Faculty of Health Sciences

Basal Statistik

Sammenligning af grupper, Variansanalyse med SPSS

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 SPSS-vejledning

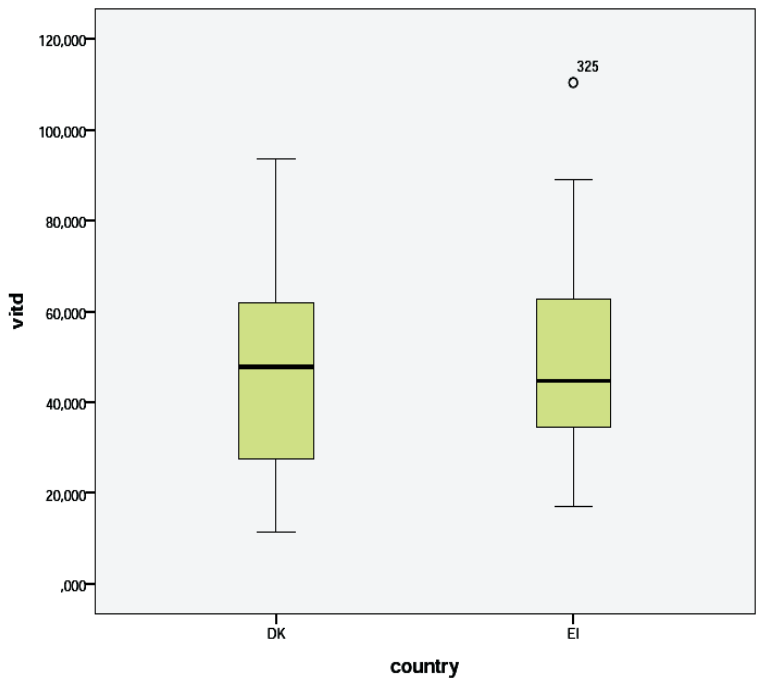
E-mail: ltsk@sund.ku.dk

$*$: 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 og figur,
se s. 99-101

Benyt
`Graphs/Graph Builder`
eller

`Analyze/Descriptive`
`Statistics/Explore`

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), udstyret med labels med landekoder (se evt. forrige forelæsning)
- Vitamin D status, `vitd` (Serum 25(OH)D, nmol/l)

Summary statistics, opdelt efter land: se nærmere s. 102 eller under T-test s. 7 (som benyttes her):

Group Statistics

	country	N	Mean	Std. Deviation	Std. Error Mean
Vitamin D	DK	53	47,16604	22,782922	3,129475
	EI	41	48,00732	20,222121	3,158165

Model for uparret sammenligning

Antagelser:

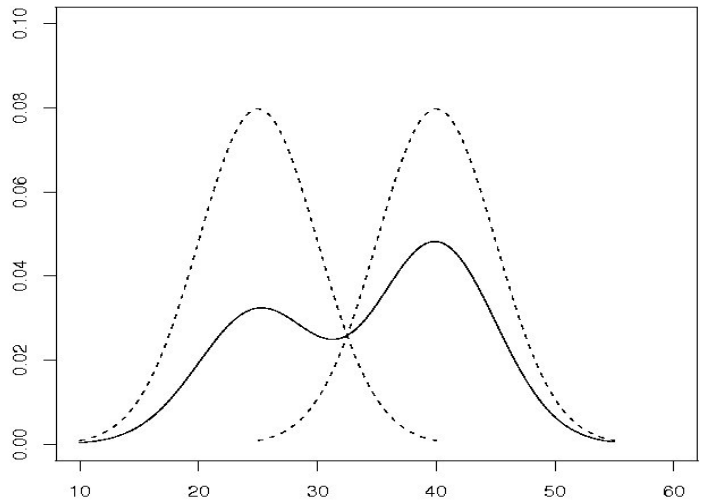
- Alle observationerne er uafhængige
  – kun 1 måling pr. person
  – personerne har ikke noget med hinanden at gøre
- Der er samme spredning(varians) i de to grupper
  – bør checkes/sandsynliggøres
- Observationerne følger en normalfordeling i hver gruppe, med hver deres middelværdi, μ_1 hhv. μ_2

– og det er disse 2 middelværdier (μ_1 og μ_2),
vi gerne vil sammenligne, altså $H_0 : \mu_1 = \mu_2$

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

Vi skal kun se på kvinder fra Irland og Danmark (`country=1` eller `country=4`), så vi benytter `Data/Select Cases`, hvor der afkrydses i `If` og skrives `category=2 | (country=1 | country=4)`

Herefter benytter vi `Analyze/Compare Means/Independent Samples T-test`, hvor vi sætter `vitd` over i `Test Variable(s)` og `country` i `Grouping Variable`.

Under `Define Groups` vælges `Group1` til 1 og `Group2` til 4.

Typisk output fra et uparret t-test

udover tabellen fra s. 4

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means	
		F	Sig.	t	df
Vitamin D	Equal variances assumed	1,364	,246	-,186	92
	Equal variances not assumed			-,189	90,213

		t-test for Equality of Means			
		Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence ... Lower
Vitamin D	Equal variances assumed	,853	-,841279	4,514686	-9,807835
	Equal variances not assumed	,850	-,841279	4,446079	-9,673908

		95% Confidence Interval of the ...
		Upper
Vitamin D	Equal variances assumed	8,125276
	Equal variances not assumed	7,991349

Kommentarer til output

- Først nogle summary statistics for hvert land (ses s. 4)
- Derefter et test for ens varianser (spredninger), som ikke forkastes, idet P=0.25
- 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 estimat for forskellen på de to midelværdier, med tilhørende standard error og konfidensinterval, igen i 2 forskellige udgaver, afhængig af, om spredningerne(varianserne) kan antages at være ens eller ej.

∗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):

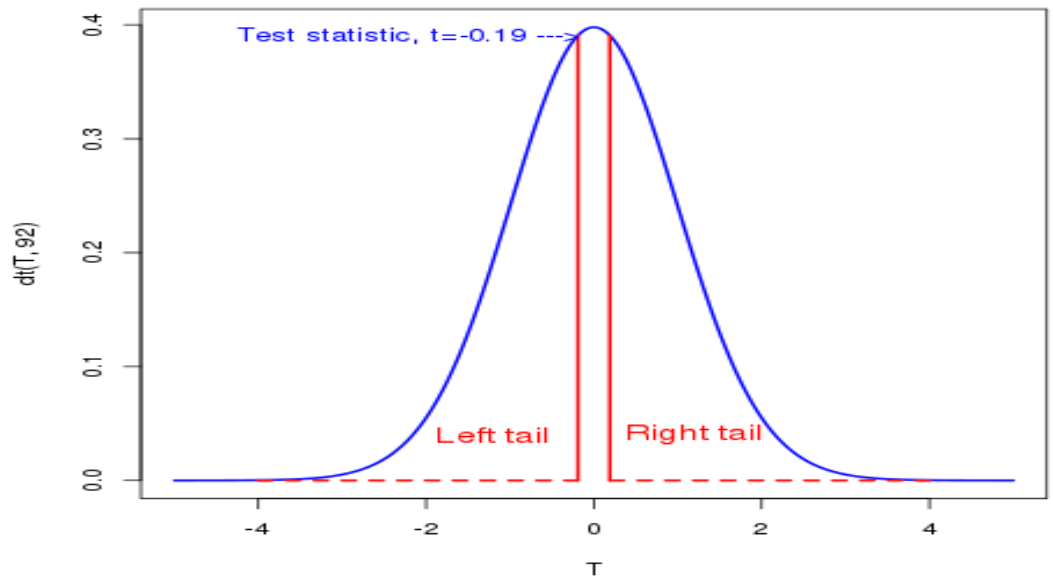
1. Fordel tilfældigt 53 i en gruppe, 41 i en anden, $\Longrightarrow t_1$
2. Fordel tilfældigt 53 i en gruppe, 41 i en anden, $\Longrightarrow t_2$
3. Fordel tilfældigt 53 i en gruppe, 41 i en anden, $\Longrightarrow t_3$
   osv. osv.

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.

