

**Forskningsprojektet  
FEL I UNDERSÖKNINGAR  
Rapport nr 60**

**MODELLBASERADE SURVEY-FÖRFARANDEN  
av  
Tore Dalenius och Anders Klevmarken  
Stockholms universitet**

**Stockholms universitet  
Statistiska institutionen  
Stockholm 23**

**30 mars 1973**

## MODELLBASERADE SURVEY-FÖRFARANDEN

av

Tore Dalenius och Anders Klevmarcken  
Stockholms universitet0. Rapportens tema och syfte

Ett viktigt inslag i den utveckling, som teorin och metodiken för survey sampling undergått efter andra världskrigets slut utgörs av förfaranden för att effektivt utnyttja tillgänglig à prioriinformation.

Den under 1960-talet livliga debatten om de teoretiska grundvalarna för survey sampling kan - i vissa delar - ses som en debatt om hur ett visst slag av à prioriinformation (representerad av urvalsenheternas "labels") bör utnyttjas.

När det gäller survey sampling är möjligheterna att effektivt utnyttja à prioriinformation icke begränsad till inferensteorin; som kortfattat diskuteras i Dalenius (1971) finns det "i praktiskt taget varje undersöknings-situation en mängd information, som kan utnyttjas för konstruktioner av en bättre samplingplan än som eljest kan konstrueras", (a.o. sid. 34-35).

Om den tillgängliga à prioriinformation är omfattande, ligger det nära till hands att utnyttja den för konstruktion av en modell, som ges en instrumental roll i samplingplanen. Det primära syftet med denna rapport är att skapa underlag <sup>1)</sup> för forskning och utvecklingsarbete rörande just nämnda sätt att utnyttja à prioriinformation för konstruktion av vad vi vill kalla modellbaserade survey-förfaranden.

---

1)

Jämför förordet till rapport nr 53.

Vi söker förverkliga detta syfte genom att diskutera två tänkbara tillämpningsområden, nämligen:

- i) mätning och estimation vid surveys; och
- ii) evalvering av surveys.

I kapitel I gives en mycket schematisk redogörelse för det första av dessa tillämpningsområden. Tyngdpunkten i rapporten ligger vid möjligheterna att omvända modellbaserad evalvering, vilket behandlas i kapitel II.

## I MODELLBASERAD MÄTNING OCH ESTIMATION

Vi skall diskutera möjligheterna att använda modellbaserade förfaranden i anslutning till tre exempel:

- i) indirekt mätning;
- ii) reduktion av frågeformulär; och
- iii) estimation.

Dessa exempel företer sinsemellan vissa metodologiska likheter; vi diskuterar emellertid exemplen oberoende av varandra.

### 1. Indirekt mätning.

I framställningar av "sannolikhetskalkylen och några av dess tillämpningar" diskuteras två slags mätningar: direkta och indirekta. Vid direkt mätning observeras storheterna  $m_1 \dots m_k$ ; vägning av föremål för att bestämma deras vikt är ett exempel. Vid indirekt mätning observeras vissa funktioner av dessa storheter; på grundval av de sålunda erhållna värdena beräknas motsvarande värden på storheterna  $m_1 \dots m_k$ . Indirekt mätning förekommer bl a inom geodesin.<sup>1)</sup>

Det är näraliggande att söka tillämpa ett med indirekt mätning analogt framgångssätt i vissa surveysituationer. Antag att vi vill göra observationer på en variabel Y. Erfarenheten har visat att detta är möjligt endast med mycket stora mätfel (eller till mycket hög kostnad). Det är däremot möjligt att observera en variabel X med ringa mätfel (eller till låg kostnad).

Om relationen mellan Y och X icke är känd, kan vi överväga att tillgripa tvåfasurval. Om emellertid nämnda relation är känd, är det naturligt att söka utnyttja detta förhållande.

---

1) Inom bl a ekonometrin, psykometrin och sociometrin<sup>2</sup> tillämpas en metodik som företer stora likheter med indirekt mätning. Inom dessa områden är det vanligt att man använder modeller i vilka ingår storheter, som inte är direkt mätbara - ibland kallade latent variabler - men som genom reduktion av modellen kan mätas indirekt. Goldberger rapporterar t ex från en konferens om "Structural Equation Models" vid University of Wisconsin - Madison år 1970: "..... many of the models contained hypothetical constructs which, while not directly observed, had operational implications for relationships among observable variables." (Goldberger (1972)).

Det är måhända överväganden av detta slag som ligger till grund för utvecklingen av vissa nya metoder för att beräkna tidningars räckvidd. För en översikt av denna utveckling hänvisar vi till Frank (1968); vi nöjer oss här med att kortfattat antyda en metod.

En annonsör behöver för sin mediaplanering en uppskattning av den genomsnittliga ackumulerade räckvidden  $C(m)$  av  $m$  konsekutiva nummer av en viss tidning. För stora värden på  $m$  är insamling av observationer rörande räckvidden icke - med hänsyn till kostnader och mätfel - en praktiskt framkomlig väg. I Agostini (1962) redovisas en modell, som gör det möjligt att estimeras  $C(m)$  på grundval av observationer rörande endast två nummer. Modellen förutsätter sålunda att

$$C(m) = C(m-1) + \left[ 1 - C(m-1) \frac{a}{m^b} \right]$$

Parametrarna  $a$  och  $b$  bestäms med hjälp av följande relationer:

- i) för  $m = 1$  fås  $a = C(1)$ ;
- ii) för  $m = 2$  fås

$$\frac{C(2) - C(1)}{1 - C(1)} = \frac{C(1)}{2^b} .$$

## 2. Reduktion av frågeformulär.

Det är vid surveys ofta nödvändigt att av hänsyn till uppgiftslämnarna och uppgifternas kvalitet begränsa uppgiftsinsamlingen i större grad än som kan förefalla önskvärt med hänsyn till konsumenternas önskemål.

Antag att det vid en folkräkning gäller att insamla observationer rörande en viktig variabel  $V$  och vissa övriga variabler  $X$ ,  $Y$ ,  $Z$  och  $T$ . En i sådant fall ofta använd möjlighet är att insamla observationer på  $V$  från hela populationen och observationer på  $X$ ,  $Y$ ,  $Z$  och  $T$  från ett sample.

