

Econometria 1

Pedro Henrique Rocha Mendes *

Lista 5

1)

Para resolver essa questão utilize, no R, o objeto `db_covs`, ou, em outro software, o arquivo `dados_db_covs.txt`.

```
# install.packages("tidyverse")
# install.packages("estimatr")
# install.packages("car")
# install.packages("fastDummies")
# install.packages("knitr")

library(tidyverse)
library(estimatr)
library(car)
library(fastDummies)
library(knitr)

load("listas/lista_5/lista_5_dados.Rdata")
```

1) a.

Aplique o procedimento do teste do multiplicador de Lagrange (nota de aula 7) aos quatro testes de restrições de exclusão discriminados na tabela abaixo, preenchendo-a. Adote o nível de significância de 5%.

```
# criando dummies

db_covs <- fastDummies::dummy_columns(
  db_covs,
  select_columns = "Continent",
  remove_first_dummy = T, # criando n-1 dummies
  remove_selected_columns = T) # removendo a coluna que gerou as dummies

db_covs <- db_covs |>
```

*RA: 11201811516

```

tibble::as_tibble() |>
dplyr::rename(d_asia = Continent_Asia,
              d_europe = Continent_Europe,
              d_n_america = Continent_NorthAmer,
              d_oceania = Continent_Oceania,
              d_s_america = Continent_SouthAmer)

# estimando regressão

eq <- obitos_por_milhao ~
  casos_por_milhao + fem_head + fertility_ +
  airpoll_exposure_ + d_lockdown + d_asia +
  d_europe + d_n_america + d_oceania +
  d_s_america

modelo <- estimatr::lm_robust(eq,
                             data = db_covs)

# testando significância conjunta - características nacionais

h0 <- c("airpoll_exposure_ = 0",
        "casos_por_milhao = 0")

car::linearHypothesis(modelo,
                      h0,
                      test ="Chisq")

## Linear hypothesis test
##
## Hypothesis:
## airpoll_exposure_ = 0
## casos_por_milhao = 0
##
## Model 1: restricted model
## Model 2: obitos_por_milhao ~ casos_por_milhao + fem_head + fertility_ +
##          airpoll_exposure_ + d_lockdown + d_asia + d_europe + d_n_america +
##          d_oceania + d_s_america
##
##   Res.Df Df Chisq Pr(>Chisq)
## 1      169
## 2      167  2 12.26  0.002177 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

vcrit_1 <- round(qchisq(0.05, 2, lower.tail = F), 2) # valor crítico

# testando significância conjunta - exposição à causas da doença

h0 <- c("fem_head = 0",
        "fertility_ = 0",
        "d_lockdown = 0")

car::linearHypothesis(modelo,
                      h0,
                      test ="Chisq")

## Linear hypothesis test
##

```

```
## Hypothesis:
## fem_head = 0
## fertility_ = 0
## d_lockdown = 0
##
## Model 1: restricted model
## Model 2: obitos_por_milhao ~ casos_por_milhao + fem_head + fertility_ +
##      airpoll_exposure_ + d_lockdown + d_asia + d_europe + d_n_america +
##      d_oceania + d_s_america
##
## Res.Df Df Chisq Pr(>Chisq)
## 1      170
## 2      167  3 5.1081      0.164
```

```
vcrit_2 <- round(qchisq(0.05, 3, lower.tail = F), 2) # valor crítico

# testando significância conjunta - continente

h0 <- c("d_asia = 0",
        "d_europe = 0",
        "d_n_america = 0",
        "d_oceania = 0",
        "d_s_america = 0")

car::linearHypothesis(modelo,
                      h0,
                      test = "Chisq")
```

```
## Linear hypothesis test
##
## Hypothesis:
## d_asia = 0
## d_europe = 0
## d_n_america = 0
## d_oceania = 0
## d_s_america = 0
##
## Model 1: restricted model
## Model 2: obitos_por_milhao ~ casos_por_milhao + fem_head + fertility_ +
##      airpoll_exposure_ + d_lockdown + d_asia + d_europe + d_n_america +
##      d_oceania + d_s_america
##
## Res.Df Df Chisq Pr(>Chisq)
## 1      172
## 2      167  5 30.24 1.323e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
vcrit_3 <- round(qchisq(0.05, 5, lower.tail = F), 2) # valor crítico
```

