Kompetensgruppen för granskning och mätning

Selektiv granskning

1 Introduktion

Granskningsarbete är resurskrävande. Flera studier har visat att granskning av data i ekonomisk-statistiska undersökningar tar omkring en tredjedel av resurserna. Studier har också visat att produktionsgranskningen ofta är ineffektiv, till men för statistikbyrån och uppgiftslämnarna. Ett problem är att ”enkla” kontroller, som söker fel i data, sällan beaktar vägningstalen i undersökningen, än mindre förväntad effekt på den statistik som produceras.

Selektiv granskning är en samlingsterm för metoder som väljer poster som troligt har inflytelserika fel för manuell/interaktiv utredning, post för post.

Syftet med selektiv granskning är att minska produktionsgranskningens omfattning, dvs. minska listan över misstänkta data (fellistan), genom att beakta de misstänkta felens möjliga effekt på den statistik som produceras.

En generellt viktig förutsättning för effektiv granskning är att se till att kontrollerna som flaggar data som misstänkt felaktiga har hög träffsäkerhet, oavsett om man granskar traditionellt eller selektivt.

Vi har aspekter att avväga vid selektiv granskning:

- *misstanke* om att ett värde är felaktigt och
- *effekt* på statistiken av ett felaktigt värde.

Lägg därtill att det behövs värderingar – ur ett användarperspektiv – av vad i statistikens hela output som är mer och mindre viktigt, så har vi i stort sett förutsättningarna för att designa selektiv granskning.

2 Olika metoder

Följande presentation baseras i viss mån på denna artikel.
Dan Hedlin (2003) ”Score Functions to Reduce Business Survey Editing at the U.K. Office for National Statistics”, JOS Vol 19. I artikeln görs en uppdelning i *estimate-related* och *edit-related* metoder, dvs metoder som avgör effekten på den statistik som produceras respektive metoder som endast bygger på hur observerade data felsignaleras av kontrollerna.

2.1 Enbart kontroll-baserad selektiv granskning

Selektiv granskning baserad enbart på utfallen av kontroller är enkla att tillämpa men ger inte alltid den effekt vi önskar.

2.1.1 Räkna antalet ”flaggor”

Allra enklaste selektiva granskningen som skapar en prioritering bland objekten baseras på antalet felsignaler från alla kontroller för objektet. Många felsignaler är en indikation på många fel. Då har vi digitaliserat signaleringen i endera ”acceptabel” och ”flaggad”.

2.1.2 Summera ”misstankar”

Genom att beakta hur mycket ett datavärde avviker från acceptansområdet kan vi tilldela varje objekt en *misstanke*. Om värdet ligger inom acceptansområdet låter vi misstanke = 0. Utanför acceptansområdet sätter vi misstanken till ett tal mellan noll och ett enligt någon lämplig funktion. Misstanken vid en avvikelsekontroll borde kanske aldrig vara fullt upp lika med hundra procent, möjligen då för tusenfel som kan hanteras på detta sätt. Det är för övrigt inte viktigt att misstanken inte får vara 100%, skillnaden mellan 90% och 100% blir i praktiken försumbar. En selektiv prioritering av objekt baseras nu på summan av misstankarna från samtliga kontroller.

2.1.3 Generalisering

Det vore fullt möjligt att tillämpa olika vikter för olika variabler för att fokusera mer på felsignaleringar av viktiga variabler.

Man kan också använda vikter för kontroller så att man till exempel får mindre bidrag från kontroller som har jämn spridning av misstankar i intervallet 0 - 1 och kontroller som har hög andel stora misstankar och resten noll.

2.2 Effekt-baserad selektiv granskning

Varje observerat värde, misstänkt eller inte, jämförs med ett förväntat värde. Det förväntade värdet kan vara det insamlade och redan granskade värdet från senaste insamlingsomgång, men vi kan använda mer komplicerade algoritmer som blir nödvändiga när nya urval tas i bruk och det inte finns något tidigare värde.

