

# Lista 1 Prática Heterocedasticidade

Giovanna Pavlovic Quintão

Setembro 2020

## 1 Instruções

Você deverá entregar dois arquivos: 1 em word e 1 em excel.

- Para o documento em word:
  1. O nome do documento deve conter **apenas** seu número de matricula (ex: "999999.docx")
  2. **Todas** as questões deverão ser respondidas neste documento, independentemente se as respostas já estejam no excel.
  3. Você deverá copiar ou dar print nas saídas do R relativas a **cada** questão, e deverá escrever uma breve descrição do que foi feito.
  4. Por favor, faça as questões na ordem.
- Para o documento em xlsx:
  1. Você preencherá a planilha que disponibilizei no Stoa como o nome: "lista\_1\_tabelas\_respostas.xlsx"
  2. As questões que deverão ser respondidas **também** nesta planilha estão em verde com o nome da coluna identificadora.
  3. O nome do documento deve conter apenas seu número de matricula (ex: "999999.xlsx")
  4. Não delete coluna alguma, não altere o nome da coluna e não altere a ordem das colunas!
  5. Regras de preenchimento:
    - (a) Preencha todos os números separando os decimais por pontos (.) e não separe os milhares. Não coloque unidades de medida. (ex: 3 mil reais e 95 centavos ficariam assim: 3000.950)
    - (b) Use três casas decimais. (dica: o comando `round(x,3)` já arredonda o número x automaticamente para você !)
    - (c) Não use caracteres especiais (~ ç, etc).
    - (d) Sempre use caracteres minúsculos (ex: se a pergunta for de sim ou não, responda : sim ou nao)

## 2 Preparação

- Requisite os pacotes que você utilizará (comando library)
- Abra a base de dados no R que vamos utilizar: "PNAD-2015-ajustada.csv"
- Filtre o(s) estado(s) que você vai trabalhar. **Atenção, cada pessoa tem um conjunto de estados diferentes para trabalhar**
  1. Olhe seus estados na planilha de acordo com seu número de matrícula "sorteio\_estados.xlsx"
  2. Escreva o código para filtrar sua base de dados. Por exemplo, se meus estados fossem o 33, 11 e o 17, eu escreveria o seguinte no código:

```
df<-df %>%
  filter (uf %in% c(33,11,17))
```
  3. Preencha na planilha seu número de matrícula (`n.matricula`) e as ufs (`uf.s`). As ufs deverão ser separadas por vírgulas.

## 3 Regressão Linear e Teste

1. Faça a regressão múltipla da renda como função de: anos de estudo, raça, sexo, idade, idade ao quadrado. Interprete o resultado da dummy de gênero.
2. Faça o teste de Breusch Pagan manualmente utilizando a estatística F. Para isso, você deverá responder às seguintes questões:
  - Qual é o valor de  $R^2_{u^2}$ ? (`r_squared`)
  - Qual é o valor da estatística F? (`f_statistic`)
  - Há heterocedasticidade? (resposta: sim ou nao) (`is_heter`)

## 4 Correção de Erro Padrão Robusto White

3. Vamos calcular o erro padrão robusto de White para a variável de raça. Não se esqueça que todos os resultados das regressões têm que estar no seu arquivo word.
  - Qual é a soma dos quadrados dos resíduos da regressão auxiliar? ( $SQR_j$ ) (`sqr`)
  - Qual é o erro padrão robusto heterocedasticidade calculado para esta variável? (lembre-se que você pode checar se sua resposta está correta com o pacote) (`ep_robust`)

## 5 Modelo de Probabilidade Linear com mínimos quadrados ponderados

Vamos estimar um modelo de MPL, e vamos tratar a heterocedasticidade com o MQP. Dica: Volte na seção 8.5 do Wooldridge

1. Vamos estimar a probabilidade conclusão do ensino fundamental com base nas características do indivíduo.
  - (a) Para isso, primeiro crie uma variável de "concluiu\_EF" a partir da variável já existente de "anos\_estudo". Esta nova variável assumirá valor 0 caso  $\text{anos\_estudo} < 8$  e 1, caso contrário.
  - (b) Tire a média desta variável nova para termos uma intuição de como está a distribuição desta variável (`mean_concluiu_EF`)
  - (c) Estime a regressão com as seguintes variáveis explicativas: idade, idade2, sexo, raca. Lembre-se de colocar no arquivo de word a saída da regressão.

2. Salve o  $\hat{y}$ .

3. Verifique se todos  $\hat{y}$  estão entre (0,1). Caso não esteja, faça os ajustes necessários.

4. A partir de  $\hat{y}$ , construa as variâncias  $\hat{h}_i$ :

$$\hat{h}_i = \text{var}(\widehat{y}_i | x_i) = \hat{p}_i(1 - \hat{p}_i) = \hat{y}_i(1 - \hat{y}_i)$$

Tire a média da variância estimada  $\hat{h}$  e coloque na planilha (`mean_var`).

5. Estime a regressão com as variáveis transformadas usando como peso  $\frac{1}{\sqrt{\hat{h}_i}}$

Escreva qual é o coeficiente estimado da variável "sexo" na planilha (`coef_sexo`). Não se esqueça de colocar o sinal e arredondar!

6. Há diferenças entre a estimação inicial (sem MQP) e com MQP? Porque?