

Gabarito - Lab 2 Aula 5

2022-09

Para esta lista, será preciso usar o R. O grupo precisa que ao menos um aluno saiba os comandos.

Usando o arquivo que está disponibilizado no Moodle para esta aula (“Base_Lista_Aula_3”), faça o seguinte:

Considere que esta base indica uma variável dependente (Y), uma explicativa (X1) e dois controles (C1 e C2).

Primeiro, carregue a base e os pacotes necessários usando:

```
# install.packages("readxl")
library(readxl)

# install.packages("tidyverse")
library(tidyverse)

# install.packages("stargazer")
library(stargazer)

dados <- read_excel("Aula 5 - Lab 2 - Base.xlsx")
```

1. Rode um modelo bivariado $Y = f(X1)$. Escreva a equação estimada;

```
dados %>%
  lm(Y ~ X1, data = .) %>%
  stargazer(header = FALSE)
```

Resultado na tabela 1.

$$Y = \alpha + \beta X1 + \epsilon$$

$$Y = -56.87 + 3.45 * X1 + \epsilon$$

2. Interprete os resultados, tanto em termos dos valores dos coeficientes, como em termos da significância estatística;

O modelo bivariado indica que para cada aumento de 1 unidade em X1, há um aumento médio de 3,45 em Y. Este resultado é estatisticamente significativo, apresentando um p-valor inferior a 0,01.

3. Rode um modelo multivariado $Y = f(X1, C1, C2)$. Escreva a nova equação estimada;

```
dados %>%
  lm(Y ~ X1 + C1 + C2, data = .) %>%
  stargazer(header = FALSE)
```

Tabela 1:

	<i>Dependent variable:</i>
	Y
X1	3.453*** (0.307)
Constant	-56.875*** (12.051)
Observations	200
R ²	0.391
Adjusted R ²	0.387
Residual Std. Error	60.228 (df = 198)
F Statistic	126.889*** (df = 1; 198)
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01

Tabela 2:

	<i>Dependent variable:</i>
	Y
X1	4.154*** (0.351)
C1	-1.604*** (0.487)
C2	4.052*** (1.520)
Constant	44.836 (41.575)
Observations	200
R ²	0.443
Adjusted R ²	0.434
Residual Std. Error	57.887 (df = 196)
F Statistic	51.898*** (df = 3; 196)
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01

Resultado na tabela 2.

$$Y = \alpha + \beta X1 + \beta C1 + \beta C2 + \epsilon$$

$$Y = 44.83 + 4.15 * X1 - 1.6 * C1 + 4.05 * C2 + \epsilon$$

4. Nas publicações, geralmente encontramos tabelas como a que está abaixo:

	(1)	(2)
Variáveis	Modelo 1	Modelo 2
X1	1.564*** (0.346)	1.554*** (0.227)
X2		0.408 (0.325)
Constante	-1.975*** (0.395)	-3.119*** (0.382)
Observações	372	616
Erros padrão em parêntesis		
*** p<0.01, ** p<0.05, * p<0.1		

Monte, mesmo que apenas em seu caderno, um rascunho de como ficaria uma tabela como a ilustrada aqui com os dados que você acabou de estimar.

Como estamos interessados no efeito que X1 tem sobre Y, vamos nos ater aos coeficientes estimados desta variável.

5. Compare os valores obtidos em cada um dos modelos do coeficiente de X1 e da constante. Comente;

```
modelo_bivariado <- dados %>%
  lm(Y ~ X1, data = .)

modelo_multivariado <- dados %>%
  lm(Y ~ X1 + C1 + C2, data = .)

stargazer(modelo_bivariado, modelo_multivariado, header = FALSE)
```

Resultado na tabela 3.

6. Se esses números forem diferentes entre si, a que você atribui a mudança entre os modelos? Considere a situação em que os valores dos coeficientes tenham mudado entre os modelos 1 e 2.

Adicionando as variáveis de controle 1 e 2, o valor estimado do efeito de X1 em Y aumenta de 3,45 para 4,15, ambos coeficientes com significância estatística ao nível de 99,9%. Entretanto, essa diferença nas estimativas de ponto podem ter sido causadas por “sorte”, precisamos realizar um teste estatístico para avaliar se são realmente diferentes.

7. Proponha um teste de hipóteses para avaliar se há diferença estatística relevante entre os coeficientes obtidos dos modelos.

Adicionando o parâmetro “ci = TRUE” do pacote e função *stargazer*, vemos o intervalo de confiança ao nível de 95% abaixo dos coeficientes estimados. Um teste possível para avaliar diferença estatística relevante entre os coeficientes é a comparação destes intervalos de confiança, para verificar se há interseção entre os intervalos.

8. Faça o teste e reporte os resultados. Qual sua conclusão?

Tabela 3:

	<i>Dependent variable:</i>	
	Y	
	(1)	(2)
X1	3.453*** (0.307)	4.154*** (0.351)
C1		-1.604*** (0.487)
C2		4.052*** (1.520)
Constant	-56.875*** (12.051)	44.836 (41.575)
Observations	200	200
R ²	0.391	0.443
Adjusted R ²	0.387	0.434
Residual Std. Error	60.228 (df = 198)	57.887 (df = 196)
F Statistic	126.889*** (df = 1; 198)	51.898*** (df = 3; 196)
<i>Note:</i> *p<0.1; **p<0.05; ***p<0.01		

```
stargazer(modelo_bivariado, modelo_multivariado, ci=TRUE, header = FALSE)
```

Resultado na tabela 4.

Ainda, caso o objetivo for somente obter os valores destes intervalos de confiança, podemos usar a função *confint*, que traz os valores mínimos e máximos de um intervalo de confiança de 95% para cada coeficiente do modelo.

```
confint(modelo_bivariado)
```

```
##                2.5 %    97.5 %
## (Intercept) -80.640647 -33.109209
## X1           2.848745   4.057846
```

```
confint(modelo_multivariado)
```

```
##                2.5 %    97.5 %
## (Intercept) -37.155709 126.8267856
## X1           3.461502   4.8460610
## C1          -2.563374  -0.6438162
## C2           1.054380   7.0495765
```

Tabela 4:

	<i>Dependent variable:</i>	
	Y	
	(1)	(2)
X1	3.453*** (2.852, 4.054)	4.154*** (3.466, 4.842)
C1		-1.604*** (-2.557, -0.650)
C2		4.052*** (1.073, 7.031)
Constant	-56.875*** (-80.495, -33.254)	44.836 (-36.649, 126.321)
Observations	200	200
R ²	0.391	0.443
Adjusted R ²	0.387	0.434
Residual Std. Error	60.228 (df = 198)	57.887 (df = 196)
F Statistic	126.889*** (df = 1; 198)	51.898*** (df = 3; 196)
<i>Note:</i> *p<0.1; **p<0.05; ***p<0.01		