This document is an historical remnant. It belongs to the collection Skeptron Web Archive (included in
Donald Broady's archive) that mirrors parts of the public Skeptron web site as it appeared on 31 December 2019, containing material from the research group Sociology of Education and Culture (SEC) and the research programme Digital Literature (DL). The contents and file names are unchanged while character and layout encoding of older pages has been updated for technical reasons. Most links are dead. A number of documents of negligible historical interest as well as the collaborators’ personal pages are omitted. The site's internet address was since Summer 1993 www.nada.kth.se/~broady/ and since 2006 www.skeptron.uu.se/broady/sec/.

### 1. Deskriptiv statistik

Ta fram frekvenstabeller för variablerna i datasetet. Fundera över vad du vet om data, och vad du inte vet, men skulle vilja veta.

Vilka variabler lämpar sig för korrespondensanalys? Vilka variabler skulle gå att använda efter smärre förändringar av kodningen?

Gör korstabeller för att lära dig mer om data.

Kan du med hjälp av korstabeller verifiera samband som du gissade skulle finnas i data? Ser du något oväntat?

Om det tillför något, använd även histogram, stapeldiagram, linjediagram, scatterplottar och andra enkla deskriptiva tekniker för att lära dig mer om datasetet.

### 2a. Urval av variabler

Ta fram en preliminär uppsättning variabler som du vill studera. (Mellan 4 och 40 variabler.)

Hur kan din frågeställning användas för att hitta en lämplig avgränsning? Vilka variabler befinner sig på gränsen mellan att vara med och inte? Vilka samband är du mest intresserad av?

### 2b. Strukturerande faktorer

Dela (eventuellt) in variablerna i ”aktiva variabler” och ”strukturerande faktorer”.

Går det att dra en skiljelinje mellan variabler vars inbördes relationer du vill studera, och variabler som skall ”sprida ljus” över dessa relationer? Ett exempel är om du har frågor som rör individers ställningstaganden i ett antal frågor, samt uppgifter om individernas kön och ålder. Då är en möjlig indelning att låta ställningstagandena vara aktiva variabler, och övriga fungera som strukturerande faktorer.

En aspekt av variabeluppsättningen är dess ”homogenitet”. Med detta menas i vilken utsträckning de aktiva variablerna handlar om ”samma sak”, eller är av ”samma typ”. En uppfattning är att det är bra om uppsättningen aktiva variabler är någorlunda homogen. Detta gör det lättare att tolka korrespondensanalysens resultat.

Vilka variabler är lika varandra? Finns det udda variabler, för vilka det inte är klart om de bör vara aktiva variabler eller strukturerande faktorer?

### 3. Rubriker

Dela in de aktiva variablerna i innehållsmässigt sammanhängande grupper. (Mellan 1 och 7 grupper.)

### 4. Kodning av data

Koda de aktiva variabler enligt följande riktlinjer:

1. Ingen modalitet hos någon aktiv variabel skall innehålla mindre än 5 % av totalpopulationen. Det bästa sättet att komma dit hän är att slå samman små modaliteter så att de blir tillräckligt stora. Om detta inte är möjligt är en *nödlösning* att göra modaliteten passiv (=illustrativ i SPAD).
2. Ingen modalitet hos någon aktiv variabler skall innehålla mindre än 5 individer.

Ett strängare krav är att inget element i någon korstabulering mellan de aktiva variablerna skall innehålla mindre än 5 individer. Uppfyller dina (omkodade) data detta krav?

Om flera variabler är snarlika kan det vara lämpligt att skapa en eller flera helt nya variabel vars värden baseras på kombinationer av ursprungsvariablerna. Ett exempel är klassificering av socialt ursprung, som kan baseras på en rad uppgifter om individens bakgrund, som till exempel utbildning, arbete och inkomst.

Kontrollera (med hjälp av frekvenstabeller och korstabeller) att kodningarna du gjort blivit som du tänkt dig, och att de uppfyller de två kraven ovan.

### 5. Korrespondensanalys

Genomför den korrespondensanalys du planerat

### 6. Tolkning

Undersök resultaten med utgångpunkt från:

**Balans.** Är det ett fåtal modaliteter, variabler, rubriker som ger mycket större bidrag till korrespondensanalysens första axlar än övriga?

**Förklarande variabler.** Vilka rubriker, variabler respektive modaliteter ger störst bidrag till var och en av analysens (2, 3 eller 4) första axlar? Utgå från modifierade egenvärden.

**Rummets tekniska relevans.** Hur stor del av variationen i materialet förklarar de axlar du studerar? Om andelen är liten, kan du på något annat sätt motivera analysens relevans?

**Rummets innehållsmässiga relevans.** Är resultatet av analysen relevant i förhållande till din frågeställning? Det är inte ovanligt att man missbedömer sambanden i data och får ganska triviala resultat. Om detta inträffat, så får man överväga att göra ett annat urval av variabler, och eventuellt även ändra på kodningen av variablerna.

**Representation av enskilda punkter.** Hur väl är enskilda, intressanta, punkter representerade i de axlar du studerar?

**Individerna fördelning.** Hur ser individmolnet ut? Finns tendenser till hästskoeffekt?

**Strukturerande faktorer.** Hur fördelar sig de strukturerande faktorernas modaliteter? (Ligger de nära origo?) Hur ser de delmolnen för de strukturerande faktorernas modaliteter ut?

**Ordnade modaliteter.** Ibland kan en modaliteterna hos en variabel ordnas på ett naturligt sätt (till exempel ålderskategorier). Hur visar sig denna ordning i de axlar som korrespondensanalysen resulterat i?

**Interaktionseffekter.** Studera samband mellan variabler genom att låta kombinationer av par av variabler ingår som strukturerande faktorer i analysen.

Lägg mest krut på att tolka de två första axlarna, men titta även på den tredje och den fjärde axeln, och lyft in dem i resonemangen om det tycks motiverat.

### 7. Presentation

Välj ut några grafer *och tabeller* som tillsammans underbygger dina resonemang kring data. Utnyttja de möjligheter till visualisering som programmet du använder erbjuder, till exempel genom att:

Låta symbolernas **form och färg** säga något om vad de representerar.

Låta **symbolernas storlek** säga något om antalet individer de representerar, eller, om det är motiverat, hur mycket de bidrar till axlarna eller hur väl de är representerade.

Variera **textens typsnitt, storlek och stil** för att skilja mellan olika variabler, samt mellan aktiva variabler och strukturerande faktorer.

Visa ordningsrelationer mellan modaliteter, med hjälp av **linjer eller pilar**.

Visa **interaktionseffekter** med hjälp av kombinationer av form, färg, linjer och pilar.

Göra ett förtydligande **urval av modaliteter** med utgångspunkt från deras bidragsvärden.

Visa **individmolnet**, strukturerat av de strukturerande faktorerna.

Lägga in **grafiska spridningsmått**, till exempel ellipser eller polygoner, för individmolnets olika delar.

Lägga in **förklarande texter och frihandsgrafik** i grafen. (Ett förlag är att ange hur stor del av variationen i material som de respektive axlarna representerar.)

Det är ofta lämpligt att låta den horisontella och den vertikala axeln ha samma skala.