

PROMEMORIOR FRÅN P / STM

NR 5

UTREDNING ANGÅENDE STATISTISK
ANALYSVERKSAMHET VID SCB:
SLUTRAPPORT

INLEDNING

TILL

Promemorior från P/STM / Statistiska centralbyrån. – Stockholm : Statistiska centralbyrån, 1978-1986. – Nr 1-24.

Efterföljare:

Promemorior från U/STM / Statistiska centralbyrån. – Stockholm : Statistiska centralbyrån, 1986. – Nr 25-28.

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

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

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

Promemorior från P/STM 1980:5. Utredning angående statistisk analysverksamhet vid SCB : slutrapport.
Digitaliserad av Statistiska centralbyrån (SCB) 2016.

PROMEMORIOR FRÅN P / STM

NR 5

**UTREDNING ANGÅENDE STATISTISK
ANALYSVERKSAMHET VID SCB:
SLUTRAPPORT**

P/STM, Analysprojektet

...

Utredning angående statistisk analysverksamhet vid SCB:
Slutrapport

...

INNEHÅLLSFÖRTECKNING

- 1 Inledning
- 2 Policyfrågor: Bör statistikproducenter utföra statistisk analys?
- 3 Redogörelse för projektarbetet
 - 3.1 Bakgrund
 - 3.2 Organisation, arbetsformer
 - 3.3 En kartläggning av analysbehovet vid SCB
 - 3.4 Empiriska analysstudier
 - 3.5 Metodutveckling: Tillämpad forskning i anslutning till analysverksamheten vid SCB
 - 3.6 Implementering av vissa analysmetoder vid SCB
 - 3.7 Metodutvecklingsarbete i anslutning till SCBs programbibliotek
- 4 Projektgruppens utvärdering
- 5 Förslag till fortsatt verksamhet
 - 5.1 Löpande verksamhet
 - 5.2 Satellitprojekt

PROJEKTDOKUMENTATION

ÖVRIGA REFERENSER

BILAGOR

P/STM

1980-10-31

1 Inledning

Vid enheten för statistiska metoder har sedan några år tillbaka bedrivits försöksverksamhet i projektform med det huvudsakliga syftet att på olika sätt verka för ökad användning av statistisk analys vid SCB. Projektverksamheten har omfattat såväl teoretiska studier som empiriska fallstudier knutna till någon speciell undersökning eller statistikgren. Informationsverksamhet om statistiska analysmetoder har bedrivits bland annat via metodseminarier. Projektgruppen har bestått av handläggare från ett antal olika enheter vid verket. Till gruppen har även varit knuten en referensgrupp med representanter från ämnesavdelningarna.

Projektet har avslutats i och med utgången av budgetåret 1979/80 och arbetet redovisas i föreliggande slutrapport. Avsnitt 2 behandlar policyfrågor i samband med statistisk analys vid statistikproducerande myndigheter. Avsnitt 3 innehåller en redogörelse för projektarbetet, medan avsnitt 4 redovisar synpunkter från projektgruppen på innehåll, arbetsformer samt på de förutsättningar för arbetet som förelegat. I avsnitt 5 slutligen, lämnas ett förslag till fortsatt analysverksamhet vid SCB.

Vid utarbetandet av slutrapporten har följande personer deltagit:

Claes Cassel
Per Gunnar Cassel
Catarina Elffors
Gunnar Gelin
Margareta Holgersson
Åke Holmén
Harry Lütjohann
Anders Norberg
Bo Rennermalm
Bengt Ringnér
Jan Wretman

Planering, samordning och redigering av rapporten har utförts av Margareta Holgersson och Jan Wretman.

1980-10-31

2 Policyfrågor: Bör statistikproducenter utföra statistisk analys?

Frågan om statistikproducenter bör bedriva konsumentorienterad analys av den offentliga statistikproduktionen har under 1970-talet aktualiserats både inom och utom landet. Detta tema var exempelvis föremål för diskussion vid en av sessionerna vid ISI-mötet i Warszawa 1975, se Norwood (1975), Boreham (1975) och Balint (1975). Även vid flera CES-möten¹ har frågan diskuterats, se exempelvis Duncan (1978). Frågan diskuterades även vid SCBs forskarkonferens 1980.

Vid dessa diskussioner har man i allmänhet varit ense om att statistikproducenterna bör bedriva s k producentorienterad analys, "analys av statistiken". Här ingår kvalitetsanalyser, medelfelsberäkningar, bortfallskorrigeringar etc. Syftet med denna typ av analyser är att avslöja begränsningar i det insamlade materialet. Avsikten är både att "varna" statistik-användarna och att ge underlag för förbättring av statistiken. Beträffande den s k konsumentorienterade analysen, "analys av verkligheten" har dock meningarna i allmänhet varit delade. Från statistikproducenterna har man alltmer börjat känna en ökad efterfrågan från de statistikkonsumerande myndigheternas och organisationernas sida avseende bättre beslutsunderlag i form av mer "vidareförädlad" statistik. Man är från användarnas sida exempelvis intresserad av översikter och sammanställningar av data från olika källor samt av "konsekvensanalyser", dvs prognoser för följderna av olika tänkbara samhällsliga beslut. Inte minst i ett ansträngt budgetläge kommer krav att resas på högre utnyttjande av ett kanske reducerat antal undersökningar. De moment i statistikprocessen som omfattar insamling, kodning, stansning och granskning är dyra medan analyser av redan insamlat material är förhållandevis billiga. Dock upplevs "neutralitetsproblemen" som starka på många håll. Vid fördjupad bearbetning och analys löper statistikproducenten risk att bli anklagad för att ta politisk ställning. De skador sådana anklagelser och misstankar kan vålla en statistikproducerande myndighet uppvägs enligt vissas mening inte av de fördelar som kan medföra en ökad analysverksamhet.

Vad är då SCBs officiella ståndpunkt i denna fråga? Med åren har man kunnat registrera en alltmer positiv inställning till en ökad grad av bearbetning och "vidareförädling" av statistikprodukterna från verkets sida. I samband med omorganisationen av SCB i början av 1970-talet utgavs 1972 en handbok: "Verksamheten vid SCB - nuläge och utvecklingslinjer". I kapitlet Konsumentbehov och slutprodukter, avsnittet Statistikprodukternas egenskaper (2.3) kan man bl a läsa följande:

"Innehållet (i den statistik som produceras av SCB) har också en annan dimension i bearbetningsgraden. ADB-teknikens utveckling har här medverkat till nya variationsmöjligheter. Praktiska förutsättningar har skapats att maskinellt utföra mer omfattande bearbetningar för fördjupad analys. Ett högre utnyttjande av ADB kan beräknas leda till att SCB konfronteras med nya analysproblem, för att kunna möta efterfrågan på en längre driven bearbetningsgrad än som tidigare varit vanligt.

¹ CES = Conference of European Statisticians

1980-10-31

... Gränsen för hur långt analysen skall gå måste bedömas från fall till fall. Det fordras speciella skäl för att en mer djupgående analys skall göras inom SCB. Ett ämnesområde kanske saknar särskilda forskningsinstitutioner eller konsumenter med analytisk expertis och statistiken blir därför inte effektivt använd. Sådana ämnesområden kan inom SCB tas upp till analys, och då helst i samarbete med universitet och högskolor eller andra forskningsinstitutioner.

En utveckling av analysverksamheten kan emellertid i och för sig vara gynnsam för SCBs statistikproduktion och förbättra kvaliteten på slutresultaten genom att ge underlag för förbättrad produktutveckling och kvalitetskontroll. Av speciell betydelse är här den analys som är inriktad på produkternas kvalitet och den statistiska produktionsprocessens utformning. Denna kan innehålla en rad element. Den kan t ex bestå i variansberäkningar, i jämförelser som belyser aktualiteten i ett register eller i konsistensprövning av resultat genom studium av tidsserier eller genom tvärsnittsanalys."

Andan i ovanstående citat kan med viss ansträngning betecknas som positiv, om än blandad med en tydlig försiktighet. Man talar om "gränser för hur långt analysen skall gå", vilka "måste bedömas från fall till fall" och om att "speciella skäl" erfordras för att "mer djupgående analys skall göras inom SCB".

Åtta år senare, i riktlinjerna för 1980 års verksamhetsplanering uttrycks emellertid verkets policy för ökad analysverksamhet med större eftertryck (SCB-meddelande 1980-01-31):

"Statistiken behöver vidare utvecklas så att den i högre grad än nu kan användas som underlag för utvärdering, omprövning och rationalisering av olika aktiviteter i samhället. Den behöver bli mer problemorienterad och åtgärdsinriktad och det bör finnas en beredskap att snabbt kunna belysa olika problem. Detta innebär bl a att vi skall sträva efter att ge mera av helhetsbilder över större områden. Statistiken skall vidare i ökad utsträckning belysa samband, resultat och effekter. Analys, tidsserier, longitudinella studier och prognoser bör utgöra ett ökat inslag i verksamheten. Aktualiteten behöver förbättras."

Detta betonas ytterligare i Handbok för verksamhetsplaneringen (1980-01-21):

"Ett väsentligt inslag i strävandena att kunna tillgodose de mest angelägna statistikbehoven med begränsade resurser är att noga penetrera och ta till vara möjligheterna att utnyttja befintlig statistik och befintligt grundmaterial bättre. Det kan gälla kombinationer av olika material vid SCB och externa administrativa material, nya och anpassade former för presentation och spridning av resultat, mer inslag av analys osv."

Vid omprioriteringar av verksamheten vid SCB bör man enligt riktlinjerna verka för att resurser görs tillgängliga för ökad analysverksamhet:

P/STM

1980-10-31

...
...
"Analysverksamhet. Som angivits i avsnitt 2.2 bör inslagen av analys av SCBs statistiska information öka i vår verksamhet. En utveckling mot ökad analysverksamhet måste ske successivt och i takt med att resurser görs tillgängliga och kompetens utvecklas.

...
Samverkan med forskningen kan ge upphov till och stimulera en utvidgad analysverksamhet. Kontakter med forskningen kan t ex ske genom att forskningen representeras i SCBs konsumentgrupper och genom konferenser eller seminarier som riktar sig till forskningen. Den kan också ske genom vetenskapliga rådets verksamhet eller via forskarservicegruppen (FS-gruppen)."

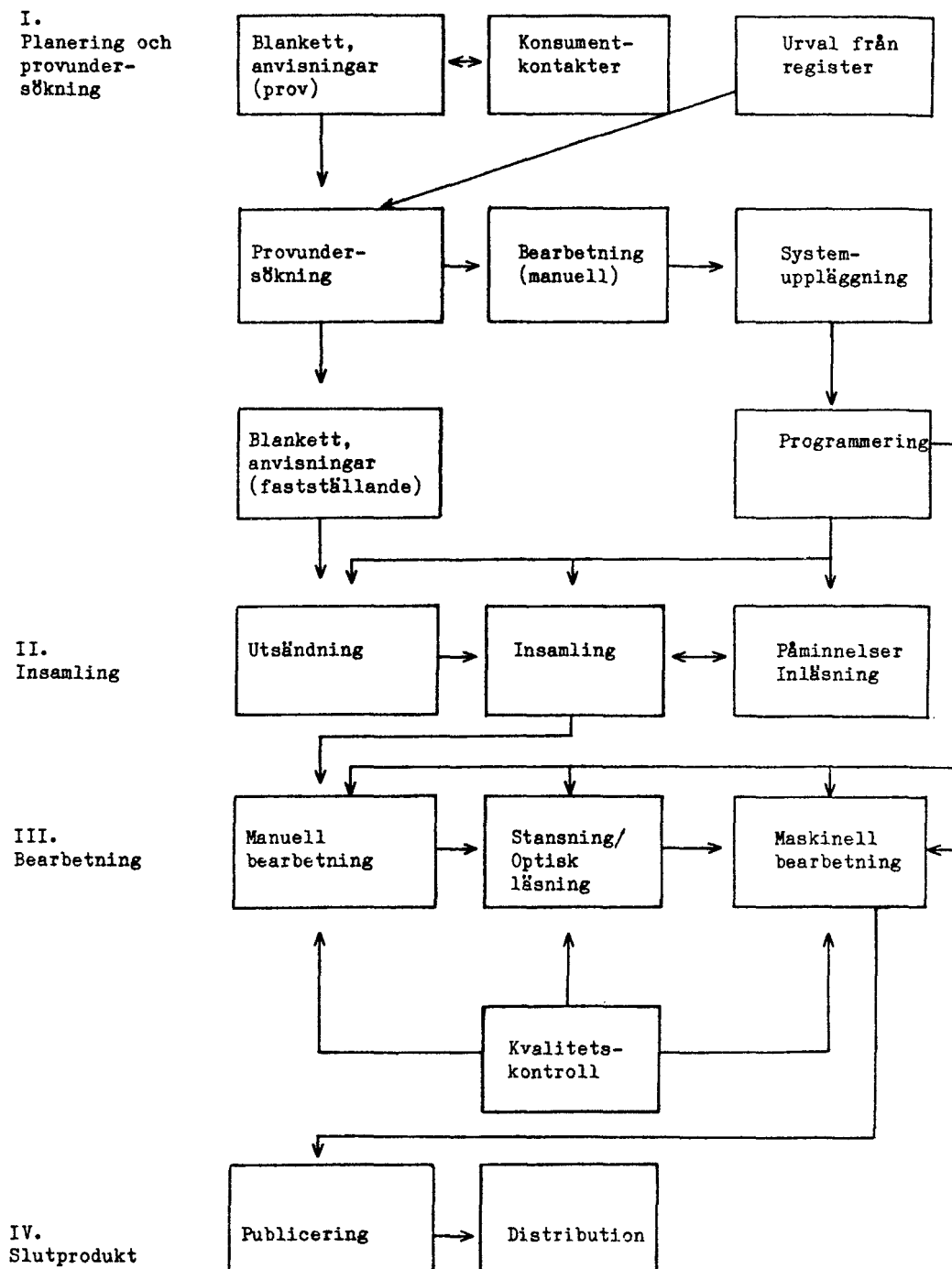
Slutligen betonas vikten av vidareutbildning av SCBs personal på analysområdet:

"Vidare ställer ett fördjupat arbete med behovsanalys och användaranpassning inkl analysinsatser delvis nya krav på personalen. Den måste få bättre kunskap om och egen erfarenhet av de sakområden som statistiken belyser resp utnyttjas inom. Detta gäller särskilt den personal som arbetar med konsumentkontakter och innehållsinriktad planering vid ämnesavdelningarna, liksom den personal som kan komma att arbeta med analys av olika slag. Det är ytterst angeläget med konkreta insatser på detta område."

Trots denna bevisligen positiva inställning till ökad analysverksamhet, en inställning som dessutom accentuerats på senare år kan vi konstatera att "förädlingsgraden" hos de allra flesta av SCBs produkter än så länge är tämligen låg. Med ett par undantag ligger fortfarande tyngdpunkten vid presentationen av SCBs statistik på redovisning av undersökningsmetod samt "råtabeller". På de allra senaste åren har tillkommit inslag av kvalitetsredovisningar, "analys av statistiken". En av orsakerna till att graden av vidareförädling inte har ökat i takt med att policyn förändrats är troligen att man inte följt upp policyförändringen genom att i statistikprocessen planera in analysmoment med samma självklarhet som man planerar in momenten kodning, stansning och granskning etc. Detta illustreras på ett slående sätt i broschyren "Något om SCB", utgiven av informationsenheten 1978, där den statistiska produktionsprocessen vid SCB illustreras i nedanstående flödesdiagram. Var finns den statistiska analysen? Det finns en annan sak som också på ett slående sätt framgår av diagrammet, nämligen att man utelämnat all statistisk metodplanering avseende urval och estimation m m.

