MODELAGEM E INFERÊNCIA ESTATÍSTICA

Modelo de regressão múltipla
Exercícios parte 2

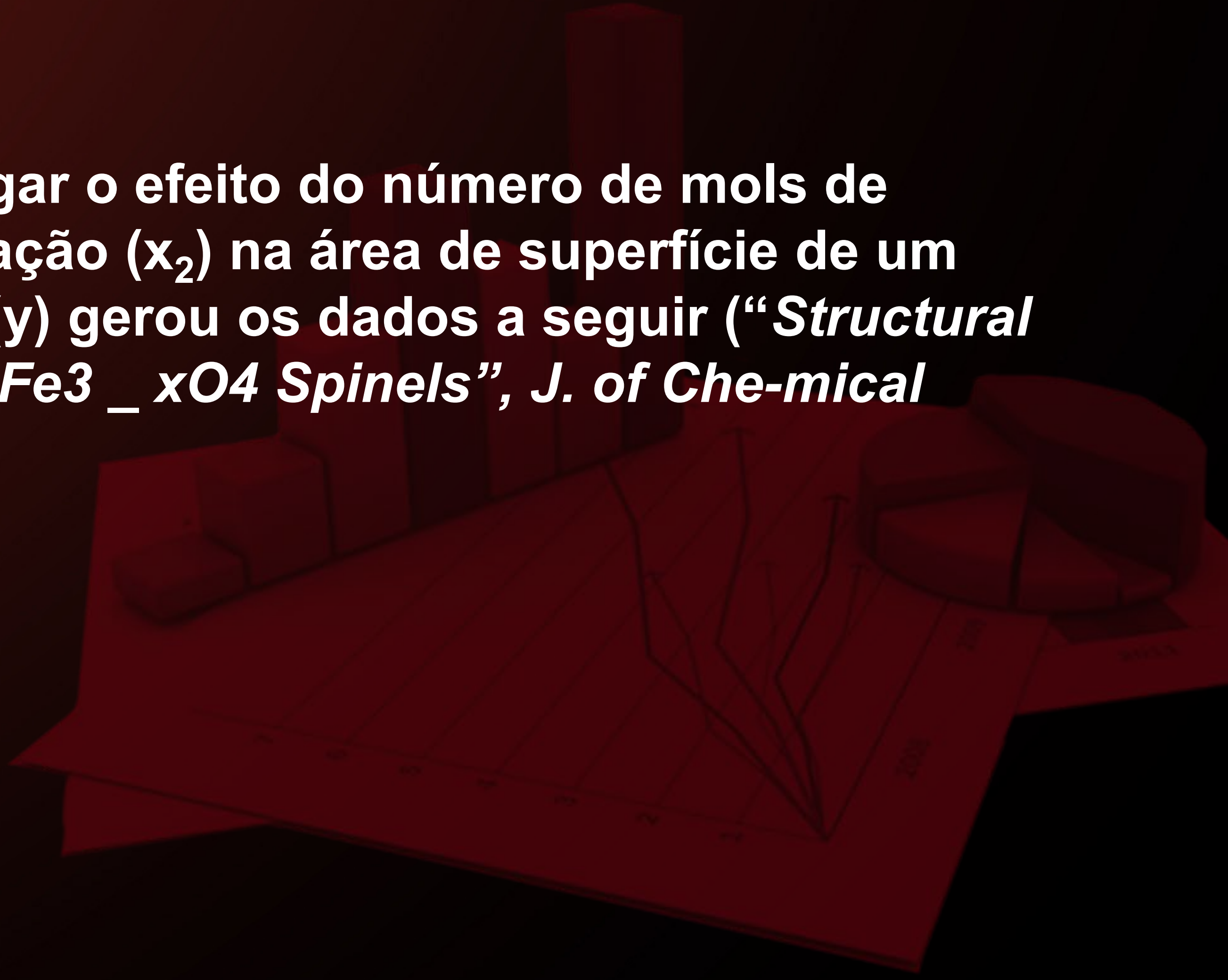
O QUE VOU ESTUDAR HOJE?

