

PC-programvaror för estimation och statistisk analys

Dan Hedlin



R&D Report
Statistics Sweden
Research - Methods - Development
1995:2

INLEDNING

TILL

R & D report : research, methods, development / Statistics Sweden. – Stockholm : Statistiska centralbyrån, 1988-2004. – Nr. 1988:1-2004:2.

Häri ingår Abstracts : sammanfattningar av metodrapporter från SCB med egen numrering.

Föregångare:

Metodinformation : preliminär rapport från Statistiska centralbyrån. – Stockholm : Statistiska centralbyrån. – 1984-1986. – Nr 1984:1-1986:8.

U/ADB / Statistics Sweden. – Stockholm : Statistiska centralbyrån, 1986-1987. – Nr E24-E26

R & D report : research, methods, development, U/STM / Statistics Sweden. – Stockholm : Statistiska centralbyrån, 1987. – Nr 29-41.

Efterföljare:

Research and development : methodology reports from Statistics Sweden. – Stockholm : Statistiska centralbyrån. – 2006-. – Nr 2006:1-.

R & D Report 1995:2. PC-programvaror för estimation och statistisk analys / Dan Hedlin.
Digitaliserad av Statistiska centralbyrån (SCB) 2016.

urn:nbn:se:scb-1995-X1010P9502

PC-programvaror för estimation och statistisk analys

Dan Hedlin



R&D Report
Statistics Sweden
Research - Methods - Development
1995:2

Från trycket
Producent
Ansvarig utgivare

Mars 1995
Statistiska centralbyrån, utvecklingsavdelningen
Lars Lyberg

Förfrågningar

Dan Hedlin
tel 08-783 43 34
telefax 08-661 45 99

© 1995, Statistiska centralbyrån, 115 81 STOCKHOLM
ISSN 0283-8680

Printed
Producer
Publisher

March, 1995
Statistics Sweden
Lars Lyberg

Inquiries

Dan Hedlin
telephone +46 08 783 43 34
telefax +46 8 661 45 99

© 1995, Statistics Sweden, S-115 81 STOCKHOLM, Sweden
ISSN 0283-8680

Innehåll

SAMMANFATTNING	2
1. VARFÖR DENNA RAPPORT	3
2. ANDRA ÖVERSIKTER	5
3. ANDRA PROGRAMVAROR	6
4. CLAN	8
5. EGRET	10
6. EXTRACT	11
7. ETT PROGRAM FÖR FRAKTILSKATTNINGAR	13
8. MINITAB	13
9. ROCANOVA	16
10. SAS/INSIGHT	17
11. SAS/STAT	20
12. SPSS	23
13. SUDAAN	26
14. SAS: DETALJER	30
BILAGA 1. EXEMPEL.	37
BILAGA 2. FÖRKLARING AV VISSA TEKNISKA TERMER.	43
REFERENSER	47

PC-programvaror

för

estimation och statistisk analys

Sammanfattning

Den här rapporten ger korta presentationer av vissa PC-programvaror som kan användas till estimation och statistisk analys. Förutom dem som själva använder den här typen av programvaror torde den här rapporten kunna intressera dem som behöver överblick. Syftet är inte att klassificera programvaror i "onda" och "goda". Det vore barnsligt. Dessutom går det inte. Syftet är inte heller att jämföra programvarornas statistiska redskap med varandra. Det vore alltför tidsödande.

Rapporten tar upp främst programvarorna i Tabell 1, sid 4: deras tillämpningsområden och användning inom SCB i stora drag, användargränssnitt och möjligheter till in- och utflöde. Dessutom ger rapporten referat av andra översikter av eller jämförelser mellan den här typen av programvaror. Ett annat avsnitt ger korta beskrivningar av vissa andra programvaror än de som finns med i Tabell 1. I bilagor finns en del teknisk information som är värd att spridas:

- Kortfattade förklaringar av vissa tekniska begrepp (bilaga 2). Bilagan vänder sig till icke-tekniker som vill ha en chans att förstå vad som sägs i diskussionen av nya programvaror för PC.
- Exempel som främst gäller SAS (bilaga 1), exempel som många skulle ha nytta av men troligen få känner till.

Trots att det idag finns oerhört många programvaror för statistiska tillämpningar, finns det inte så många som är helt anpassade till urvalsundersökningar. Statistiska centralbyråer i flera länder och en del universitet utvecklar därför egna program för urvalsundersökningar.

En ny situation kan skönjas inom programvaruområdet: Programvarukonstruktörerna satsar på öppenhet (man ska kunna använda även konkurrenters programvaror i samverkan) och lättillgänglighet (bort från traditionell programmering, mot standardiserade användargränssnitt och menystöd). Det börjar därför bli alltmer intressant att använda mer än en programvara, allteftersom tillämpningarna kräver specialprogram.

Några anspråkslösa prov visar att öppenheten ännu inte är fullständig. Det går visserligen i de flesta fall bra att föra över data mellan programvaror, men det är svårare att föra över metadata (t ex att för en viss tillämpning ska alla observationer i datamaterialet med SNI-koder 10 till 37 uppfattas som en grupp som benämnes "Utvinning av material och tillverkning" osv).

Troligen gör vi en ny, liknande översikt när läget hunnit förändra sig tillräckligt mycket. Därför önskar vi fortlöpande information om nyheter och annat som bör ingå i den uppdaterade versionen. Delge sådan info Dan Hedlin, U/STM, SCB, 115 81 Stockholm.

1. Varför denna rapport

1. Vi behöver löpande bevakning av programvaror. Särskilt behöver vi uppmärksamma nyheter.
2. En ny situation kan skönjas inom programvaruområdet:
 - Programvarukonstruktörerna satsar på hantering av metadata.
 - De satsar på öppenhet, man ska kunna använda även konkurrenters programvaror i samverkan. Kunder börjar kräva att såväl data som metadata utan problem ska kunna föras över mellan programvaror.
 - Programvarorna blir allt mer lättanvända. Konstruktörerna rör sig bort från traditionell programmering, mot standardiserade användargränssnitt, mera menystöd och standardiserade programspråk (jfr SQL, se bilaga 2 nedan). Programvaror kan alltmer betraktas som utbytbara verktygslådor, snarare än något som användaren ingår i symbiotisk förening med.

Det börjar därför bli alltmer intressant att använda mer än en programvara, alltefter-
som tillämpningarna kräver specialprogram. När den icke-tekniske användaren av
Word startar Equation Editor för att skriva en formel, uppfattar han förmodligen inte
att han använder två olika programvaror. Så smidigt fungerar användargränssnitten
och utbytet av data. På grund av öppenheten kan en arbetsgrupp använda olika pro-
gramvaror efter personlig smak med - i vissa fall - goda möjligheter till utbyte av
data och metadata. Menystöd och andra hjälpfunktioner gör det dessutom lättare för
sällan-användaren att utnyttja en speciell programvaras fördelar.

Vad rapporten innehåller

Rapporten omfattar programvaror för estimation och statistisk analys¹ som har an-
vänts inom SCB. Floran är tämligen omfattande inom SCB. Det vore intressant att
studera även andra programvaror, men ämnet är redan tillräckligt omfattande. Även
tidsserieanalys undviker vi av samma skäl. För programvaror som skapar tabeller, se
Björkman (1994). ADB-enheten inom SCB driver ett projekt om tabellframställning.
Vidare begränsar vi oss till:

- PC-programvaror. Nyutveckling sker sedan ett antal år främst på PC.
- Av skäl som framgår under "Varför denna rapport" håller vi oss till program-
varor som uppfyller endera:
 1. "Färdigförpackade" programvaror, sådana som inte kräver programmerings-
kunskaper eller omfattande kunskaper i en programvaras speciella syntax.
eller
 2. Programvaror som kan hantera metadata.

Syftet är inte att klassificera programvaror i "vilken som är bäst" och "de andra". Det
går inte. Tex behöver naturligtvis inte programvaror som kräver programmerings-
kunskaper vara "dåliga". De kräver mer av användaren, men han vinner troligen en
hel del i flexibilitet. En kraftfull men "svår" programvara kan vara lämplig inom en
del av organisationen som är van vid den, men olämplig inom en annan del som är
ovan och inte behöver styrkan.

¹ Det kanske måste påpekas att statistisk analys använder vi på många sätt i vitt skilda situationer:
mätfelsstudier, bortfallsstudier, blankettexperiment, jämförelse av utrustningar för scanning,
utredning av utnyttjande av stordatorn, analys av svaren från SCBs personalenkät, etc, etc.

Jag har inte jämfört likartade programvaror, för att se vilka redskap som den ena har som den andra saknar. Allt annat en ytlig jämförelse vore alltför tidsödande.

Rapporten tar upp främst programvarornas tillämpningsområden och användning inom SCB i stora drag, användargränssnitt och möjligheter till in- och utflöde. Bilaga 2 förklarar vissa tekniska termer som har med kommunikation mellan programvaror att göra. I bilaga 1 finns detaljer och exempel som gäller SAS och i någon mån SUDAAN och Excel.

I denna rapport ingår programvarorna i Tabell 1. De tas upp i alfabetisk ordning. Som synes blandas därmed myror med elefanter. Tabellen skiljer på programvaror som har redskap för urvalsundersökningar och dem som inte har det. I de flesta av SCBs urvalsundersökningar väljs objekt i ramen med sinsemellan olika urvalssannolikheter. Det påverkar de beräkningar som görs på basis av observationerna. Väljs samtliga objekt med samma sannolikhet, blir beräkningarna enklare. Med "breda" programvaror avses här sådana som inte bara har redskap för ett smalt tillämpningsområde. De behöver däremot inte ha en mängd kringutrustning (t ex redskap för att läsa in data i udda format) för att kallas breda. SUDAAN är ett gränsfall mellan "bred" och "speciell".

Det har inte varit möjligt för mig att skaffa mig egna erfarenheter av samtliga programvaror. Just utbytet av data är något där många manualer lovar mer än vad programvaran håller. Jag har haft mycket värdefull hjälp av en lång rad personer. Förutom de personer vars namn finns i texten kan nämnas John Josefson, SPSS Scandinavia AB, Gunnar Ahlqvist, SAS Institute AB och Joakim Skalin, Handelshögskolan. Samarbetet med ADB-enheten har dock varit mindre omfattande än vad som vore önskvärt. Det är inte alltid så lätt att pressa ner arbete över avdelningsgränserna i den administrativa kvarnen. Det är självklart att det är jag personligen som bär ansvaret för alla tillkortakommanden.

Programvaror som saknar redskap för urvalsundersökningar		Programvaror gjorda för urvalsundersökningar	
breda programvaror	program/programvaror för speciella tillämpningar	breda programvaror	program/programvaror för speciella tillämpningar
Minitab	Egret	SUDAAN	CLAN
SAS/INSIGHT	RocaNova		Ett program för fraktilskattningar
SAS/STAT			Extract
SPSS			

Tabell 1. Programvarorna/programmen i rapportens huvuddel uppdelade efter typ. Se även avsnittet "Andra programvaror".

Programvaror rekommenderade inom SCB

Hur förhåller sig då det här till SCBs fastställda programvarupolicy? I ADB-rapport 1994:1 står bl a att *"på SCB får endast rekommenderad ... programvara användas för sådana tillämpningar där det av kommunikationsskäl är av särskild vikt att endast en gemensam programvara används..."*. Det diskuteras inte i policyn vad som menas

med "kommunikationsskäl", det ges bara ett exempel på ordbehandlare och utbyte av dokument. En rimlig tolkning är att om ett oförutsett behov av datautbyte skulle uppstå, är det den som använder icke rekommenderad programvara som har att tillse att utbytet fungerar. I policyn står inget om att vissa programvaror skulle vara förbjudna, förutom i just det fall som läggs under "kommunikationsskäl". Av programvarorna nedan är det endast SAS (inklusive INSIGHT och STAT) som är rekommenderad. CLAN nämns varken i ADB-rapporterna 1994:1 eller 1994:09, men det finns ett äldre beslut av dåvarande ADB-direktionen att den ska höra till de rekommenderade programvarorna. Dessutom är det så att flera av programvarorna nedan, inklusive CLAN, är skrivna i SAS och omfattas därför i tillämpliga delar i någon mån av beslutet att SAS ska vara rekommenderad.

Den här rapporten diskuterar inte om programvarorna som tas upp är *lämpliga* att använda och i så fall hur.

2. Andra översikter

"The Study Group on Computers in Survey Analysis" (SGCSA) har i tjugio år gett förteckningar över programvaror som kan vara intressanta för urvalsundersökningar. Den senaste omfattar drygt 100 PC-programvaror. En introduktion finns i Rowe (1992).

Varje nummer av **Computational Statistics & Data Analysis** har en sektion "Statistical Software Newsletter" som innehåller bl a korta genomgångar av programvaror. Koch och Haag (1994) publicerar en lista över "statistical software & related systems". Den omfattar ca 80 programvaror. Sawitzki (1994) rapporterar en serie tester av den rent numeriska tillförlitligheten hos 9 programvaror, däribland Excel, SPSS och SAS. Testerna gäller ganska enkla och vardagliga beräkningar för vilka det finns kända numeriska algoritmer. Genomgången visar brister hos såväl Excel, SPSS som SAS och även hos nästan alla av de övriga sex programvarorna. Excel räknar bl a helt fel på variansen och standardavvikelsen för dels nio stora tal, dels nio små tal, som sinsemellan skiljer föga åt ("these results are most surprising": Sawitzki, 1994, sid. 291). Provet av vissa regressionsmodeller klarar Excel av bra i de flesta fallen. Det fel som Sawitzki rapporterar är ett fall där de två oberoende variablerna är linjärt beroende vilket användaren av programvaran borde bli upplyst om. Excel levererar istället endast en tabell över skattningar av koefficienterna. Ingen varning alls. Även SPSS, version 5 för Windows, gör vissa fel på beräkningar: parvisa korrelationer mellan variabler som är linjärt beroende men som har kraftigt olika storleksordning blir allt från ca 0.60 till exakt 1 (Spearman's 'rank correlation'). Även SPSS missade det linjära beroendet i regressionsmodellen. SAS upptäckte i det fallet att det var något skumt, men lämnade ut en felaktig varningstext. Sawitzki rapporterar inga fall där SAS inte kan räkna rätt. Däremot kan man få felaktiga beräkningar om man använder olämpliga redskap. SAS har tre procedurer för regressionsmodeller varav en är gjord för illa-konditionerade problem och skall användas i fall som de här testerna är exempel på. Används de vanliga procedurerna kan man få felaktiga resultat, ibland med varning, ibland utan. Om procedurerna används för data där alla tre borde fungera lika bra, kan man få olika resultat likafullt.

Sawitzki (1994, sid. 300) är kritisk: "... quality awareness needs to be improved [hos programvarukonstruktörer]" och "We cannot rely on correctness of results of the current data analyses systems. We [people in the field] have to invest additional work, provide plausibility checks and recalculations to verify the results." Testerna är

gjorda med gamla versioner av programvarorna, men det är versioner som var i bruk under 1991-1993 då testerna gjordes.

Svenska Statistikersamfundet gjorde 1991 en sammanställning av programvara som medlemmar då använde. Listan omfattar ett fyrtital programvaror.

PC Magazine (1993) gjorde en trevlig men inte alltför djup genomgång av breda, kommersiella programvaror för "statistisk analys": BMDP, Genstat, Minitab, NCSS, P-Stat, SAS, S-Plus, SPSS, Stata, Statistica, Statgraphics och Systat.

Wolter (1985) har ett appendix om med 14 programvaror för variansskattningar samt synpunkter på hur sådana programvaror ska utvärderas. OSIRIS, SUDAAN, SUPER CARP är de av de 14 som är "state-of-the-art, well-supported, portable programs". Idag finns OSIRIS för stordator (operativsystem MVS), SUDAAN för PC (se nedan) och SUPER CARP i PC-version (se "Andra programvaror" nedan).

"**The American Statistician**" som ges ut av ASA har i varje nummer en sektion med namnet "Statistical Computing". I de senaste numren finns bl a:

- PC-programvaror för presentationsfärdig grafik;
- moduler för grafik som kan kombineras med eller byggas in i andra programvaror;
- en översikt av några gratis- eller "nästan-gratis-program" för diverse statistiska tekniker;
- provkörningar av SAS, BMDP och SPSS för variansanalys med balanserade, blandade modeller (SPSS visar sig ej tillräckligt, de två andra ganska svåra att använda på korrekt sätt);
- en granskning av det generella paketet Genstat.

Frigge m. fl. (1989) visar att välkända programvaror ger olika resultat när man beräknar fraktiler och ritar box-plottar.