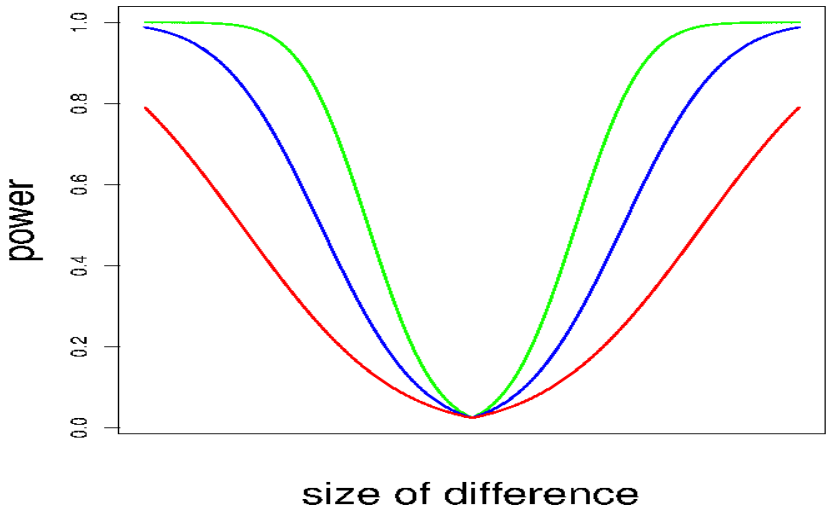
Styrke

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, men her har vi oplysninger om spredningsestimater fra T-testet. Disse er dog også behæftet med usikkerhed, som jeg ikke få SPSS til at udregne.....
På s. 104 er angivet en formel for konfidensintervaller for spedninger, og med frihedsgraderne hhv. 52 (DK) og 40 (EI), giver det nedenstående intervaller (som godt nok her ses i form af en udskrift fra SAS):

```
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 er noget, der hedder GPOWER....?? men man kan også benytte denne hjemmeside:
`http://sampsize.sourceforge.net/iface/s2.html#means`
hvorved vi f.eks. får:

```
Sample size results
Estimated sample size for two-sample comparison of means

Test Ho:   m1 = m2, where m1 is the mean in population 1
and m2 is the mean in population 2

Assumptions:
        alpha =          .05 (two-sided)
        power =           .9
           m1 =            50
           m2 =            55
          sd1 =            20
          sd2 =            20
        n2/n1 =             1

Estimated sample size:
           n1 =           337
           n2 =           337
```

Dimensionering i praksis, II

eller

```
Sample size results
Estimated sample size for two-sample comparison of means

Test Ho:   m1 = m2, where m1 is the mean in population 1
and m2 is the mean in population 2

Assumptions:
        alpha =          .05 (two-sided)
        power =          .80
           m1 =           50
           m2 =           55
          sd1 =           28
          sd2 =           28
        n2/n1 =            1

