

Introduksjon til IBM SPSS Statistics 20

av
Stian Lydersen

NTNU

Revidert 13 aug 2012

http://folk.ntnu.no/slyderse/medstat/SPSS/Introduksjon_SPSS.pdf

http://folk.ntnu.no/slyderse/medstat/SPSS/Introduction_SPSS.pdf

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

2

Noen statistikk-programpakker

- Generelle:
 - R⁰⁾, SAS, Stata²⁾,
 - SPSS¹⁾, MINITAB¹⁾, Statistica
- Spesielle:
 - SamplePower¹⁾, Amos¹⁾, LISREL, M-plus, StatXact, LogXact
- Grafikk:
 - SigmaPlot¹⁾
- Regneark:
 - Excel¹⁾

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

3

SPSS

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

4

Litteratur

- Kinear & Gray: ”IBM SPSS Statistics 19 Made Simple”, 2011
- Bowers, David: ”Medical Statistics from Scratch. An Introduction for Health Professionals” 2nd ed, Wiley, 2008. ISBN 978-0-470-51301-9.

5

Dokumentasjon fra SPSS Inc.

- Statistics Base / Regression / Advanced Models / Categories / Conjoint / Exact Tests / Missing Values / Bootstrapping / samt mange flere
 - http://download.spss.no/SPSS_Statistics_19_Doc_en.zip
 - Noen finnes også i bokform
- Help -> Topics
- Help -> Algorithms: Teknisk beskrivelse / definisjoner av metoder

6

Filtyper

- Datafil (*.sav)
- Viewer-fil (*.spv)
- Syntax-fil (*.sps)
- ... og noen andre mindre brukte

7

Flytting av filer mellom SPSS-versjoner

- Datafil:
 - Små format-endringer. Går vanligvis bra
- Viewer fil:
 - Kan være vanskelig eller umulig
 - Viktige resultater bør lagres som syntax og/eller pdf fil
- Syntax-fil:
 - Går vanligvis bra

8

Struktur for datafil: Cases, variable og verdier

- Case:
 - Eks: Person
- Variable:
 - Eks: identifikasjonsnr, alder, høyde, kjønn
- Verdier:
 - Eks: 205, 45, 178,2, "kvinne"

9

Repeterte målinger - 2 alternative format:

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

10

Vinduer

- Data Editor:
 - Data View
 - Variable View
- Viewer
- Syntax Editor
- ... og noen andre mindre brukte

11

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"

12

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
- Role: Input, Target, Both, None, Partition, Split

Missing:

- “System missing” - ingen innlest verdi. Enklest og nesten alltid OK! (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

14

Measure, eksempler:

- Scale:
 - høyde i cm
- Ordinal:
 - Hvordan er helsa di? 1 = “Dårlig”, 2 = “Ikke helt god”, 3 = “God”, 4 = “Svært god”
- Nominal:
 - Siviltstand: 1 = “ugift”, 2 = “gift”, 3 = “samboer”, 4 = “skilt”, 5 = “enke(mann)”

15

Tid mellom to datoer

- Transform -> Date and Time Wizard -> Calculate with dates and times
 - Default: Gir tid rundet av ned (trunkert) til f.eks antall hele år
 - Anbefalt: Du kan velge å beholde desimaltall

16

Beregning av nye variable

- Eksempel: Kroppsmasseindeks
- $BMI = (\text{vekt i kg}) / (\text{høyde i meter})^2$
- Transform -> Compute variable

17

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)
 - vha Copy - Paste i Windows (Nyttig ved enkle/små filer eller data-områder. Kan miste noe informasjon)

18

Deskriptiv statistikk - tabeller

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

19

Deskriptiv statistikk – tabeller (forts)

- En skalavariabel og en kategorisk variabel:
 - Compare means -> Means
- To skalavariabeler:
 - Kategoriser den ene variabelen
 - alternativt: enkel lineær regresjon

20

Feilsøking og feilretting

- Descriptive statistics
 - Frequencies, Descriptives, Crosstabs
- Oppretting av funne feil:
 - Finn feilen(e):
 - Edit - Find (I Data Editor - Data View), eller
 - Data -> Sort cases
 - Rett opp eller slett feil verdier

21

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

- I SPSS:
 - Rediger ferdig objektet (tabellen/ grafen)
 - Edit -> Copy
- I Word:
 - Sett markøren på rett sted
 - Edit -> Paste special -> (velg et passende format)
- I Excel (tabell)
 - Sett markøren på rett sted
 - Edit -> Paste

22

Hypotesetesting

- Sett opp nullhypotese og alternativ hypotese.
Eksempel:
 - H_0 : Forventet respons er lik i gruppene
 - H_1 : 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)

23

Valg av statistisk testmetode.

- 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)

24

Eksakt versus asymptotisk

- Eksakt:
 - $0.215 * 0.529 = 0.113735$
- Tilnærmet:
 - $0.215 * 0.529 \approx 0.2 * 0.5 = 0.1$
- Asymptotisk betyr tilnærmet, med bedre tilnærming jo større n

25

Beregning av p-verdier i SPSS:

- Asymptotiske og eksakte metoder finnes i
 - Crosstabs
 - Nonparametrics (Wilcoxon-Mann-Whitney, Kruskal-Wallis m.fl.)
- Asymptotisk er default
- Bør velge eksakt ved små datasett
- Eksakt er for tidkrevende eller umulig ved store datasett

26

Noen nyttige grafer

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

27

Noen nyttige grafer (forts.)

- Én skalavariabel:
 - Histogram
 - Sammenlikne data med normalfordeling: Q-Q plot lettere å lese og tolke enn “normal curve overlay” i histogram
- To skalavariabel:
 - Scatterplot

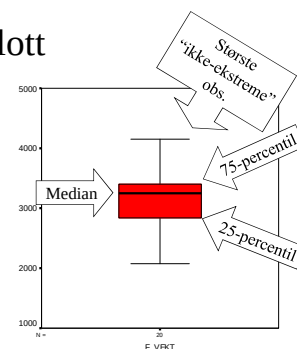
28

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)

29

Box-plott



30