



Análise e Transformação de Dados

Ficha Prática nº 3 - Parte B de ATD 2022

Alberto Cardoso ©DEI2021/2022

edited by: Bruno Leitao

Objetivo: Pretende-se continuar a análise de séries temporais, efetuando a identificação e a estimação do modelo a considerar (AR, ARMA e ARIMA), bem como realizar o teste de diagnóstico do modelo e a sua utilização para previsão de valores futuros da série.

Linguagem de Programação:

- **MATLAB (Econometrics Toolbox)**
- Python (módulos: numpy, scipy, statsmodels.tsa, matplotlib.pyplot)

Exercício 3.

Tendo por base a decomposição da série temporal nas suas componentes: tendência, sazonal, cíclica e errática/irregular, as fases seguintes correspondem à determinação do modelo mais adequado para representar o comportamento da série e possibilitar a previsão de valores futuros. Neste trabalho considera-se que a série temporal pode ser descrita por um processo univariado Auto-Regressivo (AR), Auto-Regressivo de Médias Móveis (ARMA) ou Auto-Regressivo Integrado de Médias Móveis (ARIMA). A escolha do modelo poderá ser estruturada nas seguintes 5 fases:

- Fase 1. Verificação da estacionaridade da série: Verificar se a série é estacionária e, caso não o seja, proceder a sucessivas diferenciações até atingir a estacionaridade.
- Fase 2. Identificação do Modelo: Determinar os critérios de definição do comportamento da série. Procura-se averiguar se é possível identificar a ordem do modelo usando os métodos de Função de Autocorrelação (FAC) e da Função de Autocorrelação Parcial (FACP).
- Fase 3. Estimação do Modelo: Estimar os modelos candidatos a serem selecionados após a identificação, procedendo-se à análise dos modelos mais adequados com base em critérios de escolha. Um dos critérios poderá ser o critério da soma do Erro Quadrático.
- Fase 4. Teste de Diagnóstico: Consiste em verificar se o modelo descreve adequadamente a série de dados objeto da análise.
- Fase 5. Previsão: Consiste em fazer a previsão, isto é, prever os valores futuros da série.

As seguintes tarefas devem ser aplicadas à série temporal considerada na Parte A desta ficha.

Exercício 3.1 Verificar a estacionaridade da série regularizada e da série sem a tendência de ordem 1, usando a função **adftest** do Matlab (*Econometrics Toolbox*). Se o resultado desta função for o valor 1, a série deverá ser considerada estacionária..

```
In [ ]: %--- Usar série temporal regularizada (xlro) e a série sem a tend
        ência de grau 1 (xlro_t1)
        %--- e a tendência de grau 1 (tr1_1) que resultaram da parte A de
        sta ficha prática
```

```
In [ ]: %--- Ex. 3.1

        % ver slide 2-30

        % [hint:] help adftest

        % representa um teste estatístico!

        % Since the null hypothesis assumes the presence of unit root,  $y$ 
        % ( $t$ ) =  $c + d*t + \alpha*y(t-1)$ 
        % that is  $\alpha=1$ , the p-value obtained should be less than the signi
        % ficance
        % level (say 0.05) in order to reject the null hypothesis.
        % Thereby, inferring that the series is stationary.
```

Exercício 3.2 Particionar a série estacionária (sem tendência de grau 1) em duas subséries: a série de treino que deve considerar as temperaturas entre 1980 e 2000 e deve ser usada para determinar os parâmetros dos modelos **AR**, **ARMA** e **ARIMA**; a série de teste que deve ser usada para posteriormente testar o desempenho dos modelos identificados para prever valores futuros entre 2001 e 2018. (**Nota:** as funções Matlab **ar**, **armax** e **arima** (*System Identification Toolbox*) necessitam que se crie um objeto **iddata(y,[],Ts,'TimeUnit','months')** para determinar os parâmetros dos modelos.)

```
In [ ]: %--- Ex 3.2
        %tr_data=; %Temp entre Jan 1980 e Dez 2000
        %tst_data=; %Temp entre Jan 2001 e Dez 2018

        % trends dos datasets de treino e tst?
        %tr_trend=;
        %tst_trend=;

        % Criar objeto IDDATA com os dados de treino
        % [hint:] help iddata
        %id_treino = ;
```

Exercício 3.3 Identificar o modelo **AR** para a série sem a tendência de ordem 1, considerando o resultado da Função de Autocorrelação Parcial (FACP) para definir o valor de **na** (ordem do modelo que se quer estimar) e definindo uma abordagem adequada (por exemplo, o método dos mínimos quadrados). Para isso, usar a função **ar** para obter o modelo e aceder aos coeficientes **a_n**.

```

In [ ]: %--- Ex 3.3
%=====MODELO AR=====
% ver slide 2-32 / 2-33
% [hint:] help parcorr - indicador para ordem *na* do modelo AR
%to_do [Create and check parcorr figure and define na1_AR]

% na1_AR=;

% estimate model AR

% [hint:] help arOptions
%opt1_AR =;

% [hint:] help ar
%modell_AR = ; % modelo AR

% [hint:] help polydat
%pcoef1_AR =; % parâmetros do modelo AR

