

**RAPPORTER**

**85/34**

**STATISTISKE EGENSKAPER VED  
BYRÅETS STANDARD UTVALGSPLAN**

AV  
TOR HALDORSEN

---

STATISTISK SENTRALBYRÅ  
CENTRAL BUREAU OF STATISTICS OF NORWAY

RAPPORTER FRA STATISTISK SENTRALBYRÅ 85/34

STATISTISKE EGENSKAPER VED  
BYRÅETS STANDARD UTVALGSPLAN

AV  
TOR HALDORSEN

STATISTISK SENTRALBYRÅ  
OSLO — KONGSVINGER 1985

ISBN 82-537-2271-0  
ISSN 0332-8422

**EMNEGRUPPE**

**Økonomisk og statistisk teori og analyse**

**ANDRE EMNEORD**

**Utvalgsmetoder**

## FORORD

Statistisk Sentralbyrå har en standard utvalgsplan for landsomfattende utvalg fra befolkningen. Denne brukes ved de fleste av Byråets utvalgsundersøkelser, bl.a. ved arbeidskraftundersøkelsene (AKU). Fullstendige tellinger som folke- og bolig tellingen 1980 gir en sjelden anledning til å vurdere sider ved utvalgsplanen som ikke lar seg gjøre ved hjelp av resultatene fra de løpende undersøkelsene. I rapporten brukes folketellingsdata til en vurdering av statistiske egenskaper til utvalgsplanen og enkelte metoder Byrået bruker i forbindelse med utvalgsundersøkelser.

Tor Haldorsen, Erling Siring og Rolf Aaberge planla prosjektet, Brynjulv Hauksson gjorde programmeringsarbeidet og Tor Haldorsen stod for gjennomføringen og har utarbeidet denne rapporten.

Statistisk Sentralbyrå, Oslo, 18. oktober 1985

Arne Øien



## INNHOOLD

Side

1. Innledning .....	7
1.1. Bakgrunn .....	7
1.2. Utvalgsplanen .....	7
1.3. Data .....	7
2. Utvalgsvarians .....	8
2.1. Opplegg .....	8
2.2. Designeffekt .....	9
2.3. Stratifiseringseffekt .....	18
2.4. Effekt av to-trinnstrekking .....	20
2.5. Effekt av endret befolkningssammensetning .....	23
3. Betinget forventning gitt trekkingen på første trinn .....	27
3.1. Om usikkerhetsbegrepet .....	27
3.2. Avvik som en del av usikkerheten .....	28
3.3. Konsekvenser av usikkerhetsberegningene .....	31
4. Klyngeeffekt .....	32
4.1. Trekkenhet og analyseenhet .....	32
4.2. Beregninger av klyngeeffekten .....	32
5. Effekt av ulike trekkesannsynligheter .....	36
5.1. Bakgrunn .....	36
5.2. Beregninger av effekten .....	36
5.3. Konsekvenser .....	40
Appendiks	
1. Formelverk .....	41
Litteratur .....	43
Utkommet i serien Rapporter fra Statistisk Sentralbyrå (RAPP) .....	44



## 1. INNLEDNING

### 1.1. Bakgrunn

Statistisk Sentralbyrås nåværende utvalgsplan for landsomfattende undersøkelser ble tatt i bruk i 1975. Forarbeidet er beskrevet i Thomsen og Rideng (1974). En beskrivelse av planen er gitt i SØS nr. 33. Beregning av enkelte statistiske egenskaper finner en i Sæbø (1976), (1977). I den opprinnelige utvalgsplanen er det foretatt bare små endringer fram til i dag.

En begrenset vurdering av utvalgsplanen kan gjøres i forbindelse med den enkelte undersøkelse. Resultatene kan brukes til å estimere størrelser som belyser ulike sider ved planen. Men estimeringen vil være forbundet med en viss usikkerhet og i noen tilfelle savner vi gode estimeringsmetoder. Ideelt sett trenger en opplysninger om hele befolkningen for å evaluere utvalgsplanen. Slik informasjon har vi i folketellingene. Datagrunnlaget for mine beregninger er folke- og bolig tellingen i 1980.

Det er flere forhold som har betydning når en skal lage en utvalgsplan. Det er praktiske, økonomiske og statistiske hensyn som må samordnes. Det er derfor grunn til å understreke at det er de statistiske egenskaper som vil bli behandlet i denne rapporten.

Selv om opplysninger om hele befolkningen har vært tilgjengelige, har det vært nødvendig å foreta enkelte tilnærmelser og forenklingen. Disse er beskrevet i den løpende tekst. Formelverket er gjengitt i Appendiks. De fleste formler vil en finne på grunnlag av boka til Des Raj (1968).

### 1.2. Utvalgsplanen

I utvalgsplanen er landet delt i 102 strata (områder). 24 av disse består av bare én kommune (dog er Hole slått sammen med Ringerike og Siljan med Skien). Dette er store kommuner med som regel mer enn 30 000 innbyggere og de fleste er bykommuner. De andre strata består av to eller flere kommuner. Ved stratifiseringen tok en hensyn til geografisk beliggenhet og dernest kommunetype (næring) og innbyggertall. De 24 strata dekker om lag 40 prosent av befolkningen. I disse trekkes personer/-familier direkte eller vi kan si at kommunen trekkes med sannsynlighet 1.00 på første trinn og at det så trekkes personer/familier. De andre strata er delt i to eller flere primære utvalgsområder som hver består av én eller flere kommuner. Her trekkes først ett primærområde i hvert stratum, så trekkes personer/familier fra dette i annet trinn. Trekkingen på første trinn foregikk proporsjonalt med folketallet i 1970. For hele landet omfatter de trukne primærområdene om lag 60 prosent av befolkningen. For praktiske formål regner vi med at det på annet trinn trekkes tilfeldig.

Siden 1975 har en stort sett holdt fast på primærområdene som ble trukket på første trinn. Et fåtall områder er byttet ut. Beregningene bygger på de områder som nå er i bruk. På annet trinn trekkes det stort sett på nytt for hver undersøkelse.

Når det trekkes et utvalg, bestemmes antall enheter fra det enkelte område slik at utvalget blir selvveiende (dvs. alle enheter har samme trekkesannsynlighet når vi regner med trekkingen på begge trinn). Det gjør at gjennomsnittet i utvalget er en forventningsrett estimator. Det er egenskapene til denne estimatoren som vil bli behandlet. I kapittet 5 vil en spesielt se på konsekvensene av at en i noen tilfeller strengt tatt ikke har selvveiende utvalg.

I noen mindre undersøkelser trekkes utvalget egentlig i tre trinn da personer/familier trekkes fra én og bare én tredel innen utvalgsområdene. Egenskapene til slik trekking blir ikke behandlet i denne rapporten.

### 1.3. Data

Som nevnt før er utgangspunktet for disse beregningene Folke- og bolig telling 1980. Denne omfatter alle personer (også utenlandske statsborgere) som ifølge Forskrifter om føringen og ordningen av folkeregistrene fastsatt av Statistisk Sentralbyrå den 26. november 1979, var registrert bosatt i Norge på tellingstidspunktet den 1. november 1980.

Folketellingen omfattet 4 091 132 personer. I beregningene er 45 170 holdt utenfor. Disse var ikke bosatt i privathusholdning eller de bodde i husholdning med mer enn fem familiekjerner. Hensikten er å skaffe en masse som er nær målpopulasjonen i de fleste undersøkelser og der det er lett å regne på de innsamlingsprosedyrer som brukes.



Folketellingsopplysningene stammer fra offentlige registre eller skjema som den enkelte har fylt ut. Ved en utvalgsundersøkelse vil en bruke besøk- eller telefonintervju eller intervju pr. post. Svar på samme spørsmål vil kunne variere med innsamlingsmetode. Beregningene bygger på at folk vil svare som i Folketellingen. En må også være oppmerksom på at det kan være forskjell i definisjonen av samme begrep i Folketellingen og andre undersøkelser f.eks. Arbeidskraftsundersøkelsen (AKU).

## 2. UTVALGSVARIANS

### 2.1. Opplegg

Utvalgsvarians er det tradisjonelle mål på usikkerhet til anslagene fra en utvalgsundersøkelse. I publikasjoner fra utvalgsundersøkelser har som regel Byrået tatt med en tabell for å antyde størrelsen på utvalgsvariansen. Grunnlaget for tabellen er en tommelfingerregel som sier at utvalgsvariansen er 50 prosent større enn når en benytter et tilfeldig utvalg fra hele befolkningen. Dette er en grov regel da vi vet at forholdet varierer med utvalgsstørrelse og geografisk fordeling av kjennetegnet. For eksakte variansberegninger må en kjenne variabelverdiene i hele målpopulasjonen. Varianser kan også estimeres i løpende undersøkelser, men det krever spesielle beregninger og enkelte trekk i vår utvalgsplan gjør at det ikke fins særlig gode estimatorene for variansen. Laake (1974) viser at den største vanskeligheten er at det trekkes bare ett primærområde pr. stratum. Eksakte beregninger med visse mellomrom bør kunne gi grunnlag for bedre brukerveiledning enn det som nå er tilfelle.

Beregningene bygger på at vi har et selvveiende utvalg og at gjennomsnittet i utvalget brukes som estimator. Det er flere forhold som påvirker variansen til gjennomsnittet. For det første er det stratifiseringen, for det andre er det to-trinnstrekkingen og videre er det forholdet mellom aktuelle befolkningstall og folketellingstall i 1970 for primærområdene innen hvert stratum. Rapporten prøver å vise hvordan hver av disse forhold påvirker utvalgsvariansen.

Når en bruker spesielle utvalgsplaner, er det vanlig å bruke tilfeldig utvalg fra hele befolkningen som målestokk. Designeffekten er forholdet mellom varians til aktuell utvalgsplan og tilfeldig utvalg. I dette kapitlet vil jeg splitte opp designeffekten og se på den som produktet av tre ulike effekter; stratifiseringseffekt, to-trinnseffekt og effekt av befolkningsendring fra 1970 til 1980. Dette gjør jeg ved å tenke meg fire ulike planer for å gjennomføre en utvalgsundersøkelse. Strata og primærområder er hele tiden som i Byråets standardplan. For gitt størrelse på totalutvalget er planene:

- A: Tilfeldig utvalg fra hele populasjonen.
- B: Stratifisert utvalg med proporsjonal allokering. Dette innebærer at det trekkes tilfeldig innen hvert stratum og antallet som trekkes er proporsjonalt med strataenes befolkning i 1980.
- C: To-trinnstrekking i aktuelle strata med 1.trinns sannsynligheter proporsjonale med befolkningstallene for primærområdene i 1980. Utvalget fordeles på strata som i B.
- D: Byråets standardplan. Atskill seg fra C ved at 1.trinns sannsynligheter beregnet med 1970-tall. Fordelingen av utvalget på strata vil kunne avvike fra den i C.

Alle utvalg er selvveiende og utvalgsvariansene betegnes med var A, var B, var C og var D.

Designeffekten til Byråets standardplan blir da var D/var A. Denne kan skrives

$$\frac{\text{var D}}{\text{var A}} = \frac{\text{var B}}{\text{var A}} \times \frac{\text{var C}}{\text{var B}} \times \frac{\text{var D}}{\text{var C}}$$

Her kaller jeg var B/var A stratifiseringseffekt, var C/var B to-trinnseffekt og var D/var C effekt av befolkningsendring. Disse vil bli behandlet i hvert sitt underkapittel, men først skal vi se på designeffekten.

## 2.2. Designeffekt

Denne måles altså som forholdet mellom varians til Byråets utvalgsplan og varians ved tilfeldig trekking fra hele landet. Hvis effekten er større enn 1, viser det at det aktuelle opplegg er mindre effektivt i statistisk forstand enn tilfeldig utvalg. Grovt sagt er variansen ved Byråets utvalgsplan av formen  $a/n + b$  og ved tilfeldig trekking av formen  $c/n$ , der  $n$  er utvalgsstørrelsen og  $a$ ,  $b$  og  $c$  er parametre som bl.a. avhenger av hvilket kjennetegn vi behandler. Det gjør at designeffekten blir av formen  $a/c + (b/c) \cdot n$ , altså en lineær funksjon i  $n$ . Siden  $b$  og  $c$  alltid er positive så vil designeffekten øke lineært med utvalgsstørrelsen. Videre ser en at variansen ved Byråets utvalgsplan (som ved alle to-trinns utvalgsplaner) avtar med  $n$ , men ikke bli null selv om en trekker samtlige enheter innen de uttrukne primære utvalgsenheter.

I tabellene har jeg tatt med tall for tre ulike masser. Det er alle personer, personer 16 - 74 år og alle (privat-) husholdninger. Byrådet gjennomfører undersøkelser for alle tre typer enheter. Aldersbegrensningen 16 - 74 år er den som brukes i arbeidskraftundersøkelsene (AKU). Siden denne har en dominerende plass blant Byråets utvalgsundersøkelser, har jeg tatt med mange variable som gjelder arbeidsforhold. Vi skal i et senere kapittel se at det er spesielle vanskeligheter med å ha husholdning som trekkeenhet, men i denne omgang forutsetter vi at det trekkes selvveiende utvalg med husholdninger.

Tabell 1. Designeffekt for utvalg av ulike størrelser. Alle personer

Variabel	Utvalgsstørrelse				
	1 000	2 500	5 000	12 000	30 000
Bor i leilighet med					
Bad .....	1,14	1,36	1,73	2,76	5,45
WC .....	1,14	1,37	1,75	2,82	5,58
Telefon .....	1,03	1,13	1,28	1,72	2,85
Kjønn .....	1,01	1,04	1,08	1,19	1,47
Yrkesaktiv .....	1,01	1,04	1,09	1,23	1,58
Bor i tettbygd område .....	0,99	1,50	2,35	4,74	10,91
Viktigste kilde til livsopphold					
Inntekt av arbeid .....	1,08	1,22	1,44	2,08	3,72
Pensjon, trygd .....	1,00	1,02	1,05	1,13	1,35
Stipend, lån, formue e.l. ....	1,00	1,00	1,01	1,02	1,05
Verneplikt/siviltjeneste .....	1,00	1,00	1,00	1,01	1,01
Udefinert .....	1,00	1,01	1,01	1,03	1,08
Hustype					
Våningshus .....	0,97	1,14	1,41	2,17	4,15
Frittliggende enebolig .....	0,99	1,17	1,47	2,32	4,49
Rekkehus .....	0,99	1,09	1,24	1,68	2,80
Horisontal tomannsbolig .....	1,00	1,02	1,06	1,18	1,48
Annet under 3 etg. ....	1,00	1,02	1,06	1,15	1,40
Blokk, leiegård .....	0,77	0,86	1,03	1,48	2,65
Forretningsbygg .....	1,00	1,01	1,02	1,06	1,15
Hotell, aldershjem .....	1,00	1,01	1,02	1,05	1,11
Udefinert .....	1,00	1,01	1,03	1,08	1,21
Alder .....	1,01	1,03	1,08	1,22	1,58
Antall rom .....	1,03	1,10	1,20	1,51	2,30