Categoria	Variáveis	Valor da estatística do teste	Graus de liberdade	Valor crítico	p-valor	Rejeitar H_0 (S/N)?
Características nacionais	País é dirigido por mulher, taxa de fertilidade, lockdown foi implementado	12,26	2	5.99	0,002177	S

Categoria	Variáveis	Valor da estatística do teste	Graus de liberdade	Valor crítico	p-valor	Rejeitar H_0 (S/N)?
Exposição à causas de doença	Exposição à poluição atmosférica, infecções por COVID-19	5.1081	3	7.81	0,164	N
Continente	Dummies continentais	30,24	5	11.07	$1,3 \cdot 10^{-05}$	S

1) b.

Responda: os resultados ora obtidos contradizem os obtidos no laboratório 1 indicando insignificância estatística da explicativa poluição atmosférica?

Sim, contradiz, pois o coeficiente para poluição atmosférica no teste de significância individual é significativo, com p-valor menor que 5%. Quando tomada em conjunto com variáveis não significativas como `fem_head` e `fertility_`, a variável não contribui para a explicação da variância de `obitos_por_milhao`.

```
estimatr::tidy(modelo) |>
  dplyr::select(term, p.value) |>
  dplyr::filter(term == "airpoll_exposure_" |
                 term == "fem_head" |
                 term == "fertility_") |>
  dplyr::mutate(alfa = dplyr::case_when(
    p.value > 0.05 ~ "N",
    TRUE ~ "S"
  )) |>
  knitr::kable(col.names = c("Variável",
                             "P-valor",
                             "Significativo a 5%?"))
```

Variável	P-valor	Significativo a 5%?
fem_head	0.6730017	N
fertility_	0.1895275	N
airpoll_exposure_	0.0367167	S

2)

Para resolver essa questão utilize, no R, o objeto `db_gee`, ou, em outro software, o arquivo `dados_db_gee.txt`.

Um pesquisador está interessado em estimar uma equação explicando emissões de gases de efeito estufa em função de variáveis macroeconômicas e binárias indicando nível de desenvolvimento e continente. A fórmula é a seguinte:

$$\text{geepc} = \beta_0 + \beta_1 \log(1 + \text{pibpc}) + \beta_2 \log(1 + \text{epc}) + \beta_3 \log(1 + \text{apc}) + \beta_4 \text{d_dev_ing} + \beta_5 \text{d_dev_tra} + \beta_6 \text{d_reg_afr} + \beta_7 \text{d_reg_asi} + \beta_8 \text{d_reg_cam} + \beta_9 \text{d_reg_nam} + \beta_{10} \text{d_reg_oce} + u$$

Aplicando-se as definições na tabela abaixo:

Variável	Categoria	Nome sucinto
Emissões de gases de efeito estufa per capita	Variável dependente	geepc
PIB per capita	Macro	pibpc
Consumo energético per capita	Macro	epc
Área territorial per capita	Macro	apc
País em desenvolvimento	Desenvolvimento	d_dev_ing
País em transição	Desenvolvimento	d_dev_tra
Continente Africano	Continente	d_reg_afr
Continente Asiático	Continente	d_reg_asi
Continente Americano Central e Caribenho	Continente	d_reg_cam
Continente Americano do Norte	Continente	d_reg_nam
Continente Americano do Sul	Continente	d_reg_sam
Continente Oceania	Continente	d_reg_oce

2) a.

Como preâmbulo, estime a regressão linear simples apenas com o consumo energético per capita (epc) como variável explicativa. A partir dos resultados, preencha a tabela abaixo.

```
# estimando regressão

eq <- geepc ~ log(1 + epc)

modelo <- estimatr::lm_robust(eq,
                              data = db_gee)

# função para parâmetros

parametros <- function(var, col) {
  x <- modelo |>
  dplyr::filter(term == var) |>
  dplyr::select(col) |>
  dplyr::pull() |>
  round(2)

  return(x)
}
```

```

}

# tabela

modelo <- tibble::as_tibble(estimatr::tidy(modelo))

table_a <- tibble::tibble(var = c("(Intercept)", "log(1 + epc)")) |>
  dplyr::group_by(var) |>
  tidyr::nest() |>
  dplyr::mutate(
    data = purrr::map(
      data,
      ~ tibble::tibble(
        col = c("estimate", "std.error", "p.value")
      )
    )
  ) |>
  tidyr::unnest(cols = c(data)) |>
  dplyr::ungroup() |>
  dplyr::mutate(
    value = purrr::map2_dbl(
      var,
      col,
      ~ parametros(.x, .y)
    )
  ) |>
  tidyr::pivot_wider(names_from = col,
    values_from = value) |>
  dplyr::rename(`Variável` = var,
    `Estimativa` = estimate,
    `Erro padrão` = std.error,
    `P-valor` = p.value) |>
  dplyr::mutate(`Significativo a 5%?` = dplyr::case_when(
    `P-valor` < 0.05 ~ "S",
    TRUE ~ "N"
  ))

table_a |>
  knitr::kable()

```

Variável	Estimativa	Erro padrão	P-valor	Significativo a 5%?
(Intercept)	7.36	0.19	0	S
log(1 + epc)	6.38	0.54	0	S

2) b.

Estime a regressão linear múltipla descrita pela fórmula acima, a qual corresponde a uma regressão “longa”, ou seja, com um número considerável de explicativas. Com base nisso preencha a tabela abaixo.

```
# estimando regressão

eq <- geepc ~
  log(1 + epc) + log(1 + pibpc) + log(1 + apc) +
  d_dev_ing + d_dev_tra + d_reg_afr +
  d_reg_asia + d_reg_cam + d_reg_nam +
  d_reg_sam + d_reg_oce

modelo <- estimatr::lm_robust(eq,
                             data = db_gee)

# tabela

modelo <- tibble::as_tibble(estimatr::tidy(modelo))

table_b <- tibble::tibble(var = c("(Intercept)",
                                   "log(1 + epc)", "log(1 + pibpc)",
                                   "log(1 + apc)", "d_dev_ing",
                                   "d_dev_tra", "d_reg_afr",
                                   "d_reg_asia", "d_reg_cam",
                                   "d_reg_nam", "d_reg_sam",
                                   "d_reg_oce")) |>

  dplyr::group_by(var) |>
  tidyr::nest() |>
  dplyr::mutate(
    data = purrr::map(
      data,
      ~ tibble::tibble(
        col = c("estimate",
                "std.error",
                "p.value")
      )
    ) |>
  tidyr::unnest(cols = c(data)) |>
  dplyr::ungroup() |>
  dplyr::mutate(
    value = purrr::map2_dbl(
      var,
      col,
      ~ parametros(.x, .y)
```

```

)
) |>
tidyr::pivot_wider(names_from = col,
                    values_from = value) |>
dplyr::rename(`Variável` = var,
              `Estimativa` = estimate,
              `Erro padrão` = std.error,
              `P-valor` = p.value) |>
dplyr::mutate(`Significativo a 5%?` = dplyr::case_when(
  `P-valor` < 0.05 ~ "S",
  TRUE ~ "N"
))

table_b |>
knitr::kable()

```

Variável	Estimativa	Erro padrão	P-valor	Significativo a 5%?
(Intercept)	-50.76	3.94	0.00	S
log(1 + epc)	4.73	0.47	0.00	S
log(1 + pibpc)	4.11	0.32	0.00	S
log(1 + apc)	1.62	0.17	0.00	S
d_dev_ing	4.76	0.69	0.00	S
d_dev_tra	3.13	0.67	0.00	S
d_reg_afr	1.57	0.75	0.04	S
d_reg_así	4.39	0.52	0.00	S
d_reg_cam	-1.40	0.60	0.02	S
d_reg_nam	6.73	1.00	0.00	S
d_reg_sam	0.40	0.73	0.58	N
d_reg_oce	7.44	1.25	0.00	S

2) c.

Quanto à regressão longa, responda:

1. *Quais variáveis são estatisticamente não-significativas?*

```

table_b |>
dplyr::filter(`Significativo a 5%?` == "N") |>
knitr::kable()

```

Variável	Estimativa	Erro padrão	P-valor	Significativo a 5%?
d_reg_sam	0.4	0.73	0.58	N

2. Há alguma variável cuja insignificância estatística é contra-intuitiva?

Sim, *epc*, que era significativa na regressão simples e na múltipla não é mais. Além disso, *pibpc* não ser significativo nem influenciar o nível de emissões *per capita* é contra-intuitivo, pois espera-se que tal variável seja significativa e com o coeficiente positivo.

2) d.

Compare os resultados das duas estimações nos itens “a” e “b” no tocante ao coeficiente do consumo energético per capita. Caso tenha havido alteração da estimativa pontual ou do valor observado da estatística *t*, explique por que isso ocorreu. Em sua resposta explore a diferença entre regressão simples e múltipla.

```
df1 <- table_a |>
  dplyr::mutate(Variável = dplyr::case_when(
    Variável == "epc" ~ "epc_a"
  )) |>
  dplyr::slice(2)

df2 <- table_b |>
  dplyr::mutate(Variável = dplyr::case_when(
    Variável == "epc" ~ "epc_b"
  )) |>
  dplyr::slice(2)

dplyr::bind_rows(df1, df2) |>
  knitr::kable()
```

Variável	Estimativa	Erro padrão	P-valor	Significativo a 5%?
NA	6.38	0.54	0	S
NA	4.73	0.47	0	S

A alteração na significância da variável se dá por conta do viés de variável omitida. Esse viés surge quando se omite uma variável explicativa x_k no modelo que é relevante na explicação da variação da variável dependente. Tal viés leva a conclusões equivocadas sobre a função de regressão populacional, como neste caso, onde a variável *epc* se mostrou significante no modelo viesado, o que leva a conclusão de que o consumo energético per capita é relevante para explicar as emissões de gases do efeito estufa.

2) e.

Compare o R^2 ordinário da regressão simples com o R^2 ajustado da regressão múltipla e explique por que os dois diferem em magnitude.

2) f.

Teste a hipótese de que nível de desenvolvimento explica fração irrelevante da emissão de gases de efeito estufa. Para isso aplique um teste de significância conjunta às explicativas d_dev_ing e d_dev_tra , utilizando o procedimento do teste de multiplicador de Lagrange da nota de aula 7. Em sua resposta informe:

1. Valor observado da estatística do teste
2. Valor crítico ao nível de significância de 5%
3. P-valor
4. Decisão apropriada entre rejeitar ou não a hipótese nula
5. Qual conclusão pode ser retirada do resultado do teste? Justifique.

3)

Ao estimar a equação de Mincer para o logaritmo natural da remuneração, foram obtidas estimativas pontuais de 0,071 e 0,026, respectivamente, para os coeficientes das explicativas captando nível educacional e experiência. Diversas outras explicativas foram incorporadas. As duas explicativas são medidas em anos e, portanto, trata-se de variáveis quantitativas discretas. Nenhuma delas foi incorporada ao modelo em forma logarítmica, mas sim em nível, i.e., sem qualquer transformação matemática de seus valores originais. Tendo em mente estes detalhes e também a definição do conceito do coeficiente de uma explicativa discreta, interprete economicamente, com a máxima clareza e precisão possíveis, os valores das estimativas pontuais, i.e., explique o significado econômico das estimativas pontuais considerando não apenas os sinais delas, mas também as magnitudes. Para isso, utilize a aproximação $\log(1 + w) \approx w$.

4)

Seja reconsiderado o problema de estimar o efeito da qualidade do ensino fundamental no desempenho no ensino médio, estudado na lista 3. Há duas especificações possíveis, sendo apenas a segunda a FRP verdadeira.

- FRP 1: $Taxa_aprov_i = \alpha_0 + \alpha_1 \text{quali_pub}_i + e_i$
- FRP 2: $Taxa_aprov_i = \beta_0 + \beta_1 \text{quali_pub}_i + \beta_2 \text{educ_pais}_i + u_i$

Demonstre que a estimação da FRP 1 é sujeita à inconsistência, assumindo, em conformidade com a lista 3, que a covariância populacional entre quali_pub e educ_pais é não-nula. A solução

pode ser apresentada tanto com notação não matricial como com notação matricial. Sendo a última utilizada, haverá acréscimo de meio ponto à nota final dessa questão.