3. Andra programvaror

Här räknas upp en del programvaror som inte kommit med i det här svepet men som ändå verkar intressanta.

Av tidsskäl saknas **Paradox** och **Excel** i den här utredningen. Eftersom de två hör till SCBs rekommenderade programvaror, är information om dem jämförelsevis lättillgänglig inom SCB. Båda innehåller en hel del statistiska redskap. Excel kan vara intressant för statistiker av även andra skäl: många tycker om att använda tyngre programvaror för beräkningar och Excel för presentation av resultaten i diagram och tabeller (se Björkman, 1994, för Excels tabellverktyg). Excel ger goda möjligheter att ta in data från andra programvaror.

Lisrel är både en metod och en statistisk modell. Lisrel inkorporerar faktor-, stiganalys och linjära regressionsmodeller som beskrivs i strukturerade system av variabler. Dessa kan vara observerade variabler i intervall- eller ordinalskala, eller s k latent variabler som inte kan observeras annat än indirekt. Ett exempel på modell är (från Jöreskog, Sörbom 1993): x_1 och x_2 är oberoende variabler, y_1, y_2, y_3 beroende och sambanden är, bortsett från feltermen och koefficienter, $y_1 = x_1$, $y_2 = x_2$, y_1 , $y_3 = x_1 y_1 y_2$, där y -variabler alltså finns även i högerled. Att vissa av variablerna kallas oberoende och de andra beroende betyder i det här sammanhanget att variationen för de beroende delas upp i ("förklaras med") variationen för de oberoende variablerna och för feltermerna. Programvaran LISREL 8W för Windows, samt

PRELIS som bl a hanterar partiellt bortfall, utför beräkningar för olika Lisrelmodeller. Nytt för version 8 är framför allt ett förenklat kommandospråk samt interaktiv modellbyggnad via stigdiagram som visas på skärmen. Användare inom SCB är bland andra Ingrid Munck, som årligen ger en kurs tillsammans med Dag Sörbom i SCB-skolans regi. Modellen LISREL hör till familjen kovariansmodeller som i SAS täcks av bl a Proc Calis. Programvaran LISREL har vissa redskap som Proc Calis i SAS/STAT saknar. Leverantör av version 8 är Scientific Software International, medan version 7 levereras av SPSS Inc.

Många som arbetar med analyser inom SCB tycker om matrisspråk såsom Gauss eller SAS/IML. Det finns flera andra sådana programvaror på marknaden. Fördelen är att analysmetoderna ofta uttrycks i matrisform i statistisk litteratur. Användaren kan därför utnyttja samma sorts språk när han läser och skriver metodrapport som när han utför beräkningar. Dessutom höjs nivån, när han väl lagt upp data i matriser, behöver han inte bekymra sig om deras detaljinnehåll.

Statistiska centralbyråer och en del universitet utvecklar egna program för urvalsundersökningar. Här kommer en lista över några sådan program.

TANJA är ett program för variansberäkningar och ingår i en familj av program för datagranskning, sortering, samsortering o d, tabellframställning och -redigering. De är utvecklade av Göran Jostin, SCB, och används flitigt inom fr a V-avdelningen. Programmens in- och utflöde är ASCII-filer. En utvärdering finns i Hedlin (1993), främst m a p variansskattningar vid tvåstegsurval.

PC CARP från Statistical Laboratory, Iowa State University, är ett program för flerstegsurval (framför allt variansskattningar och en del andra typer av analys av data från sådana urval). En introduktion finns i Pettersson, Selén (1990). Carlson m fl (1993) jämför PC CARP med SUDAAN (dock något föråldrade versioner) och finner att PC CARP har en del begränsningar: den kan endast ta emot textfiler och som resultat ge tabeller men inte filer som kan användas i fortsatta bearbetningar. Vidare ersätter PC CARP partiellt bortfall med nollor; saknar möjlighet att i utskriften ersätta siffervärden för kategoriska variabler med text; och kan hantera endast 50 variabler åt gången. Sättet att arbeta med PC CARP skiljer sig åt från SUDAAN. Den förra är helt menystyrd medan SUDAANs användare skriver program. Dessutom finns i PC CARP hjälptexter tillgängliga på skärmen, vilket SUDAAN saknar. Det är därför troligt att PC CARP har en låg ingångströskel för den nye användaren. Den som däremot vill göra ett stort antal liknande analyser i PC CARP tvingas upprepa samtliga kommandon för varje analys.

TREES. Skattning av medelvärden och totaler med tillhörande varianser om designen är flerstegsurval, se Rylett, Bellhouse (1988). TREES kan även ge en skattad populationskovariansmatris.

US Bureau of the Census:

CPLX. Analys av kontingenstabeller med data som kommer från flerstegsurval.

VPLX. Punkt- och variansskattningar för flerstegsurval.

IMPS. Dataregistrering, granskning, aggregering och presentation av resultaten i tabeller. Via FN används i IMPS framför allt i tredje världen. Gratiskopiering av IMPS uppmuntras.

Centraal bureau voor de statistiek, Holland:

- **Blaise**. Gör elektroniska blanketter, inkl. kontroller av indata.
- **Abacus**. Ritar tabeller med trevligt utseende.

- **Bascula.** Väger observationerna för att minska variansen eller ev. bortfallsbias. Vägningen påminner om metoden "raking ratio" och beskrivs i Bethlehem, Keller (1987).
- **CBSview.** Ett menystystem som externa statistikanvändare kan välja ut data med. En introduktion till holländska centralbyråns system av programvaror finns i Bethlehem, Hundepool (1992).

Statistics Canada:

Generalized Estimation System, GES, beräknar skattningar med den generella regressionsestimatoren (se Statistics Canada, 1993). Det går att kalibrera urvalsvikterna (jfr CALMAR nedan). De urvalsdesigner GES f n har är stratifierat enstegsurval av kluster. GES är nytt och befinner sig i ett uppbyggnadsskede. Målen för GES ges i Estevao (1992). GES är konstruerat i SAS, som användaren styr via ett menysystem (som möjligen inte är gjort i SAS).

INSEE, Frankrike:

CALMAR modifierar de vikter som motsvarar urvalssannolikheterna i en urvalsundersökning. Modifieringen görs dels för att få skattningar i tabeller att överensstämma med t ex kända marginalsammor, dels för att höja precisionen. CALMAR är skrivet i SAS makrospråk. Namnet CALMAR är en ordlek, det är bläckfisk på franska men syftar också på "kalibrering med avseende på marginaler". Programmet är inte helt testat och kräver kontrollberäkningar: rent numeriska precisionsproblem kan uppstå. Sixten Lundström har provat CALMAR och själv skrivit motsvarande program i APL. Exempel på hur CALMAR kan användas finns i Lundström (1993) som annars främst innehåller en översikt över de statistiska metoderna kalibrering och raking ratio. I ett prov av kalibrering av estimat i KSP (Kortperiodisk sysselsättningsstatistik för privat sektor) användes CALMAR samt APL-program (Lundström, 1994). CALMAR kan endast användas inom SAS. Användaren anropar programmet med i princip följande:

```
%calmar(data=s.urv, m=1, ..., poids=vikter)
```

där S.URV är filen med de ursprungliga vikterna (VIKTER) och m=1 är precisering av kalibreringsmetod (Lundström, 1993). Att skapa ett menysystem som kan ge användaren visst stöd vid anropet av CALMAR vore okomplicerat. CALMAR är konstruerat av Oliver Sautory vid den franska statistiska centralbyrån, som använder programmet i vissa hushållsundersökningar. Manualen är Sautory (1993). Teori och en del erfarenheter beskrivs i Sautory (1992). För uppgifter om SAS, se "SAS: detaljer".

4. CLAN

Bakgrund och användningsområde

CLAN kan beräkna nivå- och variansskattningar för varje rationell funktion av total-skattningar, för godtyckliga redovisningsgrupper. Även den generella regressionsestimatoren kan skrivas som en rationell funktion av totalskattningar (se Särndal, Swensson, Wretman 1992, avsnitt 6.4), men inte en medianskattning. De urvalsdesigner och bortfallsmodeller som CLAN klarar är de fyra strategier som introducerades i SMED 83, se Tabell 2. Alla fyra bygger på stratifierat urval med OSU utan återläggning inom strata. För alla fyra ska antal ramelement och antal utvalda ele-

ment anges. Ur dessa räknas vikterna ut. Det finns inget sätt att ange kalibrerade vikter (se ovan "CALMAR"). CLAN är skrivet i SAS makrospråk.

	"Rak uppräknings". Svarssannolikheterna antas vara lika inom strata. Dessutom görs vissa antaganden om oberoende och övertäckningens andel av bortfallet, se närmare t ex Rosén (1990).	"Svarshomogenitetsgrupper". Varje stratum kan underindelas i grupper inom vilka bortfallsantagandena för "rak uppräknings" ovan antas gälla.
Urval av intresseobjekt	strategi 1	strategi 2
Urval av kluster av intresseobjekt	strategi 3	strategi 4

Tabell 2. Strategi 1-4.

CLAN skattar varianser m h a taylorutveckling som beräknas med stegvisa woodrufftransformationer. Den funktion, som användaren begär variansberäkning för, förenklas därmed till en total. Att woodrufftransformationerna görs stegvis är en nydaning och gör det möjligt för användaren att få varianser för varje rationell funktion, som han uttrycker med en svit av de fyra räknesätten (för en beskrivning se Andersson, Nordberg, 1992).

Användning inom SCB

CLAN används i produktionen av kortperiodisk lönestatistik i den privata sektorn, och i övrigt kanske av ett 20-tal personer inom SCB. CLAN håller på att införas i fler produkter. Strategi 4 används troligen sällan, möjligen inte alls. I Andersson, Nordberg (1992) finns exempel på hur CLAN kan användas.

Användargränssnitt

CLAN kan endast användas inom SAS. Användaren måste skriva program i SAS' makrospråk. Programmet kräver alltså vissa kunskaper i detta. Det är den obegränsade flexibiliteten i val av redovisningsgrupper och rationella funktioner för vilka varianser ska skattas som tvingar fram det jämförelsevis komplexa användargränssnittet. Inom SCB-projektet "statistiska programvaror" pågår utveckling av ytterligare ett gränssnitt, där användaren ska i princip kunna ge samma kommandon men via menyer.

In- och utflöde

CLANs utflöde är alltid en SAS-fil med ett bestämt namn som användaren inte kan påverka. I SAS använder man gärna värdeförrädsbeskrivningar ("format") för gruppering, (se bilaga 2). På ett indirekt sätt går det att använda format även i CLAN ("SAS: detaljer", Andra programvarors filer).

Leverans, konstruktion, underhåll

CLAN är konstruerat av Lennart Nordberg och Claes Andersson, båda SCB. CLAN kan göras tillgängligt inom SCB över nätverket, så att programmet kan användas utan att den enskilde användaren behöver installera något. För allmänna uppgifter om SAS, se avsnittet "SAS: detaljer".

Pris

Gratis även för användare utanför SCB.

Möjliga plattformar och övriga hårdvarukrav

Se avsnittet "SAS: detaljer".

Övrigt

CLAN innehåller inget redskap för tabellering, istället används vanliga SAS eller någon annan programvara till vilken CLANS resultat kan exporteras. SCB-skolan ger en kurs i användningen av CLAN.

5. EGRET

Bakgrund och användningsområde

EGRET (förkortning av Epidemiological Graphics, Estimation and Testing) är en vidareutvecklad version av gratisprogrammet PECAN. EGRET är gjort för estimation och analys inom epidemiologi och biostatistik, vilket inte hindrar att det kan vara användbart inom andra verksamheter. Programmet omfattar logistisk regression (betingade eller obetingade modeller; systematiska eller slumpmässiga faktorer), regression med Cox-, Poisson- eller exponentialmodeller; betabinomialmodeller med slumpmässiga faktorer; Kaplan-Meier-estimatoren (en icke-parametrisk estimator för analys av överlevnadsdata) och analys av $2 \times K$ -tabeller (homogenitet, trend, oddskvot).

En tilläggsmodul för enklare grafisk presentation finns. EGRET har ett eget filformat där även vissa metadata lagras. Enklare omkodningar och transformationer av data går att utföra.

Användning inom SCB

Peeter Fredlund, tidigare SCB, har använt EGRET inom ett samarbetsprojekt mellan SCB, Statens institut för psykosocial miljömedicin och Johns Hopkins University, USA. Datamaterialet omfattande delar av ULF (undersökningarna om levnadsförhållanden).

Användargränssnitt

EGRET är helt menystyrt, men menyerna ser inte ut som de i Windows. Användarens kommandon sparas i en log-fil.

In- och utflöde

Data läses in från textfiler i ASCII-format (se bilaga 2) eller LOTUS-filer i WK1-format eller SYMPHONY-filer i WRK-format. Resultaten sparas i en fil med ett förutbestämt namn. Denna fil är en ASCII-fil.

Leverans, konstruktion, underhåll

Programmet är utvecklat av Statistics and Epidemiology Research Corporation, Seattle, USA, med anslag från the National Cancer Institute. EGRET började säljas i slutet på 80-talet. En utmärkt manual som även fungerar som lärobok medföljer.

Pris

1991 kostade EGRET \$900 i USA. Tekniskt stöd ingår.

Möjliga plattformar och övriga hårdvarukrav

EGRET går att köra på en PC med 286-processor, DOS 2.1 eller högre och med minst 512K RAM ledigt. Allt extra minne upp till 640K underlättar bearbetningarna. Matematikprocessor är inget krav men snabbar upp.

Övrigt

EGRET är betydligt snabbare än t ex motsvarande SAS-procedurer. Koder för partiellt bortfall definieras av användaren. De måste dock vara tal. Ingen imputeringsstrategi finns, partiellt bortfall i en variabel för en observation gör att hela observationen exkluderas från beräkningarna.

6. EXTRACT

Bakgrund och användningsområde

Extracts stordatorversion konstruerades i mitten på 80-talet, medan PC-versionen är helt ny. Programmet beräknar skattningar för proportioner, medelvärden, totaler, fraktiler och ginikoefficient (en skattning av en parameter som definieras via populationens fördelningsfunktion, ofta använd till att ge mått på ojämlikhet i inkomstfördelningar) och varianser för dessa. Strategierna kan vara 1 eller 3 (se Tabell 2, sid 9). Dessutom har Extract ett redskap för en sorts nätverksurval, då individer väljs ur en ram med individer som element men de objekt som är intressanta i undersökningen är de hushåll individerna hör till. Varianser av fraktilskattningarna planeras. Redovisningsgrupper kan anges på ett sätt som ger stor flexibilitet.

Manualen är Gilbertsson (1994). Den är mera en bruksanvisning än en beskrivning av vad programvaran exakt gör. För att få veta detta bör man fråga Berndt Gilbertsson.

Användning inom SCB

Extract används i produktionen av Hushållens Inkomster och inom Fördelningsmodellprojektet. Även vissa andra användare finns inom och utanför SCB.

Användargränssnitt

Extract är ett SAS-program och har alltså samma användargränssnitt som övriga SAS. En skillnad är sättet att ange "styrkort" som är hämtat från Extracts stordatorversion. Ett exempel hämtat från Hushållens Inkomster visas i Figur 1 (bl a skattad medelinkomst inom fraktilgrupper för olika typer av hushåll).

```
PARM= TYP=2,3,4 ST=H KOL=61-70,5,43-45,3 XLD
IF 1
ISU741H
KEG702H=20-64
KEG702H=65-999
    (<> 0) /SAMTLIGA
    (1) /20-64 år
    (2) /65- år
RUB=SM21a ISU741 Disponibel inkomst i decilgrupper mm
RUB=, för år 1991
SET P.HR91
RENAME BSTRAT=STRATUM BLILLAN=LILLAN
RENAME BSMEDV=SMEDVIKT BSTORAN=STORAN
/
```

Figur 1. Exempel på program i Extract.

In- och utflöde

Extract kan exportera Excel- och PC-AXIS-filer (PC-AXIS är en databashanterare som utvecklats av SCB). En standardtabell, som är en textfil, skrivs alltid ut. Den är inte tabseparerad (se bilaga 2) och dess utseende går inte att styra. I SAS använder man gärna värdeförrådsbeskrivningar ("format") för gruppering, (se bilaga 2). På ett indirekt sätt går det att använda format även i Extract ("SAS: detaljer", Andra programvarors filer).

Leverans, konstruktion, underhåll

Extract är konstruerat av Berndt Gilbertsson, SCB. Inom SCB är programmet är installerat tillsammans med SAS: den som har tillgång till SAS för Windows kommer också åt Extract utan extra installation.

