

STATSVIDENSKABELIG KANDIDATEKSAMEN 2000 II

Udvalgt fag

STIKPRØVETEORI

Skriftlig prøve 4 timer med hjælpemidler

I de senere år har der været stor fokus på danskernes tilbagetrækningsalder.

Socialforskningsinstituttet (SFI) har undersøgt danskernes forventede tilbagetrækningsalder i flere omgange. I 1997 blev der foretaget en større stikprøveundersøgelse i Danmark, blandt personer der var i alderen 20 år til 70 år. Et af spørgsmålene lød : ”I hvilken alder synes De selv, at det er passende at holde op med at arbejde?”

På grund af de mange undersøgelser SFI har udført vedrørende tilbagetrækningsalder, kan det i det følgende antages at variationskoefficienten (CV) herfor er 0,07.

Stikprøven i 1997 var $n=1.310$ og populationen i alderen 20 år til 70 år kan antages at være $N=3.502.000$ personer.

Den gennemsnitlige tilbagetrækningsalder for de 1.310 udtrukne personer var på 61,2 år.

- 1) Brug ovennævnte oplysninger til at vise at spredningen (s) i stikprøven er 4,3. Beregn spredningen på estimatet for den gennemsnitlige tilbagetrækningsalder, der er baseret på en stikprøve på 1.310 personer, der er udtaget simpelt tilfældigt fra en population på $N=3.502.000$ med en spredning på $S=4,3$ og udregn det tilsvarende 95% konfidensinterval.
- 2) Hvor skal stikprøven være, hvis længden af det tilsvarende 95% konfidensinterval skal være 1 år ?

Af de 1.310 udtrukne personer svarede 2 af dem, at alder større end eller lig 90 år var passende.

- 3) Estimer andelen af personer der synes at den passende tilbagetrækningsalder skal være mindst 90 år og angiv det tilsvarende 95% konfidensinterval.

Blandt en del forskere hersker der den opfattelse, at den passende tilbagetrækningsalder er afhængig af svarpersonens alder.

Den gennemsnitlige alder for de 1.310 svarpersoner er 50,1 år og den tilsvarende spredning er på 6,2 år. For den samlede population på 3.502.000 personer gælder at den gennemsnitlige alder er på 42,6 år og den tilsvarende spredning på 13,8 år. Korrelationskoefficienten mellem svarpersonens alder og den passende tilbagetrækningsalder er på 0,32265.

- 4) Estimer den gennemsnitlige tilbagetrækningsalder dels med en ratioeknik dels med en regressionsteknik.
- 5) Beregn de tilsvarende spredninger og sammenlign de to estimatorer.

I nedenstående Tabel 1 vises den gennemsnitlige tilbagetrækningsalder geografisk fordelt.

	n_k	$\bar{y}_k$	s_k	N_k
Hovedstadsregion	248	61,5	3,3	812.000
Øvr. Sjælland & øer	433	61,4	3,3	1.093.000
Jylland	629	60,8	5,1	1.597.000
I alt	1.310			3.502.000

- 6) Med udgangspunkt i Tabel 1, skal du angive den proportionale og den optimale stikprøve allokering af en stikprøve på 1.310 personer samt beregne de tilhørende spredninger.
- 7) Hvor stor skal den optimalt allokerede stikprøve være, hvis længden af det tilsvarende 95% konfidensinterval skal være 1 år ?

De personer, som i stikprøven er lønmodtagere, er blevet spurgt om deres månedlige bruttoindkomst. Hermed bliver stikprøven reduceret til 1.091 personer. Betragt disse 1.091 personer som udtaget simpelt tilfældigt blandt 2.406.000 lønmodtager i Danmark.

I nedenstående Tabel 2 er vist mænd og kvinders gennemsnitlige månedlige bruttoindkomst.

Tabel 2 Månedlig bruttoindkomst.

	n_k	$\bar{y}_k$	s_k	N_k
Mænd	560	22.789	9.648	1.262.000
Kvinder	625	15.894	4.798	1.144.000
I alt	1.091			2.406.000

- 8) Estimer den gennemsnitlige månedlige bruttoindkomst blandt lønmodtagere i Danmark og beregn den tilsvarende spredning.
- 9) Kommenter beregningsformen for udregningen af spredningen.

SFI har lavet flere undersøgelser hvori spørgsmålet ”I hvilken alder synes De selv, at det er passende at holde med at arbejde?” indgår. I en husstandsundersøgelse for EUROSTAT er spørgsmålet også stillet.

I det følgende betragtes populationen af 2 personers husstande i Danmark. Denne population udgøres af 800.000 husstande og dermed af 1.600.000 personer. I denne population af husstande er der udtrukket simpelt tilfældigt en stikprøve på 386 husstande. Begge personer i husstanden har besvaret spørgsmålet ”I hvilken alder synes De selv, at det er passende at holde med at arbejde?”

Gennemsnittet af de 772 personers svar er 60,9 og den tilsvarende varians er på 14,8 (dermed er spredningen på 3,9).

Endvidere er der indenfor hver af de 386 husstande udregnet en varians. Gennemsnittet af disse 386 varianser er 10,1 (dermed er spredningen 3,2).

De 60,9 år er et estimat for populationens ønskede tilbagetrækningsalder.

10) Vis at variansen (s_m^2) mellem de 386 husstande er 19,5.

11) Estimer spredningen for den gennemsnitlige tilbagetrækningsalder.

12) Sammenlign en stikprøve på 1000 personer udtrukket simpelt tilfældigt blandt de 1.600.000 personer med en stikprøve på 500 udtrukket simpelt tilfældigt blandt de 800.000 husstande. Idet det antages at begge personer i husstanden også udspørges.

Udgifterne til transport er i gennemsnit kr. 50,00 pr. adresse. Lønudgifterne til interviewererne er i gennemsnit kr. 200,00 pr. interview. Når begge personer i husstanden interviewes er den gennemsnitlige lønudgift pr. interview kr. 150,00. Dette skyldes at nogle spørgsmål kan udelades i det 2. interview.

13) Beregn omkostningerne ved en stikprøveplan, hvor der udtages 1.000 personer simpelt tilfældigt. Beregn omkostningerne ved en stikprøveplan, hvor der udtages 500 husstande og hvor begge personer i husstanden udspørges. Hvilken stikprøveplan vil du anbefale? Du bedes begrunde valget af stikprøveplan.