validarcae
Utility tool to deal with the Portuguese classification of economic activities (CAE)

Marta Silva

2020 Portuguese Stata Conference

Portuguese Classification of Economic Activities

- Framework to organize and classify statistical units producing goods and services

- Allows to present statistical information by economic activity

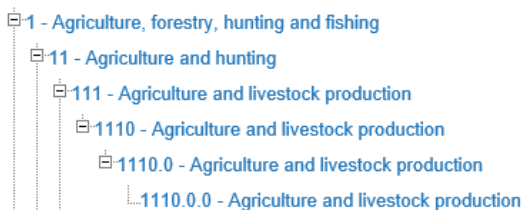| Level	Classification	
World Level	ISIC	United Nations' International Standard Industrial Classification of all Economic Activities
European Level	NACE	Statistical classification of economic activities in the European Communities
National Level	**CAE**	**Portuguese Classification of Economic Activities**

Source: Eurostat (2008)

CAE Revisions

- CAE suffered several revisions over time aiming the harmonization with European classification systems:

Revision	Period	1 L	2 L	1 D	2 D	3 D	4 D	5 D	6 D
1	1973 - 1993	NA	NA	NA	10	34	80	201	602
2	1994 - 2002	17	31	NA	60	222	503	715	NA
2.1	2003 - 2007	17	31	NA	62	224	515	719	NA
3	2008 -	21	NA	NA	88	272	616	850	NA

- The classification has an hierarchical structure and several levels of aggregation
  - The number and scope of the levels of aggregation changed with each revision

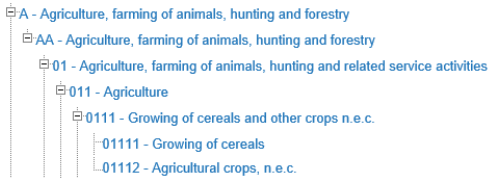# Portuguese Classification of Economic Activities - CAE Rev.1

- CAE Rev.1 contains 6 levels of aggregation:
  1. Division - represented by 1 digit
  2. Subdivision - represented by 2 digits
  3. Class - represented by 3 digits
  4. Group - represented by 4 digits
  5. Subgroup - represented by 5 digits
  6. Detail - represented by 6 digits

1 - Agriculture, forestry, hunting and fishing
  11 - Agriculture and hunting
    111 - Agriculture and livestock production
      1110 - Agriculture and livestock production
        1110.0 - Agriculture and livestock production
          1110.0.0 - Agriculture and livestock production

Source: Statistics Portugal

Portuguese Classification of Economic Activities - CAE Rev.2 and CAE Rev.2.1
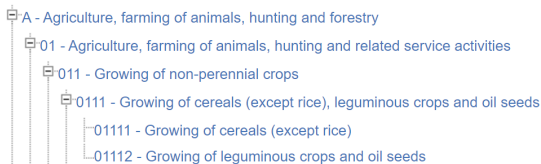
- CAE Rev.2 and CAE Rev.2.1 contain 6 levels of aggregation:
  1. Section - represented by a letter
  2. Subsection - represented by 2 letters
  3. Division - represented by 2 digits
  4. Group - represented by 3 digits
  5. Class - represented by 4 digits
  6. Subclass - represented by 5 digits

A - Agriculture, farming of animals, hunting and forestry
  AA - Agriculture, farming of animals, hunting and forestry
    01 - Agriculture, farming of animals, hunting and related service activities
      011 - Agriculture
        0111 - Growing of cereals and other crops n.e.c.
          01111 - Growing of cereals
          01112 - Agricultural crops, n.e.c.

Source: Statistics Portugal

Portuguese Classification of Economic Activities - CAE Rev.3

- CAE Rev.3 contains 5 levels of aggregation:
  1. Section - represented by a letter
  2. Division - represented by 2 digits
  3. Group - represented by 3 digits
  4. Class - represented by 4 digits
  5. Subclass - represented by 5 digits

A - Agriculture, farming of animals, hunting and forestry
  01 - Agriculture, farming of animals, hunting and related service activities
    011 - Growing of non-perennial crops
      0111 - Growing of cereals (except rice), leguminous crops and oil seeds
        01111 - Growing of cereals (except rice)
        01112 - Growing of leguminous crops and oil seeds

