

Making the PNADs Compatible

PUC-Rio / Department of Economics

This document describes the variables obtained through the `datazoom_pnad` STATA package, which makes the information contained in the PNAD - Brazilian National Household Survey - compatible year-by-year, from 1981 to 2012, and explains which procedures were taken to do so.

There is relative consistency between the different sets of PNAD microdata. Yet, an important change occurred in 1992, as a more ample definition of labor was adopted and the themes of migration and fertility were introduced and remained in the survey ever since.¹ Also, variables' names were modified between the decades of 1980 and 1990 that were not necessarily associated to methodological changes.

Due to the reasons above, there were two different procedures in order to make the surveys' data compatible. The first one is based on surveys of the 1980s, and aims to make the versions of the following two decades compatible with the first. This required that only those variables which already existed in the 1980s were maintained. On the other hand, even in the 1980s, some issues such as race or ethnicity were only incorporated in the end of the decade. These variables are listed in the final databases, even though they might not have been available for a specific year.

The second process of conciliation follows the same pattern as the first, but in correspondence with the 1990s and the following decade. In this case, there are few differences throughout the years, compared to the ones found on the procedure on the 1980s' standardization. It was therefore possible to maintain most part of existing variables in the final databases.

The conciliations performed by the `datazoom_pnad` package consist in creating variables which can be directly compared from one year to another. The databases elaborated by the package attempted to compile all the information that could possibly be obtained in a conceptually consistent way, for households and the individuals, in most of the years. Among other things, the procedure standardized how the categories were registered (e.g., registering answers of the type "yes/no" in the same way, using the same codes to register states of residency and birth etc). For 1980s standardization, variables names (in Portuguese) are somehow informative of their contents. Supplemental information was disconsidered due to its low frequency.

¹ More details on the evolution of the themes which pertain to the PNAD can be found in the survey's Technical Notes.

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.

1. Standardizing the 1980s Data

In the 1980s, the each survey's year wasn't available as a variable in the microdata. This variable is created independently of the compatibility option chosen.

It is worth noting that, until 1990, there was only one type of registry. As a consequence, the information pertaining to households and individuals is gathered in the same microdata file. The variable that allows separating each type of information was used in order to create standardized files of households and people separately.

A.1. State or Federative Unit

The states codes were standardized only after 1992. Nowadays, each code possesses two digits: the first indicates the macro-region of the country, while the second indicates the state. Besides the modification of the states codes, a variable was created to indicate the macro-region. Finally, the state of Tocantins was aggregated with Goiás for those years in which the separation had already occurred. These procedures resulted in the following variables:

```
* regioao      = 1 - North
*              = 2 - Northeast
*              = 3 - Southeast
*              = 4 - South
*              = 5 - Midwest

* uf           = 11 - Rondônia
*              = 12 - Acre
*              = 13 - Amazonas
*              = 14 - Roraima
*              = 15 - Pará
*              = 16 - Amapá
*              = 21 - Maranhão
*              = 22 - Piauí
*              = 23 - Ceará
*              = 24 - Rio Grande do Norte
*              = 25 - Paraíba
*              = 26 - Pernambuco
*              = 27 - Alagoas
*              = 28 - Sergipe
*              = 29 - Bahia
*              = 31 - Minas Gerais
*              = 32 - Espírito Santo
*              = 33 - Rio de Janeiro
*              = 35 - São Paulo
*              = 41 - Paraná
*              = 42 - Santa Catarina
*              = 43 - Rio Grande do Sul
```

```

*           = 50 - Mato Grosso do Sul
*           = 51 - Mato Grosso
*           = 52 - Goiás
*           = 53 - Distrito Federal

```

1.1. Standardizing the Household Data

B. Control and Serial numbers

Household identification also differs in this decade and can easily be mistaken with others. The variable `v0101` in the 1980s represents the household number, while in the 1990s e 2000s, `v0101` indicates the survey year. Moreover, `v0101` does not exist in 1983 or in 1990. However, the usual variables which identify the household since 1992 are available in 1983 and 1990, but not for any other year. Thus, the variable `id_dom` was constructed to identify households in a unique way in all PNAD editions and facilitate the junction between the household and individuals' bases.

```

* ano       = year of the survey
* id_dom    = household identification number

```

C.1. Household location

Until 1990, in accordance with the dictionaries provided by IBGE, the household status could be 'urban', 'rural', 'urban conglomerate' and 'rural conglomerate'. In practical terms, however, only 'urban' and 'rural' appear in the database. Since 1992 the status can be classified in eight different categories, three of them urban and five rural. The urban categories were aggregated, as were the rural, so what remained was:

```

* urbana    = 1 - urban
*           = 0 - rural

```

C.2. Censitary status

The variable which indicates de censitary area code remained stable over time, having only its name altered throughout the decades.

```

* area_censit = 1 - metropolitan region
*             = 2 - self-representative municipality
*             = 3 - non-self-representative

```

Also, a dummy variable for metropolitan area was created:

```

* metropol   = 1 - metropolitan area
*           = 0 - self-representative and non-self-representative

```

C.3. Sampling Weights

For the data of 1992 until 2001, sampling weights contain the code '-1', which corresponds to an observation that is apparently missing. All codes were altered to *missing* (.) in order to prevent any misinterpretations of the code as a negative weight.

* peso - sampling weight

C.4. Total number of residents and Total number of residents 10-year-old or older

Similarly to the procedure with weights, the code '-1' was used between 1992 and 2001. The same procedure above was adopted.

For 2001 survey, instead of the number of inhabitants who were 10-years-old or older, the number of inhabitants five-year-old or older was counted. For this reason, this variable does not exist in the 2001 database. However, it is possible for the user to create this variable using the individuals file.

* totpess - total number of people living in the household

* totpess_10_mais - total number of 10-years-old or older people living in the household

C.5. Kind of Household

The variable was only recoded to the 1980s.

* especie_dom = 1 private permanent
* = 3 private improvised
* = 5 collective

C.6. Type of Household

Until 1990, there was additionally the category 'rustic' as one of the options to classify the type of household. This option was then incorporated to 'house'.

* tipo_dom = 2 house
* = 4 apartment
* = 6 room

C.7. Walls

Since 1992, 'straw' has been one of the options. To make it compatible, 'straw' and 'other material' options were aggregated.

* parede = 1 brickwork
* = 2 equipped wood
* = 3 uncoated mud
* = 4 recycled wood
* = 5 other material

C.8. Roof

As was the case with walls, 'straw' was added as an option in 1992. Again, this option and 'other material' were aggregated.

* cobertura = 1 tiles
* = 2 concrete
* = 3 equipped wood

```

*           = 4 zinc
*           = 5 recycled wood
*           = 6 other material

```

C.9. Water Supply

Until 1990, this theme was represented through one variable only. After 1992, that one variable was disaggregated into four other variables. In order to standardize them, a variable was created indicating whether the water (canalized or not) came from a main distribution network.

```

* agua_rede = 1 main distribution network
*           = 0 others (fountain, well, others)

```

C.10. Sanitary Drains

Until 1990, the type of sanitary drain was investigated for all households. Since 1992, this theme has been limited to households with bathroom or sanitary installment inside the property for their exclusive use. For this reason, the variable is valid only for the households that comply with that condition following 1992s.

```

* esgoto = 0 main network
*         = 2 septic tank
*         = 4 rudimentar tank
*         = 6 other

```

C.11. Sanitary Installments

Until 1990, there was only one question that combined the existence of a sanitary and its private or collective use. After 1992, this theme began to be addressed by two questions. Two variables were created, one indicative of the existence of sanitary installments and the other indicating its use. The second question is applicable only when the households possess sanitary installments (private or collective).

```

* sanit = 1 possesses sanitary installment
*       = 0 does not have sanitary installment

```

```

* sanit_excl = 1 exclusive sanitary installment
*            = 0 sanitary installment shared by more than one
household

```

C.12. Garbage

Since 1992, the options of garbage disposal have been altered. 'Collected' was dismembered into 'directly' and 'indirectly'; 'burned' and 'buried' were unified; 'thrown at sea or river' was introduced. A dummy variable was created to indicate garbage disposal.

```

* lixo = 1 garbage is collected
*      = 0 garbage is burned, buried, left in empty lot,
thrown in river, ocean or lake, others

```

C.13. Electric Lighting

Before 1992, the question was whether the household possessed electric lighting. In 1992, the question was altered to whether the lighting was 'electric', 'oil, kerosene, or bottled gas' and 'others'. It is believed that there is no relevant difference between adopting a 'yes/no' variable after 1992.

```
* ilum_eletr = 1 has electric lighting
*           = 0 does not have
```

C.14. Number of rooms and bedrooms

The only difference was that values '-1' and '99' were recoded to missing.

```
* comodos - number of rooms
* dormit   - number of bedrooms serving as bedrooms
```

C.15. Household's Occupation Condition

The alternatives changed from 1980s to the following decades. The option 'owned – not paid for' was substituted for 'owned – in acquisition'. In fact, the previous formulation might prevent those who were still paying for their houses from answering that they had not yet paid for them. The two alternatives were therefore combined. In addition to that, the option 'ceded' was dismembered into 'ceded by employer' and 'ceded by other'. The standardized variable is a dummy:

```
* posse_dom = 1 owned
*           = 0 rented, ceded, others
```

C.16. Value of Rent and Installments

The values that indicated non-declaration, which differed over the years, were recoded to 'missing'.

```
* aluguel - monthly rent paid in the month of the survey
* prestacao - monthly installment paid in the week of the survey
```

C.17. Water Filter

The codes of the alternatives have changed throughout the years. They were standardized into:

```
* filtro = 1 available
*        = 0 not available
```

C.18. Stove

In the 1980s, the question was whether the household possessed a stove. After 1992, this question was divided into two: whether there was a single burner stove or whether there was a stove with two or more burners. These two questions were combined in order to obtain a dummy variable.

* fogao = 1 available
* = 0 not available

C.19. Refrigerator

In 1992, in order to inspect the availability of a refrigerator, the previous yes/no question was replaced by one that allowed the type of refrigerator to be discriminated as having one or two doors. Again, a dummy variable was created.

* geladeira = 1 available
* = 0 not available

C.20. Radio

This item was inspected only in the years of 1982, 1988, 1989 and 1990. The codes have been altered over the years.

* radio = 1 available
* = 0 not available

C.21. Television

This item's inquiry only took place in the same years as the radio's was (that is, 1982, 1988, 1989, and 1990). In 1992, the yes/no question was replaced by another that would investigate the type of the television: "black and white" or "color". Again these variables were combined into a dummy variable:

* tv = 1 available
* = 0 not available

C.22. Household Monthly Income

Household income is the sum of incomes earned by all families living in the household. Family income is the total sum of each family member's income, apart from lodgers, domestic employees and employees' relatives. Until 1990, this was how household income had been calculated: disregarding each family member's position. If two people were lodgers in the household but constituted a second family in the same household, lodgers' incomes were computed in the household income, since they were not considered lodgers in their respective families. This has changed since 1992. If the second family was exclusively constituted by lodgers, domestic employees and their relatives, second families' incomes were not computed as household income. On the other hand, if the second family is constituted by the reference person's relatives or friends ("agredados" in Portuguese), these members' income was computed under household income.

In the standardization process, the variable from 1980s PNADs was kept unaltered and a second one was created similarly to 1990 and 2000's format. For all other PNADs, only the existing variable was kept².

```
* renda_dom      - household monthly income
* renda_domB     - household monthly income, compatible with
                  PNADs after 1992
```

1.2. Standardizing the Individual Data

In the individuals' files, some variables are constant to all household members, such as the household identification, federative unit and censitary status. In this section, those variables that exist in both the households' and individuals' files will not be discussed, due to the fact that the procedure adopted with them is the same as the one used with households.

B. Control and serial number and number associated with household member

For the 1980s the variable 'orden' is created even when no compatibilization is chosen, for this variable does not exist in the original files. From 1992 on, it was only necessary to rename this variable. For 'control number' and 'serial number', see section 1.1-B.

C.1 Household setting

In the 1980s, there were only two options for household location: urban or rural. Since 1992, it is possible to classify a household location into five different types of rural areas, and three types of urban area. These types were coded to construct a dummy variable.

```
* urbana = 1 urban area
*        = 0 rural area
```

C.2 Metropolitan region

The census area variable was recoded in order to create a dummy variable. Note that the variable below does not take into account changes in the composition of metropolitan region eventually implemented by IBGE.

```
* metropol = 1 metropolitan region
*          = 0 otherwise
```

C.3. Sampling weights

² In 1984 PNAD, the constructed variable differs from the existing one for some households. It is not possible to know whether this difference is due to an error in adding the incomes or the whether some of incomes are not available in the database.

Unlike household files, PNADs from 1992 to 2001 lack value “-1”. Therefore, the variables were only renamed as “peso”.

C.4. Gender

The variable was recoded in order to create a dummy variable.

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

C.5. Date of Birth

For the day of birth, zero indicates that the age was presumed or estimated. For the month of birth, 20 or 30 indicates that the age was presumed or estimated. For the year of birth, the variable renders the presumed or estimated age (from 0 to 98) in the cases that the date of birth is not known. All these codes were recoded to missing.

Only after the 2000s, the variable year of birth’ codes began to have four digits. Before, the chiliad was not mentioned, so, as an example, 1949 was coded as 949.

For the 1980s, in some cases, the year of birth appears switched with the variable of age. This was corrected in the standardized variable.

```
* ano_nasc = year of birth, with 4 digits
* dia_nasc = day of birth, from 1 to 31
* mês_nasc = month of birth, from 1 to 12
```

C.6. Age

The only necessary recodification was of (999), which represents an age not stated and was transformed to ‘missing’

```
* idade     = age in years
```

C.7., C.8. – Household and Family Condition

There was only a modification in the variables’ names.

```
* cond_dom = household condition
* cond_fam = family condition
```

C.9. – Family Number

There was only a modification in the variable’s names.

```
* num_fam = family number
```

C.12. – Color or Race

In the 1980s, the inquiry in color or race did not occur in 1981, 1983 and 1985. The option 'indigenous' was available for the first time as late as 1992. It is impossible to infer which category those who declared to be indigenous would have chosen before and, then, this category was not incorporated into any other. Therefore, this variable is not completely comparable among decades.

```
* cor      = 2 white
*          = 4 black
*          = 6 yellow
*          = 8 brown/pardo
*          = 0 indigenous
```

C.11. – Education Variables

In the 1980s, every individual who was over five years old had to answer all questions in the education section. In 1992, the questions to be answered started to depend on what alternative was chosen in the previous question, so it began to be possible not to answer all questions. Given how the survey was organized previously, there may be inconsistencies between answers of the 1980s surveys.