Estimated sample size:
           n1 =          493
           n2 =          493
```

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 (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

Gå ind i menuen
`Analyse/Nonparametric Tests/Legacy Dialogs` og vælg
`2 Independent Samples`, hvorefter `vitd` sættes i
`Test Variable List` og `country` sættes i `Grouping Variable`.
Under `Test Type` vælges `Mann-Whitney U`.

Output s. 30 giver P-værdien 0.91 og altså absolut ingen signifikans.

△Output fra Mann-Whitney test

Mann-Whitney Test

Ranks

	country	N	Mean Rank	Sum of Ranks
Vitamin D	DK	53	47,22	2502,50
	EI	41	47,87	1962,50
	Total	94		

Test Statistics[a]

	Vitamin D
Mann-Whitney U	1071,500
Wilcoxon W	2502,500
Z	-,114
Asymp. Sig. (2-tailed)	,909

a. Grouping Variable: country

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:

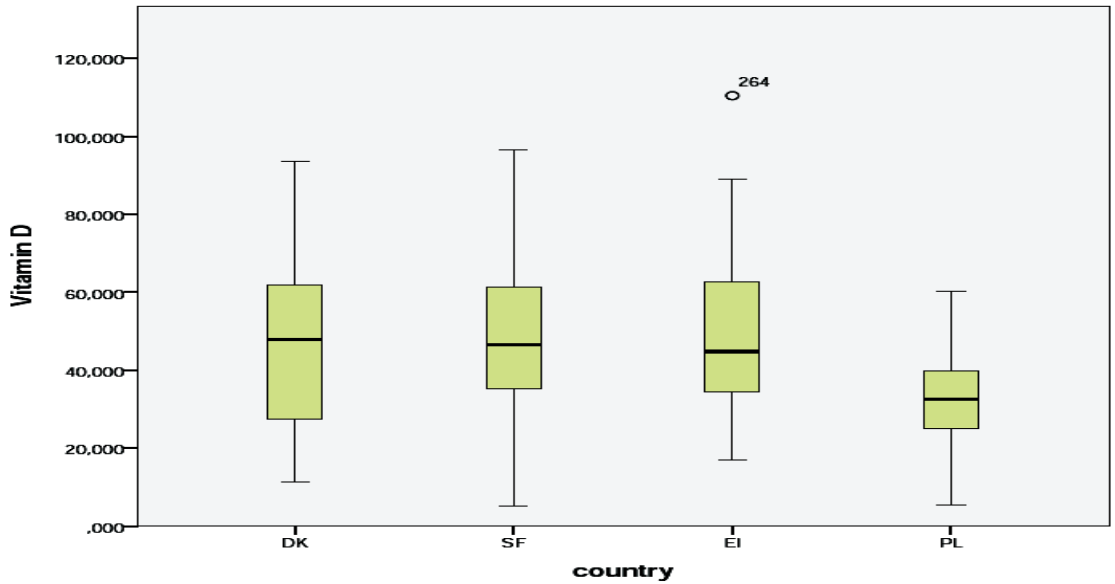
(-12.71 cm^3, 13.19 cm^3)

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

Foretages ganske som s. 3, bare med `Data/Select Cases` ændret til kun at være `category=2`.



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* (`Std. Deviation` nedenfor)

Summary statistics for de 4 grupper:

Descriptive Statistics

country		N	Minimum	Maximum	Mean	Std. Deviation
DK	Vitamin D	53	11,400	93,600	47,16604	22,782922
	Valid N (listwise)	53				
SF	Vitamin D	54	5,200	96,600	47,99074	18,724711
	Valid N (listwise)	54				
EI	Vitamin D	41	17,000	110,400	48,00732	20,222121
	Valid N (listwise)	41				
PL	Vitamin D	65	5,400	60,300	32,56154	12,464483
	Valid N (listwise)	65				

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
  Poolet varians (vægtet gennemsnit): $343.897 = 18.54^2$
- 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

Her kan benyttes

- `Analyze/Compare Means/One-Way ANOVA`:
  - Fordele: kan lave Welch test i tilfælde af uens varianser (se s. 48)
  - Ulemper: Giver ikke parameterestimater og kan derfor kun bruges i simple situationer
- `Analyze/General Linear Model/Univariate`:
  - Fordele: er meget fleksibel, og kan klare generelle lineære modeller (senere forelæsninger)
  - Ulemper: lidt mere indviklet opsætning

Begge kan lave multiple sammenligninger, med korrektion for massesignifikans (se s. 55-56)

Ensidet ANOVA i praksis, II

Den generelle lineære model har som specialtilfælde en ensidet variansanalyse:

Brug `Analyze/General Linear Model/Univariate`, sæt `vitd` i `Dependent Variable` og `country` i `Fixed Factor(s)`.

Under `Options` afkrydses `Parameter estimates` for at sikre, at der kommer parameterestimater i output.

Output fra GLM (ensidet ANOVA)

Den typiske start på outputtet (den mindre brugbare del):

Tests of Between-Subjects Effects

Dependent Variable: Vitamin D

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	10373,991[a]	3	3457,997	10,055	,000
Intercept	400194,107	1	400194,107	1163,704	,000
country	10373,991	3	3457,997	10,055	,000
Error	71874,406	209	343,897		
Total	477557,370	213			
Corrected Total	82248,397	212			

a. R Squared = ,126 (Adjusted R Squared = ,114)

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:

Parameter Estimates

Dependent Variable: Vitamin D

Parameter	B	Std. Error	t	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Intercept	32,562	2,300	14,156	,000	28,027	37,096
[country=1]	14,604	3,432	4,255	,000	7,839	21,370
[country=2]	15,429	3,415	4,519	,000	8,698	22,161
[country=4]	15,446	3,698	4,176	,000	8,155	22,737
[country=6]	0[a]	.	.	.	.	.

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

med fortolkning på de følgende sider

Bemærkninger til output s. 39

- Estimatet s for spredningen σ (poolet for de 4 grupper=lande) er ikke umiddelbart angivet.
  Man skal selv udregne den som kvadratroden af `Mean Square Error`, som her ses at være 343.897, så vi finder
  $s = \sqrt{343.897} = 18.54$
- F Value: Teststørrelsen for test af ens middelværdier i de 4 grupper, med tilhørende P-værdi. Her er P angivet som 0.000, 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 s. 40

- `Intercept` svarer til niveauet for referencegruppen (sidste gruppe, alfabetisk eller numerisk), dvs. Polen (`country=6`)
- Estimatet ud for f.eks. `[country=1]` er forskellen i niveau mellem DK (`country=1`) og referencegruppen PL (`country=6`)

Bemærk:

Man kan få vilkårlige forskelle frem:

- ved omkodning af grupper
- ved at foretage multiple sammenligninger (se s. 55-56)

Modelantagelse 1: Uafhængighed

Dette er noget, man skal vide

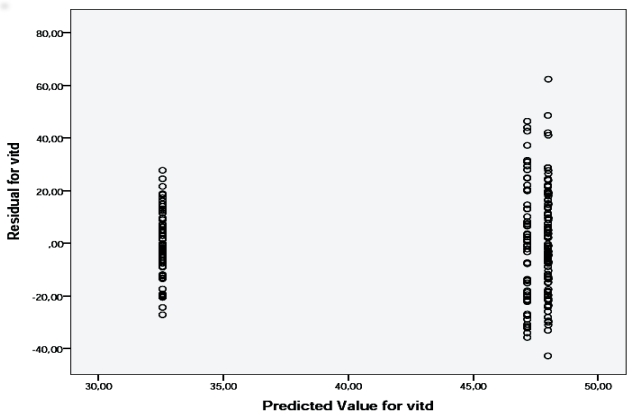
- ingen tvillinger, søskende etc.
- kun *en* observation for hver person
  (ellers hører det hjemme under emnet
  "Korrelerede målinger", kursets sidste emne)

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 plot fås ved at gemme residualer og predikterede værdier fra analysen, se. s. 110

Spredningen stiger vist lidt med den predikterede værdi

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, så vi tester med Levenes test (se s. 109)

Test of Homogeneity of Variances

Vitamin D

Levene Statistic	df1	df2	Sig.
6,837	3	209	,000

ANOVA

Vitamin D

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	10373,991	3	3457,997	10,055	,000
Within Groups	71874,406	209	343,897		
Total	82248,397	212			

Ved sammenligning af de 4 variansestimater fås en P-værdi angivet som P=0.000, og altså kraftig signifikans!

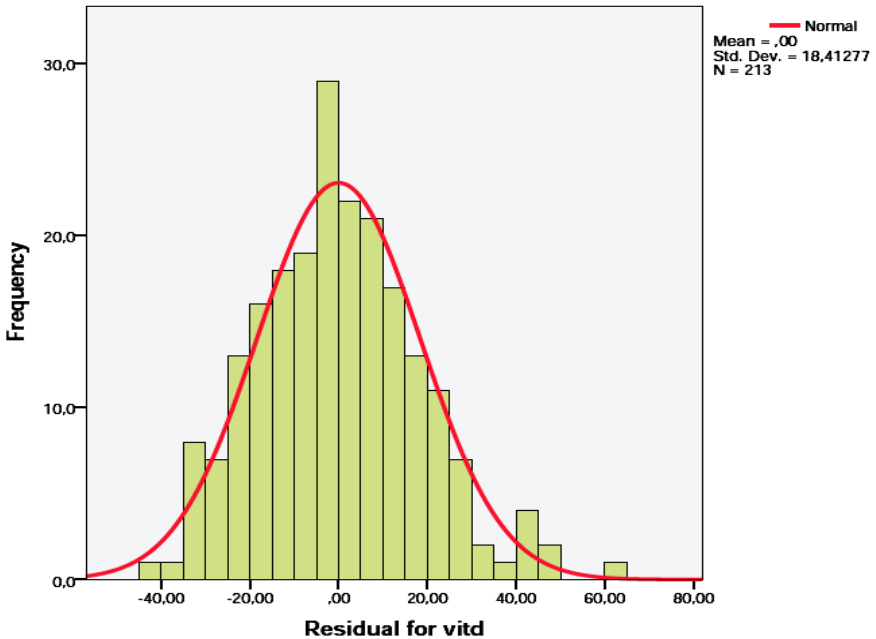
Dette vil vi gerne gøre noget ved lige om lidt (s. 48).

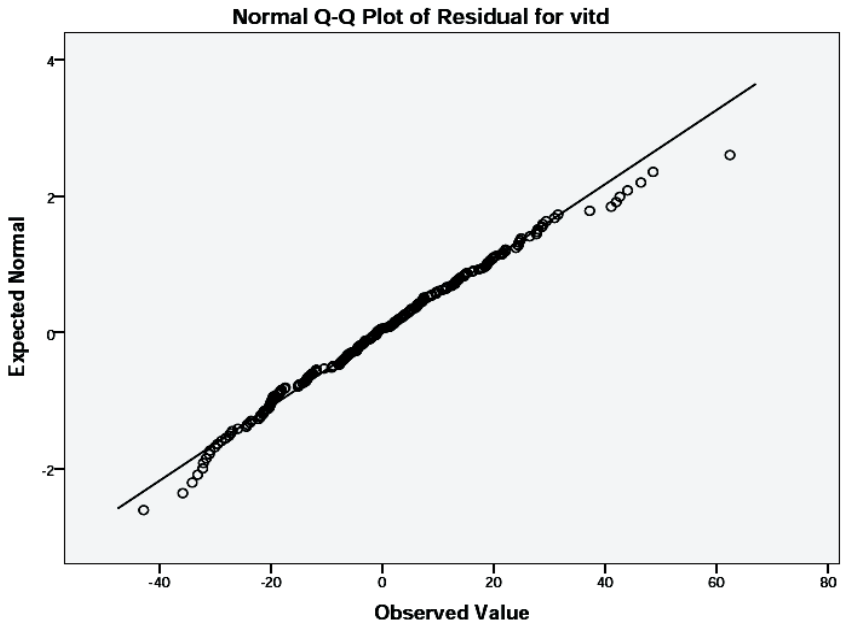
Modelantagelse 3: Normalfordelingsantagelsen

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: Check af normalfordelte residualer





Ser residualerne normalfordelte ud?

Næsten, dog lidt "omvendt hængekøje"=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:
  Her er man *undtagelsesvis* nødt til at bruge
  `Analyze/Compare Means/One Way ANOVA`, se s. **??**

Robust Tests of Equality of Means

Vitamin D

	Statistic[a]	df1	df2	Sig.
Welch	14,642	3	102,329	,000

a. Asymptotically F distributed.

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, pga de uens spredninger
- 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

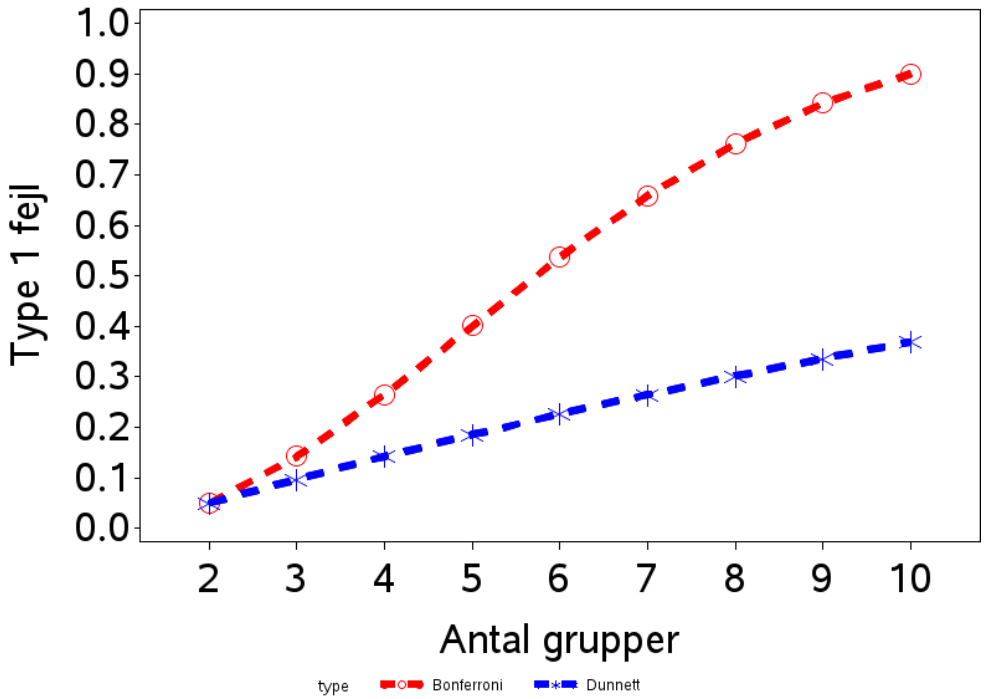
Hvis man sammenligner k grupper (lande) parvist, er der $m = k(k-1)/2$ mulige test, hver med signifikansniveau $\alpha = 0.05$.

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:

Dette kan gøres ved at gå ind i
`Analyze/Compare Means/One Way ANOVA` eller
`Analyze/General Linear Model/Univariate`, sætte `vitd` i `Dependent List` og `country` i `Factor`.

Efterfølgende går man så ind i `Post Hoc`, sætter `country` i `Post Hoc Tests for` og foretager relevante afkrydsninger, f.eks. `Tukey` eller `Games-Howell`

Output ses s. 56-58

Tukey korrektion for sammenlingning af lande

Multiple Comparisons

Dependent Variable: Vitamin D

Tukey HSD

(I) country	(J) country	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
DK	SF	-,824703	3,585676	,996	-10,11096	8,46155
	EI	-,841279	3,856986	,996	-10,83018	9,14762
	PL	14,604499*	3,432104	,000	5,71597	23,49303
SF	DK	,824703	3,585676	,996	-8,46155	10,11096
	EI	-,016576	3,841377	1,000	-9,96505	9,93190
	PL	15,429202*	3,414553	,000	6,58612	24,27228
EI	DK	,841279	3,856986	,996	-9,14762	10,83018
	SF	,016576	3,841377	1,000	-9,93190	9,96505
	PL	15,445779*	3,698438	,000	5,86749	25,02407
PL	DK	-14,604499*	3,432104	,000	-23,49303	-5,71597
	SF	-15,429202*	3,414553	,000	-24,27228	-6,58612
	EI	-15,445779*	3,698438	,000	-25,02407	-5,86749

*. The mean difference is significant at the 0.05 level.

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

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:

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

Homogeneous Subsets

Vitamin D

Tukey HSD[a,b]

country	N	Subset for alpha = 0.05	
		1	2
PL	65	32,56154	
DK	53		47,16604
SF	54		47,99074
EI	41		48,00732
Sig.		1,000	,996

Means for groups in homogeneous subsets are displayed.

a. Uses Harmonic Mean Sample Size = 51,839.

b. The group sizes are unequal. The harmonic mean of the group sizes is used. Type I error levels are not guaranteed.

Games-Howell korrektion

er bedre, når varianserne er forskellige:

Multiple Comparisons

Dependent Variable: Vitamin D

Games-Howell

(I) country	(J) country	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
DK	SF	-,824703	4,035651	,997	-11,36805	9,71864
	EI	-,841279	4,446079	,998	-12,47921	10,79665
	PL	14,604499*	3,490533	,000	5,43762	23,77138
SF	DK	,824703	4,035651	,997	-9,71864	11,36805
	EI	-,016576	4,057939	1,000	-10,65692	10,62377
	PL	15,429202*	2,980448	,000	7,62599	23,23241
EI	DK	,841279	4,446079	,998	-10,79665	12,47921
	SF	,016576	4,057939	1,000	-10,62377	10,65692
	PL	15,445779*	3,516278	,000	6,15100	24,74056
PL	DK	-14,604499*	3,490533	,000	-23,77138	-5,43762
	SF	-15,429202*	2,980448	,000	-23,23241	-7,62599
	EI	-15,445779*	3,516278	,000	-24,74056	-6,15100

*. The mean difference is significant at the 0.05 level.

Her finder vi Finland vs. Polen: 15.43 (7.63, 23.23), altså noget smallere end den sædvanlige Tukey-korrektion (pga Polens mindre varians).

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:
Brug menuen `Analyze/Nonparametric Tests/Legacy Dialogs` og vælg `K Independent Samples`, se s. 112

Kruskal-Wallis Test

Ranks

	country	N	Mean Rank
Vitamin D	DK	53	117,96
	SF	54	124,73
	EI	41	121,34
	PL	65	74,28
	Total	213	

Test Statistics[a,b]

	Vitamin D
Chi-Square	26,682
df	3
Asymp. Sig.	,000

a. Kruskal Wallis Test

b. Grouping Variable: country

Bemærk:

- Dette er et approksimativt test
- Man kan (i princippet) også få en eksakt vurdering af teststørrelsen, men her giver det desværre fejlmeddelelsen `insufficient memory`

Sammenligning af Finland og Polen

...som om vi kun havde disse to lande, dvs. et T-test (som s. 7-8):

Group Statistics

	country	N	Mean	Std. Deviation	Std. Error Mean