DEN STATISTISKA PRODUKTIONSPROCESSEN VID SCB

Översikt



P/STM

1980-10-31

3 Redogörelse för projektarbetet

3.1 Bakgrund

I anslutning till den senaste omorganiseringen av SCB i början av 70-talet tillsattes 1972 en arbetsgrupp, kallad SAMSAS¹, med uppgift att utreda frågor sammanhängande med SCBs metodarbete för statistisk analys. Bland annat skulle arbetsgruppen beskriva forsknings- och serviceuppgifter på det statistiska analysområdet och utarbeta förslag till arbetsfördelning och samordning mellan vissa berörda enheter vid verket. Resultaten redovisades i en slutrapport daterad 1973-05-02.

Mer i detalj var gruppens arbetsuppgifter att:

- definiera begreppet statistisk analys
- kartlägga behovet av analyserviceverksamhet
- ange och beskriva arbetsuppgifter inom analyservice
- utarbeta förslag till arbetsfördelning och arbetssamordning mellan vissa enheter vid SCB för analyserviceverksamheten
- kartlägga behovet av metodutveckling för statistisk analys
- utarbeta förslag till hur samordnings- och prioriteringsfrågor vad gäller metodarbete skall behandlas i framtiden
- utarbeta förslag till formerna för uppläggning av ett programbibliotek för SCBs statistiska analysverksamhet.

Med utgångspunkt från SAMSAS-utredningens resultat inrättades två utvecklingsprojekt

1 Uppbyggnad av programbibliotek (knutet till enheten för databehandlingsmetoder, P/DBM)

2 Utredning angående statistisk analysverksamhet vid SCB (knutet till enheten för statistiska metoder, P/STM), i fortsättningen kallat Analysprojektet.

Det förstnämnda projektet har numera övergått i löpande verksamhet med namnet "Vård av programbibliotek" och är knutet till systemenheten, S/SYS.

Starttidpunkt för analysprojektet var 1974-07-01. I de genom GD-beslut fastställda riktlinjerna med samma datum (se bilaga 1) sammanfattades projektets syfte på följande sätt:

"Syftet med projektet är att skapa en organisation för att få ett samordnande grepp över metodutvecklingen inom det statistiska analysområdet och att bemästra samarbetsproblemen inom nämnda område vid SCB. Tyngdpunkten i projektarbetet skall

¹ SAMSAS = Styrgruppen och Arbetsgruppen för Metodfrågor vid Statistisk Analysverksamhet vid SCB

P/STM

1980-10-31

ligga på den statistiska metodsidan. Projektet skall dock vara förankrat i ämnesenheterens problemområden (huvudsakligen den löpande statistikproduktionen). Avsikten är att projektarbetet genom metodutveckling, lansering av metoder och metods-service skall aktivt medverka till att SCBs statistik i ökad utsträckning förädlas (analyseras)."

De ekonomiska resurser som ställts till projektgruppens förfogande har i genomsnitt över åren motsvarat ungefär 1 heltidsanställd person per budgetår.

3.2 Organisation, arbetsformer

I enlighet med SAMSAS-utredningens förslag tillsattes en projektgrupp med medlemmar från vardera enheterna P/STM, P/DBM, I/PI, I/UI (dåvarande U/UI) samt S/SYS (dåvarande D/PC). Till projektarbetet har vidare i samband med de fallstudier som genomförts tillfälligt knutits ett antal representanter från ämnesavdelningarna. Produktansvarig enhet har varit P/STM.

Vid sidan av projektgruppen inrättades dessutom en referensgrupp för att tillgodose behovet av förankring av projektet i ämnesavdelningarna. Referensgruppen har bestått av cheferna för P/STM (ordförande) och P/DBM samt en representant från vardera A, I, F och S/UC. Referensgruppen har erhållit löpande information om arbetet i projektet cirka en gång per halvår. I bilaga 2 listas namnen på de personer som vid någon tidpunkt ingått som medlemmar i projekt- respektive referensgrupp.

I riktlinjerna för projektarbetet föreslogs att den av SAMSAS föreslagna projektverksamheten skulle bedrivas enligt s k totalprojektmodell. En närmare beskrivning av innebörden av detta begrepp återfinns i Olofsson (1974), se speciellt Delrapport III, 1973-11-15.

3.3 En kartläggning av analysbehoven vid SCB

Av riktlinjerna för analysprojektet framgår att det till projektgruppens arbetsuppgifter har hört att kartlägga behovet av analys-service vid SCB:

"När det gäller såväl uppdrags- som anslagsfinansierad analys-service skall projektgruppen i samverkan med ämnesavdelningarna och S/UC ta fram och å jour-hålla en behovskatalog vad avser statistiska metoder och programvara inom området för analys-service."

Våren 1976 inleddes därför en serie intervjuer med handläggare från olika ämnesenheter. Syftet med intervjuerna var att försöka kartlägga behov och önskemål avseende såväl statistisk analys som grafisk presentation av SCBs statistik. Förslag på lämpliga intervjupersoner lämnades av medlemmarna i referensgruppen.

Vid intervjuerna diskuterade man kring följande fem frågor:

1 a) Använder Du statistisk analys och/eller grafisk presentation i Dina projekt?

P/STM

1980-10-31

b) Planerar Du att använda statistiska analysmetoder och/eller grafisk presentation?

c) Tror Du att Du själv eller någon annan på Din enhet skulle kunna ha behov av statistisk analys och/eller grafisk presentation inom något projekt?

2 Behöver Du hjälp med att utföra analys?

3 Var ligger problemen vid statistisk analys av SCBs data?

- definition av population, brist på urvalsram
- bortfall
- mätfel
- kvalitet
- urvalsdesign
- service
- övrigt

4 Saknar Du programvara för analys och grafisk presentation?

5 Upplever Du något speciellt behov av metodutveckling inom områdena statistisk analys och grafisk presentation?

Resultatet av denna kartläggning redovisas i Holgersson et al (1977).

Det visade sig vid intervjuerna att begreppet statistisk analys gavs tämligen olika tolkningar av de olika intervjupersonerna. I SAMSAS-rapporten definierades statistisk analys som "studiet av samband mellan variabler med hjälp av statistiska tekniker". Intervjupersonerna tolkade emellertid begreppet analys i en betydligt vidare mening.

Genom intervjuundersökningen erhöll projektgruppen flera konkreta förslag till fallstudier. Önskemål framförda i samband med intervjuerna bidrog även till ökade ansträngningar från projektgruppens sida att förbättra analys servicen på SCBs dator (se vidare avsnitt 3.7).

Några ytterligare, mer övergripande inventeringar av analysbehovet vid SCB har ej genomförts inom projektets ram. Sådan verksamhet har emellertid pågått kontinuerligt genom informella kontakter mellan projektgruppens medlemmar och handläggare vid ämnesenheter. Det finns dock skäl som talar för att systematiska kartläggningar bör genomföras med jämna mellanrum inom den föreslagna löpande analysverksamheten (se avsnitt 5).

Ämnesspecifika kartläggningar förekommer emellertid vid verket. Vid A-avdelningen tillsattes hösten 1979 en arbetsgrupp för att kartlägga områden där den lantbruksstatistiska analysverksamheten behöver vidareutvecklas. Gruppen har inriktat sig på den konsumentorienterade analysen, dvs den som går ut på att med olika metoder få ut mera information av statistiken eller primärmaterialet. Gruppens arbete är en uppföljning av den översyn och genomgång av den areella statistiken som tidigare genomförts vid A-avdelningen, se Medin (1975), Widén (1975), Esberger (1975 a, b) samt Jorner (1980a). Ett utkast till slutrapport föreligger, se Jorner (1980b).

3.4 Empiriska analysstudier

En rad konkreta analysstudier baserade på data från SCBs produktion har genomförts inom projektets ram. Denna verksamhet har bedrivits i samarbete med handläggare vid ämnesenheterna eller med externa uppdragsgivare. På så sätt har projektgruppen bidragit till att, som föreskrivs i riktlinjerna:

"verka för ökad användning av analys genom att i samråd med ämnesavdelningarna ta initiativ till och i form av pilotprojekt genomföra analysarbeten i anslutning till SCBs reguljära statistikproduktion."

Förutom att ha tjänat som hjälpmedel för lansering av statistisk analys - studierna har vid ett flertal tillfällen tjänat som undervisningsmaterial vid interna SCB-kurser och seminarier (se exempelvis Holgersson, 1979c) - har de dessutom givit uppslag till metodstudier (se nästa avsnitt).

Valet av undersökning/statistikgren för de empiriska studierna inom projektet har inte skett enligt någon systematisk, i förväg uppgjord plan. En nödvändig förutsättning har istället varit att vi haft möjlighet att etablera ett samarbete med en ämneskunnig person (inom verket eller externt) med intresse för statistisk analys samt vissa, nödvändiga teoretiska bakgrundskunskaper.

Ett problem har varit att bestämma i vilken form analyserna skall presenteras. Här är det svårt att göra generella rekommendationer. Några alternativ som vi provat är att förse SOS-publikationer eller SM med bilagor innehållande statistisk analys till vilka det hänvisas i huvudtexten. Ett annat alternativ har varit att publicera analyserna separat i Statistisk Tidskrift, eller i andra publikationsserier.

I följande delavsnitt ges kortfattade beskrivningar av olika fallstudier inom projektet.

3.4.1 Analys av bostadsområden med hjälp av faktor- och klusteranalys

Under våren 1974 framtog SCB på uppdrag av boendeutredningen (BOU) uppgifter rörande olika boende-, bostads- och inkomstförhållanden på viss delområdesnivå i Stor-Göteborg och Norrköping. Uppgifterna hämtades från SCBs specialprogram för kommuner grundat på FoB 70. BOUs syfte härvidlag var att studera den s k bostadssegregationen.

Den beskrivning av boendeförhållanden osv som SCB levererade var av "konventionellt" slag dvs den bestod av en listning, delområde för delområde, av frekvenser och medelvärden (procenttal) för studerade variabler rörande boendeförhållanden osv.

Under arbetet med uppdraget från BOU diskuterades frågan om andra mera avancerade metoder för beskrivning vore givande för den typ av data som nämnts. På grund av ändrade tidsplaner för BOU kunde detta inte närmare studeras inom uppdraget från BOU. Istället utfördes detta arbete inom ramen för analysprojektet. Två metoder prövades på det material som togs fram för BOU,

1980-10-31

nämligen faktor- och klusteranalys. Resultaten av analyserna redovisas i Hedström och Larsson (1975).

För att man skulle få synpunkter på tillämpbarheten av faktor- och klusteranalyser för beskrivningar av exempelvis boendeförhållanden, skickades analysresultaten ut till ett antal personer; både metodexperter och eventuella konsumenter.

Härvid framfördes från flera håll åsikten att när analyser görs för exempelvis kommuners räkning måste man vid val av analysmetod ta hänsyn till att den skall vara lätt att förklara för bl a politiker som saknar statistisk utbildning. Klusteranalys har här visat sig betydligt lättare att förklara än faktoranalys. Vid faktoranalys har det varit svårt att på ett enkelt sätt förklara innebörden av en faktor.

Från olika kommuner och från bostadsdepartementet lämnades flera förslag till hur klusteranalys skulle kunna användas; både för deskriptiva ändamål och för prognosändamål.

3.4.2 Användning av klusteranalys vid framställning av en sysselsättningsnomenklatur

Larsson (1976) redovisar ett försök med användning av klusteranalys vid framställning av en sysselsättningsnomenklatur. Arbetet skedde inom ramen för det s k SYN-projektet (Svensk Sysselsättningsnomenklatur). Detta var ett samarbetsprojekt mellan SCB och bl a arbetsmarknadens parter. Projektet finns beskrivet i "SYN-projektet 1976: Verksamhetsrapport från metodgruppen för utveckling av en generell sysselsättningsnomenklatur".

Utifrån resultaten av klusteranalyser av sammanlagt 809 typsysselsättningar med avseende på 28 grundfunktioner utarbetades ett förslag till nomenklatur av en grupp ämnesexperter. Oberoende av dessa försök hade ytterligare en ämnesgrupp utarbetat ett förslag till nomenklatur på mer konventionellt sätt.

Vid jämförelser av de båda förslagen visade det sig att det var stora likheter mellan resultaten. Den konventionella metoden gav visserligen fler grupper än den numeriska. För ett 15-tal grupper förelåg dock i det närmaste full överensstämmelse mellan de båda metoderna.

Tillämpningar av klusteranalys med liknande syfte men med något avvikande angreppssätt redovisas nedan i avsnitt 3.4.5.

3.4.3 Analys av inkomstförhållanden med hjälp av Inkomstfördelningsstudierna (HINK) 1972 och 1973

SCB publicerade 1976 en inkomstfördelningsstudie avseende förhållandena 1972. Statistik som belyser inkomstskillnader mellan olika grupper av hushåll, klassificerade efter barnantal, socio-ekonomisk indelning m m, och som beaktar såväl inkomster av arbete och kapital som transfereringar av olika slag hade efterfrågats såväl inom Sverige som internationellt. Den utförda studien ingick som en del av ett internationellt projekt inom FNs statistiska och ekonomiska kommission för Europa.

1980-10-31

Huvudrapporten redovisar resultat i tabellform med bl a decilfördelningar och snedhetsmått. I en bilaga till rapporten redovisas en studie över de anställdas inkomstförhållanden med hjälp av multivariata analysmetoder; se Elofsson och Larsson (1976) samt Elofsson (1974, 1977). För att undvika ett alltför heterogent material koncentrerade man studierna till anställda individer utan företagarinkomst. I dessa studier ingick 2 688 individer från urvalet.

AID-analys användes för att ge en uppfattning om vilka bakgrundsfaktorer som har stor betydelse för lönedifferenserna. För att ta hänsyn till olika urvalssannolikheter för individer från olika urvalsstrata använde man vägd AID-analys. Inom var och en av ett antal grupper som erhöles i AID-analysen gjordes sedan regressionsanalyser. Resultaten av dessa regressions samband jämfördes mellan de olika grupperna. För att underlätta dessa jämförelser anpassade man samma regressionsmodeller till de olika grupperna.

Analysen har kompletterats med en rad plottningar som illustrerar styrkan i sambanden.

1972 års inkomstfördelningsstudie har följts av årliga studier, s k HINKar. På datamaterialet från 1973 gjordes några metodstudier (Norberg, 1977). Stabiliteten i resultaten av AID-metoden prövade man genom att göra samma analys av båda, genom slumpmässig delning uppkomna, hälfter av datamaterialet. De båda AID-träden var tämligen lika (Norberg, 1978c). Experiment med olika plottningstekniker för att belysa sambanden mellan olika slag av inkomster gjordes. Resultaten av dessa studier presenteras i Norberg (1978a)

3.4.4 Analys av likheter och skillnader mellan tingsrätter med avseende på påföljder för viss typ av brott

Från enheten för rätts- och socialvårdsstatistik beställdes ett stort tabellmaterial av forskare vid sociologiska institutionen i Lund. Materialet omfattade för var och en av Sveriges tingsrätter (sammanlagt ett 100-tal) fördelningen av antalet domar på olika påföljder för skilda typer av brott år 1976. För en viss tingsrätt och en viss typ av brott angavs således antalet personer som dömts till respektive fängelse, böter, skyddstillsyn, villkorlig dom m fl påföljder.

Forskarna var intresserade av att studera likheter och skillnader mellan tingsrätter med avseende på fördelningen av påföljder, "påföljdsprofilen". Materialet var stort och svåröverskådligt, och man ville ha information om metoder som kunde vara till hjälp för att belysa data. Man kunde inte ge några mer preciserade frågeställningar att analysera.

