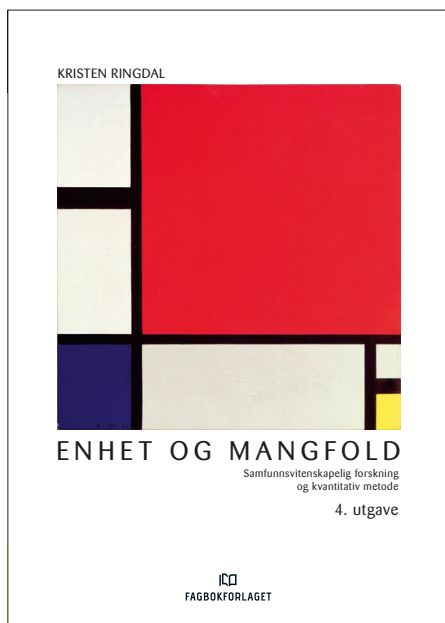


En kort innføring i SPSS

Vedlegg til Kristen Ringdal: *Enhet og mangfold*, 4. utgave



Det finnes flere håndbøker i SPSS som kan være nyttige. En av de beste på engelsk er Andy Fields *Discovering statistics using SPSS*.

Om SPSS

SPSS (Statistical Package for the Social Sciences) ble utviklet for stormaskiner i slutten av 1960-årene og ble sammen med SAS en av de mest benyttede programpakkene for statistisk dataanalyse. I begynnelsen måtte programmer legges inn på hullkort, leses inn i datamaskinen, og kjøres. Etter hvert kunne en legg inn programlinjer fra terminaler, men programmet var kommando-styrt helt fram til 1980-årene da PC-versjonen kom. Dette førte til en rask utvikling i retning av menystyring. SPSS har de beste menyene for databearbeiding og statistisk analyse. Dette er spesielt nyttig for nybegynnere. I 2009 ble SPSS kjøpt opp av IBM og er nå kjent som IBM SPSS Statistics.

Bruk av skrifttyper

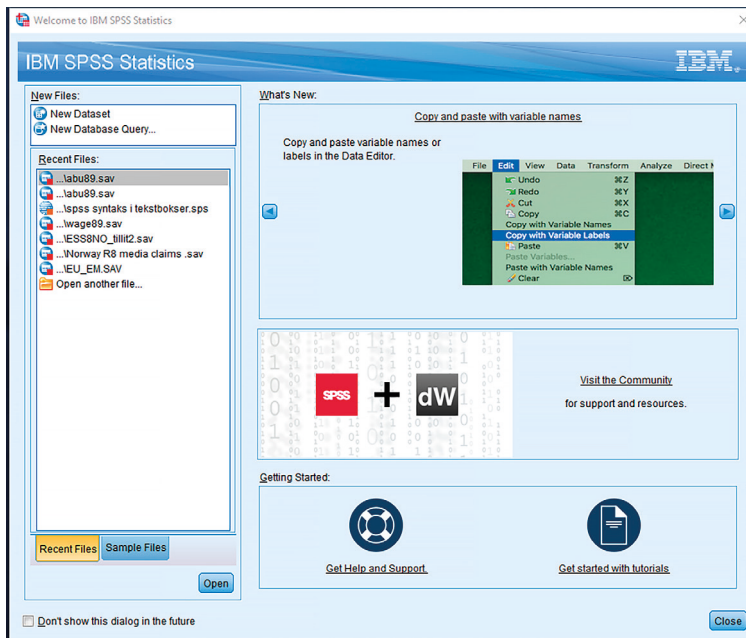
I tillegg til den vanlig teksten er kommandoer, filnavn og variabelnavn, samt utskrifter trykt i *Courier New*. Menyvalg og nødvendige trykk på tastaturet er gjengitt med **rød skrift**.

