Interne notater

STATISTISK SENTRALBYRÅ

82/25 26. juli 1982

ESTIMERING AV VARIANSER I TIDSNYTTINGSUNDERSØKELSENE: DOKUMENTASJON AV UTVALGSPLANENE

av

Erling Siring

INNHOLD

1. INNLEDNING

I Statistisk Sentralbyrå har det vært gjennomført to tidsnyttingsundersøkelser, én i 1971-72 og én i 1980-81. En av de tingene en nå ønsker å kartlegge er om folks tidsbruk har endret seg i perioden mellom de to tidsnyttingsundersøkelsene. Videre ønsker en å kartlegge hvordan tidsbruken eventuelt har endret seg. For å kunne gjøre dette må en vite hvor sikre tallene i undersøkelsene er. Dette notatet dreier seg i stor grad om hvordan en skal estimere varianser i undersøkelsene. Videre blir det presentert en regel for om en kan påstå om det har skjedd endringer eller ikke.

For å kunne si noe om sikkerheten på tallene i undersøkelsen var det naturlig å begynne med å betrakte utvalgsplanene som ble brukt. I begge undersøkelsene ble Byråets standard utvalgsplan brukt i utgangspunktet. Etter at utvalgene var trukket, ble personene fordelt på tidsperioder. Metodene for dette har vært mangelfullt dokumentert, og vi har derfor funnet det nødvendig å presentere det vi har funnet ut om dem i dette notatet. Dette er gjort i kapittel 2.

Når en skal estimere mål for usikkerhet, slik som varians, kan en betrakte problemet fra to synsvinkler, fra en "modellsynsvinkel" og fra den synsvinkel som eksisterer innen den klassiske stikkprøveteori. I kapittel 3 beskriver vi forskjellen mellom disse to synsvinklene, og begrunner hvorfor vi har valgt "modellsynsvinkelen".

I kap. 4 presenteres modellen som vi har brukt, og i kap. 5 utledes de varianser som skal estimeres. I kapittel 6 presenteres forskjellige estimatorer som kan brukes for å estimere varianser til gjennomsnittstall. I avsnitt 6.2 presenteres estimatorer som er lette å beregne, men som overestimerer variansene. I avsnitt 6.3 presenteres en estimator som er forventningsrett, men som er vanskelig å beregne.

I kapittel 7 presenteres en regel for å kunne vurdere om det har skjedd endringer eller ikke. I kapittel 8 vurderer vi en annen og bedre estimator for gjennomsnittstall i befolkningen enn gjennomsnitt i utvalget.

2. UTVALGSPLANENE

Som nevnt i innledningen, har visse sider ved utvalgsplanene for tidsnyttingsundersøkelsene vært mangelfullt dokumentert. Vi skal her presentere det vi har funnet ut om utvalgsplanene.

2.1. Tidsnyttingsundersøkelsen 1971-72

Som i de fleste intervjuundersøkelsene i Statistisk Sentralbyrå ble utvalget trukket i to trinn. I første trinn ble det trukket primære utvalgsområder, og i annet trinn ble utvalget trukket fra de utrukne områdene. I [1] er Byråets generelle utvalgsplan pr. 1971 beskrevet.

Til undersøkelsen ble det i alt trukket ut 5 215 personer i alderen 15 - 74 år. Hver person i utvalget skulle føre dagbok to eller tre dager på rad. Dagbokperiodene kunne begynne på tre forskjellige ukedager, tirsdag, torsdag eller lørdag. De som begynte på en lørdag, førte dagbok tre dager på rad. De andre førte dagbok to dager på rad. I alt var det 3 040 personer som førte dagbok.

Undersøkelsesperioden var et helt år, dvs. dagbokperiodene dekket alle dager i et helt år. Når det gjelder fordelingen av personer på dagbokperioder, ble det gjort slik at omtrent like mange personer skulle føre dagbok i hver dagbokperiode.

I [2] står det mer om hvordan utvalget ble trukket, og dagbokføringen etc.

2.2. Tidsnyttingsundersøkelsen 1980-81

Undersøkelsen var en personundersøkelse. Ca. 5 200 personer i alderen 16-74 år ble trukket fra Byråets utvalgsområder. Som i den forrige tidsnyttingsundersøkelsen varte undersøkelsesperioden et helt år. Undersøkelsesperioden begynte i uke 40 i 1980 og varte til og med uke 39 i 1981.

Hver person i utvalget førte dagbok i to dager på rad. På hver dag i undersøkelsesperioden begynte "nye" personer å føre dagbok. En sørget for at det var omtrent like mange som begynte på sin todagers-periode hver dag.

Utvalget til undersøkelsen ble trukket i henhold til Byråets standard-utvalgsplan, som er

beskrevet i [3]. Sett II av utvalgsområder ble brukt.

Etter at utvalget var trukket, ble det delt i 4 puljer, som skulle føre dagbok i forskjellige tidsperioder. Puljeinndelingen foregikk ved at en delte utvalgsområdene i 4 grupper. Ved grupperingen tok en hensyn til folketall, næringsstruktur, folketetthet og geografisk beliggenhet. De 4 gruppene ble laget så like hverandre som mulig m.h.t. de nevnte variablene. Hver av de 4 gruppene (puljene) skulle være mest mulig "representativ" for landet.

Hensikten med puljeinndelingen var å få samlet arbeidet til intervjuerne til enkelte uker. De forskjellige puljene hadde hver sine undersøkelsesperioder, der hver undersøkelsesperiode besto av hver fjerde uke i undersøkelsesåret. Under er en oversikt over hvilke uker personene i de forskjellige puljene skulle føre dagbok i.

	Puljenr.			
Kvartal	1	2	3	4
	Dagbokuker	Dagbokuker	Dagbokuker	Dagbokuker
4. kvartal 1980	40	41	42	43
	44	45	46	47
	48	49	50	51
	52			
1. kvartal 1981		1	2	3
	4	5	6	7
	8	9	10	11
	12	13		
2. kvartal 1981			14	15
	16	17	18	19
	20	21	22	23
	24	25	26	
3. kvartal 1981				27
	28	29	30	31
	32	33	35	34
	36	37	38	39

Pulje 1 besto av følgende utvalgsområder: 202, 207, 210, 216, 217, 221, 223, 226, 232, 237, 244, 248, 254, 257, 261, 263, 268, 277, 282, 286, 290, 295, 297, 415, 416, 417, 418, 521, 525, 623, og 821.

Pulje 2 besto av utvalgsområdene 204, 211, 215, 218, 220, 227, 228, 230, 235, 238, 239, 246, 252, 258, 264, 271, 273, 275, 281, 284, 287, 289, 296, 298, 419, 420, 421, 523, 524, 622, 722 og 922.

Pulje 3 besto av utvalgsområdene 201, 205, 208, 212, 222, 224, 229, 231, 241, 242, 243, 245, 249, 253, 259, 260, 262, 267, 270, 274, 276, 280, 288, 292, 294, 299, 422, 423, 424, 425 og 522.

Pulje 4 besto av utvalgsområdene 206, 209, 213, 214, 219, 225, 233, 234, 236, 240, 247, 251, 255, 265, 266, 269, 272, 278, 279, 283, 285, 291, 293, 426, 427, 428, 526, 621, 721, 822 og 921.

Etter at puljeinndelingen var foretatt, ble hvert IO tildelt ukenr. og ukedag for første dagbokdag. Dette foregikk systematisk innenfor hver pulje, etter at utvalget var blitt sortert på utvalgsområde x familienummer x personkode . Når det gjelder sorteringen etter familienummeret, var denne dag x mnd. x år x personnr.. Siden dag og måned kom først i sorteringen, kan en forutsette at utvalget ble tilfeldig sortert innen hvert utvalgsområde.

Hvordan tildelingen av ukenr. og ukedag for første dagbokdag foregikk, kan best vises ved et eksempel. Vi tar for oss pulje 2.

Omr.	IO-nr.	Ukenr.	Ukedag
1	1	1	Sø
"	2	5	Ma
"	3	9	Ti
"	4	13	On
"	5	17	To
"	6	21	Fr
"	7	25	Lø
"	8	29	Sø
"	9	33	Ma
"	10	37	Ti
"	11	41	On
"	12	45	To
"	13	49	Fr
"	14	1	Lø
2	15	5	Sø
2	16	9	Ma
"	17	13	Ti

Den systematiske tildelingen av dagbokdager førte til at like mange personer ble valgt ut til hver dag i undersøkelsesperioden, og at den geografiske fordelingen i utvalget var omtrent den samme fra dag til dag.

Hvis vi forutsetter at sorteringen av personer var tilfeldig innenfor hvert utvalgsområde, og at hvert utvalgsområde hadde like stor sannsynlighet for å komme i den ene som den andre puljen, hadde vi et selvveiende utvalg for hver dag i undersøkelsesperioden.

3. UTVALGSVARIANS KONTRA MODELLVARIANS

De vanlige mål for usikkerhet er varians og standardavvik, der det siste er kvadratroten av det første. Det er også disse målene vi skal konsentrere oss om her, og i første rekke varians.

Når en skal estimere varians til tall fra en utvalgsundersøkelse, kan en komme fram til to forskjellige resultater ettersom en bruker to forskjellige betraktningsmåter. Den ene betraktningsmåten er den som finnes innen den klassiske stikkprøveteori, og den andre er generert ut fra modellbetraktninger.

Vi skal i dette avsnittet komme litt inn på forskjellen mellom de to betraktningsmåtene, og hvorfor vi i tidsnyttingsundersøkelsene har valgt å estimere "modellvarians".

Vi skal anta at det til hver person i populasjonen er knyttet en størrelse X, som kan måles uten nevneverdig målefeil. X kan være både en kontinuerlig variabel, slik som tid brukt til husarbeid, eller en såkalt dikotom variabel,

$$\text{der } X = \begin{cases} 1 \text{ hvis personen har utført en bestemt aktivitet} \\ 0 \text{ ellers} \end{cases}$$

Vi antar at vi trekker et utvalg av personer for å kunne si noe om X-ene i populasjonen.