Differensen mellan observerat och förväntat värde, multiplicerat med designvikten (vid urvalsundersökning), ger en *förväntad mätfeleffekt* på den skattade summan av variabeln i fråga för hela undersökningspopulationen. Effekten kan dessutom beräknas per redovisningsgrupp och relateras till skattade summan alternativt relateras till andra osäkerheter i statistiken.

Metoden förutsätter att man på förhand kan förvänta sig ungefär vilket värde som respektive variabel kan anta för alla objekt. Ju bättre för-

väntade värden man har, desto effektivare blir den selektiva granskningen. Metoden passar därför bäst i återkommande undersökningar och i synnerhet i de fall flertalet objekt har ingått i tidigare insamlingsomgångar.

2.2.1 Effekt utan misstanke

En enkel variant av selektiv granskning vore att inte längre använda traditionella kontroller utan bara skapa en poäng baserat på att aggregat av de förväntade mätfelseffekterna. Nu inses genast att om undersökningens variabler har varianser av olika storleksordning så måste effekterna normaliseras genom att divideras med standardavvikelser. Nu går det bra att bilda summa per objekt eller summa av till exempel kvadrerade normaliserade effekter till en poäng.

Om man inte använder kontroller som ger misstankemått så kommer de mycket stora företagen att prioriteras med denna metod. Deras observerade värden behöver inte alls vara misstänkta för att skillnaden mellan aktuell och föregående månads värden ska ha tydlig effekt.

Enklast är att göra en prioritering med avseende på skattade totaler i undersökningen. Vill man uppnå kvalitet för redovisningsgrupper (branscher) växer komplexiteten.

2.2.2 Effekt med misstanke, utan förväntat värde

En riktigt enkel poängfunktion bygger på observerat värde multiplicerat med designvikten (vid urvalsundersökning), inte på differensen mellan observerat och förväntat värde. Sedan kan man multiplicera med misstanken. Metoden fungerar bra när det ogranskade värdet felaktigt är för stort. Men om det ogranskade värdet är felaktigt litet blir poängen särskilt låg och kommer inte att prioriteras alls. Därför riskerar metoden att systematiskt hitta fel uppåt men inte nedåt, vilket rent av introducerar bias i undersökningen.

Det finns undersökningar där förväntade värden helt enkelt inte går att finna. Ett exempel är investeringar som kan vara stora ett år fast det inte var någon investering alls året innan och vice versa. Det kan rent av vara omvänt samband att ett företag som investerat mycket ena året därför investerar litet året därpå.

2.2.3 SCB:s rekommenderade metod

Det blir inte mycket mer komplicerat att multiplicera den förväntade mätfelseffekten per objekt och variabel med misstanken. Misstanken beräknas av kontroller och får vara digital (noll och ett) eller kontinuerlig 0 – 1 och även båda typerna får förekomma samtidigt.

På SCB används selektiv granskning med poängfunktion. Metoden innebär att

- för varje urvalsobjekt (uppgiftslämnare) och

- för varje observationsobjekt i urvalsobjektet (kluster) och
- för varje statistiskt mått (kopplat till en/flera variabler) och
- för varje redovisningsgrupp (domän) som objektet bidrar till
- skapas en lokal poäng.

Den lokala poängen är produkten av misstanke och effekt. *Misstanke* för variabelvärdet kan vara 0/1 till följd av en/flera vanliga granskningskontroller där felsignalering ger värdet 1. Det rekommenderas dock att misstanken anges på en kontinuerlig skala mellan 0 och 1, det utnyttjar mer information. *Effekt* är differensen mellan ogranskat och förväntat värde för observationsobjektet (anställd, produkt) uppräknat med urvalsobjektets designvikt.

De lokala poängen för redovisningsgrupper aggregeras till varje mått (variabel), till varje observationsobjekt och vidare till en global poäng för varje urvalsobjekt.

De urvalsobjekt som har en global poäng över ett visst tröskelvärde går till manuell utredning. De objekt som har ett lägre värde lämnas utan åtgärd eller automatändras om träffsäkerheten i kontrollen är hög, eftersom de sammantaget inte påverkar skattningarna mer än marginellt.

