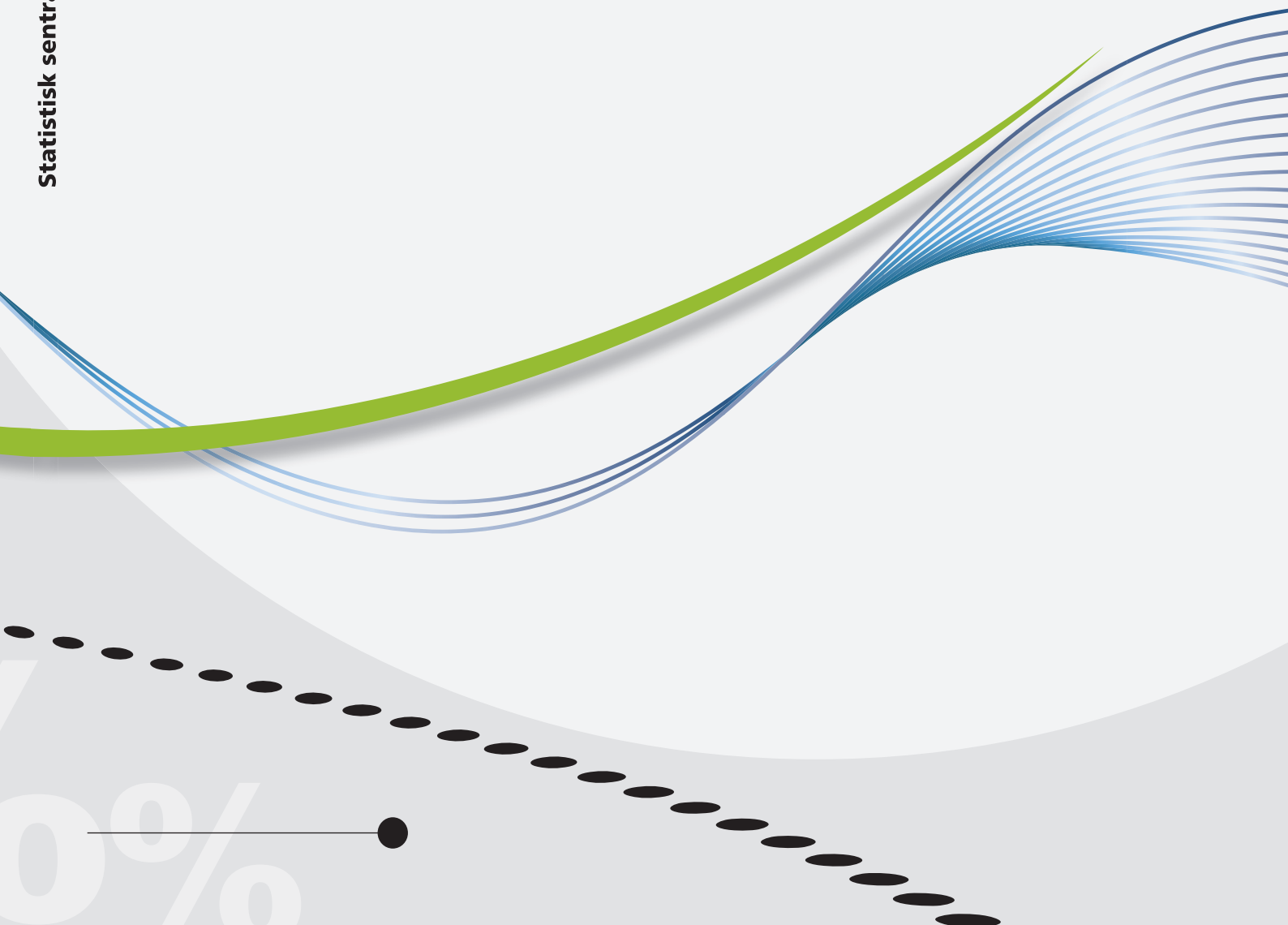


*Anne Brit Thorud, Torild Fløysvik, Dag  
Abrahamsen, Harald Tønseth, Gisle Berge, Aslaug  
Hurlen Foss og Jon Ole Johansen Hagemo*

## **System for beregning av nasjonale tall i KOSTRA del III**





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

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

	<b>Standardtegn i tabeller</b>	<b>Symbol</b>
© Statistisk sentralbyrå	Tall kan ikke forekomme	.
Ved bruk av materiale fra denne publikasjonen skal Statistisk sentralbyrå oppgis som kilde.	Oppgave mangler	..
Publisert desember 2012	Oppgave mangler foreløpig	...
	Tall kan ikke offentliggjøres	:
	Null	-
ISBN 978-82-537-8525-7 (trykt)	Mindre enn 0,5 av den brukte enheten	0
ISBN 978-82-537-8526-4 (elektronisk)	Mindre enn 0,05 av den brukte enheten	0,0
ISSN 1891-5906	Foreløpig tall	*
Emne: 00.00.20	Brudd i den loddrette serien	—
	Brudd i den vannrette serien	
Trykk: Statistisk sentralbyrå	Desimaltegn	,

## Forord

Denne rapporten dokumenterer prosjektet ”System for beregning av nasjonale tall, del III”. En egen sluttrapport beskriver de mer prosjektinterne forhold. Prosjektet er en oppfølger til prosjektene ”System for beregning av nasjonale tall i KOSTRA” fra 2010 og ”System for beregning av nasjonale tall i KOSTRA\_del II” fra 2011. Resultater og anbefalinger til videre utvikling fra de første prosjektene, er dokumentert i Interne dokumenter 30/2010 og Notater 50/2011.

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. Denne 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. Dette delprosjektet har fokusert på kvalitetskontroll av bruken av tidligere utviklede løsninger samt på teknisk videreutvikling.

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 levekårsstatistikk, Gisle Berge fra seksjon for naturressurs- og miljøstatistikk, Aslaug Hurlen Foss fra seksjon for statistiske metoder og standarder 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 Elisabetta Vassenden, fra seksjon for levekårsstatistikk Elisabeth Rønning, fra seksjon naturressurs- og miljøstatistikk Jørn Kristian Undelstvedt, forskningssjef ved seksjon for statistiske metoder og standarder, Jan Bjørnstad og fra seksjon for datafangstsystemer, Ann Christine Westling.

Det har ikke vært noen referansepersoner/-gruppe eller kvalitetslos knyttet til prosjektet.

Prosjektet er egenfinansiert av SSB gjennom statsoppdraget. Ressurser er tilført prosjektet fra KOSTRAs utviklingsmidler.

Statistisk sentralbyrå, 3. desember 2012

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 hadde grunnlagstallene for fylkeskommuner fått estimerte tall for fylkesgrupper, mens de for kommunene hadde fått estimater for kommunegrupper og fylker. Også disse ble beregnet bakover til og med regnskapsåret 2008. Prosjektet har 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.

I rapportens kapittel 1 beskrives formålet med og avgrensningen av prosjektet, mens kapittel 2 gjennomgår noen sentrale begreper.

Fjorårets prosjekt identifiserte og testet alternative metoder for estimering. Dette prosjektet ønsket å sikre at de ulike fagområdene hadde funnet en best mulig løsning for sine data. Kapittel 3 gir en kort oversikt over de valg som er gjort tilgjengelig. Her vises også prosjektets vurdering og anbefaling av usikkerhetsmål. Dette er foreløpig ikke implementert verken internt eller eksternt.

I kapittel 4 beskrives hvilken estimeringsfunksjonalitet som nå foreligger i KOSTRA, med hovedvekt på det som er utviklet i dette delprosjektet.

Flere av fagområdene som utfører estimering i den generelle løsningen har tatt en ny vurdering av hvilke metoder, modeller og forklaringsvariabler som egner seg best for de ulike typene data på området. Kapittel 5 oppsummerer erfaringene med estimeringen både fra bistandsrunden hos de statistikkansvarlige og fra de utvalgte statistikkområdene sosialtjenesten, pleie og omsorg, vann, avløp og renovasjon og kommuneregnskap. Blant annet har en nå funnet en løsning for å skille ikke rapporterte data fra verdien ”0” (null)/blank, slik at kun verdier fra kommuner som ikke har rapportert blir estimert. Når det gjelder estimering av data med stor variabilitet fra år til år (”rykk- og nappdata”) eller data som ikke passer med noen av modellene, så anbefaler prosjektgruppa at det ikke blir utviklet en generell løsning for dette, men at den enkelte fagseksjon, evt. med støtte fra seksjon for statistiske metoder, finner en metode som passer til dataene.

I kapittel 6 gis en del anbefalinger om videre utvikling. Blant annet bør det etableres tidsserier med landstall tilbake til og med regnskapsåret 2001. Usikkerhetsmål bør innføres, i første omgang internt, men etter hvert også for eksterne brukere.

## Innhold

<b>Forord</b> .....	<b>3</b>
<b>Sammendrag</b> .....	<b>4</b>
<b>Innhold</b> .....	<b>5</b>
<b>1. Innledning</b> .....	<b>6</b>
1.1. Bakgrunn og problemstilling .....	6
1.2. Formål og avgrensninger .....	7
1.3. Oversikt over innholdet i rapporten.....	7
<b>2. Begreper og definisjoner</b> .....	<b>8</b>
<b>3. Metode</b> .....	<b>9</b>
3.1. Estimeringsmodeller og metoder.....	10
3.2. Regresjonsmetodene: .....	11
3.3. Usikkerhetsmål.....	13
<b>4. Estimeringsfunksjonaliteten</b> .....	<b>13</b>
<b>5. Resultater og diskusjon</b> .....	<b>16</b>
5.1. Erfaringer og innspill i prosjektet .....	16
5.2. Egne gjennomsnittsberegninger.....	22
5.3. Behovet for å skille mellom null, blank og ennå ikke innsendte data.....	23
5.4. Estimering av "rykk og napp" -data.....	24
<b>6. Anbefalinger til videre utvikling</b> .....	<b>27</b>
<b>Vedlegg A: Prosjektskriv</b> .....	<b>29</b>
<b>Vedlegg B: Metodebeskrivelse</b> .....	<b>30</b>
<b>Vedlegg B: Metodebeskrivelse</b> .....	<b>31</b>
<b>Vedlegg B: Metodebeskrivelse</b> .....	<b>32</b>
<b>Vedlegg C: Landstall i Kostra – krav til konsistens</b> .....	<b>37</b>
<b>Vedlegg D: Detaljert oversikt over utarbeidelse av estimater i faktaark I. Vann og I.</b>	
<b>Avløp</b> .....	<b>41</b>

## 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 heller ikke alle områdene som har fulltelling til 15. juni, slik at tidsseriedataene også kan være mangelfulle. Det har derfor vært vanskelig å si noe om utviklingen i ressursbruk og tjenesteproduksjon nasjonalt og på regionale nivåer for mange fagområder. Estimerte nasjonale og regionale verdier har særlig 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 SSB.

Med dette som bakgrunn er det utviklet et system for beregning av nasjonale og regionale verdier i KOSTRA. Det har vært jobbet med utviklingen av dette i to 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. Det gjenstod dog fremdeles en del arbeid før man hadde et fullstendig system med estimerte verdier.

I prosjektets fase tre skulle gjennomsnittsberegninger som baseres på de innrapporterte dataene erstattes med estimerte gjennomsnitt. I tillegg savnet man tidsserier for landstall samt estimerte verdier på lavere regionale nivåer (fylker og KOSTRA-grupper).

*Tidsserier* For flere fagområder finnes det mangler i datasettene for tidligere år. Løsningen for estimering burde utvides til å gjelde også for tidligere årganger. Rent teknisk lå mye tilrette allerede, men det måtte settes av tid til kvalitetssikring av dataene.

*Gjennomsnitt basert på estimerte tall* Gjennomsnittsberegningene har vært basert på data som er levert. De ville kunne avvike fra landsestimatene og burde erstattes med gjennomsnitt som baseres på de estimerte dataene. Også her lå mye tilrette teknisk sett. Det måtte settes av tid til kvalitetssikring av de nye gjennomsnittene. Det måtte legges inn begrensning på valg av kommuner i beregningen av "Egne gjennomsnitt" for å sikre at brukerne ikke kunne beregne data, og i sær ikke sensitive data, for enkeltkommuner.

*Estimering på lavere regionalt nivå* Estimering på lavere regionale nivåer enn landet innebærer en større utfordring enn på nasjonalt nivå. Store mangler i datasettene vil kunne innebære at estimeringen blir usikker. Det måtte settes av ekstra tid til kvalitetssjekk i forkant av publiseringen. For å kunne beregne gjennomsnitt basert på estimater fra disse nivåene måtte vi ha estimerte tall på fylkesnivå og på kommunegruppenivå på plass.

*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 variere etter hva slags data en ser på. Et mål på usikkerhet ville gi brukerne en pekepinn på i hvor stor grad de kan ha tillit til det estimerte tallet.

*Estimering som ivaretar sammenhenger i regnskap* Den tidligere utviklede løsningen estimerte ett og ett regnskapsbegrep for seg. Dette medførte at ulike regnskaper ble vurdert som avvikende og grunnlaget for estimering ble forskjellig fra begrep til begrep. Målet var å finne fram til en løsning som valgte felles estimeringsgrunnlag, ved at data fra de samme kommunene inngikk i beregningene av alle variablene.



*Estimering av data som ikke følger naturlig utvikling* Det var ønskelig å etablere en løsning for estimering av data som ikke følger en naturlig utvikling fra år til år, de såkalte "rykk- og napp"- dataene, eksempelvis antall kloakkstopp.

Dette prosjektet videreførte arbeidet fra de to tidligere delprosjektene.

## 1.2. Formål og avgrensninger

*Effekt mål:* 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 på nasjonalt og lavere regionalt nivå. Dette vil være med på å oppfylle KOSTRAs målsetting om å vise aktuell, relevant, pålitelig og sammenlignbar styringsinformasjon.

*Resultat mål:* Rent konkret skulle prosjektet føre til:

- 1) Estimerte verdier på lavere regionale nivåer 15.3.2012
- 2) Gjennomsnitt basert på estimerte nasjonale verdier 15.3. 2012, ev 15.6.2012
- 3) Estimerte nasjonale og regionale verdier for tidligere årganger 15.3. 2012
- 4) Estimerer for regnskap som ivaretar de nødvendige sammenhenger i regnskapet, på alle detaljerings- og aggregeringsnivåer 15.3.2012
- 5) Funksjonalitet for egne gjennomsnittberegninger 15.6.2012 (ev. 15.3.2013)
- 6) En løsning for å utstyre estimatene med et usikkerhetsmål for internt bruk
- 7) Løsning for estimering av data som ikke følger en naturlig utvikling fra år til år ("rykk- og napp"- data)

*Avgrensning* Resultatmål 3 var opprinnelig tenkt å gjelde estimerte verdier for hele tidsperioden i KOSTRA, fra 2001 og fremover. For resultatmål 6 var det ønskelig at prosjektet skulle implementere et usikkerhetsmål på estimatene til internt bruk. For resultatmål 7 var det ønskelig at prosjektet skulle gjøre tilgjengelig en metodisk og it-messig løsning for estimering av data som ikke har en naturlig utvikling fra år til år.

Prosjektet og styringsgruppa innså ganske tidlig i prosessen at det ikke ville være rom for å gjennomføre disse nevnte resultatmålene fullt ut. For usikkerhetsmål og estimering av "rykk og napp"-data ble det derfor bestemt at dette prosjektet skulle finne metodiske og IT-messige løsninger, mens selve implementeringen utsettes til et oppfølgingsprosjekt. Estimerer for tidligere årganger er implementert for årene 2008 og fremover, mens årgangene før det er utsatt. Det vises også til rapportens kapittel 6 hvor prosjektets anbefalinger til videre utvikling er beskrevet.

## 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. I tillegg skisseres en løsning for å måle og vise usikkerheten knyttet til den enkelte variabel.

Kapittel 4 beskriver hvilken estimeringsfunksjonalitet som finnes i KOSTRA, med hovedvekt på det som er utviklet i dette delprosjektet.

Kapittel 5 beskriver erfaringene SSB har med estimering av nasjonale og regionale verdier så langt.

Kapittel 6 beskriver prosjektgruppas anbefalinger i forhold til temaer som fremdeles kan videreutvikles og forbedres med hensyn til beregning av estimater i KOSTRA.

## 2. Begreper og definisjoner

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

**Tabell 2.1. Begreper og definisjoner**

Beta	Er regresjonskoeffisienten og uttrykker forholdet mellom avhengig variabel (den variabelen som vi skal estimere totalen for fordi det mangler noen verdier) og forklaringsvariabelen.
DS	Dagens statistikk – publisering av artikler og tabeller knyttet til et eller flere fagområder på ssb.no Artikkelen gir en beskrivelse av dataene og sier gjerne noe om utviklingen over tid.
Ekstremverdi	Dataverdi som avviker betydelig fra andre verdier i datasettet og som kan mistenkes for å være feil.
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 som sammenligner kommunale tall for utvalgte kommuner 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. Kalles også FA-admin.
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	Enhetsfracfall: Enheter i undersøkelsesbestanden som mangler fullstendig. Partielt fracfall: Enheter som er med i undersøkelsesbestanden, men hvor enkelte opplysninger mangler.
Grunnlagsdata	Vil i stor grad være antall enheter 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.
Konfidensintervall	En måte å angi feilmarginen av en måling eller en beregning på. Et konfidensintervall angir intervallet som med en spesifisert sikkerhet inneholder den sanne (og ukjente) verdien av variabelen man har målt. Således inneholder et 95 % konfidensintervall den sanne verdien med en sikkerhet på 0,95. En verdi som ligger utenfor 95 % konfidensintervallet kan altså sies å avvike signifikant fra forventningen.
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.

Poissonregresjon	Regresjon der vi antar at observasjonene er poissonfordelt. Antagelsen om poissonfordeling blir ofte brukt på hendelser som er sjeldne og tellbare og uten en øvre grense.
Nøkkeltall	Det samme som indikator.
Prikkede tall	Tall som er blanket av personvern hensyn. Dette gjelder sensitive opplysninger basert på tre tilfeller eller færre (altså verdiene 0, 1, 2 og 3), eller fire tilfeller eller færre i IPLOS.
Regresjonskoeffisient	Beta er regresjonskoeffisienten og uttrykker forholdet mellom avhengig variabel (den variabelen som vi skal estimere totalen for fordi det mangler noen verdier) og forklaringsvariabelen.
Regresjonslinje	Den lineære funksjonen/graf som passer best med innsamlede data.
Residual	Feilledet i en regresjonsanalyse. Residualene er det som ikke kan forklares når vi søker å forklare den avhengige variabelen opp mot den eller de uavhengige variablene.
Standard avvik	Standardavviket er et mål for spredningen av verdiene i et datasett eller av verdien av en stokastisk variabel. Den er definert som kvadratroten av variansen.
Standardfeil	Standardfeil er en måte å angi feilmarginen til et estimat på. Den er definert som det estimerte standardavviket til estimatoren.
Stokastisk variabel	Stokastisk variabel, tilfeldig variabel, en observert variabel som kan anta tilfeldige tallmessige verdier innenfor et nærmere angitt utfallsrom når observasjonene gjentas flere ganger. Begrepet stokastisk benyttes av og til som motsetning til deterministisk, der man alltid forventer å finne samme resultatet ved gjentatte observasjoner.
Studentisert residual Rstudent	Residualet justert ved å dividere det med et estimat av standardavvik. Samtidig er selve observasjonen ikke med i beregningen av predikert verdi og standardavvik. Studentisering av residualer er en viktig teknikk for å avdekke enkeltobservasjoner som avviker fra data. Forkortes ofte som rstudent.
Traversere	Forflytning i beregningene i faktaark-admin for å lese/estimere en mellomregningsformel på laveste nivå.
Variabel	Egenskap ved de statistiske enhetene i en undersøkelse, f.eks. alder eller omsetning.
Varians	Varians (empirisk) er et mål på variasjonen i et utvalg fra en statistisk fordeling. Den empiriske variansen er et estimat av den teoretiske variansen.
Variasjonskoeffisienten	Angir hvor stor utvalgsusikkerheten utgjør av det 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 og metoder og forklaringsvariable som kan benyttes ved estimeringen. For mer utførlig beskrivelse henvises til rapporten fra det forrige prosjektet, Notater 50/2011.

### 3.1. Estimeringsmodeller og metoder

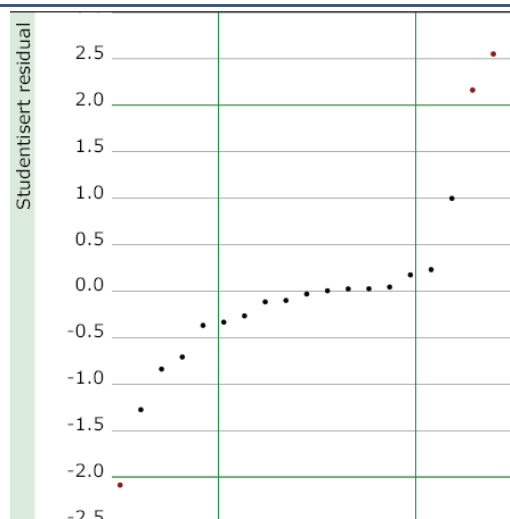
#### Regresjon

Det er mulig i KOSTRA å velge tre regresjonsmetoder: Rate, rate med konstantledd og vanlig lineær regresjon. Verdiene som oppgavegiverne rapporterer inn til KOSTRA kan betraktes som et tilfeldig utvalg av alle de verdiene som skulle vært rapportert. Enkelte kommuner rapporterer ikke skjemaet i det hele tatt (enhetsfrfall) og noen rapporterer bare deler av skjemaet (partielt frfall). Det partielle frafallet kan i enkelte tilfeller skyldes at verdien egentlig skal være null og ikke missing. Dette kan skape problemer for estimeringen. Hvordan dette er løst for enkelte statistikker er beskrevet i kapittel 5.3. Verdiene som blir rapportert inn blir delt inn i tre kategorier:

1. Verdiene er riktige og representative
2. Verdiene er riktige, men ikke representative
3. Verdiene er feil

Verdiene som faller i den første kategorien skal være med i datagrunnlaget for estimeringen. Verdiene i den andre kategorien holdes utenfor estimeringen, men beholder sin verdi. Verdiene i den siste kategorien får sin verdi fjernet og behandles som frfall. Når alle KOSTRA-tall som ikke har blitt rapportert har fått estimert en verdi, summeres enhetene i og utenfor utvalget og vi får en estimert totalverdi, for hele landet eller for et annet regionalt nivå.

Figur 3.1. Studentiserte residualer



Verdiene fordeles i de tre kategoriene automatisk ut i fra størrelsen på det studentiserte residual (rstudent), det vil si hvor langt observasjonen er fra regresjonslinja og standardisert i forhold til variansen. Dersom rstudent er veldig stor, det vil si over 50, betraktes verdien som feil (kategori 3). Denne grensen kan også settes høyere eller lavere gjennom valgmuligheter i faktaark-admin. Dersom rstudent er mindre enn 50, men over 2,5, anses verdiene som riktige, men ikke representative (kategori 2). Også den nedre grensen er det mulig å sette lavere eller høyere. Verdiene som anses som feil, er veldig ekstreme. Disse kan påvirke rstudent til alle enheter som ikke anses som feil. Rstudent beregnes derfor på nytt etter at enhetene med feil verdi er fjernet.

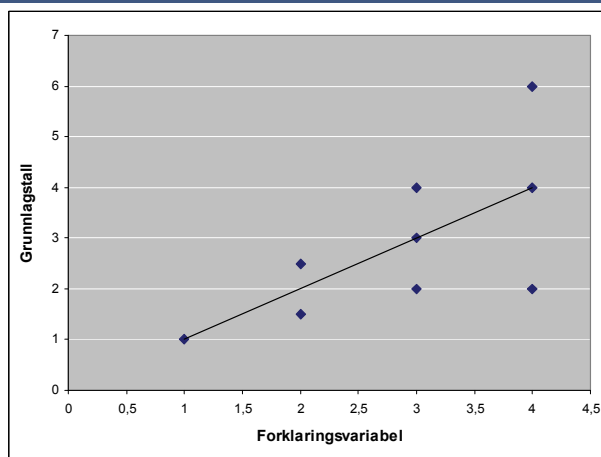
Alle regresjonsmetodene trenger en forklaringsvariabel. I utgangspunktet kan alle mellomregninger brukes, men folketallet i kommunen er satt som standard valg. Det er viktig at det er en god sammenheng mellom forklaringsvariabelen og den variabelen som det skal estimeres en verdi for. Forklaringsvariabelen må være rapportert for alle enheter (fulltelling) og alle verdiene må være positive tall (større enn 0).

### 3.2. 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 konstantlegg” (jfr. neste punkt).

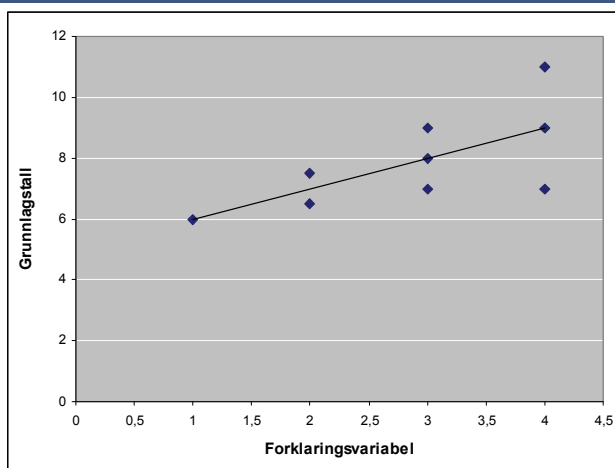
**Figur 3.2. Stilisert modellskisse på ratemodellen**



*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 likt null, ville man likevel forvente at grunnlagstallet avviker signifikant fra null som verdi. Dette kan skyldes oppstartskostnader og/eller stordriftsfordeler.

**Figur 3.3 Stilisert modellskisse på rate med konstantledd**

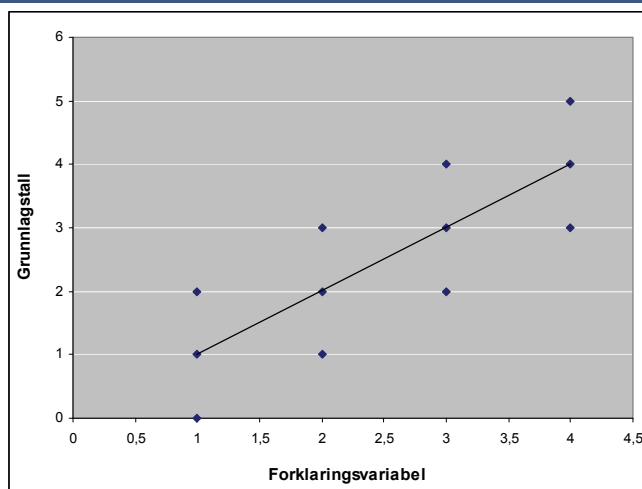


*Regresjon - enkel lineær regresjon*

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.

**Figur 3.4. Stilisert modellskisse på regresjon**



### 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 estimater/imputering*

Eksterne estimater betyr at statistikkansvarlig estimerer totalene 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 på kommunenivå.

### Bruk av metodene ved publiseringen i juni 2012

Det blir produsert omtrent 20 000 estimater for grunnlagstallene i KOSTRA. Omtrent 95 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. Det er bare 1 prosent som har valgt en annen metode. 4 prosent av tallene blir bare summert og ikke estimert. Summering benyttes på data som har fulltelling, typisk innen barnehage og grunnskole, men også innen KNNM og VAR og for noen typer data fra kultursektoren. 83 KOSTRA-tall har blitt imputert med forrige års verdi, dette brukes i hovedsak på variabler innen barnevern, KNNM og sosialtjenesten. Eksterne estimater er brukt på 157 tall innen pleie og omsorg og VAR-sektoren. Bare en variabel er estimert med vanlig regresjon og kun en ved hjelp av rate med konstantledd. 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. Det er også fremmet forslag om å vurdere om ratemodellen med konstantledd kan være en bedre metode å legge inn som default. Se prosjektets anbefalinger til videre arbeid i kapittel 6.

**Tabell 3.1. Bruk av metoder ved estimeringen i juni 2012**

Metode	Antall	Prosent
Rate .....	19 073	94,75
Rate med konstantledd .....	1	0,00
Regresjon .....	1	0,00
Imputering .....	83	0,41
Eksterne estimater .....	157	0,78
Summering .....	789	3,92
Ingen metode valgt .....	25	0,12
I alt .....	20 129	100

### 3.3. Usikkerhetsmål

I prosjektet har det blitt vurdert flere usikkerhetsmål og på forskjellige nivåer. Det har blitt vurdert om usikkerhetsmål skal bli gitt for både nøkkeltall og grunnlagsdata, men foreløpig finner prosjektgruppa det mest hensiktsmessig å bare gi usikkerhetsmål på grunnlagsdataene. I tillegg har det blitt vurdert om usikkerhetsmålet skal omfatte utvalgsusikkerhet (som følge av frafall) eller også målefeil (data som er rapportert feil). I prosjektgruppa ble det enighet om at det er tilstrekkelig å vise utvalgsusikkerheten. Det er bare regresjonsmetodene som kan gi mål på utvalgsusikkerheten. Det er to usikkerhetsmål som er aktuelle, variasjonskoeffisient og konfidensintervall (prediksjonsintervall). De to målene er forklart under. Disse målene tar imidlertid bare hensyn til at vi estimerer verdier for manglende kommuner, og ikke hensyn til målefeil som kan bli rettet i revisjonen. Variasjonskoeffisienten er noe enklere å tolke for brukerne og dette usikkerhetsmålet ble derfor valgt. Utrekningen av variasjonskoeffisienten er gitt i vedlegg B Metodebeskrivelse.

#### Variasjonskoeffisient

Standardfeilen til den estimerte totalverdien forteller oss noe om hvor usikkert estimatet er. Ofte er det slik at jo større estimatet er, jo større standardfeil kan vi akseptere. Det kan derfor være hensiktsmessig å uttrykke standardfeilen som en andel av selve estimatet. Estimaten av denne andelen kaller vi for estimert variasjonskoeffisient eller bare variasjonskoeffisient, og den er ofte oppgitt i prosent. Hvor stor prosentandel som er akseptabel, kan variere fra statistikk til statistikk. For variasjonskoeffisienten blir ofte forkortelsen CV ("*coefficient of variation*") brukt.

#### Konfidensintervall

Et 95 % konfidensintervall for den sanne totaleverdien er vanligvis på formen: estimat  $\pm$  2standardfeil. Det betyr at sannsynligheten for at intervallet dekker den sanne verdien er 0,95 (95 prosent). Eller sagt på en annen måte, ved gjentatte utvalg så vil 95 prosent av intervallene dekke den sanne verdien. Vi sier da at det er 95 prosent *sikkerhet* for at det *beregnete* intervallet dekker den sanne verdien.

#### Visning av usikkerhetsmål

Foreløpig anbefaler prosjektgruppa at usikkerhetsmålene legges inn i estimattesteren. Dermed kan de statistikkansvarlige finne variasjonskoeffisienten samtidig med at de tester selve estimeringen. Variasjonskoeffisienten vil da være gitt per KOSTRA-gruppe og totalt. Det kan være mulig å lage variasjonskoeffisient for fylkestall, men da må det beregnes standardavvik for disse, siden KOSTRA-grupper ikke kan summeres til fylker. Det må dessuten bestemmes hvor disse usikkerhetsmålene for fylkestall skal vises.

## 4. Estimeringsfunksjonaliteten

I dette kapitlet beskrives hvilken estimeringsfunksjonalitet som nå foreligger i KOSTRA, med hovedvekt på det som er utviklet i dette delprosjektet.

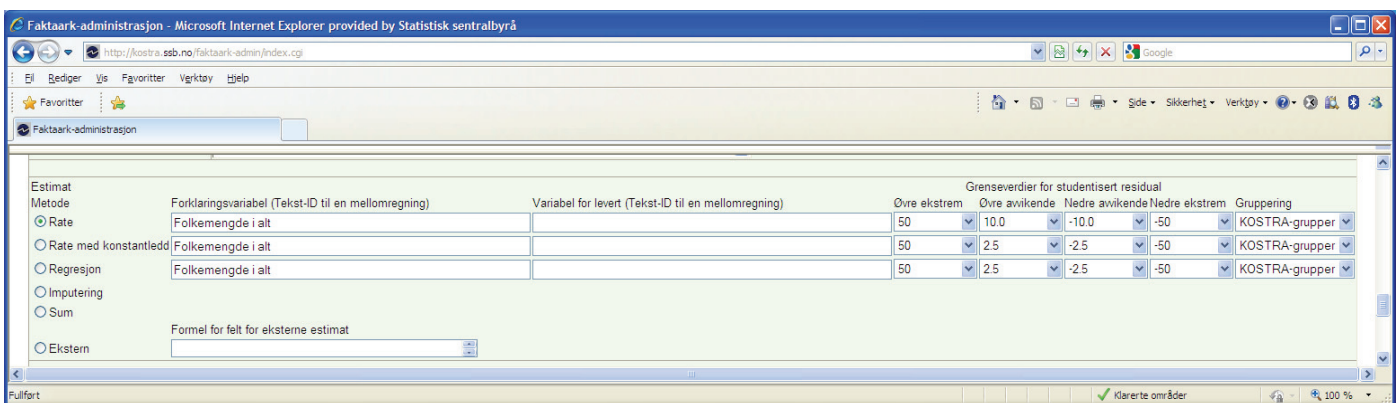
Generelt omfatter estimering kombinasjonen av alle regioner som merket med talltype = estimat, årgangene for 2008 og nyere, samt indikatorer og mellomregninger som er merket med estimat = ja.