Ett alternativt förfarande är att från alla enheter i populationen insamla uppgifter om dels  $V$ , dels en kombination av 2 av de 4 övriga variablerna; var och en av de 6 möjliga kombinationerna användes inom ramen för något lämpligt design. Programmet skulle alltså (med användningen av ett design för systematiskt val av kombination) se ut sålunda:

Enhet	V	X	Y	Z	T
1	✓	✓	✓		
2	✓	✓		✓	
3	✓	✓			✓
4	✓		✓	✓	
5	✓		✓		✓
6	✓			✓	✓
.					
.					

Detta program har i jämförelse med det tidigare redovisade den nackdelen, att det försvårar multivariat analys på grundval av alla fem variablerna. Å andra sidan har det vissa förtjänster. Vi nöjer oss här med att peka på följande förtjänster:

- (1) Den omständigheten, att man vid detta förfarande observerar färre variabler (t ex endast V, X och Y) än vid en samplingundersökning (då V, X, Y, Z och T observeras) kan förväntas främja observationernas kvalitet <sup>1)</sup>.
- (2) Det alternativa förfarandet är "rättvist" i den meningen, att alla enheter får medverka i samma utsträckning; alla får lämna uppgifter om tre variabler, icke en eller fem; och
- (3) Det alternativa förfarandet kan i vissa fall ge ett bättre skydd än vad fallet är i det första programmet mot intrång i privatlivet - ingen enhet behöver lämna en fullständig redovisning!

Även om man icke känner relationerna mellan de här aktuella variablerna V, X, ... T, är det möjligt att med tillgång till observationer insamlade enligt det alternativa förfarandet estimerar t ex totaler och medelvärden. Förutom enkla uppräkningsestimater kan bli kvotestimater (med användning av V som hjälpvariabel) ifrågakomma.

Om man emellertid känner ovan nämnda relationer, bör det emellertid vara möjligt att konstruera bättre estimater. En möjlighet är sålunda att utnyttja

<sup>1)</sup> I ett realistiskt fall kan den reduktion av frågeformuläret, som sålunda äger rum, vara mycket betydande.

det i Dalenius (1962) angivna förfarandet. En annan möjlighet är att estimeras t ex  $\bar{X}$  med estimat av typen

$$\hat{X} = C_v \bar{v} + C_x \bar{x} + C_y \bar{y} + C_z \bar{z} + C_t \bar{t}$$

enligt diskussionen i Gurney och Dalenius (odat), till vilken vi också refererar för en utförligare diskussion av problemet med reduktion av frågeformulär.

### 3. Estimation.

Vi ger två exempel.

Ex. 1 Det kan vid en survey bli aktuellt att försöka bestämma variabelvärdet för en urvalsenhet, som icke ingår i urvalet. En illustration av detta fall diskuteras i Hansen et al (1953), vol I, s. 483, som också anvisar en regressionsteknik att använda i detta sammanhang.

Ex. 2 I Amundsen (1960) redovisar författaren argument för att vid 'reduktion' av ekonomiska data för analys välja metoder härför med beaktande av ekonometriska synpunkter. Tillämpat på en hushållsbudgetundersökning leder detta synsätt till att basera estimationen av t ex efterfrågan (konsumtionen) på en regressionsmetod, som speglar relationen mellan efterfrågan och dess förklaringsvariabler.

## II MODELLBASERAD EVALVERING

### 4. En översikt av evalveringsförfaranden.

Statistikerna har länge insett värdet - för konsumenter såväl som producenter - av att kritiskt granska resultaten av statistiska undersökningar ur synpunkten av deras tillförlitlighet. För detta slag av granskning användes numera termen evalvering.

Vad som just sagts har länge observerats i fråga om folkräkningar. Det var sålunda där tidigt praxis att komplettera redovisningen av de egentliga folkräkningsresultaten med data, som tjänade syftet att belysa folkräkningens tillförlitlighet.

Under de tre senaste decennierna har en analog evalveringspraxis utvecklats också i fråga om andra undersökningar, totalundersökningar (t ex jordbruks- och företagsräkningar) lika väl som stickprovsundersökningar (t ex arbetskraftsundersökningar).

För evalvering av en folkräkning är det naturligt att tillgripa vad som kommit att kallas demografisk metodik; en bättre term är analytisk metodik. Det för analytisk metodik utmärkande är att man söker utnyttja redan tillgängliga data för att pröva tillförlitligheten i undersökningsresultaten. Om dessa data hämtas från den undersökning, som är föremål för evalvering, talar man ibland om "intern evalvering". Om de däremot hämtas från någon annan undersökning (t ex en föregående folkräkning), brukar man tala om "extern evalvering"; se Jaffe (1951).

Utvecklingen av en evalveringspraxis för andra undersökningar än folkräkningar aktualiserade tidigt behovet att skapa ny metodik för evalvering. Som en konsekvens därav utvecklades så småningom metoder, som krakteriserades av att nya data insamlades genom särskilda evalveringsundersökningar; för en översikt av använda metoder hänvisas till US Bureau of the Census (1963).

Den just antydda utvecklingen kom möjligen för en tid att medföra en viss underskattning av potentialiteten hos ansatsen med analytisk evalvering. Under de senaste åren - icke minst i anslutning till folk- och bostadsräkningar omkring 1970 - har ett ökat intresse för analytisk evalvering förmärkts.



I avsnitten 5-8 skall vi ge ett exempel på modellbaserad evalvering, nämligen evalvering på grundval av en kombinerad tidsserie- och tvärsnittsanalys samt illustrera detta numeriskt.

