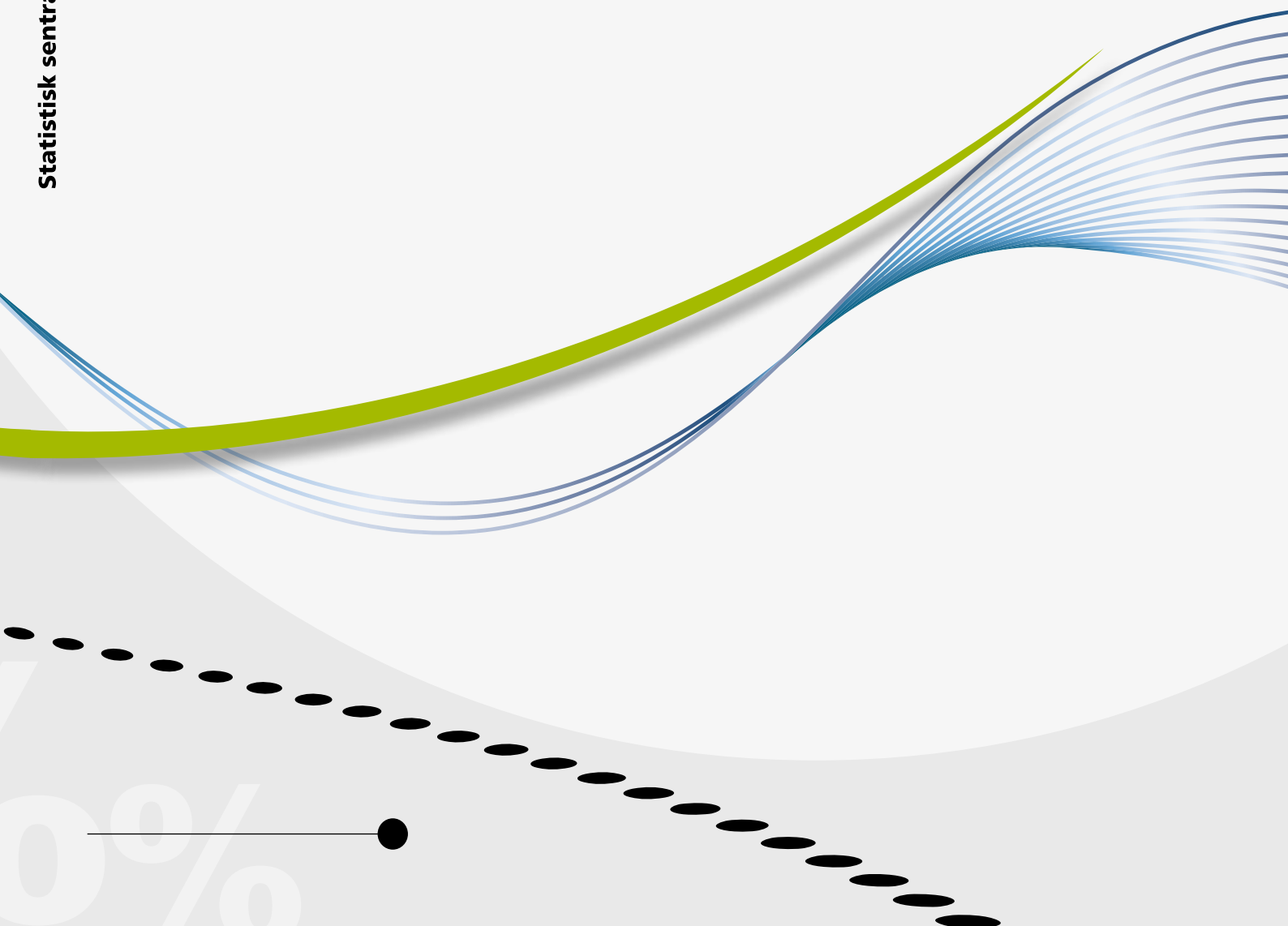
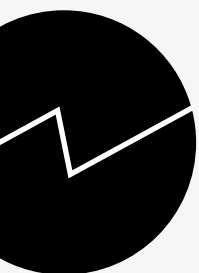


*Anne Brit Thorud, Torild Fløysvik, Dag Abrahamsen,
Harald Tønseth, Aslaug Hurlen Foss og Jon Ole
Johansen Hagemo*

System for beregning av nasjonale tall i KOSTRA del IV



*Anne Brit Thorud, Torild Fløysvik, Dag
Abrahamsen, Harald Tønseth, Aslaug Hurlen Foss
og Jon Ole Johansen Hagemo*

**System for beregning av nasjonale tall i
KOSTRA del IV**

© Statistisk sentralbyrå Ved bruk av materiale fra denne publikasjonen skal Statistisk sentralbyrå oppgis som kilde. Publisert april 2014 ISBN 978-82-537-8918-7 (elektronisk) Emne: Offentlig sektor	Standardtegn i tabeller	Symbol
	Tall kan ikke forekomme	.
	Oppgave mangler	..
	Oppgave mangler foreløpig	...
	Tall kan ikke offentliggjøres	:
	Null	-
	Mindre enn 0,5 av den brukte enheten	0
	Mindre enn 0,05 av den brukte enheten	0,0
	Foreløpig tall	*
	Brudd i den loddrette serien	—
	Brudd i den vannrette serien	
	Desimaltegn	,

Forord

Denne rapporten dokumenterer prosjektet ”System for beregning av nasjonale tall, del IV”. Prosjektet er en oppfølger til prosjektene ”System for beregning av nasjonale tall i KOSTRA” fra 2010, ”System for beregning av nasjonale tall i KOSTRA_del II” fra 2011 og ”System for beregning av nasjonale tall i KOSTRA_del III” fra 2012. Resultater og anbefalinger til videre utvikling fra de første prosjektene er dokumentert i Interne dokumenter 30/2010, Notater 50/2011 og Notater 68/2012.

Bakgrunnen for å opprette et system for beregning av nasjonale tall er at man ved publiseringene i KOSTRA (KommuneSTatRApportering) har en del frafall i rapporteringen, især til publiseringen 15. mars, slik at det er vanskelig å si noen om utviklingen i ressursbruk og tjenesteproduksjon på nasjonalt nivå. Estimerte nasjonale verdier har derfor vært etterspurt fra flere departementer og andre eksterne aktører som har behov for slike data i styringsøyemed, i tillegg til at de dekker interne behov i Statistisk sentralbyrå (SSB).

Det første prosjektet utarbeidet et automatisert system for estimering av nasjonale verdier i KOSTRA basert på en generell estimeringsmodell. Den generelle metoden passer ikke på alle typer data i KOSTRA, og i delprosjekt II ble omfanget av statistiske metoder og modeller og forklaringsvariabler utvidet, slik at flest mulig av variablene i KOSTRA kunne estimeres. Delprosjekt III hadde fokus på kvalitetskontroll av bruken av tidligere utviklede løsninger samt på teknisk videreutvikling. I dette delprosjektet har hovedfokus vært på finne beste default estimeringsmetode samt kvalitetssikring av bruken av metoder og forklaringsvariabler generelt. Planen var også å ferdigstille utvikling av det som har vært ansett som nødvendige it-rutiner. På grunn av reduksjon i tildelte it-ressurser har det vært nødvendig å utsette ferdigstillingen til et senere tidspunkt.

Prosjektgruppa har bestått av prosjektleder Anne Brit Thorud fra seksjon for offentlige finanser og prosjektdeltakerne Torild Fløysvik fra seksjon for offentlige finanser, Dag Abrahamsen fra seksjon for helsestatistikk, Harald Tønseth fra seksjon for leveårsstatistikk, Aslaug Hurlen Foss fra seksjon for statistiske metoder og Jon Ole Johansen Hagemo fra seksjon for datafangstsystemer.

Styringsgruppa har bestått av seksjonssjefene ved de berørte seksjonene. Fra seksjon for offentlige finanser Irene Arnesen (leder), fra seksjon for helsestatistikk Ann Kristin Brændvang, fra seksjon for leveårsstatistikk Elisabeth Rønning, fra seksjon for statistiske metoder Jan Bjørnstad og fra seksjon for datafangstsystemer, Ann Christine Westling.

Statistisk sentralbyrå, 5. mars 2014

Hans Henrik Scheel

Sammendrag

Rapporten dokumenterer arbeidet SSB har gjort med å videreutvikle løsningen for beregning av nasjonale tall i KOSTRA. Ved siden av de publiserte kommune- og gjennomsnittstall hadde brukerne av KOSTRA-data lenge etterlyst estimater på både nasjonalt og regionalt nivå. Dette gjelder især ved publiseringen 15. mars, da det pga manglende innsending fra mange kommuner kan være vanskelig å danne seg et bilde av hva det sannsynlige ”sluttresultatet” blir.

Arbeidet resulterte i publisering av et stort antall estimerte landstall 15. mars 2011. Før publiseringen 15. juni samme år var en del metodiske vansker blitt løst slik at enda flere estimater kunne frigis. Ved publiseringen 15. mars 2012 var det bare et lite mindretall av KOSTRA-indikatorene som ikke hadde fått estimerte landstall. Estimatenes var også beregnet tilbake til og med regnskapsåret 2008. I publiseringen 15. juni 2012 ble estimater for aggregerte grunnlagstall for fylkesgrupper publisert for fylkeskommuner, mens kommunene hadde fått tilsvarende estimater for aggregerte grunnlagstall for kommunegrupper og fylker. Også disse ble beregnet bakover til og med regnskapsåret 2008. Prosjektet hadde i tillegg klart å få tilpasset løsningen for beregning av egne gjennomsnitt slik at den kan benyttes på alle estimerte tall, bortsett fra sensitive tall og data som lastes inn ferdig estimert. 2013-publiseringen inneholdt ingen nye nyvinninger teknisk sett. Tilgangen på it-ressurser ble skåret betraktelig ned, og hovedfokus har vært på å sjekke ut om bruken av rate med konstantledd ville ha gitt bedre estimater i marspubliseringen enn rate som forhåndsvalgt estimeringsmodell.

