

STATSVIDENSKABELIG KANDIDATEKSAMEN 1999 II

Udvalgt fag

STIKPRØVETEORI

Skriftlig prøve 4 timer med hjælpemidler

EUROSTAT (Den Europæiske Unions statistiske kontor) fik i 1994 foretaget en stikprøve i samtlige medlemslande. Formålet var blandt andet at belyse indkomstfordelingerne internt i landene og mellem medlemslandene. I de efterfølgende år er familierne i stikprøven blevet geninterviewet og dermed er der opstået et panel, som benævnes ECHP (= European Communities Household Panel). Alle over 16 år i familien interviewes, deraf ordet household.

Et centralt emne er indkomst, og der spørges derfor om den månedlige indkomst før skat for det pågældende år.

I Danmark er det SFI-SURVEY, der står for stikprøvedesign og gennemførelse af interviewene.

På nuværende tidspunkt er oplysninger for 2. interviewrunde evalueret for samtlige lande i EU, der er derfor taget udgangspunkt i stikprøven gennemført i året 1995.

I det følgende ses på en delmængde af stikprøven, nemlig den del der består af familier med præcis 2 individer (og hvor begge er over 16 år).

Denne delmængde består af 824 familier omfattende 1.648 personer, og alle er blevet spurgt om deres månedlige indkomst i 1995.

Stikprøven er inddelt i 3 aldersklasser : De unge (16 år til og med 44 år), de midaldrende (45 år til og med 65 år) og de ældre (fra og med 66 år).

I nedenstående Tabel 1 vises fordelingen af indkomstoplysninger på de tre aldersklasser:

	n_k	$\bar{y}_k$	s_k
De unge	437	17.328	4.869
De midaldrende	696	18.280	6.830
De ældre	515	17.959	1.589

- 1) Brug oplysninger i Tabel 1 til at vise, at den gennemsnitlige indkomst for de 1.648 personer er 17.927 kr. Vis endvidere, at den tilsvarende spredning (s) er 5.186 kr.
- 2) Beregn spredningen på en estimator, der er baseret på en stikprøve på 1.648 personer, der er simpelt tilfældigt udtaget fra en population på $N=1.700.000$ med en spredning $S=5.186$ og udregn det tilsvarende 95% konfidensinterval.
- 3) Hvor stor skal stikprøven være, hvis længden af 95% konfidensintervallet skal være 896 kr. (svarende til 5%) ?

Fordelingen for hele landets befolkning (over 16 år), for de tre aldersgrupper er:

Procentfordeling
Opgjort af Danmarks Statistik

De unge: 16 år til og med 44 år.	60%
De midaldrende: 45 år til og med 65 år.	22%
De ældre : fra og med 66 år.	18%

- 4) Igen med udgangspunkt i Tabel 1, skal du angive den proportionale og den optimale allokering af en stikprøve på 1.648 personer blandt $N = 1.700.000$. De tilhørende spredninger skal også beregnes.
- 5) Hvor stor skal stikprøven være med en proportionalt allokeret stikprøve, hvis længden af det tilsvarende 95% konfidensinterval skal være 896 kr. ?

Den gennemsnitlige indkomst for de 1.648 personer i året 1994 var på 16.847 kr. Den tilsvarende spredning var 5.125 kr., mens korrelationskoefficienten mellem indkomsten for 1994 og 1995 var på 0,69241. På basis af registerkørsler i Danmarks Statistik kan det oplyses, at den gennemsnitlige indkomst for hele populationen i 1994 var på 14.400 kr.

- 6) Estimer den gennemsnitlige indkomst for 1995 dels med en ratioteknik, dels med en regressionsteknik.
- 7) Beregn endvidere de tilhørende spredninger, idet N stadig er 1.700.000.
- 8) Sammenlign de to estimatorer

I det følgende antages, at det er de 824 familier (som udgør de 1.648 personer), der er udtaget simpelt tilfældigt blandt 850.000 familier. Betraget på denne måde bliver den gennemsnitlige husstandsindkomst 35.854 kr. og den tilsvarende spredning 7.484 kr.

- 9) Beregn under disse forudsætninger spredningen på estimatorerne både for den gennemsnitlige personlige indkomst og den gennemsnitlige familieindkomst.
- 10) Estimer spredningen (s_m) mellem de 824 klynger
- 11) Vis, at den tilsvarende (s_i) er lig med 5.077,8.

Nu betragtes hele stikprøven for 1995. I alt er udtaget 2.990 familier med forskellige antal medlemmer over 16 år, familierne kan antages at være udtaget simpelt tilfældigt blandt alle 2.400.000 familier i Danmark. Alle medlemmer i familien over 16 år er blevet spurgt om indkomster. Antallet af personer i stikprøven er 4.938. Antallet af personer over 16 år i Danmark kan sættes til 4.300.000.

For hele stikprøven gælder, at den samlede indkomst er 88.335.162 kr.

- 12) Angiv to estimater for den gennemsnitlige månedlige indkomst for personer over 16 år i Danmark i 1995.
- 13) Beregn spredningerne for de to estimatorer, idet spredningen mellem husstandenes samlede indkomst er 13.644 kr. og $\tau_z = 7.990$ kr.
- 14) Giv en samlet kommentar til hele opgavesættet. F. eks. kan der foretages en sammenligning mellem den anvendte allokering og den proportionale samt den optimalt allokerede stikprøveplan i spørgsmål 4. Endvidere kan estimatorne og deres spredninger i spørgsmål 12 og 13 sammenstilles og efterfølgende kommenteres.