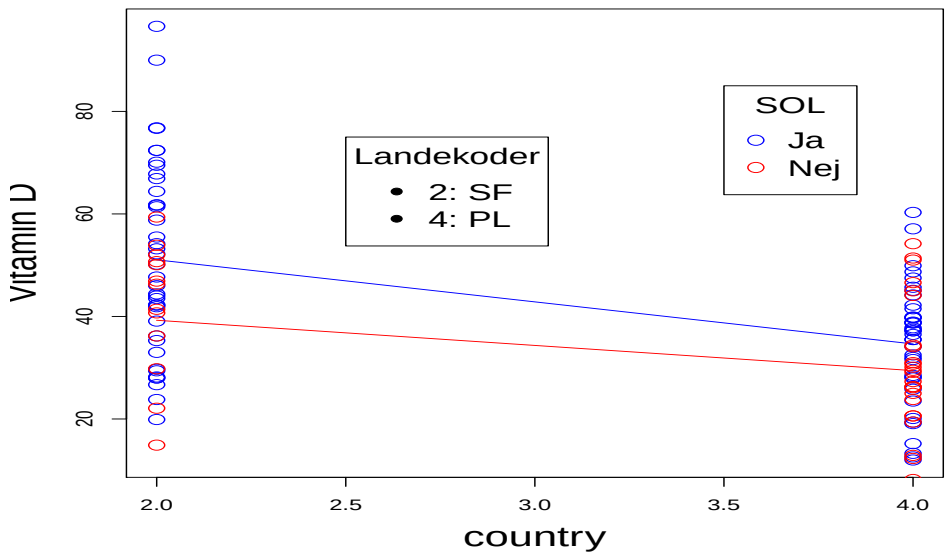
Vitamin D for solelskere vs. solhadere:

- 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



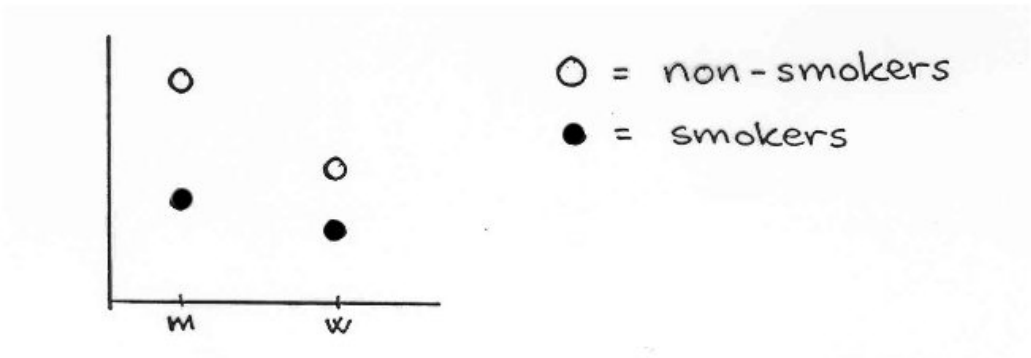
Er der forskel på sol-effekterne i de to lande?

