

Introduksjon til SPSS

Inger Johanne Bakken

Enhet for anvendt klinisk forskning, NTNU
og
Avdeling for forebyggende helsearbeid, SINTEF

Mål

- Deltakerne skal få innblikk i:
- Hvordan komme i gang med SPSS, ulike datavindu.
- Kompatibilitet med andre dataformatet.
- Oppretting av datafil. Innlesing eller inntasting av data. Redigering.
- Presentasjon av data – deskriptiv statistikk.
- Tabeller og grafer.
- Enkle analyser av data

Noen programpakker statistikk

Generelle:

– S-plus¹⁾, R²⁾, 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 (Statistical Package for the Social Sciences)

- Første versjon allerede i 1968
- Lav kom-i-gang terskel
- Enkel innlegging, strukturering og redigering av datafil
- Enkelt å hente inn data fra andre dataformat (Excel, tekst, SAS)
- Enkelt å restrukturere hele datasettet
- Begrenset metodedokumentasjon (?)
- Kan benyttes til de aller fleste "vanlige" og "mindre vanlige" analysemetoder, men er begrenset på avanserte/svært spesielle metoder
- "Dominerer" innen medisinsk forskning i Norge

Litteratur SPSS

- Arthur Griffith: "SPSS for Dummies"
- Andy Field: "Discovering statistics using SPSS"
- Sarah Boslaugh: "An Intermediate Guide to SPSS Programming: Using Syntax for Data Management"

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

- Mange gode hjelpemidler på nettet, eksempler:
<http://www.ats.ucla.edu/stat/>
<http://www.spsstools.net/>

Dokumentasjon fra SPSS Inc.

Help -> Topics

Help -> Algorithms: Teknisk beskrivelse / definisjoner av metoder

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\SPSS15\Manualer

Medisinsk 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.
- Andy Field: "Discovering statistics using SPSS"



Åpner SPSS – Hva ser du?

- Datavindu
For innlegging/visning av data
- Variabelvindu
For definisjon/visning av variabelformat
- Outputvindu
For presentasjon av analyser og feilmeldinger
- Syntaxvindu (ikke påkrevd hvis du bare bruker rullegardinmenyene)
For alle slags kommandoer:
 endring av dataformat
 rekoding
 lagring av fil
 kjøring av analyser

SPSS filtyper

- Datafil (*.sav)
Data + variabelinformasjon ("dictionary")

(svarer til datavindu og variabelvindu)
- Output (*.spo)
Resultater fra kjøringer
- Syntax (*.sps)
Må brukes aktivt – program som svarer til økta

Hva er syntax?

- Når du bruker vindusmenyene, genererer du syntax – men ser det ikke med mindre du ber om det!
- Når vi nå åpner SPSS igjen skal vi se
 - 1) Paste (lim inn syntax via rullegardin)
Eks. frekvensanalyse: Analyze – Descriptive statistics - Frequencies PASTE
 - 2) Slå på syntax loggen i options
Edit – Options – Viewer – Display commands in the log

Hvorfor Syntax?

- SPSS gir mye output - > kan være vanskelig å holde orden (bokholderi) basert på output alene
- Men:
- Lav terskel for å komme i gang med programmering ("syntax")
 - Syntax er en effektiv måte å holde orden på hvordan man har "ryddet" dataene og hvilke analyser som er benyttet
 - Syntax gjør det også enkelt å kjøre de samme analysene om igjen med små modifikasjoner på samme datasett, eller på et nytt datasett
- Eksempel:
- 1) Kjørt alle analysene, oppdager at én variabel er kodet feil
 - 2) Gjennomført en spørreundersøkelse 2008, skal gjennomføre tilsvarende undersøkelse 2009

Mange versjoner SPSS

Nå: 16.0

Innhenting fra eldre versjoner:

- Datafil går (nesten?) alltid bra
- SYNTAX går alltid bra, er i prinsippet bare en tekstfil.
Men må kanskje redigeres noe.
- Output
Blir ofte krøll. Store endringer fra versjon 11 til 12 og også problemer mellom 13 og 14.0

Struktur i datafilen

- Case:
 - Øftest Person innen medisinsk forskning
- Variable:
 - Eks: identifikasjonsnummer, dato for registrering, kjønn, fødselsår, høyde, vekt
- Verdier
 - Eks: 1, 2008-05-26, kvinne, 1975, 170, 65

Vi tok for oss: Klargjøring av datafil

- Åpne eksisterende fil Open – finn frem fila, åpnes som i andre applikasjoner
- Ved manuell innlegging av data: Starter med å definere variablene
- Korte variabelnavn (enkelte regler for valg av navn, feil gir feilmelding)
- Definisjon av variabeltypen er viktig. Streng er "hva som helst". Width angir antall tegn som faktisk lagres, column det som vises på skjermen og i utskrifter. Lett å trekke ut.
- To desimaltegn er standard for nye numeriske variable (kan endres)
- Kan kopiere informasjon fra en variabel til en annen
- Label: Her kan du legge inn et hensiktsmessig variabelnavn (informativt men ikke for langt). Hvis du har lagt inn label er det dette som kommer frem i output
- Kan slå sammen to eksisterende filer ved å bruke Data - Merge data

Til SPSS!

Missing

- "System missing" - ingen innlest verdi. (Ikke for variabel type "string")
- Brukerdefinert missing:
 - Kan skille mellom årsaker til missing
 - Bruk verdier som ligger langt unna de reelle verdiene
eks. nektet å oppgi f_aar: '-100'
uleselig f_aar: '-1000

Til 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)"

Vi tok for oss: Sortering og endring av strengvariable

- Sortering av data ved høyreklikk variabelnavn eller via Data – Sort
- Oppdage duplikater ved Data – Identify duplicate cases
- Slette data, enten marker og delete, eller via Data – Select cases
- Enkelt å gjøre om en dato i streng format til en dato som SPSS gjenkjenner som dato
- Strengvariable er generelt tungvinte å arbeide med, vi så hvordan
- Transform – Automatic recode kan brukes til å gjøre om til numerisk format, og at all informasjon i strengen blir ivaretatt gjennom "Labels"
- Kontrollerte dette i vår aller første analyse: En krystabell mellom gammel og ny variabel for kjønn.

Vi tok for oss: Beregning av nye variable og litt om SYNTAX

- Hvordan beregne en ny variabel basert på eksisterende variable:
I vindusmenyen: Transform – Compute variable
 - Legg inn riktige beregninger (eks høyde i meter og BMI)
- Vi så også hvordan man ved å trykke på Paste ikke kjører kommandoen (i dette tilfellet beregning av ny variabel) men limte kommandoen inn i syntax:

```
COMPUTE BMI = vekt/(hoyde_m*hoyde_m) .  
EXECUTE .
```
- SYNTAX kjøres ved å markere og trykke på "play".
- SYNTAX kan enkelt redigeres hvis vi oppdager feil kan lagres til ei senere økt

Vi tok for oss:

Fordel med datoformat fremfor streng

- Datoer kan man gjøre beregninger med (mer komplisert med dato som streng)
- Eks:
Ny variabel for år registrert?
- Transform Trykk på function group og velg date extraction så kommer bare aktuelle opp. Velg XDATE.YEAR og legg inn dato for registrering. Resultat: ny variabel som angir året for registreringen
- Vi brukte denne til å beregne alder i år.
- Andre aktuelle: Tid mellom to datoer, ukedag for datoen, antall dager/uker siden datoen
- Nå har vi en datafil som vi kan bruke til noen enkle analyser.

Deskriptiv statistikk - tabeller

- En skalavariabel
– Descriptive statistics -> Descriptives

Eksempel: alder og BMI

- En kategorisk variabel:
– Descriptive statistics -> Frequencies

Eksempel: Kjønn

- To kategoriske variable:
– Descriptive statistics -> Crosstabs

Eksempel: kjønn og røyking (la på valget at vi ville ha prosentvis fordeling ved å trykke på cells)

Deskriptiv statistikk – tabeller (forts)

- Deskriptiv statistikk – tabeller (forts)
- En skalavariabel og en kategorisk variabel:
– Compare means -> Means

Eksempel: alder for kvinner og menn
- To skalavariabel:
– Kategoriser den ene variabelen
– alternativt: enkel lineær regresjon

(kommer tilbake til dette i øvingen)
- Forlater nå den lille testfilen og åpner en større 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

Hypotesetesting

- Sett opp nullhypotese og alternativ hypotese.
- Eksempel:
 H_0 : Forventet respons er lik i gruppene
 H_0 : Forventet respons er forskjellig

– P-verdien (signifikanssannsynlighet, sig.) er sannsynligheten for å få de observerte verdier eller noe mer ekstremt, gitt at H_0 er sann.
– Forkast H_0 hvis p-verdi er mindre enn signifikansnivå (f.eks 0,05 eller 0,01)

Valg av statistisk metode

- Enkelte tester forutsetter en bestemt modell.
 - Eksempel: T-test forutsetter at data er (tilnærmet) normalfordelt
- Ikke-parametriske tester er mer fleksible
 - Eksempel: Sammenlikning av to medianer:

Nonparametric tests > 2 independent groups (Mann-Whitney)

Beregning av p-verdier i SPSS

Asymptotisk og eksakt finnes i

- Nonparametrics (Wilcoxon-Mann-Whitney, Kruskal-Wallis m.fl.)

Asymptotisk er default når denne opsjonen er gitt

- Bør velge eksakt ved små datasett
- Eksakt er for tidkrevende ved store datasett

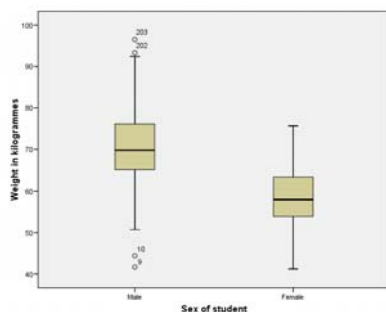
Grafer i SPSS

- De fleste grafer finnes som standard og interaktivt
- Standard holder for de fleste formål
- Chart Builder, mange muligheter
- Interactive
- Legacy Dialogs, standard grafer (som i SPSS 14)
- Analyse: Kan ofte velge grafer her også, eksempler:
 - Descriptive statistics: Q-Q plot
 - Descriptive statistics: Frequencies – Charts - Options

Noen nyttige grafer

- Scatterplot (to skalavariabel)
- Histogram
- Stolpediagram (bar chart)
- Box plot

Box plot



Oppsummering

- SPSS er lett å komme i gang med
- For etablering av filer:
 - Legg inn data manuelt, hent inn fra andre program, kombiner filer, restrukturer filer
- Transformer og kombiner variable
 - Kategorisere, beregne, kombinere ulike variable
- Svært mange analysemuligheter
- Lett å klikke og dra i vindusmenyene
- Lim inn ("paste") fra vindusmenyer til syntax og du er i gang med programmering!

Lykke til!