Tabell 2. Designeffekt for utvalg av ulike størrelser. Personer 16-74 år

Variabel	Utvalgsstørrelse				
	1 000	2 500	5 000	12 000	30 000
Yrkesaktive .....	1,04	1,11	1,23	1,55	2,40
Ekteskapelig status					
Ugift .....	1,00	1,01	1,02	1,06	1,16
Gift .....	1,02	1,07	1,16	1,43	1,98
Enke/enkemann .....	1,00	1,00	1,01	1,03	1,07
Skilt .....	0,99	1,00	1,01	1,03	1,09
Separert .....	1,00	1,00	1,01	1,02	1,04
Utdanning					
1 Barneskolenivå .....	1,00	1,00	1,00	1,00	1,01
2 Ungdomsskolenivå .....	0,99	1,03	1,10	1,28	1,76
3 Gymnasnivå I .....	1,01	1,02	1,03	1,09	1,22
4 Gymnasnivå II .....	1,00	1,01	1,04	1,10	1,26
5 Universitets- og høgsolenivå I ...	0,99	1,00	1,02	1,07	1,20
6 Universitets- og høgsolenivå II ..	1,00	1,00	1,01	1,03	1,08
7 Universitets- og høgsolenivå III .	0,99	1,00	1,02	1,05	1,15
8 Forskernivå .....	1,00	1,01	1,01	1,04	1,10
9 Uoppgitt .....	1,00	1,00	1,00	1,00	1,00
Udefinert .....	1,00	1,00	1,01	1,04	1,10
Yrkesfelt					
1 Administrasjons- og forvaltnings- arbeid, bedrifts- og organisasjons- ledelse .....	1,00	1,00	1,01	1,04	1,10
2 Kontorarbeid .....	0,99	1,00	1,02	1,06	1,18
3 Handelsarbeid .....	1,00	1,01	1,02	1,06	1,15
4 Jordbruks-, skogsbruks- og fiske- arbeid .....	0,99	1,06	1,18	1,51	2,36
5 Gruve- og sprengingsarbeid m.m. ...	1,02	1,06	1,14	1,34	1,86
6 Transport- og kommunikasjonsarbeid	1,00	1,01	1,03	1,07	1,18
7 Industri-, bygge- og anleggsarbeid	1,00	1,03	1,07	1,18	1,47
8 Industri-, bygge- og anleggsarbeid	1,01	1,04	1,09	1,22	1,58
9 Servicearbeid .....	1,00	1,01	1,02	1,05	1,14
Næring					
1 Jordbruk, skogbruk, fiske og fangst ..	0,99	1,07	1,20	1,58	2,55
11 Jordbruk .....	1,00	1,11	1,28	1,76	3,01
12 Skogbruk .....	1,02	1,07	1,14	1,36	1,93
13 Fiske og fangst .....	1,01	1,12	1,30	1,81	3,12
2 Oljeutvinning og bergverksdrift .....	1,10	1,32	1,68	2,70	5,33
3 Industri .....	1,01	1,07	1,16	1,43	2,11
31 Produksjon av næringsmidler, drikke- varer og tobakksvarer .....	1,03	1,09	1,20	1,50	2,29
32 Produksjon av tekstilvarer, bekled- ningsvarer, lær og lærvarer .....	1,03	1,09	1,20	1,51	2,31
33 Produksjon av trevarer .....	1,06	1,21	1,44	2,12	3,85
34 Treforedling, grafisk produksjon og forlagsvirksomhet .....	1,02	1,08	1,17	1,45	2,15
35 Produksjon av kjemiske produkter, mineralolje-, kull-, gummi- og plastprodukter .....	1,05	1,19	1,41	2,03	3,65
36 Produksjon av mineralske produkter	1,04	1,11	1,24	1,61	2,55
37 Produksjon av metaller .....	1,21	1,67	2,44	4,60	10,18
38 Produksjon av verkstedprodukter ...	1,04	1,14	1,31	1,78	2,99
4 Kraft- og vannforsyning .....	1,01	1,03	1,07	1,18	1,45
5 Bygge- og anleggsvirksomhet .....	1,01	1,03	1,06	1,15	1,39
6 Varehandel, hotell- og restaurant- virksomhet .....	1,00	1,02	1,05	1,15	1,38
7 Transport, lagring, post og tele- kommunikasjoner .....	1,01	1,02	1,05	1,13	1,32
8 Bank- og finansieringsvirksomhet, for- sikringsvirksomhet, eiendomsdrift og forretningsmessig tjenesteyting .....	0,99	1,00	1,01	1,04	1,14
9 Offentlig, sosial og privat tjeneste- yting .....	1,01	1,04	1,08	1,22	1,56

Tabell 2 (forts.). Designeeffekt for utvalg av ulike størrelser. Personer 16 - 74 år

Variabel	Utvalgsstørrelse				
	1 000	2 500	5 000	12 000	30 000
<b>Yrkesstatus</b>					
Ansatt .....	1,02	1,09	1,19	1,50	2,28
Selvstendig .....	1,00	1,02	1,07	1,19	1,51
Familiemedlemmer .....	1,00	1,02	1,05	1,15	1,41
Udefinert .....	1,01	1,02	1,06	1,15	1,38
<b>Arbeidstid i året</b>					
100- 499 timer .....	1,00	1,00	1,01	1,03	1,07
500- 999 " .....	1,00	1,01	1,01	1,03	1,08
1 000- 1 299 " .....	1,00	1,00	1,01	1,02	1,04
1 300- .....	1,02	1,05	1,10	1,25	1,65
Udefinert .....	1,01	1,03	1,06	1,15	1,40
<b>Arbeidstid i uka 25.10.-31.10. 1980</b>					
0 timer .....	1,00	1,01	1,01	1,04	1,09
1- 9 timer .....	1,00	1,00	1,01	1,02	1,04
10-19 " .....	1,00	1,00	1,01	1,02	1,05
20-29 " .....	1,00	1,00	1,01	1,02	1,06
30+ " .....	1,02	1,05	1,11	1,27	1,69
Udefinert .....	1,01	1,02	1,06	1,14	1,37
<b>Inntekt .....</b>	<b>1,01</b>	<b>1,06</b>	<b>1,15</b>	<b>1,38</b>	<b>1,97</b>

Tabell 3. Designeffekt for utvalg av ulike størrelser. Husholdninger

Variabel	Utvalgsstørrelse				
	500	1 500	3 000	6 000	15 000
Eier leiligheten .....	1,00	1,04	1,10	1,22	1,57
Bor i leilighet med					
Bad .....	1,02	1,09	1,19	1,39	2,00
WC .....	1,03	1,11	1,23	1,47	2,20
Telefon .....	1,00	1,05	1,12	1,26	1,68
Kjøkken .....	1,00	1,04	1,10	1,23	1,60
Ovn .....	0,90	0,95	1,03	1,19	1,68
Disponerer bil .....	0,99	1,04	1,12	1,28	1,75
Antall rom .....	0,93	1,00	1,12	1,35	2,05
Inntekt.....	1,00	1,04	1,10	1,22	1,59
Antall personer .....	0,97	1,01	1,08	1,21	1,60
Antall familier .....	1,06	1,19	1,38	1,76	2,93
Hustype					
Våningshus .....	0,91	1,00	1,13	1,40	2,20
Frittliggende enebolig .....	0,90	0,97	1,08	1,30	1,95
Rekkehus .....	0,97	1,03	1,10	1,26	1,73
Horisontal tomannsbolig .....	0,99	1,01	1,03	1,09	1,26
Annet under 3 etg. ....	0,99	1,01	1,04	1,10	1,27
Blokk, leiegård .....	0,71	0,77	0,87	1,07	1,65
Forretningsbygg .....	1,00	1,00	1,01	1,02	1,07
Hotell, aldershjem .....	1,00	1,01	1,02	1,04	1,10

Tabellene illustrerer til fulle at en ikke kan snakke om én designeffekt, men må spesifisere variabel og utvalgsstørrelse. Tallene varierer fra snaut 0,8 til nesten 11,0. Her må dog innskytes at utvalgsstørrelsene 30 000 (15 000) er tatt med pga. diskusjonen med utvidelse av AKU. I de undersøkelser som er gjennomført har tallet på personer i utvalget ligget mellom 1 000 og 12 000. For små undersøkelser finnes det variable der Byråets utvalgsplan er mer effektiv enn tilfeldig utvalg, men disse er i absolutt mindretall. Som regel er designeffekten nær 1,0 for de små utvalg. Vi ser også at virkningene av økende utvalgsstørrelse er vidt forskjellig for de ulike variable.

Det finnes ingen absolutt målestokk for å avgjøre hvilke designeffekter som er "store". Men for å ha et holdepunkt kan en velge 1,5 som brukes i Byråets publisering. Siden 12 000 personer er en øvre grense for de utvalg som faktisk er trukket, vil vi begrense oss til utvalg mindre eller lik 12 000. Designeffekter av en viss størrelse har vi for noen leilighetsvariable (bad, wc, telefon) og husvariable (hustype, antall rom). Videre for "inntekt av eget arbeid" og framfor alt for andelen som bor i tettbygd område. For personer 16-74 år er det andel yrkesaktive, andel yrke i primærnæringene og spesielt enkelte næringsgrupper som skiller seg ut. Når husholdningen er enhet i utvalget, er det bare for "antall familier" at designeffekten er større enn 1,5 for utvalg mindre eller lik 6 000 enheter.

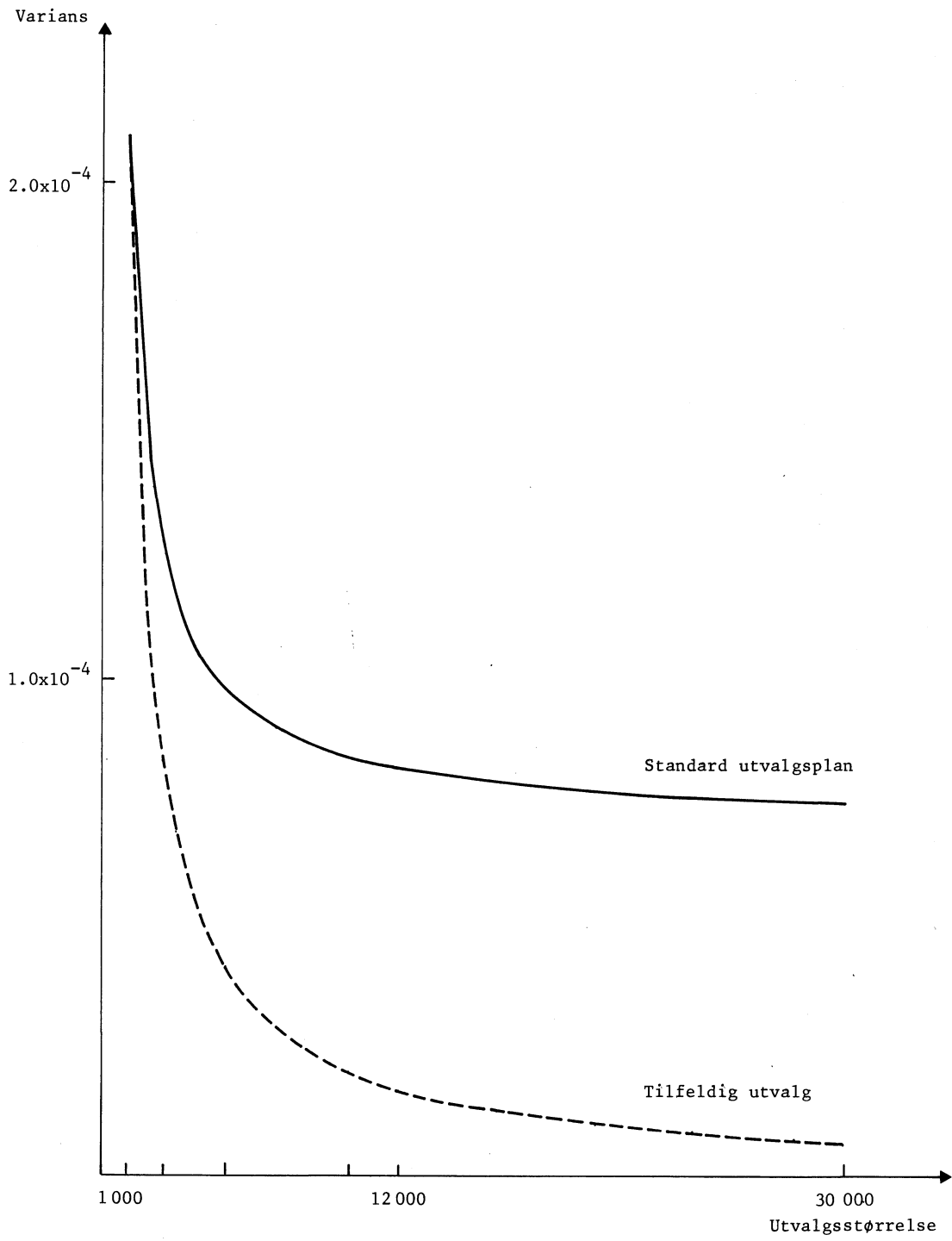
Av variablene er det andelen bosatt i tettbygd strøk som har en spesielt stor designeffekt. Denne er så stor at det er naturlig å forvente en viss designeffekt for andre variable som har en viss sammenheng med denne. Hustype og antall rom er slike variable.

Sæbø (1976) har også beregnet designeffekter for næringer. Han brukte data fra folketellingen i 1970 og kunne forutsette den forenklingen at trekkesannsynlighetene på første trinn var proporsjonale med befolkningstallene. Mine resultater for næring er i meget bra samsvar med hans. Det viser bl.a. at de strukturelle endringer i næring fra 1970 til 1980 ikke har påvirket egenskapene til utvalgsplanen når det gjelder å anslå andelen i ulike næringer.

Når en diskuterer å øke utvalget for en bestemt utvalgsplan, er det visse begrensninger i tolkingen av designeffekt som jeg vil nevne. Hvis en for utvalgsstørrelse  $n$  har en designeffekt  $d$ , så kan det utlegges som at et tilfeldig utvalg av størrelse  $n/d$  gir samme presisjon som et utvalg på  $n$  fra utvalgsplanen. Men en kan ikke si at et utvalg på størrelse  $n \cdot d$  fra utvalgsplanen gir samme presisjon som et tilfeldig utvalg på  $n$ .

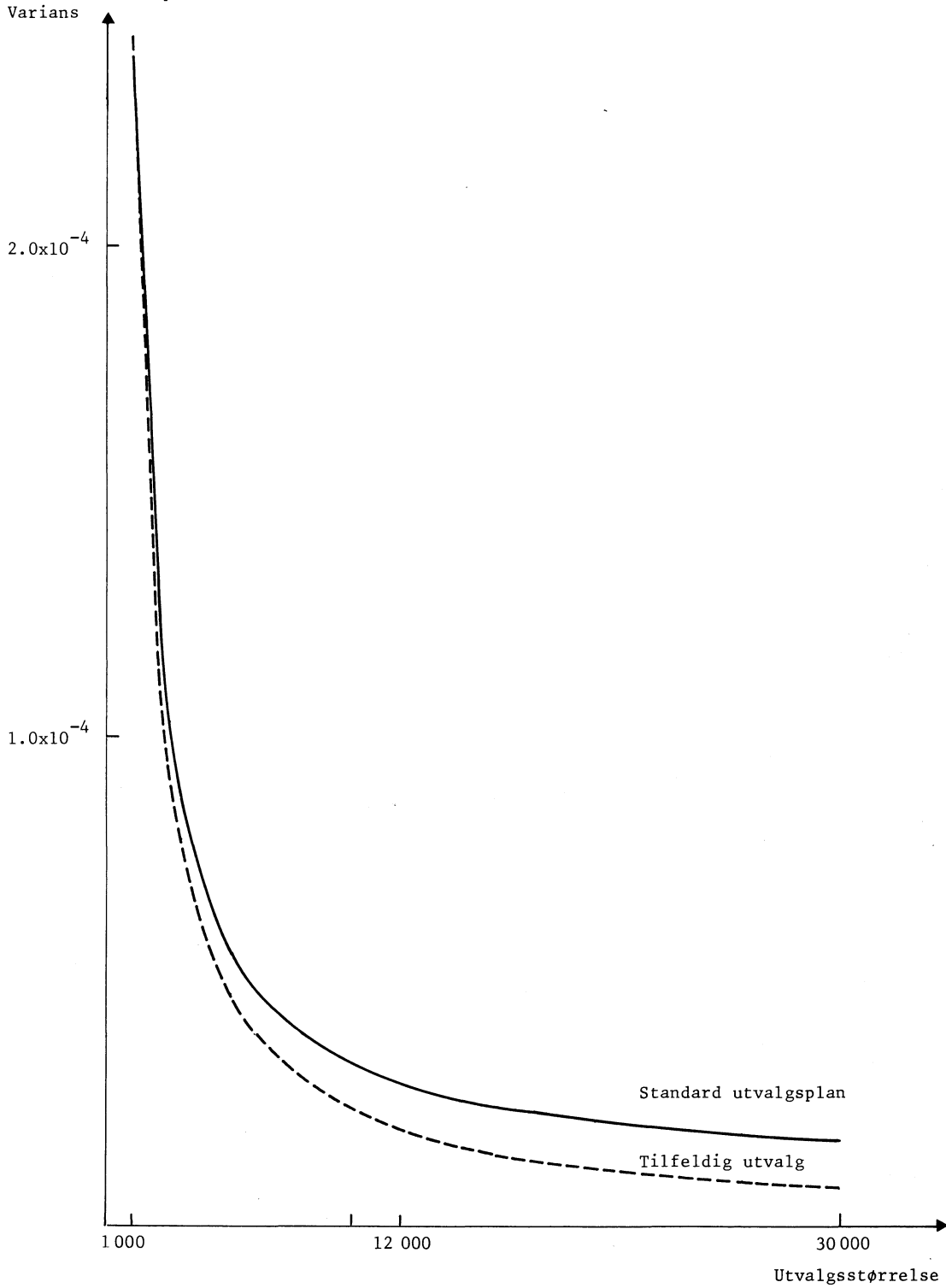
Jeg vil illustrere dette poenget nærmere ved å skissere hvordan variansen for Byråets utvalgsplan og tilfeldig utvalg avhenger av utvalgsstørrelse for tre variable.

