PROGRAM *DOMEST* FOR ESTIMATION FOR DOMAINS AND SMALL AREAS

PART 2: Worked examples

Dr. Ari Veijanen
Statistics Finland

Prof. Risto Lehtonen
University of Helsinki

BNU Workshop

August 2010

(Last update 19 August 2010)

This document supplements the document
PROGRAM *DOMEST* FOR ESTIMATION FOR DOMAINS AND SMALL AREAS
PART 1: Technical documentation
Program Domest is developed at Statistics Finland by Dr. Ari Veijanen together with Prof. Risto Lehtonen (University of Helsinki).
Domest is freely available from the authors.
Contact: ari.veijanen@pp1.inet.fi

Domest – Worked examples

Preliminaries

Domest is organized into 7 functional pages:

1. Page 'Introduction'
2. Page 'Session'
3. Page 'Data'
4. Page 'Transformations'
5. Page 'Fixed Effects Model'
6. Page 'Linear Mixed Model'
7. Page 'Domain Estimation'

Pages 5, 6 and 7 represent the most important phases in Domest estimation practice. Page 3 is usually visited in an early phase when importing the population and sample data sets and for the specification of sampling design and domain structure. Also at an early stage, Page 4 can be used for variable transformations, e.g. to create the domain indicators. Page 2 is usually visited when restoring Domest session with materials from previous session. In addition to the main pages, Domest includes Help.pdf file and facilities to save and print the results of the analysis. The underlying theory is presented in Lehtonen and Veijanen (2009) and Saei and Chambers (2004) and is summarized in Part 1 of Domest documentation.

The examples are worked out with test data sets prepared for pedagogical purposes. A population data set and three sample data sets have been created. The original SAS data sets have been imported into Domest on page 'Data'. Domest transforms the SAS data sets into a format manageable by Domest.

The population data set is at unit level with $N = 966$ records. There are $D = 10$ domains in the population. The three sample data sets (drawn with SAS/SURVEYSELECT) are:

1. The data set STR_SRSWOR_sample is a stratified SRSWOR sample of $n = 100$ elements drawn with equal allocation ($n_d = 10, \ \ d = 1,...,10$) from the population, where the variable domain is the stratum variable. This sample represents the *planned domains* case where the strata are the domains of interest.
2. The data set STR_PiPS_sample is a stratified fixed-size πPS sample of $n = 100$ elements drawn with equal allocation ($n_d = 10, \ \ d = 1,...,10$) from the population, where the variable domain is the stratum variable. This sample also represents the *planned domains* case where the strata are the domains of interest.
3. The data set SRSWOR_sample is an $n = 100$ element SRSWOR sample from the population, representing the *unplanned domains* structure in domain estimation.

The continuous response variable is *y*, continuous *x* represents an auxiliary variable and *z* is the size variable in the πPS design. We are interested in the estimation of domain totals of *y* and accuracy measures standard error or root MSE. The variable domain identifies the domains of interest.

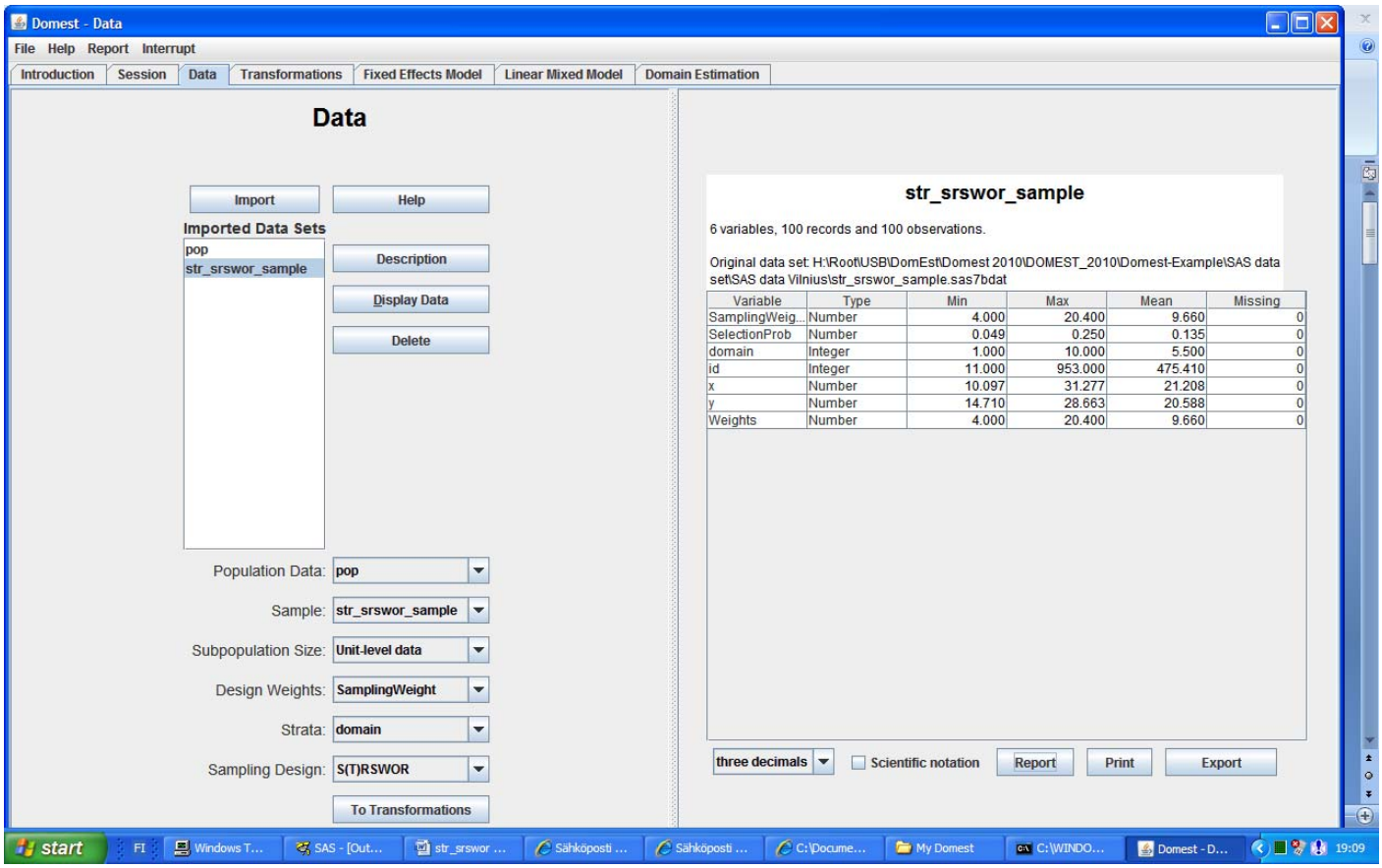
The following estimator types are discussed (see Part 1 for details):

1. Direct design-based estimators
    a. Horvitz-Thompson (HT) estimator with ordinary and Hájek variance estimators
2. Indirect design-based estimators
    a. GREG (Generalized Regression) estimator with linear fixed-effects assisting model
    b. MGREG, mixed-model assisted GREG estimator
3. Indirect model-based estimators
    a. EBLUP (Empirical Best Linear Unbiased Predictor) type estimator.

1. ESTIMATION FOR PLANNED DOMAINS (strata)

1.1. STR_SRSWOR_sample

Stratified SRSWOR sampling with domains as strata and equal allocation, $n_d = 10,\ \ d = 1,...,10$



STR_SRSWOR_sample

The CORR Procedure (Unweighted results)

Variable	N	Mean	Std Dev	Sum	Minimum	Maximum
y	100	20.58784	2.95298	2059	14.70985	28.66262
x	100	21.20821	4.84223	2121	10.09723	31.27745

Pearson Correlation Coefficients, N = 100
Prob > |r| under H0: Rho=0

	y	x
y	1.00000	0.83556 <.0001
x	0.83556 <.0001	1.00000

1. Estimation for planned domains (strata)

1.1. STR_SRSWOR_sample

1.1.1. HT estimator SAS/SURVEYMEANS

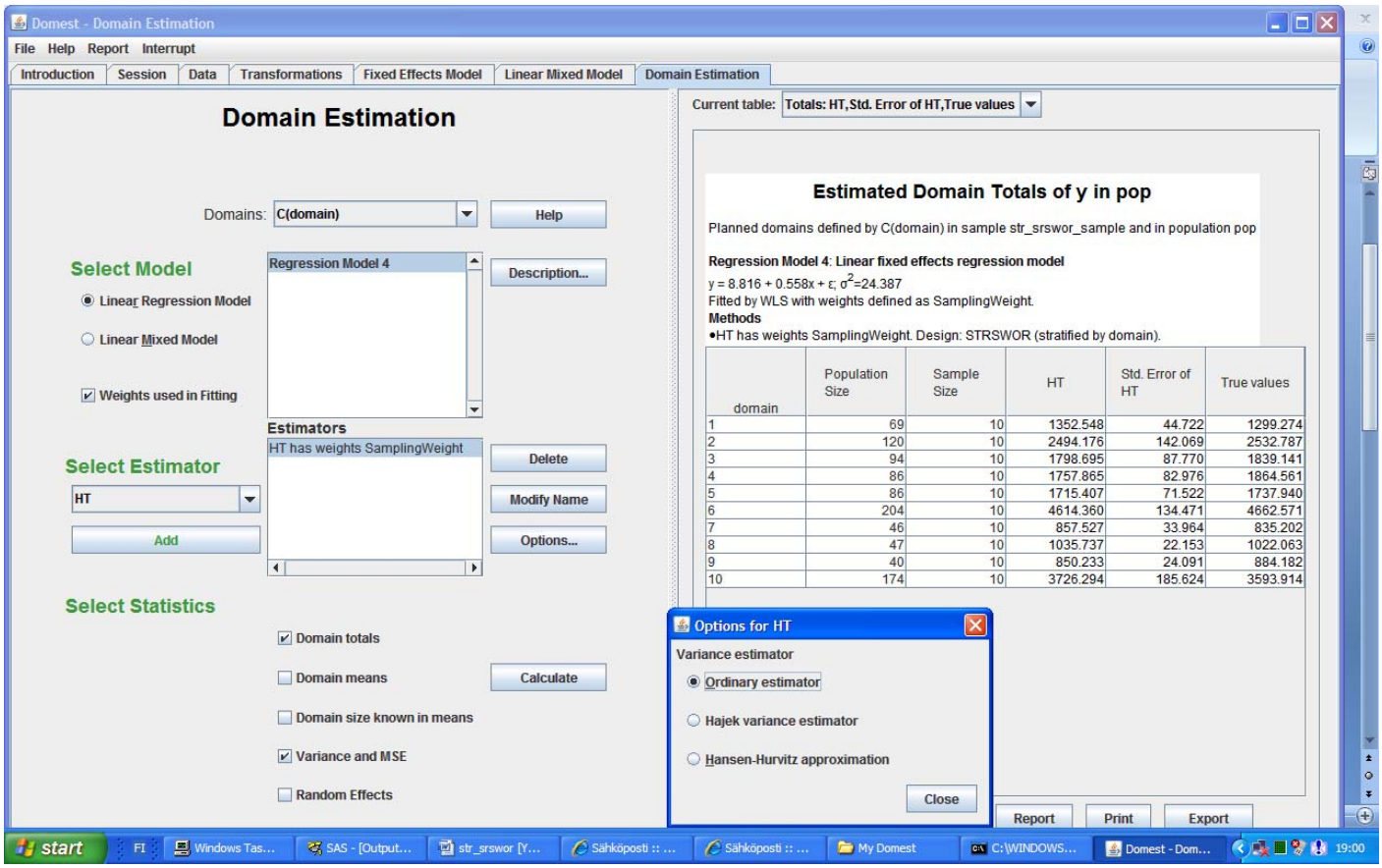
```
proc summary data=pop nway;
class domain;
output out=str(drop=_type_
    rename=(_freq_=_total_));
run;

proc surveymeans
data=STR_SRSWOR_sample sum
    total=str;
TITLE "Planned domains /
STR_SRSWOR_sample";
by domain;
var y;
weight samplingweight;
run;
```

```
Planned domains / STR_SRSWOR_sample

domain=1
Number of Observations            10
Sum of Weights                    69
Variable             Sum         Std Dev
_____________________________________________
y           1352.547628       44.721506
_____________________________________________
domain=2
Number of Observations            10
Sum of Weights                   120
_____________________________________________
y           2494.175809      142.068588
_____________________________________________
domain=3
Number of Observations            10
Sum of Weights                    94
_____________________________________________
y           1798.695374       87.770331
_____________________________________________
domain=4
Number of Observations            10
Sum of Weights                    86
_____________________________________________
y           1757.864797       82.976090
_____________________________________________
domain=5
Number of Observations            10
Sum of Weights                    86
_____________________________________________
y           1715.406742       71.522025
_____________________________________________
domain=6
Number of Observations            10
Sum of Weights                   204
_____________________________________________
y           4614.359667      134.470845
_____________________________________________
domain=7
Number of Observations            10
Sum of Weights                    46
_____________________________________________
y            857.527496       33.964165
_____________________________________________
domain=8
Number of Observations            10
Sum of Weights                    47
_____________________________________________
y           1035.737192       22.153194
_____________________________________________
domain=9
Number of Observations            10
Sum of Weights                    40
_____________________________________________
y            850.232647       24.091342
_____________________________________________
domain=10
Number of Observations            10
Sum of Weights                   174
_____________________________________________
y           3726.293657      185.624344
```

1. Estimation for planned domains (strata)

1.1. STR_SRSWOR_sample

1.1.2. HT estimator OPTION: Ordinary variance estimator (Formulas (1) and (2) in Part 1)



Estimated Domain Totals of y in pop

Planned domains defined by C(domain) in sample str_srswor_sample and in population pop

Regression Model 4: **Linear fixed effects regression model**
$y = 8.816 + 0.558x + \varepsilon$; $\sigma^2=24.387$
Fitted by WLS with weights defined as SamplingWeight.
Methods
•HT has weights SamplingWeight. Design: STRSWOR (stratified by domain).

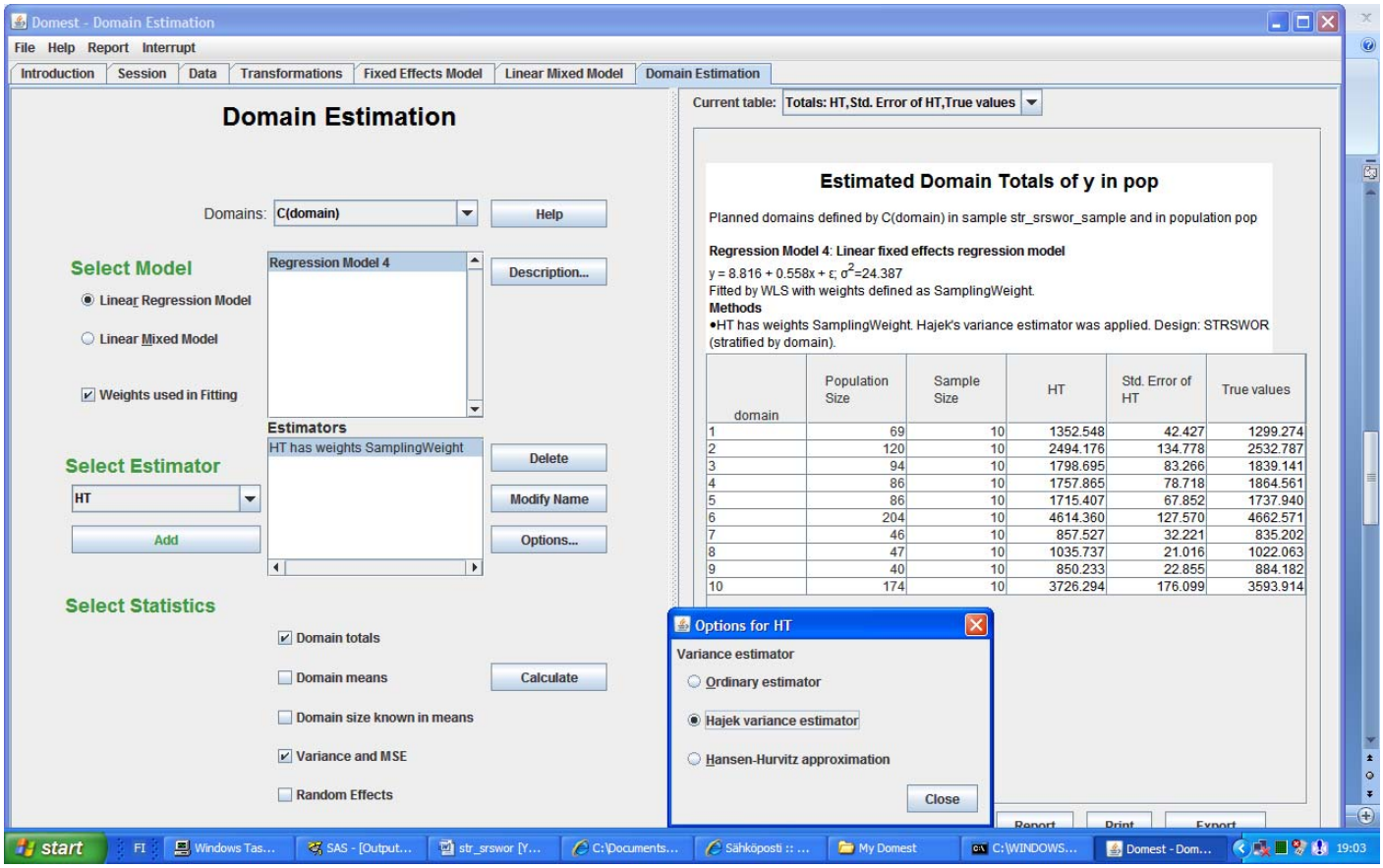
domain	Population Size	Sample Size	HT	Std. Error of HT	True values
1	69	10	1352.548	44.722	1299.274
2	120	10	2494.176	142.069	2532.787
3	94	10	1798.695	87.770	1839.141
4	86	10	1757.865	82.976	1864.561
5	86	10	1715.407	71.522	1737.940
6	204	10	4614.360	134.471	4662.571
7	46	10	857.527	33.964	835.202
8	47	10	1035.737	22.153	1022.063
9	40	10	850.233	24.091	884.182
10	174	10	3726.294	185.624	3593.914

1. Estimation for planned domains (strata)

1.1. STR_SRSWOR_sample

1.1.3. HT estimator OPTION: Hájek variance estimator (Formulas (1) and (3) in Part 1)

NOTE: Information on domain sizes in population N_d are incorporated in the estimation procedure



Estimated Domain Totals of y in pop

Planned domains defined by C(domain) in sample str_srswor_sample and in population pop

Regression Model 4: Linear fixed effects regression model

$y = 8.816 + 0.558x + \varepsilon$; $\sigma^2=24.387$

Fitted by WLS with weights defined as SamplingWeight.

Methods

•HT has weights SamplingWeight. Hájek's variance estimator was applied. Design: STRSWOR (stratified by domain).

domain	Population Size	Sample Size	HT	Std. Error of HT	True values
1	69	10	1352.548	42.427	1299.274
2	120	10	2494.176	134.778	2532.787
3	94	10	1798.695	83.266	1839.141
4	86	10	1757.865	78.718	1864.561
5	86	10	1715.407	67.852	1737.940
6	204	10	4614.360	127.570	4662.571
7	46	10	857.527	32.221	835.202
8	47	10	1035.737	21.016	1022.063
9	40	10	850.233	22.855	884.182
10	174	10	3726.294	176.099	3593.914

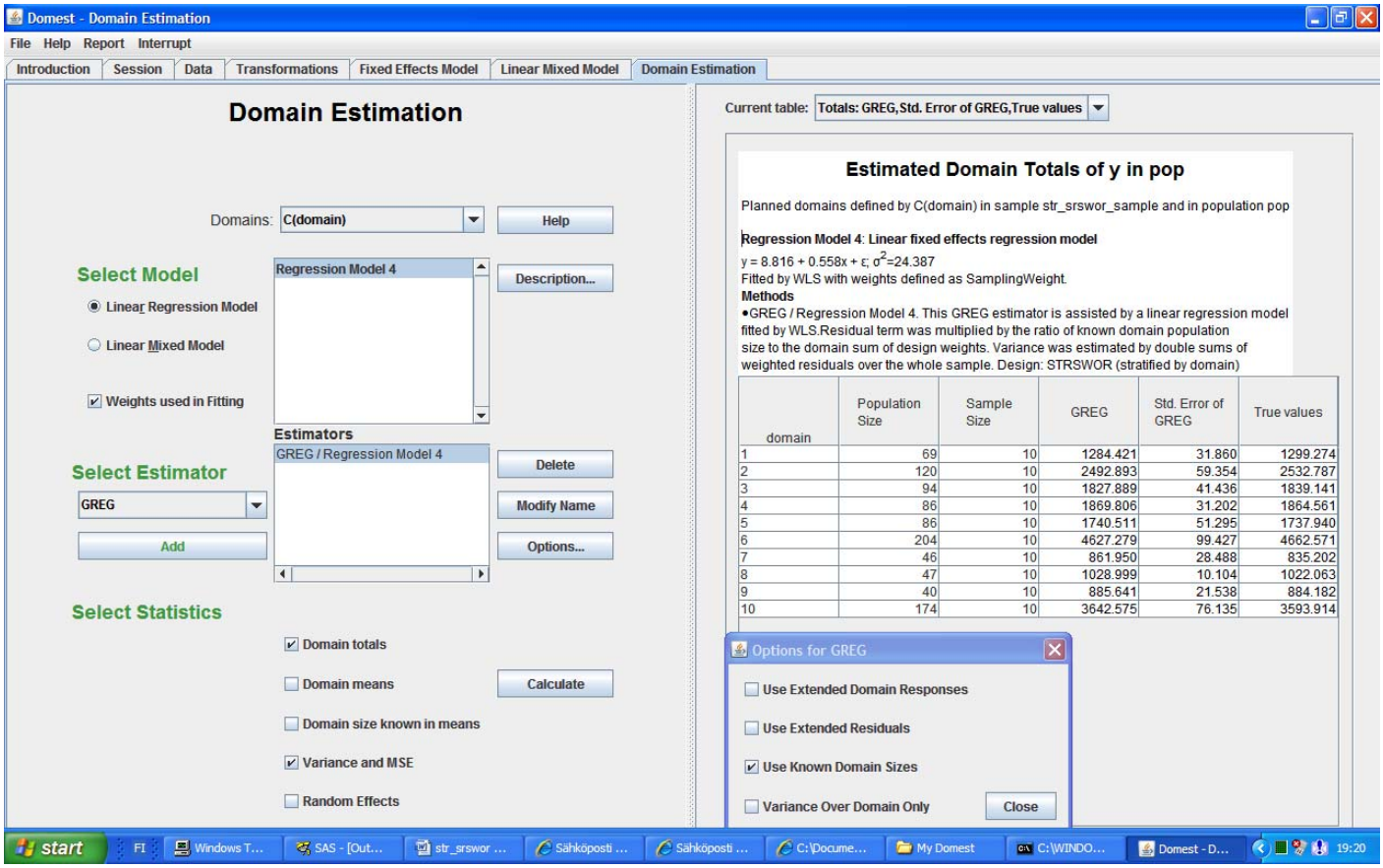
1. Estimation for planned domains (strata)

1.1. STR_SRSWOR_sample

1.1.4. GREG (Generalized Regression) estimator OPTION: Use Known Domain Sizes (Formula (11))

Assisting model: Linear fixed effects model $Y_k = \beta_0 + \beta_1 x_k + \varepsilon_k$

NOTE: Information on domain sizes in population N_d also are incorporated in the estimation procedure



Estimated Domain Totals of y in pop

Planned domains defined by C(domain) in sample str_srswor_sample and in population pop

Regression Model 4: Linear fixed effects regression model

$y = 8.816 + 0.558x + \varepsilon$; $\sigma^2 = 24.387$

Fitted by WLS with weights defined as SamplingWeight.

Methods

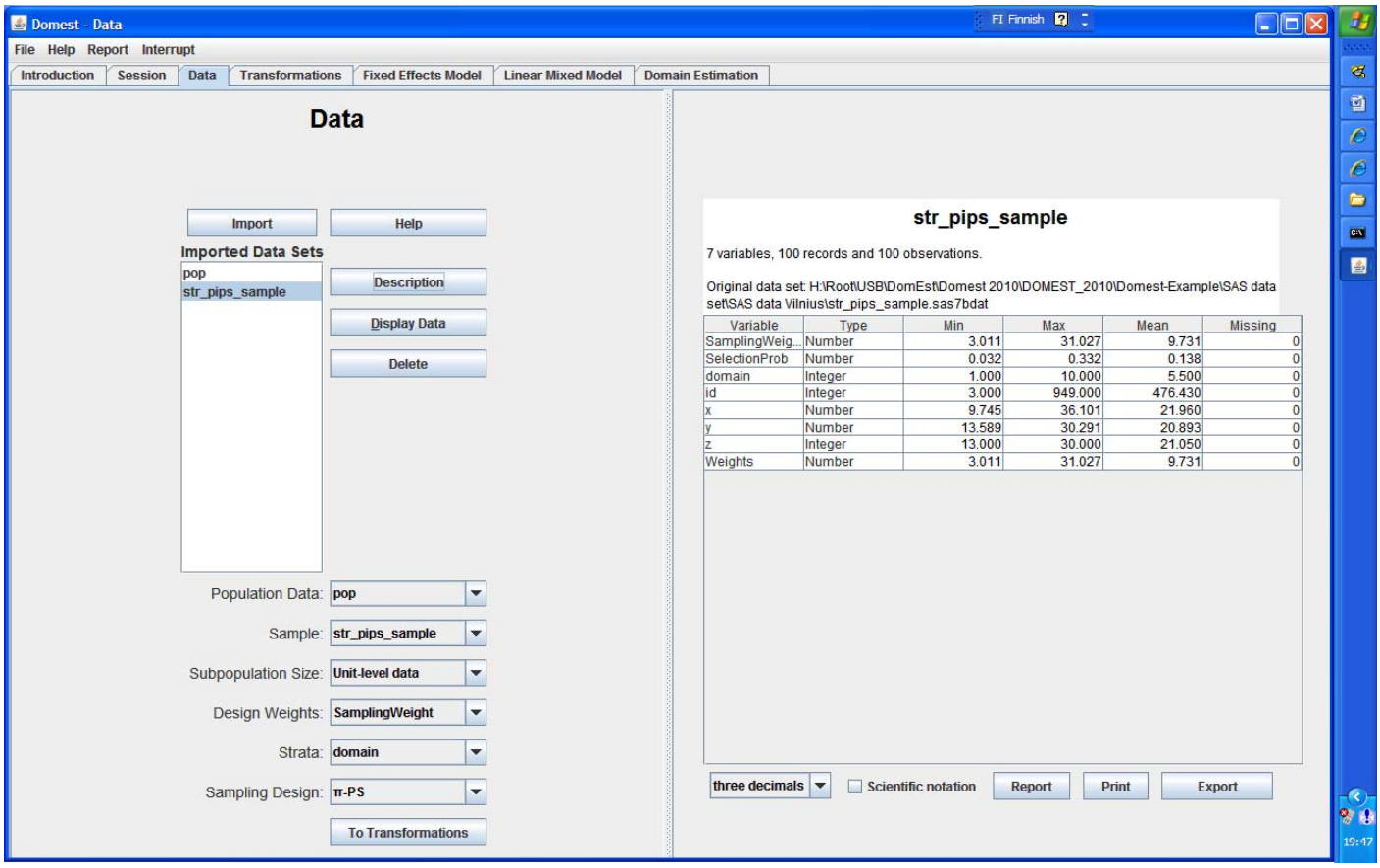
•GREG / Regression Model 4. This GREG estimator is assisted by a linear regression model fitted by WLS. Residual term was multiplied by the ratio of known domain population size to the domain sum of design weights. Variance was estimated by double sums of weighted residuals over the whole sample. Design: STRSWOR (stratified by domain)

domain	Population Size	Sample Size	GREG	Std. Error of GREG	True values
1	69	10	1284.421	31.860	1299.274
2	120	10	2492.893	59.354	2532.787
3	94	10	1827.889	41.436	1839.141
4	86	10	1869.806	31.202	1864.561
5	86	10	1740.511	51.295	1737.940
6	204	10	4627.279	99.427	4662.571
7	46	10	861.950	28.488	835.202
8	47	10	1028.999	10.104	1022.063
9	40	10	885.641	21.538	884.182
10	174	10	3642.575	76.135	3593.914

1. Estimation for planned domains (strata)

1.2. STR_PiPS_sample

Stratified πPS with domains as strata and equal allocation, $n_d = 10, \quad d = 1,...,10$



Variable	Type	Min	Max	Mean	Missing
SamplingWeig...	Number	3.011	31.027	9.731	0
SelectionProb	Number	0.032	0.332	0.138	0
domain	Integer	1.000	10.000	5.500	0
id	Integer	3.000	949.000	476.430	0
x	Number	9.745	36.101	21.960	0
y	Number	13.589	30.291	20.893	0
z	Integer	13.000	30.000	21.050	0
Weights	Number	3.011	31.027	9.731	0

STR_PiPS_sample

The CORR Procedure (Unweighted results)

Variable	N	Mean	Std Dev	Sum	Minimum	Maximum
y	100	20.89329	3.20676	2089	13.58878	30.29122
x	100	21.95957	5.19386	2196	9.74501	36.10130
z	100	21.05000	3.30404	2105	13.00000	30.00000

