

## Quality indicators for registers in Statistics Portugal

The lack of a common identifier (ID), to link different data sources, constitutes a constraint to the updating process of the National Dwelling Register (FNA) in Statistics Portugal.

In this context a validation and quality assessment system was created in Statistics Portugal Data Warehouse, where a set of indicators for buildings and households has been defined to measure and monitor the quality of the National Dwelling Register (FNA) and some of the attributes and their completeness. Several indicators have been developed however this case focusses on the **Address Quality Degree Indicator (AQD)**. The indicator is explained in detail below.

### Indicator "Address Quality Degree" (AQD)

The indicator was created to assess the syntactic quality of an address. The AQD variable is based on the features described below, in which the following concepts were adopted:

- Mono-household building - Building with a single household;
- Multi- household building - Building with two or more households.

The AQD includes the following features:

- Structure of syntactic particles of address:
  - the absence of street name (R);
  - the existence of street name but the absence of the building door, number or name (name = pair of attributes: building type prefix and building name) locator (P);
  - the lack of households or the existence of duplicated households in multi-households buildings (A);
  - Non unique Address, till the buildings door number or name descriptor, between household and building in mono-household buildings (U);
- Absence of the designated family representative;
- The geographical area of the household location:
  - The 2011 Census Place. Characteristics of three different types of geography were studied: the Parish (2014 TIPAU), the 1Km<sup>2</sup> GRID (household density); the 2011 Census place (buildings). Given the more detailed level of spatial analysis that it allows, it was decided to feature in the AQD the 2011 Census Place.

The AQD variable is intended for use in the following situations:

- Monitor and measure the update/quality status of the households addresses for mailing purposes;
- Possible use, in the sample selection process, in order to prevent the selection of households whose addresses do not have a minimum quality that will allow addressing the circular (postal notification);
- Contact the municipalities whose geographic places have a large number of buildings with no street names in order to validate this information and to try to raise awareness to this reality.

### Type of 2011 Census Place

Statistics Portugal has proceeded to create a classification of 2011 Census Places regarding the existence of street names in the addresses. The 2011 Census Places typology has been defined on the basis of the following actions:

- Identify, for each building, its 2011 Census Place;
- Count the total number of buildings of each 2011 Census Place (N);
- Count, for each 2011 Census Place, the number of buildings with street names in the address (M);

- Calculate the percentage of buildings with street names in the address (PE) regarding the total number of buildings of the 2011 Census Place ( $PE = M * 100/N$ );
- Set a percentage scale with intervals of 5% of PE (PE);
- Represent, in a table, the 2011 Census Places on the PE scale;
- Identify, on a map, the 2011 Census Places on the PE scale;
- Define a threshold on the PE scale from which it is considered that the existence of addresses without street names is a real characteristic of the 2011 Census Place and not the result of poor quality addresses. In other words, we intent to know if the lack of street name in the address of a building results from the poor quality of its address or if it's a real characteristic of the 2011 Census Place were the building is located (“Insufficient Address - Real”)

The results of this indicator application are presented in the table below, which represents the number and percentage of 2011 Census Places at the PE Scale:

PE Scale	N° of 2011 Census Places	% of 2011 Census Places	Accumulated % of 2011 Census Places
A-0%	3083	11,58%	11,58%
B-0% a <5%	446	1,67%	13,25%
C-5% a <10%	376	1,41%	14,67%
D-10% a <15%	208	0,78%	15,45%
E-15% a <20%	203	0,76%	16,21%
F-20% a <25%	149	0,56%	16,77%
G-25% a <30%	173	0,65%	17,42%
H-30% a <35%	137	0,51%	17,93%
I-35% a <40%	126	0,47%	18,41%
J-40% a <45%	135	0,51%	18,91%
K-45% a <50%	100	0,38%	19,29%
L-50% a <55%	141	0,53%	19,82%
M-55% a <60%	123	0,46%	20,28%
N-60% a <65%	163	0,61%	20,89%
O-65% a <70%	135	0,51%	21,40%
P-70% a <75%	173	0,65%	22,05%
Q-75% a <80%	186	0,70%	22,75%
R-80% a <85%	217	0,81%	23,56%
S-85% a <90%	326	1,22%	24,79%
T-90% a <95%	593	2,23%	27,01%
U-95% a <100%	2699	10,14%	37,15%
V-100%	16736	62,85%	100,00%
<b>Total</b>	<b>26628</b>	<b>100,00%</b>	

Table 1. 2011 Census Places distribution at the PE Scale

Definition of a two value typology, with a threshold at 75 percent, to classify the 2011 Census Places regarding the existence or absence of street names:

- T – 2011 Census Places with street names: 2011 Census Places with PE > 75% - it means that more than 75% of the 2011 Census Place buildings have street names and it is acceptable to assume that addresses of buildings without street names are the result of errors that must be corrected.
- S – 2011 Census Places with no street names: 2011 Census Places with PE or <= 75% PE – it means that 75% or less of the 2011 Census Place buildings have no street name and it is acceptable to assume that buildings addresses without street names are considered as real,

as a characteristic of the 2011 Census Place itself, and not as the result of an error at the buildings address.

### **Household's Addresses Classification Algorithm (Algorithm and AQD categories)**

For the AQD categories the following algorithm was set:

- **(R) "Address without street name"** – Address just with Place or Locality attributes filled (Note: Validation Rule – If Place is filled the Locality must also be filled - there are no address registers with the Place filled and without the Locality filled):
  - **T** (2011 Census Place with street names) -> **C.R** (Insufficient address - without street name - Correctable)
  - **S** (2011 Census Place without street names) without the designated family representative -> **R.R.** (Insufficient address – Real - without Street name)
  - **S** (2011 Census Place without street names) with the designated family representative -> **S.R.** (Sufficient address without street name)
- **(P) "Address with street name and without building door number/name"** – Address with street name filled but without building door number or name (pair of attributes: building type prefix and building name):
  - **T** (2011 Census Place with street names) -> **C. P.** (Insufficient address – with street name but without building door number or name - Correctable)
  - **S** (2011 Census Place without street names) unnamed family representative -> **R. P.** (Insufficient address – Real - without building door number/name and family representative name)
  - **S** (2011 Census Place without street names) with the name of the family representative -> **S.P.** (Sufficient Address - Real - without building door number/name but with the family representative name)
- **(U) Building with 1 single household but with different addresses between building and household:**
  - **C. U.** (Insufficient Address - Correctable - without unique address between building and household)
- **(A) Building with more than one household without indication of floor number and/or side (right, left, etc.) or with inconsistent floor or side information** (duplicated floor and side, etc.):
  - **C.A.** (Insufficient Address - Correctable - without floor number)
- **(B) Remaining addresses, good:**
  - **B** (good address)

Notes:

- Without family representative name – Field "Name" = "FAMILY REPRESENTATIVE"
- With family representative name – Field "Name" <> "FAMILY REPRESENTATIVE"
- To aid the correction of all the potentially correctable Addresses it have been created specific excel files
- In the reports documents it only be presented the classification at one level. For example:

- If IR.R it will only be reported the IS.

#### **AQD Households Addresses Classification:**

**B** – Good -> Selectable and Addressable

**S** – Sufficient -> Selectable and Addressable (despite not all of the address attributes - locators - are filled)

**S.R.** (Sufficient address without street name)

**S.P.** (Sufficient Address - Real - without building door number/name but with the family representative name)

**C** – Insufficient correctable -> non selectable and not addressable (Address with errors that must be corrected in order to be selectable and addressable)

**C.R.** (Insufficient address - without street name - Correctable)

**C. U.** (Insufficient Address - Correctable - without unique address between building and household)

**C.A.** (Insufficient Address - Correctable - without floor number)

**C. P.** (Insufficient address – with street name but without building door number or name - Correctable)

**R** – Insufficient real -> selectable and non-addressable

**R.R.** (Insufficient address – Real - without Street name)

**R. P.** (Insufficient address – Real - without building door number/name and family representative name)

#### **Further work**

The application of these quality indicators will be revised and updated under the 2021 Census' final results.