How to use the SAS program for the identification of Industrial Districts

Silvia Lombardi

ISTAT - Italian National Institute of Statistics

lombardi@istat.it

Istat

Outline

1. The ISTAT mapping methodology of industrial districts: the theory

    i.   Classification variables

    ii.   The algorithm

2. The application of the algorithm

    i.   Creation of the input dataset

    ii.   SAS code: outlook

    iii.   Execute SAS program

Istat

1. The ISTAT mapping methodology of IDs

The methodology is based on the application of **location quotient (LQ)** within a hierarchical procedure

It relies on **employment data at the local unit level** provided by 2011 Business Census (count variable)

Economic units involved are: **enterprises, not for profit institutions and public institutions**

Persons employed within all economic units are taken into account because they represent all the economic agents involved into LMAs

Employment data classified on the basis of structural variables is calculated **by LMA**

Istat

i) Classification variables

Two sets of variables involved: the **economic activity** and **enterprise class size**

As a proxy of production *filiére* the algorithm utilizes aggregations of the Statistical Classification of Economic Activities in the European Community, NACE Rev.2 (2007)

Economic activities (NACE - 5 digit) are grouped into **8** types of economic activities (**Table 1**) and, among manufacturing activities, in **11** manufacturing industries (**Table 2**).

The economic activity (1/2)

All economic activities (NACE Rev.2) are grouped into the following 8 typologies:

- ➢ Agricultural activities
- ➢ Extractive industry
- ➢ Construction
- ➢ Manufacturing
- ➢ Business services
- ➢ Consumer services
- ➢ Social services
- ➢ Traditional services

Example: Electricity, gas, steam and air conditioning supply, Water collection, treatment and supply, Wholesale and retail trade and repair of motor vehicles and motorcycles, Wholesale of grain, unmanufactured tobacco, seeds and animal feeds, Wholesale of grain, unmanufactured tobacco, seeds and animal feeds, Wholesale of flowers and plants, Wholesale of live animals, … , Water transport, Air transport, Postal and courier activities, …. , Telecommunications, Repair of consumer electronics, Repair of footwear and leather goods, Repair of watches, clocks and jewellery

Istat

Table 1 - ECONOMIC ACTIVITIES *Variable T1*

TYPES OF PRODUCTION ACTIVITIES	NACE Rev. 2 code
Agri-industrial activities	01, 02, 03, 81300
Extractive industry	05 – 09
Manufacturing	10 - 33, 383, 581, 59201, 59202, 95120, 95220, 95240, 95290
Construction	412, 42, 43
Business services	461, 4624, 4651, 4661 - 4666, 46692, 46693, 494, 521, 5224, 5229, 582, 62, 631, 6391, 64, 6611, 6612, 66192, 66194, 662, 663, 69 - 73, 741, 743, 749, 7712, 773, 78, 80, 8121, 8122, 81291, 821, 823, 82912, 89921, 89922, 82994, 82999, 941, 95110
Consumer services	411, 55, 56, 591, 59203, 60, 68, 742, 7711, 772, 79, 811, 90, 92, 932, 96 - 99
Social services	37, 381, 382, 39, 65,66191, 66193, 66195, 75, 81299, 85 - - 88, 91, 931, 949
Traditional services	35, 36, 45, 4621 - 4623, 463, 464, 4652, 46691, 46699, 467, 469, 47, 491, 492, 493, 495, 50, 51, 5221, 5222, 5223, 53, 61, 6399, 774, 822, 82911, 82991 - 82993, 84, 942, 95210, 95230, 95250

Istat

The economic activity (2/2)

Manufacturing activities are then grouped into 11 categories:

- ➢ Textile and clothing;

- ➢ Leather and footwear;

- ➢ Households goods (wooden furniture, tiles and other glass and ceramics items);

- ➢ Jewellery, musical instruments and toys;

- ➢ Food and beverages;

- ➢ Machinery, electrical and optical equipment;

- ➢ Manufacture of basic metals and fabricated metal products; chemicals and plastics;

- ➢ Transport equipment; paper, publishing and printing;

- ➢ Other manufacturing activities.

