



*Margareta Stålnacke, Anne Gro Hustoft og
Leiv Solheim*

Vurdering av kvalitet i statistikk

En oversettelse av notater fra Eurostat
om kvalitetsrapportering

Innhold

1. DEFINISJON AV KVALITET I STATISTIKK.....	4
1.1 Innledning	4
1.2 Definisjon av kvalitet.....	4
1.3 Nøyaktighet	6
1.4 Aktualitet og punktlighet	6
1.5 Klarhet og tilgjengelighet i statistikken	7
1.6 Sammenlignbarhet.....	7
1.7 Sammenheng	8
1.8 Kostnadskrav	8
1.9 Noter	9
1.10 Referenser	11
1.11 Vedlegg	12
2. FORSLAG TIL STANDARD FOR KVALITETSRAPPORTERING	13
2.1 Innledning	13
2.2 Relevans.....	13
2.3 Nøyaktighet	13
2.3.1 Utvalgsfeil	14
2.3.2 Ikke-utvalgsfeil	16
2.4 Planlegging og punktlighet.....	22
2.5 Tilgjengelighet og klarhet	22
2.6 Sammenlignbarhet.....	23
2.6.1 Romlig sammenligning	23
2.6.2 Sammenligning over tid	23
2.7 Sammenheng	24
2.7.1 Sammenheng mellom foreløpig og endelig statistikk	24
2.7.2 Sammenheng mellom årlig og korttids statistikk	24
2.7.3 Sammenheng mellom statistikk fra samme sosioøkonomiske område	25
2.7.4 Sammenheng mellom statistikk for nasjonalregnskap	25
2.8 Kostnad og budsjett.....	26
2.8.1 Kostnader for respondenten	26
2.8.2 Kostnader og budsjett for statistikkbyrået	26
3. ORDLISTE FOR KVALITETSBEGREP I STATISTIKK	27

VEDLEGG

Assessment of the quality in statistics, Eurostat

A. Item 5.1 of the agenda, Definition of quality in statistics	38
B. Item 5.3 of the agenda, Standard quality report.....	46
C. Item 5.2 of the agenda, Glossary on quality in statistics	60
D. Norsk – Engelsk ordliste	70
E. Engelsk – Norsk ordliste	73

STIKKORDREGISTER.....	76
------------------------------	-----------

De sist utgitte publikasjonene i serien Notater	77
--	-----------

Forord

Eurostat har utarbeidet 3 arbeidsnotater knyttet til kvalitetsrapportering for innsendte statistikker. Disse notatene er "Assessment of the Quality in Statistics" (Doc. Eurostat/A4/Quality/98/General/Definition, Doc. Eurostat/A4/Quality/98/General/Glossary og Doc.Eurostat/A4/Quality/98/General/Standard report). Arbeidsnotatene er utkast, men representerer likevel et viktig grunnlag for SSBs kvalitetsrapporteringsprosjekt ("Årlig kvalitetsrapport"), der målet er å komme fram til et opplegg som kan brukes i en årlig kvalitetsrapportering av SSBs statistikker. SSBs kvalitetsrapport skal gi nødvendig informasjon både til Eurostat og til andre brukere av vår statistikk. Som et første trinn innenfor kvalitetsrapporteringsprosjektet presenteres her en norsk oversettelse av de tre Eurostatnotatene.

1. Definisjon av kvalitet i statistikk

(Assessment of the quality in statistics, Item 5.1 of the agenda, Definition of quality in statistics, Eurostat.)

1.1 Innledning

Dette notat tar opp de hovedpunktene som skal inkluderes i kvalitetsrapporter. Detaljer om hva som skal inngå i kvalitetsrapporter, er presentert i dokumentet 'Forslag til standard for kvalitetsrapporting' ('Standard quality report').

1.2 Definisjon av kvalitet

Kvalitet er i ISO 8402 - 1986 definert som: " de karakteristiske trekk ved et produkt eller en tjeneste som har noe å gjøre med produktet/tjenestens evne å tilfredsstille uttalte og implisitte behov". Dette kan analyseres i rammeverket av 'Total Quality Management'. Vi vill gå inn for en mer begrenset tilnærming til kvalitet gjennom å konsentrere oss om *produktet* heller enn prosessen som helhet. Det vil imidlertid gjøres to unntak: relevans er en verdivurdering og spørsmålet om aktualiteter relatert til dato, noe som er prosessinformasjon.

Statistikkens kvalitet kan defineres med hensyn til flere kriterier:

- de statistiske begrepenes relevans
- estimatenes nøyaktighet
- aktualitet og punktlighet knyttet til spredning av resultater
- informasjonens tilgjengelighet og klarhet
- statistikkens sammenlignbarheter
- sammenhengen mellom statistikker.

De tilgjengelige ressursene i statistikkproduksjonen er en ramme for kvaliteten, selv om de ikke er et mål på kvalitet. Når en vurderer en medlemsstats evne til å etterkomme retningslinjene for kvalitet, er det nødvendig å ta de tilgjengelige ressurser med i betraktningen.

Relevans

En undersøkelse er relevant hvis brukerens behov blir ivaretatt. En identifisering av brukerne og deres forventninger er derfor nødvendig.

Estimatenes nøyaktighet

Nøyaktighet defineres som forskjellen mellom estimert verdi og den (ukjente) sanne populasjonsverdien. For å fastsette et estimats nøyaktighet, må en analysere den totale feilen knyttet til estimatet.

Aktualitet og punktlighet i resultatspredning

De fleste brukere ønsker oppdaterte tall som publiseres ofte og punktlig i forhold til fastsatte publiseringsdatoer.

Tilgjengelighet og klarhet i informasjonen

Statistiske data er mest verdifulle når de er lett tilgjengelige for brukere, finnes på en form brukeren ønsker og er hensiktsmessig dokumentert. Hjelp til bruk og tolkning av statistikken skal også være tilgjengelig hos produsenten.

Statistikkenes sammenlignbarhet

Statistikk for et bestemt fenomen er mest nyttig når den kan brukes til pålitelige sammenligninger på tvers av tid og rom. Det skal være mulig å sammenligne statistikk landene imellom for å kunne vurdere betydningen av aggregert statistikk på europeisk nivå.

Sammenheng

Når statistikker kommer fra en enkelt kilde, er det sammenheng mellom dem slik at de kan settes sammen på flere komplekse måter. Når statistikkene kommer fra ulike kilder, særskilt fra statistiske undersøkelser med ulik hyppighet, er det sammenheng mellom statistikkene dersom de er basert på felles definisjoner, klassifikasjoner og metodiske standarder. Den informasjon disse statistikkene formidler til bruker vil da ha en klar sammenheng, eller i det minste ikke være motstridende. Sammenheng mellom statistikker er rettet mot sammenligning av ulike statistikker, der disse som oftest produseres på ulike måter og for forskjellig primær bruk.

Angående kostnader

Der finnes helt klart en sterk sammenheng mellom kvaliteten på statistikken og de ressurser som er tilgjengelige for å produsere dem. Kostnadsvurderinger skal tas hensyn til når kvaliteten evalueres.

1.3 Nøyaktighet

Denne side av kvalitet har blitt studert inngående i mange statistikkbyråer og av forskere. Nøyaktighet fokuserer på analyse av feil, inndelt i utvalgsfeil og ikke-utvalgsfeil.

Feil kan klassifiseres som følger:

- Utvalgsfeil
- Ikke-utvalgsfeil
 - Trekkeregisterfeil
 - Målefeil
 - Bearbeidingsfeil
 - Feil pga. frafall
 - Modellforutsetningsfeil

Groves (1989), Lessler og Kalsbeek (1992), Särndal, Swensson og Wretman (1992), Biemer og Fesco, 'Quality Measurement Model of the BLS' (1995) bruker lignende klassifiseringer.

Til nå har statistikere først og fremst konsentrert seg om utvalgsfeil. Publiserte nøyaktighetsestimat fra statistikkbyråer dekker vanligvis kun utvalgsfeil¹. Imidlertid utvikler også forskningen på ikke-utvalgsfeil seg raskt, og noen metoder finnes nå for en første vurdering av slike feil. Se for eksempel Biemer et al. (1991).

1.4 Aktualitet og punktlighet

Vanligvis krever brukere at statistisk informasjon produseres på minst mulig tid, at statistikkene frigis så fort som mulig og at publiseringen skjer regelmessig der dette er relevant. For å holde produksjonstiden lavest mulig må en bruke effektive produksjonsteknikker, noe som ofte koster mye. Hele prosessen fra datainnsamling, editering, imputering og estimering til spredning av informasjon, må holdes under kontroll for å redusere bearbeidingsperioden. En deløsning på problemet er å beregne foreløpige estimater basert på et delutvalg av respondenter. En må være forsiktig når en presenterer disse foreløpige tallene, og likeså ved den følgende spredningen av de endelige tallene.

¹ Faktisk er den estimerte variansen i praksis avhengig av svarprosenten og visse ikke-utvalgsfeil - for eksempel tilfeldige målefeil - men den forklarer ikke alle ikke-utvalgsfeil. Særlig er de fleste systematiske feil ikke inkludert i dette målet på nøyaktighet.

I blant vil det være konkurranse mellom statistisk informasjon (ofte privat) som produseres og spres raskt, men med en lavere pålitelighet, og den mer tidkrevende og presise statistiske informasjonen fra de statistiske byråene. Avveiningen mellom aktualitet og nøyaktighet har svært tydelige følger for visse brukere. Dette utfordrer de statistiske sentralbyråene til å forbedre aktualiteten til statistikken uten å redusere de høye kravene til nøyaktighet.

1.5 Klarhet og tilgjengelighet i statistikken

Spredningen er en meget viktig del av informasjonskjeden. Det er ikke nok å ha 'pålitelig statistikk' lagret et sted i statistikkbyrået. Statistikken må gjøres tilgjengelig for alle potensielle brukere på en passende form. For det første skal brukeren ha mulighet til å finne lett fram til hva slags statistikk som er tilgjengelig. For det andre skal den fysiske tilgjengeligheten til statistikk være enkel, og for det tredje skal nødvendig informasjon om begreper, prinsipper og metoder følge med statistikken. Ulike nivåer på forklaringer kan i visse tilfeller være passende for å skille mellom de som er fagspesialister og andre. Endelig kan analysen av statistikk legge vekt på servicedimensjonen knyttet til statistiske produkter.

1.6 Sammenlignbarhet

Det kan være forskjell mellom nasjonale begrep og europeiske definisjoner. Anta for eksempel at en medlemsstat har en definisjon på bruttoinvesteringer som utelukker verdien på kapitalvarer når de er anskaffet gjennom leie, mens slike kapitalvarer inngår i den europeiske definisjonen. Hvis det ikke er mulig å samle inn data på individnivå for bedrifter om kapitalvarer som er leid, må et estimat lages på et aggregert nivå. Sammenligningen er ikke begrenset slik at den kun skal fungere innenfor EU: Eurostat må kunne vurdere sammenlignbarheten mellom EU-statistikk og statistikk fra andre land eller grupper av land (USA, Japan).

Sammenlignbarhet skal også eksistere over tid. Endringer som skyldes forandringer i begreper, eller der årsaken er selve prosessen, skal dokumenteres, og virkningene av disse forandringene vurderes. På samme måte skal det tas hensyn til generelle forandringer i samfunnet (f.eks. ny lovgivning, fusjoner og fisjoner) som virker inn på kontinuiteten.

1.7 Sammenheng

Dersom det finnes samsvarende statistikk fra ulike kilder² skal disse identifiseres og eventuelle forskjeller om mulig kvantifiseres. Uoverensstemmelse mellom to sett av statistikker som er et produkt av ulike undersøkelser, kan skyldes forskjell i datainnsamlingsprosessen, eller ulikheter i rapportering som fører til forskjeller i estimatene. Situasjonen kan bli forbedret gjennom f.eks. å hente inn månedlige eller kvartalsvise tall som rapporteres på årsbasis eller gjennom å kombinere resultater fra ulike undersøkelser. I alle tilfeller skal det brukes forskjellige formuleringer for ulike begreper for å unngå at brukerne misforstår.

I tillegg skal sammensatte statistiske mål (forhold, elastisitet, osv.) være basert på sammenhengende elementære statistiske mål (med sammenlignbare definisjoner fra referanspopulasjonen(e), kjennemerke, referanseperiode og statistiske enheter).

1.8 Kostnadskrav

Kvalitet er forbundet med kostnader. Informasjon om hvilke ressurser de ansvarlige for undersøkelser rår over, gjør det mulig å vurdere hvordan budsjettet kan optimaliseres. De ressurser som finnes hos et medlemsland, skal tas med i beregningen ved vurderingen av dets evne til å tilfredsstille kvalitetskravene. To kostnadskomponenter må vurderes, *statistikkbyråets kostnader* og *oppgavegivers kostnader* (bedrifter eller deler av bedrifter, husholdninger, personer). Kostnadene for oppgavegiver varierer ofte med størrelsen på den statistiske enheten, hovedsakelig etter størrelse på bedrift og aktivitetsområde³. Kostnadene for en ny undersøkelse kan bli mye høyere enn kostnadene for en pågående undersøkelse der bare deler av utvalget byttes ut fra gang til gang. Bruk av administrative data reduserer belastningen på statistiske enheter kraftig.

Kostnader for personer eller husholdninger er i alminnelighet ikke tatt hensyn til. Effekten av oppgavebyrden for personer/bedrifter tas med i beregningene for frafallsfeil.

Kostnad er mer en begrensning på kvalitetsforbedring enn en delkomponent i selve kvalitetsbegrepet.

² Med 'kilde' menes undersøkelser eller statistiske beregninger fra administrative data på nasjonal nivå laget av det offisielle statistiske systemet.

³ Ettersom ulike spørsmål kan bli brukt i ulike sektorer.

1.9 Noter

Note 1: Det er en avveining mellom forskjellige kvalitetskomponenter, spesielt:

aktualitet / nøyaktighet

nøyaktighet / sammenlignbarhet i rom⁴

relevans / sammenlignbarhet over tid

relevans / nøyaktighet

sammenheng for store grupper (områder) / relevans for små grupper (under-grupper)

Note 2: Oppsplittingen av kvalitet i komponenter er ikke entydig. Andre organisasjoner har foreslått andre komponenter for felles bruk i statistikk:

1) *Forente Nasjoner* (1983). Ni punkter i alt, inkludert:

- 'Sammenlignbarhet over tid'
- 'Sammenlignbarhet med andre statistikker'.

2) *Statistikkbyrået i Canada/Statistics Canada* (1992). Ni punkter i alt, inkludert:

- 'Sammenlignbarhet over tid'
- 'Sammenlignbarhet med data fra andre kilder'.

3) *Beekman og Struijs* (1993). Tre hovedkategorier som inkluderer:

- 'Samsvar mellom statistiske resultatdata'; sammenligne eller kombinere data fra en statistikk med data fra andre med hensyn på sammenlignbarhet over tid (kontinuitet i tidsserier) og sammensvar med andre statistikker for samme periode/tidspunkt.

⁴ Sammenligning i rommet kan medføre bruk av begreper som ikke er tilpasset det enkelte land. På den annen side kan bruk av begreper som er relevante i det enkelte land føre til at sammenligning blir vanskelig.

