

Interne notater

STATISTISK SENTRALBYRÅ

83/32

5. desember 1983

VALG AV ESTIMATORER FOR PARAMETRE I VAREHANDELSSTATISTIKK OG REGNSKAPSSTATISTIKK

Av

Knut Fredrik Strøm

INNHOOLD

	Side
1. Innledning	1
2. Rateestimatoren	1
3. Variansen	1
4. Estimering av nøkkeltall (forholdstall)	2
5. Resultater fra bruk av rateestimatoren	3
6. Effisiens	4
7. Konklusjon	5
8. Dokumentasjon	5

1. INNLEDNING

Regelmessig gjennomføres det utvalgsundersøkelser blant bedrifter/foretak i varehandelen i tillegg til den årlige varehandelsstatistikken som bygger på totaltelling gjennom navnekortundersøkelsene til det sentrale bedrift- og foretaksregisteret. Dette gjelder blant annet undersøkelser av varefordelingen i engros- og detaljhandel, bruttofortjeneste, varebeholdning ets. Fra og med statistikkåret 1981 gjennomføres det også en årlig regnskapsundersøkelse i detaljhandel basert på utvalg. I engroshandel bygger regnskapsstatistikken på totalundersøkelse av store foretak.

Ved fastsettelse av utvalgsstørrelse, publisering av resultatene og i analysesammenheng er det viktig å kjenne variansen til de sentrale variablene som inngår i undersøkelsene. For å få så liten usikkerhet som mulig, er det viktig å velge rett estimator.

Fra den årlige varehandelsstatistikken (bedrifts- og foretaksregisteret) kjenner vi sysselsetting og omsetning for hver bedrift. En rekke av de variablene vi skal estimere verdien til er tilnærmet proporsjonale med omsetningen (innenfor hvert stratum). Strata er bestemt av næring og størrelse. Dette proporsjonalitetsforholdet representerer a priori informasjon som vi kan utnytte ved å velge en såkalt rateestimator. En rateestimator har mindre varians enn en vanlig gjennomsnittsestimator¹⁾

($\hat{a} = N\bar{X}$) dersom $\sum_{i=1}^N (a_i - \frac{a}{b} b_i)^2 < \sum_{i=1}^N (a_i - \bar{a})^2$. $\frac{a}{b}$ er proporsjonalitetsfaktoren (raten). I vårt tilfelle vil b være omsetning, a kan f.eks. være totalkapitalen.

2. RATEESTIMATOREN

Hvis a og b er populasjonstørrelser ($a = \sum_{i=1}^N a_i$ og $b = \sum_{i=1}^N b_i$), kan vi formulere rateestimatoren

for a slik:

$$\hat{a} = \sum_{j=1}^m \left(\sum_{i=1}^{n_j} \frac{b_j}{\sum_{i=1}^{n_j} Y_{ij}} X_{ij} \right)$$

der n_j = antall observasjoner i stratum j, $j = 1, 2, \dots, m$

og m = antall strata

N_j = antall i populasjonen i stratum j.

X_{ij} og Y_{ij} er observasjoner av a og b i stratum j i utvalget

b_j er omsetning i stratum j

\hat{a} vil være en konsistent, men ikke forventningsrett estimator for a. Frafall kan føre til skjevhet, men dersom proporsjonalitetsforutsetningen er oppfylt i rimelig grad, skaper ikke dette større problemer.

3. VARIANSEN²⁾

Estimert varians for rateestimatoren:

$$\widehat{\text{Var}}(\hat{a}) = \sum_{j=1}^m \left(\frac{N_j^2}{n_j} \left(1 - \frac{n_j}{N_j} \right) \frac{1}{n_j} \sum_{i=1}^{n_j} \left(X_{ij} - \frac{\sum_{i=1}^{n_j} X_{ij}}{\sum_{i=1}^{n_j} Y_{ij}} Y_{ij} \right)^2 \right)$$

Estimatoren forutsetter trekning med samme sansynlighet innenfor hvert stratum.

I stedet for de observerte ratene $\frac{\sum_{i=1}^{n_j} X_{ij}}{\sum_{i=1}^{n_j} Y_{ij}}$ i utvalget kan vi bruke $\frac{1}{n_j} \sum_{i=1}^{n_j} X_{ij} / \frac{1}{N_j} \sum_{r=1}^{N_j} Y_{rj}$ der

nevneren (\bar{b}_j) som er kjent fra varehandelsstatistikken, er en ikke-stokastisk størrelse. Mer generelt

1) E. Sverdrup: Lov og tilfeldighet. Bind I. 2) Rolf Aaberge har hjulpet til med å finne et uttrykk for variansen, og for øvrig gitt kommentarer til notatet. Se ellers Des. Raj: Sampling Theory.

kan vi kalle raten for R_j , slik at vi senere kan velge hvilken form vi vil gi proporsjonalitetsfaktoren.

Det generelle uttrykket for estimert varians for rateestimatoren blir da:

$$\widehat{\text{Var}}(\hat{a}) = \sum_{j=1}^m \left(\frac{N_j}{n_j}\right)^2 \left(1 - \frac{n_j}{N_j}\right) \frac{1}{n_j} \sum_{i=1}^{n_j} (X_{ij} - R_j \cdot Y_{ij})^2$$