I rapportens kapittel 1 beskrives formålet med og avgrensningen av prosjektet, mens kapittel 2 definerer noen sentrale begreper som er brukt i rapporten.

Kapittel 3 gir en kort oversikt over de estimeringsmodeller og metoder som er gjort tilgjengelig samt en oversikt over hvilke som faktisk blir benyttet.

Kapittel 4 oppsummerer erfaringene fra bistandsrunden hos de statistikkansvarlige på de gjenstående statistikkområdene, samt sammenligningen av datakvalitet ved bruk av metodene rate og rate med konstantledd på statistikkområdene kommune-helse, kultur, bolig, samferdsel og kommuneregnskap. Her beskrives også alternative løsninger til håndtering av skillet mellom null, blank og ennå ikke innsendte data, i prioritert rekkefølge, samt håndtering av estimering bakover for 2007-2001. Utvikling av løsning for usikkerhetsmål er utsatt inntil videre.

I kapittel 5 anbefaler prosjektgruppa at arbeidet med kvalitetssikring fortsetter i form av et arbeidsutvalg, og at ferdigstilling av systemet for estimering hviler ett år.

Innhold

Forord	3
Sammendrag	4
Innhold	5
1. Innledning	6
1.1. Bakgrunn og problemstilling	6
1.2. Formål og avgrensninger	7
1.3. Oversikt over innholdet i rapporten	8
2. Begreper og definisjoner	9
3. Metode	10
3.1. Estimeringsmodeller og metoder	10
4. Resultater og vurderinger	12
4.1. Kvalitetssikring av bruken av metoder og forklaringsvariabler til estimering	12
4.2. Vurdering av default estimeringsmodell	14
4.3. Håndteringen av skillet mellom null, blank og ennå ikke innsendte data	18
4.4. Estimering av sensitive data	18
4.5. Aritmetiske snitt	19
4.6. Usikkerhetsmål	19
4.7. Estimerte verdier for 2001-2007	20
4.8. Nye ssb.no og føringer for publisering av estimerte verdier	20
5. Videre utvikling	21
Vedlegg A. Prosjektskriv	22
Vedlegg B. Rate og rate med konstantledd – testing av metodene på utvalgte variabler	25

1. Innledning

I dette kapitlet beskrives bakgrunnen for prosjektet og dets målsettinger. Sist i kapitlet gis en veiledning til leserne over innholdet i de enkelte kapitlene i rapporten.

1.1. Bakgrunn og problemstilling

KOSTRA-publiseringen 15. mars er gjenstand for stor oppmerksomhet fra mange hold blant annet på grunn av sin høye aktualitet. Det er dessverre ikke fullstendig rapportering på alle områder fra alle (fylkes-)kommuner på dette tidspunktet. Det er imidlertid et stort behov for også å kunne si noe om utviklingen i ressursbruk og tjenesteproduksjon nasjonalt og på regionale nivåer så tidlig som mulig.

Det har vært jobbet med utviklingen av et system for estimering av nasjonale verdier i KOSTRA i tre tidligere prosjekter. Del I utviklet i 2010 en generell automatisert løsning med en estimeringsmetode og en forklaringsvariabel tilgjengelig, tilpasset flest mulig variabler. Løsningen ble videreutviklet gjennom del II i 2011 slik at det ble publisert estimerte nasjonale verdier for de aller fleste grunnlagsdata i KOSTRA. I prosjektets fase III ble gjennomsnittsberegninger basert på de innrapporterte dataene erstattet med estimerte gjennomsnitt.

Til prosjektets fase fire gjenstod følgende oppgaver:

Skille mellom null, blank og ikke innsendte data

Det mangler generelle løsninger for å håndtere problemene med blanke verdier i innsendte skjemaer, som egentlig skulle vært rapportert med verdien 0. Disse blir oppfattet som "ikke-innsendt" og derfor estimert, noe som gir for høye estimer. Det bør også finnes en metode for å unngå at det estimeres verdier for kommuner som ikke har en bestemt tjeneste/ordning.

Generell kvalitetssikring av bruken av metoder og forklaringsvariabler til estimering

Selv om estimeringsfunksjonaliteten er på plass, bør det jobbes videre med kvalitetssikring av bruken av metoder og forklaringsvariabler for å sikre best mulige estimer. Opptellinger viser at de aller fleste variablene estimeres ved hjelp av ratemodellen. Om dette skyldes at det ikke er behov for de andre tilgjengelige metodene, ratemodellen med konstantledd og vanlig regresjon, eller at de som skal bestemme estimeringen ikke er trygge nok på å velge en alternativ metode i forhold til standardmetoden, burde undersøkes nærmere.

Vurdering av default estimeringsmetode

Ratemodellen er den absolutt mest brukte estimeringsmodellen i KOSTRA. Den er brukt gjennomgående for alle regnskapsvariablene, og også på de fleste andre variablene i KOSTRA. Det bør vurderes om ratemodellen med konstantledd kan være en bedre modell å ha som default fordi den også fanger opp at det forekommer visse faste, indirekte kostnader og en viss produksjon uavhengig av den direkte tjenesteytende ressursbruken og tjenesteproduksjonen.

Aritmetiske gjennomsnitt

Enkelte indikatorer på nivå 2 er ikke forholdstall med en brøk som beregning. I den tidligere løsningen ble det beregnet aritmetiske gjennomsnitt for slike variabler. Det er ikke tilrettelagt for dette i systemet for estimering. Her blir disse betraktet som grunnlagsdata og de estimerte verdiene blir aggregert opp til summer på nasjonalt/regionale nivåer, noe som gjør at det ikke kan publiseres estimer for slike variabler. Det bør vurderes om og eventuelt hvordan man skal beregne aritmetiske gjennomsnitt for slike variabler.

Usikkerhetsmål

Hvor godt estimeringen treffer vil avhenge av hvor stor spredning det er på dataene som ligger til grunn for estimeringen. Usikkerheten i estimeringen vil derfor kunne

varierte etter hva slags data en ser på. Et mål på usikkerhet ville gi brukerne en pekepinn på i hvor stor grad de kan sette lit til det estimerte tallet. Variasjonskoeffisient er anbefalt fordi dette er et enkelt og intuitivt forståelig mål for brukerne.

Tidsserier

For flere fagområder finnes det mangler i datasettene for tidligere år. For å få mer sammenlignbare tidsseriedata på nasjonalt og regionale nivåer er det behov for å kunne estimere manglende verdier også for de tidligste årgangene i KOSTRA. Rent teknisk lå mye tilrette allerede, men det måtte settes av tid til kvalitetssikring av dataene.

Nye ssb.no og føringer på publiseringen av estimerte tall

Overgangen til ny publiseringsløsning på hjemmesidene til ssb.no kan gi noen føringer som publiseringen av estimerte tall må ta hensyn til. At det finnes ferdige tall nasjonalt og for regionale nivåer kan i seg selv være nyttig når tabellene skal genereres. Det kan også være at overgangen kan gi nye muligheter i publiseringsøymed. Blant annet kan nevnes ønsket/behovet for bedre å synliggjøre at dataene er estimerer, for eksempel med farge, muligheten for å publisere variasjonskoeffisient osv. Det er viktig at dette følges opp videre.

Estimering av sensitive data på lavere regionale nivåer og egne gjennomsnitt

Det publiseres ingen estimerer fra de sensitive KOSTRA-skjemaene på lavere nivå enn landet, verken for nøkkeltall eller grunnlagsdata, for skjemaene 11, 11b, 11c, 13 og 15. Fremtidig løsning må enten baseres på utvidet prikking når estimatet bare omfatter én prikket kommune, eller en løsning med tilfeldig avrunding. Med en løsning på plass for prikking av sensitive data kan også funksjonalitet for egne gjennomsnittsberegninger utvikles.

1.2. Formål og avgrensninger

Effektmål:

Prosjektets hovedmål har vært å jobbe videre med estimeringsfunksjonaliteten i KOSTRA slik at man også kan publisere estimerte tall for tidligere årganger samt for alle nøkkeltall og grunnlagsdata i KOSTRA, både nasjonalt og på lavere regionalt nivå. Dette vil være med på å oppfylle KOSTRAs målsetting om å vise sammenlignbar styringsinformasjon.

Resultatmål:

Prosjektet skulle:

- 1) Implementere løsning for håndtering av skille mellom null, blank og ikke innsendte data innen 15.3.2013
- 2) Kvalitetssikre bruken av estimeringsmetoder og forklaringsvariabler for de resterende fagstatistikkene innen 15.3.2013
- 3) Avklare om ratemodellen med konstantledd bør være default estimeringsmodell for alle statistikkene innen 15.6.2013
- 4) Avklare om det kan/bør beregnes aritmetiske gjennomsnitt og eventuelt implementere dette innen 15.6.2013
- 5) Avgjøre hvordan og ev. klargjøre for intern og ekstern visning av usikkerhetsmål til 15.3.2014
- 6) Klargjøre for publisering av estimerte verdier for årene 2001-2007 til 15.3.2014
- 7) Avklare muligheter/konsekvenser for publisering av estimerte verdier i nye ssb.no
- 8) Avklare om man skal publisere egne gjennomsnitt for sensitive data og ev. klargjøre for dette til publisering 15.3.2014.

