

STATSVIDENSKABELIG KANDIDATEKSAMEN 1995 I

Udvalgt fag

STIKPRØVETEORI

Skriftlig prøve 4 timer med hjælpemidler

Nielsen Marketing undersøger salget af en lang række varer i danske detailforretninger for at kunne levere opgørelser over salgsvolumener og salgspriser til interesserede kunder, der f.eks. kan være virksomheder, der vil kende effekten af deres reklameindsats. Ideelt set skulle Nielsen Marketing indsamle data i alle landets butikker, men da dette naturligvis er uoverkommeligt, anvendes stikprøver. Da strukturen i dansk detailhandel er under hastig forvandling, skal stikprøven ofte justeres, for at Nielsen kan opnå den størst mulige præcision for de mindst mulige omkostninger. Opgaven illustrerer nogle af de overvejelser, Nielsen kan gøre sig i den anledning, ud fra salget af én vare, Lätta margarine, i uge 39 i 1994. I opgaven er salget opgjort i antal solgte pakker Lätta den pågældende uge.

- a) I en mindre kæde på 39 meget ens butikker er der indsamlet data for alle butikker. Det gennemsnitlige salg var 40.1 pakker og standardafvigelsen var 18.0. Som et led i en større stikprøve blandt alle landets butikker ønsker Nielsen at anvende data for 3 simpelt tilfældigt udvalgte forretninger i kæden. Angiv variationskoefficienten for stikprøvegennemsnittet for de 3 butikker samt længden af et 95% konfidensinterval for det gennemsnitlige salg i kædens butikker.
- b) Som hjælpevariabel anvender Nielsen forretningernes samlede salgssum i 1993. Det gennemsnitlige salg i 1993 var 13.8 mill. kr. og standardafvigelsen var 5.3. Korrelationen mellem salget af Lätta i uge 39 og det samlede salg i 1993 var 0.25. Angiv estimatorernes standardafvigelse, når hjælpevariablen anvendes i en ratioestimation og en regressionsestimation i en stikprøve af størrelse 3. Kommentér de eventuelle præcisionsforbedringer i forhold til simpel tilfældig udvælgelse.
- c) I Nielsens stikprøve på 3 butikker var salget i uge 39 hhv. 31, 90 og 33 pakker Lätta. De tre butikkers gennemsnitlige salg i 1993 var 13.9 mill. kr. Sammenhold værdierne af $\bar{y}_{SI}$, $\bar{y}_R$ og $\bar{y}_{REG}$, idet populationsværdier for gennemsnit, varianser og korrelationer i a) og b) anvendes. Diskutér herudfra det mest hensigtsmæssige valg af estimator.
- d) I ingen af stikprøvens butikker var salget under 25 pakker Lätta. Angiv 95% sikkerheds-

grænser for antallet af kædens 39 butikker med salg under 25 pakker Lätta.

- e)** Vurdér eventuelle fordele og ulemper ved systematisk udvælgelse af en stikprøve på 3 af kædens 39 butikker. Kan systematisk udvælgelse overflødiggøre brugen af hjælpevariable?
- f)** Nielsen arbejder ud fra en population på ialt 3731 butikker, der er fordelt på landsdele som angivet i tabel 1, der også viser estimerede værdier for standardafvigelsen for salget af Lätta i de fire landsdele. Der gælder

$$\sum W_k S_k = 123.4 \quad \text{og} \quad \sum W_k S_k^2 = 21334.8 .$$

Angiv den proportionale og den optimale allokering af en stikprøve af samlet størrelse 100 samt estimatorernes standardafvigelser. Kommentér forskellen mellem de to allokeringer og simpel tilfældig udvælgelse, idet standardafvigelsen i hele populationen er $S = 149.5$.

- g)** Antag at en stikprøve på størrelsen 100 indsamles simpelt tilfældigt over hele landet, og at der dernæst vægtes for at gøre stikprøven repræsentativ m.h.t. landsdele. Hvorledes vil man på forhånd formode, at vægtningen påvirker præcisionen? Antag dernæst at stikprøven rent faktisk består af 10 butikker i Hovedstaden, 18 på Sjælland, 15 på Fyn og 57 i Jylland. Hvordan vil man efter at have set denne fordeling vurdere vægtningens betydning for præcisionen?
- h)** Nielsen har inddelt populationen af landets 3731 butikker i 20 kategorier, der i det følgende opfattes som klynger. Antag at der udvælges $\kappa = 4$ kategorier simpelt tilfældigt, indenfor hvilke der vælges simple tilfældige stikprøver af butikker. Beregn estimatorne $\bar{y}_{kl-1}$ og $\bar{y}_{kl-R}$ ud fra data i tabel 2.
- i)** Angiv hvorledes variansen på estimatorerne i **h)** kan estimeres ud fra oplysningerne i tabel 2. Beregningerne forventes ikke udført.

Tabel 1. Data fra fire landsdele.

Landsdel	N_k	S_k
Hovedstaden	522	188.5
Sjælland	788	250.1
Fyn	330	105.5
Jylland	2091	62.2

Tabel 2. Data fra fire kategorier.

k	N_k	n_k	$\bar{Y}_k$	s_k
1	338	27	26.5	13.5
2	29	3	54.0	17.4
3	85	4	5.3	2.9
4	306	15	13.2	13.5