

# Tabulando a PNAD-C no Stata

Prof. Francis Petterini (PPGEco/UFSC)

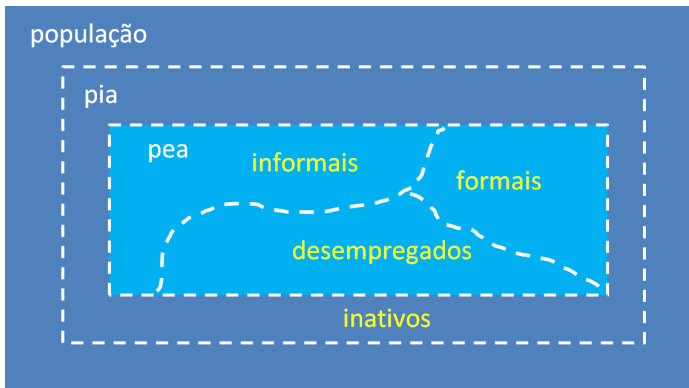
[f.petterini@ufsc.br](mailto:f.petterini@ufsc.br)

5 de novembro de 2024

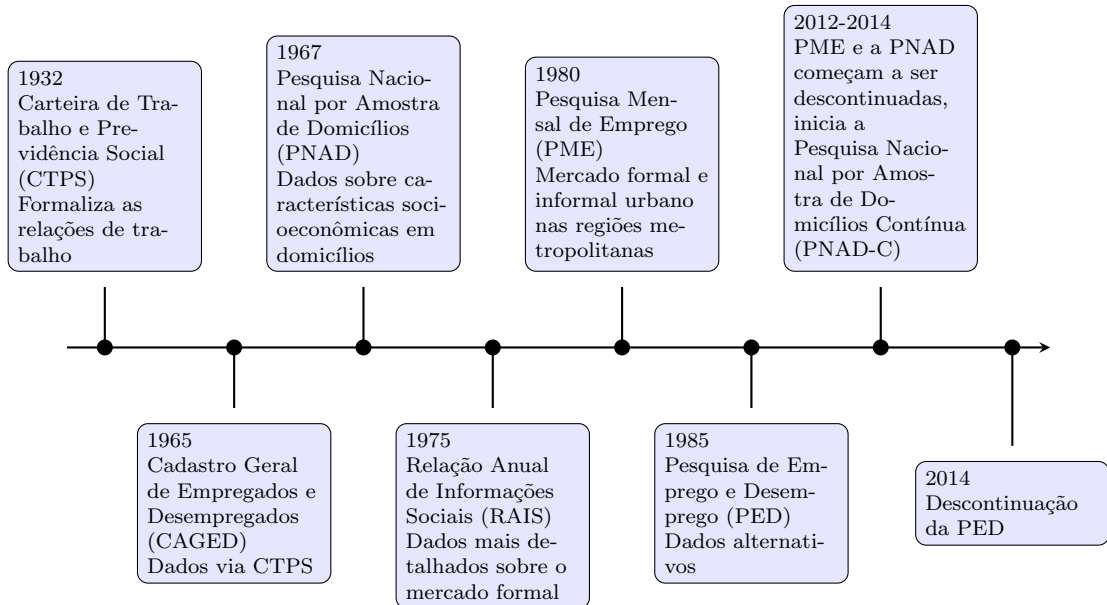
# Sumário

1. Contextualização
2. Estrutura longitudinal da PNAD-C
3. Microdados da PNAD-C
4. Exercício prático

# Mercado de trabalho



- ▶  $pea = \text{ocupados formais} + \text{ocupados informais} + \text{desempregados}$
- ▶  $pia = pea + \text{inativos}$



## O contexto em que nasce a PNAD-C

- ▶ **1999:** Desvalorização do R\$ após o abandono do câmbio fixo (esgotamento reservas internacionais que sustentavam uma paridade com o dólar) ⇒ aumento da inflação e incertezas econômicas ⇒ reformas estruturais ⇒ necessidade de políticas públicas ⇒ demanda de mais informações.
- ▶ **2002:** Eleição e a incerteza em relação ao novo governo ⇒ desvalorização do real e elevação da inflação e do risco-país ⇒ necessidade de políticas públicas ⇒ demanda de mais informações.
- ▶ **2008:** Crise financeira mundial (práticas arriscadas no setor financeiro dos EUA, especialmente no mercado imobiliário) ⇒ quedas nas exportações, restrição ao crédito e redução de investimentos ⇒ necessidade de políticas públicas ⇒ demanda de mais informações.
- ▶ **2010:** Crescente intervenção estatal, altos gastos públicos e baixa produtividade gera crise de desconfiança ⇒ demanda de mais informações.

