

Faculty of Health Sciences

Basal Statistik - SPSS

Den generelle lineære model.

Lene Theil Skovgaard

18. marts 2019



APPENDIX

med instruktioner til **SPSS**-analyse svarende til nogle af slides

- ▶ Biokemisk iltforbrug, transformationer, s. 3-10
- ▶ Udglatning og polynomie-fit, s. 11-12
- ▶ Lineære splines, s. 13-15
- ▶ Ikke-lineær regression, s. 9, 16
- ▶ Vitamin D, GLM, s. 17-22
- ▶ Interaktioner, s. 23-25
- ▶ Blodtryk og fedme, s. 26-28



Data vedr. biokemisk iltforbrug

Slide 7-8

Omstrukturering fra 6 linier med oplysning om alle 4 flasker den pågældende dag til 24 linier med kun en enkelt måling af iltforbrug pr. linie:

Benyt `/data/Restructure` og afkryds

Restructure selected variables into cases, klik Next.

Under How many variable groups? afkrydses One, klik Next,

Under Case Group Identification skiftes til

Use selected variable, og days sættes over i Variable.



Omstrukturering, fortsat, samt figur

Slide 7-8

Herefter går man ind i `Variables to be Transposed`, skriver `boc` i `Target Variable` og sætter `boc1`, `boc2`, `boc3` og `boc4` i det store felt, hvorefter der klikkes `Next`.

Endelig går man i `Create index variable` og sætter kryds ved `None` og klikker `Next`, og under `Variables to Cases: Options` klikkes blot `Finish`

Figuren slide 8

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



Analyse af biokemisk iltforbrug

Slide 11

Definition af nye variable:

Først skal vi logaritmetransformere iltforbruget. Dette gøres under Transform/Compute, hvor man som Target Variable sætter det nye variabelnavn, her logboc, og i feltet Numeric Expression skriver $\ln(\text{boc})$

Dernæst skal vi udregne den reciproke tid, igen under Transform/Compute, hvor man som Target Variable sætter invdays, og i feltet Numeric Expression skriver $1/\text{days}$

Derefter følger man opskriften på simpel lineær regression, se de næste sider



Regressionsanalyse med transformerede variable

Slide 11

Gå ind i menuen Analyze/Regression/Linear, og i boksen sættes logboc som Dependent og invdays som Independent(s)

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



Plot med regressionslinie

Slide 12

Først laves plottet ved at gå ind i Graph/Chart Builder/Scatter og i den fremkomne boks trækker man logboc over på Y-aksen, og invdays over på X-aksen.

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

Man kan endvidere vælge, om man vil have liniens ligning skrevet på linien eller ej (flueben ved Attach label to line kan fjernes) Farven på linien kan vælges under Lines: klik på farven og Apply/Close



Tilbagetransformeret fit

Slide 14-15

I regressionsopsætningen fra s. 7 klikkes på Save, hvorefter man under Predicted Values vælger Unstandardized, som så kommer til at ligge i datasættet under navnet PRE_1.

Disse tilbagetransformeres til oprindelig skala ved at benytte Transform/Compute og som Target Variable sætte f.eks. predikteret og i Numeric Expression skrive $\exp(\text{PRE}_1)$.

Gå nu ind i Graph/Chart Builder, vælg Scatter, og træk i den fremkomne boks predikteret over på Y-aksen og days over på X-aksen. Dobbeltklik på den fremkomne graf og klik på ikonet Add Fit Line at Total og derefter i Properties-boksen afkrydse Spline og klikke Apply.

Det er formentlig meget besværligt (eller umuligt) at få såvel kurve som observationer på samme figur.....



Ikke-lineær regression

Slide 16

Her benyttes Analyze/Regression/Nonlinear Regression, hvor man sætter `boc` i Dependent og i Model Expression skriver selve det ikke-lineære udtryk `gamma*exp(-beta/days)`.

Herefter klikker man i Parameters og tilføjer

- ▶ Name: `gamma`, Starting Value: 228 Add
- ▶ Name: `beta`, Starting Value: 0,8 Add

og til sidst: Continue

Jeg kan dog ikke pt få det korrekte resultat....



Plot af ikke-lineært fit

Slide 17

Hvis proceduren s. 9 fungerede korrekt, ville man benytte Save og gemme de predikterede værdier ved at afkrydse i Predicted Values.

Herefter benyttes Graphs/Chart Builder/Scatter (det simple længst til venstre), days sættes på X-aksen og PRED_1 på Y-aksen.

Efterfølgende dobbeltklikkes på grafen, og man vælger Add Interpolation Line, hvorefter man afkrydser i Spline.

