

Anne Gro Hustoft og Hans Viggo Sæbø

Nokre sentrale omgrep knytte til metadata -

til bruk i dei felles metadatasystema i SSB

Nynorsk versjon

Interne dokumenter

Innhald

Innhald	1
Bakgrunn	3
Om definisjonane	3
Statistikk	5
Statistisk mål	5
Indikator	5
Indeks	5
Kvalitet	6
Kvalitet i statistikken	6
Metadata	7
Statistikkmetadata	7
Statistikkmetadataasystem	7
Register	7
Basisregister	8
Administrativt register	8
Populasjon	9
Statistisk eining	10
Identifikator	10
Variabel	11
Måleining	12
Klassifikasjon	12
Kodeliste	13
Referansar	14
De sist utgitte publikasjonene i serien Interne dokumenter	14

Bakgrunn

I arbeidet med metadatastrategi for SSB vart det klart at sjølv om statistikken langt på veg byggjer på felles omgrep, manglar vi ei samla oversikt over sentrale omgrep knytte til metadata. Dette kan føre til at personar med ulike fagbakgrunn og knytte til ulike arbeidsoppgåver i SSB, snakkar forbi kvarandre. Dette kan igjen leie til mistydingar og effektivitets- og kvalitetstap ved utveksling av data og metadata, og ved organisering og innlegging av metadata i dei felles metadatasystema eller metadatatdrivne systema i SSB som Statistikkbanken, Stabas og Vardok.

Etablering og dokumentasjon av omgrep knytte til metadata, er ein viktig del av arbeidet med oppfølging av metadatastrategien. Dette notatet er eit første innspel i dette arbeidet. Det er utarbeidd i nær kontakt med metode, IT, formidling og fagstatistikarar, og revidert etter kommentarar frå dei fleste avdelingane i SSB, innhenta i ein høyringsrunde. Notatet er òg diskutert i Metadataforum, i Styringsgruppa for metadatastrategien og i Standardutvalet.

Ei drøfting med tilråding om korleis desse omgrepa bør tolkast i det praktiske arbeidet med metadatasystem, vil vere ei nyttig vidareutvikling.

På eit meir detaljert nivå trengst det presise definisjonar også for kommunikasjon mellom maskinar. Her vil det bli utført eit definisjonsarbeid i prosjektet "Tenestebibliotek for metadatasystem". Dette skal henge saman med det meir overordna arbeidet med omgrep.

Ein av grunnane til at folk med ulike arbeidsoppgåver kan snakke forbi kvarandre, er at tolkinga av sentrale omgrep avheng av føremål, ståstad i organisasjonen og detaljeringsnivå. Statistikken vert tradisjonelt publisert i form av tabellar, og fleire av omgrepa er utvikla for å karakterisere innhaldet i tabellar (sjå publiseringshåndboka, håndbok 64). Dei same omgrepa kan ha eller få eit litt anna innhald på mikronivå, det vil seie når vi karakteriserer innsamla data som vert lagde i datafiler. Ei kjelde til mistydingar kan også vere at nokre omgrep som er nært knytte til, og utvikla kvar for seg innanfor personstatistikk og makroøkonomisk statistikk (som nasjonalrekneskapen), vert tolka ulikt innanfor desse fagdisiplinane. Nokre omgrep bør difor ikkje inngå i eit felles omgrepsapparat, for andre bør bruksområdet spesifiserast. Men for dei fleste er det meir snakk om å presisere definisjon og tolking enn å endre omgrepa. I det følgjande har vi freista å gjere dette.

Om definisjonane

I arbeidet har vi teke utgangspunkt i internasjonale definisjonar og sedvane i SSB og elles. Mange av definisjonane er samla av SDMX-initiativet (Standard Data and Metadata eXchange) som er eit samarbeid mellom Eurostat, den europeiske sentralbanken, BIS, OECD, IMF, FN og Verdsbanken, sjå SDMX i referanselista. Definisjonane her er ofte henta frå andre kjelder. Der det er relevant, er originalkjelda gjeven saman med SDMX som kjelde, mens andre kjelder som gjev same definisjon, ikkje er tekne med.

Alternativt kunne ein ta utgangspunkt i tradisjonelle definisjonane i SSB som er dokumenterte i "Innføring i praktisk statistikk" (håndbok frå 1980). Nokre av definisjonane herfrå er tekne med i publiseringshåndboka. Desse er som regel ikkje gjevne som klart formulerte definisjonar, men kjem implisitt fram i teksten. *Det er i alle høve ikkje noko her som er i motstrid med dei internasjonale definisjonane.* Dei eksisterande handbøkene gjev ein god bakgrunn for norsk omsetjing av internasjonale omgrep. Ofte finst det synonym, og desse er gjevne i notatet.

Omgrepa som vert definerte i notatet kan grovt sett grupperast som omgrep knytte til overordna termar som statistikk, metadata, kvalitet og register, og som omgrep knytte til dei einskilde einingane statistikken byggjer på, og deira eigenskapar. Dei overordna omgrepa er definerte først. Der det er aktuelt med krysstilvisingar, er dette løyst ved lenking i det elektroniske notatet. På papir er lenkjene understreka.

