

Faculty of Health Sciences

Basal Statistik

Logistisk regression mm.

Lene Theil Skovgaard

25. marts 2019



APPENDIX vedr. SPSS

svarende til diverse slides:

- ▶ To-gange-to tabeller, s. 3
- ▶ Plot af binære data med loess-udglatning, s. 4
- ▶ Logistisk regression, s. 5-7
- ▶ Modelkontrol, s. 8-10, 14
- ▶ Diagnostics, s. 11
- ▶ Alternative links, s. 12-13
- ▶ Ordinale data, s. 15-16
- ▶ Tælleletal, s. 17-20



To-gange-to tabeller

Slide 7-10

Da vi ikke har individ-data, men en kolonne med antal, skal vi lave en vægtet analyse. Der benyttes Data/Weight Cases/, afkrydses i Weight cases by, hvorefter antal flyttes over i Frequency Variable.

Herefter tabelanalysen i

Analyze/Descriptive Statistics/Crosstabs, hvor gender sættes i Row(s) og farveblind sættes i Column(s).

Det er vigtigt, hvad der er rækker og søjler

Herefter går man i Statistics og afkrydser Chi-square og Risk samt i Cells, hvor man afkrydser Expected samt under Percentages: Row



Figurer

Slide 13

Benyt Graph/Chart Builder/Scatter/Grouped Scatter, sæt infektion i Y, alder i X og over2timer i Set color. Herefter OK, hvorved figuren fremkommer uden udglattet kurve.

For at lægge den udglattede kurve på ser det ud til at være nødvendigt først at lave variablen `infektion` om til en `scale-variable`, hvis den ikke var det i forvejen. Dette gøres i `Variable View`. Når man så laver grafen, kan man efterfølgende dobbeltklikke på den, klikke `Add Fit Line at Subgroups`, afkrydse `Loess` og `Apply/Close`.

Det kommer dog ikke helt til at ligne den tilsvarende SAS-figur



Logistisk regression i SPSS

Slide 23-24

For at lave den logistiske regression benytter vi `Analyze/Regression/Binary Logistic`, hvor vi sætter `infektion` som `Dependent`, begge forklarende variable (`alder` og `over2timer`) i `Covariates`.

Herefter går vi ind i `Categorical Covariates` og sætter `over2timer` ind her, da den ikke indeholder numerisk information. Her kan man vælge, om referencen skal være første eller sidste kategori. Vælg f.eks. `First` og klik `Change/Continue`.

Herefter går vi ind i `Options` og afkrydser `CI for  $\exp(\beta)$` .



Andre metoder til logistisk regression

Slide 23-24

Der finde i SPSS en meget generel indgang til analyse af mange forskellige slags data, nemlig Generalized Linear Models, hvor man skal udfylde en del faneblade (dem foroven i det grå felt)

Først udfylder man på indgangssiden, hvor man afkrydser Custom, vælger Binomial under Distribution og Logit under Link function.

Derefter benyttes de andre *faneblade*:

- Response: Her sættes infektion ind under Dependent Variable, og under Reference category vælges First (lowest value), fordi dette bevirker, at vi vurderer udfald 1 mod udfald 0, altså risikoen for at *få* en infektion, ikke for at gå fri af en infektion.



Andre metoder til logistisk regression, II

Slide 23-24

Fortsættelse af faneblade fra s. 6

- ▶ Predictors: Her sættes alder i Covariates og over2timer sættes i Factors (og her kan man i Options afkrydse descending under Category Order for Factors for at få 0 som reference.
- ▶ Model: Her sættes begge variable ind som Main Effects
Senere også som Interaction
- ▶ Statistics: Afkryds
Include exponential parameter estimates



Goodness of fit

Slide 36-38

I opsætningen af den logistisk regressionsanalyse (se s. 5) benyttes Options, hvor man afkrydser Hosmer-Lemeshow Goodness-of-fit

Check af linearitetsantagelsen for alder

Slide 39-42

- ▶ Alderseffekt modelleret med både alder og $\ln(\text{alder})$ (output s. 39-40). Definer:
 $\ln\text{alder50} = \ln(\text{alder}) / \ln(1.1) - \ln(50) / \ln(1.1);$
og indsæt denne variabel i Covariate(s) sammen med de øvrige, se s. 5
- ▶ Alderseffekt modelleret som lineær spline (output s. 41-42):
Definer $\text{alder_over60} = (\text{alder} > 60) * (\text{alder} - 60)$ og
 $\text{alder_over75} = (\text{alder} > 75) * (\text{alder} - 75)$
og indsæt denne variabel i Covariate(s) sammen med de øvrige, se s. 5



Plot af standardiserede residualer

Slide 43

Modellen er den på s. 23, og i opsætningen af den logistiske regression (se s. 5 her i appendix),
og man gemmer diverse størrelser ved at klikke Save og afkrydse de ønskede størrelser, f.eks. Predicted probabilities, Cook's, Standardized residuals og DfBeta(s)