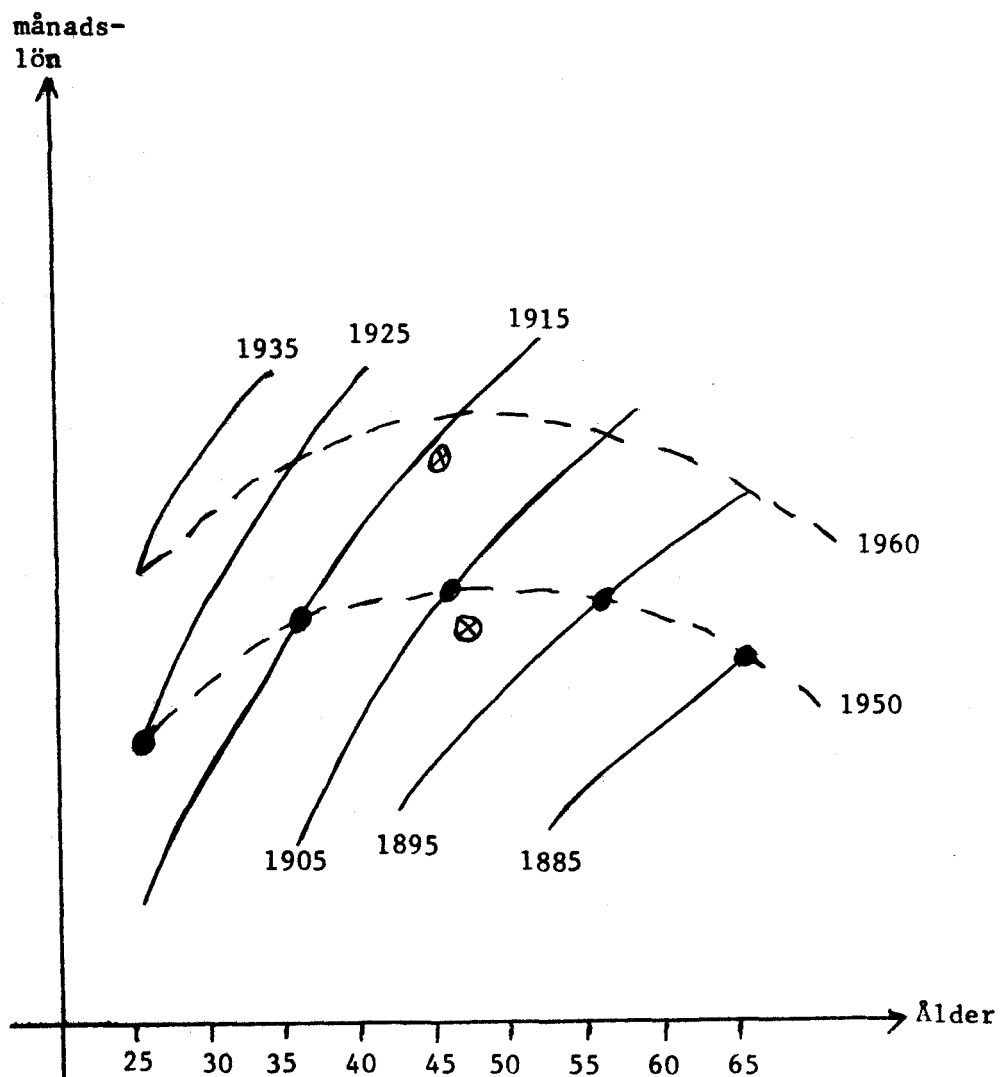
##### 5. Tidsserier och tvärsnitt - en begreppsöversikt.

Syftet med kapitel II motiverar, att vi kortfattat redogör för skillnaden mellan en tidsserie och ett tvärsnitt. Antag att vi på ett eller annat sätt definierat en grupp av individer. För varje individ kan vi vid flera tidpunkter göra observationer på en undersökningsvariabel. Om vi observerar en individ vid upprepade tillfällen, får vi en tidsserie av individdata. Gör vi i stället observationer på flera individer vid en och samma tidpunkt får vi ett tvärsnitt av individdata.

Om den för individerna gemensamma egenskap, som definierar gruppen är daterbar, brukar vi tala om en kohort, t ex kohort av födda, examinerade, värnpliktiga, etablerade företag, bilmodeller osv. Att följa en kohort innebär att man gör observationer på individer med denna gemensamma egenskap vid upprepade tidpunkter. Individerna behöver inte nödvändigtvis vara identiska. De observationer man då får (ibland, för varje tidpunkt, omräknade till t ex en summa eller ett medelvärde) utgör en tidsserie av kohortdata. Om man definierar flera kohorter och observerar dessa vid en och samma tidpunkt erhåller man ett tvärsnitt av kohortdata. Detta illustreras i figur 1.

Vi antages observera månadslönerna för kohorter av födda 1885-1935. De heldragna kurvorna beskriver den genomsnittliga löneutvecklingen över livscykeln för respektive kohort. År 1950 var de som föddes 1895 55 år gamla, de som föddes 1905 45 år, de som föddes 1915 35 år, osv. De genomsnittliga lönerna år 1950 för respektive kohort har utmärkts med en fylld ring och dessa har sammanbundits med en streckad kurva, en tvärsnittskurva. Följer man kohorterna i ytterligare tio år och observerar dem på nytt år 1960, får man en ny tvärsnittskurva; i figuren illustrerad med den övre streckade kurvan. Bildar man summan, genomsnittet eller någon annan funktion av alla observationerna i var och en av dessa tvärsnitt, erhåller man en aggregerad tidsserie av kohortdata; de korsfyllda ringarna i figuren. Kohortdata kan således sägas utgöra basen för såväl tvärsnitt som aggregerade tidsserier.

Inom många tillämpningsområden har tillgången på kohortdata hittills varit mycket sparsam. En alltmer utbyggd statistikproduktion kombinerad med rationella register- och lagringsmetoder ger emellertid en rikare tillgång



Figur 1. Schematisk illustration av sambandet mellan kohortdata, tids-serier och tvärsnitt.