Istat

Table 2 – MANUFACTURING INDUSTRIES   *Variable T2*

MANUFACTURING INDUSTRY	Nace Rev. 2 code
Textile and clothing	13, 14
Leather and footwear	15
Households goods	16, 23, 31, 3291, 32994, 9524, 9529
Jewellery, musical instruments and toys	264, 3211,3212, 322 - 324
Food and beverages	10, 11, 12
Machinery	182, 2453, 2454, 25, 261 -- 263, 265 -- 267, 2711, 2712, 2720, 2731, 2732, 274, 275, 279, 28, 29310, 304, 325, 3311 -- 3314, 332, 9512, 9522
Manufacture of basic metals and fabricated metal products	241 -- 243, 2441 -- 2445, 2451, 2452
Chemicals and plastics	19, 201 -- 204, 2052 -- 2060, 21, 22, 2446, 268, 2733, 32991
Transport equipment	291, 292, 29320, 301 -- 303, 30911, 30912, 30921, 30922, 30923, 30990, 3315 -- 3317, 38312
Paper, publishing and printing	17, 181, 581, 59201, 59202
Other manufacturing activities	20510, 30924, 3213, 32992, 32993, 32999, 3319, 38311, 3832

Istat

The Enterprise size class

➢ Four sizes - European classification   *Variable f_size*

- **Micro enterprises**: with less than 10 persons employed;
- **Small enterprises**: with 10-49 persons employed;
- **Medium-sized enterprises**: with 50-249 persons employed;
- **Large enterprises**: with 250 or more persons employed.

➡ **MSMEs**

Small and medium sized enterprises (SMEs), that is, with three classes was the old methodology (up to 2001).

ii) The algorithm*

The algorithm is composed of a hierarchical **four steps** procedure which determines the dominant industry and the presence of MSMEs

Statistical measures: LQ and employment prevalence indicators at the LMA level

It identifies LMAs specialized in manufacturing based on MSMEs and a dominant manufacturing specialization mainly based on MSMEs

* for Italian version, see:https://www.istat.it/it/archivio/150320

STEP 1: identification of LMAs specialized in manufacturing (among the 8 macro categories of all economic sectors – variable"T1")

❖ **1.a)** For each LMA the algorithm calculates the LQ on T1 by utilizing the following formula:

1.A $LQ_m = (LMA_{emp, T1} / ITA_{emp, T1}) / (LMA_{emp, tot} / ITA_{emp, tot})$ [1]

n1: Persons employed by LMA, economic activity
d1: Persons employed in ITALY, by economic activity
n2: Persons employed by LMA, ALL economic activities
d2: Persons employed in ITALY, ALL economic activities

LMAs with **LQ > 1** are further investigated

Istat

❖ **1.b)** For LMAs with **LQ > 1** in Manufacturing, Business services or Consumer services are compared by using the prevailing employment value in order to detect manufacturing LMAs

$$[(LMA_{emp,\ T1} / ITA_{emp,\ T1}) - (LMA_{emp,\ tot} / ITA_{emp,\ tot})] * ITA_{emp,\ T1} \quad [2]$$

Specialization index is multiplied by Persons employed in ITALY by economic activity

LMAs with highest value of manufacturing employment in [2] are defined as a manufacturing LMA and go to Step 2

In 2011, **240** manufacturing LMAs have been identified

STEP 2: identification of manufacturing LMAs featured by MSMEs

The algorithm searches for LMAs where there is a concentration of employment in manufacturing micro, small and medium sized enterprises: it calculates the LQ on enterprise size class by utilizing the following formula:

$$(LMA_{emp\ (class),\ manif} / ITA_{emp\ (class),\ manif}) / (LMA_{emp,\ manif} / ITA_{emp,\ manif})\ [3]$$

n1: Persons employed in manufacturing by LMA, class size

d1: Persons employed in manufacturing in ITALY, class size

n2: Persons employed in manufacturing by LMA, ALL class sizes

d2: Persons employed in manufacturing in ITALY, ALL class sizes

LMAs which show **LQ>1** in MICRO, SMALL, MEDIUM enterprise size class are MSMEs-based manufacturing LMAs and go to next step. 151 manufacturing MSMEs LMAs are identified in 2011

Istat

STEP 3: identification of *dominant industry* of manufacturing SMEs-based LMAs (among the 11 manufacturing categories of Table 2 or "T2")