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

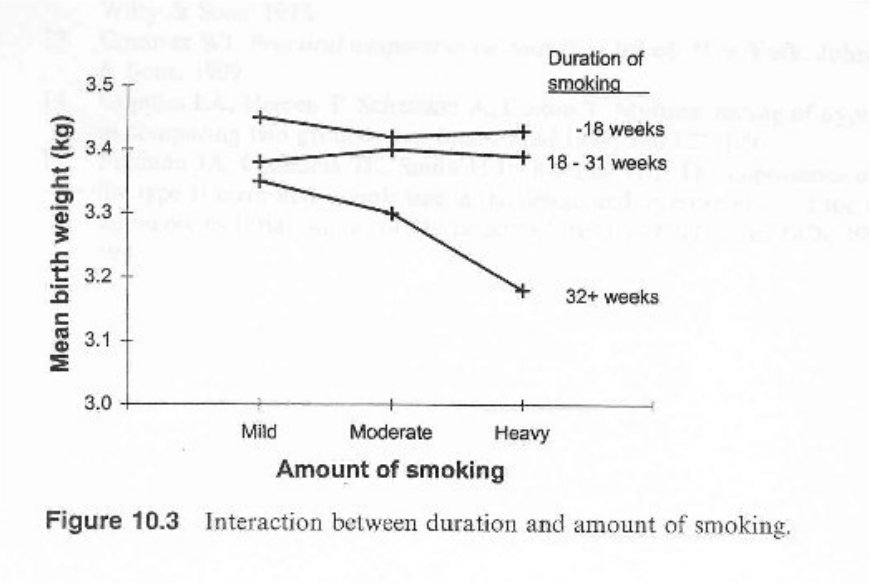


Figure 10.3 Interaction between duration and amount of smoking.

- 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:

```
model3 = lm(VitaminD ~ relevel(country,ref="SF")*sol,
         data=vitd3, na.action=na.exclude)
```

som giver outputtet

```
> anova(model3)
Analysis of Variance Table

Response: VitaminD
                                  Df  Sum Sq Mean Sq F value    Pr(>F)
relevel(country, ref = "PL")       1  7021.8  7021.8 30.2872 2.291e-07 ***
sol                                1  1594.1  1594.1  6.8760  0.009919 **
relevel(country, ref = "PL"):sol   1   270.1   270.1  1.1652  0.282652
Residuals                        115 26661.6   231.8
---
```

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

Output vedr. interaktion, fortsat

```
> summary(model3)

Call:
lm(formula = VitaminD ~ relevel(country, ref = "PL") * sol, data = vitd3,
    na.action = na.exclude)

Coefficients:
                                    Estimate Std. Error t value Pr(>|t|)
(Intercept)                           29.438      2.986   9.858   <2e-16 ***
relevel(country, ref = "PL")SF         9.819      5.047   1.945   0.0542 .
solJa                                  5.205      3.855   1.350   0.1796
relevel(country, ref = "PL")SF:solJa   6.585      6.101   1.079   0.2827
---

> confint(model3)
                                        2.5 %   97.5 %
(Intercept)                         23.523532 35.35339
relevel(country, ref = "PL")SF      -0.179374 19.81674
solJa                               -2.431013 12.84127
relevel(country, ref = "PL")SF:solJa -5.498914 18.66937
```

Fortolkningen af estimaterne følger på de næste sider.

Fortolkning af estimater

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

- `(Intercept)`=29.438:
  Det estimerede niveau (her blot gennemsnittet) af vitamin D for referencegruppen, dvs. solhadere fra Polen.
- `relevel(country, ref = "PL")SF`=9.819:
  Finlands forspring frem for Polen for sol-referencegruppen, dvs. for solhadere
- `solJa`=5.205:
  Effekten af at kunne lide sol vs. at hade den, for country-referencegruppen, dvs. for polakker

Estimater, fortsat

- `relevel(country, ref = "PL")SF:solJa`=6.585:
  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="PL", sol="Nej"`

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 + 5.21 + 9.82 + 6.59 =51.05	29.44 + 5.21 =34.64
nej	29.44 + 9.82 =39.26	29.44

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

Estimationsvenlig kode

Separate effekter af land: Udelad country:

```
model3.land = lm(VitaminD ~ sol + country:sol,
          data=vitd3, na.action=na.exclude)
```

```
> summary(model3.land)

Call:
lm(formula = VitaminD ~ sol + country:sol, data = vitd3, na.action = na.exclude)

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)        39.257      4.069   9.647  < 2e-16 ***
solJa              11.790      4.728   2.494   0.0141 *
solNej:countryPL   -9.819      5.047  -1.945   0.0542 .
solJa:countryPL   -16.404      3.426  -4.787 5.07e-06 ***
---

> confint(model3.land)
                       2.5 %    97.5 %
(Intercept)        31.196453 47.317833
solJa               2.424682 21.156032
solNej:countryPL -19.816737  0.179374
solJa:countryPL  -23.191061 -9.616760
```