Følgende funksjonalitet tilbys:

Faktaarkene på nivå 3:

- Estimerte verdier for grunnlagstallene
  - Alle mellomregninger kan estimeres ved hjelp av en metode. I faktaark-admin fins det 6 forskjellige metoder å velge i: Rate, rate med konstantledd, regresjon, imputering, eksterne estimater og summering. Noen av metodene har parametre som settes slik at estimeringen blir best mulig.
  - Hovedregelen er at hvis beregningsformelen til en mellomregning kun består av andre mellomregninger skal de underliggende estimeres og estimatet blir da summen. Hvis en mellomregning refererer direkte til et felt i et e-skjema eller lignende estimeres denne mellomregningen ut i fra metode og parametre.

Del av skjermbilde for metode og parametre for estimering av en mellomregning:



Faktaarkene på nivå 2:

- Estimerte verdier for forholdstallene
  - Hovedregelen er at en indikator er et forholdstall, og at underliggende mellomregninger estimeres og summene settes inn i beregningsformelen til indikatoren for så å bli regnet ut.
  - Unntak er at noen indikatorer bare refererer til én mellomregning, som i prinsippet er for å vise grunnlagsdata på nivå 2. I disse tilfellene vil estimatet for den underliggende mellomregningen bli vist.
- Egne estimat
  - Utregningen er basert på ferdig utregnede estimat da alle kommuner enten har tall fordi de har levert eller estimat pga manglende innrapportering. Egne estimat kan ikke lagres da deres grunnlag er forskjellig fra gang til gang, dvs. at man ikke vet på forhånd hvilke kommuner det skal beregnes eget estimat for.
  - Brukeren (internett/internt) velger seg ut en gruppe kommuner som man ønsker estimat for.
  - Brukeren får et eget estimat per indikator i faktaarket basert på de parametre som faktaark-ansvarlig har satt på indikatoren og de underliggende mellomregningene.
  - Unntak for egne estimat er de tilfeller der det fins prikkede tall. Unntaksreglen er at det skal være 3 eller flere som har data som ikke er prikket. Ellers blir estimatet i seg selv prikket pga sensitivitet.



Knapp for eget gjennomsnitt/estimat:

**A. Finansielle nøkkeltall og adm., styring og fellesutgifter - nivå 2**  
 Reviderte tall per 29.06.2012

Fylkeskommune 
  Kommune 
  Bydel 
  Institusjon: Videregående opplæring

Utvalgte Nøkkeltall 
  Detaljerte nøkkeltall 
  Kvalitet 
  Grunnlagsdata

	Halden	Kostragruppe 13	østfold	Landet uten Oslo	Landet
	2011	2011	2011	2011	2011
<b>1. Finansielle nøkkeltall</b>					
<b>1.1 Finansielle nøkkeltall i prosent av driftsinntektene</b>					
Brutto driftsresultat i prosent av brutto driftsinntekter Enhet : Prosent	f i d s	-6,7	1,5	0,8	1,8
Netto driftsresultat i prosent av brutto driftsinntekter Enhet : Prosent	i d s	-8,1	1,4	0,2	1,8
Overskudd for lån og avsetninger i % av brutto driftsinntekter Enhet : Prosent	f i d s	-12,4	-4,6	-4,1	-4,6
Netto finans og avdrag i % av brutto driftsinntekter Enhet : Prosent	i d s	5,0	3,8	4,3	3,7
Netto avdrag i prosent av brutto driftsinntekter Enhet : Prosent	i d s	2,4	2,9	2,9	2,9
Netto finans i prosent av brutto driftsinntekter Enhet : Prosent	i d s	2,7	0,9	1,4	0,8

Velg regioner som skal inngå i estimatet:

**Velg egne regioner**

- 010100 : Halden
- 010400 : Moss
- 010500 : Sarpsborg
- 010600 : Fredrikstad
- 011100 : Hvaler
- 011800 : Aremark
- 011900 : Marker
- 012100 : Rømskog
- 012200 : Trøgstad
- 012300 : Spydeberg
- 012400 : Askim
- 012500 : Eidsberg
- 012700 : Skiptvet
- 012800 : Rakkestad
- 013500 : Råde
- 013600 : Rygge
- 013700 : Våler
- 013800 : Hobøl
- 021100 : Vestby
- 021300 : Ski

Se estimat som en egen region/kolonne:

**A. Finansielle nøkkeltall og adm., styring og fellesutgifter - nivå 2**  
Reviderte tall per 29.06.2012

Fylkeskommune 
  Kommune 
  Bydel 
  Institusjon: Videregående opplæring

Utvalgte Nøkkeltall 
  Detaljerte nøkkeltall 
  Kvalitet 
  Grunnlagsdata

A. Finansielle nøkkeltall og adm., styring og fellesutgifter - nivå 2

	Halden	Kostragruppe 13	Østfold	Landet uten Oslo	Landet	Egen kommunegruppe
	2011	2011	2011	2011	2011	2011
<b>1. Finansielle nøkkeltall</b>						
<b>1.1 Finansielle nøkkeltall i prosent av driftsinntektene</b>						
Brutto driftsresultat i prosent av brutto driftsinntekter Enhet : Prosent	f i d s	-6,7	1,5	0,8	1,8	1,7
Netto driftsresultat i prosent av brutto driftsinntekter Enhet : Prosent	i d s	-8,1	1,4	0,2	1,8	2,0
Overskudd før lån og avsetninger i % av brutto driftsinntekter Enhet : Prosent	f i d s	-12,4	-4,6	-4,1	-4,6	-3,3
Netto finans og avdrag i % av brutto driftsinntekter Enhet : Prosent	i d s	5,0	3,8	4,3	3,7	3,2
Netto avdrag i prosent av brutto driftsinntekter Enhet : Prosent	i d s	2,4	2,9	2,9	2,9	2,9
Netto finans i prosent av brutto driftsinntekter Enhet : Prosent	i d s	2,7	0,9	1,4	0,8	0,4
- herav Netto tap/gevinst på finansielle instrumenter i prosent av brutto driftsinntekter Enhet : Prosent	i d s	-	0,1	-0,2	0,1	0,1

For ytterligere teknisk beskrivelse av estimeringsfunksjonaliteten henvises til rapporten fra forrige delprosjekt, Notater 50/2011, kapittel 4.

## 5. Resultater og diskusjon

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

### 5.1. Erfaringer og innspill i prosjektet

#### Metodebistand til fagseksjonene

Alle fagseksjonene fikk tilbud om metodebistand for å velge estimeringsmetode. Fagområdene Kirke, Landbruk og Fysisk planlegging, kulturminne, natur og nærmiljø takket ja til tilbudet.

Fagområdet kirke hadde selv testet standardmetoden; ratemodellen med folke- mengden som forklaringsvariabel. Figurene i faktaark-admin og tester utført i SAS, viste at denne metoden fungerte greit. Det var likevel rom for forbedring og istedenfor å velge folke- mengde for å estimere antall begravelser ble det gjort forsøk på å bruke antall døde som forklaringsvariabel. Tilsvarende for antall døde ble det gjort forsøk på å bruke antall fødte istedenfor folke- mengde som forklarings- variabel. Endringen av forklaringsvariabel til fødte og døde ga en bedre sammen- heng med døde og begravelser og ble derfor valgt som forklaringsvariabler.

For fagområdet Fysisk planlegging, kulturminne, natur og nærmiljø ble det hovedsakelig sett på variablene "Antall søknader om dispensasjon for motorferdsel i utmark behandlet" og "Antall avslåtte søknader om disp. fra motorferdsel forbud". Det var ikke særlig stor sammenheng mellom disse variablene og folke- mengde. Figurene viste at det var nesten omvendt proporsjonalt; jo flere innbyggere i kommunen jo færre søknader og omvendt. Sannsynligvis ville areal for utmark gi en bedre forklaring på antall søknader om dispensasjon for motorferdsel i utmark, men dette finnes ikke tallfestet ennå. Antall søknader og avslåtte søknader er heltall som kan variere veldig fra år til år. Folke- mengde kan bare forklare omtrent 0,5

prosent av variasjonen i antall søknader når vi ser på hele landet under ett. Det er noe bedre sammenheng mot fjoråret. Der det er rapportert to år på rad, kan forrige års verdi forklare 89 prosent av variasjonen i antall søknader. Siden ikke alle rapporterer årlig, kan vi ikke bruke ratemodell med forrige års verdi, men det er mulig å imputere fjorårets eller tidligere års verdier. Noen av de som ikke har rapportert skulle vært rapportert som null, f.eks. Utsira. Det er mulig å legge inn at hvis skjema er levert, men at et felt ikke er fylt inn, så skal verdien settes til null.

For fagområdet landbruk ble det i hovedsak bare en diskusjon av variablene som ble samlet inn og om de egnet seg for estimering. Mange av tallene som blir samlet inn er av typen antall søknader, antall bedrifter og areal dyrka jord. Slike tall vil sannsynligvis ha størst sammenheng med det kommunen rapporterte forrige år og ikke folkemengde. Så hvis det skal gjøres beregninger for disse tallene ble det anbefalt enten å bruke imputering av forrige års verdi eller en regresjonsmodell med forrige års verdi som forklaringsvariabel.

Det er altså kun et fåtall av fagseksjonene som har tatt i mot tilbudet om metodebistand. Som vi beskrev i kapittel 3.1.3 brukes defaultinnstillingene, ratemodellen med folkemengde som forklaringsvariabel, på de langt fleste variablene, uten at dette nødvendigvis er det som passer best på alle typer data. Det er derfor grunn til å følge opp fagseksjonene ytterligere, og prosjektet anbefaler at det jobbes videre med dette. Se prosjektets anbefalinger til videre arbeid i kapittel 6.

### **Sosialtjenesten**

Erfaringene fra sosialtjenesten tilsier fortsatt at de statistikkansvarlige bør gjennomgå de estimatene som måtte blankes ved siste publisering. Man bør peke på hva begrunnelsen var for å blanke estimatet (hva feilen bestod i), og angi hva som må gjøres for å kunne levere estimat for 2012-data neste år, eventuelt angi hvorfor det ikke er realistisk.

Det er fra og med 15.3.2012 funnet en løsning for å unngå estimering av ikke-utfylte nuller. Det er også fra og med innrapporteringen av 2012-data foreslått en løsning som unngår at det blir estimert tall for kommuner som ikke har bestemte ordninger (for eksempel introduksjonsstønad). Det gjenstår å se om sistnevnte løsning fungerer. Se nærmere om dette under kap. 5.3

Vi ser ellers at en del brukere har vansker med å forstå at det estimerte landstallet ikke er å forstå som det endelige tallet (m.a.o. at det ikke er snakk om en fulltelling).

### **Pleie og omsorg**

Ved publiseringene 15.3 og 15.6.2012 var det estimat for alle variabler på landsnivå, også bakover i tid til 2008. Fylkes- og KOSTRA-gruppeestimer var først klare til juni-publiseringen, og da bare for 2011, tiden strakk ikke til mer. Det har imidlertid kommet svært få henvendelser på at tidligere årganger mangler. Tilgjengelige ressurser tatt i betraktning, kan det derfor synes som om det bør være tilstrekkelig å oppdatere tidligere årganger i forbindelse med publiseringen i mars 2013. Da har en forhåpentligvis også kunnet optimalisere produksjonslinjen og justere estimat for variabler med slikt behov.

Bortsett fra meldinger på manglende fylkestall, er det ikke mottatt reaksjoner på estimat-tilnærmingen i 2012.

Nedenfor er angitt en del feilsituasjoner som bør følges opp – de fleste er løselige og bør være korrigert til publiseringen i 2013.

*Manglende verdi i mellomregninger som inngår i indikatorer* Det viser seg at enkelte komplekse estimatberegninger ikke fungerer tilfredsstillende på konserntallenes regionale nivå. Dersom en komponent er tom i en mellomregning, f.eks. at et fylke ikke har plasser i barnebolig, vil estimattall for fylket mangle dersom variabelen ”plasser i barnebolig” inngår i en mellomregning som brukes i indikatoren. Mens dette gikk bra på landsnivå, kan en risikere ikke å få estimat i enkelte regioner. Løsningen er mellomregninger som alltid har verdi større enn 0. En enkel måte å omgå dette problemet, er å legge til ”+ 0” i formelen slik at tallet i det minste blir 0 i stedet for udefinert/blank/missing.

*Konsern vs kommunekasse* Det er en ekstra stor utfordring når fagdata er ulike på konsern og kommunekasse. Dette dreier seg om kun 3-4 kommuner, men slår inn på mange variabler (mange plass- og årsverksvariabler). I en del tilfeller har en ikke klart å opprettholde identiske beregningsmetoder i de to arkene, hvilket har ført til manglende tall i konsernarket av typen ”manglende verdi i mellomregning” som nevnt over.

*Regnskap* Regnskapsdataene innen pleie og omsorg estimeres etter ratemetoden – uten konstantledd. Det er ikke vurdert om rate med konstantledd kunne fungert bedre, men særlig for institusjonstjenesten må en regne med store oppstartskostnader og stordriftsfordeler som det ikke blir tatt høyde for nå. Kvaliteten på enkeltestimater må dog vurderes opp mot behovet for sammenheng i alle regnskapsbegrepene, jf. kap. 5.1.4

*Institusjoner* Det er en stor utfordring at antall institusjonsplasser ikke nødvendigvis avspeiler folketallet i den enkelte kommune; Kommunene står fritt i å velge institusjonsgrad, så lenge tilbudet finnes. For plasser i institusjon brukes derfor fjorårstallet der en har frafall. Det er ikke gjort analyser av hvor presis regnskapstallene knyttet til institusjonsfunksjonene (f253 og 261) treffer, disse er jo estimert etter ratemodellen uten konstantledd. Fjorårstallet med vekstfaktor vil sannsynligvis treffe best. Også her kan det være motstrid mellom best mulig kvalitet på estimatet for enkeltvariabler og behovet for sammenheng og konsistens i alle regnskapsbegrepene, slik det er beskrevet i kap. 5.1.4.

*Personell* Når det gjelder personell, blir tallene i mars estimert eksternt ved seksjon for arbeidsmarkedsstatistikk. Publiseringen av årsverkstall i mars baserer seg på færre datakilder enn den tilsvarende publiseringen i juni. I marstallene mangler blant annet data fra lønns- og trekkoppgaveregisteret (LTO), fra PAI-registeret og andre lønnsregistre. Det hentes inn data fra disse for å sikre at totaltallet for sysselsatte for landet skal bli det samme i registerstatistikken som i Arbeidskraftundersøkelsene (AKU). I praksis betyr dette at tallene for årsverk i mars, basert kun på Aa-registeret blir liggende for lavt jamført med tall som publiseres i juni. SSB foretar derfor en oppblåsing av registertallene i marspubliseringen med forholdstallet som var mellom tall for mars- og junipubliseringen året før og endring i forholdstall mellom AKU og tall fra Aa-registeret for inneværende år. For kommuner som har fulgt reglene for meldingsgang til Aa-registeret begge år, skal de oppblåste tallene være av relativt god kvalitet. I kommuner som enten har rettet opp feil eller hvor feilene har økt, vil oppblåsing ofte forsterke feil i endringstallene.

Både i mars og juni er det fulltelling, slik at en ikke skal trekke inn ytterligere estimering. Faktaarkadmin administrerer dette smidig.

*Brukerdata* Av brukerdata er det inntil videre kun beboere i heldøgnsbemannet bolig som hentes fra skjema (KOSTRA skjema 4 - Pleie og omsorg samleskjema). Variabler som mangler her imputeres med fjorårsverdi.

Øvrige brukerdata hentes fra IPLOS-registeret, og estimeres eksternt ved imputering av fjorårstall. I tillegg til frafall som følge av manglende innrapportering, hender det også at åpenbart feilaktige data fjernes og erstattes av fjorårsverdien ved beregning av estimater.

*Brukerbetaling* For brukerbetaling estimeres det nå med rate og folketall. Brukerbetaling er ikke knyttet til folketall, så her er nok ratemodellen med fjorårsverdi som forklaringsvariabel mer egnet. Da får en med vekstfaktoren, i motsetning til ved ren imputering.

### VAR-området

På VAR-området omtales her kun vann og avløp – ikke renovasjon/avfall.

Det er i 2012 frigitt betydelig flere estimater sammenlignet med hva som var tilfelle til junipubliseringen i fjor. Det har vært gjort en relativt stor innsats i 2012 på å få gjennomført overføring fra ”gamle gjennomsnitt” til ”nye estimater”.

**Tabell 5.1. Estimerte tall innen vann og avløp juni 2011 og juni 2012**

Emneområde	Nivå	Totalt antall tall	Antall estimert juni 2011	Antall estimert juni 2012	Prosentandel estimert juni 2012
Vann	Nivå 2 - nøkkeltall	46	0	32	70 %
	Nivå 3 - grunnlags	68	5	34	50 %
Avløp	Nivå 2 - nøkkeltall	42	0	25	60 %
	Nivå 3 - grunnlagstall	71	6	42	59 %

Det gjenstår riktignok noen estimater på både grunnlags- og nøkkeltallene, så helt ferdig er man ennå ikke for perioden 2008..2012. Detaljer over hvilke tall som fremdeles mangler estimat og hvilke som er beregnet eksternt finnes i vedlegg D. Eksterne estimater er beregnet utenfor faktaark-admin ved hjelp av SAS, og importert inn igjen i det tradisjonelle produksjonsløpet i form av ”eksterne estimater” før publiseringen.

