Econometria 1

Pedro Henrique Rocha Mendes *

Prova 2

```r
# bibliotecas utilizadas

library(tidyverse)
library(knitr)
library(janitor)
library(kableExtra)

# removendo notação científica
options(scipen = 999)
```

- ☒ 1
- ☒ 2
- ☒ 3
- ☐ 4
- ☐ 5

1)

Um pesquisador estimou a FRP que se encontra na tabela 1. Trata-se de uma equação minceriana em que o salário horário é a variável dependente e as variáveis independentes correspondem a características dos(as) trabalhadores(as). Foi utilizada uma subamostra aleatória da PNAD anual de 2015.

1) a.

Os resultados podem ser visualizados na tabela 1. Preencha as três últimas colunas, inserindo, para cada variável independente, (i) o valor observado da estatística t de Student, (ii) o p-valor correspondente para o teste bilateral de significância individual e (iii) o símbolo de significância adequado. A última informação deve ser incorporada de acordo com a seguinte simbologia: (a)

*RA: 11201811516, Turno: Noite

+, para significativo a um nível de significância de 10% mas não para níveis de significância de 5% ou 1%, (b) $\star$, para significativo a 5% mas não para nível de significância de 1%, e (c) $\star\star$, para significativo a nível de significância igual ou menor do que 1% (OBS: o erro padrão é a raiz do estimador não-viesado para a variância do estimador de mínimos quadrados ordinários para o coeficiente). Em caso de significância a nível superior a 10%, deixar a célula em branco.

```r
tabela1 <- readr::read_delim("provas/p2/tabelas/tabela1.csv",
                             delim = ";") |>
  janitor::clean_names()

k <- nrow(tabela1) - 1

n <- 1000

tabela1 |>
  dplyr::mutate(estat_t = estimativa_pontual/erro_padrao) |>
  dplyr::mutate(p_valor = purrr::map_dbl(
    estat_t,
    ~ round(2 * pt(abs(.x), df = n-k-1, lower.tail = F), 4)
  )) |>
    dplyr::mutate(signific = dplyr::case_when(
    p_valor < 0.01 ~ "**",
    p_valor < 0.05 & p_valor != "**" ~ "*",
    p_valor < 0.1 & (p_valor != "**" | p_valor != "*") ~ "+",
    p_valor > 0.1 ~ " "
  )) |>
  dplyr::mutate(
    dplyr::across(
      2:5,
      ~ round(.x, 4)
    )
  ) |>
  knitr::kable(
    caption = "Resultados da estimação da FRP",
    booktabs = T,
    linesep = "",
    format.args = list(big.mark = ".", decimal.mark = ','),
    col.names = c("Variável",
                  "Estimativa pontual",
                  "Erro padrão",
                  "Estatística t",
                  "p-valor",
                  "Significância")) |>
  kableExtra::kable_styling(font_size = 9,
                            latex_options = "HOLD_position")
```

Tabela 1: Resultados da estimação da FRP

Variável	Estimativa pontual	Erro padrão	Estatística t	p-valor	Significância
Intercepto	0,5705	1,4192	0,4020	0,6878	
Etnia preta ou parda	-0,7696	0,4036	-1,9067	0,0569	+
Educação (anos de estudo)	0,6733	0,0582	11,5752	0,0000	**
Experiencia no mercado de trabalho	0,2660	0,0545	4,8771	0,0000	**
Experiência no mercado de trabalho ao quadrado	-0,0045	0,0011	-4,0035	0,0001	**
Experiência no emprego atual	0,1705	0,0309	5,5132	0,0000	**
Atividade agropecuária (binária)	0,1489	1,2283	0,1212	0,9035	
Atividade industrial (binária)	0,3496	0,7398	0,4726	0,6366	
Atividade construção civil (binária)	0,9667	0,9529	1,0145	0,3106	
Atividade serviços sociais (binária)	0,8554	0,7203	1,1875	0,2353	
Atividade outros serviços (binária)	-0,8838	0,6300	-1,4028	0,1610	
Atividade administração pública (binária)	3,6075	0,9053	3,9850	0,0001	**
Pessoa de referência na família (binária)	0,8102	0,3961	2,0452	0,0411	*
Realiza afazeres domésticos (binária)	-0,8771	0,4064	-2,1580	0,0312	*
Região Norte (binária)	-2,3885	0,7543	-3,1665	0,0016	**
Região Nordeste (binária)	-2,3135	0,6708	-3,4489	0,0006	**
Região Sudeste (binária)	-0,7238	0,6420	-1,1273	0,2599	
Região Sul (binária)	-0,7514	0,7116	-1,0560	0,2912	
Área urbana (binária)	1,5331	0,7833	1,9574	0,0506	+

1) b.

O R^2 ordinário foi de 0,2951, com tamanho (sub)amostral equivalente a 1.000. Com base nisso e na tabela acima, aplique o teste de significância global a 5%, informando (e justificando):

- *O valor observado da estatística;*
- *O p-valor;*
- *Se é correto ou não rejeitar a hipótese nula.*

```r
r2 <- 0.2951

f <- round((r2/k)/((1-r2)/(n-k)), 4)

pvalor <- round(pf(f, df1 = k, df2 = n-k-1, lower.tail = F), 4)

tibble::tibble(parametro = c("$R^2$", "F", "p-valor"),
               valores = c(r2, f, pvalor)) |>
  knitr::kable(
    booktabs = T,
    escape = F,
    col.names = c(),
    format.args = list(decimal.mark = ",")) |>
  kableExtra::kable_styling(font_size = 9,
                            latex_options = "HOLD_position") |>
  kableExtra::footnote(general = "Conclusão: Rejeita-se $H_0$",
                       general_title = "",
                       escape = F)
```

R^2	0,2951
F	22,8392
p-valor	0,0000
Conclusão: Rejeita-se H_0	

2)

A tabela 2 apresenta parte dos resultados da estimação de duas equações que explicam a extensão de terra ocupada com culturas agrícolas exceto por soja, milho ou cana-de-açúcar ("outras culturas"), em microrregiões brasileiras.

2) a.

Informe, para o teste de significância individual bicaudal, a região crítica para os seguintes níveis de significância, arredondando após a segunda casa decimal:

```r
tabela2 <- readr::read_csv("provas/p2/tabelas/tabela2.csv") |>
  janitor::clean_names()

# k = k + 1, já que o número de linhas na tabela incorpora o intercepto

k <- nrow(tabela2)-1

n <- 558

f <- 31.85

t <- tibble::tibble(alfa = c(0.1, 0.05, 0.01)) |>
  dplyr::mutate(limite_inferior = purrr::map_dbl(
    alfa,
    ~ round(qt(.x/2, df = n-k-1, lower.tail = T), 2)
  ),
  limite_superior = abs(limite_inferior))

t |>
  dplyr::mutate(
    dplyr::across(
      2:3,
      as.character
    ),
    a = "$(-\\infty;",
    b = "]",
    c = "[",
    d = ";\\infty)$") |>
  tidyr::unite(a, limite_inferior, b,
               col = "limite_inferior",
               sep = "",
               remove = T) |>
  tidyr::unite(c, limite_superior, d,
               col = "limite_superior",
               sep = "",
               remove = T) |>
```

```
  tidyr::unite(limite_inferior, limite_superior,
               col = "rc",
               sep = "\\cup",
               remove = T) |>
  dplyr::mutate(
    rc = purrr::map_chr(
      rc,
      ~ stringr::str_replace_all(
        string = .x,
        pattern = "\\.",
        replacement = "{,}")
    )
  ) |>
  knitr::kable(
    format.args = list(big.mark = ".", decimal.mark = ','),
    booktabs = T,
    linesep = "",
    col.names = c("$\\alpha$", "Região crítica"),
    escape = F) |>
  kableExtra::kable_styling(font_size = 11,
                            latex_options = "HOLD_position")
```

α	Região crítica
0,10	$(-\infty; -1{,}65] \cup [1{,}65; \infty)$
0,05	$(-\infty; -1{,}96] \cup [1{,}96; \infty)$
0,01	$(-\infty; -2{,}58] \cup [2{,}58; \infty)$

2) b.

Informe, para o teste de significância global bicaudal[1]*, a região crítica para os seguintes níveis de significância, arredondando após a segunda casa decimal:*

```
f <- tibble::tibble(alfa = c(0.1, 0.05, 0.01)) |>
  dplyr::mutate(f = purrr::map_dbl(
    alfa,
    ~ round(qf(.x, df1 = k, df2 = n-k-1, lower.tail = F), 2)
  ))

f |>
  dplyr::mutate(
    dplyr::across(
      2,
      as.character
    ),
    a = "$[",
    b = ";\\infty)$") |>
  tidyr::unite(a, f, b,
               col = "rc",
               sep = "") |>
  dplyr::mutate(
    rc = purrr::map_chr(
```

[1] Imagino que o correto seja o teste F unicaudal, já que a distribuição F é assimétrica e sempre positiva.

```r
      rc,
      ~ stringr::str_replace_all(
        string = .x,
        pattern = "\\.",
        replacement = "{,}")
    )
  ) |>
  knitr::kable(
    format.args = list(big.mark = ".", decimal.mark = ','),
    col.names = c("$\\alpha$", "Região crítica"),
    booktabs = T,
    linesep = "",
    escape = F) |>
  kableExtra::kable_styling(font_size = 11,
                            latex_options = "HOLD_position")
```

α	Região crítica
0,10	$[1{,}47; \infty)$
0,05	$[1{,}64; \infty)$
0,01	$[2; \infty)$

2) c.

Considerando (i) testes de significância individual bicaudais e (ii) o teste de significância global com base na estatística F bicaudal, preencha a quarta coluna da tabela 2 ("Simbologia") com a seguinte simbologia:

- Explicativa significativa a 1% ou menos: escrever o número "1";
- Explicativa significativa a 5% mas não significativa a 1%: escrever o número "5";
- Explicativa significativa a 10% mas não significativa a 5% e nem a 1%: escrever o número "10";
- Explicativa não significativa a 10% ou a 5% ou a 1%: escrever a letra "N";

```r
tabela2 |>
  dplyr::mutate(simbologia = case_when(
    abs(estatistica_t) > t$limite_superior[3] ~ "1",
    dplyr::between(abs(estatistica_t),
                   t$limite_superior[2],
                   t$limite_superior[3]) ~ "5",
    dplyr::between(abs(estatistica_t),
                   t$limite_superior[1],
                   t$limite_superior[2]) ~ "10",
    abs(estatistica_t) < t$limite_superior[1] ~ "N"
  )) |>
  knitr::kable(
    caption = "Regressão explicando a área de outras culturas,
    microrregiões brasileiras",
    format.args = list(big.mark = ".", decimal.mark = ','),
    booktabs = T,
```

```
    linesep = "",
    col.names = c("Variável", "$\\beta$", "Estatística t", "Simbologia"),
    escape = F) |>
  kableExtra::kable_styling(font_size = 9,
                            latex_options = "HOLD_position")
```

Tabela 2: Regressão explicando a área de outras culturas, microrregiões brasileiras

Variável	β	Estatística t	Simbologia
Preço de soja	9.894,47	0,68	N
Preço de milho	-16.909,80	-0,85	N
Preço de cana-de-açúcar	40.212,66	2,19	5
Preço de outras culturas	-0,19	-0,51	N
Preço de produtos florestais	-33.198,00	-2,09	5
Preço da terra	-268,26	-0,24	N
Preço do trabalho	-1.020,26	-3,31	1
Temperatura Dez-Jan-Fev	-31.752,74	-3,80	1
Temperatura Mar-Abr-Mai	35.987,36	3,59	1
Temperatura Jun-Jul-Ago	-14.920,47	-1,99	5
Temperatura Set-Out-Nov	4.685,87	0,69	N
Precipitação Dez-Jan-Fev	-133,71	-1,12	N
Precipitação Mar-Abr-Mai	176,81	1,44	N
Precipitação Jun-Jul-Ago	-81,37	-0,69	N
Precipitação Set-Out-Nov	64,61	0,51	N
Área total	0,00	1,53	N
Valor médio da variável dependente na vizinhança	0,05	1,70	10
Constante	171.399,00	2,34	5

3)

Com dados da PNAD 2015 foram rodadas regressões de Mincer separadamente para as cinco regiões brasileiras. A hipótese de que a etnia tem influência sobre a remuneração foi submetida à refutação a partir de um teste de significância conjunta para variáveis binárias indicando autodeclarados brancos (`d_bca`), negros (`d_negro`) e pardos (`d_pardo`). Formalmente, trata-se do teste abaixo:

- H_0: $\beta_{\text{d\_bca}} = \beta_{\text{d\_negro}} = \beta_{\text{d\_pardo}} = 0$
- H_1: $\beta_{\text{d\_bca}} \neq 0$ ou $\beta_{\text{d\_negro}} \neq 0$ ou $\beta_{\text{d\_pardo}} \neq 0$

Os valores da estatística F para o teste unicaudal e respectivos graus de liberdade constam na tabela 3. Preencha as duas últimas linhas da tabela 3 conforme indicado. Considere para isso um nível de significância de 5%. Arredonde os valores críticos a partir da quinta casa decimal. Na última linha da tabela 3, marque "S" caso o valor da estatística F seja desfavorável, a um nível de significância de 5%, à hipótese nula, e marque "N" se o valor da estatística F for favorável à hipótese nula.

```r
tabela3 <- readr::read_csv("provas/p2/tabelas/tabela3.csv") |>
  janitor::clean_names()

#' s = número de restrições impostas ao modelo restrito, e
#' como são cinco regressões, cada uma com três variáveis
#' e não tem informações sobre o número total de coeficientes
#' no enunciado, s é igual ao número de coeficientes listados
#' na hipótese nula e o modelo se torna de significância global

s <- 3

tabela3 |>
  dplyr::mutate(
    valor_critico = purrr::map_dbl(
      gl_do_denominador_de_f,
      ~ round(qf(0.05, df1 = s, df2 = .x, lower.tail = F), 4)
    ),
    signific = dplyr::case_when(
      estatistica_f > valor_critico ~ "S",
      TRUE ~ "N"
    )
  ) |>
  knitr::kable(
    caption = "Resultados dos testes de significância conjunta
    para as três medidas de etnia, cinco regiões brasileiras",
    booktabs = T,
    linesep = "",
    format.args = list(big.mark = ".", decimal.mark = ','),
    col.names = c("Estado",
                  "Estatística F",
                  "G. L. do denominador de F",
                  "Valor crítico",
                  "Rejeita $H_0$?"),
    escape = F) |>
  kableExtra::kable_styling(font_size = 9,
                            latex_options = "HOLD_position")
```

Tabela 3: Resultados dos testes de significância conjunta para as três medidas de etnia, cinco regiões brasileiras

Estado	Estatística F	G. L. do denominador de F	Valor crítico	Rejeita H_0?
Norte	17,5949	12.294	2,6056	S
Nordeste	23,7029	21.734	2,6053	S
Sudeste	104,4111	30.219	2,6052	S
Sul	20,9448	16.125	2,6055	S
Centro-Oeste	17,5349	10.886	2,6057	S

4)

Em cada um dos cinco itens a seguir há afirmações sobre propriedades assintóticas do estimador de MQO para o vetor de parâmetros da FRP, $\hat{\beta}_{\text{MQO}}$, e também para procedimentos de inferência relacionados. Em cada item, selecione a única afirmação correta e justifique sua escolha.

4) a.

Quanto às diferenças entre as propriedades de ausência de viés e consistência, referentes à $\hat{\beta}_{\text{MQO}}$, *é correto afirmar que:*

- ☐ *Apesar das duas propriedades serem assintóticas, ou seja, válidas apenas para amostra de tamanho infinito, elas diferem por dizerem respeito à estatísticas distintas. Ausência de viés diz respeito à expectativa do estimador, enquanto consistência diz respeito ao limite em probabilidade ("plim") do estimador;*
- ☒ *Apenas a propriedade de consistência é assintótica, ou seja, válida apenas para amostra de tamanho infinito, e não há diferença entre as duas propriedades em termos das estatísticas a que se referem, pois ambas dizem respeito, exclusivamente, à expectativa do estimador;*
- ☐ *Apenas a propriedade de consistência é assintótica, ou seja, válida apenas para amostra de tamanho infinito e, além disso, as duas propriedades diferem por dizerem respeito à estatísticas distintas. Ausência de viés diz respeito à expectativa do estimador, enquanto consistência diz respeito ao limite em probabilidade ("plim") estimador.*

A inexistência de viés de MQO é uma propriedade de amostra finita, pois é válida para qualquer amostra de tamanho n. Porém, segundo Wooldridge,[2] a inexistência de viés dos estimadores não pode ser conseguida sempre. Assim, caso um estimador de MQO não seja viesado, as hipóteses MCRL1-MCRL4 garantem que o valor médio do parâmetro populacional. A consistência do parâmetro de dá quando a distribuição do estimador se torna mais estreitamente distribuída ao redor do parâmetro, e quando n tende ao infinito, tal distribuição se torna um único ponto no parâmetro populacional. Porém, caso um estimador, mesmo que não viesado, não leve, em geral, para um valor mais próximo do parâmetro de interesse, tal estimador não é consistente. Assim, conclui-se que só a consistência é assintótica.

4) b.

Examine em detalhe a demonstração para a consistência do estimador de MQO exibida a seguir e responda, posteriormente, as perguntas quanto às passagens lógicas numeradas.

$$\begin{aligned}
\text{plim}\left(\hat{\beta}_{\text{MQO}}\right) &= \beta + \text{plim}\left[\left(\sum_{i=1}^{N} x_i x_i^{\top}\right)^{-1} \sum_{i=1}^{N} x_i u_i\right] \\
&= \beta + \text{plim}\left[\left(\sum_{i=1}^{N} x_i x_i^{\top}\right)^{-1}\right] \text{plim}\left(\sum_{i=1}^{N} x_i u_i\right) = \{1\} \\
&= \beta + E\left(x_i x_i^{\top}\right)^{-1} E\left(x_i u_i\right)
\end{aligned}$$

[2]WOOLDRIDGE, Jeffrey M., **Introductory econometrics: A modern approach**, [s.l.]: Cengage learning, 2015.

$$\operatorname{plim}\left(\hat{\beta}_{MQO}\right) = \beta + E\left(x_i x_i^{\top}\right)^{-1} E\left(x_i u_i\right)\ (A)$$

$$E\left(x_i u_i\right) = \{2\} = E\left[E\left(x_i u_i \mid x_i\right)\right] = \{3\} = E\left[x_i E\left(u_i \mid x_i\right)\right] = \{4\} = E\left(x_i 0\right) = 0$$
$$\rightarrow E\left(x_i u_i\right) = 0\ (B)$$

Combinando (A) e (B), tem-se: $\operatorname{plim}\left(\hat{\beta}_{MQO}\right) = \beta$.

***(1)** A passagem {1} faz uso do seguinte fato (assinalar apenas uma alternativa):*

- ☒ *O "plim" da média amostral é equivalente à expectativa;*
- ☐ *A fórmula do estimador de mínimos quadrados ordinários é* $\hat{\beta}_{MQO} = \beta + \left(\sum_{i=1}^{N} x_i x_i^{\top}\right)^{-1} \sum_{i=1}^{N} x_i u_i$.

$$\operatorname{plim}(\hat{\beta}_{mqo}) = \beta + \frac{N}{N}\left\{\operatorname{plim}\left[\left(\sum_{i=1}^{n} x_i x_i^{T}\right)^{-1}\right]\operatorname{plim}\left(\sum_{i=1}^{n} x_i u_i\right)\right\}$$
$$= \beta + \operatorname{plim}\left[\frac{1}{N}\left(\sum_{i=1}^{n} x_i x_i^{T}\right)^{-1}\right]\operatorname{plim}\left(\frac{1}{N}\sum_{i=1}^{n} x_i u_i\right)\ (a)$$

desenvolvendo (a)

$$\begin{cases} \operatorname{plim}\left[\frac{1}{N}\left(\sum_{i=1}^{n} x_i x_i^{T}\right)^{-1}\right] = E\left(x_i x_i^{T}\right)^{-1} \\ \operatorname{plim}\left(\frac{1}{N}\sum_{i=1}^{n} x_i u_i\right) = E\left(x_i u_i\right) \end{cases}$$

***(2)** As passagens {2} e {3} consistem em:*

- ☐ *Aplicação da lei das expectativas iteradas e consideração de que as variáveis explicativas, e, portanto, o vetor x_i são não-aleatórios, sendo, pois, constantes para a expectativa;*
- ☐ *Aplicação da expectativa a uma constante nula e consideração de que as variáveis explicativas e, portanto, o vetor x_i, deixam de ser aleatórios ao condicionar-se a expectativa em um valor específico (e, pois, constante) para tal vetor (e, consequentemente, em valores constantes para todas as explicativas);*
- ☒ *Aplicação da lei das expectativas iteradas e consideração de que as variáveis explicativas e, portanto, o vetor x_i, deixam de ser aleatórios ao condicionar-se a expectativa em um valor específico (e, pois, constante) para tal vetor (e, consequentemente, em valores constantes para todas as explicativas);*

(1) Lei das expectativas iteradas: $E[E(Y|X)] = E(Y)$

(2) Propriedade do valor esperado condicional: $E[c(X)|X] = c(X)$, para qualquer função $c(X)$.

Assim:

$$E(x_i u_i) = E[E(x_i u_i|x_i)] \tag{1}$$

$$E[E(x_i u_i)|x_i] = E[x_i E(u_i|x_i)] \tag{2}$$

***(3)** A passagem {4} consiste em:*

- ☒ *Hipótese de exogeneidade do modelo clássico de regressão linear;*
- ☐ *Hipótese de homocedasticidade do modelo clássico de regressão linear;*
- ☐ *Fato de que a expectativa do termo de perturbação é sempre nula, ou seja, $E(u_i) = 0, i = 1, \ldots, N$;*

(MCRL 4) Exogeneidade

A expectativa condicional do termo de perturbação é nula, isto é, $E[u_i|x_i] = 0$, $i = 1, \ldots, N$

4) c.

Conhecer a função de distribuição de probabilidade do estimador de MQO referente à situação em que o tamanho amostral é infinito é necessário pois:

- ☒ *Mesmo sendo plenamente conhecida a função de distribuição de probabilidade do estimador de MQO para um tamanho amostral finito, é necessário, de maneira a ter conhecimento completo de tal função, também estudar a situação de tamanho amostral infinito;*
- ☐ *A função de distribuição de probabilidade do estimador de MQO para um tamanho amostral finito é desconhecida, a menos que seja assumida a hipótese de Gauss-Markov segundo a qual o termo de perturbação tem distribuição normal. É exatamente por conta da necessidade de conhecer a função em questão, no caso geral em que a hipótese de Gauss-Markov não necessariamente se aplica, que o estudo da situação com amostra infinita é necessário;*

Os estimadores de MQO, quando padronizados adequadamente, têm distribuições normais padrão aproximadas para grandes amostras. Isso se dá porque os operadores de MQO usam médias amostrais, e o teorema do limite central garante essa aproximação. Porém, se o tamanho da amostra não é muito grande, os testes de hipóteses para estimadores de MQO que se baseiam na normalidade assintótica pode ficar comprometida.[3] Assim, é necessário ter conhecimento das propriedades assintóticas para casos onde as hipóteses de amostra finita não se sustentam.

[3] *Ibid.*

4) d.

Tenha-se em vista o teste em que se procura determinar se algumas das variáveis independentes, referidas doravante como "variáveis-alvo" são conjuntamente significativas. Seja denominada por regressão irrestrita aquela com o conjunto completo de variáveis explicativas e por regressão restrita aquela em que são excluídas as variáveis-alvo. Nesse sentido:

- ☐ *O teste com base na estatística F compara as somas dos quadrados dos resíduos das regressões restrita e irrestrita. Neste caso, para calcular o valor crítico, é necessário saber apenas o número de variáveis-alvo. Já o teste do multiplicador de Lagrange consiste em calcular o R^2 dos resíduos da regressão restrita contra o conjunto completo de explicativas. Este último teste também toma por base a função de distribuição de probabilidades F de Snecedor e, para calcular o valor crítico, é preciso saber apenas o número de variáveis alvo.*
- ☒ *O teste com base na estatística F compara as somas dos quadrados dos resíduos das regressões restrita e irrestrita. Neste caso, para calcular o valor crítico, é necessário saber tanto o número de variáveis-alvo como o número de graus de liberdade calculado como subtração do tamanho amostral pelo número de parâmetros do modelo irrestrito. Já o teste do multiplicador de Lagrange consiste em calcular o R^2 dos resíduos da regressão restrita contra o conjunto completo de explicativas. Este último teste toma por base a função de distribuição de probabilidades qui-quadrado e, para calcular o valor crítico, é preciso saber apenas o número de variáveis-alvo.*
- ☐ *O teste com base na estatística F consiste em calcular o R^2 dos resíduos da regressão restrita contra o conjunto completo de explicativas. Neste caso, para calcular o valor crítico, é necessário saber tanto o número de variáveis-alvo como o número de graus de liberdade calculado como subtração do tamanho amostral pelo número de parâmetros do modelo irrestrito. Já o teste do multiplicador de Lagrange compara as somas dos quadrados dos resíduos das regressões restrita e irrestrita. Este último teste toma por base a função de distribuição de probabilidades qui-quadrado e, para calcular o valor crítico, é preciso saber apenas o número de variáveis-alvo.*

Estatística de teste $\hat{F}\left(H_0\right) = \frac{(SQR_R - SQR_{IR})/s}{SQR_{IR}/[N-(K+1)]} \sim F_{s,\ N-(K+1)}$, onde:

- s = número de restrições (de exclusão) impostas ao modelo, ou seja, é preciso saber o número de variáveis alvo;
- N - (K + 1) = tamanho amostral pelo número de parâmetros do modelo irrestrito.

Estatística de teste $LM = NR_u^2 \sim \chi_s^2$, onde:

- N = tamanho amostral;
- R^2_u = R^2 da regressão em que a variável dependente é o resíduo da regressão irrestrita e todas as outras variáveis independentes são utilizadas.

5)

Utilize o teorema de Frisch-Waugh para demostrar que, para um modelo de regressão múltipla com K variáveis explicativas, a fórmula do estimador do coeficiente da k-ésima explicativa é:

$$\hat{\beta}_{MQO_k} = \frac{\sum_{i=1}^{N} \tilde{x}_{ik} y_i}{(1 - R_k^2) \sum_{i=1}^{N} (x_{ik} - \bar{x}_k)^2}$$

Sendo $\tilde{x}_{ik}$ *o i-ésimo resíduo da regressão em que a k-ésima explicativa é a variável dependente e as variáveis independentes são as demais explicativas e* ${R^2}_k$ *é o coeficiente de determinação (não-ajustado) dessa regressão. Desconsidere, ao aplicar o teorema, o estágio em que y é regredida contra todas as explicativas exceto pela k-ésima.*

Referências

WOOLDRIDGE, Jeffrey M. **Introductory econometrics: A modern approach**. [s.l.]: Cengage learning, 2015.