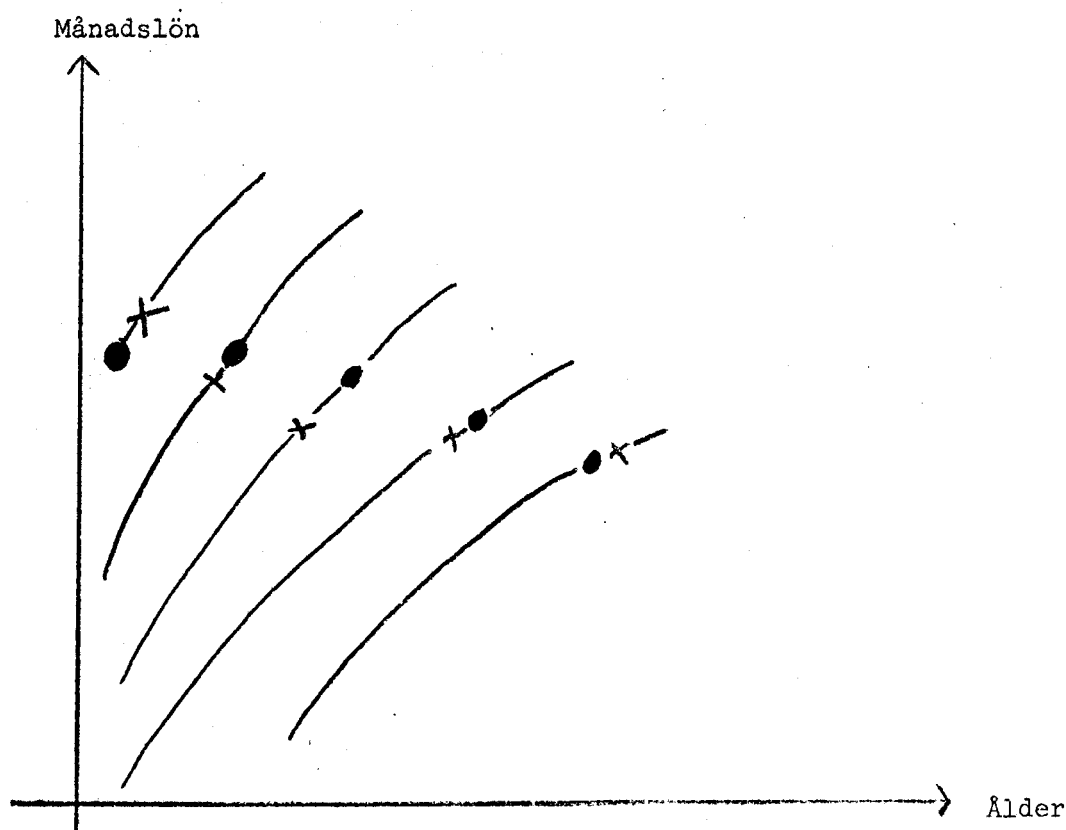
på denna typ av data. Detta tillsammans med förbättrade metoder att behandla och analysera stora datamängder ökar givetvis möjligheterna att estimeras och testa modeller. Modeller som ytterst bygger på antaganden om individers beteende, men som tidigare endast kunde testas på aggregerade tidsserier eller på "statiska" tvärsnitt, kan med kohortdata göras till föremål för mera kraftfulla test. Genom att uppmärksamma det faktum att data för ett antal kohorter genererar tvärsnitt och att ett antal tvärsnitt åtminstone delvis beskriver kohorter kan man också analysera hur ett tvärsnitt övergår till ett annat. Därigenom är det möjligt att undvika mer eller mindre egendomliga antaganden om relationerna mellan tidsserier och tvärsnitt.

## 6. Tre typer av evalvering.

Kohortdata av det slag som vi redovisat i föregående avsnitt kan användas för tre typer av evalvering. Vi behandlar dessa i tur och ordning.

### 6.1 Evalvering mot befintliga observationer (typ 1)

Figur 2 illustrerar en situation då man använder tidigare gjorda observationer för att evalvera nya. Varje heldragen kurva avses illustrera observationer på en tidsserie av kohortdata.



Figur 2 Evalvering mot befintliga observationer (typ 1)

De fyllda ringarna är observationer tillhöriga ett och samma tvärsnitt. Vi tänker oss nu att vi får nya observationer på i princip samma storheter i tvärsnittet, eventuellt med en ny observationsmetod. De nya observationerna (kryss i figuren) avviker något från de gamla. Har vi nu anledning att tro att den gamla observationsmetoden är pålitligare i någon mening än den nya kan vi använda de gamla observationerna för att bedöma kvaliteten hos de nya (och därmed hos metoden). Givetvis kan man också tänka sig den omvända situationen, nämligen att de nya observationerna erhållits med en metod som är bättre (t ex i form av mindre bias och spridning). De nya observationerna tjänar då som bas för en utvärdering av de gamla. Detta är t ex situationen vid vissa typer av särskilda evalveringsundersökningar.

Det krävs givetvis kriterier för evalveringen. Ibland kan dessa formuleras i traditionella statistiska termer. Antag t ex att vi har observationen  $x_c$  på ett antal kohorter  $c = 1, \dots, n$ , som vi vill evalvera med hjälp av en ny undersökning som ger observationerna  $y_c$ ,  $c = 1, \dots, n$ . Egenskaperna hos  $y$ -metoden är kända såtillvida att den ger unbiased estimat av de sökta storheterna  $\mu_c$ ,  $c = 1, \dots, n$ . Vidare antar vi att såväl  $x$ - som  $y$ -observationerna kan betraktas som approximativt normalfördelade samt att observationer på olika kohorter är oberoende. Problemet skulle kunna formuleras så att man vill undersöka om  $E(x_c) = \mu_c$ . Vi testar därför nollhypotesen  $H_0: E(x_c - y_c) = 0$ . Testvariabel är  $\bar{d} = \bar{x} - \bar{y}$  och variansen för denna kan uppskattas med  $s^2(d_c)/n$ , där  $d_c = x_c - y_c$  och

