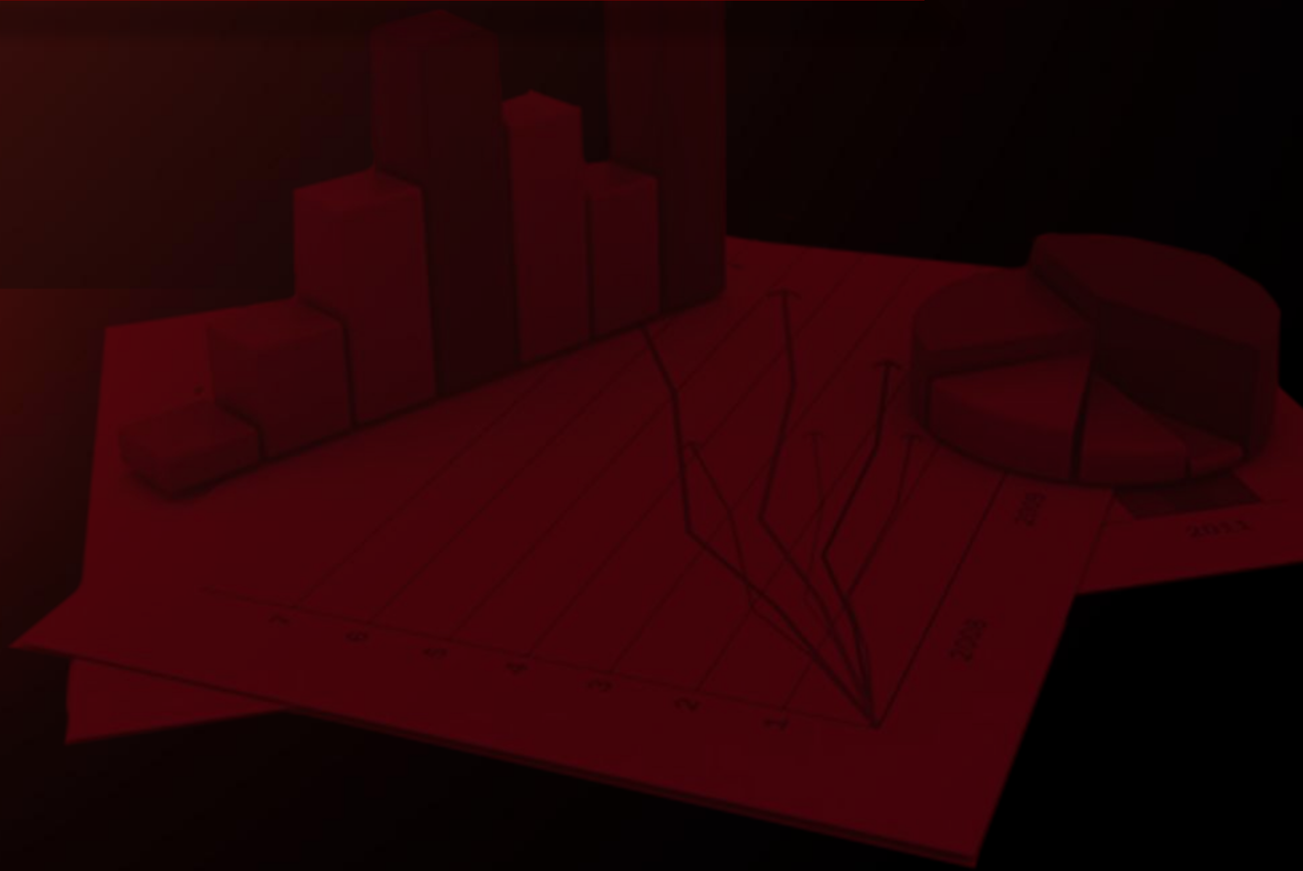
MODELAGEM E INFERÊNCIA ESTATÍSTICA



Adequações do modelo e modelos não lineares
Exemplos

O QUE VOU APRENDER HOJE?

- Exemplos práticos
- Adequação do Modelo
- Modelo intrinsecamente linear

ADEQUAÇÃO DO MODELO: EXEMPLO 1

Dados sobre x = taxa de liberação na área do queimador e y = emissões de NOx. Observe os dados e os valores ajustados, os resíduos e os resíduos padronizados. A reta de regressão estimada é y = –45,55 + 1,71x e r2 = 0,961.

```
[4]
lstx2 = (100, 125, 125, 150, 150, 200, 200, 250, 250, 300, 300, 350, 400, 400)
lsty2= (150, 140, 180, 210, 190, 320, 280, 400, 430, 440, 390, 600, 610, 670 )
# Construir o DataFrame e nomear as colunas
df = pd.DataFrame(list(zip(lstx2, lsty2)),
                  columns =["x","y"])
x=df['x']
y=df['y']
```

```
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const        -45.5519     25.468     -1.789      0.099    -101.041       9.938
x              1.7114      0.100     17.168      0.000       1.494       1.929
==============================================================================
```

ADEQUAÇÃO DO MODELO: EXEMPLO 1

	x	y	yp	e	e*
0	100	150	125.591327	24.408673	0.745861
1	125	140	168.377135	-28.377135	-0.843355
2	125	180	168.377135	11.622865	0.345426
3	150	210	211.162943	-1.162943	-0.033841
4	150	190	211.162943	-21.162943	-0.615821
5	200	320	296.734560	23.265440	0.660343
6	200	280	296.734560	-16.734560	-0.474977
7	250	400	382.306176	17.693824	0.500064
8	250	430	382.306176	47.693824	1.347926
9	300	440	467.877792	-27.877792	-0.800463
10	300	390	467.877792	-77.877792	-2.236128
11	350	600	553.449409	46.550591	1.388367
12	400	610	639.021025	-29.021025	-0.924322
13	400	670	639.021025	30.978975	0.986682

```
# Construir o DataFrame e nomear as colunas
df = pd.DataFrame(list(zip(x, y, y_pred,resi,standardized_residuals)),
                 columns =["x","y","yp","e","e*"])
df.head(14)
```

e* não é múltiplo de e devido ao fato de que as variâncias residuais são diferentes umas das outras.

ADEQUAÇÃO DO MODELO: EXEMPLO 2

Base de dados GT Auto

```
#Observar os dados
import io
#import pandas as pd
df = pd.read_csv(io.StringIO(uploaded['carst.csv'].decode('utf-8')))
df.head(30)

```

	brand/model/year	cap_vol	consumo	power	weight	cemm	nu_cy	Etype
0	Gurgel BR800 0.8 1991	792	12.0	33	650	34.4	2	0
1	FIAT UNO Mille EP 1996	994	10.4	58	870	18.6	4	0
2	Hyundai HB20 Sense 2020	1000	12.8	80	989	14.5	3	0
3	FIAT Strada 1.4 2016	1368	10.3	86	1084	12.5	4	0
4	VolksWagen GOL 1.6 2015	1598	10.5	104	961	9.8	4	0
5	Chevrolet Cruze LTZ 1.8 2016	1796	8.5	144	1427	10.2	4	0
6	Honda Civic EXR 2016	1997	9.5	155	1294	10.9	4	0
7	Ford Focus 2.0 GLX 2012	1999	9.2	148	1347	10.4	4	0
8	BMW 325i 3.0 2012	2996	6.5	218	1460	7.1	6	0
9	AUDI A4 3.2 V6 Fsi 2011	3197	7.1	269	1610	6.4	6	0
10	Mercedes-Benz CLS 350 3.5 V6 2012	3498	6.6	306	1735	6.1	6	0
11	Mercedes-Benz CLS 500 5.5 V8 2007	5461	4.2	388	1760	5.4	8	0

ADEQUAÇÃO DO MODELO: EXEMPLO 2

Base de dados GT Auto

11	Mercedes-Benz CLS 500 5.5 V8 2007	5461	4.2	388	1760	5.4	8	0
12	Chevrolet Camaro SS 6.2 V8 2018	6162	6.4	461	1709	4.2	8	0
13	Pagani Zonda F 7.3 V12 2006	7291	3.0	602	1230	3.6	12	0
14	Volkswagen Polo Highline 1.0 Tsi 2021	999	11.4	128	1147	9.6	3	1
15	Chevrolet Tracker Premier 1.2 T 2021	1199	11.2	133	1271	9.4	3	1
16	Chevrolet Cruze LTZ 1.4 T 2021	1399	11.0	153	1321	9.0	4	1
17	Honda Civic Touring 1.5 T 2021	1498	11.0	173	1329	8.6	4	1
18	Peugeot 306 Griffe 1.6 THP 2019	1598	10.6	173	1392	8.3	4	1
19	Audi A4 Attraction 1.8 TSFi 2015	1798	8.7	170	1470	8.3	4	1
20	BMW 320i Sport 2.0 T 2020	1998	10.5	184	1460	7.1	4	1
21	Porsche Panamera 3.0 V6 2018	2995	7.8	330	1815	5.7	6	1
22	Honda NSX 3.5 V6 T 2017	3493	7.0	507	1725	3.3	6	1
23	Nissan GTR Nismo 3.8 V6 T 2015	3799	5.0	608	1729	2.7	6	1
24	Mercedes-Benz E63 S AMG 4.0 V8 2021	3982	6.1	612	1935	3.4	8	1
25	Aston Martin DB11 5.2 V12 T 2017	5204	2.6	608	1770	3.9	12	1
26	Pagani Huayra 6.0 V12 T 2014	5980	3.0	730	1350	3.3	12	1
27	Bugatti Chiron 8.0 W16 2016	7993	3.8	1500	1998	2.5	16	1

ADEQUAÇÃO DO MODELO: EXEMPLO 2

Base de dados GT Auto Motor aspirado

```
1 df1 = df[df.Etype != 1]
2 df1.head(15)
```

	brand/model/year	cap_vol	consumo	power	weight	cemm	nu_cy	Etype
0	Gurgel BR800 0.8 1991	792	12.0	33	650	34.4	2	0
1	FIAT UNO Mille EP 1996	994	10.4	58	870	18.6	4	0
2	Hyundai HB20 Sense 2020	1000	12.8	80	989	14.5	3	0
3	FIAT Strada 1.4 2016	1368	10.3	86	1084	12.5	4	0
4	VolksWagen GOL 1.6 2015	1598	10.5	104	961	9.8	4	0
5	Chevrolet Cruze LTZ 1.8 2016	1796	8.5	144	1427	10.2	4	0
6	Honda Civic EXR 2016	1997	9.5	155	1294	10.9	4	0
7	Ford Focus 2.0 GLX 2012	1999	9.2	148	1347	10.4	4	0
8	BMW 325i 3.0 2012	2996	6.5	218	1460	7.1	6	0
9	AUDI A4 3.2 V6 Fsi 2011	3197	7.1	269	1610	6.4	6	0
10	Mercedes-Benz CLS 350 3.5 V6 2012	3498	6.6	306	1735	6.1	6	0
11	Mercedes-Benz CLS 500 5.5 V8 2007	5461	4.2	388	1760	5.4	8	0
12	Chevrolet Camaro SS 6.2 V8 2018	6162	6.4	461	1709	4.2	8	0
13	Pagani Zonda F 7.3 V12 2006	7291	3.0	602	1230	3.6	12	0

ADEQUAÇÃO DO MODELO: EXEMPLO 2

Reta de regressão

```
#adicionar uma constante preditora
x = sm.add_constant(x)
#fit linear regression model
modelo = sm.OLS(y, x).fit() #atenção esta variável será usada no IC
print(modelo.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                consumo   R-squared:                       0.834
Model:                            OLS   Adj. R-squared:                  0.820
Method:                 Least Squares   F-statistic:                     60.33
Date:                Tue, 15 Feb 2022   Prob (F-statistic):           5.08e-06
Time:                        17:34:00   Log-Likelihood:                -21.429
No. Observations:                  14   AIC:                             46.86
Df Residuals:                      12   BIC:                             48.14
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const         11.9686      0.566     21.145      0.000      10.735      13.202
cap_vol       -0.0013      0.000     -7.767      0.000      -0.002      -0.001
==============================================================================
Omnibus:                        1.442   Durbin-Watson:                   1.883
Prob(Omnibus):                  0.486   Jarque-Bera (JB):                1.002
Skew:                           0.615   Prob(JB):                        0.606
Kurtosis:                       2.548   Cond. No.                     6.12e+03
==============================================================================
```

```
# parâmetros da reta
b=modelo.params
b0=b[0] #intercepto
b1=b[1] #inclinação
print('y= {0}{1} x'.format(b0,b1))
```

```
y= 11.968578736702085-0.0012593116220535796 x
```

$$y = 11,9686 - 0,001259x$$

x: Capacidade volumétrica.

y: Consumo.

ADEQUAÇÃO DO MODELO: EXEMPLO 2

Resíduos e resíduos padronizados

	x	y	yc	e	e*	e/e*
0	792	12.0	10.971204	1.028796	0.923383	1.114160
1	994	10.4	10.716823	-0.316823	-0.281988	1.123532
2	1000	12.8	10.709267	2.090733	1.860422	1.123795
3	1368	10.3	10.245840	0.054160	0.047584	1.138197
4	1598	10.5	9.956199	0.543801	0.474716	1.145529
5	1796	8.5	9.706855	-1.206855	-1.048677	1.150836
6	1997	9.5	9.453733	0.046267	0.040048	1.155287
7	1999	9.2	9.451215	-0.251215	-0.217440	1.155327
8	2996	6.5	8.195681	-1.695681	-1.457151	1.163696
9	3197	7.1	7.942559	-0.842559	-0.724684	1.162657
10	3498	6.6	7.563507	-0.963507	-0.831048	1.159388
11	5461	4.2	5.091478	-0.891478	-0.821422	1.085286
12	6162	6.4	4.208701	2.191299	2.119047	1.034097
13	7291	3.0	2.786938	0.213062	0.232426	0.916688