<b>Característica</b>	<b>PNAD</b>	<b>PME</b>	<b>PNAD-C</b>
Abrangência	Nacional	6 regiões metropolitanas (São Paulo, Rio de Janeiro, Belo Horizonte, Porto Alegre, Recife e Salvador)	Nacional
Frequência	Anual (setembro)	Mensal	Contínua (microdados trimestrais)
Amostra	130k dom./ano	45k dom./mês	210k dom./trimestre
Objetivo	Emprego, renda, educação, moradia etc.	Limitado a emprego e renda	Emprego, renda, educação, moradia etc.
Longitudinal	Não	Sim	Sim

# Estrutura longitudinal da PNAD-C

Mês	Painel																														
	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12	A13	A14	A15	B1	B2	B3	B4	B5	B6	B7	B8	B9	B10	B11	B12	B13	B14	B15	
jan/09	5			4			3			2			1																		
fev/09		5			4			3			2			1																	
mar/09			5			4			3			2			1																
abr/09				5			4			3			2			1															
mai/09					5			4			3			2			1														
jun/09						5			4			3			2			1													
jul/09							5			4			3			2			1												
ago/09								5			4			3			2			1											
set/09									5			4			3			2			1										
out/09										5			4			3			2			1									
nov/09											5			4			3			2			1								
dez/09												5			4			3			2			1							
jan/10													5			4			3			2			1						
fev/10														5			4			3			2			1					
mar/10															5			4			3			2			1				
abr/10																5			4			3			2			1			
mai/10																	5			4			3			2			1		
jun/10																		5			4			3			2			1	

# Blocos da PNAD-C

- ▶ <https://www.ibge.gov.br/estatisticas/sociais/trabalho/17270-pnad-continua.html?=&t=microdados>
- ▶ [https://www.ibge.gov.br/arquivo/projetos/sipd/Questionario\\_PNAD\\_Continua.pdf](https://www.ibge.gov.br/arquivo/projetos/sipd/Questionario_PNAD_Continua.pdf)
- 1. Identificação e controle
- 2. Características gerais dos moradores (todos)
- 3. Características de educação (5 anos ou mais)
- 4. Características de trabalho (14 anos ou mais)
  - 4.1 Identificação da situação de ocupação
  - 4.2 Pessoas ocupadas
  - 4.3 Pessoas não ocupadas
- ▶ Suplementos
- ▶ Variáveis derivadas
- ▶ Pesos



## Padrão FWD

As variáveis são lidas por sequências definidas das posições das colunas, o que poupa espaço físico de armazenamento.

```
1 | 20221111 11000001611100110209311000103.17156451000156.52116685000554024007304041111210050101299999
2 | 20221111 11000001611100110209311000103.17156451000156.52116685000554024008515086111106050205199999
3 | 20221111 11000001611100110209311000103.17156451000156.52116685000554024009328304111104050305110062
4 | 20221111 11000001611100110209311000103.17156451000156.52116685000554024008288320111205050407299999
5 | 20221111 11000001611100110209311000103.17156451000156.52116685000554024007490024111101050510113062
6 | 20221111 11000001611100110309311000103.17156451000098.11953307000554024005264794111213040101219011
7 | 20221111 11000001611100110309311000103.17156451000098.11953307000554024002554486111115040212128021
8 | 20221111 11000001611100110309311000103.17156451000098.11953307000554024005466609111112040312101081
9 | 20221111 11000001611100110309311000103.17156451000098.11953307000554024007304041111210040412220071
10 | 20221111 11000001611100110509311000103.17156451000167.98615866000554024006117429111111040101199999
11 | 20221111 11000001611100110509311000103.17156451000167.98615866000554024008969441111204040205299999
12 | 20221111 11000001611100110509311000103.17156451000167.98615866000554024008506710111105040307199999
13 | 20221111 11000001611100110509311000103.17156451000167.98615866000554024007490024111101040410199999
14 | 20221111 11000001611100110609311000103.17156451000122.29889553000554024006792603111110050101126051
15 | 20221111 11000001611100110609311000103.17156451000122.29889553000554024007304041111210050202227041
16 | 20221111 11000001611100110609311000103.17156451000122.298895530005540240093283041111040503041119042
17 | 20221111 11000001611100110609311000103.17156451000122.298895530005540240059307081111103050404111092
18 | 20221111 11000001611100110609311000103.17156451000122.298895530005540240017911761111117050507199999
19 | 20221111 11000001611100110709311000103.17156451000167.57138771000554024008969441111204020101226062
20 | 20221111 11000001611100110709311000103.17156451000167.57138771000554024008506710111105020202199999
21 | 20221111 11000001611100110809311000103.17156451000145.31010633000554024008690546111208040101228071
```

