

SPSS introduktion

Enes Al Weswasi, Olof Bäckman, Anders Nilsson och Fredrik Sivertsson

2022-07-19

Contents

Om SPSS Introduktion	7
Komma igång med SPSS	11
1 Installera och starta SPSS	11
1.1 Ladda ner SPSS	11
1.2 Installera SPSS	11
1.3 Starta SPSS	12
2 Viktiga fönster i SPSS	13
2.1 Datamatrixfönstret (Data view)	13
2.2 Variabelfönstret (Variable view)	13
2.3 Output-fönstret (Output)	14
2.4 Syntax (Syntax editor)	14
Förbereda datamaterialet	19
3 Databearbetning	19
3.1 Select Cases: analysera enbart vissa observationssenheter	19
3.2 Recode: att koda om befintliga variabler	20
3.3 Compute: att skapa nya variabler utifrån befintliga variabler . .	21
4 Variabelkonstruktion med hjälp av syntax	23

Analysera data I	27
5 Beskrivande statistik	27
5.1 Frekvenstabeller, central- och spridningsmått	27
5.2 Kort om grafiska tekniker	28
Analysera data II	31
6 Bivariat analys: Att studera samvariationen mellan två variabler	31
6.1 Sambandsmått	32
Analysera data III: Hypotesprövning	37
7 Att välja rätt test vid hypotesprövning	37
8 Chi2-test	39
9 T-test	41
10 Enkel regression	43
11 Multipel regression	45
Visualisera data	49
12 Bearbeta tabeller och figurer	49
12.1 Tabeller	49
12.2 Figurer	49
Övrigt	53
13 Presentation av dataset	53
13.1 NTU 2017-2021	53

<i>CONTENTS</i>	5
13.2 NTU 2013-2015	53
13.3 Skolundersökning 2005	54
13.4 Pathways to desistance	54
14 Dataövningar	57
14.1 Dataövning 0	57
14.2 Dataövning 1	57
14.3 Dataövning 2	57
14.4 Dataövning 3	57
14.5 Dataövning 4	57

Om SPSS Introduktion

Denna SPSS-introduktion är avsedd för dig som läser kursen Metod 2 på Kriminologiska institutionen vid Stockholms universitet. Introduktionen inleds med en SPSS-guide som rymmer en genomgång av SPSS olika fönster, hur man lägger in data och hur man öppnar en redan befintlig datafil. Därefter följer grundläggande tillvägagångssätt för bearbetning och beskrivning av data (beskrivande statistik). Guiden avslutas med en genomgång av några grundläggande analysmetoder för bivariata och multivariata samband.

De tre dataset som vi kommer att jobba med finns på kurssajten På Athena. Det första datasetet är Brottsförebyggande rådets skolundersökning 2005. Det andra datasetet är en fil baseras på fem årgångar (2017-2021) av Brottsförebyggande rådets Nationella trygghetsundersökning. Det tredje och sista datasetet är från en amerikansk självdeklarationsstudie vid namn Pathways to desistance innehållande uppgifter från 1 300 brottsbelastade ungdomar. Observera att dessa filer enbart är till för att användas på denna kurs. Efter kursens avslutande ska alla eventuella lokala kopior raderas. En kortare beskrivning av dessa datamaterial finns med som bilaga i slutet av denna guide. För mer utförliga beskrivningar hänvisas till de Brå-rapporter som finns på Athena samt Pathways to desistance-artikeln.

SPSS-guiden ger en introduktion till att arbeta med SPSS Version 28 vilket är den senaste versionen och som vi rekommenderar att ni använder. Några mer ingående redogörelser av statistiska metoder ges inte, för sådana får ni gå till kurslitteraturen. Ni får här inte heller någon utförlig beskrivning av datamaterialen som används på kursen. Sådan information hittar ni via Athena.

Dataövningarna under kursen gång ger ytterligare träning i att använda SPSS. Uppgifterna till dataövningarna hittar ni i anslutning till respektive dataövningsplanering i Athena. Här finns också korta introduktionsfilmer till uppgifterna.

Komma igång med SPSS

Chapter 1

Installera och starta SPSS

1.1 Ladda ner SPSS

Nedanstående instruktioner är hämtade från Stockholms universitets sida om SPSS. Om någon länk ej fungerar så prova att besök SU:s SPSS sida. Där hittar ni även information om hur man avinstallerar SPSS eller förnyar licensnyckeln. Installationsfiler för SPSS finns tillgängliga här.

1. Välj vilket operativsystem du vill installera SPSS på. SPSS finns till Windows, Mac, och Linux. Observera att SPSS inte finns till Chromebook.
2. Olika versioner av SPSS finns tillgängliga men vi rekommenderar version 28 eftersom "SPSS Introduktion" är anpassad efter denna version. Välj vilken version du vill installera (samt om du kör x32 eller x64 på Windows)
> Klicka på .exe- .pkg-, respektive .dmg-filen > Välj Spara fil/Save File.
3. För SPSS på Windows, finns ZIPade installationsfiler. Spara ner den ZIPade filen på skrivbordet > högerklicka på mappen och välj Extract All
> välj Extract > dubbelklicka på setup.exe > välj Run.

1.2 Installera SPSS

4. Gå igenom installations-wizarden.
 - Om du får ett val mellan Authorized user license och concurrent user license > välj Authorized user license.
 - Om du ska ange Organization > skriv Stockholms universitet.

5. När du har gått igenom wizarden > välj Install (detta kan ta några minuter).
6. När installationen är genomförd > välj License Product. Du kommer att se en lista på de applikationer som är temporärt licensierade i 14 dagar. Klicka på Next > se till att alternativet Authorized User License är valt > välj Next.
7. För att få åtkomst till aktuella behörighetskoder, klickar du här. Klistra in koden för den version du har valt att licensiera > välj Next. Du ser nu en progressruta och denna avslutas med “**End of Transaction**” Successfully processed all codes > välj Next. Du ser nu en ruta som bekräftar de licensierade modulerna. Klicka på Finish för att avsluta licensieringen. Verifiera
8. Verifiera installationen genom att kontrollera att det går bra att starta programmet.