❖ **3.a)** MSMEs-based manufacturing LMAs are further investigated: for each LMA the algorithm calculates the LQ on T2 by utilizing the following formula:

$$(LMA_{emp,\ T2} / ITA_{emp,\ T2}) / (LMA_{emp,\ Tot\ man} / ITA_{emp,Tot\ man}) \qquad [4]$$

n1: Persons employed by LMA, T2

d1: Persons employed in ITALY, T2

n2: Persons employed by LMA, ALL T2(= total manufacturing)

d2: Persons employed in ITALY, ALL T2 (=total manufacturing)

❖ **3.b)** Define the dominant manufacturing industry of LMA

For LMAs which show **LQ>1** in one or more modalities of T2 the employment prevalence is calculated in order to detect which modality presents the highest value.

$$[(LMA_{emp,\ T2} / ITA_{emp,\ T2}) - (LMA_{emp,\ Tot\ man} / ITA_{emp,\ Tot\ man})] * ITA_{emp,\ T2} \quad [5]$$

The correspondent manufacturing activity is defined as the dominant industry (dom_ind) of the LMA

STEP 4: identification of LMAs based on MSMEs where the dominant industry is featured by MSMEs

- ❖ **4.a)** The MSMEs based manufacturing LMA is defined as an ID if its dominant industry is featured by MSMEs, that is, if both the following conditions hold:

$$(LMA_{emp\ (mpmi),\ dom\_ind} / LMA_{emp\ (tot),\ dom\_ind}) > 50,0\% \qquad [6]$$

n1: Persons employed in dominant industry in MSMEs

d1: Persons employed in dominant industry in ALL enterprises

Istat

In case of LMAs with only one medium enterprise, the latter has not to concentrate employment:

❖ **4b. 1)**

$$(\text{LMA}_{emp\,(small),\,dom\_ind} / \text{LMA}_{emp\,(med),\,dom\_ind}) > 50,0\% \qquad [7]$$

n1: Persons employed in dominant industry in SMALL enterprises
d1: Persons employed in dominant industry in MEDIUM enterprises

❖ **4b. 2)**

$$(\text{LMA}_{emp\,(micro),\,dom\_ind} / \text{LMA}_{emp\,(med),dom\_ind}) > 50,0\% \quad [8]$$

n1: Persons employed in dominant industry in MICRO enterprises
d1: Persons employed in dominant industry in MEDIUM enterprises

Istat

… To sum up

❖ Step 1
- Test 1a: calculate LQ on economic activities
- Test 1b: chooses the manufacturing LMAs by using prevailing employment value

❖ Step 2: calculate LQ on enterprise size class

❖ Step 3: assessing dominant industry
- Test 3a: LQ on manufacturing industries
- Test 3b: chooses the dominant industry of each LMAs by using prevailing employment value

❖ Step 4: final refinements
- Test 4a: The share of employment in dominant industry in MSMEs on total of employment in dominant industry must be higher than 50%
- Test 4b: case of only one medium enterprise

