

## Statistik 1

### Övning 5

1. Hämta följande datafil: <http://www.abo.fi/fak/mnf/mate/jc/statistik1/Socialsurvey.sav>

Datamaterialet omfattar diverse variabler anknutna till graden av den upplevda livskvaliteten, samt bakgrundsfaktorer för individerna. Använd Descriptive statistics – Crosstabs för att utföra följande analyser (sätt kryss i 10 översta rutor i fliken Statistics och i Observed, Expected, Row & Column i fliken Cells):

1.1 Studera sambandet mellan 'General happiness' och 'Region of US'.

1.2. Studera mellan 'General happiness' och 'Region of US' skilt för män och kvinnor och jämför resultatet med det i 2.1. Vad tror Du är orsaken till skillnaden?

1.3. Studera sambandet mellan 'General happiness' och 'Take active part in world affairs', återigen först för hela materialet och sedan skilt för män och kvinnor. Slutsatser?

1.4. Studera sambandet mellan 'General happiness' och 'Illegal drugs', först för hela materialet och sedan skilt för män och kvinnor. Titta speciellt på värden på sambandsmåten. Är det motiverat här att använda sambandsmått avsedda för ordinalskalevariabler?

1.5. Studera sambandet mellan 'One's spouse being unemployed' och 'Is life exciting or dull' skilt för män och kvinnor. Slutsatser?

2. Vi jobbar vidare med samma datamaterial och undersöker sambandsstrukturen bland följande tre variabler: kön (Sex), intresse för omvärlden (Take active part in world affairs) och arbetslös och jobbsökande (Unemployed and looking for Work...). Man kan tänka sig att intresset har en koppling både till könet och om man är arbetslös.

Börja med att definiera samtliga tre variabler som kategoriska. För att leta reda på en sambandsstruktur som verkar rimlig för datat, använder vi oss av sk log-linjära modeller (Analyze, Loglinear, Model selection). Dessa modeller betraktas dock mer i detalj under kursen Statistik 2. Modellerna omfattar interaktionstermer med beteckningen '\*', exempelvis  $X * Y$  och  $X * Y * Z$ . Interaktionen  $X * Y$  innebär att variablerna  $X$  och  $Y$  inte är oberoende. Interaktionen  $X * Y * Z$  innebär att de tre variablerna alla har ett samband med varandra. En modell som saknar interaktionsterm mellan vissa par av variabler implikerar att de variabler är marginellt eller betingat oberoende (eller både och under vissa förutsättningar). Till exempel, modellbeteckningen  $X * Y, Z$  motsvarar det att variabeln  $Z$  är oberoende av både  $X$  och  $Y$ , samt att  $X$  och  $Y$  har ett samband.

Termen 'saturated model' innebär att samtliga variabler anses ha ett samband med alla övriga variabler som inkluderats i modellen.

Vi anger först 'Range' (Min = 1, Max = 2) för kön (Sex), intresse för omvärlden (Take active part in world affairs) och arbetslös och jobbsökande (Unemployed and looking for Work...). Klicka på Model och försäkra att 'saturated' har valts. Välj Continue och Ok. SPSS letar bland lämpliga modeller genom att starta från den modell där alla tre variabler har ett samband och stryker samband (=interaktionstermer) då de inte visar sig vara signifikant avvikande från 0 (=nollhypotes). Från utskrift ser vi till vilken modell sökningen konvergerar. Chi-square testet anger dess anpassningsgrad till datat.

För att förstå effekten av vald modell för jämförelser, startar vi nu sökningen istället från modell Unemployed \* Take part, Sex. Denna modell har den tolkning att arbetslöshetsfrekvensen och intresset för omvärlden inte skiljer sig mellan män och kvinnor, samt att intresset för omvärlden hänger ihop med det om man är arbetslös eller ej. Modellen definieras genom att lägga till Unemployed \* Take part som interaktionsterm och Sex endast som 'main effekt' -term (syftar till att variabeln endast betraktas för sig i modellen. Från utskrift ser vi till vilken modell SPSS nu konvergerar och att den skiljer sig från den tidigare modellen. Datat stödjer ett samband mellan Sex och Unemployed (arbetslöshetsfrekvensen skiljer sig bland män och kvinnor) och eftersom vår starthypotes enligt Unemployed \* Take part, Sex negligerar sambandet, kommer vi i den senare analysen felaktigt till slutsatsen att alla variablerna är oberoende av varandra. Vi måste alltså noggrant överväga hurdana hypoteser används som utgångsläge i sambandsanalys.