1.3 Starta SPSS

Öppna SPSS genom att dubbelklicka på ikonen på skrivbordet eller gå via startmenyn När du öppnar SPSS kommer du att, som på bilden nedan, se en data-matris – ett rutnät bestående av rader och kolumner.

Raderna i denna matris motsvarar observationsenheter (t.ex. individer) medan kolumner motsvarar variabler (t.ex. frågor i en surveyundersökning).

Om du vill öppna ett befintligt dataset väljer du:

1. File > Open > Data
2. Hitta datafilen i katalogen och välj ”Open”
3. I exemplet nedan öppnas filen Skol05.sav från datorns skrivbord

Chapter 2

Viktiga fönster i SPSS

SPSS innehåller flera olika fönster med skilda funktioner.

2.1 Datamatrixfönstret (Data view)

Detta fönster redovisar variablerna (t.ex. en fråga i en enkäteundersökning) kolumnvis och observationsenheterna (exempelvis de skolelever som svarat på en enkät) radvis. En enskild cell redovisar alltså en specifik observationsenhets värde på en specifik variabel.

2.2 Variabelfönstret (Variable view)

I detta fönster visas de olika variablerna och deras egenskaper. Variablerna redovisas radvis och egenskaperna kolumnvis. Byte av fönster görs enkelt genom att klicka nere i vänstra hörnet, där man kan välja mellan Data View och Variable View.

2.2.1 Mer om variabelfönstret (Variable view)

När du öppnar ett befintligt datamaterial får du upp detta i datamatrixfönstret (Data View), antingen med variabelvärden i text (bilden till vänster) eller som siffror (till höger). Hur man väljer att se värdena kan man ange via menyns "View"):

I exemplet har datafilen Skol05.sav öppnats och varje rad motsvarar här en skolelev och dennes svar på de olika frågorna i enkäten. Vid kodningen har varje svar kodats in med en siffra och dessa värden har getts etiketter/labels.

Om du byter till Variable View (detta görs enkelt längst ner till vänster) ser du hur materialet kodats in med etiketter (labels) som anger både en precisering av de olika variablerna och de olika variabelvärdena:

I variabelfönstret redovisas alltså variablerna radvis och dess egenskaper kolumnvis. I detta fönster finns ett antal kolumner som är viktiga att känna till. Under den första kolumnen "Name" listas variabelnamnet på varje variabel i ditt dataset. Variabelnamnet hålls vanligtvis kort och det blir därför ofta svårt att utläsa vad variabeln mäter. Information om vad variablerna mäter och hur de är konstruerade finns ibland i en separat kodbok eller variabellista men du kan även förtydliga dina variabler i själva datafilen. Under kolumnen "Label" anges variabeletiketten, som är en kort beskrivning av variabeln. Under kolumnen "Values" anges vad variabelns olika värden svarar mot.

Variabeln "kon" har till exempel värde 0 eller 1, där värde 0 betyder "flicka" och värde 1 betyder "pojke". För att undersöka detta markerar du cellen för den aktuella variabeln under kolumnen "Values" och klickar därefter på den lilla rutan till höger. Nu kommer fönstret "Value Labels" upp. Om variabeln saknar etiketterade variabelvärden kan du ange dessa genom att föra in värdet i rutan "Value" och vad värdet står för i rutan "Label". Klicka därefter "Add". När du etiketterat samtliga variabelvärden klickar du "OK".

Tänk på att variabelns värden endast behöver specificeras för variabler som befinner sig på nominaleller ordinal skalnivå. Om du till exempel har en kontinuerlig variabel som mäter ålder i antal år (kvotskala) är variabelvärdena i sig informativa och du behöver inte förtydliga vad dessa står för.

2.3 Output-fönstret (Output)

I detta fönster visas resultat. Det kan röra sig om frekvenstabeller, korstabeller eller olika typer av diagram. Till vänster finns en översiktsmeny där du enkelt kan bläddra mellan dina resultat.

2.4 Syntax (Syntax editor)

Vissa bearbetningar av data där man vill kombinera olika variabler kan vara enklare att göra i syntax. I syntax kan du skriva kommandon, dvs. det du vill att SPSS ska göra åt dig – en frekvenstabell, korstabell, omkodning av variabler mm. Alla sådana moment har sina egna kommandon.

Du kan öppna syntax via File > New > Syntax:

Kommandon kan antingen skrivas direkt i syntax eller så kan du gå via menysystemets funktioner och klick på Paste

Klickar du på denna knapp så kommer syntax kommandot för denna omkodning att öppnas i syntaxfönstret:

“Syntaxspråket” tar ett tag att lära sig, och är heller inte nödvändigt för er att kunna, men att använda sig av syntax (direkt eller via ”paste”) kan många gånger underlätta arbetet i SPSS och om du sparar syntax får du en loggbok över det du har gjort, på så sätt är det enkelt att gå tillbaka för att se hur du gått tillväga vid analyser, omkodningar etc. (Om du sparar de kommandon och körningar du har gjort och har användning för i syntax, skriv gärna rubriker/kommentarer som anger vad och varför du gjort olika moment. Obs: För att inte programmet ska missta rubriker/kommentarer för kommandon måste du skriva en asterisk (*) före dessa och avsluta med en punkt.) Se även Djurfeldt m.fl. 2010/2018 Appendix 2.

Förbereda datamaterialet

Chapter 3

Databearbetning

En stor del av arbetet består av att bearbeta sitt datamaterial inför analys. Det finns ett antal funktioner i SPSS med vilka man enkelt bearbetar materialet efter kodning. Kanske är du bara intresserad av att studera vissa observationenheter, t.ex. bara kvinnor? Eller så kanske du vill koda om variabler (slå ihop svarskategorier eller klassindela) eller skapa nya variabler som bygger på information i två eller fler befintliga variabler (t.ex. skapa ett index). Några funktioner för detta presenteras nedan.

3.1 Select Cases: analysera enbart vissa observationssenheter

Funktionen används när du endast vill undersöka vissa observationssenheter, exempelvis endast de flickor som ingår i skolundersökningen. Hur variabeln kön har kodats framgår av variabelförteckningen (du hittar denna som pdf fil i Athena), alternativt kan du markera variabeln i SPSS genom att välja Utilities > Variables, leta reda på Kön i listan och titta i rutan "Variable information".

Följande kan då utläsas: Du finner att flickor har värdet 0 på variabeln "kon". För att enbart välja ut de observationssenheter som är flickor (alltså har värdet "0" på variabeln "kon") gör du följande:

```
Data > Select Cases > If condition is satisfied > If
```

Lyft in den aktuella variabeln i fönstret och ange villkor (variabelvärde för att tas med), därefter:

```
Continue > OK
```

Tänk på att du nu har angett att kommande analyser endast ska göras för de med värde "0" på variabeln "kon". Om du vill återgå till att analysera samtliga

observationsenheter (både pojkar och flickor), klickar du på alternativet "All cases". Alltså:

```
Data > Select Cases > All cases
```

Notera att det är variabeletiketten ("Kön") och inte variabelnamnet ("kon") som står i rullistan till vänster i bilden ovan. Standardinställningen är att visa variabeletiketten om en sådan finns, men ibland är det smidigare att istället visa variabelnamnet. För att göra detta högerklickar du på variabellistan till vänster och markerar "Display Variable Names".

3.2 Recode: att koda om befintliga variabler

Funktionen Recode kan användas till att koda om variabler (koda om befintlig variabel: "Recode into Same Variables", eller till en ny variabel: "Recode into Different Variables"). "Recode" används ofta då variabeln har många kategorier/variabelvärden som forskaren vill sammanfatta till färre kategorier/variabelvärden. Det används om man vill kategorisera en kontinuerlig variabel (t.ex. klassindela) eller om man vill slå ihop kategorier i en kategorisk variabel. Variabeln offgrov i skolundersökning (Skol05.sav) har tre värden: 0, 1 respektive 2, se frekvenstabellen nedan:

Anta att du vill skapa en variabel som enbart skiljer på utsatt respektive ej utsatt för grövre våld. Du vill alltså skapa en ny variabel där de som svarat ja hamnar i samma kategori oavsett om varit utsatta en gång (variabelvärde 1) eller två gånger eller fler (variabelvärde 2). Du vill med andra ord att den nya variabeln enbart ska ha två kategorier; Nej och Ja:

```
Transform > Recode into Different Variables
```

För över den variabel som du vill koda om till den stora rutan med rubriken "Numeric Variable->Output Variable" och ange vad den nya variabeln ska ha för variabelnamn ("Name") och variabeletikett ("Label"). Klicka därefter på "Change". I detta fall har den nya variabeln fått namnet offgrov2.

Gå sedan vidare och öppna "Old and new values". Ange variabelvärdet i den ursprungliga variabeln ("Old Value") och vilket värde detta ska bli i den nya variabeln ("New Value"). Koda om ett värde i taget och klicka på "Add" efter att du har angett det gamla och det nya variabelvärdet. I rutan "Old > New:" får du en översikt på hur du har valt att göra omkodningen. Bilden nedan visar att både värde 1 och värde 2 i den gamla variabeln ska ha värde 1 i den nya variabeln. När du har kontrollerat att omkodningen ser riktig ut väljer du "Continue" och sedan "OK". Kontrollera att du har fått den nya variabeln tillagd längst ner i variabellistan (Variable view).

I exemplet ovan är det två variabelvärden i den gamla variabeln som slagits samman. Ibland har dock de variabler du vill klassindela betydligt fler variabelvärden, och det är då osmidigt att ange varje enskilt variabelvärde under

3.3. COMPUTE: ATT SKAPA NYA VARIABLER UTIFRÅN BEFINTLIGA VARIABLER²¹

"Value:". I dessa fall markerar du istället "Range:" och anger i den övre rutan den lägsta klassgränsen och i den nedre rutan den högsta klassgränsen. Tänk dig till exempel att du ska klassindela variabeln ålder i Nationella trygghetsundersökningen. Istället för att under "Value:" ange vad varje enskild ålder ska tillhöra för klass kan du använda "Range". Ett tips: För att förtydliga vad dina variabelvärden i den nya variabeln betyder letar du upp denna i variabelfönstret och anger detta under "Values".

3.3 Compute: att skapa nya variabler utifrån befintliga variabler

Med hjälp av funktionen Compute kan du skapa nya variabler. Antag att du vill slå samman – summera – två variabler till en variabel. Att summera variabler kan användas för att skapa summaindex. Index innebär alltså att man summerar värdena på flera variabler till en totalsumma. Istället för flera variabler som mäter samma underliggande fenomen (i exemplet: brott) får vi en sammanfattande variabel. Utifrån skolundersökningen kan index exempelvis skapas utifrån frågor om brottslighet, betyg, attityder osv (se vidare Djurfeldt m.fl. 2010/2018, Appendix 3).

I detta exempel består datamaterialet av 6 individer och två variabler – antal stöldbrott ("Stöldbrott") och antal våldsbrott ("Våldsbrott").

Anta att du vill skapa en variabel som anger det totala antalet brott:

Transform > Compute Variable

Namnge den nya variabeln i Target variable. I detta fall döps den nya variabeln till "Totbrott". I rutan "Numeric expression" anges funktionen för att skapa den nya variabeln, i detta fall anges vilka variabler som skall summeras (+):

Klicka därefter "Continue" och "Ok". Den nya variabeln placeras längst till höger i datamatrixen ("Data view") och längst ner i variabelfönstret ("Variable view"). Summering kan också göras genom att ange funktionen "sum(Våldsbrott, Stöldbrott)", skillnaden mellan de två sätten är att missing values (internt bortfall) behandlas olika: Med funktionen "sum" får inte de observationsenheter som har missing på någon av de variabler som summeras missing på den nya variabel du skapar.