Figur 1. Varians ved standard utvalgsplan og tilfeldig utvalg. Andel bosatt tettbygd område

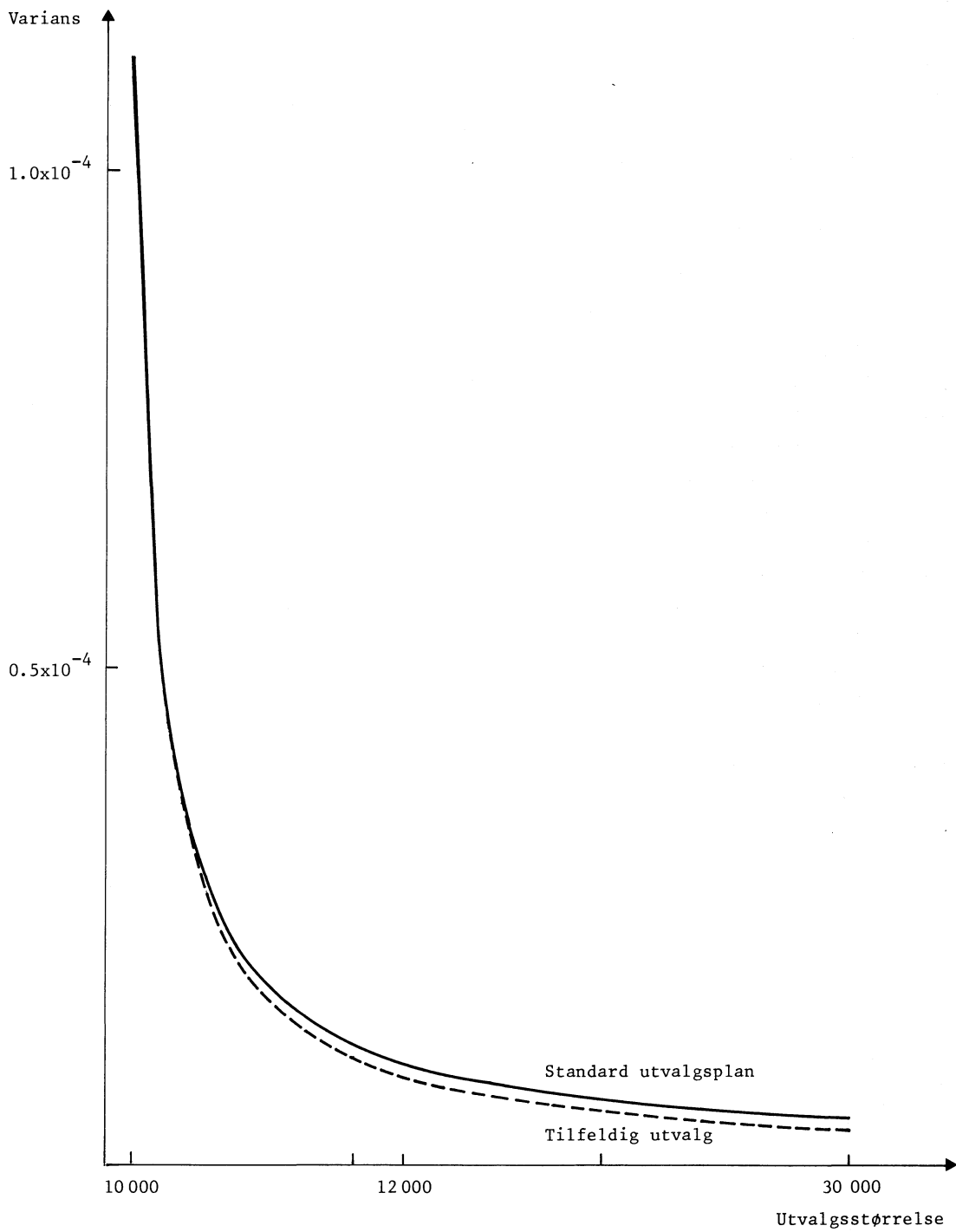




Figur 2. Varians ved standard utvalgsplan og tilfeldig utvalg. Andel med yrkesstatus ansatt blant personer 16-74 år



Figur 3. Varians ved standard utvalgsplan og tilfeldig utvalg. Andel i varehandel, hotell- og restaurantvirksomhet blant personer 16-74 år



Den første av andelen bosatt i tettbygd område blant alle personer, de to andre gjelder personer 16-74 år og er andelen som har yrkesstatus "ansatt" og andelen som arbeider i "Varehandel, hotell- og restaurantvirksomhet", næring 6. Det karakteristiske er at i det aktuelle området (<30 000 observasjoner) avtar variansen til tilfeldige utvalg raskere enn variansen til utvalgsplanen for økende utvalg. For tilfeldige utvalg er variansen direkte omvendt proporsjonal med utvalgsstørrelsen. Fordobling av utvalget gir halvering av variansen. For økning i utvalget fra 12 000 til 30 000 får vi en varians som er  $(12/30) \times 100 = 40$  prosent av den opprinnelige varians. For utvalgsplanen er de tilsvarende tall 92,61 og 48 prosent for de tre variablene. Heldigvis representerer andelen bosatt i tettbygd område det ekstreme. For de fleste variable vil resultatet ligge mellom de to andre. Men disse tallene viser også at en investering i større utvalg for AKU, bør kombineres med justering av utvalgsplanen for å oppnå maksimal effekt. Jeg vil komme nærmere inn på hvordan en bør justere i slutten av neste kapittel.

### 2.3. Stratifiseringseffekt

Dette er den første av deeffektene i den multiplikative dekomponeringen av designeffekten som jeg viste i underkapittel 2.1. Den måles som forholdet mellom varians til stratifisert utvalg med proposjonal allokering og tilfeldig utvalg fra hele landet. Strata er som i utvalgsplanen. Forholdet avhenger ikke av utvalgsstørrelsen.

Resultatene er avhengige av mitt valg av dekomponering. Denne er tiltalende fordi den kan knyttes til veldefinerte og rimelige utvalgsplaner. Dekomponeringen gir absolutt innsikt, men denne kan selvsagt kompletteres med beregninger for andre dekomponeringer.

I SØS nr. 33 er det redegjort for alle de hensyn som spilte inn da en lagde stratifiseringen. Her skal vi se på en spesiell side, hva den betyr for variansen til anslaget for ulike størrelser. Fra teorien for utvalgsundersøkelser vet vi at en ikke kan tape på å stratifisere hvis en fordeler utvalget proporsjonalt med befolkningen i de enkelte strata. Vi skal se på gevinsten en får.

Tabell 4. Forholdet mellom variansen til stratifisert og tilfeldig utvalg. Alle personer

---

Bor i leilighet med	
Bad .....	0,99
WC .....	0,99
Telefon .....	0,98
Pipe .....	0,88
Kjønn .....	1,00
Yrkesaktiv .....	1,00
Bor i tettbygd område .....	0,69
Viktigste kilde til livsopphold	
Inntekt av eget arbeid .....	0,99
Pensjon, trygd .....	0,99
Stipend, lån, formue .....	1,00
Verneplikt/siviltjeneste .....	1,00
Udefinert .....	1,00
Hustype	
Våningshus .....	0,88
Frittliggende enebolig .....	0,88
Rekkehus .....	0,94
Horisontal tomannsbolig .....	0,98
Annet under 3 etg. ....	0,99
Blokk, leiegård .....	0,71
Forretningsbygg .....	1,00
Hotell, aldershjem .....	1,00
Udefinert .....	0,99
Alder .....	0,99
Antall rom .....	0,99

---

Tabell 5. Forholdet mellom variansen til stratifisert og tilfeldig utvalg. Personer 16-74 år

Næring		
1	Jordbruk, skogbruk, fiske og fangst ..	0,94
	11 Jordbruk .....	0,94
	12 Skogbruk .....	0,99
	13 Fiske og fangst .....	0,95
2	Oljeutvinning og bergverksdrift .....	0,98
3	Industri .....	0,98
	31 Produksjon av næringsmidler, drikkevarer og tobakksvarer .....	0,99
	32 Produksjon av tekstilvarer, bekledningsvarer, lær og lærvarer .....	0,99
	33 Produksjon av trevarer .....	0,98
	34 Treforedling, grafisk produksjon og forlagsvirksomhet .....	0,99
	35 Produksjon av kjemiske produkter, mineralolje-, kull-, gummi- og plastprodukter .....	0,98
	36 Produksjon av mineralske produkter	0,99
	37 Produksjon av metaller .....	0,95
	38 Produksjon av verkstedprodukter ...	0,99
4	Kraft- og vannforsyning .....	1,00
5	Bygge- og anleggsvirksomhet .....	1,00
6	Varehandel, hotell- og restaurantvirksomhet .....	0,99
7	Transport, lagring, post og telekommunikasjoner .....	1,00
8	Bank- og finansieringsvirksomhet, forsikringsvirksomhet, eiendomsdrift og forretningsmessige tjenesteyting .....	0,98
9	Offentlig, sosial og privat tjenesteyting .....	0,99

Jeg har bare tatt med to tabeller med variable der det er en viss gevinst. For de andre variable vi brukte fra folketellingen var det ingen gevinst dvs. forholdet var svært nær 1,00. Det er bare i forbindelse med dimensjonene tettbygd - spredtbygd at det er en effekt å spore. Jeg tror at årsaken til denne bl.a. er de 24 strata med én kommune. Disse har en relativt stor andel bosatt i tettbygd område. De gir en grovsortering av landet som nettopp er hensikten med å stratifisere; å samle enhetene i grupper som er mer homogene enn landet som helhet.

Stratifiseringseffektene kan synes små i forhold til hva en er vant med å oppnå i forbindelse med gjennomføring av utvalgsundersøkelser. Da finner en ofte effekter i området 0,7 - 0,9. Jeg tror det har sin årsak i at en har gått veien om de overordnede enheter kommuner/kommunegrupper ved stratifiseringen. Det er kjennetegn ved disse som er brukt for å lage strata, av personer/husholdninger. En må også ta i betraktning at da en samlet kommuner i strata, hadde en også krav om en viss geografisk nærhet. Det foreligger ikke beregninger av stratifiseringseffekten med 1970-tall, men det er neppe forandringer i 10-årsperioden som har redusert en opprinnelig gevinst.

For å ha en annen stratifisering å sammenligne med, delte jeg personer i alder 16-69 år i 16 grupper mht. kjønn og alder. Med denne stratifiseringen har vi en effekt på om lag 0,83 når vi estimerer andelen som er yrkesaktive. Med stratifiseringen i utvalgsplanen var det ingen effekt dvs. forholdstallet var lik 1,00. Tilsvarende tall ville vi trolig få for mange sysselsettingsvariable.

Noen vil være forbauset over de små effekter som gjelder for næring og næringsområder, da nettopp næringen er med i grunnlaget for stratifiseringen. Men en må huske at næringen ble brukt på en grov og indirekte måte via kommunetypeklassifiseringen og at det dermed ikke er gitt at det vil gi noen effekt på estimerer for detaljerte næringsvariable.

#### 2.4. Effekt av to-trinnstrekking

Trekkingen av enheter gjøres i de fleste strata i to trinn. Dette øker variansen og det er denne økningen som mange tenker på når de snakker om designeffekt. Avgjørende for hvor stor økningen blir, er fordelingen av det aktuelle kjennetegn på de områder som kan trekkes i første trinn. Hvis det er stor variasjon i gjennomsnittet for områdene, vil det gi en stor økning i variansen. Denne variasjonen vil være til stede uansett hvor stort utvalget gjøres. Ved store utvalg vil den være den dominerende del av variansen.

I tabellene gir jeg tall for to-trinnseffekten på estimatet for hele landet. Siden 40 prosent av utvalget trekkes i ett trinn, vil effekten av to-trinnstrekkingen være større enn tabellene viser i de strata der det trekkes i to trinn. Dette kan ha betydning for regional estimater fra utvalgsplanen.

Tabell 6. To-trinnseffekt for utvalg av ulike størrelser. Alle personer

Variabel	Utvalgsstørrelse				
	1 000	2 500	5 000	12 000	30 000
Bor i leilighet med					
Bad .....	1,02	1,04	1,09	1,21	1,53
WC .....	1,03	1,07	1,14	1,35	1,87
Telefon .....	1,04	1,10	1,22	1,53	2,33
Kjønn .....	1,00	1,00	1,00	1,00	1,01
Yrkesaktiv .....	1,01	1,02	1,03	1,08	1,19
Bor i tettbygd område .....	1,40	2,10	3,25	6,49	14,88
Viktigste kilde til livsopphold					
Inntekt av arbeid .....	1,01	1,04	1,08	1,20	1,50
Pensjon, trygd .....	1,01	1,04	1,08	1,20	1,51
Stipend, lån, formue .....	1,00	1,00	1,01	1,02	1,06
Verneplikt/siviltj. ....	1,00	1,00	1,00	1,01	1,02
Udefinert .....	1,00	1,01	1,01	1,03	1,09
Hustype					
Våningshus .....	1,11	1,29	1,59	2,44	4,65
Frittliggende enebolig .....	1,08	1,22	1,44	2,08	3,73
Rekkehus .....	1,06	1,15	1,32	1,77	2,95
Horisontal tomannsbolig .....	1,02	1,04	1,09	1,22	1,56
Annet under 3 etg. ....	1,01	1,04	1,08	1,19	1,48
Blokk, leiegård .....	1,08	1,21	1,43	2,05	3,66
Forretningsbygg .....	1,00	1,01	1,03	1,07	1,17
Hotell, aldershjem .....	1,00	1,01	1,02	1,05	1,12
Udefinert .....	1,01	1,02	1,04	1,09	1,23
Alder .....	1,02	1,05	1,11	1,27	1,68
Antall rom .....	1,02	1,05	1,11	1,26	1,65

Tabell 7. To-trinnseffekt for utvalg av ulike størrelser. Personer 16 - 74 år

Variabel	Utvalgsstørrelse				
	1 000	2 500	5 000	12 000	30 000
Yrkesaktive .....	1,01	1,03	1,06	1,15	1,39
<b>Ekteskapelig status</b>					
Ugift .....	1,00	1,01	1,02	1,05	1,13
Gift .....	1,01	1,02	1,03	1,07	1,19
Enke/enkemann .....	1,00	1,01	1,01	1,04	1,10
Skilt .....	1,00	1,01	1,02	1,04	1,10
Separert .....	1,00	1,00	1,01	1,02	1,04
<b>Utdanning</b>					
1 Barneskolenivå .....	1,00	1,00	1,00	1,00	1,01
2 Ungdomsskolenivå .....	1,02	1,05	1,11	1,26	1,66
3 Gymnasnivå I .....	1,00	1,01	1,02	1,05	1,12
4 Gymnasnivå II .....	1,01	1,02	1,03	1,08	1,20
5 Universitets- og høghskolenivå I ...	1,00	1,01	1,03	1,07	1,17
6 Universitets- og høghskolenivå II ..	1,00	1,01	1,01	1,03	1,08
7 Universitets- og høghskolenivå III .	1,00	1,01	1,03	1,06	1,16
8 Forskernivå .....	1,00	1,01	1,02	1,04	1,11
9 Uoppgitt .....	1,00	1,00	1,00	1,00	1,00
Udefinert .....	1,00	1,01	1,01	1,04	1,09
<b>Yrkesfelt</b>					
1 Administrasjons- og forvaltnings- arbeid, bedrifts- og organisasjons- ledelse .....	1,00	1,01	1,01	1,04	1,09
2 Kontorarbeid .....	1,00	1,01	1,02	1,06	1,14
3 Handelsarbeid .....	1,00	1,01	1,02	1,05	1,12
4 Jordbruks-, skogsbruks- og fisker- arbeid .....	1,04	1,12	1,24	1,60	2,51
5 Gruve- og sprengingsarbeid m.m. ...	1,03	1,07	1,14	1,35	1,88
6 Transport- og kommunikasjonsarbeid	1,01	1,01	1,03	1,06	1,16
7 Industri-, bygge- og anleggsarbeid	1,01	1,03	1,07	1,17	1,42
8 Industri-, bygge- og anleggsarbeid	1,02	1,04	1,09	1,22	1,56
9 Servicearbeid .....	1,00	1,01	1,02	1,05	1,12
<b>Næring</b>					
1 Jordbruk, skogbruk, fiske og fangst ..	1,05	1,14	1,28	1,69	2,75
11 Jordbruk .....	1,06	1,17	1,35	1,86	3,19
12 Skogbruk .....	1,03	1,08	1,15	1,38	1,96
13 Fiske og fangst .....	1,07	1,18	1,38	1,93	3,36
2 Oljeutvinning og bergverksdrift .....	1,13	1,36	1,74	2,80	5,56
3 Industri .....	1,03	1,08	1,17	1,42	2,08
31 Produksjon av næringsmidler, drikke- varer og tobakksvarer .....	1,04	1,11	1,22	1,54	2,37
32 Produksjon av tekstilvarer, bekled- ningsvarer, lær og lærvarer .....	1,04	1,11	1,23	1,56	2,42
33 Produksjon av trevarer .....	1,08	1,22	1,45	2,11	3,81
34 Treforedling, grafisk produksjon og forlagsvirksomhet .....	1,03	1,09	1,19	1,47	2,19
35 Produksjon av kjemiske produkter, mineralolje-, kull-, gummi- og plastprodukter .....	1,08	1,21	1,43	2,04	3,63
36 Produksjon av mineralske produkter	1,04	1,12	1,25	1,61	2,55
37 Produksjon av metaller .....	1,30	1,81	2,67	5,07	11,30
38 Produksjon av verkstedprodukter ...	1,06	1,16	1,32	1,78	2,97
4 Kraft- og vannforsyning .....	1,01	1,04	1,07	1,18	1,45
5 Bygge- og anleggsvirksomhet .....	1,01	1,03	1,05	1,13	1,32
6 Varehandel, hotell- og restaurant- virksomhet .....	1,01	1,02	1,05	1,12	1,30
7 Transport, lagring, post og tele- kommunikasjoner .....	1,01	1,02	1,05	1,11	1,29
8 Bank- og finansieringsvirksomhet, for- sikringsvirksomhet, eiendomsdrift og forretningsmessig tjenesteyting .....	1,00	1,01	1,02	1,05	1,13
9 Offentlig, sosial og privat tjeneste- yting .....	1,02	1,04	1,08	1,21	1,52