Avgrensning

Begrenset tilgang på it-ressurser til prosjektet har gjort at ambisjonene om å kunne ferdigstille systemet for estimering ikke har kunnet realiseres i 2013. Prosjektet har derfor vurdert/avklart en del av oppgavene som gjenstod fra de forrige

delprosjektene, men ikke lagt opp til it-messig implementering av disse i denne omgangen.

1.3. Oversikt over innholdet i rapporten

Kapittel 1 beskriver bakgrunnen for og formålet med dette prosjektet.

Kapittel 2 gir en ordliste med forklaringer på de begreper og definisjoner som brukes i rapporten.

Kapittel 3 gir en kort oversikt over de estimeringsmodeller og metoder som er tilgjengelige og resultatet fra en gjennomgang/veiledningsrunde prosjektet har hatt med de statistikkansvarlige med hensyn til ”beste valg” av estimering for den enkelte variabel.

Kapittel 4 beskriver resultatene og vurderingene som er gjort i dette delprosjektet.

Kapittel 5 beskriver prosjektgruppas anbefalinger til oppfølging i 2014.

Denne rapporten omtaler ikke de tekniske løsningene for estimeringen som gjøres i KOSTRA. Det henvises til beskrivelsene i rapporten fra et av de tidligere delprosjektene, Notater 50/2011.

2. Begreper og definisjoner

I dette kapitlet forklares de begreper og definisjoner som brukes i denne rapporten.

Figur 2-1. Begreper og definisjoner

Default estimeringsmetode	Den estimeringsmetoden som er lagt inn som standard valg. I KOSTRA er dette ratemodellen.
Driller	En editeringsmodul i ISEE med muligheten for å drille mellom forskjellige aggregeringsnivåer i dataene.
Dynarev	En editeringsmodul i ISEE som brukes til editering av dataene i KOSTRA.
Ekstremverdi	Dataverdi som avviker betydelig fra andre verdier i datasettet og som kan mistenkes for å være feil.
Estimattester	Visning av utregningene/beregningene av estimatene, i faktaaark-administrasjon.
Estimering	Anslå verdi for en ukjent størrelse ut fra data ved hjelp av statistisk metode. Beregne verdien for en ukjent størrelse med data fra et utvalg.
Estimerte nasjonale tall	Anslag på verdier på variabler på nasjonalt nivå, for grunnlagsdata og indikatorer.
Faktaark	Presentasjon av variabler for sammenligning av kommunale/ fylkeskommunale tjenester og ressursbruk for utvalgte kommuner/fylkeskommuner og gjennomsnitt for landet, fylke og KOSTRA-grupper, eller selvdefinerte grupper. Grunnlagstallene har uttak som for Statistikkbanken. Også indikatorene kan alternativt hentes direkte fra Statistikkbanken.
Faktaark-administrasjon	Intern applikasjon hvor alle variablene i KOSTRA defineres og programmeres.
Forklaringsvariabel	Annet navn er avhengig variabel. En variabel som blir brukt i en modell til å forklare variasjonen i det vi ønsker å estimere.
Frafall	Enhetsfrafall: Enheter i undersøkelsesbestanden som mangler fullstendig. Partiell frafall: Enheter som er med i undersøkelsesbestanden (data er rapportert), men hvor enkelte opplysninger mangler.
Grunnlagsdata	Vil i stor grad være antall enheter/verdier summert for en periode eller per en bestemt telledato og rapportert som absolutte tall.
Imputering	Å sette inn verdier for manglende opplysninger. Man forsøker å utnytte annen informasjon til å finne rimelige verdier for de manglende variablene.
Indikatorer	Et forholdstall basert på grunnlagsdata. Publiseres ofte som prosent eller kroner per innbygger.
IPLOS	Individbasert pleie og omsorgsstatistikk.
ISEE	Integrert System for Editering og Estimering. Modulbasert teknisk rammeverkløsning i SSB for editering (revisjon) av data. Se også Driller og Dynarev.
Mellomregning	Ledd som inngår i en indikatorberegning, gjerne absoluttall. Er ofte selve grunnlagstallet.
Nivå 2 / 3	Angir faktaarkets "sider". Nivå 2 er indikatorer, oftest forholdstall. Nivå 3 er grunnlagsdata i absolutte tall.
Nøkkeltall	Det samme som indikator.
Prikkede tall	Tall som er blanket av personvern hensyn. Dette gjelder sensitive opplysninger basert på tre tilfeller eller færre (verdiene 0, 1, 2 og 3) for barneverns- og sosialdata, eller fire tilfeller eller færre i IPLOS.
Regresjonsmetoder	Statistiske metoder som brukes for å estimere verdier i statistikker som har frafall i rapporteringen.
SAS	Dataprogrammer som brukes blant annet til analyse.
Variabel	Egenskap ved de statistiske enhetene i en undersøkelse, f.eks. alder eller utgifter.
Variasjonskoeffisient	Angir hvor stor utvalgsusikkerhet det er i den estimerte totalen, ofte oppgitt i prosent.
Vekstfaktor	En størrelse som er nyttig for å regne ut hvor mye noe øker eller minker f.eks. per år.

3. Metode

I dette kapitlet beskrives kort de modeller, metoder og forklaringsvariabler som kan benyttes ved estimeringen. For mer utførlig beskrivelse henvises til rapporten fra et av de tidligere delprosjektene, Notater 50/2011.

3.1. Estimeringsmodeller og metoder

3.1.1. Regresjon

Regresjonsmetodene:

- Ratemodell uten konstantledd

Ratemodellen passer best på grunnlagstall som tenderer mot økende variasjon (eller spredning) med økende verdi for forklaringsvariabelen. Dersom folketall er forklaringsvariabel, betyr det at grunnlagstallet varierer mer for store kommuner enn det gjør for små kommuner. Man forutsetter en lineær sammenheng (dvs. en modell i form av en rett linje) mellom forklaringsvariabelen og grunnlagstallet og at linjen går gjennom null (origo). Dersom linjen forventes å gå ovenfor eller under origo, bør man i stedet vurdere å benytte metoden "Rate med konstantledd" (jfr. neste punkt).

- Ratemodell med konstantledd

Ratemodell med konstantledd baserer seg på de samme forutsetningene som den rene ratemodellen beskrevet i punktet over. Forskjellen er at her forventes linjen å avvike fra kryssing med origo på y-aksen (loddrett akse for grunnlagstallet). Det betyr eksempelvis at dersom folketall er forklaringsvariabel vil man ofte forvente at den estimerte linjen på sammenhengen mellom forklaringsvariabel og grunnlagstallet vil krysse y-aksen litt over eller under null. I en liten kommune hvor folketallet er svært lavt, ville man likevel forvente at grunnlagstallet avviker signifikant fra null som verdi. Dette kan skyldes oppstartskostnader og/eller stordriftsfordeler.

- Regresjon - enkel lineær regresjon

Forskjellen mellom rate med konstantledd og regresjon er at variasjonen i datasettet er antatt å være konstant for regresjon, mens den for ratemodellen er antatt å være økende. Med folketallet som forklaringsvariabel forutsetter regresjonsmodellen at små og store kommuner har like stor variasjon i grunnlagstallet ut fra den modellerte linjen, mens ratemodellen forutsetter stor variasjon i grunnlagstallet for kommuner med høy befolkning, og liten variasjon for kommuner med lav befolkning.

3.1.2. Andre estimeringsmetoder

Imputering

I tillegg til regresjonsmetodene er det tilbud om imputering, det vil si innsetting av tidligere års verdier i årets beregning av totalene. Hvis det er ønsket å ta hensyn til vekstfaktoren mellom to år, må statistikkansvarlige velge en regresjonsmetode med forrige års verdi som forklaringsvariabel.

Imputering henter den siste tilgjengelige rapporterte observasjonen for et gitt grunnlagstall for kommunene og summer disse verdiene sammen med de observasjonene som finnes i årets datasett. Denne siste tilgjengelige rapporterte observasjonen kan være fjorårets verdi, men den kan også hentes fra lenger tilbake i tidsserien. Finnes det ingen tall bakover i tidsserien i KOSTRA, brukes tallet blank/null for den aktuelle kommunen i det endelige estimatet.

Eksterne estimerer/imputering

Eksterne estimerer betyr at statistikkansvarlig estimerer aggregerte tall selv, enten ved noen av metodene over eller ved imputering. Disse verdiene blir lagt inn i KOSTRA som eksterne data. Ekstern imputering som metode er særlig aktuell for data med frafall på lavere nivå enn kommune, f.eks. på institusjonsnivå, eller der

en av sensitivitetshensyn må prikke data for enkelte kommuner. Seksjonen må selv produsere og eventuelt aggregere data til kommunenivå.

3.1.3. Bruk av metodene ved publiseringen i juni 2013

Det blir produsert omtrent 20 000 estimater for grunnlagstallene i KOSTRA. 93 prosent av alle grunnlagsdataene blir estimert ved hjelp av ratemodellen. Denne metoden er satt som default metode, og er derfor et enkelt valg å ta for de statistikkansvarlige. 5 prosent av tallene blir bare summert og ikke estimert. Summering benyttes på data som har fulltelling, typisk innenfor personelltall fra register, barnehage og grunnskole, men også innenfor KNNM og VAR og for noen typer data fra kultursektoren. 192 KOSTRA-tall har blitt imputert med forrige års verdi, dette brukes i hovedsak på variabler innenfor barnevern, KNNM og sosialtjenesten. Eksterne estimater er brukt på 175 tall innenfor pleie og omsorg og VAR-sektoren. Kun en variabel er estimert med rate med konstantledd og ingen benytter vanlig regresjon. Dette kan tyde på at det enten ikke er behov for alle de metodene som er tilgjengelig, eller at de som skal bestemme estimeringen ikke er trygge nok på å velge en alternativ metode. Det bør derfor jobbes videre med dette.

Metode	Antall	Prosent
Rate	19 380	92,9
Rate med konstantledd	1	0,0048
Regresjon	0	0,0
Imputering	192	0,9
Eksterne estimater	175	0,8
Summering	1 092	5,2
I alt	20 863	100,0

4. Resultater og vurderinger

I dette kapitlet beskrives erfaringene med estimering av nasjonale tall i KOSTRA så langt.

4.1. Kvalitetssikring av bruken av metoder og forklaringsvariabler til estimering

Dette kapitlet refererer gjennomgangen av estimeringsmetoder og forklaringsvariabler på de statistikkområdene som gjenstod fra forrige prosjektperiode.

Barnevern

Data

Dataene baserer seg hovedsakelig på filuttrekk fra fagsystemer i kommunene. Filene er på personnivå, med personnummer, kommunenummer, saksidentitet og variabler om saken. Kontrollprogram for filene er nå innskjerpet slik at færre feil slipper gjennom, det sparer mye tid i revisjonsarbeidet. Alle kommuner får feillister og hovedtabell som de må gi tilbakemelding på. Av sikkerhetshensyn sendes dette som rekommandert post. Statistikken skal legges om slik at sakene blir bedre datert og det lettere kan kjøres forløpsanalyser. Statistikkssystemet skal legges om fra Kires til Driller, men dette er utsatt ett år. I forbindelse med omlegging kan enkelte kommuner kan få problemer med å rapportere og det kan bli større behov for estimering for å få landstall/fylkestall.

Dataene blir summert opp til kommunenivå og lastet inn i KOSTRA-systemet. Det er fulltelling 15. juni, 15. mars er det fulltelling noen år og noen år ikke. Dette avhenger av hva slags oppgaver som blir prioritert. Null blir lastet inn som blank i systemet, dette må bli endret hvis en estimeringsmetode skal bli benyttet.

Estimering

Det er hovedsakelig valgt summering eller imputering. Variabelen 'undersøkelser i alt' var er testet ut i SAS. Verdien fra forrige år forklarte 99 prosent av verdien i år. Veksten var på omtrent 7 prosent. Det ble derfor anbefalt å bruke ratemodellen med forrige års verdi som forklaringsvariabel.

Bolig

Data

Datagrunnlaget bygger på både KOSTRA-skjema og tall fra Husbanken. Statistikken var overført til seksjon for levekår fra seksjon for helse i 2012. I KOSTRA-skjemaet er det omtrent 3 prosent frafall i juni. Ventelister er det forskjellige tolkninger av i kommunene og derfor varierende kvalitet på rapporteringen, det er derfor valgt å ikke lage landstall av dette.

Estimering

Omtrent alle indikatorene bygd på KOSTRA-skjema ble gjennomgått. På hovedtall fungerte metoden rate med folkemengde godt. For tall som var svært detaljerte fungerte ikke denne metoden særlig godt, og estimeringsmetoden ble endret til imputering. Siden det ikke var fulltelling var det ikke mulig å legge på en vekstfaktor (rate med forrige år som forklaringsmetode).

Følgende variabler ble endret fra default ratemodell til imputering etter gjennomgangen:

- Av dette, antall brukereide boliger med kommunal disposisjonsrett
- Antall boliger kjøpt eller bygget siste år
- Antall boliger solgt siste år
- Av dette, antall boliger solgt til leietakere siste år
- Antall nye boliger som følge av oppdeling av eksisterende boliger
- Antall tapte boliger som følge av sammenslåing av eksisterende boliger
- Antall avslag på søknad om kommunal bolig

- Av dette, antall avslag på nye søknader
- Antall husleiekontrakter på ett år eller kortere
- Av dette, antall husleiekontrakter kortere enn 6 md.
- Garantier for depositum til leie av bolig, antall
- Lån til depositum, antall
- Kommunal bostøtte, antall husstander som mottar støtte i alt
- Antall husstander i midlertidige botilbud, i alt
- Antall husstander i midlertidige botilbud i mer enn 3 måneder
- Antall husstander med barn i midlertidige botilbud
- Antall husstander med barn i midlertidige botilbud i mer enn 3 md.
- Antall husstander i natthjem

Nøkkeltall på nivå 3

På nivå 3 ligger noen variabler som er nøkkeltall og ikke grunnlagsdata. I utgangspunktet bør de legges på nivå 2 som indikatorer. Fagseksjonen oppfordres til å vurdere å endre dette.

- Gjennomsnittsbetrag for startlån videretildelt av kommunen, i kroner
- Gjennomsnittsbetrag for boligtilskudd til tilpasning, i kroner
- Gjennomsnittsbetrag for boligtilskudd til etablering, i kroner

Introduksjonsstønad

Det har ikke vært gjort estimering på disse tallene tidligere. Alle kommuner skal nå levere skjema for introduksjonsstønad for flyktninger. De som ikke har ordningen trenger bare å svare at de ikke har ordningen, ikke fylle ut resten av skjemaet. Disse kan kodes til å ha null på alle variablene slik at det kan skilles mellom null og uoppgitt. Til neste år kan det da velges imputering eller rate med fjorårets verdi som forklaringsvariabel.

Kommuneregnskapet

Alle mellomregningene fra kommuneregnskapet ble satt til imputering. Det løser problemet ved at det er mange kommuner som ikke har ordningen og vi ikke ønsker å estimere tall for disse. Ved bruk av imputering vil kommuner som ikke hadde flyktninger i fjor heller ikke få noen tall på dette i år. Det ble antatt at den ene kommunen som ikke har rapportert data de siste årene ikke har flyktninger slik at det er greit at det er uoppgitt på denne kommunen.

Følgende variabler ble endret fra default ratemodellen til imputering etter gjennomgangen:

- Netto driftsutgifter til introduksjonsstønad
- Brutto investeringsutgifter til introduksjonsstønad
- Netto driftsutgifter f. 275 og f. 273, art 089
- Netto driftsutgifter introduksjonsstønad, korr. for art 089
- Netto driftsutgifter, art, korr. for art 089

Skjema for introduksjonsordning

For følgende mellomregninger skal seksjonen laste inn egne estimater med imputeringsmetoden:

- Mottakere av introduksjonsstønad
- Mottakere av introduksjonsstønad i alderen 18-55 år
- Samlet utbetaling til introduksjonsstønad

For følgende mellomregninger skal seksjonen bruke summen av de som har levert:

- Gjennomsnittlig utbetaling pr. stonadsmåned
- Gjennomsnittlig deltakelseslengde (måneder)

Samferdsel

Registertall

For registertall med fulltelling skal estimeringsmetoden settes til **sum** for å få beregnet landstall.

Eksempel:

- Antall personbiler
- Motorkjøretøyer
- Europa- og riksveier, lengde i kilometer
- Fylkesvei, lengde i kilometer

Skjemadata

For km vei og antall lyspunkter er det lite endring fra år til år i en kommune.

Imputering er da et godt valg for disse variablene. Folkemengden kan forklare km vei og antall lyspunkt, men sjekk på enkelt kommuner viste at forrige års verdi lå nærmere enn rate med folkemengden som forklaringsvariabel.

Eksempler:

- Antall lyspunkter langs kommunale veier og gater
- Antall lyspunkt langs riks- og fylkeveier som kommunen betaler for
- Private veier i km det ytes kommunalt tilskudd til
- Gang- og sykkelvei i km som er et kommunalt ansvar
- Kommunale veier og gater i km med fartsgrense 40 km/t eller lavere
- Kommunal vei og gate inkl. gang/sykkelvei. Lengde i km
- Kommunal vei og gate. Antall kilometer
- Private veier i km
- Kommunal vei og gate med fast dekke. Antall kilometer
- Antall parkeringsplasser skiltet for handikappede (HC)
- Antall utstedte parkeringstillatelser for handikappede

Regnskapstall

Defaultmetode (rate med folkemengde) beholdes, unntaket er hvis regnskapstallet skal være null for enkelte kommuner/fylke. Da bør imputering brukes eller hvis fulltelling rate med forrige års verdi som forklaringsvariabel.

Fylkestall

Alle leverer skjema innen 15. juni-publiseringen, men det kan være noe som ikke er fylt inn. Estimering er derfor mest aktuelt for 15. mars-publiseringen.

Prinsippene som er nevnt for kommunetall kan også bli fulgt for fylkesskjema.

4.2. Vurdering av default estimeringsmodell

For å teste metodene er det tatt ut fire datasett i en egen testbase. Datasettene består av kombinasjonene av grunnlagstall for 2012 i mars og juni og rate med og uten konstantledd som default metode. Dette betyr at variablene som benytter en annen metode enn rate også er med i disse datasettene. Det er prøvd å ta disse ut av analysen, men det må gjøres for hånd. Det kan være at det er noen igjen i datasettene, men dette bør ikke ha noe å si siden det gir den samme forskjellen mellom mars og juni. Noen variabler er fjernet, enten fordi de ikke hadde tall i mars eller fordi de manglet verdi i et av datasettene av en eller annen grunn. For kommune-regnskapet er det bare benyttet ratemodellen. I juni er det estimert verdier for tre små kommuner (1857, 1928 og 1929) på grunn av manglende regnskaps-rapportering. For tjenestetallene er det ukjent hvor mange tall som mangler på det enkelte området per 15. mars.