Estimationsvenlig kode, II

Separate effekter af sol: Udelad `sol`:

```
model3.sol = lm(VitaminD ~ country + country:sol,
         data=vitd3, na.action=na.exclude)
```

```
> summary(model3.sol)

Call:
lm(formula = VitaminD ~ country + country:sol, data = vitd3,
    na.action = na.exclude)

Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)      39.257      4.069   9.647   <2e-16 ***
countryPL        -9.819      5.047  -1.945   0.0542 .
countrySF:solJa  11.790      4.728   2.494   0.0141 *
countryPL:solJa   5.205      3.855   1.350   0.1796
---
```

```
> confint(model3.sol)
                     2.5 %    97.5 %
(Intercept)      31.196453 47.317833
countryPL       -19.816737  0.179374
countrySF:solJa   2.424682 21.156032
countryPL:solJa  -2.431013 12.841269
```

Fokus på effekt af sol

Soldyrkere vs. solhadere, stadig kun Finland og Polen:

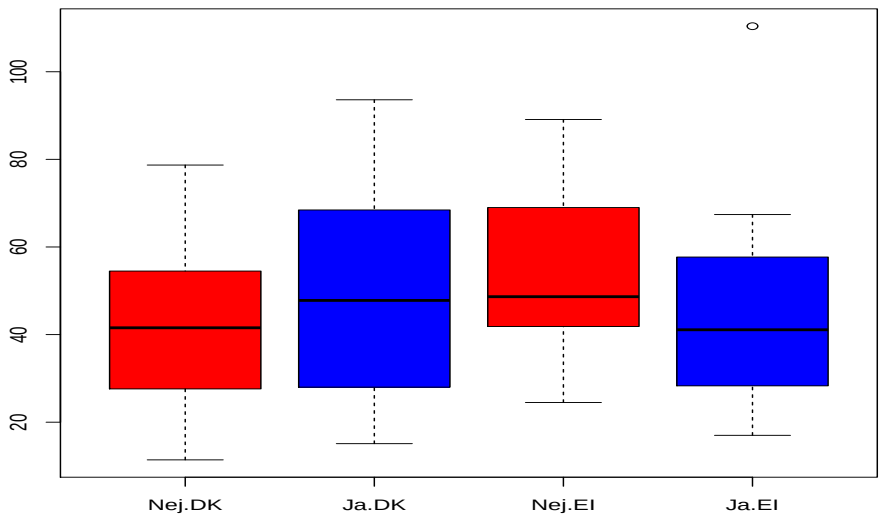
Model indeholder	estimat	CI
kun solvaner (T-test)	10.07	(3.63, 16.51)
solvaner og land	7.83	(1.91, 13.76)
solvaner, kun SF	11.79	(0.48, 23.10)
solvaner, kun PL	5.21	(-1.01, 11.42)

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

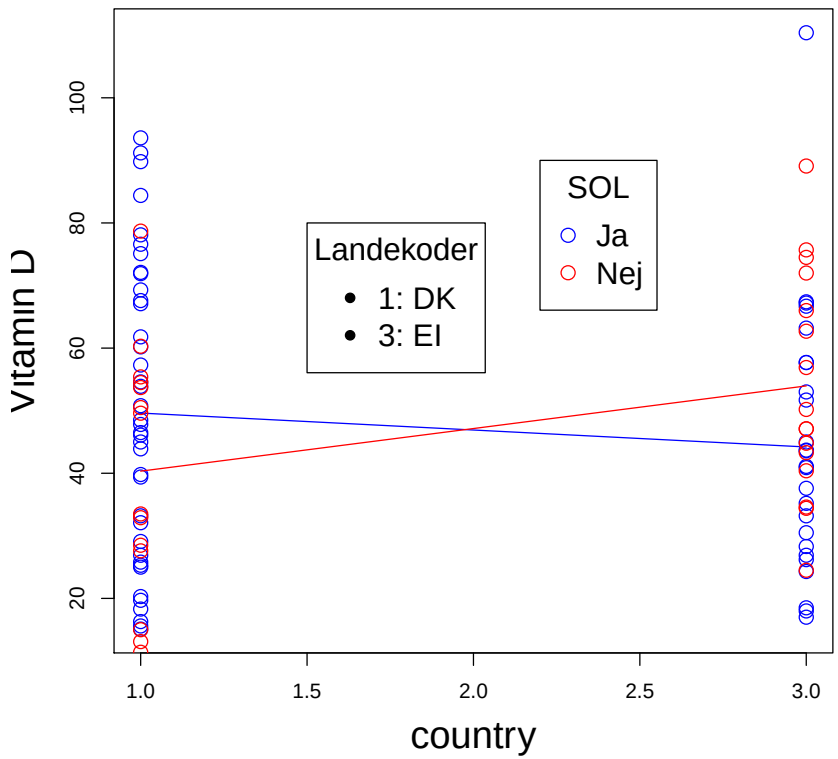
(ganske som Finland vs. Polen, s. 66 og 112)



