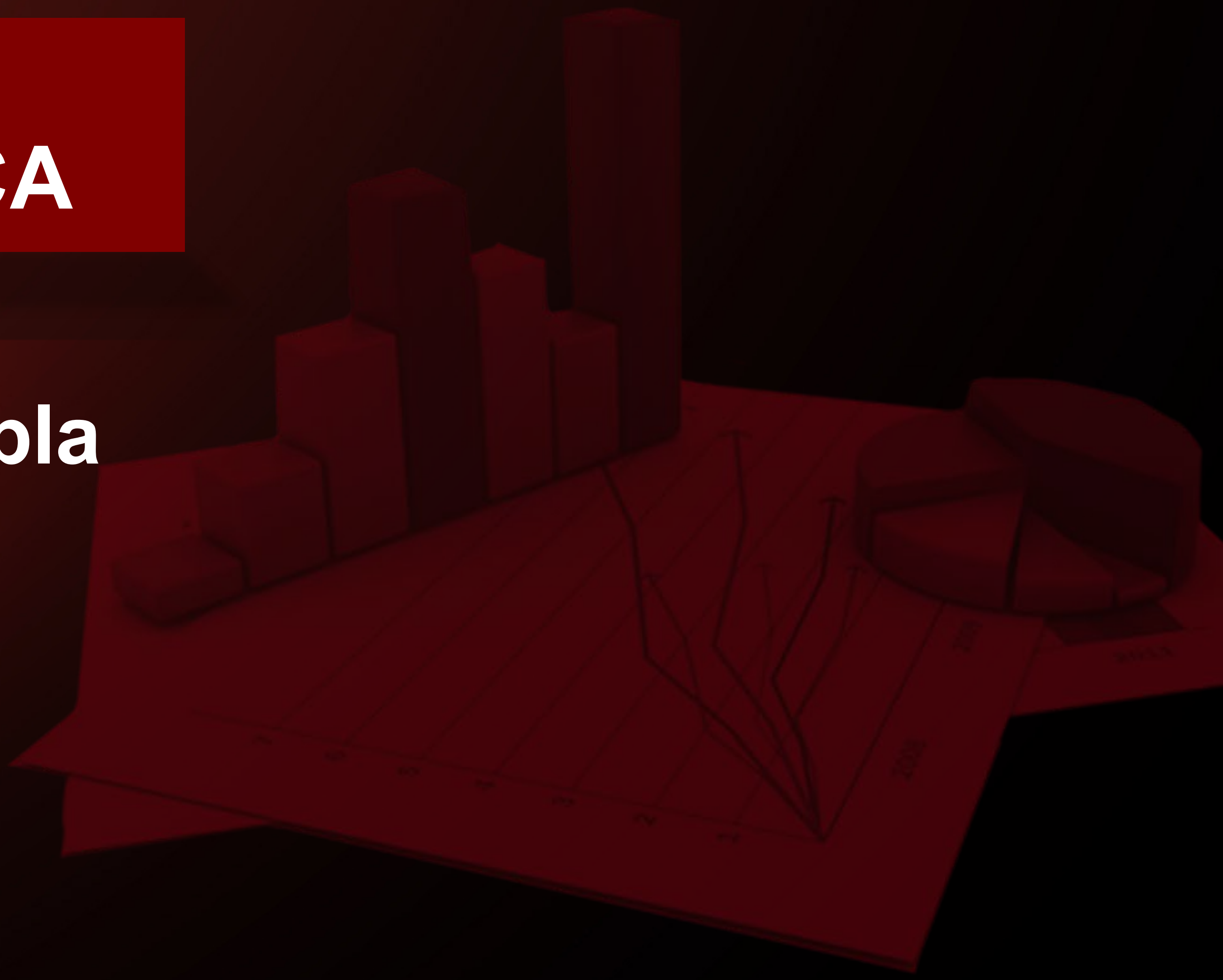
MODELAGEM E INFERÊNCIA ESTATÍSTICA



Modelo de regressão múltipla
Exercícios parte 1

O QUE VOU ESTUDAR HOJE?