Forskjell mellom mars- og junipublisering for de to modellene

Et vanlig mål på forskjellen mellom metoder er RMSE (root mean square error). Den summerer forskjellen mellom estimatene kvadrert. Problemet med dette målet er at hvis avviket er stort, men lite i prosent, vil det gi et stort utslag på målet. Det er derfor laget et samlemål til kalt RSPE (root squar percent error- roten til den kvadrerte prosentvise feilen) for å finne ut hvilken metode som er best. Denne måler den metoden som har minst prosentvis feil. Det vil si at små og store tall

teller likt i denne metoden, fordi vi måler den prosentvise endringen. Dataene i KOSTRA kan også være på forskjellige skalaer, for eksempel areal og kroner, av den grunn er prosentvise forskjellen et bedre mål. Den metoden som har lavest verdi på målet RSPE er den metoden som har den minste prosentvise forskjellen mellom mars og juni tallene. Et problem med dette målet er små tall ofte kan ha stor prosentvis endring. Et annen svakhet med metoden er at det kan være noen få variabler som påvirker mye og som utgjør det meste av forskjellen mellom metodene rate og rate med konstantledd. Enkelte variabler som var ekstreme er derfor fjernet fra analysen.

E_{juni} - Estimert tall i juni, enten rate eller rate med konstantledd

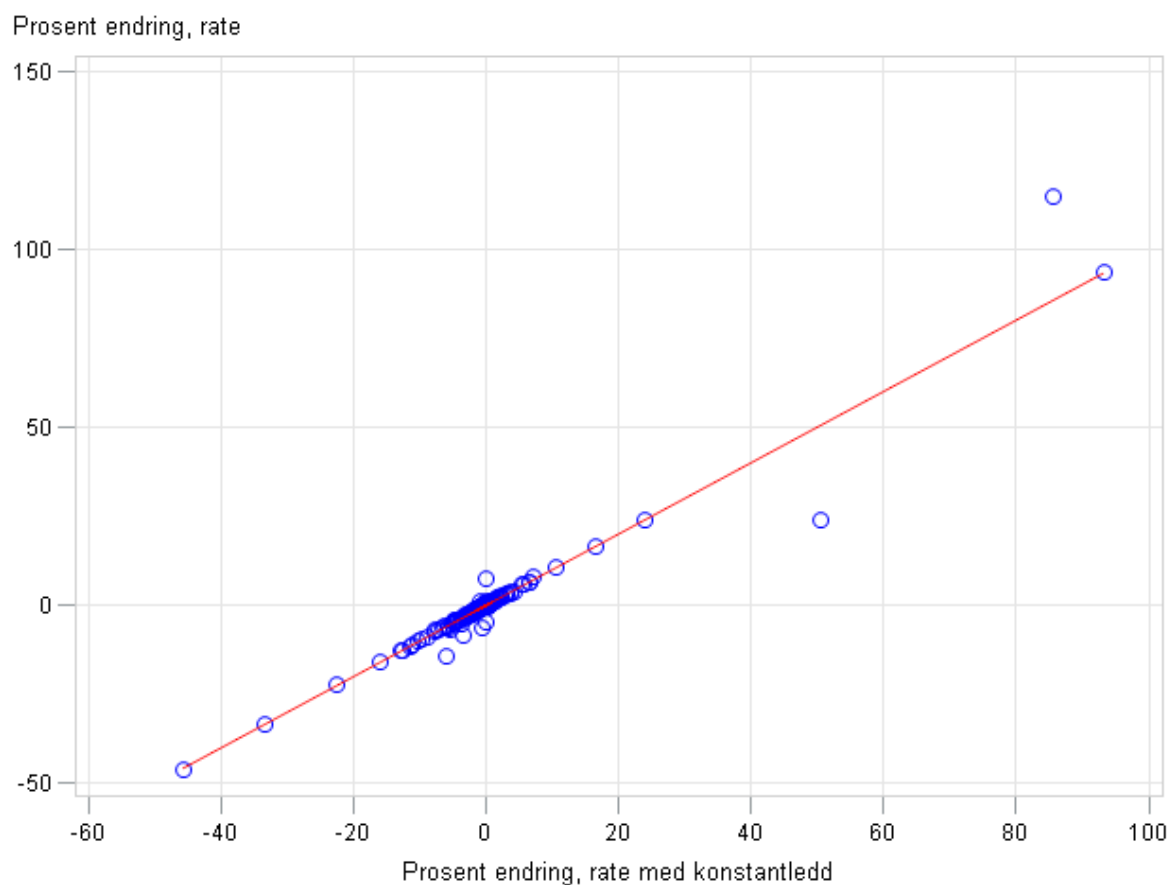
E_{mars} - Estimert tall i mars enten rate eller rate med konstantledd

$$RMSE = \sqrt{\sum \left(\frac{(E_{\text{mars}} - E_{\text{juni}})^2}{n} \right)}$$

$$RSPE = \sqrt{\sum \left(\frac{(E_{\text{mars}} - E_{\text{juni}}) * 100}{E_{\text{juni}}} \right)^2}$$

	Antall	RSPE		RMSE	
		Rate	Rate med konstantledd	Rate	Rate med konstantledd
Alle variabler som er med i testen	342	173	160	3 141 785	3 159 366
Alle økonomiske variabler	266	171	159	3 595 398	3 616 209
Ikke-økonomiske variabler	77	25	16	73 273	80 157
4.1. Konsern - Eiendomsforvaltning for utvalgte kommunale formålsbygg	67	9,4	9,7		
A1. Konsern - Finansielle nøkkeltall og adm., styring og fellesutg.	142	126,3	123,2		
E1. Konsern - Kommunehelse - nivå 3 etter Variabel, Kommune og Tid	49	41,9	42,0		
K1. Konsern - Kultur - nivå 3 etter Variabel, Kommune og Tid	46	13,2	13,6		
M1. Konsern - Samferdsel - nivå 3 etter Variabel, Kommune og Tid	12	3,8	3,9		
N1. Konsern - Bolig - nivå 3 etter Variabel, Kommune og Tid	27	11,0	8,2		

Figur 5.2.1. Prosentvis endring mellom mars og juni plottet for rate modellen mot rate med konstantledd, alle variabler som er testet



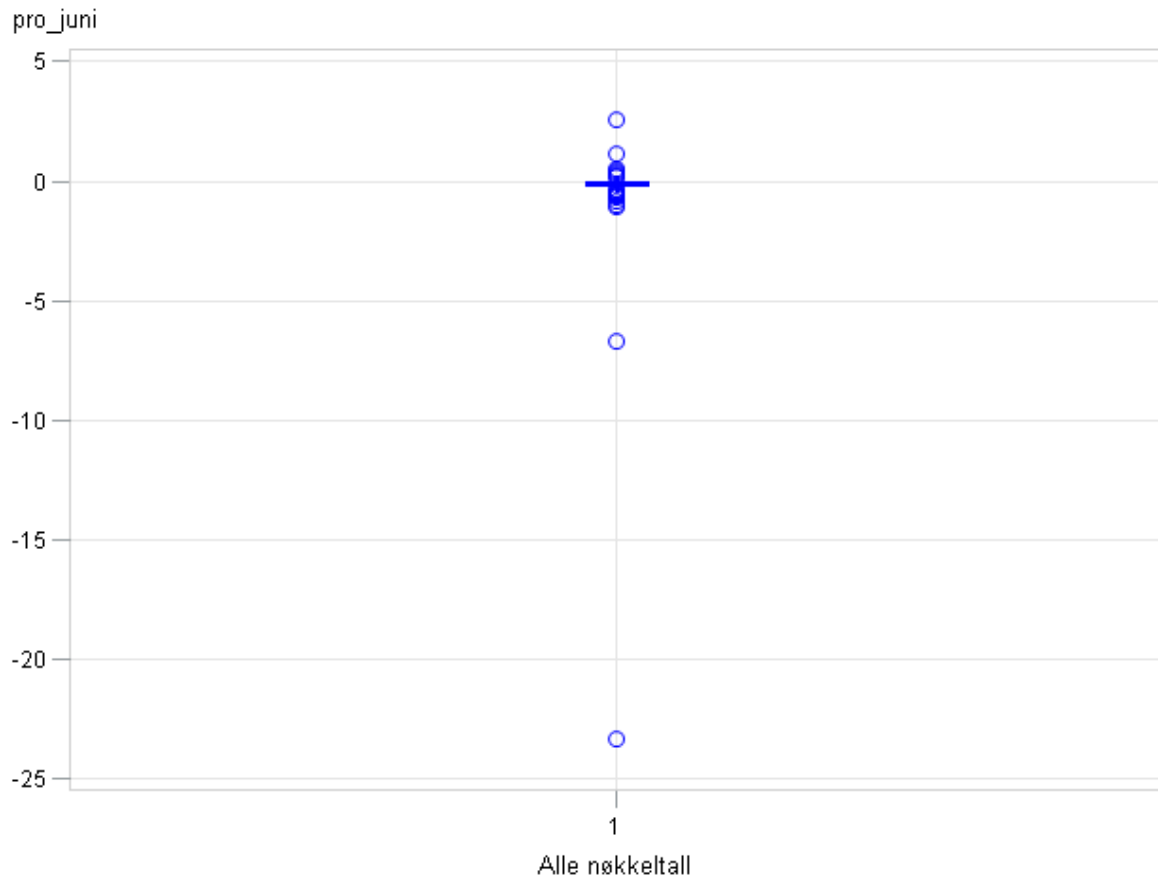
Forskjellen i tallene for juni mellom rate og rate med konstantledd er i snitt -0,2 prosent. Det vil si at hvis vi hadde valgt rate med konstantledd så ville tallene blitt i snitt 0,2 prosent høyere enn det som er publisert.