Dei fleste omgrepa er gjevne i ei overskrift med utheva skrift (som det første omgrepet "statistikk"). Nokre omgrep kan lettast definerast i samanheng med andre. I ein slik samanheng kan det vere lett å

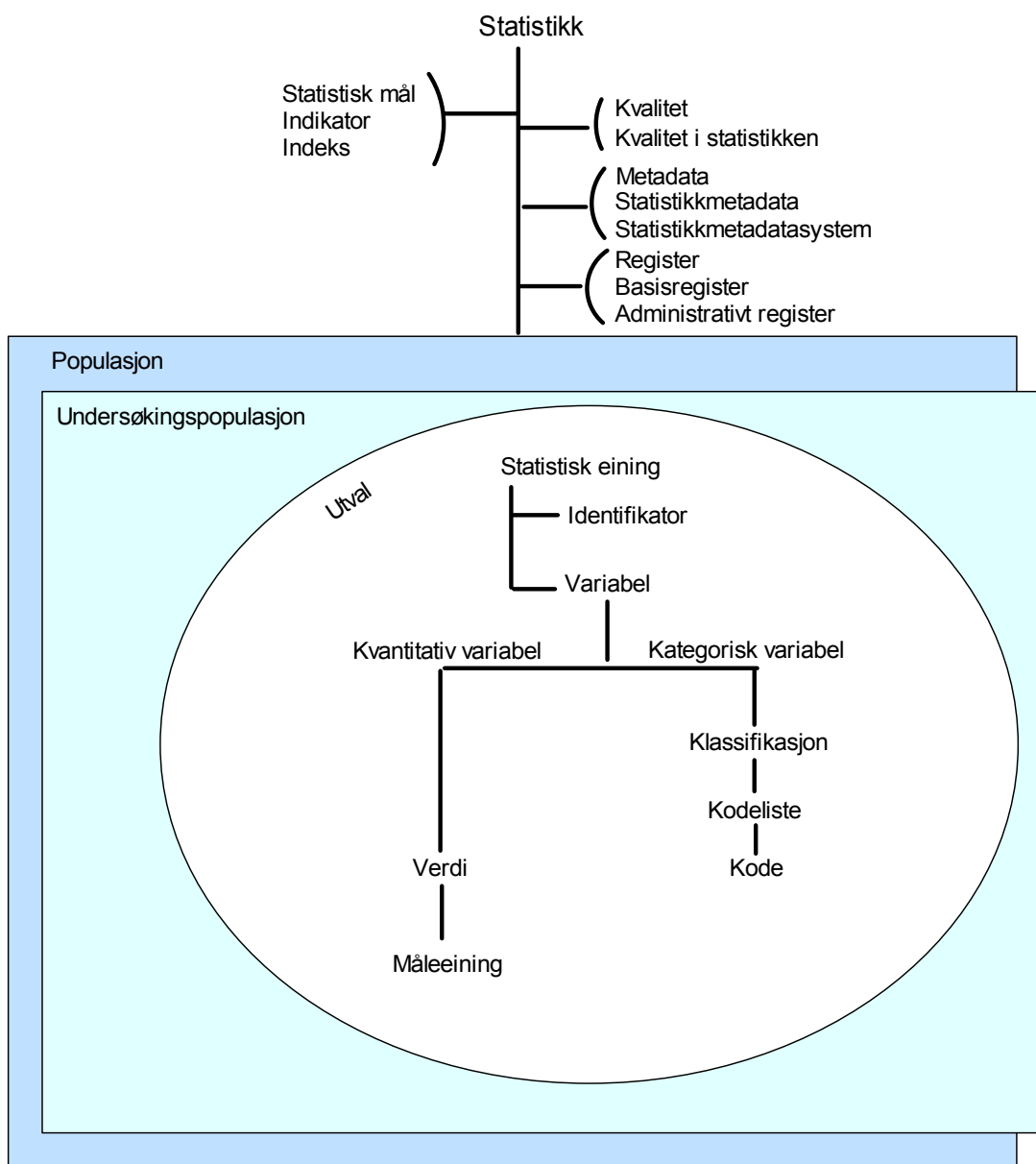
forstå kva som er meint med eit omgrep, mens det er vanskeleg å definere det presist under ei eiga overskrift (eksempel på et slikt omgrep er "data" som berre er omtala i avsnittet om "statistikk").

Definisjonane med kommentarar dekkjer to spalter. Venstre spalte inneheld sjølve definisjonen på norsk og, viss ein finn han i internasjonale kjelder, også på engelsk. Kjelda er òg gjeven. I venstre spalte finn ein elles korte eksempel og synonym. Høgre spalte er ei kommentarspalte. Her nemner vi til dømes tilgrensande omgrep som vert nytta på spesielle område. Nokre omgrep er, som nemnt i førre avsnitt, berre definerte gjennom omtale her.

Omgrepa som er med, er dei der det erfaringsmessig kan oppstå mistydingar i forhold til kva og korleis metadata skal leggjast inn i dei sentrale metadatasystema i SSB, samt nokre generelle definisjonar knytte til statistikk og metadata. Mange andre omgrep kunne vore definerte. I denne omgangen vil vi vise til SDMX som i Metadata Common Vocabulary (MCV) gjev meir enn 300 definisjonar på engelsk. Kjelder i SSB med generelle omgrepsdefinisjonar omfattar i tillegg til dei nemnde handbøkene (om praktisk statistikk og publisering), den nye revisjonshandboka (håndbok 84). I tillegg kjem meir fagspesifikke kjelder som manualar for nasjonalrekneskapen, definisjonar i kredittmarknadsstatistikken og omgrepsdefinisjonane som er med i NOS-publikasjonar. Desse omfattar i hovudsak metadata, mens dette notatet berre handlar om generelle omgrep for å karakterisere desse ("meta metadata").

Figuren viser omgrepa i den rekkjefølgja dei er definerte i dette notatet.

Figur 1. Definerte omgrep



DEFINISJONAR	KOMMENTARAR
<p>Statistikk <i>Statistikk er talfesta opplysingar om ei gruppe eller eit fenomen. Desse vert skaffa til vegar ved samanstilling og arbeid med opplysingar om dei enkelte einingane i gruppa eller eit utval av desse einingane, eller ved systematisk observasjon av fenomenet.</i></p> <p>Kjelde: Statistikklova</p>	<p>Statistikk baserer seg på, og inneber normalt ei samanstilling av <i>data</i>. I ein statistikk der til dømes den <u>statistiske eininga</u> er føretak, vil data vere verdiane som vert henta inn for kvar <u>variabel</u> (til dømes tal på tilsette) for kvart enkelt føretak. Denne typen data, som ofte vert kalla mikrodata, vert lagra i filer eller register, og arbeidde med vidare til statistikk.</p>
<p>Statistisk mål <i>Eit statistisk mål er ei samanstilling av dei (kvantitative) variabelverdiane dei <u>statistiske einingane</u> har.</i></p> <p>Kjelde: SDMX: <i>A summary (means, mode, total, index etc.) of the individual quantitative variable values for the statistical units in a specific group (study domains).</i></p> <p>Eksempel: Gjennomsnitt, medianverdi, samla tal, feilmargin, <u>indeks</u>.</p>	<p>Statistikken vert uttrykt gjennom statistiske mål.</p>
<p>Indikator <i>Ein indikator er eit mål som er avleidd av data og/eller statistikk og indikerer status eller utvikling på eit spesifikt område.</i></p> <p>Eksempel: Indikatorar for berekraftig utvikling, sosiale indikatorar, konjunkturindikatorar, strukturindikatorar (i Europa)</p>	<p>Ein indikator vil ofte vere basert på statistikk og kan vere eit tal frå ein statistisk tabell eller eit <u>statistisk mål</u>.</p> <p>Indikatorverdien må ofte samanliknast med noko (som er gjeve på førehand) for å ha meining. Informasjonsnivået knytt til ein indikator, skal vere høgt. Det vil seie at ein indikator, i større grad enn statistikken han baserer seg på, kan tolkast som kunnskap. Det er viktig at området eller temaet indikatoren skal kaste lys over, er definert presist.</p>
<p>Indeks <i>Ein indeks er eit tal som syner endringar i storleik over tid eller rom.</i></p> <p>Kjelde: SDMX og ISI (2003): <i>A quantity that shows by its variations the changes of a magnitude over time or space</i></p> <p>Eksempel: Prisindeks (som konsumprisindeksen), produksjonsindeks, jamstellingsindeks</p>	<p>Ein indeks kan vere av ulike typar. Når ein konstruerer ein indeks, er det viktig å definere kva han skal dekkje, basisperiode, vektning og metode for gjennomsnittsberekning. Verdien av ein indeks i basisperioden, eventuelt referanseområdet, vert ofte sett lik 100.</p>

DEFINISJONAR

KOMMENTARAR

Kvalitet

Kvalitet er heilskapen av eigenskapar ved eit produkt eller ei teneste som fortel om kva evne produktet eller tenesta har til å tilfredsstille krav eller behov.

Kjelde: ISO8402 og Noregs standardiseringsforbund (NS)

SDMX:

Quality is defined as the totality of features and characteristics of a product or service that bear on its ability to satisfy stated or implied needs

Det finst fleire vis å formulere ein definisjon av kvalitet på, men felles for alle er at dei tek *utgangspunkt i behova til brukarane* for dei aktuelle produkta eller tenestene. Men det er mange brukarar med ulike behov, og produsenten av ei vare eller ei teneste må ta omsyn til og balansere desse mot kvarandre. Ofte set produsenten opp eit sett med krav.

Når det gjeld kvalitet i statistikken, har dei fleste statistikkbyråa laga ei liste med faktorar som også tek omsyn til heilskapen i statistikken (sjå kvalitet i statistikken).

Kvalitet i statistikken

SSB definerer *kvalitet i statistikken* ved å vise til ulike kriterium/dimensjonar. Ein vurderer om statistikken

- *er relevant*
- *er nøyaktig*
- *er aktuell og punktleg*
- *kan samanliknast med andre og har samanheng*
- *er tilgjengeleg og klar*

Kjelde: SDMX og Eurostat (2003):

- *Relevance*
- *Accuracy*
- *Timeliness and punctuality*
- *Comparability and coherence*
- *Accessibility and clarity*

Desse dimensjonane svarer til ulike eigenskapar i den generelle kvalitetsdefinisjonen.

Heilskap/bredde i statistikken vert dekt av relevans. Samanlikning går på tid og rom, mens samanheng går på ulike statistikkar som må vere baserte på felles standardar for å kunne samanliknast og passe saman i system som nasjonalrekneskapen. Samanheng refererer også til samanhengen mellom førebels og endeleg statistikk. At statistikken skal vere klar, tyder at det skal vere lett å forstå og bruke statistikken, altså at statistikken har følgje av god dokumentasjon eller metadata.

Denne lista omfattar kvalitetskrav til *produktet* offisiell statistikk. Desse krava inngår i Eurostat sin "Code of Practice" for offisiell statistikk og statistikkbyrå. I tillegg finst det krav til *prosessane* og til *institusjonane* som produserer statistikken. Viktige krav til prosessane er:

- Statistikken skal byggje på dei beste metodane og gode rutinar
- Kostnadseffektivitet
- Låg oppgåvebyrde

Krav til institusjonelle tilhøve omfattar m.a.:

- Offisiell statistikk skal utarbeidast på eit uavhengig grunnlag.
- Datatryggleik og personvern

Faktorane som er knytte til statistikkprodukta og definerte i venstre spalte, vert ofte omtala som *produktkvalitet*. Vi snakkar elles om *prosesskvalitet* som dekkjer krava til prosessane, og *strukturkvalitet* som dekkjer institusjonelle forhold. IMF, FN og OECD opererer med litt ulike kvalitetsdimensjonar, men hovudelementa er dei same som hjå Eurostat.

DEFINISJONAR	KOMMENTARAR
<p>Metadata <i>Metadata er data om data.</i></p> <p>Kjelde: SDMX og ISO/IEC FCD 11179-1: <i>Metadata is data that defines and describes other data.</i></p> <p>Synonym: Dokumentasjon, informasjon (om data)</p>	
<p>Statistikkmetadata <i>Statistikkmetadata er strukturert informasjon om statistikk. Dette omfattar informasjon som vert brukt til å produsere, formidle, forstå, finne eller bruke statistikk (opp att).</i></p>	<p>I forhold til den generelle ISO-definisjonen av metadata har vi for statistikken inkludert eit krav om <i>struktur</i>. Struktur er naudsynt for å bruke informasjon i eit IT-system, men også ein føresetnad for effektiv bruk av metadata, til dømes for å finne statistikk på Internett blant store informasjonsmengder, og forstå han.</p>
<p>Statistikkmetadatasystem <i>Eit statistikkmetadatasystem er eit databehandlingssystem som nyttar, lagrar og produserer statistikkmetadata.</i></p> <p>Kjelde: SDMX og UNECE (2000): <i>A statistical metadata system is a data processing system that uses, stores and produces statistical metadata.</i></p>	
<p>Register <i>Eit register gjev ei (ideelt sett) komplett oppgåve over dei <u>statistiske einingane</u> innanfor ein spesiell <u>populasjon</u>, og beskriv desse med hjelp av ulike <u>variablar</u>. Alle dei statistiske einingane i eit register har ein identifikator som gjer det mogleg å oppdatere registeret med nye opplysingar om dei statistiske einingane.</i></p> <p>Kjelde: Ein kombinasjon av UNECE (2000) og definisjon fra SCB (2004)</p> <p>I Sikkerhåndboka står denne definisjonen: <i>Eit register består av oppgåver, lister, oversikter m.m. der informasjonen er lagra systematisk slik at opplysingar enkelt kan finnast igjen.</i></p>	<p>Registra til Statistisk sentralbyrå skal meldast til tryggingsrådgievar og filene skal dokumenterast i Datadok. For dette føremålet vert registerdefinisjonen i Sikkerhåndboka nytta. Datatilsynet og Brønnøysundregistra nyttar også denne definisjonen. Dei to definisjonane vi her har referert, er ulike, men står ikkje i motsetnad til kvarandre.</p> <p>Ein spesiell type register er <i>hendingsdatabasar (bokmål: forløpsdatabaser)</i>, der data er strukturerte og tilrettelagde for analyse av samanheng over tid (flyttingar, sivilstandsendingar osv.).</p>