Pearson Correlation Coefficients, N = 100
Prob > |r| under H0: Rho=0

	y	x	z
y	1.00000	0.87600 <.0001	0.95467 <.0001
x	0.87600 <.0001	1.00000	0.81410 <.0001
z	0.95467 <.0001	0.81410 <.0001	1.00000

1. Estimation for planned domains (strata)

1.2. STR_PiPS_sample

1.2.1. HT estimator SAS/SURVEYMEANS (by DOMAIN)

domain=1

Number of Observations	10
Sum of Weights	64.5091041

Variable	Sum	Std Dev
y	1284.800982	11.623807

domain=2

Number of Observations	10
Sum of Weights	118.867495

Variable	Sum	Std Dev
y	2476.339500	39.822977

domain=3

Number of Observations	10
Sum of Weights	101.797542

Variable	Sum	Std Dev
y	1831.623605	25.574899

domain=4

Number of Observations	10
Sum of Weights	87.696286

Variable	Sum	Std Dev
y	1856.489824	38.500505

domain=5

Number of Observations	10
Sum of Weights	87.3118074

Variable	Sum	Std Dev
y	1729.143337	24.558489

domain=6

Number of Observations	10
Sum of Weights	206.538879

Variable	Sum	Std Dev
y	4707.658301	72.225238

domain=7

Number of Observations	10
Sum of Weights	44.7585345

Variable	Sum	Std Dev
y	843.379702	13.887967

domain=8

Number of Observations	10
Sum of Weights	47.5680378

Variable	Sum	Std Dev
y	1007.974748	14.472507

domain=9

Number of Observations	10
Sum of Weights	39.3016818

Variable	Sum	Std Dev
y	874.070212	9.953448

domain=10

Number of Observations	10
Sum of Weights	174.710277

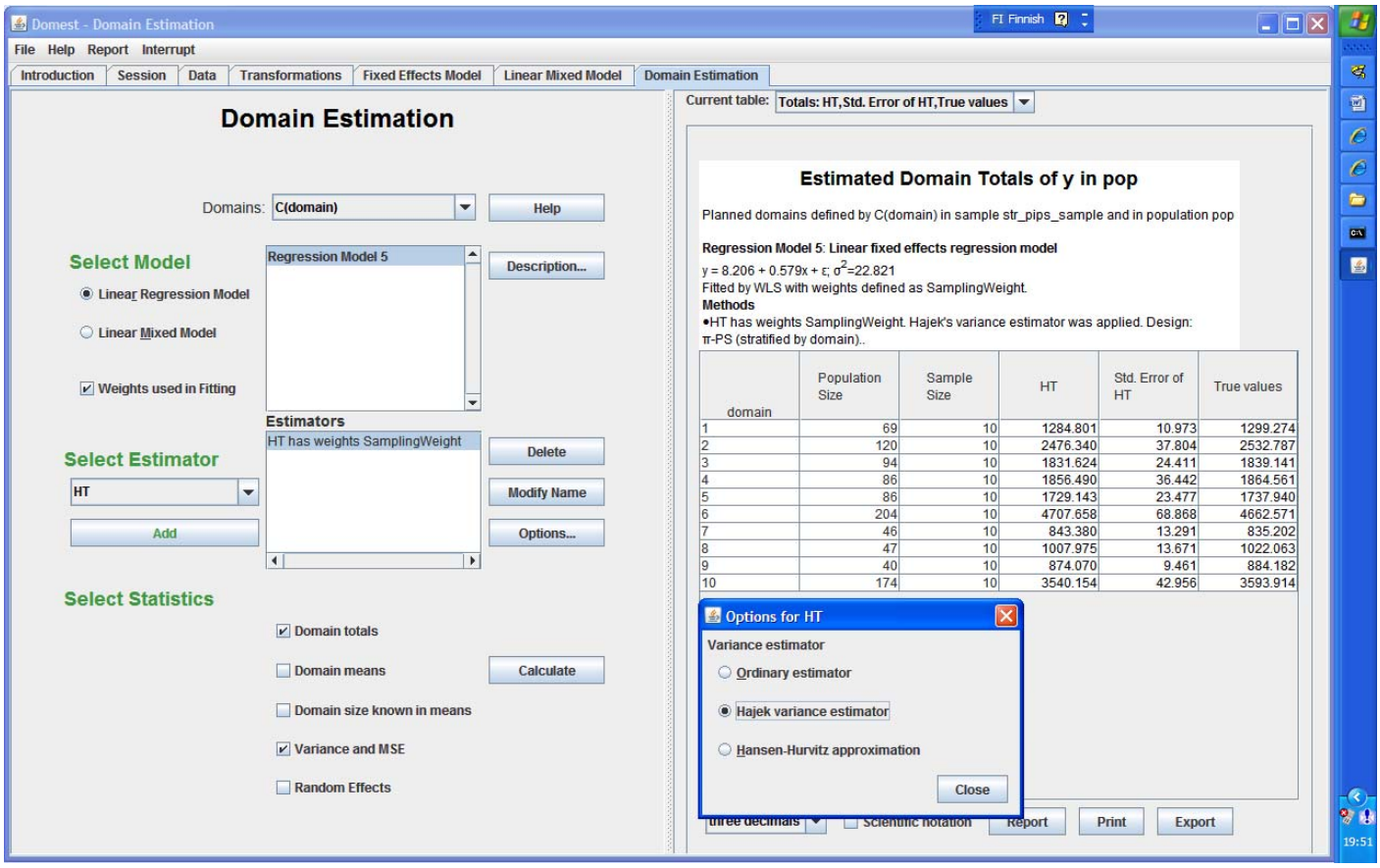
Variable	Sum	Std Dev
y	3540.154351	45.075031

1. Estimation for planned domains (strata)

1.2. STR_PiPS_sample

1.2.2. HT estimator OPTION: Hájek variance estimator (Formulas (1) and (3) in Part 1)

NOTE: Information on domain sizes in population N_d are incorporated in the estimation procedure



Estimated Domain Totals of y in pop

Planned domains defined by C(domain) in sample str_pips_sample and in population pop

Regression Model 5: **Linear fixed effects regression model**

$y = 8.206 + 0.579x + \varepsilon$; $\sigma^2=22.821$

Fitted by WLS with weights defined as SamplingWeight.

Methods

•HT has weights SamplingWeight. Hájek's variance estimator was applied. Design: π-PS (stratified by domain)..

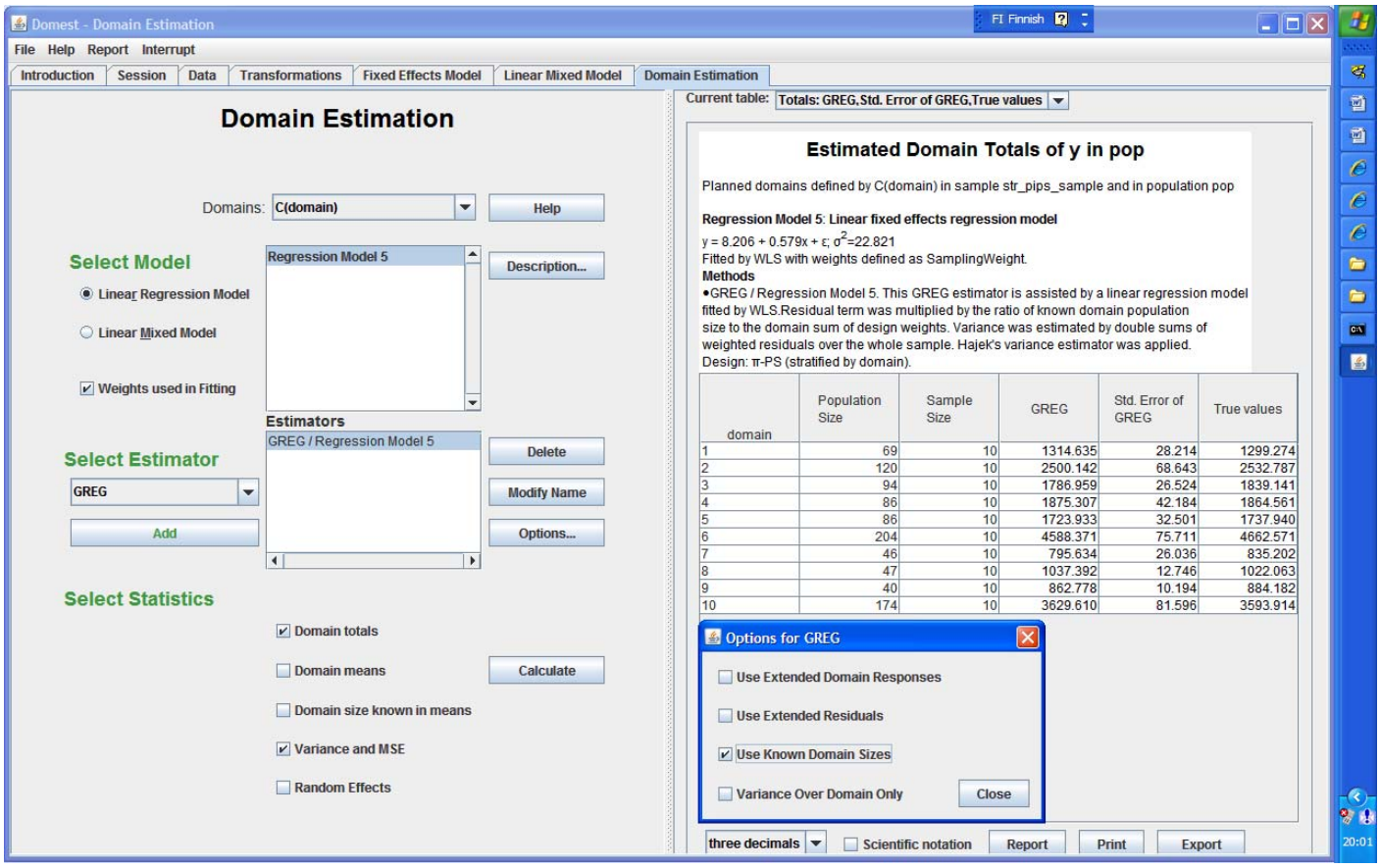
domain	Population Size	Sample Size	HT	Std. Error of HT	True values
1	69	10	1284.801	10.973	1299.274
2	120	10	2476.340	37.804	2532.787
3	94	10	1831.624	24.411	1839.141
4	86	10	1856.490	36.442	1864.561
5	86	10	1729.143	23.477	1737.940
6	204	10	4707.658	68.868	4662.571
7	46	10	843.380	13.291	835.202
8	47	10	1007.975	13.671	1022.063
9	40	10	874.070	9.461	884.182
10	174	10	3540.154	42.956	3593.914

1. Estimation for planned domains (strata)

1.2. STR_PiPS_sample

1.2.3. GREG estimator OPTION: Use Known Domain Sizes (Formula (11))

Assisting model: Linear fixed effects model $Y_k = \beta_0 + \beta_1 x_k + \varepsilon_k$



Estimated Domain Totals of y in pop

Planned domains defined by C(domain) in sample str_pips_sample and in population pop

Regression Model 5: Linear fixed effects regression model

$y = 8.206 + 0.579x + \varepsilon$; $\sigma^2 = 22.821$

Fitted by WLS with weights defined as SamplingWeight**.**

Methods

•GREG / Regression Model 5. This GREG estimator is assisted by a linear regression model fitted by WLS. Residual term was multiplied by the ratio of known domain population size to the domain sum of design weights. Variance was estimated by double sums of weighted residuals over the whole sample. Hájek's variance estimator was applied.

Design: π-PS (stratified by domain).

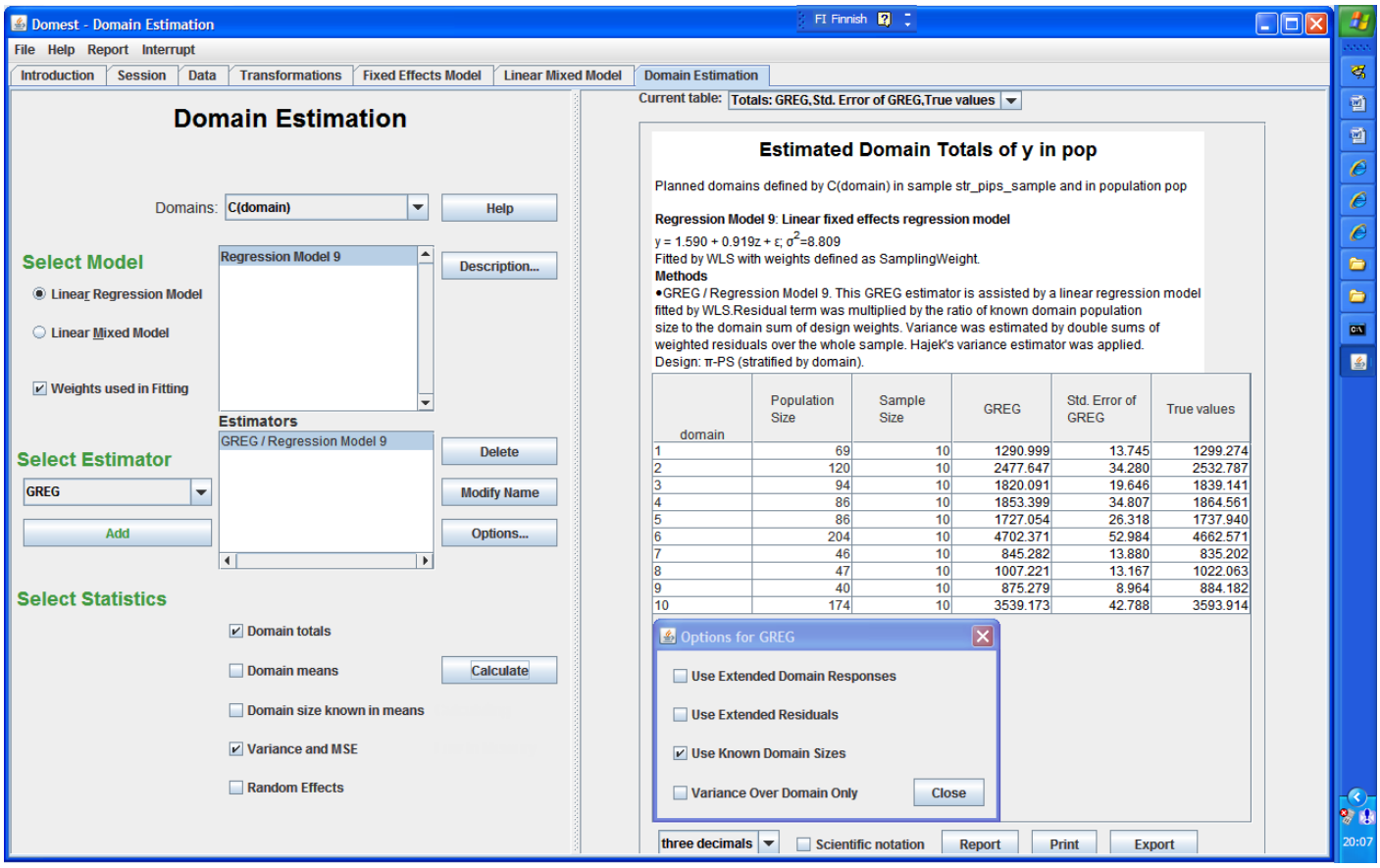
domain	Population Size	Sample Size	GREG	Std. Error of GREG	True values
1	69	10	1314.635	28.214	1299.274
2	120	10	2500.142	68.643	2532.787
3	94	10	1786.959	26.524	1839.141
4	86	10	1875.307	42.184	1864.561
5	86	10	1723.933	32.501	1737.940
6	204	10	4588.371	75.711	4662.571
7	46	10	795.634	26.036	835.202
8	47	10	1037.392	12.746	1022.063
9	40	10	862.778	10.194	884.182
10	174	10	3629.610	81.596	3593.914

1. Estimation for planned domains (strata)

1.2. STR_PiPS_sample

1.2.4. GREG estimator OPTION: Use Known Domain Sizes (Formula (11))

Assisting model: Linear fixed effects model $Y_k = \beta_0 + \beta_1 z_k + \varepsilon_k$



Estimated Domain Totals of y in pop

Planned domains defined by C(domain) in sample str_pips_sample and in population pop

Regression Model 9: Linear fixed effects regression model

$y = 1.590 + 0.919z + \varepsilon$; $\sigma^2=8.809$

Fitted by WLS with weights defined as SamplingWeight**.**

Methods

•GREG / Regression Model 9. This GREG estimator is assisted by a linear regression model fitted by WLS. Residual term was multiplied by the ratio of known domain population size to the domain sum of design weights. Variance was estimated by double sums of weighted residuals over the whole sample. Hájek's variance estimator was applied.

Design: π-PS (stratified by domain).

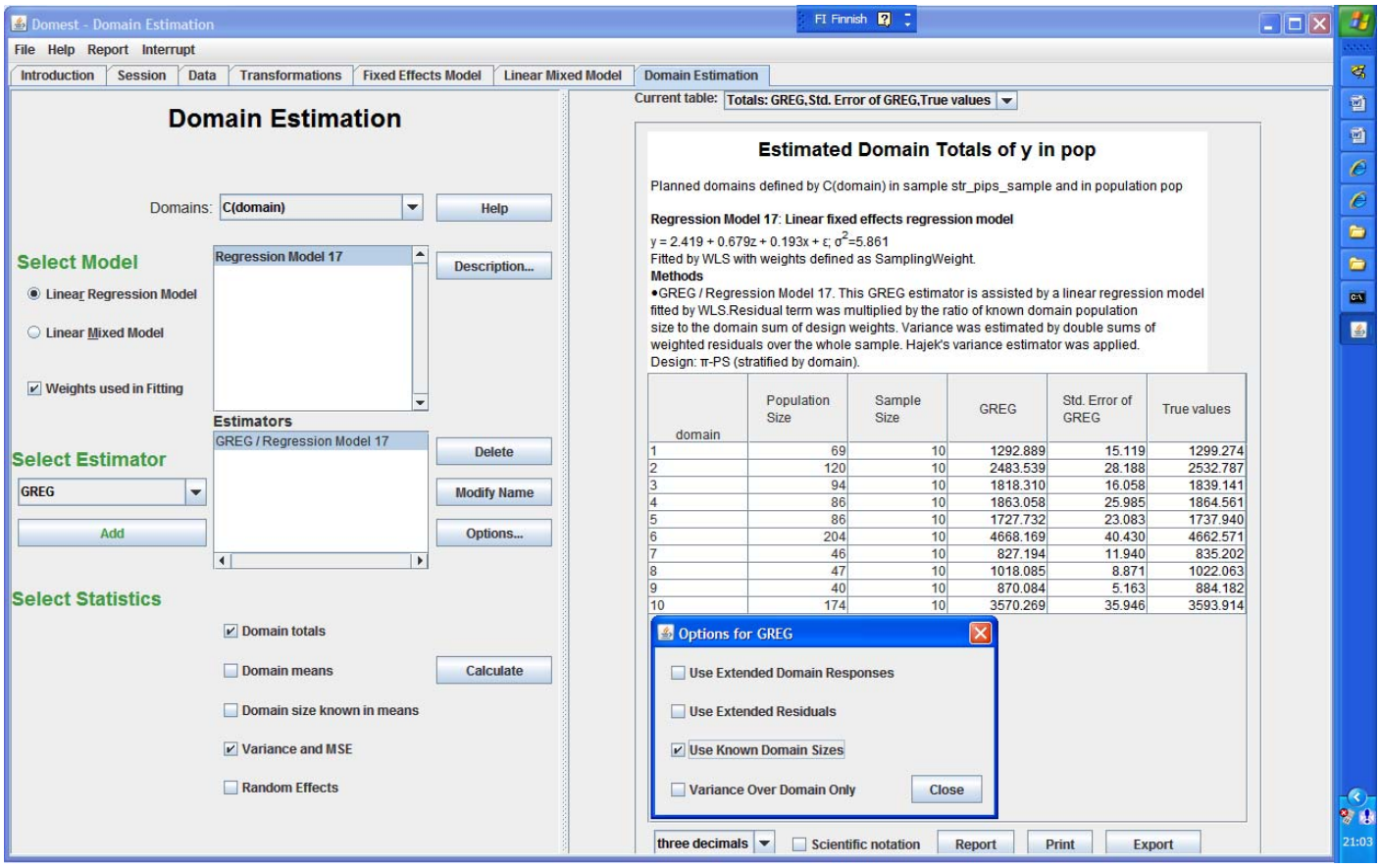
domain	Population Size	Sample Size	GREG	Std. Error of GREG	True values
1	69	10	1290.999	13.745	1299.274
2	120	10	2477.647	34.280	2532.787
3	94	10	1820.091	19.646	1839.141
4	86	10	1853.399	34.807	1864.561
5	86	10	1727.054	26.318	1737.940
6	204	10	4702.371	52.984	4662.571
7	46	10	845.282	13.880	835.202
8	47	10	1007.221	13.167	1022.063
9	40	10	875.279	8.964	884.182
10	174	10	3539.173	42.788	3593.914

1. Estimation for planned domains (strata)

1.2. STR_PiPS_sample

1.2.5. GREG estimator OPTION: Use Known Domain Sizes (Formula (11))

Assisting model: Linear fixed effects model $Y_k = \beta_0 + \beta_1 z_k + \beta_2 x_k + \varepsilon_k$



Estimated Domain Totals of y in pop

Planned domains defined by C(domain) in sample str_pips_sample and in population pop

Regression Model 17: Linear fixed effects regression model

$y = 2.419 + 0.679z + 0.193x + \varepsilon$; $\sigma^2=5.861$

Fitted by WLS with weights defined as SamplingWeight**.**

Methods

•GREG / Regression Model 17. This GREG estimator is assisted by a linear regression model fitted by WLS. Residual term was multiplied by the ratio of known domain population size to the domain sum of design weights. Variance was estimated by double sums of weighted residuals over the whole sample. Hájek's variance estimator was applied.

Design: π-PS (stratified by domain).

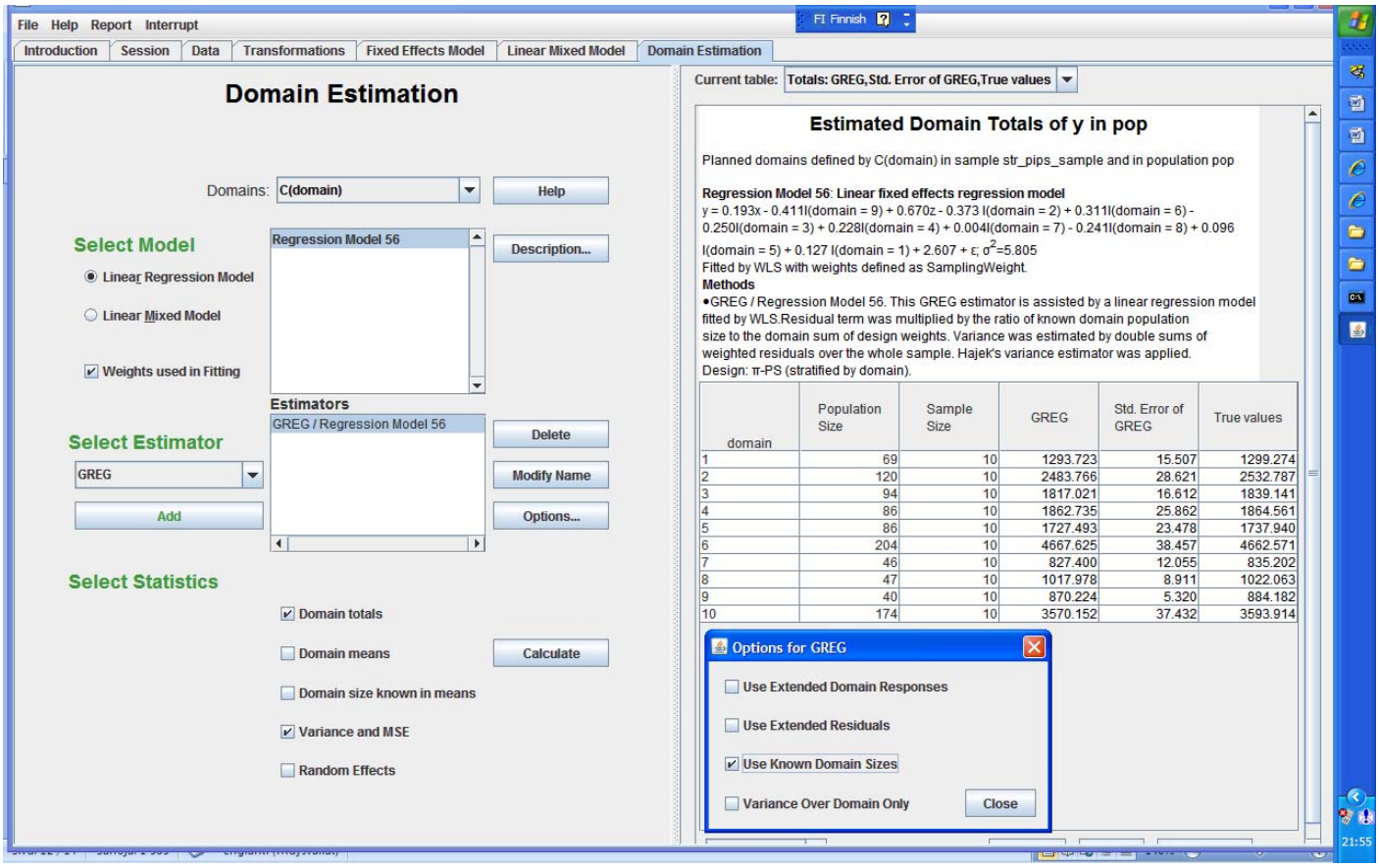
domain	Population Size	Sample Size	GREG	Std. Error of GREG	True values
1	69	10	1292.889	15.119	1299.274
2	120	10	2483.539	28.188	2532.787
3	94	10	1818.310	16.058	1839.141
4	86	10	1863.058	25.985	1864.561
5	86	10	1727.732	23.083	1737.940
6	204	10	4668.169	40.430	4662.571
7	46	10	827.194	11.940	835.202
8	47	10	1018.085	8.871	1022.063
9	40	10	870.084	5.163	884.182
10	174	10	3570.269	35.946	3593.914

1. Estimation for planned domains (strata)

1.2. STR_PiPS_sample

1.2.6. GREG estimator OPTION: Use Known Domain Sizes (Formula (11))

Assisting model: Linear fixed effects model $Y_k = \beta_0 + \beta_1 z_k + \beta_2 x_k + \gamma_1 I_{1k} + ... + \gamma_9 I_{9k} + \varepsilon_k$



Estimated Domain Totals of y in pop

Planned domains defined by C(domain) in sample str_pips_sample and in population pop

Regression Model 56: **Linear fixed effects regression model**

y = 0.193x - 0.411I(domain = 9) + 0.670z - 0.373 I(domain = 2) + 0.311I(domain = 6) - 0.250I(domain = 3) + 0.228I(domain = 4) + 0.004I(domain = 7) - 0.241I(domain = 8) + 0.096 I(domain = 5) + 0.127 I(domain = 1) + 2.607 + ε; σ^2=5.805

Fitted by WLS with weights defined as SamplingWeight.

Methods

•GREG / Regression Model 56. This GREG estimator is assisted by a linear regression model fitted by WLS. Residual term was multiplied by the ratio of known domain population size to the domain sum of design weights. Variance was estimated by double sums of weighted residuals over the whole sample. Hájek's variance estimator was applied.

Design: π-PS (stratified by domain).

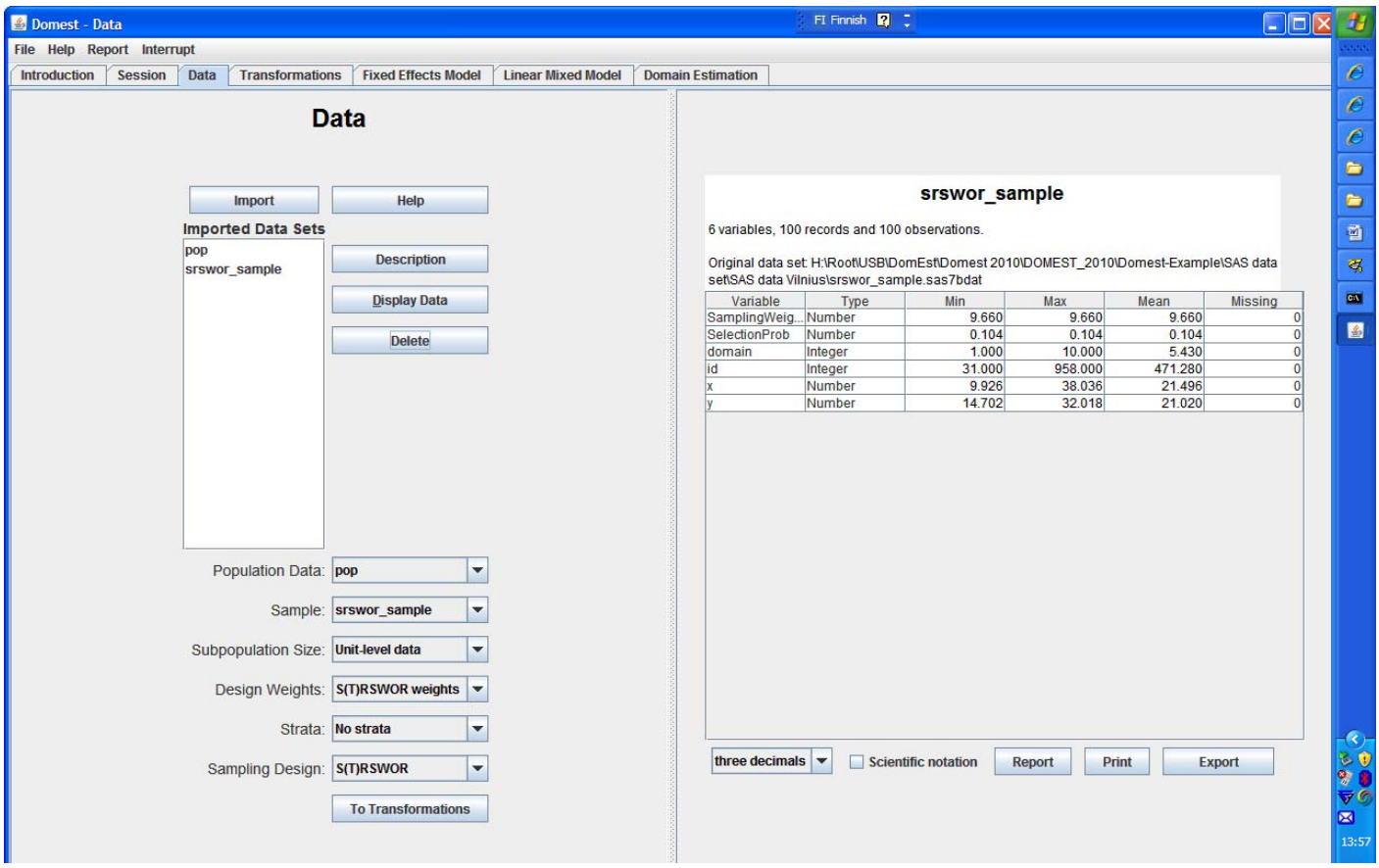
domain	Population Size	Sample Size	GREG	Std. Error of GREG	True values
1	69	10	1293.723	15.507	1299.274
2	120	10	2483.766	28.621	2532.787
3	94	10	1817.021	16.612	1839.141
4	86	10	1862.735	25.862	1864.561
5	86	10	1727.493	23.478	1737.940
6	204	10	4667.625	38.457	4662.571
7	46	10	827.400	12.055	835.202
8	47	10	1017.978	8.911	1022.063
9	40	10	870.224	5.320	884.182
10	174	10	3570.152	37.432	3593.914

2. ESTIMATION FOR UNPLANNED DOMAINS

2.1. SRSWOR_sample

SRSWOR sampling

NOTE: domain sample sizes n_d are random variates, the overall sample size $n = \sum_{d=1}^{10} n_d = 100$




```
SRSWOR_sample

The CORR Procedure (Unweighted results)

Variable          N          Mean       Std Dev           Sum       Minimum       Maximum

y               100      21.01992       3.22337          2102      14.70173      32.01814
x               100      21.49628       5.16094          2150       9.92621      38.03629

Pearson Correlation Coefficients, N = 100
        Prob > |r| under H0: Rho=0

                y             x

y       1.00000       0.87281
                       <.0001

x       0.87281       1.00000
         <.0001
```

2. Estimation for unplanned domains
2.1. SRSWOR_sample
2.1.1. HT estimator SAS/SURVEYMEANS

NOTE: Information on domain sizes in population N_d **are not used**

```
proc surveymeans data=SRSWOR_sample sum;
domain domain;
var y;
weight samplingweight;
run;
```

SRSWOR_sample

The SURVEYMEANS Procedure

Data Summary

Number of Observations	100
Sum of Weights	966

Statistics

Variable	Sum	Std Dev
y	20305	311.377232

Domain Analysis: domain

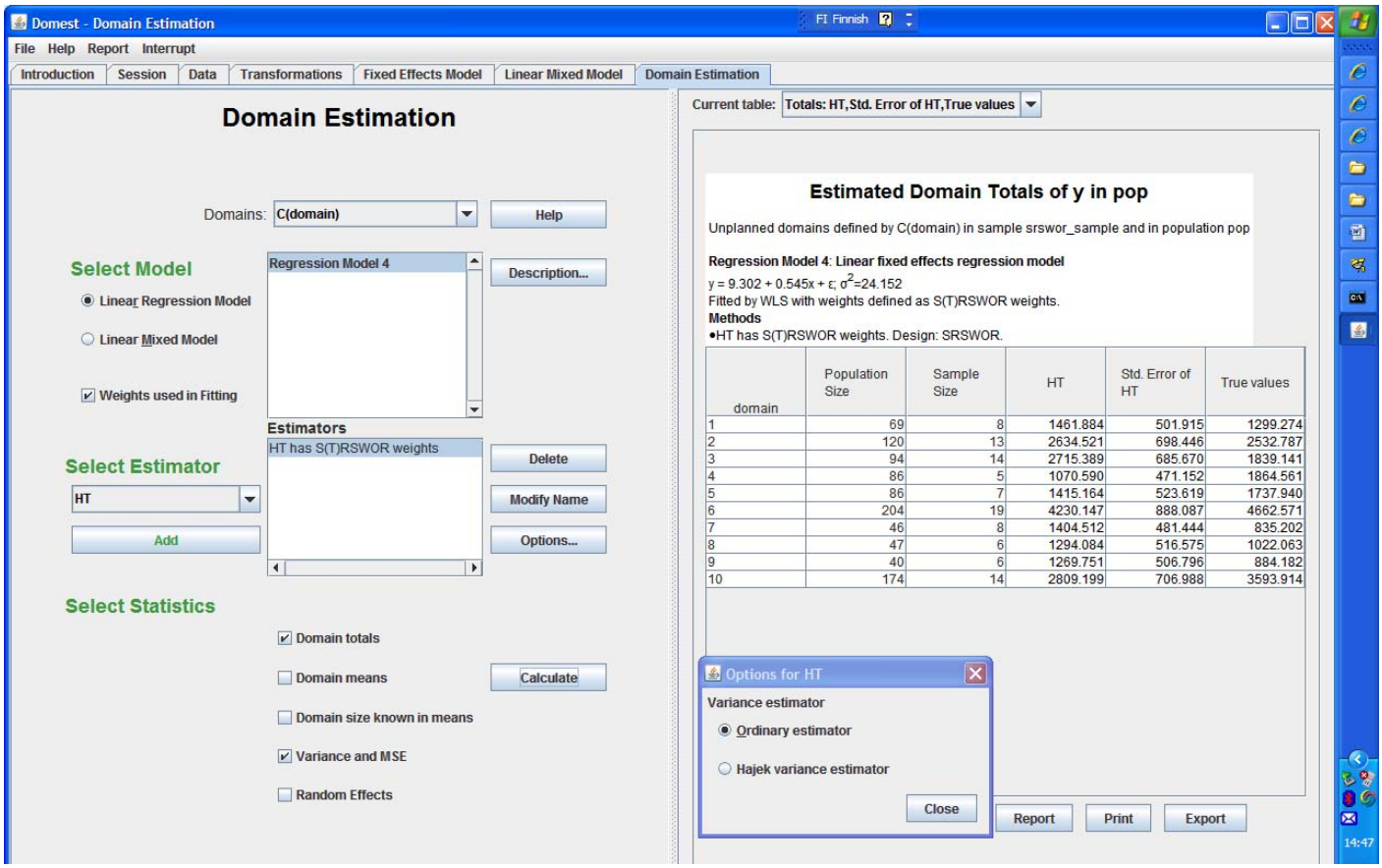
domain	Variable	Sum	Std Dev
1	y	1461.883611	501.915499
2	y	2634.520747	698.445513
3	y	2715.389185	685.669522
4	y	1070.590128	471.152357
5	y	1415.164344	523.619371
6	y	4230.147155	888.087280
7	y	1404.511876	481.444290
8	y	1294.083773	516.574859
9	y	1269.751284	506.796472
10	y	2809.199247	706.988188

2. Estimation for unplanned domains

2.1. SRSWOR_sample

2.1.2. HT estimator OPTION: Ordinary variance estimator (Formulas (1) and (5))

NOTE: Same results as in 2.1.1.



Estimated Domain Totals of y in pop

Unplanned domains defined by C(domain) in sample srswor_sample and in population pop

Regression Model 4: **Linear fixed effects regression model**

$y = 9.302 + 0.545x + \varepsilon$; $\sigma^2=24.152$

Fitted by WLS with weights defined as S(T)RSWOR weights.

Methods

•HT has S(T)RSWOR weights. Design: SRSWOR.

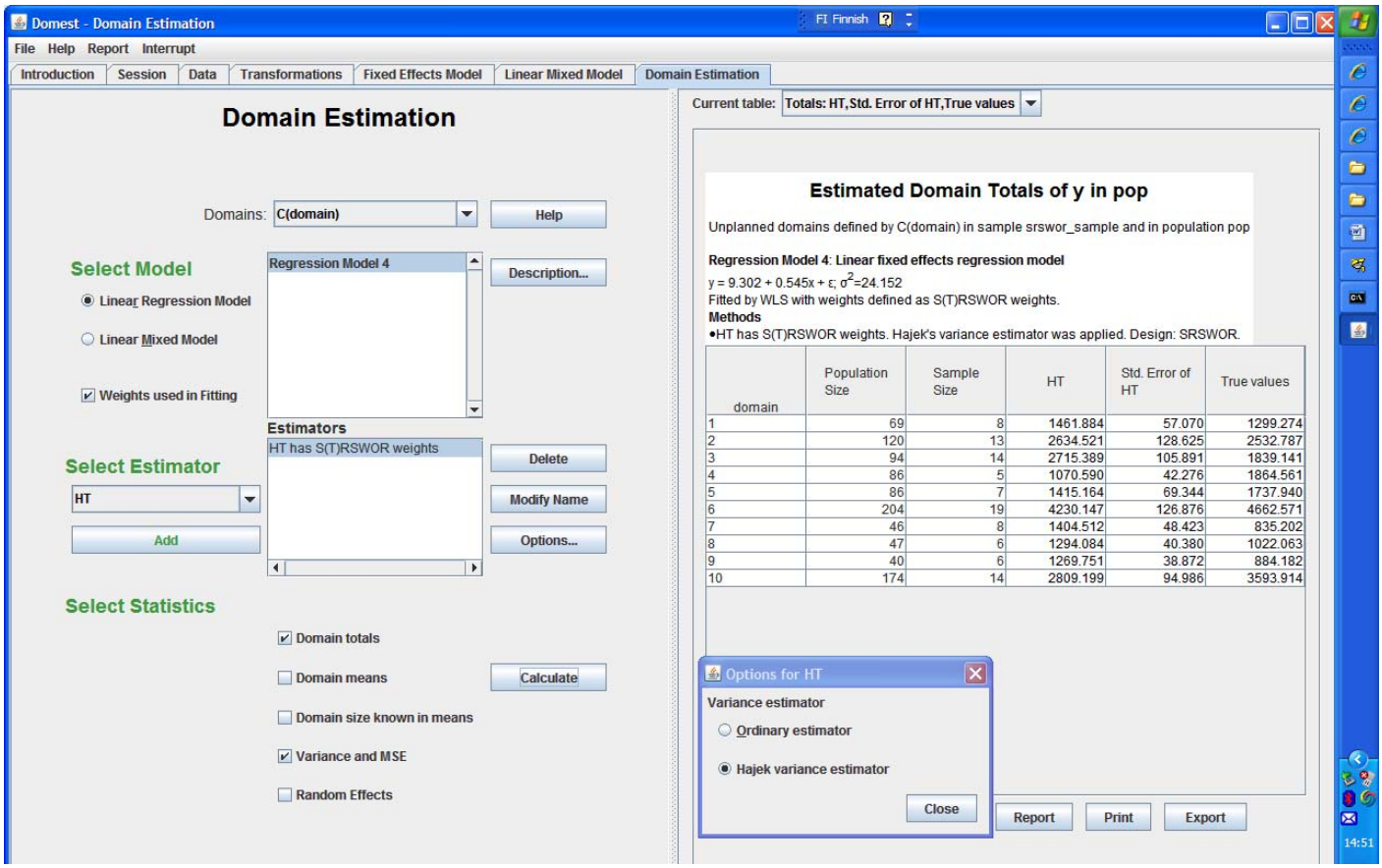
domain	Population Size	Sample Size	HT	Std. Error of HT	True values
1	69	8	1461.884	501.915	1299.274
2	120	13	2634.521	698.446	2532.787
3	94	14	2715.389	685.670	1839.141
4	86	5	1070.590	471.152	1864.561
5	86	7	1415.164	523.619	1737.940
6	204	19	4230.147	888.087	4662.571
7	46	8	1404.512	481.444	835.202
8	47	6	1294.084	516.575	1022.063
9	40	6	1269.751	506.796	884.182
10	174	14	2809.199	706.988	3593.914

2. Estimation for unplanned domains

2.1. SRSWOR_sample

2.1.3. HT estimator OPTION: Hájek variance estimator (Formulas (1) and (3))

NOTE: Information on domain sizes in population N_d are now used



Estimated Domain Totals of y in pop

Unplanned domains defined by C(domain) in sample srswor_sample and in population pop

Regression Model 4: Linear fixed effects regression model

$y = 9.302 + 0.545x + \varepsilon$; $\sigma^2=24.152$

Fitted by WLS with weights defined as S(T)RSWOR weights.

Methods

•HT has S(T)RSWOR weights. Hájek's variance estimator was applied. Design: SRSWOR.

domain	Population Size	Sample Size	HT	Std. Error of HT	True values
1	69	8	1461.884	57.070	1299.274
2	120	13	2634.521	128.625	2532.787
3	94	14	2715.389	105.891	1839.141
4	86	5	1070.590	42.276	1864.561
5	86	7	1415.164	69.344	1737.940
6	204	19	4230.147	126.876	4662.571
7	46	8	1404.512	48.423	835.202
8	47	6	1294.084	40.380	1022.063
9	40	6	1269.751	38.872	884.182
10	174	14	2809.199	94.986	3593.914

2. Estimation for unplanned domains
2.1. SRSWOR_sample
2.1.4. GREG estimation
Assisting model: Linear fixed effects model $Y_k = \beta_0 + \beta_1 x_k + \varepsilon_k$

SAS Macro EBLUPGREG
EURAREA Project

GREG estimation
Linear fixed-effects model

```
%eblupgreg(sample=SRSWOR_sample,
          population=pop,
          weights=SamplingWeight,
          y=y,
          xlist=x,
          regionIdentifier=domain,
          modules=modules.eurarea,
          estimatemeans=0,
          eblup=0,
          greg=1,
          synthetic=0,
          test=1,
          output=GREG1);
          run;
```

SRSWOR_sample

PARAMETERS

parameter	in GREG
intercept	9.3016304098
X	0.5451310135

GREG estimation / Unplanned domains / SRSWOR_sample

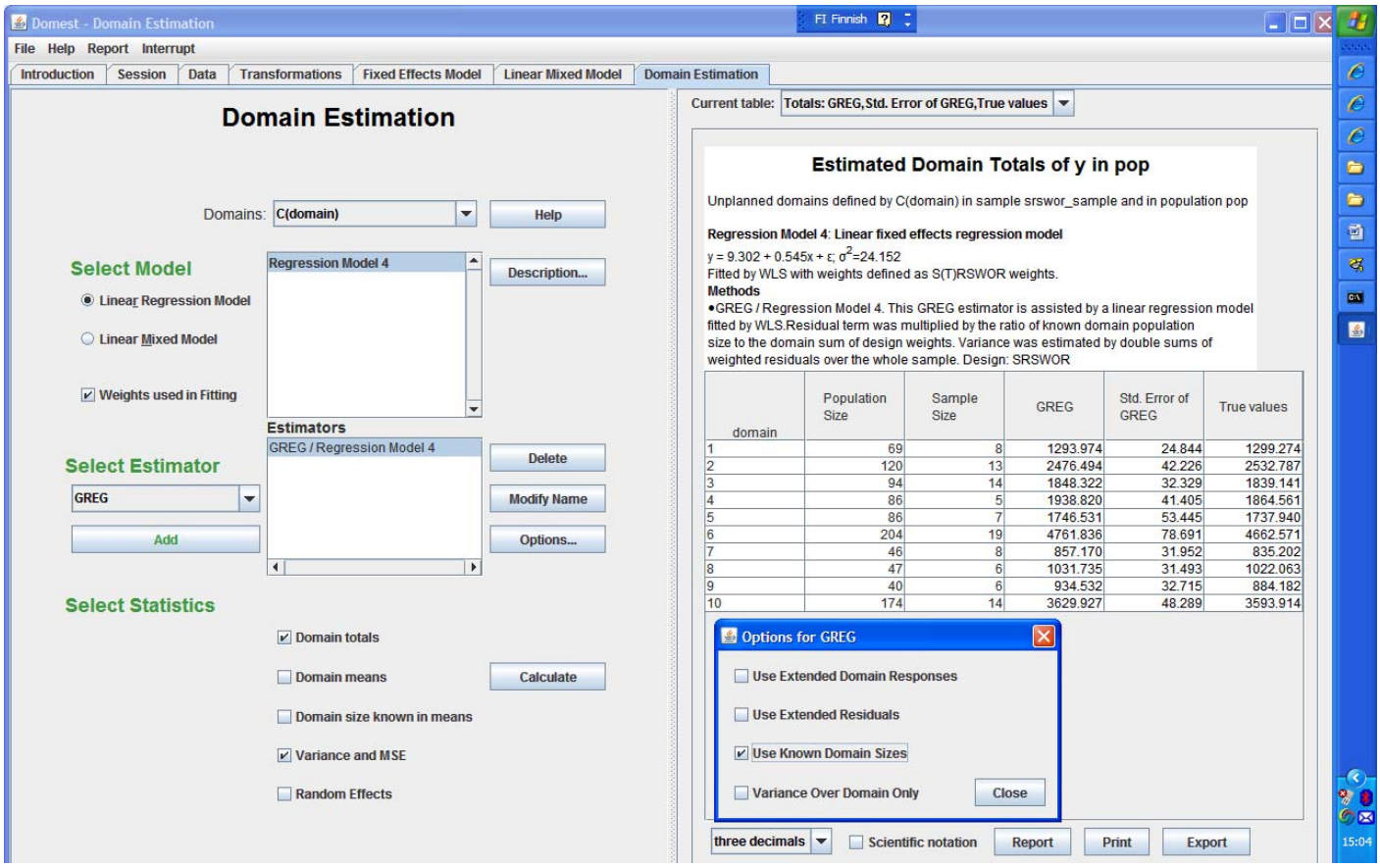
Obs	region	GREG	stdGREG	true Value
1	1	1292.93	24.8442	1299.27
2	2	2472.54	42.2263	2532.79
3	3	1816.11	32.3286	1839.14
4	4	1913.76	41.4054	1864.56
5	5	1747.49	53.4452	1737.94
6	6	4745.43	78.6909	4662.57
7	7	874.72	31.9516	835.20
8	8	1026.49	31.4930	1022.06
9	9	939.52	32.7146	884.18
10	10	3630.78	48.2893	3593.91

2. Estimation for unplanned domains

2.1. SRSWOR_sample

2.1.5. GREG estimation OPTION: Known Domain Sizes (Formula (11))

Assisting model: Linear fixed effects model $Y_k = \beta_0 + \beta_1 x_k + \varepsilon_k$



Estimated Domain Totals of y in pop

Unplanned domains defined by C(domain) in sample srswor_sample and in population pop

Regression Model 4: Linear fixed effects regression model

$y = 9.302 + 0.545x + \varepsilon$; $\sigma^2=24.152$

Fitted by WLS with weights defined as S(T)RSWOR weights.

Methods

•GREG / Regression Model 4. This GREG estimator is assisted by a linear regression model fitted by WLS. Residual term was multiplied by the ratio of known domain population size to the domain sum of design weights. Variance was estimated by double sums of weighted residuals over the whole sample. Design: SRSWOR

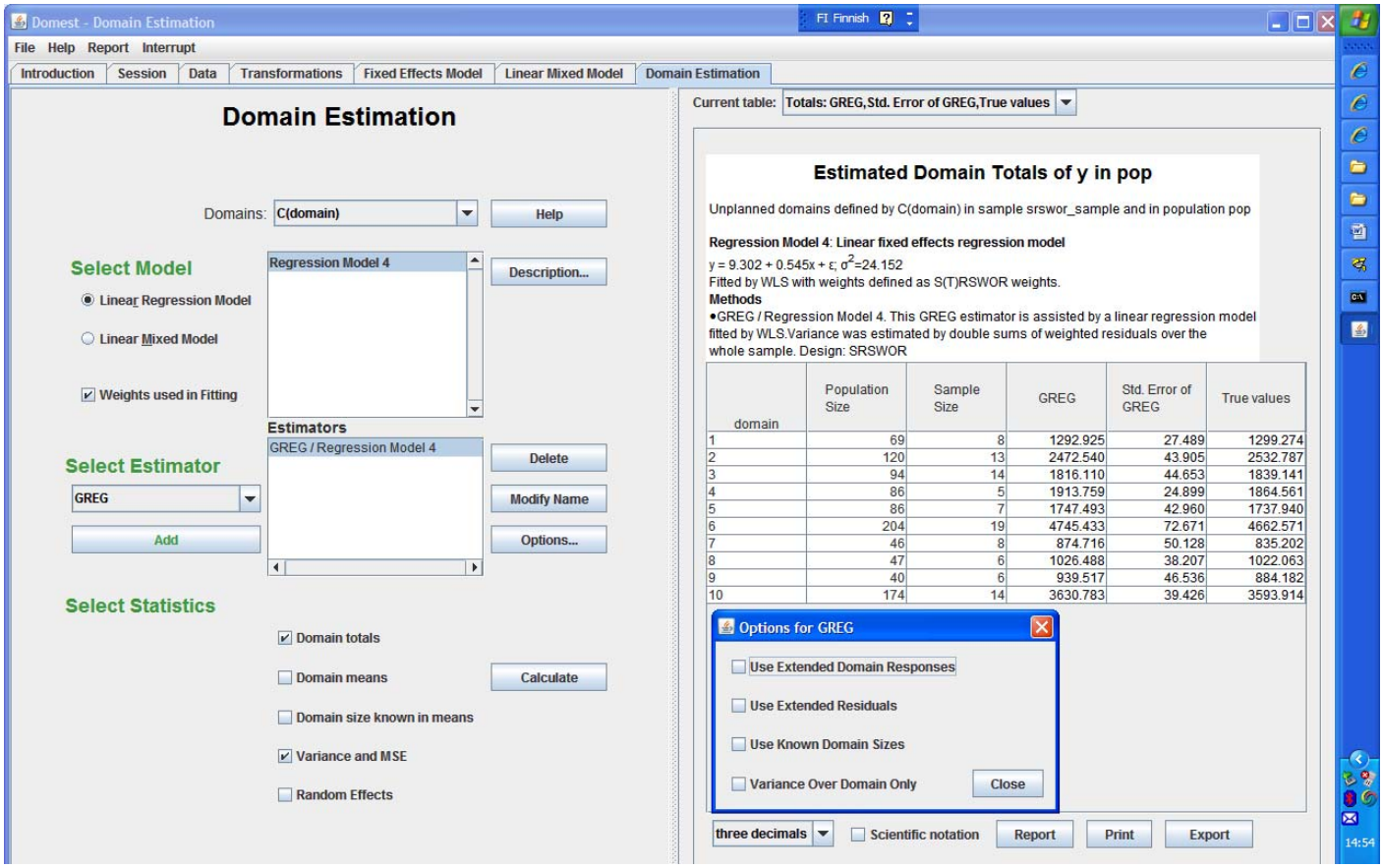
domain	Population Size	Sample Size	GREG	Std. Error of GREG	True values
1	69	8	1293.974	24.844	1299.274
2	120	13	2476.494	42.226	2532.787
3	94	14	1848.322	32.329	1839.141
4	86	5	1938.820	41.405	1864.561
5	86	7	1746.531	53.445	1737.940
6	204	19	4761.836	78.691	4662.571
7	46	8	857.170	31.952	835.202
8	47	6	1031.735	31.493	1022.063
9	40	6	934.532	32.715	884.182
10	174	14	3629.927	48.289	3593.914

2. Estimation for unplanned domains

2.1. SRSWOR_sample

2.1.6. GREG estimator OPTION: Ordinary variance estimator (Formulas (8) and (9))

Assisting model: Linear fixed effects model $Y_k = \beta_0 + \beta_1 x_k + \varepsilon_k$



Estimated Domain Totals of y in pop

Unplanned domains defined by C(domain) in sample srswor_sample and in population pop

Regression Model 4: Linear fixed effects regression model

$y = 9.302 + 0.545x + \varepsilon$; $\sigma^2=24.152$

Fitted by WLS with weights defined as S(T)RSWOR weights.

Methods

•GREG / Regression Model 4. This GREG estimator is assisted by a linear regression model fitted by WLS. Variance was estimated by double sums of weighted residuals over the whole sample. Design: SRSWOR

domain	Population Size	Sample Size	GREG	Std. Error of GREG	True values
1	69	8	1292.925	27.489	1299.274
2	120	13	2472.540	43.905	2532.787
3	94	14	1816.110	44.653	1839.141
4	86	5	1913.759	24.899	1864.561
5	86	7	1747.493	42.960	1737.940
6	204	19	4745.433	72.671	4662.571
7	46	8	874.716	50.128	835.202
8	47	6	1026.488	38.207	1022.063
9	40	6	939.517	46.536	884.182
10	174	14	3630.783	39.426	3593.914

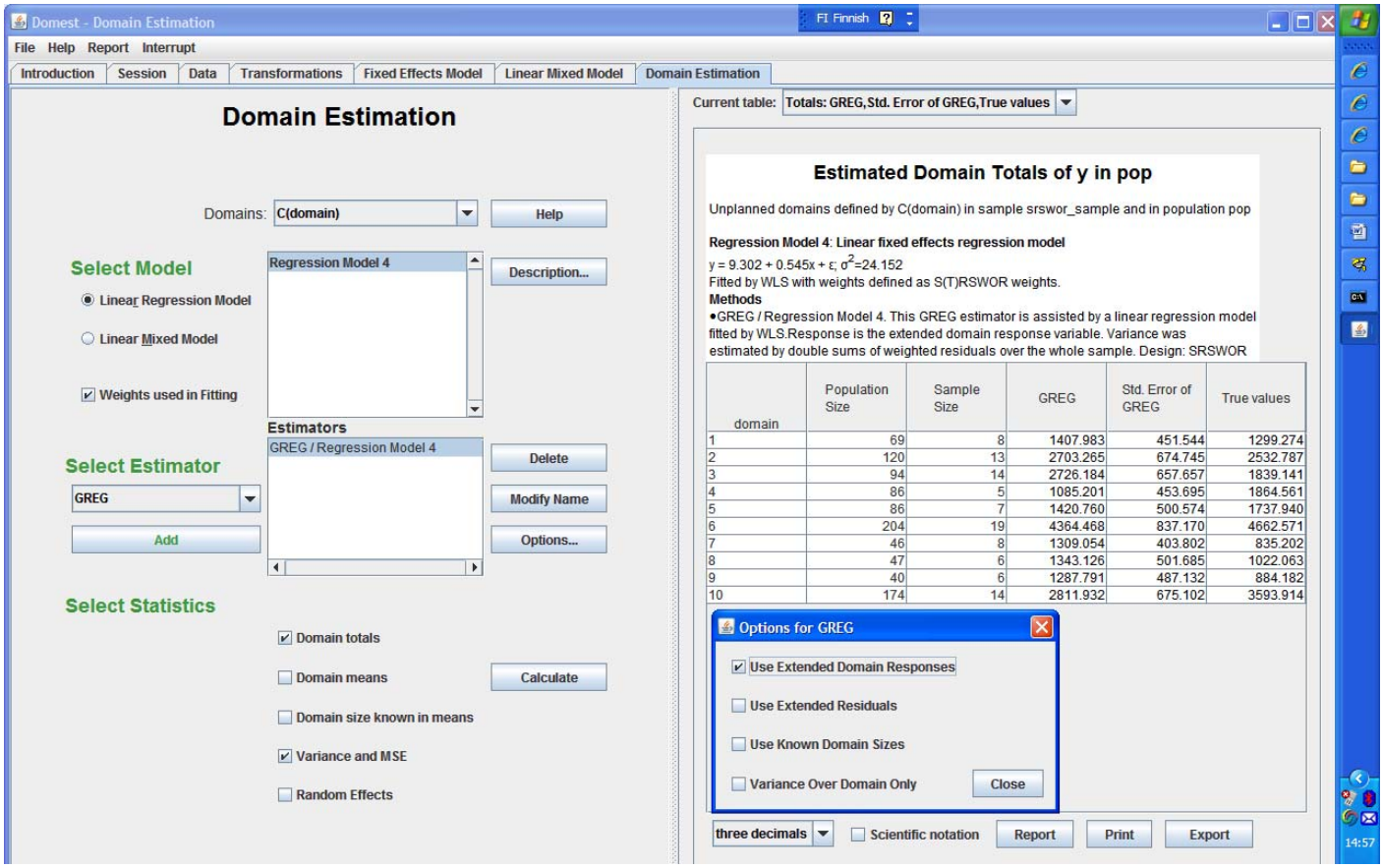
2. Estimation for unplanned domains

2.1. SRSWOR_sample

2.1.7. GREG estimator OPTION: Extended Domain Responses (Option 2 on p. 17, Part 1)

NOTE: Information on domain sizes in population **not used**

Assisting model: Linear fixed effects model $Y_k = \beta_0 + \beta_1 x_k + \varepsilon_k$



Estimated Domain Totals of y in pop

Unplanned domains defined by C(domain) in sample srswor_sample and in population pop

Regression Model 4: Linear fixed effects regression model

$y = 9.302 + 0.545x + \varepsilon$; $\sigma^2=24.152$

Fitted by WLS with weights defined as S(T)RSWOR weights.

Methods

•GREG / Regression Model 4. This GREG estimator is assisted by a linear regression model fitted by WLS. Response is the extended domain response variable. Variance was estimated by double sums of weighted residuals over the whole sample. Design: SRSWOR

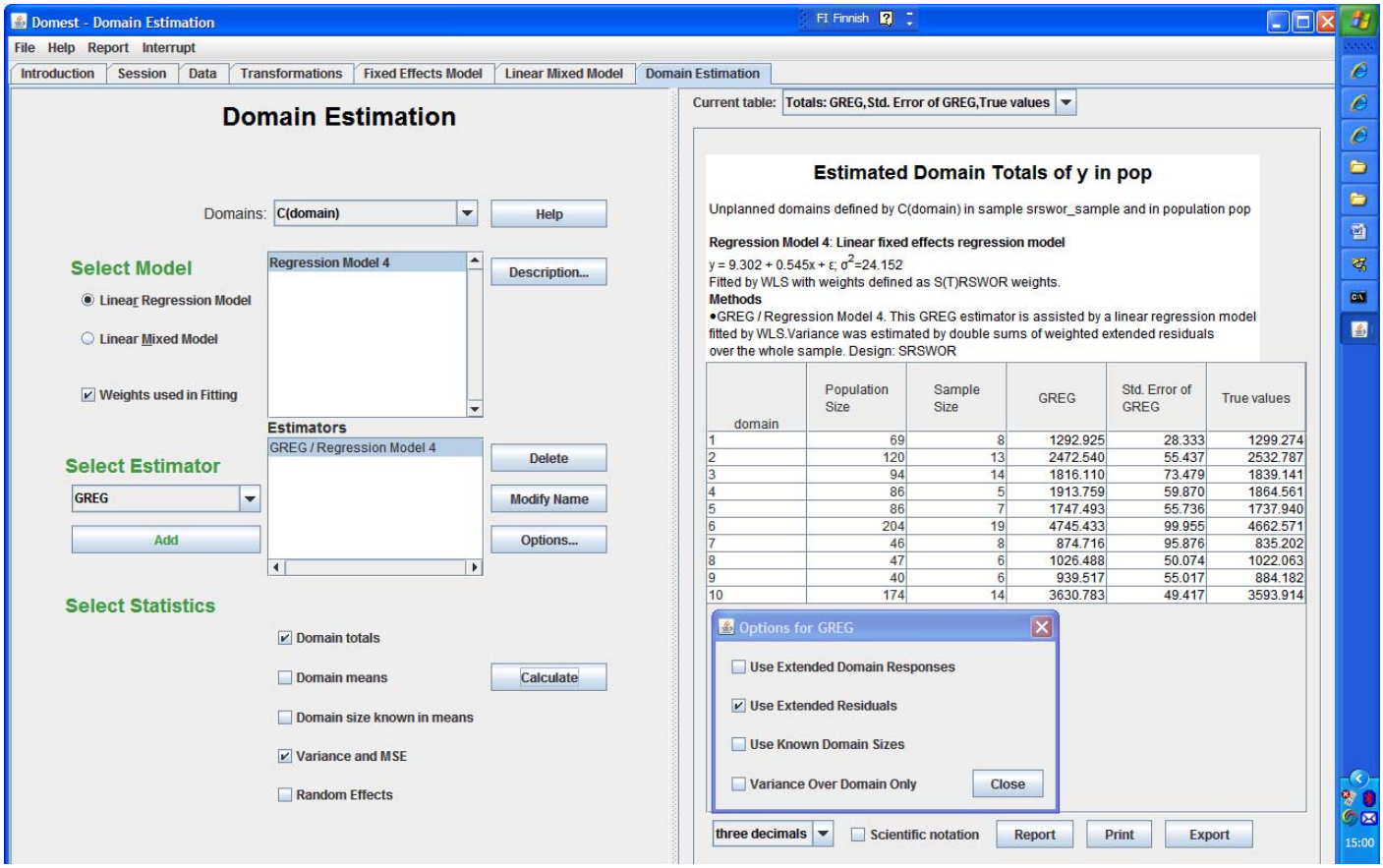
domain	Population Size	Sample Size	GREG	Std. Error of GREG	True values
1	69	8	1407.983	451.544	1299.274
2	120	13	2703.265	674.745	2532.787
3	94	14	2726.184	657.657	1839.141
4	86	5	1085.201	453.695	1864.561
5	86	7	1420.760	500.574	1737.940
6	204	19	4364.468	837.170	4662.571
7	46	8	1309.054	403.802	835.202
8	47	6	1343.126	501.685	1022.063
9	40	6	1287.791	487.132	884.182
10	174	14	2811.932	675.102	3593.914

2. Estimation for unplanned domains

2.1. SRSWOR_sample

2.1.8. GREG estimator OPTION: Extended Residuals (Option 3 on p. 17, Part 1)

Assisting model: Linear fixed effects model $Y_k = \beta_0 + \beta_1 x_k + \varepsilon_k$



Estimated Domain Totals of y in pop

Unplanned domains defined by C(domain) in sample srswor_sample and in population pop

Regression Model 4: Linear fixed effects regression model

$y = 9.302 + 0.545x + \varepsilon$; $\sigma^2=24.152$

Fitted by WLS with weights defined as S(T)RSWOR weights.

Methods

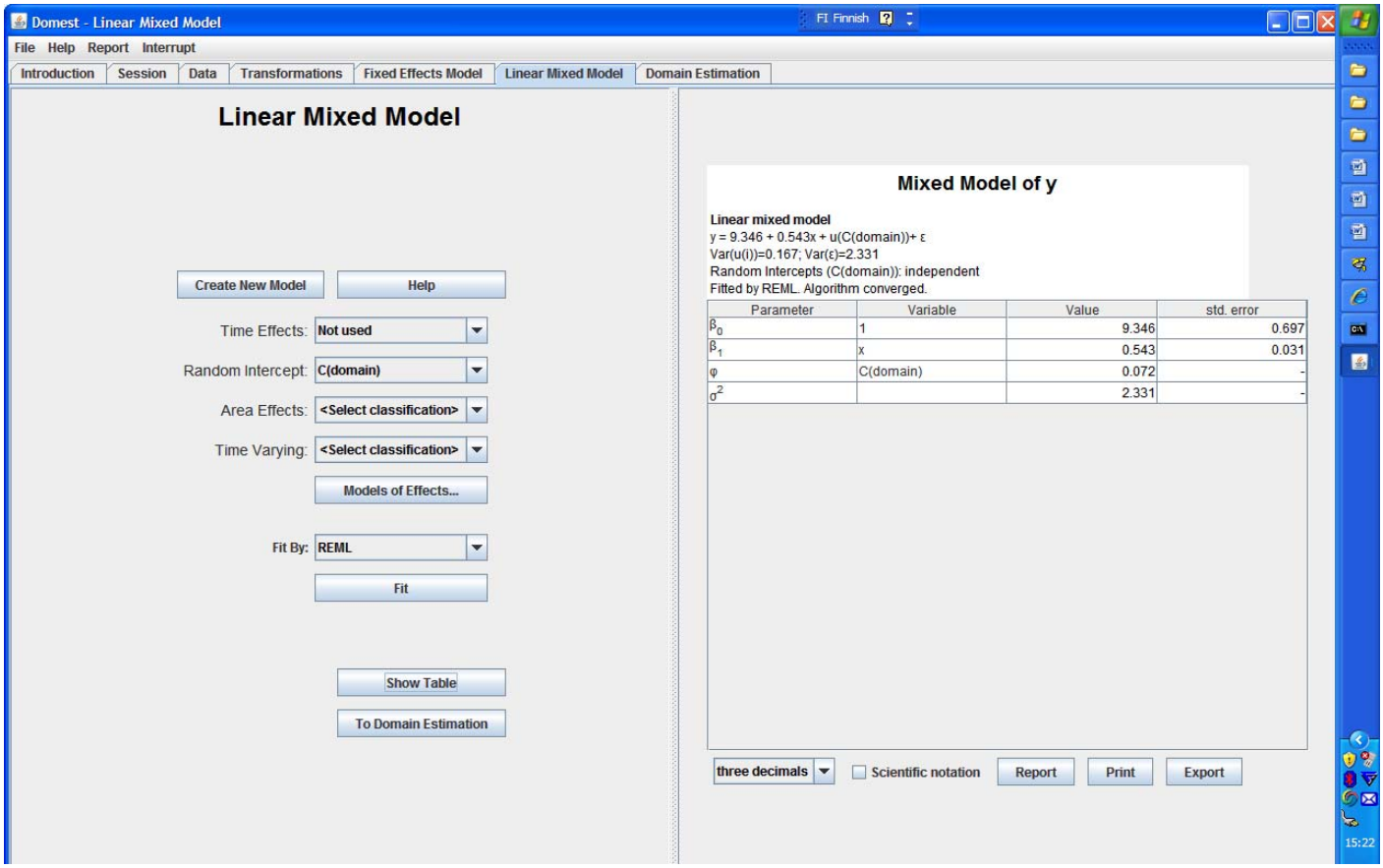
•GREG / Regression Model 4. This GREG estimator is assisted by a linear regression model fitted by WLS. Variance was estimated by double sums of weighted extended residuals over the whole sample. Design: SRSWOR

domain	Population Size	Sample Size	GREG	Std. Error of GREG	True values
1	69	8	1292.925	28.333	1299.274
2	120	13	2472.540	55.437	2532.787
3	94	14	1816.110	73.479	1839.141
4	86	5	1913.759	59.870	1864.561
5	86	7	1747.493	55.736	1737.940
6	204	19	4745.433	99.955	4662.571
7	46	8	874.716	95.876	835.202
8	47	6	1026.488	50.074	1022.063
9	40	6	939.517	55.017	884.182
10	174	14	3630.783	49.417	3593.914

2. Estimation for unplanned domains

2.1. SRSWOR_sample

2.1.9. Linear mixed model $Y_k = \beta_0 + u_d + \beta_1 x_k + \varepsilon_k$ with domain-specific random intercepts u_d



Mixed Model of y

Linear mixed model

y = 9.346 + 0.543x + u(C(domain))+ ε

Var(u(i))=0.167; Var(ε)=2.331

Random Intercepts (C(domain)): independent

Fitted by REML. Algorithm converged.

Parameter	Variable	Value	std. error
β_0	1	9.346	0.697
β_1	x	0.543	0.031
φ	C(domain)	0.072	-
σ^2	.	2.331	-

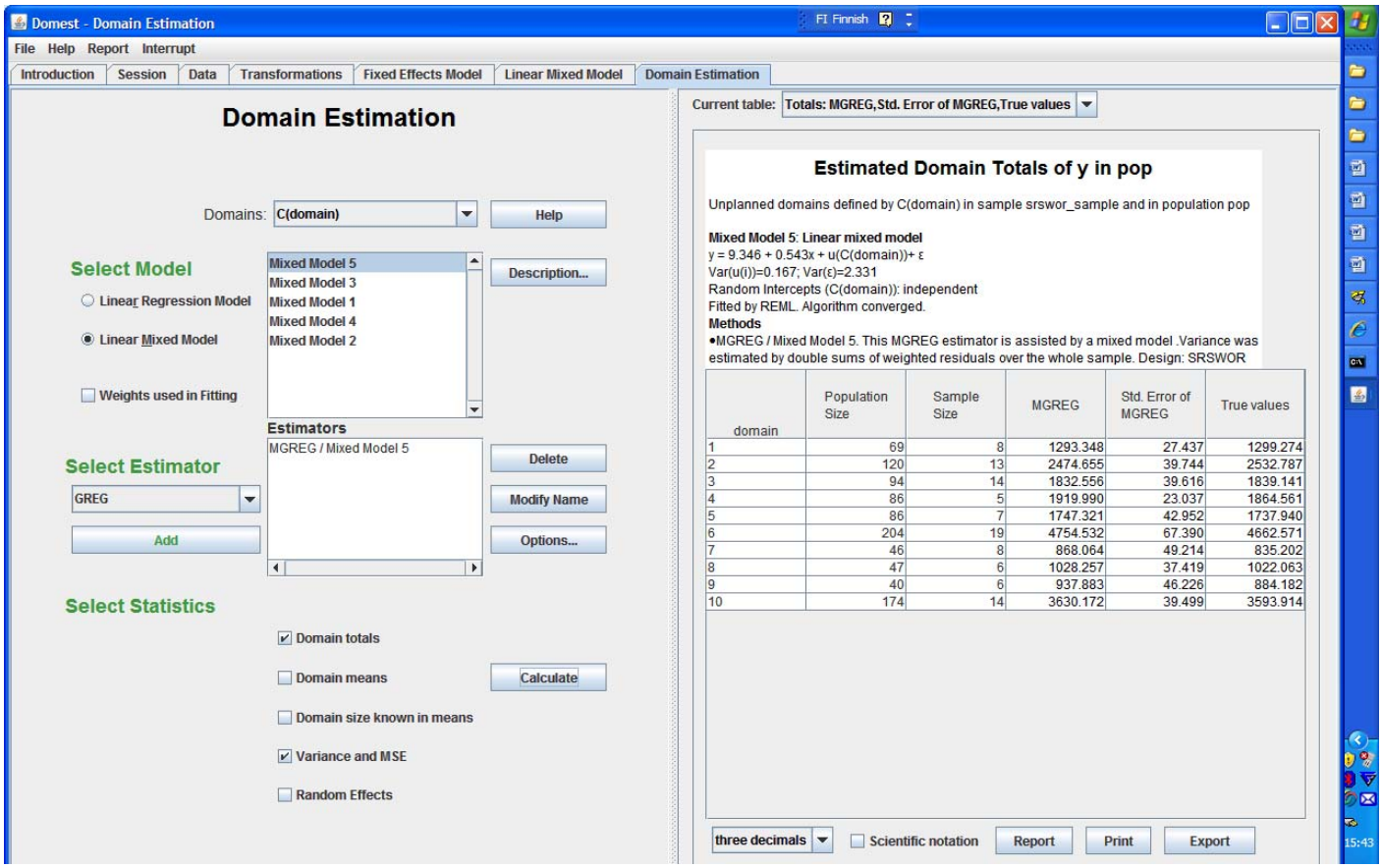
```
SAS
The Mixed Procedure

 Covariance Parameter Estimates

Cov Parm      Subject    Estimate

Intercept     domain       0.1674
Residual                   2.3307


                  Solution for Fixed Effects

                          Standard
Effect      Estimate        Error     DF    t Value

Intercept     9.3456       0.6972      9      13.40
x             0.5428      0.03120     89      17.40
```

2. Estimation for unplanned domains

2.1. SRSWOR_sample

2.1.10. MGREG estimation (Mixed-model assisted GREG) OPTION: Ordinary variance estimator (Formulas (8), (9) and (10))

Assisting model: Linear mixed model $Y_k = \beta_0 + u_d + \beta_1 x_k + \varepsilon_k$



Estimated Domain Totals of y in pop

Unplanned domains defined by C(domain) in sample srswor_sample and in population pop

Mixed Model 5: Linear mixed model

y = 9.346 + 0.543x + u(C(domain))+ ε

Var(u(i))=0.167; Var(ε)=2.331

Random Intercepts (C(domain)): independent

Fitted by REML. Algorithm converged.

