



Análise e Transformação de Dados

Ficha Prática nº 3 - Parte A de ATD 2022

Alberto Cardoso ©DEI2021/2022

Edited by: Bruno Leitão

Objetivo: Pretende-se iniciar a análise de séries temporais, efetuando o seu pré-processamento e a sua decomposição em componentes que traduzem os movimentos estruturais e erráticos.

Exercício 1.

Uma série temporal é uma sequência temporalmente ordenada de dados. O estudo estatístico de Séries Temporais envolve, em geral, dois aspetos: a) Análise e Modelação da Série Temporal – para descrever a série, verificar as suas características mais relevantes e investigar as possíveis relações com outras séries; b) Previsão da Série Temporal – determinar boas previsões de valores futuros da série, num dado horizonte de previsão, a partir de valores passados da série.

Antes de iniciar a análise de uma série temporal deve-se proceder à sua preparação através do pré-processamento dos dados que envolve, normalmente, as seguintes operações:

- Detecção e regularização do espaçamento dos dados, envolvendo a deteção de dados em falta (por exemplo, identificados pelo valor NaN) e a sua substituição por um valor estimado usando, por exemplo, um método de interpolação ou de extrapolação;
- Detecção e regularização de valores atípicos (*outliers*), envolvendo a sua deteção considerando, por exemplo, o critério $|x_i - \mu| > 3\sigma$, sendo x_i o valor da série no índice i , μ a média da série e σ o desvio padrão da série, e a sua substituição por um valor adequado. Dependendo do outlier ser aditivo ou subtrativo, o valor a usar poderá ser, por exemplo, $x_i = \mu + 2.5\sigma$ no caso aditivo e $x_i = \mu - 2.5\sigma$ no caso subtrativo.

De referir que o pré-processamento dos dados é muito importante porque a existência de dados em falta e/ou de *outliers* pode comprometer os procedimentos de análise e de modelação da série temporal, podendo, nomeadamente, induzir uma identificação incorreta do modelo e uma estimação enviesada dos seus parâmetros.

Neste trabalho, pretende-se tratar a série temporal que representa a evolução da temperatura média numa dada localização. No caso, considera-se o dataset com os dados de Lisboa de 1980 a 2018.

Cada amostra da série corresponde a 1 mês, sendo a primeira referente a janeiro de 1980.

Exercício 1.1 Ler e representar graficamente a série temporal existente no ficheiro de dados "lisbon_temp_fmt.mat" (temperaturas com espaçamento temporal em meses).

```
In [ ]: %--- Ler ficheiro com a série temporal
        %--- Contém: temp, months, years

        % [hint:] help load (1 linha)
        %

        % [hint:] ver variáveis importadas (alterar nome se necessário) e
        % criar vector n (escala temporal)
        %x1=;
        %n=;;

        %--- Representação gráfica

        figure(1)

        % hint: help plot (1 linha)
        plot(n,x1,'-+');

        legend('Série temporal','Location','northwest')
        xlabel('n [meses]');
        ylabel('T [°C]')
        title('Temperatura em Lisboa desde 1980');
```

Exercício 1.2 Verificar a existência de valores não recolhidos/medidos, identificados com NaN (*Not a Number*). Identifique-os, e substitua cada um desses valores por valores que resultam de um processo de extrapolação e represente graficamente a série temporal modificada, comparando-a com a inicial. Sugestão: Reconstruir os valores em falta usando extrapolação com o método '*pchip*' (**interp1**).

```

In [ ]: % Ex 1.2
        %---Verifica a existência de NaN
        % [hint:] help isnan
        % [hint2:] help any

        %TODO
        %haNaN=

        if haNaN
            fprintf('Foram encontrados elementos NaN.\n');
        else
            fprintf('Não foram encontrados elementos NaN.\n');
        end

        %---Elimina linhas com NaN e reconstrói
        % [hint:] ver slide 16 da aula teórica
        % [hint2:] help find
        % [hint3:] help interp1 (3a função)
        % [hint4:] ver, em particular, o que é preciso para cada índice q
        % ue queremos reconstruir

        %TODO
        %indNaN=

        % ciclo for (índices NaN) para substituir por valores obtidos int
        % erp1
        %x1r=;

        figure(2)
        plot(n,x1,'-+',n,x1r,'-o');
        legend('Série temporal','Série temporal sem NaN','Location','nort
        hwest')
        ylabel('T [°C]')
        xlabel('n [meses]');

```

Exercício 1.3 Determinar os valores da média (**mean**) e do desvio padrão (**std**) da série temporal. Determinar a correlação (**corrcoef**) entre a temperatura na década de 80 e a temperatura na década de 90. Comentar os resultados.

```

In [ ]: % Ex 1.3
%---Média, desvio padrão e correlação – usar sinal reconstruído x
1r!
% [hint:] help mean, std
%mul=;
%sigma1=;

% Determinar temperaturas na década de 80 e 90
%temp80=;
%temp90=;

% [hint:] help corrcoef
%corr=;
fprintf('A correlação entre as duas séries temporais sem NaN é:
%.3f ', corr)