DEFINISJONAR	KOMMENTARAR
<p>Basisregister Eit basisregister er eit <u>register</u> som definerer og identifiserer grunnleggjande <u>statistiske einingar</u>.</p> <p>I Strategi 2002- står det: Det finst i dag tre basisregister:</p> <ul style="list-style-type: none"> • Det sentrale folkeregistret i Skattedirektoratet, DSF (omfattar personar og familiar) • Einingsregisteret i Brønnøysund (juridiske einingar/føretak og bedrifter) • Registeret over Grunneigedom-, Adresse- og Bygning, GAB, i Statens kartverk 	<p>Dei tilsvarande <i>statistiske registra</i> eller <i>populasjonsregistra</i> som vi bruker i SSB er:</p> <ul style="list-style-type: none"> • BeReg (DSF pluss nokre SSB-spesifikke tilleggseiningar og variablar) • BoF: Det sentrale bedrifts-og føretaksregisteret (Einingsregisteret pluss nokre SSB-spesifikke tilleggseiningar og variablar) • SSB-GAB (GAB-registeret pluss hovudsakleg FoB-variablar) <p>Sjå elles omtale på Byrånettet under Fagleg arbeid/Register</p>
<p>Administrativt register Eit <u>register</u> som er oppretta av eit forvaltingsorgan til bruk i organet si verksemd, vert kalla eit administrativt register eller datasystem.</p> <p>Definisjonen byggjer på den meir omfattande definisjonen som står i føreskrifta som er knytt til statistikklova:</p> <p><i>Med eit administrativt datasystem forstår vi opplysingar som organ i statsforvaltinga og landsomfattande kommunale organisasjonar samlar inn og oppbevarer på ein slik måte at opplysingane kan hentast fram til bruk i organet eller organisasjonen si verksemd.</i></p>	<p>Det er vanleg at ein annan offentleg institusjon enn SSB eig denne typen register , men vi kan etter statistikklova nytte slike datasystem som grunnlag for offisiell statistikk. Ofte lagar SSB sin eigen variant av slike register ved uttrekk og/eller tillegg som grunnlag for statistikkproduksjonen. Vi snakkar då om <i>statistiske register</i> (som versjonane av <u>basisregistra</u> i SSB).</p>

Populasjon

Ein populasjon er samlinga av dei statistiske einingane statistikken skal gje informasjon om.

Ein kan skilje mellom to typar populasjonar:

- *Målpopulasjonen er populasjonen ein ønskjer å lage statistikk for*
- *Undersøkingpopulasjonen er populasjonen ein kan skaffe informasjon om i undersøkinga*

I publiseringshandboka tek ein også med at populasjonen er tid- og stadfesta.

Kjelde: SDMX og FN sitt statistiske kontor:
Population is the total membership or population or "universe" of a defined class of people, objects or events.

- *Target population is the population outlined in the survey objects about which information is to be sought*
- *Survey population is the population from which information can be obtained in the survey*

Eksempel på målpopulasjonar: Alle bedrifter i Noreg i 2000, alle adressene og eigedomane i Noreg i 2001, alle personar i Noreg i aldersgruppa 16-74 år i 2005

Tilhøyrande undersøkingpopulasjonar: Alle aktive bedrifter i BoF i 2000, alle adresser og eigedomar i GAB i 2001, alle personar i aldersgruppa 16-74 år i BeReg i 2005.

Synonym: Masse, univers

Eventuelle avvik mellom målpopulasjon og undersøkingpopulasjon gjev opphav til såkalla *dekningsfeil* i statistikken. Desse avvika kan til dømes kome av at nokre element manglar pga. forseinka oppdatering i eit register.

Frå undersøkingpopulasjonen vil ein i utvalsundersøkingar trekkje eit *utval* der dei statistiske einingane vert valde ut frå ein prosess basert på trekkjesannsyn.

Ei liste eller ein annan spesifikasjon av dei statistiske einingane som tilhøyrer populasjonen som skal teljast, eller trekkjast utval frå, vert kalla *ramma* for undersøkinga. *Ramma spesifiserer* undersøkingpopulasjonen, men har eit litt anna innhald enn denne, som er *samlinga* av dei statistiske einingane. Forskjellen er semantisk, og ein ser ofte at omgrepa undersøkingpopulasjon og ramme vert bruka synonymt. Men ramma kan i nokre samanhengar også inkludere annan informasjon som er naudsynt for å trekkje eit optimalt utval (som adressene til einingane og transportkostnader for å nå einingane).

DEFINISJONAR

KOMMENTARAR

Statistisk eining

Ei statistisk eining ber dei statistiske eigenskapane i ei statistisk undersøking eller eit register

Ei statistisk eining skal vere unik og mogleg å identifisere.

Kjelde: SDMX og UNECE (2000):

An object of statistical survey and the bearer of statistical characteristics. The statistical unit is the basic unit of statistical observation within a statistical survey.

I Nasjonalrekneskapen vert *eining*, *økonomisk objekt* og *hending* definerte som eigne omgrep. *Eining* er i Nasjonalrekneskapen definert som ein økonomiske aktør (person, selskap osv.), *økonomisk objekt* er ein verdi (vare, teneste, finansobjekt osv.), mens *hending* er noko som normalt krev ei handling utført av eininga (produksjon, ulike transaksjonar osv.).

Eksempel: Person, hushald, bedrift, føretak og vegtrafikkulykke

Ein kan skilje mellom tre typar statistiske einingar:

- *Observasjonseininga er den eininga ein mottek informasjon for*
- *Rapporteringseininga er den eininga som rapporterer inn data*
- *Analyseeininga er ei verkeleg eller konstruert eining som statistikken er utarbeidd eller formidla for.*

Ei observasjonseining er ikkje alltid "observert" av ein som samlar inn data i eller for SSB. Informasjonen ligg ofte i eit administrativt register, eller kjem via andre, til dømes kommune, fiskemottak eller formann i eit elgvald. Vi kan då snakke om ei *rapporteringseining* som kan vere forskjellig frå observasjonseininga. Eit anna eksempel er eit føretak (rapporteringseining) som gjev data for bedriftene sine (observasjonseiningar).

Kjelde: SDMX og UNECE (2000):

- *Observation unit are those entities on which information is received and statistics are compiled.*
- *Reporting unit is a unit that supplies the data for a given survey instance*
- *Analytical units represent real or artificially constructed units, for which statistics are compiled.*

Det kan vere fleire analyseeiningar knytte til eitt datagrunnlag, og observasjonseininga er ofte også analyseeining. Ei analyseeining kan òg bestå i summering over eller (berekna) oppdeling av observasjonseininga.

Ei statistisk eining må ikkje forvekslast med ei måleeining (kg, meter, osv).

Synonym for statistisk eining:
Teljeeining

Identifikator

Ein identifikator gjev ein eintydig identifikasjon av dei statistiske einingane som statistikken omfattar.

Eksempel: Fødselsnummer og organisasjonsnummer

Synonym: Ident, identifikatorvariabel

Ein identifikator vert nokre gonger kalla identifikatorvariabel. Han er likevel ikkje ein variabel i vanleg forstand, fordi poenget ikkje er å beskrive ein "naturleg" variasjon mellom statistiske einingar, eller over tid for éi statistisk eining (slik ein til dømes gjer med alder), men å gje den statistiske eininga ein verdi som skal skilje denne eininga frå alle andre einingar av same type.

DEFINISJONAR

Variabel

Ein variabel er ein eigenskap ved ei statistisk eining, og kan ta meir enn ein verdi, i form av ein numerisk verdi (kvantitativ variabel) eller ein kategori frå ein klassifisering (kategorisk variabel)

Kjelde: SDMX og FN sitt statistiske kontor:
A variable is a characteristic of a unit being observed that may assume more than one of a set of values to which a numerical measure or a category from a classification can be assigned (e.g. income, age, weight, etc. and "occupation", "industry", "disease" etc).

Eksempel: Inntekt, alder, mengd olje eller fisk, yrke, næring, dødsårsak.

Synonym: kjennemerke, eigenskap, attributt

KOMMENTARAR

I forhold til den engelske definisjonen har vi fjerna kravet om at den statistiske eininga vert observert ("unit being observed"). Vi nyttar i tillegg "verdi" i vid forstand av ordet, slik at til dømes spørsmål i skjema med avkryssingsboksar kan ha verdien 1 viss boksen er fylt ut, og verdien "ingenting" viss han ikkje er fylt ut.

Omgrepa *statistikkvariabel* og *klassifikasjonsvariabel* vert i SSB nytta i Statistikkbanken for å strukturere tabellane.

- *Statistikkvariablar tek kvantitative verdier (til dømes inntekt)*
- *Klassifikasjonsvariablar tek verdier knytte til ein klassifisering (til dømes yrke)*

Verdiane som ein klassifikasjonsvariabel kan ta, vert gjevne i ein klassifisering, som til dømes Standard for kommuneinndeling, eller berre ved eit sett kodar. Viss ein gjev opp inntekt ved å bruke ein klassifisering, til dømes høg inntekt, middels inntekt og låg inntekt, kan inntekt gå over frå å vere ein statistikkvariabel til å bli ein klassifikasjonsvariabel. Slik kan ein variabel skifte mellom "rollene" som statistikkvariabel og klassifikasjonsvariabel alt etter kva framstilling som er relevant i ein gjeven samanheng.

Som oftast vil tala i ein tabell vise verdiane for ein statistikkvariabel, mens forspalta og/eller tabellhovudet syner ein eller fleire klassifikasjonsvariablar.

Når verdien av ein variabel er eit tal, er det ofte ei måleining knytt til verdien. Ein verdi i form av ein kategori er ei nemning *eller* ein kode som svarer til denne. Når det gjeld verdien for kategoriske variablar (klassifikasjonar), kan ein gje han både som nemning (som "mann") eller som kode ("1" i dette tilfellet).

Eksempel på verdier: "300 000 kroner" (inntekt), "1 000 tonn" (oljeproduksjon), "42 år" (alder), "mann" med kodeverdien 1 (kjønn).

Mange variablar finst også som *tidsseriar*. Tidsseriar oppstår ved at ulike fenomen (variablar) vert observerte over tid.

Ein spesiell type variablar er dato- eller gyldigheitsvariablar. Dei står i samanheng med ein annan variabel (til dømes næring) og viser den datoen variabel verdien er gyldig frå.