Det ser dog ikke umiddelbart ud til at være muligt at få usikkerheder på disse prediktioner, så grænserne må man undvære. Ligeledes kommer selve observationerne heller ikke med på figuren.



Udglattet kurve, en såkaldt *Loess*-kurve

Slide 22

Her skal vi begrænse det datamateriale, vi ser på. Dette gøres i Data/Select Cases, hvor der afkrydses i If og skrives `5<age<20 & sex='male'`.

Herefter laver vi en tegning med Graphs/Chart Builder, hvor vi vælger det simple Scatter. Sæt age over på X-aksen og sigf1 over på Y-aksen.

Ved efterfølgende at dobbeltklikke på grafen, kan man lægge en udglattet kurve på ved at klikke på Add Fit Line at Total og derefter i Properties-boksen afkrydse Loess/Apply:

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



Polynomiale fit

Slide 26

De 3 forskellige polynomier (af grad hhv 1,2 og 3) kan formentlig kun (på rimelig simpel vis) tegnes hver for sig. Ellers skal man til at have flere versioner af datasættet sat i forlængelse af hinanden.....

Først laver vi en tegning med Graphs/Chart Builder, hvor vi vælger det simple Scatter, sætter age over på X-aksen og sigf1 over på Y-aksen.

Ved efterfølgende at dobbeltklikke på grafen, og videre på Add Fit Line at Total kan man i Properties-boksen afkrydse enten Linear, Quadratic eller Cubic

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



Lineære splines

Slide 27-32

Definition af nye variable:

Benyt Transform/Compute, hvor man som Target Variable sætter det nye variabelnavn, her f.eks. `extra_age12`, og i feltet Numeric Expression skriver `max(age-12,0)`

Ligeledes med de 3 øvrige variable.

Fit derefter en model med disse 4 ekstra kovariater, sammen med den oprindelige age ved at benytte menuen

Analyze/Regression/Linear, og i boksen sættes `ssigf1` som Dependent og age samt alle 4 extra-variable i Independent(s)

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



Illustration af fit med lineære splines

Slide 29

Det er nok virkelig indviklet (eller umuligt?) at få data og splines på samme figur....

Her er man nødt til først at gemme de predikterede værdier ved at benytte Save og vælge Predicted Values. De kommer til at hedde PRE_1 i datasættet.

Herefter går man i Herefter laver vi en tegning med Graphs/Chart Builder, hvor vi vælger Lines (det simple længst til venstre), og dobbeltklik det op i det store felt. Sæt PRED_1 over på Y-aksen og age over på X-aksen.



Alternativ parametrisering af lineære splines

så man i stedet får estimater for
hældningerne i de enkelte aldersintervaller

Slide 30-31

Definition af de alternative variable:

Benyt Transform/Compute, hvor man som Target Variable sætter det nye variabelnavn, her f.eks. `ny_age12`, og i feltet Numeric Expression skriver `min(extra_age12,1)`

Ligeledes med de 3 øvrige variable: `min(age,10)`,
`min(extra_age10,2)` og `min(extra_age13,2)` og
`ny_age15=extra_age15`

Fit derefter en model med disse 5 kovariater, i analogi med den beskrevne fremgangsmåde på forrige side.



Ikke-lineær regression

Slide 37-40

Her benyttes Analyze/Regression/Nonlinear Regression, hvor man sætter konc i Dependent og i Model Expression skriver selve det ikke-lineære udtryk $\text{beta} * (1 - \exp(-\text{gamma} * \text{tid}))$.

Herefter klikker man i Parameters og tilføjer

- ▶ Name: gamma, Starting Value: 0,05 Add
- ▶ Name: beta, Starting Value: 2000 Add

og til sidst: Continue

Mht figur af fit, se s. 10



ANOVA, sammenligning af D-vitamin i 4 lande

Slide 46

Først skal vi udregne 2-tals logaritmen af vitamin D. Dette gøres under Transform/Compute, hvor man som Target Variable sætter det nye variabelnavn, her lvitd, og i feltet Numeric Expression skriver $\text{LG10}(\text{vitd})/\text{LG10}(2)$

Til variansanalysen benyttes

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

Her kan vi vælge at se parameterestimer under Options, hvor vi afkrydser Parameter estimates)



Model med bmi som kovariat, ANCOVA

Slide 49

Til kovariansanalyse benyttes

Analyze/General Linear Model/Univariate, hvor vi sætter lvidt i Dependent Variable, country i Fixed Factor(s) og bmi i Covariate(s)

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

Husk også i Options at vælge Parameter Estimates



Model med 4 kovariater

Slide 53-54

Selve analysen foregår som skitseret på s. 19, altså: Benyt Analyze/General Linear Model/Univariate, hvor vi sætter lvitd i Dependent Variable, country og sunexp i Fixed Factor(s) og såvel bmi som lvitdintake i Covariate(s)

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

Husk også i Options at vælge Parameter Estimates



Modelkontrol for model med 4 kovariater

Slide 62-63

Den automatiske modelkontrol i SPSS er ikke særlig god.....

Brug i stedet nedenstående:

I regressionsopsætningen fra s. 20 klikkes på Save, hvorefter man typisk vil vælge

- ▶ Under Predicted Values: vælg Unstandardized, som så kommer til at ligge i datasættet under navnet PRE_1
- ▶ Under Residuals er der 5 muligheder, f.eks.
 - ▶ Unstandardized: De *sædvanlige*, altså observeret minus forventet, kaldet RES_1 i datasættet
 - ▶ Studentized deleted: Normerede Press-residualer (altså hvor den pågældende observation er udeladt ved fit af modellen), kaldet SDR_1 i datasættet



Modelkontrol, II

Slide 62-63

Ud fra det udbyggede datasæt dannet s. 21 kan vi nu lave diverse figurer:

- ▶ Benyt Graph/Chart Builder, vælg Scatter, klik Unstandardized Predicted Values over på X-aksen og Unstandardized Residuals over på Y-aksen.
- ▶ Vælg Analyze/Descriptive Statistics/Q-Q Plots og sæt Unstandardized Residuals over i Variables.
- ▶ Benyt Graph/Chart Builder, vælg Histogram og klik det ønskede residual (f.eks. RES_1) ned på X-aksen. Klik så på Distribution Curve og afkryds Normal, klik Apply og Close
- ▶ Evt. yderligere check af varianshomogenitet, se i appendix fra regression, s. 11



Udeladelse af flere kovariater samtidig

Slide 67-68

Det er ikke så let at gennemskue, hvordan sette foregår i SPSS....

Hvis alle kovariater var kvantitative (altså i en multipel regressiosanalyse) kan man i `Analyze/Regression/Linear` definere flere modeller på en gang og få dem sammenlignet med et F-test.

Men når vi som her også har kategoriske kovariater, er vi nødt til bruge `Analyze/General Linear Model/Univariate`, og denne opsætning tillader ikke definition af flere modeller på en gang.

I så fald er man nødt til at lave sine kategoriske variable om til dummy-variable, f.eks. skal country så blive til 3 dummy-variable....



Interaktionen *sunexp*lvitdintake*

Slide 70

Følg opskriften fra s.20 men i Model markeres *sunexp* og *lvitdintake* samtidig, der skiftes fra Main Effects til Interaction og klikkes på pilen og Continue.

For at få de enkelte effekter af *lvitdintake* skal denne udelades af Model, medens interaktionen bibeholdes.



Interaktionen *country*sunexp*

Slide 71-73

Følg opskriften fra s.20 men i Model markeres sunexp og country samtidig, der skiftes fra Main Effects til Interaction og klikkes på pilen og Continue.

For at få de enkelte effekter af sunexp skal denne udelades af Model

Interaktionen *country*lvitdintake*

Slide 74

Følg opskriften fra s.20 men i Model markeres *country* og *lvitdintake* samtidig, der skiftes fra Main Effects til Interaction og klikkes på pilen og Continue.

For at få de enkelte effekter af *lvitdintake* skal denne udelades af Model, medens interaktionen bibeholdes.



Blodtryk vs. fedme

Sammenligning af mænd og kvinder, figurer

Slide 79

Selve figuren laves i Graph/Chart Builder, hvor man vælger Scatter plot (nr. 2 fra venstre), og i den fremkomne boks trækker man $\log_{10}bp$ over på Y-aksen, $\log_{10}obese$ over på X-aksen og sex over i Set Color

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

Disse linier vil ikke blive parallelle, og er ikke vist i forelæsningsnoterne.



Blodtryk vs. fedme og køn: Kovariansanalyse

Slide 80

Her benyttes Analyze/General Linear Model/Univariate, hvor vi sætter log10bp over i Dependent Variable, sex i Fixed Factor(s) og log10obese i Covariate(s)

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

Husk også i Options at vælge Parameter Estimates



ANCOVA-plot, med parallelle linier

Slide 81

Her er man nødt til først at gemme de predikterede værdier ved at benytte Save og vælge Predicted Values. De kommer til at hedde PRE_1 i datasættet.

Herefter går man igen ind i Graph/Chart Builder/Scatter, vælger Scatter plot (nr. 2 fra venstre), og i den fremkomne boks udskiftes log10bp med PRE_1.

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

Desværre får man ikke samtidig selve observationerne på figuren.....

Og de to lodrette linier er heller ikke forsøgt indtegnet.