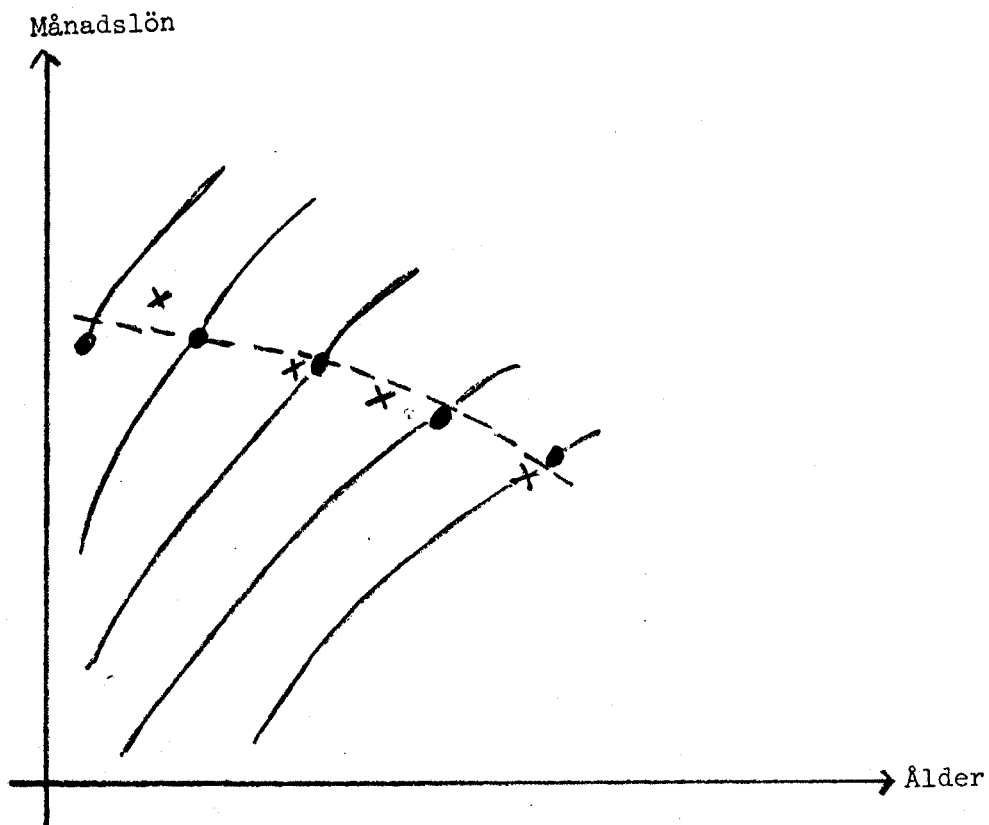
$$s^2(d_c) = \frac{n}{n-1} \sum_1^c (d_c - \bar{d})^2 .$$

Under de givna förutsättningarna kan hypotesen prövas med ett vanligt t-test.

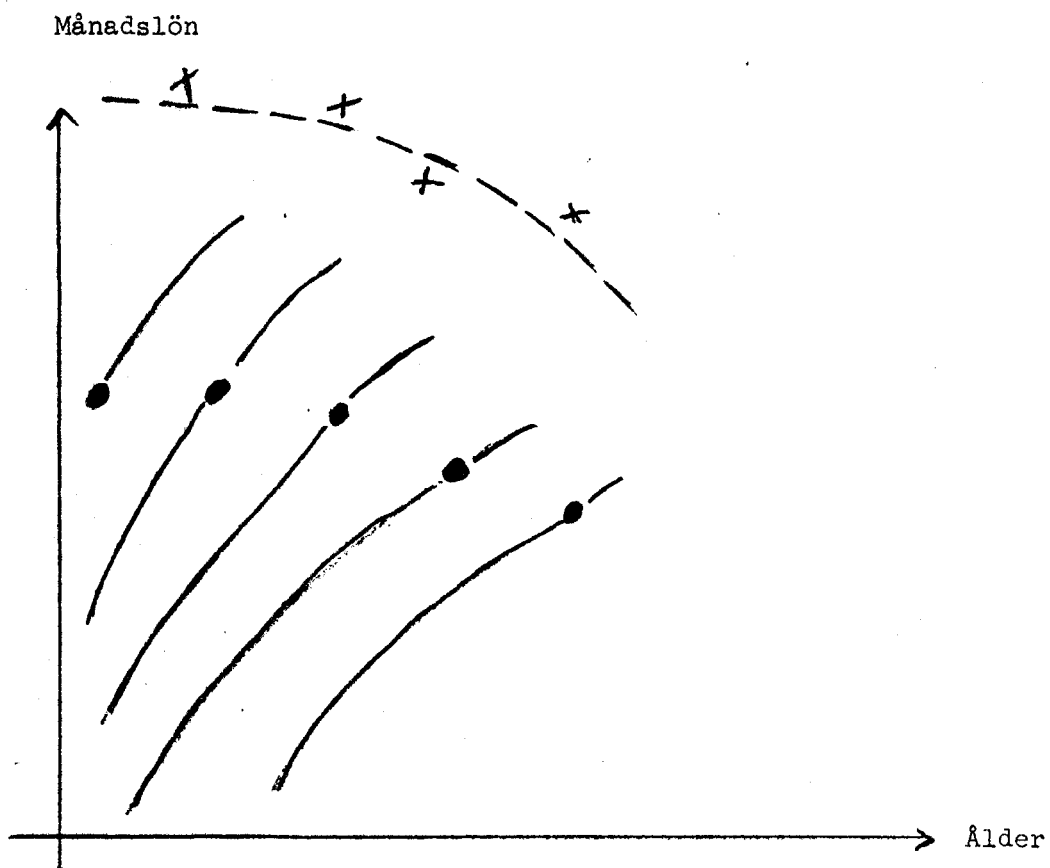
Det bör kanske påpekas att denna typ av evalvering inte är något speciellt för tvärsnitts- eller längdsnittsdata, utan den kan i princip utföras för vilken datamängd som helst.

## 6.2 Modellbaserad evalvering (typ 2)

Evalvering på en något högre ambitionsnivå illustreras i figur 3. Vi antas nu disponera en modell som förklarar hur tvärsnittet förskjuts och som estimerats på grundval av de ursprungliga kohortdata. Modellen kan användas



Figur 3. Modellbaserad evalvering (typ 2)



Figur 4. Modellbaserad evalvering med prognoser (typ 3).

för att predicera ett tvärsnitt - den streckade kurvan i figuren - mot vilket de nya observationerna - kryss - evalveras <sup>1)</sup>.

Det är naturligt att kriterierna för evalveringen väljes på grundval av modellens egenskaper. En stokastisk modell kan användas för att definiera en kritisk region för ett test av hypotesen att de nya observationerna genererats av samma modell som de ursprungliga observationerna.