Åpningsvinduet og hovedvinduet i SPSS

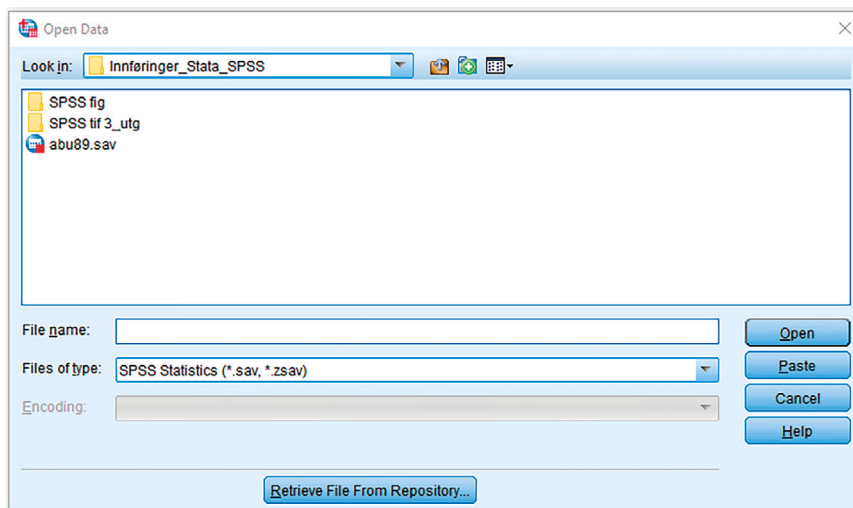
Åpningsvinduet i SPSS er gjengitt på figur 1. Hvis vi vil lage et nytt datasett, velger vi **New Datasett** fra feltet øverst til venstre. Dette vil åpne en tom datafil, der data kan legges (eller tastes) inn, som på figur 3a. Feltet til venstre viser de siste datasettene (datafilene) som har vært åpnet. Når en starter SPSS første gang vil dette feltet være tomt. Hvis vi ønsker å åpne en datafil som ikke er listet, velger vi **Open another file...**, og finner den fil vi ønsker å åpne. Det er også mulig å velge fanene **Sample Files** som vil vise en liste over datasett som følger installasjonene av SPSS.

Nederst i figur 1 finner vi et innrammet felt under **Getting Started**. Her kan en få hjelp eller starte opp med øvinger (tutorials).

La oss åpne `abu89.sav`. Dette forutsetter at datasettet er lastet ned fra bokas nettsted og lagret lokalt. Siden vi ikke har åpnet den tidligere, må vi velge **Open another file...**. Dette åpner «Open Data»-dialogboksen (figur 2) filen nå er funnet. Vi kan velge `abu89.sav` og trykke **Open** for å åpne filen.