Pris

Ingen extra kostnad inom SCB.

Möjliga plattformar och övriga hårdvarukrav

Se avsnittet "SAS: detaljer".

7. Ett program för fraktilskattningar

Bakgrund och användningsområde

Lennart Nordberg, SCB, har skrivit ett program i SAS för estimation av fraktiler och deras varianser. Urval och metod för bortfallskompensation motsvarar strategi 1 (se Tabell 2, sid 9). Den variansestimater som programmet använder är "Woodruffs metod", som är ganska grov. Se vidare t ex Särndal, Swensson, Wretman (1992).

Användning inom SCB

Programmet är anspråkslöst, och bör ses som ett exempel, inte som en uttestad programvara.

Användargränssnitt

Användaren fyller i värden på vissa parametrar (jfr CALMAR).

8. Minitab

Bakgrund och användningsområde

Minitab omfattar ett stort område av statistiska tillämpningar. Namnet tycks antyda att Minitab är en liten programvara och för inte så många år sedan var det verkligen så. Den hade en lättillgänglig uppsättning redskap som ledde fram till enkla tabeller eller diagram men det fanns en snäv begränsning av datamaterialets storlek oavsett hur stark dator man hade. Idag är det inte så. Minitab har i praktiken inga begränsningar av storleken av det arbetsark som programvaran kan hantera, andra än de hårdvaran medför. Det maximala antal variabler är dock 1000, vilket kanske inte alltid räcker eftersom man i Minitab hellre lagrar data i en gles matris än i ett antal länkade filer. Minitab är bred m a p tillämpningarna men inte särskilt djup inom alla områden. Den saknar dessutom vissa saker som man förväntar sig av en större statistisk programvara: tal lagras i enkel precision, den saknar stöd för metadata och kan inte foga två filer i Minitabformat med varandra så att poster med samma identitetskod matchas (därför lagrar man hellre data i en stor, gles matris än i många mindre filer). Fogningen kan dock göras med Minitabs makrospråk, se nedan. En fullständig uppsättning av SPSS kostar ungefär tre gånger mer och tar sex gånger mer plats på hårddisken, och SAS är ännu större och dyrare (jämförelsen gäller för en enstaka licens och för idag något gamla versioner). Hurvida "mini" i namnet "Minitab" idag är överspelat eller ej beror alltså på synvinkel. Däremot är det inte så att "tab" skulle syfta på något avancerat tabellverktyg: de olika tabeller som kommer ut ur Minitab är arbetsmaterial, inte publiceringsfärdiga tabeller.

Minitab har redskap för tidsserieanalys, explorativ dataanalys, diagram, kvalitetsstyrning (inkl s k styrdiagram med test för "special causes"), försöksplanering, regressions- och variansanalys, och en del multivariat analys och icke-parametriska test. Även en liten dataregistreringsdel finns. Minitab är inte indelat i moduler utan köparen får hela programvaran på en gång.

Minitab har framför allt två manualer: "Minitab Reference Manual", som beskriver kommandona och som är ca 5 cm tjock, och "Minitab User Guide" som ger en introduktion och dessutom förklarar menysystemet. Dessutom finns Technical Reports med tekniska specialämnena eller exempel på tillämpningar.

Användning inom SCB

Minitab används frekvent av kanske ett 20-tal personer inom SCB. Det främsta tillämpningsområdet inom SCB är tidsserieanalys. SCB-skolan ger kurser i tidsserieanalys, där Minitab används. Kurserna hålls av Anders Wallgren och Britt Wallgren. Många har erfarenhet av Minitab från högskolan, men då i många fall en äldre och enklare version.

Användargränssnitt

Minitab finns numera i en Windowsversion. Även DOS-versionen styrs via ett menysystem. Minitab är helt anpassat till Windows (det finns åtminstone inga för mig tydliga avvikelser).

Ett sätt att arbeta som är naturligt i Minitab för Windows är att ha två fönster synliga på skärmen, ett arbetsark med data och ett med användarens kommandon och Minitabs respons på dessa: utskrifter av tabeller m.m. samt felmeddelanden och andra meddelanden. Det senare kallas Session Window. Längst upp har man i vanlig ordning en menyrad. Om ett kommando beställs från denna skrivs motsvarande kommando in automatiskt i "Session Window". Användaren kan även själv skriva kommandon direkt i "Session Window". Vissa kommandon kan avges på endast ett av sätten. Användaren har full frihet att ändra vad som helst i utskrifterna i "Session Window". De är alltså editörbara.

Arbetsarket fungerar mera som ett "skrivbord" än en fil. Användaren kan lagra hur många filer i logisk mening som helst i arbetsarket. Dessutom kan han bestämma att cellen x:y ska innehålla en viss konstant, eller att ett rektangulärt område av cellerna ska innehålla en matris, på vilken Minitabs matriskommandon kan tillämpas. Det gör att användare av Minitab har stor frihet att utveckla en egen stil för datalagring, något som kan vara trevligt men även försvåra överföring av data till andra användare och andra programvaror. Till arbetsarket hör ett kalkylarksliknande datafönster och ett "informationsfönster" som underlättar för användare att hantera och överblicka data i arbetsarket.

In- och utflöde

Minitab kan importera och exportera filer i bl a följande format:

- Excel,
- Lotus 1-2-3 för Windows,
- dBase.

Manualen varnar för att partiellt bortfall kanske inte läses in rätt.

Vid inläsning av en textfil kan man ange vilken typ av tecken som avgränsar kolumnerna från varandra i den aktuella filen. Någon IF-sats finns inte i "vanliga" Minitab (jfr makrospråk nedan) varför det inte finns något färdigt redskap för att läsa in "lok och vagnar" (se Bilaga 2).

Minitab har inte anammat ODBC (se bilaga 2 för en allmän beskrivning av ODBC). Data från SQL-databaser kan importeras till Minitab och inte heller exporteras. Enligt uppgift planerar inte Minitab Inc. någon utveckling på det här området.

Minitab stöder inte OLE. Minitab har däremot DDE (se bilaga 2) och kan fungera som såväl server som klient.

Leverans, konstruktion, underhåll

Konstruktör är amerikanska Minitab Inc. Den leverantör som Minitab Inc. pekar ut för Sverige är CLECOM, The Research Park, Vincent Drive, Birmingham B15 2SQ, England. Fax +44 21 471 5169. Windowsversionen kallas "release 10", medan både den senaste DOS- och MAC-versionen heter "release 8".

Pris

Leverantören CLECOM säljer Minitab för Windows för ca 700 brittiska pund, något mindre per styck om man köper fler. DOS och Mac-versionerna kostar ca 500 var-dera. Manualer och tekniskt stöd är då inkluderade. CLECOM kan lämna offert på företagslicens.

Möjliga plattformar och övriga hårdvarukrav

Minitab finns för PC (DOS och Windows), Macintosh och vissa mini- och datorer. Användare av Minitab för Windows säger att det går bra att köra på en måttligt kraftfull PC, t ex en med 386-processor och arbetsminne på 8 MB.

Övrigt

Minitab innehåller sedan gammalt ett makrospråk (som ganska starkt påminner om SAS makrospråk). Det är ett programmeringsspråk; resultatet av ett makroprogram är en serie av vanliga Minitabkommandon, som körs automatiskt. Ett makro är en procedur eller subrutin med parametrar och de satser som är nödvändiga i ett programmeringsspråk: If Then--Else If--Else, While och Do. Ett makro kan anropa ett annat makro. Det går att ha en vektor med ett i förväg ej bestämt antal element som parametervärde. Med makrospråket kan den programmeringskunnige användaren komplettera "vanliga" Minitab och få en mycket kraftfull programvara.

Minitab har satsat hårt på grafik som analysredskap. Diagrammens utseende har förbättrats på senare tid. Det finns diagram som stöder förståelsen av resultaten från bl a variansanalys. Man får då bilder över residualer och samspel. Det finns även gott om speciella diagram för explorativ dataanalys och kvalitetsstyrning. I många av punkt- och linjediagrammen kan man lägga in en lowesskurva (en approximerande funktion till punktmängden).

De flesta diagram i Minitab är länkade med varandra. Markerar man en punkt i t ex ett punktdiagram lyses motsvarande punkt upp automatiskt i andra diagram som är framme på skärmen. Det går även att med en rektangel ("pensel") markera alla observationer som faller inom rektangeln. Det här fungerar för de flesta diagramtyper i Minitab, dock ej för bl a stapel- och cirkeldiagram och låddiagram (boxplots).

Tabellverktyget är layoutmässigt mycket begränsat. Med tre eller fler klassificeringsvariabler kan endast vissa speciella korstabeller åstadkommas. Ur analysynpunkt går det ändå bra att få fram en lista av tabeller som tillsammans visar den information som behövs, men poängen med ett tabellverktyg är ju att kunna presentera uppgifterna tydligt och komprimerat.

Utskrifterna är inte tabseparerade (se Bilaga 2, "Tabseparerade tabeller"). Det går dock att kopiera utskriften i "Session Window" till datafönstret, där mellanslag fungerar som avgränsare. Det finns nu även ett redskap, som ännu är på experimentstadiet, varmed man får utskriften direkt inlagd i en fil.

Det finns inte möjlighet att förse redskapen med egna kovariansmatriser med skattningar av kovarianser på basis av urval (jfr avsnitt "SAS: detaljer" under "analys av data från urval").

9. RocaNova

Bakgrund och användningsområde

RocaNova är ett program för intensitetsregression. Intensiteten är en parameter för en process. Ett exempel på en skattning av en intensitet är antal arbetsolyckor genom total arbetad tid under en viss tidsperiod. En sådan kan analyseras med avseende på bakgrundsfaktorer som representerar "förklarande tillstånd" (t ex typ av arbetsplats) m h a någon statistisk modell, t ex en loglinjär modell. Se t ex Qvist m fl (1991) för exempel på tillämpningar inom SCB. I dessa arbeten användes dock inte RocaNova, utan en familj av program kring programmet LOGLIN. Sten Martinelle, SCB, konstruerade RocaNova för att ersätta LOGLIN. Senaste version har nummer 1.2. RocaNova är gjort m h a Visual Basic. Ingen bearbetning av data kan göras inom RocaNova. Istället rekommenderas i första hand PC-AXIS, som är en databashanterare som utvecklats av SCB.

Användning inom SCB

Bland andra har Klas Rydenstam använt RocaNova, i tidsanvändningsundersökningen.

Användargränssnitt

RocaNova är helt menystyrt. Användaren kan spara all information om den modell han valt.

In- och utflöde

In- och utflödet kan vara textfiler eller filer i PC-AXIS-format. Det senare rekommenderas.

Leverans, konstruktion, underhåll

Programmet tillhandahålls av U/SDB, SCB.

Pris

Gratis inom SCB. 1400 kr exkl moms för kunder utanför SCB (gäller en enstaka användare).

Möjliga plattformar och övriga hårdvarukrav

RocaNova körs under Windows. Hårdvarukraven är mycket måttliga. Det krävs en 386-processor eller bättre och ett arbetsminne på minst 4 MB.

10. SAS/INSIGHT

Bakgrund och användningsområde

SAS/INSIGHT är en modul av SAS. INSIGHTs område är främst explorativ dataanalys, deskriptiv statistik och generaliserade linjära modeller. Det senare finns även i en ny procedur i SAS/STAT (se detta avsnitt nedan). INSIGHT innehåller knappast något som inte finns någon annanstans inom SAS-systemet. Poängen med INSIGHT är framför allt att det är enklare och snabbare att använda. För de flesta torde det vara avsevärt lättare att hitta fram till en intressant analys med INSIGHT än med SAS äldre delar.

Den här avsnittet bygger på version 6.08. Version 6.10 är klar men vi har inte fått den i skrivande stund.

Liksom en del andra moderna programvaror för statistiska tillämpningar söker INSIGHT förena beräkningar med bilder. Om användaren pekar på en punkt i ett punktdiagram över data kan han se vilken punkt det är genom att observationens nummer eller den etikett som är knuten till observationen visas i diagrammet (se Figur 2, som för innehåller en bugg: antalet besök borde markeras med 1 milj resp 2 milj på y-axeln). Han väljer att ta bort den observationen ur alla beräkningar och på så sätt se hur den påverkar dataanalysen. Han kan också markera observationer i datafilen och se var de dyker upp i punktdiagram och andra bilder, och vice versa. I ett tvådimensionellt punktdiagram kan värden på en tredje variabel representeras med att punkterna får olika färger. Användaren kan rita ett par eller fler bilder, och föra en rektangel ("pensel") fram och tillbaks över en av bilderna för att se hur observationer som faller inom rektangeln är fördelade i de andra bilderna. Andra sätt att få en överblick över förhållandet mellan tre eller fler variabler är roterande tredimensionella prickdiagram och rutnät av tvådimensionella prickdiagram över alla parvisa kombinationer av variablerna. Det går även att dela upp datafilen i två delar (texttårtor och glesbyggd) och sida vid sida visa diagram över vardera delen.

De typer av diagram som INSIGHT kan rita är stapeldiagram, låddiagram (boxplots) och punktdiagram. Låddiagram ritas enligt Tukeys definition. I prickdiagram kan approximerande funktioner läggas in (funktioner som är polynom av den grad användaren sätter, eller splines och kernels som båda räknas som icke-parametriska anpassningsmetoder). I stapeldiagram kan en skattad täthetsfunktion läggas in på olika sätt. Låddiagram kan få medelvärde inritat, så att man kan jämföra det med medianen. Vissa redskap finns för att få fram speciella punkt- och linjediagram:

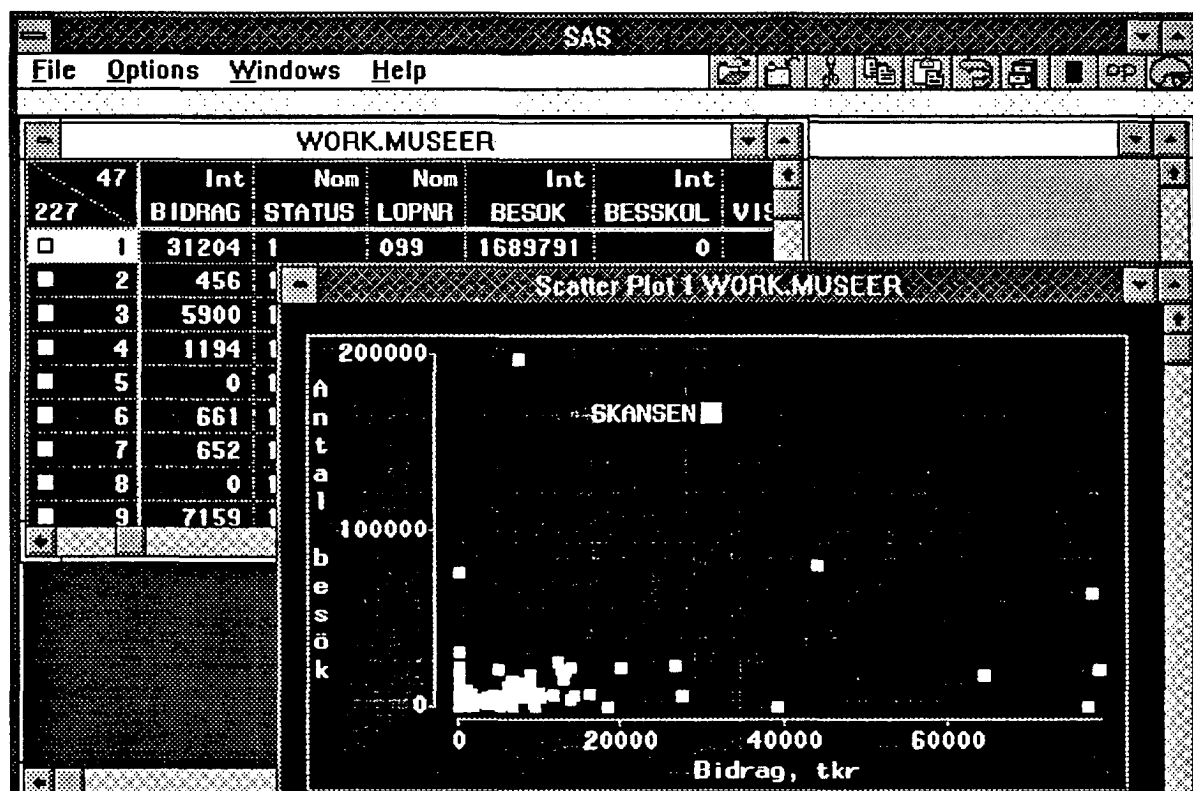
skattad empirisk fördelingsfunktion eller någon teoretisk fördelingsfunktion anpassad till data (normal-, lognormal-, exponential-, eller Weibullfördelning). I en QQ-plot prickas de sorterade värdena för en variabeln av efter y-axeln och efter x-axeln fraktiler för den teoretiska fördelingsfunktion för att man ska få en uppfattning om datafördelningen liknar den teoretiska fördelningen.

