

Kompetensgruppen för granskning och mätning

Hantera coronaeffekter i granskningen

1 Inledning

Med hänsyn till coronapandemin så kommer mycket av det data SCB samlar in att skilja sig åt jämfört med tidigare perioder. Det kommer att påverka flera av SCB:s produktionsprocesser. Granskning av data från företag och organisationer är en aktivitet som påverkas mycket.

I granskningen används ofta tidigare perioders värden för att fel-signalera nyinkomna värden om de skiljer sig mycket åt. Vid en händelse som coronapandemin, där stora delar av det ekonomiska systemet satts på paus, kommer historiska värden att vara dåliga prediktorer för att hitta mätfel.

Risken blir då att granskningen genererar långa listor med misstänkta data utan att det går att påvisa fel. Det blir stor arbetsbörda internt på SCB och stor uppgiftslämnarbörda om vi gör många återkontakter. Det gör att arbetssättet kan behöva justeras beroende på produktens förut-sättningar och behov, vilket det här dokumentet är tänkt att stödja.

Coronaeffekter som kan påverka granskningsprocessen:

- Minskat inflöde, stort objektsbortfall och stort partiellt bortfall.
- Variabelvärden som avviker stort från vad som varit historiskt normalt.
- Uppgiftslämnare som inte är nåbara vid återkontakt.

Problem som kan uppstå i granskningen till följd av dessa effekter:

- Uppgiftslämnarna hindras från att skicka in uppgifter på grund av rigida kontroller i uppgiftslämnargranskningen som inte accepterar partiellt bortfall eller stor förändring gentemot föregående period(er).
- Produktionsgranskningen överbelastas av alltför många signaler om misstänkta fel.
- Uppgiftslämnaren finns inte tillgänglig vid utredning av misstänkt felaktiga värden.
- Outputgranskningen kan inte förlita sig på jämförelser med tidigare perioder.

Rent generellt kan man säga att man i granskningen inte längre kan förlita sig på empiriska historiska slutsatser om vad som är normalt. Fokus i granskningen kan därför behöva förflyttas från historiska jämförelser till jämförelser med data eller prognoser som refererar till aktuell referensperiod.

Man kan också anta att många produkter kommer att få avsätta mer resurser till outputgranskningen, och dessutom påbörja denna tidigare, än som förut varit fallet.

Det här dokumentet kommer att ta upp generella problem i uppgiftslämnargranskning, produktionsgranskning och outputgranskning vid kraftiga svängningar i indata, samt beskriva åtgärder som kan vidtas för att minimera effekterna.

Naturligtvis kan inte dokumentet vara heltäckande för alla undersökningars problem. Det är därför viktigt att problemen identifieras och diskuteras i en bred konstellation där insamling, ämne, metod (inkl. mätteknik) deltar (t.ex. prodsamteamet). Det är upp till prodsamteamet att avgöra när en återgång till ett normalläge i granskningen är aktuellt.

För stöd vid mer specifika granskningsproblem kan man vända sig till *kompetensgruppen för granskning och mätning*, t.ex. via gruppbrevlådan Granskning@scb.se.

2 Uppgiftslämnargranskning

En fördel med uppgiftslämnargranskning är att respondenten redan vid registrerandet av värdet får återkoppling om att SCB bedömer att värdet kan innehålla mätfel.

I och med att ekonomin har förändrats så pass mycket pga. coronapandemin kan nuvarande gränser vara irrelevanta och felsignalering av värden grovt missa målet att uppmärksamma uppgiftslämnaren på eventuella mätfel. Det är då viktigt att kontrollerna ger uppgiftslämnaren "en väg ut" via en kommentar och/eller att kontrollen är mjuk så att ett felmeddelande kan passeras utan större besvär.

Mjuk kontroll utan kommentar – kontrollen flaggar variabelvärden som misstänkt fel men felmeddelandet kan ignoreras av uppgiftslämnaren.

Mjuk kontroll med kommentar – kontrollen flaggar variabelvärden som misstänkt fel men felmeddelandet kan ignoreras. Uppgiftslämnaren har även möjlighet att lämna en kommentar.