```

Exercício 3.4 Fazer a previsão de valores futuros da temperatura entre **2001 e 2018** usando os parâmetros (coeficientes) do modelo e usando também a função **forecast**. Validar o modelo comparando graficamente a previsão com a série temporal regularizada. Se necessário, repetir as tarefas 3.3 até obter o modelo adequado para a série. Considerar como métrica, por exemplo, a soma do quadrado do erro entre os valores medidos e estimados da série.

```

In [ ]: %--- Ex 3.4
%Simulação do Modelo AR
% ver slide 2-34, 2-35, 2-36
% [Careful:] look closely at the function that describes the mode
l - particularly the order of the coefficients.
% https://www.mathworks.com/help/econ/autoregressive-model.html;
%y1_AR=;

% Adicionar as tendências para comparar com a série original
%y1_AR=;

% Simulação do Modelo AR com forecast
% [hint:] help forecast
%y1_ARf=;

% Determinação dos vetores tempo para a série original e para a p
revisão
to=[1:numel(serie_teste)]; %Tempo para a série original (teste)

figure % Compara a série com a sua estimação
plot(to,tst_data+tst_trend,'-+',to,y1_AR,'-o',to,y1_ARf+tst_tren
d,'-*');
xlabel('n [meses]');
ylabel('T [°C]')
title('Série medida e regularizada (-+), Estimação com o modelo A
R (-o), Estimação com o modelo AR e usando a função forecast (-
*)')

% Métrica para análise (Soma dos quadrados do Erro)
%E1_AR=s;

```

Exercício 3.5 Identificar o modelo **ARMA** para a série sem a tendência de ordem 1, considerando valores adequados para **na** (ordem do modelo que se quer estimar) e para **nc** (histórico do ruído branco a considerar). Verificar a utilidade da Função de Autocorrelação (FAC). Deve ser utilizada a função **armax** que permite definir um método de procura adequado (por exemplo, a opção automática definida por omissão).

```

In [ ]: %--- Ex. 3.5
% =====Modelo ARMA=====
% ver modelo slide 2-28, 2-29, 2-30

% [hint:] help autocorr - indicador para ordem *nc* do modelo AR
%to_do
% nc1_ARMA=;

opt1_ARMAX = armaxOptions('SearchMethod', 'auto');
na1_ARMA=24;

% [hint:] help armax
%modell_ARMA=;

% Obter os coeficientes an e bn
%[pa1_ARMA,pb1_ARMA,pc1_ARMA]=;

```

Exercício 3.6 Fazer a previsão de valores futuros da temperatura entre **2001 e 2018** usando os parâmetros do modelo e usando também a função **forecast**. Validar o modelo comparando graficamente a previsão com a série temporal regularizada. Se necessário, repetir as tarefas 3.5 até obter o modelo adequado para a série. Considerar como métrica, por exemplo, a soma do quadrado do erro entre os valores medidos e estimados da série.

```
In [ ]: %--- Ex. 3.6

% histórico da série - termo AR

% inicializar termo ruído branco.
%e =;

% Simulação do Modelo ARMA
%y1_ARMA=;

% Adicionar as tendências para comparar com a série original
%y1_ARMA=;

% Simulação do Modelo ARMA com forecast
%y1_ARMAf=;

figure % Compara a série com a sua estimacão
plot(to,tst_data+tst_trend,'-+',to,y1_ARMA,'-o',...
     to,y1_ARMAf+tst_trend,'-*');
ylabel('T [°C]')
xlabel('n [meses]');
title('Série medida e regularizada (-+), Estimacão com o modelo A
RMA (-o), Estimacão com o modelo ARMA e usando a função forecast
(-*)')

% Métrica para análise
E1_ARMA=sum(((tst_data+tst_trend)-(y1_ARMAf+tst_trend)).^2)
```

Exercício 3.7 Em alternativa às abordagens anteriores, identificar o modelo **ARIMA** da série regularizada (eventualmente, não estacionária), considerando valores adequados para **p** (grau do histórico da variável a considerar), **D** (número de operações de diferenciação até obter a série estacionária) e para **q** (grau do histórico do ruído branco a considerar). Para isso, usar a função **arima** para criar a estrutura do modelo e a função **estimate** para estimar os parâmetros do modelo da série.

```

In [ ]: %-- Ex 3.7

% teste de estacionaridade da série regularizada com uma diferenciação
%[hint: ] help adftest, diff

if adftest(diff(x1ro))==0
    disp('Série diferenciada não estacionária')
else
    disp('Série diferenciada estacionária')
end

%===== Estimação de um Modelo ARIMA=====
% Parâmetros do modelo ARIMA
p1_ARIMA=24; % na
D1_ARIMA=1; % nº de vezes que a diferenciação é feita
q1_ARIMA=5; % nb

% Obtém a estrutura do modelo ARIMA
Md1=arima(p1_ARIMA,D1_ARIMA,q1_ARIMA);

```

Exercício 3.8 Fazer a previsão de valores futuros da temperatura entre **2001 e 2018** usando a função **simulate**. Validar o modelo identificado, comparando graficamente a previsão com a série original medida e regularizada. Se necessário, repetir as tarefas 3.7 até obter o modelo adequado para a série. Considerar como métrica, por exemplo, a soma do quadrado do erro entre os valores medidos e estimados da série.

```

In [ ]: %--- Ex. 3.8
% Estimação do modelo ARIMA

%tr_Data=;
%tst_data=;

% [hint: ] help estimate
%EstMd1=;

% Simulação do modelos
% [hint: ] help simulate
%y1_ARIMA =;

figure % Compara a série com a sua estimacão
plot(to,tst_data,'-+',to,y1_ARIMA,'-o');
ylabel('T [°C]');
xlabel('n [meses]');
title('Série (-+) e Estimacão com o modelo ARIMA (-o)')

% Métrica para análise
E1_ARIMA=sum(((tst_data)-(y1_ARIMA)).^2)

```

```

In [2]: ! jupyter nbconvert --to html ATD2022_Ficha3_ParteB_Matlab_todo_w_hints.ipynb

```

```

[NbConvertApp] Converting notebook ATD2022_Ficha3_ParteB_Matlab_todo_w_hints.ipynb to html
[NbConvertApp] Writing 549832 bytes to ATD2022_Ficha3_ParteB_Matlab_todo_w_hints.html

```

In []: