

Methods Library

as part of the modernization of the
statistical production in Norway



Statistisk sentralbyrå
Statistics Norway

UNECE - Work Session on Statistical Data Editing
24-26 April 2017
Susie Jentoft on behalf of Aslaug, Øyvind & Ane



Fellesløsning

ISEE

Kompis



Statistikkbanken



Metode



Kommune-fakta/profil (ssb.no)

Methods Library



IT: Administrative system

Methods: Standardised R-functions in Git

R-package: Kostra

RateModel

HB

TopDown

...

...

...

R-package 3

...

...

...

R-package: SSBtools

...

...

...



R functions and packages

The screenshot shows the RStudio interface with the following components:

- Source Editor:** Contains R code for the `ImputeRegression` function. The code includes comments and parameter definitions. A callout box highlights the first 15 lines of code.
- Environment:** Shows the current environment with variables like `data`, `idName`, `strataName`, `xName`, `yName`, `method`, and `newNames`. A callout box highlights the text "Git (internal) is used for sharing and collaboration".
- Files:** Shows a file explorer view of the `kostra` directory. A callout box highlights the text "Roxygen2 is used for documentation in R".
- Console:** Shows the R version (3.3.3) and copyright information.

```
1 #' Imputation of a single variable (y) by a regression model using a single explanatory variable
2 #'
3 #' Impute missing and wrong values (group 3) by the model based on representative data (group 2)
4 #' Some data are considered correct but not representative (group 2).
5 #'
6 #' @encoding UTF8
7 #'
8 #' @param data Input data set of class data.frame
9 #'
10 #' @param idName Name of id-variable(s)
11 #' @param strataName Name of strata-variable. Single strata when NULL (default)
12 #' @param xName Name of x-variable
13 #' @param yName Name of y-variable
14 #' @param method The method (model and weight) coded as a string: "ordinary" (default), "ratioconstant"
15 #' "mean" or "ratioconstant".

20 #' Can also be used to specify additional variable names that will be included in output
21 #' @param fun Function as input to ImputeRegressionNewNames for more general applications.
22 #' @param oldNames Vector of output names to be changed.
23 #' @param newNames Corresponding vector of new names.
24 #'
```



Git

The screenshot shows a web browser window displaying the Bitbucket interface. The address bar shows the URL: `https://git-adm.ssb.no/projects/RMET/repos/kostra/browse/R/ImputeRegression.R#app-switcher`. The page title is "Source" and the repository path is "R-metode / KOSTRA". The file path is "KOSTRA / R / ImputeRegression.R". The commit information shows "Langsrud committed 239e48d5d48 2 days ago".

The left sidebar contains navigation options: "ACTIONS" (Clone, Create branch, Create pull request, Fork) and "NAVIGATION" (Source, Commits, Branches, Network, Pull requests, Settings). The "Source" view is selected, showing the content of the file "ImputeRegression.R".

```
1 #' Imputation of a single variable (y) by a regression model using a single explanatory variable (x).
2 #'
3 #' Impute missing and wrong values (group 3) by the model based on representative data (group 1).
4 #' Some data are considered correct but not representative (group 2).
5 #'
6 #' @encoding UTF8
7 #'
8 #' @param data Input data set of class data.frame
9 #'
10 #' @param idName Name of id-variable(s)
11 #' @param strataName Name of strata-variable. Single strata when NULL (default)
12 #' @param xName Name of x-variable
13 #' @param yName Name of y-variable
14 #' @param method The method (model and weight) coded as a string: "ordinary" (default), "ratio", "noconstant",
15 #' "mean" or "ratioconstant".
16 #' @param limitModel Studentized residuals limit. Above limit -> group 2.
17 #' @param limitIterate Studentized residuals limit for iterative calculation of studentized residuals.
18 #' @param limitImpute Studentized residuals limit. Above limit -> group 3.
19 #' @param ... Simplified specification of the above arguments and possibly the five arguments below.
20 #' Can also be used to specify additional variable names that will be included in output (micro).
21 #' @param Fun Function as input to ImputeRegressionNewNames for more general applications.
22 #' @param oldNames Vector of output names to be changed.
23 #' @param newNames Corresponding vector of new names.
24 #' @param id When non-NULL a new variable ID will be created (see details).
25 #' @param keep When non-NULL Only variables listed in keep will be kept.
26 #' This is input to ImputeRegressionNewNames and for more general applications keep apply to the three
27 #' first list elements.
28 #' @param addName NULL or vector of strings used to name columns according to origin frame.
29 #' @param characterString Character string to separate when addName apply
```



Administrative system

The screenshot shows the KOMPIS administrative system interface. The main content area is titled "METODEBIBLIOTEKET" and "Metoder i biblioteket". A table lists several methods, including "Kvartilmetode", "Historisk imputering", "Tusenfeil", "Regresjonsimputering", and "HB-metoden". A callout box highlights the text: "Each method is a separate building block and follows CSPA".