Det går inte att inkludera vikter som motsvarar urvalssannolikheter i INSIGHTs skattningar. Behöver man inte sådana vikter, kan bl a följande parametrar skattas, förutom de vanliga såsom medelvärden och standardavvikelser:

- Vissa fixa fraktiler (medianen, övre och nedre kvartilen etc). Till skillnad från basmodulen av SAS kan man inte välja mellan olika definitioner av fraktiler (jfr Proc Univariate).
- Ginis medeldifferens (ginikoefficienten multiplicerad med $2\bar{Y}$, jfr "Extract" ovan). SAS räknar den som medelvärdet av absolutbeloppet av differensen mellan samtliga par av variabelvärden.
- Två typer av robusta skattningar av medelvärden, trimmade eller winsorizerade, som båda bygger på att de extrema värdena tas bort ur beräkningen.

Ett annat sätt att undersöka strukturen i data är att försöka anpassa en känd modell till data. Insight stöder generaliserade linjära modeller. Flexibiliteten är inte lika stor som för Proc GenMod (jfr SAS/STAT nedan), INSIGHT tillåter endast vissa fixa val för modellens beståndsdelar.

Med principalkomponentanalys kan man banta ner antalet variabler i data genom att bilda linjärkombinationer av dem och sedan studera dessa istället för hela variabelmängden.



Figur 2. Ett punktdiagram med en punkt markerad (Skansen). Motsvarande observation blir då markerad i datafönstret som syns bakom diagrammet.

Manualen SAS (1993 a) ingår i SAS manualsamling. Den har två delar: en för hur man använder INSIGHT och en med definitioner och litteraturhänvisningar. Vill man kunna utnyttja alla finesser, behöver man läsa manualen i stort sett från pärm till pärm.

Användning inom SCB

SAS/INSIGHT används i nuläget såvitt vi vet av ganska få personer inom SCB. Det är dock troligt INSIGHT kommer till användning framöver som analysredskap för metodare och även som instrument i den vanliga gransknings- och rättningsprocessen i statistikproduktionen. INSIGHT är lämpligt för dataanalys, även makrogranskning. I Hedlin (1995) finns konkreta exempel på hur INSIGHT kan användas. SCB-skolan ger en kurs i INSIGHT.

Användargränssnitt

SAS/INSIGHT kräver inte någon djup kännedom och vana av "vanliga" SAS, eftersom INSIGHT är helt menystyrt och befriat från kraven på detaljkunskaper som utmärker basmodulen av SAS, på det sätt som den används inom SCB. Det är dock en stark fördel att kunna vanliga SAS, eftersom användaren i de flesta fall behöver komplettera INSIGHT med t ex beräkningsredskap.

In- och utflöde

Speciellt för INSIGHT gäller att resultaten kan sparas som "*output objects*" (se avsnittet SAS: detaljer, "ODS"). Datafilen visas i ett enkelt kalkylark. I nuvarande version går det inte att mata in data i INSIGHT, de måste komma från övriga delar av SAS-systemet (i nästa version kommer det att gå att registrera data direkt i INSIGHT). Bryggan mellan INSIGHT och övriga SAS går via SAS-filer (jfr diskussionen i Bilaga 2, "Överföring av data mellan programvaror"). Modifieringar gjorda i INSIGHT kan sparas i en ny fil, och ett i princip obegränsat antal SAS-filer kan samtidigt läsas in i INSIGHT, tills datorns minne tar slut.

Leverans, konstruktion, underhåll

Se avsnittet "SAS: detaljer".

Pris

Se avsnittet "SAS: detaljer".

Möjliga plattformar och övriga hårdvarukrav

För SAS för Windows går det en gräns vid ca 1200 observationer som inte tycks vara beroende av minnets storlek. Ovanför den gränsen kan inte punkterna i ett punktdiagram vara markerade på olika sätt (t ex en sorts markering för de observationer som tillfälligt utesluts ur beräkningar och en annan för dem som ska ingå). Det går dock att använda olika färger. Eftersom hela datamaterialet läses in i arbetsminnet bör nog detta vara minst 10 MB. En fil på 100 variabler och 10.000 observationer (i

sig ca 2 MB) hanteras ganska bekvämt med ett minne med 16 MB storlek. Då kan även vissa andra programvaror vara aktiva, såsom Microsoft Mail. Att rita en matris av 100 punktdiagram tar dock lång tid. INSIGHT ritar dessutom om *hela* matrisen vid varje liten förändring. Det tog ca 3 min på såväl en bärbar Twinhead med 8 MB minne och en PS/2 modell 77 486DX2 med 16 MB minne. I det senare fallet användes nätverket med den elektroniska posten aktiv. Båda apparaterna hade en processor som går i takten 66 MHz.

En liknande produkt finns tillgänglig på Macintosh, men kallas där JMP (uttalas som engelskans "jump").

Övrigt

De operationer man gör under ett arbetspass kan inte sparas. SAS Institute har utlovat en "bandspelare" till version 6.10. Datafilen visas i något som ytligt ser ut som ett kalkylark, men som har en starkt begränsad redskapslåda för datamanipulering. Text får $x+y+z$ "värde saknas" om någon av termerna saknar värde, medan man i SAS basmodul kan välja mellan detta och att räkna partiellt bortfall som noll (m h a funktionen Sum). Version 6.08 är den första versionen av INSIGHT för Windows, och som man skulle kunna förmoda finns det ganska gott om buggar (ett ex visas i Figur 2, ett annat är att den metod att rita fler än en pensel i ett diagram som finns tydligt beskriven i manualen inte fungerar i praktiken).

11. SAS/STAT

Bakgrund och användningsområde

SAS/STAT är en av SAS-systemets "statistiska" moduler. Andra är

- BASE, (basmodulen)
- ETS (tidsserieanalys)
- QC (kvalitetsstyrning och en del deskriptiv statistik)
- LAB (regressions- och variansanalys med kontroller av förutsättningarna för modellerna och annan handledning, intressant för dem som saknar egna check-listor för den här typen av arbete)
- INSIGHT.

Dessa moduler finns inom ett stort och snabbt växande system av programvaror som utgör "the SAS System". Vissa aspekter på detta system ges i avsnittet nedan "SAS: detaljer".

Systemet av manualer är omfattande och svåröverskådligt. De inledande kapitlen i SAS (1989 b) ger en översikt över vissa användningsområden och motsvarande redskap i SAS/STAT. Indelningen i områden överensstämmer ungefär med den i Figur 3, sid 21. Det finns i manualen även ett kapitel för icke-parametriska metoder och ett för "Lisrel" (se Proc Calis nedan). Varje kapitel innehåller rikligt med litteraturreferenser och hänvisningar till andra delar av SAS, inklusive "SUGI Supplemental Library User's Guide, Version 5", som tar upp procedurer skrivna av användare av SAS. Det finns dessutom manualer för vissa områden som söker fylla igen luckan mellan programvarumanualer och teoretiska böcker i ämnet (framför allt tidsserie-

analys och analys av linjära modeller) och en rad "Technical Reports" om detaljer i beräkningssätt och om vissa nyare ämnen som tydligen inte fått plats i någon annan manual, t ex conjointanalys (Technical Report R-109, "Conjoint Analysis Examples").



Figur 3. Procedurer i SAS/STAT. Kopia av hjälpfönstret under "Data Analysis".

Det här antyder något om programvaran: SAS-systemet är omfattande och mångfacetterat. Det innehåller redskap för snart sagt varenda tänkbar tillämpning, men de är ofta halvfabrikat: användaren måste själv lägga till något innan de kan användas. Det ger stor flexibilitet man samtidigt en ganska hög ingångströskel för de flesta. Det här avsnittet tar upp procedurer kan användas direkt som de är, givet att data föreligger i lämplig form. I praktiken använder många statistiker SAS även på ett mer avancerat sätt, och fler moduler än bara SAS/STAT.

Bortfall

Partiellt bortfall behandlas olika av procedurerna. För varje procedur finns i manualerna ett avsnitt "Missing Values" som redogör för detta och vilka valmöjligheter som finns. Om användaren matar in en vanlig SAS-fil i proceduren, gäller för de flesta procedurer att observationer med något partiellt bortfall utesluts helt ur beräkningarna. Det framgår dock inte att om proceduren har en CLASS-sats, utesluts en observation ur beräkningarna om någon av variablerna som användaren räknar upp i satsen har partiellt bortfall, *även* om den variabeln inte ingår i modellen. Man ska alltså inte ta med onödiga variabler i CLASS-satsen. Matar användaren in en annan typ av SAS-fil än den vanliga, t ex en kovariansmatris, hanteras bortfall på annat sätt. Dessa andra typer av filer ger ytterligare möjligheter att kompensera för bortfallet: man kan

tex skapa en kovariansmatris med någon typ av imputering, och använda den i proceduren.

I SAS-filer kan partiellt bortfall lagras på olika sätt, om man vill skilja mellan tex att uppgiften inte är relevant för just den observationen, att den är relevant men ännu ej registrerad eller att man misslyckats med att mäta den. Procedurerna gör dock ingen skillnad mellan dessa olika representationer av bortfall, allt behandlas som "missing values".

Några nyheter

Här tas endast några nyheter upp, kortfattat. SAS Institute har under de senaste åren återigen satsat stort på statistikdelen av SAS-systemet.

Den mest aktuella, samlade beskrivningen av procedurerna nedan finns f n framför allt i SAS (1992) samt tillägg som gäller version 6.08 för procedurerna Calis, Lifetest och Mixed, se Technical Report P-242. Ytterligare tillägg för version 6.10 finns i ännu en manual.

Calis (Covariance Analysis and Linear Structural Equations)

Proc Calis berör samma område som programvaran LISREL (se "Andra programvaror" ovan). Kapitlet "Introduction to Structural Equations with Latent Variables" i SAS (1989 a) ger en överblick över CALIS användningsområde. Manualen SAS (1992) jämför estimatorer o d i Calis med motsvarande i LISREL version 6 eller 7. Calis introducerades i version 6.04 av SAS och är något utbyggd i SAS för Windows.

GenMod (Generalized Linear Models)

Proc GenMod täcker generaliserade linjära modeller och är en utbyggnad av Proc GLM som täcker linjära modeller såsom regressionsmodeller och variansanalys. Exempel på generaliserade linjära modeller är förutom den "vanliga" linjära modellen som används i regressions- och variansanalys, även logit- och probitmodeller i analys av proportioner och log-linjära modeller i analys av frekvenser. GenMod ger användaren stor frihet att definiera beståndsdelarna i den generaliserade linjära modellen. Resultaten från Proc GenMod kan antingen läsas av i OUTPUT-fönstret eller skickas till det s k Output Delivery System (se "SAS: detaljer"). Proceduren är nu på experimentstadiet och lär komma i färdigt skick i version 6.10. Den är nu dokumenterad i en separat "Technical Report" och dessutom finns en beskrivning av syntaxen i Hjälpfunktionen i SAS version 6.08. Även Proc NLin kan användas för generaliserade linjära modeller med en kraftig utbyggnad (se SAS 1989 b). De procedurer som behandlar ett specialfall av den generaliserade linjära modellen (tex Proc Logistic för logistisk regression) innehåller fler redskap för just det fallet än GenMod.

GLMmod (GLM model-building)

GLMmod är ett stöd vid användning av SAS-procedurer för regressionsanalys (tex Logistic), eller egenhändigt skrivna program i SAS matrispråk IML (se "Andra programvaror", sid 6). Man har som inflöde en SAS-fil med försöksdata och anger mha GLMmod en modell (tex "model utfall = x1 x2*x3;") och utflödet blir en SAS-fil med designmatrisen. Ett exempel är att om en regressionsmodell anpassas till data från en urvalsundersökning, kan det vara lämpligt att prova modeller med termer av typen πx , där π är urvalssannolikheter och x en förklarande variabel. Proc Reg kan

inte bilda produkten om infilen innehåller dels π , dels x . Ett annat exempel är om man vill göra om en kategorisk variabel till dummyvariabler. Använder man IML, vars inflöde alltid är matriser, kan GLMmod ge ett enkelt sätt att ur en vanlig SAS-fil skapa en lämplig matris.

Lattice

Proc Lattice utför varians- och kovariansanalys av experiment med vissa typer av ofullständiga försöksplaner (t ex romerska kvadrater).

Mixed (Mixed Linear Models)

Proc Mixed liknar Proc GLM (linjära modeller). Namnet antyder väl att Mixed klarar även slumpmässiga effekter, förutom fixa effekter, vilket dock även GLM gör. Skillnaden är att Mixed beräknar parametrarna som ML-skattningar eller "restricted ML". Resultaten från Proc Mixed kan antingen läsas av i OUTPUT-fönstret eller skickas till det s k Output Delivery System (se "SAS: detaljer").

MDS (Multidimensional Scaling)

Multidimensionell skalning är en rubrik varunder en mängd metoder finns. Ur en matris av skattningar av distanser mellan objekt ritas en karta över relationer mellan objekten. Kan användas i sammanhang där distanser (i vid mening) mellan varje par av objekten kan gissas eller skattas, och man vill ha en bild (bokstavligen) av objektens förhållande till varandra. Förefaller mest vanligt inom psykologin.

MultTest (Multiple Testing)

Proc MultTest justerar de värden på sannolikheter som man får vid hypotesprövning. Tanken är att om 5 % anses vara en lämplig kritisk gräns för ett enda hypotestest, ska den gränsen justeras nedåt om flera liknande test utförs efter varann eftersom sannolikheten annars är hög att ett eller par av säg 10 test får signifikant utfall.

PHreg (Proportional Hazards Regression)

PHreg utför regressionsanalys med en modell som kallas "Cox-regression" eller "proportional hazards" och som används vid analys av överlevnadsdata.

För övriga aspekter av SAS/STAT se "SAS: detaljer".

12. SPSS

Bakgrund och användningsområde

Förkortningen SPSS står numera för "Statistical Products and Service Solutions". Ursprungligen innehöll namnet "social sciences". SPSS kan väl fortfarande anses vara en programvara för främst statistiska tillämpningar, men den är inte längre begränsad till socialvetenskap. SPSS är ett stort paket av statistiska redskap, inklusive tidsserieanalys och överlevnadsanalys. Möjlighet att förse redskapen med egna kovariansmatriser med skattningar av kovarianser på basis av urval finns (jfr avsnitt "SAS: detaljer" under "analys av data från urval"). Jag har inte jämfört redskapen hos SPSS och t ex SAS för att se vilka redskap den ena har som den andra saknar. PC Magazine (1993) anger att SPSS saknar endast regressionsanalys med censurerade data och robust regression, men har inte heller gjort någon djupare genomgång.

Svenska Statistikersamfundets tidning *Qvartilen* innehöll i nr 1994:3 två artiklar om SPSS.

SPSS säljs i moduler. "Professional Statistics" innehåller bl a faktor- och diskriminantanalys, "Advanced Statistics" överlevnadsanalys, "Trends" tidsserieanalys. Alla dessa kräver att basmodulen är installerad. Ytterligare moduler finns.

De omvittnat lättillgängliga manualerna är skrivna av Marija Norušis, som varit manualförfattare hos SPSS sedan starten 1968. Manualerna innehåller dock inte så många utförliga exempel. Det finns en särskild formelsamling, SPSS (1991). Trots titeln "Statistical Algorithms" innehåller den inte algoritmer, endast formler.

SPSS används i undervisningen i flera högskolor. Ett bra kompendium för nybörjaren är Brännström, Skalin (1994).

Användning inom SCB

Ola Hagfall, AM/UA, använder SPSS i statistikproduktionen. En bredare användning av SPSS i UAs produktion är på försöksstadiet. Ett antal SCB:are har använt SPSS i sin verksamhet vid Handelshögskolan, Stockholms universitet eller någon annan högskola, e d.

Användargränssnitt

SPSS är helt anpassat till Windows. Användargränssnittet följer den standard, CUA, som numera väl får sägas vara allmänt accepterad.

In- och utflöde

SPSS kan importera och exportera filer i följande format:

- Excel (även kalkylark sparade i SYLK-format).
- Lotus 1-2-3 rel. 3.0, dBase IV. Dock av allt att döma problem med konvertering av ASCII och ANSI, om överföring sker mellan DOS-programvaror och SPSS. Vid utflöde maximalt 256 variabler.