4. ESTIMERING AV NØKKELTALL (FORHOLDSTALL)

Hver populasjonsparameter (i hvert stratum) har en estimator med tilhørende varians. Mange av de interessante parametrene er dessuten definert som forholdet mellom to populasjonsparametre. Vi skal vise et eksempel på framgangsmåten ved variansberegning til en estimator for et slik forhold. r = avlastning Q = overskudd i kroner K = kapital i kroner

$$r = \frac{Q}{K}, \text{ og vi ønsker å estimere } r.$$

Først estimerer vi Q med $\hat{Q} = \sum_{j=1}^m \sum_{i=1}^{n_j} \frac{b_j}{\sum_{i=1}^{n_j} y_{ij}} Q_{ij}$, der b_j = Omsetning i stratum j i populasjonen
 y_{ij} = Omsetning til bedrift i i utvalget fra stratum j
 $i = 1, 2, \dots, n_j$
 $j = 1, 2, \dots, m$

og likeledes estimerer vi K med $\hat{K} = \sum_{j=1}^m \sum_{i=1}^{n_j} \frac{b_j}{\sum_{i=1}^{n_j} y_{ij}} K_{ij}$

og de tilhørende variansene (tilnærmet):

$$\widehat{\text{Var}}(\hat{Q}) = \sum_{j=1}^m \left(\frac{N_j}{n_j}\right)^2 \left(1 - \frac{n_j}{N_j}\right) \frac{1}{n_j} \sum_{i=1}^{n_j} (Q_{ij} - R_{Qj} \cdot y_{ij})^2$$

$$\widehat{\text{Var}}(\hat{K}) = \sum_{j=1}^m \left(\frac{N_j}{n_j}\right)^2 \left(1 - \frac{n_j}{N_j}\right) \frac{1}{n_j} \sum_{i=1}^{n_j} (K_{ij} - R_{Kj} \cdot y_{ij})^2$$

Vi velger: $R_{Qj} = \frac{\bar{Q}_j}{\bar{Y}_j}$ og $R_{Kj} = \frac{\bar{K}_j}{\bar{Y}_j}$

der $\bar{Q}_j = \frac{1}{n_j} \sum_{i=1}^{n_j} Q_{ij}$ $\bar{K}_j = \frac{1}{n_j} \sum_{i=1}^{n_j} K_{ij}$ $\bar{Y}_j = \bar{b}_j = \frac{1}{n_j} \sum_{i=1}^{n_j} b_{ij}$

Avkastningen på kapitalen estimerer vi med $\hat{r} = \frac{\hat{Q}}{\hat{K}}$ og variansen til \hat{r}

$$\widehat{\text{Var}}(\hat{r}) = \text{Var}\left(\frac{\hat{Q}}{\hat{K}}\right) = \frac{1}{\hat{K}^2} \text{Var}(\hat{Q}) + \left(\frac{\hat{Q}}{\hat{K}^2}\right)^2 \text{Var}(\hat{K}) - \frac{2}{\hat{Q}\hat{K}} \text{Cov}(\hat{Q}, \hat{K}) \text{ der } \text{Var}(\hat{Q}) \text{ og } \text{Var}(\hat{K}) \text{ er som}$$

vist på foregående side.

A priori er Q of K positivt korrelerte for absolutte verdier av Q og K .

Da blir uttrykket $\frac{2}{\hat{Q}\hat{K}} \text{Cov}(\hat{Q}, \hat{K}) > 0$.

Ved å bruke $\frac{1}{\hat{K}^2} (\widehat{\text{Var}}(\hat{Q}) + \frac{\hat{Q}^2}{\hat{K}^2} \widehat{\text{Var}}(\hat{K}))$ som estimator for $\widehat{\text{Var}}(\hat{r})$ blir variansen til \hat{r} overestimert.

Ved nøkkeltallsberegninger inngår ofte variable som er forholdsvis sterkt korrelerte. Kovariansen vil da bli så stor at det er meningsløst å se bort fra den. En bør da unnlate å gi variansen til nøkkeltallet dersom denne estimatoren er nyttet.

5. RESULTATER FRA BRUK AV RATEESTIMATOREN

I tabell 1 finner vi estimert forbruk av varer og totalkapital i detaljhandel 1981 og estimert relativt standardavvik (variasjonskoeffesienten) basert på rateestimatoren.

Tabell 1. Forbruk av varer og totalkapital. Estimert verdi og variasjonskoeffisient. Detaljhandel 1981

Næringsundergruppe	Tallet på foretak i populasjonen	Tallet på foretak i utvalget	Forbruk av varer		Totalkapital		
			Estimert verdi Mill.kr	Var. koeff Pst.	Estimert verdi Mill.kr	Var. koeff Pst.	
62	Detaljhandel	29 950	1 619	59 583	0,69	27 883	4,61
621	Varehushandel	82	76	3 364	0,48	2 101	2,08
622	Detaljhandel med nærings- og nytelsesmidler	11 528	495	23 872	0,88	10 272	11,31
6222	Detaljhandel med dagligvarer .	6 842	377	20 734	0,58	9 174	10,30
6223	Det.h. med kjøtt og kjøttvarer	762	33	780	5,36	276	15,60
6224	Det.h. med fisk og vilt	352	12	238	1,93	74	16,93
6225	Det.h. med bakervarer	97	1	97	.	4	.
6226	Det.h. med tobakk, sjokolade, frukt og iskrem	2 930	58	1 748	9,32	637	11,25
6229	Det.h. med nærings- og nytelsesmidler ellers	545	14	275	16,61	107	60,21
623	Detaljhandel med beklednings- og tekstilvarer	4 609	274	5 791	0,94	3 831	6,11
6231	Det.h. med skotøy	617	70	834	1,26	544	8,16
6232	Det.h. med dame-, og herre- og barnekonfeksjon	2 468	132	3 475	1,26	2 252	9,30
6239	Det.h. med beklednings- og tekstilvarer ellers	1 524	72	1 474	2,27	1 034	11,18
624	Detaljhandel med møbler og innbo	2 824	171	4 371	1,83	2 982	8,95
6241	Det.h. med møbler og tepper .	820	57	2 012	1,69	1 346	8,16
6242	Det.h. med radio, fjernsyn, lamper og elektriske husholdningsapparater	1 112	81	1 951	2,45	1 314	13,63
6249	Det.h. med innbo ellers	892	33	408	5,00	322	14,83
625	Detaljhandel med jern- og fargevarer, glass, steintøy og sportsartikler	2 240	203	3 022	1,02	1 839	5,49
6251	Det.h. med jernvarer, kjøkkenutstyr, glass og steintøy ...	577	50	999	1,59	523	7,24
6252	Det.h. med fargevarer, tapeter og golvbelegg	511	62	766	1,66	432	8,22
6253	Det.h. med sportsutstyr	1 152	91	1 258	1,51	835	10,26