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:

```
> model4 = lm(VitaminD ~ relevel(country, ref="EI")*sol,
+           data=vitd4, na.action=na.exclude)

> summary(model4)

Call:
lm(formula = VitaminD ~ relevel(country, ref = "EI") * sol, data = vitd4,
    na.action = na.exclude)

Coefficients:
                                    Estimate Std. Error t value Pr(>|t|)
(Intercept)                           53.956      5.370  10.047 2.28e-16 ***
relevel(country, ref = "EI")DK       -13.621      7.862  -1.733   0.0866 .
solJa                                 -9.756      6.878  -1.419   0.1595
relevel(country, ref = "EI")DK:solJa   19.038      9.597   1.984   0.0503 .
---

> confint(model4)
                                        2.5 %    97.5 %
(Intercept)                          43.28688377 64.625616
relevel(country, ref = "EI")DK      -29.23888864  1.997817
solJa                               -23.41970551  3.907206
relevel(country, ref = "EI")DK:solJa  -0.02691043 38.103879
```

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

Med den estimationsvenlige kode (svarende til s. 88) får vi variansanalysetabellen

```
model4.sol = lm(VitaminD ~ country + country:sol,
         data=vitd4, na.action=na.exclude)
```

fås variansanalysetabellen

```
> anova(model4.sol)
Analysis of Variance Table

Response: VitaminD
            Df Sum Sq Mean Sq F value Pr(>F)
country      1     16   16.36  0.0355 0.8511
country:sol  2   1816  908.12  1.9679 0.1457
Residuals   90  41532  461.47
```

Testet `country:sol` er her et test for den *samlede* effekt af `sol`, for begge lande på en gang.

△Fokus på effekt af sol, II

men til gengæld får vi de lande-specifikke effekter af sol:

```
> summary(model4.sol)

Call:
lm(formula = VitaminD ~ country + country:sol, data = vitd4,
    na.action = na.exclude)

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)       40.336      5.741   7.026 3.94e-10 ***
countryEI         13.621      7.862   1.733   0.0866 .
countryDK:solJa    9.282      6.693   1.387   0.1689
countryEI:solJa   -9.756      6.878  -1.419   0.1595
---
```

Effekt af sol, alle 4 lande

Er der interaktion?

```
model5 = lm(VitaminD ~ relevel(country,ref="PL")*sol,
         data=vitd1, na.action=na.exclude)
```

```
> anova(model5)
Analysis of Variance Table

Response: VitaminD
                                 Df Sum Sq Mean Sq F value    Pr(>F)
relevel(country, ref = "PL")      3  10374  3458.0 10.3952 2.134e-06 ***
sol                               1    903   903.4  2.7156    0.1009
relevel(country, ref = "PL"):sol  3   2777   925.7  2.7828    0.0420 *
Residuals                       205  68194   332.7
---
```

Her er estimaterne ret svære at fortolke, som vi tidligere så, og vi fokuserer her på test af interaktion.

