

Universidade Federal de Santa Maria
Centro de Ciências Naturais e Exatas
Departamento de Estatística

Notas de Aula

STC1086 - Séries Temporais I

Prof. Dr. Cleber Bisognin



Sumário

1	Introdução	7
2	Séries Temporais	9
2.1	Séries Temporais: Definição Formal	9
2.1.1	Processos Estocásticos	9
2.2	Médias e Covariâncias	11
2.3	Estacionariedade	12
2.3.1	Estacionariedade Forte ou Estrita	14
2.3.2	Estacionariedade Fraca ou de Segunda Ordem	15
2.3.3	Operador de Defasagem ou Operador <i>Lag</i>	17
2.3.4	Ruído Branco	17
2.3.5	Passeio Aleatório (<i>Random Walk</i>)	20
2.4	Média Amostral	22
2.5	Função de Autocovariância Amostral	24
2.6	Função de Autocorrelação Amostral	25
2.6.1	Teste para significância das autocorrelações	26
3	Análise de Tendência	29
3.1	Tendência Determinística	29
3.1.1	Médias Móveis simples (MMS)	31
3.1.2	Tendência Polinomial	36
3.1.3	Tendência Não Linear	45
3.2	Testes de Tendência Determinística	50
3.2.1	Teste de Cox-Stuart	50
3.2.2	Teste de Wald-Wolfowitz	52
3.2.3	Teste de Mann-Kendall	53
3.3	Testes de Raiz Unitária	58
3.3.1	Modelo Geral	58
3.3.2	Teste de Dickey Fuller (DF)	60
3.3.3	Teste Aumentado de Dickey Fuller (ADF)	63
3.3.4	Teste de Phillips - Perron (PP)	65
3.3.5	Teste de Kwiatkowski, Phillips, Schmidt e Shin (KPSS)	66
3.4	Regressão Local (LOESS)	70
4	Análise de Sazonalidade	79
4.1	Testes para Sazonalidade	79
4.1.1	Teste de Kruskal-Wallis	79
4.1.2	Teste de Friedman	80
4.1.3	Teste de Autocorrelação em <i>lags</i> Sazonais	80
4.1.4	Aplicação	81

4.2	Decomposição	82
4.2.1	Modelo Aditivo	84
4.2.2	Modelo Multiplicativo	84
5	Modelos de Suavização	89
5.1	Introdução	89
5.2	Avaliação da Precisão das Previsões	93
5.2.1	Seleção de Modelos Baseada no Erro de Previsão	94
5.3	CrITÉRIOS de Seleção de Modelos	96
5.3.1	CrITÉRIO da Informação de Akaike	96
5.3.2	CrITÉRIO da Informação de Akaike Corrigido	96
5.3.3	CrITÉRIO da Informação Bayesiana de Schwarz	97
5.4	Modelos de Suavização Exponencial	97
5.4.1	Suavização Exponencial Simples	97
5.4.2	Estimação do Parâmetro	98
5.4.3	Aplicação	99
5.5	Suavização Exponencial de Holt	102
5.5.1	Suavização Exponencial de Holt Clássico	102
5.5.2	Suavização Exponencial de Holt Amortecido	102
5.5.3	Estimação dos Parâmetros	103
5.5.4	Aplicação 1	103
5.5.5	Aplicação 2	106
5.6	Modelo de Holt-Winters	107
5.6.1	Holt-Winters Aditivo	108
5.6.2	Holt-Winters Multiplicativo	109
5.6.3	Holt-Winters Multiplicativo com fator de amortecimento	109
5.6.4	Holt-Winters Aditivo com fator de amortecimento	109
5.6.5	Estimação dos Parâmetros	109
5.6.6	Aplicação 1	110
5.6.7	Aplicação 2	114
5.7	Taxonomia de Modelos de Suavização Exponencial	119
5.8	Modelos de Espaço de Estado de Inovações para Suavização Exponencial	122
5.8.1	ETS(A,N,N): Suavização Exponencial Simples com Erros Aditivos	122
5.8.2	ETS(M,N,N): Suavização Exponencial Simples com Erros Multiplicativos	123
5.8.3	ETS(A,A,N): Suavização Exponencial de Holt com Erros Aditivos	123
5.8.4	ETS(M,A,N): Suavização Exponencial de Holt com Erros Multiplicativos	123
5.9	Modelos de Espaço de Estados para todos os ETS	124
5.10	Estimação dos Parâmetros dos Modelos ETS	125
5.11	CrITÉRIOS de Seleção de Modelos	126
5.11.1	CrITÉRIO da Informação de Akaike	126
5.11.2	CrITÉRIO da Informação de Akaike Corrigido	126
5.11.3	CrITÉRIO da Informação Bayesiana de Schwarz	126
5.12	Previsão com os Modelos ETS	127
5.12.1	Intervalos de Confiança de Previsão	127
5.13	A função ets() no R Core Team (2020)	127
5.13.1	Aplicação 1	130
5.14	SÉRIES Temporais com Dupla Sazonalidade	133
5.14.1	Decomposição mstl()	133
5.14.2	Modelo de Holt-Winters com Dupla Sazonalidade Multiplicativa	133
5.14.3	Estimação dos Parâmetros	135

