

Statistisk röjandekontroll

av tabeller, databaser och kartor

Område	0 - 999	1000 - 1999	2000 - 2999	3000+	Totalt
A	20	2	2	1	25
B	15	12	8	15	50
C	2	4	5	1	12
D	7	10	16	2	35
Totalt	44	28	31	19	122

Undertryckning av små cellvärden är *en* aspekt av statistisk röjandekontroll, men det finns så många fler ...

Förfrågningar: **Hans Block**,
SCB, Utvecklingsavdelningen, IT- enheten,
telefon 08-506 940 90,
hans.block@scb.se

Ann-Marie Flygare,
SCB, Utvecklingsavdelningen, Metodenheten,
telefon 019-17 64 16,
ann-marie.flygare@scb.se

Innehåll

Generaldirektörens förord	5
Förord av gruppen	5
1 Sammanfattning	7
2 Beskrivning av röjande och röjandekontroll	7
2.1 Ett första exempel	7
2.2 Definition av röjande	8
2.3 Röjanderisker och röjandeprocessen	8
2.4 Målet för statistisk röjandekontroll	9
3 Behov av röjandekontroll	10
3.1 Behov av sekretess	10
3.2 Intressekonflikten	13
4 Tabeller	15
4.1 Två grundtyper av tabeller	15
4.2 Riskbedömning	16
4.3 Skyddsmetoder 1: förändringar i tabeller	18
4.4 Skyddsmetoder 2: förändringar i mikrodata	23
4.5 Bedömning av informationsförluster	24
5 Flexibla databasuttag	24
5.1 Problemställningen	24
5.2 Skyddsmetoder	25
5.3 Praxis och riskbedömning	25
6 Kartor	27
6.1 Karttyper	27
6.2 Riskbedömning och skyddsmetoder	29
7 Metoder och programvaror i praktiskt bruk	31
7.1 Sammanfattning av utländska kontakter	31
7.2 Översikt av tillgänglig mjukvara	33
7.3 ARGUS	34
7.4 SuperCROSS	35
7.5 Läget på SCB	36
8 Riktlinjer	37
8.1 Giltighet	37
8.2 Känslighet och röjandekontroll	37
8.3 Organisatoriska åtgärder	37
8.4 Tabeller	37
8.5 Flexibla databasuttag	39
8.6 Kartor	39
Referenser	39

Generaldirektörens förord

Statistisk röjandekontroll har tagits fram inom ramen för SCB:s arbete med att skydda individens integritet vid publiceringen av statistiska material. Den redovisar bakgrunden till röjandeproblematiken och presenterar det regelverk som skall tillämpas vid SCB:s publicering. Den kan naturligtvis också tjäna som riktmärke vid annan publicering av statistik. Liksom med andra Current Best Methods (CBM) vid SCB kommer denna handbok att uppdateras då nya metoder kommer fram.

Svante Öberg

Förord av gruppen

Röjandeprojektet initierades av Lars Lyberg och Per Samuelsson och har drivits inom metodenheten på SCB under tiden oktober 1999 – juni 2001. I projektet har deltagit Hans Block, Ann-Marie Flygare, Ulf Jorner, Lars Lyberg, Gösta Nilsson, Lars Nordbäck, Per Samuelson och Elisabeth Sennvall.

Målet för projektet har varit att ta fram enhetliga regler och metoder för röjandekontroll och att undersöka möjligheterna att stödja arbetet med tekniska hjälpmedel.

Avsikten var att rapporten skulle täcka alla statistiska material, även mikrodata. Emellertid är röjandegranskningen av avidentifierade mikrodata helt beroende av SCB:s utlämnandepolicy. Denna är för närvarande under diskussion. Nästa utgåva av denna rapport avses även omfatta metoder för röjandekontroll av mikrodata och regler för när dessa metoder skall tillämpas.

En väsentlig del av projektet har syftat till kunskapsuppbyggnad, där den första aktiviteten var litteraturstudier. En enkät för att ta reda på hur statistisk röjandekontroll för närvarande utförs på SCB har skickats ut till samtliga statistikprogram inom SCB. Kontakter med ungefär samma uppläggningsomfattning som enkäten har också tagits med några andra statistikansvariga myndigheter. Programvaran ARGUS har provats ut. Förfrågningar om metoder och regler för röjandekontroll har gått ut till centralbyråer eller motsvarande i Nederländerna, USA, Kanada och Australien.

Hans Block och Ann-Marie Flygare står för huvuddelen av texten i denna handbok. Kapitel 3 har skrivits av Per Samuelsson och uppdaterats av Birgitta Pettersson.

1 Sammanfattning

All statistikproduktion skall ske så att röjande av enstaka objekt undviks. Rapporten börjar med att definiera röjande och utreda behovet av röjandekontroll. I de följande avsnitten beskrivs tekniken för röjandekontroll för olika presentationsformer: tabeller, flexibla databasuttag och kartor. Metoder och regler för avidentifierade mikrodata kommer att ges i en senare upplaga. En genomgång görs av praxis vid några ledande statistikproducenter i världen och inom SCB. Funktionerna i programvaran ARGUS beskrivs.

Sammanfattningsvis gäller följande riktlinjer:

- Cheferna skall ansvara för att röjandefrågor hanteras på ett lämpligt sätt.
- Tekniska röjandefrågor behandlas i en särskild grupp med representanter från U-avdelningen, ämnesavdelningarna och IP.
- Röjandekontroll skall genomföras för alla variabler, alltså även för icke känsliga variabler.
- Nedanstående regler och parametrar för att identifiera riskceller i tabeller och kartor skall följas.
- Utvecklingen av programvaran ARGUS skall följas aktivt med avsikt att ARGUS eller dess efterföljare på sikt skall kunna rekommenderas till såväl tabeller som mikrodata.

2 Beskrivning av röjande och röjandekontroll

2.1 Ett första exempel

Innan vi börjar beskriva problematiken, ger vi ett första exempel, direkt taget ur företagsstatistiken. SCB skulle publicera den totala omsättningen fördelad på bransch enligt SNI 4-siffernivå. I en tabellcell dominerade två företag och tre var mycket små. Summan var 3 295 000. Det största företagets omsättning var 2 329 000, det näst största 921 000. VD:n i det näst största företaget visste givetvis sin egen omsättning, vem som var den störste konkurrenten, och kunde gissa att de andra konkurrenterna tillsammans hade mindre än hälften av hans egen omsättning. Om SCB hade publicerat totalen, så hade VD:n kunnat dra slutsatsen, att största företagets omsättning låg mellan

$$3\,295\,000 - 1,5 \cdot 921\,000 \text{ och } 3\,295\,000 - 921\,000$$

eller i intervallet

$$(1\,913\,000, 2\,374\,000)$$

Dessa uppskattningar ansågs vara alldeles för noggranna, så SCB undertryckte därför denna tabellcell före publiceringen.

2.2 Definition av röjande

Innan behoven av röjandekontroll diskuteras är det lämpligt att definiera begreppet röjande. Vi avstår tills vidare från den juridiska definitionen och ger i stället följande förklaring, som anknyter till statistisk forskning i en rad länder med skilda rättssystem. Förklaringen har sedan länge använts inom SCB:

Ett *röjande* föreligger när en ”angripare¹” med hjälp av statistiskt material – t.ex. tabeller eller avidentifierade filer – egen bakgrundskunskap och logiska slutledningar, med eller utan maskinell hjälp, får ny kunskap – säker eller med viss sannolikhet – om känsliga egenskaper hos enskilda objekt.

Definitionen kräver förtydliganden.

2.3 Röjanderisker och röjandeprocessen

Angriparen kan vara en vanlig statistikanvändare eller någon som har ett bestämt uppsåt att söka den känsliga kunskapen. Röjandet kan uppstå vid publicerandet av *tabeller*, särskilt när det är få objekt som bidrar till en viss tabellcell. Även tabeller som inte innehåller rena antals- eller mängdvariabler kan medföra röjanden. *Statistiska mått*, exempelvis medelvärden och spridningsmått, kan också medföra risker.

Större risk för röjande föreligger dock när statistikproducenten låter användaren själv *välja sina tabeller från en databas* av finfördelade makrodata eller t.o.m. mikrodata. Ännu lättare är det att dra slutsatser från en avidentifierad *mikrodatafil*. Detaljerade *kartor* över statistiska variabler kan också medföra röjanden.

Den *bakgrundskunskap* som angriparen har varierar från fall till fall. Inom företagsstatistiken kan det gälla en företagare som vet sin konkurrents lokalisering, bransch och ungefärliga storlek på omsättning eller antal anställda. Inom individstatistiken känner en vanlig statistik Konsument ofta till

¹ Termen ”angripare” har valts för att det skall vara lätt att förstå vad som avses. Den engelska termen är *intruder*, *inkräktare*, vilket låter mindre illvilligt, men lockar till felaktiga associationer. Termen *obehörig* kan inte heller användas, eftersom alla är behöriga att ta del av publicerat statistiskt material. Trots den negativa klangen ansluter denna rapporten till en svensk tradition och brukar ordet angripare för den som utnyttjar möjligheterna till röjande.

en bekants yrke, utbildning, ungefärliga ålder för honom² själv och hans familjemedlemmar. Någon gång kan angriparen förutsättas ha ett helt register till sitt förfogande.

Angriparen kan dra sina *slutsatser* på olika sätt. I de enklaste fallen rör det sig om att hitta ett enskilda objekt bakom en tabellcell. I svårare fall får statistikproducenten räkna med risken att angriparen ser på rad- och kolumnsummor och gör omfattande jämförelser och räkningar mellan olika tabeller för att till slut återskapa små tabellceller. Om angriparen har fått en oidentifierad mikrodatafil för statistisk bearbetning, så måste man också räkna med att han kan göra omfattande matchningar.

Ett röjande är inte alltid *säkert*. Identifieringen av en enskild sker inte alltid med full visshet. Statistiska material innehåller alltid osäkerheter, inaktuella uppgifter eller rena fel. Angriparen kan också dra felaktiga slutsatser från sitt material. Från den röjda individens synpunkt kan det vara värre att en felaktig uppgift kommer ut än att sanningen kommer fram. Även osäkra och felaktiga röjanden måste därför försvåras.