För att demonstrera hur man med hjälp av klusteranalys kan genomlysa ett större tabellmaterial utarbetades med resurser från analysprojektet ett förslag, baserat på data för brottet stöld, se Holgersson (1978 a). Några resultat:

P/STM

1980-10-31

... - Vid en uppdelning av materialet i tio kluster, visade det sig att sex tingsrätter, alla belägna i sydvästra Skåne hamnat i samma kluster. Detta kluster hade hög genomsnittlig andel domar med påföljden fängelse.

... - Det visade sig att samtliga tingsrätter inom storstockholmsregionen hamnat i kluster med låg genomsnittlig andel fängelsedomar.

Finns det regionala skillnader i praxis och tradition inom svenskt rättsväsende? Klusteranalys kan kanske på detta sätt ge idéer och uppslag till fortsatta studier av denna fråga.

3.4.5 Klusteranalys av befattningsbyten inom industri-tjänstemannasektorn 1970-1975

I Gelin (1978) redovisas resultat av en klusteranalys av befattningsbyten inom industri- och handelstjänstemannasektorn åren 1970-1975. Klusteranalysen syftar för det första till att ordna och sammanfatta bytesmönstren till en grundstruktur som visar till vilka huvudsakliga områden man kan gruppera verksamheten inom sektorn under perioden. För det andra syftar analysen till att ge underlag för grupperingar av de detaljerade befattningskoderna till aggregat som lämpar sig för statistisk redovisning. För det tredje syftar analysen till att ge underlag för bedömningar av klusteranalysens värde som grupperings- och analysmetod. I detta senare avseende kan studien vara av intresse främst därför att en bytesmatris använts som underlag för framställningen av en olikhetsmatris.

Materialet utgörs av ungefär 470 000 individuppgifter hämtade ur lönestatistiken för industri- och handelstjänstemän med befattningar kodade enligt Befattningsnomenklatur för tjänstemän (3e upplagan)

De grupperingar som framkommit verkar meningsfulla som strukturbeskrivningar och som underlag för kodaggregeringar inför statistisk redovisning.

Avslutningsvis diskuteras utvärderingsproblem och användaraspekter, bl a klusteranalys som ett hjälpmedel vid framtagning av underlag för en mer rörlighetsinriktad utbildningsplanering.

3.4.6 Regressionsanalys av sjukfrånvaron hos barn på daghem

Våren 1977 genomfördes vid enheten för rätts- och socialvårdsstatistik vid SCB en undersökning: "Barns och personals närvaro och frånvaro i daghem 1977". Syftet med undersökningen var att belysa daghemsbarnens och daghemspersonalens närvaro och frånvaro samt orsaker till denna under en mätperiod som sträckte sig från februari till april 1977. Undersökningen avsåg också att beskriva daghemspersonalens värderingar gentemot sjuka barn på daghem.

Undersökningens resultaten redovisas i serien Statistiska meddelanden¹ med hjälp av ett femtiotal korstabeller samt en kort verbal sammanfattning. Från produktansvarigas sida var man emellertid intresserad av att ytterligare belysa sambanden mellan sjukfrånvaron hos barnen och vissa, utvalda bakgrundsfaktorer, såsom barnens ålder, typ av bostad, avståndet från hemmet till daghemmet m fl faktorer.

Med resurser från analysprojektet genomfördes en sambandsstudie med hjälp av regressionsanalys. Resultaten av denna finns redovisade i en bilaga till ovan nämnda SM; se även Holgersson (1978 b). Analysobjekten utgjordes av ca 6 000 barn från sammanlagt 477 daghemsavdelningar. Den beroende variabel som studerades var andelen dagar med sjukfrånvaro av antalet dagar med överenskommen närvaro under hela mätperioden. Innan numeriska beräkningar genomfördes, företogs en vägning av materialet med vikter proportionella mot uppräkningsfaktorerna för objekten. På så sätt skedde en kompensation för effekterna av en viss överrepresentation av småbarnsavdelningar i materialet.

Resultaten av regressionsanalysen tyder på att av de variabler som här undersökts, har främst ålder, men även typ av region (storstad vs landsort), bostadstyp, personaltäthet och avstånd

¹ SM S 1979:1

1980-10-31

till daghemmet, störst betydelse för förklaring av sjukfrånvaron hos daghemsbarn. Dessa resultat bygger på förutsättningar, vilkas uppfyllelse i viss mån kan ifrågasättas. I samband med presentationen av resultaten i SM-bilagan diskuteras detta kortfattat: Exempelvis bygger analysen på att ett obundet slumpmässigt urval föreligger ur undersökningspopulationen, barn på daghem i Sverige. Det faktum att analysobjekten (barnen) här inte utgör urvalsobjekt kan påverka resultaten i högre eller lägre grad beroende på hur stor homogenitet som föreligger inom urvalsobjekten (daghemsavdelningar). Bortfallet som i denna analys utgjorde ungefär tio procent av det totala urvalet, kan naturligtvis också ha haft en snedvridande effekt på analysresultaten. Exempelvis skulle fallet kunna vara att daghemsavdelningar med speciellt låg personaltäthet genomgående ansåg sig sakna tid för ifyllande av enkäten. Någon studie av bortfallet genomfördes dock ej.

Vidare bygger resultaten självfallet på att förutsättningarna för regressionsmodellen är riktiga. Detta innebär bl a att sambandet mellan den beroende och de förklarande variablerna är linjärt och additivt. Det senare innebär att, t ex, sjukfrånvaron för barn som bor i lägenhet i förort eller ytterområde till tätort ligger ett konstant antal procentenheter över eller under sjukfrånvaron för andra barn. Detta skall gälla oavsett barnets ålder, daghemmets personaltäthet osv, det vill säga oavsett värdet på samtliga övriga förklaringsvariabler.

Man kan tänka sig att sambandet inte är additivt, utan innehåller samspel. I vårt fall kan man exempelvis misstänka att det finns en samspelseffekt mellan variabeln ålder och vissa av de övriga förklaringsvariablerna, t ex bostadstyp. Ett sådant samspel innebär att effekten på sjukfrånvaron av olika boendeformer (eller snarare av de underliggande faktorer som påverkar boendet) ökar - eller minskar - med stigande ålder.

AID-analys kan användas för att belysa samspelseffekter, se avsnitt 3.6.1. AID-analyser av daghemsundersökningen redovisas i Norberg (1980 b).

3.4.7 Kommunindelning för individualstatistiken: Faktoranalys som ett hjälpmedel vid stratifiering av kommuner

I samband med att SCB 1969 från skolöverstyrelsen övertog ansvaret för skolstatistiken beslöts att SCB som ett komplement till insamlingen av antalsuppgifter - dvs uppgifter om antalet elever, klasser etc - även skulle insamla uppgifter som belyste hur ett urval av elever gick genom skolväsendet. Denna statistik benämndes individualstatistiken. Två årskullar av grundskoleelever har hittills blivit föremål för longitudinella studier av detta slag. En dokumentation av individualstatistikens primärmaterial har publicerats i Promemorior från SCB 1976:8 "Individualstatistiken. Skoldata för ett urval elever födda 1948 och 1953".

P/STM

1980-10-31

År 1976 påbörjades planeringen av ett nytt individualstatistikprojekt. I det tidigare projektet ingick samtliga födda ett visst datum ett visst år i urvalet. Denna typ av direkt urval av elever har det nya projektet velat frångå; dels av administrativa skäl, dels därför att man ville ha större möjligheter till studier av elevgrupper, t ex klasser. Man har därför föreslagit ett flerstegsurval, där det första stegets urvalsobjekt skulle utgöras av kommuner. För detta ändamål sökte man en stratifiering av kommunerna, där strata skulle vara homogena med avseende på sådana egenskaper som kunde antas ha samband med förhållanden i skolorna. Förutom att utgöra utgångspunkt för urvalsdesignen, skulle strata tjäna som redovisningsgrupper innehållande olika "skolmiljöer".

I Holgersson (1979 a) redovisas ett förslag till hur metoden faktoranalys kan användas som ett hjälpmedel att finna en sådan lämplig gruppindelning av Sveriges kommuner. En stor mängd information hade insamlats för varje enskild kommun rörande befolkningsförhållanden, ekonomiska, kulturella, politiska och arbetsmarknadsmässiga förhållanden etc. Ur denna information valdes 29 variabler som ansågs ha betydelse för förhållanden i skolorna. Med hjälp av faktoranalys söktes en sammanfattning av informationen i de 29 uppmätta variablerna i ett mindre antal latent variabler (faktorer). Av både teoretiska och saklogiska skäl valdes en faktorlösning omfattande sex faktorer. Faktorerna betecknades på följande sätt:

Faktor nr Namn

- | | |
|---|-----------------------|
| 1 | Storleksfaktor |
| 2 | Politisk faktor |
| 3 | Sysselsättningsfaktor |
| 4 | Ekonomisk faktor |
| 5 | Invandrarfaktor |
| 6 | Tillväxtfaktor |

Denna komprimerade information om kommunerna kan användas på flera sätt. En metod, som leder till en enkel beskrivning av de resulterande redovisningsgrupperna, kan beskrivas på följande sätt: För var och en av de faktorer som bedöms vara av störst intresse för undersökningen, väljs en variabel ut som är högt korrelerad med denna faktor (hög faktorladdning). Dessa variabler används därefter som stratifieringsvariabler. I föreliggande undersökning ansågs fyra av de sex faktorerna vara de mest betydelsefulla för projektet, nämligen storleksfaktorn, politisk faktor, sysselsättningsfaktorn och invandrarfaktorn. De stratifieringsvariabler som valdes var folkmängd, andelen socialistiska mandat, andelen sysselsatta inom offentlig förvaltning och andelen invandrarelever.

Utfallsrummet för varje stratifieringsvariabel delades sedan i två delar vid följande variabelvärden:

Folkmängd: 25 000 invånare
Andel socialistiska mandat: 50 %
Andel sysselsatta i offentlig förvaltning: 25 %
Andel invandrarelever: 8 %

P/STM

1980-10-31

Då stratifiering genomförts efter denna uppdelning, visade det sig att några av de på detta sätt bildade strata var ytterst små eller egendomligt sammansatta, varför stratifieringen delvis justerades. Därvid beaktades även variablerna total kommunal utdebitering, andelen behöriga lärare samt andelen deltagare i studiecirkel utan statliga bidrag. De sistnämnda variablerna hade nämligen i faktoranalysen inte visat sig korrelera högt med någon av de sex faktorerna. Den slutliga grupperingen kom att innehålla sammanlagt 13 strata, se Lundström (1979).

3.4.8 Medieutlåning och arrangemang i kommunala bibliotekssystem - en sambandsanalys

I folkbiblioteksstatistiken för år 1976 (Statistiska meddelanden (SM) U 1977:31) redovisar SCB de enskilda kommunernas folkbiblioteksverksamhet med variabler avseende mediebestånd, utlåning, programverksamhet, personal, kostnader etc. Dessutom redovisas ett antal s k bakgrundsvariabler, t ex folkmängd, andel barn 0-12 år, medelinkomst per hushåll, andelen socialdemokratiska och vänsterpartistiska mandat. I Kulturstatistik 1970-1975: Appendix (SOS 1977) redovisar statens kulturråd och SCB bl a varje kommuns anslag per invånare för kulturverksamhet totalt 1975 och anslag till folkbiblioteken 1975.

I bägge dessa statistikprodukter görs vissa beskrivande kommentarer kring samband mellan två eller högst tre olika variabler åt gången. Syftet med den analys som nu har gjorts är att åstadkomma en sammanfattande beskrivning av samband mellan ett helt variabelkomplex med hjälp av relationsmodeller. Detta syftar till att öka kunskapen om relationerna mellan variablerna och att därigenom göra statistiken mer användbar. Vidare är syftet att ge underlag för bedömning av vilka bakgrundsvariabler som har starka samband med variablerna för folkbiblioteksverksamheten och som därför borde införlivas i en ordinarie statistikredovisning av folkbibliotekens verksamhet, därmed ökande användbarheten av statistiken.

Relationsmodeller har formulerats baserade på hypoteser om samband mellan variabler. Dessa samband antas vara linjära. Metoden för att skatta de linjära sambandens utseende och styrka är multipel regressionsanalys. De centrala sambanden mellan variablerna har åskådliggjorts med plottningar. Modellen har tolkats som tre delsystem: kommunens satsning på kultur, driftskostnadens uppdelning och konsumtion av bibliotekskulturen.

Vad först beträffar kulturanslagen så visar resultaten att ungefär hälften av variationen i dessa anslag förklaras av kommunernas skattekraft, invånartal och politiska ställning i fullmäktigeförsamlingarna, medan tätortsgrad och hushållens medelinkomst inte ger ytterligare tillskott till förklaringen utan resten beror på faktorer utanför modellen.

Driftkostnaden delas i medie-, personal- och övriga kostnader. Uppdelningen varierar uppenbarligen kraftigt från kommun till kommun. Analysen visar att kommuner med låg driftkostnad, mätt per invånare satsar i genomsnitt större andel av pengarna på medier än andra kommuner.

Driftkostnaden sammantaget förklarar "kulturkonsumtionen" på bibliotek (här främst medieutlåningen) bättre än vad kostnaden

för medier och för personal gör. Analysresultaten presenteras i Norberg och Gouiedo (1980).

3.4.9 Analyser av uppfattad och förväntad inflation med hjälp av data från Hushållens inköpsplaner (HIP)

SCB-undersökningen "Hushållens inköpsplaner" ger regelbundet information om bl a hushållens uppfattning om hur priserna utvecklats under de senaste tolv månaderna (uppfattad inflationstakt) och hur man tror att prisutvecklingen kommer att bli under de närmaste tolv månaderna (förväntad inflations-takt).

Med hjälp av regressionsanalys undersökte Holgersson bl a hur den förväntade inflationstakten varierade som en funktion av uppfattad inflationstakt, ålder, kön och inkomst. Signifikanta samband erhöles genomgående för variablerna uppfattad inflationstakt, ålder och (i vissa analyser) inkomst. Resultaten har använts av Lars Jonung vid Nationalekonomiska institutionen i Lund och publicerats i rapporter till Prisregleringskommittén samt i rapporter inom forskningsprojektet "Arbetslöshet, priser och löner i Sverige under 1900-talet". Lütjohann och Gelin (1978) har använt data från HIP för att belysa hur kontingenstabellanalys fungerar jämfört med regressionsanalys.

3.4.10 Olika försök med grafisk framställning i SCBs publikationer

I den officiella statistik som utges av SCB och andra statistikproducerande myndigheter är än så länge den vanligaste formen för publicering korstabeller försedda med varierande mängd av förklarande text och kommentarer till innehållet. På senare år har man emellertid börjat förse allt fler publikationer med någon form av grafisk framställning som komplement till tabellerna. Grafisk framställning ställer emellertid nya krav på produktansvariga.

Man måste ta speciella metodologiska och (kanske) etiska hänsyn vid diagramritning:

- Diagrammen måste följa vissa vedertagna konstruktionsprinciper för att missförstånd skall undvikas (exempelvis skall ytan av en rektangel i ett histogram, inte nödvändigtvis höjden, vara proportionell mot frekvensen i motsvarande klass).

- Försiktighet måste iakttas vid s k stympning av skalor m m.

Man måste vara estet:

- Diagrammen måste göras tilltalande för ögat. Bortsett från noggrannhet vid val av format, förhållande mellan horisontell och vertikal skalstorlek, val av linjegrovheter etc (för att inte nämna de ytterligare estetiska överväganden som krävs vid framställande av färgdiagram) är man tvungen att med stor omsorg välja typ av "skraffering" så att inte optiska villor uppstår.

Man måste dessutom vara något av psykolog. Trots att man följt alla vedertagna regler kan ett diagram vara obegripligt. En viktig fråga vid diagramkonstruktion är följande:

P/STM

1980-10-31

...
- Kan det diagram jag konstruerat tydas av någon annan än jag själv? Inte sällan kan ett diagram vara alltför informationsspackat för att överhuvudtaget ge någon upplysning.

...
Man bör slutligen beakta de ekonomiska aspekterna vid grafisk framställning i offentlig statistikproduktion:

- Är detta diagram en engångsföreteelse eller skall det uppdateras och återkomma år från år i en viss publikation?

I båda fallen, men speciellt i det senare kan det vara skäl att överväga om det kan vara ekonomiskt och praktiskt lämpligt med datorritade diagram. För närvarande har SCB viss utrustning för att framställa datorritade diagram, och kompletteringar av denna planeras.

Råd och tips om konstruktion och ritning av diagram ges i Informationsenhetens broschyr "Rita diagram" av Wahlén (1980).

I samband med den kartläggning av analysbehovet vid SCB som genomfördes 1976 (jfr avsnitt 3.3) framfördes ett starkt intresse för ökad grafisk presentation av SCBs produkter. Inom projektets ram har vi därför i några tillämpningar ägnat oss åt metodologiska och praktiska aspekter på grafisk framställning.

År 1977 publicerades den första utgåvan av Rättsstatistisk Årsbok, en publikation för samordning av den rättsstatistiska redovisningen. Årsboken innehöll ett stort antal diagram som illustrerade innehållet i vissa av tabellerna. Av produktansvariga för årsboken ombads vi att ge synpunkter på diagramval, -konstruktion etc, se Holgersson (1977 c). Detta arbete resulterade bl a i att diagramframställningen överfördes till dator och att i samband därmed ett generellt subrutinpaket för datorritning av färgdiagram utvecklades, se Andersson och Cassel (1978). Även ett program för framställning av koroletkarter (karter som visar ett områdes variabelvärde med färgton eller skraffering) på länsnivå har i detta sammanhang utvecklats, se Rättsstatistisk Årsbok 1978.

Ett förslag till grafisk presentation av statistik över varustransporter med lastbil har utarbetats, se Berglund (1977 a).

I samband med analyserna av Inkomstfördelningsstudien 1973, se Norberg (1977) utarbetades en speciell plottningsteknik för att belysa de mestadels komplicerade sambanden mellan olika slag av inkomster för hushållen. Man brukar särskilja mellan inkomstbegrepp som faktorinkomst (summan av löneinkomst, företagarinkomst och kapitalinkomst), negativa transfereringar (skatt m m), positiva transfereringar (olika former av bidrag), bruttoinkomst (faktorinkomst + positiva transfereringar) samt disponibel inkomst (bruttoinkomst minus negativa transfereringar). Denna metodik, som på ett mycket tydligt sätt kan illustrera hur skatteskalorna fungerar för olika grupper av människor i samhället, redovisas i Norberg (1978 a).

För vissa dödsorsaker, exempelvis ischemiska hjärtsjukdomar föreligger ett tydligt säsongsmönster i dödlighetstalen. För årsboken Dödsorsaker har utarbetats förslag till grafisk presentation av sådana säsongsmönster, se Dödsorsaker 1977, 1978 och 1979.

1980-10-31

3.4.11 Vädrets effekt på elförbrukningen i hushåll, handel m m

SCB publicerar en månatlig statistik över produktion och förbrukning av elenergi i Sverige. Elförbrukningen fördelas på fyra konsumentgrupper samt förluster. Uppgifterna är preliminära och baseras på urvalsundersökningar. Då slutliga uppgifter enligt den årliga industristatistiken och den årliga elstatistiken föreligger anpassas den månatliga elstatistikens värden till dessa statistikgrenar.

Ca 40 % av all elförbrukning redovisas under posten "hushåll, handel m m". En inte obetydlig del av elförbrukningen i Sveriges hushåll går till bostadsuppvärmning. Den totala elförbrukningen (inkl hushållsel och el för varmvatten) i eluppvärmda småhus svarar för en fjärdedel av elförbrukningen i gruppen "hushåll, handel m m". I runda tal går därvid 60 % till bostadsuppvärmning.

En viss del av variationerna i elförbrukningsstatistiken kan förklaras av väderförhållandena. Sambanden mellan väder och elförbrukning har analyserats med en icke-linjär regressionsmodell (Norberg, 1980c). För gruppen "Hushåll, handel m m" är sambandet klart medan det för de flesta industribranscher är alltför obetydligt för att användas i en väderkorrigerad elstatistik.

Modellen består av två delar: en väderberoende del med linjär tidstrend multiplicerad med månadens graddagstal, och en icke väderberoende del med en annan linjär tidstrend och ett skattat säsongmönster. Graddagstalet är summan av differensen mellan + 17° C och utomhustemperaturen för varje dag i månaden. Vinddata har testats i modellen men gav ej signifikant minskning av de oförklarade variationerna.

Med hjälp av modellen blir det nu möjligt att i månadsstatistiken uppskatta vädrets effekt på elförbrukningen i jämförelse med samma månad året innan.

En ARIMA-modell har också prövats med hjälp av Box-Jenkins teknik. Båda modellerna hade ungefär lika god anpassning till data. Regressionsmodellen ansågs vara mer lättolkad och ger bättre möjligheter att med diagram beskriva elförbrukningens historia.

3.4.12 Användning av Box-Jenkinsansatsen i syfte att förbättra preliminärstatistiken för industri- produktionsindex

Syfte: Att undersöka om Box-Jenkinsansatsen kan användas till att förbättra preliminärstatistiken för vissa branscher, där det för närvarande blir otillfredsställande stora skillnader mellan det preliminära indexet och det reviderade.

P/STM

1980-10-31

Genomförande: Ett stort antal Box-Jenkinsmodeller har testats.

1 Univariat modell

2 Modifierad univariat modell (där dataproblemen med semester-
månaderna juli/augusti är eliminerade)

3 Transferfunktionsmodell med arbetade timmar som oberoende
variabel

4 Transferfunktionsmodell med arbetade timmar och leveranser
som oberoende variabler

5 Transferfunktionsmodell med konjunkturbarometerdata som
oberoende variabel

Resultat: En viss förbättring skulle kunna uppnås, dock ej
så stor att man blir tillfredsställd. Resultatet (avseende
produktionen inom maskinindustri, SNI 382) visas i tabell 1
nedan.

Tabellen har erhållits på följande sätt. För var och en av de
21 månaderna jan 78 - sep 79 har man på två alternativa sätt
beräknat preliminära värden på den procentuella förändringen
jämfört med samma månad året innan. Det ena sättet innebär att
man använt den nuvarande metoden för beräkning av preliminärt
värde. Det andra sättet innebär att man använd Box-Jenkins-
teknik. Man har härvid eftersträvat realistiska förhållanden,
såtillvida att man bara utnyttjat data som i verkligheten var
tillgängliga vid den tidpunkt då det preliminära värdet normalt
skulle ha beräknats. De båda erhållna preliminära värdena har
sedan jämförts med det reviderade värdet för samma månad, var-
vid det reviderade värdet betraktats som sant värde. Man har
sett efter vilket av de båda preliminära värdena som legat
närmast det reviderade värdet, och man har vidare noterat hur
många procentenheter man kommit närmare det reviderade värdet
genom att använda det "bättre" preliminära värdet jämfört med
det "sämre" preliminära värdet. (Om förbättringen understigit
en procentenhet, har de båda metoderna betraktats som likvärdiga.)

Exempel: Det i preliminärstatistiken publicerade värdet för en
viss månad är 10 %, vilket säger att industriproduktionen skulle
ha ökat med 10 % jämfört med samma månad föregående år. Antag
att Box-Jenkinstekniken ger det preliminära värdet 12 % för
samma månad. Antag vidare att det reviderade värdet för månaden
i fråga är 5 %. Då konstateras att den nuvarande metoden ger
det bästa preliminära värdet för den aktuella månaden, närmare
bestämt ett värde som är två procentenheter bättre (dvs två
procentenheter närmare det reviderade värdet) än det värde som
Box-Jenkinstekniken ger.

P/STM

1980-10-31

Tabell 1 Jämförelse mellan nuvarande metod och Box-Jenkins-ansats

	Nuvarande metod bättre			Likvärdiga	Box-Jenkinsansatsen bättre		
	5-7	3-5	1-3	0-1	1-3	3-5	5-7
Antal procentenheters förbättring							
Antal månader	0	3	4	4	2	5	3

För närvarande publiceras det preliminära indexet med ca 1,5 månaders fördröjning. Om man reducerar tidsförskjutningen till 0,5 månader och alltså beräknar preliminära värden på grundval av data som är tillgängliga vid denna tidigare tidpunkt, erhålls följande resultat.

Tabell 2 Jämförelse mellan nuvarande metod och Box-Jenkins-ansats. Uppsnabbad publicering

	Nuvarande metod bättre			Likvärdiga	Box-Jenkinsansatsen bättre		
	5-7	3-5	1-3	0-1	1-3	3-5	5-7
Antal procentenheters förbättring							
Antal månader	2	1	5	4	2	4	3

Både tabell 1 och 2 ger vid handen att skillnaden mellan de båda metoderna för beräkning av preliminära värden inte är speciellt stor. Varför resultatet ej blev bättre kan självfallet diskuteras, men en av orsakerna är att hjälpserien (arbetade timmar) innehåller många diskreta hopp och är utsatt för stora ej förklarbara variationer. Vidare är det tveksamt om serien numer är speciellt bra som indikator för industriproduktionsindex. Det är troligtvis dessa problem som gör att den nuvarande metoden ej fungerar speciellt bra. (Man använder arbetade timmar som huvudindikator i preliminärstatistiken).

För vidare detaljer hänvisas till Holmén (1980b).

P/STM

1980-10-31

3.5 Metodutveckling: Tillämpad forskning i anslutning till analysverksamheten vid SCB

Speciella problem är förknippade med användningen av analysmetoder när data härrör från en survey, vilket oftast är fallet beträffande SCBs produktion. Detta kan innebära följande:

- komplicerade urvalsförfaranden har använts
- data innehåller fel av typen mätfel, bearbetningsfel och bortfallsfel.

Dessa problem behandlas otillfredsställande - eller inte alls - i den statistiska analyslitteraturen. Samtidigt berörs inte analysfrågor närmare i samplinglitteraturen.

Projektgruppen har arbetat med att kartlägga och belysa dessa problem samt konsekvenser av att genomföra olika åtgärder vid analyser som bygger på data från surveyer. Arbetet har bl a resulterat i en översikt över problem vid statistisk analys av surveydata, se Wretman (1980). Översikten avser att på ett enkelt sätt presentera vissa grundläggande metodstatistiska problem, som man bör tänka igenom innan man sätter igång med en analys. Framför allt diskuteras inferens- och tolkningsproblem. Det framhålls hur man ur metodstatistisk synpunkt kan ha olika ambitionsnivåer vid analys av surveydata, nämligen (i) deskription av givna data, (ii) inferens om den ändliga populationens utseende och (iii) inferens om en modell som antagits generera den ändliga populationen.

Vidare har framtagits en förteckning över litteratur rörande statistisk analys i surveysituationer, se Bergholm (1978).

I följande delavsnitt beskrivs metodutveckling som bedrivits inom projektet i anslutning till regressionsanalys vid urval från ändliga populationer (3.5.1), regressionsanalys i situationer med mätfel (3.5.2) samt i anslutning till analys av undersökningar med bortfall (3.5.3).

Sammanfattningsvis kan det projektarbete som skildras i avsnitt 3.5 sägas utgöra en första början till vad som i riktlinjerna beskrivs som upprättande och å jourhållande av "en problemkatalog för metodutvecklingsarbetet inom det statistiska analysområdet vid SCB". Mycket återstår ännu att göra på detta område. Varken internationellt eller vid svenska forskningsinstitutioner sysslar man mer än undantagsvis med dessa frågor. Det förefaller rimligt att man från SCBs sida känner ett speciellt ansvar för detta forskningsområde. Förutom att bedriva egen forskning bör SCB på olika sätt verka för att forskningsinstitutioner inom och utom universiteten behandlar problem knutna till analys av surveydata (se vidare avsnitt 5).

1980-10-31

3.5.1 Regressionsanalys vid urval från ändliga populationer

Wretman (1978) ger en kort översikt över olika ansatser vid regressionsanalys när man har stickprov från en ändlig population. Följande ansatser behandlas: (i) Fixpopulationsansats: skattning av karakteristika i den ändliga populationen, (ii) Modellansats: skattning av karakteristika i den ändliga populationen och (iii) Modellansats: skattning av modellparametrar. Godtycklig urvalsdesign antas och grundläggande resultat från olika författare sammanställs.

Jönrup (1975) och Jönrup och Rennermalm (1974 och 1976) behandlar skattning av regressionskoefficienter under ansatsen (i) ovan. Rennermalm (1980) härleder uttryck för biasen hos dessa skattningar.

Ansatzerna (i) och (iii) kan leda till olika estimatorer, eftersom man dels skattar olika saker, och dels använder olika kriterier vid val av estimator. Under ansats (iii) kan man ofta härleda estimatorer av modellparametrar, som (enligt vissa kriterier) är optimala, förutsatt att den antagna modellen är sann. Under ansats (i) kan man härleda estimatorer av populationskarakteristika, som (enligt andra kriterier) har vissa önskvärda egenskaper utan att något modellantagande behöver göras.

Vid inferens rörande modellparametrar under ansats (iii) kan man emellertid ibland av robusthetsskäl vilja använda en estimator, som härletts under ansats (i), speciellt om modellantagandet betraktas som mindre säkert. Tankegången är då i korthet följande: En estimator av typ (iii), som är optimal under en antagen modell, kan visa sig dålig eller svår att tolka ifall den antagna modellen inte skulle gälla. Å andra sidan kan en estimator av typ (i), ehuru icke optimal, ändå vara tolkbar och behålla vissa hyggliga egenskaper även om den antagna modellen inte gäller. Ringnér (1979 och 1980) utvecklar detta närmare. Bl a visas att mycket små avvikelser från linearitet i modellen räcker för att förstöra optimalitetsegenskaperna hos estimatorer härledda under ansats (iii).

I en simuleringsstudie (Norberg 1978d) har estimatorer, vägda resp ej vägda med inverterade urvalssannolikheter, jämförts under ansats (i). Konstruerade populationer och urvalsigner har valts enkla för att resultaten skall vara lättolkade.

Resultat av studien är att man bör väga med inverterade urvalssannolikheter för att få bra skattningar av regressions- och korrelationskoefficienter i den ändliga populationen. När stratifieringen baseras på beroende variabeln är det nödvändigt att väga på detta sätt.

3.5.2 Regressionsanalys i situationer med mätfel

Projektets syfte är att undersöka vilken effekt som mätfel har på skattning av parametrar i regressionsanalys, se Cassel och Råbäck (1979). Som exempel på tillämpningsområden för SCBs del kan nämnas regressionsanalys vid dikotoma variabler och analys av regressionssamband i frekvenstabeller. Båda exemplen kan ses som regressionsanalys som störs av mätfel. Man tänker sig då att det existerar ett regressionssamband mellan två variabler x och y , men man kan inte observera x och y direkt utan man kan bara se i fall observationerna hamnar i vissa klasser som då motsvarar t ex klassindelningen i frekvenstabellerna. Om man nu beräknar regressionssambandet på klassmitterna så begår man ett fel som kan sägas vara orsakat av "mätfelet". Inom projektet har en viss litteraturgenomgång gjorts för att finna metoder som behandlar problemet. En simuleringsstudie har utförts för att man skulle få en uppfattning om den möjliga effekten av mätfelen. Det visar sig att regression beräknad på klassmitter i frekvenstabeller kan avvika kraftigt från den riktiga regressionen.

3.5.3 Analys i undersökningar med bortfall

Vid de flesta undersökningar drabbas man i praktiken av bortfall. Det kan i sådana situationer bli en svår bedömningsfråga hur man skall gå tillväga för att genomföra en statistisk analys. Några klara regler för hur man bör göra finns inte.