Det er to forhold som slår en når man ser oppsettet, samt vedlegg D. Det første er at det fremdeles er mangler i estimater på både vann og avløp. Det andre er at det er relativt mange eksterne estimater i faktaarkene. Noe av bakgrunnen for begge forholdene er beskrevet i prosjektrapporten til delprosjekt II av dette prosjektet (kapittel 5.1.3, Notat 50/2011 – System for beregning av nasjonale tall i KOSTRA II).

En kort oppsummering følger her.

Grunnen til at det fremdeles *mangler estimat* for noen grunnlags- og nøkkeltall skyldes at det er tidkrevende å kvalitetssjekke variablene for å sikre best mulig estimat. En del av grunnlaget har også endret seg opp igjennom årene, noe som tilsier at estimatformler ikke nødvendigvis kan kopieres bakover i tidsserien, men krever justering. Dette kan medføre en del testing siden formelapparatet henger sammen på veldig mange måter og en oversikt over alle sammenhenger ikke alltid er like enkelt å se i skjermbildet i faktaark-admin. Man må derfor bruke en del tid på å få en fullstendig oversikt før man eventuelt gjør justeringer.

Arbeidet med estimater har medført at tidligere formler for grunnlags- og nøkkeltall har måttet justeres slik at estimater i det hele tatt lar seg beregne. Det er et annerledes system man nå har gått over til, og det krever justeringer – justeringer man som fagperson på forhånd ikke nødvendigvis er forberedt på.

Når det gjelder utstrakt bruk av *eksterne estimat*, så har det vist seg hensiktsmessig å estimere flere av grunnlagstallene i SAS, dvs. utenfor faktaark-admin. Det genereres da en resultatfil eksternt for publisering som importeres inn igjen til publisering via en tilrettelagt importrutine i faktaark-admin.

Der er i hovedsak to hovedgrunner til at eksterne estimater er foretrukket metode for estimering på vann- og avløpsområdet i KOSTRA: (1) tilfeller av ”rykk og napp”-data og (2) avhengigheter mellom de rapporterte og publiserte tallene for en gitt årgang. Problemstillingene utledes i de påfølgende avsnittene.

De etablerte metodene i faktaark-admin (automatisk imputering, rate- eller regresjonsmodell) vil være naturlige førstevalg når det gjelder valg av estimeringsmetode. Disse følger et kontinuerlig løp og er mer fullautomatiserte inn mot publisering enn beregning fra eksterne estimater. Ulempen med de etablerte metodene er at de ikke tar tilstrekkelig hensyn til tilfeller av ”rykk og napp” som vi har noen eksempler på innenfor vann- og avløpsområdet. Med ”rykk og napp”-data menes her data som kan, men ikke nødvendigvis skjer årlig.

**Tabell 5.2. Eksempler på ”rykk- og napp data” på vann- og avløpsområdet. Grunnlagstall.**

Faktaark	Grunnlagstall
I. Vann	Antall meter utskiftet/rehabiliteret ledningsnett
I. Vann	Antall lekkasjereparasjoner på nettet (antall/km ledningsnett)
I. Avløp	Lengde nylagt ledningsnett
I. Avløp	Lengde fornyet ledningsnett
I. Avløp	Antall kloakkstopper i avløpsledninger og kummer
I. Avløp	Antall avsluttede saker over kjelleroversvømmelser, der kommunen har erkjent erstatningsansvar

De etablerte estimat metodene i faktaark-admin vil i slike tilfeller søke å generere et tallestimat for alle kommuner som ikke har levert skjema eller har blank verdi. Beregnet estimat vil i slike tilfeller kunne bli for høyt, og derfor virker det mer naturlig at deler av vann- og avløpstallene i KOSTRA beregnes som eksterne estimater også i umiddelbar framtid. Eksterne estimater på vann- og avløpsområdet beregnes i SAS, og her gis det mer muligheter til tilpasning i beregning av estimat for ”rykk og napp”-data, og dermed trolig også mer korrekt resultat enn om man skulle benyttet etablerte metoder i faktaark-admin.

Slik de tidligere tabellene viser, så er det de ”ikke-økonomiske” grunnlags- og nøkkeltallene hvor eksterne estimater er benyttet hittil på vann og avløp. For disse finnes det nå et helhetlig beregningsopplegg i SAS, og tidsbruken på denne loopen innom eksterne estimater, tar normalt minimalt med tid nå som beregningsmodulen er ferdig programmert.

Imputering av verdier fra tidligere års rapportering er en hensiktsmessig måte å få anslag på en del av grunnlags- og nøkkeltall på vann og avløp. Antall innbyggere tiknyttet, utslipp og lengde på ledningsnett er variabler som ikke endres spesielt mye fra et år til et annet.

Imputering er allerede et alternativ og utgjør en av de etablerte metodene i faktaark-admin. Den har også vært vurdert anvendt på en del av de grunnlags- og nøkkeltallene hvor de i dag har endt opp som eksterne estimater. Eksterne estimater bruker imidlertid også i hovedsak imputering som metode, men den oppfører seg noe mer ”raffinert” enn faktaark-admin har mulighet til. Bakgrunnen skyldes punkt to ovenfor, at der er avhengigheter mellom rapporterte og publiserte tall for en gitt årgang, og at man derfor har behov for å imputere ikke bare ett tall men flere og det på tvers av grunnlagstallene. Dette illustreres best med et eksempel.

Dersom man ønsker å bruke imputering som metode i faktaark-admin, så vil denne metoden lete seg bakover i tidsserien for et bestemt grunnlagstall inntil det finner en verdi som kan benyttes. Man har imidlertid ingen kontroll med hvilken årgang man havner på for et gitt grunnlagstall i forhold til et annet. For flere av grunnlagstallene på vann og avløp kan dette være problematisk databehandling da man kan miste linkene mellom total mengde og delmengde for en del grunnlagstall. På avløp har man for eksempel publiserte grunnlagstall på total lengde, men også mer detaljert for hvilke periode den totale lengden fordeler seg utover (se liste under).

- Lengde ledningsnett totalt
- Lengde ledningsnett med ukjent alder
- Lengde ledningsnett etter 2000 (meter)

- Lengde ledningsnett 1980-99 (meter)
- Lengde ledningsnett 1960-79 (meter)
- Lengde ledningsnett 1940-59 (meter)
- Lengde ledningsnett før 1940 (meter)

Total lengde skal kunne summeres opp av de ulike lengdene som man har for før 1940, 1940-59 osv.

Dersom man imidlertid lar automatikken råde grunnen i faktaark-admin med etablert imputering-metode, så har man ikke noen garanti for at man ender opp med data fra samme årgang for totalt antall ledningsmeter og la oss si perioden før 1940. Dersom kommunen har endret i rapporteringen sin slik at feltet perioden ”før 1940” er blanket/tom fra et år til neste – noe som ikke er utenkelig, selv om det logisk høres urimelig ut – vil man kunne ende opp med at total lengde hentes ut i fra 2010-rapporteringen, mens for grunnlagstallet lengde ledningsnett før 1940 hentes fra 2009-rapporteringen. Så lenge total og delmengde ikke kan behandles avhengig av hverandre i imputering-metoden i faktaark-admin, så vil det kunne oppstå mulige skjevheter mellom genererte estimater for total og delmengde.

I eksterne estimater vil man kunne håndtere dette problemet med avhengighet mellom ulike grunnlagstall av samme årgang ved å imputere flere tall samlet før estimat beregnes. Dette er også gjort i SAS på vann og avløp, som har flere slike tilfeller av total-delmengde problematikker knyttet til seg. Derfor er også her eksterne estimat vurdert som den mest hensiktsmessige og korrekte måten å beregne estimater for denne type grunnlagstall.

Det kan imidlertid legges til avslutningsvis at der hvor det er benyttet eksterne estimater på vann og avløp, så er det hittil ikke frigitt estimater for kommunegrupper, kun fylke og landet (med og uten Oslo). Bakgrunnen til dette er at kommunegrupperingen endrer seg relativt hyppig, omtrent hvert 5. år, og en tilbakeberegning av eksterne estimater for alle årgangene da vanskelig lar seg prioritere å gjennomføre desentralisert på en liten fagseksjon.

### **Regnskapsdata**

Standard estimering av regnskapsdata, på nasjonalt og regionale nivåer, skjer ved at en og en variabel estimeres uavhengig av hverandre, med folkemengde som forklaringsvariabel og med standard nedre grenseverdier på +/- 2. Metoden undersøker om noen kommuner avviker i forhold til modellen og de andre kriteriene. Hvilke enheter (kommuner) som blir betraktet som avvikende og tatt ut av estimeringsgrunnlaget varierer derfor fra variabel til variabel. Dette resulterer i manglende sammenhenger mellom enkeltvariabler og resultatbegrep og totalbegrep i regnskapet. Eksempelvis ble estimerte verdier for eiendomsskatt på boliger og fritidseiendommer og for eiendomsskatt på annen eiendom ikke lik estimert sum eiendomsskatt. Tilsvarende gjaldt gjennomgående for alle enkeltvariabler som inngår i resultatbegreper og totalbegreper.

Løsningen på problemet er å sette et felles kriterium for alle variablene for å kaste data ut av estimeringsgrunnlaget. Her innebærer det at grenseverdiene for avvikende enheter er satt såpass høyt (+/- 10) at samtlige data inngår i estimeringsgrunnlaget, og at en og samme estimeringsmetode (ratemodellen) og forklaringsvariabel (folkemengde) benyttes på alle variablene.

Det er ikke gitt at folkemengde er den beste forklaringsvariablen på alle variablene. Blant annet er det ingen nødvendig sammenheng mellom for eksempel eiendomsskatt og folkemengde. Eiendomsskatt er en kommunal skatt som den enkelte kommune kan velge å skrive ut, og de kan velge å skrive ut eiendomsskatt til bare næringslivet eller til både næringslivet og privatboliger.

Det er sjelden vi ser ekstremverdier i regnskapet, og det burde derfor ikke være noe stort problem at alle rapporterte data inngår i estimeringsgrunnlaget. Den valgte løsningen medfører dog at kvaliteten på estimatet på enkeltvariabler ikke er optimal, da det godt kan være at andre estimeringsmetoder og forklaringsvariabler i enkelte tilfeller vil passe bedre enn det felleskriteriet tilsier. Samtidig er hensynet til interne sammenhenger mellom enkeltvariabler og resultat- og totalbegrepene så viktig at det må gå foran. Se også vurderingene i kap. 5.1.2.

Med disse kriteriene lagt til grunn estimeres og publiseres samtlige regnskapsvariabler.

Se også vedlegg C: Notat Iso 16.november 2011 ”Landstall i KOSTRA – krav til konsistens”

## 5.2. Egne gjennomsnittsberegninger

I KOSTRA var det ønskelig å få på plass igjen funksjonalitet for å lage egne gjennomsnitt. Egne gjennomsnitt vil si at brukerne selv kan velge hvilke regioner (kommuner/fylkeskommuner) som de vil skal inngå i snittberegningen.

Tidligere løsning for beregning av egne gjennomsnitt (og alle andre snitt) var basert på innrapporterte data. Sensitive data ble prikket fra visning, og inngikk ikke i snittene.

I dag estimeres verdier på nasjonalt og regionale nivåer for å håndtere frafall i rapporteringen.

Ved utvikling av ny funksjonalitet for egne snittberegninger var det viktig å avklare hvilke spilleregler som skal gjelde for visning av snittene. Det måtte tas hensyn til at sensitive data ikke skal vises direkte eller indirekte, selv om de er estimerte, især dersom estimatet ganske enkelt består i at (et sensitivt) fjorårstall er direkte imputert. Det er heller ikke ønskelig å vise estimerte verdier for enkeltkommuner uansett type data. Det måtte derfor settes noen kriterier slik at det ble umulig å utlede estimert verdi for enkeltkommuner.

Funksjonaliteten for egne gjennomsnittsberegninger vil ikke kunne omfatte data som er estimert utenfor faktaark-admin, Dette er fordi datagrunnlaget ikke er tilgjengelig i faktaark-admin, eller at estimering skjer med forbehold som ikke finnes i de tilgjengelige metodene.

### Generell løsning for egne gjennomsnittsberegninger

Funksjonaliteten for egne gjennomsnitt for alle typer data som ikke er sensitive fungerer slik at egne gjennomsnitt nektes utført dersom det i den valgte gruppen mangler data for en (og bare en) kommune. Dersom to eller flere av de valgte kommunene mangler data vil det ikke være mulig ut fra summen å avlede verdien for hver av de manglende kommunene.

### Egne gjennomsnittsberegninger for sensitive data

For sensitive data generelt og for helsestatistikken og i sær IPLOS-dataene spesielt, vil den generelle løsningen ikke fungere fordi det kun er noen få personer som har tilgang til uprikkede tall. All estimering av absolutte grunnlagstall for å lage regionale summer gjøres på seksjonen. Funksjonaliteten for egne snitt i KOSTRA forutsetter tilgang til estimerte og uprikkede verdiene på kommunenivå.

I folketellingen har man brukt et avrundingsprogram, hvor verdiene 1 og 2 (mindre enn 3) avrundes til 0 eller 3 tilfeldig. Fra tabellen trekkes ut alle celler som har verdi mindre enn 3, og regner ut totalen for disse små tallene, f.eks 27. Så avrundes verdien til 24 eller 30 tilfeldig. Deretter settes tre av verdiene til 3 og de andre til 0 – hvilke som skal få hhv. 3 og 0 velges tilfeldig. Avvik sjekkes mot originaltabellen til man synes at avrundingene er ok. Tallene legges så tilbake i tabellen.



En slik løsning kan muligens være metodisk bedre enn dagens praksis med å prikke tall, men det vil få store konsekvenser dersom man skal gå over til en slik løsning i KOSTRA for alle sensitive data, både fordi de korrekte totaltallene da uthules/bortfaller og fordi det da vil dukke opp ”fiktive” tall for enkeltkommuner som disse ikke vil kjenne seg igjen i. Å vurdere alle effektene av en slik endring, og eventuelt utvikle dette til en generell løsning, faller klart utenfor rammene for dette prosjektet.

For pleie- og omsorgsdata, er det planlagt oppstart av overgang til tilfeldig avrunding, fortrinnsvis ved bruk av standard programvare, i løpet av høsten 2012.

For sosialhjelp, introduksjonsstønad, barnevern og kvalifiseringsstønad er det ennå ikke tatt stilling til eventuelle metoder for å beregne egne snitt. I utgangspunktet har det vært skepsis mot å benytte avrundede små tall, da dette innebærer en feilinformasjon om det faktiske nivået i de berørte kommunene.

### **5.3. Behovet for å skille mellom null, blank og ennå ikke innsendte data**

Det finnes enkelte tjenester/ordninger som ikke alle kommunene har. Disse kommunene skal naturlig nok ikke inngå i estimeringsgrunnlaget. Det kan være utfordrende å få tak i opplysninger om hvilke kommuner dette gjelder, og så bruke disse opplysningene for å unngå at estimatene blir for høye for de aktuelle variablene.

Ved publisering av data for 2011 var det ikke funnet noen løsning på dette. Dersom ikke statistikkansvarlig selv satt med en dekkende oversikt over slike kommuner, måtte estimerer droppes eller de ble liggende for høyt. For 2012-data ble det først vurdert å registrere slike frafall på skjema for sensitive opplysninger (KOSTRA-skjema 999). I stedet er det nå foreslått at kommunen skal gjøre en avkryssing på de aktuelle skjema for at ordningen ikke finnes hos dem, og så sende skjemaet inn.

Et annet problem er at det finnes ikke-utfylte felt i de innsendte dataene. Dette innebærer at enkelte felt(er) i et skjema, der det skulle stått verdien ”null”, ikke er fylt ut. Ved estimering er det vanskelig å skille mellom ikke utfylt felt i et innrapportert skjema og skjemaer som ennå ikke er innsendt. Ved estimering må skjemaer som er innsendt ha tall i alle felt. Ikke utfylte felt må gis verdien 0 for å ikke inngå i estimeringsgrunnlaget.

For sosialtjenestedata er dette håndtert slik:

Ved årets første prøvekjøring lå årsverkstall fra skjema 7 og 8 fortsatt alt for høyt, fordi KOSTRA (som i fjor) tolket ”blankt” som ”ikke rapportert”, og derved estimerte verdier som skulle vært satt til 0.

Problemet ble først vurdert løst ved å legge inn de aktuelle ”nuller” i et eget excel-ark ved tilbakeføringen av data til KOSTRA. Løsningen ble i stedet å endre de aktuelle formlene i faktaark-admin. For kommunale sysselsettingstiltak (funksjon 273) innebar dette for eksempel at formelen ble endret fra ”`ssb7_sumhe273`” til ”`if (antskjema("7")){ 0+ ssb7_sumhe273 }`”

Etter dette ble estimatene 15.03 riktige.

Det finnes også andre typer data som har de samme problemstillingene, hvor dette ennå ikke er løst innenfor rammene av faktaark-admin.

