

Prova de Coleta e Análise de Dados Secundários

Frederico Bertholini

Parte 1

1. Carregue o arquivo `deciso.es.rds` que está na pasta `aulas_ENAP/CADS2018/Exercícios/dados` em um objeto chamado `deciso.es` (ou outro nome à sua escolha).

Obs.: o arquivo pode ser baixado:

do OneDrive <https://goo.gl/5xVnvn>

ou do GitHub https://github.com/fredbsr/aulas_ENAP/tree/master/CADS2018/Exerc%C3%ADcios/dados

2. Separe a coluna `classe_assunto` em duas colunas, uma contendo a `classe` e outra contendo o `assunto`
3. Elabore um `data.frame` em que as linhas sejam a classe, as colunas sejam os meses e os valores sejam as quantidades de decisões

Dica 1: crie uma variável `mes` e exclua os casos em que não há informação sobre data da decisão

Dica 2: agrupar por classe e mes e fazer o spread

Parte 2

1. Leia o conjunto de dados ‘api’ do pacote `survey` usando o comando `data(api)`
2. Elimine os objetos ‘apiclus1’, ‘apiclus2’, ‘apipop’ e ‘apisrs’ mantendo apenas o objeto `apistrat`
3. crie o objeto ‘pesos’ selecionando as colunas ‘stype’ e ‘pw’ do objeto ‘apistrat’
4. crie o objeto ‘escolas_notas’ selecionando as colunas ‘cds’, ‘stype’, ‘api99’ e ‘api00’ do objeto `apistrat`
5. Remova as duplicatas (linhas em duplicidade) do arquivo `pesos` que você criou no passo 4
6. Quantas linhas tem o novo objeto `pesos`, sem as duplicidades?
7. Traga a variável `pw` para `escola_notas`, criando um novo objeto `amostra` resultado da união (join) do objeto `pesos` ao objeto `escolas_notas`

Dica: use `left_join`, com `escola_notas` na esquerda.

8. Crie o objeto `tbl_svy amostra_expandida` expandindo a amostra aleatória simples (`amostra`) usando a variável (coluna) “pw”, contendo o peso amostral como o peso e a variável `stype` como o estrato.

Dica: `as_survey(strata=stype,weight=pw)`

9. Usando a variável `stype` crie uma nova variável indicando se a escola é de nível fundamental (categorias **E** e **M** de `stype`) ou de nível médio (categoria **H** de `stype`).

Dica: use `mutate` e `case_when`.

10. Faça um gráfico de barras comparando a variação média das notas de 1999 (`api99`) e 2000 (`api00`) por nível de escola (`nivel`) utilize as estimativas intervalares para construir barras com o intervalo de confiança
11. Pode-se dizer que a variação média das notas entre 1999 e 2000 por nível de ensino (Médio e Fundamental) foi diferente?