Tabell 7 (forts.). To-trinnseffekt for utvalg av ulik størrelse. Personer 16 - 74 år

Variabel	Utvalgsstørrelse				
	1 000	2 500	5 000	12 000	30 000
<b>Yrkesstatus</b>					
Ansatt .....	1,02	1,04	1,09	1,21	1,54
Selvstendig .....	1,02	1,04	1,09	1,21	1,53
Familiemedlemmer .....	1,01	1,03	1,07	1,17	1,42
Udefinert .....	1,01	1,03	1,07	1,16	1,41
<b>Arbeidstid i året</b>					
100- 499 timer .....	1,00	1,00	1,01	1,02	1,05
500- 999 " .....	1,00	1,00	1,01	1,02	1,04
1 000- 1 299 " .....	1,00	1,00	1,00	1,01	1,02
1 300- .....	1,01	1,02	1,04	1,10	1,26
Udefinert .....	1,01	1,03	1,07	1,17	1,42
<b>Arbeidstid i uka 25.10.-31.10. 1980</b>					
0 timer .....	1,00	1,01	1,01	1,03	1,08
1- 9 timer .....	1,00	1,00	1,00	1,01	1,03
10-19 " .....	1,00	1,00	1,01	1,01	1,03
20-29 " .....	1,00	1,00	1,01	1,02	1,04
30+ " .....	1,01	1,02	1,05	1,11	1,28
Udefinert .....	1,01	1,03	1,06	1,15	1,39
<b>Inntekt .....</b>	1,01	1,02	1,04	1,09	1,22

Tabell 8. To-trinnseffekt for utvalg av ulik størrelse. Husholdninger

Variabel	Utvalgsstørrelse				
	500	1 500	3 000	6 000	15 000
<b>Eier leiligheten .....</b>	1,01	1,03	1,05	1,11	1,28
<b>Bor i leilighet med</b>					
Bad .....	1,01	1,02	1,05	1,10	1,25
WC .....	1,01	1,04	1,07	1,15	1,39
Telefon .....	1,01	1,04	1,09	1,18	1,45
Kjøkken .....	1,00	1,02	1,04	1,08	1,20
Ovn .....	1,02	1,08	1,17	1,35	1,89
<b>Disponerer bil .....</b>	1,01	1,05	1,11	1,22	1,56
<b>Hustype</b>					
Våningshus .....	1,04	1,14	1,29	1,60	2,53
Frittliggende enebolig .....	1,03	1,12	1,25	1,51	2,29
Rekkehus .....	1,02	1,07	1,14	1,28	1,72
Horisontal tomannsbolig .....	1,01	1,02	1,05	1,10	1,26
Annet under 3 etg. ....	1,01	1,02	1,05	1,10	1,26
Blokk, leiegård .....	1,03	1,11	1,24	1,49	2,24
Forretningsbygg .....	1,00	1,01	1,01	1,02	1,06
Hotell, aldershjem .....	1,00	1,01	1,02	1,04	1,09
<b>Antall rom .....</b>	1,02	1,06	1,13	1,26	1,66
<b>Inntekt .....</b>	1,01	1,02	1,05	1,10	1,25
<b>Antall personer .....</b>	1,01	1,04	1,09	1,18	1,46
<b>Antall familier .....</b>	1,00	1,01	1,01	1,03	1,08

To-trinnseffekten er størst for andelen som bor i tettbygd område. For mange strata må det være forholdsvis stor forskjell mellom de primære utvalgsområdene m.h.t. andelen som er bosatt i tettbygd område. Dette gjør at vi også får to-trinnseffekter for variable som har sammenheng med tettbygdvariabelen. Det dreier seg om hustype, yrke og primærnæringene. Effekten er langt større enn den positive effekten stratifiseringen hadde og gir dermed forklaring på den designeffekt som er påvist for disse variable.

Det er også to-trinnseffekt i estimeringen for andre næringer og næringsområder. Her mener jeg forklaringen må være at det er variable med stor geografisk konsentrasjon. Det er næringer som utøves i et fåtall bedrifter/steder. De ansatte vil for en stor del være bosatt i vertskommunen eller nærliggende kommuner. Når en slik kommune kommer i strata sammen med andre kommuner, vil det gi relativ stor variasjon i andelen mellom de primære utvalgsområdene.

## 2.5. Effekt av endret befolknings sammensetning.

Trekkesannsynlighetene på første trinn er proporsjonale med befolkningstallene i 1970. For den enkelte undersøkelse beregnes antallet som trekkes fra utvalgsområdene slik at utvalget blir selvveiende, dvs. alle enheter i den aktuelle befolkning har samme trekkesannsynlighet. Det gjøres ved å justere antallet som trekkes fra det enkelte utvalgsområde slik at produktet av de to trekkesannsynlighetene er likt for alle primære utvalgsområder. Antallet som trekkes fra et stratum blir altså bestemt av utviklingen i utvalgsområdet og ikke av utviklingen i hele stratomet. Bare hvis befolkningen i utvalgsområde har avtatt/økt relativt like mye som befolkningen i hele stratomet, vil stratomet ha en proporsjonal representasjon i totalutvalget.

Nå har en av flere grunner betraktet dette som et lite problem. Det er vesentlige er å ha et selvveiende utvalg. Det letter bearbeidingen og er et greit grunnlag for spesielle estimeringsmetoder. Videre er det i denne sammenheng små forandringer som skjer i befolkningssammensetningen. Vi vet også at proporsjonal allokering av utvalget ikke er optimalt m.h.t. varians. En endring kan derfor slå begge veier. Dette er et resultat fra teorien om stratifisert utvalg der det forutsettes at utvalget trekkes i ett trinn. Forholdet kan være noe annerledes i forbindelse med to-trinnstrekking.

For et stratum vil bidraget til variansen som skyldes variasjon mellom primærområdene kunne skrives

$$\sum_j \Pi_j \left( \frac{C_j}{\Pi_j} \bar{y}_j - \bar{y} \right)^2 = \sum_j \frac{C_j^2}{\Pi_j} \bar{y}_j^2 - \bar{y}^2$$

$\Pi_j$  = Trekkesannsynlighet 1. trinn område nr j

$C_j$  = Aktuell befolkningsandel " " "

$\bar{y}_j$  = Gjennomsnitt for kjennetegn " " "

$\bar{y}$  = " " " " i stratomet

k = Antall primærområder

Vi har  $\sum_j \Pi_j = \sum_j C_j = 1$  og  $\bar{y} = \sum_j C_j \bar{y}_j$

Det som skal sammenlignes er størrelsen av uttrykket for  $C_j = \Pi_j$  for alle j og for forskjell mellom andelene. Av Cauchy-Schwarz ulikhet og betingelsene har vi

$$\frac{C_1^2}{\Pi_1} + \frac{C_2^2}{\Pi_2} + \dots + \frac{C_k^2}{\Pi_k} \geq \frac{(C_1 + C_2 + \dots + C_k)^2}{\Pi_1 + \Pi_2 + \dots + \Pi_k} = 1 = C_1 + C_2 + \dots + C_k$$



Det viser at summen av koeffisientene foran de gitte, positive  $\bar{y}^2$ -ene er minst når alle  $C_j$  er lik  $\pi_j$ . Det betyr ikke at varianskomponenten alltid er minst ved likhet, men er en indikasjon på at det er slik i mange tilfeller.

Jeg vil derfor undersøke forholdene for en del variable. Effekten beregnes som forholdet mellom varians som den er i utvalgsplanen i dag og hva den hadde vært hvis trekk sannsynlighetene på første trinn var beregnet ved hjelp av de aktuelle befolkningstall.

Effekten er slik at uansett hvilken retning den går, så forsterkes den ved økende utvalgsstørrelse.

Som for to-trinnseffekten gir jeg tall for effekten på estimatet for hele landet. Effekten "dempes" av de 40 prosent av utvalget der det trekkes i ett trinn. Den kan ha større betydning for enkelte regionale estimater.

Tabell 9. Effekt av endret befolkningssammensetning for utvalg av ulike størrelser. Alle personer

Variabel	Utvalgsstørrelse				
	1 000	2 500	5 000	12 000	30 000
Bor i leilighet med					
Bad .....	1,13	1,32	1,61	2,31	3,60
WC .....	1,12	1,29	1,54	2,12	3,01
Telefon .....	1,02	1,04	1,08	1,16	1,25
Kjønn .....	1,02	1,04	1,08	1,18	1,46
Yrkesaktiv .....	1,01	1,03	1,06	1,14	1,33
Bor i tettbygd område .....	1,02	1,04	1,05	1,06	1,07
Viktigste kilde til livsopphold					
Inntekt av arbeid .....	1,07	1,18	1,34	1,75	2,49
Pensjon, trygd .....	1,00	0,99	0,98	0,95	0,90
Stipend, lån, formue e.l. ....	1,00	1,00	1,00	1,00	0,99
Verneplikt/siviltjeneste .....	1,00	1,00	1,00	1,00	1,00
Udefinert .....	1,00	1,00	1,00	1,00	0,99
Hustype					
Våningshus .....	1,00	1,01	1,01	1,02	1,02
Frittliggende enebolig .....	1,04	1,09	1,15	1,26	1,36
Rekkehus .....	1,00	1,00	1,00	1,01	1,01
Horisontal tomannsbolig .....	1,00	1,00	0,99	0,98	0,97
Annet under 3 etg. ....	1,00	1,00	0,99	0,98	0,96
Blokk, leiegård .....	1,00	1,00	1,01	1,01	1,02
Forretningsbygg .....	1,00	1,00	1,00	0,99	0,99
Hotell, aldershjem .....	1,00	1,00	1,00	1,00	0,99
Udefinert .....	1,00	1,00	1,00	1,00	0,99
Alder .....	1,00	0,99	0,99	0,97	0,95
Antall rom .....	1,02	1,05	1,10	1,21	1,41

Tabell 10. Effekt av endret befolknings sammensetning for utvalg av ulike størrelser. Personer 16-74 år

Variabel	Utvalgsstørrelse				
	1 000	2 500	5 000	12 000	30 000
<b>Næring</b>					
1 Jordbruk, skogbruk, fiske og fangst ..	1,00	1,00	1,00	0,99	0,99
11 Jordbruk .....	1,00	1,00	1,00	1,00	1,00
12 Skogbruk .....	1,00	1,00	1,00	1,00	0,99
13 Fiske og fangst .....	1,00	0,99	0,99	0,98	0,97
2 Oljeutvinning og bergverksdrift .....	1,00	0,99	0,99	0,98	0,98
<b>3 Industri</b>					
31 Produksjon av næringsmidler, drikkevarer, og tobakksvarer .....	1,00	1,01	1,01	1,02	1,03
32 Produksjon av tekstilvarer, bekledningsvarer, lær og lærvarer .....	1,00	1,00	0,99	0,98	0,98
33 Produksjon av trevarer .....	1,00	0,99	0,99	0,98	0,97
34 Treforedling, grafisk produksjon og forlagsvirksomhet .....	1,00	1,01	1,01	1,02	1,03
35 Produksjon av kjemiske produkter, mineralolje-, kull-, gummi- og plastprodukter .....	1,00	1,00	1,00	1,00	1,00
36 Produksjon av mineralske produkter .	1,00	1,01	1,02	1,02	1,03
37 Produksjon av metaller .....	1,00	1,00	1,00	1,00	1,01
38 Produksjon av verkstedprodukter ....	0,99	0,97	0,97	0,96	0,95
4 Kraft- og vannforsyning .....	1,00	1,00	1,00	1,00	1,00
5 Bygge- og anleggsvirksomhet .....	1,00	1,01	1,01	1,02	1,05
6 Varehandel, hotell- og restaurantvirksomhet .....	1,00	1,01	1,01	1,03	1,07
7 Transport, lagring, post og telekommunikasjoner .....	1,00	1,00	1,01	1,01	1,03
8 Bank- og finansieringsvirksomhet, forsikringsvirksomhet, eiendomsdrift og forretningsmessige tjenesteyting .....	1,00	1,00	1,00	1,01	1,02
9 Offentlig, sosial og privat tjenesteyting .....	1,00	1,00	1,01	1,02	1,04
<b>Yrkesstatus</b>					
Ansatt .....	1,03	1,06	1,12	1,25	1,50
Selvstendig .....	1,00	1,00	1,00	1,00	1,00
Familiemedlem .....	1,00	1,00	1,00	1,00	1,00
Udefinert .....	1,00	1,00	1,00	0,99	0,99
<b>Arbeidstid i året</b>					
100 - 499 timer .....	1,00	1,00	1,00	1,01	1,02
500 - 999 " .....	1,00	1,00	1,01	1,01	1,04
1 000 - 1 299 " .....	1,00	1,00	1,00	1,01	1,02
1 300 - " .....	1,01	1,03	1,06	1,14	1,32
Udefinert .....	1,00	1,00	1,00	1,00	0,99
<b>Arbeidstid i uka 25.10. - 31.10. 1980</b>					
0 timer .....	1,00	1,00	1,00	1,01	1,01
1 - 9 " .....	1,00	1,00	1,00	1,00	1,01
10 - 19 " .....	1,00	1,00	1,00	1,01	1,02
20 - 29 " .....	1,00	1,00	1,00	1,01	1,02
30 + " .....	1,01	1,03	1,07	1,15	1,33
Udefinert .....	1,00	1,00	1,00	1,00	0,99
Inntekt .....	1,03	1,06	1,13	1,29	1,64

Tabell 11. Effekt av endret befolningssammensetning for utvalg av ulike størrelser. Husholdninger

Variabel	Utvalgsstørrelse				
	500	1 500	3 000	6 000	15 000
Eier leiligheten .....	1,01	1,03	1,06	1,11	1,25
Bor i leilighet med					
Bad .....	1,03	1,08	1,15	1,28	1,62
Wc .....	1,03	1,08	1,15	1,29	1,60
Telefon .....	1,01	1,02	1,05	1,09	1,18
Kjøkken .....	1,01	1,04	1,08	1,16	1,36
Øvn .....	1,00	1,00	1,01	1,01	1,02
Disponerer bil .....	1,01	1,02	1,04	1,08	1,16
Antall rom .....	1,02	1,06	1,11	1,20	1,38
Inntekt .....	1,01	1,04	1,07	1,13	1,29
Antall personer .....	1,01	1,02	1,04	1,07	1,15
Antall familier .....	1,06	1,18	1,36	1,72	2,73

Tabellene viser at det er i få tilfelle denne effekten er av nevneverdig størrelse. Men for noen variable viser det seg at det er denne effekten, og ikke to-trinnseffekten, som er årsak til at vi hadde en designeffekt. For utvalg blant alle personer gjelder det andel som bor i leilighet med bad og ditto med WC, og det gjelder andel som har inntekt av arbeid som viktigste kilde til livsopphold. For utvalg av husholdninger er det tegn til effekt for kjennetegnet antall familier.