Hård kontroll med kommentar – kontrollen flaggar variabelvärden med misstänkta fel. För att komma vidare måste uppgiftslämnaren korrigera värdet eller lämna en kommentar som bekräftar att värdet är korrekt.

Hård kontroll utan kommentar – uppgiftslämnaren måste korrigera värdet för att komma vidare. Denna variant ska **aldrig** användas för kontroller som signalerar misstänkta fel, t.ex. då värden hamnar utanför ett specificerat acceptansområde.

Våra rekommendationer angående uppgiftslämnarkontroller skiljer sig åt mellan första insamling av data där coronapandemin kan ha haft effekt och senare insamlingsomgångar.

Att göra inför datainsamlingen avseende den första referensperioden där corona kan ha haft effekt:

- Använd inte hårda kontroller utan kommentar för avvikelsekontroller (med acceptansområde).
- Ge uppgiftslämnaren möjlighet att lämna kommentar till alla misstänkta fel, även vid mjuk kontroll utan kommentar, för att minska behovet av återkontakter.
- Vidga acceptansområden, antingen uppåt eller nedåt beroende på hur respektive variabel och delpopulation påverkas av coronapandemin.

Exempel 1: Förändring av omsättning jämfört med föregående insamlingsomgång har acceptansområdet $\pm 10\%$, ändra till $-20\% - 10\%$

Exempel 2: Omsättning under månaden dividerat med antal anställda har i en bransch acceptansområdet 100 tkr – 300 tkr, ändra till 50 tkr – 300 tkr

Att göra inför datainsamling avseende den andra referensperioden där corona kan ha haft effekt:

- Överväg att ta bort kontroller som blev ett problem för den första referensperioden där corona hade en effekt.
Exempel 1: Kontroller som leder till många kommentarer av samma typ.
Exempel 2: Kontroller som uppgiftslämnarna lämnat förklaringar till avseende den första referensperioden, där dessa förklaringar kan väntas gälla så länge coroneffekterna kvarstår.
- Återställ acceptansområden för kontroller som jämför ett värde enbart mot föregående referensperiod, om coroneffekten bedöms ha haft fullt genomslag föregående referensperiod.

3 Produktionsgranskning

En väl fungerande produktionsgranskning bidrar till att snabbt fånga felaktiga värden löpande under insamlingsperioden. Det förkortar produktionstiden och minskar outputgranskningen. En del av förutsättningarna är att granskningen sker kontinuerligt under insamlingen vilket innebär att kontrollerna ofta baseras på objektets tidigare

rapporter. Nu när tidigare perioder är en dålig prediktor av framtida värden så kan granskningsrutinerna behöva justeras.

Man kan förvänta sig att fler variabelvärden kommer att flaggas av granskningskontroller vid kraftiga svängningar i ekonomin. Då fler objekt hamnar på listan behövs en prioritering av vilka som ska utredas, eventuellt med återkontakt. Att kartlägga de uppgiftslämnare vars data påverkar statistiken mest är nödvändigt.

Det kan vara lämpligt att använda sig av en poängfunktion som rangordnar de flaggade objekten. Man kan behöva acceptera att det inte är möjligt att granska alla de objekt vilka normalt skulle ha granskats om det inte funnits några coronaeffekter. Om selektiv granskning med SELEKT görs och scorefunktion används bör indikatorn SELEKTOR , alternativt global poäng, fungera tillfredsställande, om än inte perfekt.

Ett alternativ till att felsignalera objekt med hjälp av historiska inrapporterade värden är att använda sig av fördelningen av inkomna svar för att studera vilka av dessa som avviker mer än snittutvecklingen. Nackdelen är att en sådan granskning endast kan göras först när en större mängd svar kommit in vilket gör att produktionsgranskningen förväntas kunna påbörjas senare. Ett exempel på en sådan metod är HB-metoden vilket beskrivs i *Guide till granskning* som finns tillgänglig under 5.2 *Granska mikrodata* i verksamhetsstödet.