Du kan även välja att skapa en variabel som anger medelvärdet på de ingående variablerna i indexet. Detta gör du enklast med funktionen "mean()". En fördel med detta val är att du även kan ange hur många av de ingående variablerna som måste ha valida värden för att individen ska få ett värde på indexet. Anta att du vill skapa ett index som mäter vänners antisociala attityder och du vill inkludera följande fem variabler (Skol05): Har någon av dina kompisar 1) tagit något utan att betala i en affär? ("snattkom"), 2) förstört någonting? ("kompvand"), 3) brutit sig in någonstans? ("kompinbro"), 4) slagit ner någon? ("kommissh"), 5)

åkt fast för polisen? ("komaktfa"). Den svarande kan välja att svara ja eller nej där 1 är ja och 0 är nej. Ett medelvärdesindex kommer därför att variera mellan 0 och 1 där 1 betyder att individen har svarat ja på de ingående enkätfrågorna och 0 betyder att individen har svarat nej på de ingående enkätfrågorna. Följande funktion skulle först summera alla individens värden på de fem ingående variablerna och därefter dividera summan med fem, d.v.s. antalet variabler:

$(\text{snattkom} + \text{kompvand} + \text{kompinbro} + \text{kommissh} + \text{komaktfa})/5$

Men ovanstående funktion kräver att individen har valida värden på samtliga fem ingående variabler, om en enda variabel saknar ett valitt värde så kommer SPSS inte att beräkna ett indexvärde för den individen. Detta gör inte så mycket om de ingående variablerna har ett litet internt bortfall för ungefär samma individer. Problemet uppstår när det finns ingående variabler med stort internt bortfall och/eller om det interna bortfallet är fördelat på en stor andel av individerna. Detta resulterar i att ditt index får ett mycket stort internt bortfall vilket kan skapa osäkerhet och precisionsförlust i kommande analyser. En smart lösning på problemet är att bestämma ett visst antal variabler som individen måste ha svarat på för att ingå i indexet och beräkna indexvärdet endast på dessa variabler. Säg att du väljer att endast tre av de fem ingående enkätfrågorna måste ha besvarats för att ett indexvärde ska beräknas. Detta gör du genom att placera en punkt och siffran 3 efter själva mean-funktionen enligt bilden nedan:

Om du anger mean-funktionen utan att specificera hur många variabler som ska ha valida värden så är utgångspunkten att åtminstone två variabler har valida värden. Men detta kan vara lite väl skakigt (försämra begreppsvaliditeten, se Bryman) på ett index som består av många variabler. Tänk därför på hur många variabler som det är rimligt att en individ ska ha valida värden på för att ingå i indexet. Detta är ett val som forskaren ska göra och inte SPSS.

Bra att känna till är att oavsett om du väljer en summa- eller medelvärdesfunktion så kommer fördelningen i indexet (när du t.ex. vill beskriva variabeln i en frekvenstabell) att vara identisk förutsatt att beräkningen tar hänsyn till lika många variabler. Skillnaden kommer endast att vara skalan på vilket indexet mäts.

Chapter 4

Variabelkonstruktion med hjälp av syntax

Ett exempel på när bearbetningar är bra att göra i SPSS är när man vill kombinera olika variabler och svaralternativ för att skapa en ny variabel.

En funktion för detta är Compute (se exemplet ovan), en annan är funktionen If. Med If kan man precisera villkor för variabelvärden utifrån en eller flera variabler.

I skolunderökningen (Skol05.sav) finns tre variabler som anger födelseland: För eleven själv ("fodland1"), för dennes mor ("mfodland") respektive för dennes far ("pfodland"). Utifrån dessa tre variabler kan man skapa en variabel som anger utländsk bakgrund – "utl_bkgr" - som får värdet 0 om eleven själv och dennes föräldrar är födda i Sverige och värdet 1 om eleven själv och någon av dennes föräldrar är födda i annat land.

Öppna syntax: File / New / Syntax

Skriv ett kommando som anger hur variabeln ska skapas (granska gärna kommandot nedan och fundera över dess innebörd, täcker villkoren in alla? Finns andra möjliga definitioner?):

Markera programsatsen och klicka på den gröna pilen. I datafilen finns nu den nya variabeln "utl_bkgr". Nästa steg är att kontrollera så att den konstruerade variabeln blivit korrekt: Ta fram en frekvens (se vidare nedan), ser fördelninge rimlig ut om du jämför med ursprungsvariablerna? Kontrollera också gärna ett par av dina observationsenheter – utifrån de värden de har på de tre ursprungliga variablerna, har det blivit korrekt? Därefter går du vidare med att sätta labels på variabeln.

Obs: När du konstruerar nya variabler är det självfallet viktigt att dessa är genomtänkta och välmotiverade.

Analysera data I

Chapter 5

Beskrivande statistik

5.1 Frekvenstabeller, central- och spridningsmått

Det är vanligt att man inleder en studie med att studera hur observationsenheterna fördelar sig med avseende på en enskild variabel. Viktiga verktyg i detta ändamål är frekvenstabeller, centralmått och spridningsmått. Anta att du studerar den Nationella trygghetsundersökningen (NTU 2013-15 M2.sav) och vill ha information om oro över brottsligheten i samhället:

Analyze > Descriptive statistics > Frequencies

Börja med att söka upp den variabel du är intresserad av i rullistan till vänster (kom ihåg att du kan välja att visa variabelnamn eller variabeletiketter genom att högerklicka på listan). Därefter markerar du variabeln och flyttar över den till den högra rutan genom att använda pilen mellan rutorna

alternativt dubbelklicka på variabeln. Om du nu väljer alternativet "OK" kommer SPSS att producera en frekvenstabell på variabeln. Detta är standardvalet (som du kan se är valet "Display frequency tables" markerat), men ofta vill man sammanfatta sin variabel lite mer utförligt. Längst till höger finns möjligheter att ytterligare specificera vad du vill få fram för statistik. Till exempel kan du genom att klicka på "Statistics" välja central- och spridningsmått

Skalnivån (mätnivån) för frågan om oro är på ordinal nivå varför vi i detta fall nöjer oss med en frekvenstabell:

Av frekvenstabellen kan vi utläsa mer specifikt hur många individer och hur stor andel som uppger sig vara oroliga för brottsligheten i samhället. Vi kan till exempel se att endast 24 procent svarar att de inte alls är oroliga.