4) *Statistikkbyrået i Nederland/Statistics Netherlands* (1995). Seksjon II.2 i 'The need of coherence' tar opp syv aspekter:

- additiv evne hos statistikk knyttet til ulike felt innenfor økonomisk aktivitet, f.eks. fast kapitaldannelse i handel og industrinæring
- sammenlignbarhet hos statistikk innenfor ulike områder, f.eks. arbeidsledighetdata fra arbeidskraftundersøkelser og statistikk over endring i antall ansatte fra produksjonsstatistikken
- samsvar mellom tall fra korttidsstatistikk og årsstatistikk
- samsvar mellom resultater fra regional statistikk og nasjonal statistikk
- kontinuitet i resultatene for periode t og periode $t+1$
- sammenlignbarhet mellom nasjonal statistikk og internasjonal statistikk
- overensstemmelse i begreper og resultater mellom bedriftsstatistikken og 'the System of National Accounts (SNA)' og 'the European System of Accounts (ESA)'.

5) *Statistisk centralbyrå i Sverige/Statistics Sweden* (1994). Fire komponenter er valgt ut: Innhold, Tid, Pålitelighet (Relialibilitet) og Tilgjengelighet. Innhold inkluderer sammenlignbarhet med andre statistikker og Tid inkluderer sammenlignbarhet over tid. En meget brukervennlig tilnærming med stor verdi knyttet til en kvalitetsdeklarasjon. Imidlertid foreslår E. Elvers og B. Rosén i sitt bidrag (1997) til 'Encyclopedia of Statistical Science', Wiley and Sons, en oppsplitting i fem komponenter. Disse svarer til de første seks komponentene hos Eurostat, men hos Elvers/Rosén grupperes sammenheng og sammenlignbarhet sammen.

1.10 Referenser

Biemer, P.P. and Fesco, R. S. (1995). *Evaluating and controlling measurement error in business surveys*, in Cox and al. (eds) Business Survey Methods, New York: John Wiley, pp. 257-281.

Biemer, P., Groves, R.M., Lyberg, L.E., Mathiowetz, N.A. and Sudman, S. (eds) (1991). *Measurement Errors in Survey*; New York: Wiley.

Beekman, M. M. and Struijs, P. (1993). *The quality of economic concepts and definitions*. Statistical Journal of the United Nations ECE 10, 1-15.

Bureau of Labour Statistics (1995). Quality Measurement Model, presented at the International Conference on Survey Measurement and Process Quality; Bristol (UK).

Elvers, E. and Rosén, B. (1997). Quality Concept for Official Statistics, Entry in the forthcoming update of the Encyclopedia of Statistical Sciences, Wiley & Sons.

Groves, R. M., (1989). *Survey Errors and Survey Costs*, New York: John Wiley.

Lessler, J. T. and Kalsbeek, W. D. (1992). *Nonsampling Errors in Surveys*, New York: John Wiley.

Särndal, C. E., Swensson, B. and Wretman, J. (1992). *Model Assisted Survey Sampling*, New York: Springer-Verlag.

Statistics Canada (1992). *Policy on Informing Users of Data Quality and Methodology*. Policy manual, Statistics Canada, April 7, 1992.

Statistics Netherlands (1995). *Reference Manual in Design and Implementation of Business Surveys*. Koeijers, E. and Willeboordse, A. (editors). First draft, on request of Eurostat.

Statistics Sweden (1994). *Quality definition and recommendations for quality declarations of official statistics* (authors: Chris Denell, Eva Elvers and Bengt Rosén).

United Nations (1983). *Guidelines for quality presentations that are prepared for users of statistics*. Statistical Commission and Economic Commission of Europe, Conference of European Statisticians, Meeting on Statistical Methodology, 21-24 November 1983.

1.11 Vedlegg

Begreper som er brukt

Et grunnleggende begrep i statistikken er det statistiske kjennemerket som defineres som følger: et gitt statistisk mål anvendes for å summere verdiene av en gitt variabel for enhetene i en spesifikk gruppe. Det totale antall enheter i gruppen kalles vanligvis populasjon. Ofte er ulike delpopulasjoner av interesse. Populasjon, enheter og variabler har alle et referansetidspunkt, et spesifikt tidspunkt eller en tidsperiode.

For å svare på spørsmål om fagrelaterte ting, finnes det en mengde passende/interessante statistiske kjennemerker. I blant er det vanskelig å måle disse i en statistisk undersøkelse. Undersøkelsen er rettet mot å bestemme kjennemerker som det er mulig å måle i undersøkelsen, og som godt nok dekker det som er interessant.

En undersøkelse gir estimater for verdiene til de valgte kjennemerkene. Et estimat kan være basert på data fra en totalundersøkelse, observasjoner fra et utvalg, en modell eller en kombinasjon av de nevnte eksemplene. Hjelpvariabler brukes i visse tilfeller sammen med utvalgsdata.

2. Forslag til standard for kvalitetsrapportering

(Assesment of the quality in statistics, Item 5.3 of the agenda, Standard quality report, Eurostat.)

2.1 Innledning

I dette avsnittet behandles et forslag til hva en kvalitetsrapport for statistikk burde inneholde. Det gis en beskrivelse av de emner som skal behandles i rapporten, og det pekes spesielt på den informasjonen som Eurostat forventer å få fra medlemslandene. En slik rapport vil vise hvordan medlemslandene avviker fra Eurostats normer. Noe av bakgrunnen for dette forslaget finnes i avsnittene 'Definisjon av kvalitet i statistikk' og 'Ordlister for kvalitetsbegrep i statistikk'.

2.2 Relevans

For å sikre at statistikken som produseres, er relevant for brukerne, skal medlemslandene utføre regelmessig brukerundersøkelser.

Undersøkelsene skal gjentas med noen års mellomrom:

For hver statistikk skal det gis en brukerbeskrivelse, bakgrunn for og tilfredsstillelse av brukernes behov og statistikkens relevans for brukerne.

2.3 Nøyaktighet

Kunnskap om statistikkens nøyaktighet er svært viktig for brukeren. Denne burde framkomme ved publisering av konfidensintervall for alle tall. Konfidensintervallet burde dekke effekten av korrigerede og ikke-korrigerede feil, samt variansen. Variansen kan skyldes:

- skjevhet i estimatet, avkuttet (cut-off) utvalg, imputeringsmetodene for frafall og andre systematiske feil (som kan skyldes nyvekting, målefeil osv.)
- sannsynlighetsutvalg, feilklassifisering, målefeil, bearbeidingsfeil, imputering, statistiske modeller (f.eks sesongjustering) og vekter korrigeret på grunn av frafall.

Definisjonsmessig er variansen forbundet med et estimat. Ved hjelp av statistiske metoder kan en korrigere for å minske skjevheten. Korreksjonen kan innarbeides i estimatoren. Andre feil som gjøres i datainnsamlingen eller ved dataregistreringen, kan ikke korrigeres for verken ved redigering eller ved estimering, men de skal også regnes med i konfidensintervallet.

For å kunne sammenligne konfidensintervallen fra medlemsland til medlemsland, må de samme komponentene inkluderes i variansberegningen. Som en første tilnærming, foreslår vi å begrense disse komponentene til varians som skyldes tilfeldig utvalg (knyttet til den estimatoren som er brukt), feilklassifisering og imputering på grunn av frafall, selv om noen ikke-utvalgsfeil ikke kan utelukkes og dermed vil bidra til konfidensintervallet. Medlemslandene skal gi informasjon om de forskjellige typer feil.

2.3.1 Utvalgsfeil

Utvalgsfeil kan deles opp etter om de oppstår i sannsynlighetsutvalg eller ikke-sannsynlighetsutvalg. I det første tilfellet, som er den vanligste typen av utvalg, er det spesielt interessant å se på skjevhet og varians.

2.3.1.1 Sannsynlighetsutvalg

Anta at estimering skjer på grunnlag av et tilfeldig utvalg, og hvis andre feilkilder ikke er tatt hensyn til (f.eks. trekkeregisterfeil, målefeil osv.), er kvadratrotten av middelkvadratfeilen et mål for nøyaktigheten. Middelkvadratfeilen er summen av variansen til estimatet og kvadratet av skjevheten. Dette betyr at dersom et utvalg gir et estimat som ikke er skjevt, blir middelkvadratfeilen lik variansen til estimatet.

Tilfeldige utvalg brukes når det finnes et register, og nøyaktigheten kan beregnes for kjennemerker knyttet til de statistiske enhetene i populasjonen. Utvalg som trekkes i flere trinn, kan bruke ikke-tilfeldig trekking av enheter på et av trinnene. Dette er ofte tilfellet i utvalgsundersøkelser knyttet til konsumpriser, der en i første trinn trekker markeder, og i andre trinn spesielle varer i disse markedene.

Skjevhet

Estimatorer for totaler er ofte forventningsrette, f.eks. er de estimat som lages med hjelp av Horvitz-Thompson-estimatoren, er forventningsrett. Det er allikevel vanlig at estimatene er beregnet ved bruk av forhold- eller regresjonsestimatorer. Disse er skjeve, men konsistente. Av og til brukes mer robuste, men skjeve, estimatorer.

Følgende skal rapporteres med noen års mellomrom:

Årlige skjevheter dersom dette er mulig å måle.
Den årlige medianen og verdiområdet for skjevheten ved kvartals- og månedens statistikker.

Varians

For de fleste utvalgsplaner som bruker standardmetoder (f.eks. enkelt tilfeldig utvalg eller stratifisert tilfeldig utvalg), og som dessuten bruker standardmetoder for estimering (f. eks. Horvitz-Thompson- eller rate-estimatorer), finnes formler for å beregne variansen og standardavviket til estimatoren. Når et trinn i trekkeprosessen ikke er et sannsynlighetsutvalg, er det mulig å komme nær den riktige variansen dersom visse forutsetninger gjelder. Imidlertid fører dette ofte til en underestimering av variansen, noe som må vurderes. Det trengs en harmonisert metode for å estimere variansen.

Skal rapporteres	Hyppighet		
	Måned- eller kvartalsvis	Årlig	Med noen års mellomrom
Standardavvik for:			
• Total (Nivå for en indeks)	Ja	Ja	Ja
• Vekstrater påfølgende periode	Ja	Ja	
• Vekstrater samme periode for påfølgende år	Ja		
Oversending til Eurostat	Måned eller kvartalsvis	Årlige	Med noen års mellomrom

Estimeringen må ta hensyn til usikkerhet på grunn av imputering knyttet til frafall (enhets- og partielt frafall). Dessuten må en ta hensyn til feil som skyldes feilklassifisering.

Denne informasjonen er avgjørende for å vurdere muligheten for å kunne formidle pålitelig statistikk på Eur15-nivå⁵, selv om statistikken på medlemslandnivå ikke er pålitelig nok. Dessuten kan den brukes ved publisering av konfidensintervall.

2.3.1.2 Ikke-sannsynlighetsutvalg

I visse tilfeller brukes ikke-sannsynlighetsutvalg. Statistikkbyråene kan bruke ekspertutvalg (også kalt skjønnsutvalg) som er basert på høy dekingsgrad for relevante kjennemerker (f.eks. produksjon, sysselsetting, omsetting).

I denne type utvalg er det umulig å oppnå en objektiv vurdering av estimatenes nøyaktighet. Likevel kan grove indikatorer for nøyaktighet konstrueres vha. følsomhetsanalyse. Når enheter er trukket med en sannsynlighet knyttet til en strukturell hjelpevariabel, slik som merverdi, kan en mer avansert

⁵ Eur15 nivå er totaltall for alle de 15 medlemslandene.

metode etableres på grunnlag av en økonometrisk modell som estimerer effekten av de bedrifter som ikke er trukket ut.

Skal rapporteres	Hyppighet
<ul style="list-style-type: none"> • Den årlige dekningsgraden • Indikator(er) på nøyaktighet • Metoder som er brukt for å få de pålitelige indikatorene 	<p>Årlig</p> <p>Månedlig</p> <p>Med noen års mellomrom</p>

2.3.2 Ikke-utvalgsfeil

Under hele prosessen fra planlegging til estimering og publisering oppstår ikke-utvalgsfeil. I dette avsnittet ser vi på de ulike typer av ikke-utvalgsfeil som finnes, og hvorfor de oppstår.

2.3.2.1 Registerfeil

Med registerfeil menes overdekning, underdekning og feilklassifisering. En rask innføring av nye enheter, og stryking av "døde" enheter, fra trekkeregistret er nødvendig for å begrense disse feilene. *Overdekning* betyr enten enheter som er feilaktig medregnet fordi de egentlige ikke hører til populasjonen, eller enheter som faktisk ikke eksisterer.

Skal rapporteres	Hyppighet
<ul style="list-style-type: none"> • Årlige rater for overdekning med hensyn til de viktigste karakteristikken av/kjennemerkene i populasjonen • Metoder brukt for å tallfeste ratene 	<p>Årlig</p> <p>Med noen års mellomrom</p>

Underdekning gjelder (nye) enheter som ikke er inkludert i registret, enten på grunn av nyetablering eller fisjon, eller fordi enheter er klassifisert feil. Underdekning kan også oppstå når medlemslandene bruker avkuttet utvalg. I dette tilfellet bør skjevheten beregnes.

Skal rapporteres	Hyppighet
<ul style="list-style-type: none"> • Årlige rater for underdekning med hensyn til de viktigste karakteristikken av/kjennemerkene i populasjonen • Årlig skjevhet på grunn av avkuttet utvalg • Metoder brukt for å beregne tallene ovenfor 	<p style="text-align: center;">Årlig</p> <p style="text-align: center;">Årlig</p> <p style="text-align: center;">Med noen års mellomrom</p>

Feilklassifisering betyr at enhetene som hører til målpopulasjonen, blir feilaktig plassert i en klasse. Disse feilklassifiseringene kan vise tilbake på økonomiske fenomener eller spesielle tilfeller, og en må vurdere om feilen skal rettes opp eller ikke. Hvilke regler gjelder for å ta hensyn til feilklassifiseringer? Hvordan påvirker disse korreksjonene statistikken? Oppstår feilen når statistikken er revidert?

Skal rapporteres	Hyppighet
<ul style="list-style-type: none"> • Metodisk opplegg for å behandle feilklassifiseringer • Årlig rate for feilklassifiseringer • Metoder brukt for å beregne tallene 	<p style="text-align: center;">Med noen års mellomrom eller når endring oppstår</p> <p style="text-align: center;">Årlig</p> <p style="text-align: center;">Med noen års mellomrom</p>

Feilklassifisering av enheter kan påvirke estimater for en delpopulasjon over flere strata. Den tilsvarende variansen bør ta hensyn til variansen til alle enheter i denne delpopulasjonen, summert over alle de opprinnelige strata.

2.3.2.2 Målefeil

Feil som oppstår ved datainnsamlingen kalles målefeil. Det finnes flere kilder til målefeil:

- “måleinstrumentet” - (spørre)skjemaet
- oppgavegiver
- informasjonssystemet hos oppgavegiver
- type datainnsamling (personlig intervju, telefonintervju, postintervju, dagbok intervju, administrative poster, direkte observasjon og elektronisk observasjon)
- intervjueren.

Ved datainnsamling kan feil forekomme, spesielt når EDI (electronic data interchange) ikke er brukt.

Feil som skyldes rapporteringsenheten

Feil som skyldes rapporteringsenheten, er viktigere for husholdningsundersøkelser enn for bedriftsundersøkelser. Spesialstudier av slike feil skal rapporteres til Eurostat. For eksempel kan det dreie seg om effekter knyttet til erindring av tidligere hendelser, tendenser til å overdrive populære standpunkter (sosialt ønskelige effekter), mangel på oppmerksomhet fra respondentene, samt effekter på grunn av alder og utdanning eller andre personlige kjennemerker (etc).