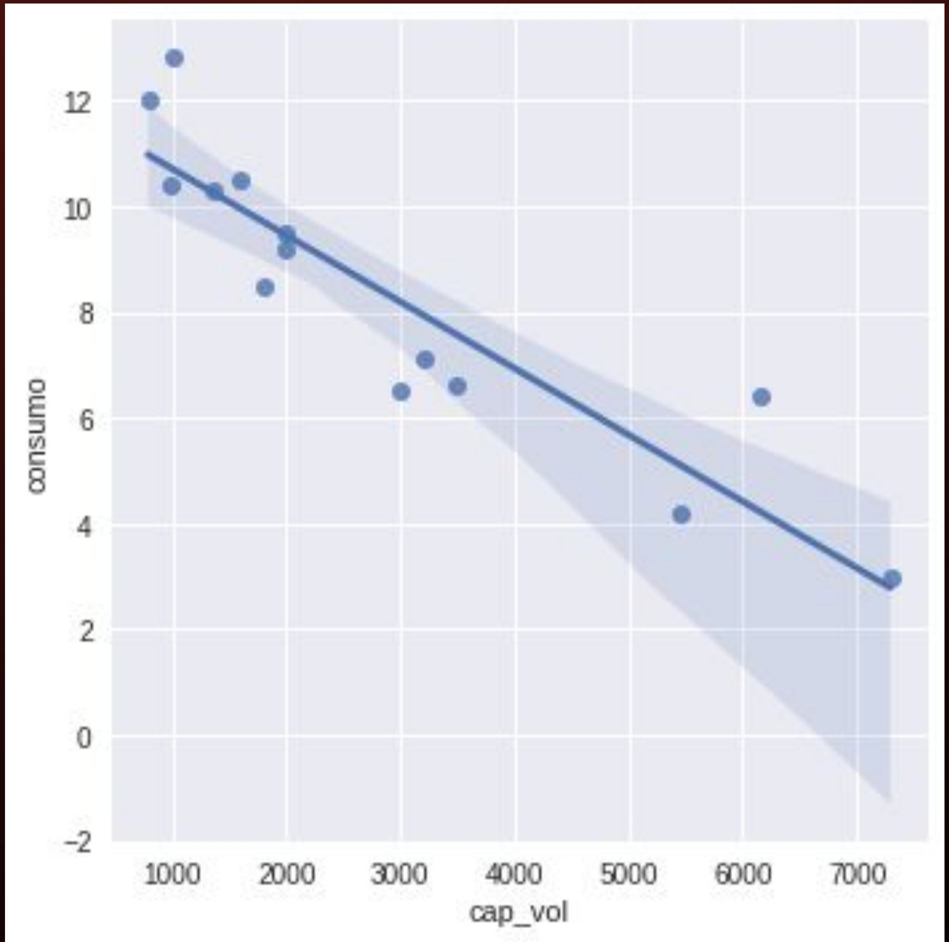
x: Capacidade volumétrica.
y: Consumo.
$\hat{y}$: Valores previsto de y.
e: Resíduos.
e*: Resíduos padronizados. e/e*: proporção

ADEQUAÇÃO DO MODELO: EXEMPLO 2

Gráficos de diagnóstico

1. Reta de regressão e valores observados

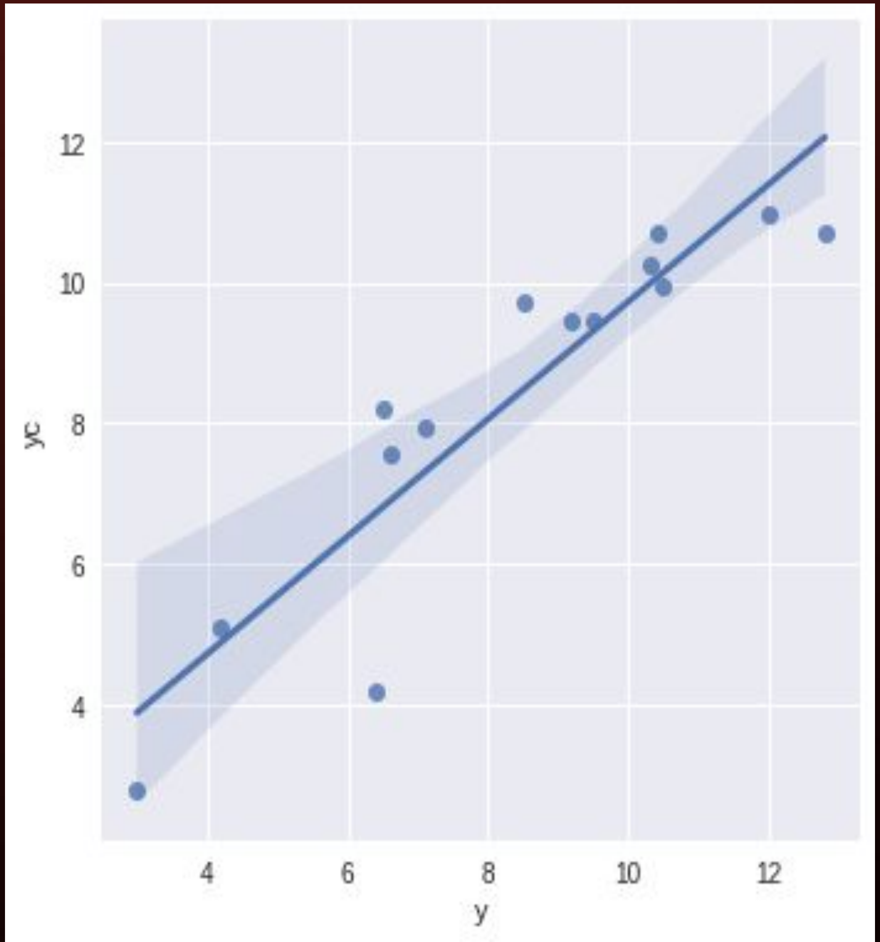
```
#Gráficos de diagnóstico 1 (y vs. x)
sns.lmplot(x='cap_vol', y='consumo', data=df1);plt.grid(True)
```



ADEQUAÇÃO DO MODELO: EXEMPLO 2

Gráficos de diagnóstico

2. Valores previstos vs. valores observados

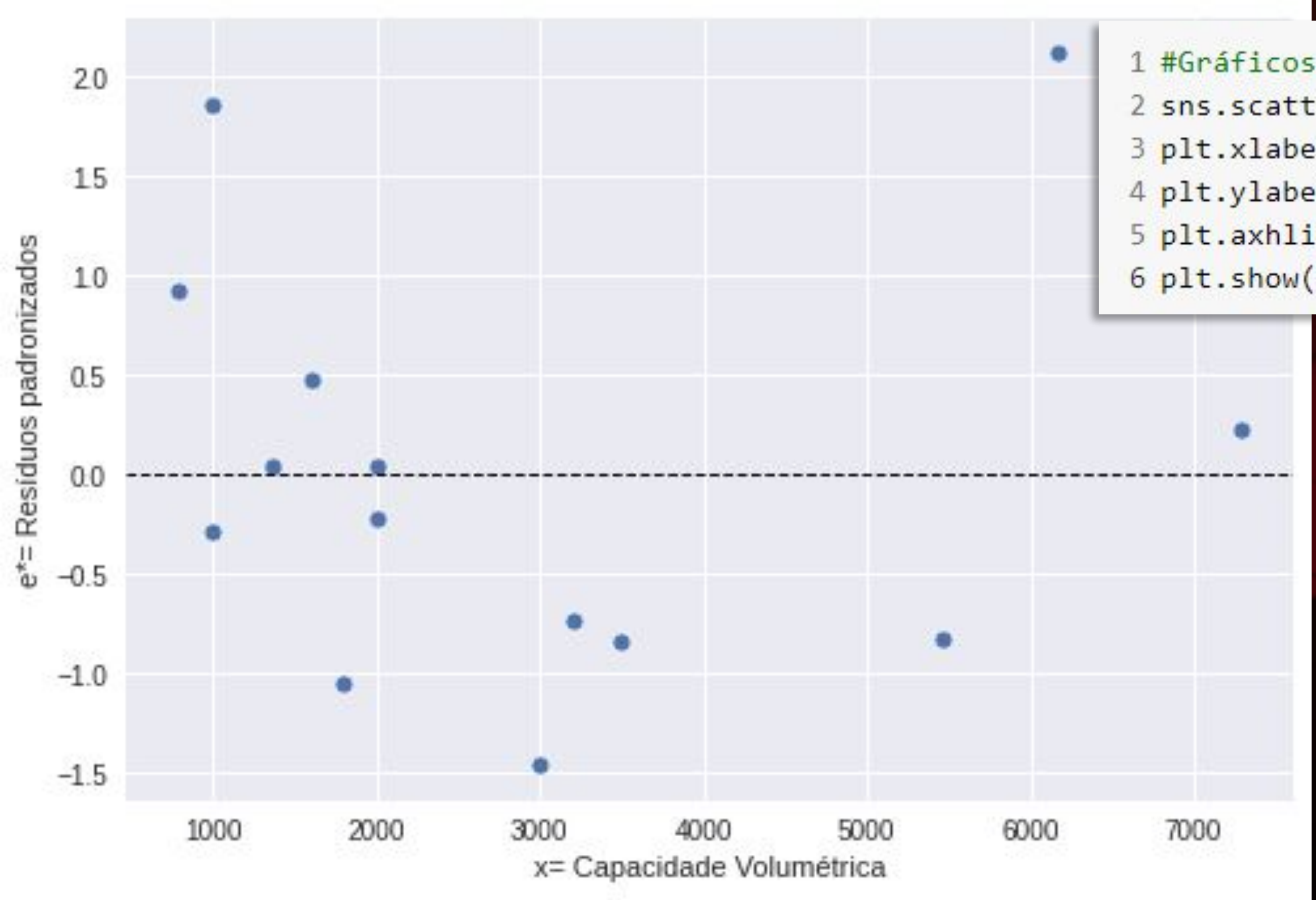



```
#Gráficos de diagnóstico 2 (y vs. yc)
sns.lmplot(x='y', y='yc', data=dft);plt.grid(True)
```

