

Statistik 2 2010, 3.-7.5.2010

Stansens PC-klass ASA-huset.

Material del II

Mera om partiella korrelationer och skattning av kovariansstrukturer

Vi fortsätter med mattedatat från förra modulen och analyserar det med programmet MIM som är skräddarsytt för analys av grafiska modeller av olika slag. Datat finns formaterat för MIM i filen MATHMARK.DAT

Programmet utgår alltid i en sökning först från den modell som användaren senast deklarerat. Deklarera nu den fullständiga modellen (alla variabler beror på alla andra) genom att välja Model-Saturated model. Gör nu en sökning på en optimal modell med Select-Stepwise (acceptera default-valen). När sökningen är avslutad, skriv upp modellbeskrivningen (efter Selected model står det vilka grupper av variabler som bildats). Rita sedan upp en graf av den valda modellen (Graphics, Graph).

Deklarera nu igen en modell där samtliga variabler är oberoende av varandra (Model-Main effects model). Gör sedan sökningen med Forward förkryssat och jämför resultaten (modellbeskrivning samt bilden) med de som erhöles genom bakåt-sökningen. Finns det skillnader mellan resultaten?

Gör till sist en totalsökning med Select-Global (kryssa för Verbose output). Jämför resultatet med de tidigare. Finns det någon skillnad? Jämför även med insikter vi fick tidigare om datat med B-Course (datat diskretiserades för analysen) och parvisa samt partiella korrelationer som sådana.

I följande MIM datafil BoneMineralContent.dat finns det 137 observationer på 6 variabler från en estrogen-effekt studie. Syftet är att se vilka faktorer som påverkar mineralhalten i benstommen. Datat innehåller som kovariater BMI (body mass index), ålder och enzyminhalt (alkaline).

Definiera den fullständiga modellen och tvinga sedan kopplingarna mellan kovariaterna genom kommandot *fix UVW*. Gör sedan en bakåtsökning (Select-Stepwise) med default-värden och titta på resultaten med hjälp av grafen. Upprepa proceduren (deklarerat och fixa först) genom att använda BIC och birectional sökning. Titta igen på resultaten. Finns det någon skillnad till de tidigare erhållna länkarna mellan variabler?

Logistisk regression och ordinal regression

En 'vanilla' förklaring av logistisk regression finns här:

<http://luna.cas.usf.edu/~mbrannic/files/regression/Logistic.html>

Logistisk regression är en mycket vanlig metod för regressionsanalyser där responsvariabeln är dikotom (representerar två kategorier).

Öppna filen bd1.sav. Den innehåller data om cancerfall och kontrollindivider m.a.p. ålder och alkohol- samt tobaksförbrukning. Anpassa först en logistisk regressionsmodell med dessa tre förklarande variabler och case som en responsvariabel. Hur blir slutsatserna? Betrakta även graden av korrekt klassificerade fall p.b.a. modellen. Anpassa sedan en modell där alc2 och tob2 har även interaktionsterm.

Principen bakom ordinal regression som en generalisering av logistisk regression beskrivs på ett icke-tekniskt sätt här:

http://www.norusis.com/pdf/ASPC_v13.pdf

Vi använder nu ordinal regression i SPSS på datat i filen FYSISKNÄRHET.sav. Syftet är att undersöka om upplevd brist på fysisk närhet verkar ha en koppling till negativa självkänslor som vuxen (BSI scores). Vi börjar med att titta på responsvariabeln 'Känslor av värdelöshet'. Sätt könet i modellen som en faktor och de variabler som representerar upplevd fysisk närhet under barndomen som förklarande variabler. Vilka variabler verkar ha en effekt och hur ser effekten ut? Verkar det finnas skillnad mellan könen? Anpassa nu en ny modell där könet och endast de förklarande variabler som hade en signifikant effekt tas med. Jämför effekterna med de som erhöles tidigare. Upprepa analysen för övriga BSI scores.

Mixed models

'Mixed models' är ett begrepp som syftar till regressionsmodeller där det ingår både fixa och slumpmässiga förklarande variabler (faktorer/kovariat). Dessa modeller kan vara linjära, generaliserade linjära eller icke-linjära. Med den kategoriseringen syftar man till ett hårdant samband det antas finnas mellan responsvariabler och förklarande variabler. Modellerna används i många olika sammanhang, bl a är longitudinella studier ett viktigt tillämpningsområde. Följande länk bjuder på en icke-teknisk presentation av en mixed model i dylikt sammanhang:

<http://lme4.r-forge.r-project.org/slides/2009-07-07-Rennes/3LongitudinalD.pdf>

En allmän introduktion till mixed models för experimentella data av Douglas Bates finns här:

http://web.abo.fi/fak/mnf//mate/jc/miscFiles/LME_Bates.pdf

Bates har gjort ett R-program (lme4) för anpassning av mixed models som klarar i regel av mer krångliga och stora data än mixed models modulen i SPSS.