DEFINISJONAR	KOMMENTARAR
<p>Måleining <i>Ei måleining er det ein <u>variabel</u> vert gjeve i.</i></p> <p>Eksempel på måleiningar: 1 000 NOK, Euro, tonn, liter</p>	<p>Ofte har ein òg ein faktor inkludert i måleininga slik at måleininga kan vere anten NOK eller 1 000 NOK. Måleiningar kan grupperast i ulike typar som vekt, valuta, lengd osv., og ei måleining kan reknast om til ei anna måleining så lenge dei høyrer til same type.</p>
<p>Klassifikasjon <i>Ein klassifikasjon er eit sett av diskrete, uttømmande og gjensidig utelukkande kategoriar som ein <u>variabel</u> kan ha.</i></p> <p>Kjelde: SDMX og FNs statistiske kontor: <i>A classification is a set of discrete, exhaustive and mutually exclusive observations, which can be assigned to one or more variables to be measured in the collation and/or presentation of data. The terms "classification" and "nomenclature" are often used interchangeably, despite the definition of a "nomenclature" being narrower than that of a "classification".</i></p> <p>Eksempel på klassifikasjonar: Standard for næringsgruppering (SN), Standard for yrkesklassifisering (STYRK), Norsk standard for utdanningsgruppering (NUS), Standard for familietype</p>	<p>I ein klassifikasjon vil kategoriane ofte vere identifiserte ved ein kode og alltid ved ei nemning. (Sjå eigen definisjon av <u>kodeliste</u>).</p> <p>Ein klassifikasjon kan bestå av fleire nivå (hierarkisk).</p> <p>I praksis er det <i>klassifikasjonsversjonar</i> vi nyttar i statistikk. Ein klassifikasjonsversjon er gyldig i eit visst tidsrom, og vert, når det er naudsynt, erstatta av ein ny versjon. Det er viktig med nøklar (ofte kalla korrespondansetabellar) mellom ulike versjonar slik at ein kan sjå korleis ein kategori i ein versjon har endra seg i neste versjon (til dømes at "29.100 Produksjon av jern og stål" i SN94 endra seg til "29.100 Produksjon av jern og stål samt ferrolegeringar" i SN2002).</p> <p>Det er ulik praksis når det gjeld å utarbeide nye versjonar av klassifikasjonar. For nokre av klassifikasjonane vil det vere tillate å gjere mindre endringar (til dømes leggje til element) utan at ein får ein ny versjon. Det er berre større revisjonar som gjer noko med sjølv strukturen i klassifikasjonen, som fører til ein ny versjon. Dette gjeld Norsk standard for utdanningsgruppering (NUS). I andre klassifikasjonar vil alle endringar (bortsett frå korrigering av skrivefeil o.l.) føre til ein ny versjon, til dømes fekk vi ein ny versjon av Kommuneinndelinga då 1842 Skjerstad forsvann frå inndelinga når kommunen blei innlemma i 1804 Bodø i 2005.</p> <p>Ein klassifikasjon kan vere ein formell standard klassifikasjon (som SN2002), men kan også representere ei gruppering av ein kvantitativ <u>variabel</u>, som til dømes alder (sjølv om vi freistar å standardisere på visse grupperingar) og jordbruksareal. Desse grupperingane vert ofte laga for å forenkle presentasjonen eller unngå problem med konfidensialitet.</p> <p>Omgrepet <i>standard</i> har eit vidare innhald, men i SSB bruker vi ofte <i>standard</i> som forkorting for standard klassifikasjon. Alle klassifikasjons-eksempel i venstre kolonne er standard klassifikasjonar. Desse er vedtekne av administrerande direktør, byggjer ofte på internasjonale klassifikasjonar og gjer det mogleg å samanlikne statistikken til SSB på tvers av statistikkområde.</p>

DEFINISJONAR

Kodeliste

Ei kodeliste er ei liste som ein variabel hentar verdiane sine frå.

Kjelde: SDMX: A code list is a predefined list from which some statistical concepts (coded concepts) take their values.

Eksempel på kodeliste (brukt i Mediebruksundersøkinga):

- 1 Ja
- 2 Nei
- 8 Vil ikkje svare
- 9 Veit ikkje

KOMMENTARAR

Kvar kategori i ei kodeliste har ein unik kode og ei unik nemning. Det er ikkje krav om, som for klassifisering, at kategoriane skal vere uttømmande og gjensidig utelukkande. Ei kodeliste i SSB kan vere ei liste over kategoriar som ikkje er uttømmande for eit område, men som dekkjer kategoriar som er aktuelle for ei spesiell undersøking. Desse kategoriane kan vere gjensidig utelukkande, men treng ikkje vere det.

Vi har også i spørjeundersøkingane våre ein del kodelister av typen "smørbrøddister" der kategoriane ikkje er gjensidig utelukkande, men uttømmande.

I daglegtale nyttar ein ofte kodeliste om (korte) klassifiseringar der kodane er tekne med, til dømes kodelista for kjønn (1=mann, 2=kvinne).

Referansar

Datarevisjon, Statistisk sentralbyrås håndbøker, nr. 84, 2005

Eurostat (2003): *Assessment of Quality in Statistics: Definition of quality in statistics*, WG, Luxembourg, October 2003

Eurostat (2005): *European Statistics Code of Practice*. Adopted by the SPC 24 February 2005

FNs statistiske kontor: *United Nations Glossary of Classification Terms*, publisert på http://unstats.un.org/unsd/class/family/glossary_short.htm.

Innføring i praktisk statistikk, Statistisk sentralbyrås håndbøker 44, 1980

ISI (2003): *The Oxford Dictionary of Statistical Terms* - Oxford University Press 2003

Publiseringshåndboka, Statistisk sentralbyrås håndbøker, nr. 64, 1997

SCB: Report 2004:2- Registerstatistikk - administrativa data för statistiska syften.