5.2 Kort om grafiska tekniker

I syfte att beskriva våra resultat i grafisk form kan man även välja "Graphs" i huvudmenyn. Genom alternativet "Chart Builder" kan du välja på en mängd olika diagramtyper som på bästa sätt beskriver din/a variabel/ler. Alltså:

Graphs > Chart Builder

Vanliga diagramtyper för att beskriva enskilda variabler är stapeldiagram ("Bar chart"), cirkeldiagram ("Pie chart") och histogram. Med detta verktyg kan du på grafisk väg även studera eventuella samband mellan två variabler. Vi återkommer senare till att åskådliggöra samvariationen mellan två kontinuerliga variabler genom att ta fram ett så kallat spridningsdiagram ("Scatter plot").

Analysera data II

Chapter 6

Bivariat analys: Att studera samvariationen mellan två variabler

Under denna kurs kommer du vilja undersöka huruvida det finns ett samband mellan två variabler. Under förutsättning att variablernas skalnivåer är nominal- eller ordinalskala (ej intervall- eller kvotskala) görs detta vanligtvis genom att studera de båda variablerna i en korstabell ("Crosstab"). Gör följande:

Analyze > Descriptive statistics > Crosstabs

I detta fönster har du likt tidigare en rullista till vänster som innehåller samtliga variabler i datamaterialet. Innan du fortsätter är det viktigt att du, med hänvisning till din frågeställning, har gjort klart vilken variabel som är tänkt att påverka den andra. Beroende variabel placeras i radled ("Row(s)") och oberoende variabel placeras i kolumnled ("Column(s)"). Vi kan t.ex. vara intresserade av huruvida oro för brottsligheten i samhället skiljer sig åt efter kön. Vi gör då en korstabell med variablerna S4 (Oro brottslighet) Kön. Identifiera variablerna i rullistan till vänster och för sedan över dessa till "Row(s)" respektive "Column(s)" genom att använda pilarna. I detta fall gjordes alltså antagandet att Kön är oberoende. Som framgår finns ytterligare funktioner/alternativ. För kursen relevanta rutor är här "Statistics", "Cells" och "Format". Under "Statistics" kan man välja mellan ett flertal olika sambandsmått och signifikanstest. Vi återkommer till sambandsmått och signifikanstest, nu ligger fokus på att konstruera en korstabell som kan möjliggöra tolkningen av om och i så fall hur våra variabler är relaterade till varandra. För att underlätta denna tolkning väljer du först alternativet "Cells".

Att sammanställa tabellen endast med antal observationer i varje cell gör en

jämförelse svår. Under rubriken "Percentages" är det möjligt att markera om korstabellen ska sammanställas med rad- ("Row"), kolumn- ("Column"), och/eller totalprocent ("Total"). Radprocent innebär att summera varje rad till 100 procent, medan totalprocent innebär att redovisa hur stor andel varje cell utgör av samtliga observationer. Här har vi valt att markera kolumnprocent, med vilket avses att kolumnerna summeras upp till 100 procent.

När den oberoende variabeln är placerad i kolumnled och den beroende variabeln i radled möjliggör valet av kolumnprocent tolkningen av huruvida det verkar finnas ett samband mellan variablerna – om oro för brottsligheten i samhället skiljer sig åt mellan män och kvinnor. Eftersom vi konsekvent väljer att placera våra variabler på detta sätt kommer vi alltså i syfte att utreda ett eventuellt samband alltid vilja redovisa korstabellen med kolumnprocent. Klicka på "Continue" när du gjort ditt val. Slutligen ska vi ta en titt på alternativet "Format". Här kan du välja om radledet ska redovisas i stigande ("Ascending") eller fallande ("Descending") ordning. I syfte att tolka riktningen på sambandet kan valet av stigande eller fallande ordning underlätta, men detta är en smaksak och du kommer oavsett val kunna göra samma tolkning av korstabellen i syfte att utreda ett eventuellt samband. I detta fall har standardalternativet stigande ordning valts, vilket innebär att korstabellen i radled kommer att sammanställas med det lägsta värdet överst. Klicka på "Continue" när du gjort ditt val.

Klicka därefter "OK" så producerar SPSS en korstabell över relationen mellan kön och oro för brottsligheten. Resultatet visas i output fönstret. Så som vi har valt att sammanställa vår korstabell (oberoende variabel i kolumnled och beroende variabel i radled och sammanställd med kolumnprocent) är det nu möjligt att se om det verkar finnas ett samband mellan kön och oro för brottsligheten i samhället. Nästa steg är att tolka korstabellen. Tekniken är att jämföra kategorierna i den oberoende variabeln radvis.

I korstabellen kan vi se att 31 procent av männen jämfört med 17,5 procent av kvinnorna svarat att de inte alls är oroliga. Det verkar alltså som att det finns ett samband i den meningen att kvinnor är mer oroliga för brottsligheten än män.

Eftersom vi har en variabel på ordinal nivå (oro) och en på nominal nivå (kön), kan vi inte uttala oss om riktningen på sambandet, dvs. om det rör sig om ett positivt eller negativt samband (hur vi kodat variabeln kön, dvs. vilket kön som kodats som 1 eller 2, är ju godtyckligt).

6.1 Sambandsmått

I ovanstående exempel kunde vi, genom att tolka korstabellen, se att ett samband verkar föreligga mellan kön och oro för brottslighet. Ibland vill man även uttala sig om sambandets styrka och i detta syfte är användningen av sambandsmått bra. I de fall som sambandets riktning är tolkningsbart ger

sambandsmåtten även denna information. Statistiker har tagit fram olika sambandsmått som gäller för variabler som befinner sig på olika skalnivåer. För att välja sambandsmått börja med följande:

Analyze > Descriptive statistics > Crosstabs

Placera din oberoende variabel i kolumnled och din beroende variabel i radled. Välj även, precis som tidigare, att sammanställa korstabellen med kolumnprocent under alternativet "Cells". Klicka därefter på "Statistics".

Här får vi en viss vägledning av SPSS när det gäller vilka sambandsmått som är lämpliga att använda för våra variabler beroende på mät/skalnivå. Se vidare Djurfeldt m.fl. 2018: 149.

6.1.1 Samband mellan två numeriska variabler

När vi har att göra med variabler som befinner sig på intervall- eller kvot-skala är varken korstabell eller ovan nämnda sambandsmått lämpliga verktyg för att utreda ett eventuellt samband. Föreställ dig till exempel att undersöka sambandet mellan ålder och brott i en korstabell, där respondenten i enkätundersökningen har fått ange sin ålder och även självskatta antalet begångna brott under det senaste året – det skulle resultera i en enorm korstabell eftersom varje specifik ålder- och brottkombination kräver sin egen cell.

I detta fall är istället ett spridningsdiagram ("Scatter plot") lämpligt att använda för att studera huruvida ett samband verkar föreligga. Gör följande:

Graphs > Chart builder

Under "Gallery", klicka på "Choose from", välj "Scatter / Dot" och dra "Simple Scatter" upp till rutan "Chart Preview". Dra din oberoende variabel till rutan för x-axeln och din beroende variabel till rutan för y-axeln. Klicka därefter "OK".

Vi kan även använda sambandsmått för att beräkna styrka och riktning på sambandet. När vi har att göra med två kontinuerliga variabler är sambandsmåtten Pearson's r (korrelationskoefficienten r) lämpligt att använda. Gör följande:

Analyze > Correlate > Bivariate

För över de variabler du vill korrelera till rutan "Variables" och markera Pearson's r . Klicka därefter "OK". Precis som tidigare nämnda sambandsmått varierar Pearson's r på en skala mellan -1 och +1, där 0 indikerar att det inte finns ett samband medan -1 anger ett perfekt negativt samband och +1 anger ett perfekt positivt samband.

Ett alternativ till att studera relationen mellan ålder och brott på ovanstående vis är att klassindela de båda kontinuerliga variablerna till kategoriska variabler med hjälp av recode-kommandot (se under Databearbetning). Man kan tänka sig att klassindela ålder till de tre klasserna "ungdom", "ung vuxen" samt

"vuxen", samt brott till de tre klasserna "inga brott", "1-2 brott", "3 eller fler brott". På det sättet skulle vi konstruera två variabler på ordinal skalnivå utav de två ursprungliga variablerna på kvotskala. Därmed kan vi med de nya variablerna studera relationen mellan ålder och brott i en korstabell (i detta fall med nio celler) och med de sambandsmått som är lämpliga för variabler på ordinal skalnivå. Tänk på att variabler på högre skalnivå alltid kan transformeras till variabler på lägre skalnivå.

6.1.2 Samband mellan två kategoriska variabler

Analysera data III: Hypotesprövning

Chapter 7

Att välja rätt test vid hypotesprövning

Samhällsvetare arbetar nästan alltid med urval dragna ur en population men vill generalisera till hela populationen. Den gren inom statistik som hjälper oss att göra detta kallas statistisk inferens. Arbetsgången att pröva de samband vi är intresserade av är att ställa upp nollhypotes och mothypotes, där nollhypotesen uttrycker att det inte finns en skillnad medan mothypotesen uttrycker att det finns en skillnad. Det vi prövar är något förenklat om skillnaderna och de samband vi undersöker är tillräckligt stora för att kunna antas gälla i populationen. Om så är fallet förkastar vi nollhypotesen.

Under detta avsnitt följer två vanliga metoder för hypotesprövning för att avgöra om ett samband är signifikant: chi2 och t-test. Chi2 används när oberoende och beroende variabel befinner sig på nominal eller ordinal skalnivå, medan t-test används när den oberoende variabeln befinner sig på nominal eller ordinal skalnivå och har två kategorier medan den beroende variabeln är kontinuerlig och befinner sig på intervall- eller kvotskala.

Oavsett vilket av dessa signifikanstest som är lämpligt kommer ni på liknande sätt att tolka det p-värde som SPSS beräknar för att avgöra om ett samband är signifikant eller inte. Är en uppmätt skillnad "verklig" eller kan den bero på slumpen? Ett p-värde under 0.05 är signifikant på fem procents nivå, ett värde under 0.01 på en procents nivå och ett värde under 0.001 på en promilles nivå. Vanligt är att ställa upp fem procent signifikansnivå som gräns för vad som ska anses vara ett signifikant resultat. Det innebär att endast i fem fall av hundra skulle urvalet visa på en skillnad som egentligen inte existerar i populationen. Ett annat sätt att uttrycka detta på är att risken att vi förkastar en sann nollhypotes är fem procent (kallas även typ 1-fel).

Vilket statistiskt test man behöver använda för att utföra en hypotesprövning bestäms helt utifrån de variabler man använder sig av. Listan på vilka statistiska

test som finns att tillgå är lång men det finns några få som förekommer ofta eller som är praktiska att känna till eftersom de underlättar förståelsen för andra statistiska test. I kursen Metod II kommer ni att stifta bekanskap med tre statistiska test: t-test, χ^2 och regression. Den sistnämnda kan i sin tur delas in i två kategorier: enkel och multipel.