Exercícios

Modelo de primeira ordem

Modelo de primeira ordem com interação

Variáveis numéricas
Variáveis categóricas

EXERCÍCIO 1

Sejam y = vendas em uma lanchonete (milhares de $), x_1 = número de lanchonetes concorrentes no raio de 1 km, x_2 = população no raio de 1 km (milhares de pessoas) e x_3 uma variável indicadora que vale 1 se a lanchonete tiver drive-thru e 0 se não tiver. Suponha que o modelo de regressão verdadeiro seja:

$$y = 10,00 - 1,2x_1 + 6,8x_2 + 15,3x_3 + \varepsilon$$

a) Qual é o valor médio das vendas quando há 2 lojas concorrentes, 8000 pessoas no raio de 1 km e a lanchonete dispõe de drive-thru?
b) Qual é o valor médio das vendas para uma lanchonete que não dispõe de drive-thru e que tem três concorrentes e 5000 pessoas no raio de 1 km?
c) Interprete β_3 .

EXERCÍCIO 1

Antes de resolver o exercício

Modelo de primeira ordem

$$y = 10,00 - 1,2x_1 + 6,8x_2 + 15,3x_3 + \varepsilon$$

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon$$

Variável resposta → y = vendas em uma lanchonete (milhares de $),

Variáveis preditoras:

- x_1 e x_2 ← numéricas
  - x_1 = número de lanchonetes concorrentes no raio de 1 km,
  - x_2 = população no raio de 1 km (milhares de pessoas) e
- x_3 ← categórica
  - x_3 =1 se a lanchonete tiver drive-thru
  - x_3 = 0 se não tiver drive-thru

EXERCÍCIO 1

a) Qual é o valor médio das vendas quando há 2 lojas concorrentes, 8000 pessoas no raio de 1km e a lanchonete dispõe de drive-thru?

Modelo $y = 10,00 - 1,2x_1 + 6,8x_2 + 15,3x_3 + \varepsilon$

- y = vendas em uma lanchonete (milhares de $),
- x_1 = número de lanchonetes concorrentes no raio de 1 km
  - x_1 =2
- x_2 = população no raio de 1 km (milhares de pessoas) e
  - x_2 = 8 (veja que na descrição indica que o valor é por milhares)
- x_3 =1 → a lanchonete possui drive-thru

Valor esperado: $\mu_{y \cdot x_1^* \cdot x_2^* \cdot x_3^*} = \mu_{y \cdot 2 \cdot 8 \cdot 1}$?

$$\mu_{y \cdot 2 \cdot 8 \cdot 1} = 10,00 - 1,2(2) + 6,8(8) + 15,3(1)$$

$\mu_{y \cdot 2 \cdot 8 \cdot 1}$ =77,3 isto é, 77300$ em vendas

EXERCÍCIO 1

b) Qual é o valor médio das vendas para uma lanchonete que não dispõe de drive-thru e que tem três concorrentes e 5000 pessoas no raio de 1km?

Modelo $\mathbf{y = 10,00 - 1,2x_1 + 6,8x_2 + 15,3x_3 + \varepsilon}$

- **y = vendas em uma lanchonete (milhares de $),**
- **x_1 = número de lanchonetes concorrentes no raio de 1 km**
  - **x_1 =3**
- **x_2 = população no raio de 1 km (milhares de pessoas) e**
  - **x_2 = 5 (veja que na descrição indica que o valor é por milhares)**
- **x_3 =0 → a lanchonete não tem drive-thru**

Valor esperado: $\mu_{y\cdot x_l^*\cdot x_2^*\cdot x_3^*} = \mu_{y\cdot 3\cdot 5\cdot 0}$?

$$\mu_{y\cdot 2\cdot 8\cdot 1} = 10,00 - 1,2(3) + 6,8(5) + 15,3(0)$$

$\mu_{y\cdot 2\cdot 8\cdot 1}$ **=40,4 isto é, 40400$ em vendas.**

EXERCÍCIO 1

b. Interprete β_3

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon$$

Modelo $\mathbf{y = 10,00 - 1,2x_1 + 6,8x_2 + 15,3x_3 + \varepsilon}$

Se manter fixos os valores de x_1 (quantidade de lanchonetes) x_2 (população no raio de 1 km), o que ocorre com as vendas?

$$y = \underbrace{\beta_0 + \underbrace{\beta_1 x_1 + \beta_2 x_2}_{\text{constante}}}_{\text{constante}} + \beta_3 x_3$$

Elas incrementam em 15,3 (isto é \$15300) se a lanchonete tiver drive-thu

EXERCÍCIO 2

Pesquisadores realizaram um estudo para verificar como y = deflexão máxima (mm) de vigas de cimento compostas com dureza ultrarreforçada era influenciada por x1 = proporção do alcance do cisalhamento e x2 = resistência à tensão de divisão (MPa), e uma variável x3= x1x2, resultando nos dados a seguir:
(“*Shear behavior of reinforced ultrahigh toughness cementitious composite beams without trans-verse reinforcement*”, J. of Materials in Civil Engr., 2012: 1283-1294):

x_1	x_2	y	x_1	x_2	y
2.04	3.55	3.11	3.08	3.62	3.36
2.04	6.07	3.26	3.08	5.89	6.49
3.06	3.55	3.89	4.11	3.62	2.72
3.06	6.07	10.25	4.11	5.89	12.48
4.08	3.55	3.11	2.01	6.18	2.82
4.08	6.16	13.48	3.02	6.18	5.19
2.06	3.62	3.94	4.03	6.18	8.04
2.06	6.16	3.53			

EXERCÍCIO 2

a) Se R^2=0,825 observe a seguir alguns resultados obtidos no python, descreva o modelo resultante e faça um teste de utilidade do modelo.

	coef	std err	t	P>|t|
Intercept	17.2787	7.167	2.411	0.035
x1	-6.3678	2.260	-2.817	0.017
x2	-3.6584	1.364	-2.682	0.021
x3	1.7067	0.431	3.956	0.002

b) O preditor de interação deve ser retido no modelo? Faça um teste de hipóteses utilizando um nível de significância de 0,05.

c) Se o modelo $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1^2 + \beta_4 x_2^2 + \beta_5 x_1 x_2$ fosse ajustado aos dados. Considerando que os preditores x1, x2 e x1x2 permanecem no modelo, os preditores quadráticos fornecem informações adicionais úteis? Declare e teste as hipóteses apropriadas.

EXERCÍCIO 2

a) Se R^2=0,825 observe a seguir alguns resultados obtidos no python, descreva o modelo resultante e faça um teste de utilidade do modelo.

	coef	std err	t	P>|t|
Intercept	17.2787	7.167	2.411	0.035
x1	-6.3678	2.260	-2.817	0.017
x2	-3.6584	1.364	-2.682	0.021
x3	1.7067	0.431	3.956	0.002

O resultado da regressão é uma reta do tipo:

$$y = 17{,}3 - 6{,}37x_1 - 3{,}66x_2 + 1{,}71x_3$$

De acordo com a descrição do enunciado

$$y = 17{,}3 - 6{,}37x_1 - 3{,}66x_2 + 1{,}71x_1x_2$$

EXERCÍCIO 2

a) Se R^2=0,825 observe a seguir alguns resultados obtidos no python, descreva o modelo resultante e faça um teste de utilidade do modelo.

O modelo com preditores de primeira ordem e interação:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + \varepsilon$$

$$y = 17,3 - 6,37x_1 - 3,66x_2 + 1,71x_1x_2$$

EXERCÍCIO 2

a) Se R^2=0,825 observe a seguir alguns resultados obtidos no python, descreva o modelo resultante e faça um teste de utilidade do modelo.

O modelo com preditores de primeira ordem e interação:

$$y = 17,3 - 6,37x_1 - 3,66x_2 + 1,71x_1x_2$$

Teste de utilidade do modelo:

- **Hipótese nula** $H_0: \beta_1 = \beta_2 = \beta_3 = 0$
- **Hipótese alternativa** H_a **pelo menos um** $\beta_i \neq 0$

EXERCÍCIO 2

a) Se R^2=0,825 observe a seguir alguns resultados obtidos no python, descreva o modelo resultante e faça um teste de utilidade do modelo.

Estatística de teste:

$$f = \frac{R^2/k}{(1 - R^2)/[n - (k + 1)]}$$

n=15, k=3 e R^2=0,825

$$f = \frac{0,825/3}{(1 - 0,825)/[15 - (3 + 1)]}$$

$$f = 17,29$$

EXERCÍCIO 2

a) Se R^2=0,825 observe a seguir alguns resultados obtidos no python, descreva o modelo resultante e faça um teste de utilidade do modelo.

- $F_{k,n-(k+1)} \rightarrow F_{3,11}$
- Analisar se $f \geq F_{crit}$ rejeitar H_0
- f=17,29 e $F_{crit} = 6,49$
- 17,29 $\geq 3,59$ SIM, portanto rejeitar a hipótese nula H_0 isto é, que existe relação entre y e pelo menos um dos parâmetros.

	α	1	2	3	4	5	6
10	.050	4.96	4.10	3.71	3.48	3.33	3.22
	.010	10.04	7.56	6.55	5.99	5.64	5.39
	.001	21.04	14.91	12.55	11.28	10.48	9.93
11	.100	3.23	2.86	2.66	2.54	2.45	2.39
	.050	4.84	3.98	3.59	3.36	3.20	3.09
	.010	9.65	7.21	6.22	5.67	5.32	5.07
	.001	19.69	13.81	11.56	10.35	9.58	9.05
12	.100	3.18	2.81	2.61	2.48	2.39	2.33
	.050	4.75	3.89	3.49	3.26	3.11	3.00
	.010	9.33	6.93	5.95	5.41	5.06	4.82
	.001	18.64	12.97	10.80	9.63	8.89	8.38

EXERCÍCIO 2

b) O preditor de interação deve ser retido no modelo? Faça um teste de hipóteses utilizando um nível de significância de 0,05.

Preditor de interação? 🤔 $\rightarrow \beta_3$

$x_3 = x_1 x_2$

k=3, n=15 e α=0,05

Do enunciado $\beta_3 = 1{,}71$ e $s_{\hat{\beta}_3} = 0{,}431$

	coef	std err	t	P>|t|
Intercept	17.2787	7.167	2.411	0.035
x1	-6.3678	2.260	-2.817	0.017
x2	-3.6584	1.364	-2.682	0.021
x3	$\boxed{1.7067}$	$\boxed{0.431}$	3.956	0.002

EXERCÍCIO 2

a) O preditor de interação deve ser retido no modelo?
Faça um teste de hipóteses utilizando um nível de significância de 0,05.

$\beta_3 = 1,71$ e $s_{\hat{\beta}_3} = 0,431$

$$t = \frac{\hat{\beta}_3 \cancel{- \beta_{30}}}{s_{\hat{\beta}_3}} = \frac{1,71}{0,431} = 3,968$$

	coef	std err	t	P>|t|
Intercept	17.2787	7.167	2.411	0.035
x1	-6.3678	2.260	-2.817	0.017
x2	-3.6584	1.364	-2.682	0.021
x3	$\boxed{1.7067}$	$\boxed{0.431}$	$\boxed{3.956}$	0.002

EXERCÍCIO 2

a) O preditor de interação deve ser retido no modelo?
Faça um teste de hipóteses utilizando um nível de significância de 0,05.

$\beta_3 = 1{,}71\ ,\ s_{\widehat{\beta}_3} = 0{,}431,\ t = 3{,}968$

$$t_{crit} = \mathbf{t}_{\frac{\alpha}{2},(\mathbf{n}-(\mathbf{k}+\mathbf{1}))} = \mathbf{t}_{0{,}025{,}11} = 2{,}201$$

$|t| \geq t_{crit} \rightarrow 3{,}968 \geq 2{,}201$ SIM, portantorejeitarhipótesenula

$\beta_3$deve ser mantido no modelo

	coef	std err	t	P>|t|
Intercept	17.2787	7.167	2.411	0.035
x1	-6.3678	2.260	-2.817	0.017
x2	-3.6584	1.364	-2.682	0.021
x3	1.7067	0.431	3.956	0.002

EXERCÍCIO 2

c) Se o modelo $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1^2 + \beta_4 x_2^2 + \beta_5 x_1 x_2$ fosse ajustado aos dados. Considerando que os preditores x1, x2 e x1x2 permanecem no modelo, os preditores quadráticos fornecem informações adicionais úteis? Declare e teste as hipóteses apropriadas.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1^2 + \beta_4 x_2^2 + \beta_5 x_1 x_2$$

- Hipótese nula $H_0: \beta_3 = \beta_4 = 0$
- Hipótese alternativa H_a pelo menos um $\beta_i \neq 0$
- Se $|t| \geq t_{crit}$ rejeitar H_0
  - n=15 k=5
  - $t_{crit} = t_{\frac{\alpha}{2},(n-(k+1))} = t_{0{,}025\,,\,9} = 2{,}262$

EXERCÍCIO 2

c) Se o modelo $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1^2 + \beta_4 x_2^2 + \beta_5 x_1 x_2$ fosse ajustado aos dados. Considerando que os preditores x1, x2 e x1x2 permanecem no modelo, os preditores quadráticos fornecem informações adicionais úteis? Declare e teste as hipóteses apropriadas.

```python
lstx1 = (2.04, 2.04, 3.06, 3.06, 4.08, 4.08, 2.06, 2.06,
         3.08, 3.08, 4.11, 4.11, 2.01, 3.02, 4.03)
lstx2 = (3.55, 6.07, 3.55, 6.07, 3.55, 6.16, 3.62, 6.16,
         3.62, 5.89, 3.62, 5.89, 6.18, 6.18, 6.18)
lstx3= np.multiply(lstx1, lstx1)
lstx4= np.multiply(lstx2, lstx2)
lstx5= np.multiply(lstx1, lstx2)
lsty= (3.11, 3.26, 3.89, 10.25, 3.11, 13.48, 3.94, 3.53,
         3.36, 6.49, 2.72, 12.48, 2.82, 5.19, 8.04)
# Construir o DataFrame e nomear as colunas
df = pd.DataFrame(list(zip(lstx1, lstx2, lstx3, lstx4, lstx5,lsty)),
                  columns =["x1","x2","x3","x4","x5","y"])
```

EXERCÍCIO 2

c) Se o modelo $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1^2 + \beta_4 x_2^2 + \beta_5 x_1 x_2$ fosse ajustado aos dados. Considerando que os preditores x1, x2 e x1x2 permanecem no modelo, os preditores quadráticos fornecem informações adicionais úteis? Declare e teste as hipóteses apropriadas.

```
regmul = smf.ols('y ~ x1 + x2 + x3 + x4 + x5', data = df)
res = regmul.fit()
print(res.summary())
```

EXERCÍCIO 2

c) Se o modelo $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1^2 + \beta_4 x_2^2 + \beta_5 x_1 x_2$ fosse ajustado aos dados. Considerando que os preditores x1, x2 e x1x2 permanecem no modelo, os preditores quadráticos fornecem informações adicionais úteis? Declare e teste as hipóteses apropriadas.

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                      y   R-squared:                       0.845
Model:                            OLS   Adj. R-squared:                  0.759
Method:                 Least Squares   F-statistic:                     9.807
Date:                Fri, 01 Apr 2022   Prob (F-statistic):            0.00192
Time:                        02:35:54   Log-Likelihood:                -26.215
No. Observations:                  15   AIC:                             64.43
Df Residuals:                       9   BIC:                             68.68
Df Model:                           5
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept    -34.3227     48.933     -0.701      0.501    -145.017      76.372
x1            -6.5682      6.364     -1.032      0.329     -20.964       7.827
x2            19.3469     21.662      0.893      0.395     -29.657      68.351
x3             0.0585      0.954      0.061      0.952      -2.100       2.217
x4            -2.3586      2.217     -1.064      0.315      -7.373       2.656
x5             1.6549      0.452      3.661      0.005       0.632       2.677
------------------------------------------------------------------------------
```

EXERCÍCIO 2

c) Se o modelo $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1^2 + \beta_4 x_2^2 + \beta_5 x_1 x_2$ fosse ajustado aos dados. Considerando que os preditores x1, x2 e x1x2 permanecem no modelo, os preditores quadráticos fornecem informações adicionais úteis? Declare e teste as hipóteses apropriadas.

Se $|t| \geq t_{crit}$ rejeitar $H_0 \rightarrow |t_i| \geq 2,262$? ? NÃO rejeitar H_0

Os preditores β_3 e $\beta_4$não proveem informação relevante

	coef	std err	t	P>|t|
Intercept	-34.3227	48.933	-0.701	0.501
x1	-6.5682	6.364	-1.032	0.329
x2	19.3469	21.662	0.893	0.395
x3	0.0585	0.954	0.061	0.952
x4	-2.3586	2.217	-1.064	0.315
x5	1.6549	0.452	3.661	0.005

MODELAGEM E INFERÊNCIA ESTATÍSTICA

Modelo de regressão múltipla
Exercícios parte 1