### 6.3 Modellbaserad evalvering med prognoser (typ 3)

Figur 4 illustrerar en evalvering på en ännu något högre ambitionsnivå.

I denna tredje illustration tänkes vi observera ett tvärsnitt utanför sample-perioden, vilket evalveras mot en prognos gjord med hjälp av modellen. Precis som i den förra illustrationen kan man tänka sig att utnyttja modellens stokastiska egenskaper för att utforma ett testförfarande.

## 7. Några numeriska illustrationer.

I detta avsnitt konkretiseras de tre fallen ovan genom några numeriska illustrationer. De data som därvid användes har hämtats från Svenska Arbetsgivareföreningens (SAF) lönestatistik för tjänstemän och från Sveriges Civilingenjörskörbundet (CF) lönestatistik. En närmare beskrivning av dessa data finns bl a i Klevmarken (1972).

Den första numeriska illustrationen anknyter till figur 2 ovan. I tabell 1 redovisas medianlönen efter ålder enligt de två källorna för tjänstemän i enskild tjänst med teknisk högskoleutbildning. Den aktuella frågeställningen är om det finns någon systematisk skillnad mellan de två källornas uppskattning av civilingenjörernas ålders-löneprofil som måste beaktas i förhandlingarna mellan SAF och CF. Man lägger genast märke till att med undantag för 24-åringar ligger de av CF redovisade medianlönerna genomgående över de av SAF redovisade och skillnaden växer med stigande ålder. Något test av om denna skillnad kan anses slumpmässigt betingad eller ej förefaller helt onödig i detta fall, men genomförda beräkningar ger en illustration till valet av testmetod i en situation som är mindre

<sup>1)</sup> En fördel med modellevalvering är att de data som skall evalveras inte nödvändigtvis behöver redovisas i exakt samma kategorier som de ursprungliga. Detta förhållande illustreras i figur 3.

självlklar än denna. Med ett t-test (beräknat på logaritmiska medianlöner) av det slag som nämndes i föregående avsnitt kan nollhypotesen inte förkastas på 5 %-nivån. Detta visar svagheten hos detta test, då skillnaden mellan de två tvärsnitten uppvisa en trend eller något annat systematiskt förlopp. Den nollhypotes man egentligen bör testa är att skillnaderna mellan de två tvärsnitten kan betraktas som en serie av oberoende slumpmässiga observationer med förväntan noll. En tänkbar testvariabel är en von Neumann kvot <sup>1)</sup>. Tillämpad på de två tvärsnitten i tabell 1 skulle nollhypotesen förkastas på 5 %-nivån <sup>2)</sup>. Andra tänkbara testmetoder skulle man kunna finna inom familjen av run-test.

Nästa illustration ansluter till figur 3. För 1970 förelåg SAF:s lönestatistik endast redovisad i vissa åldersklasser, inte för varje åldersår. Statistiken från CF och dess delförbund fanns däremot redovisad för varje åldersår. I en jämförelse mellan de två statistikmassorna skulle man vara tvungen att lämna en del av den information som fanns i CF-statistiken outnyttjad om jämförelsen enbart skulle baseras på de publicerade medianlönerna. En alternativ ansats, som tillåter att all information utnyttjas, är att basera jämförelsen på prediktioner av ett fullständigt tvärsnitt från en modell.

Den modell som använts i det följande exemplet är relativt primitiv. Den förutsätter att tvärsnittsprofiler av logaritmiska löner i förväntan förskjutes uppåt lika mycket varje år och att profilernas form inte förändras. För att inte binda modellen i en viss funktionsform har profilerna approximeras med polygoner <sup>3)</sup>.

1)

$$\text{von Neumann kvot} = \frac{\frac{1}{n-1} \sum_{t=2}^n (D_t - D_{t-1})^2}{\frac{1}{n} \sum_{t=1}^n D_t^2}$$

där  $D_t$  är skillnaden mellan tvärsnitten vid åldern  $t$ . I detta numeriska exempel skulle testvariabeln antaga värdet 0,131.

2) Samtliga test i det följande genomföres på 5 %-nivån.

3) Modellen för ett tvärsnitt  $T$  kan skrivas

$$\ln L_{Tj} = \exp \left[ \alpha + \beta T + \sum_{i=1}^n (\gamma_i - \beta) D_i + u_{Tj} \right],$$

där  $j$  är åldern,  $D_i$  en variabel som anger hur många år som tillbringats i det  $i$ :te åldersintervallet och  $u_{Tj}$  en slumpterm oberoende normalfördelad med förväntan noll.  $\alpha$ ,  $\beta$  och  $\gamma_i$  är parametrar som skattas med generaliserad minsta-kvadratmetod. En närmare redogörelse för modellen och skattningsmetoderna finns i Klevmarken (1972).

I tabell 2 kolumn 2 redovisas de prediktioner som erhållits sedan modellen estimerats från klassindelad lönestatistik från SAF. I kolumn 3 återges medianlöner från ett delförbund inom CF, nämligen Svenska Mekanisters Riksförening (SMR). Den sista kolumnen visar att SMR-lönerna ligger över SAF-lönerna i varje åldersklass. Återigen är skillnaden så stor att något test egentligen inte behövs. Med von Neumann-testen tillämpad på skillnaderna mellan (de logaritmiska) lönerna förkastar man således hypotesen att skillnaden mellan (de logaritmiska) lönerna är oberoende observationer med förväntan noll. (Testvariabeln antar i detta exempel värdet 0,237).