Statistikere prøver å redusere disse feilkildene ved å forbedre spørreskjema og gi god opplæring av intervjuerne. Feil som skyldes respondenten, er vanskelige å skille fra spørreskjema- og intervjuereffekter. For å måle variasjonen i svar, må respondenten svare på samme spørsmål to ganger, noe som kan gjøres for et delutvalg.

Feil som skyldes spørreskjema

Det er vanskelig å vurdere kvaliteten av spørreskjema på en objektiv måte. Det er nødvendig å bruke et kognitivt laboratorium for å utføre en pilotundersøkelse og så analysere resultatene av undersøkelsen. Hovedformålet med bruk av et kognitivt laboratorium er å gjennomføre en nøyaktig analyse av spørsmålene for å identifisere strategier som innbringer mer presise svar. For eksempel er det kjent at målefeil oppstår når personer spørres om alder. Det er en tendens til å runde av alder oppover. Derfor er det bedre å spørre om fødselsdato.

Feil som skyldes intervjueren

Det er enklere å måle effekten av feil som skyldes intervjueren, siden utvalget av intervjuere kan gjøres tilfeldig. En spesiell utvalgsplan kan konstrueres slik at en kan måle variasjonen i svar med hensyn på intervjuerne.

Feil som skyldes måling av en variabel/størrelse

Målefeil kan f.eks. skyldes at, referanseperioden for statistikken ikke er den samme som den perioden dataene er innsamlet for (kalenderår er ikke alltid det samme som regnskapsår), eller at en ved måling av ren prisutvikling ikke tar hensyn til kvalitetsforandringer.

Spesielle teknikker kan brukes for å vurdere målefeil (gjentatt intervju, estimering av intervjuereffekten osv.). Men kostnaden for dette er ofte uoverkommelig høy, både for oppgavegiveren og de som samler inn data.

Det er vanlig at de nasjonale statistikkbyråene begrenser sine undersøkelser til å sjekke samsvar mellom kjennemerker i spørreskjema (eller mellom kjennemerker i spørreskjema og annen informasjon av høy kvalitet).

Revisjonsarbeidet beskrives gjennom en oppsplitting i makrorevisjon / mikrorevisjon, der den sistnevnte deles inn i frafall, feil og avvik.

Det rapporteres med noen års mellomrom:

- Metoder som brukes for å redusere ulike typer av målefeil
- Varians og skjevhet som skyldes intervjueren
- Feilrater

2.3.2.3 Bearbeidingsfeil

Feil som oppstår i datainnsamlingsprosessen ved koding, inntastning, redigering og ved framstilling av tabeller, kalles bearbeidingsfeil.

For koding kan noen mål for feilraten finnes gjennom standardteknikker for kvalitetskontroll, f.eks. gjennom kvalitetsundersøkelser på et delutvalg fra det bearbeidede spørreskjemaet (for å sjekke nivået på feil enten under inntastingsfasen eller under revisjonsarbeidet).

Når det gjelder revisjon, vekting, estimering og tabullering, er feilen vanskeligere å måle.

Skal rapporteres	Hyppighet
<ul style="list-style-type: none">• Rate for bearbeidingsfeil, brutt ned på de enkelte bearbeidingstrinn• Metoder brukt for estimering av ratene	Med noen års mellomrom
	Med noen års mellomrom

2.3.2.4 Frafallsfeil

Enhetsfracfall og partielt frafall er ofte en viktig kilde til feil i undersøkelser. De kan føre til skjevheter i estimatene på en ukontrollerbar måte. En tradisjonell og enkel indikator for frafall er svarrate (svarprosent), som kan lages på individnivå eller på spørsmålnivå. Denne raten kan enten vektas (med utvalgsvekter eller andre variabler som f.eks. antall sysselsatte eller omsetning), eller ikke vektas. Skillet mellom det faktiske frafallet, og individer som ikke hører til i målpopulasjonen, er meget viktig. Det anbefales at en skiller ut individer der en ikke vet om manglende observasjoner skyldes frafall eller registerfeil.

Et viktig spørsmål er hvordan individer som har svart, skiller seg fra dem som ikke har svart. En årsak til frafall er forsinkelse i levering av data fra respondenten. Effekten av denne type frafall kan måles ved sammenligning a posteriori. Andre typer frafall kan granskes gjennom en tilleggsundersøkelse av delutvalg av individer med frafall, der en bruker mer effektive datainnsamlingsmetoder (f.eks. intervju istedet for postalt spørreskjema).

Både enhets- og partielt frafall må bearbeides under produksjonsprosessen, enten gjennom imputering (delvis eller total) eller gjennom korrigerede vekter for de som svarer. Denne behandlingen fører til mer usikkerhet (tilfeldighet), noe som må integreres i estimatene av skjevhet og varians.

Skal rapporteres	Hyppighet		
	Måned og kvartalsvis	Årlige	Med noen års mellomrom
<ul style="list-style-type: none">• Svarrater på individnivå og spørsmålnivå• Beskrivelse av de metoder som er brukt for imputering og/eller korrigerede vekter for frafall• Varians på grunn av imputering• Skjevhet på grunn av frafall	Ja	Ja	Ja
	Ja	Ja	Ja
<ul style="list-style-type: none">• Metoder brukt i estimering i punktene ovenfor	Etter hyppighet i overføring for estimatet		

2.3.2.5 Modellfeil

Det kan skje at en feilaktig eller mindre god modell velges, og det fører til modellfeil. Eksempler på metoder det velges blant, for å øke modellnøyaktigheten er; kalibrering, generalisert regressionsestimater, beregninger basert på hele populasjonen eller konstant populasjon, avstemming og

sesongjustering. I tillegg til dette finnes det modeller som ikke spesielt brukes for å øke nøyaktigheten, men for å få fram totaltall og indekser.

Kvalitet på modellene skal vurderes med hensyn til varians og skjevhet. Denne vurderingen burde integreres i de konfidensintervall som hører sammen med statistikken, men det finnes ikke noen standardmetoder for en slik integrering.

Sammenligning

Et viktig punkt i forbindelse med kvalitet knyttet til sesongjusteringsmodeller, er å finne fram til og korrigere for ekstreme verdier før en sesongjusteringer. Avvikende verdier, som skyldes en uvanlig hendelse (f.eks. streik), for aggregerte data, må korrigeres før en kan estimere sesongkomponentene.

Avstemming av kvartalsvise data mot årlige data

Avstemming brukes vanligvis for å revidere tall for kortere tidsperioder (f.eks. kvartal) slik at de stemmer overens med tall fra årsstatistikken. Denne metoden skal beskrives når den brukes. Videre skal en sammenligning av utgangsverdien og den endelige verdien rapporteres med en vurdering av hvordan metoden virker. Konsekvensen for nøyaktigheten til de sammenlignede seriene skal også evalueres (reduksjon av skjevhet, økning av varians).

Andre metoder kan brukes for å estimere deler av en statistikk når det mangler data for en delpopulasjon, når aggregerte administrative data er brukt eller når en nødvendig del av dataene ennå ikke er tilgjengelig (f.eks. vektorer til beregninger av indekser).

Valg av modell skal rapporteres til Eurostat (hvorfor en modell er foretrukket framfor en annen), gjerne sammen med tilhørende estimaters estimeringsfeil.

Rapporteringskriterier, på flerårig basis, for hver modell som er brukt til å justere seriene :

- Bekreftelser/verifisering av forutsetningene for den underliggende modellen
- Test av modellens prediksjonsevne gjennom bruk av 'historiske' data
- Sammenligning mellom resultater produsert av modellen og data fra beslektede kilder
- Bruk av kontrollundersøkelser og kryssvalidering
- Følsomhetstester av modellen med hensyn på de estimerte parametre
- Gyldigheten av data som er brukt i modellen.

2.4 Planlegging og punktlighet

For å analysere virkningen av hvert trinn i produksjonsprosessen, og for å forenkle diskusjonen vedrørende forsinkelser med medlemslandene eller utvikle avanserte estimater på Europa-nivå, er følgende informasjon nyttig:

- bruk av elektronisk dataoverføring for statistiske data innsamlet i medlemslandet
- lovpålagt svarfrist for respondenten i medlemslandet
- innsamlingsdato for nøkkeldata, der datoer for utsending av spørreskjema og oppfølgingsundersøkelser/purringer spesifiseres, sammen med tidspunkt for feltarbeidet
- en graf over svar med hensyn på antall spørreskjema som innkommet per uke og per arbeidsdag
- startdato og sluttdato for revisjonsfasen
- dato for imputeringsfasen
- detaljer og dato for når datainnsamlingsfasen kan regnes som ferdig
- dato for beregning og spredning av foreløpige resultater
- dato for kvalitetskontroll av samsvarende resultater og kontroll av mål som ikke røper noe
- dato for når publikasjonen er tilgjengelig
- dato for publisering.

Evalueringsav punktlighet vedrørende dataoverføring til Eurostat vil foregå etter forordningens moduler som spesifiserer periodisitet og forsinkelser ved dataoverføring.

2.5 Tilgjengelighet og klarhet

Eurostat er interessert i følgende informasjon om spredningsprosessen:

- en kopi av publikasjonen / publikasjonene
- informasjon om hva som er sendt, hvis det er gjort, til enheter som har deltatt i undersøkelsen
- informasjon om spredning av resultatene (f.eks. til hvem resultatene er sendt og hvem brukerne er)
- Kopi av metodeartikler som er brukt som grunnlag for statistikkproduksjonen.

Denne informasjon er nyttig til planlegging av en spredningsplan på Eurostat-nivå, for å kunne gi bedre svar på henvendelser fra medlemslandene og for å få med nasjonale kommentarer til publikasjonene som kan være til hjelp for Eurostat når de tolker statistikkene. Videre kan informasjonen være til hjelp når nytten av statistikktilgjengelighet skal framheves.

2.6 Sammenlignbarhet

Sammenlignbarhet som fungerer både i rom og tid kan gi et bilde av utviklingen både i medlemslandet og innenfor EU. Her er det viktig med god dokumentasjon for å vurdere sammenlignbarheten.

2.6.1 Romlig sammenligning

Dersom det er en forskjell mellom nasjonale begreper og europeiske begreper (definisjoner av statistiske enheter, referansepopulasjon, klassifisering, definisjoner av observerte kjennemerker), skal disse forskjeller nevnes⁶ både med hensyn til forskjell i begreper og med estimat for konsekvensene for statistikken.

På lignende måte skal medlemslandene rapportere om de følger dokumentet 'European Recommendations for Measurement', og om mulig forklare avviket fra normene i dokumentet og vurdere konsekvensene for estimatene.

Den ovenfornevnte informasjonen gjør det mulig å vurdere hvor sammenlignbare statistikkene er. Noen frivillige tilleggsundersøkelser på små delutvalg burde vise avviket fra de europeiske standardene.

En sammenligning mellom medlemsstatenes praksis og Eurostats anbefalinger skal rapporteres når betydelige endringer skjer, ellers skjer rapporteringen med noen års mellomrom. Om mulig skal også forskjellene i praksis vurderes.

2.6.2 Sammenligning over tid

Noen brukere, spesielt de som arbeider innenfor det makroøkonomiske området, benytter seg av tidsserier istedet for punktestimater for en dato. Stabilitet i begreper og metoder er viktig for disse brukerne. Detaljer om forandringer i definisjoner, dekningsgrad og metoder skal dokumenteres, likeledes hvordan fusjoner og fisjoner påvirker behandlingen av statistiske enheter. Fusjoner og fisjoner av enheter (foretak, husholdninger) har mange ganger stor betydning for visse typer av statistikk, særskilt når den brytes ned i henhold til spesielle klassifiseringer. Når sammenligningen av statistikk fra to etterfølgende perioder er mer påvirket av en lovmessige endringer enn av den sosio-økonomiske trenden, er det viktig å få med dette aspektet. All endring av nasjonale lovgivningen som er av betydning for sammenligningen, skal rapporteres.

⁶ Når nasjonal statistikk er utarbeidet ikke bare på bakgrunn av en utvalsundersøkelse, men også vha. en modell som skal justere for manglende samsvar mellom statistikken fra undersøkelsen og det mål medlemslandet hadde med undersøkelsen, skal justeringen som skyldes modellvalg inngå i punkt 2.3.2.5 'Modellfeil'. Kun forskjellen mellom europeiske normer og normer som oppnås etter den potensielle nasjonale modelltilpassningen skal rapporteres under dette punktet. For mer informasjon, se dokument om 'Generell sammenlignbarhet'.

For sesongjusterte serier, skal sammenligningen ta hensyn til forandringer i justeringsprosedyrer, men den metoden som brukes for sesongjustering, skal rapporteres under punkt 2.3.2.5 'Modellfeil'.

Når signifikante endringer fører til brudd i tidsserier, skal medlemslandene rapportere disse sammen med statistikken. Estimater for disse endringene overføres til Eurostat ved et seinere tidspunkt.

2.7 Sammenheng

For å få et materiale som kan sammenlignes på en god måte, må en kjenne sammenhengen mellom statistikker og begreper.

2.7.1 Sammenheng mellom foreløpig og endelig statistikk

Sammenhengen mellom foreløpige og endelige indekser skal maksimeres på en slik måte at foreløpige tall er tilstrekkelig nøyaktige.

Den prosentvise årlige gjennomsnittsforskjellen mellom den ikke sesongsjusterte endelige statistikken/indeksen og den foreløpige statistikken/indeksen skal leveres sammen med all statistikk, det skal også variasjonen rundt gjennomsnittet og medianen.

2.7.2 Sammenheng mellom årlig og korttids statistikk

For flere kjennemerker produseres statistikk både i løpet av året og som årsstatistikk. Det er viktig for statistikkbyråenes omdømme hos brukerne at det er en viss sammenheng i informasjonen som formidles. En enkel metode for å vurdere dette, er å sammenligne estimater for de samme kjennemerkene fra statistikk med ulik hyppighet. Det virker mest naturlig å sammenligne årlige gjennomsnitt eller totaler dersom statistikkene av ulik hyppighet gir estimater av nivå-typen, og sammenligne rater når minst en av statistikkene er en indeks. Sammenligningen skal gjøres for en eventuell avstemming utført av det nasjonale statistikkbyrået. Videre må medlemslandene, hvis de har utført avstemming (laget sammenligningsgrunnlag) for sine korttids statistikker som er basert på den årlige statistikken, rapportere også dette spesielt.

For hver sektor skal medlemsstatene sammenligne følgende vekstrater:

- beregnet fra korttids undersøkelser (*GRst Growth rate short term*)
- beregnet fra årlige undersøkelser (*GRyv Growth rate year value*)
- beregnet fra årlige undersøkelser, med eliminering av effekter for fødsler, dødsfall og endringer i klassifisering av enheter⁷. (*GRyc Growth rate year changes*)

Praksis som avviker fra Eurostats normer og anbefalinger, og som har innvirkning på utviklingen i den produserte statistikken slik som basisperiode, type av indeks (Laspeyres; Paasche, Fischer) - skal også tas med i betraktningen.

2.7.3 Sammenheng mellom statistikk fra samme sosioøkonomiske område

Ofte representerer en gruppe statistikker, muligens av ulike typer (i pengeverdi, i volumpris eller som konstant pris, prisindikatorer), samme fenomen, men utfra ulike vinklinger. Det er meget viktig å kontrollere at statistikkene er sammenlignbare på en måte som tar hensyn til brukernes spørsmål, og som legger til rette for korreksjoner.