3.1. Klassisk stikkprøveteori

Innen den klassiske stikkprøveteori betraktes alle X-verdiene i populasjonen som faste, ikke-stokastiske verdier. Ved utvalgsundersøkelser forutsettes at det eneste usikkerhetsmomentet er det som følger av at en observerer bare en del av populasjonen.

All inferens i den klassiske stikkprøveteori bygger på at utvelgingsprosessen er stokastisk, dvs. at alle individer har en på forhånd fastsatt sannsynlighet for å komme med i utvalget. Videre forutsettes det at den stokastikken som innføres når utvalgsplanen lages, er den eneste form for stokastikk i en utvalgsundersøkelse. Alle målinger betraktes som faste størrelser uten usikkerhet.

Når en skal beregne usikkerheten i resultatene, tar en følgelig kun utgangspunkt i selve utvelgingsmetoden. Et mål for usikkerhet beregnet på dette grunnlaget, er utvalgsvarians. Vi skal definere hva utvalgsvarians er.

Anta at det er N enheter i populasjonen, og at vi trekker et utvalg på n. Anta videre at vi er interessert i å estimere $\bar{X} = \frac{1}{N} \sum_{i=1}^{N} X_i$, der $\bar{X}$ er gjennomsnittet i hele populasjonen. X_i betegner X-verdien til person nr. i. La $\bar{X}_s = \frac{1}{n} \sum_{i \in s} X_i$ være gjennomsnittet i utvalget. Vi antar at utvelgingsprosessen er stokastisk, og slik at $E\bar{X}_s = \bar{X}$, dvs. at $\bar{X}s$ er forventningsrett.

Utvalgsvarians defineres da som:

$$\sigma^2 = \text{var } \bar{X}_s = \sum_s p(s)(\bar{X}s - \bar{X})^2,$$

der summasjonen går over alle mulige utvalg s av størrelse n, og p(s) er sannsynligheten for at utvalget s skal bli trukket.

For å gjøre det klarere hva σ^2 er, kan vi anta at vi gjentar utvalgsundersøkelsen k ganger. Vi antar videre at vi observerer gjennomsnittet $\bar{X}_s$ i undersøkelse nr. s. $\bar{X}_1, \bar{X}_2, \ldots, \bar{X}_s, \ldots, \bar{X}_k$ vil være observasjoner generert fra en sannsynlighetsfordeling. σ^2 er variansen i denne fordelingen, og kan skrives på følgende form:

$$\sigma^2 = \lim_{k \to \infty} \frac{1}{k} \sum_{s=1}^{k} (\bar{X}_s - \bar{X})^2$$

Hvis utvalget s er trukket ved en enkel, tilfeldig trekning, er følgende estimator for σ^2 forventningsrett.

$$\hat{\sigma}^2 = (1 - \frac{n}{N}) \frac{1}{n(n-1)} \sum_{i \in s} (X_i - \bar{X}_s)^2 .$$

Hvis en trekker utvalget i flere trinn slik som i tidsnyttingsundersøkelsene, vil $\hat{\sigma}^2$ i de fleste tilfeller underestimere utvalgsvariansen. I de fleste praktiske situasjoner har vi erfaring for at utvalgsvariansen har en tendens til øke når en innfører flere trinn i trekkingen.

3.2. Modellbasert tankegang

Innenfor den modellbaserte tankegangen tenker en seg at X-verdien knyttet til en person ikke er fast, men stokastisk. Dvs. en tenker seg at X er realisasjonen av en stokastisk variabel. De $(X_1, \ldots, X_N)$ verdiene i populasjonen forutsettes generert av den simultane fordelingen til vektoren $(X_1, \ldots, X_N)$ av stokastiske variable.

Om $(X_1, \ldots, X_N)$ gjøres det modellforutsetninger som skal reflektere bakgrunnskunnskaper. Slike modellforutsetninger kan være både kompliserte og svært enkle. Modellforutsetningene kan f.eks. være så enkle som å anta at $X_1, \ldots, X_N$ er uavhengige, identisk fordelte med forventning u og varians τ^2.

Anta at vi har gjort en utvalgsundersøkelse og at $\bar{X}_s$ er gjennomsnittet i utvalget. Med modellvariansen til $\bar{X}_s$ mener vi variansen gitt modellspesifikasjonene og utvalget.

Hvis $X_1, \ldots, X_N$ er uavhengige med varians τ^2, blir modellvariansen til $\bar{X}_s$ lik $\frac{\tau^2}{n}$. En forventningsrett estimator for var $\bar{X}_s = \frac{\tau^2}{n}$ er:

$$v = \frac{1}{n(n-1)} \sum_{i \in S} (X_i - \bar{X}_s)^2$$

Når n er liten i forhold til N, har vi at $v \approx \hat{\sigma}^2$ (jfr. avsnitt 3.1) til tross for at v og $\hat{\sigma}^2$ er estimatorer for to forskjellige ting. Det er forøvrig bare i enkelte tilfeller at utvalgsvariansen er omtrent like stor som modellvariansen.

Ved estimering av modellvarians tar en ikke hensyn til hvordan utvalget er trukket. En ser på utvalget som "fast", og estimerer den variansen som er en del av modellspesifikasjonen.

Modelltankegangen er svært mye brukt. Ved regresjonsanalyse f.eks., forutsettes en modell av følgende type:

$$X_i = \sum_{j=1}^{p} \beta_j Y_{ij} + V_i \qquad \text{for } i=1, \ldots, N$$

$Y_{i1}, \ldots, Y_{ip}$ er et sett av p forklaringsvariable, $\beta_1, \ldots, \beta_p$ er regresjonskoeffisienter og V_i er en stokastisk variabel med forventning 0 og varians τ^2 for alle i.

Fordelen ved å bruke en modell er at en kan få utnyttet kunnskaper som en har fra før av. Hvis en f.eks. ut fra tidligere erfaring vet at det er en tilnærmet lineær sammenheng mellom to eller flere variable, kan en bruke en regresjonsmodell som den over og oppnå større sikkerhet i resultatene.

3.3. Drøfting av de to betraktningsmåtene

De to betraktningsmåtene har begge gode og dårlige egenskaper. Om en bør bruke målet utvalgsvarians eller modellvarians som mål for usikkerhet, mener vi bør variere med problemstillingen. Noen statistikere mener at en nesten alltid bør bruke modellvarians, mens andre mener at en nesten alltid bør bruke utvalgsvarians. Det er derfor ikke alltid opplagt når en bør bruke det ene eller det andre.

En av fordelene med den modellbaserte tankegangen er at en kan få utnyttet eventuelle kunnskaper som en har fra før av. Hvis modellen er riktig og gjelder for alle personer i populasjonen, både i og utenfor utvalget, er modellframgangsmåten uangripelig.

Problemet ved den modellbaserte framgangsmåten er å lage en god modell. Hvis modellen er dårlig, kan inferensen bli feil. Dette kan skje hvis en ved analyse av data bruker metoder som ikke er robuste overfor avvik i modellen. I denne sammenheng må det imidlertid nevnes at det har vært jobbet en del med å utvikle metoder som er robuste overfor hele klasser av modeller.

Fordelen med å bruke utvalgsvarians som mål for usikkerhet er at en slipper å gjøre modellforutsetninger som kan være mer eller mindre feil.

I visse situasjoner synes imidlertid utvalgsvarians ikke å være særlig godt egnet som mål for usikkerhet. Vi skal gi et eksempel på dette. Eksempel: Det skal nå gjøres en undersøkelse blant kvinnelige akademikere for å kartlegge de problemene de har møtt i arbeidslivet. I undersøkelsen vil en bl.a. være interessert i å studere utviklingen som har skjedd over tid. Spesielt vil en være interessert i yrkeshistorien til mindre faggrupper, slik som sosiologene. Tabell 1 viser hvor mange kvinnelige sosiologer som er utdannet i Norge i forskjellige 5-årsperioder i tiden fra 1935 - 74.

Tabell 1. Antall kvinnelige sosiologer utdannet i forskjellige femårsperioder

	Tidsperiode				
I alt	1935-54	1955-59	1960-64	1965-69	1970-74
39	0	5	2	7	25

Vi ser at i tiden 1935 - 1974 ble det utdannet i alt 39 kvinnelige sosiologer. Alle disse skal være med i undersøkelsen. Fra et klassisk stikkprøveteori - synspunkt vil det da ikke være noen usikkerhet knyttet til de utsagnene en måtte komme med om kvinnelige sosiologers yrkeshistoriske utvikling. Dette er også logisk riktig når en kun ønsker å kartlegge yrkeshistorien til de 39 personene.

Hvis en derimot ønsker å si noe om den generelle utviklingen over tid av kvinnelige sosiologers arbeidssituasjon, er det klart at det må være feil at det ikke er noen usikkerhet knyttet til resultatene. Anta f.eks. at de 25 sosiologene som ble utdannet i 1970 - 1974, hadde blitt utdannet i tiden 1960 - 1964. Er det da sikkert at de ville ha fått den samme yrkeshistorien som de to som ble utdannet i det nevnte tidsrommet? Svaret må bli nei.

Vi ser at når en ønsker å trekke konklusjoner om den generelle utviklingen over tid, dukker det opp en form for usikkerhet i resultatene som ikke kommer til uttrykk ved utvalgsvarians. Ved slike problemsstillinger blir det naturlig å forutsette at det er tilfeldigheter knyttet til hva hver person i populasjonen opplever eller foretar seg. Dette er nettopp modelltankegang.

Når det gjelder valget mellom modellvarians og utvalgsvarians som mål for usikkerhet, ønsker vi ikke her å spesifisere nøyaktig i hvilke situasjoner en bør bruke det ene eller andre målet. Vi mener imidlertid at utvalgsvarians er mest tiltalende hvis en ikke har en god modell til rådighet, og hvis målsettingen med utvalgsundersøkelsen er en av følgende:

i) Å estimere antall (eller andelen) personer i populasjonen med et bestemt kjennemerke på en bestemt dag.

ii) Å estimere en gjennomsnittsverdi (f.eks. gjennomsnittsinntekt) i populasjonen på en bestemt dag.