Istat

Flows and intermediate results
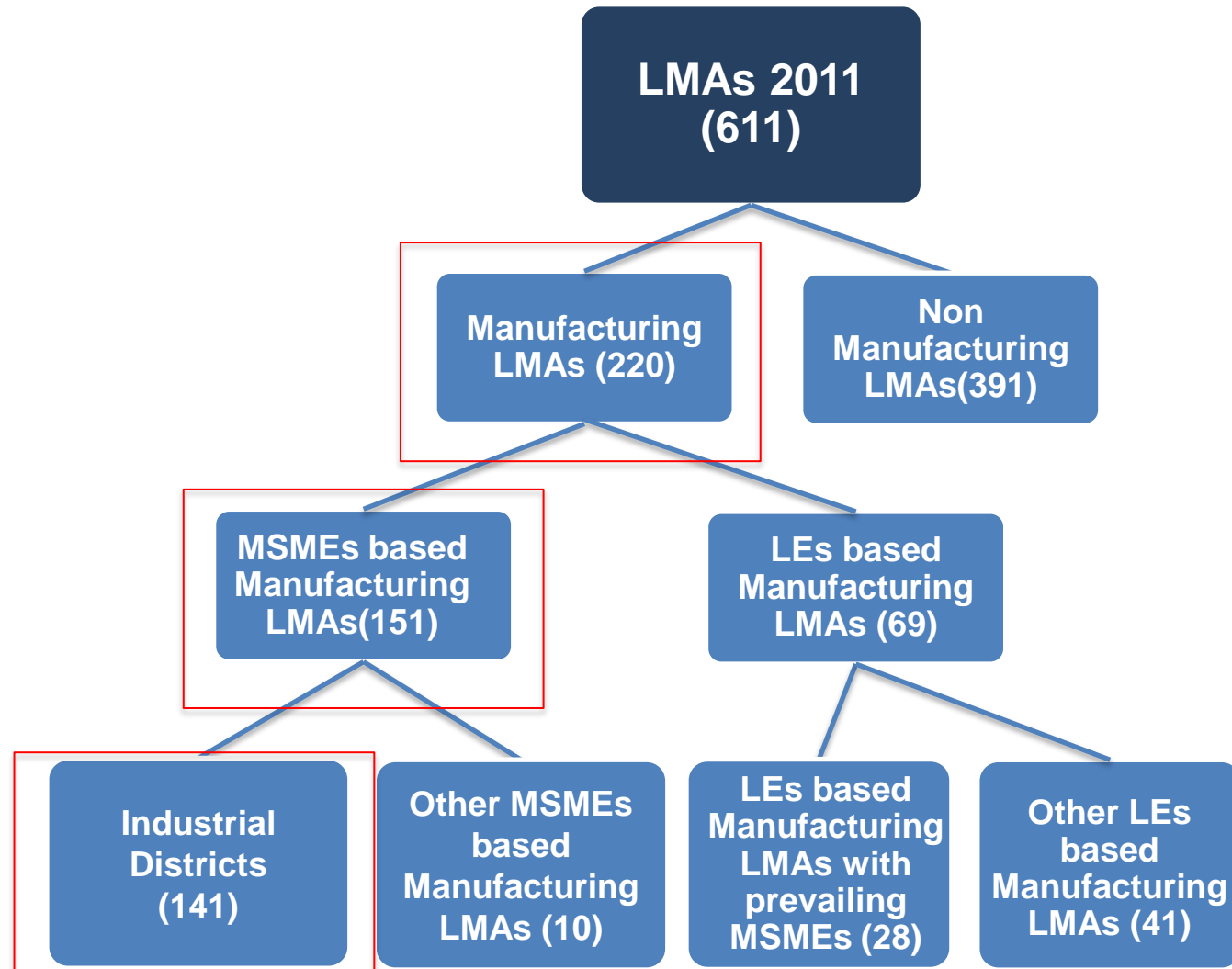
2. The application of the algorithm

i. Creation of the dataset – SAS code

ii. SAS code: outlook

iii. Execute SAS program

Istat

Creation of the dataset

CONSTRUCTION OF VARIABLES (1/2)

1. ECONOMIC ACTIVITY
If you have your own business census micro data, use the SAS dataset NACE.sas7bdat in the folder 'Industrial Districts' which contains the coding from Nace Rev. 2 (5 digit) classification into Table 1 and Table 2 aggregations.
Just merge it with your dataset before calculating the values by LMAs.

2. CLASS SIZE
Secondly, build the variable class size on your business census data.
3. Run PROC FORMAT for the three variables

Istat

```
proc format;

/* Table 1 */
value $t1f
    '01'='Agri-industrial activities'
    '02'='Extractive industry'
    '03'='Manufacturing'
    '04'='Construction'
    '05'='Business services'
    '06'='Consumer services'
    '07'='Social services'
    '08'='Traditional services';

/* Size class */
value f_size
    1='from 0 to 9 persons employed'
    2='from 10 to 49 persons employed'
    3='from 50 to 249 persons employed'
    4='250 persons employed or more';
```

Istat

```
/* Table 2 */

value $t2f
    '99'='Not manufacturing'
    '01'='Textile and clothing'
    '02'='Leather and footwear'
    '03'='Households goods'
    '04'='Jewellery, musical instruments and toys'
    '05'='Food and beverages'
    '06'='Machinery'
    '07'='Manufacture of basic metals and fabricated metal products'
    '08'='Chemicals and plastics'
    '09'='Transport equipment'
    '10'='Paper, publishing and printing'
    '11'='Other manufacturing activities';
run;
```