For eksempel må korttidsstatistikk for bedrifter, som omsetning, merverdi og variasjoner i aksjer, på årsbasis kunne sammenlignes med produksjonsstatistikk basert på deflaterte produksjonspriser.

Statistikk for befolkning, sysselsetting og arbeidsledighet skal også kunne sammenlignes. På liknende vis bør meningsmålinger og kvantitative indikatorer kunne sammenlignes. Sysselsetting som er målt som sosialstatistikk (f.eks. folketellinger), eller som bedriftsstatistikk, skal ideelt sett vise en høy grad av samsvar. Undersøkelser av personers ferievaner skal i stor grad være forenlige med turiststatistikken, eller korttids indikatorer for servicesektoren osv.

2.7.4 Sammenheng mellom statistikk for nasjonalregnskap

For å kunne råde brukere mht. hvilke informasjonskilder som passer deres behov best, kan det være interessant å kunne sammenligne statistikk fra undersøkelser med nasjonalregnskapet. Et totalmål for sammenhengen mellom disse undersøkelsene og andre informasjonskilder vil nyttig for brukerne. Metodene som brukes i nasjonalregnskapet må derfor beskrives i den aktuelle statistikken, inklusive primærdatakilder og justeringer som er brukt. Uoverensstemmelser i begreper bør også behandles.

⁷ Mer eksakt tilsvarende dette målpopulasjonen definert som:

- enheter klassifisert etter sin hovedaktivitet i år n
- populasjonen utgjør den felles delmengden av populasjonen i år n og populasjonen i år n+1.

Utvalget må etterstratifiseres slik at det kan tas hensyn til de ovenfor nevnte restriksjoner. I de tilfeller der delvis rotert utvalg brukes, består et pålitelig estimat av den felles delmengden for utvalget i år n og utvalget år n+1, forutsatt at utvalget ikke er helt skiftet ut i de tilsvarende strata.

En summering av de i avsnitt 2.7 nevnte sammenhengene skal rapporteres til Eurostat hvert år.

2.8 Kostnad og budsjett

Kvalitet henger nøye sammen med kostnader. Medlemsstatene skal oppgi sine kostnader da disse inngår som en viktig del i vurderingen av kvalitet.

2.8.1 Kostnader for respondenten

Medlemslandene skal årlig sende Eurostat oversikt over oppgavegivers gjennomsnittskostnader. Den totale kostnaden for populasjonen, beregnet som gjennomsnittskostnaden per bedrift multiplisert med utvalgsstørrelsen, skal også leveres til Eurostat. Foruten gjennomsnittskostnader for respondenten, skal det angis om det er brukt samlede eller roterte utvalg i undersøkelsen, dette for å måle spredningen av oppgavebyrden mellom de enkelte enhetene.

2.8.2 Kostnader og budsjett for statistikkbyrået

Medlemslandene skal oversende Eurostat de totale kostnadene for undersøkelsen. Med flere år mellom skal det informeres om antall sysselsatte (statistikere, metodestatistikere, programmerere og analytikere, intervjuere og kontorpersonale). Detaljert informasjon om datautstyr (hardware og software) og annet verktøy (f.eks. utstyr til optisk lesning) må også beskrives.

3. Ordliste for kvalitetsbegrep i statistikk

(Assesment of the quality in statistics, Item 5.2 of the agenda, Glossary on quality in statistics, Eurostat.)

Denne ordlisten begrenser seg til begreper som er nødvendige for å unngå uklarheter ved lesing av de to andre notatene i denne serien, 'Definisjon av kvalitet i statistikk' og 'Forslag til standard for kvalitetsrapporting', og andre forslag til kvalitetsrapportering på spesifikke områder. Ordlisten er nødvendig for å skape et felles grunnlag for analyse av kvalitet i statistikk.

Disse definisjonene er hentet fra et liknende arbeid utført av 'the US Department of Commerce (1978) og av Lessler og Kalsbeek (1992). Disse dokumentene gir en oversikt og analyse av tilgjengelig litteratur om feilproblemer innenfor statistikkundersøkelser.

Definisjoner som er relatert til utvalgsundersøkelser, er tilpasset Särndal et al. (1992).

Aktuell periode

Med dette menes den periode som statistikken beskriver.

Aktuell prisindeks eller verdiindeks

Denne indeksen brukes til å måle den gradvise endringen i aktuelle priser ved å nytte konstante vekter.

Avkuttet utvalg

Trekksansynligheten er lik null under en bestemt størrelse.

Basisår

Dette refererer seg til det året da vektene for varer er fastsatt. Vektene er et resultat av forbrukernes forbruk prioritert dette bestemte året som er kartlagt gjennom en forbruksundersøkelse. Det benyttes for å beregne en fast basisindeks (i motsetning til kjedeindeks).

Bearbeidingsfeil

Disse feilene er knyttet til de innsamlede data og er et resultat av mangelfull gjennomføring av en riktig planlagt undersøkelsesmetode. Bearbeidingsfeil omfatter alle etterinnsamlings-operasjoner, og inkluderer også trykking av spørreskjema. De fleste bearbeidingsfeil oppstår i data for individuelle enheter, selv om de også kan forekomme i tabeller og estimer.

Død enhet

En enhet som opphører å eksistere i den aktuelle perioden, kalles en død enhet. For bedrifter gjelder dette når virksomheten er opphørt, og for personer gjelder dette ved dødsfall.

Elektronisk datautveksling, EDI

Dette er den elektroniske overføringen av informasjon om forretningstransaksjoner mellom forretningsinteressenter, som gjøres på et standardformat. EDI gir bedriftene mulighet til å gjenfinne informasjon fra egne systemer og sende denne informasjonen til handelspartnere, leverandører, kunder og statlige etater gjennom et elektronisk kommunikasjonsnettverk.

Enheter som ikke tilhører målpopulasjonen

Med dette menes enheter som ikke skal inngå i trekkefilen fordi de ikke tilhører målpopulasjonen i den aktuelle perioden. Enheter som er registrert som dødsfall, fusjoner og fisjoner hører til målpopulasjonen når de er registrert i den aktuelle perioden.

Enhetsfracfall (totalfracfall)

Se fracfall.

Erindringsfeil

Skylde at respondenten kan ha glemt hendelser, blandet sammen datoer osv.

Estimat

Den beregnede verdien basert på en estimator.

Estimator

Den matematiske funksjonen som brukes for å estimere en verdi for en spesiell parameter.

Fastprisindeks eller volumindeks

Denne indeksen måler den gradvise utviklingen i mengde ved at ta hensyn til konstante priser.

Feil

Forskjellen mellom en variabels verdi estimert i en undersøkelse og den sanne verdien.

Feilklassifisering

Feilklassifisering skjer når enheter fra målpopulasjonen blir plassert i feil klasse (f.eks. næring, geografisk område eller størrelse).

Forventet verdi

Det hypotetiske gjennomsnitt over alle tenkte utvalg fra en undersøkelse, der alle utvalg gjøres under de samme betingelser.

Frafall

I en undersøkelse vil som regel noen enheter ikke svare, og dette kalles frafall. Frafall er også forårsaket av dødsfall, at enheter befinner seg utenfor rammen av undersøkelsen og at enhetene ikke kan kontaktes. Det finnes to typer frafall. Den første typen oppstår når en trukket enhet som er kontaktet, ikke svarer, og dette kalles enhetsfracfall eller totalfracfall. Den andre er når respondenten unnlater å svare på et eller flere spørsmål. Dette siste kalles partielt frafall eller svarfracfall.

Frafallsfeil

Frafallsfeil oppstår når manglende data må imputeres på grunn av frafallet.

Frafallsskjevhet

Dette er en skjevhet som skyldes at de som svarer ikke er representative for dem som ikke svarer.

Fullstendig populasjon

Refererer til hele populasjonen. Den skal inneholde fødsler, dødsfall, fusjoner og fisjoner for den periode som f.eks. indekser er beregnet for. Det motsatte er konstant populasjon.

Hedonisk metode

Den hedoniske tilnærmelsen antar at en vare kan bli sett på som en samling med fysiske karakteristikk, at kvalitet er av disse karakteristikkene, at forandring av markedets pris på en vare kan deles opp i en priskomponent for kvalitet og en ren priskomponent, og dessuten at prisen på et produkt er korrelert med varens karakteristikk.

Ikke-utvalgsfeil

Dette er feil som ikke er knyttet til selve utvalget. Ikke-utvalgsfeil oppstår for alle typer undersøkelser ved planlegging, gjennomføring, databearbeiding og endelig estimering. F.eks. er feil i register en ikke-utvalgsfeil.

Imputering

Imputering er å erstatte en manglende observasjon gjennom en logisk redigering eller ved bruk av statistiske metoder. Det brukes både om enhetsfrfall (totalfrfall), f.eks. når en person nekter å delta og ved partielt frfall (svarfrfall), dvs. når det er frfall på ett eller flere spørsmål, men ikke på alle

Inndeling i statistiske enheter

Inndelingen i statistiske enheter er i hovedsak knyttet til en inndeling av juridiske enheter som gjelder EUs harmoniserte regler og/eller ved bruk av klassifiseringer. En dårlig inndeling kan oppstå når to enheter er tenkt som en og omvendt (for eksempel en kafé som finnes i et kjøpesenter). Et annen type enheter som er vanskelig å dele inn i statistiske enheter er meget små bedrifter, der den totale mengden av tjenester/arbeid er mindre enn et årsverk.

Interessepopulasjon

Punktestimater er utledet på grunnlag av den del av populasjonen som er av interesse. En interessepopulasjon kan være hele populasjonen eller en spesifikk delpopulasjon. Interessepopulasjonen kan følge utvalgsstrata helt eller gå på tvers av flere strata.

Intervjuerfeil

Feil som skyldes intervjueren (f.eks. når spørsmål stilles eller ved rapportering av svar).

Kjedeindeks

Kjedeindeks beregnes gjennom å sammenligne inneværende periode med en nærliggende periode (som gjerne holdes fast et år om gangen). Dessuten er varevektene for kjedeindeksen oppdatert for denne perioden, i motsetning til faste basisindekser, se basisår.

I praksis er det ofte en tidsforskjell mellom den nåværende perioden og den perioden som vektene er beregnet for. Vektene er mange ganger avhengige av resultater fra den siste strukturundersøkelsen, og de beregnes på årsbasis. Tidsforskjellen kan reduseres ved bruk av en predikativ modell ikke brukes for å framskrive vektene.

Kodefeil

Feil koding av svaret på et spørsmål.

Koding

Proessen å gjøre om muntlig og/eller skriftlig informasjon til tall eller andre symboler for lett å kunne framstille tabeller.

Konsistent

Et estimat er konsistent når sannsynligheten for at feilen er større enn gitt størrelse, går mot null når utvalget øker.

Konstant populasjon

Denne populasjonen refererer seg til grunnlaget for å trekke et utvalg. Det inneholder ikke fødsler, dødsfall, fusjoner og fisjoner. Populasjonen skal være identisk for begge perioder der indeks er estimert. Konstant populasjon er det motsatte av fullstendig populasjon.

Kontroll

En studie av data der de individuelle enhetene fra en undersøkelse kontrolleres mot datafelter fra en annen undersøkelse (f.eks. sammenligning av alder som er rapportert i en folketelling med alder som er rapportert fra fødselsregistre/attester).

Lagringsfeil

Feil som oppstår på grunn av unøyaktighet i lagringen av svar.

Laspeyres indeks

En Laspeyres indeks for pris bruker mengde som vekt og indeks for mengde bruker pris som vekt. Vektene er beregnet for basistidspunktet (se basisår), i motsetning til Paasche indeks, der vektene er beregnet for observasjonstidpunktet.

Makroredigering

Dette er en prosedyre som peker på mistenkelige data for alle kjennemerker eller deler av dem. Prosedyren gjennomføres ved kontroll av aggregater eller gjennom bruk av statistiske metoder.

Middelkvadratfeil

Middelkvadratfeil beregnes som variansen til estimatet pluss kvadratet av skjevheten.

Mikroredigering

En meget grundig kontroll der hver enkelt observasjon blir undersøkt for å finne feil.

Målefeil

Målefeien er forskjellen mellom den innsamlede verdien fra undersøkelsen og den sanne (individuelle) verdien. Disse feilene kan være et resultat av:

- respondenten har ikke gitt et sant svar (respondentfeil),
- intervjueren har ikke registrert svaret korrekt (intervjuerfeil), eller
- det er feil i måleinstrumentet (f.eks. i spørsmålstillingen).

Målpopulasjon

Dette er den del av populasjonen som omfatter de elementer som skal analyseres. På grunnlag av denne skal undersøkelsen produsere statistikken.

Nekterprosent

Andel av oppgavegivere (bedrifter eller personer) som, til tross for kontakt, vegrer å gi den ønskede informasjon. Prosenten beregnes gjennom å dividere antallet nektinger i forhold til den opprinnelige utvalgsstørrelsen.

Overdekning

De enheter som er med i utvalget, men som ikke inngår i målpopulasjonen, kalles overdekning. Disse oppdages vanligvis når enheten kontaktes, men dette skjer ikke alltid fordi noen enheter ikke blir kontaktet og noen utelukkes fra utvalget. Grunnen til overdekning er at trekkeregistret ikke er oppdatert. Overdekning fører som regel til skjevheter i estimatene.

Paasche indeks

En Paasche indeks for pris bruker mengde som vekt, og indeks for mengde bruker pris som vekt. Vektene er observert i den aktuelle perioden / tidspunktet, noe som er det motsatte av Laspeyres indeks.

Partielt frafall (svarfracfall)

Se *fracfall*.

Purring / oppfølging

En prosedyre (purring) der de enhetene i et utvalg som ikke har svart i undersøkelsen (f.eks. telefonintervju, personintervju eller postalintervju) blir kontaktet igjen med samme eller en annen metode for å øke svarprosenten. Uttrykket (oppfølging) kan også brukes for å betegne gjentatte undersøkelser basert på et panel av respondenter.

Referanseperiode

I sammenheng med indeks:

Viser til den periode der en indeksserie har verdien 100.0. Referanseperiode og basisperiode er identiske for fast basisindeks, men skiller seg for kjedeindeks.

I andre sammenhenger:

Referanseperiode er den periode som statistikken rapporteres for. Kalles også aktuell periode for indekser.

Register

Dette skal være en oppdatert fil hvorfra alle statistiske enheter til trekkingfiler velges ut. Register er som regel et uttak fra administrative data.

Relativt standardavvik

Se *variasjonskoeffisient*.

Revekting

Revekting er en justeringsprosedyre som brukes for å korrigere for frafall. Den kan også brukes for å øke presisjonen ved bruk av hjelpeinformasjon. Standardmetoder inkluderer etter-stratifisering, kalibrering og svartendensmodellering.

Revisjon

Revisjon betyr feilsøking og oppretting av feil som oppstått i datainnsamlingen. Kontroller for å identifisere manglende, ukorrekte eller mistenkelige verdier gjennom datastyrt revisjon, kalles revisjonsregler. En feiloppretting skjer når en verdi endres som følge av oppdaget feil.

