

Lista 8

Gilberto Sassi

Primeiro semestre de 2021.

Nos exercícios abaixo, use *bootstrap*, *jackknife* e *validação cruzada*.

1. Considere a amostra de mil candidatos no ENEM em sua edição de 2019 na cidade de Palmas cujas informações estão no conjunto de dados `equipe_sample__Palmas.csv`. Construa um intervalo de confiança usando *bootstrap* para média, mediana e desvio padrão usando coeficiente de confiança 1% para a nota em matemática dos candidatos (`NU_NOTA_MT`).
2. Responda cada uma das perguntas abaixo ao nível de significância 5%
 - i. A nota média de matemática do ENEM em palmas (`NU_NOTA_MT`) é maior que 500?
 - ii. A nota mediana de matemática do ENEM em palmas (`NU_NOTA_MT`) é maior que 500?
 - iii. O desvio padrão da nota de matemática do ENEM em palmas (`NU_NOTA_MT`) é maior que 100?
3. Para o conjunto de dados `carros.xlsx`, qual o valor-p no teste de hipóteses bilateral com $H_0 : \mu = 15$ ao nível de significância 1%, em que μ é a velocidade média populacional dos carros?
4. Para o conjunto de dados `carros.xlsx`, use o procedimento de Neyman-Pearson para o teste de hipóteses bilateral com $H_0 : \mu = 15$ ao nível de significância 1%, em que μ é a velocidade média populacional dos carros.
5. Para o conjunto de dados `empresa.xlsx`, construa o intervalo de confiança para o número de filhos dos funcionários com coeficiente de confiança 99%.
6. Para o conjunto de dados `empresa.xlsx`, construa o intervalo de confiança para a média salarial em salários mínimos com coeficiente de confiança 95%. Baseando sua resposta neste intervalo de confiança, a média salarial é maior que 10 salários mínimos?
7. Para o conjunto de dados `iris.ods`, existe evidência de que o desvio padrão do comprimento de sépala é maior que 0,1 centímetros? Use o procedimento de Neymann-Pearson e nível de significância igual a 5%.
8. Para o conjunto de dados `mtcars.csv`, existe evidência de que o desvio padrão da variável `mpg` (milhar

por galão) é maior que 5 ao nível de significância 1%? Use o procedimento de Neymann-Pearson.

9. Para as seguintes variáveis calcule o primeiro e o terceiro quartis e o seus respectivos desvios padrões e vícios usando *jackknife*:

- i. NU_NOTA_MT do conjunto de dados `equipe_sample__Palmas.csv`;
- ii. `vel` do conjunto de dados `carros.xlsx`;
- iii. `salario` do conjunto de dados `empresa.xlsx`;
- iv. `comprimento_sepala` do conjunto de dados `iris.ods`;
- v. `mpg` do conjunto de dados `mtcars.csv`.

10. Para o conjunto de dados `marketing.csv`, decida entre os seguintes modelos usando *validação cruzada* com $K = 1$:

- i. **Modelo 1:** $venda \sim 1 + gasto + \epsilon$;
- ii. **modelo 2:** $\log(venda) \sim 1 + gasto + \epsilon$;
- iii. **Modelo 3:** $venda \sim 1 + gasto + gasto^2 + \epsilon$;
- iv. **Modelo 4:** $\log(venda) \sim 1 + gasto + gasto^2 + \epsilon$;