Som alternativ till att ha ett förutbestämt tröskelvärde för den globala poängen kan man manuellt utreda de objekt som har höga globala poäng och avsluta granskningen när tid eller resurser inte räcker längre.

Uppenbara fel kan också hanteras genom att man väljer i vilken utsträckning som de ska vara grund för återkontakt, beroende på effekten på statistiken.

För en närmare genomgång av den metodik för selektiv granskning som används på SCB, hänvisas till avsnitt 3 om SELEKT nedan och A General Methodology for Selective Data Editing.

2.2.4 Finns en lätt-version?

Den metod som rekommenderas av SCB kan tyckas ambitiös. För det första vill vi skapa en global poäng som gör det möjligt att begränsa antalet återkontakter med uppgiftslämnare. I vissa undersökningar är detta kanske inte ett starkt krav, vi vill endast begränsa listan över felsignalerade variabelvärden. Lite enklare är detta nog, men den slutliga aggregeringen är ett litet problem. Vi måste ju ändå räkna lokala poäng för varje variabel i relation till annan statistisk osäkerhet för respektive variabel.

Kan vi förenkla genom att inte bedöma effekten relativt annan statistisk osäkerhet i alla redovisningsgrupper, enligt olika indelningar? Ja, beräkna effekten endast mot en totalskattning för hela populationen. Då har vi åtminstone beaktat designvikterna som kan vara mycket

varierande i urvalsundersökningar. Små redovisningsgrupper har naturligt stora medelfel (urvalsosäkerhet) och effekten av ett mätfel blir också stor om urvalsstorleken är liten. Det kan fungera i vissa situationer.

2.3 SSB:s ISEE är en top-down-metod

Statistisk Sentralbyrå i Norge har utvecklat ett verktyg för selektiv granskning som heter Integrated system for Editing and Estimation, ISEE.

De två processerna granskning och estimation är helt integrerade i verktyget ISEE. Det innebär förutsättningar för den så kallade top-down-metoden för granskning. Effekten på de slutliga skattningarna av populationssummor och andra skattningar p g a alla ändringar som görs i data kan ses omedelbart.

Man inleder en ny produktionsomgång med att imputera alla sökta data. Det kan egentligen vara enligt samma metoder som vi skulle framställa de förväntade värdena i metoderna som beskrivits ovan. Nu kan man ta fram de första pseudo-skattningarna för omgången.

När en blankett kommer in från en uppgiftslämnare så byts de fejkade värdena mot de inkomna. Genast kan man då se effekten, dvs hur skattningarna av all statistik ändras. Vi går inte här in på hur man skapar en ordning för vad som ska prioriteras.