Den sista numeriska illustrationen är en jämförelse mellan en prognos på tvärsnittsprofilen för civilingenjörer i SAF-anslutna företag år 1970 och den i efterhand observerade profilen av medianlönen; således ett exempel i anslutning till figur 4. Prognosen gjordes med hjälp av samma modell som i den föregående illustrationen. I tabell 3 återges prognosen, den observerade profilen såväl som den procentuella skillnaden dem emellan. Inom större delen av åldersintervallet är överensstämmelsen ganska god och ett testförfarande är befogat för att avgöra om den observerade profilen för 1970 kan anses ha genererats av samma modell som de tidigare profilerna. Under förutsättning att den använda modellen **är realistisk kan skillnaderna mellan de prognosticerade och observerade logaritmiska profilerna betraktas som en serie av oberoende slumpmässiga observationer med förväntan noll.** Von Neumann-kvoten blir i detta fall 1,39 och vi måste förkasta hypotesen att de nya observationerna har genererats av modellen. Möjligen kan man invända att man med detta testförfarande är "för snar" att förkasta nollhypotesen, då man inte utnyttjar någon information om hur modellen anpassade sig till sample-periodens observationer. Om anpassningen var dålig under sample-perioden kan man naturligtvis inte vänta sig att den ska vara bättre eller ens lika god under prognosperioden. Om nollhypotesen gäller är kvoten mellan medelresidualkvadratsummorna för prognosperioden resp för sample-perioden F-fördelad. Kvoten blir i detta fall 0,174 och vi kan således med **detta** test inte förkasta nollhypotesen. Detta illustrerar den grundläggande svårigheten vid utvärdering av data med hjälp av en modell, nämligen att resultatet av utvärderingen är helt beroende av vilken modell man har och hur mycket man "tror" på den. Med det första testförfarandet förlitar man sig helt på modellen och resultatet blir inte oväntat att man måste förkasta den nya datamängden. Den risk



man löper är att felaktigt förkasta den nya datamängden därför att modellen varken förmår beskriva data under prognosperioden eller under sampleperioden med tillräcklig noggrannhet; (den förnuftiga slutsatsen vore i stället att förkasta modellen). Med det andra test-förfarandet demonstrerar man ett visst tvivel beträffande modellens egenskaper och begär endast att avvikelser mellan prognosen och den nya datamängden inte skall vara större än vad anpassningen under sampleperioden motiverar. Med den senare ansatsen får man i allmänhet svårare att förkasta en avvikande datamängd. Evalveringens styrka blir helt beroende på hur "bra" modellen är.

#### 8. Principiella problem vid modellbaserad evalvering.

Den grundläggande idén - att verkställa evalvering av data med hjälp av en modell - kan uppenbarligen ges en vidare tillämpning. Vi skall här kortfattat beröra den principiella problematik, som i så fall aktualiseras. Användbarheten av modellbaserad evalvering ställer uppenbarligen vissa krav på modellen; denna måste vara "bra". Men den ställer också vissa krav på analysen. Ett principiellt och praktiskt huvudproblem vid evalvering av data med hjälp av modeller är sålunda hur man tolkar en observerad skillnad mellan prediktion och utfall, dvs de data som skall evalveras.

Orsakerna till en skillnad kan vara flera och man kan tänka sig att dela upp den i motsvarande komponenter, t ex fel i utfallsdata, fel i modellstrukturen och fel i de exogena komponenter som insättes i modellen för att göra prediktioner. I detta sammanhang kan vi kanske nöja oss med att betrakta två felkomponenter, nämligen fel i utfallsdata och fel i prognosmetodiken.

Man kan nu särskilja två olika frågeställningar:

- a) givet en viss modell kan utfallet ha genererats av denna modell?
- b) givet ett visst utfall vilka fel i prognosmetodiken finns det?

I den första frågeställningen har man en fast tro på sin modell och tänker sig att utfallet kan avvika från prediktionen på grund av mätfel (i vid bemärkelse). Man tänker sig således att det är insamlingsmetoderna som har förändrats (är annorlunda) snarare än den mekanism som forskaren inte behärskar och som antages ha genererat observationerna.

I den andra frågeställningen är förhållandet det omvända. Insamlingsmetoderna och därmed observationernas "egenskaper" antages vara oförändrade medan däremot den prognosmetodik som använts antages vara otillräcklig. Ofta brukar en stor del av avvikelserna mellan prognos och utfall under dessa antaganden kunna förklaras med "felaktigheter" i prognosmodellens uppbyggnad. Inom ekonometrin brukar man i detta sammanhang tala om strukturella förändringar. Den modellstruktur man valt för att beskriva sample-perioden måste förändras för att man på ett nöjaktigt sätt skall kunna beskriva prognosperioden.<sup>1)</sup> Bakom antagandet om observationernas "oförändrade egenskaper" och om förändringen i modellstrukturen måste det finnas en föreställning om en annan modell än den man faktiskt har, som skulle kunna beskriva observationerna under såväl sample- som prognosperioden.

Beklagligtvis kan man knappast räkna med att kunna separera de två komponenterna i skillnaden mellan prognos och utfall. Inom t ex ekonometrin har man oftast koncentrerat sig på fel i strukturerna givet ett visst utfall, medveten om den bristande teoretiska underbyggnaden och de stora förenklingar man vanligen måste göra (och samtidigt möjligen omedveten om de brister som finns i den ekonomiska statistiken). Inom andra tillämpningsområden kan man kanske förlita sig mera på de teorier som ligger till grund för prognoserna, varvid det kan vara naturligare att ställa den första av de två frågorna ovan.

Den enda tillämpning vi kan finna i vilken det är möjligt att separera de strukturella felen från felen i utfallsdata är i simuleringar av en evalveringssituation, då modellstrukturerna fastlägges från början.

Förutsatt att man finner det meningsfullt att ställa den första av de två frågorna återstår ändå några statistiskt tekniska problem. På grundval av vilka kriterier skall man avgöra om utfallet kan anses avvika så mycket från prediktionerna att "fel" i statistiken föreligger? Detta problem berördes ovan i anslutning till de numeriska exemplen.

---

1) Ett statistiskt mått för att undersöka strukturella förändringar har föreslagits av Gadd och Wold; se Wold (1964).