Vad som räknas som *känsligt* finns det många åsikter om. Åsikterna varierar såväl mellan länder som över tiden. Statistikproducenterna har sina uppfattningar, uppgiftslämnarna sina. I 13 § personuppgiftslagen definieras känsliga personuppgifter som uppgifter om ras eller etniskt ursprung, politiska åsikter, religiös eller politisk övertygelse, medlemskap i fackförening, samt personuppgifter om hälsa eller sexualliv. Sekretesslagen talar dock inte om vad som är känsligt, utan bara om att den som uppgiften rör eller honom närstående inte skall lida skada eller men. SCB skyddar därför *alla* uppgifter som rör enskilda, även sådana som inte upplevs som särskilt känsliga.

Definitionen ovan talar enbart om enskilda objekt. Statistiken skall inte skydda s.k. gruppintegritet. Om statistiken t.ex. skulle visa att oväntat många i en viss grupp av individer har ofördelaktiga egenskaper, så är detta ett intressant rön, och inte något som skall döljas.

2.4 Målet för statistisk röjandekontroll

En statistikproducent kan i praktiken aldrig helt eliminera risken för röjanden. Däremot kan han bedöma riskerna och vidta åtgärder som skall minska riskerna. Därefter kan han bedöma åtgärdernas effekt i form av ökad säkerhet och förlust av information. Sådana bedömningar och åtgärder bör rutinmässigt ingå i den statistiska produktionsprocessen.

Denna rapport beskriver risker och åtgärder för de olika publiceringsformer som tillämpas i SCB:s statistikproduktion. En förhoppning är att framställ-

² I denna rapport används pronomina som han/honom/hans även när dessa skulle kunna syfta på en kvinna. Varken SCB eller någon av författarna till rapporten har således velat antyda att uppgiftslämnare, statistikproducenter eller s.k. angripare oftare skulle vara män än kvinnor. Dessa pronomina har valts enbart för att göra texten mera kompakt.

ningen skall öka medvetenheten om röjandeproblemen såväl inom SCB som inom övriga statistikansvariga myndigheter. På sikt kan detta leda till åtgärder som kan utföras på ett rationellt sätt i statistikproduktionen, är rimligt effektiva mot röjanden och inte förorsakar alltför stora informationsförluster. Som följd av detta minskar risken för att den statliga statistikproduktionen utsätts för förtroendekriser.

3 Behov av röjandekontroll

3.1 Behov av sekretess

Behovet av sekretess gäller från två sidor. För det första är det ett oavvisligt krav från de **fysiska och juridiska personer som omfattas** av de uppgifter som skall användas för att framställa statistiken, att uppgifterna om dem inte avslöjas. Att skydda uppgifterna är ett åtagande som SCB och andra statistikansvariga myndigheter (SAM) gjort i förhållande till dem som omfattas av uppgifterna. Särskilt tydligt är detta vid undersökningar där uppgifterna hämtas in direkt från den som berörs av uppgiften.

Den andra sidan är behovet från **statistikproducentens** sida att faktiskt få den som avses med dessa uppgifter att känna förtroende för att det utlovade skyddet håller.

Båda dessa sidor samverkar till att uppnå hög svarsfrekvens i statistiska undersökningar och därmed bra statistik, vilket alla är betjänta av. I detta sammanhang diskuteras inte hur höga svarsbortfall kan kompenseras för och andra åtgärder som sammanhänger med detta, utan här konstateras endast att hög svarsfrekvens är utslagsgivande för kvaliteten i en statistisk undersökning, givet förstås att svaren är korrekta.

Det är som nämnts ett viktigt villkor att allmänheten – och därmed uppgiftslämnarna – har förtroende för SCB:s handhavande av de uppgifter som verket förfogar över. Hur detta förtroende förvärvas och vidmakthålls är en fråga som inte är alldeles lätt att besvara. Ett förtroende måste byggas från grunden, vilket innebär att redan allmänhetens tillit till statsapparaten är en faktor i detta avseende. Nästa nivå är att allmänheten har förtroende för den verksamhet som SCB bedriver, och här kanske också förtroendet för andra institut av motsvarande slag spelar in. Sista nivån är förtroende för det faktiska handlandet vid SCB.

Det som avhandlas här är förtroendet för SCB:s hantering av de uppgifter som berör enskilda personer och företag och som används för statistikframställning. På detta finns många infallsvinklar. En är att gällande rättsregler innebär ett sekretesskydd för uppgifter som kan knytas till enskilda (personer eller företag). En annan är att säkerheten är utformad så, att obehöriga inte kan komma åt dessa uppgifter. En ytterligare är att statistiken inte

redovisas så, att man i en tabellcell med en till visshet gränsande sannolikhet kan avslöja en viss fysisk eller juridisk person. Även de etiska regler som finns för statistiker är en aspekt som är relevant i sammanhanget, eftersom dessa också innebär att statistikerna inte någonsin skall yppa något som kan vara menligt för de enskilda personer som ingår i ett visst material som skall användas för statistikframställning.

3.1.1 Rättsregler

Skyddet för enskilda fysiska och juridiska personer finns i olika författningar. Det som skiljer SCB och andra statistikproducenter inom systemet för officiell statistik från privata statistiska institut är att de uppgifter som samlas in av de statistikansvariga myndigheterna omfattas av skydd enligt sekretesslagen. Detta skydd är ett undantag från offentlighetsprincipen, som annars gäller generellt inom statsförvaltningen. Offentlighetsprincipen, som i och för sig är en viktig del av demokratin, kan alltså inte gälla blint utan måste förses med vissa inskränkningar för att andra intressen tar över. Hur Sverige planerar sitt försvar är kanske det tydligaste exemplet på att ett undantag från offentligheten är motiverat. Men även hänsynen till enskilda personers personliga och ekonomiska förhållanden tar i vissa fall över offentlighetsprincipen. Det är således dessa hänsyn som ligger bakom sekretessen för uppgifter till statistiken.

Enligt 9 kap. 4 § sekretesslagen (1980:100) gäller sekretess i sådan särskild verksamhet hos en myndighet som avser framställning av statistik för uppgift om enskilda personliga eller ekonomiska förhållanden. Sekretessen innebär ett förbud att röja uppgiften. Med uppgifter om *enskilda* avses både fysiska och juridiska personer. Sekretessen är enligt huvudregeln *absolut*, vilket innebär att uppgiften inte får lämnas ut. Sekretesstiden är sjuttio år när det gäller uppgifter om personliga förhållanden och annars tjugo år. Eftersom sekretessen är begränsad till att skydda enskilda, gäller den inte uppgifter om statliga eller kommunala myndigheter.

I den nämnda sekretessbestämmelsen finns dock vissa undantag från sekretessen. Uppgiften får t.ex. lämnas ut i oidentifierad form och för forsknings- och statistikändamål. Det krävs dock att *det står klart* att uppgiften kan lämnas ut utan att den som avses med uppgiften eller någon närstående lider skada eller men. Det får alltså inte finnas skuggan av ett tvivel om att uppgiften kan lämnas ut riskfritt för den som berörs av uppgiften. I detta sammanhang är undantag från sekretessen i fråga om oidentifierade uppgifter av särskilt intresse. Det står så här i lagtexten: ”Uppgift ... som inte genom namn, annan identitetsbeteckning eller därmed jämförbart förhållande är direkt hänförlig till den enskilde, får dock lämnas ut, om det står klart att uppgiften kan röjas utan att den som uppgiften rör eller honom närstående lider skada eller men”. Dyliga uppgifter (d.v.s. ”oidentifierade” uppgifter eller uppgifter i tabeller) är alltså i princip sekretesskyddade om det på något sätt går att hänföra dem till en enskild, och de får bara lämnas ut

om det kan ske riskfritt. Det gäller förstås även vid publicering eller annat offentliggörande.

För röjandeprojektet är denna reglering något av fundamentet för arbetet. Det är ett åliggande att se till att material som görs tillgängligt för utomstående inte innehåller uppgifter som kan hänföras till någon enskild. Skulle materialet innehålla sådana uppgifter, måste det som sagt *stå klart* att det inte är till men eller skada för den som avses med uppgiften. Därmed kan man slå fast att varje publicering av statistik måste föregås av en sådan kontroll av materialet att det är säkerställt att det inte kan ge upphov till några röjanden!

I sammanhanget bör nämnas att sedan den 1 april 2001 gäller en ny regel som innebär att statliga myndigheter skall lämna de uppgifter som behövs för framställning av officiell statistik (6 § statistikförordningen 2001:100). Uppgiftsskyldigheten gäller även för statistikansvariga myndigheter, bl.a. SCB, och bryter således sekretessen vid SCB.

Skyddet av den personliga integriteten har diskuterats under många år nu, delvis som en följd av datoriseringen i samhället. Sverige var tidigt ute med en nationell lagstiftning på detta område: en *datalag* fanns redan 1973. Även EU har tagit fram ett regelverk med avsikt att skydda integriteten för EU-medborgarna. Konstruktionen gjordes så att det skulle ankomma på medlemsländerna att införa dessa regler i nationell lagstiftning inom tre år. Sverige lät en kommitté svara för denna ”nationalisering” och en ny lag, *personuppgiftslagen* (1998:204), trädde i kraft den 24 oktober 1998.

Personuppgiftslagen innehåller olika bestämmelser om när behandling av personuppgifter får ske. Ändamålet med en behandling är alltså centralt, men även känsligheten hos de uppgifter som behandlas är viktig och får olika konsekvenser för vad som är tillåtet beroende på graden av känslighet i materialet. Enligt personuppgiftslagen är framställning av statistik ett sådant allmänt intresse som gör att behandlingen av personuppgifter är tillåten. Vidare är framställning av statistik alltid förenlig med de andra ändamål för vilka uppgifterna ursprungligen samlats in. De uppgifter som man lämnar till skattemyndigheten för taxering får alltså också användas för statistikändamål.

Officiell statistik har reglerats särskilt i Sverige. Bestämmelserna finns i lagen (2001:99) om den officiella statistiken och en förordning som ansluter till lagen.

Det som sägs är att den officiella statistiken skall framställas och offentliggöras med beaktande av skyddet för enskilda. Här skall ”enskilda” tolkas som både fysiska och juridiska personer. Vidare finns en särskild bestämmelse som innebär att det är förbjudet att sammanföra uppgifter i den officiella statistiken med andra uppgifter i syfte att utröna enskildas identitet. Till förbudsregeln är kopplad en straffbestämmelse. Bakgrunden till denna

bestämmelse är att erfarenheten har visat att det i några enstaka fall gjorts försök att bryta igenom anonymiteten och att detta har kunnat leda till att en identitet avslöjas. Detta framgår av regeringens proposition när lagen om den officiella statistiken infördes.

För röjandeprojektets del innebär den sist nämnda bestämmelsen att den sätter en gräns för hur långt granskningen av eventuella röjanden skall sträcka sig. Någon systematisk kontroll görs således inte av att det är omöjligt att plocka ihop ett antal tabeller så att det i något gränssnitt mellan dessa – i en tabellcell eller dylikt – skulle kunna falla ut en enskild enhet.

Det finns också en yrkesetisk deklARATION som Internationella Statistiska Institutet antog 1985 och som SCB senare också beslutat skall gälla för de anställda. Deklarationen innehåller rättesnören i snart sagt samtliga sammanhang där en statistiker utövar sin yrkesverksamhet. Det som förtjänar att särskilt lyftas fram i detta sammanhang är följande uttalande: ”Statistiker bör vidta lämpliga åtgärder för att förhindra att deras data publiceras eller på annat sätt släpps ut i sådan form att någon undersökningsspersonens identitet skulle kunna röjas eller gissas.” (Översättning till svenska av SCB.) Återigen finns här ett berättigat intresse som grund för röjandeprojektets arbete.

3.2 Intressekonflikten

I SCB:s roll ingår att så långt möjligt tillgängliggöra statistik, så länge man inte ger avkall på kvalitetskraven. En del av denna roll är att offentliggöra statistik som verket har statistikansvar för. I egenskap av producent av statistik för andra statistikansvariga myndigheter kan det också ligga i producentrollen att offentliggöra statistiken.

SCB:s policy beträffande offentliggörandet innebär att det är kravet på snabbhet och användarnas behov som sätts i fokus. Och användarnas behov gäller också, som framgått, bl.a. detaljeringsgraden. Det är just i detta avseende som en intressekonflikt seglar upp. Ju större detaljeringsgraden är i de tabeller som offentliggörs är, desto större är risken att uppgifter om en enskild kan röjas. Hur stor detaljeringsgraden kan vara utan att äventyra den enskildes skydd går inte att ange generellt, utan detta måste avgöras från fall till fall – från tabell till tabell.

En trygg lösning på konflikten är att minska detaljeringsgraden i en statistisk redovisning när man förutser en risk för röjande. En redovisning på församlingsnivå ändras till redovisning på kommunnivå, kommunnivå till länsnivå, länsnivå till riksnivå. Och problemet är borta! Men med det också en del av användbarheten. Det är alltså inte en möjlig lösning att helt sonika ändra redovisningsnivån, eftersom informationsförlusten blir så stor. Det säger sig självt att en användare som skall studera vissa förhållanden inom en kommun inte är betjänt av att få en tabell för hela det län inom vilket kommunen

ligger. Därför måste konflikten lösas på annat sätt, och det är här röjandeprojektet skall tillhandahålla lösningar.

Statistikansvariga myndigheter tillhandahåller också material som inte tillhör den officiella statistiken till utomstående. Det är både tabeller och av-identifierat material och det är till olika användare. När det gäller tillhandahållande generellt av denna typ av material, gäller samma regler som för officiell statistik i det avseendet att något material som kan innehålla röjanden inte får förekomma. Generellt gäller också att det material som tillhandahålls med samma villkor är tabellmaterial.

SCB tillhandahåller inte villkorslöst avidentifierat material till allmänheten. Verket har inte beträtt den väg som är vanlig i t.ex. USA, som innebär att man tillhandahåller s.k. Public Use Files. Dessa är både avidentifierade och fråntagna andra identifierande uppgifter i sådan grad att det överhuvudtaget inte är möjligt att knyta de kvarvarande uppgifterna till någon enskild. I Sverige har vi inte upplevt någon större efterfrågan på denna typ av filer. Det skulle också innebära en risk för oro av mer allmänt slag hos uppgiftslämnare till statistiken om sådana filer blev spridda i samhället och med det en känsla av att *"jag finns i materialet och kanske ändå kan avslöjas!"* Det är just uppkomsten av denna typ av oro som måste förhindras och som nämnts under avsnitt 3.1.

SCB har fastställt en sekretesspolicy i september 1999. Denna innebär ingen kursändring från tidigare, utan kan sägas vara en beskrivning av nuvarande praxis vid SCB:s tillämpning av statistiksekretessen.

Innehållet i sekretesspolicyn är i korthet att SCB normalt inte lämnar ut identitetsangivet material. Avidentifierat material kan efter prövning lämnas ut till myndigheter till forskning och för statistikändamål. Sådant material omfattas av sekretess hos mottagaren och skyddas alltså på samma sätt som det skyddas hos SCB. För utredningsarbete kan avidentifierat material också lämnas ut, men om mottagaren inte har en "egen" sekretess omfattas materialet inte av sekretess och prövningen måste alltså ske med tanke på att vem som helst har rätt att få del av materialet – bara på pappersmedium visserligen – enligt offentlighetsprincipen. I policyn sägs om detta att om prövningen leder till att ett utlämnande skall ske "... måste materialet granskas särskilt för att säkerställa att skada eller men inte kan uppkomma till följd av att materialet inte kommer att vara sekretesskyddat hos mottagaren". Det är alltså en typ av röjandegranskning som skall till. För att göra denna granskning och sedan bedöma vilka förändringar i materialet som behövs kan röjandeprojektets förslag följas.

4 Tabeller

4.1 Två grundtyper av tabeller

Innan någon metod för röjandeminimering kan användas, måste man identifiera celler där eventuella röjanden skulle kunna ske. Vilken regel som skall användas för identifiering av dessa celler beror på om det är en *frekvenstabell* eller en *tabell med summerbar innehållsvariabel*, exempelvis omsättning eller total inkomst.

Exempel. En enkel *frekvenstabell* (enbart antalsuppgifter) i två dimensioner:

Folkmängd den 1 november 2000

Län	Totalt	därav Män	Kvinnor
Riket	8 880 620	4 391 246	4 489 374
Stockholms län	1 822 224	891 014	931 210
Uppsala län	293 939	144 678	149 261
Södermanlands län	255 821	126 625	129 196
---	---	---	---

Exempel. En tabell med summerbar innehållsvariabel (konsumtion). Uppgifterna är hämtade från SCB:s hemsida under Ekonomi, Nationalräkenskaper och används för att definiera terminologin i denna rapport.

Hushållens årliga konsumtion på ändamål. Mkr

	Löpande priser	
	1999	2000
Livsmedel och icke alkoholhaltiga drycker	122 121	125 688
Alkoholhaltiga drycker och tobak	40 033	41 907
Beklädnadsartiklar och skor	52 391	55 597
Bostad, el, gas och uppvärmning	295 967	297 497
Möbler, hushållsartiklar och rutinunderhåll	46 482	50 439
Hälsa- och sjukvård	23 379	24 751
Transporter	127 433	133 009
Kommunikation	29 032	28 212
Fritid, underhållning och kultur	106 335	112 662
Utbildning	1 704	1 680
Hotell, caféer och restauranger	49 404	52 866
Övriga varor och tjänster	72 592	79 175
Summa ändamålsfördelat	966 873	1 003 483

Spridningsvariabler är ändamål och tiden. Varje cell utgörs av en summa. Exempelvis är värdet i första kolumnen och första raden Sveriges totala konsumtion av livsmedel och icke alkoholhaltiga drycker under år 2000. I och med att värdet i varje tabellcell är en summa så är det möjligt att summera varje kolumn i tabellen. Att summera en rad över tiden är däremot sällan meningsfullt.

Summerbara variabler kan även presenteras som medelvärden i tabellcellerna. Summering rad- och kolumnvis blir då inte meningsfull. Bedömningen av riskceller är dock likartad i dessa fall. Däremot blir skyddsåtgärderna inte lika enkla som de som beskrivs i avsnitt 4.3.

4.2 Riskbedömning

4.2.1 Riskbedömning i frekvenstabeller

I frekvenstabeller bidrar alla uppgiftslämnare lika mycket till cellvärdena. Antingen är det antalet, andelen eller procentandelen objekt tillhörande en cell som publiceras. Problem med röjanderisk uppstår när celler endast består av ett fåtal uppgiftslämnare. För att undvika röjanden tillämpas röjandeminimeringsmetoder på celler med mindre än ett på förhand specificerat antal. Denna regel för identifiering av *riskceller* kallas **tröskelvärdesregeln**.

Ibland kan röjande ändå ske i celler med litet högre frekvenser, om samtliga celler i en viss kategori med värden skilda från 0 har ”känsliga” värden.

Exempel. Alla i en grupp har känsliga värden.

Antal invånare i ... fördelade på ålder och missbruksproblem

Ålder	Totalt	Därav alkoholmissbrukare

60 – 69	5	5

Om någon i en kommun har läst ovanstående lilla tabell om kommunen, och gått på 65-årskalas, så kan han dra slutsatsen att värden är alkoholist!

4.2.2 Riskbedömning i tabeller med summerbar innehållsvariabel

I tabeller med summerbar innehållsvariabel innehåller värdet i varje cell information från alla uppgiftslämnare som tillhör cellen (att jämföras mot frekvenstabellen där endast respondentens tillhörighet registreras). Värdet i varje cell kan till exempel vara summan, medelvärdet eller medianen av uppgiftslämnarnas bidrag.

Olika uppgiftslämnare ger olika stora bidrag till det slutliga cellvärdet, vilket kräver en annan regel än tröskelvärdesregeln för att upptäcka riskceller.

Regler som används för identifiering av riskceller i dessa tabeller är:

- **(n, k)-regeln (dominansregeln).** Regeln identifierar celler där de n största uppgiftslämnarna bidrar med $100k$ % eller mer till det totala cellvärdet.

Exempel. Om $(1, 0.5)$ -regeln skall användas, innebär det att en uppgiftslämnare inte får bidra med 50 % eller mer till det totala cellvärdet. Re-

geln $(2, 0.9)$ innebär att två objekt tillsammans inte får bidra med 90 % eller mer till det totala cellvärdet.