Samordning av utvalg

Å øke overlappingen mellom utvalgene i flere undersøkelser, istedet for å trekke uavhengige utvalg, kalles samordning av utvalgene. En positiv samordning søkes ofte i undersøkelser som gjentas over tid (paneler) for å øke nøyaktigheten på korrelerte variabler fra to undersøkelser. En reduksjon av overlappingen mellom utvalg i ulike undersøkelser kalles negativ samordning, og brukes når en ønsker en mer lik oppgavebyrde for individene (personer eller bedrifter) i populasjonen.

Sann verdi

Er den faktiske verdien som vi ville finne dersom vi hadde perfekt måleutstyr, og hvis ingen feil ble gjort verken ved innsamling av primær data eller ved matematiske beregninger.

Skjevhet

Forskjellen mellom den forventede verdien på et estimat og den sanne verdien på estimatet kalles for skjevhet.

Standardavvik

Beregnes som kvadratroten av variansen til estimatet.

Svarprosent

For en intervjuundersøkelse er telleren antallet oppnådde intervju og nevneren den totale utvalgsstørrelsen minus de som skal utelukkes (f.eks. de som ikke oppfyller kravet for å delta i en viss studie). Vektet svarprosent er ofte mer nyttig. Svarprosenten i en undersøkelse kalles også svarraten.

I tillegg til dette kan en *frafallsprosent* brukes der følgende vektor beregnes: utvalgsvekten multiplisert med en viktig variabel som f.eks. omsetning eller sysselsetning. Denne *frafallsprosenten* viser virkningen av frafallet.

Svarprosent for enkelte spørsmål

Denne rate beregnes som forholdet mellom antall svar og antall svar pluss antall frafall. Vektet svarprosent er også et brukbart begrep.

Trekkeregister

Trekkeregistret for en undersøkelse er den listen/fortegnelsen/registeret eller de listene med enheter som avgrensner, identifiserer og gir adgang til målpopulasjonen.

(Trekke)registerfeil

Dette kan skyldes mangler i input-data, eller forsinkelser og feil i datainnsamlingen og bearbeidingen.

Trekkeregisterfeil omfatter:

- overdekning, underdekning, duplisering
- klassifikasjonsfeil - enheter som ikke er klassifisert, feilklassifisering med hensyn til f.eks. næring, geografi eller størrelse
- kontaktadressefeil - enheter som er ufullstendige eller unøyaktige sånn at kontakt blir umulig.

Underdekning

Enheter som hører til målpopulasjonen, men som ikke er med i trekkeregistret gir underdekning. Dette gjelder f.eks. (nye) bedrifter som ikke inngår i trekkeregistret, enten fordi de er nyetablerte eller har gjennomgått en fisjon, eller fordi de er feilklassifisert. Dette leder for det meste til systematiske feil i estimatene.

Undersøkelsesplan

Undersøkelsesplanen tar med følgende ting:

- statistikker som er produsert
- populasjon
- tidspunkt for undersøkelsen
- nøyaktighet
- hvilke data som er samlet inn for spesifikke enheter i den gitte populasjonen
- og hvilke metoder som er brukt ved innsamling og bearbeiding av data (for å produsere ønskede statistikker).

Dokumentasjon av operasjonelle, organisatoriske og administrative spørsmål er vanligvis også med.

Utvalgsfeil

Disse feilene skyldes at kun en del av målpopulasjonen brukes i undersøkelsen, istedet for en totaltelling av hele målpopulasjonen.

Utvalgsplan

Utvalgsplanen skal inneholde informasjon om målpopulasjonen og den endelige utvalgsstørrelsen, definisjoner av strata og metode som er brukt for trekking av utvalget. Den skal også inkludere detaljer om forandringer i metode, og dessuten de eventuelle innvirkninger dette har for sammenligning over tid.

Varians

Variansen for en tilfeldig variabel viser spredningen omkring gjennomsnittet for den empiriske verdien. Dette er det gjennomsnittlige kvadratavviket fra gjennomsnittsverdien for variabelen.

Variasjonskoeffisient

Dette er et estimat som måler den relative spredningen, og som er lik forholdet mellom standardavviket og gjennomsnittet. I praksis er dette det estimerte standardavviket for et punkt estimat dividert med punkt estimatet for gjennomsnittet, målt i prosent.

Verdiindeks

Se aktuell prisindeks.

Volumindeks

Se fastprisindeks.



Vedlegg A

Doc. Eurostat/A4/Quality/98/General/Definition - ex97/02

Orig. : EN

Available in **EN**

RD/mv

**ASSESSMENT OF THE QUALITY
IN STATISTICS**

**Luxembourg - 29 May 1998
Jean Monnet Building - Room M1
Starting at 9.30 a.m.**

Item 5.1. of the agenda

Definition of quality in statistics

Definition of quality in statistics

Introduction

This paper lists the major subjects that should be included in a report on quality. The details to be included in the report are contained in document 'Proposal for a quality report'

1. DEFINITION OF QUALITY

Quality is defined in the ISO 8402 - 1986 as: "the totality of features and characteristics of a product or service that bear on its ability to satisfy stated or implied needs". This could be analysed in the framework of Total Quality Management. We will adopt a more restricted approach, concentrating initially on the "products" rather than on the process as a whole. Two exceptions will be made: relevance is a valuation and timeliness is related to dates, which are process information.

Quality of statistics can be defined with reference to several criteria:

- relevance of statistical concepts;
- accuracy of estimates;
- timeliness and punctuality in disseminating results;
- accessibility and clarity of the information;
- comparability of statistics;
- coherence.

Although not a measure of quality, the resources available for the production of statistics act as a constraint on quality. When assessing the ability of a Member State to comply with quality guidelines, it is necessary to take account of the resources available.

Relevance

A survey is relevant if it meets users' needs. The identification of users and their expectations is therefore necessary.

Accuracy of estimates

Accuracy is defined as the closeness between the estimated value and the (unknown) true population value. Assessing the accuracy of an estimate involves analysing the total error associated with the estimate.

Timeliness and punctuality in disseminating results

Most users want up-to-date figures, which are published frequently and on time at pre-established dates.

Accessibility and clarity of the information

Statistical data have most value when they are easily accessible by users, are available in the forms users desire and are adequately documented. Assistance in using and interpreting the statistics should also be forthcoming from the providers.

Comparability of statistics

Statistics for a given characteristic have the greatest usefulness when they enable reliable comparisons of values taken by the characteristic across space and over time. The comparability component stresses the comparison of same statistics between countries in order to evaluate the meaning of aggregated statistics at the European level.

Coherence

When originating from a single source, statistics are coherent in that elementary concepts can be combined reliably in more complex ways. When originating from different sources, and in particular from statistical surveys of different frequencies, statistics are coherent insofar as they are based on common definitions, classifications and methodological standards. The messages that statistics convey to users will then clearly relate to each other, or at least will not contradict each other. The coherence between statistics is orientated towards the comparison of different statistics, which are generally produced in different way and for different primary uses.

About Cost

There is clearly a strong link between the quality of statistics and the resources available to produce them. An assessment of costs should be kept in mind during the quality evaluation process.

2. ACCURACY

This aspect of quality has been extensively studied in many statistical agencies and by academics. Accuracy focuses on an analysis of errors, divided into sampling and non-sampling errors.

Errors can be classified as follows:

- 2.1. Sampling errors
- 2.2. Non-sampling errors
 - 2.2.1 Frame errors
 - 2.2.2 Measurement errors
 - 2.2.3 Processing errors
 - 2.2.4 Non-response errors
 - 2.2.5 Model assumption errors

Groves (1989), Lessler and Kalsbeek (1992), Särndal, Swensson and Wretman (1992), Biemer and Fesco (1995), Quality Measurement Model of the BLS (1995) use similar classifications.

Until now, statisticians have focused mainly on sampling errors. Estimates of accuracy published by statistical agencies usually cover only sampling errors¹. However, research into

¹ In fact, the estimated variance depends in practice on response rates and certain non-sampling errors - e.g. random measurement errors - but it does not account for all non-sampling errors. In particular, most systematic errors are not included in this measure of accuracy.

non-sampling error is developing rapidly, and some methods are now available for a first assessment of such errors. For instance, see Biemer et al. (1991).

3. TIMELINESS AND PUNCTUALITY

Users generally require that statistical information takes a minimum amount of time to produce, is released as soon as it is available, and where appropriate is available on a regular basis. Keeping production times to a minimum implies efficient production techniques, often at great expense. The whole process of data collection, editing, imputation, estimation and dissemination has to be kept under control in order to minimise the processing period. A partial solution to the problem is to provide early estimates, based on a sub-sample of respondents. Care needs to be taken over the presentation of these early estimates, and also with the subsequent dissemination of the revised information.

There is sometimes competition between (often private) statistical information, produced and disseminated very rapidly but with less reliability, and slower, but more accurate, statistics from government agencies. The trade-off between timeliness and accuracy has some very visible consequences for certain users. This challenges statistical agencies to improve the timeliness of their statistics with no reduction in the high standards of quality.

4. CLARITY AND ACCESSIBILITY OF STATISTICS

Dissemination is a vital step in the information chain. It is not sufficient to have "good statistics" stored somewhere inside the statistical office. They have to be made available to all potential users, in an appropriate form. Firstly, users should be in a position to know easily which kind of statistics is available. Secondly, physical access to the statistics should be convenient. Thirdly, the statistics should be accompanied by the necessary information on concepts and methods. Different levels of explanation should sometimes be envisaged, in order to differentiate between those who are subject specialists and those who are not. Finally, analysis of the statistics may emphasise the service dimension of statistical products.

5. COMPARABILITY

There may be a difference between national concepts and European definitions. For example, suppose that in a given Member State the concept of gross investment excludes the value of capital goods acquired through financial leasing but that they are included in the European definition. If it is not possible to collect at individual enterprise level the value of goods acquired through financial leasing, an estimate will need to be made at an aggregated level. Comparability is not limited to the comparability within the EU: Eurostat should be in a position to assess the comparability of EU statistics with the statistics of other countries or groups of countries (USA, Japan).

Comparability should also exist over time: changes due to modification of the reference concept or of the measurement process should be documented and the impact of these changes assessed. In the same way, changes in society in general (e.g.: new legislation, mergers and demergers) having an impact on continuity should be taking into account.

6. COHERENCE

Where similar statistics from various sources² exist, they should be identified and any differences should, if possible, be quantified. A discrepancy between two sets of statistics produced by different surveys may be due to differences in the data collection process or differences in reporting units resulting in different estimates. The situation may be improved by benchmarking (for instance monthly or quarterly statistics on annual results) or by combining different survey results. In any case, misunderstanding by users should be prevented by the use of different wording for the different concepts.

In addition, statistics estimating complex concepts (ratios, elasticity, etc.) should be based on coherent elementary statistics (with compatible definitions of the reference population(s), characteristic, reference period, and statistical unit).

7. COST CONSTRAINT

Quality is associated with costs. Information on the resources available to survey managers allows an assessment of how to optimise the budget they receive. The resources available to a Member State should be taken into account when assessing their likely ability to satisfy quality measures. Two components of cost can be considered: *the cost to the statistical office* and *the cost to the reporting units* (enterprises or parts of enterprises, households, individuals). The cost to respondents differs often with the size of the statistical unit mainly according to the size of the enterprise and the sector of activity³. Costs for a new survey may be much higher than the costs for an existing survey when the sample for existing surveys is only partially from one wave to the following. The use of administrative data drastically reduces the burden on statistical units. Cost for individuals or households are in general not considered. The effects of the burden on individuals are taken into account in non response errors.

Costs are more a constraint for quality improvement than a component of quality itself.

Note 1: There is a trade-off between the different components of quality, especially:

timeliness/accuracy
accuracy/comparability through space⁴
relevance/comparability over time
relevance/accuracy
coherence for large domains/relevance for sub-domains.

² : "Sources" mean here surveys or statistics calculated from administrative data at the national level by the Official Statistical System.

³ : Since different questions may be asked in different sectors.

⁴ : Because comparability through space may impose common concepts that are more difficult to measure than national - but non-comparable - concepts.

Note 2: The above breakdown of quality into components is not unique. Other organisations have proposed other breakdowns for the joint use of statistics:

1) *United Nations (1983)*. Nine points overall, including:

- 'Comparability over time'
- 'Comparability with other statistics'

2) *Statistics Canada (1992)*. Nine points overall, including:

- 'Comparability over time'
- 'Comparability with data from other sources'

3) *Beekman and Struijs (1993)*; three main categories including:

- 'compatibility of statistical output'; compare or combine statistics from one statistics with those of others with the aspects comparability in time (continuity of time series) and compatibility of simultaneous statistics

4) *Statistics Netherlands (1995)*. Section II.2. on 'The need for coherence' lists seven aspects:

- additivity of statistics referring to different areas of economic activity, e.g. fixed capital formation in trade and manufacturing industry;
- comparability of statistics on different subjects, e.g. employment data from labour surveys and turnover statistics from production statistics;
- consistency of the outcomes of short-term statistics with those of annual statistics;
- consistency of the outcomes of regional statistics with those of national statistics;
- continuity of the outcomes for period t with those for period t+1;
- comparability of national statistics with international statistics;
- conformity of concepts and outcomes of business statistics with the System of National Accounts (SNA) and the European System of Accounts (ESA).

5) *Statistics Sweden (1994)*: four components are singled out: Content, Time, Reliability, Accessibility. Content includes comparability with other statistics and Time includes comparability over time. A very user-oriented approach with great importance attached to a quality declaration. However, E. Elvers and B. Rosén, in their contribution (1997) to the *Encyclopedia of Statistical Sciences*, Wiley and Sons, propose a decomposition into 5 components similar to the first 6 components of Eurostat but grouping coherence and comparability.

References

- Biemer, P., Groves, R.M., Lyberg, L.E., Mathiowetz, N.A. and Sudman, S. (eds) (1991)** *Measurement Errors in Surveys*, New York: Wiley
- Beekman, M. M. and Struijs, P. (1993).** *The quality of economic concepts and definitions.* Statistical Journal of the United Nations ECE 10, 1-15.
- Biemer, P.P. and Fesco R. S. (1995)** *Evaluating and controlling measurement error in business surveys*, in Cox and al. (eds) Business Survey Methods, New York: John Wiley, pp.257-281
- Bureau of Labor Statistics (1995)**, Quality Measurement Model, presented at the International Conference on Survey Measurement and Process Quality, Bristol (UK)
- Elvers, E. and Rosén, B. (1997)**, Quality Concept for Official Statistics, Entry in the forthcoming update of the Encyclopedia of Statistical Sciences, Wiley & Sons
- Groves, R.M. (1989)** *Survey Errors and Survey Costs*, New York: John Wiley
- Lessler, J.T. and Kalsbeek, W. D. (1992)** *Nonsampling Errors in Surveys*, New York: John Wiley
- Särndal C.E., Swensson B. and Wretman J.(1992)** *Model Assisted Survey Sampling*, New York: Springer-Verlag
- Statistics Canada (1992).** *Policy on Informing Users of Data Quality and Methodology.* Policy manual, Statistics Canada, April 7, 1992.
- Statistics Netherlands (1995)** *Reference Manual on Design and Implementation of Business Surveys.* Koeijers, E. and Willeboordse, A. (editors). First draft, on request of Eurostat.
- Statistics Sweden (1994)** Quality definition and recommendations for quality declarations of official statistics (authors: Chris Denell, Eva Elvers and Bengt Rosén).
- United Nations (1983)** *Guidelines for quality presentations that are prepared for users of statistics.* Statistical Commission and Economic Commission of Europe, Conference of European Statisticians, Meeting on Statistical Methodology, 21-24 November 1983.

Basic concepts used

A basic concept of statistics is the statistical characteristic, which is defined by the following operation: a given statistical measure is applied to summarise the values of a variable for the units in a specified group. The entire group of units is usually called the population. Often, different subgroups, study domains, are of interest. Population, units, and variables all have a time of reference, a point or a period in time.

To answer questions on a subject-related issue, there is a set of suitable statistical characteristics, the characteristics of interest. Sometimes these are difficult to make operational in a statistical survey. The survey is directed towards targeting characteristics that are feasible from the survey point of view and close enough to the interests.

A survey leads to estimates of the values of the statistical target characteristics. An estimate may be based on data from an enumeration, observations in a sample, a model, or some combination thereof. Auxiliary information is sometimes used together with sample data.



Doc. Eurostat/A4/Quality/98/General/Standard report

Orig. : EN

Available in EN

RD/m

**ASSESSMENT OF THE QUALITY
IN STATISTICS**

**Luxembourg - 29 May 1998
Jean Monnet Building - Room M1
Starting at 9.30 a.m.**

**Item 5.3. of the agenda
Standard quality report**

Standard Quality Report

CONTENTS

1. RELEVANCE	48
2. ACCURACY	48
2.1 Sampling errors	49
2.1.1 Probability sampling	49
2.1.2 Non probability sampling	50
2.2 Non sampling errors	50
2.2.1 Frame errors	50
2.2.2 Measurement errors	51
2.2.3 Processing errors	52
2.2.4 Non-response errors	53
2.2.5 Model assumption errors	53
3. TIMELINESS AND PUNCTUALITY	54
4. ACCESSIBILITY AND CLARITY	55
5. COMPARABILITY	55
5.1 Spatial comparability	55
5.2 Comparability over time	56
6. COHERENCE	56
6.1 Coherence between provisional and final statistics	56
6.2 Coherence of annual and short-term statistics	56
6.3 Coherence of statistics in the same socio-economic domain	57
6.4 Comparison of statistics with national accounts	57
7. COST AND BUDGET	57
7.1 Cost to respondents	57
7.2 Costs and budget for the statistical office	58

STANDARD QUALITY REPORT

INTRODUCTION

This paper sets out the proposals for a quality report on statistics. It describes the subjects that should be covered in a report on quality, and lists particular information that Eurostat expects Member States to provide. Such a report makes Member State deviations from Eurostat norms explicit. Some background to the proposal below is included in the documents "Eurostat/A4/Quality/98/General/Definition" and "Eurostat/A4/Quality/98/General/Glossary".

1. RELEVANCE

To make sure that the produced statistics remain relevant for the users, the Member States should conduct regularly a survey of the opinion of their main users.

To be reported on a multi-yearly basis: A summary including users' description, origin and satisfaction of users' needs, and relevance of statistics for the users.

2. ACCURACY

The overall accuracy is one of the main needs of users. This information should be given by the publication of confidence intervals together with statistics. The confidence intervals should take into account all the effects of corrections of errors and remaining errors, and the variance due to probability sampling, of which:

- bias of estimators, cut-off threshold, imputation methods of non-responses and other systematic bias (due to reweighting, measurement errors etc.);
- variance due to probability sampling, misclassifications, measurement errors, processing errors, imputation and reweighting due to non-responses, statistical models (for instance, seasonal adjustment).

By definition, the variance is associated with an estimator. Corrections are often done in order to reduce bias using statistical techniques. They may be included in the estimator. Other errors occurring at the time of data collection or data entry can not be corrected neither by editing process nor by estimators, but they should be taken into account in a confidence interval.

In order to obtain comparable confidence intervals between Member States, the same components should be included in the variance estimation process. In a first step, we propose to limit the components to variance due to probability sampling (associated to the estimator used), misclassifications and imputation due to non-responses, even if some non-sampling errors can not be excluded and will contribute to the confidence intervals.

Member States (MS) should provide information on the different types of errors.

2.1 Sampling errors

2.1.1 Probability sampling

If we assume we are estimating from a random probability sample, and if we ignore all other sources of error (frame error, measurement error, etc.), then a measure of the accuracy of an estimate is the square root of the mean square error. The mean square error of an estimator is equal to the variance of the estimator plus the square of the bias. Obviously, whenever the methods of sampling and measurement produce an unbiased estimator, then the mean square error is equal to the variance of the estimator.

Probability samplings are used where registers are available and accuracy can be estimated for variables related to statistical units. Sampling in several stages may use non-probability sampling for drawing units at one stage of a multi-stage design.

For instance, it is often the case for survey on consumer prices, where first-stage units could be outlets and second stage units certain commodities within that outlet.

2.1.1.1 Bias

Estimators of totals are often unbiased. For example, estimates produced using the Horvitz-Thompson estimator are unbiased. Frequently, however, estimates are calculated using the ratio or regression estimator, and these are biased but consistent. Occasionally more robust, but biased, estimators are used.

<p>To be reported on a multi-yearly basis: Yearly biases due to the estimation method, when measurable. The yearly median and range of bias for quarterly and monthly statistics</p>
--

2.1.1.2 Variance

For most of the standard sample designs (for example simple random or stratified random sampling) used with standard estimation methods (for example Horvitz-Thompson or ratio estimator), there are standard formulae to estimate the variance and standard error associated with an estimator. When a stage of the selection procedure is not a probability sample, it is possible to approximate the variance under certain assumptions. However, this often leads to an under-estimation of the variance, which should be assessed.

To be reported:	Frequency of statistics		
	Monthly (resp. Quarterly)	Yearly	Multi-yearly
Standard error of:			
Total (resp. Level for an Index)	Y	Y	Y
Growth rates Consecutive periods	Y	Y	
Same period for consecutive years	Y		
Frequency of Transmission to Eurostat:	Monthly (resp. Quarterly)	Yearly	Multi-yearly

A harmonised way for estimating variance is needed.

The estimation has to take account of the variability due to imputation of non-responses (item and unit). The estimation has to consider as well the impact of misclassification errors.

This information is crucial to judge if by aggregating statistics at the Eur15 level a reliable information can be disseminated, even if it is not the case at the Member State level. Furthermore it could be used to publish confidence intervals.

2.1.2 Non probability sampling

In some circumstances, non-random sampling is used. National Statistical Institutes may use expert samples (also called judicious) based on a high coverage of relevant characteristics (for instance production, employment, turnover...).

In these circumstances it is impossible to obtain an objective assessment of the accuracy of the estimates. Nevertheless, some rough accuracy indicators of quality can be designed using sensitivity analyses. When units are selected with certainty following a structural auxiliary variable, such as yearly value added, a more sophisticated one could be built using an econometric model in order to estimate the effect of enterprises not selected.

To report:	Transmission frequency
• The yearly coverage rate	Yearly
• An accuracy indicator	Monthly
• The methods used to obtain these reliability indicators	Multi-yearly

2.2 Non sampling errors

2.2.1 Frame errors

Frame errors include overcoverage, undercoverage, and misclassification.

A quick inclusion of new units and exclusion of dead units in the sampling frame is necessary to limit frame errors.

Over-coverage relates either to wrongly classified units that are in fact out of scope or to units that do not exist in practice.

To be reported:	Transmission frequency
• Yearly rates of over-coverage broken-down according to the main characteristics of the population.	Yearly
• The methods used to obtain these figures	Multi-yearly

Undercoverage refers to (new) units not included in the frame, either through real birth or demergers, and to wrongly classified units.

It may also be due to the deliberate use of cut-off samples by Member States. In this case, Bias should be assessed.

To be report:	Transmission frequency
• The yearly rates of under-coverage broken-down according to the main characteristics of the population.	Yearly
• The yearly biases due to cut-off samples	Yearly
• The methods used to obtain these figures	Multi-yearly

Misclassification refers to wrong classifications that belong to the target population. These misclassifications may reflect economic phenomena or special cases, and should be corrected or not. What are the rules for taking account of misclassifications? How do these corrections affect statistics? Do they occur when statistics revised?

To report:	Transmission frequency
• Methodological notes on the processing of misclassification.	Multi-Yearly / when changes occurs
• The yearly rate of misclassification	Yearly
• The methods used to obtain these figures	Multi-yearly

The misclassified units lead to estimates for a domain over several strata. The corresponding variance should take into account the variance of all units of the considered domain over the several initial strata.

2.2.2 *Measurement errors*

Measurement errors are errors that occur at the time of data collection. There are a number of sources of measurement error, including the survey instrument (the form or questionnaire), the respondent, the information system, the mode of data collection (face to face interviewing, telephone interviewing, self-administered mail survey, diary surveys, administrative records, direct observation, and electronic observation), and the interviewer.

Reporting errors may occur, especially if EDI (electronic data interchange) is not used.

Errors due to the reporting unit

The errors due to the reporting units are more important in household survey than in business survey. Specific studies done by Member States on these errors should be reported. For example it may concern the effects of recalling past events, the tendency to over-report characteristics perceived to be valued by others (social desirability effects), the lack of attention of the respondents, the effects of age and education etc.

Statisticians try to minimise these errors by improving questionnaire design and interviewer training. Errors due to the respondent are difficult to separate from questionnaire and interviewer effects. It requires to ask again the same questions to the respondents (for a sub-sample) in order to measure the response variability.

Errors due to the questionnaire

The quality of a questionnaire is difficult to assess in an objective way.

A description of the pilot survey and the conclusion of the analysis of the questionnaire by a cognitive laboratory are necessary in order to assess the questionnaire.

The primary objective of cognitive laboratory research methods is to identify through careful analysis questioning strategies that will yield more accurate answers.

For instance, a well-known type of measurement error occurs in a survey where individuals are asked to report their age. There is a tendency to report age data that are rounded off. It is preferable to ask the date of birth.

Errors due to the interviewer

The effect of the interviewer is easier to measure than for other measurement errors, because the selection of the interviewer may be random. A sample design may be built in order to measure the variation due to the interviewer.

Errors due to the measurement of a concept

For instance:

- The reference period for which statistics are requested may differ from the period data are based on. (Calendar year and accounting year)
- The measure of pure price evolution needs to take account of quality changes.

Special techniques may be used to assess measurement error (re-interview, estimation of an “interviewer effect”, etc.). But the costs of this sort of exercise are often prohibitively high, both for data providers and data collectors.

Very often, NSIs limit their investigation to consistency edits between characteristics of the questionnaire (or between characteristics of the questionnaire and any external information of good quality).

The description of the edits should consider the breakdown into macro-edits/micro-edits, the latter being divided into (missing values/errors/anomalies).

To be reported on a multi-yearly basis:

- 1) Methods used to reduce this kind of error
- 2) The variance and bias due to the interviewer
- 3) Error rates.

2.2.3 Processing errors

Processing errors are errors in post-data-collection processes such as coding, keying, editing, weighting, and tabulating.

For coding, some measures of the error rates can be achieved through standard techniques of quality control, for instance by verifying the quality of a sub-sample of the processed questionnaires (to check the level of errors either during the keying phase or during the processing of edits by NSI staff).

For editing, weighting, estimation and tabulating processes, the errors are difficult to measure..

To report:	Transmission frequency
• The rates of processing errors broken down according to the different steps.	Multi-Yearly
• Methodological notes on their estimation	Multi-Yearly

2.2.4 Non-response errors

Item non-response and **unit non-response** are often a major source of error in surveys. They may bias the estimates in an uncontrolled way. A traditional and simple indicator of non-response is the response rate, either by unit or by item. It can be either weighted (by the sampling weights or other variables such as numbers of employees, turnover etc.) or non-weighted. The distinction between actual non-respondents and out-of-scope units is very important. It would be advisable to keep as a separate group the category of units for which the NSI does not know if the non-measurement is a non-response or a frame error.

An important question is how respondent units differ from non respondent ones. One kind of non responses is due to data coming late. The effect of such non-responses can be measured by comparison a posteriori. The second kind of non response can be investigated through a supplementary survey of a sub-sample of non-respondents, generally using more efficient data-collection techniques (for instance, interviews instead of mail inquiries).

Non-responses (item- or unit-) have to be processed during the production process, either through imputation (partial or total) or through reweighting of the responding units. This process introduces randomness, which should be integrated in the estimation of biases and variances.

To be reported:	Frequency of statistics		
	Monthly or quarterly	Yearly	Multi-yearly
	Frequency of transmission to Eurostat		
• Unit and item response rates	Monthly / Quarterly	Yearly	Multi-yearly
• A description of the methods used for imputation and/or reweighting for non-response	Multi yearly		
• The variance due to imputation	Monthly / Quarterly	Yearly	Multi-yearly
• Non-response bias	Multi yearly		
• Methodological notes on the above estimations			

2.2.5 Model assumption errors

Model assumption errors occur with the choice of methods, such as calibration, generalised regression estimator, calculation based on full scope or constant scope, benchmarking, seasonal

adjustment and other models not included in the preceding accuracy components, in order to calculate statistics or indexes.

The quality of such models should be assessed in term of variance and bias. This evaluation should be integrated in the confidence intervals related to statistics, but there is no standard method for this integration.

Seasonal adjustment model. A very important point related to the quality of seasonal adjustment is the detection and correction of outliers prior the seasonal adjustment. Outliers of aggregated data are due to irregular phenomena (such as a strike) and have to be corrected for estimating the seasonal components.

Benchmarking:

Benchmarking procedures are generally used to produce revised infra-annual series coherent with yearly statistics. This method should be described when it is applied. The comparison between initial values and final value should be reported so as to assess the impact of the method, and consequences for the accuracy of the benchmarked series should be evaluated (diminution of bias, increase of variance).

Other models may be used for the estimation of part of the statistics when no data collection for a sub-population is available, when aggregated administrative data is used, when a part of needed data are not yet available (for instance the weights for index calculation).

The selection procedure of these models (i.e. why a given model has been chosen in preference to alternative ones), together, perhaps, with the associated estimation error of the corresponding estimates, should be reported to Eurostat.

To be reported for each model used to adjust the series:
on a multi-yearly basis, comments on:

- The verification of the assumptions underlying the model
- The test of the predictive power of the model by using historical data to 'predict' known quantities
- The comparison of the results generated by the model with other related sources of data
- The use of screening and cross-validation studies
- The tests of sensitivity of the model to parameters' estimation
- The validation of the data inputs to the model

3. TIMELINESS AND PUNCTUALITY

In order to analyse the impact of each step in the production process, and with a view to facilitating the discussion of delays with the Member States or developing advance estimates at the European level, the following information would be useful:

- the use of electronic data transmission for the statistical data collection in the Member State
- the legal deadline imposed on respondents in the Member State
- key data-collection dates, specifying when the questionnaires and recalls and follow-ups were sent out and when the field work took place;
- a graph of response in terms of number of questionnaires received by week or by working day
- starting and finishing dates for the editing phase;
- dates for the imputation phase;

- details and date of the event which indicates that the data collection phase can be considered to have ended;
- date on which the advance results were calculated and disseminated;
- date of the quality check (congruency of results) and non-disclosure measures;
- date of availability of the camera-ready publication
- date of publication

Punctuality of data transmission to Eurostat will be evaluated according to the annexed module of the regulation specifying periodicity and delays for data transmission.

4. ACCESSIBILITY AND CLARITY

Eurostat is interested in the following information on the dissemination process:

- a copy of the publication(s)
- information on what, if any, results are sent to reporting units included in the sample
- information on the dissemination scheme for the results (e.g. to whom the results are sent, who the users are)
- a copy of any methodological documents relating to the statistics provided.

This information is useful not only in order to design a consistent dissemination scheme at Eurostat level and to better inform users calling from Member States, but also as a way of obtaining any national comments included in these publications which may assist in interpretation of the statistics within Eurostat. The information will also be of help in reporting on the benefits of the availability of statistics.

5. COMPARABILITY

5.1 Spatial comparability

Where there is a difference between National concepts and European concepts (definition of statistical units, reference population, classifications, definition of observed characteristic), detail of the differences in terms of concept and estimation of consequences for resulting statistics should be mentioned¹.

Similarly, Member States should report if they follow the existing European Recommendation for Measurement, and possible deviations from these norms and the corresponding assessment of consequences on the estimates.

Combination of both information will provide an assessment of the level of comparability of the statistics.

Some complementary (voluntary?) survey on a small sub-sample should assess deviations from the European Standards.

¹ When national statistics are derived not only from a survey sampling estimate but also from a model to adjust for some discrepancy between the statistic produced by the survey and the objective aimed at by the Member State, the adjustment due to choice of model should be included in paragraph 2.2.5 "model assumption errors". Only the differences between the European norms and the norms aimed at after the possible national model adjustment have to be reported in this paragraph. To get more detailed information, see document Eurostat/A4/Quality/98/General/Comparability.

A comparison between Member state practices and Eurostat recommendation should be transmitted on a multi-yearly basis and when significant changes occur. When possible, these differences should be assessed.

5.2 Comparability over time

Some users, especially those working in the macroeconomic field, use time series rather than point estimates for a date. The stability of the concepts and methods of measurement is very important to them. Details of changes in definitions, coverage or methods should be provided. A description of how mergers and demergers of statistical units are handled should be given. Mergers and demergers of units (enterprises, households) often have a significant impact on some statistics, especially when broken down by certain classification. When the comparison of statistics from two consecutive periods may be more affected by legal events than by the socio-economic trend, it is important to report on this aspect. Any change in legislation at the national level having consequences on continuity should be reported. The consequences of non-negligible changes should also be reported.

For seasonal adjusted series, comparison should take into account changes in adjustment procedures, but the seasonal adjustment model should be reported under paragraph 2.2.5 "Model assumption errors".

When significant changes occur leading to discontinuities, Member States should report the reasons together with the statistics. Estimates of these should be transmitted to Eurostat later.

6. COHERENCE

6.1 Coherence between provisional and final statistics

The coherence between provisional and final index should be maximised in the sense that the provisional statistic/ index should be sufficiently accurate.

The percentage of yearly average differences between the final not seasonally adjusted statistic/ index and the provisional statistic/index should be provided for each requested statistic, as well as the dispersion around this mean/median value.

6.2 Coherence of annual and short-term statistics

For several characteristics, statistics have to be produced with both infra-annual and annual frequencies. It is important for the image to the users to aim at a certain coherence of disseminated information. A simple validation method consists in comparing estimates of same concept but derived from these statistics of different frequencies. It seems most natural to compare average annual levels or totals when both frequencies provide estimators in level, and growth rates when at least one of these statistics is an index. These comparisons should be made before possible benchmarking by the NSI, and the Member State should indicate separately if it benchmarks its infra-annual statistics on the considered annual statistic.

For each sector, Member States should compare the following growth rates:

- calculated from the short term survey GRst
- calculated from the annual survey GRyv
- calculated from the annual survey, eliminating the effect of GRyc
births, deaths and changes of classification of the units ²

Conventions diverging from Eurostat norms and recommendations that have a specific impact on the dynamics of the produced statistics, such as base period, type of index - Laspeyres, Paasche Fischer- etc., should also be considered.

6.3 Coherence of statistics in the same socio-economic domain

Frequently, a group of statistics, possibly of different type (in monetary value, in volume or constant price, price indicators) brings a representation of the same phenomenon under different angles. It is very important to check that these representations are compatible in order to anticipate users' questions and to prepare corrective actions.

For instance, business short term statistics like turnover, value-added, variations in stocks have to be compared on a yearly basis with Prodcom statistics after deflation by production prices. Statistics on population, unemployment and employment should be compared. Similarly, opinion surveys and quantitative indicators should be compared. Employment as measured by social statistics (censuses for instance) and by business statistics should ideally have limited discrepancies. Surveys on individuals' holidays should have some consistency with quantitative indicators of tourism statistics, or short term indicators in the service sector (Horeca). Etc.

6.4 Comparison of statistics with national accounts

In order to be in a position to advise users on the source of information best suited to their needs it may also be interesting to compare statistics from the surveys to national accounts. So an indication of the overall consistency of these surveys with other sources of information could be provided to users. The methodology used in the national account statistics would need to be described for the considered statistics, including the primary data source and the adjustments that are made. Discrepancies in the concepts should also be taken into account.

A summary of all these comparisons should be sent to Eurostat every year.

7. COST AND BUDGET

7.1 Cost to respondents

Member States should provide Eurostat annually with the average cost to respondents of providing the information requested. The total cost to the population, obtained by multiplying the average cost per enterprise by the sample size, should also be provided.

²more precisely, this corresponds to the target population defined as follows:

- units are classified according to their main activity in year n
- the population is the intersection of the population in n with the population in n+1.

The sample has to be post-stratified to take into account this restriction. In the event of a partly rotating sample, a reliable estimate consists in considering the intersection of the sample in year n and the sample in year n+1 provided the sample is not completely renewed in the corresponding strata!

They should also indicate whether they co-ordinate samples for different surveys or whether they rotate the sample for the surveys, in order to measure the dispersion of the total statistical burden between individual units.

7.2 Costs and budget for the statistical office

Member States should provide Eurostat with the total costs associated with the survey. On a multi-yearly basis, information on the number of employees (broken down by statisticians, methodologists, programmers and analysts, interviewers, clerical staff) should be provided. Computer resources (hardware and software) and other processing equipment (Optical Character Reading equipment, etc.) used should also be detailed.

INDEX OF TERMS DEFINED IN THE GLOSSARY³

Bias, 3
Consistent, 3
Editing, edits, 5
Electronic data interchange, 5
Electronic data transmission, 7
Estimate, 3
Estimator, 3
Frame, 4
Frame error, 3, 4
Imputation, 7
Item non-response, 6
Macro-editing, 5
Measurement error, 3, 5
Micro-editing, 5
Misclassification, 4
Non-responses, 6
Out-of-scope, 6
Overcoverage, 4
Processing error, 6
Register, 3
Standard error, 3
Undercoverage, 4
Unit non-response, 6
Variance, 3

³ See Doc Eurostat/A4/Quality/98/General/Glossary



Vedlegg C

Doc. Eurostat/A4/Quality/98/General/Glossary - ex97/03

Orig. : EN

Available in EN

RD/mv

**ASSESSMENT OF THE QUALITY
IN STATISTICS**

**Luxembourg - 29 May 1998
Jean Monnet Building - Room M1
Starting at 9.30 a.m.**

**Item 5.2. of the agenda
Glossary on quality in statistics**

Glossary on Quality in Statistics

The scope of this glossary is limited to what is necessary to avoid ambiguities when reading its companion documents “Definition of quality”, “Proposal for a quality report on statistics” and other proposal for a quality report on specific domains. It is an important prerequisite to the design of a common analysis of the quality of statistics.

These definitions are adapted from similar work by the US Department of Commerce (1978) and by Lessler and Kalsbeek (1992). These documents are themselves a review and analysis of the available literature on errors in surveys.

Definitions related to survey sampling are adapted from Särndal et al. (1992).

BASE YEAR

It refers to the year for which the weights of commodities are fixed following the consumer preference at this particular year. It is used to calculate fixed base index (by opposition to chain index).

BIAS

The bias is equal to the difference between the expected value of the estimator and the true value being estimated.

CHAIN INDEX

Chain index is calculated by comparing the current period to the most recent period. Furthermore, by opposition to fixed base index, the commodity weights of a chain index are re-based on the most recent period.

In practice, there is often a time lag between the most recent period and the period on which weights are calculated. The weights often depend on the availability of the results of last structural survey and are calculated on a yearly basis. This time lag may be reduced using a predictive model.

CODING

Coding is the procedure for converting verbal information into numbers or other symbols which can be more easily counted and tabulated.

CODING ERROR

The assignment of an incorrect code to a survey response.

CONSISTENT

An estimator is consistent if the probability that it is in error by more than a given amount tends to zero as the sample becomes large.

COEFFICIENT OF VARIATION

The coefficient of variation of an estimator is a measure of relative dispersion equal to the ratio of standard deviation to mean. In practice, this is the estimated standard deviation of a point estimate divided by the point estimate and expressed as a percentage.

CONSTANT SCOPE

Constant scope refers to frame population, it should not include births, deaths, mergers and demergers, the population should be identical for both periods on which indexes are estimated, on the contrary of full scope.

COORDINATION OF SAMPLES

Increasing the sample overlap for some surveys rather than drawing the samples independently is known as positive coordination. A positive coordination is often searched in repeated surveys over time (panels) in order to obtain a better accuracy of statistics depending on correlated variables from two surveys. Reducing the overlap between samples for different surveys is known as negative coordination. A negative coordination is searched in order to share more equally the response burden among responding units.

CURRENT PERIOD

It refers to the period for which statistics are produced.

DEAD UNIT

A dead unit is a unit that ceases existing during the current period. For enterprise it is because of business closed down and for persons because of death.

DELINEATION

The delineation of statistical units is mainly done by considering groupings or divisions of legal units and applying EU harmonised rules and/or using classifications. Bad delineation may occur when two units are considered as one and vice versa (for instance a cafeteria attached to a supermarket). Another frequent case of difficulties in delineation is encountered with very small enterprises which represent a total employment less than the full-time equivalent of one year. Bad delineation may also occur for household too.

DOMAIN

Point estimates are derived for domains of interest. A domain may be the entire population or any specified sub-population for which separate estimates are designed. The domains may completely overlap with the sampling strata or may cut across them.

EDITING, EDITS

Editing is the procedure for detecting and adjusting individual errors in data records resulting from data collection and capture. The checks for identifying missing, erroneous, or suspicious values in computer-assisted editing are called edit rules or edits. An editing change occurs when an item (question) value is adjusted as a consequence of action taken when an error is identified.

ELECTRONIC DATA INTERCHANGE (EDI)

EDI is the electronic transfer of business transaction information in a standard format between business partners. EDI offers businesses the opportunity to retrieve information electronically from their internal systems and to forward that information to trade partners/suppliers/customers/government through a communications network.

ERROR

The difference between the survey value and the corresponding true value.

ESTIMATE

An estimate is the calculated value using an estimator.

ESTIMATOR

An estimator is the mathematical function by means of which the estimate for a particular parameter is computed.

EXPECTED VALUE

The hypothetical average from the conceived replicates of the survey all conducted under the same essential conditions.

FOLLOW-UP

A procedure whereby those members of a selected sample for whom a response is not obtained by one data collection strategy (e.g. telephone, interview or mail) are contacted by the same or another data collection strategy in order to increase response rate. The term can also be used to designate repeated surveys among a panel of respondents.

FRAME

The frame for a survey is the listing or listings of units that delimit, identify, and allow access to the elements of the target population.

FRAME ERROR

Frame error may be caused by the inherent limitations of input data, or by delays and errors in data acquisition and processing. Frame error covers:

- coverage errors - erroneous inclusions, omissions and duplications;
- classification errors - units not classified, or misclassified by industry, geography or size;
- contact errors - units with incomplete or incorrect contact data.

FULL SCOPE

Full scope refers to frame population, it should cover births, deaths, mergers and demergers, during the period on which indexes are estimated, on the contrary of constant scope.

HEDONIC METHOD

The hedonic approach assumes that a commodity can be view as a bundle of physical and performance characteristics, that quality is a transform of those characteristics, that a change in the market price of a commodity can be factored out into a quality price component and a pure price component, and that the price of a product is correlated with its characteristics. Regression methods are used to estimate hedonic price (which reflects the pure price change) to eliminate the effects of quality changes.

IMPUTATION

Imputation refers to the replacement of missing data using logical edits or statistical procedures. It concerns item non-response or unit non-response.

INDEX AT CURRENT PRICE OR INDEX IN VALUE

An index used to measure the evolution of values expressed in current prices, by considering constant weighting factors.

INDEX AT FIXED PRICE OR VOLUME INDEX

An index used to measure the evolution of quantity by considering constant prices.

INTERVIEWER ERROR

Error due to the interviewer when for instance asking the question, reporting the answer...

ITEM NON-RESPONSE

See non-response.

ITEM RESPONSE RATE

The item response rate should be calculated for an item by dividing the number of available data by the number of available and missing data. The weighted item response rate is useful too.

LASPEYRES INDEX

A price (resp: quantity) Laspeyres index used quantity (resp: price) weights which are observed in the reference situation (by opposition to Paasche index).

MACRO-EDITING

A procedure for pointing out suspicious data by checking aggregates or applying statistical methods on all records or on a subset of them.

MEAN SQUARE ERROR

The variance of the estimator plus the bias squared.

MEASUREMENT ERROR

Measurement error can be thought of as the difference between the value collected during the survey and the true (individual) value. These errors may result from: the failure of the respondent to report the true value (respondent error); the failure of the interviewer to record the value correctly (interviewer error); or the failure of an instrument (questionnaire) to measure the value correctly.

MICRO-EDITING

An exhaustive check to find errors by inspection of each individual observation.

MISCLASSIFICATION

Misclassification refers to wrongly classified units, (for example by industry, geographical area or size) which belong to the target population.

NON-RESPONSE

Non-response results from a failure to collect complete information on all units in the selected sample except non-eligible units (dead units, out-of-scope units, and non-contacts). There are two types of non-response. First, a sampled unit that is contacted may fail to respond. This represents unit non-response. Second, the unit may respond to the questionnaire incompletely. This is referred to as item non-response.

NON-RESPONSE BIAS

This is the bias resulting from limiting the survey analysis to the observed units.

NON-RESPONSE ERROR

Non-response error results from the imputation of missing data due to non-responses.

NON-SAMPLING ERROR

Non-sampling error is the error attributable to all sources other than sampling error. Non-sampling errors arise during the planning, conducting, data processing and final estimation stages of all types of survey.

OUT-OF-SCOPE UNITS

Units that should not be included in the sampling frame because they do not belong to the target population in the current period. The units that are dead, merged or demerged during the current period are not considered as out-of-scope units.

OVER-COVERAGE

Units which are included in the sampling frame but do not belong to the target population. These cases are usually observed for contacted units, but not necessarily for non-contacted units or those excluded from a sample. Reasons for over-coverage are out-of-scope units and a non-updated frame. The overcoverage in a register generally biases the estimators drawn from that sampling list.

PAASCHE INDEX

A price (res.: quantity) Paasche index used quantity (reps: price) weights which are observed in the current situation (by opposition to Laspeyres index).

PROCESSING ERROR

Processing error is the error in the final survey results arising from the faulty implementation of correctly planned survey methods. Processing errors encompass all post-collection operations, as well as the printing of questionnaires. Most processing errors occur in data for individual units, although errors can also be introduced in tabulations and estimates.

RECALL ERROR

Error due to the respondent who might forget events, shift dates, etc...

RECORD CHECK

A study in which data on individual units obtained by one method of data collection are checked against data for the same units from available records obtained by a different method of data collection (for example, comparison of ages as reported in censuses with information on ages from birth certificates).

RECORD-KEEPING ERROR

An error that arises from inaccuracy in the records used for responses.

REFERENCE PERIOD

Index context:

It refers to the period for which an index series is given the value of 100.0. The reference period and the base period are identical for fixed base index, but differ for chain index.

Other context:

Reference period is the period for which statistics are produced. (Also called current period in the framework of indices).

REFUSAL RATE

The proportion of enterprises which, though successfully contacted, refuse to give the information sought. This should be calculated by dividing the number of refusals by the original total sample size.

REGISTER

This should be an up-to-date file of all statistical units from which frames are extracted. Registers are generally derived from administrative data.

RELATIVE STANDARD ERROR

See coefficient of variation.

REWEIGHTING

Reweighting is an adjustment procedure, which copes with unit non-response. It may also be used to increase precision through the use of auxiliary information. Standard methods include post-stratification, calibration and response propensity modelling

SAMPLE DESIGN

The sample design should provide information on the target and final sample sizes, strata definitions and the sample selection methodology. It should also include details of changes in methodology and any impact these changes have on comparability over time.

SAMPLING ERROR

Sampling error is the error due to surveying only a subset of the population rather than conducting a complete census of all businesses in the target population.

STANDARD ERROR

This is calculated as the square root of the variance of the estimator.

SURVEY DESIGN

The survey design addresses the following issues: what statistics are produced, for which population, when, and with what accuracy; what data are to be collected for which units of the population of interest, and what are the methods by which those data are to be collected and processed to produce the required statistics. Operational, organisational and administrative issues are usually addressed.

TARGET POPULATION

This is the conceptual population of elements (units of analysis) at which the survey is aimed and for which statistics should be produced.

TRUE VALUE

The actual population value that would be obtained with perfect measuring instruments and without committing any error of any type, both in collecting the primary data and in carrying out mathematical operations.

UNDER-COVERAGE

Failure to include in the frame all units belonging to the target population. This mainly includes (new) enterprises not included in the frame, either through real birth or demergers, and wrongly classified units. This generally leads biases in the estimates.

UNIT NON-RESPONSE

See non-response.

UNIT RESPONSE RATE

The percentage of an eligible sample for whom information is obtained. For an interview survey, the numerator of the formula is the number of interviews. The denominator is the total sample size minus non-eligible respondents, i.e. minus those not meeting the criteria for a potential respondent as defined for that particular study. Weighted response rate is more useful. In addition to this, another non-response rate using the following weights may be computed: the sampling weights times any existing important variable from the sampling frame (turnover or size). The latter figure illustrates an impact of non-response.

VARIANCE

The variance of a random variable reflects the dispersion of the empirical values around the average. It is the mean square deviation of the variable around the average value.

Norsk - Engelsk ordliste

Aktuell periode	Current period
Aktuell prisindeks (eller verdiindeks)	Index at a current price
Avkuttet utvalg	Cut-off survey
Basisår	Base year
Bearbeidingsfeil	Processing errors
Død enhet	Dead unit
Elektronisk datautveksling, EDI	Electronic data interchange
Enheter som ikke tilhører målpopulasjonen	Out-of-scoop units
Enhetsfracfall (totalfracfall)	Unit non-response
Erindringsfeil	Recall error
Estimat	Estimate
Estimator	Estimator
Fastprisindeks (eller volumindeks)	Index at fixed price
Feil	Error
Feilklassifisering	Misclassification
Forventet verdi	Expected value
Fracfall	Non-response
Fracfallsfeil	Non-response error
Fracfallsskjevhet	Non-response bias
Fullstendig populasjon	Full scope
Hedonisk metode	Hedonic method
Ikke-utvalgsfeil	Non sampling error
Imputering	Imputation
Inndeling i statistiske enheter	Delineation
Interessepopulasjon	Domain
Intervjuerfeil	Interviewer error

Kjedeindeks	Chain index
Kodefeil	Coding error
Koding	Coding
Konsistent	Consistent
Konstant populasjon	Constant scope
Kontroll	Record check
Lagringsfeil	Record-keeping error
Laspeyres indeks	Laspeyres index
Makroredigering	Macro editing
Middelkvadratfeil	Mean square error
Mikroredigering	Micro editing
Målefeil	Measurement error
Målpopulasjon	Target population
Nekterprosent	Refusal rate
Overdekning	Over-coverage
Paasche indeks	Paasche index
Partielt frafall (svarfracfall)	Item non-response
Purring / oppfølging	Follow up
Referanseperiode	Reference period
Register	Register
Relativt standardavvik	Relative standard error
Revekting	Reweighting
Revisjon	Editing, edits
Samordning av utvalg	Coordination of sample
Sann verdi	True value
Skjevhet	Bias
Standardavvik	Standard error
Svarprosent	Unit response rate
Svarprosent for enkelte spørsmål	Item response rate
Trekkeregister	Frame
(Trekke)registerfeil	Frame error
Underdekning	Under coverage
Undersøkelsesplan	Survey design

Utvalgsfeil	Sampling error
Utvalgsplan	Sample design
Varians	Variance
Variasjonskoeffisient	Coefficient of variation
Verdiindeks (eller aktuell prisindeks)	Index in value
Volumindeks (eller fastprisindeks)	Volume index

Engelsk - Norsk ordliste

Base year	Basisår
Bias	Skjevhet
Chain index	Kjedeindeks
Coding	Koding
Coding error	Kodefeil
Consistent	Konsistent
Coefficient of variation	Variasjonskoeffisient
Constant scope	Konstant populasjon
Coordination of sample	Samordning av utvalg
Current period	Aktuell periode
Cut-off survey	Avkuttet utvalg
Dead unit	Død enhet
Delineation	Inndeling i statistiske enheter
Domain	Interessepopulasjon
Editing, edits	Revisjon
Electronic data interchange	Elektronisk datautveksling, EDI
Error	Feil
Estimate	Estimat
Estimator	Estimator
Expected value	Forventet verdi
Follow up	Purring / oppfølging
Frame	Trekkeregister
Frame error	(Trekke)registerfeil
Full scope	Fullstendig populasjon
Hedonic method	Hedonisk metode
Imputation	Imputering

Index at a current price (or index in value)	Aktuell prisindeks (eller verdiindeks)
Index at fixed price (or volume index)	Fastprisindeks (eller volumindeks)
Interviewer error	Intervjuerfeil
Item non-response	Partielt frafall (svarfracfall)
Item response rate	Svarprosent for enkelte spørsmål
Laspeyeres index	Laspeyres indeks
Macro editing	Makroredigering
Mean square error	Middelkvadratfeil
Measurement error	Målefeil
Micro editing	Mikroredigering
Misclassification	Feilklassifisering
Non-response	Frafall
Non-response bias	Frafallsskjevhet
Non-response error	Frafallsfeil
Non sampling error	Ikke-utvalgsfeil
Out-of-scoop units	Enheter som ikke tilhører målpopulasjonen
Over-coverage	Overdekning
Paasche index	Paasche indeks
Processing errors	Bearbeidingsfeil
Recall error	Erindringsfeil
Record check	Kontroll
Record-keeping error	Lagringsfeil
Reference period	Referanseperiode
Refusal rate	Nekterprosent
Register	Register
Relative standard error	Relativt standardavvik
Reweighting	Revekting
Sample design	Utvalgsplan
Sampling error	Utvalgsfeil
Standard error	Standardavvik
Survey design	Undersøkellesplan
Target population	Målpopulasjon
True value	Sann verdi

Under coverage
Unit non-response
Unit response rate
Variance

Underdekning
Enhetsfrafall (totalfrafall)
Svarprosent
Varians

Stikkordregister

A

Aktualitet · 4; 5; 6; 9
Aktuell periode · 28; 34
Aktuell prisindeks · 28
Avstemming · 20; 21; 24

B

Basisår · 28; 31; 32

D

Død enhet · 28

E

Elektronisk datautveksling, EDI ·
28
Erindringsfeil · 29
Estimat · 29
Estimator · 29

F

Fastprisindeks · 29
Feil · 29
 bearbeidingsfeil · 6; 13; 19; 28
 frafallsfeil · 6; 8; 20; 30
 ikke-utvalgsfeil · 6; 14; 16; 30
 modellforutsetningsfeil · 6
 målefeil · 6; 13; 18; 19; 33
 trekkeregisterfeil · 6; 14; 36
 utvalgsfeil · 6; 14; 37
Feilklassifisering · 13; 15; 16; 17;
29; 36
Forventet verdi · 29
Frafall · 13; 14; 15; 19; 20; 30; 34;
36
 enhetsfracfall · 20; 29; 30; 31
 partiellfracfall · 20; 30; 31; 34
Frafallsskjevhet · 30

H

Hedonisk metode · 30

I

Imputering · 6; 13; 15; 20; 31
Intervjuerfeil · 31

K

Kjedeindeks · 31
Klarhet · 4; 5; 7; 22
Kodefeil · 32
Koding · 19; 32
Konfidensintervall · 13; 15; 21
Konsistent · 32
Kontroll · 32
Kostnad · 5; 8; 26
Kostnadskrav · 8

L

Lagringsfeil · 32
Laspeyres indeks · 32; 34

M

Makroredigering · 32
Middelkvadratfeil · 14; 33
Mikroredigering · 33
Modellfeil · 20; 24

N

Nekterprosent · 33
Nøyaktighet · 4; 5; 6; 7; 9; 13; 14;
15; 36

O

Oppfølging · 34
Overdekning · 16; 33; 36

P

Populasjon · 12; 14; 16; 20; 26;
35; 36
 delpopulasjon · 12; 17; 21; 31
 fullstendig populasjon · 30; 32
 interessepopulasjon · 31
 konstant populasjon · 20; 30;
32
 målpopulasjon · 17; 20; 29; 33;
36; 37
 referansepopulasjon · 23
Punktlighet · 4; 5; 6; 22
Purring · 34
Paasche indeks · 32; 34

R

Referansetidspunkt · 12
Referanseperiode · 19; 34
Register · 14; 30; 34
 trekkeregister · 36
Registerfeil · 16; 20
Relativt standardavvik · 34
Relevans · 4; 5; 9; 13
Revekting · 34
Revisjon · 19; 22; 35

S

Sammenheng · 4; 5; 8; 9; 24
Sammenlignbarhet · 4; 5; 7; 9; 23
Sammenligning · 5; 21
Sann verdi · 35
Skjevhet · 13; 14; 16; 20; 21; 30;
33; 35
Standardavvik · 15; 35; 37
Statistiske enheter · 23; 31; 34
Svarprosent · 20; 34; 35; 36
Systematiske feil · 13; 36

T

Tilgjengelighet · 4; 5; 7; 10; 22

U

Underdekning · 16; 36
Undersøkelsesplan · 36
Utvalg · 8; 12; 14; 15; 18; 26; 29;
30; 32; 33; 34; 35; 37
 avkuttet utvalg · 13; 16; 28
 ikke-sannsynlighetsutvalg · 14;
15
 sannsynlighetsutvalg · 13; 14;
15
Utvalgsplan · 15; 18; 37

V

Varians · 13; 15; 17; 33; 35; 37
Variasjonskoeffisient · 37
Verdiindeks · 28; 37
Volumindeks · 29; 37

De sist utgitte publikasjonene i serien Notater

- 99/12 K.J. Einarsen, H. Hartvedt, A.B. Skara og C.B. Strand: Faktaark for FylkesKOSTRA-utdanning. Årsrapporteringen for 1998. Sør-Trøndelag fylkeskommune: Nøkkeltall med indikatorer for: Prioriteringer og ressursbruk, Dekningsgrad, Produktivitet og kvalitet. 41s.
- 99/13 G. Sparby: Selvangivelsesstatistikk 1993-1996: Dokumentasjon. 121s.
- 99/14 B.R. Joneid og J. Lajord: FD – Trygd: Dokumentasjonsrapport: Demografi. 1992-1993. 95s.
- 99/15 A.L. Brathaug: Evaluering av fylkes-KOSTRA, helsetjenester 1998. 64s.
- 99/16 A.J. Isachsen, S.O. Stoknes og G.H. Bjønnes: Den store gjettekonkurransen. 26s.
- 99/17 Z. Li-Chun, L. Solheim og M. Stålnacke: Lønnsomhetsundersøkelsen for fiskefartøy: Ny utvalgsplan og estimeringsmetode. 61s.
- 99/18 Å. Nossun og L. Sandberg: Harmonisert konsumprisindeks. 31s.
- 99/19 P.E. Tønjum: Dokumentasjon av edb-rutiner for kjeding av årlig og kvartalsvis nasjonalregnskap. 66s.
- 99/20 B. Halvorsen og K.R. Wangen: Dokumentasjon av utdrag fra skattestatistikken 1975-1985 for kobling mot forbruksundersøkelsen. 31s.
- 99/21 B. Halvorsen: Dokumentasjon av analysefiler til prosjektet "Fleksibel energibruk i husholdningene": Prisdatabaser for varer og tjenester (1975-1994), husholdningstariffer for elektrisitet (1975-1996) og temperaturdata (1957-1996). 34s.
- 99/22 B. Halvorsen: Dokumentasjon av analysefiler til prosjektet "Fleksibel energibruk i husholdningene": Forbruksundersøkelsen 1974-1995. 34s.
- 99/23 A. Barstad: På vei mot det gode samfunn?: Om opplegget for en publikasjon som beskriver verdilandskapet i Norge. 72s.
- 99/24 G. Dahl: FD – Trygd: Type prosjekt, organisering, forløpsdata og analysemuligheter. 34s.
- 99/25 J. Johansen og J. Lajord: FD – Trygd: Dokumentasjonsrapport: Arbeidssøkere. 1992-1993. 99s.
- 99/26 K-A. Hovland og F. Bendiksen: Foreldrebetalingundersøkelse: Rapport om betaling for heldagsopphold i kommunale og private barnehager, januar 1999. 38s.
- 99/27 G. Dahl og J. Johansen: FD – Trygd: Dokumentasjonsrapport: Sysselsetting. 1992-1993. 77s.
- 99/28 P. Bakken, H. Hagen og J.A. Osnes: Kvartalsvis lagerstatistikk. 64s.
- 99/29 E. Rønning: Fokusgrupper om opplysninger om skolegang: Dokumentasjon og resultater. 38s.
- 99/30 D. Roll-Hansen: Kompetanse i grunnskolen. 53s.
- 99/31 T.P. Bø: Klassifisering av registrerte arbeidsledige og personer på tiltak i arbeidskraftundersøkelsen (AKU). 45s.
- 99/36 M. Stålnacke, J-A. Sigstad Lie og L. Solheim: En analyse av SSBs generelle utvalgsplan fra 1995 basert på næringsvise sysselsettingstall. 83s.
- 99/37 B.O. Lagerstrøm: Trivsels- og arbeidsmiljøundersøkelse blant intervjuere i Statistisk sentralbyrå. 155s.
- 99/38 K.J. Einarsen: Evalueringsrapport for pilotforsøket for FylkesKOSTRA-utdanning. 55s.
- 99/39 L. Rogstad: FoB2000: Adressesamsvar mellom folkeregister og adresseregister i GAB: – rapport fra Lysebu-seminar 8. og 9. desember 1998, – tiltaksplan for bedre adressesamsvar. 39s.
- 99/41 R. Johannessen: Kommunale gebyrer knyttet til bolig. Januar 1999. 30s.