# Chaves de Identificação

[https://ftp.ibge.gov.br/Trabalho\\_e\\_Rendimento/Pesquisa\\_Nacional\\_por\\_Amostra\\_de\\_Domicilios\\_continua/Trimestral/Microdados/Documentacao/Chaves\\_PNADC.pdf](https://ftp.ibge.gov.br/Trabalho_e_Rendimento/Pesquisa_Nacional_por_Amostra_de_Domicilios_continua/Trimestral/Microdados/Documentacao/Chaves_PNADC.pdf)

---











Domicílio	UPA + V1008 + V1014	Confiável
Pessoa	UPA + V1008 + V1014 + V2003	Desconfiável

---

- ▶ UPA: Unidade Primária de Amostragem (9 dígitos)
- ▶ V1008: Número de seleção do domicílio (2 dígitos)
- ▶ V1014: Grupo de amostra (2 dígitos)
- ▶ V2003: Número de ordem (2 dígitos)

## Exercício

<b>Ano</b>	2022				2023				2024	
<b>Trimestre</b>	1	2	3	4	1	2	3	4	1	2

 PNADC_012022	1.614.914 KB	Documento de Texto
 PNADC_012023	1.608.600 KB	Documento de Texto
 PNADC_012024	1.635.835 KB	Documento de Texto
 PNADC_022022	1.638.448 KB	Documento de Texto
 PNADC_022023	1.612.814 KB	Documento de Texto
 PNADC_022024	1.631.203 KB	Documento de Texto
 PNADC_032022	1.657.711 KB	Documento de Texto
 PNADC_032023	1.630.819 KB	Documento de Texto
 PNADC_042022	1.624.763 KB	Documento de Texto
 PNADC_042023	1.608.162 KB	Documento de Texto

```

*mklink /D "D:\drive" "D:/G/Meu Drive/pnad_c"
clear all
tic
loc dir "D:/drive"
// fs para listar arquivos .txt no dir
fs "`dir'/*.txt", dir
di r(files)
// Loop do infix
foreach file in `r(files)' {
  clear
  di "Processando o arquivo: `file'"
  infix ///
  ano 1-4 ///
  tri 5-5 ///
  capital 8-9 ///
  upa 12-20 ///
  v1008 28-29 ///
  v1014 30-31 ///
  v2003 91-92 ///
  sexo 95-95 ///
  dia_nasc 96-97 ///
  mes_nas 98-99 ///
  ano_nas 100-103 ///
  idade 104-106 ///
  raca 107-107 ///
  cond_dom 398-399 ///
  edu 406-407 ///
  pea 409-409 ///
  ocup 410-410 ///
  w1 444-451 ///
  w2 452-459 ///
  h1 462-464 ///
  h2 474-476 ///
  using "`dir'/'`file'"
  compress
  keep if capital == 42
  drop capital
  // Extrai o nome base do arquivo para salvar dta
  loc outfile: substr loc file ".txt" "", all
  sa "`dir'/'`outfile'.dta", replace
}
toc

```

```
clear all
tic
loc dir "D:/drive"
fs "`dir'/*.dta", dir
loc dta1 : word 1 of `r(files)'
use "`dir'/\`dta1'", clear
forval i = 2/10 {
  loc dir "D:/drive"
  fs "`dir'/*.dta", dir
  loc dta : word `i' of `r(files)'
  ap using "`dir'/\`dta'"
}
sa "`dir'/data_fln.dta", replace
toc
```

```

cls
clear all
loc dir "D:/drive"
u "`dir'/data_fln.dta"
egen dom = group(upa v1008 v1014)
egen id = group(dom v2003 dia_nasc mes_nas ano_nas)
g t = yq(ano, tri)
form t %tq
drop upa v1008 v1014 v2003 dia_nasc-ano_nas ano tri
so id t
fillin id t
g entrev = cond(sexo != ., 1, 0)
bys id: egen q = count(sexo)
ta q if t == tq(2022q1)
keep if q == 5 & idade >= 14
drop _fillin q
so id t
foreach j of varlist dom cond_dom sexo raca {
  bys id: egen media_`j' = mean(`j')
  replace `j' = media_`j' if missing(`j')
  bys id: g inconsistente_`j' = `j' != media_`j'
  bys id: egen media2_`j' = mean(inconsistente_`j')
}
su inconsistente*
foreach j of varlist cond_dom sexo raca {
  bys id: drop if media_`j' != `j'
}
su inconsistente*
drop media_dom-media2_raca

```

```

so id t
foreach j of varlist idade edu {
  bys id: egen min_`j' = min(`j')
  bys id: replace `j' = min_`j' if t == tq(2022q1)
  bys id: replace `j' = `j'[_n-1] if missing(`j')
  bys id: g delta_`j' = `j'-`j'[_n-1]
  bys id: egen max_delta_`j' = max(delta_`j')
}
ta max_delta_idade if t == tq(2022q1), m
drop if max_delta_idade > 1
ta max_delta_edu if t == tq(2022q1), m
drop min_idade-max_delta_edu
egen dom2 = group(dom)
egen id2 = group(id)
g w = max(w1, w2)
g h = max(h1, h2)
replace dom = dom2
replace id = id2
g branco = cond(raca == 1, 1, 0)
g homem = sexo-1
tab pea, m
replace pea = 0 if pea == 2
tab ocup, m
replace ocup = 0 if ocup == 2
drop dom2 id2 sexo raca w1 w2 h1 h2
order dom id t entrev cond_dom idade homem branco edu ///
pea ocup w h

```

```

compress

fillin id t

drop _fillin

tab t ocup, m
tab branco t, m
tab homem t, m
tab idade t, m
tab edu t, m
tab pea ocup, m
tab t ocup, m

gsort id -t

foreach j of varlist cond_dom idade homem branco edu {
  bys id: replace `j' = `j'[_n-1] if missing(`j')
}

sa "`dir'/data_fln_filtrada.dta", replace

```

```

cls
clear all
*loc dir "D:/drive"
*u "`dir'/data_fln_filtrada.dta"

import excel "https://petterini.ufsc.br/files/2024/11/data_fln_filtrada.xlsx", firstrow

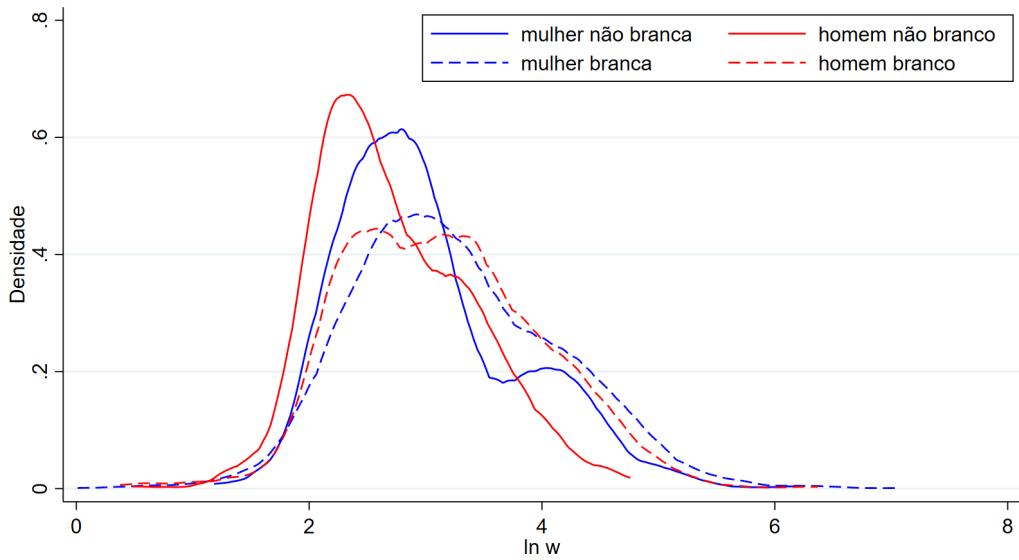
g tri = qofd(t)
form tri %tq
replace t = tri
drop tri

g lnw = ln(w/((30/7)*h))

tw (kdensity lnw if homem == 0 & branco == 0, lcolor(blue) lwidth(medium) lpattern(solid) ///
    legend(label(1 "mulher não branca") )) ///
(kdensity lnw if homem == 0 & branco == 1, lcolor(blue) lwidth(medium) lpattern(dash) ///
    legend(label(2 "mulher branca") )) ///
(kdensity lnw if homem == 1 & branco == 0, lcolor(red) lwidth(medium) lpattern(solid) ///
    legend(label(3 "homem não branco") )) ///
(kdensity lnw if homem == 1 & branco == 1, lcolor(red) lwidth(medium) lpattern(dash) ///
    legend(label(4 "homem branco") )), ///
xtitle("ln w") ytitle("Densidade") ///
legend(order(1 2 3 4) ring(0) position(1) colfirst) ///
graphregion(color(white)) bgcolor(white) xsize(16) ysize(9)

```





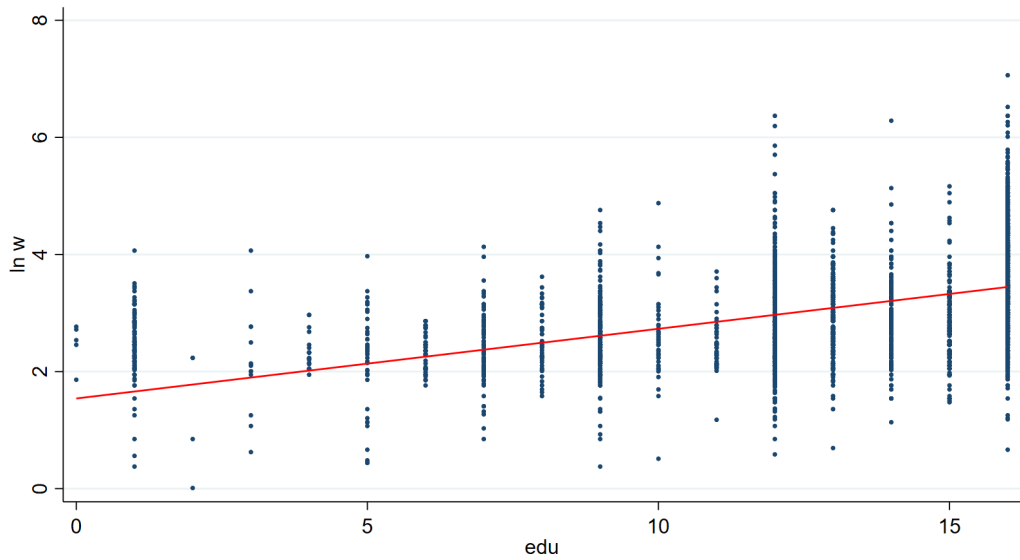
```
tw (scatter lnw edu, lcolor(black) msize(tiny)) ///
    (lfit lnw edu, lcolor(red)), ///
    xtitle("edu") ytitle("ln w") ///
    graphregion(color(white)) bgcolor(white) ///
    xsize(16) ysize(9) ///
    legend(off)

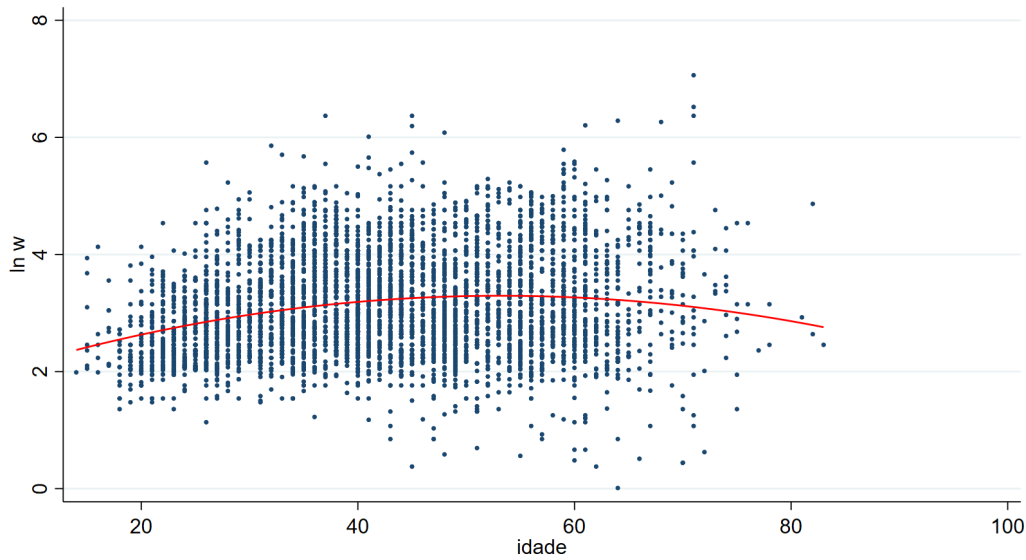
tw (scatter lnw idade, lcolor(black) msize(tiny)) ///
    (qfit lnw idade, lcolor(red)), ///
    xtitle("idade") ytitle("ln w") ///
    graphregion(color(white)) bgcolor(white) ///
    xsize(16) ysize(9) ///
    legend(off)
```

## Equação Minceriana

1. **Teoria do Capital Humano:** Educação e experiência são investimentos que aumentam a produtividade e, conseqüentemente, os salários.
2. **Retorno sobre o Investimento em Educação:** Salário futuro  $w$  depende exponencialmente dos anos de escolaridade  $S$ :  $w = w_0 e^{\beta_1 S}$ .
3. **Transformação Logarítmica:** Linearizando a relação entre educação e salários:  $\ln(w) = \ln(w_0) + \beta_1 S$  onde  $\beta_1$  é o retorno percentual da escolaridade.
4. **Inclusão da Experiência:** Adiciona-se a experiência  $X$  e o quadrado da experiência  $X^2$  para capturar o efeito crescente e depois decrescente da experiência:

$$\ln(w) = \beta_0 + \beta_1 S + \beta_2 X + \beta_3 X^2 + \varepsilon$$





```
g idade2 = idade^2

reg lnw edu idade idade2 homem branco i.cond_dom
est sto ols

heckman lnw edu idade idade2, select(ocup = edu idade idade2 homem branco i.cond_dom)
est sto he

xtset id t

xtreg lnw edu idade idade2, fe
est sto fe

ivregress 2sls D.lnw D.edu D.idade D.idade2 (D.L.lnw = L2.lnw), nocon
est sto ah

estout ols he fe ah, cells(b(star fmt(3)) se(par fmt(3)))
```

# Suplementos

[https://ftp.ibge.gov.br/Trabalho\\_e\\_Rendimento/Pesquisa\\_Nacional\\_por\\_Amostra\\_de\\_Domicilios\\_continua/Anual/Microdados/PNADC\\_Pesquisas\\_Suplementares\\_Anuais\\_20241018.pdf](https://ftp.ibge.gov.br/Trabalho_e_Rendimento/Pesquisa_Nacional_por_Amostra_de_Domicilios_continua/Anual/Microdados/PNADC_Pesquisas_Suplementares_Anuais_20241018.pdf)

[https://www.ibge.gov.br/estatisticas/downloads-estatisticas.html?caminho=Trabalho\\_e\\_Rendimento/Pesquisa\\_Nacional\\_por\\_Amostra\\_de\\_Domicilios\\_continua/Anual/Microdados/Trimestre](https://www.ibge.gov.br/estatisticas/downloads-estatisticas.html?caminho=Trabalho_e_Rendimento/Pesquisa_Nacional_por_Amostra_de_Domicilios_continua/Anual/Microdados/Trimestre)

```
cls
clear all
tic
infix ///
uf 6-7 ///
capital 8-9 ///
rural 33-33 ///
familia 576-576 ///
amigo 577-577 ///
colega 578-578 ///
vizinho 579-579 ///
guarda 580-580 ///
civil 581-581 ///
pm 582-582 ///
bombeiros 583-583 ///
justica 584-584 ///
forcas_arm 585-585 ///
using "Z:/drive/PNADC_2021_trimestre4.txt", clear
compress
drop if familia == .
replace rural = rural-1
sa "Z:/drive/PNADC_2021_trimestre4.dta", replace
toc

foreach var in familia amigo colega vizinho guarda civil pm bombeiros justica forcas_arm {
    tab `var' rural, col
}
```