UNIVERSIDADE FEDERAL DA PARAÍBA
Séries Temporais, 08/03/21
Professora: Tatiene Souza

## Algumas dicas de como trabalhar com dados no R

```
#Dias da semana para o mes de marco de 2021
format(ISOdate(year=2021, month=3, day=1:31), "%A")
# Ou de forma abreviada,
format(ISOdate(year=2021, month=3, day=1:31), "%a")
meses=format(ISOdate(year=2021, month=1:12, day=1), "%B")
meses=format(ISOdate(year=2021, month=1:12, day=1), "%b")
```

Carregando os dados no R

```
# Listando os conjuntos de dados do pacote tseries do R
data(package='tseries')

# Carregando e obtendo uma breve descrição dos dados. Por exemplo,
help(USAccDeaths)
# Numero totais mensais de mortes acidentais nos EUA.

# Em USAccDeaths há a sequencia temporal dos dados.
# Há 2 formas de mostrar os valores de uma série temporal, a saber:

USAccDeaths
print(USAccDeaths,calendar=F)

start(USAccDeaths) #inicio
end(USAccDeaths) # final
frequency(USAccDeaths) # frequencia

# Instalar pacotes

install.packages("ggplot2")
install.packages("ggfortify")
install.packages("forecast")
install.packages("urca")
install.packages("ggfortify")
install.packages("astsa")
install.packages("readxl")
```

Uma $\boxed{\text{série temporal}}$ é uma coleção de observações feitas sequencialmente ao longo do tempo. A característica mais importante deste tipo de dados é que as observações vizinhas são dependentes e estamos interessados em analisar e modelar esta dependência. Por outro lado, em modelos de regressão a ordem das observações é irrelevante para a análise.

Qualquer conjunto de observações ordenados ao longo do tempo. São exemplo de séries temporais:

- Número de infectados pelo corona vírus;
- Número de óbitos pelo corona vírus;
- Valores diários de poluição na cidade de São Paulo;
- Valores mensais de temperatura em Salvador;
- Índices diários da bolsa valores;
- Precipitação atmosférica anual;
- Número médio de manchas solares;
- Registro de marés no Porto de Santos.

Como a maior parte dos procedimentos estatísticos foi desenvolvida para analisar observações independentes o estudo de séries temporais requer o uso de técnicas específicas. Dados de séries temporais surgem em vários campos do conhecimento como Economia (preço diários de ações, taxa mensal de desemprego, produção industrial), Medicina (eletrocardiograma, eletroencefalograma), Epidemiologia (número mensal de novos casos de meningite), Meteorologia (precipitação pluviométrica, temperatura diária, velocidade do vento), número de casos e óbitos confirmados pelo corona vírus no Brasil (`https://covid.saude.gov.br/`), etc.

Exemplo : Série - AirPassengers

```
library(ggfortify)
library(ggplot2)
library(forecast)
library(urca)
library(astsa)
library(readxl)

help(AirPassengers)
# Totais mensais de passageiros de companhias aéreas, 1949 a 1960.

# tipo do objeto
str(AirPassengers)

# descritiva da série
summary(AirPassengers)

# gráfico da série AirPassengers
plot(AirPassengers)
autoplot(AirPassengers)
# linha tracejada e na cor vermelha
autoplot(AirPassengers, colour = 'red', linetype = 'dashed')
```

$\boxed{\text{Regressão} \times \text{Séries Temporais}}$ Modelos de séries temporais podem ser desenvolvidos de uma maneira bastante semelhante à modelos de regressão. A observação $y_t$ no tempo $t(t = 1,\ldots,T)$ é considerada como a soma de duas componentes, a saber: $y_t$ = componente fixa + erro aleatório. Este modelo é chamado modelo *aditivo* pois chamamos duas componentes a *componente fixa* e o *erro aleatório*. Outra possibilidade é o modelo *multiplicativo*, ou seja, $y_t$ = componente fixa $\times$ erro aleatório.