Tabell 1 (forts.). Forbruk av varer og totalkapital. Estimert verdi og variasjonskoeffisient. Detaljhandel 1981

Næringsundergruppe	Tallet på foretak i populasjonen	Tallet på foretak i utvalget	Forbruk av varer		Totalkapital		
			Estimert verdi Mill.kr	Var. koeff Pst.	Estimert verdi Mill.kr	Var. koeff Pst.	
626	Detaljhandel med ur, optisk artikler, musikkinstrumenter, gull- og sølvvarer	2 000	89	1 507	3,63	1 065	8,04
6261	Det.h. med ur og optiske artikler	715	29	359	2,26	220	14,11
6262	Det.h. med musikkinstrumenter, noter, gramofonplater og musikk- og videokassetter	564	15	380	3,84	252	32,89
6263	Det.h. med gull- og sølvvarer	421	30	451	4,26	377	6,56
6264	Det.h. med film og videokassetter	300	15	316	3,63	217	12,24
627	Detaljhandel med motorkjøretøyer og bensin	3 223	177	14 941	2,41	4 374	12,79
6271	Det.h. med motorkjøretøyer, deler og rekvisita	1 509	110	8 424	3,43	3 190	15,84
6272	Det.h. med bensin og smøreoljer	1 714	67	6 516	2,74	1 184	11,66
629	Detaljhandel ellers	3 444	134	2 516	5,02	1 418	15,46
6291	Det.h. med bøker og papir ...	590	48	1 049	2,17	646	13,58
6293	Det.h. med kosmetikk	460	13	210	2,19	160	15,58
6294	Det.h. med blomster og planter	926	28	315	25,78	123	33,47
6295	Det.h. med brensel	264	14	374	1,93	122	25,02
6299	Det.h. med varer ikke nevnt annet sted	1 204	31	568	3,32	368	18,13

6. EFFISIENS

Vi har også beregnet variansen til vanlig gjennomsnitt estimatoren ($N \cdot \hat{X}$)

$$\text{Var } \hat{X} = N_1^2 \text{var } \hat{X}_1 + N_2^2 \text{var } \hat{X}_2, \text{ der}$$

$$\text{Var } \hat{X} = \sum \frac{\hat{\sigma}^2}{n_j} \cdot \frac{N_j - n_j}{N_j - 1} N_j^2, \quad \hat{\sigma}^2 = \frac{1}{n-1} \sum (X_i - \bar{X})^2$$

I tabell 2 har vi sammenliknet de to estimatorene anvendt på data fra regnskapsundersøkelsen i detaljhandel. Effisiensen er definert som forholdet mellom variansene.

Tabell 2. Rateestimator og vanlig gjennomsnitt. Esimert verdi, varians og effisiens for tre variable. Detaljhandel 1981

Variabel	Estimator	
	Vanlig gjennomsnitt	Rateestimator
A. Forbruk av varer:		
\hat{X} (Mill.kr)	59,159	59,583
Var \hat{X}	3 451 274 000	171 072 289
Variasjonskoeffisienten $\frac{SD(\hat{X})}{\hat{X}}$	3,10%	0,69%
Effisiens	20,0	0,05
B. Totalkapital		
\hat{X}	29 745	27 883
Var \hat{X}	1 242 287 800	1 653 163 284
Var. koeff.	3,70%	4,61%
Effisiens	0,75	1,33
C. Overskudd før skatt ¹⁾		
\hat{X}	2 689	2 415
Var \hat{X}	180 496 000	36 893 822
Var. koeff.	15,80%	7,95%
Effisiens	5,0	0,20

1) Inkl. "eierlønn" i personlige foretak.

Det er ikke korrigert for frafall, cut-off (for små foretak) o.l. ved beregningen av $N\hat{X}$ (vanlig gjennomsnitt). X er derfor underestimert ved $N\hat{X}$.

Rateestimatoren er betydelig mer effektiv m.h.p. forbruket (og dermed bruttofortjenesten) enn vanlig gjennomsnitt. For totalkapitalen er imidlertid vanlig gjennomsnitt en noe mer effektiv estimator. Også for overskuddet er rateestimatoren mer effektiv. Disse resultatene for relativ effisiens for de to estimatorene er ikke overraskende, ettersom omsetningen forventes å være mer korrelert med resultatsidevariablene enn balanseidevariablene.

7. KONKLUSJON

Vi har i dette notatet sannsynliggjort at for undersøkelser nevnt i kapitel 1 har rateestimatoren bedre egenskaper enn vanlig gjennomsnitt. Ikke minst er det enklere (om ikke helt korrekt) å korrigere for frafall, cut-off o.l. ved at Y kan defineres fritt. Imidlertid må det advares mot å tro at estimatene er riktigere fordi en gjennom oppblåsingsfaktoren $\frac{Y}{X}$ kan tilpasse \hat{Y} (eksogen). Estimatorens gode egenskaper avhenger av i hvilken grad proposjonalitetsforutsetningen er oppfylt.

8. DOKUMENTASJON AV BEREGNINGSPROGRAMMET

Vi har laget et generelt program for beregning av variansen til rateestimatoren. I tabell 3. ser vi eksempel på utskrift fra variansberegningsprogrammet. 1)

1) Programmeringsarbeidet er utført av Jørn Fjæstad.