Som alternativ til løsningen som ble valgt for sosialtjenestedata, kan det utvikles en generell løsning hvor alle ikke-utfylte numeriske felt i innsendte skjema gjennomgående gis verdien 0. En evt. verdi i feltet må så eksplisitt legges til. Alternativt også en noe tilpasset løsning per skjema. Prosjektet anbefaler at man i

det videre arbeidet foretar en konsekvensutredning for å sjekke om disse løsningene kan brukes for alle typer data. I denne sammenhengen bør også de metodiske konsekvensene utredes, blant annet om alle de ”innsatte” 0-verdiene skal påvirke gjennomsnittene eller om de bør tas ut av estimeringsgrunnlaget og så legges til etterpå. En løsning på dette bør være på plass til neste publisering av data 15.3.2013.

#### **5.4. Estimering av ”rykk og napp” -data**

I dette kapitlet beskrives problematikk og løsning for data som ikke følger en naturlig utvikling fra år til år.

Noen tall i KOSTRA har blitt kalt ”rykk og napp”-data, grunnen til dette er at de er preget av stor tilfeldig variasjon både mellom kommuner og fra år til år. Eksempel på slike tall er ”fornyhet ledningsnett” (gjelder på både vann- og avløpsområdet) og investeringer. I tillegg fins det variabler som er opptellinger av sjeldne hendelser med stor variasjon og som ofte er et forholdsvis lite heltall. Et eksempel på slikt tall er antall kloakkstopper på ledningsnettet. Det er null eller positivt heltall, med stor variabilitet fra år til år.

Det forekommer også ”rykk og napp”-data innen for området ”Fysisk planlegging, byggesak, kulturminner, natur og nærmiljø”. Et eksempel kan være antall søknader om byggetillatelser, som avhenger av hvor mye byggeområde av ulike typer som til en hver tid er tilgjengelig, som igjen avhenger av hvor aktive kommunene er i sin arealplanlegging. Det er svært aktuelt å prøve ut metoder som beskrevet for VAR-området også for fysisk planlegging, byggesak etc.

#### **Estimering av ”rykk og napp” -data – dagens praksis på vann og avløp**

Problematikken rundt ”rykk og napp”-data har vært håndtert på avløp over flere år, på vann noe kortere tid. Tradisjonelt har man i DS-artiklene brukt justerte tall, og ikke ukritisk de tall som til enhver tid er blitt generert av faktaark-admin. Nå har imidlertid dette blir enklere – sett i lys av estimatene – så nå vil DS også på vann og avløp være samkjørt mot tallene i faktaarkene fullt ut.

Det benyttes eksterne estimater som metode for å lage estimater for disse ”rykk og napp”-dataene på lands- og regionnivå.

Måten ”rykk og napp”-data tradisjonelt er beregnet og som nå er grunnlag for eksterne estimater er illustrert med et eksempel nedenfor.

”Andel fornyelse av ledningsnettet i prosent” er et eksempel på ”rykk og napp”-data. Det er ikke gitt at alle kommuner fornyer ledningsnettet sitt hvert år, og da spesielt ikke mindre kommuner. Fornyelsen skjer etter et behov, som kan være uregelmessig og derfor ekstra utfordrende å estimere.

”Gamle snitt” slik fornyelsen tradisjonelt har vært beregnet i faktaark-admin ble gjort ved å summere sammen tall over fornyet ledningsnett og totalt ledningsnett, og dividert de første tallet på det andre – en enkel brøk. Feilen oppstod imidlertid siden brøken ble beregnet ut i fra en delmengde, nemlig kun de kommunene som faktisk har oppført et tall (null eller positivt tall). Resultatet blir en høyere fornyelsesprosent av ledningsnettet enn hva faktisk er tilfelle. Dette er forsøkt korrigert i eksterne estimater – dagens grunnlag for estimatene som nå er publisert for årgangene 2008..2011.

For de eksterne estimatene forutsetter man en sammenheng mellom totalt ledningsnett og lengde fornyet ledningsnett (selv om denne sammenhengen kan være noe ”tynn”). Man forutsetter også at dersom en kommune har rapportert skjema, men det er blankt i feltet for antall meter fornyet, at her mener man null meter fornyet ledningsnett. Dette er utgangspunktet.

Man deler så populasjonen i to grupper – de som har rapportert og de som ikke har rapportert. De som har rapportert vil ikke være gjenstand for justering – her antar man at rapporteringen er komplett (alle kommuner som har fornyet ledningsnettet sitt har ført opp en positiv verdi over null). Man kan videre dele inn gruppen av de som har rapportert skjema i to: de som har rapportert et tall på fornyelsen og de som ikke har det (rapportert blankt). Man deler dermed kommune-Norge inn i 3 ulike del-populasjoner eller grupper:

**Tabell 5.3. Fordeling av populasjonen i grupper for estimering av "rykk og napp"-data innen vann og avløp**

Gruppe	Rapportert skjema	Rapportert meter fornyelse
A	Ja	Ja (null eller høyere)
B	Ja	Nei (blankt)
C	Nei	-

På grunnlag av "rapportertgruppen" (A+B) beregner man to ulike faktorer som anvendes på de som ikke har rapportert skjema, altså gruppe C: (1) antall meter fornyet per total lengde ledningsnett og (2) svarandel på fornyelse blant de kommunene som faktisk har levert skjema.

Den første faktoren – meter fornyet per meter ledningsnett – er beregnet som median-verdi utelukkende blant gruppe A i tabellen ovenfor. For så å korrigere for at det er en del kommuner som har svart blankt blant de som sendt inn skjema, har man behov for faktor nummer to – svarandel på fornyelse. Denne beregnes som en brøk mellom 0 til 1, og utgjør forholdet  $A/(A+B)$ , dvs. andel kommuner som har rapportert en verdi for fornyelse blant alle de som har levert skjema.

Resultatet anvendes deretter på alle gruppe C-kommuner via følgende formel:

*Lengde fornyet ledningsnett = total lengde ledningsnett \* ant. meter fornyet per meter ledningsnett \* svarandel på fornyelse av kommuner som har levert skjema.*

Total lengde ledningsnett er normal basert på imputering av ledningsnettdata fra tidligere år, en informasjon SSB allerede har relativt god oversikt over. Det er lite sannsynlig at total lengde ledningsnett endrer seg betydelig fra et år til et annet, så denne tilnærmingen antas å være god.

Til slutt summerer man enkelt opp rapporterte verdier for de kommuner som faktisk har rapportert skjema (gruppe A+B), sammen med de estimerte verdiene kommuner som ikke har rapportert (gruppe C). Summert vil gruppe A og C normalt komme ut med en positiv verdi større enn null, mens gruppe B regnes i sin helhet som null i estimatet.

Det kan legges til at svarprosenten på skjemaet er relativt høy, slik at korrigeringen som gjøres på gruppe C ikke utgjør spesielt stor andel av det endelige estimatet på lands- eller regionnivå.

Tilsvarende metode er benyttet mot andre "rykk og napp"-data på vann- og avløpsområdet.

Hvis vil skulle legge til rette for denne metoden i faktaark admin, må det bli mulig å bruke en forklaringsvariabel der manglende verdier er imputert med forrige års verdier. I denne estimeringen er det brukt en ratemodell, der regresjonskoeffisienten er beregnet ut fra medianen og ikke summer som er det normale. Ved stor variasjon i data kan dette være lurt, da medianen ikke så lett blir påvirket av ekstreme verdier.

## Testing av lineær-regresjon og poisson-regresjon på antall kloakkstopp

I regresjonsmodellene i KOSTRA er det antatt at støyleddet er normalfordelt. For ”rykk og napp”-data kan det hende at denne antagelsen ikke holder mål. Enkel lineær regresjon og poisson regresjon er testet ut. Resultatene fra testingen av metodene er beskrevet i vedlegg B: Metodebeskrivelse.

- Lineær regresjon* Det er testet lineær regresjon for antall kloakkstopp med lengde på ledningsnett og antall kloakkstopp forrige år som forklaringsvariabler. Begge disse variablene kan hver forklare drøyt 50 prosent av variabiliteten i dataene. Det er svært mange tall som mangler i datasettet, noen på grunn av at manglende verdi skulle ha vært null. I regresjonsmodellene er det bare brukt data der det ikke er missing verdier for avhengig variabel og forklaringsvariabel, det vil si i analysen med lengde på ledningsnett, 288 observasjoner og for analyse med forrige år som forklaringsvariabel 252 observasjoner. Kontrollplottene viser at det er et mønster i residualplottet, med en klar økning i variabilitet jo lengre ledningsnett er eller antall kloakkstopp forrige år. Testen av normalfordelingsantagelsen, Q-Q plottet, viser at antagelsen om normalfordelte støyledd ikke er god.
- Poissonregresjon* I stedet for å anta at støyleddet er normalfordelt, er det mulig å modellere antall kloakkstopp som poissonfordelt. Antagelsen om poissonfordeling blir ofte brukt på hendelser som er sjeldne og tellbare og uten en øvre grense. Antall kloakkstopp er nettopp slike sjeldne tellbare hendelser. Antagelsen om poissonfordeling blir brukt i en regresjonsmodell, generaliserte lineære modeller. Det blir først testet med lengde på ledningsnett som forklaringsvariabel. Lengde på ledningsnett er signifikant i regresjonen, det vil si det er en sikker sammenheng mellom lengde ledningsnett og antall kloakkstopp. Poissonregresjonen kan også bare forklare omtrent 50 prosent av variabiliteten i data. Kontrollfigurene viser også at denne regresjonsmetoden har en noe mangelfull tilpasning til dataene.

I en ny regresjonsmodell ble det lagt til to nye forklaringsvariabler: lengde nytt ledningsnett og lengde fornyet ledningsnett. Alle forklaringsvariablene er signifikante. Det betyr at sammenhengen med antall kloakkstopp er ganske klar, jo lengre ledningsnett jo flere kloakkstopp. Hvis ledningsnettet er nytt eller fornyet minker antall kloakkstopp. Denne analysen er basert på 198 observasjoner, da det er flere missingverdier på fornyet og nytt ledningsnett enn på total lengde ledningsnett. For resultater av regresjonene, se vedlegg B: Metode. Lengde på ledningsnett er en god variabel for å estimere antall kloakkstopp, både når vi bruke lineær-regresjon og poisson-regresjon. Lengde på ledningsnettet er det ikke fulltelling på, så hvis den skal brukes i en regresjonsmetode må den først gjøres komplett ved for eksempel imputering av tidligere års verdier.

## Konklusjon

Estimering av ”rykk og napp”-data kan være vanskelig. Lineær regresjon/ratemodell kan gi adekvate resultater og bli brukt når variablene er kontinuerlige. Hvis variabelen er en tellbar sjelden hendelse kan poissonregresjon brukes og gi adekvate resultater. Dette er imidlertid en metode som få personer i statistikkproduksjonen har kjennskap til og som derfor kan være utfordrene og bruke. For vanskelige data er det alltid viktig å finne en god forklaringsvariabel, da dette kan ha stor betydning for estimeringen. For å få best mulig resultat har Miljøstatistikk valgt å bruke en forklaringsvariabel som forklarer godt, men som det ikke er fulltelling på. De har løst dette ved å imputere manglende verdier i forklaringsvariablen før de har brukt den i modellen. I tillegg bruker de en robust form for ratemodell, med justering for ”blanke” svar.

## 6. Anbefalinger til videre utvikling

I dette kapitlet beskrives de anbefalingene prosjektet har for en videre utvikling av systemet for beregning av nasjonale tall.

<i>Usikkerhetsmål</i>	Prosjektgruppen anbefaler at man utstyres grunnlagsdataene (mellomregningene) med variasjonskoeffisient som et mål på hvor treffsikkert det enkelte estimatet er. Variasjonskoeffisient er valgt fordi dette er et enkelt og intuitivt forståelig mål for brukerne. For internt bruk anbefaler prosjektet at variasjonskoeffisienten vises i estimatresten som en egen kolonne. Derved kan statistikkansvarlige vurdere usikkerheten til estimatet samtidig med at selve beregningen testes. Det er også ønskelig at usikkerhetsmålet kan vises eksternt, fordi dette antagelig kan bidra til ønsket varsomhet i bruk av tallene. Hvordan dette best kan gjøres er hittil ikke nøyte vurdert, men det bør være enkelt tilgjengelig for brukerne, gjerne vist grafisk.
<i>Estimerte verdier for årgangene 2001..2007</i>	For en del av statistikkene i KOSTRA er det mangler i datasettene for en eller flere av de tidligste årgangene også. 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. Det er en del problemstillinger knyttet til estimering av tidligere årganger som må løses, typisk skille mellom null, blank og ikke rapportert (se eget punkt). For flere av metodene bør det også bestemmes hvor langt tilbake i tid det skal letes etter sist rapportert verdi. Det må også avklares hvilke kriterier som skal ligge til grunn for å finne et eldst mulig nullpunkt, hvor det er tilstrekkelig med data med en tilstrekkelig kvalitet, til at man kan estimere eldre årganger basert på senere års rapporteringer.
<i>Kvalitetssikre bruken av metoder og forklaringsvariabler til estimering</i>	Estimeringsfunksjonaliteten er på plass, men det bør jobbes videre med kvalitetssikring av bruken av metoder og forklaringsvariabler for å sikre best mulige estimater. 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, bør undersøkes nærmere.
<i>Vurdering av default estimeringsmetode</i>	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.  Det bør i den forbindelse gjøres en nærmere vurdering av treffsikkerheten til de ulike estimeringsmetodene på noen utvalgte fagområder. Dette kan gjøres ved å sammenligne estimert verdi for kommuner som mangler data 15.3 med de data som senere sendes inn.
<i>Skille mellom null, blank og ikke innsendte data</i>	Til neste publisering bør det finnes en metode for å unngå at det estimeres verdi for kommuner som ikke har en bestemt ordning. Det legges foreløpig opp til at kommunen ved kommende innsending skal gjøre en avkryssing på de aktuelle skjemaene for at ordningen ikke finnes hos dem, og så sende skjemaene inn.  Et annen problemstilling er hvordan man kan unngå feiltolkning av blanke felt i skjemaer som er rapportert, slik at ikke den estimerte verdien blir for høy. To mulige løsninger på dette er skissert under kapittel 5.3. Også her er det ønskelig at en løsning er på plass til neste publisering av data 15.3.2013.
<i>Gjennomsnittsberegninger for indikatorer som ikke har brøk som formel</i>	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 estimater for slike variabler. Det bør vurderes om og eventuelt hvordan man skal beregne aritmetiske gjennomsnitt for slike variabler.

*Egne gjennomsnitt for sensitive data*

Det bør avgjøres om det er mulig/ønskelig å tilby funksjonalitet for egne gjennomsnittsberegninger på sensitive data, og i så fall etter hvilken metode undertrykking av små tall skal gjøres. Det er skissert en alternativ løsning med tilfeldig avrunding av små sensitive tall som kanskje kan benyttes på helsestatistikken – på data fra levekårsstatistikkene er det mindre sannsynlig at denne løsningen er tilfredsstillende.

*Nye ssb.no og føringer på publiseringen av estimerte tall*

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 estimater, for eksempel med farge, muligheten for å publisere variasjonskoeffisient osv. Det er viktig at dette følges opp videre.

## Vedlegg A: Prosjektskriv

### Prosjektskriv

Prosjektnavn	System for beregning av nasjonale tall i KOSTRA_del III		
Kortnavn	Nasjonale tall_III		
Planlagt startdato	01.11.2011	Planlagt sluttdato	15.10.2012
Oppdragsgiver	KOSTRAs styringsgruppe v/Olav Ljones		
Ansvarlig enhet	Seksjon for offentlige finanser v/Irene Arnesen		
Er prosjektet del av et større hovedprosjekt?	<input type="checkbox"/> Ja	<input checked="" type="checkbox"/> Nei	
Hvis ja, oppgi navn og prosjektnr 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 to 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. Det gjenstår fremdeles en del arbeid før vi har et fullstendig regime med estimerte verdier.

Dagens gjennomsnittsberegninger som baseres på de innrapporterte dataene må erstattes med estimerte gjennomsnitt. I tillegg savnes tidsserier for landstall samt estimerte verdier på lavere regionale nivåer.

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) Publisere estimerte verdier på lavere regionale nivåer 15.03.2012
- 2) Publisere snitt basert på estimerte nasjonale verdier 15.3. 2012, ev 15.06.2012
- 3) Publisere estimerte nasjonale verdier på tidligere årganger 15.3. 2012
- 4) Publisere estimater for regnskap som ivaretar de nødvendige sammenhenger i regnskapet, på alle detaljerings- og aggregeringsnivåer 15.03.2012
- 5) Publisere funksjonalitet for egne snittberegninger 15.06.2012 (ev. 15.03.2013)
- 6) Finne en løsning for å utstyre estimatene med et usikkerhetsmål for internt bruk
- 7) Finne en løsning for estimering av data som ikke følger en naturlig utvikling fra år til år ("rykk- og napp"-data)



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

### **Faglige problemstillinger**

#### **Tidsserier:**

For flere fagområder finnes det mangler i datasettene for tidligere år. For de områder som har fulltelling vil dette gi sumtall på nasjonalt nivå. Løsningen for estimering bør utvides til å gjelde også for tidligere årganger. Rent teknisk ligger mye tilrette allerede, men det må settes av tid til kvalitetssikring av dataene. Estimater for tidligere årganger bør kunne sjekkes i god tid før dataene for 2011 skal godkjennes.