Våra rekommendationer angående granskningskontroller i produktionsgranskningen skiljer sig åt mellan första insamling av data där coronapandemin kan ha haft effekt och senare insamlingsomgångar.

Att göra inför datainsamlingen avseende den första referensperioden där corona kan ha haft effekt:

- Skapa en "lista" över alla objekt (företag) i undersökningen med en kolumn som är designvikt (stora N dividerat med lilla n) multiplicerat med ett genomsnitt av företagets värde på en central variabel för undersökningen ett par referensperioder bakåt i tiden (eller längre period). Den centrala variabeln ska vara kvantitativ och något sätt spegla företagets betydelse i undersökningen t.ex. omsättning, investering, eller antal anställda. De objekt som har störst "uppräknad normalnivå" bör prioriteras.
- Försök utnyttja uppgiftslämnarnas kommentarer vid utredning så att återkontakt kan undvikas.
- Vidga acceptansområden , antingen uppåt eller nedåt beroende på hur respektive variabel och delpopulation påverkas av coronapandemin.

Exempel 1: Förändring av omsättning jämfört med förra

insamlingsomgången har acceptansområdet $\pm 10\%$, ändra till $-20\% - 10\%$.

Exempel 2: Omsättning under månaden dividerat med antal anställda har i en bransch acceptansområdet 100 tkr – 300 tkr, ändra till 50 tkr – 300 tkr.

Att göra inför datainsamling avseende den andra referensperioden där corona kan ha haft effekt:

- Försök utnyttja uppgiftslämnarnas kommentarer, även från tidigare perioder som påverkats av corona, vid utredning så att återkontakt kan undvikas.
- Återställ acceptansområden för kontroller som jämför ett värde enbart mot föregående referensperiod, om coroneffekten bedöms ha haft fullt genomslag föregående referensperiod.

4 Outputgranskning

Syftet med outputgranskning är att säkerställa att statistikvärden kan publiceras med tillräckligt god kvalitet. Aktiviteten utförs först när insamlingen är avslutad, eller nästintill avslutad, och syftar till att hitta kvarvarande mätfel i data samt eventuella bearbetningsfel. I princip kan liknande metoder för outputgranskning användas av företags-, organisations-, individ- och hushållsundersökningar.

Det finns olika metoder som kan användas vid outputgranskning. Vanligt förekommande är att använda sig av *aggregatmetoden* där exempelvis punktestimat kan studeras i en tidsserie för att borra sig ned till finare nivåer när något avviker. Denna metod kan, utan justering för coroneffekter, vara alltför trubbig för att hitta mät- eller bearbetningsfel. För att förbättra metoden i den aktuella situationen skulle acceptansintervallet, utanför vilket ett aggregat signaleras som misstänkt, behöva justeras.

Detta innebär i princip att man på förhand måste göra en prognos för vad som kan vara ett rimligt intervall för varje aggregat. För att göra denna prognos kan man använda information från tidigare delar av granskningsprocessen eller andra undersökningar. Kanske kan man också använda bedömningar som gjorts av andra organisationer, t.ex. Konjunkturinstitutet, eller utifrån övrig omvärldsbevakning?

Fördelningsmetoden är en metod som, med tanke på coroneffekter, är bättre lämpad för att peka ut potentiella outliers eller mätfel än en ojusterad aggregatmetod. Med denna metod studeras enbart fördelningen av inkomna värden för den aktuella referensperioden. Värden som avviker från mönstret kan identifieras och utredas. För att lättare hitta avvikelser kan detta göras i skärningar efter bransch och/eller i kombinationer av flera variabler.

Tänk på att prova ut nya strategier för outputgranskning i god tid. Tester av olika metoder kan t.ex. baseras på bara en liten andel svarande och eventuellt imputeringar för resten av målobjekten. Om dessa tester leder till att misstänkta värden identifieras kan de också ge värdefull input till produktionsgranskningen.

För mer information om aggregatmetoden och fördelningsmetoden se dokumentet *Handledningen för outputgranskning*, som finns tillgängligt under delprocess 6.2 *Granska makrodata* i verksamhetsstödet.