Each method is a separate building block and follows CSPA

Metode				Sist endr. ▾	Aktiv
Kvartilmetode	variabler er ubnormalt. Grenseverdien er en faktor ganget forholdet mellom median og kvartil. Metoden gir ekstra informasjon til hjelp for å vurdere mulige feil eller ikke.			+og+HB- 06.03.2017 Ola Nordmann	true
Historisk imputering	Imputerer ved bruk av forrige års verdi. Hvis forrige år ikke finnes settes verdien til null. Vi kan se på dette som en modell og dermed estimere usikkerheten slik som for andre imputeringsmodeller.	https://git-adm.ssb.no/projects/RMET	https://wiki.ssb.no/pages/Viewpage.action?pageId=57248232	27.02.2017 Donald Duck	true
Tusenfeil	Oppdager enheter med mulig tusenfeil eller andre store endringer. Dette gjøres ved å sammenligne verdien på samme variabel i to ulike perioder på logaritmisk skala (10-logaritmen).	https://git-adm.ssb.no/projects/RMET	https://wiki.ssb.no/pages/Viewpage.action?pageId=61145556	22.02.2017 Ola Nordmann	true
Regresjonsimputering	Setter inn manglende verdier (imputering) i en interessevariabel (y) ved bruk av en modell som benytter en forklaringsvariabel (x). Flere modeller kan velges. Standardvalg er ratemodell - regresjonslinja går gjennom origo (0,0) og økende varians.	https://git-adm.ssb.no/projects/RMET	https://wiki.ssb.no/pages/Viewpage.action?pageId=57248232	22.02.2017 Ola Nordmann	true
HB-metoden	Finner ekstremverdier basert på	https://git-adm.ssb.no/projects/RMET	https://wiki.ssb.no/pages/Viewpage.action?spacekey=KOSTRA&title=Makroeditering+%2Fkvartilmetoden+og+HB-metoden	13.02.2017 Ola Nordmann	true



Administrative system

Parametre til metoden

#	Navn	Beskrivelse	Type	Detaljer	Endre Slett	Flytt
1	Identifiseringsvariabel	Variabel som identifiserer enhetene	Variabel	Rolle: Identifiseringsvariabel, Datatype: Tekst, Påkrevd: Ja, Flere verdier tillatt: Nei	Endre Slett	⬇
2	Periode	Variabel med periodeverdier	Variabel	Rolle: Identifiseringsvariabel, Datatype: Tekst, Påkrevd: Ja, Flere verdier tillatt: Nei	Endre Slett	⬆ ⬇
3	Interessevariabel (y)				Endre Slett	⬆ ⬇
4	Forklaringsvariabel (x)				Endre Slett	⬆ ⬇
5	Gjeldende år	Gjeldende år	Liste	Antall elementer: 5, Default: [], Påkrevd: Ja, Tillat flervalg: Nei	Endre Slett	⬆ ⬇
6	Grenseverdi modell	Grenseverdi for studentiserte residualer for den	Flyttall	Min: 1, Maks: <ingen verdi satt>, Default: 2.5, Påkrevd: Ja, Flere verdier tillatt: Nei	Endre Slett	⬆ ⬇
7	Grenseverdi iterasjoner	Grenseverdi på studentiserte residualer ved iterasjon.	Flyttall	Min: 1, Maks: <ingen verdi satt>, Default: 4.5, Påkrevd: Ja, Flere verdier tillatt: Nei	Endre Slett	⬆ ⬇
8	Grenseverdi imputering	Grenseverdi på studentiserte residualer for å anse	Flyttall	Min: 1, Maks: <ingen verdi satt>, Default: 50, Påkrevd: Ja, Flere verdier tillatt: Nei	Endre Slett	⬆ ⬇
9	Strata	Navn på stratavariabel	Variabel	Rolle: Identifiseringsvariabel, Datatype: Tekst, Påkrevd: Nei, Flere verdier tillatt: Nei	Endre Slett	⬆ ⬇
10	Regresjonsmodell	Hvilken variant av imputering ønsker man å kjøre. Ratio vil bli kjørt	Liste	Antall elementer: 5, Default: [Rate], Påkrevd: Nei, Tillat flervalg: Nei	Endre Slett	⬆

Input and output methods documented in accordance with GSIM

[+ Legg til parameter](#)



