Making the Demographic Censuses Compatible

PUC-Rio / Department of Economics

This document describes the variables obtained through the datazoom_censo ***STATA package, which makes the information contained in the Brazilian Demographic Censuses compatible year-by-year, from 1970 to 2010, and explains which procedures were taken to obtain them.***

Despite the similarities within successive Demographic Censuses in regard to the general themes approached by them, there is very little consistency between the different sets of microdata. This occurs, primarily, due to the constant modifications in the set of collected data: variables were included and excluded, categories were merged or changed and the extent to which some items were surveyed was altered in each Census. In general terms, these modifications are adopted in order to amplify the individuals and households' information sets surveyed. Accordingly, the information from most recent years can be used to reconstruct the type of information collected in previous years.

Furthermore, with every new Census substantial changes occur in the way data is stored: variable names and the order of categories are frequently altered, for example. Hence, even when identical information is available in two different Censuses, some effort is necessary to obtain comparable data.

The standardization prepared by `datazoom_censo` package consists of using the microdata available in the 1970, 1980, 1991, 2000 and 2010 Censuses to construct variables which may immediately be matched and compared between years. The databases generated by the program were constructed seeking to build upon any information which was conceptually very similar for two or more years. Among other procedures, the program named variables according to their content and standardized the set of categories, causing categorical codes to be exact the same throughout the different years.

Hereafter, the variables constructed by the program will be presented alongside details regarding the procedures adopted to obtain them and, when necessary, qualifications on their degree of compatibility.

Household identification

Frequently it is necessary to associate to a specific individual the characteristics of the household in which he/she resides. Hence, the household identification variable "id_dom" is available in both individual and household files, allowing that the observations in the household file be related to household residents in the individual file. Household identification variables can be readily found available from the 1980 Census onward. In these years, "id_dom" received these variables' values, but it was necessary to construct the 1970's Census's identification (details available below).

Observation: the '*id_dom*' variable simply allows the association of households to their residents, in accordance with the data of a specific year. <u>It is not possible</u> to identify the same household or individual in different years.

Data Organization and household identification within the 1970 to 2010 Censuses

Concerning the separation between household and individual information, and how to connect them, there are various and significant differences between the original microdata in different Censuses.

In the 1970 and 1980 Censuses, each data file (which corresponds to one state) contains information both on the sampled households and the individual residents. Each line within the file corresponds to one individual, and the household data appears (repeatedly) as part of each of its residents' registry. The 1980 Census also introduces the variable "household identification", which repeats itself for each household resident. This variable is unavailable in the 1970 Census.

*The construction of the '*id_dom*' variable in the 1970 Census uses the household identification found within the data structure itself. More specifically, the residents of a same household can be found in consecutive rows of data, where in each private household the first individual is the household head or sole resident. Using these observations, a household can be identified in the following manner: (1) if an individual resides in a private household where he is the household head or the sole resident, then this household is different from the one in which the individual described in the previous row lives; (2) if an individual resides in a collective household and in the previous row there is an individual who lives in a private household, then these two households are also different; (3) in all other cases, the household is the same as that of the previous row.*

In 1991 the files also contain both information on households and individuals, although the households are presented in separate rows. The rows containing households and individuals differ in regard to their content, but possess an equal portion which contains (among other information) the set type and the "questionnaire identification" (which allows the matching between each household and its residents).

Finally, in 2000 and 2010 the microdata of households and individuals presents itself in separate files. The "control number", available in both files, matchs the household to its residents.

Standardizing the Household Data

C.1.Household Setting

In 1970, the Census households were classified, concerning their setting, in three categories, according to their location in an “urban area", “suburban area” or “rural area”, such as delimited by the municipal administration. In the following years, the “suburban area” category was incorporated into the “urban area” category. Furthermore, the household setting also began to combine the legal factor (location in urban or rural area according to county) with other geographical factors (contiguity, infrastructure, service availability).

In the standardized data, using only the 1991 and 2000 Censuses, it is possible to maintain a more specific categorization:

```
* sit_setor =  1 – Urban area of village or city
*              2 – Non-urbanized area of village or city
*              3 – Isolated urbanized area
*              4 – Rural – urban extension
*              5 - Rural – small town
*              6 - Rural - nucleus
*              7 - Rural – other conglomerates
*              8 - Rural – excluding rural conglomerates
```

To make the data compatible with the 1980 Census it is necessary to compress the categories, in such a manner that the previous first two categories become “Urban – Village or City”, while categories 4 through 7 become “Rural Conglomerate”:

```
* sit_setor_B =  1 – Urban village or city
*                2 – Isolated urbanized area
*                3 – Rural Conglomerate
*                4 – Rural – excluding rural conglomerates
```

Finally, in the data standardization with the 1970 and 2010 Censuses it is only possible to discern between urban and rural areas. The distinction between both areas is readily available from 1980 and onward, while in the 1970 Census it was necessary to join the “suburban” category with the “urban” setting.

```
* sit_setor_C =  1 – Urban
*                2 – Rural
```

C.2. Household Sort

The categorization adopted within the successive Censuses was maintained. However, since only in 1980 was it possible to differentiate a collective improvised household from a collective permanent household, these two categories were joined together.

```
* especie = 0 – private permanent
```

```
*           1 – private improvised
*           2 - collective
```

C.3. Walls

This item was investigated only in 1980, 1991 and 2010. In 1980 and 1991 the alternatives are the same, though the codes are not. 1980's codes were made compatible with the 1991's. In 2010, alternatives are more disaggregated, but there are no major problems in grouping all in previous years' format: 'coated or uncoated brickwork' = 'brickwork'; 'coated or uncoated clay' = 'clay'; 'leftover wood' = 'leftover materials'. The alternative 'no walls' was considered as missing.

```
* paredes   = 1   Brickwork
*           = 2   Equipped wood
*           = 3   Uncoated clay
*           = 4   Leftover materials
*           = 5   Straw
*           = 6   Other
```

C.4. Roof/Covering

This variable is available only to 1980 and 1991, with a slight difference in the alternatives' codes. Besides, in 1980 there is the alternative 'asbestos roof tile' and, in 1990, the alternative is 'cement-asbestos tile'.

```
*cobertura     = 1 concrete slab
*              = 2 clay roof tile
*              = 3 asbestos roof tile
*              = 4 spelter
*              = 5 equipped wood
*              = 6 straw
*              = 7 leftover materials
*              = 8 other
```

C.5. Household Type

From the 1991 Census onward, it is possible to distinguish three types of households: "house", "apartment" and "room". In the file with the last two years of the Census, this information is presented in the following manner:

```
* tipo_dom = 1 - house
*            2 - apartment
*            3 – room
```

In 1980, only the categories "house" and "apartment" were available. Since "apartment" was attributed to the households which were later described as "rooms" (IBGE, 1980, p. 32), to make the 1980 Census compatible it is necessary to aggregate these two categories.

```
* tipo_dom_B = 1 - house
*              2 – apartment (or room)
```

In 1970, the household type was not researched. In 2010, however, there are two peculiarities: for permanent private households, there was the alternative 'hut or longhouse' to indigenous areas, which was coded as 'house'. Apart from other years, the type of household was investigated for improvised ('tent', 'trailer' etc) or collective ('asylum', 'jail' etc) households. These are considered here as missing.

In 1991 and 2000, in addition to the household type, the sector in which the household is located was also investigated, known as the "underdeveloped conglomerates" (favelas and other conglomerates which lacked proper infrastructure).

```
* subnormal = 0 – no
*              1 – yes
```

In order to maintain consistency, this variable was constructed only for private permanent households such as houses and apartments, since in 1991 data was gathered only for these two categories.

C.6. Occupation Condition and Rent

In all years, the households were divided regarding their occupation condition, into "private", "rented", "ceded" and "other condition". Furthermore, from 1980 onward a distinction is also made between "ceded by employer" and "ceded by other". For these years, the following categories are displayed:

```
* cond_ocup = 1 – privately owned
*             2 - rented
*             3 – ceded by employer
*             4 – ceded by other
*             5 – other condition
```

To standardize with the 1970 Census, categories 3 and 4 were joined together:

```
* cond_ocup_B = 1 – privately owned
*               2 - rented
*               3 – ceded
*               4 – other condition
```

In 2000 and 2010, it is possible to know whether the privately owned household was already paid or the residents were still paying for. This information generated the following variable:

```
* cond_ocup_C = 1 – privately owned, already paid
*               2 – privately owned, in acquisition
*               3 - rented
*               4 – ceded by employer
*               5 – ceded by other
*               6 – other condition
```

Except for 1991, it is possible to distinguish between a privately owned household which has already been paid for and one that is being acquired (or paid for). For this reason, for 1970, 1980, 2000 and 2010, there is:

```
* dom_pago = 0 – household in acquisition
```

```
*            1 – household paid for
```

For 1991 and 2000, the land ownership is also evaluated for privately owned households. Therefore, for 1991 and 2000, the following variable is available:

```
* terr_proprio = 0 - no
*                 1 – yes, land is owned
```

The monthly household rent, available in the 1980, 1991 and 2010 Censuses, is available under the variable 'aluguel'. In 1970, this variable is available only in terms of nominal ranges. The variable 'aluguel_70' was kept in the standardized database so as the user can make the standardization, if he wishes.

C.7. Water supply

For all years, the following variable was constructed:

```
* abast_agua = 1 – main network with internal plumbing
*              2 – main network without internal plumbing
*              3 – well or fountain with internal plumbing
*              4 – well or fountain without internal plumbing
*              5 – other
```

In 2000 and 2010, censuses inquired whether piped water was available.

```
* agua_canal = 1 – Piped at least in one room
*              2 – Piped only in the property or land
*              3 – Not piped
```

C.8. Sanitary Installments

In every Census (for private permanent households) the existence of access to some type of sanitary installments was assessed, even if these installments were accessible to more than one household. To compile these data, the following variable was constructed:

```
* sanitario = 0 – no access
*             1 – has access
```

Furthermore, in every occasion it was possible to determine the type of drain used within the installments available. These were made available with the following variable:

```
* tipo_esc_san = 1 – Main network
*                 2 – Septic tank
*                 3 – Rudimentary tank
*                 4 – Other drain type
```

These categories can be matched exactly to the categories available in 1970 and 1980. In the posterior Censuses, new categories were added each year, which are identical in 2000 and 2010 (there was an extension in 1991, but not in the same way as 2000 and 2010). Henceforth, only to 2000 and 2010 there is:

```
* tipo_esc_san_B = 1 – Main network
```

```
*                   2 – Septic tank
*                   3 – Rudimentary tank
*                   4 – Ditch
*                   5 – River, lake or sea
*                   6 - Other
```

Additionally, in 1980 and 1991, the exclusivity of the sanitary installments was also researched, so that the following variable became obtainable:

```
* sanitario_ex = 0 – no exclusive access to sanitary installments
*                 1 – exclusive access to sanitary installments
```

Finally, since 1991 the number of household bathrooms (room with a shower or bathtub and a sanitation fixture) was polled. This data was used to construct the following variable:

```
* banheiros = 0          - none
*              1 to 4    - number of bathrooms, from 1 to 4
*              5         – five or more bathrooms
```

C.9. Garbage Disposal

Only researched from 1991 onward. The categories remained the same in 2000 and 2010.

```
* dest_lixo = 1 – Collected by disposal service
*              2 – Thrown in disposal/waste basket
*              3 – Burned (on property)
*              4 – Buried (on property)
*              5 – Left in wasteland or empty lot
*              6 – Thrown in river, ocean or lake
*              7 - Other
```

C.10. Electric Lighting

In all years the presence of electric lighting was investigated.

```
* ilum_eletr = 0 – not available
*               1 – available
```

Additionally, in 1980 and 1991, the presence of an energy consumption meter exclusive to the household was investigated. The criteria are not clear in relation to the classification of households with a collectively owned meter, i.e., when the meter registers the consumption of more than one household. This delineation is made in 2010, for which the collective meter was considered as a household's exclusive energy consumption meter.

```
* medidor_el = 0 – not available
*               1 – available
```

C.11. Durable consumption goods

The Census evaluates the existence of various types of durable consumption goods (and occasionally the quantity and characteristics of these) in households. Since the survey for these goods was not uniform throughout the years, little data can be found in two or more years which can be standardized. The variables below were constructed for those goods available in at least two years:

Stove – This item is not present in the 2000 and 2010 Censuses, although it was surveyed in all other Census years conjointly with the type of fuel used for cooking. While in 1991 the type of cooking fuel was investigated for any type of cooking equipment, in 1970 the type of fuel was only asked when a stove was used. In 1980, however, it is possible to separately consider the existence of stoves and of cookers, which allows the construction of the following variables for 1970-1980:

```
* fogao = 0 – not available
*         1 – available

* comb_fogao = 1 - gas
*              2 - firewood
*              3 - coal
*              4 - other
*              0 – no stove available
```

Furthermore, for 1980 and 1991, it is possible to observe:

```
* fogao_ou_fog = 0 – not available
*                1 – available (either/or)

* comb_cozinha = 1 - gas
*                2 - firewood
*                3 - coal
*                4 - other
*                0 – neither stove nor cooker available
```

Radio – This item was surveyed uniformly throughout the 1970-2000 Censuses:

```
* radio = 0 – not available
*         1 – available
```

Refrigerator – In 1970 and 1980 only the presence of refrigerators was asked, while in 2000 the presence of a refrigerator or freezer was investigated in one single variable. In 1991 it is possible to discern between the presence of freezers and the presence of refrigerators. From 1970-1991, the following comparison is possible:

```
* geladeira = 0 – not available
*             1 – refrigerator available
```

For 1991 and 2000:

```
* gelad_ou_fre = 0 – not available
*                1 – refrigerator or freezer available
```

It is also relevant to observe that in 1991 only *electric* refrigerators were considered, which excludes similar equipment using gas or kerosene. However, since in 1991 the surveying of

appliances was only done in households *with* access to electricity, the distinction seems to be of little relevance.

Telephone – This item was not researched in 1970. From 1980 onward, questions were asked about the existence of telephone lines, even when these were only extensions from other households. In 2010, besides landlines, the existence of mobiles was investigated. Only landlines were considered in the construction of the variable below.

```
* telefone = 0 – not available
*            1 – available
```

Television – In 1980 and 1991 it is possible to evaluate separately the existence of color and B&W television. The following variables were therefore constructed for those years:

```
* televisao, tv_pb, tv_cores = 0 – not available
*                              1 – available
```

Since in 1970, 2000 and 2010 this distinction between types of television was not made, only the first variable is available for these years.

Automobile – The presence of automobiles was investigated in every Census, excluding freight cars but including utility cars. In 1970 and 2010, only private automobiles were evaluated. In 1980 both private and work cars were surveyed, although the former was only analyzed when a negative response had been obtained to the latter. In 1991 both types of cars were evaluated separately (along with the quantity of private cars and the type of work car). Finally, in the 2000 Census both types were aggregated. Therefore, the two variables below are available for 1980 and 1991, although the 2000 Census only allows for the first variable and the 1970 and 2010 Census for the second.

```
* automovel, automov_part = 0 – not available
*                           1 - available
```

Washing machine – Since 1991, the possession of washing machine is investigated. And, since 2000, the possession of personal computer.

```
* lavaroupa = 0 – not available
*             1 - available


* microcomp = 0 – not available
*             1 - available
```

Observation: in the 1991 Census the survey on appliances was limited to households with electric lighting. Therefore, the variables ‘geladeira’, ‘geladeira_ou_fre’, ‘televisao’, ‘tv_pb’ e ‘tv_cores’ in the standardized file are “missing data” for households without electric lighting in 1991. Therefore, for the proper statistical comparison calculations to be drawn, it is necessary to restrict the universe to those households for which the variable ‘ilum_eletr’ is 1.

C.12. Number of rooms

For all years the variables 'tot_comodos', which indicates the number of rooms in a household, and 'tot_dorm', which indicates the numbers of bedrooms in a household, were constructed.

C.13. Household Income

In every Census the monthly income was calculated for each of a private household's residents. This data, together with the data on the relation of each resident to the head of the household (available from 1980 onward; refer to section D.2), render information on household income. Household income consists of the sum of each resident's income, aside from domestic employees, their relatives, pensioners or guests. This data is available in the variable 'renda_dom'. However, for 1980, this variable is constructed from individual information. Therefore, it is located in the individual database only.

C.14. Household sampling weight

The variable 'peso_dom' refers to the household's sampling weight. No action was taken besides changing the original names in order to have the same variable for all censuses.

D.13. Number of household residents

The variables 'n_pes_dom, 'n_homem_dom' and 'n_mulher_dom' contain the total number of household residents, the number of male residents and the number of female residents, respectively. For 2010, the information about the number of male and female residents must be calculated from individual database. Therefore, the standardized household database does not contain these variables. It can be

Standardizing the Individual Data

D.1. Gender

The following categorization was adopted for all years:

```
* sexo = 0 - female
*        1 - male
```

D.2. Household and Family Condition

In 1980, 1991, 2000 and 2010 individuals' relation to the (a) head of the household and (b) head of the family was investigated. The alternatives remained the same throughout the

different Census years, although there are some categorization changes and a more complex alteration in the concept of "family". The definition of family is approached at the end of this section.

The 1991 Census covers this question best: its categories were aggregated to become compatible with the 2000 Census, resulting in the construction of the variables 'cond_dom' and 'cond_fam', which have the same categorization, shown below.

```
* cond_*** =  1 – Head (of household or family)
*             2 – Spouse, partner
*             3 – (Step)Son/(Step)Daughter
*             4 – Parent, In-Law
*             5 – Grandchild, Great-grandchild
*             6 - Sibling
*             7 – Other relative
*             8 – Guest/Friend
*             9 - Pensioner
*            10 – Domestic employee
*            11 – Domestic employee's relative
*            12 – Individual in collective household
```

To make the information in the 1991 and 2000 Censuses compatible with the 1980 Census, a rather different categorization system must be utilized. This gives origin to the two new variables 'cond_dom_B' and 'cond_fam_B'. The same categories are applicable to the 1970 Census, aside two details. Firstly, the position within the household was not surveyed in 1970, only the position within the family (which allows only for variable 'cond_fam_B' to be measured in the 1970 Census). Secondly, the category "domestic employee's relative" was not available in 1970, and there is no clear documentation as to how these people were categorized then. In 2010 Census, there is no family condition, but both relation to the head variables can be created.

```
* cond_***_B =  1 – Head (of household or family)
*               2 – Spouse, partner
*               3 – (Step)Son/(Step)Daughter
*               4 – Parent, In-Law
*               5 – Other Relative
*               6 – Friend
*               7 – Guest, Pensioner
*               8 – Domestic employee
*               9 – Domestic employee's relative
*              10 - Individual in collective household
```

The concept of "family" in the Brazilian Census

In the 1970, 1980 and 2000 Censuses, family was described as: (i) a person who lived alone; or (ii) two or more people connected by parentage or domestic dependence (applicable to guests, friends and domestic employees); or still (iii) two or more people living in a private household who hold ties from living together. In 1970 and 1980, this last case admitted a maximum five people; when there were six or more people without parentage or domestic dependence, the household was considered "collective".

*By this definition, households in which there were no relatives to the household head, but in which there were domestic employees **with their** relatives, **were not** divided into cohabitant families, even though conditions (i), (ii) and (iii) were violated by some factor of the scenario described. In such a rare case, it was considered that the household contained only one family, composed from the household head, his domestic employee and his domestic employee's relative(s), which justifies this last category as a possible "family position", such as was done in 1980 and 2000 (but not in 1970, as prior stated). (The discrepancy in 1970, however, seems to be of little quantitative significance since, in Rio de Janeiro in 1980, for example, this situation only affected 230 people in 157 households, where 1,242,287 people were surveyed in 342,368 households). While in 1980 this is the only circumstance in which "domestic employee's relative" is part of the same family as the head of the household, in 2000 there were cases in which employee's families were not separated from the employer's, even when the employer had a family (consisting of more than simply himself) in his household.*

In 1991, on the other hand, exceptionally in the aforementioned case, it was admitted that a sole household resident without relatives and his employee and employee's relatives were considered two different families. Therefore, in this year, the question "family position" did not admit "domestic employee's relative" as a response. In any other matters, the 1991 Census family was comprised of the same elements as those seen in other years.

D.3. Age

The microdata of all Censuses allows for the evaluation of age in complete years (`idade`) and, for children with less than one year of age, in months (`idade_meses`). Furthermore, with the exception of the 1980 Census, it is also reported if age was assumed (that is, if age was given directly, instead of being calculated through birth date documentation), through the following variable:

```
* idade_presumida = 0 – no
*                   1 - yes
```

D.4. Color/Ethnicity/Race

This item wasn't investigated in 1970. In 1991, 2000 and 2010, the alternatives are the same, while in 1980 the category "indigenous" doesn't exist, and the interviewer's manual indicates that people who describe themselves as "indigenous" should be considered "brown/pardo".

Henceforth, for 1991, 2000 and 2010, the following variable was constructed:

```
* raca = 1 - white
*        2 – black/african-american
*        3 – yellow/asian
*        4 – brown/pardo
*        5 - indigenous
```

The variable which also standardizes the 1980 Census is:

```
* racaB = 1 - white
```

```
*          2 – black/african-american
*          3 – yellow/asian
*          4 – brown/pardo
```

where individuals who were previously categorized as “indigenous” in 1991 and 2000 were reclassified as “brown/pardo”.

D.5. Religion

The categories for this item were expanded each year. Since 1980, however, it remained possible to aggregate numerous categories into a few, represented in the variable below:

```
* religiao = 0 – not religious
*             1 - catholic
*             2 – traditional evangelic
*             3 – pentecostal evangelic
*             4 - kardecist
*             5 – afro-brasilian spiritist
*             6 – oriental religions
*             7 – jewish/israeli
*             8 – other religions
```

To make the information compatible with 1970 it was necessary to compress even more categories:

```
* religiao_B = 0 – not religious
*               1 - catholic
*               2 - evangelical
*               3 – spiritist (kardecist or afro-brasilian)
*               4 - other
```

D.6. Mental or Physical Disability

Only the 1991, 2000 and 2010 Censuses investigated this theme. In 1991 seven types of disability were surveyed: for individuals with only one of them, the categories were the disability itself, while those individuals with two or more disabilities were put into the category “more than one”, which made it impossible to see which, exactly, were the disabilities which an individual had.

In the 2000 and 2010 Censuses, a more comprehensive questionnaire was applied, with individual questions surveying the occurrence of each disability and the intensity to which some of them occurred. For example, while in 1991 only total blindness was evaluated, in 2000 there were two intermediary levels of “sight difficulty”. However, the 2000 questionnaire also lacked information in certain categories, such as paraplegia and member loss, wherein only the most serious disability was identified (in this case, paralysis).

Due to these differences, 1991 was not standardized. For 2000 and 2010, the following variables indicate the level of difficulty of seeing, hearing and walking, respectively:

```
* dif_enxergar, dif_ouvir, dif_caminhar = 1 totally incapable
*                                         2 great difficulty
```

```
*                                          3 some difficulty
*                                          4 no difficulty
```

Besides those three variables, there is another one which indicates mental deficiency:

```
* def_mental = 1 has mental deficiency
*               0 does not have
```

D.7. Place of Birth and Migration

Migrant Condition – Since 1980, it is possible to identify if an individual has always lived in the same city. In 1970 it wasn't possible to distinguish between individuals who always resided in the same city from those who, having lived in other locations, later returned to their birth city. The variable constructed for 1980, 1991, 2000 and 2010 is:

```
* sempre_morou = 0 - no
*                 1 – yes, always lived in the same city
```

Additionally, in 1980 and 1991 it is possible to verify if the individual migrated between urban and rural zones of the city. This information is contained in the variable:

```
* onde_morou = 0 only urban zone
*               1 only rural zone
*               2 both rural and urban zones
```

Nationality and Place of Birth – In every Census the nationality of individuals was investigated, allowing the construction of the following variable:

```
* nacionalidade = 0 – native brazilian
*                  1 – naturalized brazilian
*                  2 - foreigner
```

Other information about an individual's place of birth is also supplied in every Census, such as if he was born in the city and state in which he currently lives. The following variables can thus be constructed:

```
* nasceu_mun, nasceu_UF = 0 no
*                         1 yes, born in the current city (state)
```

Furthermore, the federative unit or country of origin in which an individual was born are also supplied. This data is available in the variables 'UF_nascim' and 'pais_nascim' (codes in annex):

```
* UF_nascim = 11-53 codes FU of birth
```

```
* pais_nascim = 30-98 codes for foreign country
```

Also, in 1991, 2000 and 2010, the year in which each foreigner fixed residence in Brazil is registered in the variable 'ano_fix_res'.

Last Migration/Time of residency – In 1970, the time an individual resided in a city was only investigated for those who were not born in that city. That is, opposed to other censuses, the

information on those who were born, migrated and eventually returned to their city of birth was not collected in 1970. In 1980, however, the time intervals related to residency differed from those observed in the previous edition. Since 1991, this variable has been registered in a 'continuous' variable: 'anos_mor_mun'. Therefore, for all years apart from 1980 the following variable was constructed:

```
* t_mor_mun_70 =  0 – less than 1 year
*                 1 – 1 year
*                 2 – 2 years
*                 3 – 3 years
*                 4 – 4 years
*                 5 – 5 years
*                 6 – 6 years to 10 years
*                 7 – 11 years or more
```

And, except for 1970:

```
* t_mor_mun_80 =  0 – less than 1 year
*                 1 – 1 year
*                 2 – 2 years
*                 3 – 3 years
*                 4 – 4 years
*                 5 – 5 years
*                 6 – 6 to 9 years
*                 7 – 10 or more years
```

Note that, in order to obtain a standardized variable with identical intervals for all years, one has to recode the variables above so as to create the alternative "6 or more years". For a compatible statistic for all years, it is necessary to consider 'nasceu_mun' equal to zero. This way it will be possible to obtain the same sample surveyed in 1970. If the 1970 observations are disconsidered, this procedure is no longer necessary.

The necessary procedures for standardizing residency time variables vary throughout years. 1970 edition only provides information on those individuals who were not born in their respective cities of residency at the time of the survey, but everyone had to choose an alternative – in the case of those who had never migrated, this option would represent that person's age.

In 1980, for this last case the option 'nasceu' (born) was available and should be answered by those who were born and had always lived in the same Federative Unit. This procedure compares to the one adopted in 2010, but through a different questionnaire. That is, in 2010, those who had been born and always lived in the FU of residency was not inquired about how long he or she had lived there.

In 1991, respondents had to indicate the number of years they had resided in their current FU. This differs from 1970 for inquiring not only those who had not been born in the city of residency about migration, but also those who were native and had lived for some time elsewhere. In 2000, the results are very comparable to 1991's even though the questionnaires

differed. That is, even those who were born and had always lived in the FU had to indicate the time of residency in that FU.

Because the 1970 variable's options are years ranges, it is not possible to separate natives who had always lived in their FU's from those who eventually migrated. Therefore, only the 1980 and 2010 Censuses were adapted to the remainder editions by considering age as the time of residency in the FU for those people on whom information is not available. Besides the continuous variable 'anos_mor_UF' – applicable to 1991, 2000 and 2010 – another was created to all years apart from 1980.

```
* t_mor_UF_70=  0 – less than 1 year
*               1 – 1 year
*               2 - 2 years
*               3 - 3 years
*               4 - 4 years
*               5 - 5 years
*               6 - 6 to 10 years
*               7 - 11 or more years
```

And, except for 1970:

```
* t_mor_UF_80 =  0 – less than 1 year
*                1 – 1 year
*                2 - 2 years
*                3 - 3 years
*                4 - 4 years
*                5 - 5 years
*                6 – 6 to 9 years
*                7 - 10 or more years
```

It is worth noting that the reason why there are two variables is the difference in time intervals. Again, in order to include 1970, only those individuals to which 'nasceu_mun' equals to zero should be considered.

The city, FU and country of prior residency were also investigated for those who had migrated in the 10 previous years. However, only the 1991 and 2010 editions surveyed all three locations. 1970 and 2000 Censuses did not inquire about prior cities of residence, and in the 1980 microdata there is no variable indicating countries of prior residency, even though this question was in the questionnaire. Moreover, in 1970 even those who had migrated in the 10 previous years or before answered the questions relating to prior residency. This, together with the fact that only non-native respondents answered the migration-related questions, requires that the 1970 variables below be used cautiously. The countries' codes are the same used in 1970, meaning that countries that went through dismemberments were regrouped in this standardization.

```
* mun_mor_ant = cities' codes
* UF_mor_ant = 11-53 codes for Federative Unit of prior residency
* pais_mor_ant = 30-98 codes for foreign country
```

In 1970, 1980 and 1991 the setting of migrants' previous households was investigated.

```
* sit_mun_ant = 0 urban zone
*               1 rural zone
```

Place of Residence 5 years ago – From 1991 onward, the place in which an individual was living five years prior to the Census was also surveyed. Therefore, it became possible to construct the variables 'mun_mor5anos', 'UF_mor5anos'and 'pais_mor5anos', which inform, respectively, the previous city, state and country the individual lived in five years before the Census. In 1991 and 2000 the household setting in which individuals resided five years prior to the Census was also surveyed:

```
* sit_dom5anos = 1 urban zone
*                0 rural zone
```

D.8. Education

In each census year, every individual who was over five years old was asked if he or she was literate. In the 2000 Census, this information was asked for all interviewees. In the standardizing process, this information was considered missing to children less than 5 years old. An individual is considered literate if he is able to read and write at least a simple note in his native language at the time the Census was conducted. This information was made compatible and available through the variable "alfabetizado".

```
* alfabetizado = 0 - no
*                1 – yes, can read and write
```

Regarding schooling, school attendance was always researched, although the content therein was modified over the years. In all years researched, short professional development and extended cultural courses were disregarded. As a variable, school attendance evaluates if a given individual is currently enrolled in an educational institution, even if absent due to illness or school break. In 1970, individuals who were enrolled in the kindergarten were excluded, although they were described as meeting the criteria for school attendance in 1980 and 1991, years in which individuals in daycare were not. Furthermore, the 1980 Census also considered "attendees" people who participated in distance education programs through radio or television. The 2000 Census's attendance criteria encompassed individuals who were in kindergarten or at daycare. Finally, in 2010, prep-courses for college admission exams were disconsidered.

For the reasons above, in order to construct a variable that is comparable for all Censuses, daycare, kindergarten and prep-courses were not considered as criteria for attending school (neither did distance education programs through radio or television in 1980). The variable 'freq_escola' was thus constructed:

```
* freq_escola = 0 – no
*               1 – yes, attends school
```

The variable 'freq_escolaB' includes those who attend kindergarten in option '1'. In this case, it is compatible for all years but 1970.

```
* freq_escolaB = 0 – no
*                1 – yes
```

In 2000 and 2010 it was also investigated whether the school attended was public or private. The variable 'rede_freq' contains this information:

```
* rede_freq = 0 – private education
*             1 – public education
```

Apart from 1991, all other waves investigated whether the school attended was in the city of residency and this information is available under the variable 'mun_escola'. It is worth noting that one must consider data for which 'freq_escola' equal to 1 (and not 'freq_escolaB') in order to compatibilize the data with 1970 information.

```
* mun_escola = 0 - no
*              1 – yes, live in the same city where attend school
```

People's level of education was investigated in a different way every year. In 1970 the highest grade concluded by each individual is asked. In 1980 and 2000, this information is only collected for individuals who no longer attend school, while those who do attend school inform their current grade and educational level. A person who enrolls in a grade level inferior to one's highest concluded grade therefore "loses" (concerning what is being measure) their real qualification. In 1991 the same procedure as 1980 and 2000 takes place, although there is an explanation of how to recover information on an individual's highest concluded grade in the interviewer's manual.

Henceforth, 1970 and 1991 Censuses has enough information for a person's total years of schooling to be calculated, considering the highest grade and course ever attended. This information is contained within the variable 'anos_estudo'. In relation to 1980 and 2000 it is possible to make a different calculation of years of schooling for those who *currently* attend an educational institution, considering as highest grade the grade level below the individual's current grade. The variable 'anos_estudoB' combines this information with the data obtained in the first variable for those who aren't currently students.

In 2010 there is a loss of information as the one observed in 1980 and 2000. However, if the degree pursued at the moment of the survey was college/university level, the interviewee was asked whether he had previously graduated from another higher education institution. However, for those who attend college/university or do not attend school, it was asked neither the grade currently attended (for the former) nor the highest grade attended (for the latter). This way, it is impossible to construct a continuous years of schooling variable for 2010. In order to have a variable to compare 2010 to other waves, it was created a group of years of schooling variable based on anos_estudoB variable (and so not valid for 1970).

```
* anos_estudoC = 0 – no or less than 3 years of schooling (incomplete
primary education)

*               1 – 4 to 7 (primary/ secondary/ 1st cycle incomplete)

*               2 – 8 to 10 (secondary/ high school/ 2nd cycle
incomplete)
```

```
*                3 – 11 to 14 (high school graduated/
college/university incomplete)

*                4 – 15 or more (college/university graduated,
master's degree, doctorate)
```

All Censuses have a variable indicating the university course concluded (engineering, economics, history etc.). Throughout the years, new courses were created, others had their name altered and there were significant changes in the content of a few. There was an attempt to group courses by field of study, using IBGE's classification. Two variables were created: one which used older classifications and another with CONCLA's classification, considerably more modern than the first one. Besides, the original variable was maintained under 'curso_concl'.

```
* cursos_c1 =     3     humanities
*                 4     biological sciences
*                 5     sciences
*                 6     agricultural sciences
*                 7     social sciences
*                 8     defense area
*                 9     other courses

* cursos_c2 =     1     Education
*                 2     Arts, Humanities and Languages
*                 3     Social sciences, Business and Law
*                 4     Science, Mathematics and Computing
*                 5     Engineering, Manufacturing and Construction
*                 6     Agriculture and Veterinary Medicine
*                 7     Health and Welfare
*                 8     Defense area
*                 9     Other
```

D.9. Conjugal Situation

The 1991 Census presents the most extensive questionnaire, investigating the duration of the current conjugal situation and the age at which the individual first entered a common-law marriage, along with the individual's current conjugal state. In 1970 and 1980, only this last piece of information was obtained, although subject to a minor conceptual discrepancy which will be discussed further on. Since the current conjugal state is known for all Census years, including 2000 and 2010, the following variable was constructed:

```
* estado_conj = 1 married (civil and religious)
*               2 married (civil only)
*               3 married (religious only)
*               4 in a consensual union/common-law marriage
*               5 single
*               6 separated
*               7 legally separated
*               8 divorced
*               9 widowed
```

To fully comprehend the extent to which this variable can be analyzed, we must look at two of the characteristics upon which it was built. Firstly, the criteria used in 1970 and 1980 to

establish the “separated” and “single” categories differ slightly from that used in more recent years. While in the first two Censuses analyzed an individual who was “separated” could only be considered separated if he or she had been married and no longer lived with his/her spouse, from 1991 onward individuals who terminated a common-law marriage were also included within this category. Hence, an individual who had been in a common-law marriage but did not marry and later stopped living with his/her partner was considered, until 1980, “single”, whereas in 1991 and afterwards, this individual was categorized as “separated”.

Another matter to keep in mind is that the conjugal state wasn’t directly researched in the 2000 Census, making an indirect obtainment of such information necessary for the construction of the `‘estado_conj’` variable. In this year, it was asked if an individual had ever lived with a partner, the nature of his last union and his current civil status. Since the civil status (defined by law) does not coincide exactly with the same criteria of the conjugal status (defined by cohabitation patterns), the following procedure was adopted to standardize the data:

- Individuals who had never lived with a partner were considered “*single*”;
- Individuals who currently live with their partner, but whose civil status calls single were considered “*in a common-law marriage/consensual union*”;
- Individuals who currently live with their partner and whose civil status is married, were considered “*married*”;
- Individuals who do *not* live, but *have lived* with a partner were considered:
    - *divorced*, *legally separated* or *widowed* if this was their civil status;
    - “*separated*”, otherwise.

The rules to attribute a conjugal state herein are exactly the same as those prescribed by the 1991 Census interviewer’s manual, with the exception that it is impossible to tell the difference between an individual who was in a common-law marriage and who separated and one whose partner died.

In the standardized data the variable “`vive_conjuge`” is also available, derived from the conjugal state, which informs if a person lives with his or her partner. In 1991, 2000 and 2010 (but not in 1970 or 1980) it was also possible to determine which individuals lived, at any given point, with a partner. This data was made available through the “`teve_conjuge`” variable.

```
* vive_conjuge, teve_conjuge = 0 - no
*                              1 – yes
```

D.10. Economic Activity and Income

Up to 1991, the focus of the survey on individuals’ economic activity was their main occupation in the 12 months preceding the Census survey. In 2000 and 2010, similar questions were asked concerning the individual’s main occupation in the previous week. Therefore, most data available in these two editions are not directly comparable with the data collected in the previous years.

Economic Activity in the Last Year (1970 – 1991 Censuses)

From 1970 to 1991, the data collected concerning employment and income was related to the individuals' labor activities in the 12 months prior to the survey. Any remunerated occupation was considered employment, alongside non-paying occupations in which an individual: (i) aided a resident of the same household with his occupation; (ii) helped charities or other institutions with social programs and benefits; (iii) was an apprentice or intern. Moreover, in 1980 and 1991 individuals were required to spend a minimum of fifteen weekly hours at their non-paid occupation to be registered as "employed".

Employed individuals were asked about general traits of their jobs and some detailed information about their habitual (or main) occupation. This corresponds to the job in which the individual was employed for the greatest part of the reference period or to his most recent job, when a switch to the new job was definitive. In case an individual had more than one job, his main job should be considered: (a) that in which he spent a greater number of weekly hours and, in case of a tie, (b) that in which he earned more.

A worker's main occupation and his main work establishment's activity were both surveyed. The information collected is present in the standardized files, under the variables 'ocup_hab' and 'ativ_hab'. The occupation and activity codes of 1970 wave are different from those of 1980 and 1991 ones. However they were kept in order to elaborate compatible aggregates. One of the possibilities is to identify the group of occupations and activity sector, which was done for the three years herein considered, sorting these into the variables 'grp_ocup_hab' e 'set_ativ_hab', found below.

```
* grp_ocup_hab =  1 administrative
*                 2 technical, scientific, artistic and similar
*                 3 agricultural or of extractive vegetal/animal prod.
*                 4 extractive mineral production
*                 5 manufacturing and construction industry
*                 6 commercial and auxiliary activities
*                 7 logistics and communications
*                 8 services
*                 9 national defense and public safety
*                10 others / ill-defined / not stated

* set_ativ_hab =  1 agric. activ., vegetal extraction and fishery
*                 2 manufacturing
*                 3 civil construction industry
*                 4 other industrial activity
*                 5 commerce
*                 6 logistics and communications
*                 7 aux. serv. concerning economic activity
*                 8 services
*                 9 social (community, medical, dental and
*                   educational)
*                10 public adm., national defense or public security
*                11 other activity
```

From 1970 to 1991 the position in occupation was also investigated. It entails the relation of the employee to the establishment in which he executes his main activity. Generally, individuals can be divided into “employers”, “self-employed or autonomous workers” and “employees”. In addition, there are workers that received a different treatment and were separated into the following categories:

- The “migrant agricultural workers”, which are people who provide help with agricultural or vegetal extraction activities but have no fixed workplace, thus earning money by task, day or hour. In 1980 they were divided into “with” and “without” intermediaries, for those who were paid directly by their employers and for those who were paid by a third party.
- The “sharecropper” was an individual who worked in the primary sector and received part of his production as payment for his work (or *paid* part of his production for access to the land). In 1980 they were subcategorized as “employers”, “self-employed” and “employees”; in 1991 the two last groups remained independently categorized, but the first group was aggregated to the other employers.
- The “domestic worker” (identified by their occupation code), were divided only in 1991, when they were classified as “employees”, when they had only one employer, or “autonomous workers/self-employed”, when they worked for several people.

Henceforth, from 1980-1991, the following variable was constructed:

```
* pos_ocup_hab = 0 not paid
*                 1 migrant agricultural worker
*                 2 sharecropper - employee
*                 3 sharecropper – autonomous/self-employed
*                 4 domestic worker - employee
*                 5 domestic worker – autonomous/self-employed
*                 6 employee
*                 7 autonomous worker/self-employed
*                 8 employer
```

In 1980 this item originally didn’t distinguish between “domestic workers” and other employees or autonomous workers. To obtain the variable above, the category is identified through the occupation code (in that year, when occupation code was 805).

In 1970 the “migrant agricultural workers” were not a separate category, thus being incorporated into “employees”. Furthermore, sharecroppers were aggregated into one category, as long as they were not employers (occupation code 813). The following variable is valid for 1970, 1980 and 1991.

```
* pos_ocup_habB = 0 not paid
*                  1 sharecropper
*                  2 domestic worker - employee
*                  3 domestic worker – autonomous/self-employed
*                  4 employee
*                  5 autonomous worker/self-employed
*                  6 employer
```

Social security contribution is registered for all positions in occupations. This information generates the following variable 'previd_A'.

```
* previd_A = 0 - no
*            1 – yes, makes contributions to social security
```

Since 1980, individuals' salary and number of hours habitually worked in their main occupation (aggregating all establishments in which they worked) and in all their occupations altogether were investigated. In 1991, the number of complete hours worked was registered (with rounding, for fractions greater than half hours), but in 1980 categories were established to gather ranges of hours worked instead. Therefore, we have available to an individual's main occupation and for all of his occupations combined the variables:

```
* hrs_oc_hab, hrs_todas_oc = 1 – less than 15 hours
*                            2 – 15-29 hours
*                            3 – 30-39 hours
*                            4 – 40-48 hours
*                            5 – 49 hours or more
```

In 1970 the hours worked by an individual were surveyed only regarding his main occupation. Furthermore, workers in the agricultural or extractive sector were excluded (they should report, instead, number of months worked per year). Since distinct intervals were used in the 1970 and 1980 variables, two variables are made available: 'hrs_oc_habB', with 1970's categories (available to '70 and '91, but not to '80); and 'hrs_oc_habC' with aggregates which can be obtained over the three years.

```
* hrs_oc_habB = 1 – less than 15 hours
*               2 – 15-39 hours
*               3 – 40-49 hours
*               4 – 50 hours or more

* hrs_oc_habC = 1 – less than 15 hours
*               2 – 15-39 hours
*               3 – 40 hours or more
```

Income, on the other hand, was calculated according to position in occupation, in the following manner:

- "Employees" had their current fixed monthly salary computed, to which was added the average variable monthly pay (tips, commissions, extra hours). Bonuses made on profit and the 13th salary were not taken into consideration;
- "Employers" and "self-employed/autonomous workers" had their income calculated as their monthly average in the reference period (or the portion of the reference period in which an individual worked), corrected to the census's month.

In 1980 and 1991, there is a distinction between main (or habitual) occupation and other occupations. The income variables constructed are, respectively, 'rend_ocup_hab'and 'rend_outras_ocup'. There is also a separation between income earned from an

occupation and income from other sources, such as rents or donations. These alternative sources of income, however, vary from year to year; for this reason the variable '`rend_total`', representing the total income earned by an individual, was created. In 1970, there is total income variable only, as there were no distinctions between income sources or between occupation types (habitual or secondary).

The Census also surveys the activity condition of individuals who were not working. In 1980 and 1991, it was investigated among individuals who were 10 years-old or older which did not have a job if they had looked for a job (actively searched) in the twelve months prior to the Census. These individuals, added to those who worked in the reference period, composed the "economically active population".

The following variable is thus available for 1980-1991. The individuals who did not integrate the economically active population were classified into categories from 3 to 9 below:

```
* cond_ativ = 0 worked in the past 12 months
*              1 job-hunted – has worked before
*              2 job-hunted – has never worked before
*              3 retired/pensioner
*              4 lives off income
*              5 prisoner
*              6 student
*              7 sick/invalid
*              8 helps with domestic chores
*              9 no occupation
```

In 1970, there is no clear distinction between those who worked and those who job-hunted, though individuals who did not have a job and did not look for one were classified into the same categories as those in 1980-1991. A unique classification was created for 1970:

```
* cond_ativB = 1 worked/job-hunted - has worked before
*               2 job-hunted – has never worked before
*               3 retired/pensioner
*               4 lives off income
*               5 prisoner
*               6 student
*               7 sick/invalid
*               8 helps with domestic chores
*               0 no occupation
```

Secondly, the classification of individuals under job-hunters depended solely on their declaration about their activities in the twelve previous months, without making any reference to their last attempt. Besides, the classification under "non-economically active population" categories depended on which activity people perceived as the most important and not on the hierarchy adopted since 1980. In order to standardize data from 1970 to 1991, the variable 'pea' was created, indicating the economically active population. It equals '1' for those whose '`cond_ativB`' equals '1'or '2' in 1970 or '`cond_ativ`' is '3' or less.

```
* pea = 1 economically active
*        0 inactive
```

In 1970 and 1980 only, it is investigated in which job the individual was working at on the week before the survey, whether the main job during the year or another job, or even both. This is registered in the variable 'trab_semana':

```
* trab_semana = 1 main job only
*               2 main job and another one
*               3 another job only
*               4 others
```

With this last variable, it is possible to discern one's occupation in the week before the survey, whether it was habitual or not. This, in principle, could be used to standardize the information with 2000 and 2010 waves. However, only those who had worked or job-hunted in the twelve previous months were asked for this information. In more recent years, this information is irrelevant for the purpose of characterizing the occupation. Therefore, it is possible that people who had been inactive in the 12 previous months worked during the reference week, which renders the comparison impossible. For this reason, this particular information from 1980 Census was not standardized.

Economic Activity in Reference Week (2000 and 2010)

The investigation concerning individuals' economic activity was drastically modified in the 2000 Census, when the labor and activity timeframe analyzed became the "reference week". This methodological change implied evaluating individuals' short term situation instead of conditions which had been established over the past year, as was done up to then. It is therefore impossible to directly compare much of the information obtained in this matter in 2000 and 2010 to similar data in the previous Census years.

All variables below refer only to 2000 and 2010 Censuses. For these years, it is possible to identify whether one exerted remunerated activity in the reference week and whether he/she was away from work due to health problems, a vacation, etc. These data are available under the following variables:

```
* trab_rem_sem = 0 - no
*                1 – yes, worked in remunerated activity

* afast_trab_sem = 0 - no
*                  1 – yes, was away from work
```

Among those who had neither worked nor been temporarily away from work, it was asked if they had helped any household member with his activity or taken part in farming or animal breeding that was destined to the household's own consumption. This information is obtained through different ways throughout years. In 2000, there is an explicit mention to work as an apprentice or intern to another household member. In addition, there is no information on whether the activity of the household member that is being helped is remunerated or not. In 2010, there was only a generic reference to remunerated

apprenticeship with another resident, but this should not impact in substantial ways the concept of non-paid work, registered under two variables:

```
* nao_remun = 0 - no
*              1 – Yes, worked without remuneration

* trab_proprio_cons = 0 - no
*                      1 – yes, exerted farming activity for
household's own consumption
```

There is also information on whether the person had more than one occupation in the reference week:

```
* mais_de_um_trab = 0 – no, had only one occupation
*                   1 – yes, had more than one occupation
```

The main occupations' codes are registered in the variables 'ocup2000' (for both waves) and 'ocup2010' (only for 2010 wave). In 2010, there are both variables because a new classification was used by IBGE in 2010. In order to make years comparable, IBGE kept both classifications. In 2000 wave, there is a variable indicating occupation codes used in 1991. The variable 'ocup1991' was kept in case the user decides to make his own standardization. It is worth noting that the time prospect in 2000 differs from that observed in 1991.

The position in the main occupation of the reference week was investigated in a different way in years 2000 and 2010. Besides the non-paid job issue, Censuses differ on how the option 'domestic employee' is designed. While in 2000 the person got to choose this category, with or without formal labor contracts, in 2010 there is not reference at all to domestic employees; they are inserted in the option 'employee' with or without formal contract. The variable below was constructed using derived variables, which are not in the 2010 questionnaire (V6930 and V6940):

```
* pos_ocup_sem  = 1 – Employee with formal contract
*                 2 – Military and Government Officials
*                 3 – Employee without formal contract
*                 4 – Domestic Employee with formal contract
*                 5 – Domestic Employee without formal contract
*                 6 – Selfemployed/Autonomous
*                 7 - Employer
*                 8 – Not remunerated
*                 9 – Produces for own consumption
```

For employers, the number of employees in the enterprise can be obtained after recoding of the alternatives in 2000 so that they became comparable to 2010 data.

```
* qtos_empregados = 1 – 1 to 5 employees
*                   2 – 6 or more
```

Social security contribution is registered for employees without formal contract (except for public employees and militaries), selfemployed and employers. However, in 2000, this was only asked for the main occupation, while in 2010 it was possible for the respondent to inform whether he had made social security contributions in other occupations as well. This last information was ignored in the construction of the variable 'previd_B'.

```
* previd_B = 0 - no
*            1 – yes, makes contributions to social security in main
occupation
```

The number of hours spent working in one's main occupation is registered under 'horas_trabprin'. In 2000, 'horas_outros_sem' contained information on the number of hours respondents worked in other occupations. This information was kept for possible future applications despite its incompatibility problem with 1980 and 1991.

Those who stated not having worked (in a paid or non-paid position) were asked whether they had taken any measure to find a job in the reference week. This is registered in:

```
* tomou_prov = 0 - no
*              1 – yes, took measures to find job
```

Income from main occupation is available in 'rend_ocup_prin' and that of other occupations is in 'rend_outras_ocup'. Gross remuneration or gross income net of expenditures is computed and benefits (related to habitation and transport, for example) are disconsidered. In case income is in the form of products and merchandises, it is calculated by their estimated or real value. If seasonal, the income is calculated using the monthly average for the reference month[1]. Besides the "continuous" variables, in 2000 and 2010 there are also variables which indicate income in terms of minimum wages for main job and other jobs: 'rend_prin_sm' and 'rend_outras_sm'.

Income unrelated to work is registered in 'rend_outras_fontes'. This variable's components were not uniform between 2000 and 2010 - neither were previous Censuses – and therefore all components were included in the same variable. In 2000, for example, there is a variable containing information on food pension, donations and allowances, while all these sources of income are grouped with social programs or transfers in 2010.

'rend_total' contains information on total income, while 'rend_total_sm' pertains to total income in minimum wages. Both represent the sum of total income derived from work and other sources.

Finally, it is possible to compute the family total income, which represents the sum of all family member incomes, excluding domestic employee, relatives of domestic employee and pensioner incomes. In 1991, this information is available in the original dataset. For the other

[1] These orientations are available in the 2000 Census' documentation, but are not explicit in the variables' description in 2010. On the other hand, there is no indication they were altered in substantial ways.

waves, it is necessary to generate it. The standardized variable is 'rend_fam'. Note that all income issues discussed above remain valid for this variable.

All variables incomes are deflated to July 2010, but variables with nominal incomes are kept in the database. Variables with deflated incomes have the same name as nominal variables, but with the suffix 'def'.

Individual's job location

Except in 1991, it was investigated in which municipality the individual performs his labor activities. The problem is that the original question is ambiguous: it refers to the labor or school attendance location. Then, if the individual works and goes to school, it is impossible to know to which location his answer refers to. It that case, it was assumed that the answer refers to the job location. The following dummy variable is valid only for those who hava a jo (independent of school attendance):

```
* mun_trab = 0 – no, live and work in different municipalities
*           1 – yes, work in the same municipality in which live
```

A Note about Income

In every Census since 1970 some measurement of individuals' monthly income was taken. This measurement typically excludes revenue obtained "sporadically" (non-regulated donations, inheritance, lottery prizes). Between 1970 and 1991, for calculation necessities, income components had distinct treatments when considered "fixed" and when considered "variable". When "fixed", components were computed considering the value observed in Census's survey month; while "variable" components were computed using their average value over the last 12 months. In 2000 and 2010, on the other hand, all income components were measured according to their value in the reference month. Individuals' total income, available any year, was registered under the variable 'rend_total'.

While in 1970 only each individual's total income was obtained, in the following years this was divided according to the income's origin. Therefore, aside from labor-originated income – discussed above –, the following are differentiated:

- since 1980, the income obtained from "retirement and pension". This concept varied with time, though. In 1991, only what was paid directly due to retirement or pension by an official pension fund was considered; in 2000 closed private pension funds were included; and in 1980 several other types of income, such as permanent allowance and the 14th salary received from PIS/PASEP, were taken into consideration.
- in 1980 and 2000: "rent and lease" and "donations, allowance and alimentary donations" (transfers received from non-residents).

D.11. Fertility

Women who were 15 or older, and, from 1991 onward, women who were 10 or older, had data registered on children had, children born alive, children born dead and children still alive. In the standardized files, these quantities correspond, respectively, to the variables 'filhos_tot', 'filhos_nasc_vivos', 'filhos_nasc_mortos' and 'filhos_vivos'.

Since 1980 these values have been computed separately according to the child's gender. From this information, the following variables were created, where suffix 'hom' refers to male children and 'mul' refers to female children.

filhos_hom, filhos_mul = number of (fe)male children had

f_nasc_vivos_hom, f_nasc_vivos_mul = (fe)male children born alive

f_nasc_mortos_hom, f_nasc_mortos_mul = (fe)male children born dead

filhos_vivos_hom, filhos_vivos_mul = (fe)male children still alive

Additionally, since 1980 it has been possible to obtain the age (or time elapsed from birth) of the last child born alive; and from 1991 onward, this child's gender. This data is registered under the variable 'idade_ult_nas_v', values in complete years, and 'sexo_ult_nasc_v', described below.

```
* sexo_ult_nasc_v = 0 - female
*                   1 – male
```

The "children had" variable covers any pregnancy which reached seven full months. These are further classified as "born alive" (or "born dead") according to the presence of any vital signs – such as breathing or heart-beating – at birth.

Finally, it is worth noting that, in order to standardize all years, it is necessary to restrict age in Censuses 1991, 2000 and 2010 for fifteen or more, since in these waves, all women over 10 years old were interviewed – that did not happen in 1970 and 1980.

D.12. Individual sampling weight

The variable 'peso_pess' refers to the individual's sampling weight. No action was taken besides changing the original names in order to have the same variable for all censuses.

D.13. Number of family residents

The variables 'n_pes_fam', 'n_homem_fam' and 'n_mulher_fam' contain the total number of family residents, the number of male residents and the number of female residents, respectively. These variables are not in the standardized household database because some households have more than one family.