

Statistik 2 2008, 15.4.08

Övning 3a

Hämta följande datafil: <http://www.abo.fi/fak/mnf/mate/jc/statistik2/EconomicActivity.dat>

Filen innehåller 665 observationer på 8 dikotoma variabler från en studie, där engelska hushåll i Rochdale blev kartlagda m.a.p. olika indikatorer (kort beskrivning av variablerna finns i filen). Starta programmet MIM och läs in filen med File-Input. Deklarera en modell där samtliga variabler är oberoende av varandra (Model-Main effects model). En beskrivning av den giltiga modellen kan alltid erhållas från Model-Show properties. Utskrift av resultat kan styras till extern fil genom att välja File-Save Output.

Vi undersöker nu vilka variabler verkar vara kopplade genom att lägga till en länk mellan varje par av variabler i taget och genom att testa om datat stödjer den länken (dvs. är den signifikant). Programmet utgår alltid i en sökning först från den modell som användaren senast deklarerat. Välj Select-Stepwise och kryssa för Forwards & One step only. Vi ser att 8 av 28 par av variabler får icke-signifikanta länkar med 5% nivå. Deklarera nu den fullständiga modellen (alla variabler beror på alla andra) genom att välja Model-Saturated model. Vi undersöker nu istället vilka länkar som kan strykas mellan variablerna (Select-Stepwise, kryssa för One step only). Observera att bakåt-sökningen är default p.g.a. att det är statistiskt mer meningsfullt. Jämför resultatet med det tidigare erhållna.

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). Spara grafen till en extern fil från File menyn i graf-fönstret. 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. Hurdana skillnader finns det?

Gör nu om de två sökningarna, så att den ena startar från Main effects model och den andra från Saturated model, men kryssa för Bidirectional & BIC. Skriv upp modellbeskrivningen från bägge fallen och jämför dem med varandra. Hurdana skillnader finns det? Observera att rummet av samtliga möjliga modeller kan totalsökas med Select-Global. För detta material finns det 2^{28} (~268 miljoner) olika grafiska modeller, dvs. genomgången tar rätt lång tid...

Övning 3b

Hämta följande datafil: <http://www.abo.fi/fak/mnf/mate/jc/statistik2/ChromosomeMapping.dat>

Filen innehåller 70 observationer på 6 bialleliska markörer från en genetisk studie. Syftet är att bilda en genetisk karta, som indikerar ordningen på markörerna. Läs in datat till MIM och undersök den optimala ordningen genom att utföra likadana analyser som ovan. Hur ser resultaten ut jämfört med situationen i den föregående uppgiften? Observera att en totalsökning kan rimligen utföras för detta material, eftersom det finns ca 32000 modeller.