- **p -procentsregeln** identifierar celler där ingen enskild uppgiftslämnarens bidrag kan uppskattas så att felets absolutbelopp blir $< p$ % av det sanna värdet. Det räcker med att titta på respondenten med det största bidraget till cellvärdet för att även skydda resterande uppgiftslämnare. Denna regel synes bäst avspegla uppgiftslämnarnas skyddsbehov, om man kan göra rimliga antaganden om bakgrundskunskaper. För en utförligare förklaring av denna regel, se *Subcommittee on disclosure limitation methodology*.

Exempel. Till en tabellcell bidrar fem företag med omsättningen 10000, 5000, 3000, 2000 och 1000 kSEK. Cellvärdet blir summan = 21000. Det näst största företaget vet sin egen omsättning, och förutsätts veta vilken konkurrent som är störst och dessutom kunna uppskatta sina övriga konkurrenters omsättning med ett fel om högst 100 % av det sanna värdet. Detta innebär, att det största företags omsättning är större än

$$21000 - 5000 - 6000 - 4000 - 2000 = 4000$$

men mindre än

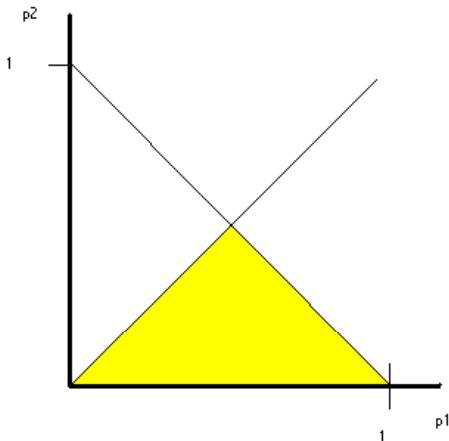
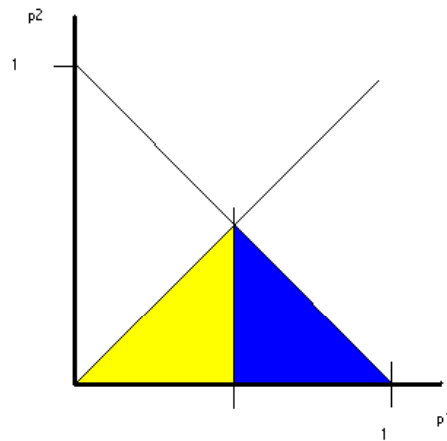
$$21000 - 5000 = 16000$$

Denne användare kan alltså uppskatta en uppgift med ett fel på högst 6000 = 60 % av det sanna värdet, och p -procentsregeln skulle förklara cellen som riskcell, om p valdes > 60 .

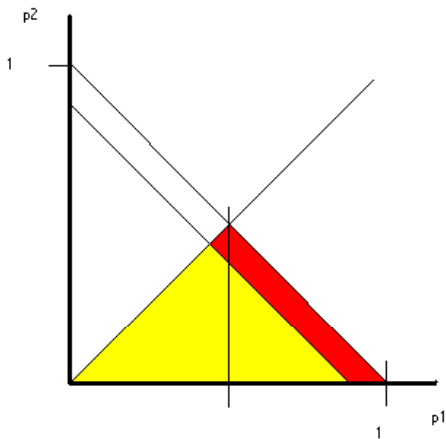
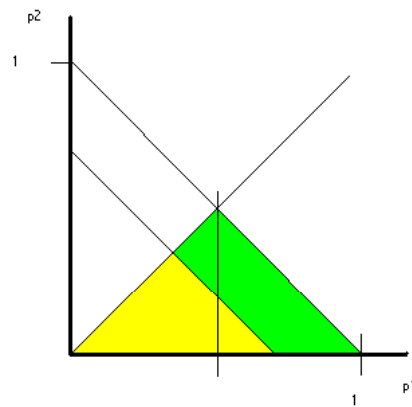
- **pq -regeln:** Parametern p är som ovan och q anger hur nära man tror att en respondent kan uppskatta en annan respondents bidrag. Denna regel är strängare än den förra för samma värde på p , vilket medför en större informationsförlust.

4.2.3 Val av parametrar till dominansregeln

För att illustrera hur olika parametrar till dominansregeln begränsar cellvärdena ritar vi några enkla figurer. Vi antar, att cellvärdet är 1 och att de två största värdena är p_1 och p_2 , $p_1 \geq p_2$ och $p_1 + p_2 \leq 1$. Det tillåtna området (gult) blir som framgår av figur 1. I figur 2 har vi tillämpat 50 %-regeln $(1, 0.50)$ och tagit bort ett blått område:

Figur 1. Ingen regel.**Figur 2.** Regel $(1, 0.5)$ (blått).

Vi illustrerar också dominansregeln med andra parametrar:

**Figur 3.** Regeln $(2, 0.90)$ (rött).**Figur 4.** Regeln $(2, 0.70)$ (grönt)

Båda reglerna medför, att minst tre element måste finnas i cellen.

SCB har tidigare använt 50 %-regeln och 3-regeln. 3-regeln ger dåligt skydd om det tredje elementet är mycket litet, och det är därför naturligt att skärpa denna regel till $(2, 0.90)$. Denna kombination rekommenderas i avsnitt 8.4.2.

Av figurerna framgår, att reglerna $(1, 0.50)$ och $(2, 0.90)$ inte överlappar varandra. Om man i stället skulle använda den ännu skarpare regeln $(2, 0.70)$, så skulle denna nästan kunna ersätta 50 %-regeln.

4.3 Skyddsmetoder 1: förändringar i tabeller

4.3.1 Undertryckning

Celler som identifierats som riskceller undertrycks, *primärundertryckningar*. Om marginaltotaler publiceras, måste även vissa andra celler undertryckas,

sekundärundertryckningar, så att det inte skall vara möjligt att med hjälp av någon marginaltotal räkna fram de undertryckta cellernas värde. Följande aspekter är viktiga vid val av celler för sekundärundertryckningar:

- En intervallskattning av ett undertryckt värde får ej vara för kort.
- Informationsförlusten skall minimeras.

Efter det att undertryckningar utförts (både primära och sekundära), är det viktigt att kontrollera att det inte går att räkna fram något enskilt undertryckt värde genom linjärkombinationer av rader och kolumner i tabellen. Exempelvis räcker det inte alltid att ha två undertryckningar i varje rad och kolumn. Ibland kan man ändå räkna fram ett enskilt värde genom att kombinera rader och kolumner.

Informationsförlusten kan mätas genom

- antalet undertryckta celler
- antalet undertryckta uppgiftslämnare
- summan av undertryckta värden.

Vilka celler som skall sekundärundertryckas bör man bestämma genom att göra något av dessa mått så litet som möjligt. Att finna det allra bästa sättet att göra sekundärundertryckningar kan kräva mycket omfattande beräkningar.

Exempel. Betrakta följande hypotetiska tabell:

Antalet socialbidragstagare efter område och bidragsstorlek (kr)

Område	0 - 999	1000 - 1999	2000 - 2999	3000+	Totalt
A	20	2	2	1	25
B	15	12	8	15	50
C	2	4	5	1	12
D	7	10	16	2	35
Totalt	44	28	31	19	122

Om vi använder tröskelvärdesregeln med parametern 3, så är cellerna med fetstil, rosa bakgrund riskceller, som skall primärundertryckas:

Område	0 - 999	1000 - 1999	2000 - 2999	3000+	Totalt
A	20	2	2	1	25
B	15	12	8	15	50
C	2	4	5	1	12
D	7	10	16	2	35
Totalt	44	28	31	19	122

Sekundärundertryckningar (markerade med kursiv fetstil, gul bakgrund) skulle kunna väljas på följande sätt:

Område	0 - 999	1000 - 1999	2000 - 2999	3000+	Totalt
A	20	2	2	1	25
B	15	12	8	15	50
C	2	4	5	1	12
D	7	10	16	2	35
Totalt	44	28	31	19	122

När man ser på denna tabell kan man lätt tro att man skyddat den tillräckligt, eftersom antingen minst två eller inga celler ur varje rad och kolumn är undertryckta. Men det finns risker. Om vi betecknar de nio undertryckta cellvärdena med U_1 till U_9 :

Område	0 - 999	1000 - 1999	2000 - 2999	3000+	Totalt
A	20	U_1	U_2	U_3	25
B	15	U_4	U_5	15	50
C	U_6	4	5	U_7	12
D	U_8	10	16	U_9	35
Totalt	44	28	31	19	122

Ser man på några linjärkombinationen av rader och kolumner:

$$\text{rad}_1 + \text{rad}_2 - \text{kolumn}_2 - \text{kolumn}_3 =$$

$$(20 + U_1 + U_2 + U_3) + (15 + U_4 + U_5 + 15) -$$

$$- (U_1 + U_4 + 4 + 10) - (U_2 + U_5 + 5 + 16) = 25 + 50 - 28 - 31$$

så får man att $U_3 = 1$ och en riskcells exakta värde har avslöjats, vilket kan leda till ett röjande.

I stället skulle följande undertryckningar föredras eftersom en mindre sammanlagd frekvens undertrycks:

Område	0 - 999	1000 - 1999	2000 - 2999	3000+	Totalt
A	20	2	2	1	25
B	15	12	8	15	50
C	2	4	5	1	12
D	7	10	16	2	35
Totalt	44	28	31	19	122

Detta mönster är dessutom mycket säkrare. Undertryckningar leder dock aldrig till fullständig osäkerhet om cellvärdena. I sista kolumnen är t.ex. summan av den undertryckta frekvensen 4, och alla okända cellers värden måste då ligga i intervallet 0 – 4. På samma sätt ger första raden att de undertryckta värdena måste ligga i intervallet 0 – 5. I tredje kolumnen undertrycks sammanlagt frekvensen 7, varav högst 5 i första raden. Man får alltså intervallet 2 – 7 för sekundärundertryckningen i tredje raden. Liknande

resonemang leder efter en del räkningar till nedanstående intervall, som efter kontroller visar sig vara kortast tänkbara. Om man dessutom vet hur undertryckningarna har gått till, så kan man dra ännu fler slutsatser. Elementet (C, 3000+) behöver inte vara sekundärundertryckt. Alltså är det primärundertryckt och därmed ≤ 2 .