Effekt af sol, alle 4 lande, II

Estimater for sol-effekt, for hvert land separat

Her bruges i stedet den estimationsvenlige kode:

```
model5 = lm(VitaminD ~ relevel(country,ref="PL") +
                       relevel(country,ref="PL"):sol,
```

```
> summary(model5)

Call:
lm(formula = VitaminD ~ relevel(country, ref = "PL") + relevel(country,
    ref = "PL"):sol, data = vitd1, na.action = na.exclude)

Coefficients:                          Estimate Std. Error t value Pr(>|t|)
(Intercept)                              29.438      3.577   8.230 2.15e-14 ***
relevel(country, ref = "PL")DK           10.897      6.046   1.802   0.0730 .
relevel(country, ref = "PL")SF            9.819      6.046   1.624   0.1059
relevel(country, ref = "PL")EI           24.518      5.795   4.231 3.51e-05 ***
relevel(country, ref = "PL")PL:solJa      5.205      4.618   1.127   0.2610
relevel(country, ref = "PL")DK:solJa      9.282      5.682   1.633   0.1039
relevel(country, ref = "PL")SF:solJa     11.790      5.664   2.082   0.0386 *
relevel(country, ref = "PL")EI:solJa     -9.756      5.839  -1.671   0.0963 .
```

Konfidensgrænser udeladt af pladshensyn

APPENDIX

med R-programbidder svarende til nogle af slides

- Figurer: s. 100, 107, 112
- T-tests mv.: s. 102, 104
- Ensidet ANOVA: s. 105ff
- Tosidet ANOVA: s. 113-115

Indlæsning af Vitamin D data

Slide 3

```
vitd <-
read.table("vitamin.txt", header=T)

vitd$country <- factor(vitd$land, levels=c(1,2,4,6),
                       labels=c("DK", "SF", "EI", "PL"))
vitd$sunexp <- factor(vitd$sun, levels=1:3,
                       labels=c("Avoid sun", "Sometimes in sun", "Prefer sun"))
vitd$log2vitd <- log2(vitd$VitaminD)
vitd$lvitdintake <- log2(vitd$vitdintake)
vitd$sol <- factor(vitd$sunexp,
                   levels=c("Avoid sun", "Sometimes in sun", "Prefer sun"),
                   labels=c("Nej", "Ja", "Ja"))

vitd1 <- subset(vitd, category==2)
```

Boxplots

Slide 3

```
vitd2 <- subset(vitd1, country %in% c("DK","EI"))

boxplot(VitaminD~factor(country), data=vitd2)
```

Slide 33

```
boxplot(VitaminD~factor(country), data=vitd1)
```

Summary statistics

Slide 4

```
install.packages("doBy")
library(doBy)

vitd2$VitD = vitd2$VitaminD

summaryBy(VitD ~ country, data = vitd2,
  FUN = function(x) { c(length=length(x),
        mean = mean(x,na.rm=T),
        min = min(x,na.rm=T),
        max = max(x,na.rm=T),
        sd = sd(x,na.rm=T))})
```

Uparret T-test

Slide 7-8

```
t.test(VitaminD ~ country, data=vitd2)
t.test(VitaminD ~ country, data=vitd2, var.equal=TRUE)

var.test(VitaminD ~ country, data=vitd2)
```

Dimensionering i R

Slide 23-25

```
install.packages("TeachingDemos")
library(TeachingDemos)

sqrt(sigma.test(vitd2$VitaminD[vitd2$country=="DK"], sigma = 1)$conf.int)
sqrt(sigma.test(vitd2$VitaminD[vitd2$country=="EI"], sigma = 1)$conf.int)

power.t.test(n = NULL, delta = 5, sd = 20, sig.level = 0.05,
             power = 0.8, type = c("two.sample"),
             alternative = c("two.sided"))
```

Se også s. 25

Nonparametrisk uparret test i SAS

Slide 29-30

Mann-Whitney test

```
wilcox.test(VitaminD ~ country, data=vitd2)

install.packages("exactRankTests")
library(exactRankTests)

wilcox.exact(VitaminD ~ country, exact=T, data=vitd2)
```

For små samples er det eksakte test påkrævet

Ensidet ANOVA i R

Slide 37ff

```
model1 = lm(VitaminD ~ relevel(country, ref="SF"),
         data=vitd1, na.action=na.exclude)
anova(model1)
summary(model1)
confint(model1)

bartlett.test(VitaminD ~ country, data=vitd1)
```