Vitamin D	SF	54	47,99074	18,724711	2,548110
	PL	65	32,56154	12,464483	1,546029

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means	
		F	Sig.	t	df
Vitamin D	Equal variances assumed	7,472	,007	5,367	117
	Equal variances not assumed			5,177	89,194

Sammenligning af Finland og Polen, II

		t-test for Equality of Means			
		Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence ... Lower
Vitamin D	Equal variances assumed	,000	15,429202	2,875055	9,735306
	Equal variances not assumed	,000	15,429202	2,980448	9,507293

		95% Confidence Interval of the ... Upper
Vitamin D	Equal variances assumed	21,123099
	Equal variances not assumed	21,351112

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

Bemærk: Samme estimat som i ANOVA'en,
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 s. 113) 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. 113) giver:

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

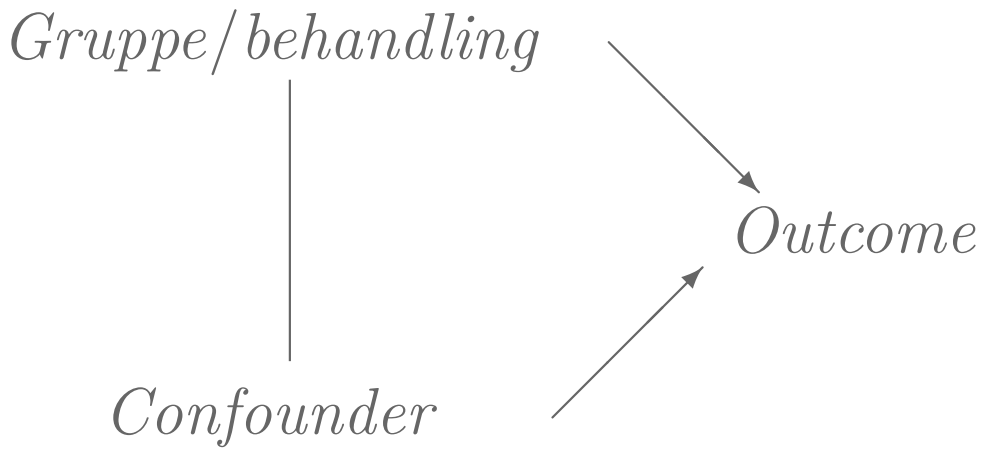
Kan denne forskel i sol-præferencer være forklaringen på forskellen i Vitamin-D?

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

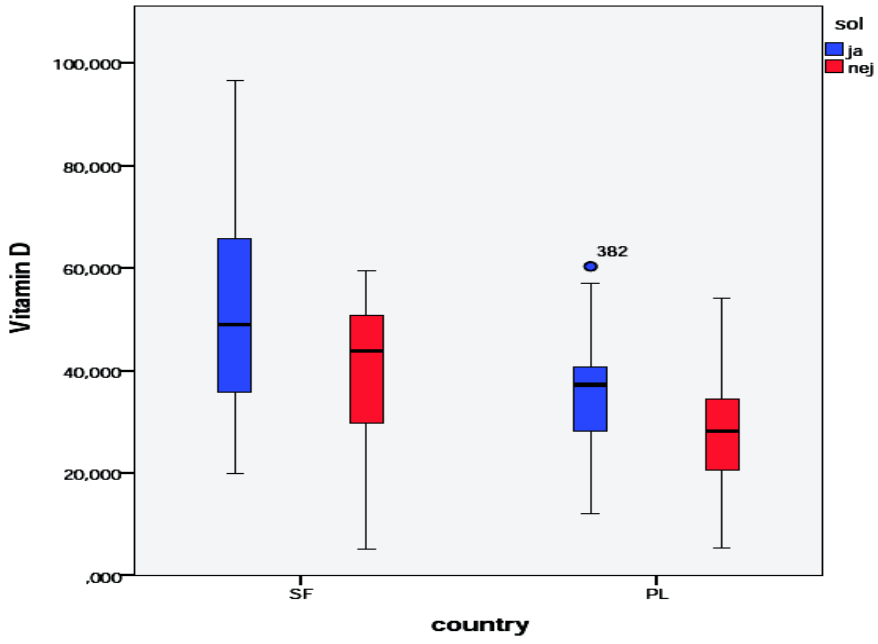
- har en effekt på outcome
- er relateret til gruppen
  som en direkte konsekvens af denne

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

Er solvaner en mediator?

Har solvanerne overhovedet en betydning?

Benyt
`Graph/Chart Builder/Bar`,
og vælg nr. 2 fra venstre.
Følg så opskriften s. 114



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. 76ff)

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

Brug `Analyze/General Linear Model/Univariate`, og følg opskriften s. 115

Herved får vi outputtet

Between-Subjects Factors

		Value Label	N
country	2	SF	54
	6	PL	65
sol	ja		79
	nej		40

fortsættes næste side...

Output fra 2-sidet ANOVA af Vitamin D, fortsat

Tests of Between-Subjects Effects

Dependent Variable: Vitamin D

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	8615,910[a]	2	4307,955	18,555	,000
Intercept	158069,165	1	158069,165	680,834	,000
country	5920,805	1	5920,805	25,502	,000
sol	1594,132	1	1594,132	6,866	,010
Error	26931,707	116	232,170		
Total	221810,340	119			
Corrected Total	35547,617	118			

a. R Squared = ,242 (Adjusted R Squared = ,229)

Parameter Estimates

Dependent Variable: Vitamin D

Parameter	B	Std. Error	t	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Intercept	27,861	2,606	10,692	,000	22,700	33,022
[country=2]	14,327	2,837	5,050	,000	8,708	19,946
[country=6]	0[a]	.	.	.	.	.
[sol=ja]	7,835	2,990	2,620	,010	1,913	13,757
[sol=nej]	0[a]	.	.	.	.	.

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

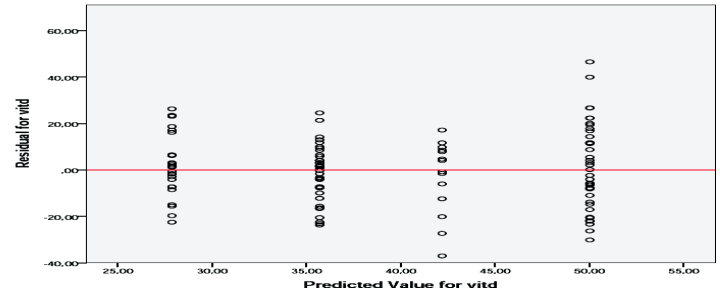
Fortolkning af 2-sidet ANOVA

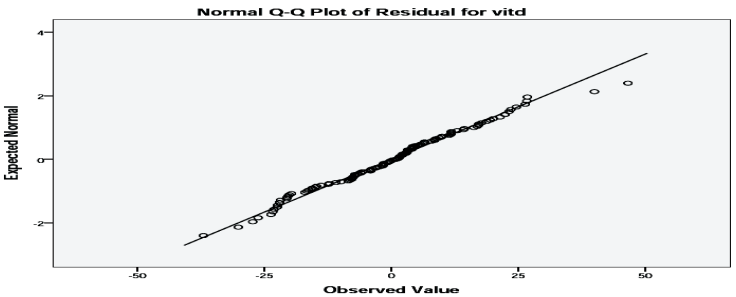
Effekter:

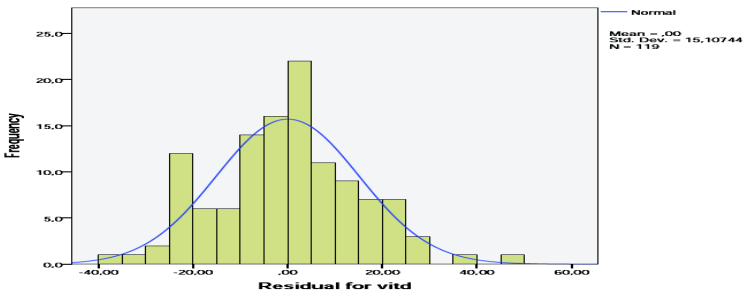
- Finland vs. Polen, for fastholdte solvaner:
  Finland estimeres til at ligge 14.33 nmol/l højere end Polen,
  med 95% CI: (8.71, 19.95)
  Dette afviger noget fra T-testet (s. 63), hvor vi fik estimatet 15.43, med 95% CI: (9.51, 21.35).
  Vi fik her indsnævret konfidensintervallet en anelse
  fordi vi fjernede noget af residualvariationen

- Folk, der kan lide sol har et 7.84 højere niveau end de fra samme land, som ikke kan lide sol,
  med 95% CI: (1.91, 13.76), P=0.01

Modelkontrolplots







Residualer mod predikterede værdier:

Intet specielt mønster her

Fraktildiagram og histogram:

Disse to figurer viser en udmærket tilpasning af normalfordelingen.

Vurdering af Finland vs. Polen

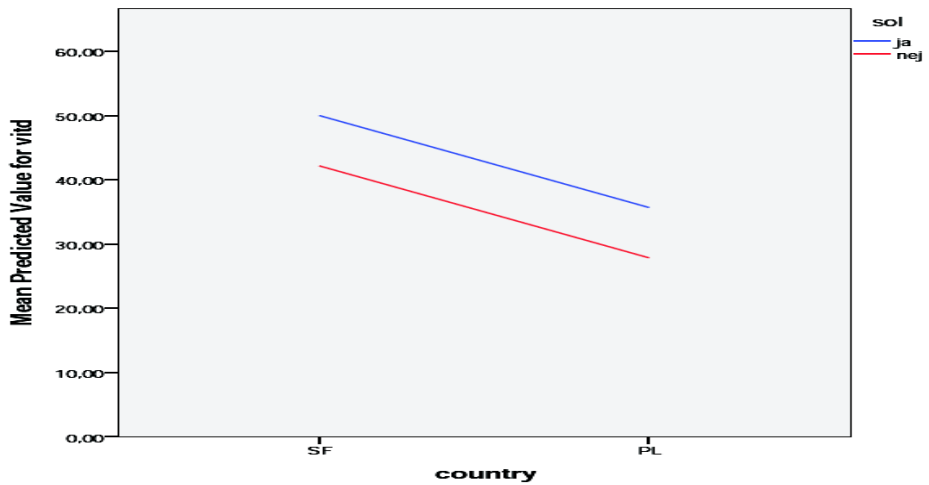
Model indeholder	estimat	CI
kun country (T-test)	15.43	(9.51, 21.35)
solvaner og country	14.33	(8.71, 19.95)
SF vs. PL, solelskere	16.40	(9.37, 23.43)
SF vs. PL, solhadere	9.82	(-0.47, 20.11)

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.

Den additive model - Predikterede værdier

Predikterede værdier svarende til samme sol-gruppe er forbundet, og hældningen viser derfor landeforskellen (se s. 116)



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 "*værdien*" 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. 75)

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

Vitamin D for solelskere vs. solhadere:

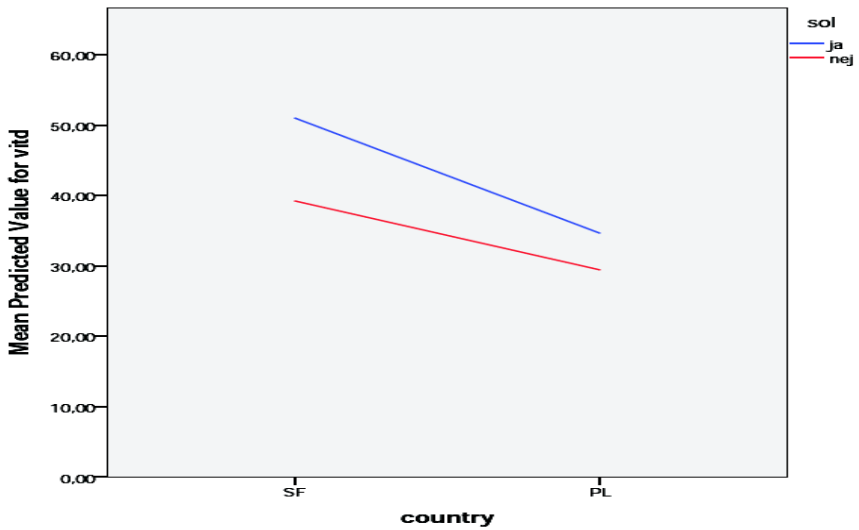
- i Finland:
  11.79 (5.64), 95% CI: (0.48, 23.10), P=0.04

- 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

Fremgangsmåde som for den additive model, se s. 116



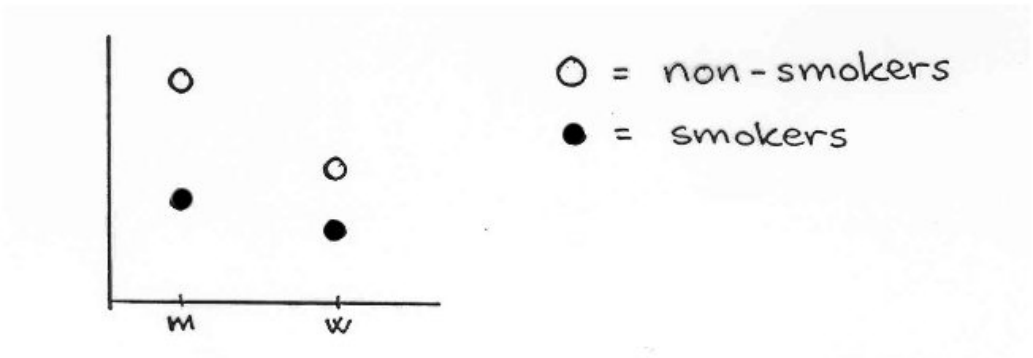
Er der forskel på sol-effekterne i de to lande?

Ikke så meget, ser det ud til....
for linierne er *næsten* parallelle (fortsættes s. 83)

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

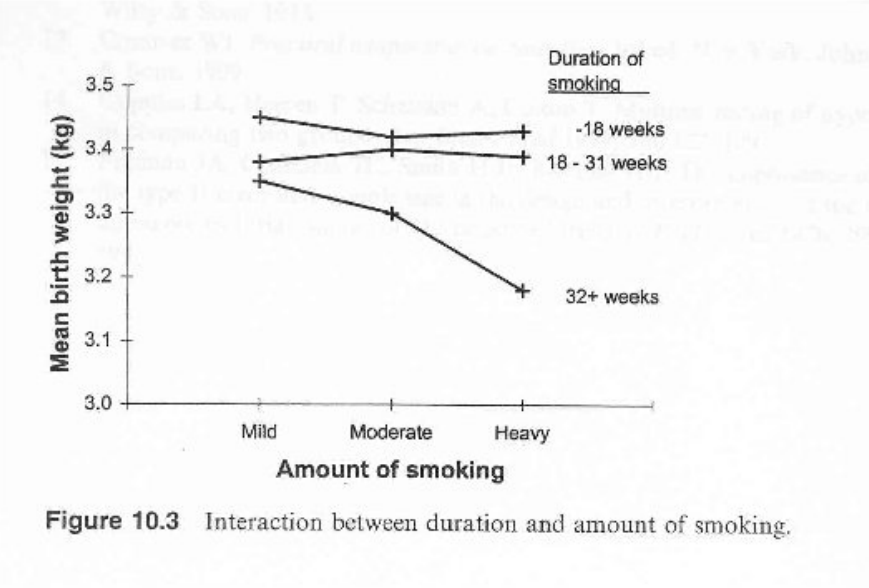


Figure 10.3 Interaction between duration and amount of smoking.

- 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:

Vi bruger opsætningen som på s. 117, idet interaktionsleddet `country*sol` skal inkluderes i `Model`

