**RAIS data set**

Source: obtained through a data agreement with the Ministry of Labor. DIME is the unit in the Bank which stores the RAIS data.

Shared folder with the data: https://worldbankgroup.sharepoint.com/sites/BRBSBFiles/Shared%20Documents/Forms/AllItems.aspx?ga=1&id=%2Fsites%2FBRBSBFiles%2FShared%20Documents%2FRestricted%2FRaisData%2FObservat%C3%B3rio%20Nacional%20do%20Mercado%20de%20Trabalho%20%2D%20RAIS%2F1%2Ddata%5Flabor%2F1%2Drais%5Fworker%2F2%2Ddata%2F03%2Dclean%2F1%2Dworker%2F1%2Dstata&viewid=fa980f94%2Ddcd5%2D47e8%2Da199%2D2043017ee085

Variables:

* Wage mass
* Average wage
* Wage threshold of the 25th percentile
* Wage threshold of the median
* Wage threshold of the 75th percentile
* Average wage among those in the 1st quartile of the wage distribution in the same year, microregion and economic sector
* Average wage among those in the 2nd quartile of the wage distribution in the same year, microregion and economic sector
* Average wage among those in the 3rd quartile of the wage distribution in the same year, microregion and economic sector
* Average wage among those in the 4th quartile of the wage distribution in the same year, microregion and economic sector

Level: worker-level data in each year since 1985

**CEMPRE data set**

READ ME: *We thought that this data would refer to both formal and informal workers, but we compared the numbers for specific cities (chosen at random) to those in RAIS, which cover only formal workers, and to our surprise the numbers in the CEMPRE data set are lower than those relative to the formal sector only. So we still need to understand whether this data set is reliable*

Source: SIDRA IBGE.

1996-2006: <https://sidra.ibge.gov.br/tabela/1735>

Obs2: the data for the 1996-2006 is only available for the variable “number of productive units”

2006-2021: https://sidra.ibge.gov.br/tabela/6449

Variables:

* Number of productive units (proxy for number of establishments)
* Population in the labor force
* Waged population in the labor force
* Wage mass (BRL 1,000 – current prices)

Level: municipality, economic sector and year. The economic sector classification is different in the data for 1996-2006 compared to that for the 2006-2021 (CNAE 2.0) period

**PNAD (1976-2015)/PNAD-C (2012 onwards)**

Source: IBGE

https://ftp.ibge.gov.br/Trabalho\_e\_Rendimento/Pesquisa\_Nacional\_por\_Amostra\_de\_Domicilios\_continua/Trimestral/Microdados/

Variables:

* Population in the labor force
* Waged population in the labor force
* Wage mass (BRL 1,000 – current prices)
* and many others

Level: individual-level data, but the most granular spatial unit at which the data is representative is metropolitan regions in each state.

**Censuses data (1991, 2000 and 2010)**

READ ME: *the classification used in the 2010 census is CNAE Domiciliar 2.0 while that in the 2000 census is CNAE Domiciliar. IBGE publishes the CNAE Domiciliar 2.0-CNAE Domiciliar. In turn, the 1991 census uses an ad-hoc classification that is not the CNAE one because CNAE became the official classification in Brazil in 1995*

Source: IBGE

Variables:

* Population in the labor force
* Waged population in the labor force
* Wage mass (BRL 1,000 – current prices)
* and many others

Level: individual-level data

Obs: data for 2000 and 2010 corresponds to a random sample of 10 percent of the census universe, while data for 1991 is a random sample of 25 percent of the census universe