Levenes test for identiske spredninger

Slide 45

```
install.packages("lawstat")
library(lawstat)

levene.test(vitd1$VitaminD, vitd1$country)
levene.test(vitd1$VitaminD, vitd1$country,location="mean")
```

Modelkontrolplots for Vitamin D eksemplet

Slide 44

Den første modelkontrolfigur (`which=1`), giver residualer mod predikterede værdier.

```
plot(model1,which=1)
```

Slide 47

To-gange-to layout med alle de automatiske modelkontrolfigurer:

```
par(mfrow=c(2,2))
plot(model1)
```

Welch test - ANOVA for uens varianser

Slide 49

```
install.packages("onewaytests")
library(onewaytests)

welch.test(VitaminD ~ country, data=vitd1)
```

Tukey korrektion for vitamin D

Slide 56-57

```
model2 = aov(VitaminD ~ relevel(country, ref="SF"),
         data=vitd1, na.action=na.exclude)

TukeyHSD(model2)
```

Non-parametrisk Kruskal-Wallis test

Slide 61

```
kruskal.test(VitaminD ~ country, data=vitd1)
```

Bemærk:
Det er uvist, hvordan man i R får en eksakt vurdering af Kruskal-Wallis testet.....

Optælling af solvaner

Slide 63

```
vitd$sol <- factor(vitd3$sunexp,
    levels=c("Avoid sun", "Sometimes in sun", "Prefer sun"),
    labels=c("Nej", "Ja", "Ja"))


soltab = table(vitd3$country,vitd3$sol)
prop.table(soltab,1)*100

percent <- function(x){x[2]/(sum(x))}
addmargins(soltab,
    FUN = list("sum", list("sum", "percent")))
```

Box-plot, opdelt efter to kategorier

Slide 66

```
vitd2$sol <- factor(vitd2$sunexp,
      levels=c("Avoid sun", "Sometimes in sun", "Prefer sun"),
      labels=c("Nej", "Ja", "Ja"))

boxplot(VitaminD~sol*factor(country),
        data=vitd3,col=c("red","blue"))
```

Slide 90

Ganske som ovenfor, blot i en anden dataframe.

Additiv tosidet ANOVA

dvs. uden interaktion

Slide 69-70

```
model2 = lm(VitaminD ~ relevel(country, ref="SF") + sol,
         data=vitd3, na.action=na.exclude)
anova(model2)
summary(model2)
confint(model2)
```

Modelkontrolplots for Vitamin D eksemplet

Slide 73

Dette svarer til modelkontrollen s. 47, blot kun de 2 øverste figurer:

```
par(mfrow=c(1,2))
plot(model2,which=1)
plot(model2,which=2)
```

Figurer af predikterede værdier

Slide 74

```
vitd3$land=as.numeric(vitd3$country)
vitd3.ja <- subset(vitd3, sol %in% c("Ja"))
vitd3.nej <- subset(vitd3, sol %in% c("Nej"))

ny.country=c("SF","SF","PL","PL")
ny.sol=c("Ja","Nej","Ja","Nej")
ny = data.frame(country=ny.country, sol=ny.sol)
pred2 = predict(model2, ny)
cbind(ny.country,ny.sol,pred2)

plot(vitd3.ja$land, vitd3.ja$VitaminD,
     ylab="Vitamin D", xlab="country",
     pch=1, col="blue", cex.lab=1.8, cex=1.5)
segments(2,pred2[1],x1=4,y1=pred2[3],col="blue")
points(vitd3.nej$land, vitd3.nej$VitaminD,
     pch=1, col="red", cex=1.5)
segments(2,pred2[2],x1=4,y1=pred2[4],col="red")
legend(3.5, 85, legend=c("Ja", "Nej"), title="SOL",pch=1,
       col=c("blue", "red"), cex=1.5)
legend(2.5, 75, legend=c("2: SF", "4: PL"),
       title="Landekoder",pch=20,
       col="black", cex=1.5)
```

Tilsvarende kode for s. 77 og 91

Tosidet ANOVA med interaktion

Slide 82, 83

```
model3 = lm(VitaminD ~
         relevel(country, ref="PL")*sol,
         data=vitd3, na.action=na.exclude)
anova(model3)
summary(model3)
```