

Statistik 2, del 7:

## Faktoranalys

Faktoranalys beskrivs i nötskal här:

<http://faculty.chass.ncsu.edu/garson/PA765/factor.htm>

Se även beskrivning av Strukturella EkvationsModeller (SEM):

<http://faculty.chass.ncsu.edu/garson/PA765/structur.htm>

Läs in filen Uscrime.sav. Filen innehåller uppgifter om diverse brottsfrekvenser i olika delstater i USA. Sammanlagt finns uppgifter om följande sju brottstyperna: murder, rape, robbery, assault, burglary, larceny, autotheft.

Vi skall använda explorativ faktoranalys för att studera datat. Notera skillnaden mellan explorativ och konfirmatorisk faktoranalys – i den senare specificerar man modellstrukturen i förväg och testar dess lämplighet för ett visst datamaterial. I explorativ faktoranalys försöker man urskilja strukturer i data med hjälp av faktormodeller som identifieras som lämpliga. Ofta utnyttjar man Scree-plot från principalkomponentanalys som ett första steg, för att avgöra hur stort antal faktorer kan vara lämpligt för datamaterialet.

Välj dimension reduction – factor analysis, metoden principal factor analysis och välj sedan antalet faktorer lika med 2. Vi bestämmer att inte använda rotering av lösningen.

Titta på skattningar av *communality* för varje variabel. Ett högt värde på detta indikerar att de allmänna faktorerna förklarar största delen av variationen för variabeln i fråga. Som en indikation på modellens anpassningsgrad, kan man betrakta resultatet från ett Chi<sup>2</sup>-test, där nollhypotesen motsvarar den specificerade faktormodellen och mothypotesen att inga begränsningar läggs på kovariansmatrisen för variablerna. Ett signifikant testresultat (på nivån  $\alpha$ ) innebär att modellens struktur inte stöds av datat. Det bör dock observeras att med väldigt stora datamaterial erhålls signifikanta resultat i regel för alla tänkbara restriktioner på kovariansmatrisen. Datat här i uppgiften är förhållandevis litet.

Titta sedan på faktorladdningarna (Factor loadings) för variablerna. Går det att urskilja något tydligt mönster? Ett intressant mönster i faktoranalyssammanhang är att delgrupp av variabler får relativt höga laddningsvärden på en viss faktor, medan alla de andra variablerna får relativt låga laddningsvärden på just denna faktor.

För att förenkla tolkningen av faktorladdningarna, upprepas nu samma analys som ovan, men med en viss rotering av faktorerna enligt sk Varimax-metoden. Upprepa stegen ovan, men använd Factor rotation – Varimax. Titta på skattningarna och laddningarna, går det nu att urskilja ett tydligare mönster bland laddningarna? Vi ser att Assault och Murder får höga laddningar på en faktor, samt att Burglary, Larceny och Autotheft får höga värden på den andra faktorn. Dessa faktorer motsvarar alltså olika dimensioner m.a.p. variation i datat, t ex den senare faktorn kunde tolkas representera egendomsbrott.

Notera att i Varimax metoden tvingas faktorerna att vara oberoende av varandra, dvs. de är ortogonala i geometrisk mening. I en sk oblique rotering, tillåts faktorerna att vara korrelerade. Det finns möjlighet att erhålla en oblique rotering med hjälp av metoden Promax. Upprepa analysen en gång med Promax metoden och jämför faktorladdningarna. Bildas här samma mönster som med Varimax roteringen?

I faktoranalys kan man skatta de icke-observerade värdena på faktorerna för varje "individ" i datamaterialet. Detta görs genom att spara Factor scores, då man valt faktoranalys från menysystemet. Upprepa de två faktoranalyserna ovan och välj att erhålla faktorpoängen (spara dem med olika namn i den andra analysen, så att värden inte skrivs över). Rita upp faktorpoängen med hjälp av Graphs-Scatterplot. Notera hur deras relation skiljer sig mellan Varimax (oberoende faktorer) och Promax (korrelerade faktorer) roteringarna.

Öppna datafilen AidsPatienter.sav. Filen innehåller uppgifter om hur aids-patienter har betraktat sina vårdande läkare i en frågeformulär med 13 olika frågor. Vi undersöker bl a hur sekventiellt test av faktormodeller kan leda till valet av antalet faktorer. Anpassa faktormodellen utan rotering med 1,2,3 och 4 faktorer och notera p-värden för  $\chi^2$ -testet. Man ser att modellen med tre faktorer blir nätt och jämt acceptabel på 5%-nivån. Använd sedan tre faktorer och anpassa modellen både med Varimax och Promax roteringen. Titta på laddningsmönstren. Det kan konstateras att inget enkelt struktur kan lätt hittas för datat med hjälp av faktoranalysen, eftersom flera variabler laddas relativt högt på fler än en faktor och för att somliga variabler laddas lågt på samtliga faktorer.