Istat

Generate the input dataset

❖ From microdata: aggregate micro data by the following variables:

LMA*, Table 1 (*T1*), Table 2 (*T2*), size class (*f_size*)

❖ Count variables are persons employed and establishments

Variable name: «emp»

Variable name: «l_units»

```
proc sql;
create table LMALIB.Business Census as
select  COD_SLL_11, T1, T2, f_size,
sum(persons_employed) as emp,
count (*) as l_units
from LMALIB..dw
group by COD_SLL_11, T1, T2, f_size
order by COD_SLL_11, T1, T2, f_size;
```

*LMA is variable
"COD_SLL_11"
in the Italian case

Istat

The input dataset

LMA_2011	Class size "f_size"	Economic activity "t1"	Manufacturing industry "t2"	Persons employed	Establishments
101	1	1	99	202	123
101	2	1	99	72	5
101	1	2	99	101	13
101	1	3	1	277	91
101	2	3	1	817	35
101	3	3	1	522	11
101	1	3	2	95	5
.....	.....	.....	.....	.....	.....
.....	.....	.....	.....	.....	.....
102	1	1	99	139	80
102	2	1	99	79	9
102	1	2	99	72	13
102	1	3	1	76	25
102	2	3	1	124	12
102	1	3	2	63	11
102	1	3	3	279	130
102	2	3	3	270	19
102	3	3	3	207	12

* Note: simulated data

Istat

The code of the algorithm

The SAS program is a set of nested queries which create a base of data and select LMAs according to selection criteria at each step. It is structured as follows:

❖ Initial part: loading the dataset and descriptive statistics

❖ Calculation of LQ by LMA within four steps

❖ Final part: visualization of results

Istat

Initial part (1/3)

❖ The .pgm assigns a SAS LIBRARY called "LMALIB" related to the folder "Industrial Districts"

❖ Macro *grabpath* which "grabs" the same path where the .pgm is located

❖ In this case, the .pgm and the datasets are in the same folder

```
%macro grabpath;

%qsubstr(%sysget(SAS_EXECFILEPATH),1,%length(%sysget(SAS_EXECFILEPATH))-%length(%sysget(SAS_EXECFILEname)))

%mend grabpath;
%let path = %grabpath;

LIBNAME LMALIB "&path";
```

Initial part (2/3)

Descriptive statistics of the dataset

Persons employed by T1 and T2 in order to check the initial values of numerators and denominators

```
ods html style=journal;
proc tabulate data =LMALIB.Business_census missing;
      class  t1 t2;
      var emp;
      table (t1='Economic activity - Table 1' all='TOTAL'),
      emp='Persons employed'*f=12.0
      *(t2='Manufacturing activity - Table 2' all='TOTAL');
      format t1 $t1f. t2  $t2f.;
      title 'Persons employed by  T1 and T2';
run;
ods html close;
```

Initial part (3/3)

Persons employed by T1 and size class in order to check the initial values of numerators and denominators

```
ods html style=journal;
proc tabulate data =LMALIB.Business_census missing;
    class  t1 t2 f_size;
    var emp;
    table (t1='Economic activity - Table 1' all='TOTAL'),
    emp='Persons employed'*f=12.0
    *(f_size='Size class' all='TOTAL');
    format t1 $t1f. t2  $t2f. f_size f_size.;
    title 'Persons employed by Economic activity and size class';
run;
ods html close;
```