Når målsettingen med en utvalgsundersøkelse er å kartlegge diverse former for sammenhenger eller trender over tid, er det i mange tilfeller nødvendig å angripe problemene fra et modellsynspunkt, fordi utvalgsteorien er mangelfull som analyseverktøy. I enkelte tilfeller, slik som i eksemplet, faller det helt naturlig å anvende modelltankegangen.

I tidsnyttingsundersøkelsene kjenner vi ikke utvalgsplanene som ble brukt i detalj. Når det gjelder alle trinnene som ble brukt i trekkingen av utvalget til Tidsnyttingsundersøkelsen 1980-81 (jfr. kap. 2.2), er det umulig å vurdere effekten av disse. Det er derfor ikke mulig å finne en forventningsrett estimator for utvalgsvariansen i tidsnyttingsundersøkelsene. Om det hadde vært mulig, ville estimatorene ha blitt svært kompliserte å beregne.

Valget har altså stått mellom å estimere modellvarians eller ikke å estimere noe mål for usikkerhet i det hele tatt. Variansestimatene skal brukes til å vurdere om det har skjedd noen endring i folks tidsbruk i perioden mellom de to tidsnyttingsundersøkelsene. I en slik problemstilling mener vi at modellvarians er et godt egnet mål for usikkerhet. Med de svake modellforutsetningene vi har gjort, mener vi at vi har valgt en forsvarlig framgangsmåte.

4. FORMULERING AV MODELLEN

Som nevnt tidligere har vi valgt å bruke modellvarians som mål for usikkerhet på gjennomsnittstall i tidsnyttingsundersøkelsene. I dette avsnittet presenteres modellen som blir brukt.

Først følger en del definisjoner.

La Y_{ij} = Tid brukt til et bestemt gjøremål på dag nr. j av person nr. i.

Y_{ij} kan også være en dikotom variabel .

$$Y_{ij} = \begin{cases} 1 \text{ hvis person nr. i har utført en bestemt aktivitet på dag nr. j} \\ 0 \text{ ellers} \end{cases}$$

Vi ser på Y_{ij} som en stokastisk variabel

Det er én Y for hver person og hver dag

I populasjonen vår har vi én enhet for hver person og hver dag. Vi skal kalle en slik enhet for en persondag. Populasjonen vår består av persondager gjennom et helt år (undersøkelsesperioden i den aktuelle tidsnyttingsundersøkelse). Vi lar (i, j) betegne person nr. i og dag nr. j.

Vi gjør følgende modellforutsetninger:

i) Populasjonen kan deles opp i grupper etter person- og dagkjennetegn, slik at Y-ene blir identisk fordelte innen hver gruppe.

ii) La K betegne en vilkårlig gruppe av identisk fordelte Y-er. Da gjelder for alle (i, j) med i K:

(1) $EY_{ij} = \xi_K$

(2) $\text{var } Y_{ij} = \sigma^2_K$

(3) $\rho(Y_{ij}, Y_{ij-1}) = \rho_K$ hvis $(i, j) \in K$ & $(i, j-1) \in K$

Med $\rho(X, Y)$ mener vi korrelasjonskoeffisienten mellom X og Y.

Videre forutsetter vi generelt at:

(4) $\rho\ (Y_{ij}, Y_{ls}) = 0$ for all j og s hvis $i \neq l$,

(5) $\rho\ (Y_{ij}, Y_{ik}) = \rho_{KL}$ hvis $k = j \pm 1$ og $(i, j) \in K$ og $(i, k) \in L$

(6) $\rho\ (Y_{ij}, Y_{ik}) = \gamma_{KL}$ hvis $k = j \pm 2$ og $(i, j) \in K$ og $(i, k) \in L$

K og L betegner to forskjellige "homogene" persondaggrupper

Vi antar altså at det er uavhengighet mellom Y-verdier fra to forskjellige personer, og at Y-verdier fra en og samme person kan være korrelerte. Vi tenker oss altså at hvis en person går på kino en dag, kan det påvirke sannsynligheten for at personen går på kino dagen etter.

I tidsnyttingsundersøkelsene førte 10-ene dagbok to eller tre dager på rad. De som førte dagbok tre dager på rad, hadde sin første dagbokdag på en lørdag.

Vi ønsker ikke å definere nøyaktig hvordan en "homogen" persondaggruppe kan se ut, men vi forutsetter at en slik gruppe ikke kan inneholde både lørdags- og mandagsobservasjoner. Av denne forutsetningen følger at det i utvalget er høyst to observasjoner fra en og samme person innenfor en "homogen" persondaggruppe.

I resten av notatet forutsetter vi at modellen gjelder.

La ξ^G betegne forventningen eller det teoretiske "gjennomsnittet" i delgruppe G i populasjonen. G lar vi betegne en gruppe av persondager som vi kan være interessert i å studere, slik som alle personer alle dager, menn hverdager, kvinner søndager osv.. Innenfor gruppe G kan det være et vilkårlig antall homogene persondaggrupper.

ξ^G betegner altså en størrelse som vi er interessert i å estimere, og kan skrives som en veid sum av ξ_K-er:

$$\xi^G = \frac{1}{D_G} \sum_{K \subseteq G} D_K \xi_K$$

D_K og D_G er antall persondager i henholdsvis gruppe K og G i populasjonen.

Som estimator for ξ^G brukes i tidsnyttingsundersøkelsene.

$$\bar{Y}_G = \frac{1}{d_G} \sum_{(i,j) \in G \cap U} Y_{ij}$$, der d_G er antall persondager i gruppe G i utvalget, U er utvalget av persondager, og $G \cap U$ er "snittet" av G og utvalget. $\bar{Y}_G$ er rett og slett gjennomsnittet av Y-verdier i gruppe G i utvalget.

La d_K betegne antall persondager i gruppe K i utvalget. Forventningen til $\overline{Y}_G$ blir:

$$E\overline{Y}_G = E[E(\overline{Y}_G \mid d_K\text{-ene})]$$

$$= E[\frac{1}{d_G} \sum_{K \subseteq G} d_K \xi_K]$$

Hvis vi forutsetter at alle personer i populasjonen hadde samme sannsynlighet for å bli trukket ut til hver dag i undersøkelsesperioden, har vi:

$$E\overline{Y}_G = E[\frac{1}{d_G} \sum_{K \subseteq G} d_K \xi_K] = \frac{1}{D_G} \sum_{K \subseteq G} D_K \xi_K = \underline{\xi}^G$$

dvs. at $\overline{Y}_G$ er en forventningsrett estimator under nevnte forutsetning.

Vi vil heretter forutsette at vi har en del observasjoner innen hver homogen persondaggruppe eller "K-gruppe". Med så omfattende stratifisering og med så store utvalg som en har brukt i tidsnyttingsundersøkelsene, kan en anta at $\frac{d_K}{d_G} \approx \frac{D_K}{D_G}$. Vi tror derfor det betyr lite at vi heretter betrakter d_K og d_G som faste, ikke-stokastiske tall. $\overline{Y}_G$ kan skrives på følgende form:

$$\overline{Y}_G = \frac{1}{d_G} \sum_{K \varepsilon G} d_K \overline{Y}_K,$$

der

$$\overline{Y}_K = \frac{1}{d_K} \sum_{(i,j) \varepsilon K \cap u} Y_{ij}$$

5. VARIANSEN TIL GJENNOMSNITTET

Vi skal i dette kapitlet finne et uttrykk for variansen til både $\overline{Y}_K$ og $\overline{Y}_G$. Vi skal først se på variansen til $\overline{Y}_K$.

I dette notatet vil vi ikke ta stilling til hvordan en skal gå fram for å få delt populasjonen opp i "homogene" persondaggrupper. Vi bare forutsetter at slike grupper finnes, og at K er en slik gruppe.

Heretter skal vi begrense oss til å betrakte Y-observasjonene i utvalget, og "overse" resten av populasjonen. Når vi skriver sumtegn, vil vi mene summering over utvalget, og med gruppe K vil vi mene den delen av gruppe K som hører til utvalget.

Vi antar nå at det er k personer i gruppe K, at k_1 av disse førte dagbok i én dag, og at k_2 førte dagbok i to etterfølgende dager. Som tidligere lar vi d_K betegne antall persondagobservasjoner i K.

Vi har

$$d_K = k_1 + 2k_2$$

La K_1 og K_2 betegne mengdene av de personer som har henholdsvis én og to dagbokdager innen gruppe K, og la Y_{i1} og Y_{i2} være det som er observert for person nr. i på henholdsvis første og andre dagbokdag innen gruppe K.

$$\bar{Y}_K = \frac{1}{d_K}\left\{\sum_{i\varepsilon K_1} Y_{i1} + \sum_{i\varepsilon K_2}(Y_{i1} + Y_{i2})\right\}$$

$$\operatorname{var}\bar{Y}_K = \frac{1}{d_K^2}\left\{\sum_{i\varepsilon K_1}\operatorname{var} Y_{i1} + \sum_{i\varepsilon K_2}(\operatorname{var} Y_{i1} + \operatorname{var} Y_{i2} + 2\operatorname{cov}(Y_{i1}, Y_{i2}))\right\}$$

$$= \frac{1}{d_K^2}\left\{\sum_{(i,j)\varepsilon K}\operatorname{var} Y_{ij} + 2\sum_{i\varepsilon K_2}\operatorname{cov}(Y_{i1}, Y_{i2})\right\}$$

$$= \frac{1}{d_K^2}(d_K\sigma_K^2 + 2k_2\sigma_K^2\rho_K)$$

$$= \frac{\sigma_K^2}{d_K^2}(d_K + 2k_2\rho_K) \qquad (7)$$

Innenfor en vilkårlig persondaggruppe G tenker vi oss at det kan være enkelte personer som har tre dagbokdager. La G_1, G_2 og G_3 betegne mengden av de personer som har henholdsvis en, to og tre dagbokdager innenfor gruppe G, og la Y_{i1}, Y_{i2} og Y_{i3} være det som er observert for person nr. i på henholdsvis første, andre og tredje dagbokdag innen gruppe G. Vi må anta at Y_{i1}, Y_{i2} og Y_{i3} alle er korrelerte med hverandre. $\bar{Y}_G$ kan nå skrives på følgende form:

$$\bar{Y}_G = \frac{1}{d_G}\left\{\sum_{i\varepsilon G_1} Y_{i1} + \sum_{i\varepsilon G_2}(Y_{i1} + Y_{i2}) + \sum_{i\varepsilon G_3}(Y_{i1} + Y_{i2} + Y_{i3})\right\}$$