Methods

•MGREG / Mixed Model 5. This MGREG estimator is assisted by a mixed model .Variance was estimated by double sums of weighted residuals over the whole sample. Design: SRSWOR

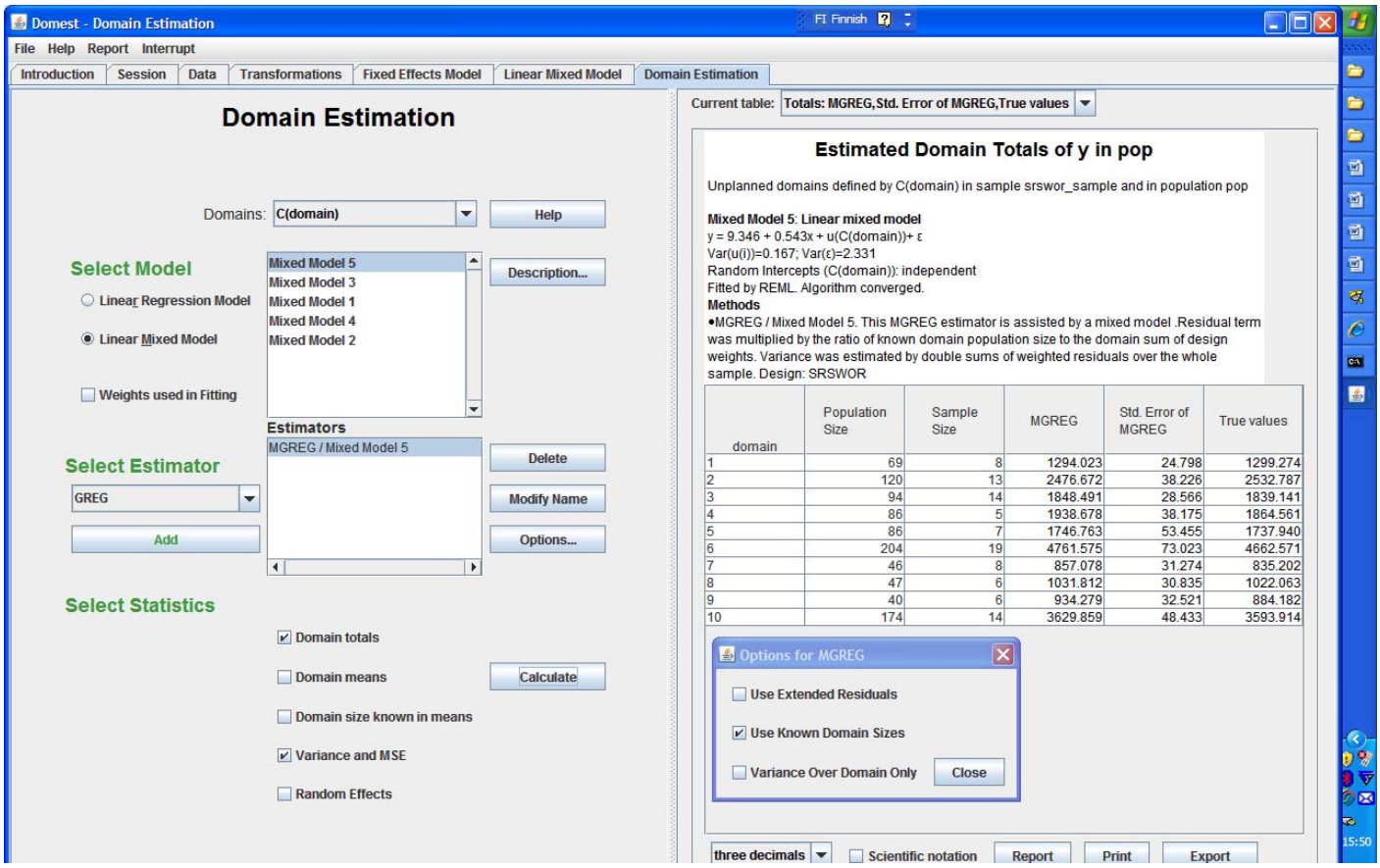
domain	Population Size	Sample Size	MGREG	Std. Error of MGREG	True values
1	69	8	1293.348	27.437	1299.274
2	120	13	2474.655	39.744	2532.787
3	94	14	1832.556	39.616	1839.141
4	86	5	1919.990	23.037	1864.561
5	86	7	1747.321	42.952	1737.940
6	204	19	4754.532	67.390	4662.571
7	46	8	868.064	49.214	835.202
8	47	6	1028.257	37.419	1022.063
9	40	6	937.883	46.226	884.182
10	174	14	3630.172	39.499	3593.914

2. Estimation for unplanned domains

2.1. SRSWOR_sample

2.1.11. MGREG estimation OPTION: Known Domain Sizes (Formulas (10) and (11))

Assisting model: Linear mixed model $Y_k = \beta_0 + u_d + \beta_1 x_k + \varepsilon_k$



Estimated Domain Totals of y in pop

Unplanned domains defined by C(domain) in sample srswor_sample and in population pop

Mixed Model 5: Linear mixed model

y = 9.346 + 0.543x + u(C(domain))+ ε

Var(u(i))=0.167; Var(ε)=2.331

Random Intercepts (C(domain)): independent

Fitted by REML. Algorithm converged.

Methods

•MGREG / Mixed Model 5. This MGREG estimator is assisted by a mixed model .Residual term was multiplied by the ratio of known domain population size to the domain sum of design weights. Variance was estimated by double sums of weighted residuals over the whole sample. Design: SRSWOR

domain	Population Size	Sample Size	MGREG	Std. Error of MGREG	True values
1	69	8	1294.023	24.798	1299.274
2	120	13	2476.672	38.226	2532.787
3	94	14	1848.491	28.566	1839.141
4	86	5	1938.678	38.175	1864.561
5	86	7	1746.763	53.455	1737.940
6	204	19	4761.575	73.023	4662.571
7	46	8	857.078	31.274	835.202
8	47	6	1031.812	30.835	1022.063
9	40	6	934.279	32.521	884.182
10	174	14	3629.859	48.433	3593.914

2. Estimation for unplanned domains

2.1. SRSWOR_sample
2.1.12. EBLUP estimation (Model based)
Assisting model: Linear mixed model $Y_k = \beta_0 + u_d + \beta_1 x_k + \varepsilon_k$

SAS Macro EBLUPGREG
EURAREA Project
EBLUP (Empirical Best Linear Unbiased Predictor)
REML estimation for the mixed model

```
%eblupgreg(sample=SRSWOR_sample,
         population=pop,
         weights=SamplingWeight,
         y=y,
         xlist=x,
         regionIdentifier=domain,
         modules=modules.eurarea,
         estimatemeans=0,
         eblup=1,
         greg=0,
         synthetic=0,
         test=1,
         output=EBLUP1);
         run;
```

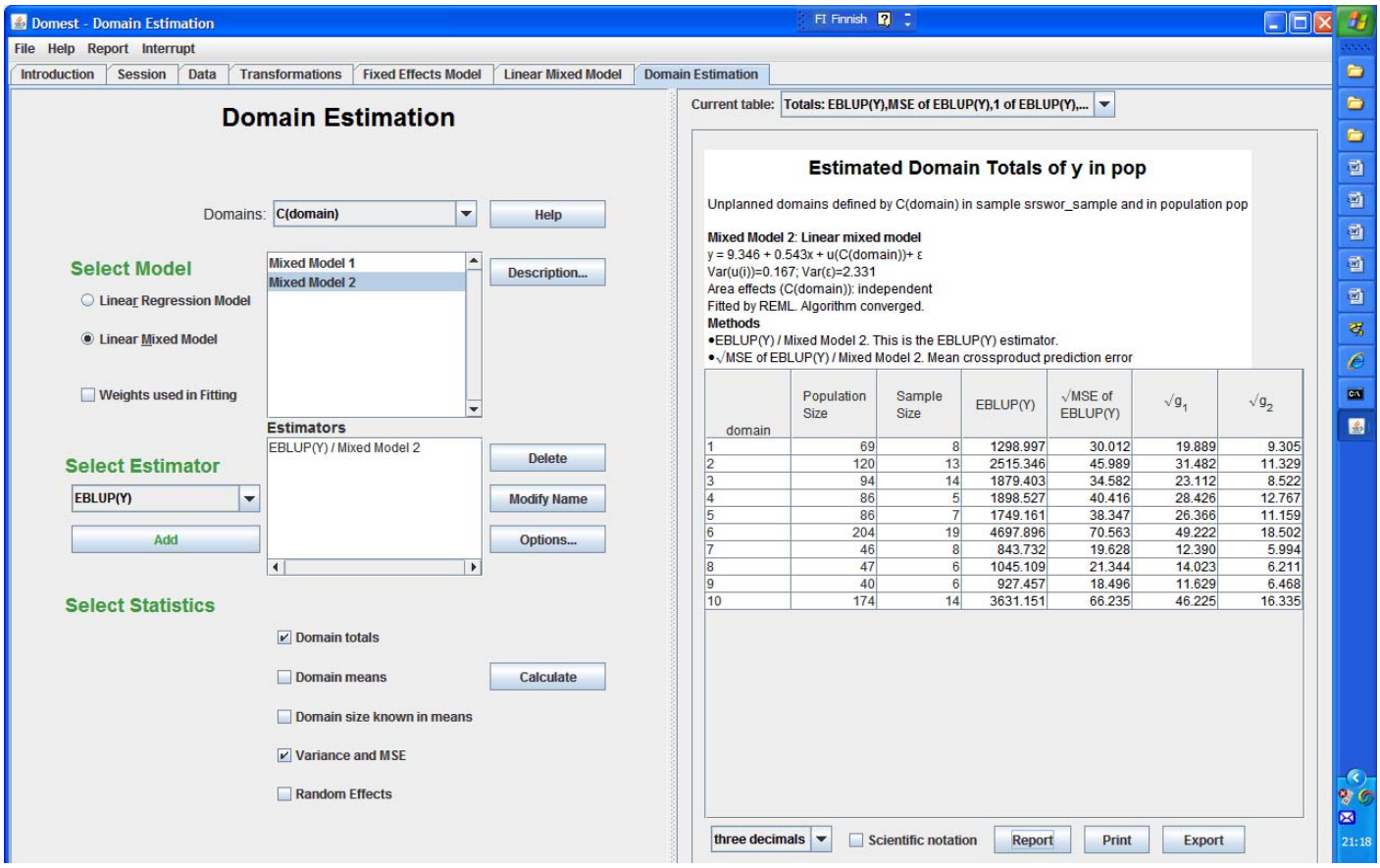
Obs	region	EBLUP	areaEffect	effectStdErr	sqrtMSE	sqrtMSENoG3	sqrtg1	sqrtg2	sqrtg3	sqrtg4	trueValue
1	1	1299.00	-0.04686	0.33689	30.0130	24.9879	19.8910	9.3044	11.7555	11.9235	1299.27
2	2	2515.34	-0.33751	0.31109	45.9891	36.9994	31.4842	11.3279	19.3138	15.7918	2532.79
3	3	1879.40	-0.38857	0.30664	34.5817	28.1660	23.1142	8.5210	14.1875	13.6548	1839.14
4	4	1898.53	0.17804	0.35523	40.4188	34.0590	28.4298	12.7663	15.3893	13.7399	1864.56
5	5	1749.16	-0.01527	0.34074	38.3482	31.6855	26.3690	11.1589	15.2744	13.5692	1737.94
6	6	4697.91	0.46976	0.29429	70.5611	56.5378	49.2251	18.4997	29.8525	20.7647	4662.57
7	7	843.73	0.20184	0.34170	19.6287	16.6742	12.3911	5.9936	7.3231	9.4109	835.20
8	8	1045.11	-0.13979	0.34905	21.3446	18.1882	14.0245	6.2103	7.8987	9.7753	1022.06
9	9	927.46	0.08648	0.34778	18.4970	16.0104	11.6301	6.4676	6.5501	8.9018	884.18
10	10	3631.15	-0.00812	0.30666	66.2346	52.6948	46.2283	16.3334	28.3749	19.3108	3593.91

2. Estimation for unplanned domains

2.1. SRSWOR_sample

2.1.13. EBLUP estimation (Formula (15))

Assisting model: Linear mixed model $Y_k = \beta_0 + u_d + \beta_1 x_k + \varepsilon_k$



Estimated Domain Totals of y in pop

Unplanned domains defined by C(domain) in sample srswor_sample and in population pop

Mixed Model 2: Linear mixed model

y = 9.346 + 0.543x + u(C(domain))+ ε

Var(u(i))=0.167; Var(ε)=2.331

Area effects (C(domain)): independent

Fitted by REML. Algorithm converged.

Methods

•EBLUP(Y) / Mixed Model 2. This is the EBLUP(Y) estimator.

•√MSE of EBLUP(Y) / Mixed Model 2. Mean crossproduct prediction error

domain	Population Size	Sample Size	EBLUP(Y)	√MSE of EBLUP(Y)	$\sqrt{g_1}$	$\sqrt{g_2}$	$\sqrt{g_3}$	$\sqrt{g_4}$	True values
1	69	8	1298.997	30.012	19.889	9.305	11.756	11.924	1299.274
2	120	13	2515.346	45.989	31.482	11.329	19.316	15.792	2532.787
3	94	14	1879.403	34.582	23.112	8.522	14.189	13.655	1839.141
4	86	5	1898.527	40.416	28.426	12.767	15.389	13.740	1864.561
5	86	7	1749.161	38.347	26.366	11.159	15.275	13.569	1737.940
6	204	19	4697.896	70.563	49.222	18.502	29.857	20.765	4662.571
7	46	8	843.732	19.628	12.390	5.994	7.323	9.411	835.202
8	47	6	1045.109	21.344	14.023	6.211	7.899	9.775	1022.063
9	40	6	927.457	18.496	11.629	6.468	6.550	8.902	884.182
10	174	14	3631.151	66.235	46.225	16.335	28.378	19.311	3593.914

References

Lehtonen R. and Veijanen A. (2009). Design-based methods of estimation for domains and small areas. Chapter 31 in Rao C.R. and Pfeffermann D. (Eds.). *Handbook of Statistics. Sample Surveys: Inference and Analysis. Vol. 29B.* New York: Elsevier.

Saei, A. and Chambers R. (2004). Small area estimation under linear and generalized linear mixed models with time and area effects. EURAREA Consortium 2004, *Project Reference Volume.* www.statistics.gov.uk/eurarea

Veijanen, A. and Lehtonen, R. (2010). Program Domest for estimation for domains and small areas. Part 1: Technical documentation. Version 1.0 (August 2010).