

# Cell Key Method

En röjandekontrollmetod för registerbaserad individstatistik

2025-02-28



# Varför måste statistiken röjandekontrolleras?

Vid SCB skyddas uppgifter av sekretess enligt 24 kap. 8 § , offentlighets- och sekretesslagen.

Sekretessen innebär att den publicerade statistiken och uppdragsleveranser måste skyddas, dvs. röjandekontrolleras, så att det inte går att identifiera eller hitta information om enskilda objekt, till exempel personer eller företag.

I dagens samhälle där den tillgängliga information om personer och företag ökar har röjandekontroll blivit ännu viktigare än tidigare.

Det är angeläget för SCB att upprätthålla uppgiftslämnarnas och statistikanvändarnas förtroende för statistiken, bland annat genom en tillförlitlig metod för röjandekontroll.

Riktlinjer för europeisk statistik kräver att absoluta garantier ges för uppgiftslämnarnas integritet och att data skyddas. Uppgifterna ska behandlas konfidentiellt och enbart användas för statistikändamål.



# Skyddsmetoder för statistikvärden

Statistikvärden redovisas ofta fördelade på redovisningsgrupper (celler) i en tabell. Skyddsmetoderna avser att skydda statistikvärden för vilka det finns risk att ett objekt kan identifieras, eller att information för ett enskilt objekt kan avslöjas. Skyddet ska vara tillräckligt och med så lite informationsförlust som möjligt.

Exempel på metoder som inte ändrar statistikvärden

- Färre redovisningsgrupper som ger mindre detaljerad statistik
- Dölja statistikvärden

Exempel på metoder som ändrar statistikvärden

- Olika varianter av avrundning
- Andra metoder som tillför slumpmässig osäkerhet (brus)

# Vad är Cell Key Method?

Cell Key Method (CKM) är en metod som tillför brus i tabeller som visar antal objekt. Metoden är utvecklad av statistikbyrån i Australien och rekommenderades av Eurostat för rökandekontrollen i den europeiska folkräkningen 2021.

Metoden innebär att statistikvärden tillförs slumpmässig osäkerhet men med restriktioner som bland annat kontrollerar brusets storlek och ser till att det inte blir någon systematisk skevhet i statistiken.

För att minska rökanderisken har SCB valt att inte kommunicera exakt vilka restriktioner som läggs på bruset, till exempel hur stort bruset kan bli. Samma restriktioner gäller för alla SCB:s tabeller där CKM används.

# Egenskaper

- Statistikvärden som är större än noll justeras med ett litet negativt eller positivt heltal, eller lämnas oförändrade
- Statistikvärdet noll ändras inte, men statistikvärden större än noll kan ändras till noll
- De skyddade statistikvärdena kan inte vara negativa
- Det blir alltid samma brus för statistikvärden där samma objekt ingår, till exempel om samma statistikvärde tas fram vid olika tillfällen
- Information kan inte röjas genom att beräkna skillnaden mellan statistikvärden för överlappande redovisningsgrupper, till exempel redovisningsgrupperna 15-74 år och 16-74 år

# Tänk på att

- Varje statistikvärde i en tabell tillförs brus separat och oberoende av bruset för andra statistikvärden i tabellen. De redovisade totalerna är därför inte alltid lika med summan av deras redovisade delar (se exempel 1 och 2).
- Om redovisade statistikvärden adderas för att skapa en ny total, så adderas även osäkerheten. Det finns därför risk för att den nya totalen innehåller större osäkerhet än vad som skulle ha tillförts om brus applicerats direkt på den nya totalen (se exempel 2).
- För att minska rójanderisken behöver alla ingående statistikvärden brusas när andelar beräknas. Andelar baserade på ett litet antal objekt bör alltid tolkas med försiktighet.

# Exempel 1 med maximalt brus 2

## Oskyddade statistikvärden och brus

Kön	Yrkesgrupp		Totalt
	A	B	
Kvinnor	199 -1	1 1	200 0
Män	796 1	4 0	800 -1
Totalt	995 -2	5 1	1000 0

## Publicerade statistikvärden

Kön	Yrkesgrupp		Totalt
	A	B	
Kvinnor	198	2	200
Män	797	4	799
Totalt	993	6	1000

Exemplet visar antal kvinnor respektive män med maximalt brus 2, dvs. bruset kan anta värdena -2, -1, 0, 1 eller 2. Alla statistikvärden, inklusive totaler, får brus.

I det här exemplet får bruset obetydlig påverkan för män och kvinnor i yrkesgrupp A men även dessa statistikvärden behöver brusas för att skydda den enda kvinnan i grupp B.

Exemplet illustrerar att de publicerade totalerna inte alltid är lika med summan av deras redovisade delar.

## Exempel 2 med hierarkisk variabel och max. brus 2

Befolkning med en viss egenskap		
	Oskyddade statistikvärden och brus	Publicerade statistikvärden
<b>Riket</b>	<b>228 - 2</b>	<b>226</b>
<b>Region Norr</b>	<b>103 + 1</b>	<b>104</b>
Kommun A	35 - 1	34
Kommun B	25 - 1	24
Kommun C	43 + 0	43
<b>Region Syd</b>	<b>125 + 2</b>	<b>127</b>
Kommun D	53 + 0	53
Kommun E	32 + 1	33
Kommun F	22 + 1	23
Kommun G	18 + 0	18

Redovisningsvariabler som är hierarkiska är vanligt förekommande i befolkningstabeller.

Exemplet illustrerar att man alltid bör hämta uppgifter ur statistikdatabasen istället för att summera själv. Summan av de publicerade statistikvärdena för region Norr och region Syd är  $104 + 127 = 231$ , och det publicerade statistikvärdet är 226. Det adderade bruset blir alltså större än vad som är nödvändigt.

På motsvarande sätt är summan av publicerade statistikvärden för kommuner i Region Norr  $34 + 24 + 43 = 101$  men det publicerade värdet är 104. För Region Syd blir det ingen skillnad.