

```
*****
***** APOSTILA DO MINICURSO MICRODADOS COM O USO DO STATA *****
***** AUTORES: IGOR VIEIRA PROCÓPIO E RICARDO DA SILVA FREGUGLIA *****
***** ECONS - FACULDADE DE ECONOMIA UFJF *****
***** SETEMBRO DE 2009 *****
*****
```

```
** Definir Diretório de Trabalho
chdir "C:\Microdados Usando Stata"
```

```
*** Rodar as linhas de 4 a 23 para imputar o banco de dados "banco1" que será utilizado para fazer os exemplos das seções 3 a 6:
```

```
clear
input str2 var1 str4 var2 str4 var3 var4 var5 var6 var7 str6 var8 var9 var10 str8 var11 var12 var13
```

| | | | | | | | | | | | | |
|------|--------|--------|---|----|----|----|----------|----|----|-------------|----|-------------|
| "MG" | "521" | "n/d" | 1 | 15 | 15 | 15 | "B" | 99 | 1 | "R\$ 45,35" | 5 | 132132136 |
| "MG" | "123" | "1360" | 1 | 51 | 51 | 51 | "BRANCO" | 2 | 0 | "R\$ 54,94" | 8 | 13549687808 |
| "MG" | "150" | "543" | 2 | 32 | 32 | 32 | "PARDO" | 3 | 0 | "R\$ 87,24" | 4 | 13549576192 |
| "MG" | "1200" | "1200" | 2 | 41 | 41 | 41 | "NEGRO" | 1 | 1 | "R\$ 12,32" | 15 | 1234968448 |
| "MG" | "410" | "1590" | 2 | 20 | 20 | 20 | "NEGRO" | 2 | 99 | "R\$ 45,36" | 6 | 1354654976 |
| "MG" | "851" | "2806" | 1 | 13 | 13 | 13 | "PARDO" | 4 | . | "R\$ 68,97" | 8 | 13216546 |
| "SP" | "1250" | "*" | 2 | 50 | 50 | 50 | "BRANCO" | 2 | 0 | "R\$ 84,58" | 12 | 16549879808 |
| "SP" | "360" | "1503" | 2 | 33 | 33 | 33 | "BRANCO" | 3 | 1 | "R\$ 61,95" | 4 | 13216549888 |
| "SP" | "200" | "2126" | 1 | 18 | 18 | 18 | "PARDO" | 2 | 1 | "R\$ 74,21" | 4 | 132465496 |
| "SP" | "980" | "3809" | 2 | 74 | 74 | 74 | "NEGRO" | 6 | 0 | "R\$ 46,26" | 8 | 123465752 |
| "SP" | "783" | "0904" | 1 | 13 | 13 | 13 | "PARDO" | 3 | . | "R\$ 65,94" | 7 | 45613551616 |
| "RJ" | "160" | "843" | 2 | 50 | 50 | 50 | "BRANCO" | 99 | 0 | "R\$ 41,30" | 1 | 16357494784 |
| "RJ" | "600" | "2303" | 2 | 64 | 64 | 64 | "NEGRO" | 4 | 1 | "R\$ 21,23" | 4 | 16549874688 |
| "RJ" | "501" | "5816" | 1 | 37 | 37 | 37 | "PARDO" | 2 | 3 | "R\$ 15,60" | 6 | 8822888448 |
| "RJ" | "203" | "2089" | 2 | 28 | 28 | 28 | "BRANCO" | 3 | 1 | "R\$ 34,18" | 0 | 135246544 |

```
***** SEÇÃO 3 - VISUALIZAR MICRODADOS NO STATA *****
```

```
*Exemplo 1: Listar todas as variáveis e observações
```

```
list
```

```
*Exemplo 2: Listar apenas as cinco primeiras observações do "banco 1"
```

```
list in 1/5
```

```
*Exemplo 3: Listar as observações de 5 a 10 das variáveis var1, var3 e var7.
```

```
list var1 var3 var7 in 5/10
```

```
*Exemplo 4: Utilizar o comando describe para descrever o "banco 1".
```

```
describe
```

```
*Exemplo 5: Descrever resumidamente o "banco 1" utilizando a opção short do comando describe.
```

```
describe, sh
```

```
*Exemplo 6: Apresentar uma tabela resumo das variáveis do "banco 1" utilizando o comando summarize.
```

```
summarize
```

```
***** SEÇÃO 4 - OPERAÇÕES BÁSICAS NO STATA *****
```

```
*Exemplo 7: Modificar o nome da variável var1 para renda
```

ren var1 uf

*Exemplo 8: Modificar o nome das variáveis var1 var2 var3 para renda idade uf.

```
ren (var2 var3 var4 var5 var6 var7 var8 var9 var10 var11 var12 var13) (renda renda_2 genero idade idade1 idade2  
raca componentes rm luz educ pib)
```