In most of the cases, each standardized variable was treated independently, since the aim was to make them compatible and not to eliminate all inconsistencies. On the other hand, the construction of a variable of years of schooling somehow corrects some inconsistencies, even though that is not its purpose.

C.11.1 – Years of Schooling

This variable was made "more continuous" in 1992. Before that year, there was one value for each of its first eight years of schooling, one for 9-11 range, and another one for 12 years of schooling or more. Since 1992, there is one value for each year of schooling: 0-16 years.

The standardization process did punctual modifications. The four-year secondary system does not account for twelve years of study in case the individual has completed this stage; standardized, this individual's years of schooling accounts for only eleven years, which is the total number of years for those who graduate from secondary school. Another modification was a dismemberment of the option '15 years or more', the highest category. The option '17 years or more' was assigned to those who concluded master's degrees or PhDs, and to those who were not attending school at the time of the survey but concluded at least six years of tertiary education. Sixteen schooling years were attributed to those who were not attending school at the time but had concluded at least the fifth year of tertiary education and those who were attending the sixth year of their tertiary education at the time. Another adjustment was made in relation to the code related to schooling years. Since there is no code zero in the original dictionary, the number of schooling years is always the number of the code subtracted by one.

```
* anoest    = years of schooling, from 0 to 17.
```

C.11.2. – School Attendance

In the 1980s, the option 'kindergarten' was not available. In the 1990s, this option exists, but it is not possible to separate 'school' from 'kindergarten'. This separation has only been possible since 2001. It is reasonable to surmise that a child who attended kindergarten in the 1980s fitted in the option 'attends school', though it is impossible to have any certainty about it. Besides this, only in 1995 people under five years old began to answer the education section. Therefore, the use of this variable requires caution.

```
* freq_escola = 1 attends school
*              = 0 does not attend school
```

C.11.3. – Literacy

The codes were recoded in order to create a dummy variable.

```
* ler_escrever = 1 knows how to read and write
*              = 0 does not know how to read and write
```

C.11.4. – Grade attended

For the 1980s, an auxiliary variable was used to make the grades compatible. If the grade attended was 'high school– 1st cycle', the codes were altered from 1- 4 to 5- 9. Therefore, the first grade of high school – 1st cycle' was considered 5th grade (of the old primary system) and so on.

Since 2007, the primary system might consist of eight or nine years. If it is the case that it consisted of nine years, the second year becomes the first grade, the third becomes the second and so on.

```
serie_freq = grade attended in school - from 1 to 8
```

C.11.5. – Education Stage Attended

Besides the insertion of 'kindergarten' in the 1990s, the option 'adults alphabetization' was also introduced in 2007. Also, the standardization process adopted the previous nomenclature in relation to the education attended, such as denoting 'ensino fundamental' as 'primary'.

```
* grau_freq = 1 primary school - regular
*            = 2 secondary school - regular
*            = 3 continuation (primary)
*            = 4 continuation (secondary)
*            = 5 college/university
*            = 6 adult education (literacy)
*            = 7 kindergarten or daycare
*            = 8 prep. school for vestibular
*            = 9 master's/PhD program
```

C.11.6 – Grade (for those not attending school)

As noted previously, this and the following questions should be answered only by those who do not attend school, so to indicate the highest grade concluded with acceptance in the last stage attended (the next question). However, in the 1980s, the survey was structured in a way

that those who attended school also answered such questions, having to select the option 'attends school' (option 0 (zero)). The value missing was ascribed in such cases.

serie_nao_freq = last grade concluded, from 1 to 8

C.11.7 Education stage (for those not attending school)

The answers to this question slightly differ from the answers of the previous one, due to the changes in education stages over time. In the 1980s surveys, it is not possible to select 'kindergarten' as the highest grade or stage attended. On the other hand, 'adult alphabetization' exists in 1981 and only after 1992.

```
* grau_nao_freq = 1 former 'primário' (1st to 4th year of 8-year
primary school)
*               = 2 former "ginásio" (5th to 8th year of 8-year
primary school)
*               = 3 former types of high school
*               = 4 primary
*               = 5 secondary
*               = 6 college/university
*               = 7 master's /PhD program
*               = 8 adult education (literacy)
*               = 9 kindergarten or daycare
```

C.12 – Characteristics of the Main Job performed in the week of the survey

In 1992, this section was reformulated, through the introduction of a more ample concept of labor. The activities related to production and construction meant for one's self-consume started to be investigated at the time. Besides that, the same restructuring of the questionnaire that had made possible that questions went unanswered occurred (i.e., the questions to be answered depend on what was the answer to the previous question). These factors imply that the variables pos-1992 and the ones in the 1980s are not exactly compatible.