Source: Statistics Portugal

- *validarcae* is a validation tool for codes of economic activity

- User-written command by BPLIM

- Why is this useful?
    - validates codes at any level of aggregation and allows to identify errors
    - helps to identify the revision when one is exploring the data and there is no metadata available
    - converts codes to higher levels of aggregation

validarcae

- accepts string or numeric variables

- reports ambiguous codes ("lost in translation" cases)
  - 011 Growing of non-perennial crops
  - 11 Manufacture of beverages

Syntax

- The syntax of *validarcae* is as follows:

```
validarcae var [if], [options]
```

Option	Description
rev(#)	specify which CAE Revision should be used
fromlabel	use the first word of the value label to retrieve the code
getlevels(#)	aggregate valid codes
dropzero	recursively drop zeros on the right from the code
keep	generate a string version of the variable

validarcae

- This command creates a new variable _valid_cae_# to identify the validity of CAE:

Code	Description
0	missing var
2	valid at 2 digits (0 + 1 digit)
10	valid at 2 digits only
20	valid at 3 digits (0 + 2 digits)
30	valid at 2 digits only or 3 digits (0 + 2 digits)
100	valid at 3 digits only
200	valid at 4 digits (0 + 3 digits)
300	valid at 3 digits only or 4 digits (0 + 3 digits)
1000	valid at 4 digits only
2000	valid at 5 digits (0 + 4 digits)
3000	valid at 4 digits only or 5 digits (0 + 4 digits)
10000	valid at 5 digits
200000	invalid code

Basic use

- By default, the command considers the most recent revision in force (CAE Rev. 3)

```
. validarcae cae
Variable cae is long
Checking compatibility with CAE rev. 3
         _valid_cae_3 |      Freq.     Percent        Cum.
----------------------+-----------------------------------
        2000 - 5d(0+4) |         57        6.90        6.90
  3000 - 4d or 5d(0+4) |          6        0.73        7.63
         10000 - 5d    |        763       92.37      100.00
----------------------+-----------------------------------
               Total  |        826      100.00
```

- this adds a variable *_valid_cae_3* to the data set

08992 Quarrying of other non-metallic mineral products, n.e.c.	2000 - 5d(0+4)
09100 Service activities incidental to oil and gas extraction, excluding surveying	2000 - 5d(0+4)
09900 Other mining and quarrying related service activities	3000 - 4d or 5d(0+4)
10110 Slaughtering of animals (production of meat)	10000 - 5d

- The code 9900 may be considered valid at two levels:
  - 5 digits: 09900 (Other mining and quarrying related service activities)
  - 4 digits: 9900 (Activities of extraterritorial organisations and bodies)

Options - read code in labels

- The command uses the first word of the value label to retrieve the code

```
. validarcae cae, fromlabel
Variable cae is long
Checking compatibility with CAE rev. 3
       _valid_cae_3 |      Freq.     Percent        Cum.
------------------------+-----------------------------------
        10000 - 5d |        826      100.00      100.00
------------------------+-----------------------------------
             Total |        826      100.00
```

Options - select the revision

- The user may also specify the revision to use when validating the codes

CAE Rev. 1	CAE Rev. 2	CAE Rev. 2.1	CAE Rev. 3
1	2	21	3

- For example, we can apply it to the years in which CAE Rev.1 was in force:

```
. validarcae cae, rev(1)
Variable cae is long
Checking compatibility with CAE rev. 1
     _valid_cae_1 |      Freq.     Percent        Cum.
------------------+-----------------------------------
           1 - 1d |          1        0.15        0.15
      100000 - 6d |        557       86.22       86.38
 200000 - Invalid |         88       13.62      100.00
------------------+-----------------------------------
            Total |        646      100.00
```

Options - drop zeros

- implements a recursive validation of invalid codes by dropping zeros on the right from the codes

```
. validarcae cae, rev(1) dropzero
Variable cae is long
Checking compatibility with CAE rev. 1
              _valid_cae_1 |      Freq.     Percent        Cum.
--------------------------+-----------------------------------
                  1 - 1d |          1        0.15        0.15
                100 - 3d |         17        2.63        2.79
          110 - 2d | 3d |          3        0.46        3.25
               1000 - 4d |         47        7.28       10.53
         1100 - 3d | 4d |          3        0.46       10.99
  1111 - 1d | 2d | 3d | 4d |        1        0.15       11.15
              10000 - 5d |         17        2.63       13.78
             100000 - 6d |        557       86.22      100.00
--------------------------+-----------------------------------
                    Total |        646      100.00
```