Med hjälp av nedanstående flödesschema kan ni avgöra vilket statistiskt test som lämpar sig bäst för de variabler som ni önskar att analysera. Observera att ett ytterligare test finns omnämnt i flödeschemat (linjär sannolikhetsmodell/logistisk regresssion) men som vi ej kommer gå igenom under kursen gång. Den finns med eftersom den introduceras i samband med den kvantitativa metodkursen på avancerad nivå.

Chapter 8

Chi2-test

I exemplet ovan kunde vi se att oro för brottsligheten i samhället skiljer sig för män och kvinnor. Utifrån vår korstabell kan vi dock endast uttala oss om förhållandet i urvalet. Vad vi nu vill ta reda på är om detta samband mellan kön och oro för brottsligheten är tillräckligt tydligt för att kunna antas gälla i populationen, dvs. i befolkningen. Eftersom vi har två kategoriska variabler är chi2 ett lämpligt signifikansmått. Proceduren i beräkningen av chi2 är att jämföra observerade frekvenser med de frekvenser som skulle förväntas om det inte fanns någon skillnad. Genom denna jämförelse får vi fram ett chi2-värde som vi i nästa steg kan jämföra med en så kallad chi2-fördelning för att ta reda på huruvida detta värde överstiger ett kritiskt värde som motsvarar en given signifikansnivå. Genom denna procedur skulle vi utifrån vår korstabell kunna räkna ut huruvida sambandet är signifikant eller inte, men som tur är har vi SPSS till vår hjälp. Arbetsgången är följande:

Analyze > Descriptive statistics > Crosstabs

Välj precis som tidigare att placera "kön" i kolumnled och "S4" i radled. Välj även, precis som tidigare, att sammanställa korstabellen med kolumnprocent under alternativet "Cells". Klicka därefter på "Statistics" och markera alternativet "Chi-square". Klicka på "Continue" och därefter "OK". Utöver korstabellen får du nu även en tabell som ger information om ditt begärda Chi2-test.

Titta på raden "Pearson Chi-Square" och kolumnen "Asymp. Sig. (2-sided)". Här kan du se att ditt p-värde är mindre än 0.01. Skillnaderna är alltså så pass stora att dessa på en procents signifikansnivå kan antas gälla i populationen. Observera att du även har ett stort antal observationer, vilket även det har betydelse för dina möjligheter att finna skillnader som går att generalisera (gör att du får ett högre Chi2 värde). Sambandet är signifikant på en procents signifikansnivå och vi kan alltså generalisera våra resultat till att gälla för hela populationen.

Observera dock att SPSS anger att 0 celler har förväntade värden som understiger fem. Som regel gäller att Chi2-testet är ogiltigt om 20% av cellerna har ett förväntat värde mindre än 5 eller en cell har ett förväntat värde mindre än 1 då korstabellen är större än 2x2. Om korstabellen är 2x2 får ingen av cellerna ha ett förväntat värde som är mindre än 5. I detta fall är dock signifikanstestet giltigt. Om chi2- testet skulle visa sig vara ogiltigt kan lösningen vara att klassindela sina variabler med hjälp av kommandot Recode (se ovan under avsnittet Databearbetning).

Chapter 9

T-test

När du vill pröva om en skillnad mellan två grupper är signifikant och utfallsvariabeln är kontinuerlig är ett oberoende t-test tillämpligt. Här är proceduren att jämföra medelvärden mellan två grupper och utifrån den uppmätta skillnaden ta reda på om den är tillräckligt stor för att antas gälla i populationen. Anta att du är intresserad av om det finns någon skillnad i pojkars och flickors medelbetyg. Du vill jämföra ett sammanfattande mått på deras betyg i kärnämnen svenska, engelska och matematik. Betygsvariablerna är på ordinalskala och antar värden 0-3, se frekvenstabell nedan för betyg i svenska:

Du skapar ett betygsindex ("Mbetyg") som varierar mellan 0-3 genom att summerna de tre betygen och dividera med tre med hjälp av funktionen compute (se ovan under databearbetning). För att göra ett oberoende t-test, där du prövar om det finns en signifikant skillnad mellan pojkars och flickors genomsnittliga betyg, är tillvägagångssättet följande:

Analyze > Compare Means > Independent Samples T-test

Oberoende variabel är i detta fall "kon" medan den beroende variabeln är "Mbetyg". Flytta den beroende variabeln till rutan "Test Variable(s)" och den oberoende variabeln till rutan "Grouping Variable". Därefter måste du definiera vilka grupper inom denna variabel som du är intresserad av. Genom att i variabelnfönstret undersöka variabeln "kon" ser du att variabelvärde 0 står för "flickor" och variabelvärde 1 står för "pojkar". Klicka på "Define Groups" och ange dessa variabelvärden i rutorna "Group 1:" respektive "Group 2:". Klicka därefter "Continue".

Klicka "OK" och gå till Output-fönstret. Du får nu fram resultaten i två tabeller. I den första tabellen redovisas deskriptiva mått som antalet i respektive grupp samt respektive grupps medelvärde och standardavvikelse på den beroende variabeln.

Vi ser att medelvärdet på betygsindex är 1,59 för flickor och 1,39 för pojkar.

Flickor har alltså högre betyg. Men är denna skillnad signifikant, d.v.s. kan anta att den även existerar mellan pojkar och flickor i populationen? För att avgöra det tittar vi på nästa tabell:

I tabellen ovan redovisas ett antal relevanta mått för det oberoende t-testet. "Mean Differance" är skillnaden mellan de båda medelvärdena. I detta fall är skillnaden i genomsnittligt betyg 0,196. Frågan är emellertid om skillnaden är tillräckligt stor för att fastställa att medelvärden skiljer sig åt i populationen? P-värdet vid ett tvåsidigt hypotestest går att finna under rubriken "Sig. (2-tailed)". Ett värde under 0,05 är signifikant på fem procents nivån. Här har vi ett värde på 0,00 vilket understiger denna gräns. Vi kan således förkasta nollhypotesen som uttryckte att det genomsnittliga betyget är samma för pojkar och flickor. (se vidare Djurfeldt m.fl. 2010/2018, s 230ff).

När vi gått igenom t-test på lektionerna har ett antagande gjorts om att de båda grupperna har samma varians. Det finns emellertid två olika test - med eller utan antagande om lika varians (equal variance assumed / equal variance not assumed). Skall man gå korrekt tillväga kontrollerar man att antagandet håller. Det görs med Levene's test for Equalite of Variances. F är kvoten av de båda gruppernas varians och om denna kvot inte är lika med 1 kan det signalera att antagandet om lika varians inte håller. Om sannolikheten för detta F-värde är mindre än 0,05 drar vi slutsatsen att skillnaden mellan urvalens varians reflekterar en skillnad i populationernas varians. Om så är fallet, vilket det emellertid inte är här ($P = 0,10$) går vi till "Equal variance not assumed". "Equal variance assumed" är mer restriktivt vilket betyder att det är svårare att få ett signifikant resultat. Därför kan ni lika gärna använda detta test.

Chapter 10

Enkel regression

Chapter 11

Multipel regression

Visualisera data

Chapter 12

Bearbeta tabeller och figurer

12.1 Tabeller

12.2 Figurer

Övrigt

Chapter 13

Presentation av dataset

13.1 NTU 2017-2021

13.2 NTU 2013-2015

Datafilen "NTU 2013-15 M2.sav" är en SPSS fil som rymmer tre hela årgångar av NTU, sammanlagt 37 118 observationer (personer som besvarat frågorna). Det går inte att identifiera vilket av de tre åren som respektive person ingått i undersökningen, varför datamaterialet behandlas som en tvärsnittsundersökning (NTU 2013-15).

Närmare information om urval, datainsamling, frågekonstruktion, kodning mm återfinns i den tekniska rapporten för NTU 2015:

Brå (2016). Den nationella trygghetsundersökningen 2015. Teknisk rapport. Brå rapport 2016:3.

Denna finns som pdf fil på kurssajten, men kan även laddas ner härifrån.

För att arbeta med NTU datamaterialet behöver ni hjälp av information från den tekniska rapporten, exempelvis ser ni där hur frågor och svarsalternativ är utformade. Utifrån variabelnamnen i datafilen går det att identifiera frågorna i frågeformuläret.

Exempel:

I datafilen finns en variabel som heter C6, av variabelns label (etikett) framgår att den rör cykelbrott: "Cykelstöld_CY_C6". Att frågan rör cykelstöld är tydligt, men vilken av alla frågor som rör cykelstöld? Genom att frågenumret i frågeformuläret anges i variabeländelsen, i exemplet "C6", kan vi koppla till frågeformuläret i den tekniska rapporten (Brå 2016, bilaga 1, sid 5):

Ni har enbart tillgång till vissa frågor i NTU – bakgrundsfrågor om t.ex. ålder och kön samt frågor om utsatthet för brott, oro för brott och förtroende för rättsväsendet.

13.3 Skolundersökning 2005

Datafilen "Skol05.sav" är en SPSS fil med data från Brottsförebyggande rådets skolundersökning om brott från 2005 (SUB2005). Undersökningen rymmer ursprungligen 7449 observationer (deltagande elever), ni har dock ett slumpurval av hälften av de svarande (3724 elever). Ni har tillgång till de flesta frågor som ingick i studien, tex om egen brottslighet och utsatthet för brott (totalt rör det sig om cirka 190 variabler).

Närmare information om urval, bortfall, datainsamling, frågekonstruktion, kodning mm återfinns i den tekniska rapporten för SUB 2005:

Brå (2008). Den nationella skolundersökningen om brott 1995–2005 Teknisk rapport. Brå rapport 2008:2.

Denna finns som pdf fil på kurssajten, men kan även laddas ner här

För att arbeta med skolundersökningen behöver ni hjälp av information från den tekniska rapporten, exempelvis ser ni där hur frågor och svarsalternativ är utformade. Utifrån variabelnamn och label i datafilen går det enkelt att identifiera frågorna, för deras exakta lydelse får ni dock gå till den tekniska rapporten.

13.4 Pathways to desistance

Datamaterialet (PATHWAYS_DO2.sav) är hämtat från en amerikansk studie vid namn Pathways to desistance. I studien så fick cirka 1 300 brottsbelastade ungdomar besvara en rad frågor om deras liv.

Kort information om datamaterialet från projektets egna hemsida:

The Pathways to Desistance study is a multi-site, longitudinal study of serious adolescent offenders as they transition from adolescence into early adulthood. Between November, 2000 and January, 2003, 1,354 adjudicated youths from the juvenile and adult court systems in Maricopa County (Phoenix), Arizona (N = 654) and Philadelphia County, Pennsylvania (N = 700) were enrolled into the study. The enrolled youth were at least 14 years old and under 18 years old at the time of their committing offense and were found guilty of a serious offense (predominantly felonies, with a few exceptions for some misdemeanor property offenses, sexual assault, or weapons offenses). Each study participant was followed for a period of seven years past enrollment with the end result a comprehensive picture of life changes in a wide array of areas over the course of

this time. The study was designed to: 1) To identify distinct initial pathways out of juvenile justice system involvement and the characteristics of the adolescents who progress along each of these pathways. 2) To describe the role of social context and developmental changes in promoting desistance or continuation of antisocial behavior. 3) To compare the effects of sanctions and selected interventions in altering progression along the pathways out of juvenile justice system involvement.

ÖVERSÄTT OCH GÖR ÖVANSTÅENDE TILL EGEN TEXT

Mer information om datamaterialet och studien finner ni här.

Chapter 14

Dataövningar

14.1 Dataövning 0

14.2 Dataövning 1

14.3 Dataövning 2

14.4 Dataövning 3

14.5 Dataövning 4

[Click here for explanation](#)