Diferentemente do modelo de regressão no qual supomos que os erros aleatórios são de uma amostra aleatória de uma distribuição, permitimos agora que os erros numa série temporal não mais sejam independentemente ou identicamente distribuídos, mas que sejam de alguma forma correlacionados.

Analisar uma séries temporal, portanto, exige um trabalho de detetive para descobrir que componentes estão na parte fixa do modelo e que tipo de estrutura de correlação está na parte aleatória ou erro. **Cautela** com alguns pontos:

- Observações correlacionadas são mais difíceis de analisar e requerem técnicas específicas;
- Precisamos levar em conta a ordem temporal das observações;
- Fatores complicadores como presença de tendências e variação sazonal ou cíclica podem ser difíceis de estimar ou remover;
- A seleção de modelos pode ser bastante complicada, e as ferramentas podem ser de difícil interpretação.

$\boxed{\text{Enfoques:}}$ Existem dois enfoques utilizados na análise de séries temporais. Em ambos, o objetivo é construir modelos para estas séries. No primeiro enfoque, a análise é feita no domínio temporal e os modelos propostos são modelos paramétricos (com um número finito de parâmetros). No segundo, a análise é conduzida no domínio de frequências e os modelos propostos são modelos não-paramétricos. O nosso enfoque será baseado no domínio temporal.

$\boxed{\text{Nossa enfâse:}}$ Geração de previsões. Dado $t = 1,\ldots,T$ queremos prever os valores da variável nos tempo $T+1,\ldots$. Incerteza: elemento básico da nossa análise. Pergunta: Como gerar previsões? A partir da observação do comportamento passado da série.

$\boxed{\text{Para que queremos saber o futuro?}}$

- Desejo de saber o que vai acontecer faz parte da natureza humana;
- Previsões são importantes para tomada de decisões.

$\boxed{\text{Questões/considerações:}}$

- O que exatamente desejamos prever?
- Qual a finalidade do exercício de previsão?
- Qual a importância de uma maior precisão nas previsões?

- Que informações se encontram disponíveis?
- Qual o custo aceitável para geração de previsões?
- Artesanalidade do processo;
- Princípio do parcimônia.

$\boxed{\text{Intuição: Processos estocásticos}}$ Suponha que observamos uma série temporal $\{y_1, y_2, \ldots, y_t\}$, decorrente de uma variável aleatória $Y$. A série temporal observada é uma possível realização do processo estocástico gerador dos dados. Pergunta: Por que *uma possível*? Para entender essa questão imagine que amanhã ou chove ou faz sol. Suponha que amanhã acabe fazendo sol, mas poderia ter chovido. Uma possível realização foi o tempo de sol, a outra possível seria chover.

Suponha que você esteja em um local em que há sol, mas possui um amigo que lhe relata chuva no lugar em que se encontra. Do ponto de vista estatístico, imagine que o mesmo processo que determinou o sol determina a chuva em outra localidade.

Um processo estocástico pode ser definido como uma coleção de variáveis aleatórias ordenadas no tempo e definidas em um conjunto de pontos $T$, que pode ser contínuo ou discreto. Iremos denotar a variável aleatória no tempo $t$ por $Y(t)$ no caso contínuo (usualmente $-\infty < t < \infty$), e por $Y_t$ no caso discreto. O conjunto de possíveis valores do processo é chamado de espaço de estados que pode ser discreto (e.g. o número de chamadas que chegam a uma central telefônica a cada 2 horas) ou contínuo (e.g. a temperatura do ar em uma localidade observada em intervalos de 1 hora).

Em análise de séries temporais a situação é bem diferente da maioria dos problemas estatísticos. Embora seja possível variar o tamanho da série observada, usualmente será impossível fazer mais do que uma observação em cada tempo. Assim, tem-se apenas uma realização do processo estocástico e uma única observação da variável aleatória no tempo $t$ denotada por $y_t$.