- this adds a variable to the data set informing how many zeros were dropped

cae	_valid_cae_1	_zerosdrop~d
122000 Logging	100000 ~ 6d	.
130100	1000 ~ 4d	2
130130 Trawl fisheries, excluding codfish	100000 ~ 6d	.
130140 Cod fishing	100000 ~ 6d	.
130150 Fishing of sardines and similar species	100000 ~ 6d	.
130160 Fishing of tuna and similar species	100000 ~ 6d	.
130190 Marine fishing, n.e.c.	100000 ~ 6d	.
130200	1000 ~ 4d	2

Options - aggregate codes

- The user may specify the level of the aggregation
- This option is only implemented for valid and unambiguous codes

	CAE Rev. 1	CAE Rev. 2	CAE Rev. 2.1	CAE Rev. 3
Section	NA	1	1	1
Subsection	NA	2	2	NA
Division	1	3	3	2
Subdivision	2	NA	NA	NA
Group	4	4	4	3
Class	3	5	5	4
Subgroup	5	NA	NA	NA
Subclass	NA	6	6	5
Detail	6	NA	NA	NA

Options - aggregate codes

```
. validarcae cae, fromlabel getlevels(1)
Variable cae is long
Checking compatibility with CAE rev. 3
```

_valid_cae_3	Freq.	Percent	Cum.
10000 - 5d	826	100.00	100.00
Total	826	100.00	

- This option adds a variable to the data set:

cae	rev3_section
23323 Manufacture of support or filler tiles	C - Indústrias transformadoras
46311 Wholesale of fruit and vegetables, except potatoes	G - Comércio por grosso e a retalho; reparação de veículos automóveis e motoc...
46422 Wholesale of footwear	G - Comércio por grosso e a retalho; reparação de veículos automóveis e motoc...
93294 Other amusement and recreational activities n.e.c.	R - Actividades artísticas, de espectáculos, desportivas e recreativas
20520 Manufacture of glues	C - Indústrias transformadoras
16295 Manufacture of other cork products	C - Indústrias transformadoras
50102 Coastal and local passenger water transport	H - Transportes e armazenagem
46214 Wholesale of grain, seed, grain legume, oleaginous and other agricultur...	G - Comércio por grosso e a retalho; reparação de veículos automóveis e motoc...
22191 Manufacture of rubber parts of footwear	C - Indústrias transformadoras
85591 Professional training	P - Educação
01491 Beekeeping	A - Agricultura, produção animal, caça, floresta e pesca
33170 Repair and maintenance of other transport equipment	C - Indústrias transformadoras

Options - aggregate codes

- The user may also opt to see the labels in English

```
validarcae cae, fromlabel getlevels(2, en)
```

cae	rev3_division
23323 Manufacture of support or filler tiles	23 - Manufacture of other non-metallic mineral products
46311 Wholesale of fruit and vegetables, except potatoes	46 - Wholesale trade (include commission trade), except of motor vehicles and motorcycles
46422 Wholesale of footwear	46 - Wholesale trade (include commission trade), except of motor vehicles and motorcycles
93294 Other amusement and recreational activities n.e.c.	93 - Sports activities and amusement and recreation activities
20520 Manufacture of glues	20 - Manufacture of chemicals, chemical products and man-made fibres, except pharmaceut...
16295 Manufacture of other cork products	16 - Manufacture of wood and of products of wood and cork, except furniture; manufactur...
50102 Coastal and local passenger water transport	50 - Water transport
46214 Wholesale of grain, seed, grain legume, oleaginous and other agricultur...	46 - Wholesale trade (include commission trade), except of motor vehicles and motorcycles
22191 Manufacture of rubber parts of footwear	22 - Manufacture of rubber and plastic products
85591 Professional training	85 - Education
01491 Beekeeping	01 - Agriculture, farming of animals, hunting and related service activities
33170 Repair and maintenance of other transport equipment	33 - Repair, maintenance and installation of machinery and equipment

Dependencies

- savesome (Nicholas J. Cox)

Where to get *validarcae*?

- To install *validarcae* run the following in Stata:

net install validarcae, from("https:
//github.com/BPLIM/Tools/raw/master/ados/General/validarcae")

- This will install the ado *validarcae*, four auxiliary adofiles and one ancillary file "caecodes.txt" to validate CAE codes.

- Thank you for the attention!