

Implementeringsgruppen  
Stödgrupp metod och process

# Anvisningar för selektiv granskning

## Inledning

I granskning på mikronivå används programmerade kontroller för att upptäcka säkra eller misstänkta fel i inkomna data. Vid traditionell produktionsgranskning sker manuell hantering av alla blanketter som innehåller poster som flaggas av kontrollerna, till exempel genom återkontakter med företagen eller manuell imputering av flaggade poster. Men att manuellt hantera alla blanketter med flaggade poster är både kostsamt, tidskrävande och i många fall onödigt. Blanketter med misstänkta fel som har liten inverkan på statistikens kvalitet kan lämnas utan åtgärd.

Syftet med selektiv granskning är att effektivisera granskningsarbetet genom att välja ett urval av blanketter med flaggade poster och avgränsa de manuella momenten till dessa. Metoden är ändamålsenlig om vi lyckas välja de blanketter som innehåller de fel som har störst effekt på statistikens kvalitet. En enkel men i vissa fall tillräcklig variant av selektiv granskning är att avgränsa den manuella hanteringen till blanketter från de största företagen.

Vid implementeringen av den nya granskningsprocessen är målet att de manuella momenten i granskningen ska minska. De undersökningar som ska ha manuella moment i granskningen på mikronivå behöver därför implementera selektiv granskning för att välja ut det som ska hanteras manuellt. Mängden manuell hantering behöver anpassas efter behovet av granskning och tillgängliga resurser. Om inga manuella moment ska göras i mikrogranskningen så behövs ingen selektiv granskning.

Rekommendationen för de undersökningar som ska ha manuella moment i mikrogranskningen men som inte redan har selektiv granskning, är att implementera en enkel men ändamålsenlig metod som inte kräver omfattande förvaltning. Om avgränsning av manuell hantering till stora företag bedöms ge efterfrågad kvalitet så bör det alternativet väljas. I de här anvisningarna ges råd för implementering av en något mer avancerad variant som ibland brukar kallas Selekt-light.

## Prioriterade målstorheter

Det första steget vid implementering av selektiv granskning är att välja ut de prioriterade målstorheter för vilka de manuella momenten förväntas ge sökt kvalitet.

- I Stinas undersökning är totalerna för variablerna x och y viktigast för hela populationen och för redovisningsgrupperna a och b. Om misstänkta fel är så stora att de riskerar att påverka tillförlitligheten i dessa skattningar så tycker Stina att de ska hanteras manuellt.

Valet av prioriterade målstorheter görs utifrån undersökningens kvalitetskrav och tillgängliga resurser.

## Granskningskontroller

De kontroller av mikrodata som flaggar poster för möjlig manuell hantering ska vara ändamålsenliga och uppdaterade. Det innebär att kontrollerna ska ha hög träffsäkerhet (sannolikheten att en flaggad post faktiskt är fel är hög) och ha god förmåga att upptäcka betydelsefulla fel. Rekommendationen är att använda ett fåtal väl fungerande kontroller och uppdatera dessa löpande.

## Val av blanketter för manuell hantering

Uppgiften är att identifiera de blanketter som innehåller fel som riskerar att påverka tillförlitligheten hos skattningarna av de prioriterade målstorheterna. Analysen görs lämpligen för varje målstorhet var för sig.

För en flaggad post som bidrar till skattningen av en av målstorheterna skulle vi idealt vilja räkna upp det faktiska felet i det observerade värdet med designvikten och sedan ställa resultatet i proportion till den färdiga statistiken för målstorheten. Det skulle ge en god bild av felets påverkan på statistiken och utgöra ett bra underlag för ett beslut om felet är värt att hantera manuellt (men om vi kände till det faktiska felet skulle vi inte behöva hantera posten manuellt).

Problemet är dock att vi under pågående insamling inte vet om en flaggad post faktiskt är fel eller hur stort felet i så fall är. Oftast har vi bara en misstanke om ett fel. Vi känner inte heller till skattningen av den berörda målstorheten då mikrogranskningen oftast görs medan insamlingen ännu pågår. Vi får därför nöja oss med att jämföra ett uppräknat *förväntat* fel med ett *förväntat* statistikvärde.