Vi finner da for variansen til $\bar{Y}_G$:

$$\operatorname{var}\bar{Y}_G = \frac{1}{d_G^2}\left\{\sum_{i\varepsilon G_1}\operatorname{var} Y_{i1} + \sum_{i\varepsilon G_2}(\operatorname{var} Y_{i1} + \operatorname{var} Y_{i2} + 2\operatorname{cov}(Y_{i1}, Y_{i2})) + \sum_{i\varepsilon G_3}(\operatorname{var} Y_{i1} + \operatorname{var} Y_{i2} + \operatorname{var} Y_{i3} + 2\operatorname{cov}(Y_{i1}, Y_{i2}) + 2\operatorname{cov}(Y_{i1}, Y_{i3}) + 2\operatorname{cov}(Y_{i2}, Y_{i3}))\right\} \qquad (8)$$

Vi tar oss ikke bryet med å skrive (8) på en parametrisk form.

6. ESTIMERING AV VARIANSEN

I dette kapitlet skal vi presentere estimatorer for variansene til $\bar{Y}_K$ og $\bar{Y}_G$. I avsnitt 6.1 presenteres en forventningsrett estimator for variansen til $\bar{Y}_K$.

I avsnitt 6.2 presenteres estimatorer for var $\bar{Y}_G$ som er lette å beregne, men som overestimerer var $\bar{Y}_G$. I avsnitt 6.3 presenteres en estimator for var $\bar{Y}_G$ som er forventningsrett, men som er vanskelig å beregne.

6.1. Estimering av variansen til gjennomsnitt for en homogen persondaggruppe

Estimatoren som presenteres i dette avsnittet, er en forventningsrett estimator for var $\bar{Y}_K$. Denne estimatoren kan også brukes som estimator for var $\bar{Y}_G$ hvis det ikke er noen personer med mer enn to dagbokdager i gruppe G. Estimatoren vil da overestimere var $\bar{Y}_G$. Dette skal vi se nærmere på i avsnitt 6.3.

Definer $X_{i1} = Y_{i1} \; \forall \; i \in K_1$ og $X_{i2} = Y_{i1} + Y_{i2} \; \forall \; i \in K_2$

La $\bar{X}_{+1} = \frac{1}{k_1} \sum_{i \in K_1} X_{i1}$ og $\bar{X}_{+2} = \frac{1}{k_2} \sum_{i \in K_2} X_{i2}$

Vi har at

$$EX_{i2} = 2\xi_K \text{ og var } X_{i2} = 2\sigma_K^2 + 2\sigma_K^2\rho_K$$

$$= 2\sigma_K^2(1 + \rho_K)$$

Videre er alle X-ene uavhengige. Påstår nå at følgende estimator er en forventningsrett estimator for var $\bar{Y}_K$

$$(9) \qquad \sigma^{*2}(\bar{Y}_K) = \frac{1}{d_K^2} \left\{ \frac{k_1}{k_1 - 1} \sum_{i \in K_1} (X_{i1} - \bar{X}_{+1})^2 + \frac{K_2}{K_2 - 1} \sum_{i \in K_2} (X_{i2} - \bar{X}_{+2})^2 \right\}$$

Bevis:

$$E\sigma^{*2}(\bar{Y}_K) = \frac{1}{d_k^2} \left\{ \frac{k_1}{k_1 - 1} (k_1 - 1) \operatorname{Var} X_{i1} + \frac{k_2}{k_2 - 1} (k_2 - 1) \operatorname{Var} X_{i2} \right\}$$

$$= \frac{1}{d_K^2} \left\{ K_1\sigma_K^2 + k_2 2\sigma_K^2 (1 + \rho_K) \right\}$$

$$= \frac{\sigma_K^2}{d_K^2} (d_K + 2k_2\rho_K) = \operatorname{var} \bar{Y}_K \qquad \text{(jfr. (7))}$$

6.2. Estimering av variansen til gjennomsnittet for en vilkårlig personddaggruppe

For å få laget homogene persondaggrupper kan det være at målpopulasjonen må deles opp i så mange grupper at d_K blir liten. I slike tilfeller vil estimatorene for var $\bar{Y}_K$ som vi har sett på, ha stor usikkerhet. Nå er det imidlertid ikke ξ_K eller var $\bar{Y}_K$ vi er interessert i å estimere, men ξ^G og var $\bar{Y}_G$, der G er en vilkårlig mengde av persondager, f.eks. kvinner hverdager eller alle persondager i et helt år. I praksis vil G være en så stor mengde at vi kan regne med at antall observasjoner fra $G(d_G)$ vil være stort. (større enn 100). Vi forutsetter her at d_G er så stor at $\frac{a_G - 1}{d_G} \approx 1$.

Vi skal nå se på estimering av variansen til $\bar{Y}_G = \frac{1}{d_G} \sum_{(i,j) \in G} Y_{ij}$.

Til en estimator for var $\bar{Y}_G$ stiller vi følgende krav:

(i) Den må ikke underestimere var $\bar{Y}_G$, og ikke overestimere var $\bar{Y}_G$ for mye.

(ii) Den må være enkel å beregne ved eksisterende dataprogrammer.

Ad (i): Variansestimatene skal blant annet brukes til å teste om det har skjedd endringer i folks tidsbruk i perioden mellom Tidsnyttingsundersøkelsen 1971-72 og Tidsnyttingsundersøkelsen 1980-81.

Siden vi ikke kan konstruere eksakte tester, velger vi å basere oss på en konservativ framfor en liberal (radikal) test. Dette er årsaken til at vi ønsker å lage en estimator som ikke underestimerer var $\bar{Y}_G$.

Ad (ii): Det skal masseproduseres variansestimater, og estimatorene må derfor være enkle å beregne.

Vi skal nå presentere estimatorer for var $\bar{Y}_G$ som vi mener oppfyller (i) og (ii) på en tilfredsstillende måte. Vi finner det forøvrig nødvendig å skille mellom tre ulike situasjoner a), b), c) etter hvordan G er sammensatt:

a) Anta at ingen personer har mer enn en dagbokdag innenfor gruppe G.

Vi foreslår da følgende estimator for var $\bar{Y}_G$:

$$(10) \qquad s_1^2 = \frac{1}{d_G(d_G - 1)} \sum_{i \in G} (Y_i - \bar{Y}_G)^2$$

Y_i betegner persondagobservasjon nr. i. s_1^2 kan brukes når en skal estimere tall for én ukedag, slik som f.eks. lørdager eller søndager.

Vi vil vise at følgende gjelder for store verdier av d_G:

$$Es_1^2 \geq \frac{1}{d_G^2} \sum_{i \in G} \text{var } Y_i = \text{var } \bar{Y}_G.$$

Vi kan først anta at G er inneholdt i en gruppe K av persondager med identisk fordelte Y-er, dvs. at vi antar at alle Y-ene er identisk fordelte. Da gjelder etter et kjent resultat fra den generelle teori:

$$Es_1^2 = E \frac{1}{d_G(d_G - 1)} \sum_{i \in G} (Y_i - \bar{Y}_G)^2 = \frac{1}{d_G^2} \sum_{i \in G} \text{var } Y_i$$

$$= \frac{1}{d_G} \sigma_K^2 = \text{var } \bar{Y}_G$$

I dette tilfellet er altså s_1^2 en forventningsrett estimator for var $\bar{Y}_G$.

Anta nå at G består av persondager fra et vilkårlig antall "homogene" persondaggrupper. Vi deler G opp i slike grupper, og lar som vanlig K betegne ei slik gruppe.

$$Ed_G (d_G - 1)s_1^2 = E \sum_{i \in G} (Y_i - \bar{Y}_G)^2 = E \sum_{K \subseteq G} \sum_{i \in K} (Y_i - \bar{Y}_G)^2$$

$$= E \sum_{K \subseteq G} \sum_{i \in K} \left[(Y_i - \xi_K) - (\bar{Y}_G - \xi_K) \right]^2$$

$$= E \left\{ \sum_{K \subseteq G} \sum_{i \in K} (Y_i - \xi_K)^2 + \sum_{K \subseteq G} d_K (\bar{Y}_G - \xi_K)^2 - 2 \sum_{K \subseteq G} d_K (\bar{Y}_K - \xi_K)(\bar{Y}_G - \xi_K) \right\}$$

$$= \sum_{i \in G} \text{var } Y_i + E \sum_{K \subseteq G} d_K (\bar{Y}_G - \xi^G)^2 + \sum_{K \subseteq G} d_K (\xi_K - \xi^G)^2$$

$$- 2 E \sum_{K \subseteq G} d_K (\bar{Y}_K - \xi_K)(\bar{Y}_G - \xi^G)$$

$$(11) \qquad = \sum_{i \in G} \text{var } Y_i - d_G \text{ var } \bar{Y}_G + \sum_{K \subseteq G} d_K (\xi_K - \xi^G)^2$$

$$= d_G (d_G - 1) \operatorname{var} \bar{Y}_G + \sum_{K \subseteq G} d_K (\xi_K - \xi^G)^2$$

Vi har da:

$$Es_1^2 = \operatorname{var} \bar{Y}_G + \frac{1}{d_G(d_G - 1)} \sum_{K \subseteq G} d_K (\xi^G - \xi_K)^2$$

$$\geq \operatorname{var} \bar{Y}_G$$

Vi ser altså at s_1^2 oppfyller (i). I praksis vil en foretrekke å estimere standardavvik i stedet for varians. Som estimator for standardavviket brukes $s_1 = \sqrt{s_1^2}$. s_1 kan beregnes ved hjelp av subprogrammet CONDESCRIPTIVE i SPSS. s_1 er identisk med "standard error" i programmet. Dette betyr at s_1 er lett å beregne og at (ii) er oppfylt.

b) Anta at noen personer har to dagbokdager innen gruppe G, men at ingen har tre dagbokdager innen gruppe G

Vi foreslår da følgende estimator for $\operatorname{var} \bar{Y}_G$.

$$s_2^2 = \frac{1.7}{d_G(d_G - 1)} \sum_{i \in G} (Y_i - \bar{Y}_G)^2 = 1.7 s_1^2$$

I tidsnyttingsundersøkelsene førte IO-ene dagbok i to dager med unntak av de som i Tidsnyttingsundersøkelsen 1971-72 hadde sin første dagbokdag på en lørdag. Disse førte dagbok i tre dager. S_2^2 kan altså brukes ved estimering av varianser i Tidsnyttingsundersøkelsen 1980-81, og ved estimering av varianser til tall for hverdager eller lørdager og søndager i Tidsnyttingsundersøkelsen 1971-72.

