Introduksjon til SPSS

Johan Håkon Bjørngaard Institutt for samfunnsmedisin, NTNU

Mål:

Deltakerne skal få innblikk i

- Oppretting av datafil. Innlesing eller inntasting av data. Redigering.
- Presentasjon av data deskriptiv statistikk. Tabeller og grafer.
- Enkle analyser av data

Noen statistikk programpakker

- Generelle:
 - S-plus¹), R⁰), SAS, Stata²), Statistica, MINITAB¹), SPSS¹)
- Spesielle:
 - LISREL, StatXact²⁾, LogXact²⁾, SamplePower¹⁾, nQuery
- Grafikk:
 - SigmaPlot¹⁾
- Regneark:
 - Excel¹⁾

⁰⁾ Gratis fra nettet, ¹⁾ NTNU-lisens, ²⁾ lisens ved DMF

SPSS

- Generell statistikk-programpakke
- Enkel å bruke
- Enkel innlegging, strukturering og redigering av datafil
- Mye output
- Begrenset metodedokumentasjon
- Begrenset på avanserte / spesielle metoder
- "dominerende" innen medisin i Norge.

Litteratur - SPSS

- Kinear & Gray: "SPSS 16 made simple", 2008, ISBN 978-1841697291
- Peat, Jennifer & Barton, Belinda: "Medical Statistics: A Guide to Data Analysis and Critical Appraisal". Blackwell publishing / BMJ books, 2005, ISBN 978-0727918123

Dokumentasjon fra SPSS Inc.

- SPSS User's Guide / Regression Models / Advanced Models / Tables / Trends / Categories / Conjoint / Exact Tests / Missing Value Analysis / Interactive Graphics
 - Bøker eller elektronisk ved NTNU.
 - DMF: P:\programvare\SPSS16\Manualer
- Help -> Topics
- Help -> Algorithms: Teknisk beskrivelse / definisjoner av metoder

Statistikk - litteratur

- Bowers, David: "Medical Statistics from Scratch. An Introduction for Health Professionals" 2nd ed, Wiley, 2008. ISBN 978-0-470-51301-9.
- Aalen, Odd. m.fl.: Statistiske metoder i medisin og helsefag. Gyldendal, 2006. ISBN: 82-05-34685-2
- Bernhard Rosner: "Fundamentals of Biostatistics", 6th ed. 2005. ISBN:0-534-41820-1.

Vinduer i SPSS

- Data editor:
 - Data View
 - Variable View
- SPSS Viewer (Output)
- Syntax Window (kommandovindu)
- ... og noen andre mindre brukte

Flytting av filer mellom SPSS-versjoner

• Datafil:

– Små format-endringer. Går vanligvis bra

- Output-fil:
 - Kan være vanskelig / umulig
 - Helt nytt grafikk-format fra versjon 11 til 12
 - Også problemer mellom versjon 13 og 14.0
- Syntax-fil:
 - Går vanligvis bra

SPSS - struktur for datafil: Cases, variable og verdier

- Case:
 - Eks: Person
- Variable:
 - Eks: identifikasjonsnr, alder, høyde, kjønn
- Verdier:

– Eks: 205, 45, 178,2, "kvinne"

• Todimensjonal

Repeterte målinger - 2 alternative format:

- 1. Hver pasient som case
- 2. Hvert måletidspunkt for hver pasient som case.
- Svitsje mellom format 1 og 2 vha Data editor -> data -> Restructure

Variable view - 1

- Name:
 - Inntil 64 tegn (bokstaver, tall, @, #, _,\$)
 - start med bokstav
 - ingen mellomrom, ingen *, ?, !, …
 - Bokstavene æ, ø, å frarådes.
 - Ikke "and", "or", "not", ...
- Type: Numeric, date, string, etc
- Width (på datafil), særlig aktuelt ved "string"

Variable view - 2

- Decimals (på skjerm/utskrift)
- Label: Inntil 120 tegn, alle tegn tillatt
- Values: F.eks 1 = mann, 2 = kvinne
- Missing: F.eks 98 = ble ikke spurt, 99=ikke besvart
- Column (på skjerm/utskrift)
- Align
- Measure: Scale, Ordinal, Nominal

Missing:

- "System missing" ingen innlest verdi. (Ikke for variabel type "string")
- Brukerdefinert missing:
 - Kan skille mellom årsaker til missing
 - Kan skille mellom missing og glemt å legge inn tall
 - Kan få problemer ved lagring i andre format enn SPSS

Measure, eksempler:

- Scale:
 - høyde i cm
- Ordinal:
 - Røyker du? 1 = "Daglig", 2 = "Av og til",
 - 3 = "Sluttet siste halvår", 4 = "Sluttet tidligere",
 - 5 = "Har aldri røkt"
- Nominal:
 - Sivilstand: 1 = "ugift", 2 = "gift",
 3 = "samboer", 4 = "skilt", 5 = "enke(mann)" ¹⁵

Tid mellom to datoer

- Transform -> Date/Time -> Calculate with dates and times
 - Gir tid rundet av ned til f.eks antall hele år
 - SPSS 16: Du kan velge å beholde desimaltall
- Hvis du ønsker desimaltall antall år:
 - Regn ut antall hele dager som ovenfor
 - Transform -> Compute -> Numeric expression: dager/365.25

Innlesing av data

- Inntasting i data editor
- Hente SPSS fil
- Innlesing fra annen fil, f.eks EXCEL
 - vha File -> Open -> Data -> Files of Type (Anbefales vanligvis. Men ikke mulig for alle versjoner)
 - vha Copy Paste i Windows
 (Nyttig ved enkle/små filer eller data-områder.
 OBS: kan miste noe informasjon)

Deskriptiv statistikk - tabeller

- En skalavariabel (eller kategorisk variabel):
 Descriptive statistics -> Descriptives
- En kategorisk variabel:
 - Descriptive statistics -> Frequencies
- To kategoriske variable:

– Descriptive statistics -> Crosstabs

Deskriptiv statistikk – tabeller (forts)

• En skalavariabel og en kategorisk variabel:

– Compare means -> Means

- To skalavariable:
 - Kategorisere den ene variabelen
 - alternativt: enkel lineær regresjon

UTVALG AV RESPONDENTER

- Av og til ønsker vi å gjennomføre noen analyser på bestemte deler av materialet. I SPSS kan vi angi hvilke respondenter vi er interessert i å velge ut i forbindelse med en kommando.
- Select if
- Split file

MANIPULERING AV DATA

- Tilrettelegger data for å kunne analysere dataene på en hensiktsmessig måte.
 - Slå sammen verdier på en variabel (omkoding)
 - Slå sammen flere variabler til en ny variabel:
- Summere sammen verdier på flere variabler (additiv indeks)
- Lage nye variabler ved å kombinere verdier på eksisterende variabler
- Konstruere nye variabler

SPSS - kommandoer:

- recode ... into
- compute
- if

SLÅ SAMMEN VERDIER PÅ EN VARIABEL (OMKODING)

• Ønsker å gruppere verdiene på en variabel i færre verdi-klasser, blant annet for å få mer oversiktlige frekvensfordelinger og krysstabeller.

Kommando: recode ... into

• Denne kommandoen lager en ny variabel med de nye verdikategoriene.

COMPUTE - kommandoen

- Benyttes generelt til å lage nye variabler, ved
 - å kopiere en eksisterende variabel under et nytt navn
 - slå sammen eksisterende variabler ved hjelp av ulike regneoperasjoner (+,-,/,*)
 - gjøre beregninger på eksisterende variabler (f.eks. transformere variabelen)

Visual binning

• Nyttig for å rekode skalavariabler

Bruk syntax-filer

- Viktig for å dokumentere hva man har gjort
- Lettere å finne ut hva man har gjort feil
- Kan gjøre samme prosedyre på nytt hvis noe går galt

Feilsøking og feilretting

- Descriptive statistics

 Frequencies, Descriptives, Crosstabs
- Oppretting av funne feil:
 - Edit Find (I Data Editor Data View)
 - eller:
 - Transform -> Recode -> Into same variables
- NB! Det er alltid lurt å ha en backup av opprinnelig fil

Kopiering av tabeller / grafer til Word, Excel eller Power Point (1):

- I SPSS:
 - Rediger ferdig objektet (tabellen/ grafen)
 - Tabell eller graf fra SPSS Viewer.
 - Merk objektet.
 - Edit -> Copy Object (eller Copy)
 - Graf fra SPSS Chart Editor:
 - Edit -> Copy Chart

Kopiering av tabeller / grafer til Word, Excel eller Power Point (2):

- I Word:
 - Sett markøren på rett sted
 - Edit -> Paste special -> Picture -> IKKE float over text
- I Excel (tabell)
 - Sett markøren på rett sted
 - Edit -> Paste

Standard vs interaktive grafer

- De fleste graftypene finnes både standard og interaktivt
- Interaktive grafer:
 - Er mer fancy(?)
 - Trenger mer tid og minne ved fremstilling, mer plass ved lagring.
- Bruk standard grafer hvis mulig(?)

Noen nyttige grafer

- Én kategorisk variabel:
 - Bar chart (stolpediagram)
 - Pie chart (kakediagram)
- To kategoriske variable:
 - Clustered bar chart (klynget stolpediagram)

Noen nyttige grafer (forts.)

- Én skalavariabel:
 - Histogram
- To skalavariable:
 - Scatterplot

Noen nyttige grafer (forts.)

- Én skalavariabel og én kategorisk variabel (sammenlikne skalavariabelen i to eller flere grupper):
 - Dot plot eller scatter plot (ved "få" observasjoner)
 - Box plot (ved "mange" observasjoner)

TIPS

- Det er viktig å prøve og feile mye!
- Så lenge du har backup-fil, kan du ikke ødelegge noe
- Bruk syntax-fil for dokumentasjon
- Lykke til!