

Økonometri 1

Efterår 2006

Ugeseddel 1

Program

Præsentation af øvelser og øvelseslærer
Ideen bag de første ugesedler
Gruppearbejde
Opsamling af gruppearbejdet og introduktion af SAS
SAS-øvelser i computerkælder

Ideen bag de første ugesedler: Estimation af Engelkurver

Ideen bag ugeseddel 1 til 3 er på baggrund af efterspørgselsteorien at estimere Engelkurver. Vi ønsker altså empirisk at analysere, hvordan efterspørgslen af enkelte varer afhænger af indkomsten ved givne priser. Til analysen anvender vi et datasæt med information om et stort antal husholdningers udgifter til forbrug. Datasættet ligger på hjemmesiden for øvelserne.

Den økonomiske teori

Økonometrikerens opgave er at udforme en empirisk analyse på baggrund af teori og data. Den empiriske analyse tager udgangspunkt i den økonomiske teori om efterspørgsel, kendt fra fx kapitel 6 i Varian "Intermediate Microeconomics" fra Mikro på 1. årsprøve.

Data

Data stammer fra canadiske forbrugsundersøgelser. Udvalgte husholdninger fra 5 regioner i Canada har ført dagbog over alle deres udgifter i en periode på to uger i 1992. Oplysningerne er slået sammen i en række kategorier, hvoraf vores data indeholder følgende 11:

fath:	Mad (<u>food at home</u>)
rest:	Restaurant
hhop:	Telefon, rengøring, børnepasning, el, vand (<u>household operations</u>)
wcloth:	Tøj til kvinder
mcloth:	Tøj til mænd
caruse:	Forbrug til bil
tran:	Transport
care:	Medicin, læge, tandlæge etc.
recr:	Rekreation
tob:	Tobak
alc:	Alkohol

I data er variablene navngivet på den måde, at *xkategori* henviser til udgifter til kategorien; fx angiver variabelen *xfath* udgifter til mad.

Data indeholder også demografiske oplysninger. På ugeseddel 1 bruger vi variabelen *dmale*. Variablen angiver kønnet af husstandens "overhoved", hvor *dmale*=1, hvis personen er en mand, og 0, hvis personen er en kvinde.

Den økonometriske analyse

Den økonometriske analyse indeholder flere trin. Første trin består i at udvælge, transformere og beskrive de relevante variable fra data. Næste trin er at estimere Engelkurverne og at relatere resultaterne til den økonomiske teori. Ugeseddel 1 fokuserer på trin et.

Et afgørende trin for økonometrikeren er at "oversætte" den økonomiske teori, så den kan relateres til data. Det betyder bl.a., at man skal overveje hvilke variable, man vil bruge i sin analyse, og om man skal transformere dem eller konstruere nye variable.

For at gøre analysen så simpel som mulig har vi foretaget følgende afgrænsninger:

1. Vi betragter et tværsnit indsamlet i 1992, som udelukkende indeholder enlige uden børn og derfor kun består af husholdninger med et medlem.
2. Vi koncentrerer os om analyser af mad, alkohol og tøj for henholdsvis mænd og kvinder.
3. Vi ser i første omgang på sammenhænge, som kan beskrives ved en simpel regressionsmodel, dvs. en forklarende variabel.

Gruppearbejde

Sæt jer sammen i grupper af 3-4 personer og diskuter nedenstående spørgsmål. Hver gruppe vælger en talsmand, som afrapporterer for gruppen i den efterfølgende plenumdiskussion. Øvelseslæreren er til stede og kan hjælpe med at afklare spørgsmål undervejs.

Antag, at vi ønsker at analysere sammenhængen mellem efterspørgslen på alkohol og husholdningens indkomst.

1. Identificér de variable, der er relevante ud fra økonomisk teori. Dvs. angiv hvilke variable, vi skal bruge for at analysere sammenhængen mellem fx efterspørgsel på alkohol og husholdningens indkomst ifølge økonomisk teori.
2. Hvilken information i data kan benyttes som empiriske mål for de ønskede teoretiske variable?
3. I empiriske analyser benyttes som regel det samlede forbrug i stedet for indkomsten. Overvej hvorfor? Hint: tænk over, hvilke varegrupper der indgår i datasættet, og hvilke der kan være udeladte.
4. Hvorfor kan det være en fordel at starte med at analysere data for husholdninger med kun ét medlem?
5. Formulér en økonometrisk model for Engelkurven. Hvad skal man bruge som afhængig variabel? Hvad skal man bruge som forklarende variabel?
6. Hvad indeholder fejlleddet?