Vi skal begrunne valget av denne estimatoren. Vi vil vise at følgende gjelder under visse forutsetninger:

$$(12) \qquad Es_2^2 \geq \frac{1.7}{d_G^2} \sum_{i \in G} \operatorname{var} Y_i \geq \operatorname{var} \bar{Y}_G$$

Vi skal først se på påstanden om at

$$\operatorname{var} \bar{Y}_G \leq \frac{1.7}{d_G^2} \sum_{i \in G} \operatorname{var} Y_i$$

Vi lar g_1 og g_2 betegne antall personer med henholdsvis en og to dagbokdager i gruppe G, og vi lar G_1, G_2, Y_{i1} og Y_{i2} betegne det samme som i kap. 5. Vi har da:

$$\operatorname{var} \bar{Y}_G = \frac{1}{d_G^2} \left\{ \sum_{i \in G_1} \operatorname{var} Y_{i1} + \sum_{i \in G_2} (\operatorname{var} Y_{i1} + \operatorname{var} Y_{i2} + 2 \operatorname{cov}(Y_{i1}, Y_{i2})) \right\}$$

I første omgang skal vi anta at G er inneholdt i en gruppe K av persondager med identisk fordelte Y-er. En får da:

$$\operatorname{var} \bar{Y}_G = \frac{1}{d_G^2} \left\{ \sum_{i \in G} \operatorname{var} Y_i + 2 \sum_{i \in G_2} \operatorname{cov}(Y_{i1}, Y_{i2}) \right\}$$

$$(13) \qquad = \frac{\sigma_K^2}{d_G^2} (d_G + 2 g_2 \rho_K)$$

Vi har beregnet en del empiriske korrelasjonskoeffisienter i Tidsnyttingsundersøkelsen 1971-72. Resultatene er presentert i tabell 2 i kapittel 8. Selv om vi har estimert korrelasjonskoeffisienter som vi har antatt skulle være spesielt store, er alle estimatene mindre enn 0,65.

Vi mener derfor at det vil være i ytterst få tilfeller, om det i det hele tatt er noen, at følgende ikke er oppfylt:

$$(14) \qquad \rho_K \leq \frac{g_1 + 2g_2}{2g_2} \cdot 0{,}7 = \frac{d_G}{2g_2} \cdot 0{,}7$$

Vi skal forutsette at (14) gjelder. Da får vi ved innsetting av (14) i (13):

$$(15) \qquad \operatorname{var} \bar{Y}_G \leq \frac{1{,}7}{d_G} \sigma_K^2 = \frac{1{,}7}{d_G^2} \sum_{i \in G} \operatorname{var} Y_i$$

Vi skal nå vise at følgende gjelder for store verdier av d_G:

$$(16) \qquad Es_2^2 = E \frac{1{,}7}{d_G(d_G - 1)} \sum_{i \in G} (Y_i - \bar{Y}_G)^2 \simeq \frac{1{,}7}{d_G^2} \sum_{i \in G} \operatorname{var} Y_i,$$

dvs. at vi må vise:

$$(17) \qquad E \frac{1}{d_G - 1} \sum_{i \in G} (Y_i - \bar{Y}_G)^2 \simeq \frac{1}{d_G} \sum_{i \in G} (\operatorname{var} Y_i)$$

Fra (11) har vi at:

$$(18) \qquad E \sum_{i \in G} (Y_i - \bar{Y}_G)^2 = \sum_{i \in G} \operatorname{var} Y_i - d_G \operatorname{var} \bar{Y}_G$$

Ved innsetting av (15) i (18) får vi:

$$E \sum_{i \in G} (Y_i - \bar{Y}_G)^2 \geq (1 - \frac{1{,}7}{d_G}) \sum_{i \in G} \operatorname{var} Y_i = \frac{d_G - 1{,}7}{d_G} \sum_{i \in G} \operatorname{var} Y_i$$

Hvis d_G er så stor at $\frac{d_G - 1{,}7}{d_G - 1} \simeq 1$,

har vi at (17) og (16) er oppfylt.

For store verdier av d_G har vi da fra (15) og (16):

$$Es_2^2 \simeq \frac{1{,}7}{d_G^2} \sum_{i \in G} \operatorname{var} Y_i \geq \operatorname{var} \bar{Y}_G$$

(i) er altså oppfylt for store verdier av d_G, og under forutsetning av at (14) gjelder.
Anta nå at G består av persondager fra et vilkårlig antall homogene persondaggrupper. Da har vi:

$$\operatorname{var} \bar{Y}_G = \frac{1}{d_G^2} \left\{ \sum_{i \in G} \operatorname{var} Y_i + 2 \sum_{i \in G_2} \operatorname{cov} (Y_{i1}, Y_{i2}) \right\}$$

$$= \frac{1}{d_G^2} \left\{ \sum_{i \in G} \operatorname{var} Y_i + 2 \sum_{i \in G_2} \sqrt{\operatorname{var} Y_{i1} \cdot \operatorname{var} Y_{i2}} \cdot \rho(Y_{i1}, Y_{i2}) \right\}$$

Hvis vi nå forutsetter at

$$(19)\qquad \rho(Y_{j1}, Y_{j2}) \leq \frac{0{,}7 \sum_{i\in G} \operatorname{var} Y_i}{\sum_{i\in G_2} (\operatorname{var} Y_{i1} + \operatorname{var} Y_{i2})} \qquad \text{for alle } j\in G_2,$$

gjelder at

$$\operatorname{var} \bar{Y}_G \leq \frac{1}{d_G^2} \left\{ \sum_{i\in G} \operatorname{var} Y_i + \frac{2 \sum_{i\in G_2} \sqrt{\operatorname{var} Y_{i1} \cdot \operatorname{var} Y_{i2}}}{\sum_{i\in G_2} (\operatorname{var} Y_{i1} + \operatorname{var} Y_{i2})} \cdot 0{,}7 \sum_{i\in G} \operatorname{var} Y_i \right\}$$

$$(20)\qquad \leq \frac{1{,}7}{d_G^2} \sum_{i\in G} \operatorname{var} Y_i$$

Vi skal nå vise at følgende gjelder for store verdier av d_G:

$$(21)\qquad Es_2^2 = E \frac{1{,}7}{d_G(d_G - 1)} \sum_{i\in G} (Y_i - \bar{Y}_G)^2 \geq \frac{1{,}7}{d_G^2} \sum_{i\in G} \operatorname{var} Y_i$$

Fra (11) har vi at:

$$Es_2^2 = \frac{1{,}7}{d_G(d_G - 1)} \left[\sum_{i\in G} \operatorname{var} Y_i - d_G \operatorname{var} \bar{Y}_G + \sum_{K\subseteq G} d_K (\xi_k - \xi^G)^2 \right]$$

Ved innsetting av (20) får vi:

$$Es_2^2 \geq \frac{1{,}7}{d_G(d_G - 1)} \left[(1 - \frac{1{,}7}{d_G}) \sum_{i\in G} \operatorname{var} Y_i + \sum_{K\subseteq G} d_K (\xi_k - \xi^G)^2 \right]$$

$$(22)\qquad \geq \frac{1{,}7(d_G - 1{,}7)}{d_G^2(d_G - 1)} \sum_{i\in G} \operatorname{var} Y_i$$

Hvis vi forutsetter at d_G er så stor at $\frac{d_G - 1{,}7}{d_G - 1} \simeq 1$, har vi at (21) er oppfylt. Fra (20) og (21) har vi da:

$$Es_2^2 \geq \frac{1{,}7}{d_G^2} \sum_{i\in G} \operatorname{var} Y_i \geq \operatorname{var} \bar{Y}_G$$

(i) er altså oppfylt under forutsetning av at d_G er "stor" og av at (19) er oppfylt.

Når en skal beregne s_2, beregner en s_1 som beskrevet under a), og multipliserer s_1 med $\sqrt{1{,}7}$ = 1,304

$$s_2 = \sqrt{1{,}7}\, s_1$$

c) Anta at en del personer har tre dagbokdager innenfor gruppe G

I denne situasjonen foreslår vi følgende estimator for var $\bar{Y}_G$.

$$(23) \qquad s_3^2 = \frac{2,4}{d_G(d_G - 1)} \sum_{i\varepsilon G} (Y_i - \bar{Y}_G)^2$$

s_3^2 tenkes brukt når en skal estimere varianser til tall for alle ukedager i Tidsnyttingsundersøkelsen 1971-72.

Vi skal vise at s_3^2 overestimerer var $\bar{Y}_G$, dvs. at $Es_3^2 \geq$ var $\bar{Y}_G$ for store verdier av d_G.