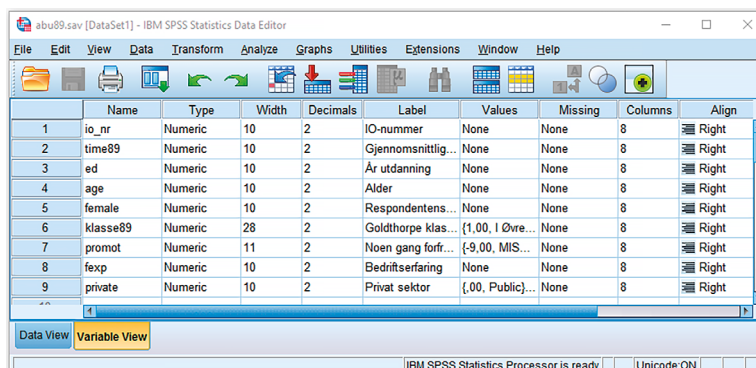
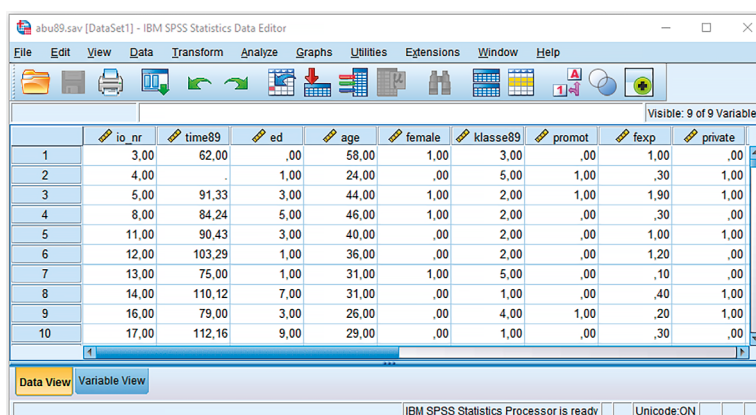
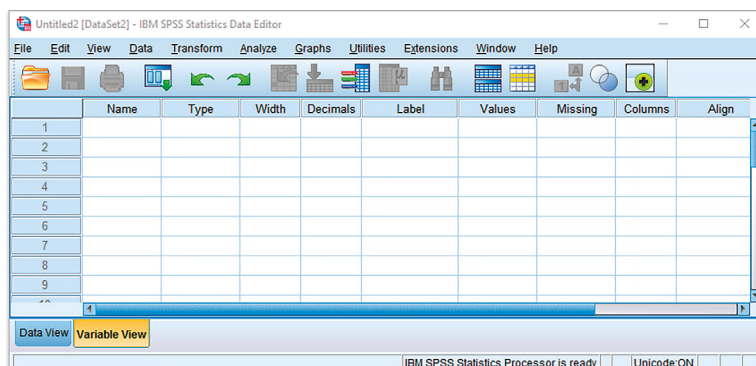


Figur 1 Åpningsvinduet i SPSS



Figur 2 «Open Data»-dialogboksen

Dette åpner hovedvinduet i SPSS som viser en del av datasettet som er åpnet med variablene i kolonnene og enhetene i linjene (figur 3 b). Verdiene som viser for hver enhet er tallkoder, normalt med to desimaler. Merk at den automatiske koden for manglende data er «.».



Figur 3 Hovedvinduet i SPSS

SPSS bygger på at vi kan gjøre det aller meste i tre vinduer: hovedvinduet, utskriftsvinduet («SPSS Viewer»), samt syntaksvinduet der kommandoer kan lagres og redigeres. Felles for de tre vinduene er menylinjen (**File, Edit, View, Data, Transform, Analyze, Graphs**, mm.), mens ikonlinjen, med snarveier til noen kommandoer, er ulike. Menyvalgene vil åpne dialogbokser der vi kan laste inn filer, omkode variabler, bestille tabeller og så videre. Menyene er utgangspunktet for det meste vi kan gjøre i SPSS.

La oss gå tilbake til figur 3b som viser utsnitt fra `abu89.sav`. Filnavnet vises i feltet øverst til venstre. Der nest følger meny- og ikonlinjene. Resten av vinduet utgjøres av SPSS sin dataeditor som har to visningsmoduser. Vinduet åpnes i **Data View** der analyseenheter utgjør linjene og variablene kolonnene i datamatriksen. Her kan dataene redigeres direkte på skjermen. Hvis vi trykker på fanen for **Variable View**, blir utvidet informasjon om alle variablene i datafilen vist med en linje for hver variabel (figur 3c). Her kan variabel- og verdinavn legges til eller forandres. Første kolonne gir variabelnavnene. Den andre viser hva slags type data som er lagret. Oftest vil variablene inneholde vanlige tall (Numeric), men det er flere andre muligheter der bokstaver (String) er den mest aktuelle. For numeriske variabler blir maksimalt antall posisjoner variabelen kan lagres i, samt antall desimaler angitt i de to neste kolonnene. Deretter følger kolonner for variabel- og verdietiketter (labels). Koder for manglende informasjon (Missing values) vises i den neste kolonnen. Bare timelønn har en slik kode definert (.). Dette er koden for «system missing» eller «sysmis». Denne koden kan suppleres med å angi en eller flere verdier som «missing». I kolonnen for **Missing** er det ingen andre koder som er definert som manglende data. Dette kan vi endre på ved å klikke på **none**, trykke på feltet til høyre i kolonnen og angi en eller flere kode som missing.

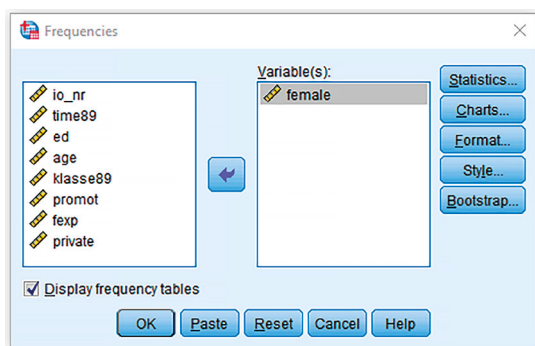
Hvis vi skal lage en ny datafil, er det første trinnet å definere variablene som skal fylles med data i **Variable View**. Det er ofttest tilstrekkelig å taste inn variabelnavnet, men også lengre variabeletiketter kan legges inn her, samt korte verdietiketter som beskriver verdiene på kategorivariabler. Når dette er gjort, velges fanene for **Data View** som nå er klar for inntasting av data.

Når vi åpner en datafil, åpnes også utskriftsvinduet, der resultatene vises. Ved å velge **Options** i «Edit»-menyen, kan vi endre formen på det som vises. La oss se på noen praktiske valg her. Velger vi **Edit → Options → General** er det under **Variable Lists** smart å merke av for visning av navn **Display names** og

File. Resultatet er at variabellister er basert på de korte variabelnavnene, samt at de listes i den rekkefølge de er lest inn eller laget. Det vi si at når vi lager en ny variabel, vil den være å finne i enden av filen. Det er også praktisk å hake av for **No Scientific Notation for small numbers in tables**. I «Viewer»-fanen haker vi av for **Display commands in the log**, som gjør at utførte kommandoer, blir lagt inn i utskriftsvinduet. Deretter går vi til fanen for **Output labels** og velger **Names and labels** og **Values and labels** under både **Output Labeling** og **Pivot Table Labeling**. Dette sikrer at tabellene i utskriftsvinduet er maksimalt lesbare.

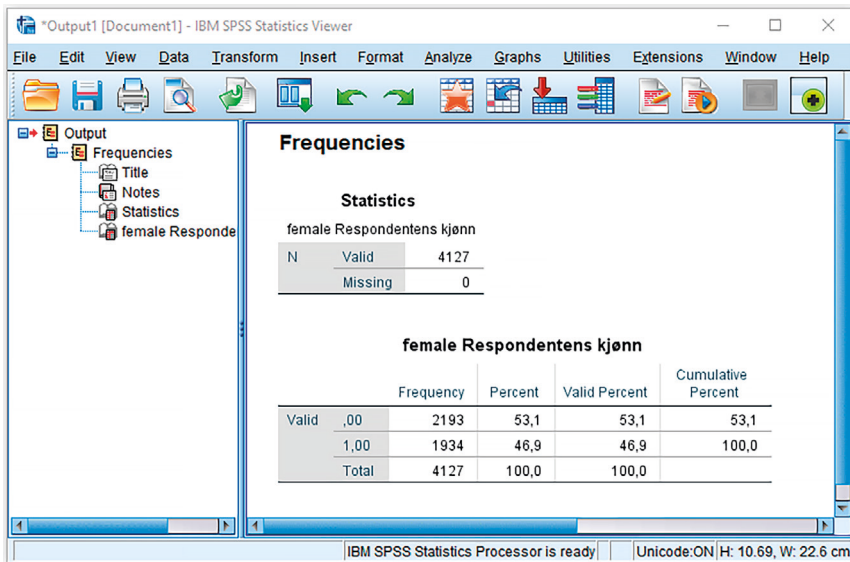
Å lage en enkel tabell i SPSS

I *Enhet og mangfold* finnes det oppskrifter på hvordan tabeller lages, hvordan korrelasjonsanalyser gjøres, og hvordan en enkel varians- eller regresjonsanalyse kan utføres i SPSS. Her skal vi bare gjennomgå hvordan enkle tabeller kan bestilles fra datafilen abu89. La oss se på fordelingen på kjønn. Menyvalgene **Analyze** → **Descriptive Statistics** → **Frequencies** åpner «Frequencies»-dialogboksen (figur 4).



Figur 4 «Frequencies»-dialogboksen

I variabellista til venstre velger vi `female` og fører denne variabelen over i feltet til høyre. Det er også mulig å bestille beskrivende statistikk og diagrammer ved å velge **Statistics** eller **Charts**. For vårt formål er det tilstrekkelig å trykke på **OK**, og tabellen med frekvensfordeling for kjønn blir sendt til utskriftsvinduet (figur 5).



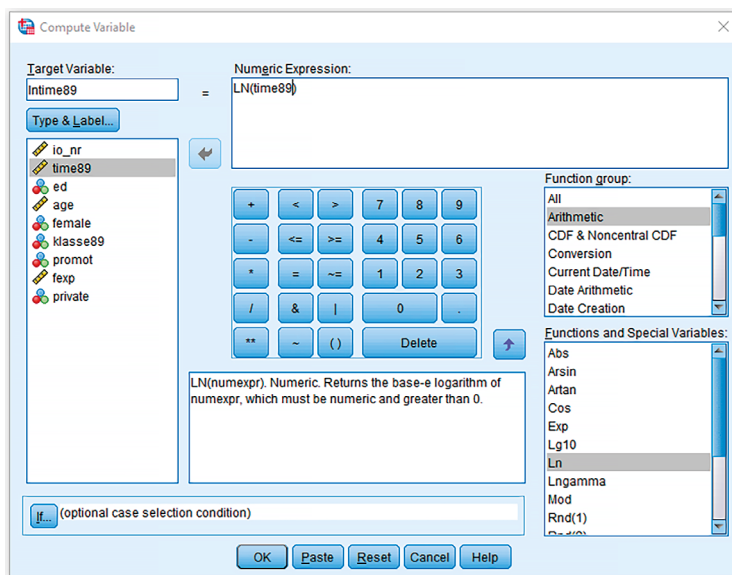
Figur 5 Utskriftsvinduet i SPSS

Feltet til venstre er en innholdsliste over elementene i utskriften til høyre. I lange utskrifter kan en lett flytte seg fra en tabell til en annen ved å trykke på en linje i innholdslista. Øverst i utskriften gjengis kommandoene som er utført, etterfulgt av en liten tabell som viser at vi har informasjon om kjønn til alle 4127 lønnstakerne i filen. Deretter følger tabellen som viser kjønnsfordelingen i utvalget. Det er ikke lagt til etiketter på variabelverdiene, men kvinner har koden 1 og menn koden 0. Den neste kolonnen viser antall menn og kvinner, så følger tre ulike prosentfordelinger. Utskriftene i SPSS har en tendens til å bli lange, og derfor er det sjelden hensiktsmessig å lagre filen i sin helhet. Vi kan lagre en redigert utskriftsfil eller vi kan overføre tabeller til andre dataprogrammer.

Normalt vil vi neppe publisere tabellen over kjønnsfordelingen, men den kan lett overføres til Word eller Excel. I disse programmene kan tabellen limes inn på to måter, ved å velge **Rediger → Lim inn**, eller **Rediger → Lim inn utvalg → Bilde (Windows Metafile)**. Den første måten er mest praktisk hvis en vil redigere tabellen. Den andre måten lagrer tabellen som et objekt som ikke kan redigeres, men som beholder formen fra SPSS-utskriften.

Kommandoer for redigering av variabler

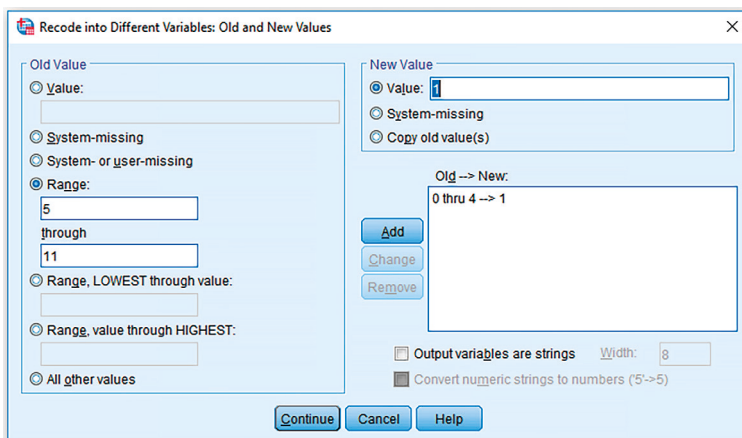
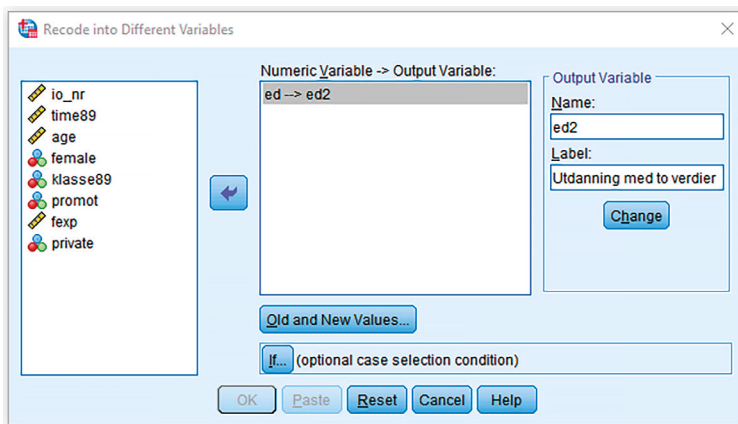
Her skal vi se hvordan vi kan endre på variablene i en datafil. De to mest interessante valgene i «Transform»-menyen er **Compute** og **Recode**. Den første kommandoen endrer eller lager nye variabler ved hjelp av matematiske operasjoner. La oss se hvordan vi kan lage ny variabel `ln_time89`, som den naturlige logaritmen av `time89`, timelønn i 1989. Figur 6 viser hvordan «Compute»-dialogboksen må fylles ut for å gjøre dette. Det nye variabelnavnet skrives inn i feltet for **Target Variable**. Etter likhetstegnet må vi først hente inn funksjonen for den naturlige logaritmen. Ved å velge aritmetiske funksjoner dukker den opp midt i feltet nederst til høyre. Ved å dobbeltklikke på funksjonen overføres den til operasjonsfeltet merket **Numeric Expression**. Etter funksjonsnavnet står det en parentes med et spørsmålstegn. Her overføres variabelnavnet ved at vi velger `time89` fra variabellista og trykker på knappen som fører det inn i funksjonen. Dermed er vinduet som vist på figur 6. Ved å trykke på **OK** blir den nye variabelen laget.



Figur 6 «Compute Variable»-dialogboksen i SPSS

Dermed gjenstår det å vise hvordan en kan omkode en eksisterende variabel i datafilen. Variabelen `ed` er utdanning målt i år ut over obligatorisk utdanning.

Verdiene varierer fra 0 til 11. Dette kan du bekrefte ved å lage en frekvenstabell på samme måte som for kjønn. La oss lage en ny utdanningsvariabel, `ed2`, som bare skiller mellom lav (0–4 år) og høy utdanning (5–11 år). Vi åpner den riktige dialogboksen ved å velge **Transform → Recode → Into Different Variables....** Først finner vi `ed` i variabellista og fører den over i feltet merket **Numeric Variable → Output Variable**. Deretter fyller vi inn `ed2` som **Output Variable** og skriver inn en variabeletikett (label) i feltet under. Merk at vi må trykke på **Change** slik at spørsmålsteget i «`ed`→?» erstattes med det nye variabelnavnet slik som vist øverst på figur 7a.



Figur 7 «Recode into Different Variables»-dialogboksen i SPSS

Dermed er det nye variabelnavnet definert, men det gjenstår å definere forholdet mellom kodene i den gamle variabelen `ed`, og den nye variabelen `ed2`. Ved å trykke på **Old and New Values** åpnes dialogboksen på figur 7b der omkodningen kan spesifiseres. Bruk av «Range»-feltet er den enkleste måten å velge hvilke verdier på den gamle variabelen som skal omkodes til 0 på den nye.

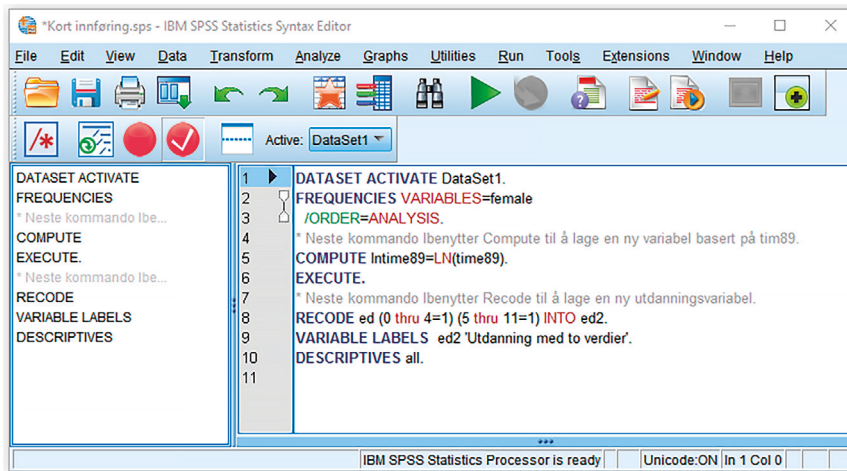
Etter å ha valgt «0 through 4» må en gå til «New Value»-feltet og skrive inn verdien 0. For å fullføre spesifiseringen av den nye verdien, må en trykke på **Add** og den første delen av kommandoen kommer til syne i «Old → New»-feltet. Figuren viser også at de gamle verdiene (5–11) som gir den nye verdien 1 er fylt inn. Det gjenstår bare å trykke på **Add**. Et trykk på **Continue** bringer oss tilbake til den første dialogboksen og et trykk på **OK** utfører kommandoen og lager den nye variabelen `ed2`.

Syntaksfiler i SPSS

Til nå har vi sett datafiler og utskriftsfiler i SPSS. Det finnes også syntaksfiler som lagrer kommandoer. Dette er praktisk hvis vi gjør store endringer i datafilen ved å omkode og konstruere nye variabler. Det er ikke uvanlig at slike kommandoer til sammen utgjør flere hundre linjer. Ved å lagre kommandoene blir de dokumentert for ettertiden, og en kan også utføre dem på nytt hvis det skulle bli nødvendig.