```

Exercício 1.4 Verificar a existência de *outliers*. Identifique-os, substitua-os por valores adequados e represente graficamente a série temporal modificada, comparando-a com as anteriores.

```

In [ ]: % Ex 1.4
%---Verifica outliers
% [hint:] ver slide 17
% [hint2:] help find
%indoutl1=;
%noutl=;

%--- Substituição dos outliers
%x1ro=;

figure(3) % representação gráfica dos resultados
% plot(n,x1,'-+',??,??,'-o',???,???,'-d');
legend('Série temporal','Série sem NaN','Série sem outliers','Loc
ation','northwest')
xlabel('n [meses]');
ylabel('T [°C]')

```

Exercício 2.

A análise da série temporal considera, habitualmente, a existência de componentes associadas a movimentos estruturais e a movimentos erráticos: a) tendência (ou tendência-ciclo, quando agrupada com a componente cíclica) – movimento subjacente de longo-prazo que caracteriza a evolução do nível médio da série; b) sazonal – movimentos estritamente periódicos, decorrentes de características ou fatores que influenciam a evolução da série; c) cíclica – movimentos oscilatórios de tipo recorrente; d) errática/irregular – movimentos aleatórios decorrentes de uma multiplicidade de fatores e de natureza imprevisível. Estas quatro componentes podem ser combinadas de forma multiplicativa ou aditiva (forma a considerar neste exercício).

Exercício 2.1 Estimar a componente da tendência para a série temporal que resulta do exercício 1, considerando aproximações polinomiais de grau 0 e 1 e usando a função **detrend**. Calcular a série sem a tendência. Representar graficamente a série temporal em bruto, a componente da tendência e a série temporal sem a tendência de grau 0 e de grau 1.

```
In [ ]: % Ex. 2.1
        %---Estimar tendência paramétrica (polinomial grau 0 e grau 1) -
        %ver slide 10, 21

        % trend de ordem 0
        % [hint:] help detrend
        % Série sem tendência
        %x1ro_t0=;
        %tr1_0=;

        % trend de ordem 1
        % [hint:] help detrend
        % série sem a tendência
        %x1ro_t1=;
        %tr1_1=;

        figure(4)
        subplot(211)
        %plot(?,?,'+',?,?,?,'-*');
        title('Série (+) e tendência (*) de grau 0')
        xlabel('n [meses]');
        subplot(212)
        %plot(?,?,'-o');
        title('Série sem tendência de grau 0')
        xlabel('n [meses]');

        figure(5)
        % TODO
```

Exercício 2.2 Estimar a tendência quadrática considerando uma aproximação polinomial de grau 2, usando as funções **polyfit** e **polyval**. Representar graficamente a série temporal em bruto, a componente da tendência e a série temporal sem a tendência quadrática.

```
In [ ]: % Ex. 2.2
        % trend de ordem 2 (recordar exercicio da aula 1), ver slide 22
        disp(' Aproximação linear de 2º grau:')
        % [hint:] help polyfit, polyval
        %TODO (2 linhas?)
        %x1ro_t2=;

        figure(6)
        % TODO
```

Exercício 2.3 Estimar a componente da sazonalidade da série temporal (sem a componente da tendência), assumindo uma sazonalidade anual. Representar graficamente a série temporal em bruto, sem as componentes da tendência e da sazonalidade.

```

In [ ]: % Ex 2.3
        %---Estimar Sazonalidade, assumindo uma sazonalidade anual. ver s
        lide 23
        %period=;

        % [hint:] estimar a sazonalidade (slide 19,23).
        %sazl=;

        %x1ro_s=;

        figure(7)
        subplot(211)
        % plot(TODO)
        title('Série (+) e sem sazonalidade (o)')
        subplot(212)
        % plot(TODO)
        title('Sazonalidade da Série')
        xlabel('n [meses]');

```

Exercício 2.4 Obter a componente irregular. Representar graficamente a série temporal regularizada, sem a componente irregular e a respectiva componente irregular.

```

In [ ]: % Ex. 2.4

        %--- Estimar a componente irregular (slide 25)
        %irregl=;

        % série temporal sem a componente irregular
        %x1ro_i=;

        figure(8)
        subplot(211)
        %plot(TODO);
        title('Série com (+) e sem (o) componente irregular')
        subplot(212)
        %plot(TODO);
        title('Componente irregular da Série')
        xlabel('n [Meses]');

```

```

In [5]: ! jupyter nbconvert --to html ATD2022_Ficha3_parteA_todo_w_hints.
        ipynb

```

```

[NbConvertApp] Converting notebook ATD2022_Ficha3_parteA_todo_w_h
ints.ipynb to html

```

```

[NbConvertApp] Writing 554866 bytes to ATD2022_Ficha3_parteA_todo
_w_hints.html

```

```

In [ ]:

```