Vi skal først vise at var $\bar{Y}_G \leq \frac{2,4}{d_G^2} \sum_{i\varepsilon G}$ var Y_i under visse forutsetninger.

var $\bar{Y}_G$ kan skrives på følgende form (jfr. (8)):

$$\text{var } \bar{Y}_G = \frac{1}{d_G^2} \left\{ \sum_{i\varepsilon G} \text{var } Y_i + 2 \sum_{i\varepsilon G_2} \sqrt{\text{var } Y_{i1} \cdot \text{var } Y_{i2}} \cdot \rho(Y_{i1}, Y_{i2}) \right.$$

$$+ 2 \sum_{i\varepsilon G_3} \left(\sqrt{\text{var } Y_{i1} \cdot \text{var } Y_{i2}} \cdot \rho(Y_{i1}, Y_{i2}) + \sqrt{\text{var } Y_{i1} \cdot \text{var } Y_{i3}} \cdot \rho(Y_{i1}, Y_{i3}) \right.$$

$$\left. \left. + \sqrt{\text{var } Y_{i2} \cdot \text{var } Y_{i3}} \cdot \rho(Y_{i2}, Y_{i3}) \right) \right\}$$

Hvis vi nå forutsetter at

$$(24) \qquad \rho(Y_{j1}, Y_{j2}) \leq \frac{0,7 \sum_{i\varepsilon G_2} (\text{var } Y_{i1} + \text{var } Y_{i2})}{2 \sum_{i\varepsilon G_2} \sqrt{\text{var } Y_{i1} \cdot \text{var } Y_{i2}}} \geq 0,7 \qquad \text{for alle } j\varepsilon G_2,$$

$$(25) \qquad \text{og at } \rho(Y_{jk}, Y_{jm}) \leq \frac{0,7 \sum_{i\varepsilon G_3} (\text{var } Y_{ik} + \text{var } Y_{im})}{2 \sum_{i\varepsilon G_3} \sqrt{\text{var } Y_{ik} \cdot \text{var } Y_{im}}} \geq 0,7 \qquad \text{for alle } j\varepsilon G_3 \text{ og for } 1 \leq k < m \leq 3,$$

er det lett å vise at

$$\text{var } \bar{Y}_G \leq \frac{1}{d_G^2} \left\{ \sum_{i\varepsilon G} \text{var } Y_i + 0,7 \sum_{i\varepsilon G_2} (\text{var } Y_{i1} + \text{var } Y_{i2}) \right.$$

$$\left. + 1,4 \sum_{i\varepsilon G_3} (\text{var } Y_{i1} + \text{var } Y_{i2} + \text{var } Y_{i3}) \right\}$$

$$\leq \frac{2,4}{d_G^2} \sum_{i\varepsilon G} \text{var } Y_i$$

Tilsvarende til under b) er det nå lett å vise at for store verdier av d_G gjelder:

$$Es_3^2 \simeq \frac{2,4}{d_G^2} \left\{ \sum_{i\varepsilon G} \text{var } Y_i + \sum_{K\varepsilon G} d_K (\xi_K - \xi_G)^2 \right\}$$

$$\geq \frac{2,4}{d_G^2} \sum_{i\varepsilon G} \text{var } Y_i \geq \text{var } \bar{Y}_G$$

(i) er altså oppfylt under forutsetning av at d_G er stor og av at (24) og (25) er oppfylt.

Når en skal beregne s_3, beregner en s_1 som beskrevet under a), og multipliserer s_1 med $\sqrt{2,4} = 1,549$

$$s_3 = \sqrt{2,4}\, s_1$$

Vi skal til slutt i dette kapitlet foreta en oppsummering.

Når en skal estimere standardavvik i tidsnyttingsundersøkelsene, foreslår vi følgende framgangsmåte:

1) "Standard error" i subprogrammet CONDESCRIPTIVE I SPSS beregnes. Vi kaller dette s_1.
2) Hvis vi er i situasjon a), brukes s_1 som estimator for standardavviket.
3) Hvis vi er i situasjon b), brukes $s_2 = s_1 \cdot \sqrt{1,7}$ som estimator for standardavviket.
4) Hvis vi er i situasjon c), brukes $s_3 = s_1 \cdot \sqrt{2,4}$ som estimator for standardavviket.

Vi foreslår framgangsmåten over under forutsetning av at d_G er stor (større enn 50). Hvis d_G er mindre enn 50, bør en multiplisere estimatoren for standardavviket med korreksjonsfaktoren $\sqrt{\frac{d_G}{d_G - 1}}$.

6.3. En forventningsrett estimator for variansen

Estimatorene s_2^2 og s_3^2 som ble foreslått i forrige avsnitt, vil i mange tilfeller overestimere var $\bar{Y}_G$ betydlig. En kan si at det er prisen en må betale for å bruke estimatorer som er lette å beregne. I spesielt interessante tilfeller kan det være viktig å få estimert var $\bar{Y}_G$ mer nøyaktig. Vi skal her se på en estimator for var $\bar{Y}_G$ som er tilnærmet forventningsrett, men vanskeligere å beregne enn estimatorene i forrige avsnitt.

Det første vi gjør er å dele personene i G opp i 3 grupper ettersom de har en, to eller tre dagbokdager innenfor persondaggruppe G. Som tidligere kaller vi de tre gruppene av personer for henholdsvis G_1, G_2 og G_3.

Tilsvarende til i avsnitt 6.1 defineres nå for person nr. i:

$$X_i = \begin{cases} Y_{i1} & \text{for } i \varepsilon G_1 \\ Y_{i1} + Y_{i2} & \text{for } i \varepsilon G_2 \\ Y_{i1} + Y_{i2} + Y_{i3} & \text{for } i \varepsilon G_3 \end{cases}$$

Med Y_{im}, m = 1, 2, 3, menes Y-verdien som ble observert for person nr. i på m'te dagbokdag. Hvis Y_{im} er tid brukt til en bestemt aktivitet på dagbokdag nr. m, er X_i samlet tid brukt av person nr. i til aktiviteten i hele dagbokperioden. I henhold til modellspesifikasjonene i kap. 4 er alle X-ene uavhengige.

Vi deler nå gruppene G_1, G_2, og G_3 opp i grupper etter person- og dagkjennetegn slik at vi får identisk fordelte X-er innen hver delgruppe. La nå X_{ijkm} betegne X-verdien til person nr. i i gruppen med dagkjennetegn j og personkjennetegn k innenfor gruppe G_m.

$\bar{Y}_G$ kan nå skrives på følgende form:

$$\bar{Y}_G = \frac{1}{d_G} \sum_{m=1}^{3} \sum_k \sum_j \sum_i X_{ijkm}$$

Hvis vi forutsetter at X-ene er identisk fordelte innen hver delgruppe vil følgende estimator for var $\bar{Y}_G$ være forventningsrett:

$$s^2 = \frac{1}{d_G^2} \sum_{m=1}^{3} \sum_k \sum_j \frac{n_{jkm}}{n_{jkm} - 1} \sum_i (X_{ijkm} - \bar{X}_{+jkm})^2$$

Med n_{jkm} mener vi antall personer med personkjennetegn j og dagkjennetegn k innenfor gruppe G_m.

Med $\bar{X}_{+jkm}$ menes $\frac{1}{n_{jkm}} \sum_i X_{ijkm}$.

At s^2 er forventningsrett følger direkte fra kjente resultater innen matematisk statistikk.

Hvis n_{jkm} er stor, er $\frac{n_{jkm}}{n_{jkm} - 1} \simeq 1$, og $s^2 \simeq v^2 = \frac{1}{d_G^2} \sum_{m=1}^{3} \sum_k \sum_j \sum_i (X_{ijkm} - \bar{X}_{+jkm})^2$.

Uttrykket $U_{km} = \sum_j \sum_i (X_{ijkm} - \bar{X}_{+jkm})^2$ kan beregnes ved hjelp av subprogrammet ONEWAY i SPSS. Det må imidlertid først lages en dagtype-variabel som indikerer hva slags type dag hver person begynte å føre dagbok på. Hvis en har X som avhengig variabel og dagtypevariabelen som uavhengig variabel, vil "Sum of squares within groups" i " ONEWAY "I være identisk med uttrykket over. Hvis n_{jkm} for alle j, k og m er så stor at $v^2 \simeq s^2$, kan en bruke v^2 som estimator for var $\bar{Y}_G$. v^2 beregnes ved å regne ut $\frac{1}{d_G^2} \sum_{k,m} U_{km}$ manuelt.

Hvis noen av n_{jkm}-ene er små, bør en bruke s^2 som estimator for var $\bar{Y}_G$. Da kan $\sum_i (X_{ijkm} - \bar{X}_{+jkm})^2$ beregnes maskinelt ved hjelp av subprogrammet CONDESCRIPTIVE i SPSS for hver kombimasjon av j, k, og m. s^2 beregnes så manuelt.

Anta at en ved inndelingen av persondagobservasjonene i grupper foretar en for grov gruppering, slik at ikke X-ene vil være identisk fordelte innen hver gruppe. Da vil en overestimere var $\bar{Y}_G$. Dette følger av det generell resultatet at:

$$\sum_{i,j} (X_{ij} - \bar{X}_{++})^2 = \sum_{i,j} (X_{ij} - \bar{X}_{+j})^2 + \sum_{i,j} (\bar{X}_{+j} - \bar{X}_{++})^2 \geq \sum_{i,j} (X_{ij} - \bar{X}_{+j})^2$$

Hvis en lar være å foreta noen gruppering etter person- og dagkjennetegn, men kun deler personene inn i de tre gruppene G_1, G_2 og G_3, vil s^2 anta følgende form:

$$s_*^2 = \frac{1}{d_G^2} \sum_m \frac{g_m}{g_m - 1} \sum_{i \in G_m} (X_{im} - \bar{X}_{+m})^2$$

Vi ser at s_*^2 er av samme form som (9) i avsnitt 6.1. s_*^2 vil overestimere var $\bar{Y}_G$, men i de fleste tilfeller vil s_*^2 overestimere var $\bar{Y}_G$ mindre enn s_2^2 og s_3^2. I situasjoner der en har behov for å bruke en bedre estimator for var $\bar{Y}_G$ enn estimatorene s_2^2 og s_3^2 i avsnitt 6.2, vil s_*^2 være et godt alternativ.