Det er svært enkelt å lagre kommandoene i en syntaksfil. Se tilbake på figur 4. Ett av valgene til høyre er **Paste**. Trykker vi på denne knappen lagres kommandoen for å lage tabellen for kjønn. Dette gjelder gjennomgående for alle analyse- og redigeringskommandoer i SPSS. På figur 8 vises de redigerings- og analysekommandoene som er benyttet i denne korte innføringen. Øverst angir det aktive datasettet. I vårt tilfelle det første datasettet som er åpnet (Dataset1). Deretter følger kommandoen for å lage en tabell med frekvensfordelingen for kjønn. Merk at alle kommandoer i SPSS må avsluttes med et punktum. Den siste delen av kommandoen er overflødig. Det er tilstrekkelig å sette punktum etter «female». Deretter er det lagt inn kommentarer før de to siste kommandoene. Kommentarløser starter med stjerne «*» og avsluttes med punktum. «Compute»-kommandoen følges av **EXECUTE** på en egen linje. Denne angir at kommandoen ovenfor skal utføres.

En alternativ måte å få en redigeringskommando utført på er å bestille en tabell der resultatvariablene inngår etter redigeringskommandoen slik som for «Recode»-kommandoen på figur 8.

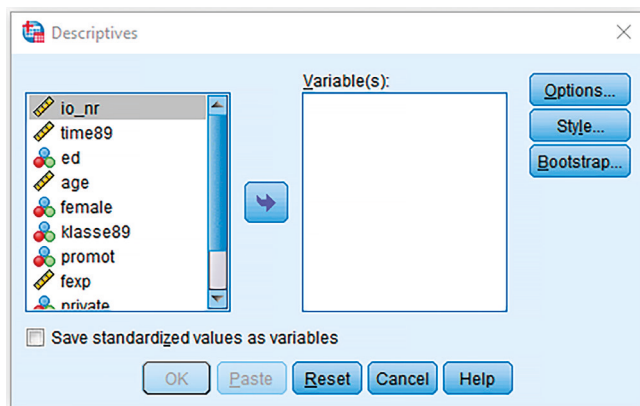


Figur 8 Syntaksvinduet i SPSS

Kommandoer i en syntaksfil kan utføres fra «Run»-menyen (hurtigvalg: **Ctrl-R**), eller ved å trykke på den grønne trekanten i ikonlinjen. Da utføres enten den kommandoen pekeren står i, eller de linjene som er valgt i syntaksfilen.

Beskrivende statistikk

Vi skal avslutte med å se på hvordan vi kan lage en tabell med alle eller utvalgte variabler i en datafil. Syntaksfilen avsluttes med kommandoen: `DESCRIPTIVES all`, som gir en tabell med beskrivende statistikk for alle variablene i filen. Det mest vanlige er å benytte en variabelliste i stedet for «all». Vi kan også gjøre dette ved bruk av menyene. **Analyze → Descriptive Statistics → Descriptives** åpner «Descriptives-dialogboksen på figur 9. Her må vi velge de variablene som skal med i tabellen. For å få tabellen på figur 10 må en velge alle variabelene og føre dem over til «Variable(s)»-feltet og avslutte med **OK**.



Figur 9 «Descriptives»-dialogboksen

IBM SPSS Statistics Processor is ready | Unicode:ON | H: 0.69, W: 16.54 cm

Descriptive Statistics					
	N	Minimum	Maximum	Mean	Std. Deviation
io_nr IO-nummer	4127	3,00	6258,00	3104,6186	1803,55460
time89 Gjennomsnittlig timelønn 1989	3759	25,00	343,75	90,1495	30,31473
ed År utdanning	4127	,00	11,00	2,6896	2,55703
age Alder	4127	16,00	74,00	39,6508	12,35712
female Respondentens kjønn	4127	,00	1,00	,4686	,49907
klasse89 Goldthorpe klasse 1989	4042	1,00	5,00	3,0210	1,18633
promot Noen gang forfremmet	4127	,00	1,00	,3778	,48488
fexp Bedriftserfaring	4127	,00	4,90	,9451	,90658
private Privat sektor	4127	,00	1,00	,6118	,48739
Intime89	3759	3,22	5,84	4,4526	,30762
ed2 Utdanning med to verdier	4127	1,00	1,00	1,0000	,00000
Valid N (listwise)	3680				

Figur 10. Utskrift med beskrivende statistikk