BIRO Academy

Distributed Statistical Analysis Software

Fabrizio Carinci, Luca Rossi

Serectrix, Italy

EUBIROD

Fragmented Analysis

Box 3.4.2. Output Logistic Model on all observations

The LOGISTIC Procedure
Model Information

Data Set	WORK._MODEL_
Response Variable	HI_HBA
Number of Response Levels	2
Number of Observations	17102
Model	binary logit
Optimization Technique	Fisher's scoring

Response Profile

Ordered Value	HI_HBA	Total Frequency
1	1	4856
2	0	12246

Probability modeled is HI_HBA=1.

Analysis of Maximum Likelihood Estimates

Standard Parameter	Wald DF	Estimate	Error	Chi-Square	Pr > ChiSq
Intercept	1	-0.6862	0.1028	44.5243	<.0001
GENDER	1	-0.2297	0.0343	44.7555	<.0001
CL_AGE2	1	0.0916	0.1092	0.7027	0.4019
CL_AGE3	1	-0.1465	0.1040	1.9842	0.1589
CL_AGE4	1	-0.2491	0.1086	5.2637	0.0218

Box 3.4.3. Output Logistic Model on aggregate data

The LOGISTIC Procedure
Model Information

Data Set	WORK.IN_SEDIS
Response Variable	HI_HBA
Number of Response Levels	2
Number of Observations	16
Weight Variable	COUNT
Sum of Weights	17102
Model	binary logit
Optimization Technique	Fisher's scoring

Response Profile

Ordered Value	HI_HBA	Total Weight	Total Frequency
1	1	8	4856.000
2	0	8	12246.000

Probability modeled is HI_HBA=1.

Analysis of Maximum Likelihood Estimates

Standard Parameter	Wald DF	Estimate	Error	Chi-Square	Pr > ChiSq
Intercept	1	-0.6862	0.1028	44.5243	<.0001
GENDER	1	0.2297	0.0343	44.7555	<.0001
CL_AGE2	1	0.0916	0.1092	0.7027	0.4019
CL_AGE3	1	-0.1465	0.1040	1.9842	0.1589
CL_AGE4	1	-0.2491	0.1086	5.2637	0.0218

Box 3.4.4. Observed/expected rates by centre using logistic regression

Centre	Den.	Num.	%Observed	% Expected	95% Lower	95% Upper
1	7699	2189	28.4	28.5	27.5	29.5
2	2360	1000	42.4	28.0	26.1	29.8
3	3422	916	26.8	28.4	26.9	29.9
4	1239	222	17.9	28.3	25.8	30.8
5	2382	529	22.2	28.4	26.6	30.2

Statistical Object

An element of a distributed information system that carries essential data in the form of embedded, partial aggregate components, required to compute a summary measure or relevant parameter for the whole population from multiple sites

Statistical Objects Meta-Data

Code	*Sequential code based on the taxonomy of the statistical objects dictionary*
Statistical Object	Name of the statistical object
Description	Short description of the principal content and output of the statistical object, ar
Variables	Type of variables (categorical, continuous)
Properties	Mathematical and statistical properties in a distributed data environment
Local Component	**OUTPUT OF THE LOCAL STATISTICAL ENGINE** Technical characteristics of the statistical object that is produced from each da Data section includes details on the format of the csv output.
Cumulative Component	**CUMULATIVE DATASET PROCESSED BY CENTRAL ENGINE** Technical characteristics of the procedure implemented in the central engine t Data section includes details on the format of the csv output.
Output	**STATISTICAL OUTPUT OF THE CENTRAL ENGINE** Includes the list of components that will be computed and stored in the statisti Defined codes are attributed to the list of electronic elements (e.g. XML tags,

Statistical Objects Dictionary

SECTION 1. FREQUENCY TABLES
Univariate Frequency Distribution, Outliers, Contingency Table

SECTION 2. MEASURES OF LOCATION
Arithmetic Mean, Percentile, Range

SECTION 3. MEASURES OF DISPERSION
Variance, Interquartile Distance

SECTION 4. GRAPHICAL ELEMENTS
Barplots, Histograms, Partial boxplots, Overall boxplots, Line plots, XY Plots, Webplots, Maps, Forest plot

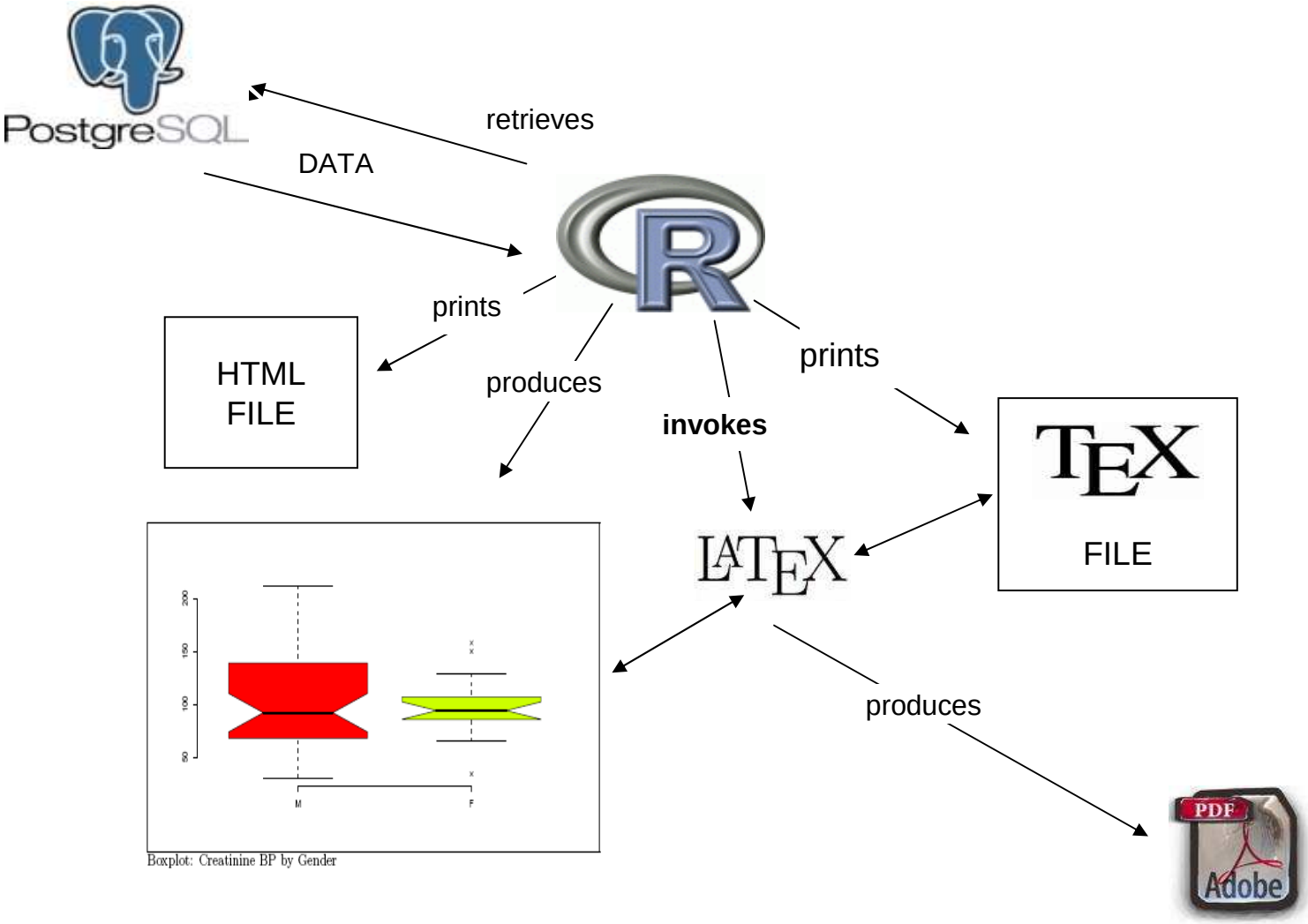
SECTION 5. REGRESSION
Linear regression,Logistic regression,Meta-analysis

