lista1_TS_fernando_papi

July 1, 2023

[172]:
```
from matplotlib import pyplot as plt
import numpy as np
import scipy
import pandas as pd
import statsmodels
import statsmodels.api as sm

# importing the style package
from matplotlib import style
# using the style for the plot
plt.style.use('seaborn-v0_8-darkgrid')
floatfmt = '.2f'
```

1) Deseja desenvolver um modelo com a finalidade de estimar as vendas médias semanais (Y - em milhares de dólares) em uma rede de lojas de acordo com o número de clientes (X). Obtenha uma regressão linear e analise a qualidade do modelo obtido. Faça análise dos resíduos. Estime a venda média semanal considerando X = 600 clientes.

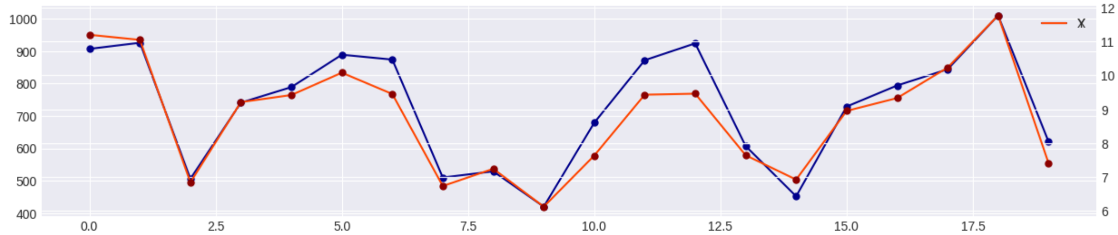
X	907	926	506	741	789	889	874	510	529	420
Y	11,20	11,05	6,84	9,21	9,42	10,08	9,45	6,73	7,24	6,12
X	679	872	924	607	452	729	794	844	1010	621
Y	7,63	9,43	9,46	7,64	6,92	8,95	9,33	10,23	11,77	7,41

[37]:
```
X = np.array([907, 926, 506, 741, 789, 889, 874, 510, 529, 420, 679, 872, 924,␣
 ↪607, 452, 729, 794, 844, 1010, 621])
Y = np.array([11.20, 11.05, 6.84, 9.21, 9.42, 10.08, 9.45, 6.73, 7.24, 6.12, 7.
 ↪63, 9.43, 9.46, 7.64, 6.92, 8.95, 9.33, 10.23, 11.77, 7.41])

X = pd.DataFrame(X, columns=['X'])
X = sm.add_constant(X)
```

[149]:
```
fig, ax = plt.subplots(figsize=(15,3))
ax2 = ax.twinx()
ax.plot(X.X, label='X', c='darkblue')
ax2.plot(Y, label='Y', c='orangered')
ax.scatter(range(len(X)), X.X, s=25, c='darkblue', zorder=3)
ax2.scatter(range(len(Y)), Y, s=25, c='darkred', zorder=3)
```

```
ax.legend()
ax2.legend()
plt.show()
```



[40]:
```
model = sm.OLS(Y, X).fit()
model.summary()
```

[40]:

Dep. Variable:	y	**R-squared:**	0.912
Model:	OLS	**Adj. R-squared:**	0.907
Method:	Least Squares	**F-statistic:**	186.2
Date:	Sat, 01 Jul 2023	**Prob (F-statistic):**	6.21e-11
Time:	13:27:21	**Log-Likelihood:**	-13.522
No. Observations:	20	**AIC:**	31.04
Df Residuals:	18	**BIC:**	33.04
Df Model:	1		
Covariance Type:	nonrobust		

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
const	2.4230	0.481	5.038	0.000	1.413	3.434
X	0.0087	0.001	13.646	0.000	0.007	0.010

Omnibus:	0.608	**Durbin-Watson:**	1.013
Prob(Omnibus):	0.738	**Jarque-Bera (JB):**	0.622
Skew:	-0.347	**Prob(JB):**	0.733
Kurtosis:	2.485	**Cond. No.**	3.22e+03

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 3.22e+03. This might indicate that there are
strong multicollinearity or other numerical problems.

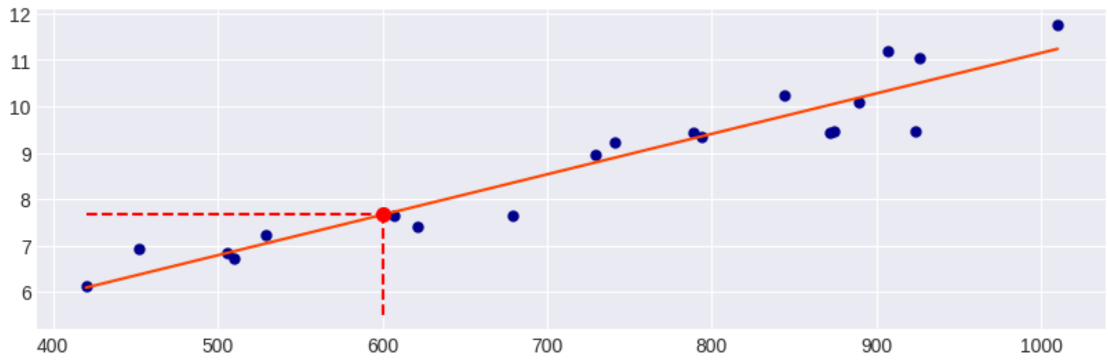
[99]:
```
fig, ax = plt.subplots(figsize=(10,3))
ax.scatter(X.X, Y, s=25, c='darkblue', zorder=3)

y_p = 600
y_h = model.predict((1,y_p))

min_x = X[X.X==min(X.X)]
max_x = X[X.X==max(X.X)]
```

```
min_y_hat = model.predict(min_x).values
max_y_hat = model.predict(max_x).values

ax.plot([min_x.X, max_x.X], [min_y_hat, max_y_hat], c='orangered', zorder=3)
ax.scatter(y_p, y_h, s=50, c='red', zorder=3)
ax.plot([y_p, y_p], [0.9*min(Y), y_h], 'r--', zorder=3)
ax.plot([min_x.X.values, y_p], [y_h, y_h], 'r--', zorder=3)
plt.show()
print('\nX=600, Y_hat =', y_h)
```