## III BILAGOR

9. Referenser

- Agostini, J.-M. (1962), How to estimate unduplicated audiences. Journal of Advertising Research, pp. 11 - 14.
- Amundsen, A. (1960), Metoder i analysen av forbruksdata. Statsøkonomisk Tidsskrift, Hefte 1, Mars 1960.
- Dalenius, T. (1962), Automatic estimation of missing values in censuses and sample surveys. Statistisk Tidsskrift, s 395 - 400.
- Dalenius, T. (1971), Principer och metoder för placering av samplingundersökningar. Intern handbok nr 4, Statistiska centralbyrån, Stockholm.
- Frank, O. (1968), Om tidningars räckvidd genom konsekutiva nummer. Forskningsrapport. Statistiska institutionen, Stockholms universitet.
- Goldberger, A.S. (1972), Structural equation models. Social Science Research Council, ITEMS, vol 26, No. 3.
- Gurney, M. och Dalenius, T. (odat.) The multivariate approach to survey design and estimation. Forskningsrapport nr 7 i projektet Fel i undersökningar.
- Hansen, M.H., Hurwitz, W.N. och Madow, W.G. (1953), Sample Survey Methods and Theory, John Wiley and Sons, New York.
- Jaffe, A.J. (1951), Handbook of Statistical Methods for Demographers. Government Printing Office, Washington, D.C.
- Klevmarken, A. (1972), Statistical Methods for the Analysis of Earnings Data with special application to salaries in Swedish industry, Industriens Utredningsinstitut och Statistiska Centralbyrån, Stockholm.
- U.S. BUREAU OF THE CENSUS (1963), Evaluation and Research Program for the U.S. Censuses of Population and Housing, 1960: Background, Procedures and Forms, Ser. ER 60, No. 1. Washington, D.C.
- Wold, H.O.A. (1964), Econometric Model Building Essays on the Causal Chain Approach. North Holland. Amsterdam.

10. Tabeller

Tabell 1. Medianlön för tjänstemän i enskild tjänst med teknisk  
högskoleutbildning efter ålder år 1970

Ålder (år)	Medianlön kronor enligt statistik		Procentuell skillnad
	från SAF	CF	
24	3000	3100	3,33
25	3105	3200	3,06
26	3290	3312	0,67
27	3474	3500	0,75
28	3700	3800	2,70
29	4000	4200	5,00
30	4225	4404	4,24
31	4501	4700	4,42
32	4850	5000	3,09
33	5050	5275	4,46
34	5275	5550	5,21
35	5588	5792	3,65
36	5600	5900	5,36
37	5875	6250	6,38
38	6200	6550	5,65
39	6250	6650	6,40
40	6450	6800	5,43
41	6500	6775	4,23
42	6633	7180	8,25
43	6671	7000	4,93
44	6800	7340	7,94
45	6800	7487	10,10
46	6900	7550	9,42
47	7200	7450	3,47
48	6900	7500	8,70
49	7500	7750	3,33
50	7175	8000	11,50

Källor: SAF-SIF-SALF-HTF tjänstemannastatistik 1970,  
SAF, Statistiska byrån  
CF-statistiken 1970, CF, tabell 5.

Tabell 2. Jämförelse av åldersfördelade civilingenjörslöner mellan SAF och SMR år 1965

(1) Ålder (år)	(2) SAF-prediktion (kronor)	(3) SMR (kronor)	(4) Procentuell skillnad
28	2714	2960	9.1
29	2881	3175	10.2
30	3059	3225	5.4
31	3242	3356	3.5
32	3436	3600	4.8
33	3641	3965	8.9
34	3858	3950	2.4
35	4089	4100	0.3
36	4188	4250	1.5
37	4290	4356	1.5
38	4394	4500	2.4
39	4501	4680	4.0
40	4610	5000	8.5
41	4722	5200	10.1
42	4837	5450	12.7
43	4954	5100	3.0
44	5075	5500	8.4
45	5198	5833	12.2
46	5187	5650	8.9
47	5177	6000	15.9
48	5167	5817	12.6

Anm Kol (1) är prediktioner från en modell (se not s 14) estimerad från åldersklassvisa medianlöner från SAF för perioden 1954-1969. Parameterestimaten är:

$$\begin{array}{ll}
 \hat{\alpha} = 4.325 & \hat{\gamma}_3(35 - 39 \text{ år}) = 0.123 \\
 \hat{\beta} = 0.065 & \hat{\gamma}_4(40 - 44 \text{ år}) = 0.089 \\
 \hat{\gamma}_1(-29 \text{ år}) = 0.069 & \hat{\gamma}_6(45 - 59 \text{ år}) = 0.063 \\
 \hat{\gamma}_2(30-34 \text{ år}) = 0.125 & \hat{\gamma}_7(60 - \text{år}) = 0.034
 \end{array}$$

Kol (2) är geometriska genomsnitt av observerade individuallöner för resp åldersår från SMR för år 1965.

Tabell 3. Jämförelse mellan prognos och utfall för civilingenjörs-  
löner inom SAF för år 1970

(1) (år)	(2) SAF-prediktion ( kronor)	(3) SAF-medianlön (kronor)	(4) Procentuell skillnad
25	3318	3105	6,9
26	3331	3290	1,2
27	3537	3474	1,8
28	3756	3700	1,5
29	3989	4000	-0,3
30	4234	4225	0,2
31	4487	4501	-0,3
32	4755	4850	-2,0
33	5039	5050	-0,2
34	5340	5275	1,2
35	5659	5588	1,3
36	5796	5600	3,5
37	5937	5875	1,1
38	6081	6200	-1,9
39	6229	6250	-0,3
40	6380	6450	-1,1
41	6535	6500	0,5
42	6694	6633	0,9
43	6857	6671	2,8
44	7023	6800	3,3
45	7194	6800	5,8
46	7180	6900	4,1
47	7165	7200	-0,5
48	7151	6900	3,6
49	7137	7500	-4,8
50	7122	7175	-0,7
51	7108	7250	-2,0
52	7094	7133	-0,5
53	7080	7475	-5,3
54	7066	7277	-2,9
55	7051	6900	2,2
56	7037	6798	3,5
57	7023	6895	3,2
58	7009	6600	6,2
59	6995	7225	-3,2
60	6981	6973	0,1
61	6768	6550	3,3
62	6562	6700	-2,1
63	6361	6364	0,0
64	<u>6167</u>	<u>6385</u>	<u>-3,4</u>
Medelvärde	6106	6074	0,5

Anm Kol (1) består av prediktioner från samma modell som användes i tabell 2.

Kol (2) är medianlöner hämtade från lönestatistiken.