

Frågor och svar om CKM

Vad är CKM?

CKM är en förkortning för Cell-Key-Method. Det är en röjandekontrollmetod som är utvecklad för frekvenstabeller som bygger på totalräknade uppgifter. Metoden har utvecklats av den Australiensiska statistikbyrån ABS. Vid användningen av CKM tillförs slumpmässig osäkerhet (även kallat brus), på ett kontrollerat sätt, i samband med framställningen av statistikvärden.

Alla objekt i det register som frekvenstabellerna produceras från får varsitt permanent slumpstal. Cellnyckeln är summan av dessa slumpstal för de objekt som bidrar till statistikvärdet – den aktuella cellen i frekvenstabellen. Cellnycklarna bidrar till att avgöra vilket brus som ska tillföras cellen.

Var används CKM?

Vid ABS används CKM bland annat i det interaktiva verktyget TableBuilder.

CKM infördes som röjandekontrollmetod i Befolkningens arbetsmarknadsstatus (BAS) från och med den första redovisningen i maj 2022.

Eurostat rekommenderar medlemsländerna att använda CKM för Census 2021.

Vad är fördelen med att använda CKM?

Producenter av officiell statistik är skyldiga att beakta röjanderisker i den redovisade statistiken. Därutöver finns ett behov att säkerställa att statistikproducenterna har fortsatt förtroende hos både användare och uppgiftslämnare.

Det är en stor fördel med en enhetlig och automatiserad metod så att separata manuella bedömningar inte behöver göras varje gång statistik ska publiceras. Det sparar resurser och ger mindre utrymme för onödig variation och subjektiva val, samt ger ett bättre skydd mot röjande. Metoden är enhetlig oavsett vad variabeln avser för typ av uppgifter eller hur detaljerade statiken är.

Innebär användning av CKM att det inte finns någon risk för röjande kvar?

CKM har bedömts ge ett fullgott skydd, i enlighet med gällande lagstiftning. Vid införandet av en röjandekontrollmetod kan de publicerade uppgifternas användbarhet påverkas. Valet av parametrar för CKM, såsom brusets storlek, har gjort så att informationsförlusten är så liten som möjligt givet ett fullgott skydd. CKM syftar till att skydda de objekt som statistiken avser, till exempel individer. Metoden skyddar inte mot röjanderisker för andra objekt som statistiken eventuellt även ger information om. Till exempel skyddar inte CKM mot risken att röja information om antal sysselsatta per företag i individbaserad sysselsättningsstatistik vid redovisning efter egenskaper kopplade till företagen, såsom näringsgren eller belägenhet.

Innebär användning av CKM att det går att redovisa hur detaljerad statistik som helst?

SCB:s bedömning är att CKM ger ett tillräckligt röjandeskydd, i förhållande till statistikens användbarhet. Det är därför i huvudsak andra aspekter än röjanderisk som ska beaktas vid framställning av mer detaljerad statistik. Det handlar om att tillförlitligheten i den detaljerade redovisade statistiken påverkas av andra osäkerhetskällor såsom ramtäckning och modellantaganden. Detta tillsammans med att informationsvärdet inte nödvändigtvis motiverar en alltför detaljerad statistik.

Vad menas med att CKM ger konsistenta statistikvärden?

För statistikvärden som bygger på uppgifter från samma objektmängd kommer samma brus att appliceras och därför kommer dessa brusade statistikvärden att överensstämma numeriskt. Exempelvis om ett statistikvärde baserat på samma objektmängd tas fram vid olika tillfällen eller förekommer samtidigt i olika tabeller.

Vad menas med att CKM ger tabeller som inte är additiva?

Bruset tillförs varje statistikvärde var för sig. En följd av detta är att de redovisade totalerna i en frekvenstabell som innehåller brusade statistikvärden inte alltid är lika med summan av deras redovisade delar. Till exempel överensstämmer inte nödvändigtvis den redovisade totalen för kvinnor och män med summan av de redovisade statistikvärdena för kvinnor respektive män. Skillnaderna blir dock i de allra flesta fall små och förändrar inte det statistiska informationsvärdet.

Vad innebär det att CKM minskar risken för differentiering?

