

Statistik 1 2007

Övning 3, 25.9.07

Hämta följande datafil: <http://www.abo.fi/fak/mnf/mate/jc/statistik1/grasparvar.sav>
Filen innehåller ett berömt dataset insamlat av Herman Bumpus år 1898. Materialet omfattar diverse morfologiska beskrivningar för 136 gråsparvar 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. Komplettera datamaterialet med 'labels' (etiketter) för variablerna Age, Sex, Survival. Titta på fördelningen för var och en av dessa variabler (Descriptive Statistics, Frequencies). Observera att en del uppgifter saknas. Hur påverkar detta fördelningen?
2. Hur stor andel av hanarna respektive honorna dog? (Descriptive Statistics, Crosstabs, Cells...).
3. Beräkna medeltalet, medianen, minimum och maximum för variablerna TL, AE och W i de fyra olika grupper som bildas av att man korsstabulerar variablerna kön och överlevnad. Använd Compare Means, Means och lägg in TL, AE och W på listan av beroende variabler (Dependent list), samt lägg kön och överlevnad på förklarande variabler som lager 1 respektive 2 (Independent List, Layer 1 & 2).
4. Dela in fåglarna i två grupper enligt vinglängd (AE), så att de fåglar vars vinglängd är kortare än medelvärdet bildar den ena gruppen och resterande fåglar den andra gruppen. Beräkna sedan fördelningen av värden på överlevnad separat för dessa två grupper. Detta kan göras på två olika sätt, antingen väljer man den aktuella delmängden av fall till vidare behandling (Data, Select cases) och beräknar sedan proportionerna, eller så bildar man en ny variabel som indikerar grupperingen (Transform, Recode, Into different variables) och använder sedan den som en faktor i analysen. Ange ett lämpligt namn för den nya variabeln.
5. Använd ett cirkeldiagram för att illustrera fördelningen för överlevnad med avseende på de två grupperna från uppgiften ovan (Graphs, Interactive, Pie...).