SDMX: *Metadata Common Vocabulary*. Edited by Marco Pellgrino (Eurostat) and Denis Ward (OECD), siste versjon oktober 2005

Sikkerhetshåndboka, Statistisk sentralbyrås håndbøker, nr. 45, 1998

SSBs Metadatastrategi - Sæbø, Andersen, Hoel, Hustoft, Linnerud, Torvbråten: SSB-rapport 2005/2

Statistikkloven av 16. juni 1989, Statistisk sentralbyrås håndbøker 54, 1989

Strategi 2002-, Planer og meldinger 2002/18 fra Statistisk sentralbyrå

UNECE (2000): *Terminology on Statistical Metadata*, Conference of European Statisticians - Statistical Standards and Studies No 53.

De sist utgitte publikasjonene i serien Interne dokumenter

- 2002/3 Kantineundersøkelse i SSB Oslo. Resultater fra fokusgrupper. 17s.
- 2002/4 Statistisk sentralbyrås organisasjon og ledelse. Juli 2002. 37s.
- 2003/1 Trivselsundersøkelsen 2002. Arbeidsmiljø og trivsel i Statistisk sentralbyrå. 54s.
- 2003/2 Bedrifts- og foretaksregisteret, Veiledning i bruk av SAKSYS. 98 s.
- 2003/3 B. Bjørlo, B. Andersen, H. Brenna, L. Solheim og J.E. Wålberg: Dokumentasjon av Landbruksundersøkelsen 2001. 73s.
- 2003/4 Statistisk sentralbyrås organisasjon og ledelse April 2003. 35s.
- 2003/5 Frigiving av statistikkfeilretting i etterkant. 21s.
- 2003/6 Overenskomst av 1. mai 2002 for intervjuere i Statistisk sentralbyrå (SSB) 19s.
- 2003/7 Teamarbeid blant individualister - om organisasjonsutvikling gjennom prosjekt - og temaarbeid. 68s.
- 2003/8 Indikatorer for produktivitet i Statistisk sentralbyrå. 18s.
- 2003/9 Statistisk sentralbyrås organisasjon og ledelse. August 2003. 34s.
- 2003/10 T. Langer. Andersen, A.B. Dahle, A.S. Abrahamsen og M.V. Moe: Temamøte mellom Danmarks Statistikk og SSB om kvalitetskontroll og indeksberegninger i utenrikshandel med varer. 82s.
- 2003/11 B. Bartsch: Kundeundersøkelse 2002. Seksjon for bedriftsregister. 39s.
- 2003/12 J.E. Skjørbotten: Proqrameksempler i SAS. Utgave 2. 46s.
- 2004/2 B. Bartsch, G. Claus, K. Henriksen, A.C. Nguyen, G-E- Wangen: Brukerundersøkelse om Bedrifts og foretaksregisteret (BoF) 2003. 72s.
- 2004/3 Trivselsundersøkelsen 2003. Arbeidsmiljø og trivsel i Statistisk sentralbyrå. 41s.
- 2004/4 K. Gaasø, S.Lien, Å. Lukerstuen og J.A. Osnes: Temaside for likestilling. 38s.
- 2004/5 Statistisk sentralbyrås organisasjon og ledelse. 34s.
- 2004/6 ILO-Arbeidskraftsregnskap, delprosjekt arbeidstid. Evalueringsrapport- med fokus på prosessene. 25s.
- 2004/8 Beregning av produktivitet i Statistisk sentralbyrå. Notat til nordisk sjefsstatistikermøte 2004. 15s.
- 2004/9 Kravspesifikasjon for elektronisk innberetning, kjennemerker og filbeskrivelse for lønnsstatistikken. Oppdatert 2004. 17s.
- 2004/10 Lønnsstatistikk for privat sektor. Dokumentasjon av datafangstrutiner 2004. 65s.
- 2005/1 Dokumentasjon av småviltjaktstatistikken. 56s.
- 2005/2 H.V Sæbø, M.Q. Andersen, T. Hoel, A.G. Hustoft, J. Linnerud og H.K. Torvbråten Metadatastrategi. Sluttrapport. 50s.
- 2005/4 G. Holta: Intern økonomiinstruks. 12s.
- 2005/5 Overenskomst av 1. mai 2004 for intervjuere i Statistisk sentralbyrå (SSB). 20s.
- 2005/6 Statistisk sentralbyrås organisasjon og ledelse. Juni 2005. 33s.
- 2005/7 G.I. Gundersen, B. Hoem, P. Løkkevik og D. Splide: Arbeidsprosesser og metoder brukt i Proteam-prosjektet "Gjennomgang av metoder og datakilder i energiregnskapet"
- 2005/8 E. Andersen, R. Dyrstad, R. Johansen og E.M. Lindahl: SSB.no som læringsressurs. 45s.
- 2006/1 Statistisk sentralbyrås organisasjon og ledelse. Februar 2006. 33s.
- 2006/2 A.G. Hustoft og H. V. Sæbø: Noen sentrale begreper knyttet til metadata - til bruk i SSBs felles metadata-systemer. 16s.