Istat

Screenshot of Step 1a as example

```sas
******************
**** STEP 1 **** ;
******************
**** Identification of LMAs specialized in manufacturing (among 8 macro categories
of all economic sectors);

proc sql;

* Calculate LQ and check prevailing employment in business services (05),
consumer services (06), manufacturing (03);

create table BASE1 as
select NUM1.COD_SLL_11, NUM1.T1, n1,d1,n2,d2, n1/d1 as NUMERATOR, n2/d2 as DENOMINATOR,
(n1/d1) / (n2/d2) as LQ,
CASE
when NUM1.T1 in ('03','05','06') and (n1/d1) / (n2/d2) > 1 then ( (n1/d1) - (n2/d2) ) * d1
else 0
END
as prev
from
(select COD_SLL_11, T1, sum(emp) as n1 label 'Persons employed by LMA, economic activity'
from LMALIB.Business_census group by COD_SLL_11, T1) as NUM1
INNER JOIN (select T1, sum(emp) as d1 label 'Persons employed in ITALY, by economic activity'
from LMALIB.Business_census group by T1) as DEN1 on (NUM1.T1=DEN1.T1)
INNER JOIN (select COD_SLL_11, sum(emp) as n2 label 'Persons employed by LMA, ALL economic activities'
from LMALIB.Business_census group by COD_SLL_11) as NUM2 on (NUM1.COD_SLL_11=NUM2.COD_SLL_11)
CROSS JOIN (select sum(emp) as d2 label'Persons employed in ITALY, ALL economic activities'
from LMALIB.Business_census) as DEN2
order by COD_SLL_11, T1
;
```

Screenshot of Step 1b as example

```
* SELECTION CRITERIA
Assess the prevailing (manufacturing) industry by LMA by using the prevailing employment value;
create table TEST1 as select COD_SLL_11,T1,prev from BASE1
where T1='03' and cats(COD_SLL_11,prev) in
(
select cats(COD_SLL_11,prev) from
(select COD_SLL_11,max(prev) as prev from BASE1 where prev > 0 group by COD_SLL_11)
)
order by COD_SLL_11;


select count (distinct COD_SLL_11) from test1; * Finds 280 manufactiring LMAs;
```

And then again for step 2, and so on…

Istat

Visualization (1/3)

%let mapFile =SLL2011_wgs84_EURO.shp;   Import the shape file

```
proc mapimport out=LMALIB.shapesll2011 datafile="&path&mapFile";
run;


data DISTRICTS2011;
    set DISTRICTS2011;                          Prepare the dataset
        dummy_DI=1;
RUN;


proc sql;
create table SLL2011 as
select distinct (SLL_2011), SLL_2011_T, DEN_SLL_2011 from
LMALIB.sll2011 ;
create table data_map as select l.*, r.t2, r.dummy_DI from SLL2011 as l
left join DISTRICTS2011 as r
on l.SLL_2011=r.COD_SLL_11;
QUIT;
```

Istat

Visualization (2/3)

```
data data_map;
    set data_map;
        if dummy_DI=. then dummy_DI=0;
        if t2='' then t2='00';
run;


goptions reset=all cback=white gunit=pct htitle=2 htext=2 ftitle="arial"
ftext="arial" xpixels=900 ypixels=700 DEVICE=png;
legend1 label=none shape=bar(2,2) position=(center) across=2;
    pattern1 v=ms c=cxFFFFFF; *white;
    pattern2 v=ms c=CX662F14; *brown;
    pattern3 v=ms c=CX4D7EBF; *light blue;
    pattern4 v=ms c=CXCC2B1B; *red;
    pattern5 v=ms c=CX44A616; *light green;
    pattern6 v=ms c=CXE5B82E; *orange;
    pattern7 v=ms c=CXD974C9; *pink;
    pattern8 v=ms c=CX3230B2; *blue;
    pattern9 v=ms c=CXFFFF00; *yellow;
    pattern10 v=ms c=CXAEADD9; *very light violet;
    pattern11 v=ms c=CXBDB2BF; *grey;
```

Istat

Visualization (3/3)

```
proc gmap data=data_map map=LMALIB.shapesll2011;

id SLL_2011;
choro t2 /coutline=grey legend=legend1;
ods html file='DI2011 by industry.html';
title 'Industrial districts 2011 by industry';
label t2='Dominant industry';
format t2 $industry.;

run;
quit;
```

Istat

RECOMMANDATIONS

❖ BUILD YOUR SAS DATASET STRICTLY AS SUGGESTED

❖ USE THE FILE NACE.sas7bdat TO BUILD THE ECONOMIC
CLASSIFICATION VARIABLES

❖ CREATE A FOLDER WITH YOUR BUSINESS CENSUS
DATA, NACE CODES, SHAPEFILES AND SAS PROGRAM
ON YOUR DESKTOP

For further information: lombardi@istat.it

Istat