Hvilken konstellasjon som gjør at det blir effekt for akkurat disse variable, har jeg ikke undersøkt i detalj. Men det er nærliggende å peke på at det dreier seg om variable der neppe variasjonen mellom primærområdene er særlig stor. Prøveregninger viser at det i slike tilfelle selv for små befolkningsendringer kan bli store utslag i "mellom"-komponenten i variansen.

Foran i dette kapitlet har jeg betraktet befolkningsendringene fra 1970 til 1980 som årsak til effekten. Strengt tatt vil en ha en variant av problemet fra første stund av. Det er fordi samme utvalgsplan brukes til å trekke utvalg av ulike enheter (familier ved siden av personer) og ulike avgrensinger (16-74 år o.l.). Førstetrinnssannsynlighetene beregnes på grunnlag av alle personer i 1970. Fordelingen av disse var ikke identisk med fordelingen av familier, personer 16-74 år selv i 1970. Dette er en bagatell som med rette har blitt underslått i en del sammenhenger.

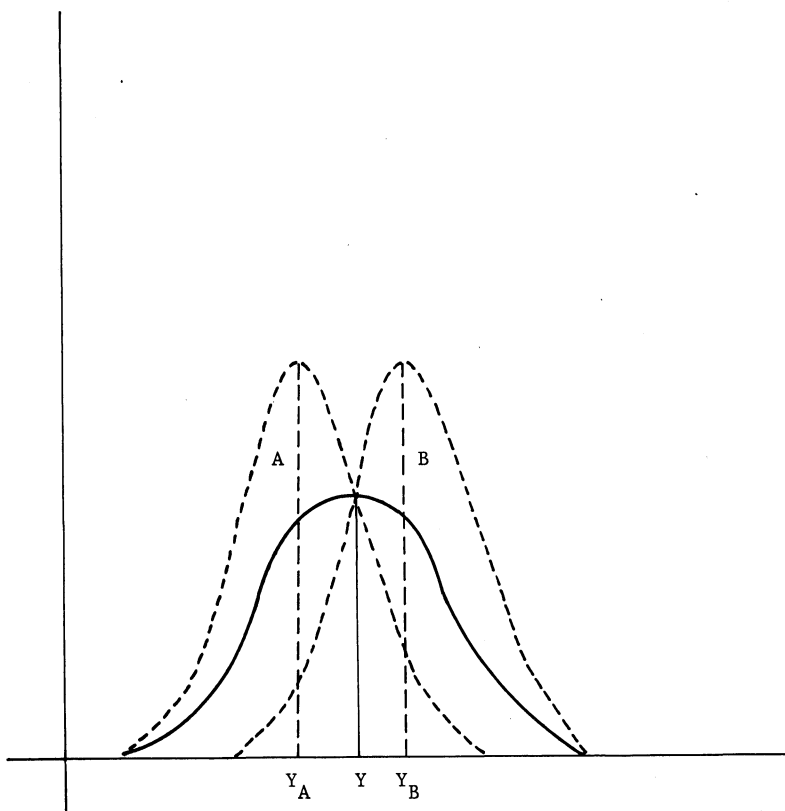
### 3. BETINGET FORVENTNING GITT TREKKINGEN PÅ FØRSTE TRINN

#### 3.1. Om usikkerhetsbegrepet

Med en utvalgsundersøkelse vil en ved hjelp av kunnskap om et begrenset antall enheter, et utvalg, anslå størrelsen på en populasjonsparameter. Denne er knyttet til hele populasjonen og vil kunne beregnes eksakt hvis en hadde kunnskap om alle enhetene i populasjonen. Siden utvalget er begrenset, vil det være usikkerhet i anslaget. For å illustrere denne usikkerheten bruker en utvalgsvarians. Denne er spredningen i de anslag en ville fått hvis en trakk mange uavhengige utvalg og hele tiden brukte samme metode for trekking og estimering. Utvalgsvariansen gir en indirekte beskrivelse av godheten til resultatet fra én undersøkelse ved å karakterisere metoden som brukes.

Når utvalget trekkes i to trinn, vil utvalgsvariansen reflektere usikkerheten som ligger i begge trekkingene. Men hvis en kjenner resultatet av første trinns trekking og har summariske opplysninger om gruppene i første trinns trekking, kan en illustrere usikkerheten i et enkelt anslag på en måte som ligger nærmere det en virkelig vil vite; hvor langt anslaget ligger fra populasjonsparameteren. Dette forklares best ved et forenklet eksempel. Vi har en populasjon som består av to like store deler A og B. Vi trekker et utvalg ved først å trekke A eller B, hver med sannsynlighet 0,5, så trekkes et utvalg personer fra den del av populasjonene som ble trukket i første trinn. Dernest estimeres en størrelse. I figur 4 har jeg med hel strek tegnet fordelingen av estimater vi får for to-trinnsprosedyren, og med stiplede strek relativ fordeling av estimatene for utvalg fra hhv. A og B.

Figur 4. Betinget og ubetinget fordeling for et eksempel på to-trinns trekking



$Y$  er populasjonsparameteren vi skal estimere. Siden forventningen i fordelingen med hel strek er lik  $y$ , har jeg forutsatt en forventningsrett estimator.  $Y_A$  og  $Y_B$  er forventningene i fordelingene for estimater fra hhv. A og B. Hvis vi trekker A i første trinn, vil estimatet vi får stamme fra den stiplede A-fordelingen. Når vi vet at estimatet stammer fra A-fordelingen, er det mer informativt å få vite  $Y_A - Y$  og spredning i A-fordelingen enn spredning i den heltrukne fordelingen. De første opplysningene sier oss mest hvor godt estimatet er.  $Y_A$  er betinget forventning gitt at A trekkes i første trinn og  $Y$  er ubetinget forventning.

Når A holdes fast over tid og forholdene er relativt stabile, vil gjentatte undersøkelser gi estimater som alle stammer fra A-fordelingen. Det er viktig å være klar over at disse ikke er sentrert om  $Y$  og at deres variasjon ikke illustrerer spredningen i den heltrukne fordelingen.

### 3.2. Avvik som en del av usikkerheten

Eksemplet gir et forenklet bilde av forholdene ved Byråets utvalgsplan. I denne er det på første trinn ikke to mulige fordelinger, men svært mange. Det blir én fordeling for hvert mulig resultat når det trekkes ett primærområde blant to eller flere i hvert av 78 (= 102 - 24) strata. Videre trekkes det ikke slik at alle fordelingene er like sannsynlige. Men prinsippet er det samme, estimatet vi får stammer fra én av de mange fordelingene. Det er selvsagt meget viktig hvordan denne fordelingen ligger i forhold til  $Y$ . Det vil vi se på ved å beregne differansen mellom betinget forventning gitt første trinns trekking og ubetinget forventning.

Siden vi hele tiden ser på gjennomsnittet som estimator, vil disse forventningene være det veiede gjennomsnitt i utvalgsområdene og gjennomsnittet i hele populasjonen. Det er differansen mellom disse som presenteres i tabellene. De er beregnet ut fra situasjonen et spesielt tidspunkt, november 1980. For de fleste variable har det vært forandring i den tid utvalgsplanen har vært i bruk. På den annen side er det for mange størrelser såpass liten relativ endring i løpet av 10 år at jeg tror resultatene gir et godt bilde for hele 10-årsperioden.

Siden 40 prosent av befolkningen bor i strata med bare én kommune, skulle en tro at avviket ble nokså lite. For disse strata gir ikke noe bidrag til avviket ettersom den betingete og ubetingete forventningen er like store.

Tabellene som følger inneholder ikke et fullstendig eller "tilfeldig" utvalg med variable. Med bl.a. AKU i tankene har jeg plukket fram variable der avviket er av en viss størrelse. Det gir grunnlag for å gjette på andre variable som ikke finnes i folketellingen, der avviket kan være stort.

Avvik er differans mellom betinget og ubetinget forventning. Positive avvik tyder på "overrepresentasjon" mens negative avvik tyder på det motsatte.

Tabell 12. Avvik mellom utvalgsområdene og hele landet. Personer 16-74 år

Variabel	Relativt avvik i prosent	Avvik Promille	Andel
Tettbygd område .....	2,3	16,5	709
<b>Utdanning</b>			
2 Ungdomsskolenivå .....	1,2	-6,5	536
3 Gymnasnivå I .....	1,3	2,4	190
4 Gymnasnivå II .....	1,1	1,5	141
5 Universitets- og høgsolenivå I .....	2,4	1,6	68
6 Universitets- og høgsolenivå II .....	0,4	-0,1	23
7 Universitets- og høgsolenivå III .....	2,5	0,6	23
8 Forskernivå .....	1,4	-0,0	1
<b>Yrkesfelt</b>			
1 Administrasjons- og forvaltningsarbeid, bedrifts- og organisasjonsledelse .....	0,6	0,2	33
2 Kontorarbeid .....	1,3	0,9	68
3 Handelsarbeid .....	1,4	0,9	64
4 Jordbruks-, skogsbruks- og fiskearbeid .....	4,2	-2,1	49
5 Gruve- og sprengingsarbeid m.m. ....	5,7	0,2	3
6 Transport- og kommunikasjonsarbeid .....	0,1	-0,1	54
7 Industri-, bygge- og anleggsarbeid .....	1,6	1,9	117
8 Industri-, bygge- og anleggsarbeid .....	2,9	-1,5	53
9 Servicearbeid .....	0,2	-0,2	85
<b>Næring</b>			
1 Jordbruk, skogbruk, fiske og fangst .....	4,4	-2,5	56
11 Jordbruk .....	1,1	-0,5	45
12 Skogbruk .....	15,1	-0,4	3
13 Fiske og fangst .....	20,2	-1,5	7
2 Oljeutvinning og bergverksdrift .....	16,2	0,6	4
3 Industri .....	2,1	2,9	143
31 Produksjon av næringsmidler, drikkevarer og tobakksvarer .....	6,1	-1,5	25
32 Produksjon av tekstilvarer, bekledningsvarer, lær og lærvarer .....	10,7	1,0	9
33 Produksjon av trevarer .....	18,7	2,5	14
34 Treforedling, grafisk produksjon og forlags- virksomhet .....	2,3	-0,4	18
35 Produksjon av kjemiske produkter, mineralolje-, kull-, gummi- og plastprodukter .....	0,7	0,1	12
36 Produksjon av mineralske produkter .....	7,0	0,4	5
37 Produksjon av metaller .....	11,4	-1,2	11
38 Produksjon av verkstedprodukter .....	4,3	2,1	48
4 Kraft- og vannforsyning .....	2,7	-0,2	6
5 Bygge- og anleggsvirksomhet .....	0,8	-0,4	55
6 Varehandel, hotell- og restaurantvirksomhet .....	0,7	0,8	125
61, 62 Engros- og agenturhandel, detaljhandel .....	0,8	0,9	107
63 Hotell- og restaurantdrift .....	0,4	-0,1	17
7 Transport, lagring, post og telekommunikasjoner ...	0,2	0,1	62
71 Transport og lagring .....	0,5	0,2	45
72 Post og telekommunikasjoner .....	0,5	-0,1	17

Tabell 12 (forts.). Avvik mellom utvalgsområdene og hele landet. Personer 16-74 år

Variabel	Relativt avvik i prosent	Avvik	
		Promille	Andel
8 Bank, og finansieringsvirksomhet, forsikringsvirksomhet, eiendomsdrift og forretningsmessig tjenesteyting .....	1,0	0,4	36
81 Bank- og finansieringsvirksomhet .....	1,0	0,1	13
82 Forsikringsvirksomhet .....	1,4	0,1	4
83 Eiendomsdrift og forretningsmessig tjenesteyting .....	0,9	0,2	19
9 Offentlig, sosial og privat tjenesteyting .....	0,1	0,2	215
91 Offentlig administrasjon, forsvar, politi og rettsvesen .....	3,9	-2,0	52
92 Renovasjon og reingjøring .....	2,0	0,1	4
93 Undervisning, helse- og andre sosialtjenester ..	1,4	1,9	136
931 og 932 Undervisning m.m. ....	0,2	-0,1	48
933 Helse- og veterinærtjenester .....	4,1	2,5	61
934 Sosial omsorg og velferdsarbeid .....	1,8	-0,4	23
94 Kulturell tjenesteyting, underholdning og sport ..	1,7	-0,1	6
95 Personlig tjenesteyting .....	1,6	0,3	17
Yrkesstatus			
Ansatt .....	0,5	3,2	612
Selvstendig .....	2,3	-1,4	61
Familiemedlem .....	1,0	0,2	15
Udefinert .....	0,6	-2,0	312

Tabell 13. Avvik mellom utvalgsområdene og hele landet. Husholdninger

Variabel	Relativt avvik i prosent	Avvik	
		Promille	Andel
Tettbygd strøk .....	2,4	17,6	724
Hustype			
Våningshus .....	4,5	-4,9	109
Frittliggende enebolig .....	0,9	-0,4	419
Rekkehus .....	4,9	5,3	107
Horisontal tomannsbolig .....	2,9	1,4	47
Annet under 3 etg. ....	1,4	0,6	40
Blokk, leiegård .....	0,5	0,9	179
Forretningsbygg .....	5,2	0,6	12
Hotell, aldershjem .....	0,8	-0,0	5
Udefinert .....	0,1	-0,0	81
Tallet på familier			
1 .....	0,1	-1,2	920
2 .....	1,4	1,0	73
3 .....	2,4	0,2	7
4 + .....	7,9	0,0	0,5
Sosioøkonomisk status			
A 11 .....	0,2	0,4	209
A 12 .....	1,5	1,0	67
A 13 .....	0,9	1,8	215
A 21 .....	3,6	-2,7	76
A 31 .....	1,8	0,2	9
A 32 .....	0,2	-0,5	298
Udefinert .....	0,2	-0,1	39

For den dikotome variabelen bosatt i tettbygd område ser vi det er et betydelig avvik. Det veide gjennomsnitt i utvalgsområdene er 16,5 promille større enn gjennomsnittet i hele landet. Det betyr at når vi bruker standard utvalgsplan til å anslå denne andelen, vil anslagene variere rundt 72,5 prosent istedenfor 70,9. Dette er en nokså stor forskjell relativt til det standardavvik vi bruker som mål på usikkerheten til anslagene. Nå brukes vel sjelden utvalgsundersøkelser for å anslå denne andelen, men resultatet har betydning fordi overrepresentasjonen av tettbygde områder kan gi utslag i estimeringen av andre mer interessante størrelser.

I forrige kapittel viste jeg at det var en stor designeffekt (to-trinnseffekt) for denne variabelen. Vi kan se på store designeffekter som en advarsel om at det ved første-trinnstrekkingen kan bli store avvik. Det skjer ikke nødvendigvis, trekkingen kan godt gi et balansert utvalg mht. den spesielle variabelen, men muligheten for avvik foreligger.

Av tabellene ser vi videre at det er avvik for enkelte næringsområder. Siden det dreier seg om nokså små andeler av totalbefolkningen, blir de relative avvik nokså store. Det gjelder primærnæringene, som er underrepresentert og bergverk som er overrepresentert. Det er også avvik for enkelte industriområder.

For gjentatte undersøkelser som AKU er grafisk framstilling av utviklingen over tid, et utmerket redskap. Men av resultatene i dette kapitlet ser vi at en må passe seg for mistolkningen av slike figurer. Det er lett å tenke seg en glatt kurve som går rett gjennom den hakkete kurve med observerte resultater. Men denne viser trenden i de trukne områdene og ikke i landet som helhet. For noen størrelser kan de observerte resultater for hele 10-årsperioden stort sett ligge på samme side av en kurve som viser utviklingen i hele landet. Når en tenker seg en kurve gjennom de hakkete resultatene, vil noen tenke seg avvikene mellom denne og resultatene som en illustrasjon av utvalgsvariansen. Det er de ikke. De illustrerer derimot den betingete varians gitt resultatet av trekkingene på første trinn. Som regel vil de underdrive den totale varians.

