

Statistik 1 2009

Övning 6, v42

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

Importerera datafilen till R Commander (*Data-Import...-From SPSS...*).

En forskargrupp arbetar med att modellera nervrespons till diverse substanser. De använder råttan som en modellorganism. Datamaterialet omfattar mätvärden på nervrespons för en försöksgrupp med råttor som kommit i kontakt med insektgiften DDT, samt motsvarande värden för en kontrollgrupp. Datamaterialet är litet och man är tveksam om normalitetsantagandet vore giltigt för en dylik respons. Använd en boxplot (*Graphs-Boxplot, Plot by groups*) och ett icke-parametriskt test för att jämföra de två grupperna. Det lämpliga testet erhålls från *Statistics-Nonparametric...-Two-sample Wilcoxon...*

2. Hämta följande datafil: <http://www.abo.fi/fak/mnf/mate/jc/statistik1/socialsurvey.txt>

Importerera filen i menyn *Data-Import...-From text...* där *Field separator* skall vara **Tabs**. Namnge datamatrixen t ex **SocialSurvey**.

Datamaterialet omfattar diverse variabler anknutna till graden av den upplevda livskvaliteten, samt bakgrundsfaktorer för individerna. REGION indikerar den del av USA där respondenten bor. TAKEPART är en variabel som indikerar om man allmänt bryr sig om vad som händer i världen. HLTH5 indikerar om man använder droger. Utför följande analyser:

2.1 Studera sambandet mellan HAPPY och REGION. Använd *Statistics-Contingency tables-Two-way table*. Lägg HAPPY som *Row variable* och REGION som *Column variable* och välj *Column percentages*. Finns det stöd för ett samband mellan variablerna?

2.2. Studera sambandet mellan HAPPY och REGION skilt för män och kvinnor och jämför resultatet med det i 2.1. Detta kan t ex göras med att skapa två nya datamatriser från SocialSurvey. Precis som i en tidigare övning, välj *Data-Active...-Subset...* och fyll det önskade villkoret i rutan *Subset expression* (**SEX == "Female"**), samt ange ett nytt namn för denna datamatrix i rutan *Name for new data set*, t ex **SocialSurveyFemale**. Välj sedan den ursprungliga datamatrixen SocialSurvey via *Data-Active...-Select active...* och upprepa proceduren ovan för att skapa en dylik datamatrix för män (använd villkoret **SEX == "Male"** och datamatrixnamnet **SocialSurveyMale**). Genom att välja den korrekta datamatrixen (*Data-Active...-Select active...*) kan analysen i 2.1. nu utföras skilt för kvinnor och män. Vad tror Du är orsaken till skillnaden mellan de två resultaten?

2.3. Studera sambandet mellan HAPPY och TAKEPART, återigen först för hela materialet och sedan skilt för män och kvinnor. Eftersom skilda datamatriser redan skapades i förra uppgiften för män och kvinnor, är det enkelt att utföra analysen på samma sätt som ovan (dvs. välj den aktuella datamatrixen och använd *Statistics-Contingency tables-Two-way table*). Slutsatser?

2.4. Omkoda HAPPY och HLTH5 till nya numeriska (ordinalskale)variabler i båda datamatriskerna **SocialSurveyFemale** och **SocialSurveyMale**. Välj först den aktuella datamatrixen (*Data-Active...-Select active...*) och använd sedan menyn *Data-Manage...-Recode...* där man väljer först HAPPY, namnger den nya variabeln t ex **HAPPYNUMERIC**, tar bort krysset från rutan *Make (each) new variable a factor(!)*, och anger sist följande omkodningsregler:

"Not Too Happy"=1
"Pretty Happy"=2
"Very Happy"=3

Var noggrann med gemener och versaler! Upprepa proceduren med variabeln HLTH5 (använd t ex **HLTH5NUMERIC** som det nya namnet) och ange omkodningsreglerna:

"No"=1
"Yes"=2

Ett ordinalt sambandsmått (Kendall's Tau) kan nu kalkyleras mellan **HAPPYNUMERIC** och **HLTH5NUMERIC** (*Statistics-Summaries-Correlation test*). Är sambandet signifikant på 5%-nivån både för män och kvinnor? Finns det någon skillnad mellan könen i detta hänseende? Vad säger koefficienten om sambandet mellan sinnessillstånd och droganvändning?

2.5. Studera effekten av en icke-konsekvent kodning gällande ordinalskalevariabler. Välj datamatrixen **SocialSurveyMale** som den aktiva. Omkoda HAPPY till en ytterligare ny numerisk (ordinalskale)variabel (*Data-Manage...-Recode...*) enligt följande. Välj först HAPPY, namnge den nya variabeln t ex **HAPPYNUMBAD**, ta bort krysset från rutan *Make (each) new variable a factor(!)*, och ange sist följande omkodningsregler:

"Not Too Happy"=2
"Pretty Happy"=3
"Very Happy"=1

Kalkylera nu Kendall's Tau mellan **HAPPYNUMBAD** och **HLTH5NUMERIC** (*Statistics-Summaries-Correlation test*). Är sambandet signifikant på 5%-nivån? Jämför koefficientens värde med det som erhöles enligt den ursprungliga omkodningen. Slutsatser?