Differentiering är i röjandekontrollsammanhang en benämning för en typ av situation då det skulle kunna bli ett röjanderisksscenario. Detta genom att använda information från minst två statistikvärden där redovisningsgruppen för det ena statistikvärdet utgör en delmängd av den andra. Ett sådant exempel är om ett statistikvärde avser åldersgruppen 15–19 år, medan ett annat avser 16–19 år. Genom att sammanvända dessa uppgifter ges information om personer för åldern 15 år. Vid tillämpning av CKM minskar risken för att ett sådant risksscenario uppstår eftersom statistiken inte ger möjlighet att röja information om 15-åringarna.

Hur bör statistikproducenten göra när den redovisade statistiken används i framställning av ytterligare statistik?

Den princip som bör tillämpas är att om de komponenter som går in i beräkningen utgörs av statistik som finns publicerad, så är det den publicerade statistiken som bör användas (dvs. brusade statistikvärden). Ett exempel är procentuell utveckling av befolkningen, när befolkningen vid de två tillfällena finns redovisade. För att säkerställa skyddet behöver då brusade statistikvärden över befolkningen vid de två tillfällena användas. Om beräkningen är mer komplicerad eller om de uppgifter som ligger till grund för beräkningen inte finns publicerade så kan motsvarande obrusade statistikvärden användas.

Varför ska vi inte informera användarna om storleken på bruset? Vilken information kan användarna få?

Att användarna av statistiken inte känner till vilka parametervärden statistikproducenten valt ökar röjandeskyddet. Det är orsaken till att information om bland annat brusets storlek inte ska kommuniceras.

SCB har gjort bedömningen att för användningen av statistiken räcker det att ge information om att samtliga statistikvärden, inklusive totaler, som är större än noll slumpmässigt justeras med ett litet negativt eller positivt heltal och att vissa statistikvärden lämnas oförändrade. Dessutom är det relevant att vara tydlig med att påverkan relativt sett är större för låga frekvenser än för höga.

I samband med implementeringen av CKM i Befolkningens arbetsmarknadsstatus togs det fram två texter. En kortare som kan användas som fotnot i anslutning till SSD och en längre som kan användas i dokumentationerna Kvalitetsdeklaration och Statistikens framställning. Se dokumentet *Texter om CKM*.

Finns det något hinder mot att använda CKM när statistiken redovisas i SSD?

Det finns inget hinder mot att använda CKM i kombination med att statistiken redovisas i statistikdatabasen (SSD). Däremot bör summeringsfunktionaliteten i SSD inte användas, utan samtliga redovisade statistikvärden bör framställas innan SSD förses med statistikvärden. Orsaken är att bruset appliceras på varje statistikvärde för sig. Summeras brusade statistikvärden kan osäkerheten bli större än vad som skulle ha tillförts om bruset hade applicerats på de summerade obrusade värdena, och större än vad som är nödvändigt för ett tillräckligt skydd.

Kan CKM användas för statistik med alla typer av objekt?

CKM är i sitt ursprungsutförande utformad för frekvenstabeller. Det spelar ingen roll vilken typ av objekt som statistiken ger information om. Förekomsten av frekvenstabeller är mindre i företagsstatistik än inom befolkningsstatistiken. Om målstorheter är definierade som sammanfattande mått av typen summa eller medelvärde behöver andra riskscenarios beaktas.

Objekt kan vara relaterade till varandra, vilket gör att statistiken också är relaterad. Exempelvis har individ- och hushållsstatistiken en stark koppling då individer ingår i hushåll och antal ensamhushåll motsvarar antalet ensamboende personer. Det här är något som statistikproducenten behöver ta i beaktan om CKM ska implementeras både i individ- och hushållsstatistik.

Var kan jag läsa mer om CKM?

Det finns flera rapporter från ABS där metoden beskrivs, se till exempel:

Fraser, B. och Wooton, J. (2005). A Proposed Method for Confidentialising Tabular Output to Protect Against Differencing. Joint UNECE/Eurostat work session on statistical data confidentiality, Geneva, Switzerland, 9-11 November 2005.

Marley, J. K. och Leaver, V. L. (2011). A Method for Confidentialising User-Defined Tables: Statistical Properties and a Risk-Utility Analysis. Proceedings of 58th World Statistical Congress, Dublin, 21-26 August 2011.

Thompson, G., Broadfoot, S. och Elazar, D. (2013). Methodology for the Automatic Confidentialisation of Statistical Outputs from Remote Servers at the Australian Bureau of Statistics.