*Exemplo 9: Acrescentar uma legenda à variável renda indicando que esta renda se refere à renda do trabalho.

```
label variable renda "Renda do Trabalho"
```

*Exemplo 9a: Acrescentar legenda às demais variáveis conforme indicação abaixo:

```
label variable uf "Unidade da Federação"  
label variable renda_2 "Renda nominal familiar"  
label variable genero "Gênero do indivíduo"  
label variable idade "Idade do indivíduo"  
label variable idade1 "Idade do indivíduo"  
label variable idade2 "Idade do indivíduo"  
label variable raca "Raça do indivíduo"  
label variable componentes "Número de pessoas na família"  
label variable rm "Indicador de Região Metropolitana"  
label variable luz "Valor da conta de luz"  
label variable educ "Anos de estudo"  
label variable pib "PIB municipal"
```

* Exemplo 10: Vamos atribuir um rótulo para a variável genero com as categorias conforme mencionado acima, ou seja, o valor 1 representa os homens e o valor 2 as mulheres. Vamos chamar o rótulo de gen:

*1ª etapa: definição do rótulo

```
label define gen 1 "Homem" 2 "Mulher"
```

*2ª etapa: atribuir o rótulo a variável:

```
label values genero gen
```

* Exemplo 10a: Atribuir um rótulo à variável rm (Região Metropolitana). O Valor 0 indica que a pessoa não reside em Região Metropolitana e o valor 1 indica que ela reside.

```
label define metro 0 "Região Não Metropolitana" 1 "Região Metropolitana"
```

```
label values rm metro
```

* Exemplo 11: A variável renda no banco de dados está em formato texto, apesar de ter todos os seus caracteres

*numéricos. Esta a forma mais simples de utilizar o comando destring. Vamos utilizá-lo de duas maneiras;

* i) criando uma nova variável; e ii) substituindo o formato na própria variável.

```
destring renda, gen(renda1)
```

```
destring renda, replace
```

* Exemplo 12: A variável renda_2 possui algumas observações com caracteres não numéricos. No entanto, estes caracteres indicam que a pessoa não tem informação para esta variável, portanto, deve ser transformada em missing.

* Se o comando for utilizado conforme exemplo acima, o Stata retornará uma informação de erro "renda_2

*contains nonnumeric characters; no replace". Se acrescentarmos a opção force as observações com caracteres

*não numéricos serão transformadas em missing.

```
destring renda_2, replace
```

```
destring renda_2, replace force
```

* Exemplo 13: A variável luz que representa o valor pago na conta de luz começa com "R\$". Se tentarmos utilizar

*o comando destring como no exemplo 1, o Stata retornará a mensagem que não a variável possui caracteres

* não numéricos, e não irá alterar a variável. Se tentarmos utilizar a opção force, como existem caracteres não numéricos em todas as observações, todas serão transformadas em missing. Para retirar o "R\$" e manter corretamente os valores numéricos, utiliza-se a opção ignore("R\$"). No nosso exemplo, além deste problema, a variável não está no formato Americano, ou seja, a separação de casa decimal está com vírgula, portanto vamos acrescentar a opção dpcomma ao final do comando.

```
destring luz, replace ignore("R$") dpcomma
```

* Exemplo 14: A variável componentes possui informações com o valor 99. No entanto, este valor indica que esta informação não foi preenchida para a respectiva observação. Para evitar problemas com o uso deste valor de forma desavisada, vamos substituir este valor por missing.

```
replace componentes = . if componentes == 99
```

* Exemplo 15: A primeira observação da variável raca possui apenas a letra B como informação. Esta letra indica a raça Branco, no entanto, para padronizar as informações da variável, substituiremos a letra B pela

* palavra BRANCO utilizando o comando abaixo:

```
replace raca = "BRANCO" if raca == "B"
```

* Exemplo 16: Podemos usar o comando replace para transformar uma variável contínua em uma variável categórica. Imagine que precisemos transformar a variável de idade em uma variável representando alguns grupos de idade. Quem tiver 18 anos ou menos de idade irá participar do grupo 1, quem tiver entre 19 e 30 (inclusive) faz parte do grupo 2, entre 31 e 50 (inclusive) grupo3, entre 51 e 65 (inclusive) grupo 4 e mais de 65 grupo 5. Para isso utilizaremos cinco linhas de comando conforme abaixo:

```
replace idade2 = 1 if idade2 >=0 & idade2 <= 18  
replace idade2 = 2 if idade2 >=19 & idade2 <= 30  
replace idade2 = 3 if idade2 >=31 & idade2 <= 50  
replace idade2 = 4 if idade2 >=51 & idade2 <= 65  
replace idade2 = 5 if idade2 >=66
```

* Exemplo 17: A variável rm (Região Metropolitana) deveria possuir apenas os códigos 0 e 1, indicando se o indivíduo vive em uma Região Metropolitana ou não. No entanto, a variável possui uma observação com o código 3 e uma com o código 99. Estes valores indicam erro na resposta da variável, e devem ser substituídos por missing.

```
recode rm (3 99 = .)
```

* Exemplo 18: Repetir o exemplo 16 do comando replace. Com o comando recode, é possível fazer o mesmo que foi feito com o comando replace em apenas uma linha de comando. Como a variável idade já foi modificada, realize o procedimento na variável idade1.

```
recode idade1 (0/18 = 1) (19/30 = 2) (31/50 = 3) (51/65 = 4) (66/max = 5)
```

* Exemplo 19: Utilizando as variáveis já existentes no banco de dados, vamos criar uma variável que representa a idade ao quadrado, outra que representa o logaritmo da renda (natural e na base 10) e uma variável de renda per capita que combina informações da renda da família e do número de componentes da família.

* Idade ao quadrado

```
gen ida2 = idade^2
```

* Logaritmo

```
gen lnrenda = ln(renda)
```

```
gen logrenda = log10(renda)
```

* Renda per capita

```
gen per_capita = renda_2 / componentes
```

*Exemplo 20: Criar uma variável constante com a renda máxima da população e outra variável com a renda máxima por raça do indivíduo.
egen renda_max = max(renda)

by raza, sort: egen renda_raca = max(renda)

* Exemplo 21: Criar uma nova variável representando grupos de idade. Quem tiver 18 anos ou menos de idade irá participar do grupo 1, quem tiver entre 19 e 30 (inclusive) faz parte do grupo 2, entre 31 e 50 (inclusive) grupo3, entre 51 e 65 (inclusive) grupo 4 e mais de 65 grupo 5.
recode idade (0/18 = 1) (19/30 = 2) (31/50 = 3) (51/65 = 4) (66/max = 5), gen(idade_grupo)

*Exemplo 22: Criar dummies de gênero usando o comando gen conforme já vimos anteriormente. Vamos criar duas variáveis, uma para indicar se a pessoa é do sexo masculino e outra para indicar que a pessoa é do sexo feminino.

```
gen masc = 1 if genero == 1  
replace masc = 0 if masc ==.
```

```
gen fem = 1 if genero == 2  
replace fem = 0 if fem ==.
```

*Exemplo 23: Criar uma variável para indicar se o individuo é do sexo masculino com o uso do comando recode.
recode genero (2=0), gen(masculino)

*Exemplo 24: Usar o comando gen para criar uma variável indicando se a pessoa tem mais de 20 anos.
gen adulto=idade>=20

*Exemplo 25: Utilizar o comando tab com a opção gen para criar uma dummy para cada categoria das variáveis genero e raza. Renomear as variáveis draca1, draca2 e draca3 para branco, negro e pardo, respectivamente.
tab genero, g(dsexo)
tab raza, g(draca)

```
ren draca1 branco  
ren draca2 negro  
ren draca3 pardo
```

***** SEÇÃO 5 - ESTATÍSTICAS DESCRITIVAS *****

* Exemplo 26: Vamos fazer uma tabela de frequência simples para a variável rm (Indicadora de Região Metropolitana). Após executar o comando, aparecerá na janela Results uma tabela conforme a tabela abaixo, com informações sobre a frequência e o percentual de cada categoria e o percentual acumulado, além da frequência total.

```
tab rm
```

* Exemplo 27: Repetir o exemplo 26 acrescentando os valores missing com sendo uma categoria.

```
tab rm,m
```

* Exemplo 28: Criar uma tabela de frequências para as variáveis rm e raza em um único comando. Serão apresentadas duas tabelas como se o comando tabulate visto anteriormente fosse executado duas vezes na sequência.

```
tab1 rm raza
```

* Exemplo 29: Fazer uma tabela cruzada entre as variáveis rm e raza. A tabela resultante do comando mostra em cada célula o número de observações que atendem a duas características simultaneamente. Ou seja, existem 3 pessoas da raça Branca vivendo em Regiões Metropolitanas, 1 pessoa da raça Negra vivendo em Regiões Metropolitanas e assim sucessivamente.

```
tab rm raza
```

* Exemplo 30: Repetir o comando utilizado no exemplo 29 acrescentando a opção col que calcula os percentuais das células em relação ao total da coluna. Neste exemplo, será calculado o percentual de pessoas brancas que vivem em Regiões Metropolitanas em relação ao total de pessoas brancas, o percentual de pessoas negras que vivem em Regiões Metropolitanas em relação ao total de pessoas negras e assim sucessivamente.

```
tab rm raza, col
```

* Exemplo 31: Repetir o comando utilizado no exemplo 29 acrescentando a opção row que calcula os percentuais das células em relação ao total da linha. Neste exemplo, será calculado o percentual de pessoas brancas que vivem em Regiões Metropolitanas em relação ao total de pessoas que vivem nas Regiões Metropolitanas, o percentual de pessoas negras que vivem em Regiões Não Metropolitanas em relação ao total de pessoas que vivem em Regiões Não Metropolitanas e assim sucessivamente.

```
tab rm raza, row
```

* Exemplo 32: Repetir o comando utilizado no exemplo 29 acrescentando a opção cel que calcula os percentuais das células em relação ao número total de observações. Neste exemplo, será calculado o percentual de pessoas brancas que vivem em Regiões Metropolitanas em relação ao total de observações não missing nas duas variáveis utilizadas, o percentual de pessoas negras que vivem em Regiões Não Metropolitanas em relação ao total de observações não missing nas duas variáveis utilizadas e assim sucessivamente.

```
tab rm raza, cel
```

* Exemplo 33: Criar tabelas cruzadas para a combinação das variáveis rm, raza e genero.

```
tab2 rm raza genero
```

* Exemplo 34: Elaborar uma tabela de frequência de raça para mulheres (variável genero igual a 2).

```
tab raza if genero ==2
```

* Exemplo 35: Elaborar uma tabela de frequência de raça para homens (variável genero diferente de 2).

```
tab raza if genero !=2
```

* Exemplo 36: Elaborar uma tabela de frequência de raça para pessoas com mais de 5 anos de estudo.

```
tab raza if educ>5
```

* Exemplo 37: Elaborar uma tabela de frequência de raça para pessoas com 5 anos ou mais de estudo.

```
tab raza if educ>=5
```

* Exemplo 38: Elaborar uma tabela de frequência de gênero para brancos e pardos.

```
tab genero if raza == "BRANCO" | raza == "PARDO"
```

* Exemplo 39: Elaborar uma tabela de frequência de gênero para brancos e pardos com mais de 5 anos de estudo.

```
tab genero if (raca == "BRANCO" | raza == "PARDO") & educ > 5
```

* Exemplo 40: Usar o comando summarize para calcular a média, desvio padrão, mínimo e máximo dos valores da variável renda.

```
sum renda
```

* Exemplo 41: Usar a opção detail do comando summarize.
sum renda, det

* Exemplo 42: Elaborar uma tabela com a média, o total de observações diferentes de missing, os valores máximo e mínimo da variável renda.
tabstat renda, stats(mean count max min)

* Exemplo 43: Executar o mesmo comando do exemplo 42 para a variável renda e variável educ.
tabstat renda educ, stats(mean count max min)

* Exemplo 44: Inverter a forma de apresentação da variável com as estatísticas. Para ilustrar vamos calcular a média, soma e total de observações diferentes de missing para a variável renda. Se a opção não for utilizada, a variável ficará na linha enquanto que as estatísticas são apresentadas nas *colunas. Acrescentando a opção columns(variables), esta apresentação fica invertida.
tabstat renda, stats (n mean sum)

tabstat renda, stats (n mean sum) columns(variables)

* Exemplo 45: Para ilustrar o problema da apresentação dos resultados para variáveis muito grandes, vamos calcular a média e a soma da variável pib.
tabstat pib, stats (mean sum)

* Exemplo 46: Conforme visto no exemplo 45, se o formato dos resultados não for controlado, o resultado será apresentado em notação científica, impossibilitando conhecer o verdadeiro valor do resultado. Vamos agora impor um formato para os resultados. Queremos que o resultado seja apresentado com 15 algarismos, sendo 2 casas decimais. Ao manter a forma de apresentação entre variáveis e estatísticas inalteradas, os resultados da média e do somatório ficarão misturados. Neste caso, é interessante usar a opção columns(variables) vista no exemplo 45.
tabstat pib, stats (mean sum) format (%15.2f)

tabstat pib, stats (mean sum) format (%15.2f) columns(variables)

* Exemplo 47: Repetir o exemplo 46 omitindo os valores depois da vírgula.
tabstat pib, stats (mean sum) format (%15.0f) columns(variables)

* Exemplo 48: Calcular a correlação entre a variável educ e renda.
cor educ renda

* Exemplo 49: Calcular a matriz de correlação entre as variáveis educ, renda e luz.
cor educ renda educ

* Exemplo 50: Calcular a média e o desvio padrão da variável educ para cada categoria da variável genero.
tab genero, sum(educ)

* Exemplo 51: Calcular a média, o somatório, o desvio padrão e o total de observações diferentes de missing da variável renda para cada categoria da variável raza.
table raza, contents (freq mean renda sum renda sd renda count renda)

* Exemplo 52: Transformar o banco de dados "banco1" em um banco de dados agregado por Unidades da Federação contendo as seguintes informações:
collapse (mean) renda per_capita idade adulto branco negro pardo (sum) total_branco=branco total_negro=negro total_pardo=pardo (count) pop=idade, by(uf)

***** SEÇÃO 7 - COMBINAR DIVERSOS BANCOS DE DADOS *****

** Executar as linhas de 273 a 291 para imputar o banco de dados para se usado como exemplo na Seção 7

```
clear
input str2 var1 str4 var2 str4 var3 var4 var5 str6 var6

"MG" "521" "n/d" 1 15 "B"
"MG" "123" "1360" 1 51 "BRANCO"
"MG" "150" "543" 2 32 "PARDO"
"MG" "1200" "1200" 2 41 "NEGRO"
"MG" "410" "1590" 2 20 "NEGRO"
"MG" "851" "2806" 1 13 "PARDO"
"SP" "1250" "*" 2 50 "BRANCO"
"SP" "360" "1503" 2 33 "BRANCO"
"SP" "200" "2126" 1 18 "PARDO"
"SP" "980" "3809" 2 74 "NEGRO"
"SP" "783" "0904" 1 13 "PARDO"
"RJ" "160" "843" 2 50 "BRANCO"
"RJ" "600" "2303" 2 64 "NEGRO"
"RJ" "501" "5816" 1 37 "PARDO"
"RJ" "203" "2089" 2 28 "BRANCO"
end
```

* Gerar uma variável sequencial que será utilizada como identificadora do indivíduo

```
gen id = _n
```

* Salvar o banco de dados

```
save "banco2", replace
```

*Exemplo 53: Para ilustrar o uso do comando append no caso em que as mesmas variáveis existem nos dois

*bancos, vamos imputar o banco de dados abaixo e logo após adicionar as observações do banco de dados

*"banco2".

```
clear
input id str2 var1 str4 var2 str4 var3 var4 var5 str6 var6

16 "BA" "421" "942" 1 16 "PARDO"
17 "BA" "651" "1190" 1 26 "BRANCO"
18 "CE" "113" "597" 2 33 "PARDO"
19 "CE" "1150" "2540" 2 12 "NEGRO"
20 "PI" "478" "1590" 2 87 "NEGRO"
21 "MT" "531" "1056" 1 56 "PARDO"
22 "MT" "1280" "987" 2 52 "PARDO"
23 "AM" "384" "2003" 2 33 "BRANCO"
24 "RR" "684" "1326" 1 18 "PARDO"
25 "RR" "320" "3009" 2 65 "NEGRO"
26 "MS" "197" "0504" 1 26 "BRANCO"
27 "RR" "130" "743" 2 76 "BRANCO"
28 "PI" "971" "1603" 2 12 "NEGRO"
end
```

```
append using "banco2.dta"
```

* Exemplo 54: Neste exemplo, vamos unir dois bancos de dados novamente, mas agora com um dos bancos *contendo três variáveis a menos. Para visualizar este procedimento, impute o banco de dados conforme *programação abaixo e depois realize o append com o banco de dados "banco2".

```
clear  
input str2 var1 str4 var2 str4 var3 var4
```

```
"BA" "421" "942" 1  
"BA" "651" "1190" 1  
"CE" "113" "597" 2  
"CE" "1150" "2540" 2  
"PI" "478" "1590" 2  
"MT" "531" "1056" 1  
"MT" "1280" "987" 2  
"AM" "384" "2003" 2  
"RR" "684" "1326" 1  
"RR" "320" "3009" 2  
"MS" "197" "0504" 1  
"RR" "130" "743" 2  
"PI" "971" "1603" 2  
end
```

```
append using "banco2.dta"
```

* Exemplo 55: Para ilustrar o comando merge na opção one-to-one vamos imputar o banco de dados abaixo e *realizar o merge com o banco de dados "banco2". Neste caso faremos um merge com correspondência um para *um, cuja variável de ligação é a id (identificadora do indivíduo):

```
clear  
input id str10 ativ status
```

```
1 "Primario" 1  
2 "Terciario" 0  
3 "Primario" 0  
4 "Secundario" 1  
5 "Secundario" 0  
6 "Terciario" 1  
7 "Terciario" 1  
8 "Secundario" 0  
9 "Secundario" 1  
10 "Secundario" 1  
11 "Terciario" 0  
12 "Terciario" 0  
13 "Terciario" 1  
14 "Primario" 1  
15 "Secundario" 0  
end
```

```
merge 1:1 id using "banco2.dta"
```

* Exemplo 56: Neste exemplo iremos acrescentar variáveis referentes a Estados ao "banco2" que tem as *informações individuais. Neste caso a variável de ligação será a var1, que contém a sigla do Estado.

- * Este vai ser um caso de várias observações de uma variável de ligação para uma. Vamos imputar o banco de dados abaixo e depois fazer o merge com o "banco2".
- * Como o banco de dados de Estado que estará aberto no Stata, a opção one-to-many será utilizada.

```
clear
input str2 var1 str14 uf_nome codigo
```

```
"MG" "Minas Gerais"      31
"SP" "São Paulo"         35
"RJ" "Rio de Janeiro"    33
end
```

```
merge 1:m var1 using "banco2.dta"
```

***** SEÇÃO 8 – UMA APLICAÇÃO À PNAD *****

* Utilizar o banco de dados da PNAD 2008 para esta seção
use "PNAD_2008_amostra.dta", clear

* Exercício 1: Mantenha no banco de dados apenas as seguintes variáveis, UF V0302 V8005 V0404 V4803 V4718 V4729 V4722 V4724 V0501 V0601 V0701 V4704 V480, que serão utilizadas nos próximos exercícios.
keep UF V0302 V8005 V0404 V4803 V4718 V4729 V4722 V4724 V0501 V0601 V0701 V4704 V4805

* Exercício 2: Renomear as variáveis UF V0302 V8005 V0404 V4803 V4718 V4729 V4722 V4724 V0501 V0601 V0701 V4704 V4805 para uf sexo idade raca educ renda peso ren_fam componentes migra analfabeto infantil ativ ocup:

```
ren (UF V0302 V8005 V0404 V4803 V4718 V4729 V4722 V4724 V0501 V0601 V0701 V4704 V4805) (uf sexo idade raca educ renda peso ren_fam componentes migra analfabeto infantil ativ ocup)
```

* Exercício 3: Acrescentar rótulos para as variáveis sexo, raca, migra, analfabeto, infantil, ativ e ocup de acordo com o dicionário de variáveis da PNAD. Ao final utilizar o comando numlabel, add para deixar junto com os rótulos os valores das variáveis:

```
label define codsexo 2 "Homem" 4 "Mulher"
label values sexo codsexo
```

```
label define codraca 2 "Branca" 4 "Preta" 6 "Amarela" 8 "Parda" 0 "Indígena" 9 "Sem declaração"
label values raca codraca
```

```
label define migrante 1 "Não migrante" 3 "Migrante"
label values migra migrante
```

```
label define analf 1 "Sabe ler" 3 "Analfabeto"
label values analfabeto analf
```

```
label define trab 1 "Trabalha" 3 "Não trabalha"
label values infantil trab
```

```
label define atividade 1 "Economicamente ativo" 2 "Não economicamente ativo"
```

label values ativ atividade

label define ocupação 1 "Ocupado" 2 "Desocupado"

label values ocup ocupacao

numlabel, add

* Exercício 4: Gerar variáveis novas com as seguintes informações: idade ao quadrado, logaritmo natural da renda

*e renda familiar per capita:

* Idade ao quadrado

gen ida2 = idade^2

* Logaritmo natural da renda

gen lnrenda = ln(renda)

* Renda familiar per capita

gen per_capita = ren_fam / componentes

* Exercício 5: A variável renda na PNAD possui o valor 999999999999 para as pessoas que não responderam a

*esta pergunta na pesquisa. Este valor influencia nas estatísticas da variável renda, puxando os resultados para

*cima. Substitua estes valores por missing. Calcule a renda média antes e depois da substituição para comparar

*os resultados.

sum renda

replace renda = . if renda == 999999999999

sum renda

* Exercício 6: Veja no dicionário de variáveis as categorias da variável de anos de estudo (V4803). Note que o

*código 1 se refere às pessoas sem instrução, o código 2 às pessoas com 1 ano de escolaridade e assim por

*diante. Desta maneira, calcular informações relativas aos anos de estudo estarão enviesadas para cima. Além

*disto, a categoria 17 se refere às pessoas que não declaram a escolaridade. Corrija esta variável para que os

*valores das categorias se refiram corretamente aos anos de estudo. Calcule a média antes e depois da correção

*dos valores.

sum educ

replace educ = . if educ == 17

replace educ = educ - 1

sum educ

* Exercício 7: Calcular a frequência relativa para cada categoria da variável raca sem considerar o peso amostral e

*considerando o peso amostral.

tab raca

tab raca [w=peso]

* Exercício 8: Calcular a frequência relativa da variável raca apenas para as mulheres (variável sexo igual a 2)

*considerando os pesos amostrais.

tab raca [w=peso] if sexo ==2

* Exercício 9: Calcular a frequência relativa de cada categoria da variável sexo para as pessoas de cor branca (raca = 2) ou amarela (raca=4) com mais de 5 anos de estudo considerando os pesos amostrais.

```
tab sexo [w=peso] if (raca == 2 | raca == 4) & educ > 5
```

* Exercício 10: Construir uma matriz de correlação com as variáveis educ, renda e ren_fam, considerando os pesos amostrais.

```
cor educ renda ren_fam [w=peso]
```

* Exercício 11: Construir uma tabela com o total e percentual de pessoas com mais de 5 anos de escolaridade para cada subgrupo de sexo e raça considerando os pesos amostrais.

```
tab raca sexo [w=peso] if educ>5, cel
```

* Exercício 12: AGREGAR BANCO DE DADOS: construir um banco de dados agregados por Estado, com o total da população, a população entre 18 e 65 anos (inclusive) e sua média de anos de estudo, o percentual de brancos, o total de pessoas com rendimentos não nulos, o percentual de trabalhadores com mais de 10 anos de estudo em relação ao total de trabalhadores, a renda média do trabalho e o percentual da população com renda familiar per capita inferior a R\$ 100,00.

* Variável de total de população

```
gen pop = 1
```

* População entre 18 e 65 anos (inclusive)

```
gen pop_18_65 =idade>=18 & idade <=65
```

* Média de anos de estudo para as pessoas entre 18 e 65 anos (inclusive)

```
replace educ = . if idade <18 | idade >65
```

* População branca

```
gen branco=raca==2
```

* Percentual de trabalhadores com mais de 10 anos de escolaridade em relação ao total de trabalhadores

```
gen trab_10 =educ>10
```

```
replace trab_10 = . if renda == .
```

* Percentual da população com renda per capita inferior a R\$ 100,00

```
gen pobres=ren_fam<100
```

```
collapse (mean) renda educ idade branco trab_10 pobres (sum) pop pop_18_65 (count) trabalhador=renda [pw=peso], by(uf)
```