Tabell 3. Variansen til rateestimatoren. Utskrift fra programmet.

Tabell over estimert varians og standardavvik for 4- 3- og 2 sifret næring

Næring (Eierf.)	Antall	Variansen	Standardavviket	Sumx
62100	74	1907122954.31	43670.62	1998326
621	74	1907122954.31	43670.62	1998326
62220	356	892178838449.37	944552.19	2381825
62230	29	1853196412.70	43048.77	33198
62240	12	156966748.02	12528.64	7865
62260	55	5136422543.03	71668.84	357130
62290	14	4149685187.03	64418.05	11121
622	466	1348692163638.77	1161332.06	2791139
62310	66	1968306213.49	44365.60	177755
62320	127	43898391552.92	209519.43	574549
62390	69	13367564877.91	115618.19	341176
623	262	54765892321.04	234021.14	1093480
62410	54	12057753599.99	109807.80	379262
62420	74	32085893687.52	179125.36	403659
62490	28	2280310890.79	47752.60	89372
624	156	71220989887.70	266872.61	872293
62510	45	1432518769.66	37848.63	128589
62520	58	1259414875.02	35488.24	119142
62530	85	7335933859.16	85650.07	239610
625	188	10181815841.30	100904.98	487341
62610	27	963514998.94	31040.54	21188
62620	15	6870126485.60	82886.23	28048
62630	29	612423409.58	24747.19	67981
62640	15	705597903.86	26563.09	66333
626	86	7326927120.28	85597.47	183550
62710	109	255328182476.94	505300.09	1332300
62720	58	19068290606.95	138087.98	106333
627	167	312864852181.33	559343.23	1438633
62910	45	7698388712.01	87740.46	164886
62930	12	621560198.54	24931.11	6195
62940	27	1695160155.73	41172.32	14596
62950	14	931698188.21	30523.73	24334
62990	30	4450147739.59	66709.43	61365
629	128	48049363760.62	219201.65	271376
62	1527	1653163283540.22	1285753.98	9136138

Sum X gjelder bare observasjonene ($\sum X_i$) og er bare med til kontrollformål.

Varians og standardavvik gjelder modellen $\hat{X} = \frac{1}{R} X_i$, der $R = \frac{\sum Y_i}{\sum X_i}$ er beregnet utifra observasjonene.

Likeledes aggregerer programmet opp n (ant. observasjoner) fra datafilen mens N_j (ant. i populasjonen) må legges inn for hvert stratum på særskilt parameterkort før hver kjøring. Programmet aggregerer N_j opp til de ulike nivåene til forspaltevariablen (eks. 2,3, og 4- siff. næringsgruppe). Programmet beregner R_j spesielt for hvert stratum,

hvilket innebærer at det er en egen programrutine for hvert nivå av forspaltevariabelen. I utskriften blir resultatene redigert til én tabell som vist i tabell 3.

KRAV TIL INPUTFILE OG PARAMETERKORT

Inputfilen må være redigert slik:

- 1 - 7 Ident. nr.
- 8 - 12 Næring
- 13 Annet kjennemerke (f.eks eierforhold)
- 14 - 15 Stratumkode
- 16 - 23 X (variabelverdi)
- 24 - 31 Y "
- 32 - 36 Blanke

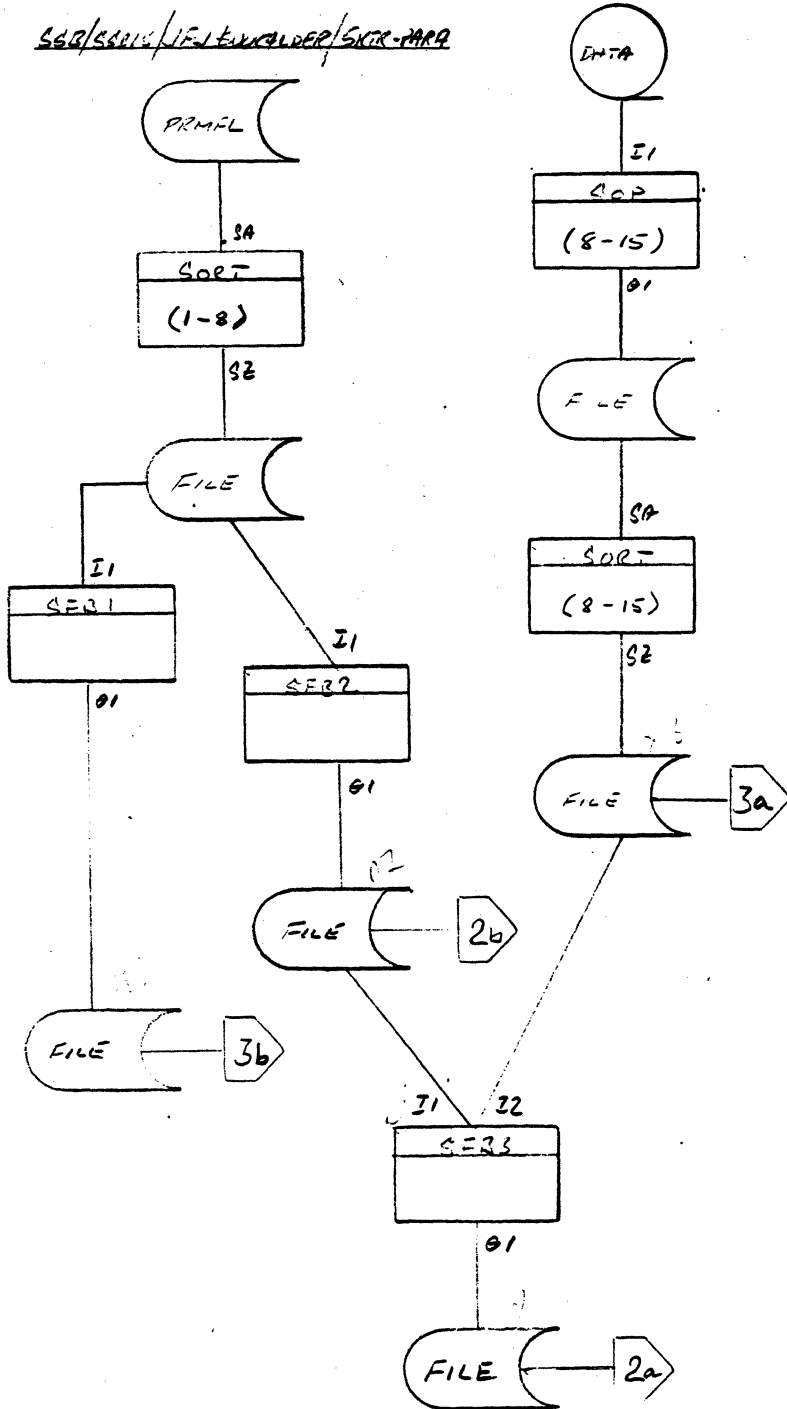
Parameterkortet spesifiserer antallet i populasjonen pr. stratum:

- 1 - 5 Næring
- 6 Eierforholdskode
- 7 - 8 Stratumkode
- 9 - 14 N (antallet i populasjonen)
- 15 - 18 Blanke

Systemkart er tatt med i vedlegg.

STANDARD DOKUMENTASJONSBLANKETT

Stat.nr. 3523	Prosjektnr. 358	Diagramnr. 41	Filnavn ST+2+3 (NF)STR-1V2	Modedefile	Side 1/6
Omfatter: <i>Beregning av varians og standardavvik for 4-3- og 2-sifret nærings</i>				Dato 12/7-83	Sign. JEL



Kommentarer

data (ikke medlysningsvis på tape)

Stratium som det brukes i beregningene for.

Sorteres på stratium

Sorteres på stratium

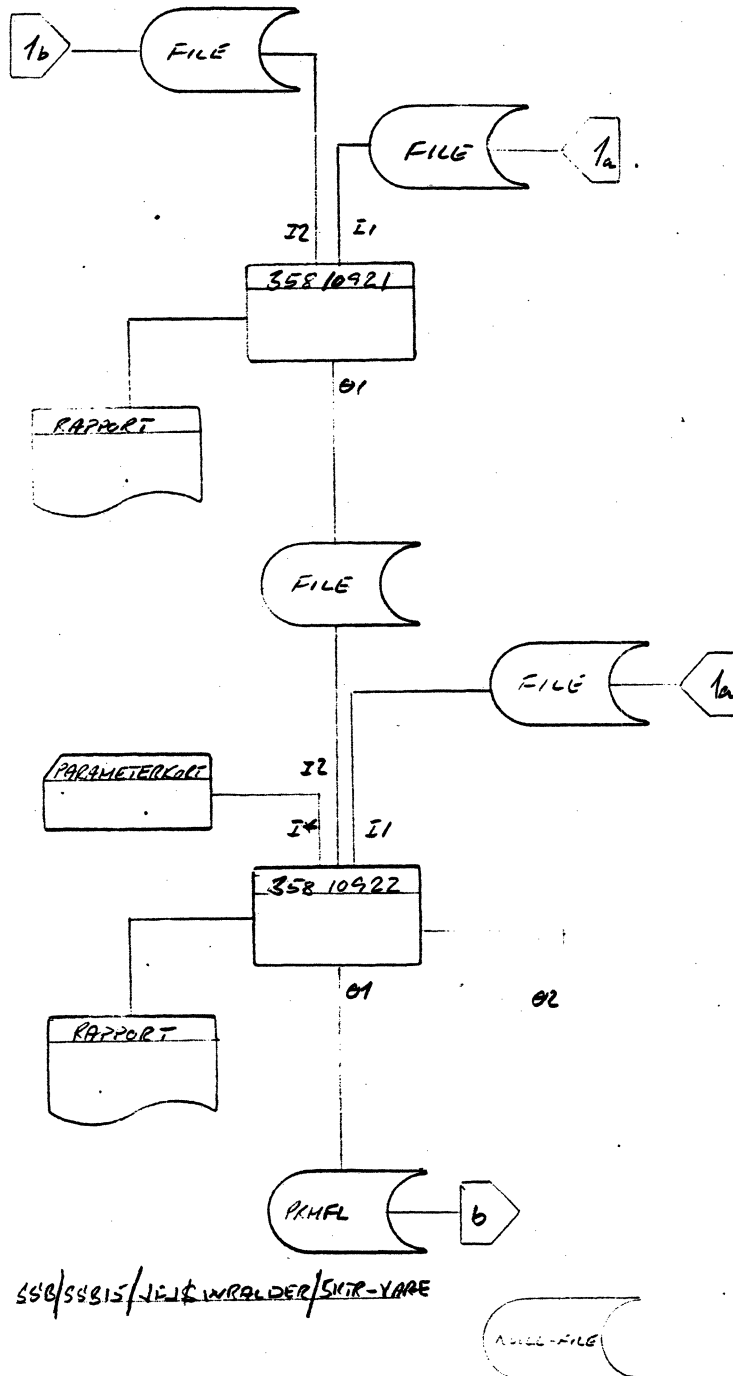
SEB1: Seleksjon av parametere uten spes. viderforh.