7. SAMMENLIGNING AV RESULTATER FRA DE TO TIDSNYTTINGSUNDERSØKELSENE

En ønsker å sammenligne resultatene fra de to tidsnyttingsundersøkelsene for å se om det har skjedd endringer i tidsrommet mellom de to undersøkelsene. Ut fra dataene ønsker en å kunne trekke én av følgende tre konklusjoner:

(i) $\xi_{G1} > \xi_{G2}$

(ii) $\xi_{G1} < \xi_{G2}$

(iii) Ikke kunne si noe

Indeksene 1 og 2 betegner undersøkelsesperiodene for henholdsvis den første og andre tidsnyttings-undersøkelsen.

Problemet over blir kalt et tredesisjonsproblem.

$$\text{La } T = \frac{\bar{Y}_{G1} - \bar{Y}_{G2}}{\sqrt{S^2(\xi_{G1}) + S^2(\xi_{G2})}}$$

der $S^2(\xi_{G1})$ og $S^2(\xi_{G2})$ er estimatorer for henholdsvis var $\bar{Y}_{G1}$ og var $\bar{Y}_{G2}$.

La ϕ_ε være ε-fraktilen i den standardiserte normalfordelingen. Vi benytter oss nå av følgende framgangsmåte for å kunne velge mellom (i), (ii) og (iii).

Hvis $T > \phi_{1-\varepsilon}$, trekker en konklusjon (i)

Hvis $T \leq \phi_\varepsilon$, trekker en konklusjon (ii)

Hvis $\phi_\varepsilon < T \leq \phi_{1-\varepsilon}$, trekker en konklusjon (iii).

Hvis antall observasjoner er forholdsvis stort, vil metoden ha nivå ε.

Dersom en f.eks ønsker å sammenligne ξ_{G1} med flere forskjellige ξ-er i siste tidsnyttings-undersøkelse, bør en bruke en multippel regel for sammenligning. I [4] blir problemet med å teste flere delhypoteser simultant behandlet.

8. EN BEDRE ESTIMATOR ENN GJENNOMSNITTET

I dette avsnittet skal vi betrakte en "homogen" persondaggruppe K, og finne en bedre estimator for ξ_K enn gjennomsnittet. Som tidligere deler vi personene som har dagbokdager innenfor gruppe K, opp i to grupper K_1 og K_2, der K_1 betegner mengden av personer med én dagbokdag og K_2 mengden av personer med to dagbokdager innen gruppe K. Antall personer i gruppene K_1 og K_2 kaller vi henholdsvis k_1 og k_2.

Vi har at dagbokobservasjoner som stammer fra personene i gruppe K_2 er parvis korrelerte med hverandre. Bakgrunnen for at vi begynte å søke etter en annen estimator enn gjennomsnittet, var at vi stilte oss tvilende til om hver enkelt av observasjonene fra K_2 gir like mye informasjon som observasjonene fra K_1.

$$\text{La } \bar{Y}_1 = \frac{1}{k_1} \sum_{i \varepsilon k_1} Y_{i1} \quad \text{og} \quad \bar{Y}_2 = \frac{1}{2k_2} \sum_{i \varepsilon k_2} (Y_{i1} + Y_{i2})$$

$\bar{Y}_1$ er gjennomsnittet av observasjonene for de som førte dagbok én dag, og $\bar{Y}_2$ er gjennomsnittet av observasjonene for de som førte dagbok to dager.

La nå

$\xi_k^* = a\bar{Y}_1 + (1-a)\bar{Y}_2$ være en estimator for ξ_k. Hvis $a = \frac{k_1}{k_1+2k_2}$ er $\xi_k^* = \bar{Y}_k$.

Forventningen til ξ_k^* blir:

$$E\xi_k^* = a\xi_k + (1-a)\,\xi_k = \xi_k.$$

ξ_k^* er forventningsrett for alle verdier av a.

$$\text{var } \xi_k^* = a^2 \text{ var } \bar{Y}_1 + (1-a)^2 \text{ var } \bar{Y}_2$$

var ξ_k^* er en funksjon av a. Vi vil nå velge den a-verdien som minimerer var ξ_k^*. Vi kaller denne a-verdien a^*. Det er lett å vise at

$$(26) \qquad a^* = \frac{\text{var } \bar{Y}_2}{\text{var } \bar{Y}_1 + \text{var } \bar{Y}_2} = \frac{\frac{1}{2k_2}\sigma_k^2(1+\rho_k)}{\frac{1}{k_1}\sigma_k^2 + \frac{1}{2k_2}\sigma_k^2(1+\rho_k)} = \frac{k_1(1+\rho_k)}{2k_2+k_1(1+\rho_k)}.$$

Vi ser at a^* øker med økende verdi av ρ_k.

Hvis $\rho_k=0$, er $a^*=\frac{k_1}{k_1+2k_2}$ og $\xi_k^*=\bar{Y}_k$.

Hvis $\rho_k=1$, er $a^*=\frac{k_1}{k_1+k_2}$ og $\xi_k^*=\frac{1}{k_1+k_2}(k_1\bar{Y}_1+k_2\bar{Y}_2)$

I det siste tilfellet gis $\bar{Y}_2$ en vekt som om $\bar{Y}_2$ skulle være basert på k_2 observasjoner og ikke på $2k_2$ observasjoner. Dette er intuitivt rimelig siden Y_{i2} ikke gir noen tilleggsinformasjon til Y_{i1}, eller omvendt, når $\rho_k=1$.

Hvis $\rho_k=-1$ er $a^*=0$ og $\xi_k^*=\bar{Y}_2$. Når $\rho_k=-1$, har en at var $\bar{Y}_2=\frac{1}{2k_2}\sigma_k^2(1+\rho_k)=0$, og at $Y_{i1}+Y_{i2}=2\xi_k$ for alle i med i K_2. At $a^*=0$ i dette tilfellet, er derfor ikke urimelig.

Vi skal nå se på hvor mye en kan tjene på å bruke ξ_K^*. Et mål for dette er effisiensen til ξ_K^* relativt til $\bar{Y}_K$, som defineres som følger:

$$D=\frac{\text{var }\bar{Y}_K}{\text{var }\xi_K^*} \tag{27}$$

(28) Vi har at var $\bar{Y}_K=\frac{\sigma_K^2}{(k_1+2k_2)^2}[k_1+2k_2(1+\rho_k)]$ og var $\xi_K^*=a^{*2}\text{var }\bar{Y}_1+(1-a^*)^2\text{ var }\bar{Y}_2$.

Insetting av (26) gir:

$$\text{var }\xi_K^*=\frac{(\text{var }\bar{Y}_2)^2\text{var }\bar{Y}_1+(\text{var }\bar{Y}_1)^2\text{var }\bar{Y}_2}{\text{var }\bar{Y}_1+\text{var }\bar{Y}_2}$$

$$=\frac{\text{var }\bar{Y}_1\cdot\text{var }\bar{Y}_2}{\text{var }\bar{Y}_1+\text{var }\bar{Y}_2}$$

$$=\frac{\frac{1}{k_1}\sigma_K^2\cdot\frac{1}{2k_2}\sigma_K^2(1+\rho_K)}{\frac{1}{k_1}\sigma_K^2+\frac{1}{2k_2}\sigma_K^2(1+\rho_K)}$$

$$=\frac{\sigma_K^2(1+\rho_K)}{2k_2+k_1(1+\rho_K)} \tag{29}$$

(28) og (29) innsatt i (27) gir:

$$D=\frac{[k_1+2k_2(1+\rho_K)][2k_2+k_1(1+\rho_K)]}{(k_1+2k_2)^2(1+\rho_K)}$$

$$=1+\frac{2k_1k_2\rho_K^2}{(k_1+2k_2)^2(1+\rho_K)}$$

Vi ser at $D\geq 1$ og at D varierer med både ρ_K og forholdet mellom k_1 og k_2. Vi skriver nå $k_2=tk_1$, og får

$$D=1+\frac{2t\rho_K^2}{(1+2t)^2(1+\rho_K)}\,. \tag{30}$$

Vi skal nå for en gitt ρ_K betrakte D som en funksjon av t, og finne hvilken verdi av t som maksimerer D. Vi deriverer D med hensyn på t

$$D'(t) = \frac{2\rho_K^2(1-2t)}{1+\rho_K(1+2t)^3}$$

$D'(t) = 0$ for $t=\frac{1}{2}$ for alle verdier av $\rho_K \neq -1$. Siden $D=1$ for $t=0$ og $D\to 1$ når $t\to\infty$, må D ha et maksimumspunkt når $t=\frac{1}{2}$.

Dette medfører at det er mest å tjene på å bruke ξ_K^* framfor $\bar{Y}_K$ når $k_2=\frac{1}{2}k_1$, dvs. når $\bar{Y}_1$ og $\bar{Y}_2$ er basert på like mange dagobservasjoner. Dette gjelder for alle verdier av ρ_K unntatt o og -1. Når $\rho_K=0$, er D lik 1 for alle verdier av t, dvs. $\xi_K^* = \bar{Y}_K$.

Tabell 2 viser en del verdier av D for forskjellige verdier av t og ρ_K.