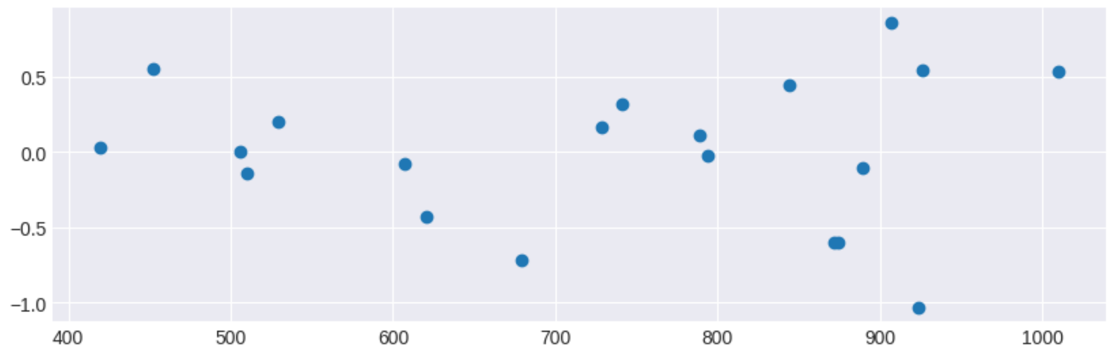


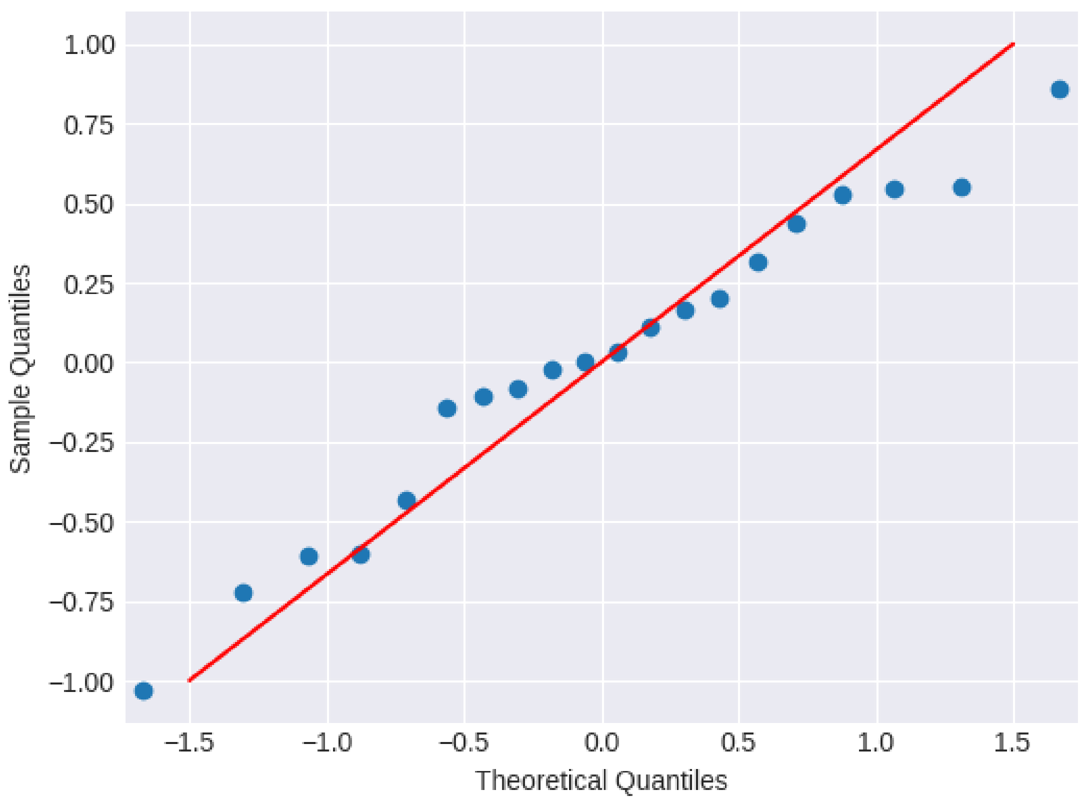

```
X=600, Y_hat = [7.6606473]
```

[120]:
```
print('mean', np.mean(model.wresid))
print('std', np.std(model.wresid))
fig, axs = plt.subplots(1,1, figsize=(10,3))
axs.scatter(X.X, model.wresid)
sm.qqplot(model.wresid)
plt.plot([-1.5, 1.5], [-1,1],'r')
pvalue = scipy.stats.shapiro(model.wresid).pvalue
print()
print(f'Shapiro-Wilk test dos residuais -> pvalue: {pvalue}.\nFalha-se em␣
↪rejeitar a hipótese nula, ou seja, os residuais são normalmente␣
↪distribuídos')
print()
```

```
mean -9.325873406851315e-16
std 0.4757601340848472

Shapiro-Wilk test dos residuais -> pvalue: 0.7665435671806335.
Falha-se em rejeitar a hipótese nula, ou seja, os residuais são normalmente
distribuídos
```


0.5
0.0
−0.5
−1.0
400
500
600
700
800
900
1000


1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
Sample Quantiles
−1.5
−1.0
−0.5
0.0
0.5
1.0
1.5
Theoretical Quantiles

[ ]:

[ ]:

2) Considere o seguinte modelo MA(3):

$$r_t = 0{,}01 + \varepsilon_t + 0{,}7\varepsilon_{t-1} + 0{,}5\varepsilon_{t-2} + 0{,}2\varepsilon_{t-3}$$

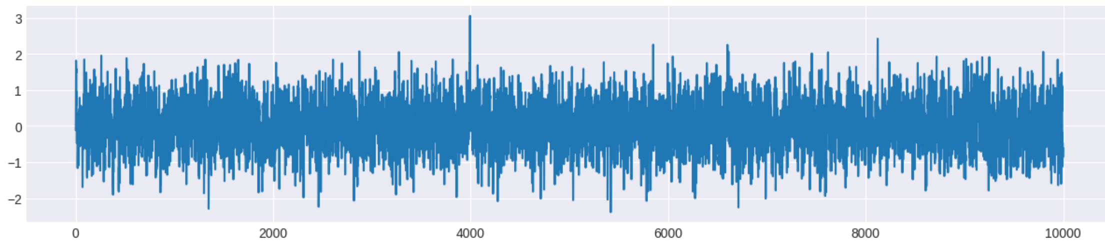
onde ε_t é ruído branco com desvio padrão $\sigma = 0{,}5$.

a. Apresente uma simulação do processo considerando 10.000 amostras.

b. Apresente o gráfico da função de autocorrelação estimada a partir dos dados da simulação do item a. Compare com os valores teóricos.

d. Apresente o gráfico da função de autocorrelação parcial estimada a partir dos dados da simulação do item a.

[176]:
```
gen = scipy.stats.norm(0, 0.5)
y = gen.rvs(10000)
y = [0.01 + y_t + 0.7*y[t-1] + 0.5*y[t-2] + 0.2*y[t-3] if t>=3 else y_t for t,␣
 ↪y_t in enumerate(y)]
fig, ax = plt.subplots(figsize=(15,3))
p = ax.plot(y)
```



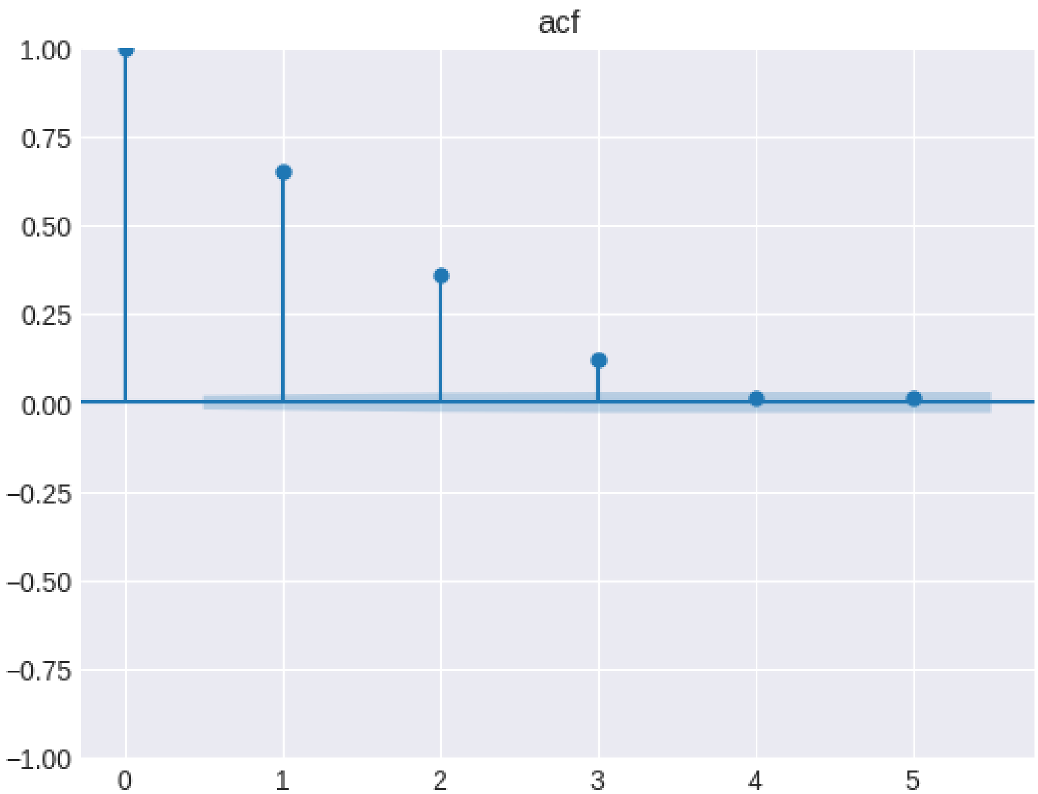
[177]:
```
from statsmodels.tsa.stattools import acf
from statsmodels.graphics.tsaplots import plot_acf
from statsmodels.graphics.tsaplots import plot_pacf
```

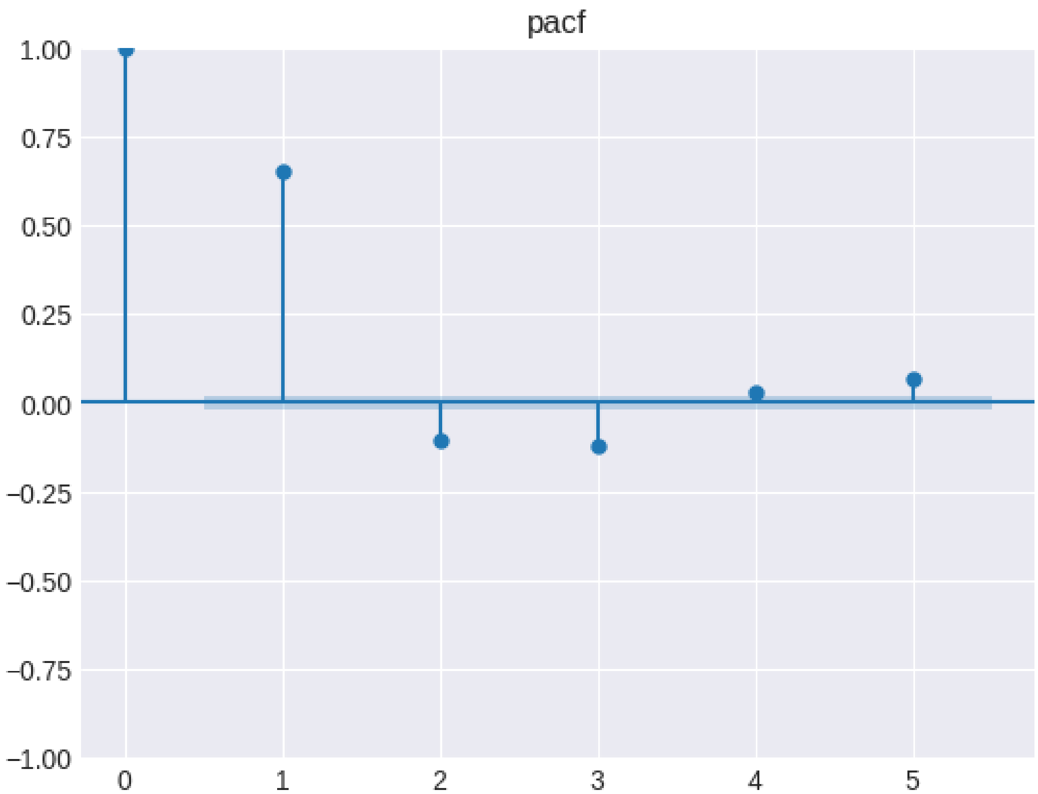
[178]:
```
lags = 5

autocorrelations = acf(y, nlags=5)
print(autocorrelations)

plot_acf(y, alpha=0.05, lags=lags, title=f'acf')
plot_pacf(y, alpha=0.05, lags=lags, title=f'pacf')
plt.show()
```

```
[1.         0.6494878  0.36045371 0.11928814 0.01272379 0.01037453]
```


acf
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
0
1
2
3
4
5



[ ]:

,

3) Seja a seguinte equação de diferenças, correspondente a um modelo AR(5):

$$r_t = 0{,}0090 + 0{,}1097 r_{t-1} - 0{,}004943 r_{t-2} - 0{,}1224 r_{t-3} + 0.02954 r_{t-4} + 0.06804 r_{t-5} + a_t$$

onde a_t represente o choque no instante t.

a. O modelo é estável? Justifique.

b. Apresente uma simulação do processo considerando como entrada a_t = [0 0 1 0 0 0 ... 0]. Gere 30 amostras da série.

c. Apresente uma simulação do processo considerando a_t como ruído branco Gaussiano com $\sigma_a = 0{,}007$. Gere 10.000 amostras da série.

d. Apresente o gráfico da função de autocorrelação a partir dos dados do item c.

e. Apresente o gráfico da função de autocorrelação parcial a partir dos dados do item c.

[ ]:
```
c = 0.0090
+ 0,1097r □1 □0,004943r □2 □0,1224r □3 + 0.02954r □4 + 0.06804r □5 +
```

ϕ

[ ]:

[ ]:

[ ]:

4) Suponha que o retorno simples mensal de um título siga o modelo MA(1) a seguir:

$$R_t = a_t + 0{,}2a_{t-1}, \qquad \sigma_a = 0{,}025$$

Assuma $a_{100} = 0{,}01$. Compute as previsões um e dois períodos à frente do retorno na origem da previsão $t = 100$. Quais são os desvios padrão dos erros de previsão associados? Calcule também as autocorrelações da série temporal de retorno com um e dois atrasos.

5) Suponha que o retorno logarítmico diário de uma aplicação siga o seguinte modelo:

$$r_t = 0{,}01 + 0{,}2r_{t-2} + a_t$$

onde $\{a_t\}$ é uma série de ruído branco gaussiano com média zero e variância 0,02. Qual é a média e a variância da série de retorno r_t? Compute as autocorrelações de r_t com um e dois atrasos. Assuma $r_{100} = -0{,}01$ e que $r_{99} = 0{,}02$. Calcule as previsões de um e dois períodos à frente da série de retorno na origem da previsão $t = 100$. Quais são os desvios padrão associados com os erros de previsão?

6) A partir de 110 observações da série temporal Y_t estimou-se o seguinte modelo:

$$Y_t = \varepsilon_t + 0{,}5\varepsilon_{t-1} + 0{,}4\varepsilon_{t-2} + 18, \quad \sigma_\varepsilon^2 = 4$$

Sabendo-se que $\varepsilon_{109} = 1{,}9$ e = 1,9 $\varepsilon_{110} = 1{,}7$, determine as previsões para os instantes 111, 112 e 113, feitas a partir da origem t =110. Obtenha também a variância do erro de previsão para os instantes 111, 112 e 113.

[ ]: