

Øvelsestime 1

Introduktion til økonometri og Stata

Agenda

- 1) Introduktion til holdundervisning og PS1
- 2) Diskussionsopgaver: Gruppearbejde og opsamling
- 3) Introduktion til Stata
- 4) Stata øvelser: Gruppearbejde og opsamling

Tobias Gabel Christiansen
tgc@econ.ku.dk
15-02-2021

KØBENHAVNS UNIVERSITET



1) Introduktion til holdundervisning

Økonometri I

Faget består af *forelæsninger* (4 timer om ugen) og *holdundervisning* (3 timer om ugen)

Holdundervisning for hold 3 ligger mandag kl. 10:00 – 13:00 i uge 7 til og med uge 21
(*dog ingen undervisning i Påsken i uge 14*)

I første omgang er al undervisning online til og med 1. april

Vi benytter *samme* zoom-link til *alle* øvelsestimer (tilgængeligt på Absalon)

Jeg sidder i et lokale på CSS så vi har en tavle ved hånden hvis det bliver nødvendigt.

1) Planen for semesteret

Dato	Opgave
Uge 7 - 15/02	PS1
Uge 8 - 22/02	PS2
Uge 9 - 01/03	PS3
Uge 10 - 08/03	OO1
Uge 11 - 15/03	PS4
Uge 12 - 22/03	PS5 + gennemgang af OO1
Uge 13 - 29/03	PS6
Uge 14 - 05/04	Påskeferie
Uge 15 - 12/04	OO2
Uge 15 - 19/04	PS7
Uge 16 - 26/04	PS8 + gennemgang af OO2
Uge 17 - 03/05	OO3
Uge 18 - 10/05	PS9
Uge 19 - 17/05	PS10 + gennemgang af OO3
	Spørgetime

← Her er vi :-)

1) Obligatoriske opgaver

- Data og opgavesæt til de obligatoriske findes på Absalon.
- Ligeledes afleveres de via Absalon

Obligatorisk opgave 1	
Tilgængelig	Afleveres
d. 1 marts	d. 15 marts
Obligatorisk opgave 2	
Tilgængelig	Afleveres
d. 29 marts	d. 19 april
Obligatorisk opgave 3	
Tilgængelig	Afleveres
d. 26 april	d. 10 maj

- Hjemmeopgaverne er en god øvelse til **eksamenen**. Arbejd sammen i Jeres grupper og brug tid på at løse dem.
- Grupper på tværs af klasser er fint (send en mail til mig)

OBS: Alle hjemmeopgaver skal godkendes for at blive indstillet til eksamenen.

1) Introduktion til holdundervisning

Økonometri I

Fokus i forelæsningserne er:

- Grundlæggende teori
- Matematiske udledninger

Fokus i holdundervisningen er:

- Anvendelsen af økonometri
- Praktiske opgaver (Diskussion og Stata)
- Fortolke og kommunikere resultater

1) Introduktion til holdundervisning

Overordnet mål med Økonometri I

Rå data



Færdigskrevet rapport

Med andre ord vil I lære at *analysere, fortolke* og *formidle* data

Udover eksamen i Økonometri I vil en stor del af Jer sandsynligvis benytte metoder fra faget i forbindelse med Jeres bacheloropgaver, seminaropgaver og/eller speciale. Potentielt også i et fremtidigt job.

TALSTÆRKE STUDERENDE SØGES TIL TRANSPORT DIVISIONEN (STUDENTERMEDHJÆLPERE)

Jobbet

Transportdivisionen søger to studenter som kan bidrage til vedligeholdelse af data og modeller samt udarbejdelse af statistiske, økonomiske og økonometriske analyser inden for transportområdet. Du kommer til at arbejde i Transportdivisionen med primær forankring i Transport økonomi- gruppen og i Center for Transport Analytics.

Dine opgaver kan blandt andet være:

- Assistance med vedligehold af de Transportøkonomiske Enhedspriser
- Assistance med vedligehold af regnearksmodellen TERESA
- Databehandling og kvalitetskontrol af data i Transportvaneundersøgelsen
- Medvirken til udarbejdelse af analyser på Transportvaneundersøgelsen
- **Udarbejdelse af statistiske og (mikro)økonometriske analyser**
- Assistance i forbindelse med brug af registerdata

KRAKA SØGER DYGTIGE STUDENTER

Om jobbet

- Vi tilbyder et spændende og fagligt udfordrende studiejob i et uformelt og afslappet arbejdsmiljø, hvor du vil komme til at arbejde med aktuelle økonomisk-politiske problemstillinger som fx finans- og strukturpolitik, klimaudfordringen, coronakrisen mm.
- Dit kernearbejdsområde vil være at udarbejde **empiriske analyser** af fx uddannelse, arbejdsmarked, produktivitet og konkurrence, herunder ved brug af registerdata.

CORONA CRISIS

New research reveals what made Danes stay home, and what didn't

2020.06.15

In a new study at the Department of Economics, researchers have shown how recommendations to maintain social distancing affect our behaviour.



LIFE EXPECTANCY

2018.10.30

Rich people don't live that much longer than the poor

New research results challenge previous findings of huge differences in life expectancy between the rich and those at the bottom of the income scale.



EQUAL PAY

Wage transparency works

2018.12.06

New research shows that wage transparency reduces gender pay gap by 7%.



1) Introduktion til holdundervisning

Arbejdsformen

En stor del af øvelserne består af gruppearbejde (zoom breakout rooms), hvorefter vi gennemgår opgaverne i fællesskab.

Stiller større krav til Jer: Nødvendigt at deltage aktivt for at få mest muligt ud af undervisningen.

I skulle gerne på nuværende tidspunkt have fundet sammen i grupper af max 3 personer. Grupperne vil fungere som ramme for:

- Gruppearbejdet i holdundervisningen (PS1-PS10)
- Obligatoriske opgaver (i alt 3)
- Eksamen (12 timers tag-hjem-opgave)

1) Introduktion til holdundervisning

Spørgsmål?



2) Introduktion til dagens øvelser

PS1 og PS2: Simpel økonometrisk analyse af Engelkurver

- **PS1:** Indledende analyse (se på data og undersøge hvordan det kan bruges til formålet samt beskrive relevante variable)
- **PS2:** Estimere og fortolke en simpel OLS regression

Overordnet formål: Analysere hvordan husholdningers forbrug af bestemte goder afhænger af indkomsten.

2) Introduktion til dagens øvelser

Hvad er det nu en Engelkurve er?

En Engel kurve beskriver hvordan en husholdnings forbrugsudgifter på et bestemt gode varierer med husholdningens indkomst.

Forskellige goder påvirkes forskelligt (luksus gode vs. nødvendigheds gode).

Desuden kan vi definere forbrugsudgifter som:

Faktisk forbrug	Forbrugsandele
$Forbrug_i = \beta_0 + \beta_1 Indkomst_i + u_i$	$\frac{Forbrug_i}{Indkomst_i} = \beta_0 + \beta_1 Indkomst_i + u_i$
β_1 måler den forventede ændring i husholdningens forbrug på varen (målt i DKK) når indkomsten stiger med 1 DKK	β_1 måler den forventede ændring i husholdningens forbrugsandel på varen når indkomsten stiger med 1 DKK
Svært at fortolke hvorvidt gode er luksusgode eller nødvendighedsgode ud fra β_1	Luksusgode: $\beta_1 > 0$ Nødvendighedsgode: $\beta_1 < 0$

2) Introduktion til dagens øvelser

Data:

Canadian Consumer Expenditure Survey

- Selv-rapporteret forbrug over to uger i 1992
- Fokus på husholdninger der består af én person
- Variablene er nærmere beskrevet i PS1.
- Vi vil fokusere på de variable der begynder med 'x' da disse angiver forbrugsudgifter. Fx 'xfath' som angiver husholdningens forbrugsudgifter på mad.
- Datasættet indeholder også en række sociodemografiske variable. Vi vil senere benytte dummy-variablen der angiver køn (*dmale*)

2) Gruppearbejde

To-do næste 30 minutter:

- Læs første del af PS1
- Diskuter spørgsmål 1-4
- Hver gruppe vælger en referent til den efterfølgende diskussion (tilfældige referenter udvælges)
 1. In this exercise we rely on demand theory. Identify the relevant variables from demand theory.
 2. Which information in the data can be used as empirical measures of the theoretical variables? Discuss why we can not estimate price effect with the current data set.
 3. In the empirical analysis, total consumption is often used instead of income. Consider why. [Hint: Observe the products included in the data set and think about which products have been omitted.]
 4. Why can it be beneficial to analyze data for households with only one member?

OBS: Brug for hjælp? Skriv Jeres gruppenummer i Padlet så kommer jeg hurtigst muligt (link i chatten)

2) Gruppearbejde - opsamling

1. In this exercise we rely on demand theory. Identify the relevant variables from demand theory.

2) Gruppearbejde - opsamling

1. In this exercise we rely on demand theory. Identify the relevant variables from demand theory.

Tænk på den ideelle situation. De relevante variable er:

Afhængig variabel: Forbrug af enkelt-goder ($p_i x_i$)

Uafhængig variabel: Indkomst (I_i)

I praksis foretrækker vi at bruge **forbrugsandele** fremfor faktisk forbrug da det:

- i) Gør det lettere at tolke hvorvidt der er tale om et luksus- eller nødvendighedsgode.
- ii) Gør det lettere at sammenligne personer med forskellig indkomst.
- iii) Undgår at personer med høj indkomst og højt forbrug (ekstreme observationer) får stor vægt og skævvrider resultatet.

Det er derfor i praksis relevant at se på $\frac{p_i x_i}{I_i}$ som den afhængige variabel.

2) Gruppearbejde - opsamling

2. Which information in the data can be used as empirical measures of the theoretical variables? Discuss why we can not estimate price effect with the current data set.

2) Gruppearbejde - opsamling

2. Which information in the data can be used as empirical measures of the theoretical variables? Discuss why we can not estimate price effect with the current data set.

Tænk på hvordan vi bedst muligt kan implementere den ideelle situation med det tilgængelige data.

Afhængige variabel: Vi observerer husholdningernes forbrug af enkelt-goder. Et eksempel er:

$$p_i^{alc} x_i^{alc} = xalc_i$$

Uafhængig variabel: Vi observerer ikke husholdningernes indkomst. I stedet kan vi bruge en husholdningens samlede forbrug som approksimation

$$x_i^{total} = \sum_k p_i^k x_i^k = xfath_i + \dots + xmcloth_i$$

Med disse kan vi beregne **forbrugsandele** for hvert gode:

$$\frac{p_i^k x_i^k}{x_i^{total}} \quad \Rightarrow \quad \frac{xalc_i}{xfath_i + \dots + xmcloth_i}$$

2) Gruppearbejde - opsamling

2. Which information in the data can be used as empirical measures of the theoretical variables? Discuss why we can not estimate price effect with the current data set.

- Vi kan ikke beregne effekterne på forbruget af et gode når prisen stiger da vi kun har adgang til forbrugsudgifter.
- Vi kender ikke priserne på de enkelte goder.

2) Gruppearbejde - opsamling

3. In the empirical analysis, total consumption is often used instead of income. Consider why. [Hint: Observe the products included in the data set and think about which products have been omitted.]

2) Gruppearbejde - opsamling

3. In the empirical analysis, total consumption is often used instead of income. Consider why. [Hint: Observe the products included in the data set and think about which products have been omitted.]

- Vi **simplificerer** analysen ved kun at analysere hvordan folk vælger at allokere den del af indkomsten de har valgt at forbruge.
- **Samlet indkomst** kan skrives som:

$$Y = C + S$$

$C = \text{forbrug}$
 $S = \text{opsparing}$

- Vi inkluderer ikke opsparing i analysen da vi kun observerer husholdninger i én periode. Opsparing er et intertemporalt fænomen (bemærk også at vi kun inkluderer ikke-varige gode)
- Kan være et problem hvis der er **systematiske forskelle** i opsparingsraterne blandt folk (fx hvis rige opsparer mere vil vi underestimerer indkomstforskelle).
- **Ikke problem hvis:**
 - i) Ingen systematiske sammenhænge mellem indkomst og opsparing
 - ii) Husker at tolke resultater som ændringer i samlet forbrug og ikke indkomst.

2) Gruppearbejde - opsamling

4. Why can it be beneficial to analyze data for households with only one member?

2) Gruppearbejde - opsamling

4. Why can it be beneficial to analyze data for households with only one member?

- Igen **simplificerer** vi analysen da vi undgår at skulle tage stilling til:

Adfærd: Husholdninger af forskellige størrelser kan have forskellig forbrugsmønstre da visse udgifter deles helt eller delvist (fx elektricitet) mens andre er upåvirkede (fx tøj)

Måling: Større husholdninger bestående af personer med normal indkomst vil blive sammenlignet mindre husholdninger/enlige med meget høj indkomst.

Alternativt: Inkluder kontrolvariabel for husholdningens størrelse.

3) Introduktion til Stata: De fem vinduer

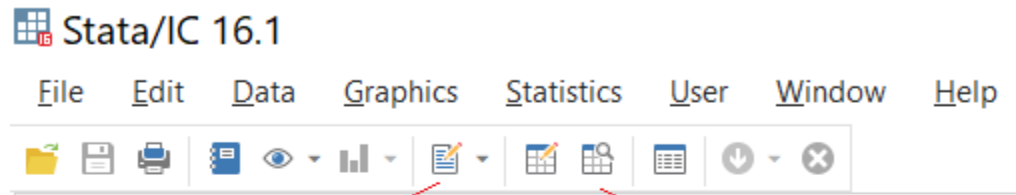
The screenshot displays the Stata/IC 16.1 software interface with five main windows:

- History:** Located on the left, it shows a list of commands entered. A red annotation states: **Historie:** Viser historik over tidligere udførte kommandoer.
- Command:** At the bottom, it contains the command prompt. A red annotation states: **Kommandovindue** Bruges til at teste kommandoer eller installere pakker. Kommandoerne vi bruger i analysen bør gemmes i en **do-file**.
- Results:** The central window showing the output of the command. It displays the Stata logo, version 16.1, copyright information, and license details. A red annotation states: **Resultater:** Viser output fra kommandoer fx tabeller med summary statistics eller regressionstabeller.
- Variables:** On the right, it lists the variables in the current dataset. A red annotation states: **Variable:** Giver et overblik over variablene i datasættet.
- Properties:** At the bottom right, it shows the properties of the selected variable. A red annotation states: **Egenskaber:** Viser egenskaber ved variablene i datasættet. Fx navn, værdi, type mv.

The status bar at the bottom indicates the file path: C:\Program Files (x86)\Stata16.

3) Introduktion til Stata: Funktioner

Udover de 5 vinduer er der adgang til en lang række funktioner og muligheder i værktøjsbjælken øverst. De to i vil benytte hyppigst er:



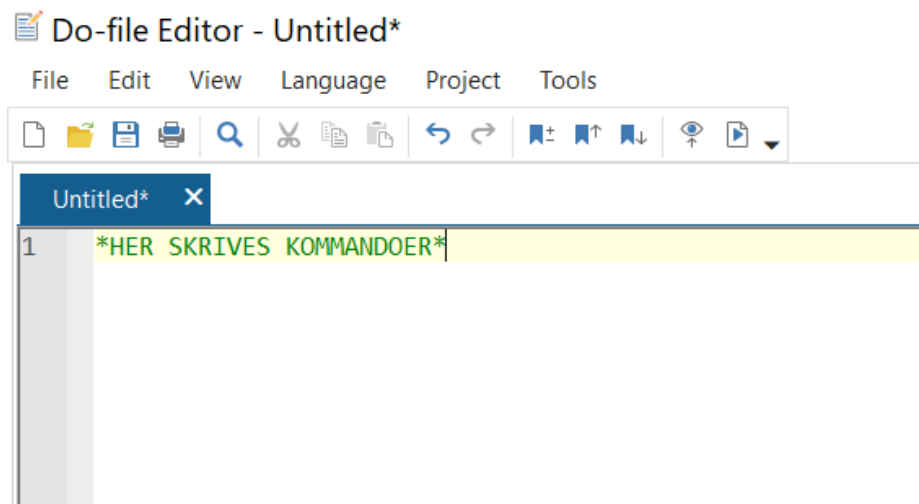
Bruges til at oprette en
do-file

Bruges til at se data

3) Introduktion til Stata: Do-file

En do-file samler og gemmer Stata kommandoer:

- Kommandoer skrives præcist som i kommandovinduet
 - Kommandoer udføres ved at trykke "CTRL"+"D", eller ved at trykke på ikonet her: 
- Som default udføres samtlige kommandoer i do-filen. For at udføre bestemte kommandoer kan disse markeres inden der trykkes "CTRL" + "D"



- Do-filer skal uploades som del af **eksamensbesvarelsen**.

3) Introduktion til Stata: Huskeregler til brug af Do-file

1. **Huske at gemme** do-filen så dit arbejde ikke går tabt.
2. **Load data gennem do-filen**, og rør derefter ikke ved datasættet (.dta-filen).
3. **Lav en mappe** hvor output-filer kan gemmes (dvs. angiv en 'current directory' eller 'cd'). Gem også do-file og .dta-fil her. På den måde er alt der hører til projektet samlet samme sted.

```
1 *GOD MÅDE AT STARTE DO-FILE
2 clear all                                // RYDDER EKSISTERENDE DATA
3 cls                                      // RYDDER RESULTATVINDUET
4 cd "STI TIL MAPPE HVOR DATA (.DTA-FILEN) ER GEMT" // ANGIVER ARBEJDSMAPPE
5 use NAVN-PÅ-DATASÆT.dta                // LOAD DATA
```

4. **Brug kommentarer** så du kan huske hvad du gjorde næste gang du åbner do-filen. Der er tre måder at kommenterer på:

```
*DEN ENNE ER VED AT STARTE LINJEN MED "*"
reg y x

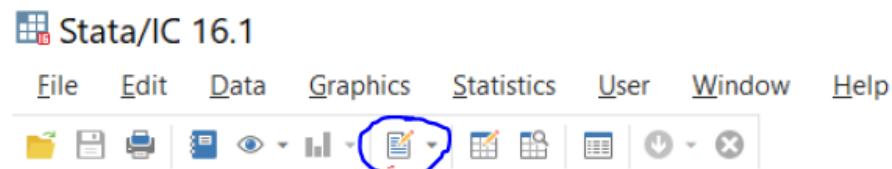
//DEN ANDEN ER VED AT SKRIVE "/"
reg y x //BEMÆRK AT MAN HER KAN KOMMENTERE EFTER KOMMANDOEN OGSÅ

/*
DEN TREDJE ER AT STARTE LINJEN MED "/*"
OG SLUTTE MED "*/". BEMÆRK HER AT MAN KAN KOMMENTERE
OVER FLERE LINJER DA STATA IGNORERER ALT IMELLEM.
*/
reg y x
```

3) Introduktion til Stata

Inden pausen:

1. Opret en mappe hvor I kan gemme data, do-file, output mv.
2. Hent data fra Absalon og gem det i denne mappe
3. Åben Stata og opret en do-file:



4. Navngiv og gem do-filen i samme mappe (File -> Save as)
5. Start nu do-filen med:

```
1 *PROBLEM SET 1
2 clear all
3 cls
4 cd "C:\Users\45414\OneDrive\Dokumenter\Økonomi mm\Holdunderviser\Økonometri I\Problem sets - Slides\PS 1"
5
6 *QUESTION 1: READ THE DATA FILE 'OPG1' INTO STATA
7 use opg1.dta //LOADS DATA
```

6. Klik "CTRL + D" og tjek at data er loaded

OBS: OneDrive eller Dropbox? Sørg for at mappen er synkroniseret/opdateret inden i forsøger at load data

OBS: Husk at gemme do-filen løbende (CTRL + S)

4) Gruppearbejde: Stata øvelser

To-do:

Arbejd i samme grupper på anden halvdel af PS1 der i korte træk går ud på at:

1. Load data
2. Se på data
3. Dan variabel for samlede forbrugsudgifter
4. Dan variable der angiver forbrugsandele for mad, tøj og alkohol.
5. Visualiser data ved at lave scatterplots
6. Kort deskriptiv analyse

Vi samler op 12:40

OBS: Brug for hjælp? Skriv Jeres gruppenummer i Padlet så kommer jeg hurtigst muligt (link i chatten)

4) Gruppearbejde - opsamling

1. Load data

```
1 *PROBLEM SET 1
2 clear all //CLEARS EXISTING DATA SO CODE CAN BE RE-RUN
3 cls //CLEAR RESULT WINDOW
4 cd "C:\Users\45414\OneDrive\Dokumenter\Økonomi mm\Holdunderviser\Økonometri I\Problem sets - Slides\PS 1" //ASSIGNS WORKING FOLDER
5
6 *QUESTION 1: READ THE DATA FILE 'OPG1' INTO STATA
7 use opg1.dta //LOADS DATA
```

- Hver gang vi kører koden ryddes alt når "**clear all**" + "**cls**" er inkluderet:
 - i) Undgår problemer hvis samme kommando gentages flere gange (fx generate var1).
 - ii) Sikre at do-filen indeholder alt hvad analysen kræver.
- **Output** (tabeller, grafer mv.) som gemmes, vil blive gemt i mappen som stien efter *cd* henviser til.
- Tjek altid **Data Browseren** efter data er loaded for at danne Jer et indtryk af strukturen:

	year	province	hgy	hage	nety	xfath	xrest	xhhop	xwcloth	xmcloth
1	92	1	20219	30	16579	1725	120	918	343	0
2	92	1	21268	39	17564	2300	956	1514	845	0
3	92	1	21570	35	17453	1300	400	523	880	0
4	92	1	27449	33	22156	620	420	435	1340	0
5	92	1	33290	38	27760	1600	2120	1178	1430	0

4) Gruppearbejde - opsamling

2. Undersøg data

i) Flere måder at undersøge data på. Typisk bruges:

- summary
- describe

```
. describe
```

Describe angiver at der er 450 observations og 38 variable i alt:

```
Contains data from opg1.dta  
obs:          450  
vars:          38
```

ii) Ingen køber både herre- og dametøj:

```
. count if xmcloth>0 & xwcloth>0
```


0

- Kan forklares ved at vi fokuserer på husholdninger der består af én person.
- Hvad med gaver? Kan skyldes at ingen køber gaver i observationsperioden (kun 14 dage)

Overordnet formål: Er data som vi forventer eller er der åbenlyse problemer? Vi kunne fx også tjekke at ingen husholdninger har et forbrug på 0

4) Gruppearbejde - opsamling

3. + 4. Lav nye variabel

```
15 *QUESTIONS 3+4: CONSTRUCT RELEVANT VARIABLES
16 generate xtot=xfath+xrest+xhhop+xwcloth+xmcloth+xtran+xrecr+xtob+xalc+xcare+xcaruse
17 generate wfath = xfath/xtot
18 generate wwcloth = xwcloth/xtot
19 generate wmcloth = xmcloth/xtot
20 generate walc = xalc/xtot
```

Budgetandele i stedet for faktisk forbrug gør det **lettere** at:

- Se om der er tale om et luksusgode eller nødvendighedsgode (hældning positiv eller negativ)
- Nemmere at sammenligne folk med meget forskellig indkomst

Budgetandele i stedet for faktisk forbrug **forbedre analysen** da:

- Ekstreme observation (ultra rige personer) påvirker resultatet mindre
- Mindre spredning i afhængige variabel kan reducere heteroskedasticitet (vender vi tilbage til)

4) Gruppearbejde - opsamling

5. Undersøg sammenhængen med scatter plots:

```
*QUESTION 5: SCATTER PLOTS
```

```
*MEN
```

```
scatter wfath xtot if dmale==1, name(men1, replace) title("FOOD (MEN)") xtitle("TOTAL EXP.") ytitle("EXPENDITURE SHARE")  
scatter wmcloth xtot if dmale==1, name(men2, replace) title("CLOTHES (MEN)") xtitle("TOTAL EXP.") ytitle("EXPENDITURE SHARE")  
scatter walc xtot if dmale==1, name(men3, replace) title("ALCOHOL (MEN)") xtitle("TOTAL EXP.") ytitle("EXPENDITURE SHARE")
```

```
*WOMEN
```

```
scatter wfath xtot if dmale==0, name(women1, replace) title("FOOD (WOMEN)") xtitle("TOTAL EXP.") ytitle("EXPENDITURE SHARE")  
scatter wwcloth xtot if dmale==0, name(women2, replace) title("CLOTHES (WOMEN)") xtitle("TOTAL EXP.") ytitle("EXPENDITURE SHARE")  
scatter walc xtot if dmale==0, name(women3, replace) title("ALCOHOL (WOMEN)") xtitle("TOTAL EXP.") ytitle("EXPENDITURE SHARE")
```

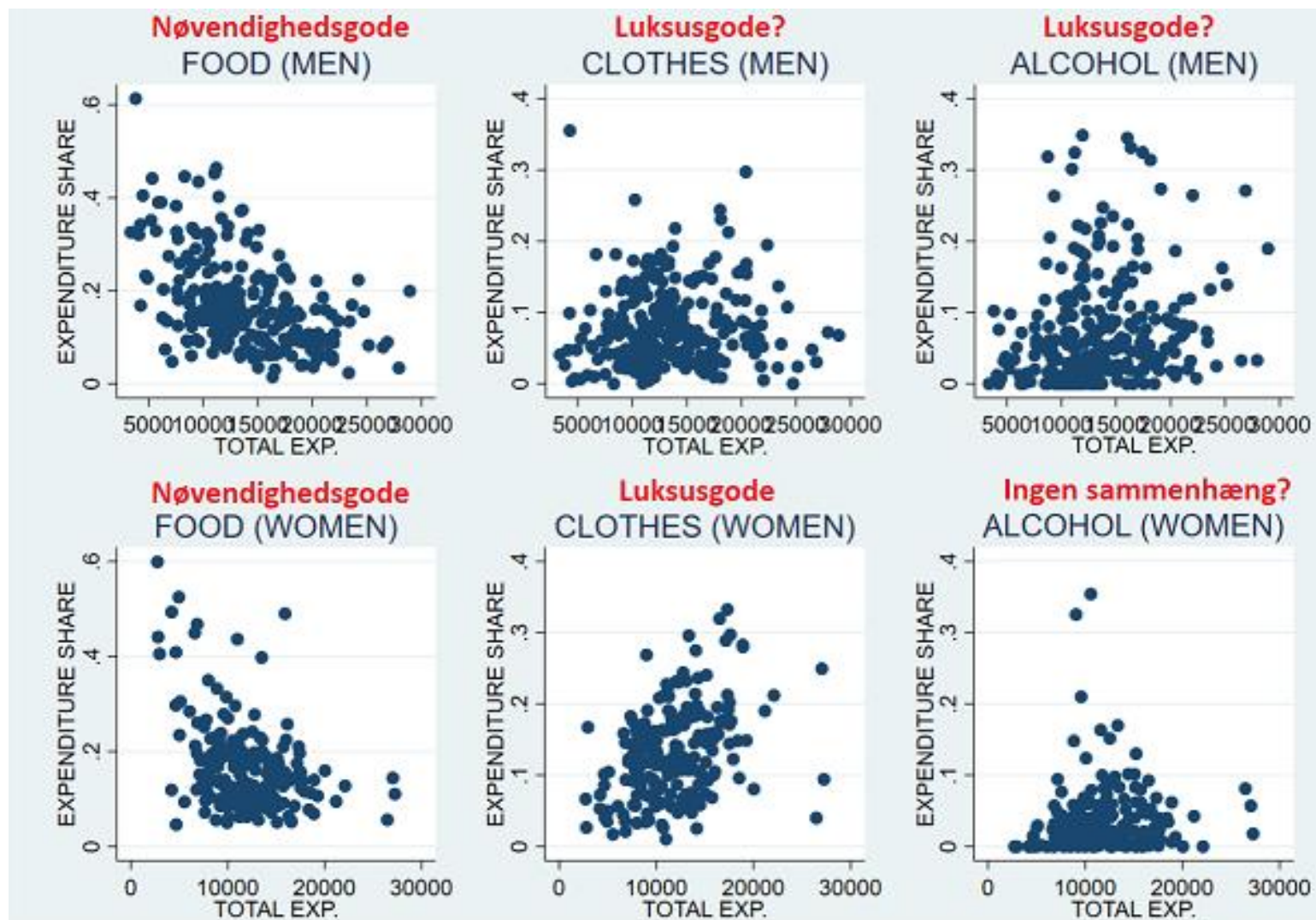
```
graph combine men1 men2 men3 women1 women2 women3
```

```
*NOTE: SCATTER PLOTS CAN BE SAVED USING GRAPH EXPORT
```

```
graph export engel_curves.png
```

- name("navn", replace) bruges til at henvise til grafen (husk at navne generelt ikke må indeholde mellemrum i Stata)
- graph combine samler grafer i ét vindue
- graph export gemmer grafen som et billede (.png) i stien angivet i begyndelsen af do-filen

4) Gruppearbejde - opsamling



4) Gruppearbejde - opsamling

6. Deskriptiv analyse

```
. *QUESTION 6: DESCRIPTIVE STATISTICS
. bysort dmale: summarize xtot wfath wwcloth wmcloth walc
```

```
-> dmale = 0
```

Variable	Obs	Mean	Std. Dev.	Min	Max
xtot	192	11855.1	4236.573	2746	27255
wfath	192	.1696893	.0965725	.0461046	.5990531
wwcloth	192	.1318719	.0662954	.0104545	.3325827
wmcloth	192	0	0	0	0
walc	192	.0364656	.047232	0	.3541864

```
-> dmale = 1
```

Variable	Obs	Mean	Std. Dev.	Min	Max
xtot	258	13645.04	4802.609	3374	28925
wfath	258	.1704464	.0962362	.0146672	.6132075
wwcloth	258	0	0	0	0
wmcloth	258	.0794321	.0566099	0	.3551402
walc	258	.0753147	.077122	0	.3485255

4) Gruppearbejde - opsamling

6. Deskriptiv analyse

- En **god** deskriptiv analyse opsummerer og fortolker kort de mest relevante sammenhænge i data.
- Citér **ikke** blot tal fra en tabel uden at tænke over det overordnede formål med analysen.

4) Gruppearbejde - opsamling

6. Deskriptiv analyse

Hvilke sammenhænge er de mest relevante?

Det vigtigste er at den deskriptive analyse lægger godt op til resten af rapporten.

Relevante ting at se efter er fx:

- Hvilke variable er vi interesserede i?
- Er der nogle grupper der er naturlige at sammenligne? Fx mænd/kvinder, control/treatment, i-lande/u-lande.
- Er der nogle ekstreme observationer (outliers) man bør tage højde for?
- Er der noget der tyder på at der er fejl i data? Fx husholdninger med forbrug på 0?

Ikke ét korrekt svar, men overvej altid:

- Om Jeres beskrivelse er relevant for resten af rapporten.
- Om I har argumenteret for at data rent faktisk kan anvendes til det ønskede formål.

4) Gruppearbejde - opsamling

6. Deskriptiv analyse

Citér ikke hovedløst tal fra tabellen. Beskriv sammenhængene med ord.

- Don't: Den gennemsnitlige værdi for x_{tot} er 11885 for kvinder og 13645 for mænd. Standardafvigelsen er 4236 og 4802.
- Do: I gennemsnit forbruger mænd mere end kvinder (\$13645 vs. \$11855), og samtidigt varierer deres forbrug mere (standardafvigelse på \$4803 vs. \$4237).

Beskriv hvorfor bestemte sammenhænge kan have indflydelse på analysen senere hen:

- Kvinder bruger en større andel af deres samlede forbrug på tøj end mænd (13 pct. vs. 8 pct.), mens mænd bruger en større andel på alkohol (8 pct. vs. 4 pct). Det indikerer at køn kan have indflydelse på forbrugsmønstre for disse goder. For andre goder, som fx mad, er der dog ingen sammenhæng.

Tak for i dag

- Husk at gemme datasættet efter I har dannet variablene. Vi skal bruge det næste gang. Gøres ved at slutte do-filen med:

```
*SAVE DATA  
save engel.dta, replace
```

- Datasættet gemmes i mappen angivet efter *cd*.