

## Statistik 1 2009

### Övning 3, v39

Hämta följande datafil: <http://www.abo.fi/fak/mnf/mate/jc/statistik1/grasparvar.sav>  
Importerera datafilen till R Commander (*Data-Import...-From SPSS...*). Filen innehåller ett berömt dataset insamlat av Herman Bumpus år 1898. Materialet omfattar diverse morfologiska beskrivningar för 136 gräsparvar som fångades efter en svår storm. Detta material användes flitigt under den multivariata statistikens begynnelsestid för att illustrera nya metoder. Här skall vi dock analysera materialet med hjälp av mer elementär beskrivande statistik.

Variablerna med nominalskala är kodade enligt följande: AGE (1 = adult, 2 = young; obs. endast för hanar), SEX (= kön, 1 = male, 2 = female), SURVIVAL (= överlevnad, 1 = dead, 2 = survived).

1. Omkoda variablerna AGE, SEX, SURVIVAL genom att ersätta sifferkoderna med de verkliga beteckningarna på utfallen. Precis som i Övning 1 görs detta i menyn *Data-Manage...-Recode...*. Välj den aktuella variabeln (t ex SURVIVAL), ange namnet **SURVIVAL** i rutan *New variable name...* och ange sedan nya värden i rutan på separata rader enligt följande: **1 = "dead"** osv.
2. Titta på fördelningen för var och en av dessa variabler (*Statistics-Summaries-Frequency distributions*). Använd *Statistics-Summaries-Count missing observations* för att se vilka uppgifter saknas och för hur många individer.
3. Hur stor andel av hanarna respektive honorna dog? Skapa en korstabell via *Statistics-Contingency tables-Two-way table*, välj SEX som en rad variable (row variable) och SURVIVAL som en . Välj även *Row percentages* för att erhålla de rätta procentandelarna.
4. Beräkna medeltalet, medianen, standardavvikelsen, minimum och maximum för variabeln AE i de fyra olika grupper som bildas av att man korstabulerar variablerna SEX och SURVIVAL. Varje statistika kan erhållas från *Statistics-Summaries-Table of statistics*. Välj **SEX** och **SURVIVAL** som *Factors* och **AE** som *Response variable*. Välj sist den statistika som önskas (proceduren kan upprepas för att få värden på olika statistikor). Minimum och maximum erhålls genom att man väljer *Other* och skriver **min** respektive **max** i rutan till höger. Ett enkelt sätt att visualisera eventuella skillnader mellan de fyra grupperna m.a.p. variabeln AE är att välja *Graphs-Plot of means*, sedan lägger man igen **SEX** och **SURVIVAL** som *Factors* och **AE** som *Response variable*. Välj även *Standard deviations*. Den resulterande bilden visar medelvärden och std.avvikelserna för grupperna.
5. Dela in fåglarna i två grupper enligt vinglängd (AE), så att de fåglar vars vinglängd är kortare än medelvärdet (245.03) bildar den ena gruppen och resterande fåglar den andra gruppen. Detta kan göras på två olika sätt: 1) man väljer den aktuella delmängden av fall till vidare behandling och beräknar sedan proportionerna, eller 2) man bildar en ny variabel som indikerar grupperingen och använder sedan den som en faktor i analysen. Den nya indikatorvariabeln kan skapas på följande sätt. Gå till menyn *Data-Manage...-Recode...* och välj variabeln **AE**. Döp den nya variabeln t ex till **AENEW** och **ta bort krysset** från rutan *Make (each) new variable a factor*. Ange sedan omkodningsreglerna i rutan *Enter recode directives* enligt följande två rader:

**lo:245.03=1**

**245.03:hi=2**

Detta innebär att samtliga värden från det minsta (**lo**) till medelvärdet kodas till 1 och samtliga värden från medelvärdet till det största (**hi**) kodas till 2. Klicka sist på **OK** för att utföra omkodningen. För att transformera AENEW till en faktor, gå till menyn Data-Manage...-Convert... och välj AENEW och klicka på OK. Acceptera sedan överskrivning och ange namnen på de två grupperna i rutan som öppnas (t ex skriv **short** för 1 och **long** för 2). När konverteringen är klar, kan man korstabulera AENEW och SURVIVAL för att undersöka om andelen fåglar som överlevde varierar m.a.p. de två grupperna (korta kontra långa vingar). Är sambandet mellan AENEW och SURVIVAL signifikant på 5%-nivån.