#### **Snitt basert på estimerte tall:**

Dagens snitt er basert på data som er levert. De vil kunne avvike fra landsestimatene og bør erstattes med snitt som baseres på de estimerte dataene. Resultatmål 2 og 3 må ses i sammenheng fordi en vil ha behov for estimerte tall på fylkesnivå og på kommunegruppenivå før en kan beregne snitt basert på estimater fra disse nivåene. Også her ligger mye tilrette teknisk sett. Det må settes av tid til kvalitetssikring av de nye snittene. Selve sjekken bør kunne gjøres i forhold til de snitt som ligger i dagens publiseringer. Det må legges inn begrensning på valg av kommuner i beregningen av "Egne gjennomsnitt" for å sikre at brukerne ikke kan beregne data for en kommune.

#### **Estimering på lavere regionalt nivå:**

Estimering på lavere regionalt nivå enn landet vil innebære en større utfordring enn på nasjonalt nivå. Store mangler vil kunne innebære at estimeringen blir usikker. Det bør settes av ekstra tid til kvalitetssjekk i forkant av publiseringen.

#### **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 variere etter hva slags data en ser på. Et mål på usikkerhet vil gi brukerne en pekepinn på i hvor stor grad de kan sette lit til det estimerte tallet.

#### **Estimering som ivaretar sammenhenger i regnskap:**

Dagens løsning estimerer ett og ett regnskapsbegrep for seg. Dette medfører at ulike regnskaper blir vurdert som avvikende og grunnlaget for estimering blir forskjellig fra begrep til begrep. Det må finnes en løsning som velger felles estimeringsgrunnlag (avvikende regnskaper) for alle begreper.

#### **Estimering av data som ikke følger naturlig utvikling:**

Det bør etableres en løsning for estimering av data som ikke følger naturlig utvikling fra år til år ("rykk- og napp"-data), eks. antall kloakkstopp, investeringer i kommunene.

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

## Vedlegg B: Metodebeskrivelse

### Beregningsmetode for variasjonskoeffisienten

Vi har følgende 3 regresjonsmetoder:

- Regresjon - enkel lineær regresjon:  $y = \alpha + \beta x_i + \varepsilon_i$ ;  $E(\varepsilon_i) = 0$ ,  $\text{VAR}(\varepsilon) = \sigma^2$
- Rate modell:  $y_i = \beta x_i + \varepsilon_i$ ;  $E(\varepsilon_i) = 0$ ,  $\text{Var}(\varepsilon_i) = x_i \sigma^2$
- Rate modell med konstantledd:  $y_i = \alpha + \beta x_i + \varepsilon_i$ ;  $E(\varepsilon_i) = 0$ ,  $\text{Var}(\varepsilon_i) = x_i \sigma^2$

#### Variasjonskoeffisienten:

Beregningsmetode for variasjonskoeffisienten er standardfeilen delt på totalen ganget med 100.

$$\text{CV}(\hat{T}) = \frac{\text{SE}(\hat{T})}{\hat{T}} \times 100$$

Hvordan totalen, T, blir estimert er beskrevet i forrige prosjekt, se notat 50/2011.

Notasjon:

i - observasjon

h - KOSTRAgruppe

x - verdi på forklaringsvariabel

y - verdi på mellomregning

T - total

T<sub>h</sub> - total verdi i kostragruppe h

$\bar{T}_{sh}$  - Estimert verdi i kostragruppe h

s - svarutvalg

s<sub>h</sub> - svarutvalg i kostragruppen h

n<sub>h</sub> - antall svar i kostragruppen h

N<sub>h</sub> - antall kommuner i kostragruppen h

x<sub>hi</sub> - forklaringsvariabel for kommune 'i' i kostragruppe h

$\bar{x}_{sh}$  - gjennomsnittet av forklaringsvariablene for de kommunene som har svart i kostragruppe h

$\bar{x}_h$  - gjennomsnittet av forklaringsvariablene for de kommunene i kostragruppe h

$\bar{y}_{sh}$  - gjennomsnittet av mellomregning for de kommunene som har svart i kostragruppe h

#### Ratemodell

Estimert parameter:

$$\hat{\beta}_h = \frac{\sum_{i \in s_h} y_{hi}}{\sum_{i \in s_h} x_{hi}} = \frac{y_{s_h}}{x_{s_h}}$$

Der

$$y_{sh} = \sum_{i \in s_h} y_{hi}$$

Og

$$x_{sh} = \sum_{i \in s_h} x_{hi}$$

Standardavvik til observasjonene:

$$\hat{\sigma}_h^2 = \frac{1}{n_h - 1} \sum_{i \in s_h} \frac{(y_{hi} - \hat{\beta}_h x_{hi})^2}{x_{hi}}$$

Standardfeil til totalen:

$$SE(\hat{T}_{S_h} - T_h) = X_h \sqrt{\frac{X_h - X_{S_h}}{X_h}} \frac{\hat{\sigma}_h}{\sqrt{X_{S_h}}}$$

$X_h = \sum_{i \in h} x_{hi}$  er summen av alle forklaringsvariablene i kostragruppe h

### Regresjon – enkel lineær regresjon

$$\text{Estimat for } \beta: \hat{\beta}_h = \frac{\sum_{i \in S_h} (x_i - \bar{x}_{sh})(y_i - \bar{y}_{sh})}{\sum_{i \in S_h} (x_i - \bar{x}_{sh})^2} = \frac{\sum_{i \in S_h} x_i y_i - n_h \bar{x}_{sh} \bar{y}_{sh}}{\sum_{i \in S_h} x_i^2 - n_h \bar{x}_{sh}^2}$$

$$\text{Estimat for } \alpha: \hat{\alpha}_h = \bar{y}_{sh} - \hat{\beta}_h \bar{x}_{sh}$$

Standardavvik til observasjonene:

$$\hat{\sigma}_h^2 = \frac{1}{n_h - 2} \sum_{i \in S_h} (y_{hi} - \hat{\alpha}_h - \hat{\beta}_h x_{hi})^2$$

Standardfeil til totalen:

$$SE(\hat{T}_{S_h} - T_h) = N_h \sqrt{\frac{N_h - n_h}{N_h} + \frac{(\bar{X}_h - \bar{x}_{sh})^2}{v_{sh}^2}} \frac{\hat{\sigma}_h}{\sqrt{n_h}}$$

$$\text{Her er } v_{sh}^2 = \frac{\sum_{i \in S_h} (x_{hi} - \bar{x}_{sh})^2}{n_h}$$

### Rate med konstantledd

$$\hat{\alpha}_h = \bar{y}_{sh} - \hat{\beta}_h \bar{x}_{sh}$$

$$\hat{\beta}_h = \frac{\sum_{i \in S_h} \{(y_i - \bar{y}_{sh}) / x_i\}}{\sum_{i \in S_h} \{(x_i - \bar{x}_{sh}) / x_i\}}$$

Standardavvik til observasjonene:

$$\hat{\sigma}_h^2 = \frac{1}{n-2} \sum_{i \in sh} \frac{(y_{hi} - \hat{\alpha}_h - \hat{\beta}_h x_{hi})^2}{x_{hi}}$$

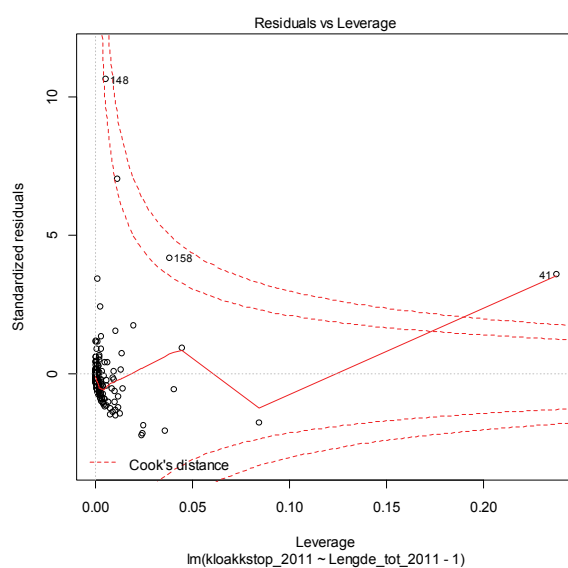
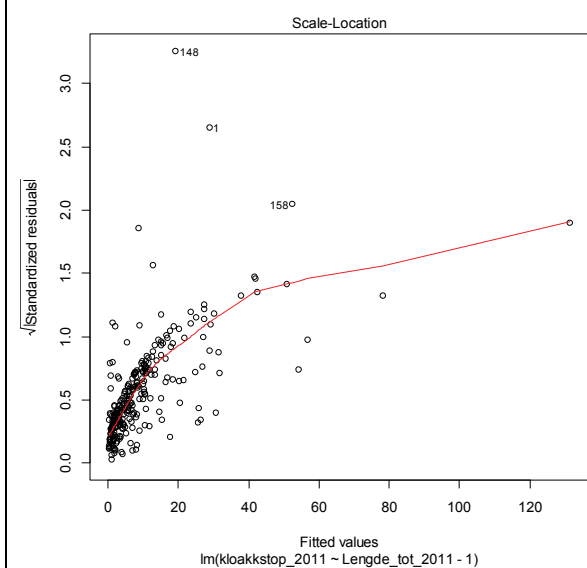
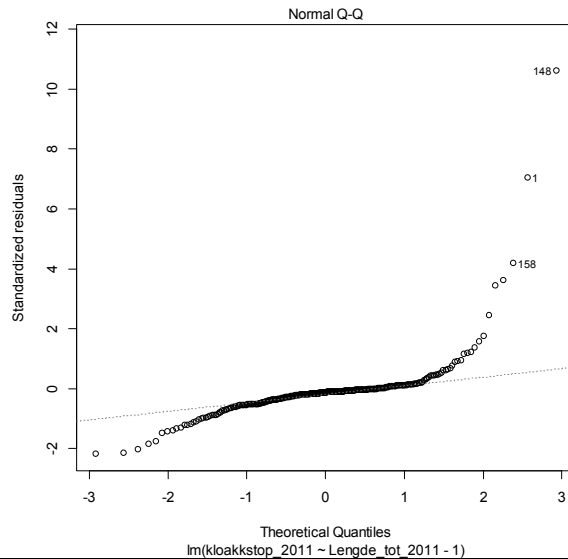
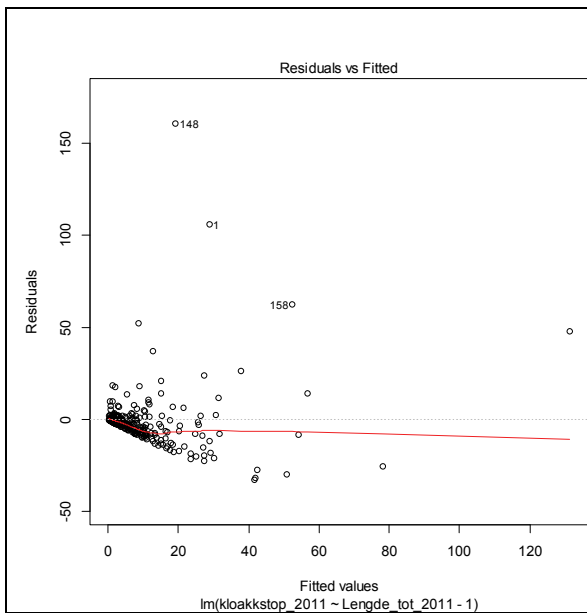
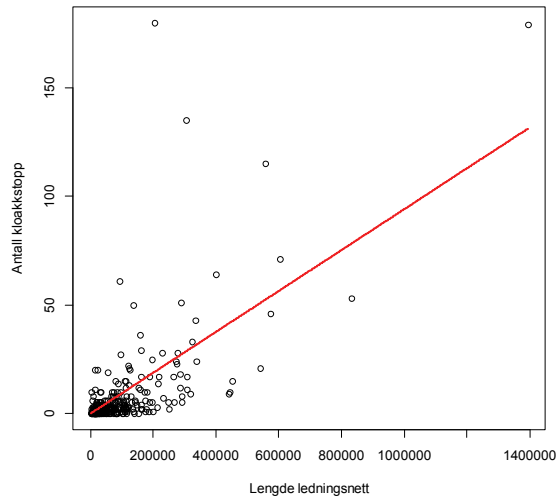
## Resultater fra estimering av 'rykk og napp' data Kloakkstopp som funksjon av lengde på ledningsnett

Call:  
lm(formula = kloakkstopp\_2011 ~ Lengde\_tot\_2011 - 1, data = avdata)

Residuals:  
Min 1Q Median 3Q Max  
-32.582 -5.629 -1.736 0.147 160.722

Coefficients:  
Estimate Std. Error t value Pr(>|t|)  
Lengde\_tot\_2011 9.410e-05 5.296e-06 17.77 <2e-16 \*\*\*  
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 15.15 on 288 degrees of freedom  
(106 observations deleted due to missingness)  
Multiple R-squared: 0.523, Adjusted R-squared: 0.5213  
F-statistic: 315.7 on 1 and 288 DF, p-value: < 2.2e-16



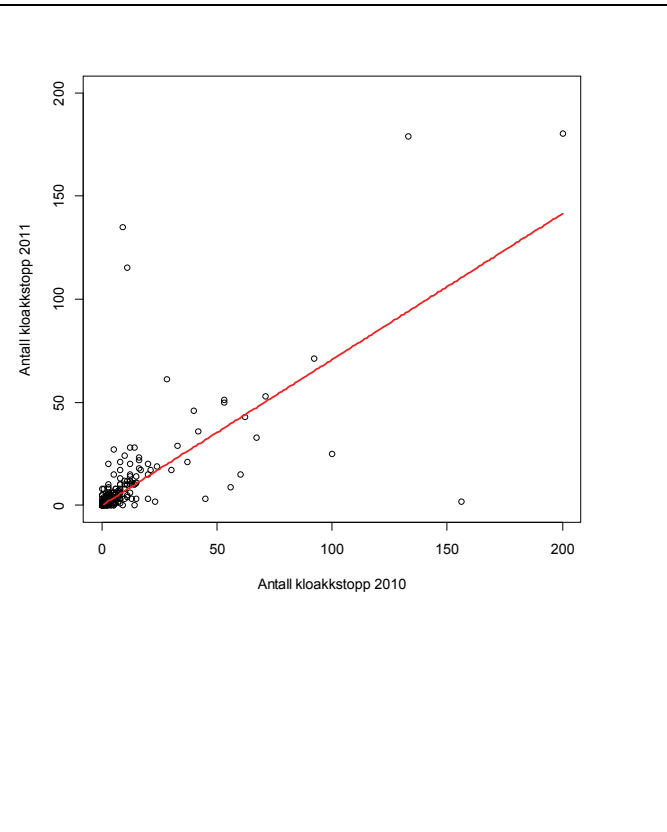
### Kloakkstopp som funksjon av kloakkstopp forrige år

**Call:**  
**lm(formula = kloakkstopp\_2011 ~ kloakkstopp\_2010 - 1, data = avdata)**

Residuals:  
 Min 1Q Median 3Q Max  
 -108.309 -0.414 0.293 2.293 128.636

Coefficients:  
 Estimate Std. Error t value Pr(>|t|)  
 kloakkstopp\_2010 0.70711 0.04041 17.5 <2e-16 \*\*\*  
 ---  
 Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 15.47 on 252 degrees of freedom  
 (142 observations deleted due to missingness)  
 Multiple R-squared: 0.5485, Adjusted R-squared: 0.5467  
 F-statistic: 306.2 on 1 and 252 DF, p-value: < 2.2e-16



### Poissonregresjon

**Call:**  
**glm(formula = kloakkstopp\_2011 ~ Lengde\_tot\_2011, family = quasipoisson(), data = avdata)**

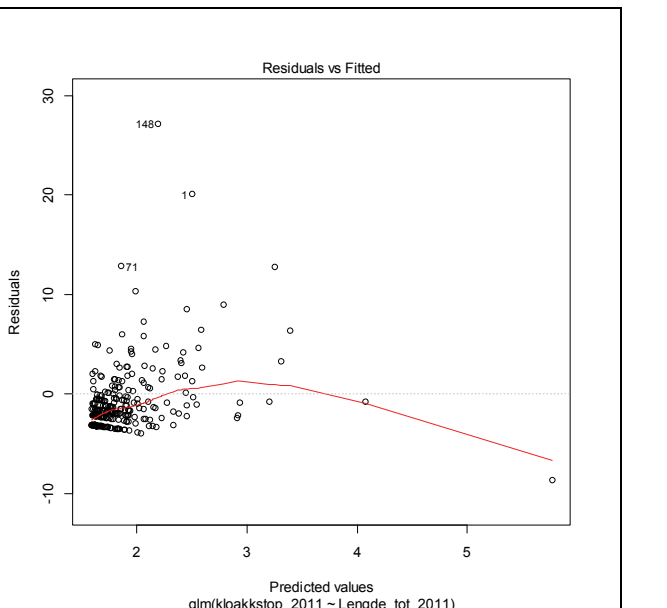
Deviance Residuals:  
 Min 1Q Median 3Q Max  
 -8.6284 -3.1398 -1.7025 -0.3191 27.1398