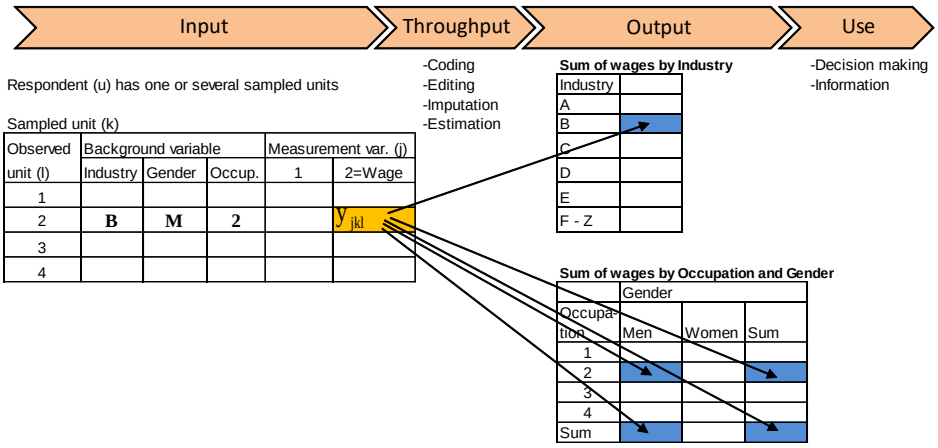
3 SELEKT

På SCB används det generiska IT-verktyget SELEKT för att flagga de objekt som ska prioriteras för manuell utredning. Verktyget är byggt i SAS och används som en fristående modul som kopplas till Triton eller det befintliga produktspecifika produktionssystemet. SELEKT innehåller också funktionalitet för att ta fram förväntade värden för beräkning av potentiell effekt. SELEKT har också funktionalitet för att definiera kontroller. Parametrar finns för att prioritera bland mått (variabler) och redovisningar. Det finns en användarhandledning som kan erhållas, skriv till *kompetensgruppen för granskning och mätning*, via gruppbrevlådan Granskning@scb.se.

Verktyget beskrivs också i
SELEKT – A Generic Tool for Selective Editing (sciendo.com)

3.1 Tabellceller

Lokala poäng beräknas för alla de tabellceller som har prioriterats.
I nedanstående exempel har markerats variabel 2 (lön) för en man i yrke 2 och bransch B. Detta variabelvärde ingår i fem tabellceller där lokal poäng beräknas. Observera att om något är fel så kan det vara lön men lika gärna kön och yrke.



3.2 Beräkning av effekt på statistiken

Poängfunktionen som beskrivs nedan sätter poäng för misstänkt avvikande värden. Följande uttryck utgör effekten på en skattning om det ogranskade värdet behålls i stället för att utreda.

faktisk effekt = $w_{k,l}$ x $({}^{ogr}y_{j,k,l} - {}^{gr}y_{j,k,l})$, där

- $w_{k,l}$ är uppräkningstalet för observationsobjekt *k,l*
- ${}^{ogr}y_{j,k,l}$ är det ogranskade värdet för variabel y_j och observationsobjekt *k,l*
- ${}^{gr}y_{j,k,l}$ är det granskade värdet för variabel y_j och observationsobjekt *k,l*.

Observationsobjekt *k,l* innebär att man har hierarkiska data på två nivåer, t.ex. om ett företag redovisar uppgifter för sina anställda. Ofta har vi dock endast en nivå.

På förhand känner vi inte ${}^{gr}y_{j,k,l}$. I poängfunktionen använder vi som ett ersättningsvärde (en proxy) det förväntade (gissade) värdet och får:

potentiell effekt = $w_{k,l}$ x $({}^{ogr}y_{j,k,l} - {}^{förv}y_{j,k,l})$,

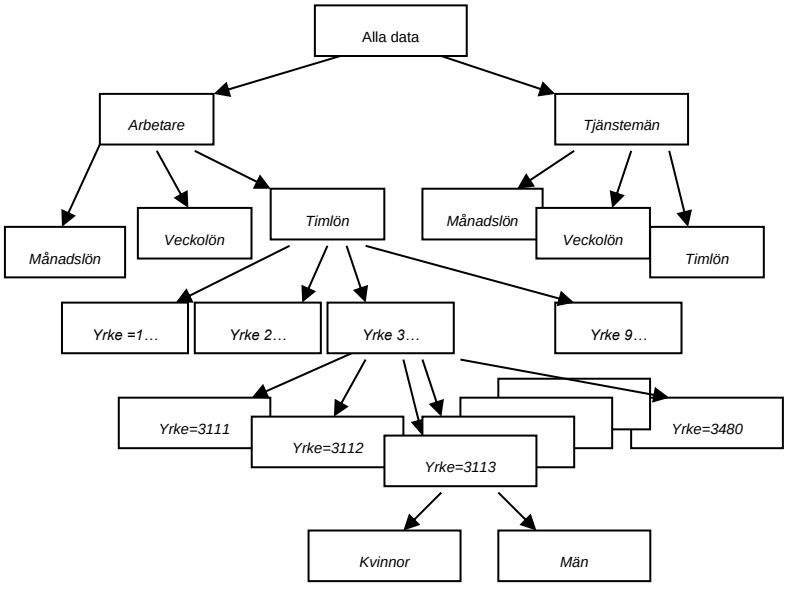
förväntad effekt = misstanke x potentiell effekt.

Misstanken att det ogranskade värdet ${}^{ogr}y_{j,k,l}$ är fel bestäms av en eller flera kontroller. Resultatet av testerna sätts ofta till 0 eller 1. Värden som ligger utanför kontrollens acceptansområde sätts traditionellt till 1, men kan också sättas till varje annat värde mellan 0 och 1, till exempel genom att anpassa det till träffsäkerheten i kontrollen. I det senare fallet får man ett bättre värde på den förväntade effekten.

3.3 Beräkning av förväntade värden

I poängfunktionen ingår, utöver de data som ska granskas, också förväntade (gissade) värden. Ju bättre man kan ”prognostisera” indata, desto effektivare blir selektiv granskning. Ofta är det bästa sättet att använda uppgifter från samma objekt från tidigare insamlingsomgångar då det är möjligt. Ett alternativ är att använda genomsnitt (från nuvarande eller tidigare undersökningsomgång) för s.k. granskningsgrupper, vilka är homogena med avseende på undersökningsvariablerna. Eftersom man får ett säkrare förväntat värde ju mer homogen gruppen är, ska ett sådant värde väljas på olika detaljeringsnivå beroende på tillgång till data i olika kombinationer av kategorivariabler. SELEKT har som inbyggd modul en funktion för att man på ett sådant sätt definiera ”bästa grupp”, men däremot söker inte SELEKT grupperna.

I exemplet nedan tänker vi oss att vi ska ta fram förväntade värden för löner i kategorierna arbetare/tjänstemän, löneform, yrke på olika nivåer och kön. SELEKT väljer då ett genomsnitt för en viss grupp längst ner i schemat i det fall antalet observationer tillåter detta, men i annat fall går man upp en nivå i pyramiden för att se om antalet observationer räcker. Detta upprepas tills man har hittat den mest homogena gruppen med tillräckligt antal observationer.



3.4 Poängberäkning

Lokal poäng beräknas för respektive redovisningsgrupp (domän) enligt viktighet x förväntad relativ effekt, eller mer specifikt uttryckt:

$$LP_{d,j,k,l} = VP_d \text{ x } VP_j \text{ x } Misstanke_{j,k,l} \text{ x } PotEff_{d,j,k,l}/SE(\hat{T}_{d,j})$$

där *d* anger domän, *j* anger mått/variabel samt *k* och *l* anger objekt då vi har två nivåer.

Lokal poäng innebär alltså förväntad effekt, relaterad till medelfelet för skattningen av måttet/variabeln i domänen och till viktigheten i övrigt. *VP* anger viktighetspoäng för domän (*d*) respektive mått/variabel (*j*). Man kan styra granskningen mot en viss variabel genom att sätta VP_j till ett högre värde än standard (förval). På motsvarande sätt kan en viss redovisningsgrupp prioriteras via VP_d. Den fullständiga beräkningsalgoritmen för lokal poäng innehåller ytterligare variabler och parametrar.

Att poängen normeras med medelfelet innebär att man nu kan se till att det tillkommande felet – beroende på att vi underlåter att utreda misstänkta mätfel med liten förväntad effekt – blir litet i jämförelse med urvalsfelet. Att använda medelfelet i skattningen fungerar dock inte då vi har en totalräknad domän. Då normeras förväntad effekt i stället med punktskattningen. Denna växling till punktskattningen sker automatiskt i SELEKT.

Målet är att beräkna **global poäng** för objektet (primärt urvalsobjekt eller uppgiftslämnare). Aggregering av lokala poäng sker över fem nivåer (som mest):

5. domän
4. mått/variabel
3. sekundärt objekt (observationsobjekt)
2. primärt urvalsobjekt
1. uppgiftslämnare

Det finns olika beräkningsmetoder för dessa aggregeringar. Man kan t.ex. summera lokala poäng, summera kvadrerade poäng eller ta maxvärdet. I SELEKT anges detta val enkelt via en särskild parameter.

3.5 Bestämning av trösklar och parametrar

I en separat modul kallad Labbet testar man inställningarna inför en implementering. Här används data från undersökningens tidigare produktionsomgång, där såväl granskade som sparade ogranskade data finns.

Beräkningar av poäng sker enligt ovan med den skillnaden att vi nu kan beräkna faktisk effekt. För alla objekt under prövat tröskelvärde behålls

de ogranskade värdena, vilka sammantaget genererar en varians och en bias (ett systematiskt fel) i skattningarna.

Relativ pseudo-bias definieras i artikeln
Lawrence, D., & McDavitt, C. (1994). Significance Editing in the Australian Survey of Average Weekly Earnings, *Journal of Official Statistics*, Vol.**10**, No.4, 1994. pp. 437–447. Här med våra beteckningar.

$$RPB_{d,j} = \frac{|\sum_{k,l} w_{k,l}({}^{ogr}y_{j,k,l} - {}^{gr}y_{j,k,l})|}{SE(\hat{T}_{d,j})}$$

Tröskelvärdet fastställs så att huvuddelen av skattningarna får en acceptabelt låg nivå på denna bias. Observera att till skillnad från processindikatorer som visar effekt av granskning så vill vi här beräkna effekt på statistiken av att inte utreda de objekt och variabler som ligger under tröskeln. Ett problem är att detta kan vi göra i samband med att selektiv granskning implementeras men därefter har vi inte tillräckliga processdata om vi inte fortsätter att i någon mån även utreda objekt med global poäng under tröskelvärdet. En annan metod är att tillämpa modeller för fördelningen av mätfel, både vad avser förekomst, storlek och riktning. Då använder man ogranskade och granskade data över tröskeln för att simulera fel i data under tröskelvärdet.

I Labbet provas också olika parametervärden för att hitta den mest effektiva kombinationen av dessa. Emellertid är antalet kombinationer mycket stort och i praktiken får man begränsa sig till ett hanterligt antal av dessa. Med mer erfarenhet av verktyget kommer kunskapen att öka om vilka kombinationer som är mest effektiva för olika typer av undersökningar.

I verktyget SELEKT sker beräkning ofta för ett stort antal tabellceller, särskilt när man har en stor output. Det är dock viktigt att man gör en prioritering av vad som är de viktigaste måtten/variablerna och redovisningarna. Att utgå från att allt är lika viktigt innebär att ingen prioritering görs.

När man har använt selektiv granskning i en undersökning under några produktionsomgångar ska en ny inställning göras. Men då saknas granskade värden under den globala tröskel som använts. För att klara en ny analys i Labbet ska man normalt, inför en ny inställning, granska värden även under tröskeln. Detta kan göras på urvalsbasis. En förutsättning för att kunna justera inställningar från mer aktuella data är därför att detta planeras i god tid.

3.6 I god tid bör följande åtgärdas

- Ogranskade data ska sparas som underlag för att man ska kunna utvärdera existerande granskningskontroller och för att implementeringen av SELEKT ska kunna testas.

- Nuvarande granskningskontroller bör analyseras och om möjligt förbättras. Ta fram processdata för att se att acceptansgränser ligger rätt. Kan man ange misstanke på en kontinuerlig skala? Kan man ta bort någon onödig kontroll? Denna analys kan dock med fördel göras som en del av implementeringen så att implementeraren ytterligare lär sig undersökningen.
- Införandet ska ske i produkten så att förändringen stör så lite som möjligt och gärna vid till exempel urvalsbyte. Samordning med IT-systemansvarig är viktigt för att få datakommunikationen att fungera väl. Produkt- och metodansvarig har viktiga roller liksom produktionsansvariga på insamlingsenheten. All denna samordning innebär att man måste ha god framförhållning inför en implementering, cirka sex månader innan granskningen ska börja.