Administrative system

Utdata fra metoden

#	Navn	Beskrivelse	Rolle	Type	Endre Slett	Flytt
1	identifiseringsvariabel	Variabel som identifiserer enhetene	Identifiseringsvariabel	Tekst	Endre Slett	⬇
2	Interessevariabel (x1)	Interessevariabel x	Målevariabel	Flyttall	Endre Slett	⬆ ⬇
3	Interessevariabel (x2)	Den variabelen vi ønsker å kontrollere og imputere på.	Målevariabel	Flyttall	Endre Slett	⬆ ⬇
4	Stratifiseringsvariabel	Stratifiseringsvariabel	Identifiseringsvariabel	Tekst	Endre Slett	⬆ ⬇
5	Kategori	Data blir delt inn i 3 grupper; 1 representativ, 2 korrekt, men ikke representativ, 3 uteligger.	Egenskapsvariabel	Heltall	Endre Slett	⬆ ⬇
6	Estimert verdi	Verdien som blir beregnet ut fra den statistiske modellen.	Målevariabel	Flyttall	Endre Slett	⬆ ⬇
7	Imputert verdi	Imputert verdi for interessevariabelen.	Målevariabel	Flyttall	Endre Slett	⬆ ⬇
8	Standardiserte residualer	Avstanden mellom den observerte og predikerte verdi dividert med estimert usikkerhet. Verdiene er fra siste iterasjon.	Egenskapsvariabel	Flyttall	Endre Slett	⬆

[+ Legg til utdata variabel](#)

Grafkonfigurasjoner til metoden

Graftype	Endre Slett
Scatterplot	Endre Slett

[+ Legg til grafkonfigurasjon](#)



Administrative system

The screenshot shows a web browser window with the URL <https://kompis-qa.ssb.no/dashboardui#!Metodek>. The browser tab is titled "Felles GUI". The page header includes "KOSTRA" and "KOMPIS".

Left sidebar:

- Profile: Susie Jentoft
- MethodLibraryAdmin
- Metodekjøring** (selected)
- Produksjoner
- Mine kataloger
 - Arbeidskatalog
- + Opprett ny katalog

Main content area:

METODEKJØRING

Kjøring av metoder fra metodebiblioteket

1. Velg metode og sett innstillinger

Metode: Regresjonsimpute ?

Grenseverdi modell: 2,5 ? Grenseverdi iterasjoner: 4,5 ? Grenseverdi imputering: 50 ? Gjeldende år: 2015 ?

Dataset Inndata: KOSTRA-grupper

Identifiseringsvariabel: REGION ? Periode: PERIODE ? Strata: ?

2. Velg målvariabler du ønsker å kjøre metoden på

Forklaringsvariabel (x)

Navn

[Martes arbeidskatalog](#)

Eiendomsforvaltning for utvalgte kommunale formålsbygg ▾ ☆

Rediger
Kopier
Slett
Endringslogg
Versjoner

[BASIS](#)
VTL
ESTIMERE
EDITERE
PUBLISERE
FORHÅNDSVISE

ID: 84759387

Navn: Eiendomsforvaltning for utvalgte kommunale formålsbygg

VTL navn: Eiendomsforvaltning for utvalgte kommunale formålsbygg

Katalog: [M](#)

Sensitiv tabell:

Opprettet: 06.09.2016
av: Marte Hvamb

Variabler:

VTL-navn ▾	Rolle
Filter	Filter
periode	Identifiseringsvariabel
region	Identifiseringsvariabel
netto driftsutgifter til kommunal eiendomsforvaltning per innbygger	Målevariabel
netto_driftsutgifter_til_kommunal_forvaltning_av_eiendommer_(f121)_per_innbygger_	Målevariabel
netto_driftsutgifter_til_kommunal_eiendomsforvaltning_i_prosent_av_samlede_netto_driftsutgifter	Målevariabel

March 2018

<https://vtl-demo.ssb.no/>

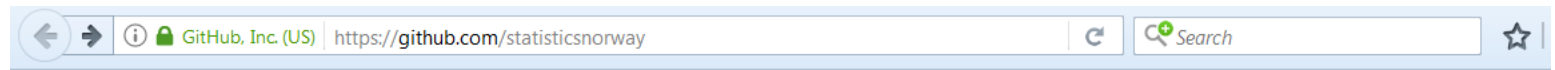
```

      /\_/\      /\_/\      /\_/\
     /::\_/      /::\_/      /::\_/
    |:::L/\_/\   |:::~\_/     |::\_/
    |:::~\_/     |::\_/~\_/   |::\_/
    L;;;/_/\     \_\_/       \_\_/
                                     \_\_/

Java VTL interpreter version 1.0-SNAPSHOT.
Type .help for command list.

vtl> 
```

https://github.com/statisticsnorway



Features Explore Pricing

This organization Search

Sign in



Statistics Norway

Oslo, Norway <http://www.ssb.no/en/>

Repositories

People 0

Search repositories...

Type: All

java-vtl

Forked from hadrienk/java-vtl

Java implementation of the Validation Transformation Language

java dsl scriptengine vtl statistics

Java ★ 4 🍴 1 Updated 22 hours ago



Top languages

Java

People

This organization has no public members.
You must be a member to see the members of this organization.



Statistisk sentralbyrå
Statistics Norway

Thank you



susie.jentoft@ssb.no