- Exercício
- Modelo de primeira ordem com interação
- Intervalos de confiança e previsão

EXERCÍCIO

Um experimento realizado para investigar o efeito do número de mols de cobalto (x_1) e a temperatura de calcificação (x_2) na área de superfície de um catalisador de hidróxido ferro-cobalto (y) gerou os dados a seguir (“*Structural changes and surface properties of CoxFe3 _ xO4 Spinels*”, *J. of Che-mical Tech. and Biotech*., 1994: 161-170).

x_1	.6	.6	.6	.6	.6	1.0	1.0
x_2	200	250	400	500	600	200	250
y	90.6	82.7	58.7	43.2	25.0	127.1	112.3
x_1	1.0	1.0	1.0	2.6	2.6	2.6	2.6
x_2	400	500	600	200	250	400	500
y	19.6	17.8	9.1	53.1	52.0	43.4	42.4
x_1	2.6	2.8	2.8	2.8	2.8	2.8	
x_2	600	200	250	400	500	600	
y	31.6	40.9	37.9	27.5	27.3	19.0	



EXERCÍCIO

a) Realize o processo de regressão no python aplicando um modelo com preditores de primeira ordem com interação.
b) Calcule uma previsão para o valor da área de superfície em relação a um teor de cobalto de 2,6, com temperatura de 250, e calcule o valor do resíduo correspondente.
c) Visto que $\beta_1 = -46{,}0$, é apropriado concluir que, se o volume de cobalto sofrer um aumento de 1 unidade, mantendo-se fixos os valores dos outros preditores, é possível esperar que a área de superfície diminua em torno de 46 unidades? Explique seu raciocínio.
d) Parece haver uma relação linear útil entre y e os preditores?
e) O desvio padrão estimado de $s_{\hat{Y}}$ quando o número de mols é 2,0 e a temperatura de calcificação é 500 é $s_{\hat{Y}} = 4{,}69$. Calcule um intervalo de confiança de 95% para o valor médio da área de superfície sob essas circunstâncias.

EXERCÍCIO

a) Realize o processo de regressão no python aplicando um modelo com preditores de primeira ordem com interação.

O modelo com preditores de primeira ordem com interação:

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

x_1 = o efeito do número de mols de cobalto

x_2 = temperatura de calcificação

$x_3 = x_1\ x_2$

y = área de superfície de um catalisador de hidróxido ferro-cobalto

EXERCÍCIO

a) Realize o processo de regressão no python aplicando um modelo com preditores de primeira ordem com interação.

Definir e condicionar os dados

```
lstx1 = (0.6, 0.6, 0.6, 0.6, 0.6, 1.0, 1.0, 1.0,
          1.0, 1.0, 2.6, 2.6, 2.6, 2.6, 2.6, 2.8,
          2.8, 2.8, 2.8, 2.8)
lstx2 = (200, 250, 400, 500, 600, 200, 250, 400,
          500, 600, 200, 250, 400, 500, 600, 200,
          250, 400, 500, 600)
lstx3= np.multiply(lstx1, lstx2)
lsty= (90.6, 82.7, 58.7, 43.2, 25.0, 127.1, 112.3,
        19.6, 17.8, 9.1, 53.1, 52.0, 43.4, 42.4, 31.6,
        40.9, 37.9, 27.5, 27.3, 19.0)
# Construir o DataFrame e nomear as colunas
df = pd.DataFrame(list(zip(lstx1, lstx2, lstx3,lsty)),
                  columns =["x1","x2","x3","y"])
df.head(20)
```

	x1	x2	x3	y
0	0.6	200	120.0	90.6
1	0.6	250	150.0	82.7
2	0.6	400	240.0	58.7
3	0.6	500	300.0	43.2
4	0.6	600	360.0	25.0
5	1.0	200	200.0	127.1
6	1.0	250	250.0	112.3
7	1.0	400	400.0	19.6

EXERCÍCIO

a) **Realize o processo de regressão no python aplicando um modelo com preditores de primeira ordem com interação.**

Realizar a regressão

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

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                      y   R-squared:                       0.780
Model:                            OLS   Adj. R-squared:                  0.739
Method:                 Least Squares   F-statistic:                     18.92
Date:                Fri, 01 Apr 2022   Prob (F-statistic):           1.64e-05
Time:                        04:34:06   Log-Likelihood:                -82.063
No. Observations:                  20   AIC:                             172.1
Df Residuals:                      16   BIC:                             176.1
Df Model:                           3
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept    185.4857     21.197      8.750      0.000     140.549     230.422
x1           -45.9695     10.612     -4.332      0.001     -68.466     -23.473
x2            -0.3015      0.051     -5.942      0.000      -0.409      -0.194
x3             0.0888      0.025      3.496      0.003       0.035       0.143
==============================================================================
```

EXERCÍCIO

a) Realize o processo de regressão no python aplicando um modelo com preditores de primeira ordem com interação.

Realizar a regressão

	coef
Intercept	$\boxed{185.4857}$
x1	$\boxed{-45.9695}$
x2	$\boxed{-0.3015}$
x3	$\boxed{0.0888}$

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

$$y = 185{,}4857 - 45{,}9695 x_1 - 0.3015 x_2 + 0{,}0888 x_1 x_2 + \varepsilon$$

EXERCÍCIO

b) Calcule uma previsão para o valor da área de superfície em relação a um teor de cobalto de 2,6, temperatura de 250, e calcule o valor do resíduo correspondente.

Calcule uma previsão para o valor da área de superfície → y=? → $\mu_{y \cdot x_1^* \cdot x_2^* \cdot x_3^*}$
em relação a um teor de cobalto de 2,6 → x_1 = 2,6
e a temperatura de 250 → x_2= 250
e calcule o valor do resíduo correspondente → e

Valor esperado: $\mu_{y \cdot x_1^* \cdot x_2^* \cdot x_3^*} = \mu_{y \cdot 2,6 \cdot 250 \cdot 650}$?

$$\hat{y} = 185,4857 - 45,9695(2,6) - 0.3015(250) + 0,0888(2,6)(250)$$

Ou $\hat{y} = 185,4857 - 45,9695(2,6) - 0.3015(250) + 0,0888(650)$

$$\mu_{y \cdot 2,6 \cdot 250 \cdot 650} = 48,31$$

EXERCÍCIO

b) Calcule uma previsão para o valor da área de superfície em relação a um teor de cobalto de 2,6, temperatura de 250, e calcule o valor do resíduo correspondente.

e calcule o valor do resíduo correspondente → e

```
#@title Resposta b)
auxi=df.iloc[10:15,:]
auxi.head()
```

	x1	x2	x3	y
10	2.6	200	520.0	53.1
11	2.6	250	650.0	52.0
12	2.6	400	1040.0	43.4
13	2.6	500	1300.0	42.4
14	2.6	600	1560.0	31.6

$$\mu_{y\cdot 2,6\cdot 250\cdot 650} = 48{,}31$$

$$e = y - \hat{y}$$

$$e = 52 - 48,31$$

$$e = 3,69$$

EXERCÍCIO

b) Calcule uma previsão para o valor da área de superfície em relação a um teor de cobalto de 2,6, temperatura de 250, e calcule o valor do resíduo correspondente.

Todos os resíduos

```
resi=res.resid
df1 = pd.DataFrame(list(zip(lstx1, lstx2, lstx3, lsty, resi)),
                  columns =["x1","x2","x3","y","e"])
auxi2=df1.iloc[10:15,:]
auxi2.head()
```

	x1	x2	x3	y	e
10	2.6	200	520.0	53.1	1.258726
11	2.6	250	650.0	52.0	3.689690
12	2.6	400	1040.0	43.4	5.682581
13	2.6	500	1300.0	42.4	11.744508
14	2.6	600	1560.0	31.6	8.006435

EXERCÍCIO

c) Visto que $\beta_1 = -46{,}0$, é apropriado concluir que, se o teor de cobalto sofrer um aumento de 1 unidade, mantendo-se fixos os valores dos outros preditores, é possível esperar que a área de superfície diminua em torno de 46 unidades? Explique seu raciocínio.

O modelo com preditores de primeira ordem com interação:

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

$$y = 185{,}4857 - 45{,}9695 x_1 - 0.3015 x_2 + \boxed{0{,}0888 x_1 x_2}$$

x_1 = o efeito do número de mols de cobalto

x_2 = temperatura de calcificação

$x_3 = x_1\, x_2$

y = área de superfície de um catalisador de hidróxido ferro-cobalto

Não é correto porque não se pode incrementar o cobalto mantendo x_3 fixo

EXERCÍCIO

d) Há uma relação linear útil entre y e os preditores?

Teste de utilidade do modelo:

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

$$y = 185,4857 - 45,9695x_1 - 0.3015x_2 + 0,0888x_1x_2$$

Usar F

```
==============================================================================
Dep. Variable:                      y   R-squared:                       0.780
Model:                            OLS   Adj. R-squared:                  0.739
Method:                 Least Squares   F-statistic:                     18.92
Date:                Fri, 01 Apr 2022   Prob (F-statistic):           1.64e-05
Time:                        04:35:44   Log-Likelihood:                -82.063
No. Observations:                  20   AIC:                             172.1
Df Residuals:                      16   BIC:                             176.1
Df Model:                           3
Covariance Type:            nonrobust
==============================================================================
```

n=20,
k=3,
R^2=0,78

```
F=res.fvalue
k=res.df_model # grau do modelo
n=res.nobs # num. amostras
```

EXERCÍCIO

d) Há uma relação linear útil entre y e os preditores?

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

n=20, k=3 e R^2=0,78

$$f = \frac{0,78/3}{(1-0,78)/[20-(3+1)]}$$

$$f = 18,90$$

```
==========================================
R-squared:                           0.780
Adj. R-squared:                      0.739
F-statistic:                         18.92
Prob (F-statistic):               1.64e-05
```

```python
#@title Resposta c)
import scipy.stats
F=res.fvalue
k=res.df_model # grau do modelo
n=res.nobs # num. amostras
dfn=k
dfd=n-(k+1)
alpha = 0.05 #nível de confiança.
F_critico=scipy.stats.f.ppf(1-alpha, dfn, dfd)
print("F_crit=",F_critico) #tabela F-dist
```

```
F_crit= 3.238871517453585
```

- $F_{k,n-(k+1)} \rightarrow F_{3,16}$
- Analisar se f $\geq$ F_{crit} rejeitar H_0
- f=18,90 e $F_{crit} = 3,24$
- 18,90 $\geq 3,24$SIM, portanto rejeitar a hipótese nula H_0 isto é, que existe relação entre y e pelo menos um dos parâmetros ($\beta_i \neq 0$).

EXERCÍCIO

e) O desvio padrão estimado de $s_{\hat{Y}}$ quando o número de mols é 2,0 e a temperatura de calcificação é 500 é $s_{\hat{Y}}$ = 4,69. Calcule um intervalo de confiança de 95% para o valor médio da área de superfície sob essas circunstâncias.

Valor da área de superfície → y=? → $\mu_{y\cdot x_1^*\cdot x_2^*\cdot x_3^*}$

Um teor de cobalto de 2,0 → x_1 = 2,0

e a temperatura de calcificação de 500 → x_2= 500

Valor esperado: $\mu_{y\cdot x_1^*\cdot x_2^*\cdot x_3^*} = \mu_{y\cdot 2\cdot 500\cdot 1000}$?

$$\hat{y} = 185,4857 - 45,9695(2) - 0.3015(500) + 0,0888(2)(500)$$

$$\mu_{y\cdot 2\cdot 500\cdot 1000}=31,5967$$

EXERCÍCIO

e) O desvio padrão estimado de $s_{\hat{Y}}$ quando o número de mols é 2,0 e a temperatura de calcificação é 500 é $s_{\hat{Y}}$ = 4,69. Calcule um intervalo de confiança de 95% para o valor médio da área de superfície sob essas circunstâncias.

Valor da área de superfície → y=? → $\mu_{y \cdot x_1^* \cdot x_2^* \cdot x_3^*}$

Um teor de cobalto de 2,0 → x_1 = 2,0

e a temperatura de calcificação de 500 → x_2= 500

Valor esperado: $\mu_{y \cdot x_1^* \cdot x_2^* \cdot x_3^*} = \mu_{y \cdot 2 \cdot 500 \cdot 1000}$?

$$\hat{y} = 185,4857 - 45,9695(2) - 0.3015(500) + 0,0888(2)(500)$$

$$\mu_{y \cdot 2 \cdot 500 \cdot 1000} = 31,5967$$

EXERCÍCIO

e) O desvio padrão estimado de $s_{\hat{Y}}$ quando o número de mols é 2,0 e a temperatura de calcificação é 500 é $s_{\hat{Y}}$ = 4,69. Calcule um intervalo de confiança de 95% para o valor médio da área de superfície sob essas circunstâncias.

$\mu_{y \cdot 2 \cdot 500 \cdot 1000}$=31,5967 e $s_{\hat{Y}}$ = 4,69.

IC: $\hat{y} \pm t_{\frac{\alpha}{2},(n-(k+1))} s_{\hat{y}}$ onde

$t_{\frac{\alpha}{2},(n-(k+1))} = t_{0{,}025,(20-(3+1))}$=2,120

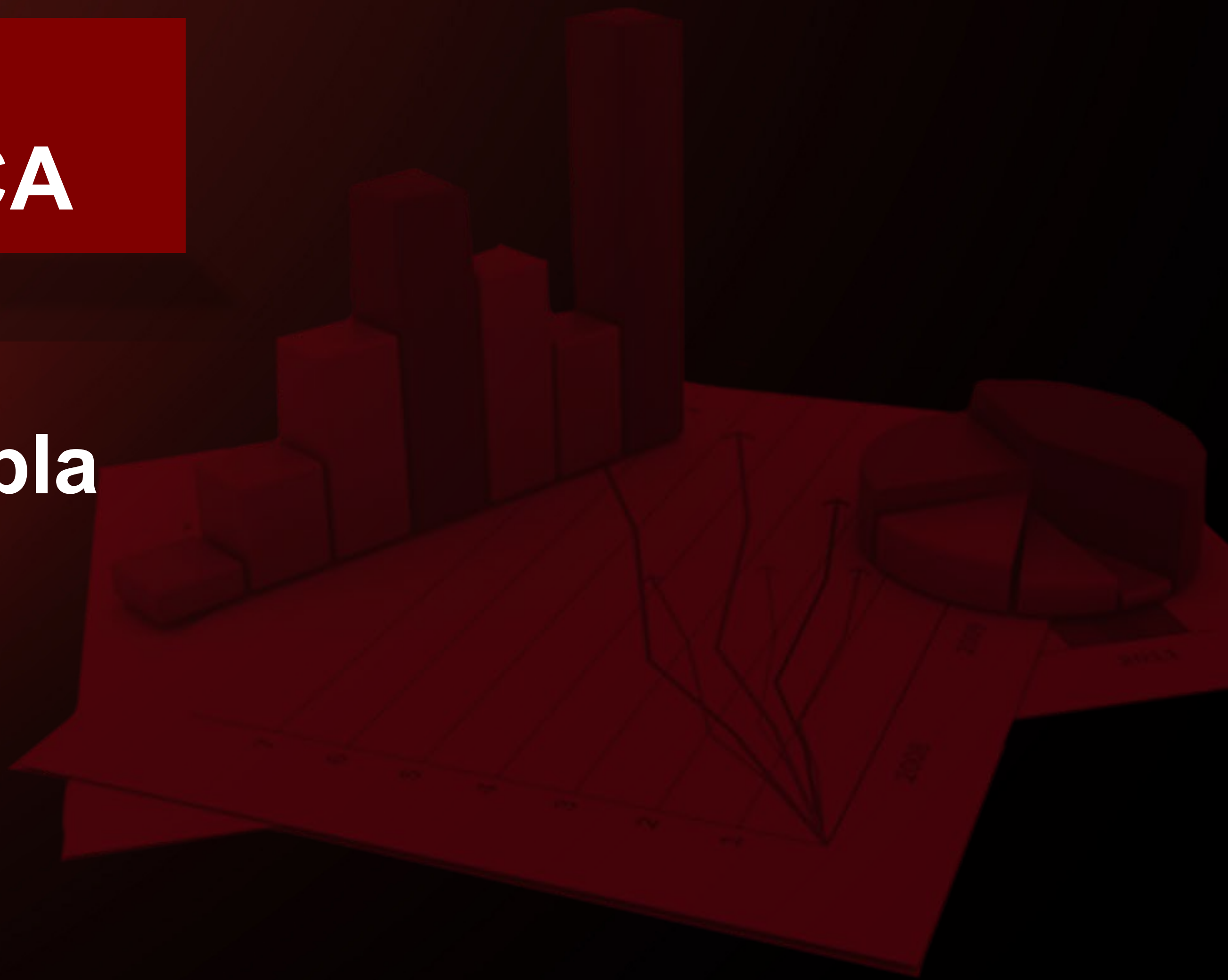
$$31{,}5967 \pm 2{,}120(4{,}69)$$

$$21{,}6539 < \mu_{y \cdot 2 \cdot 500 \cdot 1000} < 41{,}5395$$

```
#@title Resposta e)
#usar a tabela tstudent pata t
from scipy.stats import t
alpha = 0.05  # significia = 5%
df = n-(k+1) # graus de liberdade
v - t.ppf(1 - alpha/2, df)
tt-v
print(f't_crit=: {v}')
```

```
t_crit=: 2.1199052992210112
```

MODELAGEM E INFERÊNCIA ESTATÍSTICA



Modelo de regressão múltipla
Exercícios parte 2