Datasættet indeholder nu disse ekstra størrelser, så man kan nu selv styre plottene ved at benytte
Graph/Chart Builder/Scatter/Grouped Scatter

Slide 44

Plot af kumulerede residualer synes *ikke* at være tilgængeligt i SPSS



Diagnostics

Slide 46-47

Modellen er den på s. 23 (øverste del), se vejledning her i appendix, s. 5, hvor vi har yderligere har gemt diverse størrelser ved at klikke Save og afkrydse de ønskede størrelser, her Cook's, og DfBeta(s), se s. 10

Ud fra det udbyggede datasæt kan man nu selv styre plottene ved at benytte Graph/Chart Builder/Scatter/Grouped Scatter, hvor man sætter enten cook eller DfBetas i Y, alder i X og over2timer i Set color. Herefter OK, hvorved figuren fremkommer uden udglattet kurve.

For at lægge kurver på, se s. 4



Relativ risiko for farveblindhed

Slide 51-52

Da vi ikke har individ-data, men en kolonne med antal, skal vi lave en **vægtet** analyse. Der benyttes Data/Weight Cases/, afkrydses i Weight cases by, hvorefter antal flyttes over i Frequency Variable.

Risk Ratio (relativ risiko) estimeres ved at benytte **log-link**, og vi må derfor benytte den generelle metode Generalized Linear Models, (se s. 6-7), men hvor man denne gang vælger Log under Link function.



Test for trend

Slide 55-56

Her benytter vi Generalized Linear Models, hvor man afkrydser Custom, vælger Binomial under Distribution og Identity under Link function.

Under Response sættes kejsersnit over i Dependent Variable, og der afkrydses i Number of events occurring in a set of trials samt i Variable, hvor man så sætter total over i Trials Variable. Under Predictors sættes skonummer i Covariates og ellers fortsættes som s. 6-7.



Modelkontrol af lineariteten, kejsersnit

Slide 57-58

Check af linearitet kan foretages ved at lave en kopi af skonummeret: Benyt Transform/Compute og definer `skostr=skonummer`.

I modellen fra s. 14 tilføjes nu `skostr` som Factor, og under Statistics/Chi-square Statistics skiftes fra Wald til Likelihood Ratio.



Proportional odds model

Slide 66-67

Logistisk regression for hver tærskel, med **samme afhængighed** af kovariaterne, som her alle er $\log_2$ -transformerede.

Dette udføres nemt i SPSS ved at benytte Analyze/Regression/Ordinal, hvor vi sætter degree_fibr som Dependent og alle 3 forklarende variable (lha, lpiinp og lykl40) i Covariates.

For at få tilbagetransformeret til Odds Ratio'er, er man nødt til at benytte **Generalized linear models** (se s. 6-7), hvor man vælger Ordinal Response og videre Ordinal logistic



Proportional odds model, fortsat

Slide 66-67

For at få et test for *proportional odds*, går vi ind i Output og afkrydser Test of parallel lines.

Herved får vi dog ikke helt det samme, da SPSS laver et Likelihood Ratio test, mens SAS laver det som et *score*-test.

Man kan endvidere afkrydse

Estimated response probabilities og Predicted category



Poisson analyse

Slide 81-82

Vi skal her igen benytte

Analyze/Generalized Linear Models/Generalized Linear Models
hvor man går videre med Counts/Poisson loglinear.

Herefter udfylder man fanerne således

Response: total sættes i Dependent Variable

Predictors: gender sættes i Factor og years_from_2004
sættes i Covariates

Model: husk interaktion her, hvis den skal med

Estimation: giver mulighed for overspredning i modellen, hvis
man afkrydser Robust estimator

Statistics: Afkryds
Include exponential parameter estimates



Poisson analyse

Slide 84-85

Korrektion for overspredning foregår ved at afkrydse
Robust estimator under Estimation. Ellers helt som s. 17



Normalfordelingsmodel for tælledata

Slide 86

her benytter vi logaritmetransformererede data
($\log_{10}(\text{total})$):

og derefter blot Analyze/General Linear Model/Univariate,
hvor vi sætter $\log_{10}(\text{total})$ i Dependent Variable, gender i
Fixed Factor(s) og years_since_2004 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



Opdelte plot (panels)

Slide 88

Benyt Graph/Chart Builder/Groups/Point ID, afkryds Rows panel variable og sæt gender ind i denne, og sæt nu antal i Y, years_from_2004 i X og age i Set color. Herefter OK

Slide 89

Benyt Graph/Chart Builder/Groups/Point ID, afkryds Rows panel variable og sæt year ind i denne, og sæt nu antal i Y, age i X og gender i Set color. Herefter OK