SEB2: Seleksjon av plottet navn spes. viderforh

SEB3: Seleksjon av dataene med spes. viderforh som det brukes i beregningene

STANDARD DOKUMENTASJONSBLANKETT

Stat.nr. 3523	Prosjektnr. 358	Diagramnr. 42	Filnavn	Modelfile	Side 2/6
Omfatter: <i>forts.</i>				Dato 12/7-83	Sign. JFI



Kommentarer

*parameterkort med spes
sierforb*

data med spes. sierforb

*beregner R_g for alle struk
+ antallet av hvert struk*

parameterkort med R og n

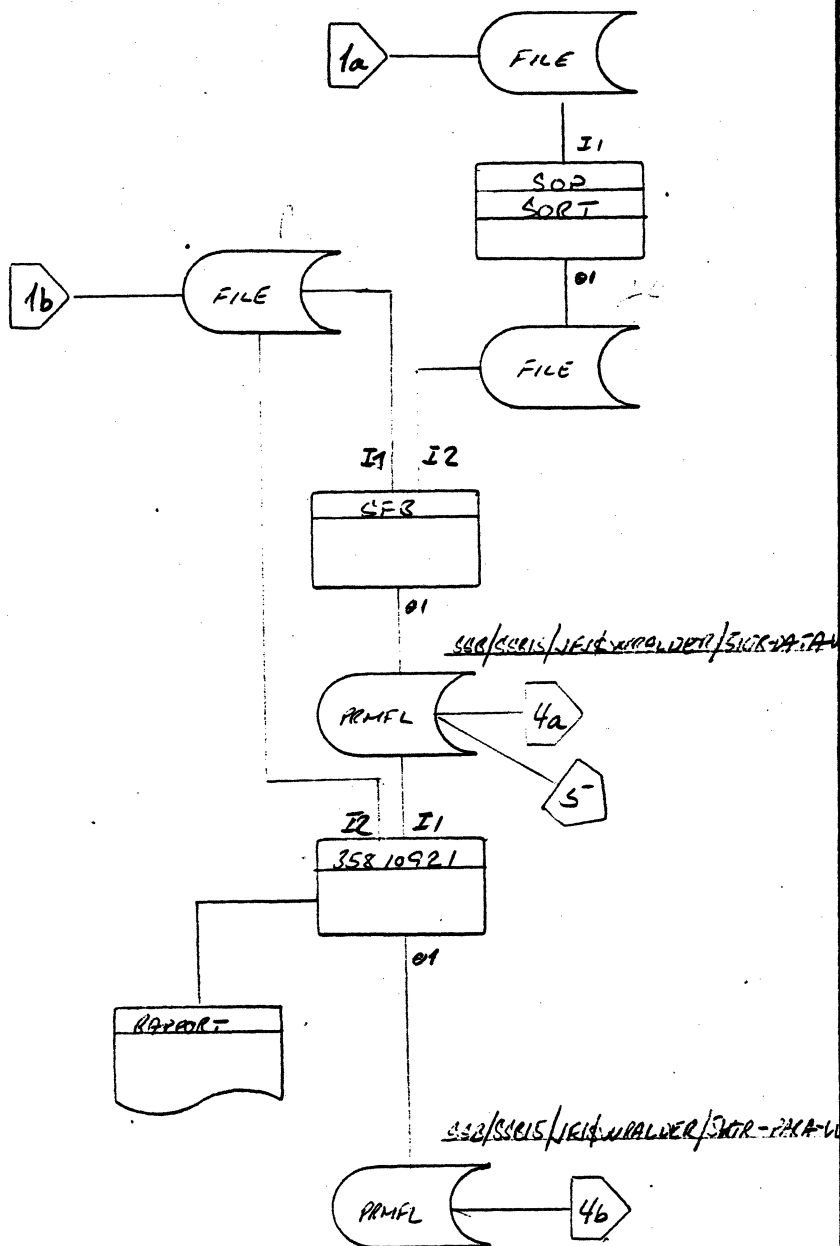
data

post = B

*beregnet var. og stansik
for struk med spes.
sierforb*

STANDARD DOKUMENTASJONSBLANKETT

Stat.nr. 3623	Prosjektnr. 358	Diagramnr. #3	Filnavn	Modelfile	Side 3/6
Omfatter: <i>frts.</i>				Dato 12/7-83	Sign. JFJ