Prosentvis avvik mellom metodene i juni blir regnet ut på følgende måte:

$$\frac{(E_r - E_{rk}) * 100}{E_r}$$

E_r - Estimert tall rate, juni

E_{rk} - Estimert tall rate med konstantledd, juni

Figur 5.2.2. Boksplot av forskjellen mellom rate og rate med konstantledd for juni tallene

Konklusjon

- For de fleste variablene er det en liten forskjell i estimeringen mellom metodene, rate og rate med konstantledd.
- Rate med konstantledd gir generelt litt bedre estimerer enn rate-modellen når vi måler marstall mot junitall.
- Gjennomgangen av estimering på fagområdene viste imidlertid at for mange nøkkeltall var imputering å foretrekke fremfor ratemodeller med folkemengde som forklaringsvariabel. Grunnen til dette var:
 - Små tall med en dårlig sammenheng med folkemengde
 - Noen tjenester (plikter) fins bare i enkelte kommuner/ fylkeskommuner. En modell med folkemengde som forklaringsvariabel gir da en dårlig tilpassing. Utgifter til fergetrafikk og mottakere av introduksjonsstønad for flyktninger er eksempler på tjenester som ikke alle kommuner/fylkeskommuner har.
 - Imputering med en vekstfaktor (rate med forrige års verdi) ble også i noen tilfeller vurdert, men forkastet på grunn av manglende fulltelling forrige år/ null-missing problemer. Imputering med siste observert verdi var god nok metode for foreløpige tall.
- For økonomitall har tidligere analyser vist at en modell med forrige års verdi som forklaringsvariabel gir en bedre tilpassing enn folkemengden. Det er også en slik modell som er anbefalt av en representant fra forskningsavdelingen. Gitt at det blir fulltelling bør man heller gå over til å bruke forrige års verdi som forklaringsvariabel enn å bytte modell til rate med konstantledd.
- Estimaten og estimeringsmetoder bør bli kontrollert kontinuerlig. For enkelte tall er det en stor forskjell mellom verdiene i mars og juni. Dette kan skyldes både editering av verdier eller metodevalg for estimering.

Estimatene i mars bør bli tatt ut og sammenlignet med juni-tallene. De største forskjellene bør så undersøkes nærmere for å finne ut hva årsaken(e) kan skyldes. Enkelte tall i mars er også så usikre at de bør vurderes prikket. Det kan for eksempel gjelde investeringer på små tjenestoområder og/eller etter KOSTRAgruppe/fylker.

4.3. Håndteringen av skillet mellom null, blank og ennå ikke innsendte data

Utfordringen består i at verdier som egentlig skulle vært satt til null, i stedet blir oppfattet som "ikke innsendt" og derfor gitt en estimert verdi større enn null. Det foreligger tre alternative løsninger:

Før/ved rapportering av skjema:

Legge inn absolutt krav/varsel om utfylling av felt i skjema – aktuelt for felt hvor det er spesielt viktig å vite om svaret er 0 eller et tall.

Legge inn automatikk i skjema med "0" som default i alle utfylte felt

I revisjonssystemet:

Via funksjonalitet i ISEE kan blanke felt i innrapporterte skjema defineres som verdien 0.

I faktaark-administrasjon:

Formel i faktaark-admin som generelt setter utfylte felt til verdien 0 på det ønskede skjemaet. Dette alternativer kan især være aktuelt for gamle årganger hvor data ikke ligger i ISEE.

Prosjektgruppa har konkludert med at den beste løsningen antagelig var å håndtere problematikken allerede ved innrapporteringen. Prosjektgruppa gikk derfor ut med en oppfordring til alle statistikkansvarlige om å være bevisst på valg av løsning allerede i utformingen av skjemaet.

De statistikkansvarlige ble bedt om å vurdere å legge inn krav om utfylling av de viktigste feltene i skjemaet. Dette må vurderes opp mot faren for økt oppgavebyrde/ forsinket innsending (ved at ikke-komplette skjema blir nektet innsendt). Et annet alternativ er å preutfylle alle feltene med verdien 0 slik at disse verdiene beholdes i de feltene hvor oppgavegiver ikke taster inn en verdi. Ulempen her kan være et "rotete" skjema med eventuell økt fare for feilregistrering, og kanskje kan det også skape en viss motvilje mot utfylling.

De statistikkansvarlige ble bedt om å vurdere de forskjellige hensynene mot hverandre, og dette er blitt gjort ved utformingen av skjema for 2013. For skjemaene 7 og 8 har vurderingen for 2013 vært at det er bedre å legge inn 0 i etterkant, enten i ISEE eller i faktaarkadmin.

4.4. Estimering av sensitive data

P.t. publiseres ingen estimer fra de sensitive KOSTRA-skjemaene 11, 11b, 11c, 13 og 15 på lavere nivå enn landet, verken nøkkeltall eller grunnlagsdata. Det er heller ikke tilrettelagt for å lage egne gjennomsnitt for slike data. Slik publisering må enten baseres på utvidet prikking når estimatet bare omfatter én prikket kommune, eller en løsning med tilfeldig avrunding.

Oppfatningen i prosjektgruppa har vært at det bør tas et prinsipielt standpunkt til om KOSTRA skal kunne benytte avrundede tall for å unngå sensitivetsproblemer (i KOSTRA og internt i SSB). Det faller klart utenfor rammene for dette prosjektet å vurdere alle effektene av en slik endring, og eventuelt utvikle dette til en generell løsning.

Seksjon for helsestatistikk vil teste en metode for ”random” innsetting av 0 og 5 for alle tall fra 0 til 5. Selv om metoden skulle være enkel å få på plass, vil det ta tid å vurdere resultatet. Metoden må ha godkjenning fra alle fagområder før den lanseres som foreslått valg av metode i KOSTRA. Ulempen med metoden er at kommunen kan finne igjen et annet tall enn det de har sendt inn (2 kan for eksempel bli til 5).

Dersom ovennevnte metode ikke velges, bør det vurderes et system for ”utvidet prikking” for at dagens situasjon med totalt fravær av summer på lavere nivå ikke skal vedvare. For eksempel ved at nest laveste verdi også prikket dersom en verdi som inngår i en sum må prikket, slik at en ikke kan regne seg fram til prikket verdi på grunnlag av summen. Ulempen her er at ikke-sensitive verdier (over 2/3/4), vil bli rammet av utvidet prikking og dermed – til forvirring for dem selv og andre – kan en kommune med i utgangspunktet ikke-sensitive verdier - bli stående uten tall. Dette kan antagelig skape minst like stor forvirring som opplegget med en tilfeldig innsetting av 0 og 5 som er begrenset til de kommunene som er konkret berørt av små tall.

Uansett valg av metode kan det få konsekvenser for vurderingen av hvor mange indikatorer det bør publiseres der mange kommuner har små tall, og der et ”skjulings”-system derfor må settes i sving.

Det ser ut for at det kan trekke ut i tid å få etablert en felles løsning.

4.5. Aritmetiske snitt

Nøkkeltallene i KOSTRA har i utgangspunktet veide gjennomsnitt. Det vil si at den enkelte mottaker/elev/ innbygger teller like mye, og at store kommuners tall derfor veier tyngre enn små kommuners i snittet. Men i KOSTRAS nøkkeltallstabeller finnes det også absolutte tall, for eksempel gebyrsatser. For disse ble det tidligere laget en løsning for aritmetisk snitt, dvs. sum av antall kroner delt på antall kommuner som hadde rapportert. Innføringen av estimeringsfunksjonaliteten har gjort det nødvendig å lage en ny løsning for beregning av aritmetiske snitt. Prosjektet har som mål å få denne på plass. For å sikre driften og gjøre det mulig å beregne snitt for tall som bygger på et innfløkt hierarki er en ny server gjort tilgjengelig. Ferdigstillelse av løsningen vil imidlertid måtte konkurrere med andre viktige prosjekter, som KOSTRA-publiserings, om it-ressurser.

4.6. Usikkerhetsmål

Det er tidligere anbefalt å legge inn variasjonskoeffisient som mål på usikkerheten i estimatene. Variasjonskoeffisienten tar bare hensyn til usikkerhet pga frafall og hvor størrelsen på frafallet da er viktig. Det unike med KOSTRA er at det ofte er fulltelling i juni.

Variasjonskoeffisient er et modellbasert teoretisk begrep. For intern bruk (for de statistikkansvarlige) kan dette være et godt mål. For eksterne brukere vil sannsynligvis det mest interessante være hvor treffsikre estimatene til 15. mars publiseringen er. Det vil være mulig ut fra en analyse av verdiene i mars og juni forrige år og tidligere år å lage et mål på generelle/normale feilmarginer i 15. mars tallene. For eksempel at det er normalt at driftsutgifter varierer med +/-1 prosent, investeringer med +/-5 prosent og investeringer etter fylke med +/-20 prosent osv.

Det har tidligere vært anbefalt fra prosjektgruppa å vise usikkerhetsmålet i estimat-testeren. Det er imidlertid få som bruker estimat-testeren. De fleste kontrollerer trolig de estimerte verdiene mot forrige år. Det vil derfor være mye mer nyttig å få det ut som en tabell per faktaark eller lignende som statistikkansvarlig kan se gjennom. Det er ikke tatt stilling til hvordan usikkerheten best kan vises for eksterne brukere.

4.7. Estimerte verdier for 2001-2007

Dagens nøkkeltall har gjennomsnitt basert på innrapporterte data for årene fram til og med 2007. Estimerte landstall/summer på grunnlagsdata og snitt basert på estimatene finnes fra og med 2008. Enkelte brukere har etterlyst estimerer bakover i tid og prosjektets mål har vært å få dette på plass for årgangene bakover til og med 2001.