C.12.0 – Work during the Week of the Survey

Since 1992, given that this is the first question, this variable is compatible with all other years. It was necessary to create its version as a dummy for the 1980s surveys.

```
* trabalhou_semana = 1 yes
*                  = 0 no
```

C.12.1 – Had work during the Week of the Survey

In this item, there is a problem between 1990 and 2000. After asking if the individual worked in the week, the following two questions in the 1990s refer to activities related to production for own consumption and to construction for their own use. The third question asks if the individual had a job in the week but was away for any reason (vacation, for example). On the other hand, in the 2000s, this last question is asked before the two questions related to production and construction activities. Then, the universe of individuals who answer those

three questions is different in the two decades. Due to the differences in the universe of the people who answer the questions, any attempt of standardizing this variable will not succeed perfectly. In spite of that, a dummy variable was created indicating who had work but was away for some reason.

```
* tinha_trab_sem = 1 yes, had a job but was away
*                = 0 had no job
```

C.12.2 – Occupation in the Week of the Survey

There has been a complete reformulation in IBGE's occupation codes beginning in the 2000's. Unlike Census microdata, however, PNADs do not gather both new and old codes in the same database, which could facilitate their standardizing process. Codes from 2002 onwards were kept in the standardized database under different names for an eventual standardizing attempt by the user. There is also a variable with occupation groups which also differs in the 2000's.

```
* ocup_sem = codes up to 2001

* grupos_ocup_sem = 1 technical, scientific, artistic and the like
*                  = 2 administrative
*                  = 3 agricultural and extractive production, plant or
animal
*                  = 4 manufacturing industry
*                  = 5 trade and auxiliary activities
*                  = 6 transport and communication
*                  = 7 services
*                  = 8 other or undeclared

* ocup_sem_nova = codes from 2002 onwards
* grupos_ocup_nova = codes from 2002 onwards
```

C.12.3- Activity or line of business

There were also changes in the activity codes. This topic is represented by two variables in the PNAD. One disaggregated, whose codes vary over the years, and another aggregated, consistent in the year 1980 and 1990, but not in 2000. Again, the 2000s variables were kept for eventual standardizing attempts by the user.

```
* ramo_negocio_sem      = 1980 and 1990 activity codes

* ramo_negocio_agreg    = 1 agriculture
*                        = 2 manufacturing industry
*                        = 3 construction industry
*                        = 4 other industrial activities
*                        = 5 trade goods
*                        = 6 services
*                        = 7 auxiliary services of economic activity
*                        = 8 transport and communication
*                        = 9 social
```

```

*           = 10 public administration
*           = 11 other activities or unreported

* ramo_negocio_sem_nova = 2000s activity codes
* ramo_negocio_agreg_nova = 2000s groups of activity codes

```

C.12.4- Position in Occupation

Before 1992, it was not possible to identify civil servants or military. Also, there were no specific categories for domestic workers. There has been research into the possession of a formal contract, but as there was no distinction between public and private, we chose not to separate the category 'employed' in 'employed with a formal contract' and 'unregistered' as it exists from 1992 as officials enter the category 'unregistered'. Thus, all these workers are considered employees. Those who worked in production activities or construction for own use or consumption is considered missing in standardized variable.

```

* pos_ocup_sem = 1 employee
*               = 2 self-employed
*               = 3 employer
*               = 4 unpaid

```

C.12.5- National Insurance Number

Although this question exists in all PNADs, due to the lack of distinction between public sector employees and private, the ownership of a formal contract is not very informative, because statutory and military employees are unregistered. Even so, the variable was retained in the based standardized.

```

* tem_carteiraassinada = 0 no
*                       = 1 yes

```

C.12.6 – Hours worked

Hours worked by those who worked in the reference week or was away from work but had an occupation (see C.12.0 and C.12.1)

```

* horas_trab_sem = hours usually worked by week in main
  occupation
* horas_trab_sem_outro = hours usually worked by week in other
  occupations
* horas_trab_todos_trab = hours usually worked by week in all
  occupations

```

C.13 – Income

In the 1980s, there are 12 variables related to income. We sought to adapt the remaining decades to the variables existing in 1980, without changing the latter. Values representing cases not applicable or skipped are converted to *missing*. All original variables are kept in the

standardized database, together with their deflated versions (as of September 2012), which carry the suffix “def”.

C.13.1 – Monthly Income in Cash

This variable refers to the main job income in cash and is compatible on all PNADs.

* renda_mensal_din = monthly income in cash in the main job of the week

C.13.2- Monthly Income in Products/Goods

Similar to cash income, but referring to products and goods received as payment.

* renda_mensal_prod = monthly income on products and goods in the main job of the week

C.13.3 – Monthly income in cash – Other Jobs

For 1980, there is a division between primary and other jobs, while in other decades, there is an intermediate category: the secondary job. The yields of the latter were added to the “other jobs”.

* renda_mensal_din_outra = monthly income on money in other jobs

C.13.4 – Monthly Income in Products/ Goods - Other Jobs

This variable contains the sum of non-monetary income of the secondary work and other jobs.

* renda_mensal_prod_outra = monthly income on products and goods in other jobs

C.13.5 – Retirement

Beginning in 1992, there is a separation of retirement payment received from Social Security Institute or the federal government from the other types. The values were added up and it was assumed that all types of retirement payments received in 1980 were declared in this item.