ADEQUAÇÃO DO MODELO: EXEMPLO 2

Gráficos de diagnóstico

3. Resíduos padronizados vs. x

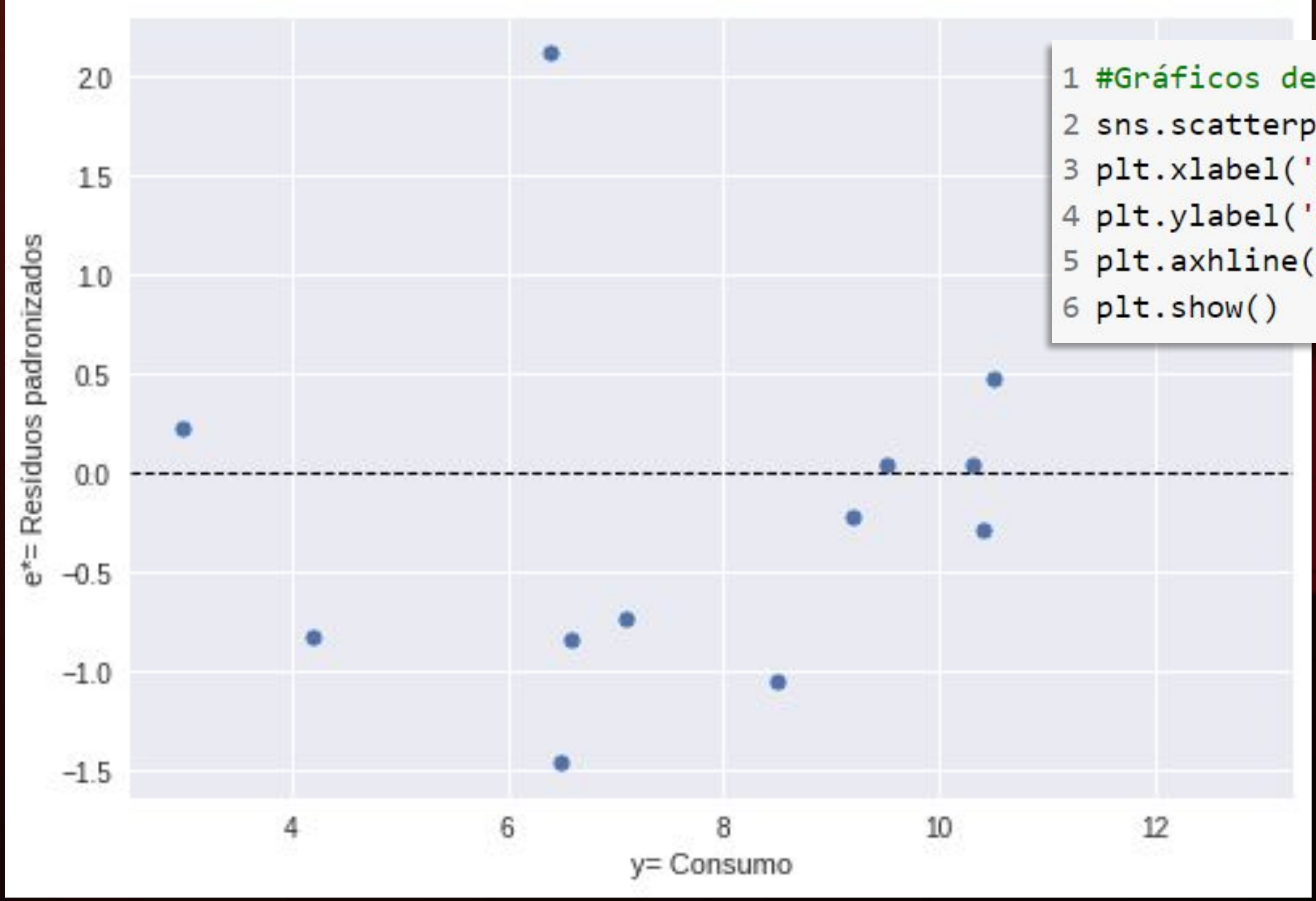



```python
#Gráficos de diagnóstico 3 (resíduos padronizados vs. x)
sns.scatterplot(x='x', y='e*', data=dft);plt.grid(True)
plt.xlabel('x= Capacidade Volumétrica')
plt.ylabel('e*= Resíduos padronizados')
plt.axhline(y=0, color='black', linestyle='--', linewidth=1)
plt.show()
```