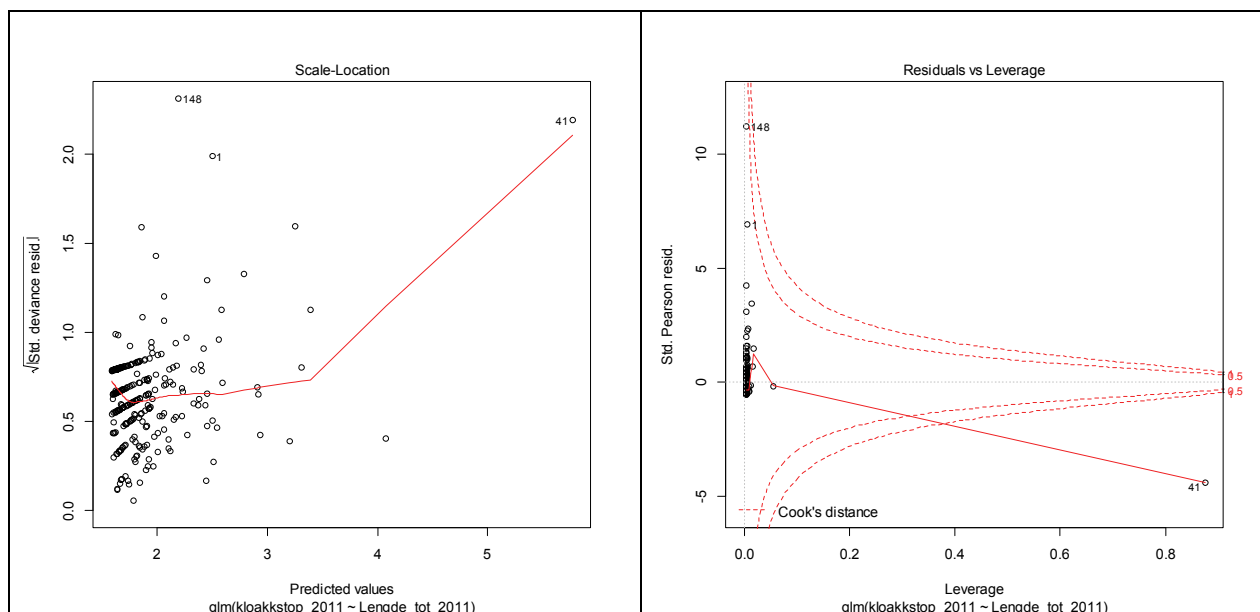
Coefficients:  
 Estimate Std. Error t value Pr(>|t|)  
 (Intercept) 1.582e+00 1.293e-01 12.23 <2e-16 \*\*\*  
 Lengde\_tot\_2011 3.003e-06 2.290e-07 13.11 <2e-16 \*\*\*  
 ---  
 Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for quasipoisson family taken to be 25.84472)

Null deviance: 5856.3 on 288 degrees of freedom  
 Residual deviance: 3614.1 on 287 degrees of freedom  
 (106 observations deleted due to missingness)  
 AIC: NA