Vi får så outputtet:

Between-Subjects Factors

		Value Label	N
country	1	DK	53
	2	SF	54
	4	EI	41
	6	PL	65
sol	ja		143
	nej		70

Output vedr. interaktion, fortsat

Parameter Estimates

Dependent Variable: Vitamin D

Parameter	B	Std. Error	t	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Intercept	29,438	2,986	9,858	,000	23,524	35,353
[country=2]	9,819	5,047	1,945	,054	-,179	19,817
[country=6]	0[a]	.	.	.	.	.
[sol=ja]	5,205	3,855	1,350	,180	-2,431	12,841
[sol=nej]	0[a]	.	.	.	.	.
[country=2] * [sol=ja]	6,585	6,101	1,079	,283	-5,499	18,669
[country=2] * [sol=nej]	0[a]	.	.	.	.	.
[country=6] * [sol=ja]	0[a]	.	.	.	.	.
[country=6] * [sol=nej]	0[a]	.	.	.	.	.

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

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=2]*[sol=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. 77

  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)

Estimationsvenlige koder, I

Forskelle på landene, for hver sol-gruppe:

Udelad `country`, så modellen kun indeholder `sol` og `country*sol` :

Parameter Estimates

Dependent Variable: Vitamin D

Parameter	B	Std. Error	t	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Intercept	29,438	2,986	9,858	,000	23,524	35,353
[sol=ja]	5,205	3,855	1,350	,180	-2,431	12,841
[sol=nej]	0[a]	.	.	.	.	.
[country=2] * [sol=ja]	16,404	3,426	4,787	,000	9,617	23,191
[country=2] * [sol=nej]	9,819	5,047	1,945	,054	-,179	19,817
[country=6] * [sol=ja]	0[a]	.	.	.	.	.
[country=6] * [sol=nej]	0[a]	.	.	.	.	.

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

Estimationsvenlige koder, II

Separate effekter af sol, for hvert land:

Udelad `sol`, så modellen kun indeholder `country` og `country*sol` :

Parameter Estimates

Dependent Variable: Vitamin D

Parameter	B	Std. Error	t	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Intercept	29,438	2,986	9,858	,000	23,524	35,353
[country=2]	9,819	5,047	1,945	,054	-,179	19,817
[country=6]	0[a]	.	.	.	.	.
[country=2] * [sol=ja]	11,790	4,728	2,494	,014	2,425	21,156
[country=2] * [sol=nej]	0[a]	.	.	.	.	.
[country=6] * [sol=ja]	5,205	3,855	1,350	,180	-2,431	12,841
[country=6] * [sol=nej]	0[a]	.	.	.	.	.

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

Fokus på effekt af sol

Soldyrkere vs. solhadere, stadig kun Finland og Polen:

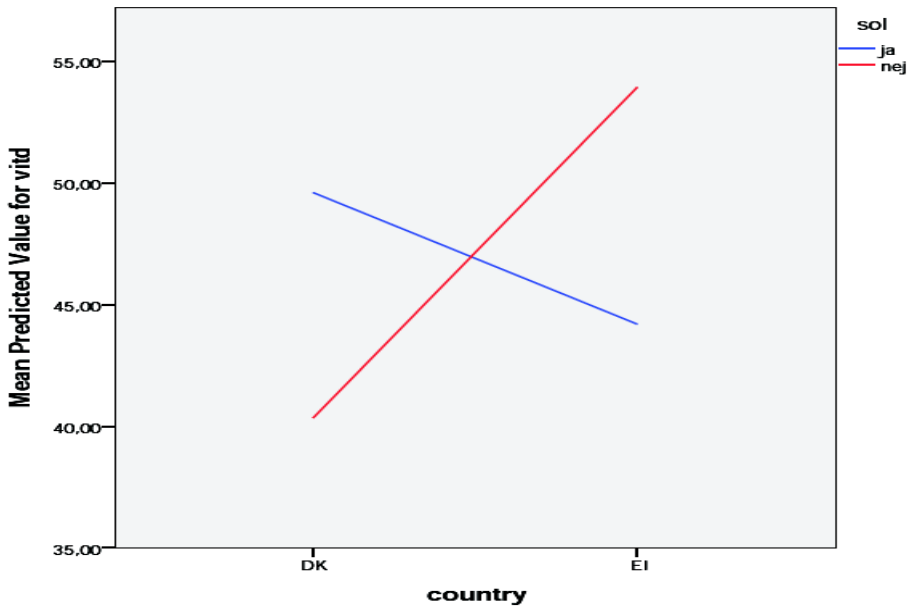
Model indeholder	estimat	CI
kun solvaner (T-test)	10.07	(3.63, 16.51)
solvaner og land	7.83	(1.91, 13.76)
solvaner, kun SF	11.79	(0.48, 23.10)
solvaner, kun PL	5.21	(-1.01, 11.42)

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

Prediktion i interaktionsmodel (dvs. gennemsnit), ganske som s. 78



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

Interaktion mellem solvaner og land?

for Danmark og Irland:

Parameter Estimates

Dependent Variable: Vitamin D

Parameter	B	Std. Error	t	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Intercept	53,956	5,370	10,047	,000	43,287	64,626
[country=1]	-13,621	7,862	-1,733	,087	-29,239	1,998
[country=4]	0[a]	.	.	.	.	.
[sol=ja]	-9,756	6,878	-1,419	,159	-23,420	3,907
[sol=nej]	0[a]	.	.	.	.	.
[country=1] * [sol=ja]	19,038	9,597	1,984	,050	-,027	38,104
[country=1] * [sol=nej]	0[a]	.	.	.	.	.
[country=4] * [sol=ja]	0[a]	.	.	.	.	.
[country=4] * [sol=nej]	0[a]	.	.	.	.	.

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

Kommentarer til Danmark vs. Irland

Der er *næsten signifikant* interaktion (P=0.050)

- 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` (`sol` under `sountry*sol`). Det bevirker, at vi stadig
  - stadig får testet for interaktion
  - får estimater for sol-effekten for hvert land
    (Se evt. forskellen fra output s. 83-84)
- Hvis man (som det er gjort her) fjerner fluebenet i `Include an Intercept`, fås estimater for middelværdien for hver af de to lande for `sol`-referencegruppen (`sol=nej`, dvs. solhadere)

△Fokus på effekt af sol, output

Parameter Estimates

Dependent Variable: Vitamin D

Parameter	B	Std. Error	t	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
[sol=ja]	44,200	4,296	10,288	,000	35,665	52,735
[sol=nej]	53,956	5,370	10,047	,000	43,287	64,626
[country=1] * [sol=ja]	5,418	5,504	,984	,328	-5,516	16,352
[country=1] * [sol=nej]	-13,621	7,862	-1,733	,087	-29,239	1,998
[country=4] * [sol=ja]	0[a]	.	.	.	.	.
[country=4] * [sol=nej]	0[a]	.	.	.	.	.
[country=1]	0[a]	.	.	.	.	.
[country=4]	0[a]	.	.	.	.	.

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

Effekt af sol, alle 4 lande, output

Between-Subjects Factors

		Value Label	N
country	1	DK	53
	2	SF	54
	4	EI	41
	6	PL	65
sol	ja		143
	nej		70

Tests of Between-Subjects Effects

Dependent Variable: Vitamin D

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Model	409363,470[a]	8	51170,434	153,825	,000
country	8875,628	3	2958,543	8,894	,000
country * sol	2777,147	3	925,716	2,783	,042
sol	758,007	1	758,007	2,279	,133
Error	68193,900	205	332,653		