### 3.3. Konsekvenser av usikkerhetsberegningene

I dette kapitlet og i kapitlet foran har jeg på ulike måter studert usikkerheten til anslag fra utvalgsplanen. Formålet har hele tiden vært å belyse egenskaper ved den plan som ble lagd i 1974. Jeg har ikke hatt som siktemål å utrede grunnlag for en helt ny plan. Men allikevel har beregningen som biprodukt gitt en del innsikt i den struktur som foreligger i Norge og som må tas hensyn til når en lager en generell utvalgsplan. Jeg skal derfor i dette avsnittet komme med visse forslag.

Jeg tror ikke at det er nødvendig med noen helt ny utvalgsplan. To faktorer er avgjørende for når det er nødvendig. I den ene ligger det som har med reisekostnader å gjøre. Hvis befolkningsutviklingen og/eller utviklingen i intervjumetode (telefon kontra besøk) gjør at reisekostnadene blir av liten betydning, så faller grunnlaget for totrinnstrekkingen bort. Da står en også friere til å lage en mer effektiv stratifisering. Den andre faktoren er bare aktuell hvis reisekostnadene fortsatt er av betydning. Da er det neppe behov for å vurdere noe helt nytt før registergrunnlaget inneholder geografiske områder langt mindre enn kommuner. Eksperimenter med folketellingslignende statistikk for disse områdene må da være grunnlaget for en ny plan.

Jeg mener at de svakheter som er avdekket med den nåværende plan, kan avhjelpes med forholdsvis enkle justeringer. Vi har sett at variabelen tettbygd/spredtbygd er en kritisk størrelse. Mellom primærområdene innen de enkelte strata er det stor variasjon og dette har i førstetrinnstrekkingen resultert i en viss skjevhet. I direkte utvalg ville en sikre seg mot slike skjevheter ved å stratifisere. Når en har totrinnstrekking kan en oppnå noe av de samme ved å trekke ut "balansert" utvalg av primær-områder, dvs. som i gjennomsnitt er nokså nær målpopulasjonen. Dette ville medføre en viss avhengighet mellom førstetrinnstrekkingene i ulike strata men ville ellers kunne ha egenskaper som utvalgsplanen i dag. Jeg vet ikke om det eksisterer noen fullt utbygd teori for dette og den ville bare være aktuell ved en fullstendig ny plan. Men en kan oppnå det samme ved å bytte ut noen få utvalgsområder med nye som har større andel bosatt i spredtbygd område. Dette burde gjøres innenfor region eller landsdel som er de minste geografiske områder det skal lages estimerer for.



Et annet problem var næringsgruppene der jeg antok det var stor geografisk konsentrasjon. En burde ta arbeidet med å finne fram til de strata dette gjelder. Forhåpentligvis dreier det seg om et fåtall. Så kunne en gå over til entrinnstrekking i disse strata.

Med disse forandringer ville en stå nokså fritt til hvordan en skulle foreta en eventuell utvidelse av AKU. Den kunne da stort sett foregå innenfor rammen av standard utvalgsplan uten at gevinsten ved å utvide utvalget ble eliminert av spesielle effekter. En annen fordel er at den kvalitetsopplysning Byrået gir i publiseringen av sine utvalgsundersøkelser ikke behøvde å være så konsentrert om designeffekten og "1,5-regelen". En kunne bruke varians ved tilfeldig trekking som usikkerhetsmål for andeler og heller supplere med spesielle beregninger for varians til numeriske variable og andre data som belyser kvaliteten. Beregningene i denne rapporten har vist at dette er noe som bør overveies for mindre undersøkelser (1 000 - 5 000 enheter) selv om utvalgsplanen ikke justeres.

#### 4. KLYNGEEFFEKT

##### 4.1. Trekkeenhet og analyseenhet

I forbindelse med utvalgsundersøkelser kan det hende at vi ikke har register over de enheter vi vil analysere. Men vi skaffer oss et utvalg ved å trekke fra et register med overordnede enheter. Eksempler kan være en analyse av syketilfelle ved hjelp av et personutvalg der personene intervjues om sine syketilfelle eller en personanalyse der en har trukket husholdninger og intervjuer alle personer som er bosatt i de trukne husholdningene. I slike tilfelle har en behov for å skille mellom trekkeenhet og analyseenhet. Vi kaller slike utvalg klyngeutvalg. De vil ha spesielle statistiske egenskaper.

I AKU er det hovedsakelig personene som analyseres men utvalget er trukket via husholdningen. Hvis en husholdning er med i undersøkelsen, skal alle personer i alderen 16 - 74 år intervjues. Personutvalget er et klyngeutvalg hvor klyngene består av personer i aktuell alder som er registrert bosatt i samme husholdning. Husholdning er trekkeenhet og person er analyseenhet. I den nåværende registersituasjon har vi visse problemer med å trekke husholdninger, men det skal vi se nærmere på i neste kapittel. Årsaken til at en har valgt denne trekkemetoden er ikke mangel på personregister men økonomi. En har alltid regnet med at et slikt klyngeutvalg gir minst kostnader når et gitt antall personer skal intervjues. I stor grad vil intervjuer kunne gjennomføre intervjuene i én husholdning på samme tur. På den annen side vet vi at klyngeutvalg svært ofte gir større usikkerhet enn direkte utvalg. Det er dette jeg skal se nærmere på i dette kapitlet.

##### 4.2. Beregninger av klyngeeffekten

Sammensetningen av klyngene er avgjørende for forholdet mellom varians ved klyngeutvalg og direkte utvalg. Hvis enhetene innen klyngene er svært like, vil variansen ved klyngeutvalg bli relativt stor. På den annen side kan variansen ved klyngeutvalg bli mindre enn variansen ved direkte utvalg, hvis enhetene innen klyngene er svært ulike. Dette kan være en retningsviser hvis en har frihet til å konstruere klyngene. I vårt tilfelle er mulighetene i så måte temmelig begrenset. Videre vil en ofte være interessert i mer enn én variabel. Da får en problemet med at effekten er avhengig av hvilket kjennetegn en ser på.

På grunn av beregningsmessige vansker skal vi ikke se på klyngeeffekt innenfor rammen av standard utvalgsplan. Vi studerer den når utvalg trekkes direkte fra hele landet. Ved vanlig tottrinnsstrekking er variansen sum av to komponenter, en for hvert trinn. En skulle dermed tro at klyngeeffekten bare virket på komponenten på annet trinn og hadde relativt mindre betydning enn ved direkte trekking. Dette er ikke sikkert. En må først vise at en har tilsvarende dekomponering når det er klyngetrekking på annet trinn. Et problem er at det i utvalget blir stokastisk antall enheter fra de ulike strata. Vi tar dog forbehold at effekten for store utvalg kan være noe mindre enn tabellene viser.

Opgaven er å estimere et gjennomsnitt eller en andel blant personer i befolkningen. I beregningene forutsettes at en trekker et tilfeldig utvalg husholdninger og som estimator brukes gjennomsnittet eller andelen blant personer i de trukne husholdningene. I denne estimatoren er både teller og nevner stokastiske og en tilnærming til forventet kvadratavvik brukes som mål for usikkerheten. Som sammenligning brukes variansen til gjennomsnittet når et personutvalg trekkes tilfeldig fra hele befolkningen. Størrelsen på utvalget er lik gjennomsnittlig antall personer pr. husholdning ganget med antall husholdninger i klyngeutvalget. Forholdet mellom usikkerhetene brukes som mål på klyngeeffekten. Forhold større enn 1,00 viser at klyngeutvalg er minst effektivt i statistisk forstand. Er forholdet 1,20 må klyngeutvalget økes med 20 prosent for å oppnå samme presisjon som et direkte personutvalg.

Jeg viser tabeller for massen 16 - 74 år. Det er de aktuelle aldere i AKU, men resultatene er også typisk for andre avgrensninger av målpopulasjonen. Klyngeutvalg brukes også i noen andre undersøkelser.

For de fleste variable er klyngeeffekten større enn 1, dvs. klyngeutvalg er mindre effektivt enn direkte personutvalg. Dette er i overensstemmelse med hva en kunne vente. At effekten i en del tilfelle er større enn 1,20 er kanskje noe bekymringsfullt. Da kan en virkelig snakke om effektivitetstap og dette forhold bør vurderes mot observerte kostnadstall ved alternative innsamlingsmetoder.

Kjennetegnet "bosatt tettsted" er tatt med for å illustrere hvor ille det kan gå når kjennetegnet vi undersøker er konstant innen husholdningen. I slike tilfelle gir tallet på personer i utvalget et helt misvisende bilde av usikkerheten. Videre ser vi at det naturlig nok er en tendens til å være flere gifte i samme husholdning og at selv med ni ulike utdanningsgrupper er personene i samme husholdning mer like enn populasjonen som helhet. For yrkesfelt er det særlig kategori 4, primærnæringene, der det er stor likhet i husholdningene. Til en viss grad er det slik at i husholdningene er det enten ingen eller flere som har yrke i primærnæringene. De samme ser vi selvsagt for næringsgruppene. Klyngeeffekt har vi også for en del andre næringer. Til slutt viser det seg at klyngeutvalg er mer effektivt enn direkte utvalg for inntekt og noen arbeidstidskategorier. Det som her reflekteres er at selv om det sikkert i noen husholdninger enten er ingen eller relativt mange fulltidsarbeidstakere, så er den kompletterende tendens sterkere. Det er i større grad slik at noen arbeider full tid og andre nesten ikke i det hele tatt.

I beregningene er det forutsatt at gjennomsnittet blant personer i utvalget brukes som estimator. Det er denne som best tilsvare det som faktisk gjøres i AKU (sett bort fra etterstratifiseringen) og andre undersøkelser med klyngeutvalg. Men den er bare én av flere mulige estimatorene. Vi definerer følgende:

H = Antall husholdninger i alt.  
 N = Antall personer i alt.  
 h = Antall husholdninger i utvalget.

For utvalget:

$Y_i$  = Sum personverdier i i-te husholdning.  
 $N_i$  = Antall personer i i-te husholdning.

En annen estimator er da:

$$(H/N) \times \left( \sum_{i=1}^h Y_i / h \right),$$

mens vår estimator kan skrives:

$$\frac{\sum_{i=1}^h Y_i}{\sum_{i=1}^h N_i}$$

Sett på husholdningsnivå er forholdet mellom de to estimatorene som forholdet mellom gjennomsnittet og rateestimatoren der antall personer i husholdningen er hjelpevariabel.

Vi har gjennomført usikkerhetsberegninger for begge estimatorene, men har bare vist resultatene for rateestimatoren da denne i overveldende grad var mest effektiv.

Som nevnt før består klyngene av personer som bor i samme husholdning. Ved noen anledninger har det vært foreslått å bruke personer i samme familie (kjerne) som klynger. Jeg har også foretatt beregninger av usikkerheten i dette tilfellet, men spanderer ikke plass på disse resultatene da de praktisk talt er identiske med resultatene for husholdningene.

Tabell 14. Klyngeeffekt for utvalgte variable. 16 - 74 år

Yrkesaktive .....	1,13	
Bosatt tettsted .....	2,61	
Sivilstand		
Ugift .....	1,04	
Gift .....	1,19	
Enke .....	1,04	
Skilt .....	1,11	
Separert .....	1,17	
Utdanning		
1. Barneskolenivå .....	1,42	
2. Ungdomsskolenivå .....	1,28	
3. Gymnasnivå I .....	1,05	
4. Gymnasnivå II .....	1,09	
5. Universitets- og høgsolenivå I .....	1,16	
6. Universitets- og høgsolenivå II .....	1,17	
7. Universitets- og høgsolenivå III .....	1,13	
8. Forskernivå .....	1,06	
9. Uoppgitt .....	1,21	
Yrkesfelt		
1. Administrasjons- og forvaltnings- arbeid, bedrifts- og organisasjons- ledelse .....	1,03	
2. Kontorarbeid .....	1,06	
3. Handelsarbeid .....	1,16	
4. Jordbruks-, skogbruks- og fiskearbeid	1,53	
5. Gruve- og sprengningsarbeid m.m. ....	1,05	
6. Transport- og kommunikasjonsarbeid .....	1,07	
7. Industri-, bygge- og anleggsarbeid .....	1,05	
8. Industri-, bygge- og anleggsarbeid .....	1,11	
9. Servicearbeid .....	1,06	
Yrkesstatus		
Ansatt .....	1,25	
Selvtendig .....	1,14	
Familie .....	1,23	
Næring		
1. Jordbruk, skogbruk, fiske og fangst	1,67	
11. Jordbruk .....		1,75
12. Skogbruk .....		1,15
13. Fiske og fangst .....		1,33
2. Oljeutvinning og bergverksdrift .....	1,24	
3 Industri .....	1,18	
31 Produksjon av næringsmidler, drikke- varer og tobakksvarer .....		1,23
32 Produksjon av tekstilvarer, bekled- ningsvarer, lær og lærvarer .....		1,19
33 Produksjon av trevarer .....		1,23
34 Treforedling, grafisk produksjon og forlagsvirksomhet .....		1,19
35 Produksjon av kjemiske produkter, mineralolje-, kull-, gummi- og plastprodukter .....		1,18
36 Produksjon av mineralske produkter		1,20
37 Produksjon av metaller .....		1,27
38 Produksjon av verkstedprodukter .....		1,14

Tabell 14 (forts.). Klyngeeffekt for utvalgte variable. 16 - 74 år

4	Kraft- og vannforsyning .....	1,13	
5	Bygge- og anleggsvirksomhet .....	1,11	
6	Varehandel, hotell- og restaurantvirksomhet .....	1,23	
	61, 62 Engros- og agenturhandel, detaljhandel .....		1,24
	63 Hotell- og restaurantdrift .....		1,20
7	Transport, lagring, post og telekommunikasjoner .....	1,13	
	71 Transport og lagring .....		1,11
	72 Post- og telekommunikasjoner .....		1,17
8	Bank- og finansieringsvirksomhet, forsikringsvirksomhet, eiendomsdrift og forretningsmessig tjenesteyting .....	1,15	
	81 Bank- og finansieringsvirksomhet .....		1,11
	82 Forsikringsvirksomhet .....		1,08
	83 Eiendomsdrift og forretningsmessig tjenesteyting .....		1,15
9	Offentlig, sosial og privat tjenesteyting ..	1,20	
	91 Offentlig administrasjon, forsvar, politi og rettsvesen .....		1,09
	92 Renovasjon og reingjøring .....		1,14
	93 Undervisning, helse- og andre sosialtjenester .....		1,19
	931, 932 Undervisning m.m., Helse- og andre sosialtjenester .....		1,22
	933 Helse- og veterinærtjenester ....		1,14
	934 Sosial omsorg og velferdsarbeid .		1,03
	94 Kulturell tjenesteyting, underholdning og sport .....		1,11
	95 Personlig tjenesteyting .....		1,11
Viktigste kilde til livsopphold			
	Inntekt av eget arbeid .....	1,40	
	Verneplikt, siviltjeneste .....	1,00	
	Pensjon, trygd .....	1,39	
	Stipend, lån, formue .....	1,16	
Arbeidstid i året			
	100 - 499 timer .....	1,06	
	500 - 999 " .....	1,01	
	1 000 - 1 299 " .....	1,03	
	1 300+ " .....	0,91	
Arbeidstid i uka			
	0 .....	1,10	
	1 - 9 timer .....	1,04	
	10 - 19 " .....	0,99	
	20 - 29 " .....	0,99	
	30+ " .....	0,93	
	Inntekt .....	0,60	

## 5. EFFEKT AV ULIKE TREKKESANNSYNLIGHETER

## 5.1. Bakgrunn

Når husholdning er trekkeenhet, har en noen spesielle problemer. Våre utvalg trekkes fra et trekkeregister som ajourføres ved hjelp av Personregisteret og folkeregistreringen. Det nåværende registergrunnlag har ikke husholdning som enhet og husholdninger kan heller ikke konstrueres på grunnlag av de opplysninger som finnes. I registeret finnes den mindre enhet familiekjerne. Når husholdning skal være trekkeenhet, trekkes et utvalg familiekjerner og så sørger intervjuerne for at alle personer/familier i husholdningene med de uttrukne familiene kommer med i utvalget. Intervjuerne foretar bare denne suppleringen av utvalget i husholdninger med inntil fem familiekjerner. Dette fører strengt tatt til at for slike utvalg har personene/familiene/husholdningene trekkesannsynlighet proporsjonal med tallet på familier i husholdningen. Hvis en ikke tar hensyn til dette, medfører det skjevhet i estimatene. En har oftest regnet med at denne skjevhet ikke har praktisk betydning. Siden både familie og husholdning er enhet i Folketellingen 1980, benytter vi anledningen til å beregne skjevheten for en del variable som inngår i tellingen.