Objekten i en undersökning kan ur bortfallssynpunkt indelas i tre kategorier: (i) Objekt för vilka man erhållit värden på samtliga variabler (s k fullständiga objekt), (ii) objekt för vilka man erhållit värden på en eller flera men ej alla variabler (s k partiellt bortfall) samt (iii) Objekt för vilka man saknar värden på alla variabler (s k objektsbortfall).

Många multivariata analysmetoder (t ex regressionsanalys) utgår från att man först skattar varianser och kovarianser för alla inblandade variabler. I litteraturen beskrivs bl a följande tre olika tillvägagångssätt för sådana beräkningar, då bortfall föreligger: (i) Endast data från de fullständiga objekten utnyttjas (s k "listwise deletion"), (ii) Varje varians- och kovariansskattning utnyttjar data från alla objekt som har värden på variabeln resp variabelparet (s k "pairwise deletion"), (iii) Saknade värden imputeras för de partiella bortfallsobjekten, varefter beräkningarna baseras på den sålunda "kompletterade" uppsättningen data från både de fullständiga objekten och de partiella bortfallsobjekten.

1980-10-31

En nackdel med "listwise deletion" är att de fullständiga objekten kan utgöra en alltför liten del av undersökningsobjekten (speciellt ifall det partiella bortfallet är utspritt över många objekt). Vid "pairwise deletion" utnyttjas å andra sidan data även från det partiella bortfallet. En nackdel är här i stället att den skattade varians-kovariansmatrisen inte säkert blir positivt semidefinit, vilket kan leda till orimligheter eller sammanbrott i analysen, se Heiberger (1977).

En översikt över förfaringssätt i samband med "missing data" vid multivariat analys ges av Kim och Curry (1977) samt Frane (1978). Statistiska metoder i samband med bortfall redovisas också i SCB-handboken "Räkna med bortfall" (1980), dock utan tonvikt på analysproblem.

3.6 Implementering av vissa analysmetoder vid SCB

Inom projektet har studerats användbarheten för SCBs del av vissa statistiska analysmetoder. Särskild uppmärksamhet har härvid ägnats åt AID-analys, Box-Jenkinsanalys, analys med hjälp av splinefunktioner, analys av kontingenstabeller med hjälp av log-linjära modeller samt säsongrensning. En närmare redogörelse för detta arbete återfinns i delavsnitten 3.6.1-3.6.6.

3.6.1 AID-analys - ett komplement till regressionsanalys

AID är en dataanalysmetod som använts i några olika undersökningar vid SCB. Exempel är Hushållens inkomster (HINK), Partisympatiundersökningarna (PSU), Hushållsbudgetundersökningarna (HBU) och Hushållens inköpsplaner (HIP), dvs i individ- och hushållsundersökningar med tämligen många undersökningsvariabler. Några principer för hur AID skall användas vid SCB finns ej, däremot har inom analysprojektet en handledning i AID-analys skrivits, se Norberg (1980 b). Denna handledning syftar till att beskriva metodens matematiska algoritm, problem vid analys av surveydata och tolkning av resultaten. Ett avsnitt behandlar tillämpningsområden med speciell tonvikt på SCBs verksamhet. Slutligen ges en beskrivning av datorprogram och tolkningsråd till utskriften.

3.6.2 Box-Jenkinsanalys

Box-Jenkinsanalys är benämning på en speciell analysstrategi vars huvudsyfte är att ge korta prognoser av tidsserier. Metodiken kan även användas för deskription och analys av tidsserier. Man kan exempelvis testa hypoteser rörande trend-förskjutningar som följd av externa händelser (strejker, lag-ändringar, energikriser m m).

Box-Jenkinsmetoden kräver mycket beräkningar. I praktiken måste man därför ha tillgång till datorer och speciella datorprogram för att kunna genomföra analysen. Box-Jenkinsanalys kan vid SCB göras på 2 nivåer.

i) Univariata modeller; där analysen enbart baserar sig på en given series egna historiska förlopp

Univariata modeller kan med fördel användas till dataanalys för att bl a

- . eliminera säsongen
- . identifiera punkter där man är osäker på datas kvalitet.
Dessa data kan i efterhand korrigeras

ii) Transferfunktionsmodeller; där en modell byggs upp av förklarande serier samt av det unika beteendet i den beroende serien.

Transferfunktionsmodeller kan bl a användas till

- . korta prognoser
- . förbättring av preliminärstatistiken med hjälp av 'leading indicators' (efterhandsprognoser)
- . skattning av effekter av interna/externa händelser och eventuell korrigering för dessa. Sådana händelser kan vara strejker, momsändringar, definitionsändringar, byte av urvalsdesign osv

Viss kritik har riktats mot Box-Jenkinsansatsen bl a för att metoden anses modellös, dvs man utgår enbart från data och ej från en given strukturell modell. Med hjälp av data bestäms ett modellsamband och detta formaliseras i en s k ARIMA-modell. Denna kritik är till viss del relevant vid t ex en ekonometrisk ansats, där man för det mesta vill göra parameterskattningar i på förhand kända eller antagna modeller. I de tillämpningar som beskrivits under i och ii är emellertid avsikten en annan, nämligen datanalis, varför det är helt naturligt att arbeta helt på basis av data.

Vår bedömning är således att Box-Jenkinsansatsen är användbar inom SCB för att försöka förbättra preliminärstatistik och för att undersöka (ev korrigera) SCB-data. Metoden kan också användas för att ge uppslag till kvalitetsförbättrande åtgärder.

Vi är dock något tveksamma till att använda Box-Jenkinsansatsen vid uppbyggandet av stora förklarande modeller, men detta är ju inte SCBs primära uppgift.

Ett flertal tester med Box-Jenkinsansatsen har gjorts, bl a vid F/IN: Test om Box-Jenkinsansatsen kan användas att förbättra det preliminära industriproduktionsindexet. Vid F/HA har metoden prövats i samband med användandet av konjunkturbarometerdata i detaljhandelns omsättningsstatistik, studier av oljekrisens påverkan på hushållens elkonsumtion samt vid en del datakvalitetsundersökningar.

Ett PM om Box-Jenkinsansatsen har utarbetats inom projektets ram, se Holmén (1979 b).

3.6.3 Utjämning av fruktsamhetskurvor med kubiska spline-funktioner

Kvinnors fruktsamhet, dvs förmåga att föda barn, är en mycket betydelsefull variabel i demografiska undersökningar. Vanligen studeras fruktsamheten med hjälp av s k fruktsamhetskurvor, vilka för varje ålder anger antalet födda per 1000 kvinnor. Kurvor över sådana observerade fruktsamhetstal uppvisar ofta ett mer eller mindre taggigt utseende. Ju mindre populationen är desto taggigare är den observerade kurvan. I diagram 1 och 2 visas ett exempel på att kurvor från stora populationer är betydligt jämnare än de från små populationer. Det är därför naturligt att antaga existensen av en "sann" jämn bakomliggande kurva och att avvikelserna från denna kurva beror på tillfälliga orsaker. Framför allt vid regionala studier är det viktigt att söka skatta de "sanna" fruktsamhetskurvorna.

I arbetet med "Regional fruktsamhetstal för perioderna 1968-1970, 1971-1973, 1974-1976" utnyttjades kubiska spline-funktioner som utjämningsfunktioner, se Rennermalm (1979). Antagandet som gjordes var att de "sanna" fruktsamhetskurvorna kunde beskrivas av kubiska splinefunktioner. Kubiska splinefunktioner, som består av bitvis definierade tredjegradspolynom, är mycket flexibla och ger trots detta trevliga jämna kurvor. Splinefunktioner är speciellt användbara vid approximation av funktioner vars uppförande i en del av definitionsområdet kan vara totalt oberoende eller mycket svagt beroende av deras uppförande i en annan del av definitionsområdet, se Hoem och Rennermalm (1979) samt Hoem et al (1980).

Kubiska splinefunktioner kan dessutom ses som en linjär modell. Eftersom fruktsamhetstal är asymptotiskt normalfördelade och oberoende kan vi om populationerna är tillräckligt stora utnyttja teorin för linjära modeller för att testa hypoteser av olika slag. I "Regionala fruktsamhetstal...." testades de skattade splinefunktionernas anpassning. För i stort sett alla regioner befanns kubiska splinefunktioner kunna tjäna som en beskrivning av de bakomliggande fruktsamhetsstrukturerna.

1980-10-31

Diagram 1. Åldersspecifika fruktsamhetstal för Sverige
1974-1976

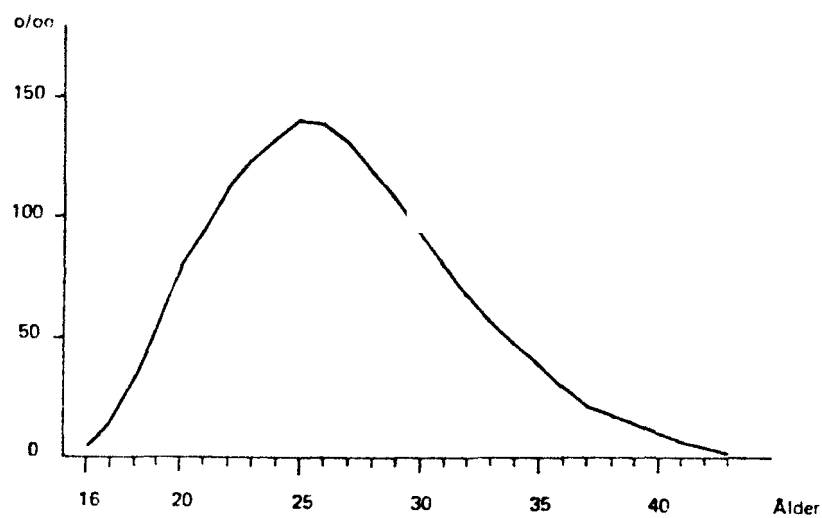
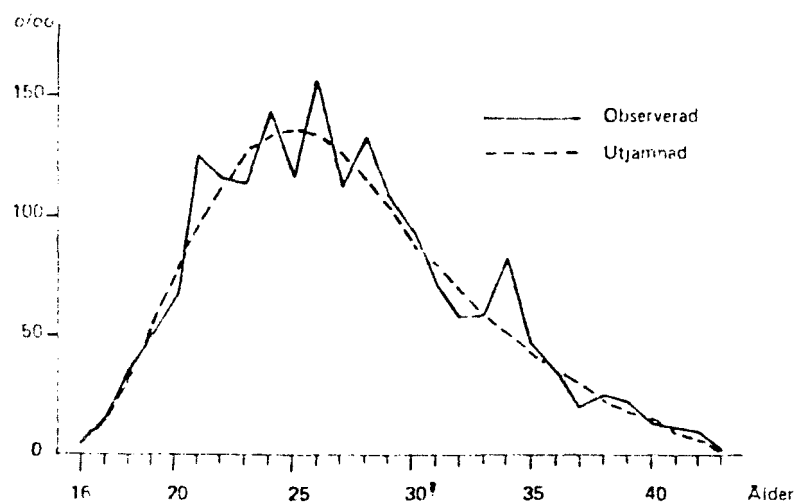


Diagram 2. Åldersspecifika fruktsamhetstal för Berg,
Härjedalen, Åre 1974-1976



1980-10-31

3.6.4 Analys av samband hos multivariata kvalitativa data

Den statistiska vetenskapen har under 1970-talet utvecklat datorstödda metoder och teori för sambandsanalys i följande situation. Samtliga variabler är kvalitativa; de mäts medelst kardinalskalor - eventuellt behandlas skalor med mer information som om de vore endast kardinala. Data sammanfattas i en kontingenstabell, där varje cell innehåller ett heltal, antalet analysenheter i cellen. Analys av sådana data kallas ofta kontingenstabellanalys.

Standardförfarandet vid kontingenstabellanalys grundas på antagandet att logaritmen för den till en cell hörande observationssannolikheten är en linjär funktion av indelningsgrunderna på samma sätt som väntevärdet i variansanalys med samspelseffekter. Det finns flera standardprogram för analys enligt denna ansats. Beräknade koefficienter kan, liksom vid regressionsanalys, tolkas även rent deskriptivt. Kontingenstabellanalys enligt standardansatsen torde för lång tid framåt komma att vara standardförfarandet vid analys av samband mellan kvalitativa variabler.

Kontingenstabellanalys har inom ramen för analysprojektet studerats av Gunnar Gelin I/PI och Harry Lütjohann P/STM. De höll tillsammans ett metodseminarium om kontingenstabellanalys i november 1978. Dokumentation:

G Gelin och H Lütjohann: Analys av kontingenstabeller.

- 1 Allmän beskrivning
 - 2 Kort om modellspecifiering, notation och ett allmänt exempel
 - 3 Ett artificiellt exempel
 - 4 Exempel på litteratur
- (Material till metodseminarium 1978-11-28)

I samarbete med Lars Jonung vid Nationalekonomiska institutionen i Lund har Gunnar Gelin inlett ett arbete på att tillämpa kontingenstabellanalys på data avseende inflationsförväntningar.

En litteraturförteckning över artiklar om kontingenstabellanalys har framställts, se Andersson (1978).

Det finns inom SCB arbete stora potentiella användningsmöjligheter för kontingenstabellanalys. Åtskilligt mer arbete än det, som har utförts inom analysprojektet, krävs dock innan metoden kan användas mer rutinmässigt. Åtminstone följande tre användningsområden är av intresse.

1 Kontingenstabellanalys kan användas för att ur multivariata kvalitativa material extrahera en sambandsstruktur, som vore svår att utläsa ur tabeller. Denna användning är deskriptiv. Den kan vara brukbar vid kommentering och publicering av data.

2 Kontingenstabellanalys kan användas för att bland många indelningsgrunder sortera fram dem, som uppvisar intressanta sambandsstrukturer inbördes. Denna användning är explorativ. Den kan vara brukbar vid val av tabellplan för publicering av data.

3 Kontingenstabellanalys kan användas i konsultationer inom och utom verket. Många av SCBs data, särskilt I-avdelningens, är av den karaktär som kontingenstabellanalysen förutsätter. De kan inte enkelt analyseras medelst äldre standardförfaranden såsom regressionsanalys. Denna tillämpning kan vara deskriptiv eller innehålla mer avancerad tolkning.

3.6.5 Metoder för säsongrensning

Säsongrensning innebär att man estimerar och korrigerar för den del av en månads- eller kvartalsseries variationer som har samband med årstidernas växlingar, infallandet av de olika högtiderna och andra under året periodiskt återkommande företeelser. Syftet med säsongrensning är att man vill eliminera säsongeffekten, för att därigenom lättare kunna observera andra komponenter i en serie såsom trenden och de slumpmässiga eller irreguljära förändringarna.

Under de senaste decennierna har metoder och datorprogram för säsongrensning utvecklats på många olika håll i världen. De program som varit aktuella för studium inom analysprojektet är Konjunkturinstitutets program SA4, Bureau of the Census program X11, en vid Statistics Canada utförd vidareutveckling av detta program, kallat X11 ARIMA samt ett program som utvecklats vid Central Statistical Office, UK. Det sistnämnda programmet benämns här CSO-programmet.

X11 ARIMA har förvärvats till SCB under augusti 1980 och har vid färdigställandet av denna rapport endast delvis utprovats. Detta program skiljer sig från X11 främst genom att det ger möjlighet till prognos av den ursprungliga serien medelst så kallade ARIMA-metoder (vilket i stort sett är det samma som Box-Jenkinsansatsen, se avsnitt 3.6.2). Säsongrensningsestimaten beräknas därefter från den med prognoserna förlängda serien. Detta förfarande ger erfarenhetsmässigt säkrare estimat på säsongkomponenterna vilket ger sig tillkänna i att revideringsfelen blir mindre när man tid efter annan gör förnyade skattningar av säsongkomponenterna.

Korrespondens med Central Statistical Office har gett vid handen att man numera använder sitt program endast för att undersöka olika serier med avseende på val av modell för säsongrensning (det är mest fråga om serier avseende sysselsättnings- och arbetslöshetsdata). Den löpande säsongrensningen utför man fortfarande med en vid CSO utvecklad variant av X11 där man har byggt in prognosförfaranden liknande dem i X11 ARIMA. De erfarenheter med CSO-programmet som gjorts inom analysprojektet leder till liknande slutsatser, nämligen att programmet har sin största tillämpning när det gäller val av modell för säsongrensning, se (Cassel 1978c).

I en tidigare undersökning (Cassel 1974) har bl.a. programmen X11 och SA4 jämförts med varandra för ett relativt begränsat datamaterial. Slutsatsen blev att utfallet av säsongrensning påverkades mera av modellvalet (multiplikativ eller additiv modell) än av valet av program (X11 eller SA4).

Teori och metodik för säsongrensning finns beskrivna i Cassel (1980). Denna PM måste dock omarbetas med avseende på de delar som berör

P/STM

1980-10-31

X11, till följd av tillkomsten av X11 ARIMA.

3.6.6 En alternativ metod för prognos av tidsserier med konjunktur eller säsongrensning av dessa

Följande metod skall ses som en mycket enkel metod att utnyttja extra information vid prognoser av tidsserier, speciellt kunskap/antaganden om framtida konjunkturförändringar. Metoden kan även användas för säsongrensning.

Någon uttömmande teoretisk utvärdering har inte gjorts, dock har vissa bitar ägnats speciellt intresse i synnerhet den bit som berör de sanna säsongkomponenterna. Vi tänker oss följande modell.

$$Y = S \cdot T \cdot K \cdot I$$

där

S är säsongen

T är trenden

K är konjunkturen

I är innovationen (slumpkomponenten)

Följande huvudkrav ställdes vid utvecklingsarbetet, där alla kraven samtidigt skall vara uppfyllda.

1 Metoden skall vara enkel att använda (i princip bara att "trycka på knappen").

2 Observationer i slutet av tidsserien skall kunna utnyttjas. (Av flera skäl anses de senaste observationerna innehålla mest information vid bedömningen av var man befinner sig i en konjunkturcykel).

3 Då säsong och konjunktur ständigt ändras både till värde och mönster, skall programmet kunna köras varje gång en tidsserie används.

4 Man skall enkelt kunna använda kunskap om framtida konjunkturutveckling eller kunna studera effekten av alternativa konjunkturantaganden.

Kraven innebär bl a följande:

- Man kan ej använda glidande medeltal vid bestämningen av säsongkomponenterna.

- T, S, I måste elimineras så att enbart K kan studeras.

P/STM

1980-10-31

Datorprogrammet som är skrivet i APL och är interaktivt innehåller följande steg.

STEG 1

Eliminering av trenden med hjälp av ett polynom av graden k eller med hjälp av en exponentiell trend.

STEG 2

Eliminering av säsongen med hjälp av bildandet av successiva kvoter (en modifierad variant av den $s-k$ link-relativemetoden används).

Det kan visas att link-relative-metoden inte alltid skattar de sanna säsongkomponenterna. För att lösa detta problem gör programmet iterationer över steg 1 och 2 tills säsongkomponenterna konvergerar.

STEG 3

Slumpen elimineras med glidande medeltal där en korrigeringsfaktor i slutet och början används så att inga observationer tappas. (Denna metod kan ej användas innan säsongen är eliminerad då den skulle påverka skattningarna av säsongkomponenterna.)

STEG 4

Konjunkturen identifieras och plottas/tabelleras.

STEG 5

Trenden och säsongkomponenterna tillfogas sedan och en prognos erhålles. De prognostiserade värdena innehåller ingen innovations(slump)faktor.

Vill man bara ha en säsongrensad serie läggs trenden på efter konvergens i steg 2 alternativt efter steg 3 beroende på om man vill ha med innovationerna eller inte.

Programmet har, som tidigare nämnts, syftet att vara användarvänligt, varför det inte är speciellt finess- eller nyansrikt. Bl a klarar det ej av plötsliga förändringar i trenden eller säsongmönstret. Det förutsätts att alla ej trendmässiga varianter som ej förklaras av de beräknade säsongkomponenterna antingen beror på konjunkturen eller slumpen. Dock klarar programmet av fallet med en successiv förskjutning av säsongkomponenternas storlek.

För en närmare beskrivning av programmet hänvisas till ett metodmeddelande om programmet (utkommer hösten 80).

3.7 Metodutvecklingsarbete i anslutning till SCBs
programbibliotek

I samband med projektgruppens kartläggning av analysbehovet vid SCB (se avsnitt 3.3) framfördes starka önskemål beträffande förbättringar av analys servicen på SCBs dator. För att tillmötesgå vissa av kraven har vi inom projektgruppen utarbetat förslag till kompletteringar av programbiblioteket, se t ex Holmén (1978) samt Norberg och Skough (1978). Delvis inom projektets ram har utarbetats ett subrutinpaket för framställning av färgdiagram (FDIA), se Andersson och Cassel (1978).

Under 1970-talet har internationellt lagts ner ett intensivt arbete på evalvering av statistisk programvara. För att tillgodose verkets behov av service avseende val mellan och användning av olika program och programpaket har denna verksamhet stöderats inom projektets ram. Detta har bl a skett via deltagande i COMPSTAT 78: 3rd Symposium on Computational Statistics i Leiden, se Holgersson och Borelius (1978) samt i symposiet "Computing and Statistics; 12th Annual Symposium on the Interface" i Waterloo 1979, se Holgersson (1979 b, 1980 a). Vidare har en uppföljning skett av verksamheten inom ett amerikanskt projekt som bedriver evalveringsstudier av statistisk programvara. Resultaten av dessa studier har dokumenterats i Holgersson (1980 c).

1980-10-31

4 Projektgruppens utvärdering

I riktlinjerna för analysprojektet (se bilaga 1) anges under punkt 9: "Avsikten är att projektarbetet genom metodutveckling, lansering av metoder och metodservice skall aktivt medverka till att SCBs statistik i ökad utsträckning förädlas (analyseras)". Vidare uppräknas under punkterna 9.1 - 9.8 ett antal arbetsuppgifter för analysprojektet.

Riktlinjerna har i stort sett följts. Metodutveckling har bedrivits på så sätt att olika analysmetoder har tillämpats på SCB-data och deras användbarhet diskuterats. Lansering av metoder har skett genom metodseminarier samt utgivande av handledningar och metodinformation. Metodservicen har påtagligt förbättrats, så tillvida att man anskaffat ett antal standardprogram och vunnit kunskap och erfarenhet av att använda dessa program för analys av SCB-data.

Har då SCBs statistik i ökad utsträckning förädlats? Kanske inte i den utsträckning man önskat. Några av orsakerna till detta är:

- i) Resursbrist på grund av att analysbehovet oftast inte förutses vid planeringen av produktionsprocessen.
- ii) Tidsbrist på grund av att man vill publicera resultat så snabbt som möjligt. Analys skulle medföra fördröjning.
- iii) Bristande metodologiska kunskaper hos ämnes- och analysansvariga.
- iv) Neutralitetsproblematiken. Man är här och var rädd för anklagelser att vid statistisk analys företräda vissa partsintressen.
- v) Bortprioritering av arbete i utvecklingsprojekt. Utvecklingsarbete av det slag som bedrivits inom analysprojektet bortprioriteras ofta av de vidtalade medarbetarna, till förmån för annat arbete, t ex konsultationsverksamhet, där tidspressen känns starkare. Projektarbetet blir lidande till följd av bristande personresurser och bristande kontinuitet.
- vi) Man kan också ifrågasätta vettigheten av den inriktning som analysprojektet från början givits. Ökad analysverksamhet förutsätter att de ämnesansvariga är ivriga att utvinna ökad information ur data. Först måste de angelägna frågorna formuleras. Därefter gäller det att hitta statistiska metoder som kan användas för att ge svar på dessa frågor. I analysprojektet har man i stor utsträckning gått den motsatta vägen: Man har först lanserat analysmetoder och därefter försökt hitta data lämpade att bearbetas med dessa metoder.

P/STM

1980-10-31

5 Förslag till fortsatt verksamhet

5.1 Löpande verksamhet

En förutsättning för framgångsrik analysverksamhet är att det finns angelägna saklogiska problem, som man genom analysen vill få behandlade. De ämnesansvariga måste mer än hittills uppmannas att (i samarbete med statistikanvändarna) ta fram sådana angelägna frågor, där SCB-data kan tänkas ha information att ge. Varje statistikgren bör själv ta initiativ till och skaffa utrymme för analys.

För att analysen sedan skall kunna utföras krävs vissa specialkunskaper i databehandling och statistisk analysmetodik, med andra ord en metodologisk beredskap. För att en sådan beredskap skall kunna hållas vid liv kan man tänka sig att ha en löpande produkt (kallad "statistisk analysberedskap" eller något sådant), se bilaga 3) som svarar för dels viss kostnadsfri rådgivning i analysfrågor, dels vidmakthållande, komplettering och samordning av de metodologiska kunskaperna på analysområdet. Det bör finnas en arbetsgrupp som ansvarar för denna produkt. Arbetsgruppen bör träffas regelbundet, och i dess arbetsuppgifter bör bl a ingå att anskaffa programvara samt att utarbeta förslag till satellitprojekt.

5.2 Satellitprojekt

Som komplettering till den ovan föreslagna arbetsgruppen föreslår vi vidare att det vid SCB fortlöpande initieras vad vi här betecknar som "satellitprojekt" med syfte att behandla olika problemområden inom statistisk analys. Dessa projekt skall vara kortvariga och ha en klar och konkret målsättning. Det problemområde som skall studeras bör vara tämligen väl avgränsat. Vaga projekt av typen: Statistisk analysverksamhet vid SCB, Studium av förändringsskattningar etc bör undvikas.

För övrigt kan satellitprojekt tänkas vara av olika natur, alltifrån empiriska studier knutna till aktuella undersökningar till mer teoretiska studier (som naturligtvis också måste vara nära knutna till den praktiska analysverksamheten vid SCB). Man bör i detta sammanhang om möjligt utnyttja möjligheter att etablera olika typer av samarbete mellan SCB och exempelvis universitetsinstitutioner, externa forsknings- och utredningsinstitut etc.

Ett projekt av detta slag som redan satts igång och till vilket analysprojektet varit en av initiativtagarna är "Tidsserieanalys inom företagsstatistiken". Arbetet inom detta projekt skall koncentreras till områdena säsongrensning, efterhandsprognoser och tidsseriebrott.

1980-10-31

En lista över problemområden lämpliga att studera inom olika satellitprojekt kan göras lång. Vi nämner här endast ett par exempel:

- Inom ämnesområdet "Olika inferensansatser vid analys av surveydata" återstår ännu mycket forsknings- och utvecklingsarbete. Ett tänkbart forskningsområde är studier av egenskaperna hos variansestimatorer i linjära modeller vid olika inferensansatser, gärna med anknytning till någon speciell statistikprodukt.
- Analys av surveydata som innehåller olika typer av bortfall är ett område där arbete endast hunnit påbörjas inom analysprojektet. Studier av hur olika typer av bortfall kan påverka analysresultat är av mycket stort intresse, men har hittills varit sällsynta i litteraturen. Vi föreslår därför att vid SCB igångsätts ett satellitprojekt, gärna med anknytning till någon speciell statistikprodukt, där denna problematik studeras - empiriskt och teoretiskt.
- I avsnitt 3.6.4 påpekas de stora potentiella användningsmöjligheterna för kontingenstabellsanalys av SCBs data. Vi föreslår därför igångsättande av ett satellitprojekt knutet till någon viss statistikprodukt, t ex någon av I/U:s intervjuundersökningar med syfte att utpröva metodikens användbarhet ur de tre aspekter som redovisas i avsnitt 3.6.4.

Förslag till satellitprojekt förutsätts i fortsättningen utarbetas av den ovan nämnda samarbetsgruppen för statistisk analysberedskap.

PROJEKTDOKUMENTATION

Slutrapport från Utredningen för metodfrågor vid statistisk analysverksamhet (SAMSAS-utredningen, 1973-05-02).

Service- och metodutveckling avseende den statistiska analysverksamheten vid SCB. Riktlinjer enligt GD-beslut 1974-07-01.

Andersson, A (1978) Litteraturförteckning över artiklar om kontingenstabeller (Promemorior från P/STM nr 2 1978-11-07) se även Förteckning över litteratur om kontingenstabeller (METODINFORMATION 79:1 1979-01-10 Delserie Statistiska Metoder).

Andersson, A, Cassel, PG (1978) FDIA - ett subrutinpaket i FORTRAN för framställning av färgdiagram (Rapport 1978-08-16).

Bergholm, P (1978) En förteckning över litteratur rörande statistisk analys i surveysituationer (PM 1978-08-01).

Berglund, L (1977a) Förslag till utvidgad presentation av SM:T "Varutransporter med lastbil" med grafisk framställning (PM 1977-08-09).

Berglund, L (1977b) Tidsserieanalys av data över lastbilstransporter 1972-1976 (PM 1977-08-15).

Cassel, C, Råbäck, G (1979) Projektet "Mätfel vid regressionsanalys" (Utkast, odaterat).

Cassel, PG (1974) Säsongrensning av arbetslöshetsdata från SCBs arbetskraftsundersökningar. (Utkast 1974-09-25).

Cassel PG (1978a) Säsongrensning och test av modell för säsongrensning med CSOs program - en jämförelse med andra program (Utkast 1978-03-12).

Cassel, PG (1978b) Säsongrensning av dödligheten i vissa dödsorsaker (material publicerat i Dödsorsaker 1977).

Cassel, PG (1978c) Säsongrensning av arbetslöshetsdata för perioden januari 1970 t o m december 1977 - test av metod att använda 1978. (PM 1978-02-21).

Cassel, PG (1979) Säsongrensning av arbetslöshetsdata för perioden januari 1970 t o m december 1979 - val av metod att användas 1979 (PM 1979-03-02).

Cassel, PG (1980) Handledning i säsongrensning (Utkast 1980-06-06).

Elofsson, G (1975) Multivariata statistiska analysmetoder. Tillämpning på material från den internationella inkomstfördelningsstudien, HINK 72 (PM 1975-09-03).

Elofsson, G (1977) Försök med analys av anställdas löner enligt inkomstfördelningsstudien 1972 (Rapport 1977-02-21).

Elofson, G, Larsson, U-B (1976) Anställdas löner studerade med multivariata analysmetoder (Bilaga 1 i Inkomstfördelningsstudien 1972; SOS, SCB).

Gelin G, (1978) Klusteranalys av befattningsbyten inom industri- och handelstjänstemannasektorn 1970-1975 (Rapport 1978-10-10).

Gelin, G (1980) Yrken som liknar varandra: klusteranalys av yrkesbyten i Sverige 1970-1975 (Rapport under utarbetande).

Gelin, G, Lütjohann, H (1978) Analys av kontingenstabeller: 1 Allmän beskrivning, 2 Kort om modellspecifiering, notation och ett allmänt exempel, 3 Ett artificiellt exempel, 4 Exempel på litteratur (material till metodseminarium 1978-11-28).

Hedström, L, Larsson, U-B (1975): En studie av bostadsområden i Göteborg och Norrköping med hjälp av faktor- och klusteranalys (metodrapport 1975-04-08).

Holgersson, M (1977a) För terminalanvändare vid STM: Användning av SPSS via GUTS-systemet (PM 1977-03-25).

Holgersson, M (1977b) Metoder för deskription och analys av data hos fyra olika programpaket för statistisk analys (SPSS, BMDP, OSIRIS och SAS76). Något om IMSL. (PM 1977-06-08).

Holgersson, M (1977c) Några synpunkter på Rättsstatistisk Årsbok 1975 avseende grafisk presentation och analys (PM 1977-10-30).

Holgersson, M (1978a) Analys av likheter och skillnader mellan tingsrätter m a p ådömda påföljder för viss typ av brott (Rapport 1978-08-02).

Holgersson, M (1978b) Regressionsanalys av sjukfrånvaron hos barn på daghem (PM 1978-12-18). Ingår som bilaga i SM S 1979:1 "Barns och personals närvaro och frånvaro i daghem 1977".

Holgersson, M (1979a) Kommunindelning för individualstatistiken: Faktoranalys som ett hjälpmedel vid stratifiering av kommuner (PM 1979-04-26). Ingår som bilaga nr 3 i Lundström, S (1979-06-20) Urvalsfrågor i individualstatistikprojektet.

Holgersson, M (1979b) Material från studieresa i USA april-maj 1979 (PM 1979-09-05).

Holgersson, M (1979c) Resursfördelning i statistikproduktionen. Material till avsnitt VI. Statistisk analys. (Kursmaterial 1979-11-08).

Holgersson, M (1980a) Studiebesök i USA och Kanada (Meddelande i Stat Tidskrift 1980:1).

Holgersson, M (1980b) Regression med och utan vägning med urvalssannolikheter. Ett numeriskt exempel (Utkast 1980-01-08).

Holgersson, M (1980c) Evalvering av statistisk programvara - en orientering om nuläget (Utkast 1980-03-10).

... Holgersson, M (1980d) Om analysverksamhet vid SCB på 1970-talet (Bidrag till SCBs forskarkonferens 1980).

... Holgersson, M, Borelius, J (1978) COMPSTAT 78: 3rd Symposium on Computational Statistics, (Meddelande, Statistisk Tidskrift 1978-09-14)

Holgersson, M, Norberg, A (1978) Rapport från studiebesök vid Statistisk Sentralbyrå i Oslo (Rapport 1978-01-18).

Holgersson, M, Norberg, A (1980) Extremvärdesbehandling hos säsongrensningssprogrammet X-11. En känslighetsanalys (Utkast 1980-02-27).

Holgersson, M, Ringnér, B, Wretman, J (1977) En kartläggning av behovet av statistisk analys och grafisk presentation av SCBs data, Intervjuer med handläggare vid ämnesenheterna (STM-rapport 1977-02-21).

Holgersson, M, Ringnér, B (1979) Analyser av hushållens el-energiförbrukning fr o m 1968 med Box-Jenkins metod (PM 1979-02-27).

Holmén, Å (1978) Till teknikergruppen PD/ADB angående inköp av programvara för tidsserieanalys enligt Box-Jenkins metoden (PM 1978-05-25).

Holmén, Å (1979a) En alternativ metod för säsongrensning (PM 1979-11-30).

Holmén, Å (1979b) En presentation av Box-Jenkins metod för analys och prognoser av tidsserier (Promemorior från P/STM nr 3 1979-12-20), se även En presentation av Box-Jenkins metod för prognoser av tidsserier (Metodinformation 79:5).

Holmén, Å (1980a) Box-Jenkins - analys. (Stat Tidskrift 1980:5)

Holmén, Å (1980b) Användning av Box-Jenkinsmetoden vid beräkning av det preliminära industriproduktionsindexet (Rapport under utarbetande).

Jonung, L (1980a) Perceived and expected rates of inflation. A note on Sweden. (Rapport, Nationalekonomiska institutionen, Lunds universitet, 1980).

Jonung, L (1980b) Uppfattad och förväntad inflation i Sverige. Bilaga utarbetad för prisregleringskommittén. (Rapport, Nationalekonomiska institutionen, Lunds universitet, feb 1980).

Jönrup, H (1975) Contribution to the theory of regression analysis in sample surveys. (I.S.I., Contributed papers, Warsaw 1975, 429-432).

Jönrup, H, Rennermalm, B (1974) Regressionsanalys vid urval från ändliga populationer (Metodinformation 74:7).

Jönrup, H, Rennermalm, B (1976) Regression Analysis in Samples from Finite Populations (Scand J Statist 1976:3).

P/STM

1980-10-31

Larsson, U-B (1976) Prov med klusteranalys vid framställning av en sysselsättningsnomenklatur (Rapport, mars 1976).

Lütjohann, H (1979a) Några grundläggande egenskaper hos linjär regressionsanalys. Anteckningar till kurs i elementär regressionsanalys 1979. (PM 1979-02-22).

Lütjohann, H (1979b) Ambitionsnivåer vid analys av data inom SCBs produktion (1979-10-09).

Lütjohann, H (1979c) En riskabel punkt i regressionsanalysen i kap 8 av rapport 15 från ULF (PM 1979-01-04).

Norberg, A (1977) Analys av IF 73 (Lägesrapport 1977-05-09).

Norberg, A (1978a) Ett dussin illustrationer av hushållens inkomster 1973 (STM-rapport 1978-01-18).

Norberg, A (1978b) Simuleringsstudie av regressionsestimat vid ändliga populationer, studie nr 1 (STM-rapport 1978-02-08).

Norberg, A (1978c) AID-analys av hushållens inkomster (Rapport 1978-08-29).

Norberg, A (1978d) Simuleringsstudier av regressionsestimat vid ändliga populationer (Metodinformation 78:4).

Norberg, A (1978e) Statistiska analysmetoder för SCB (Förslag 1978-09-15).

Norberg, A (1979a) Angående regressionsanalys för att påvisa eventuella bortfallseffekter i Sifos väljarbarometer (PM 1979-02-22).

Norberg, A (1979b) Att automatiskt upptäcka interaktioner. (PM 1979-09-05).

Norberg, A (1979c) Domstolsverkets fattigkonkurser: Ett prov med MCA-analys (Rapport nr 7 1979-12-04).

Norberg, A (1980a) Val av kritiskt värde på kriteriefunktionen vid AID-analys (PM utkommer 1980).

Norberg, A (1980b) Handledning i AID-analys (Promemorior från P/STM nr 4, utkommer 1980).

Norberg, A (1980c) Vädrrets effekt på elförbrukningen i hushåll, handel, m m (Stat Tidskrift 1980:3).

Norberg, A, Skough, A (1978) Standardprogrampaketet OSIRIS III version 2 (I PB-nytt 1978-05-09).

Norberg, A, Tillquist, B (1979) Uppföljning av kurserna i TAB68 och SPSS (PM 1979-03-12).

Norberg, A, Gouiedo, L (1980) Medieutlåning och arrangemang i kommunala bibliotekssystem - en sambandsanalys (Stat Tidskrift 1980:2).

P/STM

1980-10-31

Rennermalm, B, Arvidsson, A (1977) Utjämnning av svenska frukt-samhetskurvor 1968-1973 med kubiska splinefunktioner (Stat Tidskrift 1977:5).

Ringnér, B (1977) Anpassning av analysförfaranden till surveyundersökningar (Diskussionsunderlag vid SCBs kontaktkonferens i Grisslehamn, maj 1977).

Ringnér, B (1979) Att vikta eller inte vikta med regressions-sannolikheter vid analyser i surveysituationer (Arbetsmaterial 1979-10-09).

Ringnér, B (1980a) Analysis of Survey Data - Relations to Robust Regression and Experimental Design. (Draft 1980-05-08).

Ringnér, B (1980b) Inferens för regressionskoefficienter vid analys av surveydata. (Rapport under utarbetande).

Wretman, J (1978) Regressionsanalys vid stickprov från ändlig population (Utkast 1978-08-17).

Wretman, J (1980) Analys av surveydata (Utkast 1980-04-14).

ÖVRIGA REFERENSER

Regional fruktsamhetsstatistik för perioderna 1968-1970, 1971-1973, 1974-1976. SCB, Prognosinstitutet 1978.

Räkna med bortfall. En handbok om statistiska metoder i samband med bortfall. SCB 1980.

Balint, I. (1975) Analysis and scientific activity in the Hungarian central statistical office. Bull. I.S.I., 40:1, 448-456.

Boreham, A.I. (1975) How far should and would those who produce statistics engage in research and analysis? Bull. I.S.I., 40:1, 433-444.

Duncan, I.W., ed. (1978) Statistical services in ten years' time. Oxford: Pergamon Press.

Esberger, S.E. (1975a) Översyn av areell statistik. PM nr 3: Statistisk analys (1975-04-28).

Esberger, S.E. (1975b) Översyn av areell statistik. PM nr 4: Prognoser (1975-04-28).

Frane, I.W. (1978) Missing data and BMPD: Some pragmatic approaches. American Statistical Association. 1978. Proceedings of the Statistical Computing Section, 27-33.

Heiberger, R.M. (1977) Regression with the pairwise-present covariance matrix: A dangerous practice. University of Pennsylvania, Department of Statistics Tech Report No. 19.

Hoem, I.M. och Rennermalm, B. (1977). On the statistical theory of graduation by splines. Working paper No 14, Inst för försäkringsmatematik, Köpenhamns universitet.

Hoem, I.M. et al (1980) Experiments in modelling recent Danish fertility curves. Working paper No 29, Inst för försäkringsmatematik, Köpenhamns universitet.

Jorner, U. (1980a) Analys av lantbruksstatistik. Bidrag till SCBs forskarkonferens 1980 (1980-05-14).

Jorner, U. (1980b) Uppföljningen av utvecklingslinjer i SCBs lantbruksstatistik: Utökad analysverksamhet i anslutning till lantbruksstatistiken (Utkast okt 1980).

Kim, I.O. och Curry, I. (1977) The treatment of missing data in multivariate analysis. Sociological Methods and Research, 6, 215-240.

Lundström, S. (1979) Urvalsfrågor i individualstatistikprojektet (PM 1979-06-30).

Medin, K. (1975) Vägledande synpunkter för A-avdelningen vid framställningen av tidsserier och regional statistik samt vid statistisk analys- och prognosverksamhet. (A-meddelande 1975:14).

Norwood, J.L. (1975) Should those who produce statistics analyze them? How far should analysis go? An American view. Bull. I.S.I., 40:1, 420-432.

Olofsson, P.O. (1974) Former och metoder för projektarbete vid SCB. (Sammanfattning av utredningsresultat 1974-01-22 samt delrapporter I-IV).

Rennermalm, B. (1979) Utjämning av demografiska kvoter med kubiska splinefunktioner (Stat. Tidskrift 1979:6).

Rennermalm, B. (1980) Biasskattning för regressionskoefficienter i design-baserad inferens från ändliga populationer (Stat. Tidskrift 1980:2).

Wahlén, J. (1980) Rita diagram. Råd och tips om konstruktion och ritning. Andra utgåvan. SCB 1980.

Widén, M.L. (1975) Översyn av areell statistik. PM nr 2: Tidsserier och regional statistik (1975-04-28).

1974-07-01

Service- och metodutveckling avseende den statistiska analysverksamheten vid SCB. Riktlinjer fastställda genom GD-beslut 1974-07-01 (p 2704)

1 Den konsumentorienterade statistiska analysverksamheten bedrivs på ett flertal enheter vid SCB, bl a vid P/STM, U/UI, I/PI och I/BE. Även ämnesenheterna i övrigt ägnar sig i varierande, men vanligen mycket begränsad, utsträckning åt dylik statistisk analysverksamhet i anslutning till det löpande produktionsarbetet.

När det i fortsättningen talas om "statistisk analysverksamhet" (eller motsvarande) är det konsumentorienterad analys som avses.

I tidigare gällande organisation fanns den s k analysgruppen knuten till sektionen för statistiska metoder vid planerings- och samordningsavdelningen. Analysgruppen hade som uppgift att svara för metods-service och utvecklingsarbete på detta område. Verksamheten var speciellt inriktad på utnyttjande av ADE som hjälpmedel.

Vid övergången till den nya organisationen den 1 juli 1971 upplöstes analysgruppen såsom fristående organisatorisk enhet. Den till P-avdelningen knutna uppdragsverksamheten vad avser statistisk analys har sedan dess bedrivits i projektform på samma sätt som huvuddelen av verksamheten i övrigt vid avdelningen.

Statistiska analyser har vid SCB innefattat såväl statistisk metodtillämpning som utpräglad ADB-verksamhet, t ex system- och programmeringsverksamhet. Vidare har den omfattat både intern och extern uppdragsverksamhet. Metodutvecklingsarbetet har omfattat viss programutveckling.

2 Våren 1972 tillsattes en utredning, som antog arbetsnamnet SAMSAS, med uppgift bl a att ange och beskriva metod- och serviceuppgifterna på det statistiska analysområdet och att utarbeta förslag till arbetsfördelning och samordning mellan vissa berörda enheter vid verket. Utredningen avlämnade i maj 1973 en slutrapport (1973-05-02).

3 SAMSAS har definierat begreppet "statistisk analys" såsom "studium av samband mellan variabler med hjälp av vissa statistiska metoder". I den statistiska analysverksamheten anses även ingå både metodutveckling, t ex framställning av nya datorprogram, och metodplanering, dvs val och anpassning av metoder i enskilda analysuppdrag (av både intern och extern natur).

Med den avgränsning som sålunda gjorts har utredningen behandlat frågor rörande behov av och efterfrågan på analys-service, tillgången på och behovet av resurser härför och behovet av metodutveckling på området.

4 Med utgångspunkt från bl a SAMSAS förslag har definierats två projekt, nämligen "Uppbyggnad av programbibliotek" knutet till P/DEM och "Statistiska metodfrågor i anslutning till SCBs programbibliotek" knutet till P/STM. Det sistnämnda är fort-

1974-07-01

farande under förplanering medan i det förra programinventering m m pågår.

För budgetåret 1974/75 har SCB inom ramen för löpande verksamhet tilldelats vissa medel för uppbyggnad av och verksamhet vid ett centralt programbibliotek, bl a innehållande standardprogram för statistisk analys, ett fortsatt datatekniskt utvecklingsarbete och fortsatta studier vad avser utnyttjandet av dylika standardprogram ur statistisk metodologiska aspekter. Vidare har i verkets förslag till Kungl Maj:t föreslagits att under 1974/75 utvecklingsmedel skall få disponeras för arbetet med uppbyggnaden av programbiblioteket.

5 Utredningsförslagen behandlades i verksdirektionen den 12 november 1973. Därvid diskuterades utredningens definition av analysverksamheten ifråga. Det framhölls att statistisk analys i vid bemärkelse innefattar val och anpassning av analysmodeller och -metoder, tolkning av de statistiska resultaten, att man drar slutsatser, utarbetar rekommendationer etc på grundval av dessa. Detta förutsätter bl a en nära samverkan med uppdragsgivarna/statistikkonsumenterna. En slutsats är att statistisk analys i vidare mening bör vara en uppgift för ämnesenheter m fl som har direktkontakt med SCBs avnämare.

I diskussionen underströks önskvärdheten av att SCB ökar sina möjligheter att ge analyservice åt forskare och andra intressenter. Utnyttjande av standardprogram för analys underlättar och förbilligar verksamheten. Vikten av att det föreslagna programbiblioteket kommer till stånd poängterades i sammanhanget.

Man diskuterade också SAMSAS förslag till organisation av verksamheten. Enligt SAMSAS borde inrättas dels en projektgrupp, dels en styrgrupp till denna. Diskussionen gällde främst den föreslagna styrgruppen. Å ena sidan ifrågasattes behovet av en dylik grupp. Man menade att den föreslagna projektgruppen borde ges hela ansvaret för uppgifterna ifråga "enligt totalprojektmodell". Å andra sidan pekade man på behovet av förankring av verksamheten hos ämnesenheterna - vilket utredningen ansett motivera en styrgrupp för projektet.

Ärendet har sedermera ytterligare diskuterats i verksdirektionen den 18 februari 1974. Tyngdpunkten i diskussionen lades därvid på de organisatoriska frågorna. Det följande förslaget är utformat under beaktande av vad som framkom vid verksdirektionens båda sammanträden.

6 Inom ramen för den nyligen slutförda utredningen om former och metoder för projektarbete har framlagts vissa förslag beträffande projektorganisation och arbetsformer i projektarbetet vid SCB (se Delrapport III, 1973-11-15). Däri har bl a föreslagits ytterligare åtgärder i syfte att lansera totalprojektmodellen. Bl a har föreslagits att modellen (ytterligare) prövas i organiserade former och att det sker främst i nybildade projektgrupper. Totalprojektmodellen för samverkan innebär bl a att projektarbetet skall bedrivas i sådana former att de som deltar känner ett gemensamt ansvar för genomförandet.

1974-07-01

Inom ramen för utredningen har inventerats och redovisats olika "modeller" för organisationen av projektarbete, alternativa arbetsformer etc. För- och nackdelar har delvis diskuterats. Därvid har bl a påpekats vissa nackdelar med styrgrupper och mot den bakgrunden förordats en viss restriktivitet med inrättande av sådana grupper i projektverksamheten vid SCB.

7 Med hänsyn till bl a problemens svårighetsgrad när det gäller metodutvecklingen på det statistiska analysområdet, det nära sambandet mellan statistiska metoder och databehandlingsmetoder och behovet av erfarenhetsbyte när det gäller det statistiska analysarbetet har SAMSAS föreslagit att en projektgrupp tillsätts för att rikta och samordna verksamheten.

Förslaget synes närmast innebära en slags samarbetsgrupp med vissa övergripande funktioner, bl a "verka för ökad användning av analys", "biträda vid fördelning av serviceuppgifter mellan enheterna", "biträda vid prioritering bland metodutvecklingsprojekt" och "medverka till samordning" av dylika projekt.

8 Service- och metodutveckling avseende den statistiska analysverksamheten torde som regel förutsätta medverkan av flera funktioner/funktionella enheter. Detta talar för att verksamheten bedrivs enligt totalprojektmodell. Totalprojektmodellen är visserligen främst avsedd för projektarbete (som avser utveckling av produkter, tjänster eller metoder), men synes tillämplig även för viss löpande verksamhet. I det aktuella fallet ingår såsom framgår av förslagen i det följande både utvecklingsarbete och löpande service.

Det föreslås att den av SAMSAS föreslagna verksamheten organiseras enligt denna modell inom ramen för den i p 6 nämnda försöksverksamheten.

En viss precisering och konkretisering av projektgruppens uppgifter jämfört med SAMSAS förslag är därvid emellertid påkallad. I det följande lämnas riktlinjer beträffande inriktningen, omfattningen och organisationen av arbetet.

9 Syftet med projektet är att skapa en organisation för att få ett samordnande grepp över metodutvecklingen inom det statistiska analysområdet och att bemästra samarbetsproblem inom nämnda område vid SCB. Tyngdpunkten i projektarbetet skall ligga på den statistiska metodsidan. Projektet skall dock vara förankrat i ämnesenheterens problemområden (huvudsakligen den löpande statistikproduktionen). Avsikten är att projektarbetet genom metodutveckling, lansering av metoder och metods-service skall aktivt medverka till att SCBs statistik i ökad utsträckning förädlas (analyseras).

I enlighet med SAMSAS förslag tillsätts en projektgrupp bestående av fem personer, nämligen en representant för vardera P/STM, P/DRM, I/PI, D/PC och U/UI. Till gruppen kan tillfälligt knytas representanter även för ämnesavdelningarna och U/UC. Projektgruppen inleder sin verksamhet så snart som frågan om vilka personer som skall ingå i gruppen har lösts. Huvudman för projektet är P/STM. Till projektet knyts också en referensgrupp (se p 10

1974-07-01

nedan). Projektet benämns "Statistisk analysverksamhet vid SCB".

När det gäller såväl uppdrags- som anslagsfinansierad analyservice skall projektgruppen

9.1 i samverkan med ämnesavdelningarna och U/UC ta fram och å jour-hålla en behovskatalog vad avser statistiska metoder och programvara inom området för analyservice,

9.2 verka för ökad användning av analys genom att i samråd med ämnesavdelningarna ta initiativ till och i form av pilotprojekt genomföra analysarbeten i anslutning till SCBs reguljära statistikproduktion,

9.3 ansvara för genomförandet av interna och externa analysuppdrag som kommer till någon av enheterna P/STM, P/DEM, I/PI, D/PC och U/UI (se ytterligare förklaring nedan),

9.4 informera om och även på annat sätt främja utnyttjandet av SCBs programbibliotek (se p 9.7 nedan) såväl internt som externt.

Den i p 9.1 avsedda behovskatalogen skall tjäna som underlag dels för konsumentkontakter och andra åtgärder i syfte att utveckla servicefunktionen (särskilt de i p 9.7 nämnda utvecklingsarbetena), dels - tillsammans med den i p 9.5 nedan nämnda problemkatalogen - för metodutveckling på området.

P 9.3 innebär att analysuppdragen skall genomföras (som delprojekt) inom projektgruppens ram enligt totalprojektmodell. Det förutsätts härvid att den arbetsfördelning som behöver göras inom gruppen skall ske med hänsyn till uppdragens art, det specifika behovet av statistisk metodologiskt och ADB-tekniskt kunnande, ledig kapacitet m fl faktorer, som kan bedömas från fall till fall. Det synes praktiskt och lämpligt att uppdragen utförs av undergrupper till projektgruppen (eller i enstaka fall kanske av en enskild utredare). Även andra personer än de som (primärt) är deltagare i gruppen (jfr ovan) kan givetvis tänkas delta i genomförandet av uppdragen (eventuellt också i utvecklingsarbetet som gruppen skall bedriva). De närmare formerna för hur projektgruppen skall organisera och bedriva verksamheten skall gruppen avgöra själv. Givetvis skall flexibla och effektiva former eftersträvas.

När det gäller metodutvecklingen för analysverksamheten skall projektgruppen

9.5 upprätta och å jour-hålla en problemkatalog för metodutvecklingsarbetet inom det statistiska analysområdet vid SCB,

9.6 ta initiativ till, förplanera och samordna metodutvecklingsprojekt inom området.

9.7 Det tidigare vid P/STM initierade projektet "Statistiska metodfrågor i anslutning till SCBs programbibliotek" skall ingå som en del av projektet "Statistisk analysverksamhet vid SCB". Arbetet omfattar bl a handläggning av de statistiska metodfrågor som blir aktuella i det vid P/DEM bedrivna projektet "Uppbyggnad av programbibliotek".

1974-07-01

Vidare skall projektgruppen

9.8 i samband med SCBs årliga anslagsframställning avge förslag beträffande inriktningen och omfattningen av metodutvecklingsarbetet på det statistiska analysområdet samt beträffande omfattningen av analyserviceverksamheten för nästkommande budgetår (budgetförslag).

Projektet skall finansieras med dels för ändamålet, inom P-avdelningen, tidigare tillgängliga resurser, dels de medel som enligt vad som framgår av p 4 ovan anslagits för verksamheten under 1974/75, dels också intäkter av serviceverksamheten. Det ankommer på projektgruppen att planera verksamheten inom ifrågavarande givna ramar.

10 Vid sidan av projektgruppen inrättas som nämnts en referensgrupp för att tillgodose behovet av förankring av projektet i ämnesavdelningarna. Referensgruppen skall bestå av cheferna för STM (ordf), DEM och PC samt en representant från vardera A, I, F och U. Utöver dessa ordinarie ledamöter kan gruppen vid behov adjungera andra personer. Referensgruppen skall löpande informeras om arbetet inom projektet. Ämnesavdelningarna ges på detta sätt möjlighet att påverka projektets inriktning så att arbetet effektivt inriktas på att vidga möjligheterna till analys av SCBs material.

11 Varje metodplaneringsarbete liksom den i p 9.7 nämnda arbetsuppgiften skall betraktas som delprojekt inom ramen för projektet "Statistisk analysverksamhet vid SCB". För att debiteringsrutinerna skall fungera gäller att varje metodplaneringsarbete skall ha en separat kostnadsbärare (på kontoplan 6) med egen budget.

12 Verksamheten skall tills vidare utgöra ett försök. Försöksperioden skall omfatta två och ett halvt år, dvs t o m utgången av kalenderåret 1976. Till den 31 december 1976 skall verksamheten i projektet utvärderas beträffande projektorganisation och arbetsformer. I utvärderingen skall också ingå en kartläggning av omfattningen av ämnesenheterens egen analysverksamhet under försöksperioden. Projektgruppen skall ge underlag för denna utvärdering. Ställning skall därefter tas till formerna för ifrågavarande statistiska analysverksamhet även på längre sikt. Chefen för STM skall utarbeta förslag till principer för utvärderingen.

Förteckning över de personer som under någon tidsperiod varit medlemmar av projekt- eller referensgrupp (kronologisk ordning)

Projektgruppen:

P/STM:Bo Rennermalm (proj.ledare), Ulla Brith Larsson, Gunilla Elofsson (proj. ledare), Anders Norberg, Lars Hedstöm, Margareta Holgersson (proj. ledare), Bengt Ringnér, Harald Andersson, Per Gunnar Cassel Åke Holmén, Jan Wretman, Harry Lütjohann, Catarina Elffors.

P/DBM:Lars Nordbäck, Jonas Borelius, Hans Block

S/SYS:Staffan Nordvall, Christer Sandin

I/PI: Ulf Strömqvist, Harry Lütjohann, Gunilla Elofsson, Gunnar Gelin

I/UI: Håkan Lindström, Jan Wretman, Claes Cassel, Göran Råbäck

S/UC: Hans Dertell (adj.)

Referensgruppen:

P/STM:Per Olof Olofsson (ordförande)

P/DBM:Bo Sundgren, Lars Olsson

S/SYS:Bertil Georgson, Birgitta Lagerlöf

A/ST: Sven-Erik Esberger, Sture Lundin

F/ST: Gabriel Thulin

I/ST: Stig Lundeborg

S/UC: Dag Hallberg, Elisabeth Möller, Britt-Marie Ringmo-Ösgård

Budgetutfall åren 74/75 -- 79/80 (tkr)¹

	74/75	75/76	76/77	77/78	78/79	79/80
Analysprojektet	19	97	244	396	382	620
Metodfrågor i ansl. till SCBs programbibliotek	43	61	47			
	62	158	291			

¹ Inkl kostnader för datorkörningar enligt den taxesättning som tillämpas internt vid SCB.

Produktbeskrivning

Produkt Statistisk analysberedskap	Förslaget innebär ☒ ny produkt ☐ förändring av ☐ existerande produkt	Delprogram 14 Produktansvarig: Avd/Enhet P/STM	Nummer 1.2
---------------------------------------	--	---	---------------

Innebörden av förslagen samt motiv till förändringar, kostnader och prioritet

Bakgrund och förslag. SCB har gjort en utredning angående statistisk analysverksamhet vid verket. Viss försöksverksamhet har genomförts med syfte att öka användningen av statistiska analysmetoder vid SCB. Detta arbete har visat följande:

- Det föreligger ett uttalat behov av statistisk analys i anslutning till statistikproduktionen och uppdragsverksamheten. Det förekommer dessutom ett betydande latent sådant behov.
- Det behövs vid SCB en kontinuerlig bevakning av utvecklingen inom statistisk analys och programvara för statistisk analys, så att metodkompetens alltid finns tillgänglig inom verket.

Man kan dessutom förutsätta att efterfrågan på analys av data som publiceras eller eljest finns lagrade vid SCB kommer att öka i framtiden. Som exempel kan nämnas den via databaser ökade tillgängligheten av data på regional nivå samt tidsseriedata för bl a statliga myndigheter, större företag och näringslivets organisationer.

SCB föreslår med hänvisning till de resultat som uppnåtts att medel anslås till löpande statistisk analysverksamhet vid SCB.

Ändamål. Syftet med analysverksamheten är att löpande tillvarata och i ökad omfattning utnyttja metoder och tekniker, inklusive standardprogram, för statistisk analys och grafisk presentation. Detta är ett viktigt inslag i SCBs strävanden att rationalisera statistikproduktionen och bättre anpassa statistikprodukterna till användarnas behov.

Tillämpningen av statistiska analysmetoder ställer stora krav på såväl statistiskt-tekniskt kunnande som erfarenhetsunderlag. Det är därför viktigt att vid SCB utveckla och vidmakthålla en "pool" av erfarenheter och kunnande på analysområdet.

Exempel på användning av statistisk analysmetodik inom statistikproduktionen:

- vid planering av urvalsundersökningar kan statistiska analysmetoder användas som hjälpmedel vid stratifiering
- vid val av korstabeller för publikation kan analysmetoder användas, såväl för val av klassindelningar som för att söka lämpliga kombinationer av variabler/indelningsgrunder

Exempel på användning i syfte att ge konsumenterna mer information:

- tidsserier kan beskrivas och analyseras genom att man delar upp deras förlopp i trend och säsongvariation

Produktbeskrivning

Produkt	Förslaget innebär	Delprogram	Nummer
Statistisk analysberedskap	☒ ny produkt ☐ förändring av existerande produkt	14	1.2
		Produktansvarig	Avd/Enhet
		P/STM	

Innebörden av förslagen samt motiv till förändringar, kostnader och prioritet

- linjär och icke-linjär samvariation mellan variabler kan analyseras och beskrivas; särskilt samvariation mellan fler variabler än som lämpligen kan sammanföras i en publicerad korstabell
- grafisk framställning kan användas för att förmedla information om strukturer i data

Exempel på användning för att ge konsumenter och uppdragsgivare bättre beslutsunderlag:

- då uttalade frågeställningar finns, kan regelrätta estimations- och hypotesprövningsförfaranden användas
- prognoser kan framtagas med stöd i tidsserieanalys

Statistiska analysmetoder är alltså viktiga rationaliseringsinstrument, dels i statistikproduktionen, dels i konsumentens beslutssituation.

Huvudkonsumenter. SCBs statistikproduktion, statistikonsumenter, externa uppdragsgivare.

Konsumentkontakter. Intervjuer och diskussioner med ämnesavdelningarna och statistikonsumenter för att kartlägga och stimulera intresset för analys. Konkreta analysstudier i samarbete med produktansvariga och statistikonsumenter.

Huvudinnehåll. SCBs förslag innebär att arbete med en systematisk lansering av analysmetoder skall bedrivas löpande. Underlag för denna lansering är de studier, både metodstudier och konkreta analysstudier, som genomförts inom ramen för utredningsarbetet. Lanseringsmöjligheter är exempelvis:

- lansering genom analys av statistikprodukter vid SCB (pilotstudier)
- lansering genom seminarieverksamhet, kurser, konferenser etc
- lansering genom utarbetande och fortlöpande uppdaterande av handledningar i statistisk analysmetodik med exempel baserade på SCBs produktion och vid SCB befintlig programvara
- lansering genom artiklar i Metodinformation, Statistisk Tidskrift, PB-nytt m fl tidskrifter och informationsblad.

Som grund för lanseringen av statistisk analys enligt ovan fordras även följande verksamheter (som kan bedrivas löpande eller i projektform):

- studier av analysmetoder tillämpade på SCB-data för att lösa metodproblem som ej finns tillräckligt behandlade i litteraturen
- bevakning av internationell verksamhet beträffande evalvering av programvara för statistisk analys. En samordnad kartläggning av behoven av nya analysprogram i SCBs programbibliotek
- bevakning av litteratur inom området statistiska analysmetoder. Detta område är p g a ADB-teknologins utveckling statt i snabb expansion.

Kostnader. För det löpande arbetet med statistisk analysverksamhet beräknas för 1981/82 ett ökat medelsbehov av 100 000 kr. (U1)