Förväntat statistikvärde kan till exempel fås från tidigare undersökningsomgångar. Antingen kan det senaste statistikvärdet användas eller någon form av medelvärde från tidigare omgångar. Valet beror av hur statistiken är beskaffad. Om statistiken inte

producerats förut kan det förväntade statistikvärdet möjligen uppskattas från annan tillgänglig information.

Förväntat fel i en post kan också skattas på lite olika sätt. En enkel variant är att låta postens värde i sig utgöra en indikator för det förväntade felet, under antagandet att det är de stora posterna som rymmer de stora felen. Då undersöks den uppräknade postens värde i förhållande till det förväntade statistikvärdet för målstorheten. En brist är dock att metoden kommer ignorera poster för vilka felet är att det observerade värdet är för litet.

En vanligt förekommande variant är att skatta ett förväntat värde för posten och undersöka differensen mellan det observerade värdet och det förväntade värdet. Skattningen av postens förväntade värde kan betraktas som en imputering och det finns många olika metoder tillgängliga för imputering. Om företaget varit med i undersökningen tidigare kan tidigare lämnat data utgöra basen för skattningen. Om företaget inte varit med förut kan imputeringen till exempel göras med en närmaste-granne-metod.

Vidare behövs ett tröskelvärde för att avgöra om det uppräknade förväntade felet i proportion till det förväntade statistikvärdet för målstorheten är stort eller litet. Ett högt tröskelvärde medför att färre blanketter går till manuell hantering vilket är positivt ur kostnadssynpunkt medan ett lågt tröskelvärde ger mer manuell hantering vilket är positivt för statistikens tillförlitlighet. För att bestämma tröskelvärdet för en viss målstorhet rekommenderas en laboration på granskad och ogranskad data från tidigare undersökningsomgångar där kvalitetskrav och tillgängliga resurser för manuell hantering vägs in.

Antag att en prioriterad målstorhet är totalen av variabeln  $x$  i redovisningsgruppen  $a$  och att det observerade värdet  $x_k$  används i skattningen av målstorheten. Det uppräknade förväntade felet ges av  $|d_k(x_k - x_{k,förväntad})|$  där  $d_k$  anger designvikten för objektet  $k$ . Om tröskelvärdet för totalen av variabeln  $x$  i redovisningsgruppen  $a$  uttryckt i procent betecknas  $t_{x,a}$  och  $\hat{x}_{a,förväntad}$  betecknar förväntat statistikvärde, kan relationen som ska vara uppfylld för att blanketten ska skickas till manuell hantering uttryckas som

$$100 \cdot \frac{|d_k(x_k - x_{k,förväntad})|}{\hat{x}_{a,förväntad}} > t_{x,a}. \quad (1)$$

Rekommendationen är att fatta beslutet om manuell hantering av blanketten för var och en av de prioriterade målstorheterna individuellt. Det är även möjligt att beräkna ett sammantaget mått för blanketten baserat på samtliga prioriterade målstorheter, men då är det i många fall lättare att anpassa tröskelvärdena i laborationen för att få en bra balans mellan resursbesparing och tillförlitlighet.

- Stina har bestämt en metod för att skatta förväntade värden för variablerna x och y för de uppgiftslämnare som flaggas av undersökningens kontroller och hon har tagit fram förväntade statistikvärden för totalerna för x och y för hela populationen och för redovisningsgrupperna a och v. Efter en laboration på (total)granskade data och ogranskad data bestämmer hon sig för följande tröskelvärden:

<b>Målstorheter (totaler)</b>		<b>Tröskelvärde</b>
a	x	0,02%
a	y	0,02%
b	x	0,03%
b	y	0,03%
hela populationen	x	0,05%
hela populationen	y	0,05%

Om relationen (1) är uppfylld för en post i en blankett för någon av målstorheterna då bestämmer Stina att blanketten ska skickas till manuell hantering.