Kommentarer

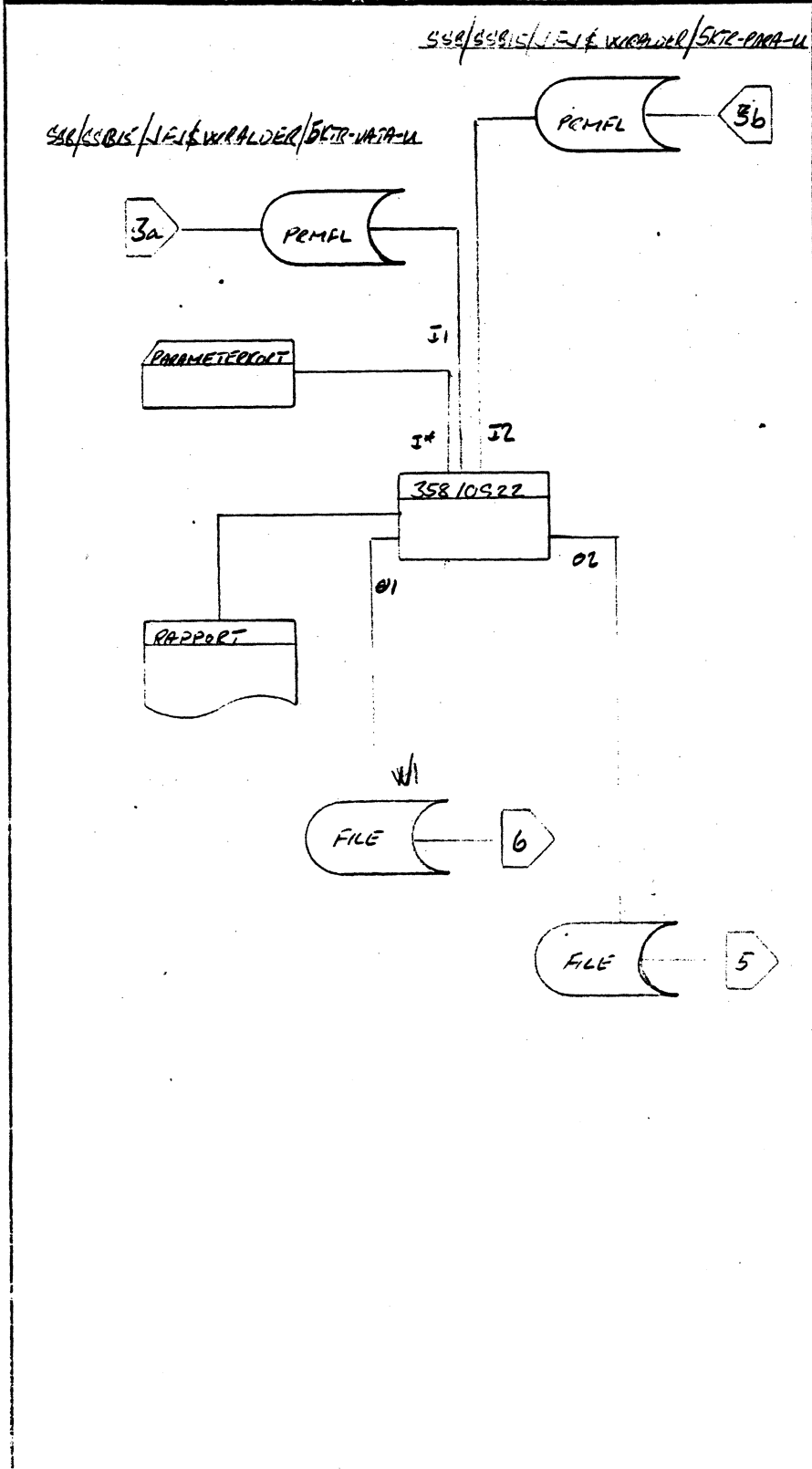
*pos 13 = "a" (eierforh. blankett)
SOP/SORT 2-15 (Sortering)
paramkont utens spes eierforh*

selektorer ut data

beveger seg til $\frac{EX}{2}$ for av hvert stratum

STANDARD DOKUMENTASJONSBLANKETT

Stat.nr. 3523	Prosjektnr. 358	Diagramnr. H4	Filnavn S47516 (JFI)SCTR-2V2	Modelfile	Side 4/6
Omfatter: <i>for.</i>				Dato 12/2-83	Sign. JFI



Kommentarer

pos 1 = U

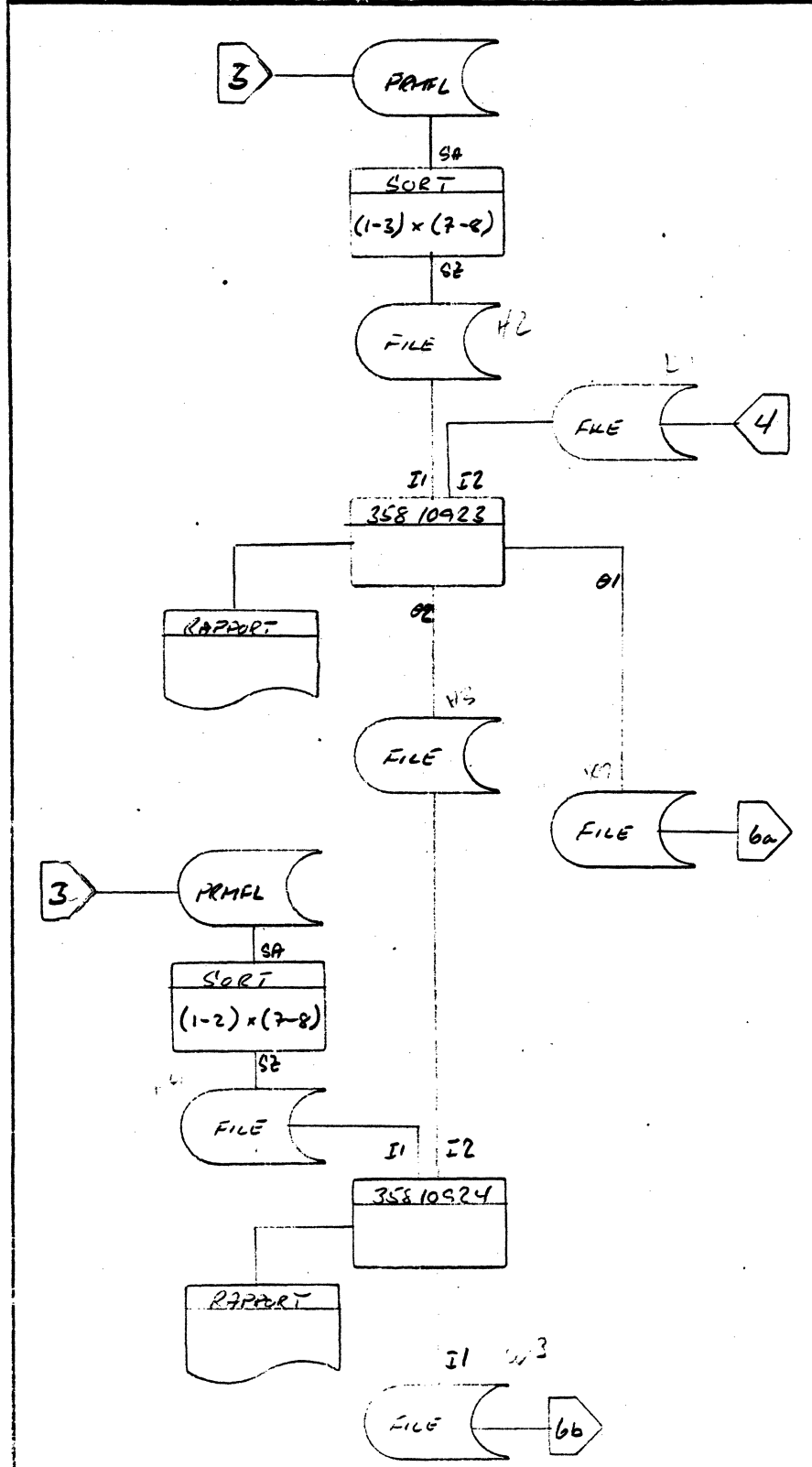
beregner var/stavrik for 4-sifret naming samt parameterkorte for 3-sifret naming