Total	477557,370	213			

a. R Squared = ,857 (Adjusted R Squared = ,852)

Output, fortsat

Estimater for sol-effekt, for hvert land separat:

Parameter Estimates

Dependent Variable: Vitamin D

Parameter	B	Std. Error	t	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
[country=1]	40,336	4,875	8,275	,000	30,725	49,946
[country=2]	39,257	4,875	8,054	,000	29,647	48,868
[country=4]	53,956	4,560	11,833	,000	44,966	62,946
[country=6]	29,438	3,577	8,230	,000	22,386	36,491
[country=1] * [sol=ja]	9,282	5,682	1,633	,104	-1,921	20,486
[country=1] * [sol=nej]	0[a]	.	.	.	.	.
[country=2] * [sol=ja]	11,790	5,664	2,082	,039	,624	22,957
[country=2] * [sol=nej]	0[a]	.	.	.	.	.
[country=4] * [sol=ja]	-9,756	5,839	-1,671	,096	-21,269	1,756
[country=4] * [sol=nej]	0[a]	.	.	.	.	.
[country=6] * [sol=ja]	5,205	4,618	1,127	,261	-3,899	14,310
[country=6] * [sol=nej]	0[a]	.	.	.	.	.
[sol=ja]	0[a]	.	.	.	.	.
[sol=nej]	0[a]	.	.	.	.	.

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

Irland opfører sig underligt....

APPENDIX

med SPSS-kommentarer svarende til nogle af slides

- Indlæsning og summary statistics: s. 99, 102
- Figurer: s. 101, 107, **??**, 114
- T-tests mv.: s. 103, 106
- Dimensionering: s. 105
- Ensidet ANOVA: s. 108ff
- Tosidet ANOVA: s. 115-**??**

Indlæsning af Vitamin D data

Slide 3

For at indlæse allerede eksisterende data, benyttes menuen `File/Open/Data`

Hvis data skal hentes fra nettet, må man først benytte `File/Open/Internet Data`, hvorefter man skriver stien `http://publicifsv.sund.ku.dk/~lts/basal/data/VitaminD.txt` i `Web location...` samt det ønskede navn på datasættet i `Dataset Name to Assign`.

I SPSS indeholder variablen `country` de numeriske værdier 1,2,4 og 6, medens der er defineret `Value Labels` (se s. 100) svarende til de mere sigende navne: `DK, SF, EI, PL`

Value labels

For at få relevante navne på værdierne for landende, kan man gå til `Data/Variable View` og klikke i `Values` under den relevante variabel (her `country`).

Herved fremkommer `Variable Labels`-box, hvor man successivt udfylder `Value` med de aktuelle værdier (her 1,2,4,6) og de dertil hørende labels (DK, SF, EI, PL). Efetr hvert par klikkes `Add`.

Boxplots

Slide 3

Vi skal kun se på kvinder (`category=2`) fra Irland og Danmark (`country=1` eller `country=4`), så vi benytter `Data/Select Cases`, hvor der afkrydses i `If` og skrives `(country=1 | country=4) & category=2`.

Herefter benytter vi `Analyze/Descriptive Statistics/Explore`, hvor vi sætter `vitd` i `Dependent List`, `country` i `Factor List` samt sætter hak i `Plots`.

De kan også laves i `Graphs/Graph Builder`, hvor `vitd` trækkes over på Y-aksen, og `country` trækkes over på X-aksen.

Summary statistics

Slide 4

For at få udregnet diverse størrelser for hvert land for sig, benyttes `Data/Split file/Compare Groups`, hvor `country` sættes ind i feltet `Groups Based on`.

Herefter går man ind under menuen
`Analyze/Descriptive Statistics`, vælger `Descriptives` og flytter de ønskede variable (her blot `vitd`) over i `Variable(s)`.

Under `Statistics` kan man så sætte flueben ved de størrelser, vi ønsker udregnet
(her `Mean`, `Std. deviation`, `Minimum` og `Maximum`).
Husk bagefter at aflyse `Split File` ved igen at sætte flueben i `Analyze all data`

Uparret T-test

Slide 7-8

Man kan evt. starte med at udvælge 2 lande, se s. 101, men det behøves ikke.

Vi benytter
`Analyze/Compare Means/Independent Samples T-test`, hvor vi sætter `vitd` over i `Test Variable(s)` og `country` i `Grouping Variable`.

Under `Define Groups` vælges `Group1` til 1 og `Group2` til 4.

Konfidensinterval for spredning

Da jeg ikke kan få SPSS til at angive et sådant konfidensinterval, kan man i stedet benytte en formel:

$$\sqrt{\frac{(n-1)s^2}{\chi^2_{\alpha/2}}} \leq \sigma \leq \sqrt{\frac{(n-1)s^2}{\chi^2_{1-\alpha/2}}}$$

hvor

n angiver antallet af observationer (her 53 hhv 41)

χ^2-størrelserne har df=antale frihedsgrader=(n-1), dvs her 52 hhv 40

α angiver den ønskede dækningsgrad, dvs. $\alpha = 0.04$ for et 95% sikkerhedsinterval

Dimensionering i SPSS

Slide 24-25

Man kan benytte denne hjemmeside:
`http://sampsize.sourceforge.net/iface/s2.html#means`
her med opsætningen svarende til output s. 25:

Sample size for a study comparing means

Mean 1	50	
Mean 2	55	%
Standard deviation 1	28	
Standard deviation 2	28	%
Power	80	% (default 90%)
Ratio n2/n1	1	(default 1)
Alpha risk	5	% (default 5%)
One-sided test	☐	
One-sample test	☐	

Cluster sampling design: ☐

Intraclass correlation

Number of clusters ____ or, (not both) ____ Number of observations

Calculate Clear

Nonparametrisk uparret test i SPSS

Slide 29-30

Mann-Whitney test eller Kruskal-Wallis test

Gå ind i menuen
`Analyse/Nonparametric Tests/Legacy Dialogs` og vælg
`2 Independent Samples`, hvorefter `vitd` sættes i
`Test Variable List` og `country` sættes i `Grouping Variable`.