Data från följande databashanterare kan importeras till SPSS (ej exporteras):

- Oracle version 5.1 eller högre
- SQL Server version 1.1 eller högre
- SQL-databaser som följer ODBC (se bilaga 2 för en allmän beskrivning av ODBC).

Export till databaserna kan ske m h a en fristående modul, DBMS/COPY, som kan översätta mellan ett 50-tal PC-programvarors filformat. En genomgång av DBMS/COPY finns i Hallahan (1992) och Wills (1992).

SPSS har tillgång till OLE version 2 (se bilaga 2). Ett praktiskt exempel av OLE kan vara att en bild som skapats med SPSS och infogats i ett Word-dokument är åtkomlig från Word, man kan dubbelklicka på bilden i Worddokumentet och därigenom starta SPSS grafikeditor. SPSS stöder endast "linking", inte "embedding". Det innebär att om originalbilden förändras i SPSS, förändras även bilden i Worddokumentet (det är nämligen samma bild). Endast grafer kan länkas, ej text eller data. Riktningen är från SPSS till annan programvara, dvs graferna måste vara skapade i SPSS.

SPSS stöder även DDE (se bilaga 2). SPSS kan bara fungera som server, inte som klient.

Från klienten kan man:

- starta SPSS (köra ett kommando som startar)
- öppna och stänga filer i SPSS
- köra procedurer i SPSS
- ta över data, tabeller och listor, grafik
- stänga ev. öppna SPSS-filer och avsluta SPSS.

Tabeller i SPSS resultatfönster innehåller osynliga tabbar. Det gör det lätt att kopiera tabellerna och klistra in dem i ett Word-dokument, där "omvandla text till tabell" beställs. Tabbarna gör att talen hamnar i rätt kolumn.

För överföring av data till och från SAS, se Bäcklund (1994) och "SAS: detaljer" nedan.

Vid inläsning av textfiler finns ett antal redskap för översättning. Text kan packade fält översättas till vanlig lagring men inte överstansade fält (se avsnittet bilaga 2). Vid export av SPSS-filer till textfiler finns det färre redskap. Programmeringsspråket är inte heller så kraftfullt, text går inte det inte inom ramen för SPSS att göra en inläsningsrutin för "lok med vagnar".

Leverans, konstruktion, underhåll

Programvaran konstrueras och underhålls av SPSS Inc, i USA. Den senaste versionen har nummer 6.0. Nordisk leverantör är SPSS Scandinavia AB, Gamla Brogatan 36-38, 111 20 Stockholm. SPSS Inc har nyligen köpt SYSTAT, som är en annan stor statistisk programvara.

Pris

En uppsättning manualer och tekniskt stöd ingår. Det tekniska stödet fungerar enligt uppgift väl. Köper man en licens för färre än 9 användare, levereras programvaran med kopieringsskydd i form av en plugg, med nio eller fler användare kan man få SPSS med tidsbegränsning (programvaran upphör att fungera efter ett visst datum).

Möjliga plattformar och övriga hårdvarukrav

Lämpligen en snabb PC (text 486-processor, 50 MHz) och arbetsminne på 8 MB eller mer. Matematikprocessor snabbar upp kraftigt. Andra plattformar är bl a Macintosh, IBM stordatorer och ett flertal UNIX-datorer, bl a IBM RS/6000 (en sådan innehas av V/IF, SCB).

Övrigt

SPSS har "chart templates", mallar för bilder. Om användaren gör text ett histogram med ett speciellt utseende (val av färger, typsnitt, speciell gruppering av staplar etc), kan detta sparas i en mall. Sedan kan han ta fram andra data och rita om histogrammet med likadant utseende med några få musklick.

Till de statistiska procedurerna kan egna anteckningar knytas. I en central installation av SPSS kan man tänka sig att skriva en text, text

"den här proceduren behandlar partiellt bortfall på det-och-det sättet" som enkelt kan visas när någon kör proceduren.

I likhet med SAS är SPSS tabelleringsverktyg något begränsat. En tabell av typen nedan får man inte alldeles enkelt:

Antal mobiltelefoner,	x
därav Stor-Stockholm	x

SPSS har inte frågespråket SQL för generellt bruk (se avsnittet bilaga 2 nedan). SQL kan endast användas för att hämta data ur en databas som stöder ODBC, inte t ex för att matcha SPSS-filer med varandra.

SPSS saknar inbyggd "client/server" (se avsnittet "SAS: detaljer", Connect).

Med SPSS' API (se även bilaga 2) kan man via menyer och dialogfönster skriva ett program som t ex startar Excel inifrån SPSS, automatiskt för över den aktuella arbetsfilen till ett kalkylark i Excel. I Excel kan ett makro skapas, som öppnar dialogfönster varmed användaren kan dirigera tillbakaflödet till Excel. Det användaren ser i SPSS är en knapp eller ett menyval där det står t ex "Excel" eller "kalkylark". Överföring sker med DDE. Man kan också tänka sig en knapp i Word som sätter igång någon SPSS-procedur och levererar svaret till Word. Användaren kan då när han skriver på sitt Worddokument låna ett redskap från SPSS genom att välja bland Words ordinarie knappar och menyer.

13. SUDAAN

Bakgrund och användningsområde

1971 inledde RTI arbetet med programvara för urvalsundersökningar. I slutet på 80-talet skrevs programvaran om i C. Ungefär samtidigt konstruerades den oberoende programvaran SUDAAN (förkortning av Survey Data Analysis), tidigare var det en samling procedurer som kördes inom SAS.

I SUDAAN är användningen av urvalssannolikheter mer integrerat än i t ex SAS. SUDAAN har färdiga redskap för skattningar ur data från ett antal typer av urvalsdesigner, bl a strategi 1-4 (se Tabell 2, sid 9) och flera typer av flerstegsurval.

SUDAAN beräknar punkt- och variansskattningar för totaler, kvoter och dubbelkvoter för icke-överlappande redovisningsgrupper. Kontraster mellan estimat olika redovisningsgrupper kan beställas. SUDAAN kan även utföra beräkningar i samband med logistisk och vanlig regression, analys av kategoriska data med linjära eller loglinjära modeller samt analys av överlevnadsdata.

Om endast en svarande finns i något stratum, kan användaren välja att beräkna variansbidraget för den observationen som väsentligen skillnaden mellan den observationens värde och hela populationens skattade medelvärde. I analys av modeller utesluts en observation helt om någon av de variabler som ingår i modellen har partiellt bortfall för den observationen.

Ett mer generellt redskap för vissa generaliserade linjära modeller är under utveckling (Shah, LaVange 1991, jfr även avsnittet SAS/STAT ovan).

SUDAAN skattar varianser i de flesta fall med en Taylorutveckling, som beräknas med Woodrufftransformationer (för en överblick av den senare tekniken se t ex Andersson, Nordberg, 1992). I vissa sammanhang kan inte den parameter som ska skattas definieras explicit som en funktion av stickprovet, utan endast implicit. I SUDAAN förekommer sådana i modeller för logistisk regression och överlevnadsanalys. För att skatta variansen av sådana parametrar använder SUDAAN Binders "implicit differentiation method" (det är en allmän metod att skatta variansen för en komplex parameter, för en beskrivning se t ex Särndal, Swensson, Wretman 1992). Konfidensintervall för medianskattningar skattas med något som är en utveckling av Woodruffs metod.

Manualen är Barnwell, Hunt and LaVange (1992). Den saknar sakregister. Den statistiska dokumentationen finns fin i ett utkast, Shah et. al. (1993). Båda kan lånas inom SCB av Dan Hedlin.

I Pettersson, Selén (1990) finns en genomgång av SUDAANs egenskaper. Delar av den här texten är hämtad från den källan. Carlson m fl (1993) går igenom SUDAAN och jämför med PC CARP, se "Andra programvaror", sid 6.

Användning inom SCB

SUDAAN har använts inom SCB endast i få fall. Ett exempel finns i rapporten "Lyberg, I. (1989). SCB har köpt några icke-överförbara exemplar av SUDAAN för PC med versionsnummer 6.20. Den här texten bygger på den versionen.

Användargränssnitt

Inflödet ska vara sorterat på ett visst sätt (t ex efter stratumkod vid stratifierat urval). Eftersom SUDAAN inte innehåller något redskap för sortering, ska detta göras innan data går in i SUDAAN. Därför, och av andra skäl, behöver SUDAAN ett komplement. Eftersom SUDAAN läser SAS-filer, är SAS det naturliga komplementet: en fil framställs med SAS och för vissa typer av beräkningar kan SUDAAN användas med denna fil som inflöde. Exempel 7 i bilaga 1 visar hur man kan arbeta med SUDAAN utan att lämna SAS. SUDAAN kommer då att för användaren framstå som en utvidgning av SAS.

Användaren skriver en textfil med kommandon, som tillsammans blir ett program i SUDAAN-språk. Se exempel i Figur 4. Rubriken och utskriften i figuren är hämtade från SUDAANs exempelprogram (något förkortade här). De numrerade raderna i utskriftens början återger det program som genererade utskriften. Sättet att ange kommandon har ytliga likheter med SAS, men inte ens en van SAS-användare kan skriva SUDAAN-program utan först lära sig SUDAAN. Ett exempel är att i SUDAAN ska namn på kataloger i DOS ha dubbelt bakåtvänt snedstreck, \, istället för enkelt, \. Ett annat är att för kategoriska variabler ska antalet olika värden anges av användaren (SUDAAN kan inte räkna dem själv). Rad 5 i programmet i Figur 4 talar om att variablerna ras och kön på rad 4 har två nivåer och "poverty status" har tre. Om detta ska anses besvärligt är en smaksak, men en viss lärotid måste man alltså räkna med. SUDAAN har inga hjälptexter som kan visas på skärmen.

"Without replacement (WOR) design, TOTALs, MEANS GEOMETRIC options, SAS input, post-stratification".

SUDAAN Survey Data Analysis Software
Copyright Research Triangle Institute September 1992
Release 6.20

```
1  PROC DESCRIPT DATA=HIS85 FILETYPE=SAS DESIGN=WOR
    TOTALS MEANS GEOMETRIC;
2  NEST COL_STR PSU_ID SUB SEGMENT;
3  TOTCNT POPPSU NSUB PSEG NPER;
4  SUBGROUP RACE SEX POV;
5  LEVELS 2*2 3;
6  WEIGHT WTF;
7  POSTVAR POV;
8  POSTWGT 700000 600000 500000;
9  TABLES RACE*SEX;
10 VAR TDV HEI ;
```

Number of observations read : 310
Denominator degrees of freedom : 4

Date: 09-23-92
Time: 13:25:20

Research Triangle Institute

Page : 1
Table : 1

Post-stratified estimates
by: Variable, RACE, SEX.

for: Variable = TDV.

RACE		SEX	
		Total	Male
Total	Sample Size	310	155
	Weighted Size	1800000.00	766263.17
	Total	15149151.15	2569163.46
	SE Total	5923388.35	753301.65
	Mean	8.42	3.35
	SE Mean	3.29	0.95
	Geom. Mean	36.18	28.44
	SE Geom. Mean	5.53	1.29
Black	Sample Size	32	20
	Weighted Size	106895.06	52095.26
	Total	1534958.72	466060.82
	SE Total	659591.15	220542.59
	Mean	14.36	8.95
	SE Mean	3.08	2.00
	Geom. Mean	37.07	32.37
	SE Geom. Mean	4.04	1.46

Figur 4. Exempel på program och utskrift i SUDAAN.

In- och utflöde

Inflödet till SUDAAN kan vara endera flata ASCII-filer eller SAS-filer. Dessa måste vara av den gamla DOS-versionen 6.04. Den som använder SAS för Windows kan lätt skapa sådana, dock med följande problem:

- Om filen är mycket stor kan det vara besvärligt att göra en kopia.
- Etiketter för variabelvärden ("format") är besvärliga att konvertera från SAS för Windows till DOS-versionen. Det torde vara enklast att skapa en textfil med den poststruktur som SUDAAN kräver för variabeletiketter (se exempel 8, bilaga 1).

SUDAAN kan inte hantera bokstäver i data (i framtidsplanerna finns att ta bort denna begränsning). SUDAAN kan använda följande metadata i en SAS-fil:

1. Variabelnamn.
2. Etikett för hela filen (som i SAS anges med satsen `DATA namn (LABEL='xxx')`; där *namn* är den SAS-fil som skapas)

3. Etiketter för variabelnamn. Skrivs automatiskt ut i tabeller.
4. Etiketter för variabelvärden, dock utan automatisk klassindelning som "format" i SAS ger.

Resultaten levereras på två sätt, dels som

- standardtabeller utformade för mänskliga ögon (Figur 4);
- textfiler med en viss given struktur för elektronisk behandling.

De senare består förutom av en fil som innehåller själva data även av en fil med postbeskrivning av datafilen samt en grupp av filer med metadata enligt punkt 1-4 ovan. Filen med postbeskrivning gör det möjligt att skapa en inläsningsrutin i en annan programvara, t ex SAS, så att SUDAANs utflöde enkelt kan föras över till den andra programvaran för vidare bearbetning eller presentation. Standardtabellerna kan modifieras med ytterligare programrader i SUDAAN-programmet, men det går inte att få diagram eller proportionellt typsnitt.

Vikter som motsvarar urvalssannolikheter anges på olika sätt beroende på design. För strategi 1 och 3 (se tabell Tabell 2, sid 9) förser användaren infilen med en variabel vars värden är de vikter som observationerna ska vägas med och en annan som innehåller antalet objekt i populationen per stratum. SUDAAN kommer själv att räkna antalet svarande. För strategi 2 och 4 ska dessutom variabler som ger grupper och vikter för poststratifieringen med. För flerstegsurval med t ex pps i första steget ska skall första och andra ordningens urvalssannolikheter finnas med.

Leverans, konstruktion, underhåll

SUDAAN utvecklas och säljs av Research Triangle Institute i USA. Information om programvaran lämnas snabbt och effektivt av Ms. Sandra M. Paschall; Research Triangle Institute, Statistical Software Center; P.O: Box 12194; Research Triangle Park, NC 27709-2194; USA; faxnummer: + 1 919 541 6178.

Möjliga plattformar och övriga hårdvarukrav

SUDAAN finns för både PC och stordatorn. Version 6 kan inte köras på PC med 286-processor eller sämre. För PC är utökat minne nödvändigt och matematik-processor rekommenderas av RTI. SUDAAN finns dessutom för DEC- och Sun-datorer.

Pris

PC-versionen kostar i enstaka ex. 995 USD (50-550 USD för uppgradering), 4000 USD (2000 för uppgradering) för installation på ett nät med obegränsat antal användare. Tekniskt stöd lämnas av RTI mot extra betalning ("small amounts of consulting" dock gratis).

Övrigt

SUDAAN har ganska enkla felkontroller. Manualen varnar på åtskilliga ställen för att om vissa saker görs fel, t ex att filen inte är sorterad, kan programvaran lämna "unpredictable results". Å andra sidan finns kontroller för de numeriska algoritmer-

nas stabilitet: procedurerna kontrollerar sina egna beräkningar och varningar skrivs ut.

Paneler finns inte med som en speciell design.

Jag gissar att någon full anpassning till Windows inte kommer att göras (t ex samma typ av användargränssnitt som windowsanpassade programvaror har). Däremot har kommunikationen med SAS förbättrats för SUDAAN på en IBM-stordator så att SUDAANs funktioner lättare kan utnyttjas för SAS-användare. Den utvecklingen kan väl förväntas även på PC.

14. SAS: detaljer

Detta avsnitt är ett försök att sammanfatta och lyfta fram en mängd detaljer om SAS som rör användargränssnittet och möjligheter till in- och utflöde.

SCB har installerat SAS för Windows, version 6.08, centralt på nätverket. Där finns modulerna

- BASE
- STAT
- GRAPH
- ETS (tidsserieanalys)
- FSP (registrering av data och system för editering)
- AF och EIS (två system menybyggnad, se "Användargränssnitt nedan")
- ASSIST (se "Användargränssnitt nedan")
- IML (matrisspråk)
- CONNECT (se nedan)
- INSIGHT
- ACCESS mot databashanteraren Sybase/SQL (se "Frågespråket SQL" nedan och "Kommunikation med SQL-orienterade databaser", sid 34)

Det finns knappast tekniska hinder mot att installera ytterligare moduler på enskilda arbets-PC.

Version 6.04 är den senaste och troligen sista DOS-versionen av SAS.

