



# Stata intro

---

*Grundlæggende introduktion til Stata*

Af Eva Sandberg, OPEN

---

## Indholdsfortegnelse

Formål.....	3
2. GRUNDLÆGGENDE GENNEMGANG .....	4
1.1 Sådan ser Stata ud (vinduer og menuer).....	5
Stata vindue.....	5
Dialogboks.....	7
1.2 Hjælpefunktion.....	9
1.2 Filplacering.....	10
1.4 Filtyper.....	12
1.5 Hent, gem og kig på data.....	13
Hent data.....	13
Gem data.....	13
Kig på data.....	13
1.6 Dokumentation.....	15
Do-filer.....	15
Log-filer.....	17
1.7 Import og eksport.....	18
Import.....	18
Eksport.....	18
1.8 Import af data fra RedCap.....	19
1.9 Overblik over data.....	20
Typisk kommandoopbygning: .....	20
Stata kommandoer.....	20
1) kommandoer som rapporterer information om data:.....	20
2) kommandoer som ændrer data (ændre kun data i memory):.....	21
2. DATAMANAGEMENT .....	24
2.1 Statas datakoncept.....	25
2.2 Variabeltyper.....	26
2.3 Stata syntakser.....	27
2.4 Variabel manipulation.....	28
Generate.....	28
Egen.....	28
Replace.....	28

Rename.....	29
Recode.....	29
Encode.....	29
Decode.....	30
Drop.....	30
Keep.....	31
2.5 Variabel konvertering.....	32
Tostring.....	32
Destring.....	32
2.6 Dato konvertering.....	33
Konvertering af datoer fra tekst til numeriske.....	33
2.7 Labels i datasæt og variabler.....	34
2.8 Kombinér flere datasæt.....	35
Hvad betyder <code>_merge</code> variabelen?.....	35
A + B ( <code>_merge 1</code> ).....	36
A + B ( <code>_merge 2</code> ).....	37
A + B ( <code>_merge 1, 2 og 3</code> ).....	38
2.9 Vend datasæt.....	39
Fra long til wide.....	39
Fra wide til long.....	40
2.10 Ekstra tips.....	41
Hjælpevariabler.....	41
Tæl alle unikke variabler.....	41
Gruppér alle unikke variabler.....	41
Trim.....	41
Store bogstaver.....	41
Små bogstaver.....	41
2.11 Oversigt over stata kommandoer.....	42
Syntakser mm.....	42
Kommandoer.....	42
Samle datasæt.....	43
Long eller wide format af data.....	43
Datoer.....	44

Find første variable i et datasæt.....	44
Tæl antal rækker inden for hver unik variable .....	44
Descriptiv statistik .....	44
Basis statistik .....	44
Kohorter / survival analyser .....	45
Case Control / Cross Sectional analyser .....	45

## Formål

Formålet med denne guide er at lave en oversigt over de grundlæggende funktioner, der findes i Stata og introducere de mest relevante og simple teknikker til at lave indledende datamanagement med.

Principperne i denne guide er direkte sammenlignelige med dem, der undervises i på Stata Intro kurset i OPEN. Vi gør derfor også opmærksom på, at denne guide langt fra er udtømmende i forhold til de funktioner Stata ellers tilbyder.

Guiden er tiltænkt som et hjælpeværktøj og opslagsværk for OPENs kursister, og er primært udarbejdet til nybegyndere og dem, som ikke er vant til at arbejde med Stata.

De eksempler som arbejdes med i denne guide stammer dels fra Stata's egne eksempel datasæt og dels fra de datasæt, som anvendes under kurset i OPEN, men principperne vil kunne overføres til egne datasæt.

Hvis der er et ønske om at øve sig på Stata's egne datasæt, findes disse i menuen under Filer File > Example datasets > Examples datasets installed with Stata.

Vi håber, I får glæde af guiden og hvis I finder fejl eller mangler, er I velkommen til at kontakte Eva Sandberg, [eva.sandberg@rsyd.dk](mailto:eva.sandberg@rsyd.dk)

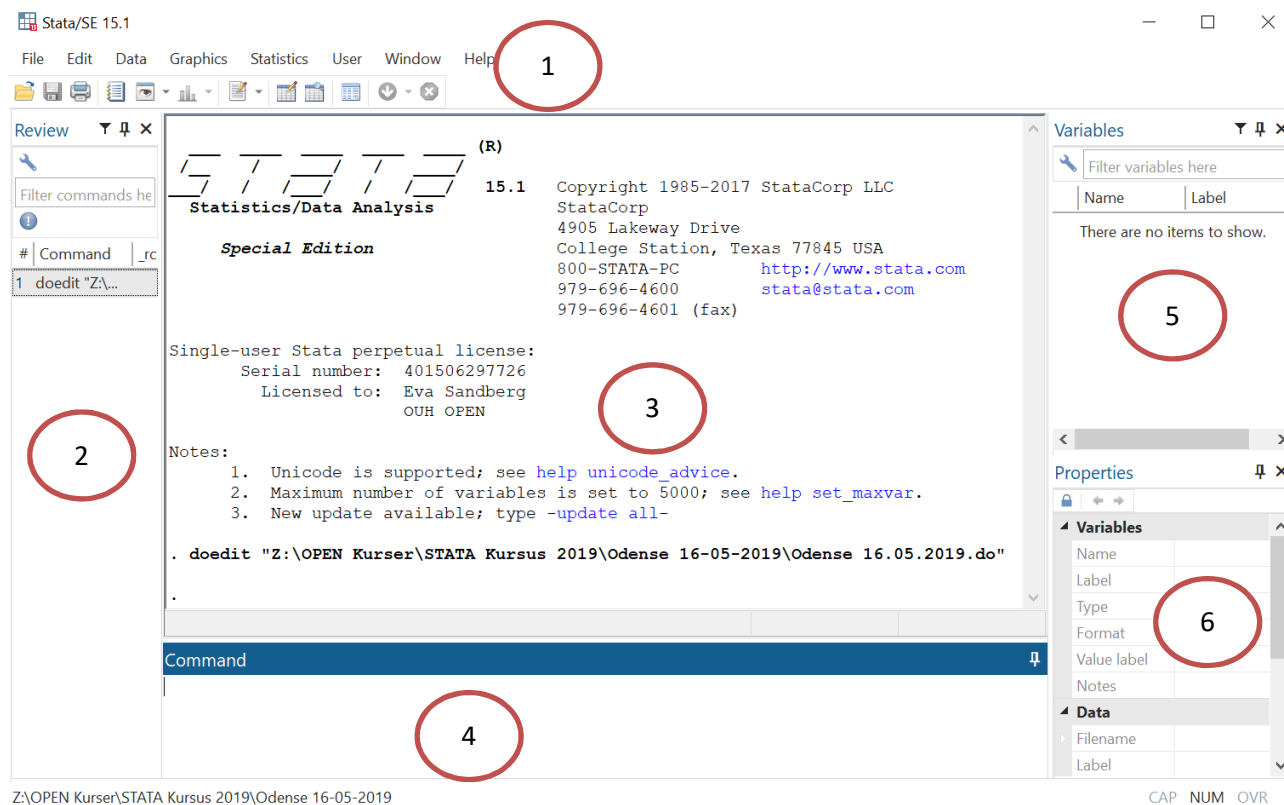
Rigtig god fornøjelse!

## **2. GRUNDLÆGGENDE GENNEMGANG**






## 1.1 Sådan ser Stata ud (vinduer og menuer)

### Stata vindue

I dette kapitel beskrives indhold og funktioner i de forskellige vinduer, der vises, når Stata åbnes. Det første billede er et skærmdump af Stata, hvor der er angivet seks cirkler med numre, som hver henviser til det samme nummer i forklaringen. Efterfølgende vises et skærmdump af en dialogboks, hvor samme forklaringsprincip er gældende.



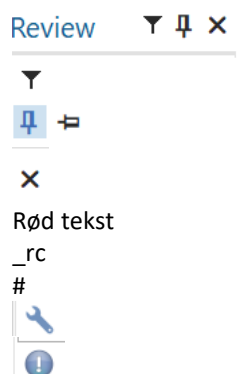
### Forklaring:

1	<b>Menuer og ikoner:</b>	<b>Stata menuer:</b> Filer : indeholder funktioner til selve arbejdet med filer Edit : indeholder standard Microsoft funktioner som også bruges i Stata Graphics : indeholder de fleste grafiske funktioner Statistics : indeholder de fleste statistikfunktioner User : viser hvilke opgaver som er aktive Window : indstilling af Stata vinduerne Help : Stata hjælp og informationer <b>Stata ikoner:</b>  : short-cut åbn fil  : short-cut gem fil  : short-cut print fil  : short-cut log  : short-cut til redigering af grafer
---	--------------------------	--



- : short-cut oversigt åbne do-filer
- : short-cut browser med mulighed for redigering
- : short-cut browser
- : short-cut variable manager
- : short-cut flere resultater
- : short-cut break af kørsel

2 **Review:** Viser hvilke kommandoer, der er kørt og hvilken status, det kørte er i



- : øverste menu
- : aktivere/deaktivere værktøj og søgefelt
- : autohide/autoview af reviewfunktionen
- : deaktiver reviewfunktionen (kan hentes under windows)
- Rød tekst : fejl i kommando
- \_rc : fejlkoder (tryk på koden og få hjælp)
- # : viser numerisk rækkefølge på kommandoerne
- Wrench icon : hjælpeværktøjer (sæt default, søg på ord, bogstaver etc.)
- ! icon : fjerner koder med fejl

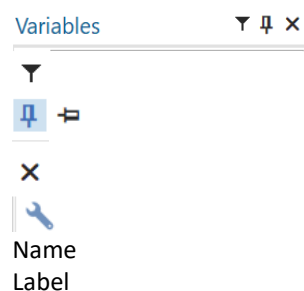
3 **Executer:** Viser udfald af kørte kommandoer, log- eller do-filer

- Højreklik på mus : for at få kommandooversigt
- Ctrl F : åbner et søgevindue til søgning i executer
- Kommandoen `cls` : rydder alt indhold i vinduet

4 **Command:** Her kan kommandoer indtastes - HUSK STATA ER CASE SENSITIV

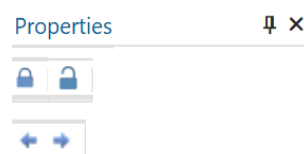
- derfor skal alle kommandoer skrives med små bogstaver
- Knappen tab (på tastaturet) : åbner variabelliste
- Knappen PageUP : finder tidligere kommandoer i Reviewlisten
- Knappen PageDown : finder efterfølgende kommandoer i Reviewlisten

5 **Variables:** Viser de variabler, som findes i det datasæt, som er indlæst



- : aktivere/deaktivere værktøj og søgefelt
- : autohide/autoview af reviewfunktionen
- : deaktiver reviewfunktionen (kan hentes under windows)
- Wrench icon : hjælpeværktøje (sæt default, søg på ord, bogstaver etc.)
- Name : viser korte navn på variabel – kun 11 karakter er tilladt
- Label : viser label/beskrivelse til variabel

6 **Properties:** **Variables:** Viser variablenes egenskaber



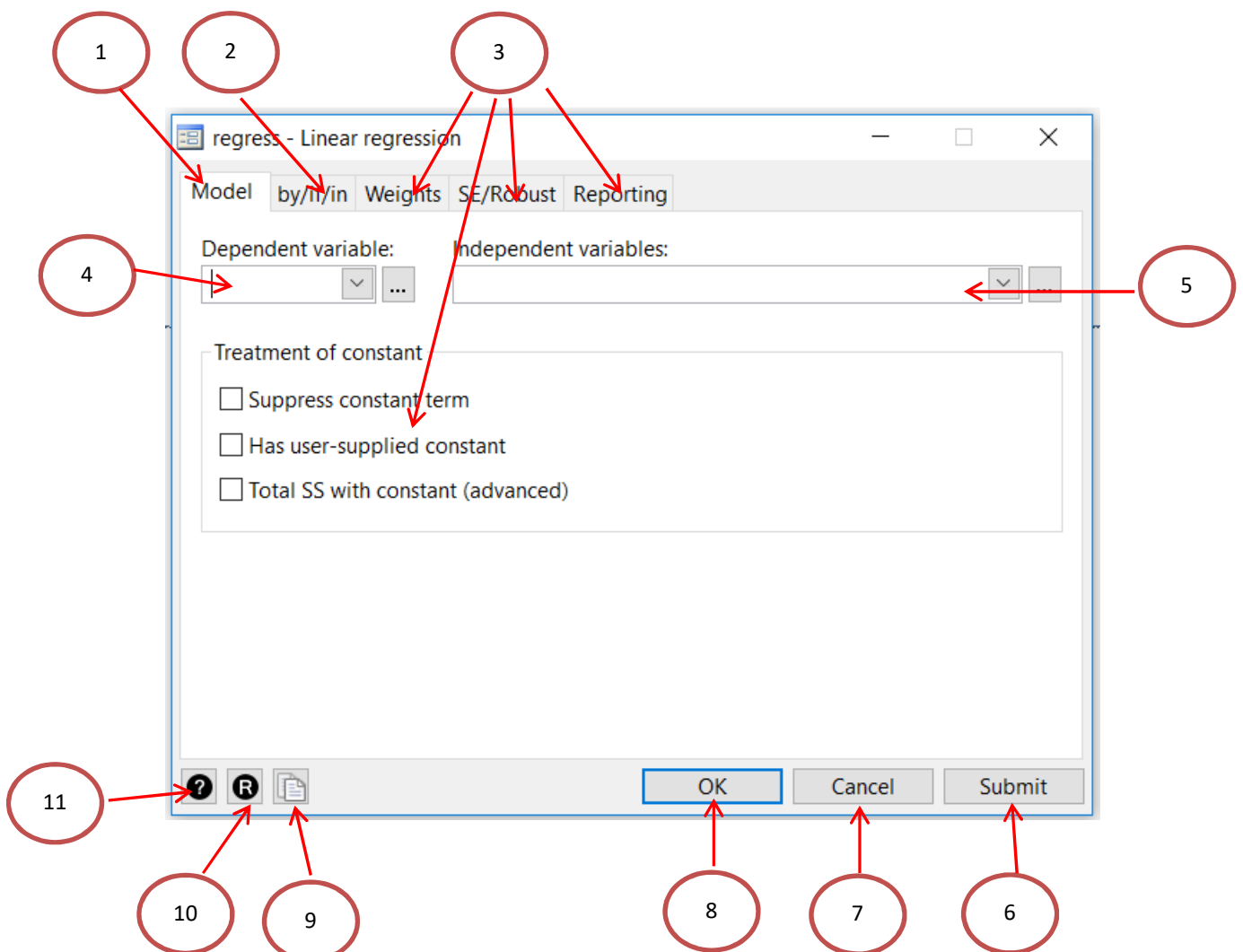
- : øverste menu
- : aktivering/deaktivering af redigering af variabler – hvis hængelås er åben, kan der redigeres direkte i boksen
- : hop frem og tilbage i variabler

## Stata intro

Name	: viser korte navn på variabel – kun 11 karakter er tilladt
Label	: viser label/beskrivelse til variabel
Type	: format type (numeriske el string)
Format	: antal format karakter
Variable label	: tildeling af variable label, hvis de er oprettet
Notes	: Tilføj en note/forklaring til variabel
<b>Data:</b>	Viser datasættets egenskaber
Filename	: Navn på datasæt
Label	: Label på datasæt
Notes	: Noter i datasættet
Variables	: antal variabler i datasæt (kolonner)
Observations	: antal observationer i datasæt (rækker)
Size	: størrelsen på datasæt
Memory	: hukommelse til rådighed
Sorteby	: variabel data er sorteret efter

## Dialogbokse




De fleste Stata kommandoer har tilhørende dialogbokse til at specificere en kommando. Det er efterfølgende muligt at se kommandoen i Result-vinduet, som kan kopieres over i en do-fil. De fleste dialogbokse har multiple funktioner, men er i princippet opbygget på samme måde.






## Stata intro

---

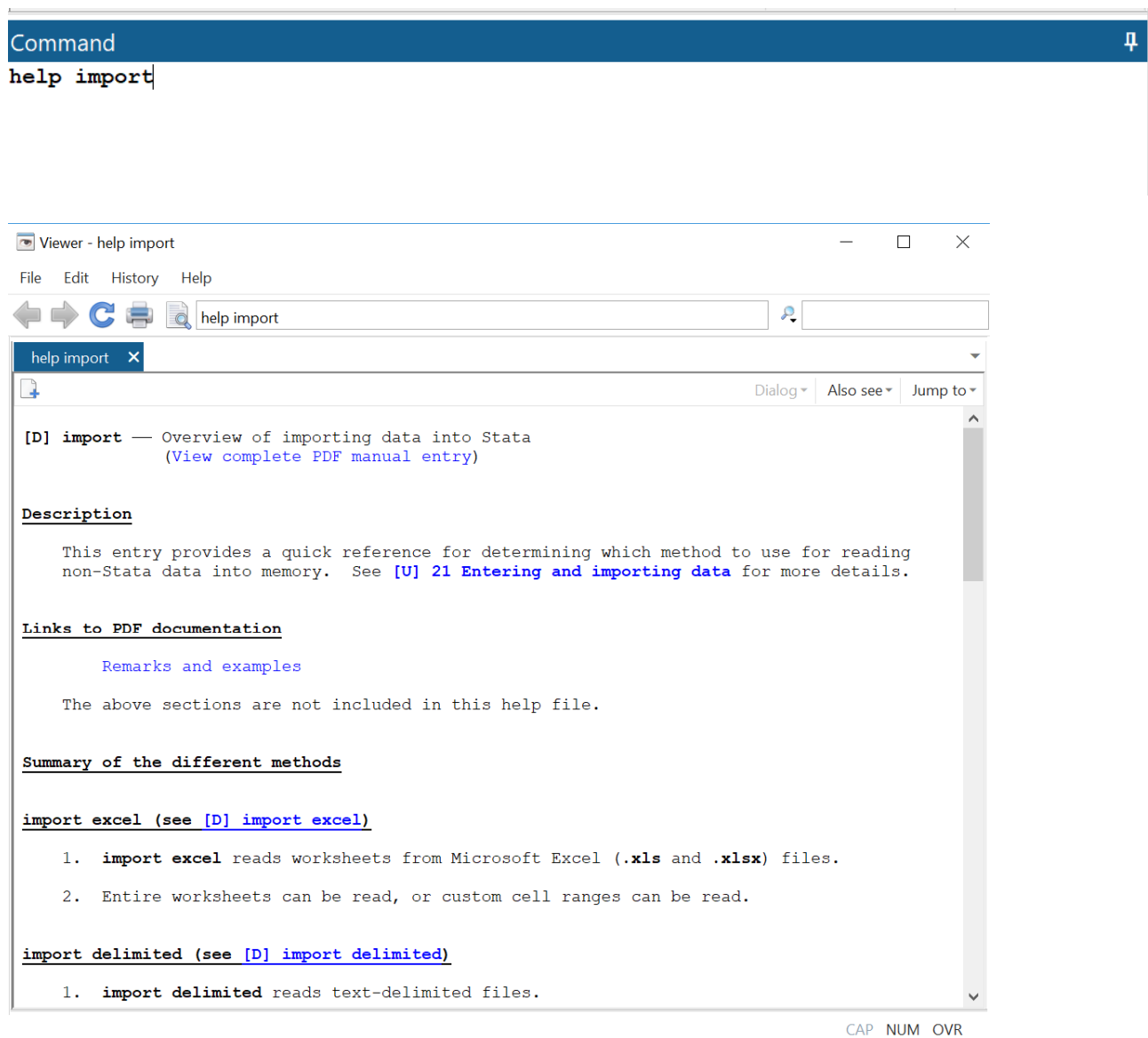
1	Main/Model	: optionsvælger
2	By/if/in	: valg af optioner for variabel
3	Options	: forskellige optioner afhængig af statistisk model
4	Dependent	: y = afhængige variabel = outcome
5	Independent	: x = uafhængige variabel = exposure
6	Submit	: udfører kommandoen, men holder dialogboksen åben
7	Cancel	: lukker dialogboksen uden at udføre kommandoen
8	OK	: udfører kommandoen og lukker dialogboksen
9		: kopierer kommandoen til udklipsholderen
10		: resætter dialogboksen til default
11		: åbner hjælpefunktionen til den kommando, som er valgt

## 1.2 Hjælpefunktion

STATA har en meget udførlig hjælpefunktion. I hjælpefunktionen findes mere udførlige forklaringer: pdf, eksempler, henvisning til menuen, metodiske forklaringer, dokumentation, referencer i litteraturen mm.

Du kan få hjælp enten ved at bruge  funktionen, hvor det er muligt. Alternativt tast help + det du ønsker hjælp til i kommandofeltet, eks.:

help import



The image shows two screenshots from the Stata software interface. The top screenshot is the Command window, where the text 'help import' is entered. The bottom screenshot is the Viewer window, which displays the help text for the 'import' command. The help text includes a description, links to PDF documentation, and a summary of different methods like 'import excel' and 'import delimited'.

Command  
help import

Viewer - help import

File Edit History Help

help import

help import

[D] **import** — Overview of importing data into Stata  
(View complete PDF manual entry)

**Description**

This entry provides a quick reference for determining which method to use for reading non-Stata data into memory. See [U] 21 [Entering and importing data](#) for more details.

**Links to PDF documentation**

[Remarks and examples](#)

The above sections are not included in this help file.

**Summary of the different methods**

[import excel](#) (see [D] [import excel](#))

1. **import excel** reads worksheets from Microsoft Excel (.xls and .xlsx) files.
2. Entire worksheets can be read, or custom cell ranges can be read.

[import delimited](#) (see [D] [import delimited](#))

1. **import delimited** reads text-delimited files.

CAP NUM OVR

### 1.2 Filplacering

Når du starter med at arbejde i Stata, er det en god idé at indstille Stata til at hente og gemme de filer, du skal arbejde med automatisk. Det vil senere spare dig for en masse arbejde og besvær med at finde rundt, hvis du arbejder med mange filer.

Med kommandoen `pwd` viser Stata dig, hvor Stata automatisk gemmer dine filer. Måske ikke lige der, du ønsker dine filer placeret.

```
pwd
```

stata executer viser:

```
. pwd  
Z:\OPEN Kurser\STATA Kursus 2019\Odense 16-05-2019
```

Dvs. Stata gemmer mine filer på:

- vores Z drev
- i mappen OPEN Kurser
- i undermappen STATA Kursus 2019
- filen hedder "Odense 16-5-2019"

MEN det er jo ikke sikkert, du synes, at din fil skal ligge der eller, at din fil skal hedde "Odense 16-5-2019"

1. Først finder du det drev og den placering, hvor din fil skal gemmes

Ved at bruge kommandoen `cd`, kan du oprette en ny sti

```
cd C:\Users\aga1ba\Desktop\
```

Nu er der oprettet en ny sti på skrivebordet, hvor Stata kan gemme filerne.

Hvis `pwd` køres igen viser Stata:

```
. pwd  
C:\Users\aga1ba\Desktop
```

## Stata intro

---

2. Derefter laver du den mappe, hvor du vil gemme din fil

```
Mkdir Stata_Kursus
```

– BEMÆRK, at jeg her har brugt \_ i stedet for mellemrum. Det gør jeg fordi, hvis jeg har mellemrum i et filnavn, så skal kommandoen omkredses af " ", ellers melder Stata fejl!

3. Derefter kører du en cd igen for at få Stata til at pege ned i mappen

```
cd C:\Users\aga1ba\Desktop\Stata_kursus
```

Hvis du kører en pwd igen, så skulle den rigtige sti være klar

```
. pwd  
C:\Users\aga1ba\Desktop\Stata_kursus
```

### 1.4 Filtyper

I det foregående kapitel, har du lært at finde ud af, hvordan du laver mapper på dit drev og hvordan du får Stata til at lægge de filer, du arbejder med det rigtige sted. I dette kapitel får du en kort oversigt over de forskellige filtyper og i de næste to kapitler lærer du om, hvordan du arbejder med dem.

Stata arbejder med tre forskellige filtyper og de genkendes på baggrund af deres ekstension, dvs. det de ender på.

Stata filtyper:

- `.dta` = stata datasæt – Filer, som indeholder det data, der arbejdes med.
- `.do` = stata do-filer – Filer, som indeholder de kommandoer og tekster, man selv skriver til at køre sine datamanagementprogrammer og analyser.
- `.log` = stata logfiler – Filer, som indeholder log på de kørte do-filer. Logfiler inkluderer også information fra executer og data og tid.

Det er en god idé at oprette en mappestruktur med separate mapper til datasæt, do-filer og log-filer. Derved får man et system, hvor man nemt kan finde sine filer igen.

Derudover introduceres også tre kommandoer, som du skal bruge, når du åbner eller gemmer filer.

- `, clear` = bruges når du åbner en fil og allerede har ét datasæt åbnet i Stata
- `, replace` = bruges når du gemmer en fil og vil erstatte det, du har lavet
- `, append` = bruges når du vil gemme en nye version af et allerede eksisterende datasæt

## Stata intro

---

### 1.5 Hent, gem og kig på data


Stata datasæt er det data, du arbejder med, dvs. det data, som i sidste ende skal danne baggrund for dine analyser og dit resultat. I det efterfølgende skal du lære, hvordan du henter, gemmer og kigger på Stata datasættet.

#### Hent data

Når du skal hente et datasæt i Stata, kan du gøre det på tre måder:

1. Skrive en manuel kommando i Command vinduet

```
use t_person, clear
```

2. Hente din fil ved at bruge Filer/Open i menuen eller ikonet 

3. Køre en do-fil, hvor kommandoen er inkluderet


**Bemærk! Efter use kommandoen skrives "clear". Det er fordi, at Stata kun kan arbejde med ét datasæt åbent ad gangen, så derfor sletter kommandoen clear alt foregående.**

#### Gem data

Når du skal gemme en fil i Stata, kan du gøre det på tre måder:

1. Skrive en manuel kommando i Command vinduet

```
save t_person, replace
```

2. Gemme din fil ved at bruge Filer/Save eller Filer/Save as i menuen eller ikonet 

3. Køre en do-fil, hvor kommandoen er inkluderet

**Bemærk! Efter save kommandoen skrives "replace". Det er fordi, at hvis du allerede har gemt en version, så kan Stata ikke overskrive uden denne kommando.**

#### Kig på data


Når du skal kigge på data, kan det gøres på tre forskellige måder:


1. Ved at bruge kommandoen br eller browser

```
br
```

## Stata intro

---

2. Ved at bruge ikonet  - når du åbner denne version, kan du KUN kigge i data

3. Ved at bruge ikonet  - når du åbner denne version, kan du RETTE i data

## 1.6 Dokumentation

### Do-filer

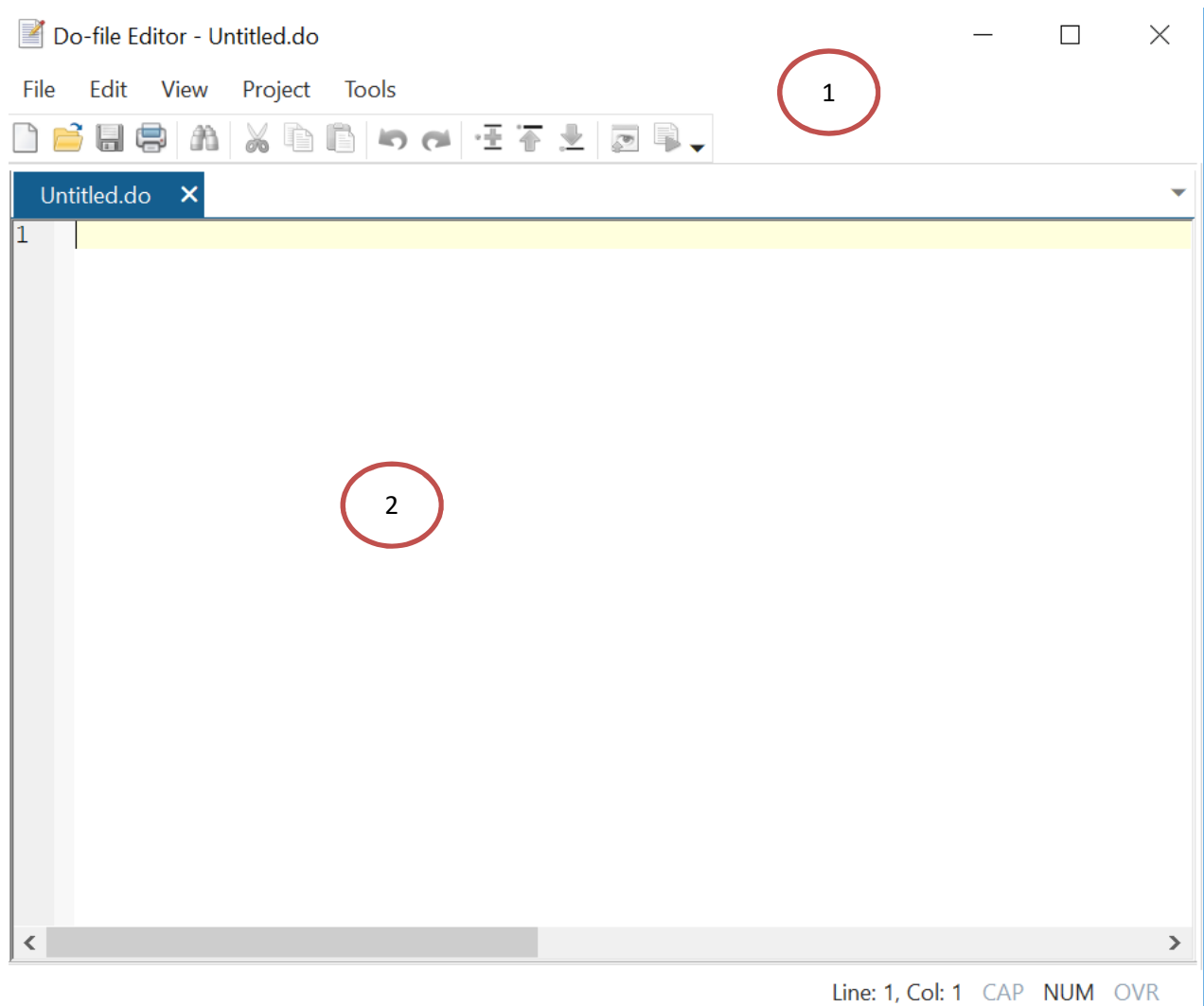
Do-filer er et meget vigtigt værktøj, når man arbejder med Stata. Det bruges til at samle, dokumentere, reproducere alle de tanker, koder og resultater, man har under sit arbejde.

En ny do-fil kan åbnes på flere forskellige måder:
















1. Åben via menu Window/Do-filer editor/ New Do-filer editor eller Ctrl+9
2. Hvis du står på en kommando i Review og højreklikker på musen, vælg do selected
3. Kopier din kommando direkte fra Command eller Execute

En eksisterende do-fil kan åbnes på to måder:

1. Åbn via menu Window/Do-filer editor/ Do-filer editor + navnet\_på\_din\_fil
2. Åbnes via din mappe





1	<b>Menuer og ikoner:</b>	<b>Do-fil menuer:</b>	
		File	: indeholder funktioner til selve arbejdet med filer
		Edit	: indeholder standard Microsoft funktioner, som også bruges i Stata
		View	: indeholder funktioner som find og zoom
		Project	: indeholder funktioner til projektet
		Tools	: indeholder funktioner til at udfører kommandoer
		<b>Do-fil ikoner:</b>	
			: short-cut ny do-fil
			: short-cut åben fil
			: short-cut gem fil
			: short-cut print fil
			: short-cut til at finde ord/tal
			: short-cut til at klippe Ctrl+X
			: short-cut til at kopier Ctrl+C
			: short-cut til at indsætte Ctrl+V
			: short-cut til undo
			: short-cut til redo
	: short-cut indsæt bogmærke		
	: short-cut find tidligere bogmærke		
	: short-cut find efterfølgende bogmærke		
	: short-cut til at finde filer i viewer		
	: short-cut til eksekvering af kommando Ctrl+D		

2	<b>Tekstområde:</b>
	Her skrives alt tekst og kommandoer.
	TIPS:
	1. Brugen af * før en tekst betyder, at Stata opfatter kommandoen som tekst og den køres ikke
	2. Brugen af /* omkring en tekst */ betyder, at Stata opfatter kommandoen som tekst og den køres ikke
	3. Brugen af 2 // efter en kommando betyder, at Stata kører kommandoen men ignorere teksten
	4. Brugen af 3 /// i en kommando betyder, at Stata ignorerer, at man skifter linje

En do-fil kan køres på flere måder:


- 1) Åbn do-filen og markér det, som skal køres og tryk Ctrl+D
- 2) brug do kommandoen i Command

```
do mindofil.do
```

### Log-filer

En log fil er den samlede dokumentation over dit arbejde og indeholder alt indhold fra din do-fil plus alle kørte tabeller og grafer. En log skal altid både åbnes og lukkes for at virke.

En log åbnes på forskellige måder:

1. Enten ved at bruge ikonet i menuen 
2. Eller ved at skrive det ind i do-filen

log using "min\_log", replace

**Bemærk! Efter kommandoen , replace. Kommandoen erstatter den eksisterende fil. Du kan også bruge kommandoen , append, så lægger Stata filerne under hinanden.**



```
Do-file Editor - Untitled.do*
File Edit View Project Tools
[Icons: Open, Save, Print, Find, Copy, Paste, Undo, Redo, Home, Up, Down, Run, Stop, Help]
Untitled.do* x
1
2 log using "minlog.log"
3
4
5 her kommer så alle mine kommandoer
6
7
8 log close
9
10
```

ATTENTION: Sæt først loggen ind, når du er helt færdig med din do-fil.

Hvis du oplever, at din log ikke kører, så er det højst sandsynligt fordi loggen allerede er åben – prøv at taste: log close.

Vær desuden opmærksom på, om do-filen indeholder personhenførbare data, hvis du arbejder på SDS eller DST.

### 1.7 Import og eksport

Det er muligt at importere og eksportere flere forskellige filtyper til og fra Stata. Excel og cvs-filer (tekstfiler) er de mest anvendte.

#### Import

Det er muligt at importere på flere forskellige måder:

1. Brug menuen File\Import
  - vælg Excel spreadsheet og følg menuen, hvis det er Excel, du skal importere
  - vælg Text data og følg menuen, hvis det er en tekstfil, du skal importere
2. Skriv kommandoen direkte i Command

```
import excel "t_person", sheet("Person") firstrow clear
```

```
import delimited "t_person", clear
```

**clear** angiver, at det aktuelle datasæt erstattes af det nye  
**firstrow** angiver, at variabelnavnene gemmes i første række i tabellen

3. Brug kommandoen via en do-fil

#### Eksport

Det er muligt at eksportere på flere forskellige måder:

1. Brug menuen File\Export
  - vælg Data to Excel spreadsheet og følg menuen, hvis det er Excel, du skal eksportere
  - vælg Data to Text og følg menuen, hvis det er en tekstfil, du skal eksportere
2. Skriv kommandoen direkte i Command

```
export excel using "t_person", firstrow(variables) replace
```

```
export delimited using "t_person", replace
```

**firstrow(variables)** angiver, at variabelnavnene gemmes i første række i tabellen  
**replace** angiver, at filer overskrives hvis den findes

3. Brug kommandoen via en do-fil

### 1.8 Import af data fra RedCap

Hvis man eksporterer data fra RedCap format, får man to filer:

- En CSV-fil, der indeholder selve data
- En DO-fil, der indeholder metainformation (labels, datatyper, ...)

For at konvertere dette til et Stata-datasæt er det letteste at:

- 1) Placér de to filer i samme mappe og vælg denne som "working directory" i Stata
- 2) Åben do-filen og kørs hele filen
- 3) Gem det resulterende datasæt som dta-fil (enten vha. menuen eller ved at tilføje en save-linje nederst i do-filen)

### 1.9 Overblik over data

#### Typisk kommandoopbygning:

Stata har en generisk måde at opbygge kommandoerne på, og hvis man har styr på den del, så bliver det nemmere at arbejde med Stata.

#### command varlist if exp in range

hvor:

- **command** beskriver hvad Stata skal gøre eks. sum, list eller save osv.
- **varlist** er den variabel ud af alle datasættets variabler, som man kan vælge
- **if exp** angiver de observationer, som Stata skal bruge. Hvis denne ikke er angivet, vil Stata udføre kommandoen på **alle** variabler i datasættet
- **in range** lidt samme funktion som if exp, men angiver begrænsninger ift. observationsrækken

Eksempel på kommando: list mpg weight if mpg>20 in 1/10

#### Stata kommandoer

Der findes to typer kommandoer i Stata:

##### 1) Kommandoer som rapporterer information om data:

Kommando:	Placering i menu:
describe	Data > Describe data > Describe data in memory or in a file
list	Data > Describe data > List data
summarize eller sum	Statistics > Summaries, tables, and tests > Summary and descriptive statistics > Summary statistics
codebook	Data > Describe data > Describe data contents (codebook)
count	Data > Data utilities > Count observations satisfying condition
browse eller br	Data > Data Editor > Data Editor (Edit) - <b>den man kan skrive i!</b> Data > Data Editor > Data Editor (Browse)
inspect	Data > Describe data > Inspect variables
table	Statistics > Summaries, tables, and tests > Other tables > Flexible table of summary statistics
tab	Statistics > Summaries, tables, and tests > Frequency tables > One-way table
sort	Data > Sort

Der er flere forskellige måder, man kan bruge kommandoerne på, og herunder er de mest brugte:

- a) describe - kommandoen helt alene, så får man en beskrivelse af alle variabler i datasættet
- b) describe var1 - kommandoen beskriver kun var1
- c) describe var1 var3 - kommandoen beskriver kun var1 og var3

- d) describe v\* - kommandoen beskriver alle variabler, som starter med v  
e) describe var1-var5 - kommandoen beskriver alle variabler fra var1 til og med var5

### 2) Kommandoer som ændrer data (ændrer kun data i memory):

Kommando:	Placering i menu:
use	File > Open
save	File > Save as
drop	Drop variables: Data > Variables Manager
	Drop observations: Data > Create or change data > Drop or keep observations
keep	Keep variables: Data > Variables Manager
	Keep observations: Data > Create or change data > Drop or keep observations
generate eller gen	Data > Create or change data > Create new variable
egen	Data > Create or change data > Create new variable (extended)
replace	Data > Create or change data > Change contents of variable
rename	Data > Data utilities > Rename groups of variables
recode	Data > Create or change data > Other variable-transformation commands > Recode categorical variable
encode	Data > Create or change data > Other variable-transformation commands > Encode value labels from string variable
decode	Data > Create or change data > Other variable-transformation commands > Decode strings from labeled numeric variable
tostring	Data > Create or change data > Other variable-transformation commands > Convert variables from numeric to string
destring	Data > Create or change data > Other variable-transformation commands > Convert variables from string to numeric
label define	Data > Variables Manager
label value	Data > Variables Manager

### Eksempler på hvordan udfald fra kommandoer ser ud:

#### 1. Codebook

```
. codebook pid
```

---

```
pid
```

---

```

      type:  numeric (long)
      range:  [30,562972]
unique values: 11,986
      mean:   283658
      std. dev: 163453
      missing.: 0/11,986
percentiles:  10%    25%    50%    75%    90%
              54422 143669 284178 425487 509754
```

# Stata intro

```
d_foddata
-----
type: numeric daily date (long)

range: [-21844,11835]          units: 1
or equivalently: [12mar1900,27may1992]  units: days
unique values: 8,776          missing .: 0/11,986

mean: -11346.1 = 07dec1928 (+ -3 hours)
std. dev: 4877.76

percentiles:    10%    25%    50%    75%    90%
               -17779  -14807  -11532  -8041  -4872
               29apr1911 18jun1919 05jun1928 26dec1937 30aug1946
```

## 1. Summarize

```
. summarize d_status_hen_start
```

Variable	Obs	Mean	Std. Dev.	Min	Max
d_status_h~t	11,986	13717.2	3755.077	3038	19358

```
. summarize d_status_hen_start, detail
```

d_status_hen_start				
Percentiles	Smallest			
1%	6670	3038		
5%	7574	3058		
10%	8424	3072	Obs	11,986
25%	10541	3193	Sum of Wgt.	11,986
50%	13957		Mean	13717.2
		Largest	Std. Dev.	3755.077
75%	17022	19358		
90%	18671	19358	Variance	1.41e+07
95%	19345	19358	Skewness	-.2058044
99%	19358	19358	Kurtosis	1.920386

## 3. Inspect

```
. inspect mpg
```

mpg: Mileage (mpg)		Number of Observations		
		Total	Integers	Nonintegers
#	Negative	-	-	-
#	Zero	-	-	-
#	Positive	74	74	-
# #	Total	74	74	-
# # #	Missing	-	-	-
# # # # .				

12 41  
(21 unique values)

## 4. Count

```
. count  
74
```

## 5. Tab

```
. tab foreign
```

Car type	Freq.	Percent	Cum.
Domestic	52	70.27	70.27
Foreign	22	29.73	100.00
Total	74	100.00	

## 6. Table

```
. table mpg
```

Mileage (mpg)	Freq.
12	2
14	6
15	2
16	4
17	4
18	9
19	8
20	3
21	5
22	5
23	3
24	4
25	5
26	3
28	3
29	1
30	2
31	1
34	1
35	2
41	1



## 2. DATAMANAGEMENT

### 2.1 Stata datakoncept

Stata kan kun arbejde med ét datasæt ad gangen. Det er derfor, vi altid starter med at skrive clear eller skriver clear efter kommandoen, som åbner en fil.

I Stata kaldes kolonnerne "variabler" og rækkerne "observationer"

Variabler har navne og/eller beskrivelser, samt en specifik datatype tilknyttet (tal, tekst, dato)

Observationer har alle samme dataformat.

Et datasæt ser nogenlunde ud som en Excel tabel.

En dependent variabel er outcome / afhængig variabel / y

En independent variabel er exposure / uafhængige variabel / x

### 2.2 Variabeltyper

byte	= numerisk heltal få cifre
int	= numerisk heltal
long	= numerisk heltal mange cifre
float	= numerisk decimaltal få decimaler
double	= numerisk decimaltal mange decimaler
.	= manglende observationer (numerisk)
"" (tom celle)	= manglende observationer (kategorisk / tekst)
string	= kategoriske/tekst variabler
Sort	= tal
Blå	= tekst variabler med numerisk værdi
Rød	= tekst

### 2.3 Stata syntakser

<	mindre end
>	større end
<=	mindre end eller lig med
>=	større end eller lig med
=	lig med (bruges når man vil ændre noget)
==	lig med (bruges når man sammenligner noget)
!	forskellig fra
!=	ikke lig med
&	og
	eller
+	plus
-	minus
*	gange
/	divider
^	opløftet i

### 2.4 Variabel manipulation

#### Generate

Kommandoen generate eller gen opretter en ny variabel. Du skal være opmærksom på, om det er en numerisk eller tekst variabel, som du opretter.

```
gen newvar=exp
```

#### Eksempel:

```
gen år ="2018"          (string variabel)
gen år= 2018           (numerisk variabel)
```

#### Hvis du vil bruge Stata's menu i stedet, finder du kommandoen her:

Menu: Data > Create or change data > Create new variable

#### Egen

Nogle gange kan man godt have brug for at lave en mere kompliceret matematisk funktion, og her er kommandoen egen mere hensigtsmæssig. Vær opmærksom på, at egen ikke tæller missing values med. Kommandoen egen laver en ny variabel afhængig af den funktion, vi har bedt den om.

```
gen newvar=function(var1)
```

#### Eksempel:

```
egen avg=mean(price)
egen sum=sum(price)
egen sd=sdt(price)
```

#### Hvis du vil bruge Stata's menu i stedet, finder du kommandoen her:

Menu: Data > Create or change data > Create new variable (extended)

#### Replace

Kommandoen replace ændrer indholdet i en variable.

```
replace var1="ny" if var1=="  → her ændres en observation til en anden
replace var1="ny", gen(var2) → her ændres en observation, og der laves en ny variabel
```

## Stata intro

---

Bemærk, hvis ikke man specificere med if her, så ændrer Stata alle observationerne i variabelen.

### Eksempel:

```
replace make="OPEL" if make=="Buick Opel"      (string)
replace trunk=1000 if trunk==11                (numerisk)
```

### Hvis du vil bruge Stata's menu i stedet, finder du kommandoen her:

Menu: Data > Create or change data > Change contents of variable

### Rename

Kommandoen rename ændrer navnet på en variabel.

```
rename oldvar newvar
```

### Eksempel:

```
rename make mærke
```

### Hvis du vil bruge Stata's menu i stedet, finder du kommandoen her:

Menu: Data > Data utilities > Rename groups of variables

### Recode

Kommandoen kan kun bruges på numeriske variabler, og den bruges til at ændre værdier.

```
recode var1 (exp), gen (newvar)
```

### Eksempel:

```
recode mpg (22=1000), gen (mpgny)           → her ændres én værdi til en ny
recode mpg (22=1000) (23=1000) (25=1000), gen (mpgny) → her ændres flere værdier til én værdi
```

Menu: Data > Create or change data > Other variable-transformation commands > Recode categorical variable

### Encode

Denne kommando giver en string variabel bagvedliggende numeriske værdier, hvilket betyder, at man visuelt har en string tekst i datasættet, men laver analyserne med den bagvedliggende numeriske værdi. Fordel med denne kommando er, at man kan lave statistik på den. Det kan man ikke på en string variabel.

## Stata intro

---

```
encode var1, gen(newvar)
```

### Eksempel:

```
encode make, gen(kategori)
```

### Hvis du vil bruge Stata's menu i stedet, finder du kommandoen her:

Menu: Data > Create or change data > Other variable-transformation commands > Encode value labels from string variable

### Decode

Kommandoen ændrer en string variabel med numeriske underværdier til en string variabel. Vær opmærksom på, at der ikke laves statistik på en string variabel.

```
decode var1, gen(newvar)
```

### Eksempel:

```
decode kategori, gen(marke1)
```

### Hvis du vil bruge Stata's menu i stedet, finder du kommandoen her:

Menu: Data > Create or change data > Other variable-transformation commands > Decode strings from labeled numeric variable

### Drop

Kommandoen drop fjerner de observationer og variable, der opfylder bestemt angivet betingelser.

```
drop var1           → her fjernes kun var1  
drop if var1==exp   → her fjernes kun de observationer i var1, som opfylder exp
```

### Eksempel:

```
drop mpg price  
dropp if make=="Buick Opel"
```

### Hvis du vil bruge Stata's menu i stedet, finder du kommandoen her:

Menu: Drop variables: Data > Variables Manager

Drop observations: Data > Create or change data > Drop or keep observations

## Stata intro

---

### Keep

Kommandoen keep beholder de observationer og variable, der opfylder bestemt angivet betingelser.

keep var1	→ her beholdes kun var1
keep if var1==exp	→ her beholdes kun de observationer i var1, som opfylder exp

### Eksempel:

```
keep mpg price  
keep if make=="Buick Opel" k
```

### **Hvis du vil bruge Stata's menu i stedet, finder du kommandoen her:**

Menu: Keep variables:      Data > Variables Manager  
      Keep observations:    Data > Create or change data > Drop or keep observations



### 2.5 Variabel konvertering

#### Tostring

Kommandoen tostring konverterer numeriske værdi til tekstfelter.

```
tostring var1, gen(newvar) → her ændres en numerisk værdi til en ny string variabel  
tostring var1, replace → her ændres en numerisk værdi til en string variabel i den eksisterende
```

#### Eksempel:

```
tostring trunk, gen (trunkny)  
tostring trunk, replace
```

#### **Hvis du vil bruge Stata's menu i stedet, finder du kommandoen her:**

Menu: Data > Create or change data > Other variable-transformation commands > Convert variables from numeric to string

#### Destring

Kommandoen destring konverterer tekstfelter til numeriske værdier. Bemærk, at der må kun være numeriske værdier i stringvariablen.

```
destring var1, gen(newvar) → her ændres en string værdi til en ny numerisk variabel  
destring var1, replace → her ændres en string værdi til en numerisk variabel i den eksisterende
```

#### Eksempel:

```
destring trunkny, gen (trunknyigen)  
destring trunkny, replace
```

#### **Hvis du vil bruge Stata's menu i stedet, finder du kommandoen her:**

Menu: Data > Create or change data > Other variable-transformation commands > Convert variables from string to numeric

### 2.6 Dato konvertering

Det kan være meget forskelligt, hvordan datoer ser ud, når man modtager et datasæt. Derfor er det vigtigt at kunne konvertere datoer, så man kan arbejde med dem.

#### Konvertering af datoer fra tekst til numeriske

Hvis en dato er angivet som string variabel, skal den konverteres til en numerisk dato, før man kan regne på datoen.

Først skal man tjekke, hvordan datoer ser ud: "dag/måned/år" eller "år/måned/dato" osv. for det skal bruges i kommandoen under konverteringen. Hvis datoen er "dag/måned/år", forkortes det med DMY. Derudover skal man også være opmærksom på, hvilket århundrede årstallet stammer fra. Det er nemlig muligt at have data fra både 1919 og 2020, hvis man arbejder med retrospektive datasæt. Efterfølgende angives den ønskede formattype, da Stata ellers kun viser en systemdato:

- ```
1.) gen dato=date(newvar, "DMY")
   format dato %td

2.) gen dato=date(newvar, "DMY", 2030)
   format dato %td
```

#### Vær obs på datoformat DMY, MDY, YMD

##### Eksempel:

```
gen datony=date(d_status, "DMY", 2030)
format datony %td
```

##### FORDI:

Nu har vi en datovariabel, vi kan arbejde med, hvis vi f.eks. skal lave et årstal:

```
gen ny=year(datony)
```

Hvis vi skal lave en alder ud fra to år:

```
gen alder =year1-year2
```

Hvis vi skal lave en alder ud fra to datoer:

```
gen alder=dato1-dato2
gen alderdage=alder/365.25           (365.25 pga. af skudår!)
```

Hvis man gerne vil have alder uden decimaler:

```
gen alderdage=int(alder/365.25)
```

## 2.7 Labels i datasæt og variabler

Der kan tilføjes beskrivelser til hele datasættet og de enkelte variabler vha. af label data og label variabler.

label variable var1 "text"

### Eksempel:

```
. describe  
  
Contains data from H:\OPEN Kurser\STATA Kursus efterår 2018\cpr_t_person.dta  
  obs:      11,986  
  vars:      12                      9 Sep 2015 16:24  
  size:     575,328
```

| variable name   | storage type | display format | value label | variable label |
|-----------------|--------------|----------------|-------------|----------------|
| pid             | long         | %12.0g         |             |                |
| d_foddato       | long         | %tdD_m_Y       |             |                |
| c_kon           | str1         | %1s            |             |                |
| c_status        | str2         | %2s            |             |                |
| d_status_hen_~t | long         | %td            |             |                |
| dato            | str9         | %9s            |             |                |
| datony          | float        | %td            |             |                |
| ny              | float        | %9.0g          |             |                |
| nyt             | float        | %9.0g          |             |                |
| alder           | float        | %9.0g          |             |                |
| Alder           | float        | %9.0g          |             |                |
| alderdage       | float        | %9.0g          |             |                |

```
Sorted by:  
Note: Dataset has changed since last saved.
```

label variable pid "Personnummer"

label variable d\_foddato "Fødselsdag"

label variable c\_kon "Køn"

label variable c\_status "Status"

label variable d\_status\_hen\_start "Dato for sidste ændring i status"

label data "Datasæt med population"

notes \_dta: Datasættet er lavet d. 17.09.2018

```
. describe  
  
Contains data from H:\OPEN Kurser\STATA Kursus efterår 2018\cpr_t_person.dta  
  obs:      11,986  
  vars:      12                      9 Sep 2015 16:24  
  size:     575,328  
  label:     Datasæt med population  
  notes:     (_dta has notes)
```

| variable name   | storage type | display format | value label | variable label                   |
|-----------------|--------------|----------------|-------------|----------------------------------|
| pid             | long         | %12.0g         |             | Personnummer                     |
| d_foddato       | long         | %tdD_m_Y       |             | Fødselsdag                       |
| c_kon           | str1         | %1s            |             | Køn                              |
| c_status        | str2         | %2s            |             | Status                           |
| d_status_hen_~t | long         | %td            |             | Dato for sidste ændring i status |
| dato            | str9         | %9s            |             |                                  |
| datony          | float        | %td            |             |                                  |
| ny              | float        | %9.0g          |             |                                  |
| nyt             | float        | %9.0g          |             |                                  |
| alder           | float        | %9.0g          |             |                                  |
| Alder           | float        | %9.0g          |             |                                  |
| alderdage       | float        | %9.0g          |             |                                  |

```
Sorted by:  
Note: Dataset has changed since last saved.
```

### 2.8 Kombinér flere datasæt

Datasæt kan kombineres på forskellige måder:

1) **append** sætter to datasæt med de samme variabler nedenunder hinanden (og udfylder manglende variabler med missing)

2) **merge** kombinerer to datasæt række-for-række på baggrund af en eller flere identifikationsvariable

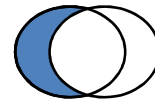
|            |                                                                                                                                    |
|------------|------------------------------------------------------------------------------------------------------------------------------------|
| 1:1        | Observationerne hører parvis sammen med én fra hvert datasæt                                                                       |
| 1:m og m:1 | Én observation i det ene datasæt hører til (op til) flere observationer i det andet                                                |
| m:m        | Flere observationer fra hvert datasæt hører sammen<br><b>(anbefales ikke at bruge, da resultat kan variere fra gang til gang!)</b> |

3) **joinby** virker som merge, men giver alle parvise kombinationer af identifikationsvariablerne

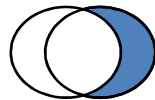
For at kunne lave et merge, skal der være en key eller unik variabel eller variabler i datasættene. Når merge bruges genereres en `_merge`-variabel, der definerer, hvordan data samles. Hvis man efterfølgende skal merge endnu et datasæt på, skal `_merge` variabelen slettes – ellers melder Stata fejl, fordi variabelen findes i forvejen!

#### Hvad betyder `_merge` variabelen?

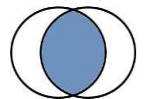
`_merge=1`; --> Ikke noget match med mergefilen



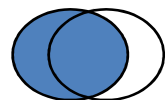
`_merge=2`; --> Ikke noget match med masterfilen



`_merge=3`; --> Match både master og merge



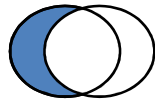
Oftest beholdes `_merge=1` og 3



I det efterfølgende er vist tre eksempler på merging af datasæt.

## Stata intro

A + B (\_merge 1)



**A**

| ID | Navn    | Køn |
|----|---------|-----|
| 1  | Ole     | M   |
| 3  | Inge    | K   |
| 4  | Kirsten | K   |
| 5  | Egon    | M   |

**B**

| ID | Telefon  |
|----|----------|
| 1  | 10203040 |
| 1  | 11121314 |
| 2  | 06050403 |
| 3  | 11223344 |
| 4  | 22334455 |
| 4  | 98766543 |
| 6  | 45234523 |

Kommer ikke med!

| ID | Navn    | Køn | Telefon  |
|----|---------|-----|----------|
| 1  | Ole     | M   | 10203040 |
| 1  | Ole     | M   | 11121314 |
| 3  | Inge    | K   | 11223344 |
| 4  | Kirsten | K   | 22334455 |
| 4  | Kirsten | K   | 98766543 |
| 5  | Egon    | M   |          |

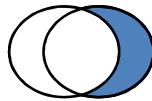
## A + B (\_merge 2)

### A

| ID | Navn    | Køn |
|----|---------|-----|
| 1  | Ole     | M   |
| 3  | Inge    | K   |
| 4  | Kirsten | K   |
| 5  | Egon    | M   |



Kommer ikke med!



### B

| ID | Telefon  |
|----|----------|
| 1  | 10203040 |
| 1  | 11121314 |
| 2  | 06050403 |
| 3  | 11223344 |
| 4  | 22334455 |
| 4  | 98766543 |
| 6  | 45234523 |

| ID | Navn    | Køn | Telefon  |
|----|---------|-----|----------|
| 1  | Ole     | M   | 10203040 |
| 1  | Ole     | M   | 11121314 |
| 2  |         |     | 06050403 |
| 3  | Inge    | K   | 11223344 |
| 4  | Kirsten | K   | 22334455 |
| 4  | Kirsten | K   | 98766543 |
| 6  |         |     | 45234523 |

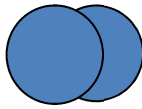
## Stata intro

---

A + B (\_merge 1, 2 og 3)

**A**

| ID | Navn    | Køn |
|----|---------|-----|
| 1  | Ole     | M   |
| 3  | Inge    | K   |
| 4  | Kirsten | K   |
| 5  | Egon    | M   |



**B**

| ID | Telefon  |
|----|----------|
| 1  | 10203040 |
| 1  | 11121314 |
| 2  | 06050403 |
| 3  | 11223344 |
| 4  | 22334455 |
| 4  | 98766543 |
| 6  | 45234523 |

| ID | Navn    | Køn | Telefon  |
|----|---------|-----|----------|
| 1  | Ole     | M   | 10203040 |
| 1  | Ole     | M   | 11121314 |
| 2  |         |     | 06050403 |
| 3  | Inge    | K   | 11223344 |
| 4  | Kirsten | K   | 22334455 |
| 4  | Kirsten | K   | 98766543 |
| 5  | Egon    | M   |          |
| 6  |         |     | 45234523 |

### 2.9 Vend datasæt

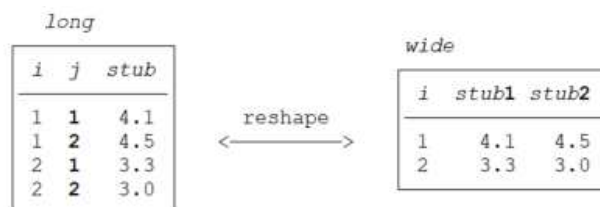
Nogen gange er det nødvendigt at vende et datasæt fra long eller wide format og omvendt:

Long betyder, at der er flere rækker per observation (dvs. flere rækker per person)

- nødvendigt i nogle statistiske modeller
- lettere at bruge for nogle typer datamanagement (f.eks. merging)

Wide betyder, at der er én række per person (dvs. flere observationer på rækken)

- oftest lettere at overskue
- lettere at bruge for nogle typer datamanagement (f.eks. variable der afhænger af flere observationer)



#### Fra long til wide

For at kunne vende et datasæt fra long til wide, skal der være en tidsvariabel.

```
reshape wide var1, i(unik variabel) j(tidsvariabel)
```

#### Eksempel:

```
gen Fødselsår= year(d_foddato)
```

Eksempel på tid vi kan lave, hvis det ikke er i datasættet

```
reshape wide c_status, i(pid) j(Fødselsår)
```

#### **Hvis du vil bruge Stata's menu i stedet, finder du kommandoen her:**

Menu: Data > Create or change data > Other variable-transformation commands > Convert data between wide and long



## Stata intro

---

### Fra wide til long

Her vil Stata selv generere en tidsvariabel.

```
reshape long var1, i(unik variabel) j(tidsvariabel)
```

### Eksempel:

```
reshape long c_status, i(pid) j(År)
```

### Hvis du vil bruge Stata's menu i stedet, finder du kommandoen her:

Menu: Data > Create or change data > Other variable-transformation commands > Convert data between wide and long

### 2.10 Ekstra tips

#### Hjælpevariabler

Der kan nogle gange være behov for at genere en variabel, der kan tælle enten rækkerne i datasættet eller bruges som indikator til at finde den første gang, et event er sket:

#### Tæl alle unikke variabler

`_n` er en indbygget variabel for nummeret på den nuværende række

```
bysort unikkevariable: gen n=_n
```

#### Gruppér alle unikke variabler

`_N` er tilsvarende en variabel for det totale antal rækker

```
bysort unikkevariable(r) : gen N=_N
```

#### Trim

Hvis man importerer fra andre kilder, kan det være nødvendigt at trimme string variabler:

```
replace varname = trim(varname)
```

#### Store bogstaver

Det kan også være, at alle karakterer skal være store bogstaver

```
replace varname = upper(varname)
```

#### Små bogstaver

Det kan også være, at alle karakterer skal være små bogstaver

```
replace varname = lower(varname)
```

### 2.11 Oversigt over stata kommandoer

#### Syntakser mm.

|   |    |                                            |
|---|----|--------------------------------------------|
| * | <  | mindre end                                 |
| * | >  | større end                                 |
| * | <= | mindre end eller lig med                   |
| * | >= | større end eller lig med                   |
| * | =  | lig med (bruges når man vil ændre noget    |
| * | == | lig med (bruges når man sammenligner noget |
| * | !  | forskellig fra                             |
| * | != | ikke lig med                               |
| * | &  | og                                         |
| * |    | eller                                      |

\*float og double er decimaltal

\*int og byte er heltal

\*str er tekst

\* . er manglernde observationer

\* independent variable = x = uafhængige variabel = exposure = indepvar

\* dependent variable = y = afhængige variabel = outcome = depvar

#### Kommandoer

|                                           |                                                                                     |
|-------------------------------------------|-------------------------------------------------------------------------------------|
| help +?                                   | hjælpefunktion i STATA - skriv: help + f.eks. dato                                  |
| cd "...\Navn_på_bibliotek_filerne_ligger" | det bibliotek, der peges på, bliver automatisk brugt                                |
| use "Navn_på_fil.dta", clear              | henter den fil, der skal bruges                                                     |
| save "Navn_på_fil.dta"                    | gemmer den fil, der skal bruges                                                     |
| save "Navn_på_fil.dta", replace           | replace overskriver den eksisterende fil                                            |
| log using "Navn_på_fil.log", replace      |                                                                                     |
| "her kommer f.eks en do. fil"             |                                                                                     |
| log close                                 | laver en log med dato over din kørsel<br>- replace overskriver den eksisterende log |
| log using "Navn_på_fil.log", append       |                                                                                     |
| "her kommer f.eks en do. fil"             |                                                                                     |
| log close                                 | laver en log med dato over din kørsel<br>- append laver en ny log                   |
| describe                                  | giver overblik over dastasæt"                                                       |
| codebook varname                          | giver overblik over den enkelte variable                                            |
| tab varname                               | giver overblik over variable - bedst til kategoriske variabler                      |
| tab varname, missing                      | missing viser, hvis der er observationer, som er tomme                              |
| summ varname                              | giver overblik over variable - bedst til numeriske variabler                        |
| count varname                             | tæller observationerne                                                              |

## Stata intro

---

|              |                                    |
|--------------|------------------------------------|
| list varname | viser observationer i listevisning |
| sort varname | sorterer variabel                  |
| browse       | viser data i browseren             |

\*\*\*\*\*

|                     |                                                                               |
|---------------------|-------------------------------------------------------------------------------|
| keep varname        | beholder variabelen                                                           |
| keep if varname ==1 | beholder variabelen, hvis den er lig med 1                                    |
| drop varname        | sletter variabelen                                                            |
| drop if varname ==1 | sletter variabelen, hvis den er lig med 1                                     |
| recode              | ændre numeriske værdier                                                       |
| replace             | erstatte værdier i observationer                                              |
| rename              | navngiver variabel                                                            |
| generate            | opretter ny variabel                                                          |
| clonevar            | kopiere eksisterende variabel                                                 |
| encode              | omdanner numerisk variabel til kategorisk - her laves samtidig en value label |
| destring, force     | omdanner tekst til tal, force tillader, at ikke-tal bliver til                |
| tostring            | omdanner numerisk til kategorisk                                              |
| label define        | definerer kategorier i en value label                                         |
| label value         | tildeler variabelen en value værdi                                            |

### Samle datasæt

merge 1:m unik\_varname using "Navn\_på\_fil.dta"

samler to datasæt, hvor én unik observation samler flere unikke observationer. F.eks. en person med flere behandlinger\*

merge m:1 unik\_varname using "Navn\_på\_fil.dta"

samler to datasæt, hvor mange unikke observationer samler én unik observation. F.eks. flere behandlinger med én person

joinby unik\_varname using "Navn\_på\_fil.dta"

samler parvise kombinationer af unikke observationer

append using "Navn\_på\_fil.dta"

lægger datasæt under hinanden

collapse (x f.eks p25) var1 (x f.eks mean) var2, by(var3)

lægger data sammen i grupper - ATTENTION - den sletter ikke anvendt data!

### Long eller wide format af data

Long format: En række per observation (dvs. flere rækker per person)

reshape long var1 var2, i(x) j(x)

Wide format: Én række per person (dvs. flere observationer på rækken)

## Stata intro

---

reshape wide var1 var2, i(x) j(x)

### Datoer

Konvertering af tekst til datoer

```
gen dato=date(datotekst, "DMY")
```

- brug ect. YMD, hvis datoen er 2017-12-24 og MDY hvis datoen er 12-24-2017

format %td dato

- %td kan erstattes af %tdYYYYNNDD eller %tdYYYY-NN-DD

```
gen datotekst2=string(dato, %td)
```

lav en variable med dd.

```
gen dato=date(current_date, "DMY")
```

format %td dato

- %td kan erstattes af %tdYYYYNNDD eller %tdYYYY-NN-DD

```
gen datotekst2=string(dato, %td)
```

### Find første variable i et datasæt

```
sort "unik id variable" "datovvariable"
```

```
bysort "unik id variable": gen n=_n
```

```
keep if n==1
```

### Tæl antal rækker inden for hver unik variable

```
bysort "unik id variable": gen "nyvariable"=_N
```

- ekskludér eller inkludér herefter det ønskede

### Tæl eller lav en unik variable i et datasæt

```
gen id =_n
```

### Descriptive statistik

Der findes et utal af grafer og histogrammer - vælg derfor fra Graphics menuen i Stata, når du skal lave:

graph

histogram

toway(scatter depvar indepvar)

### Basis statistik

ATTENTION: Ved tvivl om hvordan hele kommandoen skrives, så skriv f.eks.: help annova i kommandofeltet

anova

analyse af varians

cor

korrelation

pwcorr var1 var2,sig

parvis korrelation, sig refunderer p-værdi

oneway

one-way analyse af varians

ranksum

Wilcoxon ranksum test

tab depvar indepvar

two-by-two tabel, test af homogeneity, gamma

## Stata intro

---

|       |                                           |
|-------|-------------------------------------------|
| table | Laver sum i tabeller                      |
| ttest | Sammenligning af gennemsnit i små samples |

### Kohorter / survival analyser

ATTENTION: Ved tvivl om hvordan hele kommandoen skrives, så skriv f.eks.: help poisson i kommandofeltet

|           |                                       |
|-----------|---------------------------------------|
| stmh      | Manten-Haenzel rates                  |
| poisson   | Poisson regression                    |
| strate    | table of rates                        |
| stset     | definerer survival time rates         |
| staalen   | Aalen kummulative hazard              |
| stcox     | Cox regression (+stphtest)            |
| stsplit   | udvider data ifølge en lexis diagram  |
| sts graph | Kaplan Meyer survival funktion i graf |
| sts test  | logrank test                          |

### Case Control / Cross Sectional analyser

ATTENTION: Ved tvivl om hvordan hele kommandoen skrives, så skriv f.eks.: help mhodds i kommandofeltet

|         |                           |
|---------|---------------------------|
| mhodds  | Mantel-Haenzel odds ratio |
| partgam | partial Gamma koefficient |
| tabodds | oddstabel                 |

### Regressionsmodeller

ATTENTION: Ved tvivl om hvordan hele kommandoen skrives, så skriv f.eks.: help regress i kommandofeltet

|          |                                                |
|----------|------------------------------------------------|
| clogit   | conditional logistisk regression               |
| logit    | logistisk regresssion (output i log scale)     |
| logistic | logistisk regresssion (output i antilog scale) |
| lrtest   | likelihood ratio test                          |
| poisson  | poisson regression                             |
| predict  | prediction og indikatorer for modelfit         |
| regress  | linear regression                              |

ATTENTION - hvis: alle independent variabler er numeriske: regress depvar indepvar  
en independent variabel er kategorisk: xi: regress depvar indepvar  
en dependent variabel er categorisk bruges i. foran depvar  
regress i. depvar indepvar  
xi: regress i. depvar indepvar