Tabell 2. D's variasjon med t og ρ_K

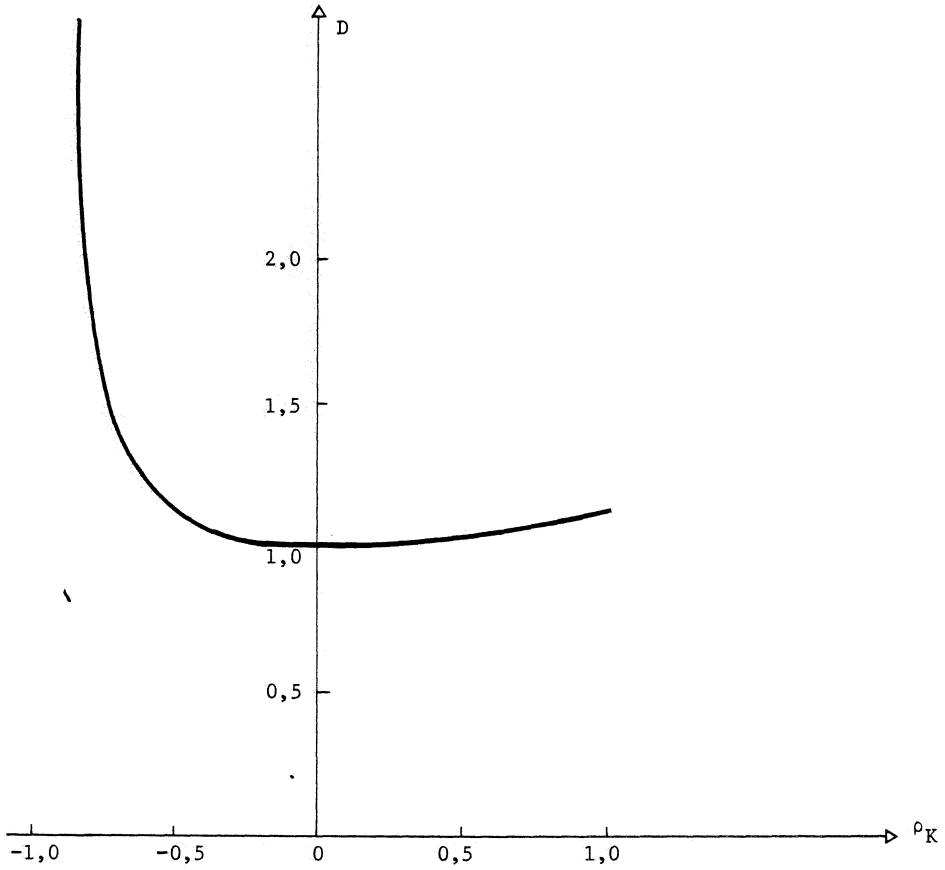
ρ_K \ t	0	$\frac{1}{4}$	$\frac{1}{2}$	1	2	∞
-0.99	1.000	22.780	25.503	22.780	16.682	1.000
-0.90	1.000	2.800	3.025	2.800	2.296	1.000
-0.75	1.000	1.500	1 563	1.500	1.360	1.000
-0.50	1.000	1.111	1.125	1.111	1.080	1.000
-0.25	1.000	1.019	1.021	1.019	1.013	1.000
0	1.000	1.000	1.000	1.000	1.000	1.000
0.25	1.000	1.011	1.013	1.011	1.008	1.000
0.50	1.000	1.037	1.042	1.037	1.027	1.000
0.75	1.000	1.071	1.080	1.071	1.051	1.000
1.00	1.000	1.111	1.125	1.111	1.080	1.000

I tabellen ser vi at det er mye å tjene på å bruke ξ_K^* i stedet for $\bar{Y}_K$ når ρ_K ligger nær -1. Av (30) ser vi at $D\to\infty$ når $\rho_K\to -1$. Videre ser vi fra (30) og tabell 2 at D avtar mot 1 når ρ_K går fra -1 til 0, og at D øker igjen når ρ_K går fra 0 til 1. For $\rho_K \geq -0.5$ er den maksimale verdi av D 1.125. En kan konkludere med at det er lite å tjene på å bruke ξ_K^* i stedet for $\bar{Y}_K$ når $-0.5 \leq \rho_K \leq 1$.

Figur 1 viser hvordan D varierer med ρ_K for $t=\frac{1}{2}$. For andre verdier av t ville kurven ha vært flatere, men ellers vist det samme variasjonsmønsteret.

I en del tilfeller har vi beregnet den empiriske korrelasjonskoeffisienten mellom 1. og 2. dags observasjoner. Tabell 3 viser den estimerte verdi av korrelasjonskoeffisienten og hvor mye en maksimalt kan tjene på å bruke ξ_K^* i stedet for $\bar{Y}_K$.

Figur 1. D's variasjon med ρ_K for t=½.



Tabell 3. Korrelasjonskoeffisienten mellom tid brukt til forskjellige gjøremål på 1. og 2. dag av dagbokperioden

Aktivitet	Persondaggruppe	Empirisk korrelasjonskoeffisient	Estimert verdi av D for t=½
Inntektsgivende arbeid	Menn, hverdager*)	0.6352	1.062
"	Gifte kvinner med hjemmeværende barn, hverdager	0.6384	1.062
"	Yrkesaktive menn, hverdager	0.4924	1.041
"	Yrkesaktive kvinner, gifte, hjemmeboende barn, hverdager	0.4833	1.039
Arbeidsreiser	De som har arbeidsreise, hverdager	0.2669	1.014
Husarbeid	Gifte menn med barn, hverdager	0.5248	1.045
"	Gifte kvinner med barn, hverdager	0.4424	1.034
Personlig pleie	Menn, hverdager	0.5160	1.044
" "	Kvinner, hverdager	0.4580	1.036
Idrett	De som har utført idrettsaktivitet 1. dag, hverdager	0.3658	1.025
Utdanning	De som har utført aktiviteten 1. dag, hverdager	0.5741	1.052

*) Med hverdager menes at både 1. og 2. dag var hverdager.

For variablene i tabellen ser vi at det er lite å tjene på å bruke ξ_G^* i stedet for $\bar{Y}_G$ som estimator for ξ_G. Var $\bar{Y}_G$ er maksimum 6.2 prosent større enn var ξ_G^*.

Ved å kombinere informasjonen fra tabell 2 og tabell 3, kan vi konkludere med at $\bar{Y}_G$ har gunstige egenskaper som estimator i tidsnyttingsundersøkelsene. Informasjonstapet i forhold til den "optimale" estimatoren ξ_G^* er ubetydelig.

9. OPPSUMMERING

I dette kapitlet skal vi oppsummere hovedresultatene i notatet.

Som mål for usikkerhet har vi valgt "modellvarians". Modellen vi har brukt er beskrevet i kapittel 4. Resultatene som er presentert i notatet, har vi utledet under forutsetning av at modellen gjelder.

I kapittel 6 presenteres estimatorer for variansen til gjennomsnittet $\bar{Y}_G$ av observasjoner fra en vilkårlig persondaggruppe G. I avsnitt 6.2 presenteres tre estimatorer for tre ulike situasjoner a), b) og c). Disse estimatorene har den fordelen at de er enkle å beregne. Ulempen ved dem er at de i mange tilfeller vil overestimere variansen en del. De tre estimatorene er følgende:

a) $$s_1^2 = \frac{1}{d_G(d_G-1)} \sum_{i\varepsilon G} (Y_i - \bar{Y}_G)^2$$

Y_i betegner persondagobservasjon nr. i, og d_G antall persondager i G. s_1^2 er beregnet til bruk under forutsetning av at ingen personer har mer enn en dagbokdag i gruppe G.

b) $$s_2^2 = 1.7\, s_1^2$$

s_2^2 er beregnet til bruk under forutsetning av noen personer har to dagbokdager i gruppe G, men at ingen har mer enn to dagbokdager i gruppe G.

c) $$s_3^2 = 2.4\, s_1^2$$

s_3^2 brukes når noen personer har tre dagbokdager i gruppe G. Dvs. at s_3^2 tenkes brukt når en skal estimere varianser for gjennomsnitt for alle ukedager i Tidsnyttingsundersøkelsen 1971-72.

Faktorene 1.7 og 2.4 i henholdsvis b) og c) er framkommet ved at vi har forutsatt at alle korrelasjonskoeffisienter er mindre eller lik 0.7 multiplisert med en (regneteknisk bekvemmelig) faktor som er større eller lik 1. Tallet 0.7 har vi kommet fram til ved at vi estimerte en del korrelasjonskoeffisienter som vi ventet skulle være spesielt store. Alle estimatene viste seg å være mindre enn 0.65.

s_2^2 og s_3^2 overestimerer variansene hvis forutsetningen om korrelasjonskoeffisientene gjelder, og hvis d_G er av en rimelig stor størrelsesorden.

I praksis vil en foretrekke å estimere standardavvik i stedet for varians. Estimatorene blir da s_1, s_2 og s_3. s_1 kan beregnes ved hjelp av subprogrammet CONDESCRIPTIVE i SPSS. s_1 er identisk med "standard error" i programmet.

s_2 beregnes ved at en først beregner s_1, og så multipliserer s_1 med $\sqrt{1.7} = 1.304$.

s_3 beregnes tilsvarende ved at en multipliserer s_1 med $\sqrt{2.4} = 1.549$.

I avsnitt 6.3 presenteres estimatorer for $\bar{Y}_G$ som er tilnærmet forventningsrette, men som er vanskelige å beregne.

I kapittel 7 presenteres en regel for å kunne si om det har skjedd endringer i bruk av tid til forskjellige aktiviteter i perioden mellom de to tidsnyttingsundersøkelsene.

I kapittel 8 vurderer vi om en i tidsnyttingsundersøkelsene bør bruke en annen estimator for ξ^G (ξ^G er definert i kap. 4.) enn $\bar{Y}_G$. Konklusjonen er at det er så lite å tjene på å ta i bruk estimatoren ξ_G^* (jfr. kap. 8), at det ikke har noen hensikt å erstatte $\bar{Y}_G$ med ξ_G^*.

10. LITTERATUR

1 Tamsfoss, S.: "Om bruk av stikkprøver ved kontoret for intervjuundersøkelser." Artikler nr. 37. Statistisk Sentralbyrå.

2 Norges Offisielle Statistikk: "Tidsnyttingsundersøkelsen 1971-72." A 692.

3 Thomsen, Ib (1977): "Prinsipper og metoder for Statistisk Sentralbyrås utvalgsundersøkelser." Samfunnsøkonomiske studier nr. 33, Statistisk Sentralbyrå 1977.

4 Haldorsen, Tor: "Testing i tabeller." Arbeidsnotater IO 77/41. Statistisk Sentralbyrå 1977.