Område	0 - 999	1000 - 1999	2000 - 2999	3000+	Totalt
A	20	0 - 5	0 - 5	0 - 4	25
B	15	12	8	15	50
C	0 - 4	1 - 6	2 - 7	0 - 4	12
D	5 - 9	10	16	0 - 4	35
Totalt	44	28	31	19	122

Dessa intervall är så långa, att ingen riskcell kan uppskattas med mindre fel än 100 % av sitt sanna värde. De mödosamma räkningarna visar också, att metoden kräver datorstöd.

Ibland kan en angripare uppskatta ett undertryckt värde, trots att det inte finns några summor. Detta kan inträffa i tidsserier, då vissa värden ändras obetydligt mellan åren.

4.3.2 Aggregering

Efter att ha tillämpat någon regel för identifiering av riskceller kan man göra en grövre indelning (aggregering), antingen genom att slå samman klasser i rader eller kolumner eller genom att göra en ny klassindelning. Därefter fortsätter granskningen av tabellen för att ta bort återstående riskceller. Nedan visas exempel på tänkbara sammanslagningar:

Område	0 - 999	1000 - 1999	2000+	Totalt
A	20	2	3	25
B	15	12	23	50
C + D	9	14	24	47
Totalt	44	28	50	122

Efter sammanslagningen av klasser återstår bara en riskcell. Ett beslut får fattas om detta är acceptabelt eller om ytterligare åtgärder måste vidtas.

4.3.3 Sluppmässig avrundning

Informationen i alla celler (känsliga och icke-känsliga) avrundas till en multipel av en förutbestämd bas b . Cellvärdet kan skrivas som

$$X = bq + r$$

där X är cellvärdet, q heltalsdelen av X/b och r resten,

$$0 \leq r < b, \quad 0 \leq q \leq \frac{X}{b}$$

Efter slumpmässig avrundning blir det nya cellvärdet

$$X' = \begin{cases} b(q+1) & \text{med sannolikheten } \frac{r}{b} \\ bq & \text{med sannolikheten } 1 - \frac{r}{b} \end{cases}$$

Exempel. Om man vill avrunda värdet 2 till en multipel av 5 ($b = 5$), är sannolikheten $2/5$ ($r = 2$) att det avrundade värdet blir 5, och 0 med sannolikheten $3/5$.

Totalerna bibehålls utan avrundning. I slumpmässig avrundning tas ingen hänsyn till att cellvärdena skall addera upp till marginaltotalen, vilket medför att summan av cellvärdena i allmänhet inte överensstämmer med tillhörande marginaltotal.

Exempel. Samma tabell som förut har avrundats slumpmässigt med basen 5.

Område	0 - 999	1000 - 1999	2000 - 2999	3000	Totalt
A	20	0	0	0	25
B	15	10	10	15	50
C	0	5	5	0	12
D	10	10	20	0	35
Totalt	44	28	31	19	122

Summeras värdena i 1:a kolumnen, $20 + 15 + 0 + 10 \neq 44$ som är marginaltotalen. Detta kan vara svårt för användarna att acceptera. Skulle man göra en ny slumpmässig avrundning får man troligen ett annorlunda utseende på grund av slumpmekanismen.

Några former av avrundning förekommer på SCB. Bland annat används deterministisk avrundning av riskceller (1 till 0, 2 till 3), utan konsekvensavrundningar eller ändring av marginaltotaler. I SuperCROSS finns olika regler för slumpmässig avrundning. Dessa metoder rekommenderas dock inte.

4.3.4 Kontrollerad avrundning

Kontrollerade avrundningar utförs som slumpmässiga avrundningar men med skillnaden att både cellvärden och marginaltotaler avrundas till en av de två närmaste multiplerna av basen. Till detta läggs bivillkoret att summorna av cellernas värden måste överensstämma med marginaltotalerna (både kolumn och rad). Kontrollerad avrundning fungerar garanterat endast för tvådimensionella tabeller. I högre dimensioner kan avvikelserna någon gång bli längre än basen för avrundningen.

Värdet på basen b för avrundning får inte vara alltför lågt, eftersom detta skulle ge en intervallskattning av det avrundade värdet som blir för kort. Å andra sidan måste detta önskemål balanseras mot kravet att informationsförlusten bör minimeras: ju bredare intervall desto större informationsförlust.

Exempel. Efter kontrollerad avrundning kan samma tabell som förut se ut som följer. Observera att vissa avrundningar har skett till en multipel av 5 som inte är den närmaste multiplern. Exempelvis har totalen 19 på sista raden avrundats till 15 trots att 20 är den närmaste multiplern av 5. Kontrollerad avrundning kräver mycket mer än enkel huvudräkning.

Område	0 - 999	1000 - 1999	2000 - 2999	3000	Totalt
				+	
A	20	5	0	0	25
B	15	10	10	15	50
C	0	5	5	0	10
D	10	10	15	0	35
Totalt	45	30	30	15	120

De slumpmässiga avrundningarna kan vara svåra att motivera för användarna. Den kontrollerade avrundningen känns mera tilltalande. Alla celler har ungefär sin riktiga storlek. Nackdelen är att den inte garanterat fungerar i tabeller med fler än två dimensioner.

4.4 Skyddsmetoder 2: förändringar i mikrodata

Ett alternativ till att använda skyddsmetoder på tabeller är att sätta in skyddsmetoder i steget före tabellframställningen, d.v.s. direkt på rådata. Efter det att mikrodata har åtgärdats med hjälp av någon skyddsmetod, skapas tabeller som då är kontrollerade.

4.4.1 Förgrovning (Recoding)

Numeriska variabler kan *avrundas*, t.ex. kan man ange inkomst i 1000-tals kronor i stället för på kronan när. Närliggande klasser kan slås samman till en klass.

Exempel. I hierarkiska variabler kan man gå från en finare indelning till en grövre. Exempelvis kan man redovisa på kommun i stället för församling, på SNI 2-siffernivå i stället för 3-siffernivå.

4.4.2 Top- eller Bottomkodning

Finns det få uppgiftslämnare i de högsta eller lägsta klasserna, slås klasser samman till en klass. Exempelvis kan höga inkomster ”topkodas” så att de högsta hamnar i en klass även om variationen är stor på dessa inkomster. På samma sätt kan låga inkomster ”bottomkodas” så att de lägsta hamnar i en klass även om de enskilda inkomsterna har stor variation.

4.4.3 Slumpmässiga störningar

Observationerna kan utsättas för en rad störningar. Man kan *ändra innehållsvariablerna* genom att addera eller multiplicera med slumpvariabler med vissa fördelningar. Man kan också byta helt korrekta värden mellan poster med likartat utseende (*data swapping*).

Slumpmässiga störningar av mikromaterial får inte göras vid SCB. Det skulle vara svårt att förklara för användare varför man frivilligt introducerade fel i sina data, särskilt som detta skulle kunna leda till att användarna drog fel slutsatser. Som exempel kan nämnas att slumpmässiga fel inte bara ökar variationen i de studerade variablerna utan också försvagar eventuella samband mellan dem. Introduktion av fel i grunddata kan t.o.m. vara olaglig, om det rör sig om personuppgifter och identifieraren bibehålls.

4.5 Bedömning av informationsförluster

Röjandekontroll innebär alltid informationsförluster. Många celler kan döljas i samband med undertryckningar, och detta förorsakar en stor informationsförlust. Å andra sidan blir intervallen ibland ganska korta, men detta syns först efter långa räkningar.

Avrundningar påverkar samtliga tabellceller.

Att jämföra informationsförlusten mellan olika röjandemetoder är inte lätt, vilket man ser redan av ovanstående konstruerade exempel. Med god programvara kan man få hjälp att inom en given metod minimera informationsförlusten.

5 Flexibla databasuttag

5.1 Problemställningen

Större risk för röjande finns om alla användare själva får *välja sina tabeller från en databas* av finfördelade makrodata eller t.o.m. mikrodata. Även om något program för röjandeskydd genomför åtgärder, exempelvis undertryckningar, så kan användaren göra sina egna sammanslagningar av klasser för att undgå undertryckningar eller för att få undertryckta värden som differenser mellan publicerade värden.

Vi betraktar fallet med en mikrodatabas. Vi ser på varje värde i databasen som en obekant, och varje uträknad tabellcell som högerledet i en linjär ekvation. Då finns det ett stort, men ändå begränsat, antal obekanta, och en användare kan i princip skapa hur många ekvationer som helst. Då kan det finnas en entydig lösning, vilket skulle innebära att hela mikrodatafilen kan avslöjas.

Exempel. Vi anknyter till det andra exemplet på undertryckningar i avsnitt 4.3.1. I detta fall räcker det att ta ut deltabeller.

Område	999 - 1999	2000 - 2999	Deltotal
B	12	8	20
C	4	5	9
Deltotal	16	13	29

Område	0 - 999	1000 - 1999	Deltotal
B	15	12	27
D	7	10	17
Deltotal	22	22	44

I dessa tabeller finns inga riskceller, och därför undertrycks ingenting. Detta innebär att sekundärundertryckningarna har avslöjats, och med dem också de primärundertryckta värdena. Denna typ av användning är dock olaglig, se avsnitt 3.1.1.

5.2 Skyddsmetoder

Flexibla databasuttag kan tekniskt skyddas på olika sätt. Givetvis bör man använda de skyddsåtgärder som beskrivs i avsnitt 4 för varje enskild tabell. Dessutom kan man notera vilka tabeller en viss användare har kommit åt, och ta med dessa i granskandet av alla senare uttag.

Detta kan dock bli mycket resurskrävande. Bearbetningstiden växer mycket fort med ökad tabellstorlek och fler dimensioner i tabellerna. Dessutom blir informationsförlusten mycket stor. Användaren kommer att få mindre och mindre data med tiden, och kommer att ångra sina första uttag. Ännu mindre information skulle komma fram, om man tog hänsyn till alla andra användares tidigare databasuttag.

Ett annat sätt är att skydda den underliggande databasen, så att denna i princip är säker, vilka uttag som än görs. Man kan exempelvis använda programvaran μ -ARGUS (se avsnitt 7.3.1) för detta ändamål.

5.3 Praxis och riskbedömning

SCB har några enstaka gånger tillåtit flexibla databasuttag. Ett exempel är data från folk- och bostadsräkningarna (FoB) på delområdesnivå i den gamla regionalstatistiska databasen. I detta fall undertrycktes riskceller i tabellerna, men ingen hänsyn togs till tidigare uttag. Uttagen blev dyra och informationsförlusterna blev stora. I stället valdes lösningen att varje kommun enbart kunde hämta uppgifter om delområden inom den egna kommunen, och dessa uppgifter omfattades av sekretess.

I några fall har aidentifierade data från registret över totalbefolkningen (RTB) med mycket magert innehåll lämnats ut som SuperCROSS-databaser. Inom hotellstatistiken tillåts sedan länge flexibla uttag, efter manuell kontroll, om varje uttag bygger på 5 eller fler enheter och om ingen enskild enhet och ingen enskild ägare dominerar.

Planer finns f.n. på att lägga ut mikrodata under SuperCROSS på Internet. Även röjanderiskabla makrodata på nivån delområden av kommuner kommer att läggas ut. Dessa databaser skall dock skyddas av lösenord, så att statistikconsumenterna, i första hand kommuner, endast kan se de data som de har rätt att se. Överföringen av tabelluttagen för dessa konsumenter kommer dessutom att ske i krypterad form. Någon närmare röjandekontroll behövs därför inte. Skulle det i en framtid bli aktuellt att vidga användningen av dessa data till andra användare, så skulle man kunna använda de hjälpmedel som utvecklats av Australian Bureau of Statistics och som finns tillgängliga med programvaran. Se avsnitt 7.4.

I övrigt finns f.n. inga andra planer på att lägga ut mikrodata med åtkomst över Internet. Innan något sådant sker, bör statistikproducenten rådgöra med röjandexpertter. Även rena datasäkerhetsskäl talar för att vara restriktiv med en sådan publicering.

6 Kartor

Att publicera statistik i form av kartor blir allt vanligare, beroende på utvecklingen av geografiska informationssystem (GIS). Att utföra statistisk röjandekontroll av kartor är oftast i princip samma uppgift som att granska tabeller. Statistik över ett önskat geografiskt område kan presenteras antingen som en tabell eller som en karta.

Kartor ger möjlighet till en mer flexibel indelning av områden jämfört med tabeller, där de geografiska områdena ofta är uppdelade i län, kommuner eller kommundelar. Går man ner på mycket låg geografisk nivå, ökar risken för röjanden. SCB skulle kunna producera statistik ända ner på fastighetsnivå. Detta kunde dock lätt ge upphov till röjanden. Eftersom en fastighet ofta ägs av en bestämd person, kan en fastighetsanknuten uppgift ibland hänföras till en enskild.

6.1 Karttyper

Statistiska kartor delas in i följande grupper: ytmönsterkartor, rutkartor, isaritmkartor, prickkartor och kartogram, se (*Wallgren m.fl. 1996*).

6.1.1 Ytmönsterkartor

Ytmönsterkartor beskriver andelar, intensiteter eller medelvärden för olika förbestämda geografiska områden. Man gör en klassindelning av den aktuella variabeln och de olika klasserna representeras med färg, raster eller mönster på kartan. Den geografiska indelningen bestäms av producenten eller beställaren. Vanliga geografiska uppdelningar är landskap, län eller kommuner.

Exempel. Ytmönsterkarta över SAMS-områden (Small Area Market Statistics) i Örebro med vägnätet pålagd.

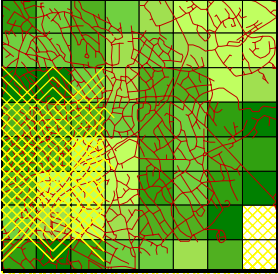


6.1.2 Rutkartor

Rutkartor presenterar statistik som ovan men delar in ett geografiskt område i rutor i stället för administrativa gränser. Rutoras storlek varierar från en

sidlängd på 200 meter och uppåt. Ju kortare sidlängd, desto högre blir riskerna för röjanden.

Exempel. Rutkarta över Huddinge, uppdelad i 1-kilometersrutor.



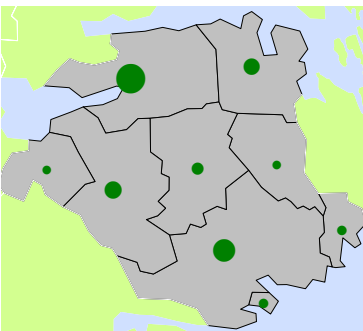
6.1.3 Isaritmkartor

Isaritmkartor har ingen förbestämd indelning av det geografiska området, utan indelningen bestäms av egenskaper hos den aktuella variabeln. Isaritmkartan består av nivåkurvor, där kurvorna styrs av de observerade värdena (mikrodata). Nivåerna bestäms efter en indelning av värdena i intervall. Exempel på isaritmkartor är kartor med variablerna snödjup och temperatur ett visst datum i Sverige.

6.1.4 Prickkartor

Prickkartor redovisar frekvenser per geografiskt område (jämför frekvenstabell). En *lägesprickkarta* eller objektsummerad prickkarta anger exakta läget på objekten, och *referensprickkartan* visar antalet i varje geografiskt område men inte det exakta läget.

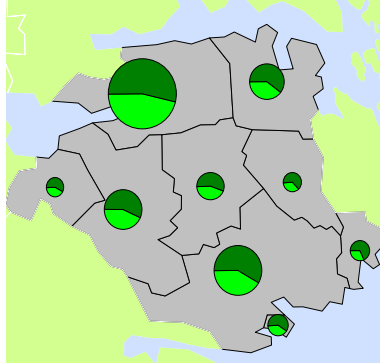
Exempel. Prickkarta över Södermanlands kommuner. Antal förvärvsarbetande mellan 20-64 år under 1998.



6.1.5 Kartogram

Kartogram är en blandning av ett traditionellt diagram och en karta. Kartan utgör underlaget på vilken diagram ritas för varje område.

Exempel. Kartogram över Södermanlands kommuner. Andel förvärvsarbetande män respektive kvinnor mellan 20-64 år under 1998.




6.2 Riskbedömning och skyddsmetoder

6.2.1 Ytmönsterkartor och rutkartor

Ytmönsterkartor och rutkartor kan ses som en form av publicering av en tabell med en summerbar innehållsvariabel (t.ex. inkomst, omsättning). Regler för identifiering av riskceller (riskområden) och metoder att dölja dessa är desamma som för tabeller med summerbar innehållsvariabel, se avsnitt 8.4 för rekommendationer. Förekommer riskområden på en rutkarta, kan man välja större rutor – för dessa områden eller generellt – (aggregering), avrundning eller undertryckning. Om det finns en beställare så styrs valet av metod ofta av dennes önskemål.

Exempel. En beställning över viss statistik önskas i form av rutor med sidlängd 200 meter. Denna geografiska indelning gör att vissa rutor har för få objekt, och därigenom skulle röjanden kunna ske. Då kan man antingen välja en karta med större rutor (aggregering) eller utelämna statistik i de rutor som innehåller för få objekt (undertryckning).

Exempel. Detta hypotetiska exempel illustrerar risken som är speciell för kartor jämfört med traditionella tabeller. En kund vill ha statistik över: födelseland Turkiet och inkomst från ett specifikt område.

	Inkomst (summa)	Antal
	1 513 786 kr	10

Därefter specificerar kunden ett annat område som är en delmängd av ovanstående:



Inkomst (summa)	Antal
1 200 000 kr	9

Från dessa två uppdrag kan kunden dra slutsatsen att det bor en individ som är född i Turkiet i det område som inte finns med i bild 2 och att den individen tjänar 313 786 kr.

6.2.2 Isaritmkartor

Om ett objekt som tillhör en extrem nivåkurva identifieras, kan man undvika detta röjande antingen genom att göra intervallen bredare eller med top/bottom kodning. Detta kan liknas vid åtgärder för mikrodata snarare än tabeller. Denna typ av kartor bygger på komplicerade algoritmer. SCB saknar dock tillräcklig erfarenhet för att i dag kunna ge rekommendationer.

Exempel. Man undersöker halten av kadmium i jord och bestämmer sig för att presentera resultatet som en isaritmkarta. Kadmium kan variera kraftigt såväl mellan stora områden som lokalt. En brant "topp" skulle kunna tyda på ett område med en industri som släpper ut höga halter av kadmium. En tänkbar åtgärd skulle kunna vara glesare isaritmer, i hela eller delar av kartan.

6.2.3 Prickkartor

Prickkartor kan ses som en publiceringsform av liknande slag som en frekvenstabell, och regler och skyddsmetoder är desamma som för frekvenstabeller. En annan tänkbar åtgärd skulle kunna vara en mindre (slumpmässig) störning av koordinaterna. Denna ansats är dock svår att genomföra i praktiken; ett antal hushåll som hamnar i Östersjön förstör trovärdigheten hos hela undersökningen.

6.2.4 Bedömning

Den fina geografiska indelningen kan alltså förorsaka röjandeproblem. Principiellt möter man emellertid inte några nya problem när det gäller att göra röjandekontroll inför publicering av kartor.

7 Metoder och programvaror i praktiskt bruk

7.1 Sammanfattning av utländska kontakter

7.1.1 Frågeställningar

För att få en uppfattning om vilka metoder och regler som används runt om i världen skickades ett antal frågor om statistisk röjandekontroll till centralbyråer och motsvarande i USA, Kanada, Australien, Nederländerna, Storbritannien och Finland. Frågorna var:

- Vilka metoder använder ni för närvarande för statistisk röjandekontroll av tabeller och av mikrodata?
- Tillåts användare definiera och skapa egna tabeller (flexibla tabelluttag)? I sådana fall, hur hanteras röjandeproblematiken?
- Hur används (hur väl tillämpas) metoderna? Har ni varierande strategier beroende på avdelning?
- Använder ni någon speciell programvara för röjandekontroll? Vilken?

Aktiviteten inom detta område varierar kraftigt mellan länderna. I USA har de federala myndigheterna ett råd som ansvarar för metoder som minimerar röjanderisker: The Subcommittee on disclosure limitation methodology. Detta råd gav 1994 ut en rapport där policy, praxis och tillvägagångssätt för röjandekontroll beskrivs för de 12 största federala statistiska myndigheterna. Flera länder, t.ex. USA, Kanada, Nederländerna och Australien, jobbar i likhet med SCB med standardisering av metoder för statistisk röjandekontroll och programvara. CBS (Nederländerna) har en handbok med regler och praxis för röjandekontroll som gäller för alla avdelningar på CBS.

Nedan följer en kortfattad sammanfattning av de erhållna svaren.

7.1.2 Frekvenstabeller

- *US Bureau of the Census.* Data swapping tillämpas på mikrodata från totalundersökningar av befolkningen (census) som skydd för röjande i frekvenstabeller. Metoder för röjandeskydd används på mikrodata i stället för på tabeller. Tabeller från stickprov anses vara skyddade av själva stickprovet. Användning av data swapping och störning tros vara framtida metoder för statistisk röjandekontroll.
- *Statistics Canada.* För individ/hushållsdata används avrundning för frekvenstabeller från totalundersökningar. Vissa resultat publiceras inte för geografiska områden som ligger under ett givet tröskelvärde. När det gäller tabeller, är trenden att man aggregerar i stället för att undertrycka.
- *ABS, Australien.* Metoder och regler varierar mellan avdelningar. Generellt gäller att celler med mindre än tre objekt inte får publiceras. Arbetet pågår med att standardisera den statistiska röjandekontrollen över avdelningsgränser.

7.1.3 Tabeller med summerbar innehållsvariabel

- *US Bureau of the Census*. Här utgör inte det faktum att materialet är ett stickprov ett fullgott skydd för företagsinformation (på grund av att vissa företag alltid kommer med i stickproven på grund av sin storlek). Rekommenderad metod är p %-regeln. Denna rekommenderas framför (n, k) -regeln på grund av att den ger ett konsistent skydd åt alla uppgiftslämnare vad gäller hur mycket information som kan härledas. Om uppgiftslämnaren behöver ytterligare skydd mot konkurrenter som lämnat bidrag till samma cell, kan detta erhållas genom en kombination av olika (n, k) -regler. Undertryckning eller sammanslagning av klasser används på riskceller.
- *Statistics Canada*. En kombination av (n, k) -regler används. Celler som identifieras som riskceller undertrycks. Parametervärdena är konfidentiella. Vissa problem med serien av (n, k) -regler förekommer. p %-regeln eller pq -regeln anses vara bättre. En utredning pågår med mål att införa enhetliga regler för företagsundersökningar vad gäller statistisk röjandekontroll. Dagens metoder ses över med syfte att ge en mängd rekommendationer.
- *ABS Australien*. Inget objekt får bidra med mer än 85 % till totala cellvärdet och två objekt får inte bidra tillsammans med mer än 90 % av totala cellvärdet.

7.1.4 Mikrodata

Flera av länderna har ett råd (motsvarande DASK på SCB) som bedömer om ett utlämnande kan ske eller inte. På US Bureau of the Census består rådet av en representant från vardera ekonomiska, demografiska och folkräkningsprogrammet och tre representanter från forsknings- och policyenheterna. Många länder skriver avtal med användare om hur data får användas. I vissa länder (bl.a. Kanada, Nederländerna, USA) finns forskningscentra i anslutning till datautlämnaren där forskare har tillgång till mikrodata på plats. Mikrodata från företag lämnas ut i betydligt mindre omfattning jämfört med individ/hushållsuppgifter på grund av den betydligt större risken för eventuella röjanden.

En del länder ger ut s.k. Public Use Micro Files (PUMF). Dessa kräver en noggrann statistisk röjandekontroll innan de kan offentliggöras.

7.1.5 Flexibla tabelluttag

I USA har US Bureau of the Census utvecklat American FactFinder för att göra data lättillgängliga för allmänheten och ge allmänheten möjlighet att skapa egna tabeller. Målet är att tillåta användare att skicka förfrågningar om önskade tabeller elektroniskt. En förfrågan passerar en brandvägg till Censusbyråns interna server, där en mikrodatafil som genomgått en statistisk röjandekontroll finns lagrad. Den efterfrågade tabellen skapas och genomgår

elektronisk granskning för eventuell röjanderisk. Om den bedöms vara säker, sänds den tillbaka elektroniskt.

Generellt på myndigheter i USA finns administrativa databaser och surveydatabaser. De administrativa är endast tillgängliga för internt bruk, medan surveydatabaserna kan bli tillgängliga för allmänheten. Den allmänt tillgängliga surveydatabasen är ofta en förkortad version som genomgått noggrann statistisk röjandekontroll. För en del databaser som innehåller känslig information används DADS (Data Access and Dissemination System) för att hantera förfrågningar om tabeller för att undvika utlämnande av mikrodata.

På ABS i Australien finns ett "data warehouse" som tillhandahåller önskade tabeller. För närvarande är det endast till för internt bruk men planer finns på att även externa kunder skall kunna skapa egna tabeller. Mikrodata som finns där är aggregerade och har genomgått statistisk röjandekontroll.

7.2 Översikt av tillgänglig mjukvara

En genomgång och jämförelse av tillgänglig programvara presenterades på Joint ECE/Eurostat Work Session on Statistical Data Confidentiality i Thessaloniki (Giessing 1999). Presentationen begränsades till programvaror för konsekvensundertryckningar i tabeller. Följande programvaror behandlades:

<i>Programvarans namn</i>	<i>Utvecklad av</i>	<i>Plattformar/Programspråk/Tillgänglighet/Kostnad</i>
GHQUAR	Landesamt für Datenverarbeitung und Statistik Nordrhein-Westfalen (Dietz Repsilber)	FORTRAN-kod för IBM and Siemens stordatorer, icke-kommersiell
USBCSUP	US Bureau of the Census (Bob Jewett)	FORTRAN-kod för DEC datorer, icke-kommersiell
CONFID	Statistics Canada (Gordon Sande/Dale Robertson)	FORTRAN- (RATFOR)-kod för IBM stordatorer, SUN SPARC arbetsstationer, icke-kommersiell
ACSSuprs	Sande and Associates, Inc. (Gordon Sande)	Förbättrad version av CONFID, kommersiell
τ -ARGUS Version 1.5	CBS-Netherlands (second prototype version)	C++ kod, WINDOWS-mjukvara (32BIT), icke-kommersiell, gratis tillgänglig från Statistics Netherlands. Ytterligare kommersiell programvara för c:a 10000 kr behövs.

Programvarorna har bedömts med avseende på beräkningsprestanda, säkerhet, informationsförluster och applicerbarhet. Skillnaden i prestanda var mycket stor, i varje fall för stora tabeller. Vissa programvaror klarade inte hierarkiska variabler. Men för SCB:are räcker det att läsa den högra kolumnen i tabellen. Det finns så liten erfarenhet om andra datormiljöer inom verket, att ARGUS är det enda tänkbara alternativet för SCB.

Givetvis finns också program på SCB som löser delproblem för speciella applikationer.

7.3 ARGUS

ARGUS är en programvara för röjandekontroll som är utvecklad inom ett EU-projekt, främst av CBS, Nederländerna. ARGUS är uppdelad i en del för kontroll av mikrodata, μ -ARGUS, och en del för kontroll av tabeller, τ -ARGUS. μ -ARGUS fungerar som ett hjälpmedel som effektiviserar, standardiserar och förenklar kontrollen av mikrodata före ett eventuellt utlämnande eller om man väljer att skydda tabeller genom kontroll av mikrodata i stället för den enskilda tabellen. τ -ARGUS utför röjandekontroll av enskilda tabeller upp till fyra dimensioner. I dagens version finns ingen möjlighet att kontrollera sammanhörande tabeller (linked tables) eller tabeller med hierarkiska spridningsvariabler.

Förändringar och åtgärder i mikrodata och tabeller dokumenteras i en loggfil som produceras samtidigt med den "säkra" mikrofilen eller tabellen.

Eurostat har startat ett projekt för utveckling av τ -ARGUS med avseende på metoder, prestanda och tillförlitlighet. Projektet kommer att ledas från CBS.

7.3.1 μ -ARGUS

Första steget i kontroll av mikrodata är att upptäcka känsliga kombinationer av variabler. μ -ARGUS försöker fånga upp kombinationer av poster (värden) som är ovanliga i det aktuella materialet i stället för att leta efter unika förekomster i hela populationen. Den som skall utföra röjandekontrollen måste först bestämma gränser för vad som är ovanliga värden för *nyckelvariabler* (*nyckel* = kombination av identifierande variabler) för att identifiera *riskposter*. Kombinationer som inträffar sällan och anses riskabla måste åtgärdas innan ett utlämnande sker.

För att bedöma röjanderisken används en metod som baseras på tröskelvärden. En kombination av värden är säker endast om den uppskattade frekvensen i populationen överstiger det angivna tröskelvärdet. Vilka nycklar som skall kontrolleras bestäms av användaren, antingen manuellt eller med hjälp av regler som finns i μ -ARGUS.

Metoder för statistisk röjandekontroll i μ -ARGUS är:

- Sammanslagning av klasser
- Top/Bottom coding
- Mikroaggregering
- Lokal undertryckning
- Störning

När mikrodata har åtgärdats, måste man även mäta informationsförlusten. Används undertryckning för att säkra celler, är måttet antalet celler som undertrycks. Vid aggregering däremot baseras måttet på en värdering av värdet på den identifierande variabeln och en värdering av varje möjlig fördefinierad kodning för varje identifierande variabel. För att minimera informationsförlusten vid aggregering och undertryckning rekommenderas först aggregering till dess att antalet riskceller nått en acceptabel nivå och därefter undertryckning av kvarvarande riskceller.

7.3.2 τ -ARGUS

I τ -ARGUS används (n, k) -regeln eller tröskelvärdesregeln för identifiering av riskceller. När riskcellerna identifierats, kan någon av nedanstående metoder användas för att skydda tabellen:

- Sammanslagning av spridningsvariablernas klasser
- Undertryckning
- Kontrollerad avrundning

Vid en körning av τ -ARGUS visas först vilka celler som är riskceller och användaren väljer därefter vilken metod som skall användas för att skydda dessa. Därefter presenteras de celler som fortfarande är riskceller.

7.3.3 Erfarenheter av ARGUS

Erfarenheterna av praktiska material gäller huvudsakligen τ -ARGUS. Beräkningarna har varit korrekta. Undertryckningar och avrundningar har skett efter förväntningar. Redskapen för förgrovningar har fungerat. Vad gäller små tabeller har det inte funnits prestandaproblem.

