



Análise e Transformação de Dados

Ficha Prática nº 3 – Parte B

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.

- 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.
- 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** necessitam que se crie um objeto **iddata(y,[],Ts,'TimeUnit','months')** para determinar os parâmetros dos modelos.)
- 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**.
- 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.
- 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).
- 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.
- 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 identificar o modelo da série.
- 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.