Størrelsen på skjevhetene vil avhenge av om det er sammenheng mellom kjennetegnet vi undersøker og tallet på familier i husholdningen. Hvis sammenhengen er sterk, vil skjevheten bli relativt stor. Videre vil skjevheten avhenge av hvor stor andel av befolkningen som bor i flerfamiliehusholdninger. Tall fra Folketellingen i 1980 viser:

Tabell 5.1. Personer, familier og husholdninger etter tallet på familier i husholdningene. Privat-husholdninger

	I alt	Tallet på familier i husholdningen			
		1	2	3	4 og flere
Personer .....	100,0	89,2	9,6	1,1	0,1
Familier .....	100,0	84,5	13,4	1,8	0,3
Husholdninger .....	100,0	92,0	7,3	0,7	0,1

Siden tallene viser at langt de fleste bor i en-familiehusholdninger, er det grunn til å håpe at skjevhetene ikke har praktisk betydning.

## 5.2. Beregninger av effekten

Forholdet er undersøkt på en litt forenklet måte. Jeg studerer skjevheten når det trekkes utvalg direkte fra hele befolkningen uten å følge trinnene i standard utvalgsplan. Det er for å redusere beregningsarbeidet og unngå en del tekniske vansker. Resultatene vil illustrere det generelle problem.

Skjevheten er beregnet som differansen mellom forventningen i det skjeve utvalget og gjennomsnittet i hele befolkningen. Skjevheten blir positiv for kjennetegn som forekommer relativt hyppig blant de som bor i flerfamiliehusholdninger og negativ når det motsatte er tilfelle.

I tabellene har jeg tatt med gjennomsnittet (andelen) for hele befolkningen dvs. den størrelse vi estimerer. Det gir grunnlag for å vurdere den praktiske betydning av skjevhet da en kan se på relativ skjevhet. Da bør en være oppmerksom på at for dikotome kjennetegn vil den relative skjevhet avhenge av hvilken av de to kategoriene en studerer. Jeg har også tatt med standardavviket ved tilfeldig trekking. Det er ofte det eneste kvantitative mål for usikkerhet vi bruker. Forholdet mellom skjevhet og standardavvik vil vise om vår praksis er holdbar.

De første tabellene er basert på personer i alderen 16 - 74 år. Dette fordi AKU er av de undersøkelser der husholdning skal være trekkeenhet.

Tabell 15. Skjevhet for utvalgte variable. Personer 16 - 74 år

Variabel	Skjevhet	Standardavvik ved tilfeldig utvalg på 12 000 personer	Gjennomsnitt i promille
Yrkesaktive .....	-0,5	4,1	708
Sivilstand			
Ugift .....	20,4	4,1	272
Gift .....	-29,9	4,4	631
Enke, enkemann .....	1,2	2,0	50
Skilt .....	5,5	1,7	34
Separert .....	2,8	1,0	13
Yrkesfelt			
1 Administrasjons- og forvaltnings- arbeid, bedrifts- og organisasjons- ledelse .....	-1,1	1,6	33
2 Kontorarbeid .....	-0,6	2,3	68
3 Handelsarbeid .....	-0,9	2,2	64
4 Jordbruks-, skogsbruks- og fisker- arbeid .....	2,1	2,0	49
5 Gruve- og sprengingsarbeid m.m. ....	0,0	0,5	3
6 Transport- og kommunikasjonsarbeid .....	0,5	2,1	54
7 Industri-, bygge- og anleggsarbeid .....	0,4	2,9	117
8 Industri-, bygge- og anleggsarbeid .....	0,6	2,0	53
9 Servicearbeid .....	0,5	2,5	85
Yrkesstatus			
Ansatt .....	-1,8	4,4	612
Selvstendige .....	0,8	2,2	61
Familiemedlem .....	0,4	1,1	15
Udef. ....	0,6	4,2	312
Viktigste kilde til livsopphold			
Inntekt av eget arbeid .....	-5,4	4,1	715
Pensjon, trygd .....	4,5	3,4	169
Stipend, lån, formue o.l. ....	1,0	1,5	29
Verneplikt/siviltjeneste .....	0,1	0,9	9
Udef. ....	-0,1	2,5	80
Arbeidstid i året			
100 - 499 timer .....	-1,1	2,8	104
500 - 999 " .....	-0,5	2,5	82
1 000 - 1 299 " .....	0,0	2,1	58
1 300 - " .....	0,8	4,5	421
Udef. ....	0,9	4,3	334
Arbeidstid i uka 25.10. - 31.10.1980			
0 timer .....	0,9	3,0	127
1 - 9 " .....	-0,6	1,5	30
10 - 19 " .....	-1,1	1,8	40
20 - 29 " .....	-1,1	2,2	60
30 - " .....	1,2	4,5	430
Udef. ....	0,6	4,2	314
Hustype			
Våningshus .....	7,3	3,0	126
Frittliggende enebolig .....	-6,8	4,5	476
Rekkehus .....	-2,0	2,9	114
Horisontaldelt tomannsbolig .....	0,5	1,9	44
Annet under 3 etg. ....	0,6	1,6	34
Blokk, leiegård .....	0,8	3,2	142
Forretningsbygg .....	0,1	1,0	11
Hotell, aldershjem .....	0,0	0,5	3
Udef. ....	-0,4	2,1	54
Inntekt (100 kr) .....	-0,95	2,11	346

Tabell 15 (forts.). Skjevhet for utvalgte variable. Personer 16 - 74 år

Variabel	Skjevhet	Standardavvik ved tilfeldig utvalg på 12 000 personer	Gjennomsnitt i promille
<b>Næring</b>			
1 Jordbruk, skogbruk, fiske og fangst .....	2,4	2,1	56
11 Jordbruk .....	2,1	1,9	45
12 Skogbruk .....	0,4	0,5	3
13 Fiske og fangst .....	0,3	0,8	7
2 Oljeutvinning og bergverksdrift .....	0,0	0,6	4
3 Industri .....	-0,1	3,2	143
31 Produksjon av næringsmidler, drikkevarer og tobakksvarer .....	0,3	1,4	25
32 Produksjon av tekstilvarer, beklednings- varer, lær og lærvarer .....	0,2	0,9	9
33 Produksjon av trevarer .....	0,0	1,1	14
34 Treforedling, grafisk produksjon og for- lagsvirksomhet .....	-0,1	1,2	18
35 Produksjon av kjemiske produkter, mineral- olje-, kull-, gummi- og plastprodukter ....	-0,1	1,0	12
36 Produksjon av mineralske produkter .....	0,0	0,6	5
37 Produksjon av metaller .....	-0,1	0,9	11
38 Produksjon av verkstedprodukter .....	-0,3	2,0	48
4 Kraft- og vannforsyning .....	-0,1	0,7	6
5 Bygge- og anleggsvirksomhet .....	0,2	2,1	55
6 Varehandel, hotell- og restaurantvirksomhet ...	-0,6	3,0	125
61 Engros- og agenturhandel .....			
62 Detaljhandel .....	-1,1	2,8	107
63 Hotell- og restaurantdrift .....	0,5	1,2	17
7 Transport, lagring, post og telekommunikasjoner	0,5	2,2	62
71 Transport og lagring .....	0,5	1,9	45
72 Post og telekommunikasjoner .....	0,0	1,2	17
8 Bank- og finansieringsvirksomhet, forsikrings- virksomhet, eiendomsdrift og forretningsmessig tjenesteyting .....	-0,5	1,7	36
81 Bank- og finansieringsvirksomhet .....	-0,2	1,0	13
82 Forsikringsvirksomhet .....	-0,1	0,6	4
83 Eiendomsdrift og forretningsmessig tjenesteyting .....	-0,2	1,2	19
9 Offentlig, sosial og privat tjenesteyting .....	-2,4	3,7	215
91 Offentlig administrasjon, forsvar, politi og rettsvesen .....	-0,7	2,0	52
92 Renovasjon og reingjøring .....	0,0	0,6	4
93 Undervisning, helse- og andre sosialtjen. .	-2,0	3,1	136
931+932 Undervisning m.m. ....	-1,4	1,9	48
933 Helse- og veterinærtjenester .....	-0,6	2,2	61
934 Sosial omsorg og velferdsarbeid ..	0,1	1,4	23
94 Kulturell tjenesteyting, underholdning og sport .....	0,0	0,7	6
95 Personlig tjenesteyting .....	0,2	1,2	17

Det viser seg at skjevhetene i mange tilfelle er relativt små i forhold til gjennomsnittet. Men det er vanskelig å sette absolutte grenser, så formålet med undersøkelsen vil være den beste rettesnor for å vurdere om skjevheten har betydning. På den annen side er det enkeltskjevheter som er så store, f.eks. for sivilstand, at de vil ha betydning i de fleste sammenhenger. Det avgjørende for skjevheten er sammenhengen med tallet på familier i husholdningen. Det er vanskelig å vite på forhånd om en har like sterk sammenheng som det tydeligvis er for sivilstand. Men en bør huske at familie i denne sammenheng følger av den strenge definisjonen som brukes i Personsregisteret. Hvis en bruker AKU til å beregne tallet på samboende, vil skjevheten opplagt ha betydning.

En annen sak er forholdet mellom skjevhet og standardavvik. Igjen kan det være vanskelig å sette absolutte grenser, men de fleste vil antagelig være enig i at hvis skjevheten er større enn 50 prosent av standardavviket, så er det dårlig forbrukeropplysning å la standardavviket være eneste mål på usikkerhet. Tabellene viser at i en del tilfelle er skjevheten så stor.

I den senere tid har det vært gjennomført tilleggsundersøkelser til AKU der husholdning har vært analyseenhet og det finnes andre husholdningsundersøkelser der utvalget er etablert ved samme metode. Hvis en bruker data om husholdning uten å veie, vil det opptre skjevheter. Jeg vil derfor presentere noen beregninger av skjevheten for husholdninger. Grunnlaget for tabellene er alle husholdninger, vi har altså ikke begrenset oss til husholdninger med "AKU-personer".

Tabell 16. Skjevhet for utvalgte variable. Husholdninger

Variabel	Skjevhet	Standardavvik ved tilfeldig utvalg på 6 000 husholdninger	Gjennomsnitt (andel) i promille
<b>Hustype</b>			
Våningshus .....	7,9	4,0	109
Frittliggende enebolig .....	-2,0	6,4	419
Rekkehus mv. ....	-1,6	4,0	107
Horisontal tomannsbolig .....	-0,1	2,7	47
Annet under 3 etg. ....	0,0	2,5	40
Blokk .....	-2,1	4,9	179
Forretningsbygg .....	0,0	1,4	12
Bygg for felleshush. ....	-0,2	0,9	5
Udefinert .....	-1,9	3,5	81
Med bad .....	0,3	4,9	822
Med WC .....	-0,4	4,9	826
Med telefon .....	0,6	6,2	624
Med kjøkken .....	3,7	5,2	793
<b>Tallet på rom</b>			
1 .....	-1,6	2,9	53
2 .....	-1,3	4,4	137
3 .....	-0,3	5,1	190
4 .....	-1,7	5,6	256
5 .....	1,1	4,4	138
6 .....	2,0	3,4	75
7 .....	1,6	2,4	35
8+ .....	2,8	2,4	36
Udef. ....	-2,7	3,5	79
<b>Tallet på familier</b>			
1 .....	-74,4	3,5	920
2 .....	61,3	3,4	73
3 .....	11,7	1,1	7
4 .....	1,4	0,3	0,5



Tabell (forts.) 16. Skjevhet for utvalgte variable. Husholdninger

Variabel	Skjevhet	Standardavvik ved tilfeldig utvalg på 6 000 husholdninger	Gjennomsnitt (andel) i promille
Tallet på personer			
1 .....	-22,6	5,8	279
2 .....	4,5	5,6	259
3 .....	7,0	4,8	163
4 .....	-1,0	4,9	179
5 .....	3,3	3,5	83
6 .....	4,3	2,0	26
7 .....	2,6	1,1	8
8 .....	1,2	0,6	2
9 .....	0,5	0,4	1
10+ .....	0,3	0,3	0,5
Antall rom .....	0,035	0,023	3,950
Inntekt (100 kr) .....	23	5	695
Antall personer .....	0,081	0,019	2,655
Antall familier .....	0,089	0,004	1,088

Resultatene ligner de vi fikk for personer. Det er ikke så ofte at den relative skjevhet (forholdet til størrelsen som estimeres) er særlig stor. Men det er også klart at standardavviket alene kan gi et helt misvisende bilde av presisjonsnivået. I flere tilfelle er skjevheten "stor" i forhold til standardavviket.

I tabellen er tatt med to spesielle variable, "tallet på familier" og "antall familier". Den første er en kategorisert utgave av kjennetegnet og den andre den direkte numeriske versjonen. For disse er sammenhengen med trekkesannsynligheten "direkte" og resultatene for disse illustrerer de "maksimale" utslag av skjevheten.

For husholdningene har vi noen flere numeriske variable enn for personer. For alle disse ligger skjevheten godt over standardavviket for utvalg på 6 000 husholdninger.

### 5.3. Konsekvenser

Tallene viser at skjevhetene nok er noe større enn det vi trodde på forhånd. Selv om det kanskje er i få tilfeller skjevheten vil ha substansiell betydning, er det større usikkerhet i tallene enn antatt. Det er mange måter å løse problemet på, jeg viser til SØS nr. 33 for noen forslag. Det beste ville være å få etablert husholdning som enhet i trekkeregisteret. Men for å få til det må en ha et langt skarpere adressebegrep med løpende ajourhold i Personregisteret. Så vidt jeg vet, er det ikke fastsatt noen dato for å gjennomføre dette. Med husholdning som enhet har en full frihet til å velge metode.

En annen løsning er å fortsette som før, men å veie resultatene omvendt proporsjonalt med trekkesannsynligheten. En må da passe på å skaffe seg opplysninger om tallet på familier i husholdningen. Frafallet skaper noen småproblemer. Den veing mht. husholdningstype som brukes i forbindelse med forbruksundersøkelsene, gir nok langt på vei en slik veing. Men etterstratifiseringen i AKU gir det neppe. I prinsippet er veing lett å gjennomføre, men i praksis er det flere problemer. For undersøkelser som skal dekke mange ulike formål og blir gjenstand for analytisk bearbeiding, er det meget fordelaktig at grunnlaget er et selvveiende utvalg. Derfor bør en foreløpig begrense bruken av utvalg trukket via husholdninger.

## 1. Formelverk

Vi innfører følgende notasjon:

$N_{ij}$  = Antall enhet i befolkningen i utvalgsområde nr. j i stratum nr. i.

$N_i$  = Antall enheter i befolkningen i stratum nr. i.

$N$  = Antall enheter i befolkningen.

La  $n_{ij}$ ,  $n_i$  og  $n$  være de tilsvarende størrelser i utvalget.

$\pi_{ij}$  = Sannsynligheten for at utvalgsområde nr. j i stratum nr. i trekkes på første trinn.

For et kjennetegn  $y$ ;

$y_{ijk}$  = Verdien til  $k$ -te enhet i utvalgsområde nr. j i stratum nr. i.

$\bar{y}_{ij}$  = Gjennomsnittet i utvalgsområde nr. j i stratum nr. i.

$\bar{y}_i$  = Gjennomsnittet i stratum nr. i

$\bar{y}$  = Gjennomsnittet i hele befolkningen.

La  $Y_{ijk}$ ,  $\bar{Y}_{ij}$ ,  $\bar{Y}_i$  og  $\bar{Y}$  være de tilsvarende størrelser i utvalget.

Det trekkes selvveiende utvalg ved at  $n_{ij}$  bestemmes slik at

$$\frac{n_{ij}}{N_{ij}} \cdot \pi_{ij}$$

er konstant for alle områder. Som estimator brukes  $\bar{Y}$ , gjennomsnittet i utvalget.

