

# Coleta e Análise de Dados Secundários

## *Prova Substitutiva*

A prova deve ser entregue por e-mail a **fredbsr@gmail**, apenas com o arquivo em formato **.R** referente ao script com as respostas.

### Parte 1

1. (0,5 ponto) Carregue o arquivo **deciso.es** que está na pasta aulas\_ENAP/CADS2018/Exercícios/dados em um objeto chamado **deciso.es** (ou outro nome à sua escolha). O arquivo pode ser baixado:
  - a. do OneDrive <https://goo.gl/5xVnvn>
  - b. ou do GitHub [https://github.com/fredbsr/aulas\\_ENAP/tree/master/CADS2018/Exerc%C3%ADcios/dados](https://github.com/fredbsr/aulas_ENAP/tree/master/CADS2018/Exerc%C3%ADcios/dados)
2. (0,5 ponto) Separe a coluna **classe\_assunto** em duas colunas, uma contendo a **classe** e outra contendo o **assunto**
3. (3 pontos) Crie uma variável indicando se determinada decisão está relacionada a drogas e faça um gráfico de linhas comparando as quantidades anuais de processos relacionados a drogas e de processos não relacionados a drogas. Cada linha do gráfico deve ter uma cor.

### Parte 2

1. (0 pontos) Leia o conjunto de dados ‘api’ do pacote survey usando o comando `data(api)`
2. (0,5 ponto) Crie o objeto do tipo `tbl_svy` chamado **amostra\_expandida** expandindo a amostra aleatória estratificada (**apistrat**), usando a variável (coluna) “pw”, como o peso amostral e a variável **stype** como o estrato.
3. (0,5 ponto) Usando a variável **stype** crie uma nova variável chamada **nivel**, indicando se a escola é de nível fundamental (categorias **E** e **M** de **stype**) ou de nível médio (categoria **H** de **stype**). Crie também a variável **api\_diff** como a variação média das notas de 1999 (**api99**) e 2000 (**api00**) (subtração de 2000 por 1999).
4. (0,5 ponto) Usando o pacote **srvyr**, crie um `data.frame` contendo as médias de **api\_diff** por **nivel** e utilize a opção de indicar o intervalo de confiança
5. (0,5 ponto) Faça o mesmo procedimento indicado na questão 3, porém agora com o dado da população, presente no `data.frame` **apipop**, criando as variáveis **nivel** e **api\_diff\_pop**. Após isso, crie um `data.frame` contendo as médias de **api\_diff\_pop** por **nivel**.
6. (0,5 ponto) Faça um join entre o objeto resultante da questão 4 e o objeto resultante da questão 5.
7. (0,5 ponto) Por que não utilizamos nomes de variável iguais para a diferença entre as notas de 1999 e 2000 no `data.frame` oriundo da população (questão 4) e no `data.frame` oriundo da expansão da amostra (questão 5)? (escreva a resposta a essa questão no próprio script, usando `#` para comentar)
8. (2 pontos) Faça um gráfico de barras comparando os resultados de por nível de ensino da escola (**nivel**) tanto para o dado amostral quanto para o dado populacional. Utilize as estimativas intervalares para construir intervalos de confiança.
9. (1 ponto) O que se pode concluir do gráfico? (escreva a resposta a essa questão no próprio script, usando `#` para comentar)