* valor_aposentadoria = income from all retirements

C.13.6 – Pension

Again, there is a separation of the pension payment received from the Institute of Social Security or the federal government from the other types. The same procedure adopted in retirement case were adopted here.

* valor_pensao = income from all pension

C.13.7 – Permanent allowances

Available for all PNADs.

```
* valor_abono = permanent income allowance
```

C.13.8 – Rent received

Available for all PNADs.

```
* valor_aluguel = income from rent received
```

C.13.9 – Other income

There is no distinction among other sources incomes for 1980. The 'other' does not exist separately in the following decades, but the items 'donations' and 'interest, dividends and other income' are inserted. All were considered 'other'.

```
* valor_outras = other sources of revenue
```

C.13.10 - Monthly Income on Main occupation

It is the sum of cash income, products and goods received in main job.

```
* renda_mensal_ocup_prin = income from main job
```

C.13.11 – Monthly income from all jobs

It is the sum of cash income, goods and merchandise received for all jobs.

```
* renda_mensal_todos_trab = income from all jobs
```

C.13.12 - Monthly Income from All Sources

It is the sum of job incomes, retirement, pension, rent, allowances and others.

```
* renda_mensal_todas_fontes = rendimento proveniente de todas as fontes
```

C.14.1 – Contribution to Social Security Institute

In the 1980s, this question was asked to individuals who worked or had a job in the reference week only. To make this question compatible through decades, this procedure was applied to 1990s and 2000s variables.

```
* contr_inst_previd    = 1 contributed  
                        = 0 did not contribute
```

C.14.2 – Type of Security Institute

Same situation described in the previous item. Also, 'undeclared' was changed to *missing*.

```
* qual_inst_prev = 2 federal
*                 = 4 state
*                 = 6 municipal
```

C.15 – Had another job

Codes were recodes to create a *dummy* variable.

```
* tinha_outro_trab = 1 Yes
                   = 0 No
```

C.16 – Previous occupation of people who were out-of-work in the week of the survey

Two alterations were made from 1992 in this subsection. In the 1980s, only those who were not working or were away from work answered this set of questions. Besides, there was an inquiry about the last job the individual had, independently of how long the individual has left the job. From 1992, every person whose occupation in the reference week was not the main occupation of the year (or the 358 days prior the week of the survey) was required to answer these questions. This means that the universe of respondents was increased and the period of the inquest was restricted. On the other hand, in the 1980s, there are variables which indicate how long the individual is out of work, so it is possible to delimitate the universe of respondents to those who left their previous work in less than twelve months prior to the survey. In account of such differences, the variables are not perfectly compatible between decades.

C.16.1 – Occupation in the prior year

There are two variables containing the occupation codes because of 1992 reform.

```
* ocup_ant_ano = 1980s and 1990s occupation codes for occupation
in the last 12 months (see dictionary)
* ocup_ant_ano_nova = 2000s occupation codes for occupation in
the last 12 months
```

C.16.2 – Activity or line of business in the prior year

There are two variables containing the occupation codes because of 1992 reform.

```
* ocup_ant_ano = 1980s and 1990s activity codes for occupation
in the last 12 months (see dictionary)
* ocup_ant_ano_nova = 2000s activity codes for occupation in the
last 12 months
```

C.16.3 – National Insurance Registration in the occupation of the previous year

```
* tinha_cart_assin_ant_ano = 1 yes, had register
```

= 0 no

C.17 – Made an attempt to find work in the week of the survey

This variable refers to individuals that were out of work - but not temporarily withdrawn from work. A dummy variable was created from the original variables.

```
* tomou_prov_cons_trab_sem = 1 yes, made attempt
*                          = 0 no
```

C.18 – Made an attempt to find work in the two months prior to the survey

This question is only valid to those who answered 'no' in the previous question. For the 1980s data, it is sufficient to recode the original variable, while for the other decades it is necessary to combine two questions related to the time before the referred week mentioned above.

```
* tomou_prov_cons_trab_2_meses = 1 yes
*                              = 0 no
```

C.19 – Attempts to find work

This variable applies only to those who attempted to find work in the two months which preceded the survey, or to those who were not working but not simply withdrawn from work in the referred week, even though since 1992 those who were employed were asked this question in the original survey.

```
* que_prov_tomou = 1 consulted an employer
*               = 2 responded to a public call for tenders
*               = 3 consulted an agency/syndicate
*               = 4 advertised for work
*               = 5 consulted relative
*               = 6 other
*               = 7 did nothing
```

2. Standardizing the 1992 data

The process of standardizing the data for 1990s and 2000s is simpler and more complete than the standardization to the 1980s, since there are fewer methodological alterations. In addition, after the major reformulation in the variables' names in 1992, the majority of them remained with the same name, suffering trivial alterations when a change in methodology occurred.

Not so uniform, however, is the evolution of the four questions that head the labor and income section. This is because of the alteration mentioned before in the order of appearance of the questions that relate to the withdrawal from work in the survey's week and to the non-remunerated activities in benefit of the household residents, between 1999 and 2001. These modifications did not cause a change in variables' original names, so it might be instructive to consider the chart that follows:

Variable's name	1992-1999	2001-2009
-----------------	-----------	-----------

V9002	Any activities in farming?	Was withdrawn from work?
V9003	Any activities in construction?	Any activities in farming?
V9004	Was withdrawn from work?	Any activities in construction?

Due to the few alterations over the years, a different approach was adopted in this standardization. The variables' original names were kept whenever it came to a punctual change, such as mark as 'missing' the codes relative to non-stated or ignored values. In the end of the household and individuals' sections, all variables which were kept in the standardized database are indicated.

2.1. Standardizing Household Data

A. – Year of survey

In the 1990s, this variable had only two digits. It was recoded so it possessed four digits instead, with no alteration in its name (v0101).

B. – Control and serial numbers series

Despite there is only one modification in these two variables (the control number has six digits in the 1990s and eight in the 2000s), they are replaced by a unique variable that identifies the household: `id_dom`. This variable is composed of `ano`, `UF`, `v0102` and `v0103` variables. It can be used to merge household and individual databases.

* `id_dom` = household identification number.

C. – Number of residents

The code '-1' (that is not mentioned in the original dictionary) was recoded to 'missing'. There are only two variables in this item, one indicative of the total number of residents and the other indicative of the total number of residents who were 10 years old or older. However, for 2001, this last variable considers 5 years-old as inferior cut-off. In the household file, it is not possible to correct this limit for 2001. But the user can use individual database to easily recover this information for that specific year.

* `v0105` = number of household residents

* `v0106` = number of household residents who are ten years old or older

D. – Telephone

Until 1999, the question was whether there was a telephone in the household. Beginning in 2001, two questions integrate the survey, one related to homephones and the other to cell

phones. The standardized variable is a dummy indicating the existence of any kind of telephone.

* `telefone` = 1 there is a homephone and/or a cellphone
 = 0 there is not a telephone

E – Other variables.

The remainder variables had their ignored or N/A codes altered to ‘missing’. It is worth noting that the monetary variables are deflated to September 2012 and carry the suffix “`def`” in their names.

2.2. Standardizing Individual Data

A.1 – Year

See item A, section 2.1.

A.2 – Control and serial numbers

See item B, section 2.1.

B – Year of Birth

Until 1999, the year of birth was available in three digits (1978 = 78). This variable was recoded in a way to provide four digits.

* `v3033` = 0 to 98 presumed age
 = 1870 to 2012 year of birth

C – Education Variables

C.1 – Years of Schooling

Similarly to the standardization to the 1980s, a variable related to years of schooling was created. For details see 1.2 Section, C.11.1.

* `anoest` = years of schooling, from 0 to 17

C.2 – Grade attended in School

From 2007, the primary school consists of 9 years. In this case, the ‘1st grade’ equals the ‘alphabetization class’.

* `serie_freq` = grade attended, from 1 to 8

C.3 – Stage attended in School

In 2007, the ‘alphabetization class’ was created. Previously it was incorporated in ‘kindergarten or daycare’.

```
* grau_freq = 1 primary school - regular
*           = 2 secondary school - regular
*           = 3 continuation school (primary)
*           = 4 continuation school (secondary)
*           = 5 college/university
*           = 6 adult education (literacy)
*           = 7 kindergarten or daycare
*           = 8 prep. school for vestibular
*           = 9 master's/PhD program
```

C.4 –Grade – for those not attending school

Similar to the previous procedure, given that an additional year was introduced in the primary system.

```
* serie_nao_freq = highest grade attended, from 1 to 8
```

C.5 – Stage – for those not attending school

Similar to the previous procedure, given that an additional year was introduced in the primary system.

```
* grau_nao_freq = 1 former "primário" (1st to 4th year of 8-year
primary school)
*               = 2 former "ginásio" (5th to 8th year of 8-year
primary school)
*               = 3 former types of high school
*               = 4 primary
*               = 5 secondary
*               = 6 college/university
*               = 7 master's/PhD program
*               = 8 adult education (literacy)
*               = 9 kindergarten or daycare
```

D – Child Labor

In 1996, 1997 and 2001, there was not a section exclusive to investigate labor activities of children of 5 to 9 years of age. However, while in the two first years there was no inquiry in work or income to those children, in 2001 the survey that was generally applied to people of ten years old or over was given to children from 5 to 9 years old.

Therefore, from 1996 to 1997, it is not possible to create a section about child labor. For 2001, this task is possible, but with restrictions. When there is a child labor section, the first questions relate to the occupation condition during the year, while the first questions of the section related to labor relate to the week of the survey specifically. This affects the way the questions are ranked and therefore the universe of respondents to each one.

Next, there is a list with the variables reconstructed to the year of 2001.

```
* v0701 = worked in that year
* v0702 = worked in production for own consumption
* v0703 = worked in construction for own consumption
```

- * v0704 = worked in that week
- * v0705 = was withdrawn from work in that week
- * v7060 = occupation codes - 358 days
- * v7070 = activity codes - 358 days
- * v0708 = occupation position - 358 days
- * v7090 = occupation codes - week
- * v7100 = activity codes - week
- * v0711 = position in that week
- * v7122 = nominal income from occupation in the week
- * v7125 = income in products and goods from occupation in the week
- * v0713 = hours worked

E – Labor and Income Characteristics

E.0 – 2001 PNAD Specificities

Only in 2001 the labor and income sections were responded by everyone five years old or older (in all other years, the minimum age is 10). For this reason, the variables related to 2001 labor section refer to the age of 5 or more. Also, there are differences in the variables' names (v4705 x v4755, for an example).

The standardized variables altered the reference age to the 2001 wave. Given that the criteria diverge for only one year, the original names of all variables were kept (exceptions are made explicit below). Next is a list of the 2001 variables that were modified and their respective names in other years:

- * v4754 para v4704 - week activity condition
- * v4756 para v4706 - position in that week's main occupation
- * v4757 para v4707 - total number of hours worked in all occupations
- * v4761 para v4711 - contribution to the social security system
- * v4763 para v4713 - occupation condition in that year
- * v4765 para v4715 - occupation position in that year
- * v4768 para v4718 - value of monthly income from main occupation
- * v4769 para v4719 - value of monthly income from all occupations
- * v4770 para v4720 - value of monthly income from other sources

E.1 – Worked in production for own consumption

As mentioned before (section 1.2., C.12 and C.12.1), the order of the first questions of this section was altered between 1999 and 2001, without altering the names of the original variables. The consequences were mentioned in those sections.

The first question, represented by the variable v9001, was kept unaltered. The variables v9002, v9003 and v9004 were transformed into the following dummy variables.

- * trab_consumo = 1 yes, worked on production for own consumption

= 0 no

E.2 – Worked in construction for own consumption

* trab_uso = 1 yes
= 0 no

E.3 – Was withdrawn from main occupation in that week

* trab_afast = 1 yes
= 0 no

E.4.1 – Main Occupation's Condition in that week

Until 2006, with the exception of 1996 and 1997, this derived variable was applicable, in the option 'occupied', to children who were five or under, while the option 'out-of-work' was applicable to people ten years old or over. For this reason, the variables' original names are slightly different (v4705 x v4805).

The variable below considers the age of ten or over.

* cond_ocup_s = 1 occupied
= 0 out-of-work

E.4.2 – Main Occupation's Condition in that year

The same observations of the previous item are valid for this item.

* cond_ocup_a = 1 occupied
= 0 out-of-work

E.5 – Position in Occupation

Beginning in 2007, the following positions in occupation were extinguished: 'employee who did not declare to have register in the social security system', and 'domestic worker who did not declare to have register in the social security system'. These categories were probably incorporated to the employees and domestic workers without registration in the social security system (formal contract), respectively. This aggregation was the only modification in the positions.

* posocup_sem = occupation position in that week
* posocup_ano = occupation position in that year

E.6.1 – Main activity of firm

Here as well occurred a change in the age of reference between 2006 and 2007, altering the variable name. Likewise, the standardization option was to use as the age of reference ten years or over.

* ativ_semana = 1 non-agricultural

= 0 agricultural

E.6.2 – Main Activity of that year's enterprise

Same as E.6.1.

* ativ_ano = 1 non-agricultural
= 0 agricultural

E.7 – Códigos de ocupação e atividade

As described in section 1.2 (items C.12.2 and C.12.3), the reformulation of occupation and activity codes since 2002 has made data between 1990 and 2000 less comparable, even though there was no alteration in post-2001 variables' original names. Again, original variables were kept for eventual standardizing attempts by the user.

* v9906, v9971, v9910, v9990 = occupation codes until 2001
(inclusive)
* v9906_novo, v9971_novo, v9910_novo, v9990_novo = occupation
codes from 2002 onward

* v9907, v9972, v9911, v9991 = activity codes until 2001
(inclusive)
* v9907_novo, v9972_novo, v9911_novo, v9991_novo = activity
codes from 2002 onward

F – Fertility

In the 1990, the fertility section was applicable only to women fifteen or older. Since 2001, the age was reduced to ten or older. This modification did not result in alteration in the variables' original names.

The standardized variables have as reference fifteen year old or older. The variables' names were slightly altered, with the inclusion of the letter 'c' (for an example, v1101c refers to v1101). No other alteration was made, with the exception of the below:

F.1 – Year of Birth of Last child born alive

In the 1990, the year of birth possessed only three digits. In the standardized variable, v1182c, all years count with four digits.

G – Deflation

The variables related to income (v4718 to v4722, and v4726) were deflated to September 2012, through an income deflator based on IBGE's INPC and created from the geometric means of the prices index from August to September of each year, in order to be

centered in the beginning of September, when, in general, the wages are paid.³ The original variables were not altered, and their nominal values were kept constant.

³ Foguel e Corseuil (2002). Uma sugestão de deflatores para rendas obtidas a partir de algumas pesquisas domiciliares do IBGE. IPEA: Texto para Discussão, n. 897.