



2024-01-16

Principer för statistikproduktion och minimering av röjanderisk vid Registercentrum Västra Götaland

Det här dokumentet innehåller principer för statistikproduktion och arbete med att förebygga och vid behov åtgärda röjanderisk vid publik visning eller utlämning av statistiska uppgifter vid Registercentrum Västra Götaland.

Riktlinjerna gäller inte:

- Åtkomst till data och statistik för behöriga användare hos den rapporterande vårdverksamheten.
- Utlämning av data där tillstånd för utlämnande av data på individnivå givits.

Statistikproduktion

De fenomen som studeras med hjälp av statistik påverkas av slumpmässiga variationer. Resultat som uppstår på grund av slumpen kan leda användaren att dra vilseledande slutsatser. Risken för att detta sker är påtaglig i statistik som är baserad på någon eller några enstaka observationer.

Statistik bygger på principen om *de stora talens lag* vilket betyder att slumpens påverkan på resultaten minskar i takt med att antalet observationer ökar. Det är därför viktigt att vid statistikproduktion sträva efter att utforma undergrupper så att de blir tillräckligt stora. Det är först när statistik håller hög kvalitet som vi kan lita på de mönster som framträder och dra korrekta slutsatser utifrån den. Det krävs som regel flera tiotals observationer för att det ska vara möjligt att uppnå statistik av hög kvalitet.

Vi vill att risken för att slumpen leder användaren till vilseledande slutsatser ska vara låg, och det är utifrån detta resonemang som Registercentrum Västra Götalands grundprincip för statistikproduktion är:

Den statistik som produceras på Registercentrum Västra Götaland ska baseras på ett så stort antal observationer att slumpens möjliga påverkan på de mönster som studeras kan bedömas vara liten.

Utöver slumpen finns det andra potentiella felkällor som kan påverka den statistiska kvaliteten. Dessa lämnas dock utanför det aktuella dokumentet.

Röjande

Vi har en skyldighet enligt lag att upprätthålla sekretess kring information om enskilda individer. Statistik från kvalitetsregister får inte utformas på ett sådant sätt att det finns risk för röjande av en enskild individs identitet.

Den grundläggande princip som vårt arbete med att förhindra röjande utgår från är:

Statistik som produceras på Registercentrum Västra Götaland ska inte kunna användas till att röja några uppgifter om enskilda individer

Vi utgår från följande definition av röjande:

Att röjande kan ske innebär att en individ med hjälp av statistik kan få veta någonting om en annan individ som den inte redan visste.¹

I vår roll som statistikproducent tar vi detta ansvar, oberoende av vår relation i övrigt till statistiken eller dess källa.

Förebyggande av röjanderisk

Risk för röjande uppstår först när en redovisningsgrupp utgörs av någon eller några enskilda individer. Principen om statistikproduktion handlar om att minska slumpens möjliga påverkan på de mönster som studeras vilket uppnås genom att basera den statistik som produceras på tillräckligt många observationer. Det betyder att principen för statistikproduktion och att förebygga röjanderisk går hand i hand.

Det enklaste och mest kraftfulla sättet att förebygga röjanderisk är således att tillämpa principen för statistikproduktion och undvika att redovisa statistik på detaljersnivåer där antalet observationer (individer) är otillräckligt.

Nedan återfinns några viktiga aspekter att ta med i arbetet med att förebygga röjanderisk:

Fasta och icke-överlappande redovisningsgrupper

Statistik ska endast redovisas för grupper som är *fasta* och *icke-överlappande*. Detta är nödvändigt för att undvika möjligheten att genom jämförelse mellan överlappande grupper identifiera mer detaljerad information.

Ännu inte avslutade tidsperioder

Statistik ska inte visas för *ännu inte avslutade tidsperioder* och statistik för en avslutad tidsperiod ska visas först när så lång period har gått sedan dess avslutande,

¹ Se bilaga 1 för en beskrivning av olika typer av röjande och hur röjande kan uppstå i praktiken.

att tidigare erfarenhet säger oss att näst intill alla registreringar för perioden ska ha hunnit komma in.

Min- och maxvärden

Vid statistikproduktion behöver hänsyn tas till att *min-* och *maxvärden* är individuella och potentiellt unika värden. Boxplots bör i normalfallet sträcka sig mellan percentilerna 5 och 95 eller 10 och 90, snarare än 0 och 100 (det ger också oftast bättre statistisk information). Individuella outliers visas inte.

Ojämna fördelningar

I vissa situationer kan det trots förebyggande insatser uppstå risk för röjande. Detta uppstår primärt när vi har att göra med variabler med tydligt *ojämna fördelningar*, som är svåra att jämma ut (till exempel genom sammanslagning av kategorier). Om vi, exempelvis, vill visa statistik separat för fyra grupper som fått olika behandlingar, och en av de fyra behandlingarna är mycket mer sällsynt än de övriga tre så uppstår röjanderisk i den fjärde gruppen. Denna typ av risk behöver då *åtgärdas*.

Åtgärder vid röjanderisk

För att åtgärda röjanderisker ska vi arbeta enligt följande princip:

Statistik på redovisningsnivåer² för vilka det inte uppstår några redovisningsgrupper där antalet individer understiger fem kan visas utan åtgärd.