Under `Test Type` vælges `Mann-Whitney U`.

Box plots

Slide 33

Boxplottet skal her vise kvinder (`category=2`) fra alle landene, så vi må ind i `Data/Select Cases` og sørge for, at der er afkrydset i `If`, hvor der skal stå `category=2`.

Herefter benytter vi
`Analyze/Descriptive Statistics/Explore`, hvor vi sætter `vitd` i `Dependent List`, `country` i `Factor List` samt sætter hak i `Plots`

og Box-plots laves i
`Graphs/Graph Builder`, hvor `vitd` trækkes over på Y-aksen, og `country` trækkes over på X-aksen.

Ensidet ANOVA i SPSS

Slide 37ff

Sædvanligvis benyttes ikke `Analyze/Compare Means/One-Way ANOVA`, da denne ikke giver estimater.

Brug i stedet `Analyze/General Linear Model/Univariate`, hvor vi sætter `vitd` i `Dependent Variable` og `country` i `Fixed Factor(s)`.

Her kan vi vælge at se parameterestimater under `Options`, hvor vi afkrydser `Parameter estimates`.

Dog benyttes `Analyze/Compare Means/One-Way ANOVA` til Welch test, se s. 48.

Test af identiske spredninger og Welch test

Slide 45, 48

Til test af identiske spredninger (Levenes test) benytter vi:
`Analyze/General Linear Model`, se s. 108, men nu går vi også ind under `Options` og afkrydser
`Homogeneity of variance test`

Hvis spredningerne afviger betydeligt fra hinanden, benyttes Welch test for identittet af middelværdier:
Her er man nødt til at bruge
`Analyze/Compare Means/One Way ANOVA`, med `vitd` i `Dependent List` og `country` i `Factor`, og under `Options` afkrydses `Welch`

Modelkontrolplots for Vitamin D eksemplet

Slide 44, 47

I opsætningen af `Analyze/General Linear Model/Univariate` (se s. 108) kan vi i `Options` vælge `Residual plot`, men...

Vi kan også vælge at få mere kontrol over tingene ved at gemme residualer og predikterede værdier. Dette gøres ved at benytte `Save`-knappen og under `Predicted Values` vælge `Unstandardized`, og under `Residuals` f.eks. vælge `Studentized`

Herefter benyttes `Graph`-menuerne.

Tukey korrektion for vitamin D

Slide 55-58

Dette kunne gøres ved at gå ind i
`Analyze/Compare Means/One Way ANOVA` eller
`Analyze/General Linear Model`, sætte `vitd` i
`Dependent List` og `country` i `Factor`.

Efterfølgende går man så ind i `Post Hoc` og foretager relevante afkrydsninger, f.eks. `Tukey` eller `Games-Howell`

Non-parametrisk Kruskal-Wallis test

Slide 29, 61

Gå ind i menuen
`Analyze/Nonparametric Tests/Legacy Dialogs` og vælg
`K Independent Samples`, hvorefter `vitd` sættes i
`Test Variable List` og `country` sættes i `Grouping Variable`.

Under `Test Type` vælges `Kruskal-Wallis`.

Optælling af solvaner

Slide 64

Først skal vi definere variablen `sol` ved at benytte `Transform/Recode into Different Variables`, hvor `sunexp` sættes i `Numeric Variable` → `Output Variable`, `sol` sættes i `Name` og der klikkes `Change`.
Herefter klikkes `Old and New Values`, hvorefter man sætter 1 i `Value` og 0 i `New value`, 2 i `Value` og 1 i `New value`, 3 i `Value` og 1 i `New value`, `Continue`

Endelig benyttes `Analyze/Descriptive Statistics/Crosstabs/`, hvor `country` sættes over i `Row(s)` og `sol` i `Column(s)`. I `Cells` afkrydses `Observed` og `Percentages: Row`.

Box-plot, opdelt efter to kategorier

Slide 67

Benyt `Graph/Chart Builder/Bar`, vælg nr. 2 fra venstre. Sæt nu `country` på X-aksen, `vitd` på Y-aksen og `sol` over i `Cluster on X: set color`.

Der klikkes nu på `X-axis` i `Element Properties`, og under `Order` markerer man `DK` og `EI`, hvorefter man trykker "`fjern`" (det røde kryds).

Additiv tosidet ANOVA

dvs. uden interaktion

Slide 70-71

Brug `Analyze/General Linear Model/Univariate`, hvor vi sætter `vitd` i `Dependent Variable` og såvel `country` som `sol` i `Fixed Factor(s)`.

Derefter går vi i `Model`, vælger `Custom` og sætter såvel `country` som `sol` over i `Model`.

For at få parameterestimater med, går vi i `Options`, hvor vi afkrydser `Parameter estimates`

Under `Save` kan vi gemme de predikterede værdier. Dette opretter en ny variabel `PRE_1` i datasettet.

Figur af additiv tosidet ANOVA

Slide 75

De predikterede værdier `PRE_1` som lige er blevet gemt (s. 115) benyttes nu i et plot ved at gå i `Graphs/Graph Builder`, hvor vi vælger `Line` (nr. 2 fra venstre). Her sætter vi `PRE_1` på Y-aksen, `country` på X-aksen og `sol` i `Color`.

Der klikkes nu på `X-axis` i `Element Properties`, og under `Order` markerer man `DK` og `EI`, hvorefter man trykker "`fjern`" (det røde kryds).

Der kan også være behov for at ændre Y-aksen, idet den indrettes efter de enkelte målinger.

Tosidet ANOVA med interaktion

Slide 83-84, 88-89

Vi starter med opsætningen som på s. 115, men under `Model` tilføjer vi nu et interaktionsled ved at markere både `country` og `sol` samtidig og under `Type` vælge `Interaction` før de tilføjes til `Model`

Rækkefølgen af effekterne `country`, `sol` og `country*sol` i `Model`-vinduet kan give forskellig information i output (men samme model).