Däremot tar det viss tid att för hand skapa de metadata som behövs. Mikrodata måste konverteras från databaser eller SAS-filer till flata ASCII-filer. Utdata kan bli en Excel-tabell, som kräver viss anpassning före utskrift. Ett försök med undertryckningar i en mycket stor tredimensionell tabell avbröts efter fem minuter.

Vi bedömer att nuvarande version av τ -ARGUS inte är tillräcklig för att implementeras i SCB:s produktionssystem. Vi bevakar därför resultaten av Eurostat-projektet.

7.4 SuperCROSS

SuperCROSS är ett tabuleringsverktyg med lättbegripligt användargränssnitt, kompakta databaser och mycket snabba och flexibla uttag. Vissa funktioner för röjandekontroll har byggts in. Bland annat finns slumpmässiga avrundningar av små celler samt vissa tröskelvärdesregler och dominansregler. Konsekvensundertryckningar sker inte. Funktionen att

läsa enstaka poster från mikrodatafilen kan stängas av. Programfamiljen SuperSTAR är ändå så attraktiv för SCB, att verket kan tvingas acceptera även mindre goda metoder för röjandeskydd i de tabeller som produceras.

7.5 Läget på SCB

För att få en uppfattning om hur man gör på SCB i dag när det gäller röjandekontroll skickades en enkät ut till c:a 200 produktansvariga. Av dessa 200 är det en del som inte berörs och de har inte heller lämnat några svar. Vissa andra statistikansvariga myndigheter har kontaktats med liknande frågor.

Slutsatser man kan dra från enkätsvaren är att det finns en stor variation vad gäller metodik och regler för röjandekontroll. Viss variation är naturligtvis berättigad, men i vissa fall sammanfaller den med att kunskapsnivån hos de produktansvariga varierar kraftigt.

Metoder och regler för kontroll av tabeller har till övervägande del använts längre än 10 år. Motiveringar till varför just de tillämpade metoderna och reglerna används varierar från väl genomtänkta motiv till att de ger minsta möjliga arbete. Metodik och regler för mikrodata har inte använts lika länge, vilket kan vara naturligt då det blivit mer aktuellt med utlämnande av mikrodata de senaste 10 åren. Litet överraskande anser majoriteten av produktansvariga att det inte behövs klarare regler, bättre metoder eller programvara för röjandekontroll. Svaren på de andra frågorna indikerar trots allt att ett sådant behov finns.

8 Riktlinjer

8.1 Giltighet

Reglerna nedan gäller för all SCB:s statistikproduktion, för leverans såväl inom som utom landet, både för anslag och uppdrag, oberoende av finansieringsform och oberoende om produkten ingår i den officiella statistiken eller ej.

Kortfattad information om verkets röjandepolicy skall lämnas på SCB:s webbplats.

8.2 Känslighet och röjandekontroll

En gradering av känslighet skall i princip inte ske. Röjandeskydd skall alltid tillämpas, så snart en sekretessbelagd variabel finns med, även om variabeln inte upplevs som särskilt känslig.

De rekommenderade parametervärdena skall ses som en miniminivå. Om röjanderisken upplevs som särskilt stor, kan strängare värden tillämpas enligt bedömning av avdelningschef.

8.3 Organisatoriska åtgärder

Inom SCB har varje avdelningschef ansvaret för att riktiga metoder används på avdelningen. Programcheferna skall se till att reglerna tillämpas och implementeras i löpande drift.

För att hålla kunskapen om röjandefrågor vid liv föreslås att verket skapar en kontaktgrupp för röjandefrågor. Gruppen skall fungera som ett nätverk för att utbyta erfarenheter och förmedla nyheter inom detta område. Förslagsvis skall gruppen träffas minst två gånger per år och ledas av någon från U-avdelningen.

8.4 Tabeller

Rekommendationerna nedan gäller alla objekttyper, t.ex. företag, individer eller fastigheter. När det gäller företag skall man i allmänhet behandla företag inom en koncern som en enhet.

Rekommendationerna gäller tabeller från både total- och stickprovsundersökningar.

Röjandekontrollen av tabeller i officiell statistik, uppdragstabeller och uppgifter lämnade till Eurostat skall såvitt möjligt samordnas.

Regler och metoder skiljer sig åt beroende på om det är en frekvenstabell eller en tabell med summerbar innehållsvariabel som skall röjandekontrolleras.

I tabeller över stickprov kan det finnas undertryckningar som motiveras av tveksam statistisk kvalitet. Detta är inte en röjandefråga och behandlas inte här.

Undertryckta celler skall markeras med ". En förklaring till denna beteckning skall finnas i alla publikationer.

SCB kan f.n. inte rekommendera hur informationsförlusten skall mätas.

8.4.1 Frekvenstabeller

Regel. Tröskelvärdesregeln skall användas för att identifiera riskceller. En cell måste innehålla minst 3 objekt för att kunna publiceras, d.v.s. tröskelvärdesregeln tillämpas med parameter 3.

Metoder. Aggregering, undertryckning eller kontrollerad avrundning skall användas för att minska röjanderisken. Slumpmässiga störningar av mikro-materialet får inte användas som skydd av tabeller.

8.4.2 Tabeller med summerbar innehållsvariabel

Regel. En kombination av dominansregler skall identifiera riskceller. Inget objekt får bidra med 50 % eller mer till cellvärdet ($(1, 0.5)$ -regel). Inte heller får två objekt tillsammans bidra med 90 % eller mer av det totala cellvärdet ($(2, 0.9)$ -regel). Om minst en av reglerna är uppfylld måste åtgärder vidtas.

Metoder. Aggregering, undertryckning eller kontrollerad avrundning skall användas för att minska röjanderisken. Slumpmässiga störningar av mikro-materialet får inte användas som skydd av tabeller.

8.4.3 Rekommenderad arbetsgång vid röjandeminimering av tabeller

- Identifiera riskcellerna.
- Om risken anses stor (många riskceller), bör någon åtgärd utföras på mikrodata och därefter måste en ny identifiering av riskceller ske. Om risken är mindre, men många riskceller återstår, bör sammanslagning av klasser (eller aggregering) användas. Återigen görs en identifiering av riskceller, och om antalet riskceller bedöms vara litet, undertrycks resterande celler eller tillämpas kontrollerad avrundning på hela tabellen.
- Om antalet riskceller är relativt litet, kan man åtgärda tabellen direkt med sammanslagning av klasser (eller aggregering). Ny identifiering av riskceller görs. Om endast få finns kvar, tillämpas undertryckning eller kontrollerad avrundning.
- Om antalet riskceller är litet kan man undertrycka eller avrunda direkt.

8.4.4 Verktyg

Det enda verktyg som är tänkbart att använda för hela kedjan är τ -ARGUS. Innan denna programvara kan rekommenderas, måste viss anpassning av den göras bland annat med hänsyn till SCB:s produktionsmiljö.

8.5 Flexibla databasuttag

Innan makrodata görs tillgängliga för flexibla databasuttag utan manuell kontroll från SCB, skall de underliggande makrodata såvitt möjligt granskas enligt metoderna i avsnitt 4.

Publicering utan restriktioner på Internet av mikrodatabaser med flexibla uttagsmöjligheter planeras inte för närvarande. Om och när detta skall ske, måste röjandefrågor beaktas. Även datasäkerhetsfrågor i anslutning bör beaktas.

Utlämnande av mikrodata skyddade av speciella programvaror förekommer däremot på SCB. Hit hör bl.a. uttag i Fasit-modellen och utlämnande av data i SuperCROSS-format. Sådana utlämnanden skall behandlas på samma sätt som utlämnande av mikrodata (beslut av avdelningschef, eventuellt efter föredragning i DASK, överklagandemöjligheter m.m.).

8.6 Kartor

Regler och skyddsmetoder för ytmönsterkartor och rutkartor är desamma som för tabeller med summerbar innehållsvariabel, se avsnitt 8.4.2.

För isaritmkartor saknar SCB i dag den erfarenhet som skulle behövas för att kunna ge riktlinjer. Åtgärder för mikrodata torde dock ligga närmast tillhands.

För prickkartor är regler och skyddsmetoder desamma som för frekvenstabeller, se avsnitt 8.4.1.

Referenser

Block, H. och Olsson, L. 1976, Bakvägsidentifiering, Statistisk Tidskrift 1976:2, Stockholm.

Block, H. 2000, Reidentification of perturbed microdata files, Indianapolis 2000.

Eurostat, 1996 Manual on Disclosure Control Methods.

Eurostat, 1999. Statistical Data Confidentiality. Proceedings of the Joint Eurostat/UN-ECE Work session on Statistical Data Confidentiality, Thessaloniki 1999.

Flygare, A-M. 2000, Statistical Disclosure Control at Statistics Sweden: Present Situation and Needs for the Future, kommer att finnas i Proceedings, JSM 2000 Indianapolis.

Giessing, S. 1999 A Survey on Packages for Automated Secondary Cell Suppression, Eurostat, Joint ECE/Eurostat Work Session on Statistical Data Confidentiality 1999, Working Paper No 12.

Giessing, S. 2001 New Tools for Cell Suppression in Tau-ARGUS: One Piece of the CASC project work draft, Eurostat, Joint ECE/Eurostat Work Session on Statistical Data Confidentiality, Working Paper No. 2

National Research Council, Social Science Council 1993. Private Lives and Public Policies, Confidentiality and Accessibility of Government Statistics. National Academy Press.

Statistics Sweden 1993. Journal of Official Statistics, Confidentiality and Data Access, Vol. 9, No. 2.

Statistics Sweden 1998. Journal of Official Statistics, Special issue on Disclosure Limitation Methods for Protecting the Confidentiality of Statistical Data, Vol. 14, No. 4.

Subcommittee on Disclosure Limitation Methodology 1994. Report on Statistical Disclosure Limitation Methodology, Statistical Policy Working Paper 22, Statistical Policy Office, Office of Information and Regulatory Affairs, Office of Management and Budget, USA.

Wallgren, A., Wallgren, B., Persson, R., Jorner, U. och Haaland, J-A. 1996. Statistikens bilder – att skapa diagram. Publica.

Willenborg, L. and de Waal, T. 1996, Statistical Disclosure Control in Practice, Lecture Notes in Statistics, Springer.

Willenborg, L. and de Waal, T. 2001, Elements of Statistical Disclosure Control, Lecture Notes in Statistics 105, Springer.