Number of Fisher Scoring iterations: 6





```
glm(formula = kloakkstop_2011 ~ Lengde_tot_2011 + Lengde_ny_2011 +
  Lengde_for_2011, family = poisson(), data = avdata)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-7.8689	-2.7015	-1.5287	0.2458	25.7259

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	1.291e+00	3.781e-02	34.159	< 2e-16 ***
Lengde_tot_2011	6.931e-06	1.796e-07	38.598	< 2e-16 ***
Lengde_ny_2011	-2.318e-05	8.371e-06	-2.769	0.00563 **
Lengde_for_2011	-2.810e-04	1.192e-05	-23.578	< 2e-16 ***

---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

Null deviance: 4655.4 on 198 degrees of freedom  
Residual deviance: 2269.0 on 195 degrees of freedom  
(231 observations deleted due to missingness)  
AIC: 2841.1

Number of Fisher Scoring iterations: 6



Iso, 16. november 2011

## Landstall i Kostra - krav til konsistens

### • Bakgrunn

I notater 2003/46 er en metode for å produsere foreløpige landstall i Kostra basert på et utvalg av kommuner beskrevet. Denne metoden er tatt i bruk for å lage landstall også for endelige tall. I bruken av metoden har en støtt på et problem i forhold til kravene om sammenheng mellom forskjellige variabler. Problemet som er oppstått skyldes at metoden også undersøker om noen kommuner avviker i forhold til den valgte modellen og dersom en kommune avviker mer enn en gitt skranke kastes kommunen ut av den gruppa kommuner som brukes til å beregne tall for kommuner som ikke har levert. Siden variablene undersøkes en for en kan vil gruppen av kommuner som avviker mer enn skranken variere fra variabel til variabel. I så fall vil ikke lenger sammenhenger (lineære) mellom variablene for kommunene enkeltvis bevares i de kommunene der vi beregner tallene og dermed vil heller ikke summen (landstallene) for kommunene totalt følge kravet til sammenheng. Løsningen på dette er å lage et felles kriterium for å kaste kommuner ut av beregningskriteriet for alle variabler.

I avsnitt 2 har vi en kort beskrivelse av problemet, mens vi i avsnitt 3 legger fram et par forslag til løsning.

### • Problemstillingen

Vi antar at ett sett av egenskaper for hver kommune måles:

$$(1.1) \quad Y_{ik} \quad , \quad i = 1, 2, \dots, I; \quad k = 1, 2, \dots, K$$

I betegner antall variabler og K betegner antallet kommuner i Norge. For å forenkle notasjon og begrepsapparat har jeg sløyfet inndelingen i Kostra grupper.

Vi antar at Y-ene følger en ratemodell med befolkningen i kommunen som forklaring. Bruker X som betegnelse for folketallet i kommunene:

$$(1.2) \quad Y_{ik} = \beta_i X_{ik} + \epsilon_{ik} \quad \text{for } i=1, 2, \dots, I \text{ og } k=1, 2, \dots, K$$

Ratemodellen i (1.2) innebærer at vi har følgende uttrykk for forventning og varians:

$$(1.3) \quad E(\epsilon_{ik}) = 0 \quad , \quad \text{Var}(\epsilon_{ik}) = \sigma_i^2 X_{ik}$$

I tillegg antar vi at det er et knippe av lineære sammenhenger mellom variablene som må være tilfredstilt:

$$(1.4) \quad \sum_{i=1}^I c_{ij} Y_{ik} = 0 \quad \text{for } j=1, 2, \dots, J$$

Det mest vanlige vil være at c-ene er -1, 0 eller 1. Det enkleste eksemplet er at en av variablene er lik summen av de andre. En litt mer komplisert struktur har vi dersom variablene er organisert en tabell der cellene unntatt en i rader og kolonner summerer til den siste cellen.. I det tillet vil en av c-ene være lik -

1 (summen i raden eller kolonnen), de andre vil enten være null (bidrar ikke til summen) eller lik 1 (de andre cellene i en rad eller i en kolonne).

Da er vi kommet til det punktet der vi skal estimere de ukjente proporsjonalitetsfaktorene i ratemodellen. La oss nå dele kommunene i to grupper, de som leverer data og de som ikke har levert data:

$$(1.5) \quad \begin{array}{l} K_r \quad - \text{antall kommuner som har levert} \\ K_n \quad - \text{antall kommuner som ikke har levert} \end{array}$$

Da kan vi beregne proporsjonalitetsfaktorene i ratemodellen:

$$(1.6) \quad \hat{\beta}_i = \frac{\sum_{k=1}^{K_r} Y_{ik}}{\sum_{k=1}^{K_r} X_k} \quad \text{for } i=1, 2, \dots, I$$

Den metoden som brukes for å vurdere om verdiene for en kommune er avvikende er å beregne raten i (1.6) på nytt ved å holde den avvikende kommunen utenfor beregningen. La oss kalle denne kommunen for l:

$$(1.7) \quad \hat{\beta}_{i(l)} = \frac{\sum_{k=1, k \neq l}^{K_r} Y_{ik}}{\sum_{k=1, k \neq l}^{K_r} X_k} \quad \text{for } i=1, 2, \dots, I$$

For å undersøke om denne kommunen er avvikende undersøker vi forskjellen mellom den observerte og den predikerte verdien. Altså beregner vi først prediksjonen for kommune l:

$$(1.8) \quad \hat{Y}_{il(l)} = \hat{\beta}_{i(l)} X_l$$

Deretter sammenliknes differansen mellom prediksjonen og observasjonen. Vi denne sammenheng sløyfer vi utregningen av usikkerheten til denne differansen, vi fastslår kun at vurderingen av denne differansen er avhengig av usikkerheten. En kommune erklæres for avvikende dersom kontrollen i (1.9) slår ut:

$$(1.9) \quad \left| Y_{il} - \hat{Y}_{il(l)} \right| > C_1$$

Legg merke til at kriteriet er avhengig av hvilken variabel som måles siden spredingen av verdiene for kommunene varierer med hvilken variabel som måles. Da må kommunen tas ut av beregningen av raten dersom kontrollen i (1.9) slår ut. Dagens opplegg der en bruker kriteriet i (1.9) for hver variabel hver for seg fører til prediksjonene i (1.8) blir avhengig av forskjellige grupper av kommuner. Før en kommune der vi predikerer verdiene basert på det samme utvalget av kommuner vil alltid sammenhenger av type (1.4) holde. Først predikerer vi verdien til en kommune som ikke har levert data:

$$(1.10) \quad \hat{Y}_{ik_n(l)} = \hat{\beta}_{i(l)} X_{k_n} \quad \text{for } i=1, 2, \dots, I$$



Da ser med litt fingerferdighet at fortsatt er (1.5) sant:

$$(1.11) \quad \sum_{i=1}^I c_{ij} \hat{Y}_{ik_n(l)} = 0$$

Dersom vi fjerner forskjellige kommuner fra de enkelte variablene holder ikke lenger kriteriet i (1.11) generelt. Det samme er resultatet viss vi fjerner en gruppe av kommuner. Dersom det er en felles gruppe for alle variabler holder kravet i (1.5), men viss gruppen av kommuner varierer fra variabel til variabel er resultatet det samme.

### • Strategi for å fjerne kommuner for flere variabler

Vi kan velge mellom to metoder for å fjerne kommuner fra utregningen av ratene slik at 1.5 holder for alle kommuner der vi predikerer verdiene. Felles for begge strategier er at vi først velger ut noen variabler der vi bruker kriteriet i (1.9) for å undersøke hvilke kommuner som avviker signifikant fra modellen. Deretter skilles metodene. I den første velger vi ut de kommunene som tas ut gjennom (1.9) for minst en variabel. Et problem med denne metoden er at vi kan risikere at en Kostragruppe ikke har normale kommuner tilbake (ut fra (1.9)). Alternativet er å utvikle et kriterium som utnytter informasjonen fra testen i (1.9) for alle variabler samtidig i en flerdimensjonal test.

Litt formalisme for å beskrive de to strategiene. La oss først sette navn på den gruppen av variabler som skal testes ved bruk av (1.9).

$$(2.1) \quad V_T = \{i \mid \text{der variabel } i \text{ testes ved (1.9)}\}$$

La oss i tillegg sette navn på de kommunene som skal fjernes fra utregningen av ratene:

$$(2.2) \quad A = \{l \mid \text{der kommune } l \text{ holdes utenfor utregning av ratene}\}$$

Da kan vi formelt skrive opp de to kriteriene for ta kommuner ut av beregningene av ratene

$$(2.2) \quad l \in A \text{ viss } \left| \hat{Y}_{il(l)} - Y_{il} \right| > C_1 \text{ for minst en } i \in V_T$$

$$(2.3) \quad l \in A \text{ viss } \sum_{i \in V_T} (\hat{Y}_{il(l)} - Y_{il})^2 > C$$

Kriteriet C i 2.3 må bestemmes.

Forskjellen mellom kriteriet i (2.2) og kriteriet i (2.3) er at det i det første tilfellet vil vi fjerne en kommune viss den er avvikende på en av variablene, mens i det andre tilfellet vil først ta ut en kommune dersom den er avvikende på flere variablene. Det første kriteriet vil for lett ta ut kommuner sammenliknet med det andre kriteriet.

- **Referanser**

Leiv Solheim (2003) Foreløpige landstall i Kostra – Prinsipper, metoder, produksjon og eksempler.  
2003/46

## Vedlegg D: Detaljert oversikt over utarbeidelse av estimater i faktaark I. Vann og I. Avløp

Fargemarkeringen i tabellen nedenfor har følgende forklaring: gule tall er estimat som ble opprettet allerede i fjor, grønne tall er tillegg i år, mens røde felter er variabler som fremdeles mangler estimater.

### Faktaark I, vann – nivå 3 (grunnlagstall). Landstall.

	2011	Eksterne estimater	Merknader
Folkemengde i alt	4 985 870		
Direkte driftsutgifter (fra skjema 23)	3 282 261		
- direkte driftsutgifter funksjon 340	:		For lite datagrunnlag for å estimere landstall
- direkte driftsutgifter funksjon 345	:		For lite datagrunnlag for å estimere landstall
Kapitalutgifter hos ekstern produsent (del av direkte driftsutgifter)	67 839		
- kapitalutgifter hos ekstern produsent funksjon 340	:		For lite datagrunnlag for å estimere landstall
- kapitalutgifter hos ekstern produsent funksjon 345	:		For lite datagrunnlag for å estimere landstall
Henførbare indirekte driftsutgifter (fra skjema 23)	202 875		
- henførbare indirekte driftsutgifter funksjon 340	:		For lite datagrunnlag for å estimere landstall
- henførbare indirekte driftsutgifter funksjon 345	:		For lite datagrunnlag for å estimere landstall
Kalkulatoriske rentekostnader (fra skjema 23)	702 130		
- kalkulatoriske rentekostnader funksjon 340	:		For lite datagrunnlag for å estimere landstall
- kalkulatoriske rentekostnader funksjon 345	:		For lite datagrunnlag for å estimere landstall
Kalkulatoriske avskrivninger (fra skjema 23)	961 017		
- kalkulatoriske avskrivninger funksjon 340	:		For lite datagrunnlag for å estimere landstall
- kalkulatoriske avskrivninger funksjon 345	:		For lite datagrunnlag for å estimere landstall
Andre inntekter (fra skjema 23)	200 746		
- andre inntekter funksjon 340	:		For lite datagrunnlag for å estimere landstall
- andre inntekter funksjon 345	:		For lite datagrunnlag for å estimere landstall
Gebyrgrunnlaget (driftsutgifter+kapitalkostnader-andre inntekter, fra skjema 23)	4 947 537		
Gebyrinntekter (fra skjema 23)	4 890 068		
Avsetning til selvkostfond / dekning av fremført underskudd (fra skjema 23)	:		Ikke mulig å estimere landstall, da forklaringsvariabel mangler
Bruk av selvkostfond / fremføring av underskudd (fra skjema 23)	:		Ikke mulig å estimere landstall, da forklaringsvariabel mangler
Saldo selvkostfond per 1.1 (fra skjema 23)	:		Ikke mulig å estimere landstall, da forklaringsvariabel mangler
Alternativkostnad ved bundet kapital på selvkostfond eller framføring av unders	:		Ikke mulig å estimere landstall, da forklaringsvariabel mangler
Saldo selvkostfond per 31.12 (fra skjema 23)	:		Ikke mulig å estimere landstall, da forklaringsvariabel mangler
Lønnsutgifter	929 330		
Kjøp av varer og tjenester som inngår i komm. egenproduksjon	1 748 691		
Kjøp av varer og tjenester som erstatter komm. egenprod	656 907		
Stipulert årsgebyr (gjelder rapporteringsåret+1)	:		Estimerte landstall basert på den benyttede gebyrtypen publisert på nivå 2.
Gebyrsats ved målt forbruk (gjelder rapporteringsåret+1) (kr/m <sup>3</sup> )	:		
Fastledd ved todelt gebyrordning (gjelder rapporteringsåret+1)	:		
Antall kommunale vannverk	1 078	Ja	
Antall vannverk med sikkerhets- og beredskapsplan	915	Ja	
Antall innbyggere tilknyttet kommunal vannforsyning	4 191 615	Ja	
Antall innbyggere tilknyttet vannverk som forsyner flere enn 1000 personer	3 990 352	Ja	
Antall personer forsynt av grunnvann el. desinf. overflatevann som hovedkilde	:		
Antall personer forsynt av desinfisert overflatevann som hovedkilde	:		
Antall personer forsynt av grunnvann som hovedkilde	:		
Antall innbyggere forsynt av behandlet vann (behandlingstype 1+ type 2)	:		
E.coli: Antall innbyggere tilknyttet kommunalt vannverk med tilfredsstillende p	4 128 458	Ja	
Intestinale enterokokker: Antall innbyggere tilknyttet kommunalt vannverk med t	4 016 986	Ja	
Farge: Antall innbyggere tilknyttet kommunalt vannverk med tilfredsstillende pr	3 377 090	Ja	
PH: Antall innbyggere tilknyttet kommunalt vannverk med tilfredsstillende prøve	3 930 643	Ja	
Antall innbyggertimer totalt	:		
Antall innbyggertimer uten avbrudd i vannforsyningen	:		
Antall innbyggertimer med avbrudd i vannforsyningen, uansett årsak	:		
Antall innbyggertimer med avbrudd i vannforsyningen, ikke-planlagt	:		
Antall meter ledningsnett	43 869 419	Ja	
Antall meter ledningsnett med ukjent alder	4 480 320	Ja	

Antall meter fornyet ledningsnett, gjennomsnitt siste tre år	276 120	Ja	
Antall meter utskiftet/rehabiliteret ledningsnett	301 548	Ja	
Reguleringsvolum i høydebasseng og krisekilde (m3)	:		
Antall trykkøkningsstasjoner	2 287	Ja	
Antall lekkasjereparasjoner på nettet (antall/km ledningsnett)	0,10	Ja	
Total vannleveranse på kommunalt distribusjonsnett	740 041 561	Ja	
A1. Mengde vann til husholdningsforbruk	:		Grunnlagstall vil trolig utgå fra faktaark
A2. Mengde vann til forbruk i næringsmiddelindustrien	:		Grunnlagstall vil trolig utgå fra faktaark
A3. Mengde vann til forbruk i annen industri og næringsvirksomhet	:		Grunnlagstall vil trolig utgå fra faktaark
A4. Mengde vann til annet forbruk (jordbruksvanning mv.)	:		Grunnlagstall vil trolig utgå fra faktaark
A5. Mengde ikke bokført vann (vanntap/lekkasje)	:		Grunnlagstall vil trolig utgå fra faktaark
B1. Mengde vann til husholdningsforbruk (boligheter/leiligheter)	306 696 454	Ja	
B2. Mengde vann til husholdningsforbruk (hytter/fritidsboliger)	12 397 668	Ja	
B3. Mengde vann til industri	92 965 029	Ja	
B4. Mengde vann til tjenesteytende næringer	47 156 979	Ja	
B5. Mengde vann til primærnæringer (jordbruk/skogbruk/fiske)	18 775 632	Ja	
B6. Mengde vann til annet forbruk	28 272 467	Ja	
B7. Mengde vann til lekkasje	233 752 334	Ja	

### Faktaark I, vann – nivå 2 (nøkkeltall). Landstall.

	2011	Eksterne estimat	Merknad
Finansiell dekningsgrad	99		
Selvkostgrad	96		
Andel av befolkningen som er tilknyttet kommunal vannforsyning	84,1	Ja	
Andel av husholdningsabonentene som har installert vannmåler	:		
Gebyrinntekter per innbygger tilknyttet kommunal vannforsyning (kr/tilkn.innb)	1167	Delvis	
Gebyrgrunnlag per innbygger tilknyttet kommunal vannforsyning (kr/tilkn.innb)	1180	Delvis	
Gebyrgrunnlag per m3 (vannleveranse) (kr/m3)	6,69	Delvis	
Driftsutgifter per tilknyttet innbygger (kr/tilkn.innb)	815	Delvis	
- herav andel driftsutgifter for produksjon av vann per tilknyttet innbygger	:		
- herav andel driftsutgifter for distribusjon av vann per tilknyttet innbygger	:		
Andel driftsutgifter av gebyrgrunnlaget	69,1		
Andel kapitalkostnader av gebyrgrunnlaget	33,6		
Andel andre inntekter av gebyrgrunnlaget	4,1		
Årsgebyr for vannforsyning (gjelder rapporteringsåret+1)	2990		
Tilknytningsgebyr vann - én sats (gjelder rapporteringsåret+1)	:		
Tilknytningsgebyr vann - lav sats (gjelder rapporteringsåret+1)	:		
Tilknytningsgebyr vann - høy sats (gjelder rapporteringsåret+1)	:		
Antall kommunale vannverk	1078	Ja	
Forsyningsikkerhet, vann	:		
Gjennomsnittlig årlig avbruddstid i vannforsyningen (totalt)	:		
Gjennomsnittlig årlig avbruddstid i vannforsyningen (ikke-planlagt)	:		
Andel fornyet ledningsnett, gjennomsnitt for siste tre år	0,629	Ja	
Beregnet gjennomsnittsalder for vannledningsnett med kjent alder	31	Ja	
Andel ledningsnett med ukjent alder	10	Ja	
Tetthet av trykkøkningsstasjoner (antall/km ledning)	0,052	Ja	
Tilknytningstetthet på distribusjonsnettet (vann) (innb/km)	96	Ja	
Andel av de kommunale vannverkene som har sikkerhets- og beredskapsplan	84,9	Ja	
Andel av innb. tilkn. komm. vannverk som er forsynt av grunnvann eller desinfis	:		
Beregnet vannlekkasje per meter ledning per år (m3/m/år)	5,3	Ja	
E.coli: Andel innbyggere tilknyttet kommunalt vannverk med tilfredsstillende pr	98,5	Ja	
Intestinale enterokokker: Andel innbyggere tilknyttet kommunalt vannverk med ti	95,8	Ja	
Farge: Andel innbyggere tilknyttet kommunalt vannverk med tilfredsstillende pro	80,6	Ja	
PH: Andel innbyggere tilknyttet kommunalt vannverk med tilfredsstillende prøver	93,8	Ja	
Total vannleveranse (spesifikk totalleveranse) per tilkn. innbygger (m3/tilkn.i	177	Ja	
A1. Andel av total vannleveranse som går til husholdningsforbruk	:		Nøkkeltall vil trolig utgå fra faktaark
A2. Andel av total vannleveranse som går til industri og næringsvirksomhet	:		Nøkkeltall vil trolig utgå fra faktaark
A3. Andel av total vannleveranse som går til annet forbruk (jordbruksvanning mv	:		Nøkkeltall vil trolig utgå fra faktaark
A4. Andel ikke bokført vann (vanntap/lekkasje)	:		Nøkkeltall vil trolig utgå fra faktaark
B1. Andel av total vannleveranse til husholdningsforbruk (boliger/leiligheter)	41	Ja	
B2. Andel av total vannleveranse til husholdningsforbruk (hytter/fritidsboliger)	2	Ja	
B3. Andel av total vannleveranse til industri	13	Ja	

B4. Andel av total vannleveranse til tjenesteytende næringer	6	Ja
B5. Andel av total vannleveranse til primærnæringer (jordbruk/skogbruk/fiske)	3	Ja
B6. Andel av total vannleveranse til annet forbruk	4	Ja
B7. Andel av total vannleveranse til lekkasje	32	Ja
Gjennomsnittlig husholdningsforbruk per tilknyttet innbygger (l/pers/døgn)	209	Ja

### Faktaark I, avløp – nivå 3 (grunnlagstall). Landstall.

	2011	Eksterne estimater	Merknader
Septiktømming - brutto driftsutgifter	319 155		
Septiktømming - antall tilknyttede innbyggere	730 268	Ja	
Direkte driftsutgifter (fra skjema 23)	3 689 179		
- direkte driftsutgifter funksjon 350	:		For lite datagrunnlag for å estimere landstall
- direkte driftsutgifter funksjon 353	:		For lite datagrunnlag for å estimere landstall
Kapitalutgifter hos ekstern produsent (del av direkte driftsutgifter)	43 935		
- herav kapitalutgifter hos ekstern produsent funksjon 350	:		For lite datagrunnlag for å estimere landstall
- herav kapitalutgifter hos ekstern produsent funksjon 353	:		For lite datagrunnlag for å estimere landstall
Henførbare indirekte driftsutgifter (fra skjema 23)	296 942		
- henførbare indirekte driftsutgifter funksjon 350	:		For lite datagrunnlag for å estimere landstall
- henførbare indirekte driftsutgifter funksjon 353	:		For lite datagrunnlag for å estimere landstall
Kalkulatoriske rentekostnader (fra skjema 23)	779 073		
- kalkulatoriske rentekostnader funksjon 350	:		For lite datagrunnlag for å estimere landstall
- kalkulatoriske rentekostnader funksjon 353	:		For lite datagrunnlag for å estimere landstall
Kalkulatoriske avskrivninger (fra skjema 23)	1 158 555		
- kalkulatoriske avskrivninger funksjon 350	:		For lite datagrunnlag for å estimere landstall
- kalkulatoriske avskrivninger funksjon 353	:		For lite datagrunnlag for å estimere landstall
Andre inntekter (fra skjema 23)	240 368		
- andre inntekter funksjon 350	:		For lite datagrunnlag for å estimere landstall
- andre inntekter funksjon 353	:		For lite datagrunnlag for å estimere landstall
Gebyrgrunnlaget (driftsutgifter+kapitalkostnader-andre inntekter, fra skjema 23)	5 684 136		
Gebyrinntekter (fra skjema 23)	5 769 717		
Avsetning til selvkostfond / dekning av fremført underskudd (fra skjema 23)	:		Ikke mulig å estimere landstall, da forklaringsvariabel mangler
Bruk av selvkostfond / fremføring av underskudd (fra skjema 23)	:		Ikke mulig å estimere landstall, da forklaringsvariabel mangler
Saldo selvkostfond per 1.1 (fra skjema 23)	:		Ikke mulig å estimere landstall, da forklaringsvariabel mangler
Alternativkostnad ved bundet kapital på selvkostfond eller framføring av unders	:		Ikke mulig å estimere landstall, da forklaringsvariabel mangler
Saldo selvkostfond per 31.12 (fra skjema 23)	0		
Lønnsutgifter	1 016 161		
Kjøp av varer som inngår i komm. egenprod	1 884 344		
Kjøp av varer og tjenester som erstatter komm. egenprod	978 551		
Stipulert årsgebyr (gjelder rapporteringsåret+1)	:		Estimerte landstall basert på den benyttede gebyrtypen publisert på nivå 2.
Gebyrsats ved målt forbruk (gjelder rapporteringsåret+1) (kr/m <sup>3</sup> )	:		
Fastledd ved todelt gebyrordning (gjelder rapporteringsåret+1)	:		
Antall avløpsanlegg i kommunen (=>50 pe)	2 618	Ja	
Antall kommunalt eide avløpsanlegg (=>50 pe)	2 209	Ja	
Antall innbyggere tilknyttet kommunalt avløp (=>50 pe)	4 114 851	Ja	
Antall innbyggere tilknyttet små kommunale anlegg	77 216	Ja	
Antall innbyggere tilknyttet anlegg (=>50pe)	4 130 320	Ja	
Antall innbyggere tilknyttet anlegg med kjemisk rensing (=> 50 pe)	1 612 532	Ja	
Antall innbyggere tilknyttet anlegg med biologisk-kjemisk rensing (=> 50 pe)	1 327 872	Ja	
Ant innb tilkn anlegg med mek, bio, naturbasert eller annen rensing (=> 50 pe)	1 013 221	Ja	
Antall innbyggere tilknyttet ledningsnett med urensset utslipp (=> 50 pe)	176 695	Ja	
Total belastning på avløpsanleggene (kg tot-P)	2 604 555	Ja	
Belastning på anlegg med kjemisk rensing (kg tot-P)	980 769	Ja	
Belastning på anlegg med biologisk-kjemisk rensing (kg tot-P)	885 217	Ja	
Belastning på anl med mek, bio, natbas og annen rensing (kg tot-P)	636 210	Ja	
Lengde ledningsnett totalt	35 714 196	Ja	
Lengde ledningsnett med ukjent alder	7 717 391	Ja	
Lengde ledningsnett etter 2000 (meter)	5 951 639	Ja	
Lengde ledningsnett 1980-99 (meter)	9 889 303	Ja	
Lengde ledningsnett 1960-79 (meter)	9 087 624	Ja	
Lengde ledningsnett 1940-59 (meter)	2 078 419	Ja	

Lengde ledningsnett for 1940 (meter)	989 822	Ja
Lengde nylagt ledningsnett	406 782	Ja
Lengde fornyet ledningsnett	155 518	Ja
Antall meter fornyet ledningsnett, gjennomsnitt siste tre år	165 312	Ja
Antall meter separat overvannsnett	15 318 509	Ja
Antall regnvannsoverløp i fellessystemet	3 134	Ja
Antall pumpestasjoner	9 044	Ja
Antall kloakkstopper i avløpsledninger og kummer	2 400	Ja
Antall avsluttede saker over kjelleroversvømmelser, der kommunen har erkjent er	401	Ja
Mengde slam disponert totalt (TTS)	:	
Mengde slam disponert til jordforbedring i jordbruket (TTS)	:	
Mengde slam disponert til jordforbedring på grøntarealer (TTS)	:	
Mengde slam levert til jordprodusent (TTS)	:	
Mengde slam disponert til forbrenning/energigjenvinning (TTS)	:	
Mengde slam disponert til toppdekke på avfallsfylling (TTS)	:	
Mengde slam deponert (TTS)	:	
Mengde slam disponert til eksport (TTS)	:	
Mengde slam disponert til annet formål (TTS)	:	
Mengde slam disponert til ukjent formål (TTS)	:	

*Faktaark I, avløp – nivå 2 (nøkkeltall). Landstall.*

	2011	Eksterne estimat	Merknad
Finansiell dekningsgrad	102		
Selvkostgrad	99		
Andel av befolkningen som er tilknyttet kommunal avløpstjeneste	84,1	Ja	
Andel av husholdningsabonentene på avløp som har installert vannmåler	:		
Gebyrinntekter per innbygger tilknyttet kommunal avløpstjeneste (kr/tilkn.innb)	1 376	Delvis	
Gebyrgrunnlag per innbygger tilknyttet kommunal avløpstjeneste (kr/tilkn.innb)	1 356	Delvis	
Gebyrgrunnlag per belastningsenhet (kr/kg tot-P)	2 182	Delvis	
Driftsutgifter per innb. tilknyttet kommunal avløpstjeneste	940	Delvis	
- herav andel driftsutgifter for avløpsrensing per tilknyttet innbygger	:		
- herav andel driftsutgifter for avløpsnett/innsamling per tilknyttet innbygge	:		
Andel driftsutgifter av gebyrgrunnlaget	69,4		
Andel kapitalkostnader av gebyrgrunnlaget	34,1		
Andel andre inntekter av gebyrgrunnlaget	4,2		
Årsgebyr for avløpstjenesten (gjelder rapporteringsåret+1)	3 343		
Tilknytningsgebyr avløp - én sats (gjelder rapporteringsåret+1)	:		
Tilknytningsgebyr avløp - lav sats (gjelder rapporteringsåret +1)	:		
Tilknytningsgebyr avløp - høy sats (gjelder rapporteringsåret +1)	:		
Årsgebyr for septiktømming (gjelder rapporteringsåret+1)	:		
Antall kommunale avløpsanlegg	2 209	Ja	
Beregnet gjennomsnittsalder for spillvannsnett med kjent alder	30	Ja	
Andel ledningsnett med ukjent alder	22	Ja	
Andel fornyet ledningsnett, gjennomsnitt for siste tre år	0,46	Ja	
Tilknytningstetthet på distribusjonsnettet (avløp) (innb/km)	117	Ja	
Tetthet av pumpestasjoner (antall/km ledningsnett)	0,25	Ja	
Antall kloakkstopper i avløpsledninger og kummer per kilometer ledningsnett	0,067	Ja	
Antall avsluttede saker over kjelleroversvømmelser, der kommunen har erkjent er	401	Ja	
Andel belastning på anlegg med kjemisk rensing	37,7	Ja	
Andel belastning på anlegg med biologisk-kjemisk rensing	34	Ja	
Andel belastning på anlegg med mekanisk, biologisk, naturbasert og annet rensep	24,4	Ja	
Andel innbyggere tilknyttet anlegg med kjemisk rensing	39	Ja	
Andel innbyggere tilknyttet anlegg med biologisk-kjemisk rensing	32,1	Ja	
Andel innbyggere tilknyttet anlegg med mekanisk, biologisk, naturbasert el. ann	24,5	Ja	
Slamtørstoff produsert per kubikkmeter avløpsvann	:		
Andel slam disponert til jordforbedring i jordbruket	:		
Andel slam disponert til jordforbedring på grøntarealer	:		
Andel slam levert til jordprodusent	:		
Andel slam disponert til forbrenning/energigjenvinning	:		
Andel slam disponert til toppdekke på avfallsfylling	:		
Andel slam deponert	:		
Andel slam disponert til eksport	:		
Andel slam disponert til annet formål	:		
Andel slam disponert til ukjent formål	:		



**B** Returadresse:  
Statistisk sentralbyrå  
NO-2225 Kongsvinger

Avsender:  
**Statistisk sentralbyrå**

Postadresse:  
Postboks 8131 Dep  
NO-0033 Oslo

Besøksadresse:  
Kongens gate 6, Oslo  
Oterveien 23, Kongsvinger

E-post: [ssb@ssb.no](mailto:ssb@ssb.no)  
Internett: [www.ssb.no](http://www.ssb.no)  
Telefon: 62 88 50 00

ISBN 978-82-537-8525-7 (trykt)  
ISBN 978-82-537-8526-4 (elektronisk)  
ISSN 1891-5906

ISBN 978-82-537-8525-7



9 788253 785257



**Statistisk sentralbyrå**  
Statistics Norway