For variansene i kapittel 2 har vi brukt følgende formler:

$$\text{Var}_A \bar{Y} = \frac{N-n}{N-1} \cdot \frac{1}{n} \cdot \frac{1}{N} \sum_{i,j,k} (y_{ijk} - \bar{y})^2$$

$$\text{Var}_B \bar{Y} = \frac{N-n}{nN^2} \sum_i \frac{N_i}{N_i-1} \sum_{j,k} (y_{ijk} - \bar{y}_i)^2$$

$$\begin{aligned} \text{Var}_C \bar{Y} &= \frac{1}{nN} \sum_i \sum_j \frac{N_{ij}}{N_{ij}-1} \left(1 - \frac{nN_i}{NN_{ij}}\right) \sum_k (y_{ijk} - \bar{y}_{ij})^2 \\ &+ \frac{1}{N^2} \sum_i N_i \sum_j N_{ij} (\bar{y}_{ij} - \bar{y}_i)^2 \end{aligned}$$

$$\begin{aligned} \text{Var}_D \bar{Y} &= \frac{1}{nN} \sum_i \sum_j \frac{N_{ij}}{N_{ij}-1} \left(1 - \frac{n}{N \pi_{ij}}\right) \sum_k (y_{ijk} - \bar{y}_{ij})^2 \\ &+ \frac{1}{N^2} \sum_i N_i^2 \sum_j \pi_{ij} \left(\frac{N_{ij}}{N_i \pi_{ij}} \bar{y}_{ij} - \bar{y}_i\right)^2 \end{aligned}$$

Den siste formelen gir strengt tatt et tilnærmet uttrykk for variansen.

I kapittel 3 beregnes differansen mellom betinget forventning gitt 1.trinns trekking og ubetinget forventning. Lar vi alle de trukne områdene ha nr.  $j = 1$  blir differansen

$$\frac{\sum_i \left( \frac{N_{i1}}{\pi_{i1}} \right) \bar{y}_{i1}}{\sum_i \left( \frac{N_{i1}}{\pi_{i1}} \right)} - \bar{y}$$

I kapittel 4 beregnes klyngeeffekter. Der ser en bort fra utvalgsplanen og forutsetter at utvalg trekkes direkte fra hele befolkningen. Vi trenger følgende notasjon

$H$  = Antall husholdninger i alt.

$N$  = Antall personer i alt.

$N_i$  = Antall personer i  $i$ -te husholdning.

$y_{ij}$  =  $y$ -verdi for  $j$ -te person i  $i$ -te husholdning.

$y_i$  = Sum personverdier i  $i$ -te husholdning.

$\bar{y}$  = Gjennomsnittet av personverdier.

For utvalget

$h$  = Antall trekkeenheter (husholdninger)

$Y_i$  = Sum personverdier i  $i$ -te husholdning

$N_i$  = Antall personer i  $i$ -te husholdning.

Estimatoren er: 
$$\sum_{i=1}^h Y_i / \sum_i N_i$$

Tilnærmet forventet kvadratavvik er:

$$\frac{H^2}{N^2} \cdot \frac{H-h}{H-1} \cdot \frac{1}{n} \cdot \frac{1}{H} \sum_{i=1}^H (y_i - N_i \cdot \bar{y})^2$$

Denne sammenlignes med et tilfeldig persontuvalg på  $n = \frac{N}{H} \times h$  personer. Gjennomsnittet har da varians

$$\frac{H-h}{h} \cdot \frac{1}{N} \cdot \frac{\sum_{ij} (y_{ij} - \bar{y})^2}{N-1}$$

I kapittel 5 beregnes forventet avvik når det er ulike trekkesannsynligheter og det uveide gjennomsnitt brukes som estimator. La

$N$  = Tallet på enheter

$y_i$  = Variabelverdi for  $i$ -te enhet (person, familie eller husholdning)

$\bar{y}$  = Gjennomsnittet i hele befolkningen

$v_i$  = Tallet på familiekjerner i husholdningen til  $i$ -te enhet.

Forventet avvik er beregnet som:

$$\frac{\sum_I v_i y_i}{N} - \bar{y}$$

$$\sum_I v_i$$

## LITTERATUR

- Laake, Petter (1974): "Estimering av totaler med en to-trinnsutvalgsplan der de primære utvalgsområder trekkes med ulik sannsynlighet i første trinn." SSB, Arbeidsnotat. IO 74/49.
- Raj, Des (1968): "Sampling Theory" McGraw-Hill, Inc.
- Statistisk Sentralbyrå (1977): "Prinsipper og metoder for Statistisk Sentralbyrås utvalgsundersøkelser." Samfunnsøkonomiske studier nr. 33.
- Sæbø, Hans Viggo (1976): "Varianser og designeffekter for sysselsettingstall estimert ved bruk av Byråets nye utvalgsplan." SSB, Arbeidsnotat IO 76/1.
- Sæbø, Hans Viggo (1977): "Varianser for endringstall som er estimert ved bruk av Byråets nye utvalgsplan." SSB, Arbeidsnotat IO 77/23.
- Thomsen, Ib og Arne Rideng (1974): "Oversikt over arbeidet med ny utvalgsplan." SSB, Arbeidsnotat IO 74/25.

Trykt 1984

- Nr. 84/1 Naturressurser og miljø 1983 Foreløpige nøkkeltall fra ressursregnskapene for energi, mineraler, skog, fisk og areal Sidetall 100 Pris kr 18,00 ISBN 82-537-1993-0
- 84/2 Torstein Bye: Energisubstitusjon i næringssektorene i en makromodell Sidetall 47 Pris kr 12,00 ISBN 82-537-2042-4
- 84/3 Trygdedes inntekts- og boforhold 1980 Sidetall 89 Pris kr 18,00 ISBN 82-537-2040-8
- 84/4 Jon Åge Vestøl: Kommunale avfallsbehandlingsanlegg Miljøstandard Sidetall 78 Pris kr 18,00 ISBN 82-537-2062-9
- 84/5 Bjørg Moen: Bibliography of Population Studies in Norway Bibliografi over befolkningsstudier i Norge Sidetall 114 Pris kr 18,00 ISBN 82-537-2045-9
- 84/6 Grete Dahl: Folketrygden. Korttidssytelser og stønad ved yrkesskade Sidetall 26 Pris kr 12,00 ISBN 82-537-2069-6
- 84/7 Tiril Vogt: Social Indicators and Environmental Dimensions Sidetall 33 Pris kr 12,00 ISBN 82-537-2060-2
- 84/8 Otto Carlsen: Pasientstatistikk 1982 Statistikk fra Det økonomiske og medisinske informasjonssystem Sidetall 61 Pris kr 18,00 ISBN 82-537-2066-1
- 84/9 Herdis Thorén Amundsen: Statistiske metoder for analyse av samvariasjon i kategoriske data Sidetall 228 Pris kr 24,00 ISBN 82-537-2074-2
- 84/10 Audun Rosland: Vannkraftutbygging - Reguleringsinngrep - Virkninger på fisk Sidetall 127 Pris kr 18,00 ISBN 82-537-2102-1
- 84/11 Skatter og overføringer til private Historisk oversikt over satser mv. årene 1970 - 1984 Sidetall 75 Pris kr 18,00 ISBN 82-537-2081-5
- 84/12 Arne Faye og Helge Heriøstad: Friluftsliv i Norge 1970 - 1982 Sidetall 77 Pris kr 18,00 ISBN 82-537-2092-0
- 84/13 Jon Paschen Knudsen: Boligstandard Variasjoner innen og mellom byer Sidetall 66 Pris kr 18,00 ISBN 82-537-2088-2
- 84/14 Erling Siring og Emil Spjøtvoll: Regresjonsanalyse med et stort antall variable Sidetall 55 Pris kr 18,00 ISBN 82-537-2122-6
- 84/15 Sindre Børke: Folke- og bolig telling 1980 Dokumentasjon Sidetall 211 Pris kr 24,00 ISBN 82-537-2112-9
- 84/16 Stein Opdahl: Aleneforeldres levekår og tidsbruk Sidetall 188 Pris kr 18,00 ISBN 82-537-2127-7
- 84/17 Alette Schreiner og Tor Skoglund: Virkninger av oljevirkosomhet i Nord-Norge Sidetall 43 Pris kr 18,00 ISBN 82-537-2118-8
- 84/18 Morten Reymert: Import- og eksportlikninger i KVARTS Utledning, estimering og simulering med likninger for utenrikshandelen Sidetall 83 Pris kr 18,00 ISBN 82-537-2123-4
- 84/19 Børre Nordby: Valg av ferietype Sidetall 53 Pris kr 18,00 ISBN 82-537-2197-8
- 84/20 Arne Liones: Energiundersøkelsen 1983 Om energibruk og energiøkonomisering i private husholdninger Sidetall 62 Pris kr 18,00 ISBN 82-537-2130-7
- 84/21 Johan Helda: Kvalitetskontrollundersøkelsen for Folke- og bolig tellingen 1980 Sidetall 115 Pris kr 18,00 ISBN 82-537-2140-4
- 84/22 Sindre Børke: Tilleggsundersøkelsen til Folke- og bolig telling 1980 Om muligheter for å erstatte skjema med registeropplysninger i senere folke- og bolig tellingen Sidetall 61 Pris kr 18,00 ISBN 82-537-2136-6
- 84/23 Roar Bergan: MINK En finansiell ettermodell til MSG En MSG-rapport Sidetall 71 Pris kr 18,00 ISBN 82-537-2138-2
- 84/24 Yngvar Holm: Engrossetningsindeks Sidetall 18 Pris kr 12,00 ISBN 82-537-2141-2
- 84/25 Morten Jensen og Morten Reymert: Kvartalsmodellen KVARTS - modellbeskrivelse og teknisk dokumentasjon Sidetall 87 Pris kr 18,00 ISBN 82-537-2139-0

Utkommet i serien Rapporter fra Statistisk Sentralbyrå (RAPP) - ISSN 0332-8422 (forts.)

Trykt 1985

- Nr. 85/1 Naturressurser og miljø 1984 Foreløpige nøkkeltall fra ressursregnskapene for miljø, energi, mineraler, skog, fisk og areal Sidetall 94 Pris kr 30,00 ISBN 82-537-2133-1
- 85/2 Aktuelle skattetal 1984 Current Tax Data Sidetall 44 Pris kr 20,00 ISBN 82-537-2142-0
- 85/3 Eva Ivås og Gunnar Sollie: MODIS IV Detaljerte virkningstabeller for 1983 Sidetall 268 Pris kr 45,00 ISBN 82-537-2153-6
- 85/4 Lorents Lorentsen og Kjell Roland: Markedet for råolje Historisk utvikling. Teorier og modeller. Prisprognoser Sidetall 58 Pris kr 20,00 ISBN 82-537-2145-5
- 85/5 Morten Reymert og Carl-Erik Schulz: Eksport og markedsstruktur Eksportutvikling og markedsandeler for Norge og andre land 1963 - 77 Sidetall 149 Pris kr 30,00 ISBN 82-537-2155-2
- 85/6 Elisabeth Fadum, Katalin Nagy og Tiril Vogt: Referansearkiv for naturressurs- og forurensningsdata: Emnekatalog for ferskvann Sidetall 313 Pris kr 50,00 ISBN 82-537-2159-5
- 85/7 Arne Rideng, Knut Ø. Sørensen og Kjetil Sørli: Modell for regionale befolkningsframskrivinger Sidetall 71 Pris kr 25,00 ISBN 82-537-2162-5
- 85/8 Kjetil Sørli: MATAUK En modell for tilgang på arbeidskraft, revidert modell og framskriving av arbeidsstyrken 1983 - 2000 Sidetall 81 Pris kr 25,00 ISBN 82-537-2163-3
- 85/9 Hilde Olsen, Morten Reymert og Pål Ulla: Det norske nasjonalregnskapet. Dokumentasjonsnotat nr. 20 - Kvartalsvis nasjonalregnskap - Dokumentasjon av beregningsopplegget Sidetall 97 Pris kr 25,00 ISBN 82-537-2167-6
- 85/10 Nordby, Børre: Feriereiser og ferieplaner. Undersøkelse i januar-februar 1985 Sidetall 60 Pris 25,00 ISBN 82-537-2170-6
- 85/11 Liv Argel: Avisenes bruk av statistikk Resultater fra en postundersøkelse i oktober 1984 Sidetall 34 Pris kr 20,00 ISBN 82-537-2185-4
- 85/12 Anders Harildstad: Det norske nasjonalregnskapet Dokumentasjonsnotat nr. 19 Arbeidskraftregnskapet - Beregning av arbeidskraftforbruket i varehandel Sidetall 45 Pris kr 25,00 ISBN 82-537-2186-2
- 85/13 Vidar Knudsen: En kvartalsmodell for boliginvesteringer estimert på norske data for perioden 1966 - 1978 Sidetall 46 Pris kr 20,00 ISBN 82-537-2206-0
- 85/14 Hogne Steinbakk og Terje Wessel: Planrekneskap for Møre og Romsdal 1984 - 1995 Hovedresultat Sidetall 56 Pris kr 20,00 ISBN 82-537-2209-5
- 85/15 Tore Høy, Terje Wessel og Hogne Steinbakk: Planrekneskap for Sogn og Fjordane 1984 - 1995 Hovedresultat Sidetall 49 Pris kr 20,00 ISBN 82-537-2210-9
- 85/16 Olav Ljønes: Utviklingen av arbeidsmarkedsmodeller i Statistisk Sentralbyrå Sidetall 65 Pris kr 25,00 ISBN 82-537-2216-8
- 85/17 Skatter og overføringer til private Historisk oversikt over satser mv. Årene 1970 - 1985 Sidetall 75 Pris kr 25,00 ISBN 82-537-2218-4
- 85/18 Elisabeth Fadum og Tiril Vogt: Referansearkiv for naturressurs- og forurensningsdata: Hefte I Arkivdel Sidetall 272 Pris kr 45,00 ISBN 82-537-2227-3
- 85/18 Elisabeth Fadum og Tiril Vogt: Referansearkiv for naturressurs- og forurensningsdata: Hefte II Registerdel Sidetall 224 Pris kr 45,00 ISBN 82-537-2227-3
- 85/19 Svein H. Trosdahl: Kommunale og fylkeskommunale utvalg oppnevnt i 1984 for perioden 1984 - 1987 Sidetall 107 Pris kr 25,00 ISBN 82-537-2235-4
- 85/20 Vidar Knudsen: INSIDENS - En modell for analyse av fordelingsvirkninger av endringer i avgifter og subsidier Sidetall 43 Pris kr 25,00 ISBN 82-537-2239-7
- 85/21 Morten Jensen: Kvartalsvise investeringsrelasjoner basert på en utvidet akseleratormodell Sidetall 55 Pris kr 25,00 ISBN 82-537-2237-0
- 85/22 Totalregnskap for fiske- og fangstnæringen 1980 - 1983 Sidetall 41 Pris kr 20,00 ISBN 82-537-2242-7
- 85/23 Arild Angelsen: Kommunale utbyggingsplaner til industriformål Sidetall 80 Pris kr 25,00 ISBN 82-537-2245-1

Utkommet i serien Rapporter fra Statistisk Sentralbyrå (RAPP) - ISSN 0332-8422 (forts.)

Trykt 1985 (forts.)

- Nr. 85/28 Paa1 Sand og Gunnar Sollie: MODIS IV Dokumentasjonsnotat nr. 23 Endringer i utgave 83 - 1 Sidetall 79 Pris kr 25,00 ISBN 82-537-2253-2
- 85/29 Roar Bergan og Øystein Olsen: Eksporttilpasning i MODAG A En MODAG-rapport Sidetall 99 Pris kr 25,00 ISBN 82-537-2255-9
- 85/30 Ingar Kristoffersen og Erik Næsset: Ressursregnskap for skog Sidetall 72 Pris kr 25,00 ISBN 82-537-2256-7
- 85/31 Frode Brunvoll: VAR Hefte I Statistikk for Vannforsyning, Avløp og Renovasjon Analyse av VAR-data Sidetall 77 Pris kr 25,00 ISBN 82-537-2258-3
- 85/32 Feriereiser og ferieplaner Undersøkelse i mai-juni 1985 Sidetall 49 Pris kr 25,00 ISBN 82-537-2262-1
- 85/33 Aktuelle skattetal1 1985 Current Tax Data Sidetall 46 Pris kr 20,00 ISBN 82-537-2265-6
- 85/34 Tor Haldorsen: Statistiske egenskaper ved Byråets standard utvalgsplan Sidetall 46 Pris kr 25,00 ISBN 82-537-2271-0



Pris kr 25,00

Publikasjonen utgis i kommisjon hos H. Aschehoug & Co. og  
Universitetsforlaget, Oslo, og er til salgs hos alle bokhandlere.

---

ISBN 82-537-2271-0  
ISSN 0332-8422