KOSTRAs faktaark /tabeller inneholder imidlertid mange ulike typer data. For enkelte vil svarprosenten være høy og dataene konsistente over tid, slik at det vil være greit å estimere bakover. For andre vil de tilgjengelige metodene for estimering ikke kunne benyttes uten videre. For eksempel blir dataene på vann, avløp og avfall/renovasjon (VAR) estimert lokalt og lastet inn i tabellene. Statistikkansvarlige går nøye igjennom de ulike typene data og kvalitetskontrollerer på detaljert nivå. Tilbakemelding fra disse fagområdene er at en kvalitetskontroll bakover vil kreve timeverk de i dag ikke har. En tilsvarende melding kommer også fra andre fagområder.

Gjennomsnitt for nøkkeltall bakover i tid bygger på data publisert i juni i regnskapsåret +1, med eventuelle endringer gjort i etterkant. For de fleste fagområder vil svarprosenten være høy, kun få og små kommuner er estimert. Gjennomsnitt basert på mottatte data vil derfor ikke skille seg stort ut ifra snitt dersom de hadde vært beregnet på estimerte tall.

For grunnlagsdataene vil valg av løsning for estimering innebære forskjell fra den løsningen som benyttes i dag: For data som har snitt basert på/bygger på innrapporterte data vil det være unaturlig å vise landstall på grunnlagsdata pga. mangler i dataene. For data som estimeres vil manglende data være beregnet for nettopp å kunne vise landstall. For data som har fulltelling, for eksempel grunnskole- og barnehagedata, bør det være greit å benytte estimeringsløsningen til å få sumtall på de ulike regionale nivåene.

Det åpnes for forskjellsbehandling med hensyn til bruk av estimering bakover for de ulike fagområdene, slik at enkelte beholder snitt for nøkkeltall basert på rapporterte data og fortsetter uten sumtall på grunnlagsdataene, mens andre benytter seg av estimeringsløsningen og får snitt basert på estimerte tall og summer på sine grunnlagsdata.

4.8. Nye ssb.no og føringer for publisering av estimerte verdier

Overgangen til ny publiseringsløsning av hjemmesidene til ssb.no kan gi noen føringer som publiseringen av estimerte tall må ta hensyn til. At det finnes ferdige tall for nasjonalt og regionale nivåer kan i seg selv være nyttig når tabellene skal genereres. Det kan også at være at overgangen kan gi nye muligheter i publiseringsøyemed. Blant annet kan nevnes ønsket/behovet for bedre å synliggjøre at dataene er estimerer, for eksempel med farge, muligheten for å publisere variasjonskoeffisient osv. Det er naturlig at dette fanges opp av SSB-prosjektet "KOSTRA publisering".

5. Videre utvikling

På grunn av andre store it-prosjekter i SSB er det gitt at prosjektet ”Beregning av nasjonale tall” ikke vil få tildelt it-ressurser i 2014. Ferdigstilling av systemet for estimering vil derfor hvile ett år. I stedet opprettes et arbeidsutvalg som fortsetter arbeidet med kvalitetssikring av metodevalg og andre faglige problemstillinger knyttet til estimering, som ikke krever it-ressurser.

Vedlegg A. Prosjektskriv

Prosjektskriv

Prosjektnavn	System for beregning av nasjonale tall i KOSTRA_del IV		
Kortnavn	Nasjonale tall_IV		
Planlagt startdato	21.01.2013	Planlagt sluttdato	10.11.2013
Oppdragsgiver	KOSTRAS styringsgruppe v/Olav Ljones		
Ansvarlig enhet	Seksjon for offentlige finanser v/Irene Arnesen		
Er prosjektet del av et større hovedprosjekt?	☐ Ja	☒ Nei	
Hvis ja, oppgi navn og projektnr på hovedprosjekt			
Prosjektleder	Anne Brit Thorud		
Kvalitetslos			
Produktnummer	3632		
Godkjent dato		Av	

A. Prosjektbeskrivelse

Beskriv kort
hvordan/ hvorfor
prosjektet har oppstått

Angi forankring i
strategier og vedtak.

Bakgrunn for prosjektet - Problembeskrivelse

KOSTRA-publiseringsen 15. mars er gjenstand for stor oppmerksomhet fra mange hold blant annet på grunn av sin høye aktualitet. Det er dessverre ikke fullstendig rapportering på alle områder fra alle (fylkes)kommuner på dette tidspunktet. Det er imidlertid et stort behov for også å kunne si noe om utviklingen i ressursbruk og tjenesteproduksjon på nasjonalt nivå så tidlig som mulig.

Det har vært jobbet med utviklingen av et system for estimering av nasjonale verdier i KOSTRA i tre tidligere prosjekter. Del I utviklet i 2010 en generell automatisert løsning med en estimeringsmetode og en forklaringsvariabel tilgjengelig. Løsningen ble videreutviklet gjennom del II i 2011 slik at det ble publisert estimerte nasjonale verdier for de aller fleste grunnlagsdata i KOSTRA. I 2012 ble gjennomsnittsberegningene som tidligere var basert på innrapporterte data erstattet av estimerte gjennomsnittsberegninger for årene 2008..2011. I tillegg ble det publisert estimerte verdier på grunnlagsdataene både på nasjonalt og regionale nivåer for de samme årene.

Det gjenstår fremdeles noe arbeid før vi har et fullstendig og kvalitetssikret system med estimerte verdier. Med dette delprosjekt_IV håper vi å komme et stykke videre i arbeidet med estimering.

Det var opprinnelig tenkt at del_IV skulle ferdigstille arbeidet med estimering i 2013. Det viser seg at knapphet på it-ressurser gjør dette umulig. Se nærmere beskrivelse av oppgavene som ønsket gjort i del_IV i punktet om Faglige problemstillinger.

Effektmålene skal angi de langsiktige virkningene av en vellykket gjennomføring. Hvilke forbedringer/ nytte vil komme som et resultat av prosjektet f. eks i et 3-5 års perspektiv.

Beskriv presist hvem som er målgruppen(e).

Antall effektmål bør være fra én til fire. Bruk gjerne **bidra** til foran målene.

Effektmål

Prosjektet skal bidra til at man kan publisere estimerte tall for tidligere årganger samt for alle nøkkeltall og grunnlagsdata i KOSTRA, både på nasjonalt og lavere regionalt nivå. Dette vil være med på å oppfylle KOSTRAS målsetting om å vise aktuell, relevant, pålitelig og sammenlignbar styringsinformasjon.

På sikt vil dette bidra til:

For SSB:

- Økt omfang av publiserte estimerte nasjonale verdier
- Høyere aktualitet på nasjonale verdier
- Mer helhetlig publisering av grunnlagsdata og nøkkeltall i KOSTRA
- Bedre kvalitet på landstallene (kan til dels erstatte dagens beregninger av landsgjennomsnitt)
- Mer brukervennlig statistikk
- Økt relevans på statistikkene

For fagdepartementene og andre, som f.eks.TBU

- Raskere tilgang til nasjonale verdier
- Økt relevans på statistikkene
- Mer brukervennlig statistikk

For kommuneadministrasjonen:

- Økt relevans på statistikkene
- Økt nytteverdi av styringsinformasjonen i KOSTRA

For innbyggere/media:

- Mer brukervennlig statistikk

Beskriv de mål som skal realiseres i løpet av prosjektperioden, altså hvilke leveranser prosjektet er ansvarlig for å frembringe. Vær konkret og kvantifiser gjerne. Resultatmålene skal i sterkest mulig grad gi tillit til at effektmålene blir nådd på lengre sikt. Antall resultatmål bør være fra ett til fire.

Resultatmål

Prosjektet skal:

- 1) Implementere løsning for håndtering av skille mellom null, blank og ikke innsendte data, innen 15.03.2013
- 2) Kvalitetssikre bruken av estimeringsmetoder og forklaringsvariabler for de resterende fagstatistikkene innen 15.03.2013
- 3) Avklare om ratemodellen med konstantledd bør være default estimeringsmodell for alle statistikkene bortsett fra regnskap, innen 15.06.2013
- 4) Avklare om det kan/bør beregnes aritmetiske gjennomsnitt og ev.implementere dette innen 15.06.2013

Hvis tid og it-ressurser er tilgjengelige høsten 2013 (innen prosjektslutt 10.11.2013):

- * Avgjøre hvordan og ev. klargjøre for intern og ekstern visning av usikkerhetsmål til 15.3.2014
- * Klargjøre for publisering av estimerte verdier for årene 2001..2007 til 15.03.2014
- * Avklare muligheter/konsekvenser for publisering av estimerte verdier i nye ssb.no
- * Avklare om man skal publisere egne gjennomsnitt for sensitive data og ev. klargjøre for dette til publisering 15.03.2014

Beskriv kort faglige forutsetninger og problemstillinger, med vekt på risiko. Metode eller løsningsforslag kan også kort beskrives her (detaljer senere i prosjektplan).

Faglige problemstillinger

- Skille mellom null, blank og ikke innsendte data
 - unngå at det estimeres verdi for kommuner som ikke har en bestemt ordning.
 - unngå feiltolkning av blanke felt i skjemaer som er rapportert, slik at ikke den estimerte verdien blir for høy.
- Kvalitetssikre bruken av metoder og forklaringsvariabler til estimering

De aller fleste variablene estimeres ved hjelp av ratemodellen. Det bør undersøkes nærmere om dette skyldes at det ikke er behov for de andre tilgjengelige metodene eller om de som skal bestemme estimeringen ikke er trygge nok på å velge en alternativ metode.
- Gjennomsnittsberegninger for indikatorer som ikke har brøk som formel
 - avklare om og ev. hvordan man kan beregne aritmetiske gjennomsnitt
- Nye ssb.no og føringer på publiseringen av estimerte tall

Overgangen til nye ssb.no kan gi noen føringer/muligheter for publiseringen av estimerte tall. Vil det bli mulig å synliggjøre bedre at dataene er estimer, f..eks med farge, vil det bli mulig å vise variasjonskoeffisienten grafisk osv.