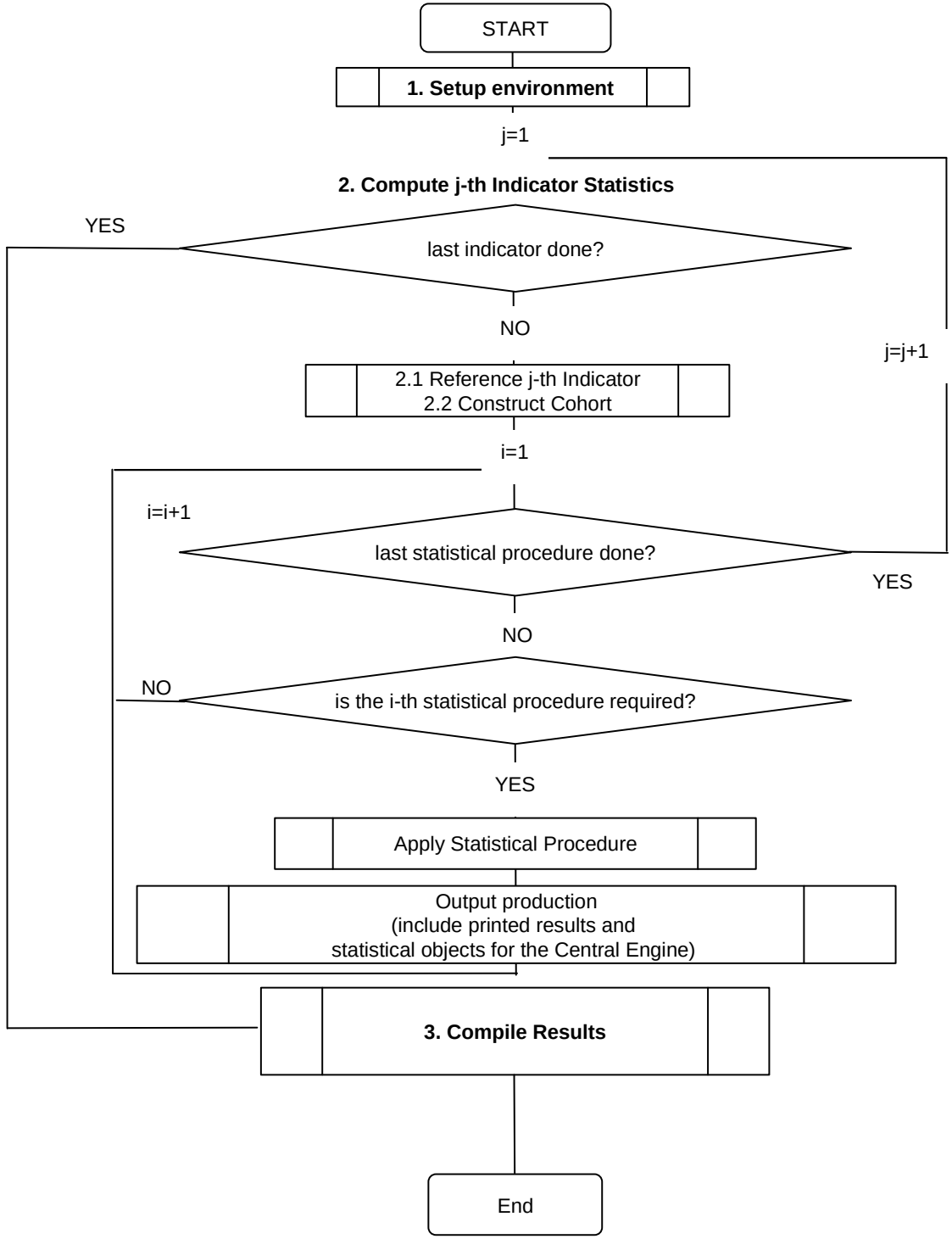
SECTION 6. STANDARDIZATION
Standardized rate, O-E

Arithmetic mean

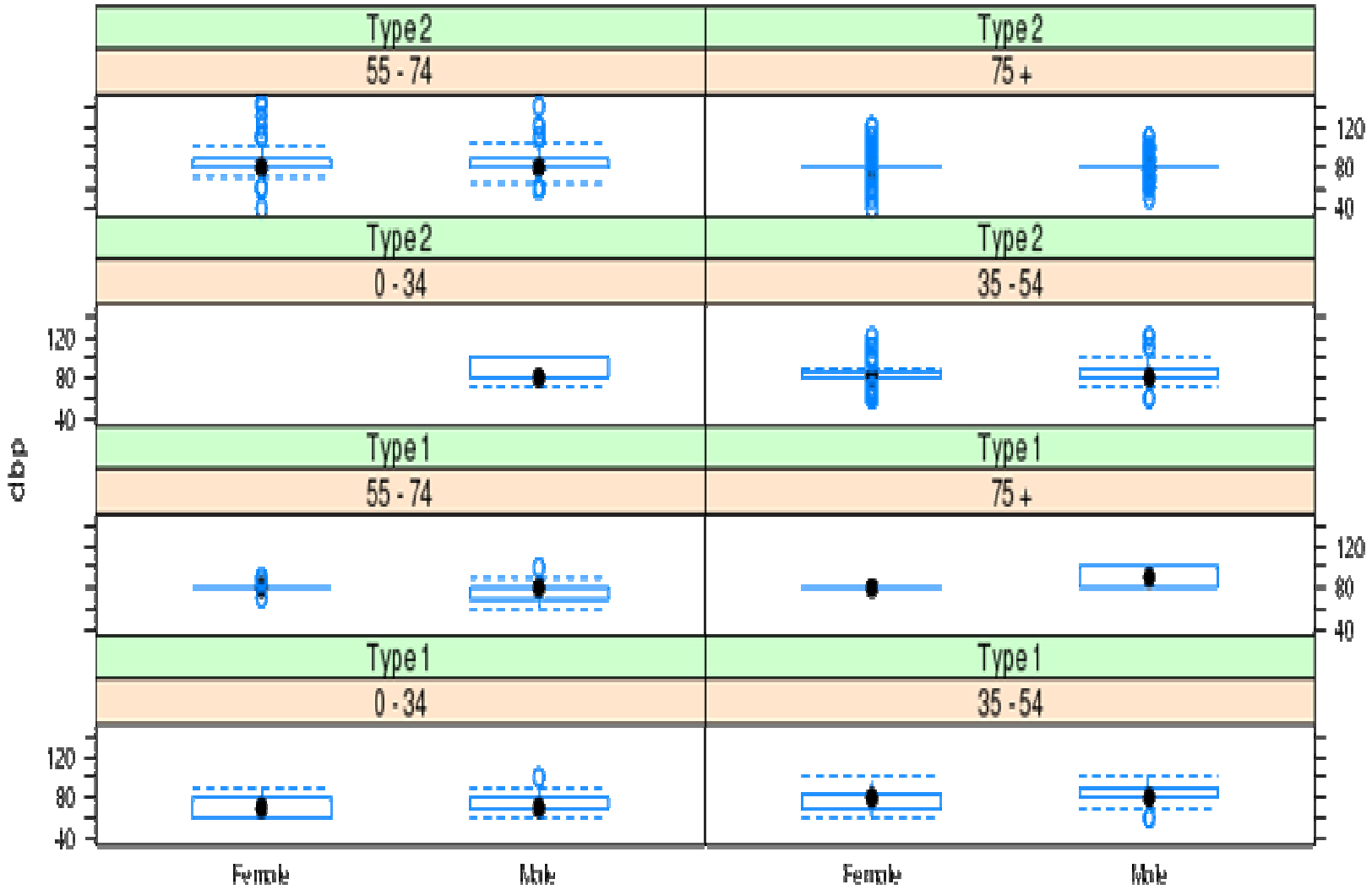
Code	2.1
Statistical Object	Arithmetic Mean
Description	Weighted average of a single characteristic, with weights equal to the number
Variables	CONTINUOUS
Properties	The mean of the overall sample is equal to the weighted mean of the arithmeti
Local Component	Data vector composed of two quantities: sum of the values of the target variab DATA: **<2.1.a>id, date, stratum, sum_x, n**
Cumulative Component	Sum of the sum of values from each local object DATA: **<2.1.a> id, date, stratum, sum_x, n**
Output	Single value of the overall arithmetic mean: cumulative object, divided by the s DATA: **<2.1.a>mean** Single value of the arithmetic mean by centre: cumulative object, divided by th DATA: **<2.1.b>id, date, stratum, mean**

BIRO Statistical Engine Design

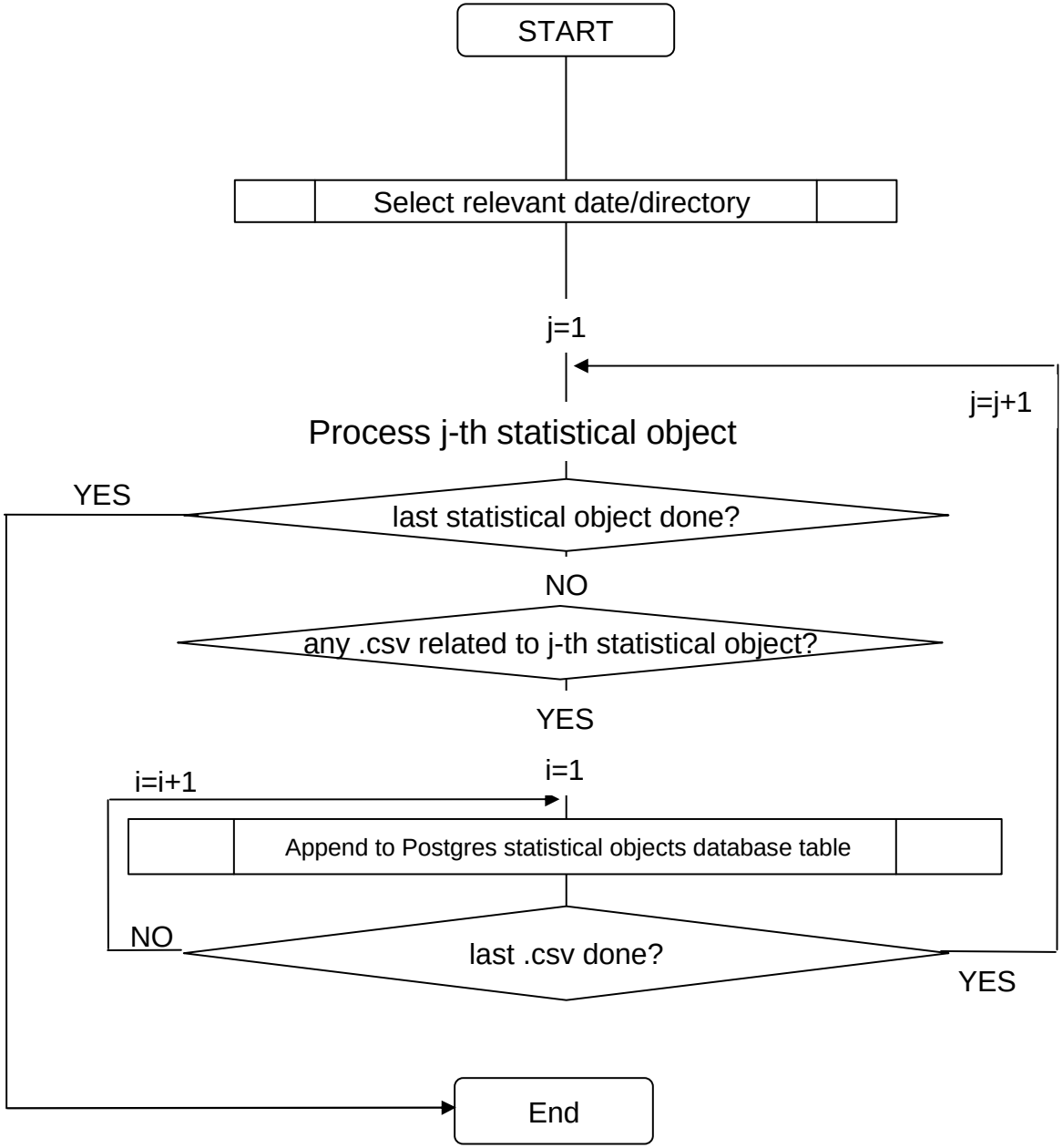



START
1. Setup environment
j=1
2. Compute j-th Indicator Statistics
YES
last indicator done?
NO
j=j+1
2.1 Reference j-th Indicator
2.2 Construct Cohort
i=1
i=i+1
last statistical procedure done?
YES
NO
NO
is the i-th statistical procedure required?
YES
Apply Statistical Procedure
Output production
(include printed results and
statistical objects for the Central Engine)
3. Compile Results
End

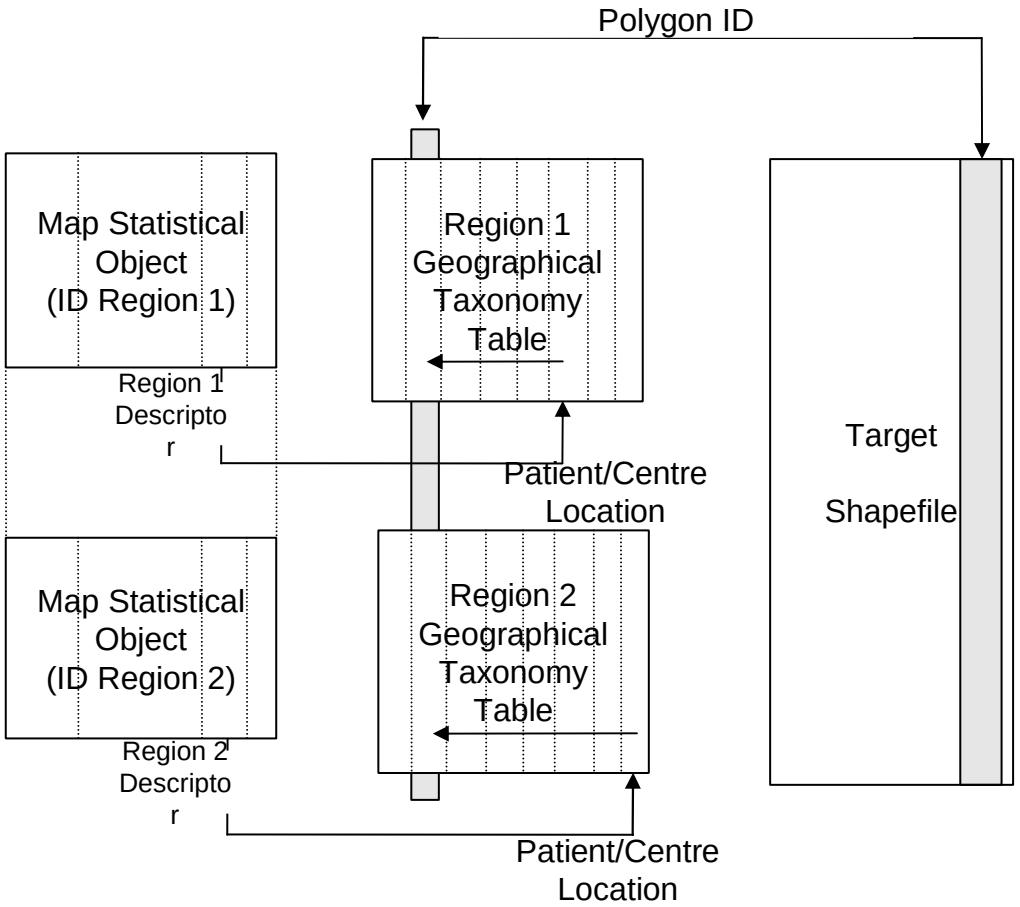
Graphics


Type 2
55 - 74
Type 2
75 +
Type 2
0 - 34
Type 2
35 - 54
Type 1
55 - 74
Type 1
75 +
Type 1
0 - 34
Type 1
35 - 54
dbp
120
80
40
Female
Male

Pile-up


START
Select relevant date/directory
j=1
j=j+1
Process j-th statistical object
YES
last statistical object done?
NO
any .csv related to j-th statistical object?
YES
i=1
i=i+1
Append to Postgres statistical objects database table
NO
last .csv done?
YES
End

Maps Central Engine



Linking geographical references to a target shapefile in BIRO central engine

Performance

LOCAL

Centre	N Patients	N episodes	Elapsed Time
1	17,552	92,237	24' 25"
2	5,315	19,434	7' 01"
3	7,846	60,274	12' 20"
4	7,827	45,345	10' 51"
5	5,008	10,994	5' 22"

GLOBAL

Centre	N Patients	N episodes	Elapsed Time
1	17,552	92,237	20' 12”
1+2	22,867	111,671	20' 54"
1+2+3	30,713	217,290	21' 33"
1+2+3+4	38,540	262,635	21' 56"
1+2+3+4+5	43,548	273,629	22' 27"