Um processo puramente aleatório é muitas vezes chamado de $\boxed{\text{ruído branco}}$. Seja $Y_t$ uma sequência de variáveis aleatórias, dizemos que $Y_t$ é um ruído branco se: $\mathrm{E}(Y_t) = 0$; $\mathrm{Var}(Y_t) = \sigma^2$ e $\mathrm{Cov}(Y_t, Y_{t+h}) = 0$. Se $Y_t$ possuir distribuição normal, dizemos que $Y_t$ é um ruído branco gaussiano.

Um importante exemplo de processo estacionário é o ruído branco, o qual é definido como uma sequência de variáveis aleatórias independente, identicamente distribuídas. Muitos processos podem ser construídos a partir do ruído branco.

Seja $Y_t$ uma sequência de variáveis aleatórias, dizemos que $Y_t$ é um $\boxed{\text{passeio aleatório}}$ se: $y_t = y_{t-1} + e_t$, em que $e_t$ é ruído branco. Note que $y_{t-1} = y_{t-2} + e_{t-1}$, e, portanto, $y_t = y_{t-2} + e_{t-1} + e_t$. Consequentemente, $y_{t-2} = y_{t-3} + e_{t-2}$, e, portanto, $y_t = y_{t-3} + e_{t-2} + e_{t-1} + e_t$ e assim sucessivamente. Por fim, $y_t = y_0 + \sum_{j=1}^{T} e_t$. Assumindo que o processo inicia na origem, temos: $y_t = \sum_{t=1}^{T} e_t$.

Exercício : Gere no R um passeio aleatório gaussiano para $T = 1000$.

```
e = rnorm(1000,0,1)
x = cumsum(e)
plot.ts(x,xlab="tempo",ylab="observações")
```

Qual o papel do "tempo"ou "defasagem"? (Gujarati p.591)

Suponha que uma pessoa receba um aumento salarial de 2000 dólares em sua remuneração anual, e que esse aumento seja "permanente", uma vez que o aumento do salário se mantém.
**PERGUNTA:** Qual será o efeito deste aumento de renda no consumo anual da pessoa?

Depois de tal ganho na renda, as pessoas não gastam tudo de uma vez. Suponha que no primeiro ano o consumo aumente em 800 dólares no primeiro ano, mais 600 dólares no segundo ano e mais 400 dólares no ano subsequente, poupando o resto. Ao final do terceiro ano, o consumo anual da pessoa terá aumentado em 1800 dólares. Como podemos escrever a função consumo?

$$Y_t = \alpha + 0.4X_t + 0.3X_{t-1} + 0.2X_{t-2} + e_t, t = 1, \ldots, n$$

em que $Y$ é consumo, $X$ é renda e $\alpha$ é uma constante.
**PERGUNTA:** Como podemos interpretar a função consumo?
Seguindo a uma aumento de 1 dólar na renda, o consumidor aumentará seu nível de consumo em cerca de 0.40 no ano do aumento, mais 0.30 no ano seguinte e mais 0.20 no ano subsequente. Assim, o impacto a longo prazo de um aumento de 1 dólar na renda é de 0.90. Dividindo cada $\beta$ (coeficiente da covariável renda) por 0.90, obtemos, respectivamente, 0.44, 0.33 e 0.23. O que indica que 44% do impacto total de uma mudança unitária em $X$ sobre $Y$ é sentido imediatamente, 77% depois de um ano e por fim, 100%.

Ao regredir uma variável de série temporal sobre outra variável de série de temporal, muitas vezes obtemos um coeficiente de determinação bem alto, embora não haja uma relação significativa entre as duas. Este problema ocorre porque, se ambas as séries envolvidas exibirem fortes tendências (movimentos ascendentes ou descedentes continuados), ou seja, o alto valor do coeficiente de determinação se deve à presença de tendência e não a verdadeira relação entre as duas. Portanto, devemos verificar se a relação entre as variáveis econômicas é verdadeira ou espúria.