

STATSVIDENSKABELIG KANDIDATEKSAMEN 1996 II

UDVALGT FAG

STIKPRØVETEORI

Alle hjælpemidler tilladt

Gallup har undersøgt indkøbsvanerne i de danske familier. Fra dette materiale betragtes i dette eksamenssæt stikprøven af danske voksne i fast parforhold - gifte eller ugift samlevende, med eller uden børn. Denne population består ifølge Danmarks Statistik af omkring 1 200 000 par altså i alt 2 400 000 personer.

Fra denne population er der over telefonen udvalgt en simpel tilfældig stikprøve på 760 par, hvorfra der er interviewet én person i hvert par - i alt 400 mænd og 360 kvinder. I eksamenssættet betragtes svarene på et spørgsmål om, hvorledes parrene fordeler de daglige indkøb mellem sig. Svarene er givet som et tal, en procentdel, mellem 0 og 100. I de par, hvor manden har svaret, angiver han i gennemsnit, at kvinden foretager 62% af indkøbene, mens kvinden, når det er hende selv, der interviewes, i gennemsnit foretager 72% af indkøbene. I denne situation repræsenterer variationen mellem familierne de forskelle, der rent faktisk er mellem fordelingen af indkøbene familierne imellem. Variationen inden for familien repræsenterer den forskel, der er i mandens og kvindens opfattelse af fordelingen af indkøbene. På nær i spørgsmål 1) opfattes disse oplysninger som repræsentative, således at sandheden om familiens indkøbsfordeling ligger midt imellem mandens og kvindens svar. Den sande procentdel af indkøb foretaget af kvinden sættes derfor til 67%.

- 1) Diskuter om den observerede forskel på mænds og kvinders svar kan være udtryk for en skævhed ved den foretagne indsamling af data, og hvad der i givet fald kan gøres for at rette op på eventuelle skævheder.
- 2) Ud fra datamaterialet sættes populationsvariansen på den sande procentdel af indkøb foretaget af kvinden (altså på det tal der i middel estimeres til 67) til 29^2 . Angiv 95% sikkerhedsgrænser udtrykt relativt til den estimerede procentdel, når en stikprøve på 760 personer indsamles simpelt tilfældigt fra populationen på 2 400 000 individer.
- 3) Af de 360 interviewede kvinder angiver 107, at de står for 100% af familiens indkøb. Angiv 95% sikkerhedsgrænser for andelen af kvinder i fast parforhold som står for alle familiens indkøb.

- 4) Af de 360 interviewede kvinder angiver 4, at manden står for 100% af familiens indkøb. Angiv 95% sikkerhedsgrænser for andelen af kvinder, hvis mand står for alle familiens indkøb.
- 5) Det er nærliggende i denne situation at benytte stratifikation efter køn for at bestemme andelen for alle (de 67%). Idet variansen for både mændenes og kvindernes svar sættes til 28^2 ønskes en sammenligning af præcisionen ved den anvendte stratifikation 360 til 400 med præcisionen ved proportional allokering 380 til 380. Er det relevant at betragte optimal allokering?
- 6) Den foreliggende kønsfordeling 360 til 400 er imidlertid ikke et udtryk for en bevidst stratifikation, men snarere en tilfældighed opstået ved indsamlingen af stikprøven. Hvad betyder det for præcisionen, hvis den fundne skæve kønsfordeling rettes op ved vægtning?
- 7) En nærliggende tanke er, at kvindens andel af indkøbene har sammenhæng med antal børn i familien. Idet antal børn opfattes som en hjælpevariabel (x) med populationsgennemsnit 0.9 og varians $S_x^2 = 1.0^2$ ønskes en vurdering af, om regressionsestimation og ratioestimation vil forbedre præcisionen. Korrelationskoefficienten mellem interessevariablen og hjælpevariablen beregnes i stikprøven til 0.03.
- 8) Er der muligheder for at forbedre præcisionen ved systematisk udvælgelse efter en passende liste over populationen?
- 9) Antag at variansen mellem svarene fra de to personer i samme par sættes til $S_k^2 = 25 = 5^2$ for alle par, således at S_{INDEN}^2 er væsentligt mindre end den samlede populationsvarians, der er $S^2 = 29^2$. Idet familierne opfattes som klynger, ønskes variansen ved en et-trins klyngeudvælgelse af 380 par, dvs. tilfældet hvor begge personer i de udvalgte par interviewes, angivet. Hvad er forskellen mellem denne stikprøveplan og simpel tilfældig udvælgelse af 760 personer?
- 10) Beregn præcisionen i stikprøven, som den faktisk er foretaget, dvs. to-trins klyngeudvælgelse af først 760 par og dernæst interview af en tilfældigt udvalgt person i hvert par. Hvad er forskellen mellem denne to-trins klyngeudvælgelse af i den sidste ende 760 personer og simpel tilfældig udvælgelse af 760 personer? Er forskellen væsentlig?