För all annan statistik ska följande göras för att förhindra röjande:

1. Alla antalsuppgifter avrundas till närmaste tiotal.
2. Statistik i form av medelvärden, medianer och liknande visas bara för grupper som uppgår till minst 15 individer (det vill säga där det visade avrundade antalet är minst 20).
3. Andelar visas som vad de faktiskt är, men alltid avrundade till hela procent.
4. Andelar visas bara när den minsta av de täljare som summerar till helheten uppgår till minst fem individer (för att till exempel någon statistik alls ska visas över andelar som har fått olika behandlingar, ska ingen behandling ha fått av färre än fem individer).
5. Vid beräkning av ett konfidensintervall för en andel beräknas och avrundas intervallets mittpunkt och halva bredd separat till hela procent, och sätts därefter ihop till ett intervall.

Om det för en redovisningsnivå är svårt att avgöra om det uppstår (eller kan uppstå) redovisningsgrupper där antalet individer understiger fem, agerar vi som om de uppstår.

Tillämpning av de fem punkterna ovan omöjliggör inte röjande men de ser till att risken för det är mycket liten.

² En ”redovisningsnivå” definieras av variabeln eller kombinationen av variabler som statistik visas separat för, till exempel **enhet**, eller **enhet*år**, eller **enhet*år*ålder**. För att statistik ska kunna visas utan åtgärd på redovisningsnivån **enhet*år *ålder**, får det inte finnas någon kombination av enhet, år och ålder som innehåller färre än fem individer.

Avslutande kommentar

Arbetet med att förhindra röjande är komplext. Detta framgår inte minst i *Handbok för statistisk röjandekontroll* som publicerades av Rådet för officiell statistik år 2015.

En fördel med principerna för statistikproduktion och minimering av röjanderisk som fastställs i detta dokument är att de är generella och kan tillämpas på all statistik som produceras vid Registercentrum Västra Götaland. Arbetsgången vid produktion av publik statistik med hänsyn till röjanderisk sammanfattas i följande punkter:

1. Statistisk kvalitet och förebyggande av röjanderisk.
 - a) Välj redovisningsnivåer på ett sådant sätt att antalet observationer är tillräckligt stort i alla redovisningsgrupper.
Flera tiotals observationer krävs för att slumpens möjliga påverkan ska kunna bedömas vara liten.
 - b) Anpassa redovisningen utifrån viktiga aspekter i arbetet med att förebygga röjanderisk.
Fasta och icke-överlappande redovisningsgrupper, ännu inte avslutade tidsperioder, min- och maxvärden samt ojämna fördelningar.
2. Bedöm om åtgärder krävs för att förhindra röjande.
Minst fem individer i alla redovisningsgrupper.
3. Vidta vid behov åtgärder för att förhindra röjande.
Tillämpbara åtgärder i fempunktslistan.

Det kan mycket väl vara så att det uppstår situationer där någon av principerna i detta dokument kan uppfattas som onödigt strikt eller överflödigt. I sådana lägen är det viktigt att komma ihåg att principerna tillsammans bildar en helhet med det övergripande syftet att producera tillförlitlig statistik av hög kvalitet med minimal risk för röjande av enskild individ.

Bilaga 1. Fördjupning - Røjande

Statistik som produceras vid Registercentrum røjer aldrig personuppgifter direkt, i form av att namn eller personnummer publiceras. Det sätt som røjande kan ske på är mer indirekt. Vi utgår från följande definition av røjande:

Att røjande kan ske innebär att en individ med hjälp av statistiken kan få veta någonting om en annan individ som den inte redan visste.

För att sådant røjande ska kunna ske, krävs att den förstnämnda individen (hädanefter: *Individ A*) redan har *någon* information om den andra individen (hädanefter: *Individ B*) och dennas vårdprocess. Om Individ A med hjälp av denna information kan identifiera Individ B i statistiken och erhålla ytterligare information om Individ B så sker ett røjande, se ett förenklat exempel nedan:

1. Gustav (Individ A) bor granne med Kalle (Individ B) i Lillstad. Gustav har hört att Kalle drabbats av lungcancer. Kalle pratar inte mycket om sin sjukdom och vill inte gärna att någon ska veta att cancer är spridd.



2. Kalle går på regelbundna uppföljningar inom vården.



3. Ett kvalitetsregister publicerar statistik om lungcancer på sin hemsida.



4. Gustav hittar statistik om lungcancer på registrets hemsida och noterar att det bara finns en person med lungcancer i Lillstad. I statistiken framkommer även att andelen med spridd cancer i Lillstad uppgår till 100 % och Gustav förstår därmed att Kalles cancer är spridd.



5. Senare samma dag träffar Gustav Johan som tycker att det är tråkigt att Kalles cancer är spridd. Johan kontaktar Kalle för att beklaga situationen.



6. Hur gick det här till? Kalle tycker inte alls att det känns bra. Tanken var ju att familjen skulle bli först att få veta.



Formulerat som i föregående stycka är røjandet *säkert*, för där hänvisas till information som ges avseende *exakt en* individ. Men vi vill också undvika att visa statistik som visserligen inte kan ge *säker* information om en individ, men ändå *tydliga ledtrådar*. Om till exempel informationen som Individ A har om Individ B stämmer på inte bara en utan två individer i statistiken, och ytterligare information kan fås om dessa två, har Individ A fått en tydlig ledtråd om vilken information som kan gälla för Individ B.

Røjande kan också uppstå genom vad som ofta kallas för *grupprøjande*. Grupprøjande består i att information om en individ røjts genom att statistiken visar att för *alla eller ingen* som en viss mängd information stämmer på, gäller också någon ytterligare information (till exempel då Individ A vet att Individ B vårdades på enhet X under en viss tidsperiod, och statistiken visar att ingen som gjorde det blev förbättrad av sin vård).