

Trabalho 1

Jose Augusto Fiorucci

10/01/2023

Instruções

1. Para cada aluno foi sorteada 1 série temporal do banco de dados da M3;
2. Conforme escolha, os alunos podem formar grupos de até 2 pessoas;
3. Cada grupo deve escolher **apenas 1 série** entre aquelas disponíveis para os membros do grupo;
4. Para a série escolhida escreva um relatório (pequeno e objetivo) com os seguintes procedimentos:
 - a. Apresentação da série selecionada com a decomposição via STL (ou automática via `mstl()`);
 - b. Seleção de 2 modelos ARIMA de forma manual, um sem transformação de Box-Cox e outro com a transformação de Box-Cox;
 - c. Seleção de 2 modelos ETS de forma automática (via `ets()`), um sem transformação de Box-Cox e outro com a transformação de Box-Cox;
 - d. Análise de resíduos de todos os modelos selecionados
 - e. Estudo de desempenho preditivo por janela deslizando comparando os 4 modelos;
 - Inicialize o estudo de janela deslizando em $n-14$ e calcule erros para horizontes até 5;
 - Repita o procedimento para os 4 modelos selecionados, então apresente os resultados em um gráfico em que o eixo-X contém os horizontes de previsão (1, 2, 3, 4, 5) e o eixo-Y contém as médias dos erros absolutos por horizonte de previsão.
 - f. Calcule previsão pontual e previsão intervalar (95%). Utilizando os dados de teste (`$xx`) calcule a métrica MAE e compare a acurácia dos modelos selecionados com os seguintes benchmarks do pacote `forecast`: `auto.arima()`, `ses()`, `holt()`, `ets()`, `stlf()`, `bats()`, `tbats()`.
5. O relatório deve ser estruturado com a seguinte sequência de seções:
 - Introdução: apresente a série selecionada, características e decomposição;
 - Modelos ARIMA: descreva o procedimento de seleção, transformações e análise de resíduos;
 - Modelos ETS: apresente os modelos selecionados e a análise de resíduos;
 - Estudo de desempenho preditivo: apresente os resultados da janela deslizando. Comente sobre a performance de cada modelo em relação a cada horizonte de previsão.;
 - Resultados: apresente as previsões em tabelas e gráficos. Apresente em uma tabela os resultados de acurácia dos 4 modelos selecionados e dos modelos benchmarks. Comente os resultados;
 - Conclusão: descreva em um ou dois parágrafos o que foi observado no estudo;
 - Anexo: inclua o código do R utilizado;
6. Entrega do relatório e apresentação:
 - O tempo de apresentação não deve passar de 10 minutos;

- A apresentação deve ser feita direto no relatório (não precisa preparar slides);
- A entrega do relatório deve ser feita na tarefa do Teams, antes do horário da apresentação;

Banco de dados

As séries temporais foram selecionadas do banco de dados da competição de previsão M3, disponibilizado no pacote *Mcomp* do R.

Para cada uma das 3003 séries temporais desse banco de dados, digamos a série com índice *id*, temos:

- `M3[[id]]$h`, para acessar o tamanho do horizonte de previsão requerido;
- `M3[[id]]$x`, para acessar a parte *in-sample* da série temporal, isto é, a parte a ser utilizada para ajustar o modelo;
- `M3[[id]]$xx`, para acessar a parte *out-sample* da série temporal, isto é, a parte a ser utilizada para medir a acurácia do modelo selecionado;

Exemplo:

```
library(Mcomp)
data(M3)
id=1000      # id da série
M3[[id]]$h   # horizonte de previsão
```

```
## [1] 8
```

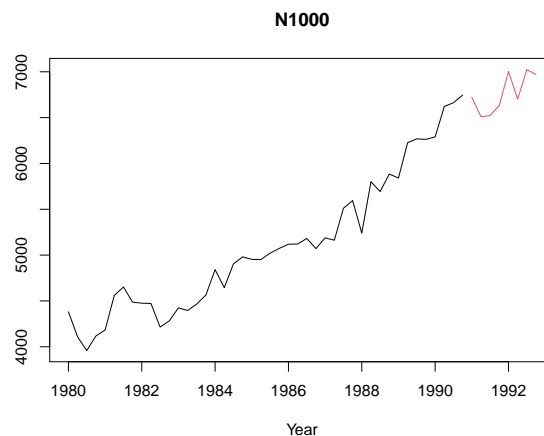
```
M3[[id]]$x    # dados de treinamento
```

```
##      Qtr1    Qtr2    Qtr3    Qtr4
## 1980 4381.5 4107.5 3959.0 4117.5
## 1981 4182.5 4559.5 4652.0 4487.0
## 1982 4475.5 4472.0 4215.0 4282.0
## 1983 4424.5 4395.5 4466.5 4565.5
## 1984 4841.0 4645.0 4904.5 4980.5
## 1985 4953.5 4951.5 5021.5 5073.5
## 1986 5119.0 5120.5 5181.5 5071.0
## 1987 5188.5 5162.5 5511.0 5594.5
## 1988 5239.5 5800.5 5694.0 5884.0
## 1989 5841.0 6226.0 6268.5 6262.5
## 1990 6290.0 6621.5 6662.5 6745.5
```

```
M3[[id]]$xx    # dados de teste
```

```
##      Qtr1    Qtr2    Qtr3    Qtr4
## 1991 6722.0 6509.5 6523.0 6633.5
## 1992 7003.5 6702.0 7023.5 6970.0
```

```
M3[[id]] %>% plot() # gráfico do treinamento e do teste
```



```
## ajuste e previsão do modelo
xx.forec <- auto.arima(M3[[id]]$x, allowdrift=FALSE) %>% forecast(M3[[id]]$h)
## calculo do erro absoluto médio da previsão
MAE <- mean(abs(M3[[id]]$xx - xx.forec$mean))
MAE

## [1] 174.75
```

Séries sorteadas:

Matricula	Nome	Id
220006067	Elaine de Oliveira Jesuino	1694
222107582	ESTEVAO SERGIO ZECO	1950
220006059	Fabiana Mariquito de Sousa	2016
222107608	JOAO VICTOR MONTEIROS DE ANDRADE	2096
220008159	José Luís Polygacz de Nazareno	2475
222107591	LEONARDO SANTOS DA CRUZ	2046
220006024	Marcos Augusto Daza Barbosa	2035
220006041	Mariana Fehr Nicácio	1893
220006172	Mathews de Noronha Silveira Lisboa	1936
222110793	MELQUISADEC DE SOUZA OLIVEIRA	2055
222108042	MOISÉS LIMA	2071
222107564	REBECA KLAMERICK LIMA	1914
221101379	SILVIO FERREIRA GOMES JUNIOR	2040
222108051	TIAGO CHANDIONA ERNESTO	1360
220006032	William Edward Rappel de Amorim	1931

Gráficos das séries:

