

Notas metodológicas - ENANI-2019

Informações para acesso e utilização dos microdados

O presente documento tem por objetivo fornecer ao usuário orientações para a utilização dos microdados do ENANI-2019. Encontram-se aqui documentadas informações importantes para o acesso, manipulação dos dados e para a produção de estimativas sobre a população de estudo.

Leitura dos microdados

Os microdados estão disponibilizados em formato *.rds*. Os arquivos em formato *.rds* devem ser lidos preferencialmente através da linguagem de programação R (cran.r-project.org). Estes arquivos já contêm os microdados e as informações do desenho amostral, assim como a calibração.

Os arquivos em podem ser exportados para outro *software* estatístico, porém, antes do uso de qualquer um dos bancos de dados selecionado pelo usuário deverá obrigatoriamente declarar o objeto de desenho e submeter à aplicação da calibração dos pesos amostrais básicos usando pós-estratificação por região, sexo e idade em anos completos, considerando 3 estágios de agregação e a variável posestrato. Ademais deve-se considerar nessa análise o total populacional por posestrato.

Parâmetro	Variável(is)
Strata	estrato_sel_anon
Identificadores (estrutura de conglomerados)	id_upa_anon+setor_anon+a06_domicilio
Weights	peso_crianca ou peso_crianca_y_XX

Após a declaração do objeto, esse deve ser submetido à calibração dos pesos, considerando 3 estágios de agregação, a variável posestrato e o total populacional por posestrato.

IMPORTANTE: As estimativas obtidas em análises usando softwares diferentes do R podem ser diferentes das estimativas publicadas pela equipe ENANI-2019, tendo em vista a limitação de parâmetros usados na construção do objeto de desenho em outros *softwares*.

A leitura dos objetos de desenho no formato .rds pode ser feita através da função `readRDS` do pacote `base`.

Exemplo:

```
> setwd("...") #Diretório que contém os microdados  
  
> dados.enani2019 <- base::readRDS(file = "data_crianca_calib_anon.rds")
```

O **banco principal** do ENANI-2019 é o `"data_crianca_calib_anon.rds"` (n=14.558). Neste banco, as variáveis coletadas (de acordo com o questionário), bem como as principais variáveis derivadas utilizadas para a construção dos indicadores descritos nos relatórios produzidos pelo ENANI (disponíveis em <https://enani.nutricao.ufrj.br/>) passaram por um processo de imputação, não apresentando, assim, valores faltantes (com exceção da idade da mãe/responsável). Dois métodos foram empregados para a imputação automática: *"hotdeck"* e *"vizinho mais próximo"*. Mais detalhes estão disponíveis no Relatório 1 (Aspectos Metodológicos) e na Nota técnica de imputação.

Para os dados de biomarcadores sanguíneos ocorreu não resposta total (falta de medidas em todos os biomarcadores analisados) e não resposta parcial (falta de medidas para algum subconjunto dos biomarcadores). Considerando a natureza dessas variáveis, decidiu-se não compensar a não-resposta neste conjunto de dados utilizando técnicas de imputação. Então, para tal, foram feitos ajustes dos pesos básicos das crianças.

Dessa forma, foram gerados outros 25 arquivos de microdados considerando diferentes subconjuntos de biomarcadores sanguíneos. Ressalta-se que todas as variáveis presentes no banco principal da pesquisa também estão reproduzidas em cada um desses outros bancos. Contudo, nestes estão selecionadas apenas as crianças com resultados para o(s) referido(s) biomarcador(es). Os subgrupos, e respectivos arquivos, estão detalhados no quadro a seguir:

Agrupamento de biomarcadores sanguíneos		n útil	Arquivo
1a	Hemoglobina, PCR, ferritina ^a	7.473	data_bioq_calib_anon_1a.rds
1b	Hemoglobina, PCR, ferritina ^b	7.293	data_bioq_calib_anon_1b.rds
1c	PCR e ferritina ^a	7.931	data_bioq_calib_anon_1c.rds
2	Hemoglobina e vitamina A	7.716	data_bioq_calib_anon_2.rds
3	Hemoglobina, vitamina A, vitamina D, e zinco	7.037	data_bioq_calib_anon_3.rds
4	PCR e zinco	7.659	data_bioq_calib_anon_4.rds
5	PCR e vitamina A	7.944	data_bioq_calib_anon_5.rds
6	PCR e selênio	7.660	data_bioq_calib_anon_6.rds
7	Ácido fólico e vitamina B12	7.418	data_bioq_calib_anon_7.rds
8	PCR, zinco, selênio, vitamina A e vitamina E	7.506	data_bioq_calib_anon_8.rds
9a	Ácido fólico, hemoglobina, ferritina, PCR, vitaminas B12, D, A, B1 e B6, zinco e selênio ^a	6.023	data_bioq_calib_anon_9a.rds
9b	Ácido fólico, hemoglobina, ferritina, PCR, vitaminas B12, D, A, B1 e B6, zinco e selênio ^c	6.532	data_bioq_calib_anon_9b.rds
10	Hemoglobina, ácido fólico, vitamina B12, ferritina e PCR ^a	6.526	data_bioq_calib_anon_10.rds
11	Hemoglobina, ácido fólico, vitamina B12, ferritina e PCR ^d	6.556	data_bioq_calib_anon_11.rds
12	Hemoglobina	8.187	data_bioq_calib_anon_12.rds
13	Ácido fólico	7.435	data_bioq_calib_anon_13.rds
14	Ferritina	8.156	data_bioq_calib_anon_14.rds
15	Vitamina B12	7.441	data_bioq_calib_anon_15.rds
16	Vitamina D	8.217	data_bioq_calib_anon_16.rds
17	Vitamina A	8.393	data_bioq_calib_anon_17.rds
18	Vitamina B1	8.328	data_bioq_calib_anon_18.rds
19	Vitamina B6	8.329	data_bioq_calib_anon_19.rds
20	Vitamina E	8.388	data_bioq_calib_anon_20.rds
21	Zinco	8.052	data_bioq_calib_anon_21.rds
22	Selênio	8.054	data_bioq_calib_anon_22.rds

Notas: Número total de crianças na amostra elegíveis para coleta de sangue (Brasil: n=12.598)

PCR = proteína C-reativa

^a 1a, 1c, 9a e 10: parâmetros coletados na mesma data.

^b 1b: Ferritina e PCR na mesma data; ferritina e hemoglobina na mesma data ou com intervalo de até 30 dias.

^c 9b: coletados em datas diferentes.

^d 11: coletados na mesma data ou com intervalo de até 30 dias.

Dessa forma, se o objetivo é realizar uma análise sobre, por exemplo, PCR e ferritina, deve-se utilizar o banco do grupo 1c (“data_bioq_calib_anon_1c.rds”). A leitura desses bancos de agrupamentos bioquímicos deve ser da mesma forma como descrita para o banco principal. Para trabalhar com algum subgrupo para o qual não tenham sido criados pesos específicos, ou seja, que não esteja descrito no quadro acima, indica-se a utilização do subgrupo com maior tamanho amostral que contenha a combinação pretendida.

Para identificar os indivíduos que fizeram parte de cada subgrupo, basta selecionar/filtrar no banco do subgrupo, através da variável indicativa do grupo (exemplo: a variável “grupo1c” indicará se o indivíduo faz parte do grupo 1c [ou seja, tem resultados para os parâmetros PCR e ferritina]).

Exemplo:

```
> setwd("...") #Diretório que contém os microdados
> data_crianca_calib_1c                                     <-
base::readRDS(file="data_bioq_calib_anon_1c.rds") %>%
  srvyr::filter(grupo1c == "Sim")
```

#Obs: É necessária a instalação prévia do pacote *srvyr*. Para instalar, utilizar o seguinte comando:

```
> install.packages("srvyr")
```

Análise de dados

Dada a amostragem complexa do ENANI-2019, para a análise dos dados e cálculo das estimativas, deve-se levar em consideração as informações do plano amostral. Cabe mencionar que, além do ajuste de não resposta, os pesos amostrais passaram por um processo de calibração para compatibilizar com os totais das estimativas populacionais de forma a recompor a estrutura demográfica dentro da pesquisa (considerando-se o cruzamento entre os totais conhecidos de macrorregião, sexo e faixa etária). Desta forma, o menor nível de desagregação geográfica da amostra é macrorregião.

Orienta-se que, no uso da linguagem de programação R para as análises, sejam utilizados os pacotes *survey* e *srvyr*, criados especificamente para análise de dados provenientes de pesquisas com estes tipos de planos amostrais. Outros pacotes R são também importantes para as análises, sendo necessária a instalação do conjunto de pacotes *tidyverse*.

No pacote *survey*, a função para a estimação de totais populacionais é a *svytotal*. Com a função *svymean* é possível estimar a média de uma variável numérica/quantitativa e também estimar frequências de variáveis categorias/qualitativas. Medianas e quantis de variáveis numéricas/quantitativas são estimados através da função *svyquantile*.

O pacote *srvyr* permite a utilização de funções análogas às apresentadas em combinação com as funções pertencentes ao pacote *tidyverse*. A função para a estimação de totais populacionais é a *survey_total*. Com a função *survey_mean* é possível estimar a média de uma variável numérica/quantitativa e também estimar frequências de variáveis categóricas/qualitativas. Medianas e quantis de variáveis numéricas/quantitativas são estimados através da função *survey_quantile*. Destaca-se que, para este pacote, estas funções devem estar sempre sendo usadas conjuntamente com a função *summarise*.

Toda a informação do desenho amostral é carregada quando da especificação dos argumentos *design* ou *data* nas funções/scripts dos respectivos pacotes acima referidos.

Exemplo:

Para variáveis quantitativas:

```
> survey::svymean(~variavelnumerica,  
                  design=dados.enani2019, na.rm=TRUE)
```

ou

```
> dados.enani2019 %>%  
  srvyr::summarise(survey_mean(variavelnumerica,  
                               na.rm=TRUE))
```

Para variáveis qualitativas:

```
> survey::svymean(~variavelcategorica,  
                  design=dados.enani2019, na.rm=TRUE)
```

ou

```
> dados.enani2019%>%  
  group_by(variavelcategorica) %>%  
  summarise(survey_mean(na.rm=TRUE))
```

Abaixo seguem exemplos para o caso de se desejar realizar análises de uma variável segundo categorias de um estratificador (variável categórica/qualitativa).

Exemplo:

Para variáveis quantitativas:

```
> survey::svyby(~variavelnumerica, ~estratificador  
               design=dados.enani2019,  
               na.rm=TRUE,  
               svymean)
```

ou

```
> dados.enani2019 %>%  
  group_by(estratificador) %>%  
  svyr::summarise(survey_mean(variavelnumerica,  
                             na.rm=TRUE) )
```

Para variáveis qualitativas:

```
> survey::svyby(~variavelcategorica, ~estratificador  
               design=dados.enani2019,  
               na.rm=TRUE,  
               svymean)
```

ou

```
> dados.enani2019%>%  
  group_by(variavelcategorica, estratificador) %>%  
  summarise(survey_mean(na.rm=TRUE) )
```

A interpretação dos resultados do ENANI-2019 deve levar em consideração a natureza amostral da pesquisa (desenho complexo de amostragem). Sendo assim, as estimativas produzidas devem ser interpretadas acompanhadas de seus respectivos intervalos de confiança (espera-se que esses intervalos incluam o resultado do fenômeno na população, com o nível de confiança especificado, em geral, de 95%).

Para a estimação dos intervalos de confiança, na utilização da função `survey_mean` deve-se adicionar o argumento `"vartype="ci"`. Para a utilização com a função `svymean`, uma função adicional deve ser adicionada `confint` (do pacote `stats`). Abaixo, seguem exemplos.

Exemplo:

Para variáveis quantitativas:

```
> stats::confint(survey::svymean(~variavelnumerica,  
                                design=dados.enani2019, na.rm=TRUE))
```

ou

```
> dados.enani2019 %>%  
  srvyr::summarise(survey_mean(variavelnumerica,  
                                na.rm=TRUE, vartype="ci"))
```

Para variáveis qualitativas:

```
> stats::confint(survey::svymean(~variavelcategorica,  
                                design=dados.enani2019, na.rm=TRUE))
```

ou

```
> dados.enani2019%>%  
  group_by(variavelcategorica) %>%  
  summarise(survey_mean(na.rm=TRUE, vartype="ci"))
```

Como a precisão da estimativa é inversamente proporcional à amplitude do intervalo de confiança (à medida que o intervalo de confiança aumenta, a precisão da estimativa diminui), ressalta-se a importância de se avaliar também o coeficiente de variação (CV) das estimativas. O CV é uma medida de dispersão obtida pela razão entre o erro padrão e o valor estimado do indicador, multiplicado por 100. Estimativas que apresentem CV elevado podem indicar que a amostra não tem tamanho suficiente para que se faça a estimação com um grau de precisão aceitável. Sugere-se o ponto de corte do $CV \leq 30\%$ como nível de precisão

adequado para as estimativas. Dessa forma, resultados com CV superior a 30% devem ser interpretados com cautela.

Para cálculo do CV, no R, pode-se proceder da seguinte forma:

Exemplo:

```
> cv(survey::svymean(~variavelnumerica,  
                      design=dados.enani2019,  
                      na.rm=TRUE))
```

ou

```
> dados.enani2019 %>%  
  srvyr::summarise(survey_mean(variavelnumerica,  
                              na.rm=TRUE,  
                              vartype="cv"))
```

Recomenda-se que, para as análises com os bancos de parâmetros bioquímicos, além de utilizar o banco do subgrupo apropriado, deve-se realizar um filtro (subset) pela variável indicativa de grupo. Por exemplo, no caso de se desejar a utilização do subgrupo 1c, os seguintes *scripts* poderiam ser utilizados para se obter uma estimativa pontual:

Exemplo:

Leitura do banco:

```
> dados.grupo1c <- base::readRDS(file =  
  "data_bioq_calib_anon_1c.rds")
```

Para variáveis quantitativas:

```
> survey::svymean(~variavelnumerica,  
                  design=subset(dados.grupo1c, grupo1c=="Sim"),  
                  na.rm=TRUE)
```


ou

```
> dados.grupo1c %>%
  subset(grupo1c=="Sim") %>%
  srvyr::summarise(survey_mean(variavelnumerica,
                                na.rm=TRUE))
```

Para variáveis qualitativas:

```
> survey::svymean(~variavelcategorica,
  design= subset(dados.grupo1c, grupo1c=="Sim"),
  na.rm=TRUE)
```

ou

```
> dados.grupo1c %>%
  subset(grupo1c=="Sim") %>%
  group_by(variavelcategorica) %>%
  summarise(survey_mean(na.rm=TRUE))
```

Para a construção de tabelas sugerem-se os comandos abaixo:

Exemplo:

Leitura do banco:

```
> dados.enani2019 <- base::readRDS(file =
"data_crianca_calib_anon.rds")
```

tabela<-

```
> dados.enani2019%>%
  srvyr::group_by(variavelcategoricax) %>%
  srvyr::summarise(prev = survey_mean(variavelcategoricay, na.rm =
TRUE, vartype = c("ci", "cv"))) %>%
  dplyr::mutate(prev = round(prev, digits = 2),
```

```
      prev_low = round(prev_low, digits = 2),  
      prev_upp = round(prev_upp, digits = 2)) %>%  
  
    dplyr::ungroup() %>%  
  
    cbind(Estratificador = c("xxx", "xxx", "xxx"), indicador="Nome do  
indicador aqui", .) %>%  
  
    dplyr::select(variavelcategoricax,   Estratificador,   indicador,  
prev, prev_low, prev_upp)  
  
tabela  
  
tabela1 = as.data.frame(tabela)
```

Para a construção de gráficos a partir das tabelas criadas sugerem-se os seguintes comandos:

Exemplo:

```
grafico <-  
tabela %>%  
  
  #dplyr::mutate(ordem = sequence(n())) %>%  
  
  ggplot(aes(x=indicador, y=prev)) +  
  
  geom_col(position = position_dodge(width=0.8), width = 0.9) +  
  
  geom_errorbar(aes(ymin = prev_low,  
                    ymax = prev_upp),  
                position = position_dodge(width=0.8),
```

```

width = 0.1)+

geom_text(family="Times New Roman",

          position=position_dodge(width=0.8),

          vjust=-1.0,

          hjust = 0.5,

          size=((9)/(14/5)),

          aes(label=formatC(round(prev,1), format='f', digits=1),
y = prev_upp), angle = 0)+

xlab("") +

ylab("Prevalence (%)") +

labs(title = expression ('Seu indicador aqui '^a)) +

scale_y_continuous(limits = c(0,100), breaks = seq(0, 100, by =
20)) +

guides(fill = guide_legend(nrow = 1, override.aes = list(shape =
22, size = 6.5)))

grafico

```

População de pesquisa / unidade de análise

A **população de pesquisa** do ENANI-2019 foi definida pelo conjunto de domicílios particulares permanentes onde residisse pelo menos uma criança com menos de 5 anos completos de idade, localizados em todo o território nacional. Não fizeram parte da população de pesquisa domicílios com crianças: (1) indígenas que viviam em aldeias; (2) estrangeiras residentes em domicílios onde não se falavam a língua portuguesa; (3) com alguma condição que as incapacitava à medição antropométrica; e (4) moradoras em domicílios coletivos (hotéis, pensões e similares, orfanatos e hospitais).

Desta forma, embora a principal **unidade de análise** no ENANI-2019 seja “crianças menores de 5 anos”, o processo de amostragem empregado também permite a estimação de parâmetros de interesse tendo o domicílio e a mãe biológica como outras possíveis unidades de análise. Assim, caso deseje-se analisar a unidade de análise “domicílio”, deve selecionar apenas uma informação de cada domicílio (a partir da variável identificadora do domicílio). Em casos onde existam mais de uma família no mesmo domicílio, deve-se atentar pois pode haver mais de um respondente (um por núcleo familiar). Para trabalhar com as mães biológicas como unidade de análise, deve-se selecionar um único filho por mãe.