ADEQUAÇÃO DO MODELO: EXEMPLO 2

Gráficos de diagnóstico

4. Resíduos padronizados vs. y

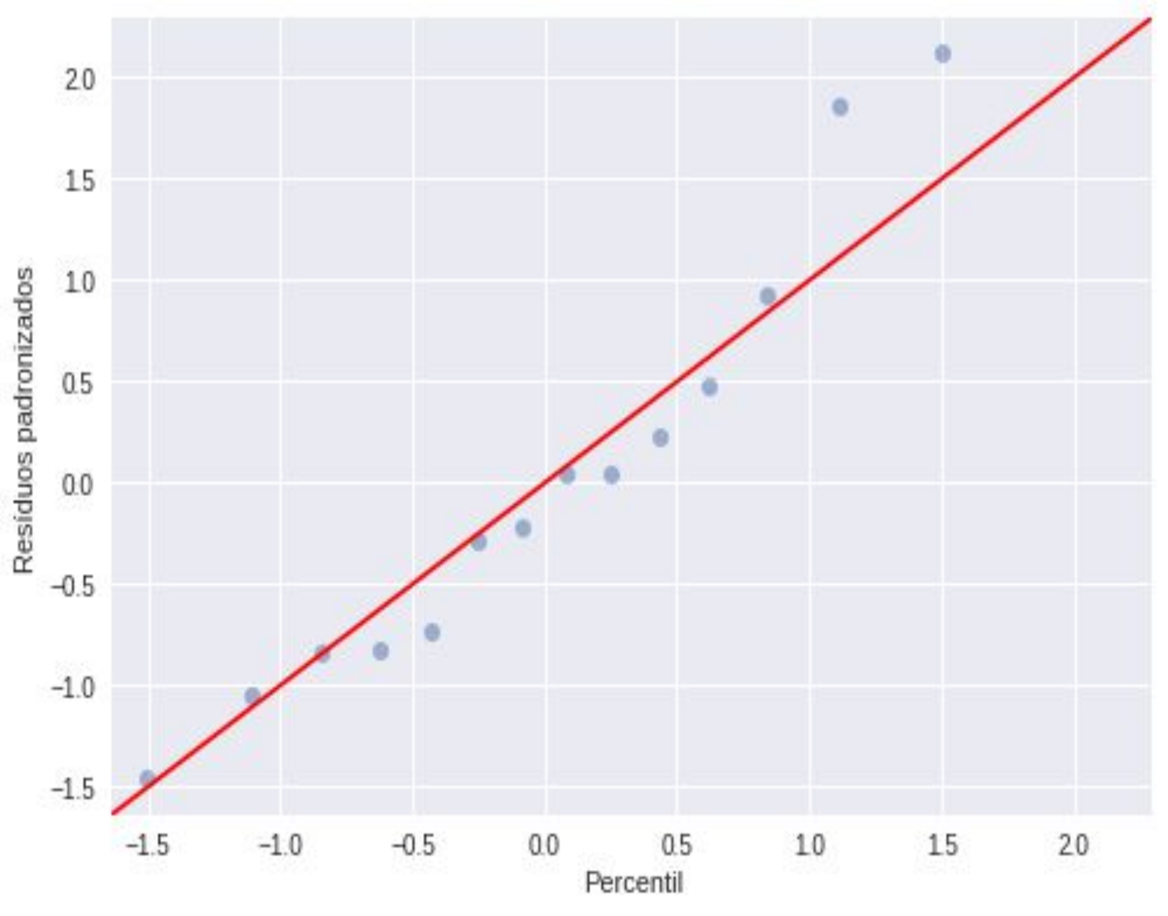



```python
#Gráficos de diagnóstico 4 (resíduos padronizados vs. y)
sns.scatterplot(x='y', y='e*', data=dft);plt.grid(True)
plt.xlabel('y= Consumo')
plt.ylabel('e*= Resíduos padronizados')
plt.axhline(y=0, color='black', linestyle='--', linewidth=1)
plt.show()
```

ADEQUAÇÃO DO MODELO: EXEMPLO 2

Gráficos de diagnóstico

5. Gráfico de probabilidade normal




```
QQ = ProbPlot(influence.resid_studentized_internal)
plot_lm_2 = QQ.qqplot(line='45', alpha=0.5, color='#4C72B0', lw=1)
plot_lm_2.axes[0].set_xlabel('Percentil')
plot_lm_2.axes[0].set_ylabel('Resíduos padronizados')
```

MODELO INTRINSECAMENTE LINEAR: EXEMPLO 3

Enunciado

Nenhum aficionado por batatas fritas gosta que elas fiquem encharcadas, por isso é importante encontrar características do processo de produção que as permitam ter uma textura atraente.
Os seguintes dados em que x = tempo de fritura (s) e y = teor de umidade (%) apareceram no artigo "Thermal and Physical Properties of Tortilla Chips as a Function of Frying Time" (J. of Food Processing and Preservation, 1995: 175–189). A partir destes dados:

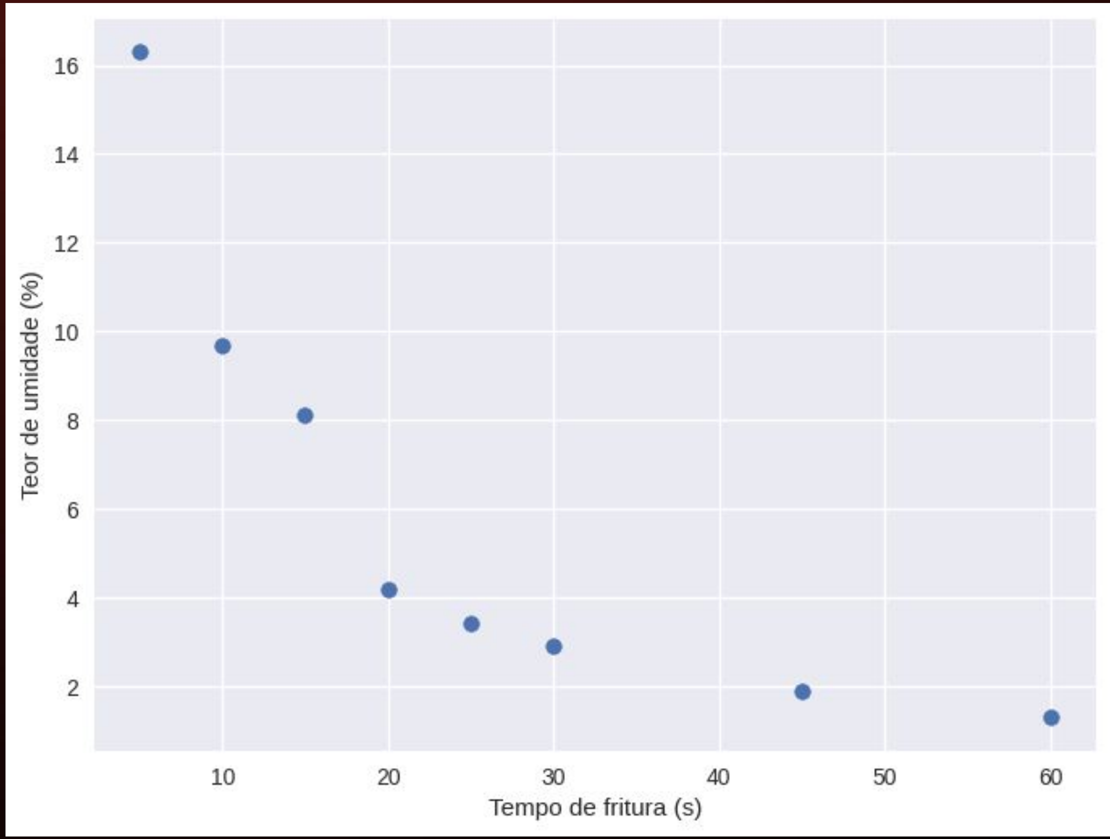
x	5	10	15	20	25	30	45	60
y	16.3	9.7	8.1	4.2	3.4	2.9	1.9	1.3

a) Construa um gráfico de dispersão de y versus x e comente.
b) Construa um gráfico de dispersão dos pares (ln(x), ln(y)) e comente.
c) Qual é a relação probabilística entre x e y sugerido pelo padrão linear no gráfico da parte (b)?
d) Preveja o valor do teor de umidade ao fritar as batatas por 20s, de uma forma que transmita informações sobre confiabilidade e precisão.
e) Analise os resíduos do ajuste do modelo de regressão linear para os dados transformados e comente.

MODELO INTRINSECAMENTE LINEAR: EXEMPLO 3

a) Construa um gráfico de dispersão de y versus x e comente.

- Os dados podem ser descritos mediante uma reta?



MODELO INTRINSECAMENTE LINEAR: EXEMPLO 3

a) Construa um gráfico de dispersão de y versus x e comente.

Confira a observação com o procedimento de regressão linear.