Frågespråket SQL

Se bilaga 2 för SQL i allmänhet. I SAS kan SQL användas såväl i kontakter med databaser som internt. Användaren kan använda SQL genom att skriva programkod (exempel 6 i bilaga 1 som visar en enkel variansskattning) eller klicka sig fram i en del av mensystemet ASSIST. Fler exempel finns i SAS (1989 c).

Analys av data från urval

Vid analys av data från urval med urvalssannolikheter som varierar över urvalsobjekten, ställer man sig frågan om observationerna ska vägas eller inte. Det finns en omfattande litteratur om det. Vi antar här att vi vill väga data. Här visas ett sätt att få in vikterna i vissa av procedurerna i SAS/STAT. Vill vi inte väga, kan vi använda procedurerna som de är (och ev. utöka modellen med variabler som korrelerar med urvalssannolikheter).

Kovariansmatrisen är den grundläggande byggstenen i många analysmetoder, t ex regressionsanalys. Om datamaterialet från en undersökning består av uppgifter från individerna $1, 2, \dots, n$, som svarat på q frågor, har vi för en enskild individ en post av data $(y_{1k}, y_{2k}, \dots, y_{qk})$. Över de n posterna kan vi skatta kovarianser för varje par av variabler genom att väga med de inverterade urvalssannolikheterna. Dessa skattningar ordnas i en kovariansmatris som är en skattning av den kovariansmatris som avser populationen och som kan användas i beräkningar av komplexa parameter-skattningar. Se vidare Särndal, Swensson, Wretman (1992, avsnitt 13.2 och framåt).

Vissa av procedurerna i SAS ger användaren möjlighet att som indata välja en egenproducerad kovariansmatris. Det kan vara ett tämligen enkelt sätt att skatta parametrar med hänsyn till att observationerna har olika inklusionssannolikheter. Däremot erbjuder inte procedurerna något redskap för att beräkna precisionen för dessa skattningar. En annan fördel kan vara hanteringen av partiellt bortfall (se "SAS/STAT" ovan). Se SAS (1989 a, "The CATMOD procedure", ex. 10) för ett exempel på variansanalys av kategoriska data från en urvalsundersökning.

De viktigaste procedurerna som kan läsa in en kovariansmatris eller motsvarande framgår av Tabell 3. Se närmare appendix 1 i SAS (1989 b).

Procedur	Användning
CALIS	se avsnitt SAS/STAT ovan
CATMOD	analys av kategoriska data, kontingenstabeller
CANCORR, FACTOR, PRINCOMP	söker samband mellan variabler utan att betrakta någon eller några som beroende och de övriga som oberoende
DISCRIM, CANDISC, STEPDISC	klassificering
REG	regressionsanalys
VARCLUS	gruppering

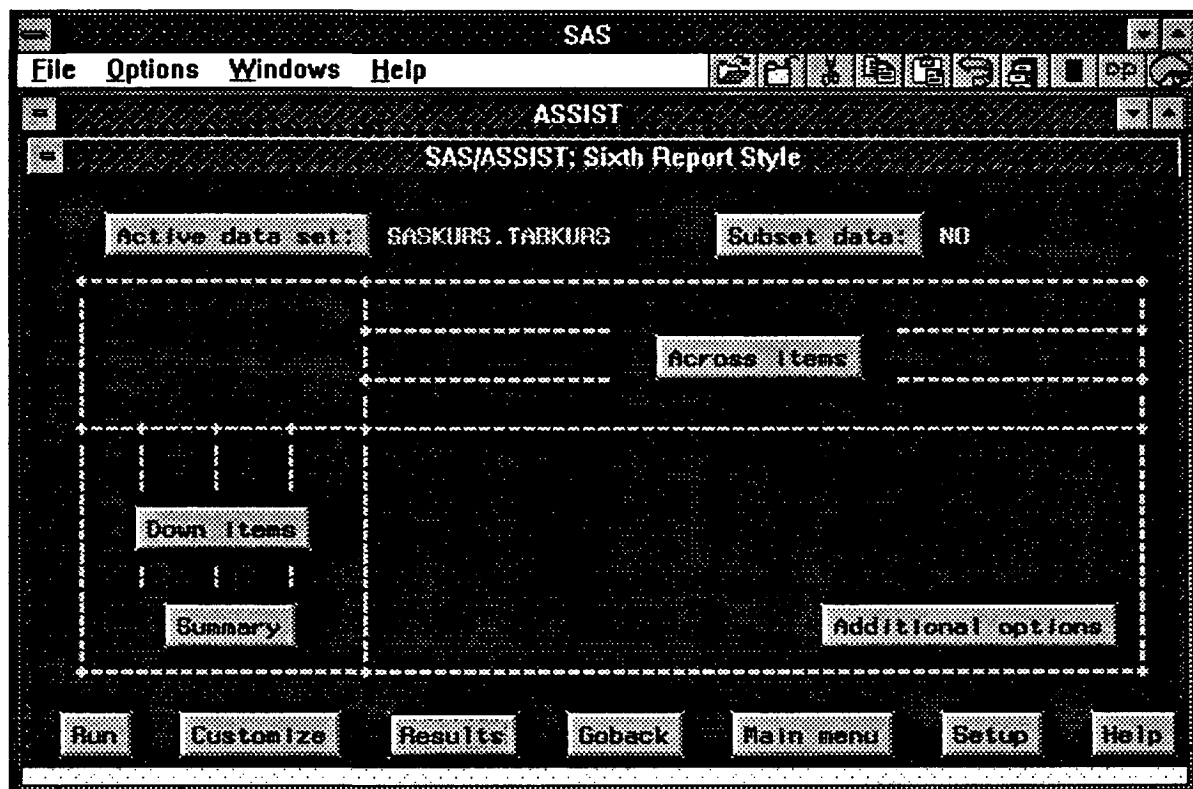
Tabell 3. SAS-procedurer som kan läsa in kovariansmatriser eller motsvarande.

Användargränssnitt

SAS 6.08 är inte fullt ut anpassat till Windows. Användargränssnittet följer inte de facto-standarderna i alla dess detaljer. Ett litet exempel bland andra är registret över framställda grafer som kallas "Graph Manager". Det enda sättet att välja en graf i listan är att skriva "S" eller "X" på ett litet streck framför namnet på den önskade grafen. Nästa version, 6.10, är bättre anpassad till Windows, bl a är problemet med Graph Manager löst. För användaren finns fyra typer av intressanta användargränssnitt:

- Progmenditorn är det som oftast används inom SCB och förmodligen även på andra håll där statistiker arbetar.
- ASSIST är huvudsakligen en programgenerator byggd i SAS eget system för konstruktion av applikationer, SAS/AF. Med ASSIST genererar användaren vanliga SAS-program genom att välja i menyer. Dessa har inte det användargränssnitt som vanliga Windowsprogram har, se Figur 5. De genererade programmen är trevligt skrivna, se Figur 6. Det går inte att få fram vilka program som helst med ASSIST. Avancerade SAS-användare kan därför tycka att ASSIST är väl begränsad.

- Vissa delar av SAS har menystöd, som påminner om ASSIST, men som ingår i den vanliga basmodulen eller i någon annan tilläggsmodul. Till skillnad från ASSIST får man dessa menysystem utan extra kostnad. Ett exempel är tidsserieanalys.
- Användargränssnitt konstruerade speciellt för någon viss tillämpning m h a av modulerna SAS/AF och SAS/FSP. En SAS-kunnig person kan skriva program och sedan konstruera en meny varmed användaren kan styra SAS-programmet (liksom den meny som visas i Figur 5). Den möjligheten har används sparsamt inom SCB. Ett prov redovisas i Hedlin (1991). Lars Johnsson, SCB, har gjort en mer modern tillämpning: ett system för Resvaneundersökningarna. Med det tar ett 20-tal användare utanför SCB fram tabeller ur data om resvanor på ett flexibelt och samtidigt lättillgängligt sätt. Antal användare kommer att öka.



Figur 5. En meny för framställning av tabeller.

ODS (Output Delivery System)

Det s k "Output Delivery System" finns i en experimentversion. Resultat som visas i Outputfönstret är textfiler som är ganska besvärliga att ta hand om. Med ODS däremot består resultaten av rutor ("output objects"). Man kan klicka på en ruta, och t ex begära att den sparas som en SAS-fil, så att resultaten enkelt kan ingå i ytterligare beräkningar. De flesta SAS-procedurer kan visserligen redan lämna ifrån sig en SAS-fil, men den har i många fall annan form och annat innehåll än utskriften i Outputfönstret. ODS ger en enkel möjlighet att föra över tabeller till t ex en ordbehandlare: tabellen sparas som en SAS-fil, och denna skrivs ut med någon symbol som avskiljer kolumnerna från varandra och som ordbehandlaren kan uppfatta (exempel 3 i bilaga 1).

Utskriften från en procedur som stöder ODS sker via mallar till skärmen. Dessa mallar kan kopieras och redigeras efter egen smak. Version 6.08 av SAS har färdiga mallar för endast Proc Mixed. De finns i SASHELP.OUTPUT.

F n kan bara SAS/INSIGHT, Proc GenMod och Proc Mixed få fram 'output objects'. Dessa hanteras med Proc Output och Proc Template som ingår i SAS/BASE men f n är dokumenterade i SAS (1992) och i någon mån i SAS (1993 a). SAS Institute kommer att bygga ut ODS kraftigt framöver.

```

/*-----*
| Summary:
|   This program creates a multilevel tabular report using the
|   data set SASKURS.TABKURS and displaying up to five across
|   and down levels.
|   Generated: 22JUN94 14:27:50
|-----*
| The OPTIONS statement specifies the dimensions of the printed
| output in characters per line and lines per page and whether or
| not the current date and page number are printed.
| The TITLE and FOOTNOTE statements clear any previously defined
| titles or footnotes.
|-----*
options linesize=75 pagesize=23 date number pageno=1;
title;
footnote;
/*-----*
| TITLE statements specify text printed at the top of the output
| page.
|-----*
title1 "Kurs i Proc Tabulate";
proc tabulate data=SASKURS.TABKURS
missing
format=6.1
;
class KON REGION CIV;
table (REGION ALL)*(CIV ALL),
      (KON)*(N PCTN)
      / misstext="." rtspc=19
;
keylabel ALL='Summa';
label KON = "Kön";
label CIV = "Civilstånd";
label REGION = "Region";
run;

```

Figur 6. Ett SAS-program framställt av ASSIST.

In- och utflöde

Textfiler

På PC har vi två teckensystem, "ASCII" och "ANSI" (se bilaga 2). Problem uppstår framför allt med å, ä, ö och heldragna linjer. SAS kan ställas in på endera teckensystemet. De flesta inom SCB har troligen SAS inställt på ANSI. ASCII-format är ett det vanliga formatet för programvaror som inte är anpassade till Windows. Läser man in sådana filer SAS, krävs en särskild åtgärd om översättning till ANSI önskas (exempel 1, bilaga 1).

Stordatorfiler

Sekventiella filer på DAFA Mittsystems stordator kan föras över "binärt" från stordator till PCn eller någon persondatorserver, dvs med filens 0:or och 1:or överförda i ograverat skick. SAS uppfattar filen som en stordatorfil trots att den ligger i PCn. Se exempel 10. Det går bra att föra över även packade fält och annat som är unikt för en stordator av IBM-typ och läsa dem på PC. En annat redskap för överföring är SAS/Connect (se avsnittet "Connect" nedan). Metoden i exempel 10 har dock två fördelar:

- billigare (lägre stordatorkostnad eftersom all hantering görs i PCn);
- inga speciella åtgärder behövs på arbets-PCn för att få igång SAS/Connect.

"Lok och vagnar"

Med SAS programmeringsspråk går det att läsa in filer med godtycklig struktur, t ex "lok och vagnar" (se bilaga 2), även om det kan kräva viss möda (exempel 5, bilaga 1).

Överstansade fält

SAS innehåller en översättningsrutin, *informat ZD*, som automatiskt skriver om överstansade tal (se bilaga 2). Se vidare SAS (1990 a).

Utskriften av resultat

En typ av ASCII/ANSI-filer i SAS är utskriften av resultat. Dessa är enkla, de saknar t ex tabbar som styr kolumnerna. Kopierar man en utskrift till t ex Word fungerar därför inte proportionella typsnitt, som Times, och inte heller Words tabellfunktion "omvandla text till tabell". Se exempel 3 för en teknik som kan få detta att fungera (jfr ODS ovan). För procedurerna Freq och Tabulate kan en speciell teknik användas.

Urklipp

SAS kan lagra följande format i Urklipp (se bilaga 2) förutom sitt eget textformat:

För såväl kopiera som klistra:

- Windows textformat (CF_TEXT)
- "*metafile format*" (bättre än bitmap, i den meningen att fler typer av information om bilden kan lagras)

Endast för kopiera, ej klistra:

- Windows bitmapformat
- "*device independent bitmap format*"

Kommunikation med SQL-orienterade databaser

SAS har byggt upp gränssnitt mot SQL-orienterade databaser som hanteras av vissa programvaror, bl a Sybase/SQL. SCB hyr en SAS-modul som innehåller redskap för överföring av data mellan SAS och Sybase-databasen. Se även ADB-rapport 1994:5 som beskriver tester där SAS används för t ex urvalsdragning ur RTB, lagt i en SQL-serverdatabas.

ODBC

SAS har byggt upp ett generellt gränssnitt mot databaser som följer ODBC (se bilaga 2 och manualerna SAS 1994a och 1994b).

DDE

I SAS används DDE (se bilaga 2) endast för överföring av variabelvärden (inte t ex bilder). SAS fungerar endast som klient och klarar bara kall och het länk. På det sätt som het länk är genomfört verkar det inte vara särskilt användbart.

Det fungerar ganska bra att föra över data mellan SAS och Excel med DDE. DDE kan användas istället för Proc DBF och Proc DIF som lyfts ut ur SAS basmodul och numera ingår i en modul som SCB inte hyr (exempel 2, bilaga 1).

OLE

SAS kan med OLE (se bilaga 2) föra över alla typer av datamängder som kan sparas i Urklipp, alltså även bilder. Liksom är fallet med DDE, fungerar SAS endast som klient. OLE i SAS går alltid över modulen SAS/AF. Denna måste vara installerad om man ska skapa en länk av OLE-typ, men om det redan är gjort kan länken användas utan SAS/AF.

Andra programvarors filer

CLAN och Extract

Varken CLAN eller Extract kan ta emot värdeförrådsbeskrivningar i SAS ("format"). Genom att omforma formaten till variabler går det dock, se exempel 9, bilaga 1.

SPSS

SAS kan läsa SPSS-filer om de är endera skapade av SPSS/DOS (ej av SPSS under Windows i "DOS-format") eller m h a SPSS "exportformat". Se Bäcklund (1994). I riktning *till* SPSS verkar variabeletiketter och användardefinierade format kunna föras över korrekt utan manuell översättning (det behövs dock något tillägg i SPSS-kommandot GET SAS). De format som ingår i SAS, till exempel *format w.d*, att talen ska skrivas ut med *d* decimaler, tycks inte gå att föra över. I riktning *från* SPSS verkar inte något annat än själva data föras över. Möjligen går det att samla metadata i en särskild SPSS-fil och översätta denna till en SAS-katalog.

Excel

SAS kan inte läsa Excelfiler direkt. SAS Institute har dock ett sådant redskap, men inte inom någon f n kommersiellt tillgänglig produkt. Kommunikation med Excel görs lämpligen med urklipp, DDE (se detta avsnitt) eller via DIF-format, vilket numera hör till en särskild modul i SAS-systemet. SCB-skolans kurs "Samverkan mellan PC-program" ger metoder för överföring av data mellan SAS och bl a Excel.

Connect

SAS har även byggt in "client/server" inuti SAS-systemet. Modulen, som kallas SAS/CONNECT, gör det möjligt för användaren att utforma program i PCn, köra de på stordatorn och läsa av resultaten på PCn. Den enda skillnaden sett från användarens synpunkt är att han ger kommandot "remote submit" istället för "submit". Förde-
lar är att stordatorns prestanda kan utnyttjas utan att belasta PCns resurser, att filer lämnas på en central plats, åtkomliga för flera, och att endast de intressanta delarna av filer behöver transporteras från stordatorn till PC:n. Möjligen fungerar detta även för en PC som är kopplad mot en server i ett nätverk.

Leverans, konstruktion, underhåll

Programvaran är gjord och underhålls av SAS Institute Inc., North Carolina, USA. Den säljs (eller snarare hyrs ut) av SAS Institute AB, Silverdals Gård, 191 40 Sollentuna. Enligt avtalet lämnas tekniskt stöd endast till en central funktion inom SCB, men man blir inte illa behandlad om man ringer ändå. F n sköter Sten Bäcklund ensam den centrala funktionen. SAS Institute AB:s policy är att framför allt satsa på stora kunder.

Dessutom finns en mindre del av SAS-systemet på Macintosh och kallas JMP (uttalas som engelskans "jump") som i Sverige säljs av Cortex Computer Center AB, Solna.

Pris

SCB hyr SAS för Windows genom ADB-enheten. För prissättning och aktuellt pris, hör med Sten Bäcklund. Dessutom har SCB tillgång till SAS på DAFA Mittsystems stordator.

Möjliga plattformar och övriga hårdvarukrav

SAS kan köras på en rad olika datorer och under åtskilliga operativsystem. Inom SCB har vi framförallt:

- SAS på stordatorn (version 6.08 installeras under våren -94);
- SAS på PC, - Windowsversionen (f n version 6.08)
 - "gamla" SAS för DOS (version 6.04).

SCB har även SAS på en IBM RS/6000 model 570 med operativsystemet AIX (dvs IBMs dialekt av UNIX).

För Windows kräver SAS främst stort arbetsminne (minst 8 MB, helst 10 MB). Det *går* att köra med mindre minne, men knappast tillsammans med den elektroniska posten. För SAS/INSIGHT se ovan.

JMP körs som sagt på Mac.

Bilaga 1. Exempel.

Exempel 1.

Satsen "infile" används när man läser in en textfil (ANSI eller ASCII) till SAS (se "Textfiler" i Bilaga 2). I satsen nedan är OEM är en förkortning av *Original Equipment Manufacturer* som står för ASCII:

```
infile 'filnamn' oem;
```

Om den fil som läses in är i ASCII-format, gör tillägget OEM att data översätts till ANSI-format. Eljest kan problem uppstå vid t ex utskrift på papper. Vill du se i vilket format dina SAS-data är lagrade kan du skriva ut dem med Proc Print, och ställa om "Fonts..." (i menyraden högst upp) till Courier. Förvandlas då alla å, ä och ö till konstiga tecken har du troligen SAS inställd på ANSI medan de data du har framme är i ASCII-format, vilket troligen medför att något annat än å, ä och ö kommer ut på papper. Se vidare manualen SAS (1993 b), appendix 4.

Exempel 2.

Följande program skickar ut variablerna LOPNR och BESOK i SAS-filen S.MUSEER till de två första kolumnerna i kalkylarket BLAD2 i Excel. Det dock finns åtminstone ytterligare tre problem att beakta:

- Olika representationerna av decimaltecken. I SAS används decimalpunkt, i de flesta SCBares installation av Excel decimalkomma (val av decimalsymbol görs i Windows och gäller alla Windowsanpassade programvaror samtidigt, dock ej SAS i version 6.08). En översättningsrutin finns numera i SAS (*informat* och *format COMMAX*, se exempel 4) som fungerar för överföring från Excel till SAS, men tydligen inte i den andra riktningen.
- Mellanslag i variabelvärden. Se vidare manualen SAS (1993 b), speciellt avsnitt "Using the NOTAB option with DDE" under "Dynamic Data Exchange" som ger exempel på överföring *till* Excel. Överföring i den andra riktningen verkar besvärligare.
- Partiellt bortfall. Se vidare manualen SAS (1993 b), speciellt avsnitt "Reading Missing data" under "Dynamic Data Exchange".

```
filename t DDE 'excel|c:\data\blad2.XLS!R1C1:R214C2'
      lrecl=300
;
data _null_;
  file t;
  set s.museer;
  put lopnr besok;
run;
```

Om observationen är längre än 132 tecken måste LRECL=x läggas till, där x är ett tal som är lika stort som eller större än observationens längd. Även i det här exemplet där LOPNR och BESOK knappast tar upp 132 tecken, skadar det inte att ange en rejält tilltagen postlängd.

Programmet i exempel 3 kan användas om man vill skriva ut samtliga variabler i S.MUSEER, utan att räkna upp namnen.

Exempel 3

Följande program skriver ut samtliga variabler i den fil som anges efter "IN=" i anropet nedan och skriver ut en textfil med det namn som anges efter "UT=". Kolumnerna som skrivs ut avgränsas med semikolon, som såväl Word som Excel uppfattar rätt. Det här sättet att skriva ut variabler kastar dock om deras inbördes ordning. Makrot kan lagras på nätet på så sätt att alla har ständig tillgång till det. Knepet att införa vektorer som innehåller samtliga variabler i infilen kan användas mer generellt än vad som görs i det här exemplet.

```
* ett makro som skriver ut en textfil;  
%macro putall(IN=,UT=,delim=);  
data _null_;  
  file &ut;  
  set &in;  
  * två vektorer skapas;  
  array chary(*) _character_;  
  array numy(*) _numeric_;  
  do i=1 to dim(numy);  
    put numy{i} "&delim" @;  
  end;  
  do i=1 to dim(chary);  
    put chary{i} "&delim" @;  
  end;  
  put; *raden avslutas;  
  drop i;  
run;  
%mend;
```

Anropet kan se ut så här:

```
%putall(IN=s.museer,UT="c:\dan\temp.txt")
```

Exempel 4.

Programmet nedan skriver ut värden på *varnamn* med två decimaler avskilda med komma-tecken.

```
proc print data=filnamn label;  
  format varnamn commax6.2;  
  var varnamn;  
run;
```

Exempel 5.

Programmet nedan läser in en fil av typen "lok-och-vagnar" och skapar en flat fil. M h a ett makro kan programmet göras om till ett i förväg obestämt antal variabler.

```
*-----*
| Inläsning av stansat material.
| Den inlästa filen kommer att bestå av:
| ID-kod          kod värde
| x              x   x
| x              x   x
| etc
| Samma ID-kod kommer att upprepas på flera rader.
*-----*
;

data temp;
  retain knr;
  infile cards missover delimiter='-';
  input knr $ kod $ @ ;
  do while ( kod^=' ');
    input varde @;
    output;
    input kod $ @;
  end;

datalines;
001-201-0000000000-201-83247-212-9472938
001-201-1235
002-201-4354715-212-57249359
003-201-715-228-23176
;
run;

*-----*
| Strukturen i TEMP ändras till:
|
| ID-kod          variabel1 variabel2 etc
| x              x       x       x
| x              x       x       x
| etc
| En ID-kod förekommer endast på en rad.
| Om en kod förekommer mer än en gång för ett och samma ID-kod,
| behålls endast den sista, utan felmeddelande.
*-----*
;

data utfil(drop=kod varde i)
;
  array y(1:5) y1-y5; * antal variabler hårdkodat;
  retain y1-y5;

  set temp; by knr;

  if first.knr then do;
    do i=1 to dim(y);
      y(i)=.;
    end;
  end;

  * avläsning av kod och tilldelning motsvarande variabel;
  if      kod='201' then y{1} =varde;
  else if kod='202' then y{2} =varde;
  else if kod='211' then y{3} =varde;
  else if kod='212' then y{4} =varde;
  else if kod='213' then y{5} =varde;
  else do;
    felkod=kod;
    felvarde=varde;
  end;
```

```

label
  y1 ='var201'
  y2 ='var202'
  y3 ='var211'
  y4 ='var212'
  y5 ='var213'
;

if last.knr then output utfil;
run;

```

Exempel 6.

Programmet nedan beräknar en totalskattning med tillhörande varians för en variabel Y för strategi 1 (se Tabell 2 på sid 9). Filen "gammalt" innehåller variablerna STRATUM och Y (intressevariabel) med en observation per individ, medan filen "strinfo" innehåller STRATUM och STORAN (N_h) med en observation per stratum.

```

proc sql;
  create table aggr as
    select stratum,
           uss(y) as kvadsumy,
           sum(y) as sumy,
           count(*) as antsvar
    from gammalt
    group by stratum
    order by stratum
;
  select aggr.stratum,
         sqrt(storan**2/ant svar*(1-ant svar/storan)/(ant svar-1)*
              (kvadsumy-sumy**2/ant svar)) as standav,
         storan/ant svar*sumy as total
    from aggr, strinfo
   where aggr.stratum=strinfo.stratum
;

```

Exempel 7.

Programmet nedan är ett SAS-program som på vanligt sätt skrivs i SAS programfönster och körs med "submit" (F8). Programmet kör TEMP1.PRC, som i sin tur är ett program skrivet i SUDAANs språk, lägger resultatet i TEMP2.OUT och hämtar in detta i programfönstret (inte i Outputfönstret). På det här sättet kan en användare av SUDAAN arbeta helt och hållet i SAS programfönster.

```

x 'c:\sudaan\sudproc
   c:\dan\templ.prc
   c:\dan\temp2.out -r
  ';

dm 'inc c:\dan\temp2.out';

```

Vill man skriva ett SUDAAN-program i SAS programfönster och köra det som om det vore ett SAS-program, genom att trycka på en tangent, kan följande skrivas in vid lämplig snabbvalstangent i SAS Keys-fönster (här nedan uppdelat på två rader av praktiska skäl, i Keys skrivs allt på en rad):

```

file 'c:\dan\templ.prc' replace;clear;
inc 'c:\dan\sudlib\sudexkv.sas'; submit;

```

där SUDEXKV.SAS är SAS-programmet ovan. För att ta tillbaks det senast körda programmet kan en annan tangent få följande text:

```

inc 'c:\dan\templ.prc';

```

Exempel 8

Programmet gör om de variabeletiketter som skapas i Proc Format till en textfil med det namn som SUDAAN kräver. Tillägget CNTLOUT skriver ut formatet som ett SAS-dataset. Filen LEVEL.DBS ska ligga i samma bibliotek som den SAS-fil som SUDAAN-programmet ska läsa. Detta bibliotek har här libref S604 som står för c:\dan\sasdlib.

```
proc format library=library cntlout=s604.stratf;
  value stratf
    1 = 'Stratum 1'
    2 = 'Stratum 2'
    3 = 'Totalundesökt stratum'
    other = 'felaktig stratumkod'
  ;
run;

data _null_;
  set s604.stratf;
  file 'c:\dan\sasdlib\LEVEL.DBS';
  sudvalue=substr(left(start),1,2);
  if fmtname='STRATF' then
    put 'STR ' @;
  else delete;
  put @9 sudvalue $1.
    @17 label $40.
  ;
run;
```

Exempel 9

Programmet gör om SAS-format till en variabel som här får namnet REDOGR.

```
proc format library=library;
  value $redogrf
    1 = 'Stratum 1'
    2 = 'Stratum 2'
    3-4 = 'Totalundersökta strata'
    other = 'felaktig stratumkod'
  ;
run;

data urval;
  set s604.urval;
  format stratum $redogrf. ;
  redogr=put(stratum,$redogrf.);
  label redogr = 'Redovisningsgrupp'
  ;
run;
```

Exempel 10

Följande program läser in en stordatorfil till en SAS-fil på PC. Stordatorfilen ska ha förts över till PC "binärt". Det innebär att filens 0:r och 1:r har behållits i ograverat skick. I exemplet ligger stordatorfilen på C:\DATA och heter REGZF. Den har fixt recordformat som är det vanliga inom SCB (det går bra att ha RECFM=F trots det SAS 1993 b, s 194 ff. säger). Det är viktigt att värdet för LRECL, postlängden, anges exakt. Variabeln REDNR i exemplet blir alfanumerisk. BRANSCH och TAL är lagrade som packade fält i stordatorfilen och blir "uppackade" och numeriska i SAS-filen på PC.

```
data prov;
  infile 'c:\data\regzf' recfm=f
        lrecl=1200 pad missover;
  input
    @1    rednr    $ebcdic8.
    @9    bransch  s370fpd2.0
    @946  tal      s370fpd4.0
    ;
run;
```

Bilaga 2. Förklaring av vissa tekniska termer.

1.1. Överföring av data mellan programvaror

Det är bara de allra enklaste programvarorna som bara har *en* möjlighet att föra in resp. ta ut data, vilket då väl alltid är textfiler utan metadata. Den äldsta metoden att föra över data är via filer: en programvara skapar en fil med data som kan läsas av ett annan programvara. Det kan vara enkla textfiler eller filer i något speciellt format som innehåller även metadata. Dessa senare kallas ofta exportfiler, "portables" e d. Nackdelar är:

- Tar tid, i synnerhet om filen måste anpassas i den mottagande programvaran.
- Manuellt arbete.
- Ej dynamiskt. Om de ursprungliga data ändras, måste användaren få vetskap om det och sedan utföra överföringen en gång till.
- Stora filer kan vara besvärliga att flytta på pga begränsningar i hårdvaran och nät.

Nedan tas olika typer av metoder för att föra över data mellan programvaror: textfiler, urklipp, DDE/OLE samt kommunikation med SQL-orienterade databaser. Trots nackdelarna är överföring via filer eller urklipp ofta utmärkt, i synnerhet som de ofta är de enda metoder som är praktiskt tillgängliga. Men visst vore det trevligt och effektivt att ha smidigare system. DDE är ett redskap i Windows för dynamisk överföring av data. OLE är ett annat Windowsredskap som täcker mer än bara datatransport, OLE suddar dessutom ut gränserna mellan programvarorna. SQL och SQL-orienterade databaser gör det lättare att hämta och foga samman data från olika filer i olika datalager.

1.2. Textfiler

"ASCII" och "ANSI"

På PC har vi två teckensystem. Datorn själv lagrar allt med nollor och ettor. Teckensystemet talar om hur t ex "ä" ska vara lagrat för att vara begripligt för datorn. De två som vi har på PC brukar i dagligt tal kallas "ASCII" och "ANSI". I en del menyer i Windows står det istället "endast text" (med vilket menas ANSI) eller "DOS-text" (ASCII). ASCII är en förkortning av *American Standard Code for Information Interchange* medan ANSI står för *American National Standards Institute*. Programvara anpassade för Windows brukar vara inställda på ANSI, medan andra PC-program vanligtvis är inställda på ASCII. Problem uppstår framför allt med å, ä, ö och heldragna linjer.

"Lok och vagnar"

På en del håll inom SCB används icke-normaliserade filstrukturer för sekventiella filer i stordatorn. En variant vanlig för data som de ser ut direkt efter registrering, är "lok och vagnar", dvs filer med utseende av typen identitet-variabelkod-värde-variabelkod-värde etc:

```
001-052-000000000-053-83247-054-9472938
001-052-1235
002-052-4354715-054-57249359
003-052-715-054-23176
```

Överstansade fält

Ett variabelformat som fortfarande är vanligt inom åtminstone ES-avdelningen är "överstansning" eller "zonat decimalt". För att undvika minustecken i data, är den sista siffran i ett negativt heltal bytt mot en bokstav, t ex skrivs -18 som 1J.

Tabseparerade tabeller

En viktig funktion hos programvaror för estimation och analys är att skapa tabeller som visas i ett resultatfönster etc (jfr Figur 4, sid 28). Dessa kan användaren vilja föra över till en ordbehandlare, kalkylprogram eller något sådant. Det är då en fördel om kolumnerna i tabellen skiljs åt med tecken som den mottagande programvaran känner igen. Kommatecken, semikolon och tabbar är vanliga sådana "listavgränsare" (*delimiters*). I Windows kan man ange vilken symbol som ska fungera som listavgränsare (under Kontrollpanel, Internationell). I Word kan man kopiera in tabellen via Urklipp, markera den och beställa "omvandla text till tabell". Det fungerar för tabeller som har lika många celler för varje rad. Andra alternativ kan vara att anpassa tabellen manuellt eller avstå från att använda Words tabellfunktion (och byta från proportionellt typsnitt till ett fast sådant, t ex Courier).

1.3. Urklipp

Windows Urklipp är ett utrymme för temporär datalagring. Urklipp kan lagra data med en mängd olika format. En speciell programvara kanske bara klarar av att lägga och läsa ett eller ett par av dem i Urklipp. Programvaror som körs under Windows stöder Urklipp, dvs "klippa och klistra". Urklipp har samma nackdelar som överföring via filer. Fördelen är att det ofta är lättare för användaren att peka ut vad som ska föras över.

1.4. DDE

Dynamic Data Exchange (DDE) är en samling regler för hur program som körs under Windows kan utbyta data med varandra. En klient och server kommunicerar genom att skicka meddelanden enligt de regler som fastställs i DDE-protokollet.

Klienten

- bestämmer vem den ska kommunicera med
- bestämmer ämne för kommunikationen
- startar kommunikationen

Servern är passiv och gör som klienten begär.

Länkarna mellan klienten och servern kan vara av två typer:

- Ej dynamisk, "kall" länk. Klienten skickar en fråga till servern, som svarar. Därefter avslutas kommunikationen.
- Dynamisk, "het" eller "varm" länk. Om data som servern har tillgång till förändras, kommer servern utan ytterligare uppmaning att skicka nya data till klienten (het länk) eller skicka en fråga till klienten av typen "vill du ha nya data?" (varm länk).

Data kan skickas i båda riktningarna. Data kan även bestå av kommandon som klienten uppmanar servern att utföra.

Bakom kommandot "klistra in special" döljer sig i många fall DDE. Det är alltså ett enkelt sätt för användaren att upprätta en DDE-länk.

1.5. OLE

Object Linking and Embedding (OLE) är en ganska ny vidareutveckling av DDE. Klienten får data från servern (den applikation som får data definieras som klient, den som sänder som server). Idén med OLE är att klienten inte bara sparar själva data utan även information om vem som skickade dessa data. Tillsammans utgör data och information om avsändaren ett objekt. OLE kan i version 1 vara endera "linking" eller "embedding". Med "object embedding" menas att objektet sparas inne i det program, dokument eller motsvarande som klienten behandlar. Objektet kan då endast förändras om klienten så begär. Med "Object

Linking" menas att objektet sparas någon annanstans än inne i klientens program e d. Objektet kan därför förändras automatiskt om motsvarande data i servern förändras. Ett exempel på "embedding" är formler i ett Worddokument i version 2, skrivna med Equation Editor. Formlerna kan bara uppdateras från Word som har information om att Equation Editor skapade dem. Equation Editor har däremot ingen information om vart formlerna tog vägen.

Med nästa version av OLE, med nummer 2, suddas gränserna ut mellan programvaror. Då kan ett stort antal typer av möjliga åtgärder föras över tillsammans med objektet, förutom bara "ändra". I praktiken kan ett arbetsmaterial, t ex ett dokument i Word, bestå av flera objekt som härrör från olika programvaror. Om användaren pekar på ett objekt som kommer från en annan programvara, t ex en bild eller en matematisk formel, förändras menyraden till att visa åtgärder som hör till denna programvara (rita in en pil, infoga integraltecken etc). På så sätt växlar programvarorna automatiskt så att vid varje tillfälle är den aktuella uppsättningen av åtgärder tillgänglig. Vill man lägga till OLE-kopplingar mellan programvaror som stöder version 2, är man inte bunden till något visst programmeringsspråk för att upprätta kopplingen. T ex kan Visual Basic användas.

1.6. Frågespråket SQL

Structured Query Language (SQL) är ett frågespråk fastlagt i ANSI-standard. Många programvaror följer denna, men har lagt till ytterligare funktioner. Även själva användargränssnittet varierar naturligtvis mellan programvarorna. SQL används främst till att hämta data ur ett datalager. Istället för att hämta hela filer extraheras just de bitar som användaren frågar efter, som kan ligga i olika filer. Även viss bearbetning kan göras, t ex summering. Datalagret kan vara en databas som ligger på någon central plats i nätverket eller det egna datamaterialet som lagras lokalt. Frågespråket SQL behöver alltså inte inskränka sig till kommunikation med SQL-orienterade databaser.

1.7. Kommunikation med SQL-orienterade databaser

Enligt centralt beslutad strategi blir SQL-orienterade databaser viktiga för SCBs statistikproduktion (ADB-rapport 1994:3).

Client/server

Client/server-arkitekturen har i sig inget med Windows att göra, det är ett sätt att organisera datalagret i ett system med olika typer av datorer. Följande två fördelar uppnås, som inte kan fås på något enkelt sätt annat sätt än med client/server (och stordator med terminaler):

- datalagret placeras på en centralt belägen dator som kan skyddas mot obehörigt intrång och kraschar;
- datalagret är ändå åtkomligt eftersom användaren bara får just de bitar han frågar efter, inte hela filer, vilket skulle leda till en oacceptabel belastning i datasystemet.

Dessutom uppnås följande:

- Kontakten mellan den programvara som hanterar datalagret (*back-end*) och den eller de programvaror som användaren beställer datauttag med (*front-ends*) följer en standard för hur kommunikationen mellan back-end och front-end ska se ut. Det gör att programvarorna är utbytbara.
- De PC som står på skrivborden avlastas eftersom en hel del bearbetningar kan göras av särskilt kraftfulla back-enddatorer.
- Frågespråket SQL kan användas för datauttag. Datauttag är idag ofta en mödosam och dyr sysselsättning, men kan med SQL och client/server skötas smidigare, eftersom det numera finns en stor mängd användarvänliga programvaror som har SQL inbyggt, och eftersom client/server tvingar fram en mer enhetlig struktur för datalager.

ODBC

Open Data Base Connectivity (ODBC) är Microsofts standard för gränssnittet mellan en databashanterare (*back end*) och de användarnära programvaror (*front ends*) som är lämpade för uttag ur databasen. En sådan programvara kan kommunicera med varje databas som följer ODBC. Data översätts till de format som gäller inne i databasen. Dessa format blir på så sätt den databasens interna bekymmer. Inte bara själva data översätts utan även frågor som programvaran riktar till databasen. Frågorna uttryckta på det sätt som är naturligt i den programvaran översätts till ett språk som följer den databasens interna regler. De typer av frågor som är möjliga att ställa mot en viss databas visas vanligen i ett dialogfönster i den användarnära programvaran. Det kan alltså finnas fler eller färre beroende på vilken ODBC-drivrutin som deltar i kommunikationen.

1.8. API

Application Program Interface (API) är den mängd av redskap som en programmerare har tillgång till när han konstruerar ett tillägg till en programvara. T ex kan man vilja lägga till ett redskap för tabellering i Word, så att en skribent enkelt kan utföra vissa beräkningar och skapa en tabell utan att behöva lämna det dokument där tabellen ska infogas. En annan programvara kan innehålla ett API med i princip funktioner av typen aggregera, tabellera, föra över data med DDE eller OLE etc. Med dessa skapar programmeraren en knapp eller ett menyval i Word som ger Wordanvändaren tillgång till lämpligt utformat tabelleringsverktyg.

1.9. Värdeförrådsbeskrivningar

Värdeförrådsbeskrivningar används för att värden på en kategorisk variabel ska ersättas med text vid utskrift. En del programvaror använder värdeförrådsbeskrivningar även då materialet ska klassindelas, t ex att för en viss tillämpning ska alla observationer i datamaterialet med SNI-koder 10 till 37 aggregeras till grupp som benämnes "Utvinning av material och tillverkning". Fördelarna med att lagra dessa metadata separat från själva data är:

- modifieringar görs på endast ett ställe;
- kan efter behov kopplas till eller bort från en eller flera datafiler;
- alla behöver inte lägga upp egna beskrivningar;
- medför standardisering av texten.

Referenser

ADB-enheten (1994:1). Användargrupper och stödnivåer för programvaror inom SCB. ADB-rapport 1994:1. SCB.

ADB-enheten (1994:3). Metodstrategi för ADB-baserad produktion och datalagring. ADB-rapport 1994:3. SCB.

ADB-enheten (1994:5). RTB-SQL/2. Stora register i SQL-databaser - projektrapport. ADB-rapport 1994:5. SCB.

Andersson, Claes; Nordberg Lennart (1992). CLAN - ett SAS-program för skattningar av medelfel. R&D Report 1992:11. SCB.

Bethlehem, Jelke G.; Keller, Wouter J. (1987). Linear weighting of sample survey data. Journal of Official Statistics. Vol 3, No 2, pp. 141-153.

Bethlehem, Jelke G.; Hundepool, Anco J. (1992). Integrated survey processing on Micro-computers. Netherlands Central Bureau of Statistics. I: Westlake, A. m fl (red.), Survey and Statistical Computing. Proceedings of an International Conference Organised by the SGCSA, Bristol, UK, 1992. Elsevier Science Publishers B.V. North-Holland, pp 405-411.

Björkman, Anna (1994). Tabuleringsprogram på PC. ADB-rapport 1994:6. SCB.

Brännström, Tomas; Skalin, Joakim (1994). SPSS för Windows - en handledning. Andra upplagan.. Handelshögskolan i Stockholm, institutionen för ekonomisk statistik.

Bäcklund, Sten. (1994). SAS till SPSS och omvänt - dataöverföring. ADB-enheten, SCB.

Carlson, Barbara L; Johnson, Ayah E.; Cohen Steven B. (1993). An Evaluation of the Use of Personal Computers for Variance Estimation with Complex Survey Data. Journal of Official Statistics. Vol 9, No 4, pp 795-814.

Estevao, V.; Hidirolou, M.A.; Särndal, C.-E. (1992). Requirements for a Generalized Estimation System at Statistics Canada. Proceedings of the Work Shop on Uses of Auxiliary Information in Surveys. Örebro, Sweden.

Frigge, M.; Hoaglin, D. C.; Iglewicz, B (1989). Some implementations of the Boxplot. The American Statistician, Vol 43, No 1, 1989.

Gilbertsson, Berndt (1994). Extract SAS. Programdokumentation version 2.1. 1994-02-17. SCB.

Hallahan, C. (1992). DBMS/COPY an DBMS/COPY Plus (version 2.0). The American Statistician, vol. 46, No. 1, pp. 49-52.

Hedlin, Dan (1991). Om att bygga menystyrda program i SAS. ADB-Rapport nr 1991:18. 91-09-19. SCB.

Hedlin, Dan (1993). PC-programvaror för variansskattningar vid tvåstegsurval. 93-07-01. SCB.

Hedlin, Dan (1995). Erfarenheter och tillämpningar av SAS/INSIGHT. Utkast. SCB.

Jöreskog, K, Sörbom, D (1993). LISREL 8: Structural Equation Modelling with the SIMPLIS Command Language. SSI Scientific Software International.

Koch, Armin; Haag, Uwe (1994). The Statistical Software Guide '93/94. Computational Statistics & Data Analysis. Vol. 18, No. 2, Sept 1994.

Lundström, Sixten (1993). Metoder att få konsistens mellan olika skattningar och samtidigt öka precisionen. R&D Report 1993:6. SCB.

Lundström, Sixten (1994). Översyn av estimationen i undersökningen *Kortperiodisk sysselsättningsstatistik för privat sektor*. 94-01-14. SCB.

Lyberg, Ingrid (1989). TA84/85 - Variansberäkningar med SUDAAN för variabeln INK, 89-11-17. SCB.

PC Magazine (1993). Stat of the Art. May 11.

Pettersson, Hans; Selén Jan (1990). Variansberäkningar på PC. En undersökning av två PC-program för analys av surveydata. Juni 90. SCB, I/Ledn.

Qvist, Jan, m fl (1991). Livsförloppsanalys. Bakgrundsmaterial från demografiska funktionen. 1991:2. SCB.

Rosén, Bengt. (1990). Teori för urvalsundersökningar. Statistiska Centralbyrån. Litteratur i kursen i urvalsteori vid Stockholms universitet, institutionen för matematisk statistik och försäkringsmatematik.

Rowe, Beverley C. (1992). The SGCSA Software Catalogue. I: Westlake, A. m fl (red.), Survey and Statistical Computing. Proceedings of an International Conference Organised by the SGCSA, Bristol, UK, 1992. Elsevier Science Publishers B.V. North-Holland, pp 395-403.

Rylett, D. T., Bellhouse, D. R. (1988). TREES: a Computer program for complex surveys. Proceedings of the Section on Survey Research Methods, American Statistical Association, pp. 694-697.

SAS (1989 a). SAS/STAT User's Guide, vol. 1. Version 6. Fourth Edition. Cary, NC: SAS Institute, Inc.

SAS (1989 b). SAS/STAT User's Guide, vol. 2. Version 6. Fourth Edition. Cary, NC: SAS Institute, Inc.

SAS (1989 c). SAS Guide to the SQL Procedure. Version 6. First Edition. Cary, NC: SAS Institute, Inc.

SAS (1990 a). SAS Language Reference. Version 6. Second Edition. Cary, NC: SAS Institute, Inc.

SAS (1990 b). SAS Guide to macro Processing. Version 6. Second Edition. Cary, NC: SAS Institute, Inc.

SAS (1992). SAS Technical Report P-229. SAS/STAT Software. Changes and Enhancements. Release 6.07. Cary, NC: SAS Institute, Inc.

SAS (1993 a). SAS/INSIGHT User's Guide. Version 6. Second Edition. Cary, NC: SAS Institute, Inc.

SAS (1993 b). SAS Companion for the Microsoft Windows Environment. Version 6, First Edition. Cary, NC: SAS Institute, Inc.

SAS (1994 a). SAS Technical Report P-262, SAS/ACCESS Interface to ODBC: SQL Pass-Through Facility. Release 6.08. Cary, NC: SAS Institute, Inc.

SAS (1994 b). SAS ODBC Driver Technical Report. User's Guide and Programmers Reference. Release 6.10. Cary, NC: SAS Institute, Inc.

Sautory, Olivier (1992). Calibration on Known Marginal Counts for Sample Surveys: Practical Experiences at INSEE. Proceedings of the Work Shop on Uses of Auxiliary Information in Surveys. Örebro, Sweden.

Sautory, Olivier (1993). La Macro CALMAR - Redressement d'un échantillon par calage sur marges. Institut National de la Statistique et des Études Économiques.

Sawitzki, Günther (1994). Report on the Numerical Reliability of Data Analysis Systems. Computational Statistics & Data Analysis. Vol. 18, No. 2, Sept 1994.

Shah, Babubhai V.; LaVange, Lisa M. (1991). Software for Generalized Linear Model Analysis of Survey Data. Presented at the International Statistical Institute Meetings in Cairo.

Shah, Babubhai V., m. fl. (1993). Statistical Methods and Mathematical Algorithms used in SUDAAN. DRAFT. Research Triangle Institute. March 1993.

Shah, Barnwell, Hunt and LaVange (1992). SUDAAN User's Manual, Release 6.0. Research Triangle Institute.

SPSS (1991). SPSS Statistical Algorithms. 2nd Ed. Chicago: SPSS Inc.

Statistics Canada (1993). User's Guide GES V2.2. GES Development Team, System Development Division, Statistics Canada.

Särndal Carl-Erik; Swensson, Bengt; Wretman, Jan. (1992). Model Assisted Survey Sampling. Springer-Verlag New York Inc.

Wills, Peter (1992). Data Use and Reuse. Mercator Computer Systems Ltd. I: Westlake, A. m fl (red.), Survey and Statistical Computing. Proceedings of an International Conference Organised by the SGCSA, Bristol, UK, 1992. Elsevier Science Publishers B.V. North-Holland, pp 297-304.

Wolter, Kirk M. (1985). Introduction to Variance Estimation. Springer-Verlag New York Inc.

R & D Reports är en för U/ADB och U/STM gemensam publikationsserie, som fr o m 1988-01-01 ersätter de tidigare "gula" och "gröna" serierna.

R & D Reports Statistics Sweden are published by the Department of Research & Development within Statistics Sweden. Reports dealing with statistical methods have green (grön) covers. Reports dealing with EDP methods have yellow (gul) covers.

Reports published 1994-95:

- | | |
|------------------|--|
| 1994:1
(gul) | Ett verksgemensamt datalager vid SCB?
(<i>Bo Sundgren</i>) |
| 1994:2
(grön) | Undersökningar med flera paneler av stratifierade nätverksurval - Estimationsförfarande vid undertäckande paneler och överlappande urvalsstratan. Med tillämpning på HINK92-undersökningen
(<i>Bengt Rosén, Håkan Lindén</i>) |
| 1994:3
(grön) | Bortfallsbarometern nr 9
(<i>Mats Bergdahl, Pär Brundell, Folke Carlsson, Dan Hedlin, Håkan Lindén, Monica Rennermalm</i>) |
| 1994:4
(grön) | Kvalitetsrapporten 1994 - Utveckling av kvaliteten för SCBs statistikproduktion |
| 1995:1
(grön) | Asymptotic Theory for Order Sampling
(<i>Bengt Rosén</i>) |

Tidigare utgivna **R & D Reports** kan beställas genom Ingvar Andersson, SCB, U/LEDN, 115 81 STOCKHOLM (telefon: 08-783 41 47, telefax: 08-783 45 99, e-mail: ingvar.andersson@scb.se)

R & D Reports listed above as well as issues from 1988-93 can - in case they are still in stock - be ordered from Statistics Sweden, att. Ingvar Andersson U/LEDN, S-115 81 STOCKHOLM SWEDEN (telephone: +46 08 783 41 47, telefax: +46 08 783 45 99, e-mail: ingvar.andersson@scb.se)