

Trabalhando com dados do censo 2010

Elias Teixeira Krainski

Universidade Federal do Paraná
Departamento de Estatística
Laboratório de Estatística e Geoinformação

Maio, 2018



1 Notícia

2 Microdados do CENSO 2010

Sumário

1 Notícia

2 Microdados do CENSO 2010

Será?

<http://g1.globo.com/globo-news/noticia/2013/01/profissao-de-estatistico-tem-segunda-maior-media-salarial-do-brasil.html>

Edição do dia 23/01/2013

23/01/2013 10h56 - Atualizado em 23/01/2013 10h56

Profissão de estatístico tem segunda maior média salarial do Brasil

Segundo diretora do ENCE-IBGE, Denise Britz, mercado está em expansão porque a sociedade não toma decisões sem se basear em informações.

No jornal OGlobo: *média no país: R\$ 5.416 por mês. Só perde para os médicos, com ganho médio mensal de R\$ 6.940.*

Vamos ver. . .

A fonte

- Os números na notícia estão baseados em estudo do INEP com dados do CENSO de 2010
- O IBGE disponibiliza os microdados do CENSO
- Ideia: considerar esses dados e estimar o rendimento dos estatísticos

Sumário

1 Notícia

2 Microdados do CENSO 2010

Microdados do CENSO 2010

- Para o ano 2010 temos um arquivo compactados (.zip) por estado (exceto SP que são dois arquivos) em:

`https://ww2.ibge.gov.br/home/estatistica/populacao/censo2010/resultados_gerais_amostra/resultados_gerais_amostra_tab_uf_microdados.shtm`

- Os microdados estão em arquivos texto com colunas de largura fixa *fixed width file* - **FWF**
 - cada linha é uma sequência de caracteres
 - uma variável é definida por trechos dessa sequência
 - Exemplo: A graduação é um código de três dígitos nas colunas 159 a 161
- Vamos considerar esses microdados e estimar a renda dos estatísticos

Prepara para leitura dos dados

```
library(readr) ## para usar read_fwf() (eficiente)

### define posicoes das colunas (ver Documentacao)
ww <- fwf_positions(
  c(29, 159, 247, 322),
  c(44, 161, 253, 327),
  c('peso', 'grad', 'rend', 'rendt'))

### define classes (opcional, facilita)
colcl <- do.call('cols', list('d', 'i', 'd', 'd'))
```


Leitura dos dados

```
setwd(diretoriodados)
(z0 <- dir()) ## arquivos presentes no diretorio atual
(zz <- z0[grep('zip', z0)]) ## seleciona com '.zip'
names(zz) <- gsub('.zip', '', zz, fixed=TRUE)
res <- lapply(zz, function(z) { ## PR e RJ: 18:19
  system(paste('unzip', z))
  uf <- gsub('.zip', '', z, fixed=TRUE)
  fl <- dir(paste0(uf, '/Pessoas'))
  r <- read_fwf(paste0(uf, '/Pessoas/', fl), ww, colcl)
  system(paste('rm -r', uf))
  r$peso <- r$peso * 1e-13
  return(r)
})
setwd(wdir)
save('res', file='pesogradrend.RData', compress='xz')
```

Verificação inicial

```
rbind(am=sapply(res[18:19], nrow), ### tam. amostra PR e RJ
### Estimativa do total populacional: soma dos pesos
  tot=c(sum(res$PR$peso), sum(res$RJ$peso)))
```

```
##           PR           RJ
## am    1293034    1143650
## tot  10444526    15989929
```

```
round(sapply(res, function(d) sum(d$peso)/1e6), 1) ## (milh.)
```

##	AC	AL	AM	AP	BA	CE	DF	ES
##	0.7	3.1	3.5	0.7	14.0	8.5	2.6	3.5
##	GO	MA	MG	MS	MT	PA	PB	PE
##	6.0	6.6	19.6	2.4	3.0	7.6	3.8	8.8
##	PI	PR	RJ	RN	RO	RR	RS	SC
##	3.1	10.4	16.0	3.2	1.6	0.5	10.7	6.2

Número de estatísticos

```
### Estatisticos na amostra e estimativa do total desses
rbind(am=sapply(res, function(x)
  sum(x$grad==462, na.rm=TRUE)),
  tot=round(etot <- sapply(res, function(x)
    sum(x$peso[which(x$grad==462)])))) ## total
```

##	AC	AL	AM	AP	BA	CE	DF	ES	GO	MA	MG	MS	MT	PA	PB
## am	0	0	9	3	53	19	16	12	9	5	49	1	1	16	18
## tot	0	0	171	17	865	363	336	135	119	71	752	6	22	302	192
##	PE	PI	PR	RJ	RN	RO	RR	RS	SC	SE	SP1	SP2-RM	T0		
## am	18	1	52	163	20	1	2	34	6	2	78		103	2	
## tot	299	18	900	2875	287	10	16	419	39	42	898		2032	15	

Salário dos estatísticos

```

stot <- sapply(res, function(x) {
  ii <- which(x$grad==462)
  sum(x$peso[ii]*x$rendt[ii], na.rm=TRUE)
})
round(c(stot/etot, BRASIL=sum(stot)/sum(etot))) ### rendimento

```

##	AC	AL	AM	AP	BA	CE	DF	ES
##	NaN	NaN	0	0	1251	1199	515	429
##	GO	MA	MG	MS	MT	PA	PB	PE
##	1109	1061	792	0	12000	1194	1624	511
##	PI	PR	RJ	RN	RO	RR	RS	SC
##	120	5803	5158	53	0	0	611	374
##	SE	SP1	SP2-RM	TO	BRASIL			
##	3134	2862	2145	84	2771			