variante av utgiver for 4-sifret

parameterkort for 3-sifret naming

STANDARD DOKUMENTASJONSBLANKETT

Stat.nr. 3523	Prosjektnr. 358	Diagramnr. 45	Filnavn	Modedefile	Side 5/6
Omfatter: <i>forb</i>				Dato 1978-83	Sign. JFI



Kommentarer

selekterte data

3-sifret x strømskode

bestemmer var og størrelse for 3-sifret samt parameterkødet for 2-sifret

skrevet for 2-sifret

var/størrelse 3-sifret

2-sifret x strømskode

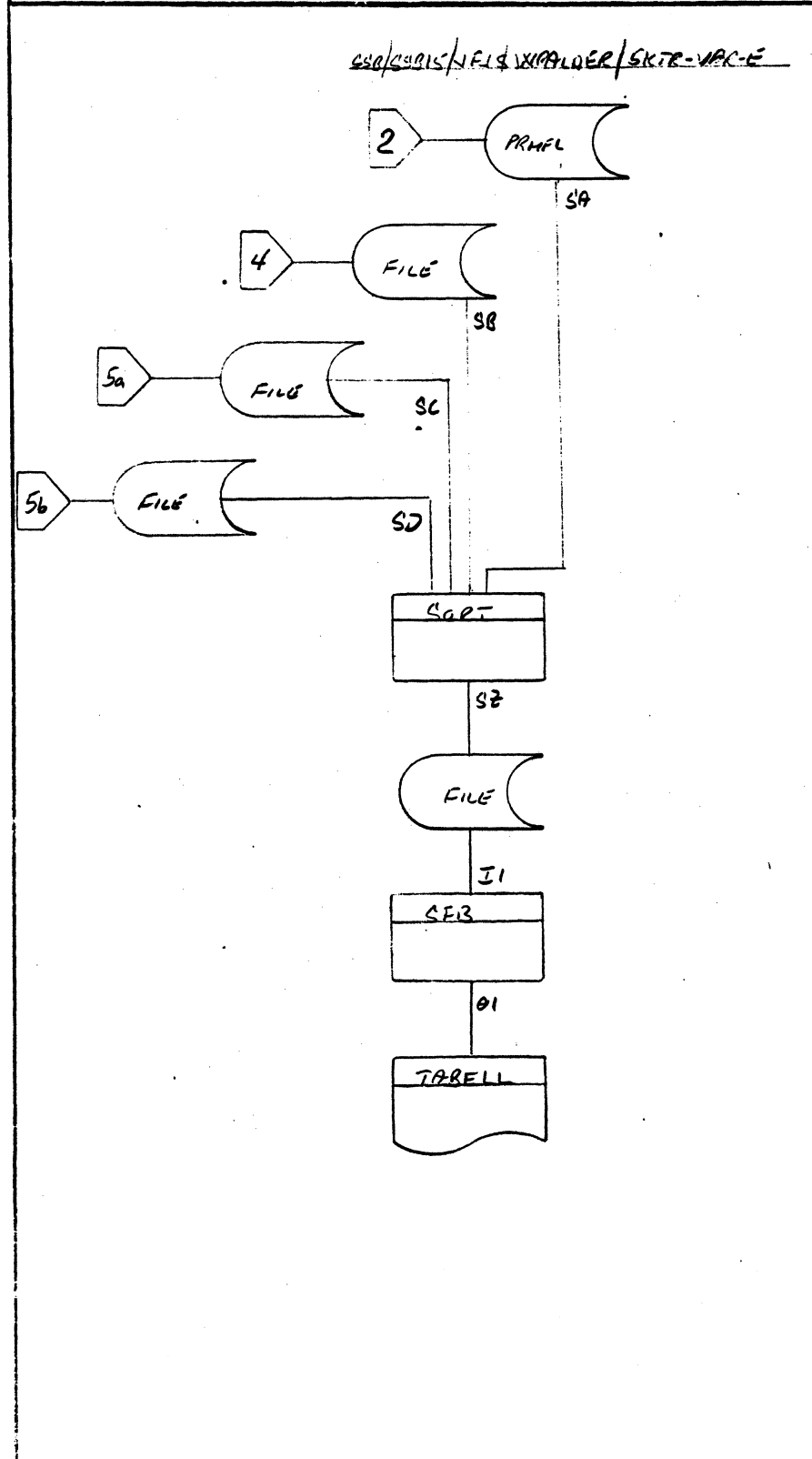
bestemmer var og størrelse for 2-sifret

var/størrelse 2-sifret

STANDARD DOKUMENTASJONSBLANKETT

Stat.nr. 3523	Prosjektnr. 358	Diagramnr. #6	Filnavn	Modefile	Side 6/6
------------------	--------------------	------------------	---------	----------	-------------

Omfatter: <i>frts.</i>	Dato 12/7-83	Sign. JFJ
------------------------	-----------------	--------------



Kommentarer

46. frib med erort

4-sifret

3-sifret

2-sifret