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                      y   R-squared:                       0.657
Model:                            OLS   Adj. R-squared:                  0.600
Method:                 Least Squares   F-statistic:                     11.48
Date:                Wed, 16 Feb 2022   Prob (F-statistic):             0.0147
Time:                        03:03:48   Log-Likelihood:                -19.587
No. Observations:                   8   AIC:                             43.17
Df Residuals:                       6   BIC:                             43.33
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const         11.8599      2.079      5.705      0.001       6.773      16.947
x             -0.2242      0.066     -3.389      0.015      -0.386      -0.062
==============================================================================
Omnibus:                        1.672   Durbin-Watson:                   0.824
Prob(Omnibus):                  0.433   Jarque-Bera (JB):                0.817
Skew:                           0.738   Prob(JB):                        0.665
Kurtosis:                       2.482   Cond. No.                         57.2
==============================================================================
```

MODELO INTRINSECAMENTE LINEAR: EXEMPLO 3

b) Construa um gráfico de dispersão dos pares (ln(x), ln(y)) e comente.

Observe os valores das variáveis transformadas.

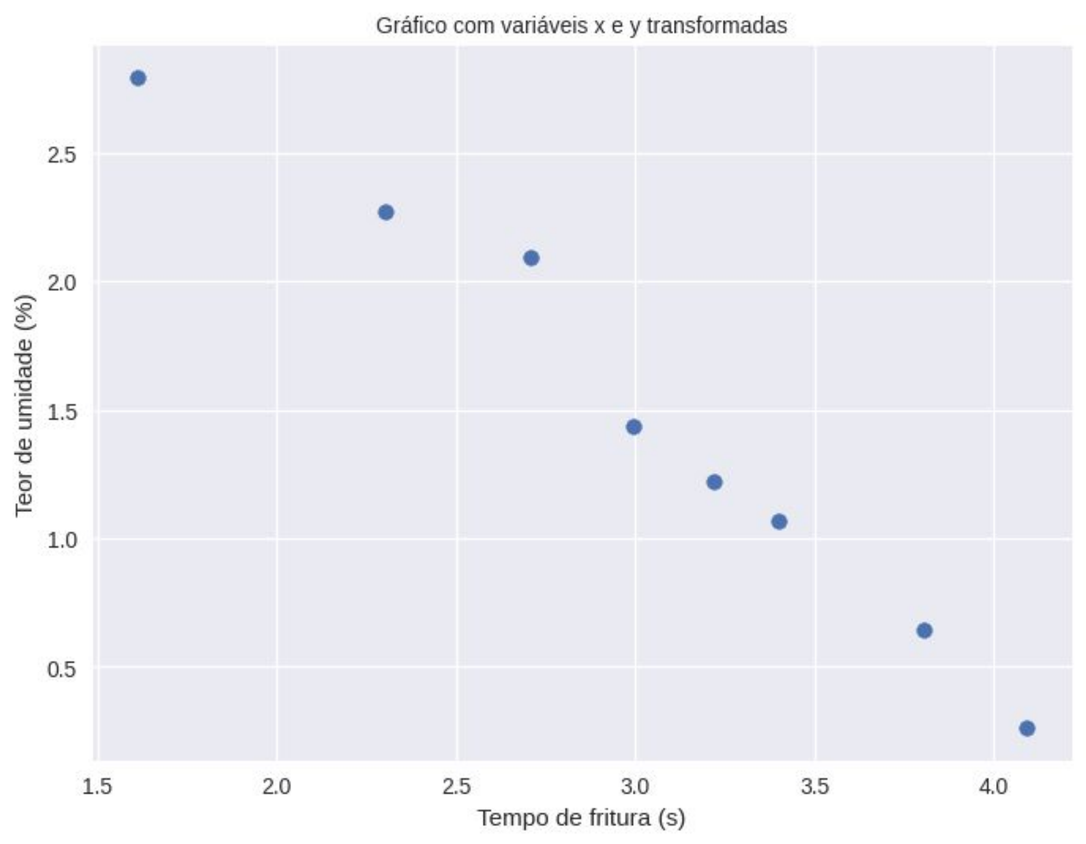
```
dfries2 = pd.DataFrame(list(zip(x, y, vtx, vty)),
                  columns =["x","y","ln_x","ln_y"])
dfries2.head()
```

	x	y	ln_x	ln_y
0	5	16.3	1.609438	2.791165
1	10	9.7	2.302585	2.272126
2	15	8.1	2.708050	2.091864
3	20	4.2	2.995732	1.435085
4	25	3.4	3.218876	1.223775

MODELO INTRINSECAMENTE LINEAR: EXEMPLO 3

b) Construa um gráfico de dispersão dos pares (ln(x), ln(y)) e comente.

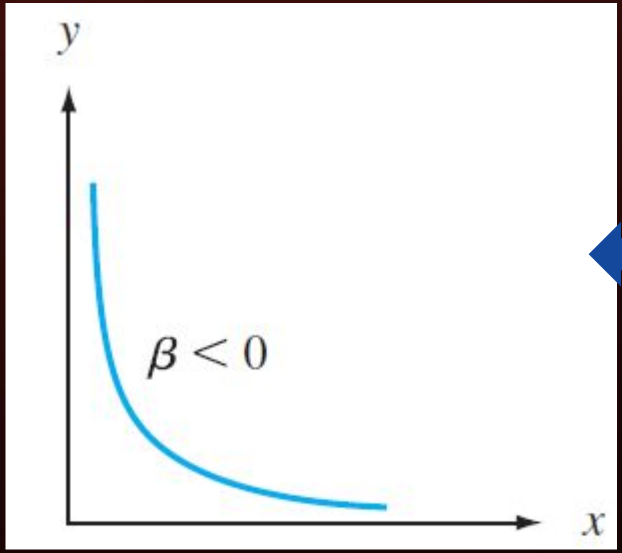
- Observe os valores das variáveis transformadas.

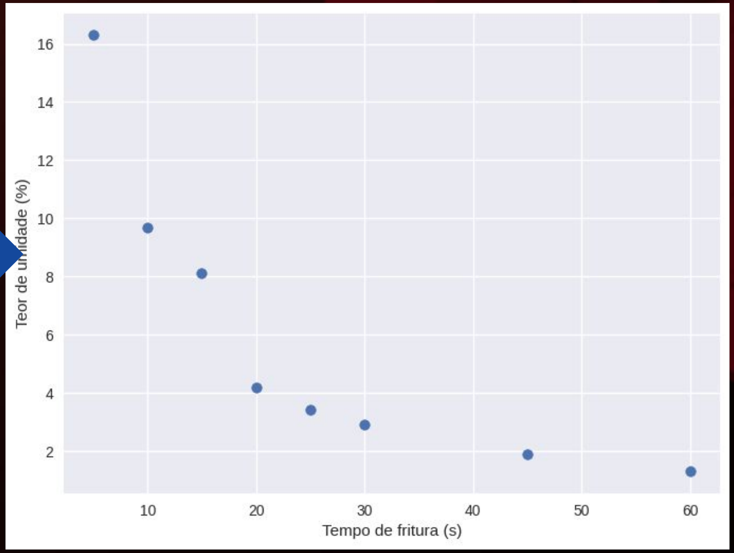


MODELO INTRINSECAMENTE LINEAR: EXEMPLO 3

c) Qual é a relação probabilística entre x e y sugerido pelo padrão linear no gráfico da parte (b)?

Função	Transformações para linearizar	Forma linear





Fonte: (DEVORE, 2018, p. 515)

MODELO INTRINSECAMENTE LINEAR: EXEMPLO 3

d) Preveja o valor do teor de umidade ao fritar as batatas por 20s, de forma que transmita informações sobre confiabilidade e precisão.

- Considerar o procedimento dos mínimos quadrados, mas aplicar às variáveis transformadas para obter a reta:

Forma linear

Depois, aplicar as transformações necessárias para obter a função potência.

Função

MODELO INTRINSECAMENTE LINEAR: EXEMPLO 3

d) Preveja o valor do teor de umidade ao fritar as batatas por 20s, de forma que transmita informações sobre confiabilidade e precisão.

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                      y   R-squared:                       0.976
Model:                            OLS   Adj. R-squared:                  0.971
Method:                 Least Squares   F-statistic:                     239.1
Date:                Wed, 16 Feb 2022   Prob (F-statistic):           4.63e-06
Time:                        03:42:03   Log-Likelihood:                 5.2525
No. Observations:                   8   AIC:                            -6.505
Df Residuals:                       6   BIC:                            -6.346
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          4.6384      0.211     21.978      0.000       4.122       5.155
x             -1.0492      0.068    -15.462      0.000      -1.215      -0.883
==============================================================================
Omnibus:                        8.964   Durbin-Watson:                   1.881
Prob(Omnibus):                  0.011   Jarque-Bera (JB):                2.783
Skew:                           1.325   Prob(JB):                        0.249
Kurtosis:                       4.154   Cond. No.                         14.1
==============================================================================
```

MODELO INTRINSECAMENTE LINEAR: EXEMPLO 3

d) Preveja o valor do teor de umidade ao fritar as batatas por 20s, de forma que transmita informações sobre confiabilidade e precisão.

Resultado inicial:

Forma linear

Voltando para as variáveis originais:

Função	

MODELO INTRINSECAMENTE LINEAR: EXEMPLO 3

d) Preveja o valor do teor de umidade ao fritar as batatas por 20s, de forma que transmita informações sobre confiabilidade e precisão.

- **Para x=20s quanto vale y?**

Forma linear

Função potência

Esta informação é suficiente para ter noção de confiabilidade e precisão?

MODELO INTRINSECAMENTE LINEAR: EXEMPLO 3

d) Preveja o valor do teor de umidade ao fritar as batatas por 20s, de forma que transmita informações sobre confiabilidade e precisão.

Para x=20s, quanto vale y?

A equação do IP:

$$\hat{y}' \pm t_{\frac{\alpha}{2},n-2}\sqrt{s^2 - s_{\hat{y}}^2}$$

Necessário calcular: $t_{\frac{\alpha}{2},n-2}$, s^2 e $s_{\hat{y}}^2$

$$\hat{\sigma}^2 = s^2 = \frac{SQE}{n-2} = \frac{\sum (y_i - \hat{y}_i)^2}{n-2}$$

$$S_Y = s\sqrt{\frac{1}{n} + \frac{(x^* - \bar{x})^2}{S_{xx}}}$$

Forma linear	Função potência

MODELO INTRINSECAMENTE LINEAR: EXEMPLO 3

d) Preveja o valor do teor de umidade ao fritar as batatas por 20s, de forma que transmita informações sobre confiabilidade e precisão.

- **Para x=20s, quanto vale y?**

A equação do IP:

$$\hat{y}' \pm t_{\frac{\alpha}{2},n-2}\sqrt{s^2 - s_{\hat{y}}^2}$$

Forma linear	Função potência

Necessário calcular:

$$t_{\frac{\alpha}{2},n-2} = 2{,}4469$$

```
#@title Resposta d)
#usar a tabela tstudent pata t
from scipy.stats import t
alpha = 0.05  # nível de significia
df = len(x) - 2 # gl (n - 2)
v = t.ppf(1 - alpha/2, df)
print(f't=: {v}')
```

```
t=: 2.4469118487916806
```

MODELO INTRINSECAMENTE LINEAR: EXEMPLO 3

d) Preveja o valor do teor de umidade ao fritar as batatas por 20s, de forma que transmita informações sobre confiabilidade e precisão.

Para x=20s, quanto vale y?

Forma linear	Função potência

```python
x=dfries['x']
y=dfries['y']
vtx=np.log(x)
vty=np.log(y)
y_pred=list(modelo.predict())
#sns.resid(modelo)
resi=(modelo.resid)
#SQE = residuos ao quadrado
sqe=np.sum(resi*resi)
gl=modelo.df_resid  #gl=n-2
s2=sqe/gl
s=pow(s2,1/2)
sqe,s2,s
```

```
(0.1259894484543931, 0.020998241409065552, 0.14490769961967348)
```

A equação do IP

$$\hat{y}' \pm t_{\frac{\alpha}{2},n-2}\sqrt{s^2 - s_{\hat{y}}^2}$$

Necessário calcular: s

$$\mathrm{SQE} = \sum (y_i - \hat{y}_i)^2$$

$$s^2 = \frac{SQE}{n-2} = \frac{0{,}12599}{8-2} = 0{,}020998$$

$$s = \sqrt{s^2} = 0{,}125989$$

MODELO INTRINSECAMENTE LINEAR: EXEMPLO 3

d) Preveja o valor do teor de umidade ao fritar as batatas por 20s, de forma que transmita informações sobre confiabilidade e precisão.

- Para x=20s, quanto vale y?
  A equação do IP:

$$\hat{y}' \pm t_{\frac{\alpha}{2},n-2}\sqrt{s^2 - s_{\hat{y}}^2}$$

Forma linear	Função potência

Necessário calcular: $s_{\hat{y}}^2$

```
vtxs=np.log(20) #x*=ln(20)
vtxm=vtx.mean()
n=modelo.df_resid + 2 #gl+2=n
xx=np.sum(vtx*vtx)
xau=np.sum(vtx)
sxx=xx-xau*xau/n
sy=s*pow(((1/n)+(vtxs-vtxm)*(vtxs-vtxm)/sxx),1/2)
sy
```

```
0.05125314354652712
```

$s_{\hat{y}} = s\sqrt{\frac{1}{n} + \frac{(x^* - \bar{x})^2}{s_{xx}}}$=0,51253

MODELO INTRINSECAMENTE LINEAR: EXEMPLO 3

d) Preveja o valor do teor de umidade ao fritar as batatas por 20s, de forma que transmita informações sobre confiabilidade e precisão.

- **Para x=20s, quanto vale y?**
  A equação do IP:

Forma linear	Função potência

$$\hat{y}' \pm t_{\frac{\alpha}{2},n-2}\sqrt{s^2 - s_{\hat{y}}^2}$$

$$\hat{y}' \pm t_{\frac{\alpha}{2},n-2}\sqrt{s^2 - s_{\hat{y}}^2} = 2{,}9957 \pm (2{,}4469)$$

$$\sqrt{0{,}020998 - 0{,}51253^2}$$

IP: $\widehat{Y}' = (1{,}119196\ ,\ 1{,}871399)$

$$1{,}119196 < \widehat{Y}' < 1{,}871399$$

Voltando para as variáveis originais:

IP: $\widehat{Y}' = (3{,}062392\ ,\ 6{,}497386)$

$$3{,}062392 < \hat{Y} < 6{,}497386$$

MODELO INTRINSECAMENTE LINEAR: EXEMPLO 3

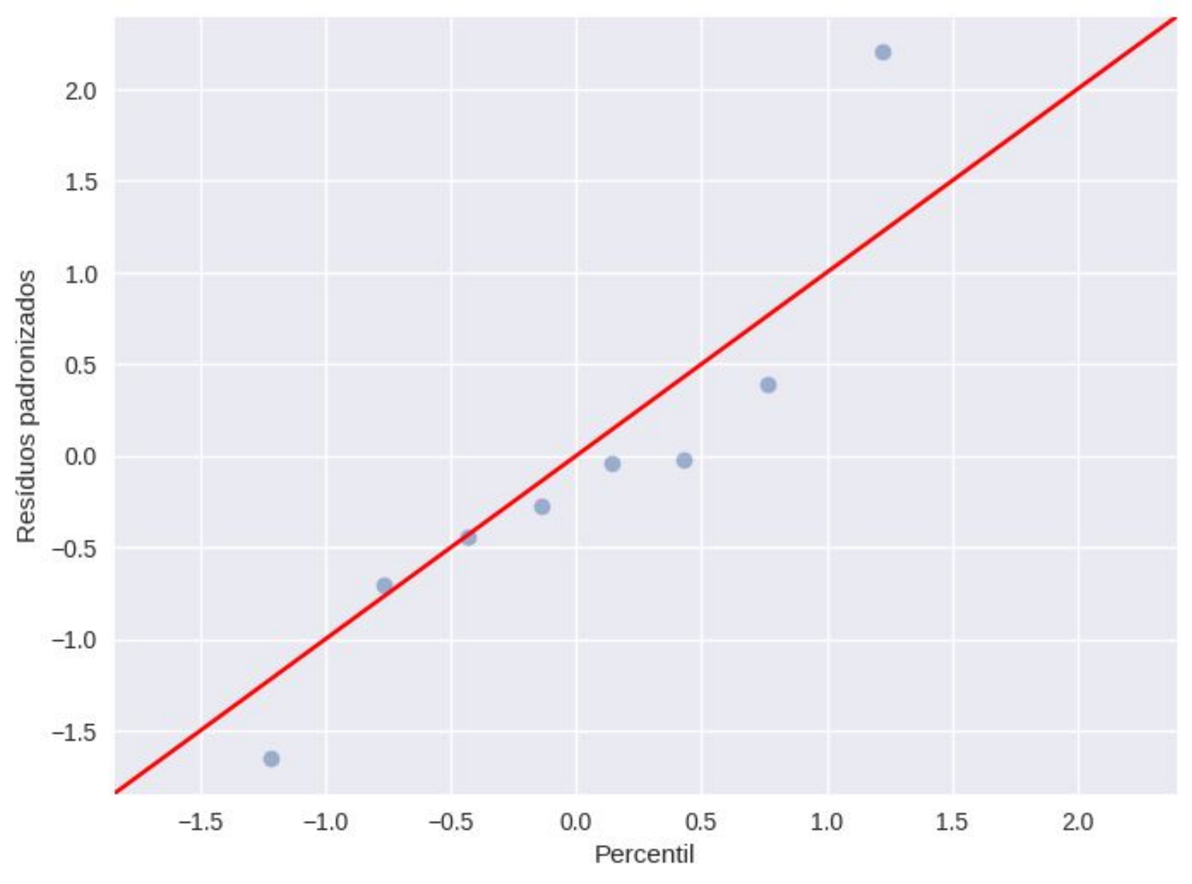
e) Analise os resíduos do ajuste do modelo de regressão linear para os dados transformados e comente.

- Para as variáveis transformadas ln(x) e ln(y)

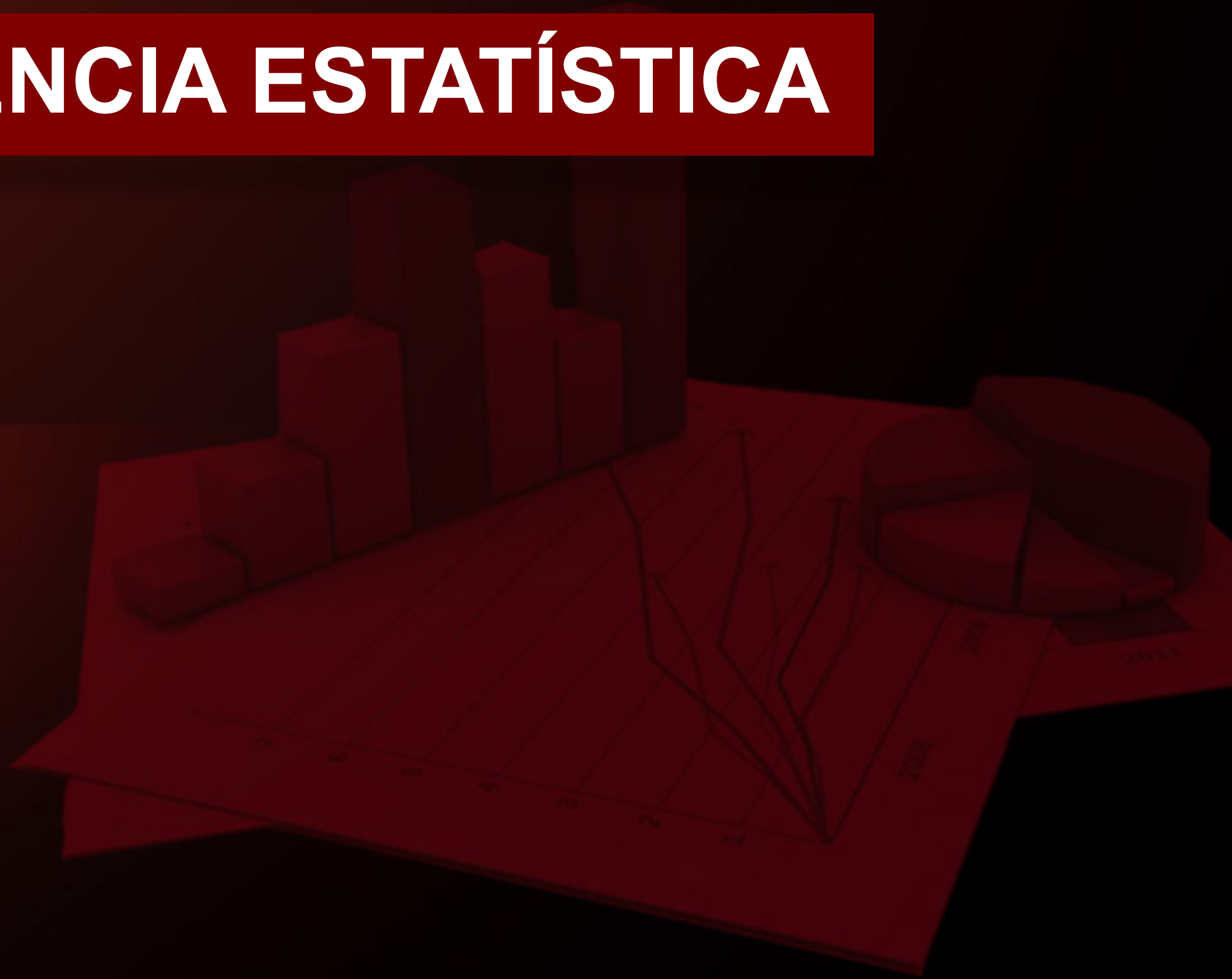
	x	y	yc	e	e*	e/e*
0	1.609438	2.791165	2.949796	-0.158631	-1.649456	0.096171
1	2.302585	2.272126	2.222547	0.049579	0.391681	0.126580
2	2.708050	2.091864	1.797134	0.294730	2.200852	0.133916
3	2.995732	1.435085	1.495298	-0.060214	-0.444247	0.135541
4	3.218876	1.223775	1.261176	-0.037401	-0.277341	0.134855
5	3.401197	1.064711	1.069885	-0.005174	-0.038898	0.133019
6	3.806662	0.641854	0.644472	-0.002618	-0.021025	0.124511
7	4.094345	0.262364	0.342636	-0.080272	-0.703222	0.114149

MODELO INTRINSECAMENTE LINEAR: EXEMPLO 3

e) Analise os resíduos do ajuste do modelo de regressão linear para os dados transformados e comente.



MODELAGEM E INFERÊNCIA ESTATÍSTICA



Adequações do modelo e
modelos não lineares
Exemplos