Efter at have identificeret de relevante variable, er det tid til at kigge på data. Ofte er data sat sammen af flere kilder eller trukket fra et større datasæt. En beskrivende analyse bør derfor være det første skridt i enhver empirisk analyse. Formålet er bl.a. at få et overblik over strukturen i data, at tjekke alle observationer er kommet med, og at data "ser fornuftige ud" sådan, at der fx ikke optræder fejlobservationer som negativt forbrug. Alt dette er formålet med den kommende SAS-øvelse.

SAS-øvelserne diskuteres ved næste øvelsesgang. Noter derfor dine observationer undervejs og print de nødvendige output og grafer.

En af grupperne bliver bedt om kort at svare på spørgsmål 8 skriftligt (½-1 side). Svaret sendes til Alice.Klynge@econ.ku.dk og din øvelseslærer senest mandag kl. 12. Det vil så blive lagt på fagets hjemmeside, under "øvelser" og "diskussionssiden".

SAS-øvelse

1. Indlæs data. Data ligger i en fil (opg1.sas7bdat), som kan hentes på hjemmesiden for øvelserne. Følgende programstump kan bruges til at indlæse data (se også "Elementær indføring i SAS" side 43):

```
LIBNAME ind "c:\wrk\";  
DATA ENGEL;  
SET ind.opg1;  
RUN;
```

2. Undersøg data. Man kan få et overblik over data ved at anvende faciliteten VIEWTABLE. Klik på "Explorer" nederst i SAS skærbilledet. Under "Libraries" og "Work" findes det genererede SAS datasæt "ENGEL". Ved dobbeltklik på ikonet åbner VIEWTABLE, som ligner et regneark.

Hvor mange husholdninger er der i datasættet? Hvor mange oplysninger er der for hver husholdning? Findes der husholdninger, der har udgifter til både dame- og herretøj, og er det realistisk? [Hint: skab et nyt datasæt, som kun indeholder de personer, der både har udgifter til dame- og herretøj.]

3. Konstruer de samlede udgifter. Følgende programstump kan måske hjælpe:
 $xtot = xfath + xrest + \dots + xwcloth;$

4. Lav en graf af udgiften til mad mod de samlede udgifter. Følgende program kan være til hjælp:

```
PROC GPLOT data=navn;  
PLOT xfath*xtot;  
TITLE "overskrift";  
RUN;
```

Hvilken sammenhæng er der mellem udgiften til mad og de samlede udgifter?

5. Konstruér udgiftsandelene for mad, tøj og alkohol:

```
....  
wfath=xfath/xtot;
```

```
.....
```

Hvorfor kan det være lettere at arbejde med udgiftsandelene for en vare frem dens udgift i absolutte størrelser?

6. Udvalg de relevante data. Lav et datasæt for enlige mænd og ét for enlige kvinder. Følgende SAS kommando kan bruges:

```
...
IF variabel=værdi;
...
```

7. Undersøg data ved hjælp af grafer. Plot for hhv. mænd og kvinder andelen til mad (samt tøj og alkohol) mod de samlede udgifter (se Statistik med SAS side 34 eller trin 3).

Hvilken sammenhæng er der mellem udgiftsandelen til mad (tøj og alkohol) og de samlede udgifter?

8. Lav en beskrivende analyse af udgiftsandelene til hver af de analyserede varegrupper og af totalforbruget for hvert køn. Det kan fx ske via PROC MEANS.

```
PROC MEANS data=navn n mean std median min max;
VAR xtot wfath .....
TITLE "overskrift";
RUN;
```

Proceduren beregner antal, gennemsnit, standardafvigelse, median, minimum og maximum for hver variabel. Opstil en tabel, som opsummerer resultaterne fra den beskrivende analyse. Kommentér kort på tabellen idet du fremhæver hovedtræk i data og evt. fejlobservationer.

Hjemmearbejde til næste øvelse:

Gør SAS-øvelserne færdige og løs følgende teoretiske opgaver:

1. Formuler den lineære regressionsmodel (et konstantled og en forklarende variabel) på matrix form for n observationer.
2. Opskriv OLS-estimatoren på matrix form. Vis, at når man udregner OLS-estimatoren, fås:

$$\hat{\beta}_0 = \bar{y} - \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} \bar{x}, \quad \hat{\beta}_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}, \text{ hvor } \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \text{ og } \bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$$

Hint: Brug regneregler for invertering af matrix: $\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad-bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$ samt regnereglerne (A.7) og (A.8) s. 709 i Wooldridge.