Angi dokumenter, rapporter eller annet materiale som inneholder viktig informasjon for prosjekt-deltakerne. Det kan være metodebeskrivelser, forprosjektrapporter eller utredninger, vedtak, instruksjoner eller mandat fra ledelsen.

Grunnlagsdokumentasjon, aktuelle linker og andre prosjekter

Notater 46/2003 Foreløpige landstall i KOSTRA. Prinsipper, metoder, produksjon og eksempler
Referat fra møte i styringsgruppa for KOSTRA 19.03. 2009
ath/iae, 05. mars 2009, Notat: Estimering av nasjonale tall (KOSTRA)
Interne dokumenter 30/2010 System for beregning av nasjonale tall i KOSTRA
Interne dokumenter 31/2010 Sluttrapport fra prosjektet System for beregning av nasjonale tall i KOSTRA
Notater 50/2011 System for beregning av nasjonale tall i KOSTRA_II
Interne dokumenter 24/2011 Sluttrapport for prosjektet System for beregning av nasjonale tall i KOSTRA_II
Notater 68/2012 System for beregning av nasjonale tall i KOSTRA del III
Interne dokumenter 35/2012 Sluttrapport for prosjektet System for beregning av nasjonale tall i KOSTRA_III

Vedlegg B. Rate og rate med konstantledd – testing av metodene på utvalgte variabler

Fra faktaark A1. Konsern - Finansielle nøkkeltall og adm., styring og fellesutg.

	Juni 2012. Rate m.k.	Mars 2012. Rate m.k.	Juni 2012. Rate	Mars 2012. Rate	Rate m.k. forskjell mars- juni. Prosent	Rateforskjell mars-juni. Prosent	Forskjell mellom rate og rate m.k. Juni. Prosent
Netto driftsutgifter, bolig, konsern	-17 430	-32 325	-14 137	-30 386	85,5	114,9	10,8
Disposisjonsfond, konsern	20 826 012	40 227 192	20 805 568	40 235 755	93,2	93,4	0,1
Salgs- og leieinntekter, tjenester utenfor ordinært kommunalt ansvar, konsern	193 657	104 718	192 509	103 725	-45,9	-46,1	1,1
Netto driftsutgifter, kirke, konsern	2 814 760	2 811 237	2 809 842	2 811 100	-0,1	0,0	0,2
Brutto investerings- utgifter, grunnskole- opplæring, konsern	9 191 285	9 179 309	9 187 982	9 186 048	-0,1	0,0	0,0
Mva-kompensasjon påløpt i investerings- regnskapet, konsern	4 994 633	4 986 744	4 989 747	4 988 763	-0,2	0,0	0,1

Fra faktaark 4.1. Konsern - Eiendomsforvaltning for utvalgte kommunale formålsbygg.

	Juni 2012. Rate m.k.	Mars 2012. Rate m.k.	Juni 2012. Rate	Mars 2012. Rate	Rate m.k. forskjell mars- juni. Prosent	Rateforskjell mars-juni. Prosent	Forskjell mellom rate og rate m.k. Juni. Prosent
Brutto investerings- utgifter, administrasjons- lokaler, konsern	967 122	918 208	960 332	912 926	-5,1	-4,9	-0,7
Brutto investerings- utgifter, kommunale kulturbygg, konsern	894 763	922 688	894 067	923 229	3,1	3,3	-0,1
Utgifter til drifts- aktiviteter, kommunale idrettsbygg, konsern	1 320 673	1 279 783	1 320 670	1 280 869	-3,1	-3,0	0,0
Utgifter til vedlikeholds- aktiviteter, administra- sjonslokaler, konsern	157 527	157 635	157 300	157 497	0,1	0,1	-0,1
Utgifter til drifts- aktiviteter, administra- sjonslokaler, konsern	871 730	870 514	869 244	869 930	-0,1	0,1	-0,3
MVA-refusjon drift, kommunal eiendoms- forvaltning, konsern	1 774 911	1 773 632	1 772 367	1 773 340	-0,1	0,1	-0,1

Fra faktaark E1. Konsern - Kommunehelse

	Juni 2012. Rate m.k.	Mars 2012. Rate m.k.	Juni 2012. Rate	Mars 2012. Rate	Rate m.k. forskjell mars- juni. Prosent	Rateforskjell mars-juni. Prosent	Forskjell mellom rate og rate m.k. Juni. Prosent
Brutto investerings- utgifter, funksjonene 232, 233, 241, konsern	173 401	115 360	173 401	115 493	-33,5	-33,4	0,0
Brutto driftsutgifter, akutthjelp helse- og omsorgstjenester, f256, konsern	39 574	48 972	39 574	48 972	23,7	23,7	0,0
Antall avtalehjemler for private fysio- terapeuter, deltid (f.o.m. 40-100 %) ...	1 817	1 934	1 816	1 936	6,4	6,6	0,1
Timer pr. uke av fysioterapeuter. Funksjon 232	6 393	6 374	6 371	6 366	-0,3	-0,1	0,4
Timer per uke av leger. Funksjon 120 Administrasjon.	3 098	3 093	3 098	3 100	-0,2	0,1	0,0
Brutto driftsutgifter funksjon 232, 233, 241, konsern	11 609 291	11 583 908	11 591 376	11 585 540	-0,2	-0,1	0,2

Fra faktaark K1. Konsern – Kultur

	Juni 2012. Rate m.k.	Mars 2012. Rate m.k.	Juni 2012. Rate	Mars 2012. Rate	Rate m.k. forskjell mars- juni. Prosent	Rateforskjell mars-juni. Prosent	Forskjell mellom rate og rate m.k. Juni. Prosent
Driftstilskudd til lag og foreninger	522 069	475 411	520 640	474 436	-8,9	-8,9	0,3
Netto driftsutgifter Idrett (funksjon 380), konsern ...	937 389	971 939	936 817	972 072	3,7	3,8	0,1
Sum årlig åpnings- timer i kommunale fritidssenter 6-20 år	533 478	512 888	533 485	513 784	-3,9	-3,7	0,0
Totalt antall kinobesøk	13 221 697	12 754 506	13 223 578	12 753 051	-3,5	-3,6	0,0
Bokbestand, skjønnlitteratur for voksne	6 395 052	6 355 273	6 385 218	6 361 081	-0,6	-0,4	0,2
Lydbøker, totalt utlån lydbøker fra folkebibliotek	1 970 122	1 960 626	1 967 294	1 960 790	-0,5	-0,3	0,1
Netto driftsutgifter aktivitetstilbud barn og unge (funksjon 231), konsern ...	983 284	985 678	982 654	985 876	0,2	0,3	0,1

Fra faktaarkene M1. Konsern – Samferdsel og N1. Konsern - Bolig

	Juni 2012. Rate m.k.	Mars 2012. Rate m.k.	Juni 2012. Rate	Mars 2012. Rate	Rate m.k. forskjell mars- juni. Prosent	Rateforskjell mars-juni. Prosent	Forskjell mellom rate og rate m.k. Juni. Prosent
Netto driftsutgifter i 1 000 kr, samferdsel i alt, konsern	3 565 533	3 575 942	3 558 913	3 575 383	0,3	0,5	0,2
Netto driftsutgifter i 1 000 kr, kommunale veier og gater, konsern	3 728 332	3 710 498	3 721 714	3 710 046	-0,5	-0,3	0,2
Brutto investeringsutgifter i 1 000 kr, samferdsel i alt, konsern	4 143 490	4 042 314	4 129 540	4 033 170	-2,4	-2,3	0,3
Brutto investeringsutgifter i 1 000 kr, kommunale veier og gater, konsern	3 372 844	3 375 742	3 358 894	3 365 982	0,1	0,2	0,4
Kostnader til gatebelysning i 1 000 kr langs kommunale veier og gater	442 754	440 160	442 572	440 631	-0,6	-0,4	0,0
Kostnader til gatebelysning i 1 000 kr langs fylkes-/riksveier, kommunalt ansvar	112 525	110 543	112 514	110 694	-1,8	-1,6	0,0
Brutto driftsutgifter i 1 000 kr for kommunale veier og gater, konsern	5 008 381	4 984 047	5 001 304	4 983 564	-0,5	-0,4	0,1
	Juni 2012. Rate m.k.	Mars 2012. Rate m.k.	Juni 2012. Rate	Mars 2012. Rate	Rate m.k. forskjell mars- juni. Prosent	Rateforskjell mars-juni. Prosent	Forskjell mellom rate og rate m.k. Juni. Prosent
MVA-refusjon drift, bolig, konsern	356 209	355 533	356 229	355 532	-0,2	-0,2	0,0
Antall nyinnflyttede rusmiddelmissbrukere med en psykisk lidelse	608	609	608	610	0,2	0,3	0,0
Antall nyinnflyttede husstander	12 495	12 433	12 538	12 582	-0,5	0,4	-0,3
Privat eide boliger med kommunal disposisjons rett	10 531	10 139	10 531	9 988	-3,7	-5,2	0,0
Kommunalt innleide boliger	10 724	10 065	10 747	9 175	-6,1	-14,6	-0,2
Netto driftsutgifter, kommunalt disponerte boliger og boligvirkemidler, konsern	-17 430	-26 204	-17 335	-21 481	50,3	23,9	0,5

Statistisk sentralbyrå

Postadresse:
Postboks 8131 Dep
NO-0033 Oslo

Besøksadresse:
Akersveien 26, Oslo
Oterveien 23, Kongsvinger

E-post: ssb@ssb.no
Internett: www.ssb.no
Telefon: 62 88 50 00

ISBN 978-82-537-8918-7 (elektronisk)



Statistisk sentralbyrå
Statistics Norway