5.14.4	Aplicação	137
6	Outliers	139
6.1	Introdução	139
6.2	Estatística Descritiva	140
6.2.1	Máximo e Mínimo	140
6.2.2	Histograma	140
6.2.3	Boxplot	141
6.2.4	Percentis	143
6.2.5	Filtro de Hampel	144
6.3	Testes Estatísticos	144
6.3.1	Teste de Normalidade de Doornik-Hansen	145
6.3.2	Teste de Normalidade de Anderson-Darling	146
6.3.3	Teste de Grubbs	147
6.3.4	Teste de Dixon	150
6.3.5	Teste de Rosner	152
6.3.6	Observações Adicionais	155
6.4	Transformação de Box-Cox	156
6.5	Aplicação Transformação Box-Cox	157
7	Modelo Vetorial Autorregressivo	159
7.1	Introdução	159
7.2	Regressão Espúria	159
7.3	Modelo Vetorial Autorregressivo de Ordem 1 - VAR(1)	161
7.3.1	Média Constante	162
7.3.2	Matriz de Covariância Constante	162
7.3.3	Matrizes de Covariância e Correlação Cruzada Dependente apenas de 1	163
7.4	Modelo Vetorial Autorregressivo de Ordem p - VAR(p)	164
7.5	Estimação	164
7.5.1	Mínimos Quadráticos Ordinários	165
7.5.2	Verossimilhança Condicional	165
7.6	Definindo a Ordem do Modelo	166
7.7	Diagnóstico do Modelo Estimado	167
7.8	Causalidade de Granger	167
7.9	Função Impulso-Resposta	168
7.10	Função Impulso-Resposta Ortogonalizada	169
7.11	Decomposição da Variância do Erro de Previsão	170
7.12	Modelo Vetorial Autorregressivo Estrutural	170
7.12.1	Do VAR para o SVAR	170
7.12.2	do SVAR para o VAR	171
7.12.3	Identificação	171
7.12.4	Modelo A	172
7.12.5	Modelo B	172
7.12.6	Modelo AB	173
7.13	Processo de Estimação	174
	Referências Bibliográficas	177
8	Adendo A	183

Capítulo 1

Introdução

Esbarramos com séries temporais o tempo todo no dia-a-dia, elas normalmente aparecem nos meios de comunicação na forma de gráfico mostrando índices ou medições como taxas de desemprego, porcentagem de ibope, flutuações nos preços de um produto, etc. Essa estatística é muito utilizada dentro das ciências econômicas (mais precisamente na econometria) com aplicação em diversas outras áreas como demografia, meteorologia, epidemiologia e mais recentemente na ecologia. (<https://bit.ly/3ghSBZs>)

Séries temporais podem ter várias definições, mas todas equivalentes.

Definição 1: Uma série temporal é uma sequência ordenada por tempo de valores de observação de uma variável física ou financeira feita em intervalos de tempo igualmente espaçados Δ_t , representada como um conjunto de valores discretos x_1, x_2, x_3, \dots , etc. Na prática, a sequência de valores é obtida dos sensores por amostragem dos sinais contínuos relacionados. Baseando-se em valores medidos e geralmente corrompidos por ruído, os valores de séries temporais geralmente contêm um componente de sinal determinístico e um componente estocástico representando a interferência de ruído que causa flutuações estatísticas em torno dos valores determinísticos. (<https://bit.ly/2z14k8N>)

Simplificando temos:

Definição 2: Uma Série Temporal é um conjunto de dados ordenados ao longo do tempo, em um intervalo de tempo bem definido. (<https://bit.ly/2z14k8N>)

Definição 3: Uma série temporal é, basicamente, uma sequência de observações tomada ao longo de um período de tempo. Em termos um pouco mais formais. Dizemos que uma série temporal $\{X_t\}_{t \in \mathbb{T}}$ é uma possível realização de um processo estocástico, sendo este uma sequência de variáveis aleatórias. Diversos conjuntos de dados se apresentam como uma série temporal, como a taxa de desemprego, os juros básicos de uma economia, o PIB, a taxa de inflação, etc, o que torna esse campo da econometria extremamente importante. Para além da economia, há também aplicações do que chamamos de econometria de séries temporais na engenharia, nos negócios, nas ciências naturais, nas ciências sociais, etc. (<https://bit.ly/3ghSBZs>)

Sendo um pouco mais formal, podemos verificar como os conceitos de série temporal e processo estocástico se relacionam.

Definição 4: Um processo estocástico é uma sequência ordenada de variáveis aleatórias, podendo ser definido formalmente como

$$\{Y(s, t), s \in S, t \in \mathbb{T}\}$$

onde, para cada $t \in \mathbb{T}$, $y(\cdot, t)$ é uma variável aleatória no espaço amostral S , e a realização desse processo estocástico é dada por $Y(s, \cdot)$ para cada $s \in S$ com respeito a um ponto no tempo $t \in \mathbb{T}$. Consequentemente, a *série temporal que observamos é uma realização particular de um processo estocástico desconhecido*. Ela pode ser representada como segue

$$\{Y_t\}_{t=1}^n = \{Y_1, Y_2, \dots, Y_t, \dots, Y_{n-1}, Y_n\}$$

(<https://bit.ly/3ghSBZs>)

Nesses termos, esse processo estocástico desconhecido é o que chamamos de processo gerador dos dados que formam uma série temporal. Cada Y_t em $\{Y_t\}_{t=1}^n$ é, desse modo, apenas um valor possível de uma variável aleatória. Uma variável aleatória, por seu turno, é aquela que assume valores numéricos e tem um resultado que é determinado por um experimento. O desafio do estatístico e/ou econometrista será, portanto, tendo acesso apenas à série temporal, buscar compreender esse processo estocástico desconhecido. Quanto melhor for essa compreensão, melhor será a modelagem e, com efeito, a previsão de observações futuras. (<https://bit.ly/3ghSBZs>)

A análise de séries temporais leva em conta o fato que os pontos de dados tomados no decorrer do tempo possam ter uma estrutura interna (tal como correlação, tendência ou variações sazonais) que deverá ser considerada. (<https://bit.ly/3bUjwXB>)

A análise de séries temporais tem basicamente dois objetivos principais: (I) Descrição e Modelagem: Envolve caracterização da série de dados, ajuste de modelos matemáticos, extração de índices e comparação entre variáveis ou séries obtidas em sítios distintos. (II) Previsão (*forecasting*): A previsão se baseia na ideia de que o que acontece no futuro próximo dependerá, em grande parte, do que aconteceu ou está acontecendo agora. Em outras palavras, a “memória temporal” existente entre observações próximas é um aspecto essencial quando se deseja aplicar técnicas de previsão. A partir do entendimento de padrões passados é possível fazer previsões anuais, trimestrais ou mensais (descobrir os próximos valores da série). (<https://bit.ly/3cTeonW>)

A Análise de Séries Temporais é usada para muitas aplicações tais como:

- Previsões Econômicas
- Previsões de Vendas
- Análise Orçamentária
- Análise de Mercado de Ações
- Projeções de Rendimentos
- Processos e Controles de Qualidade
- Estudos de Estoques
- projeções de cargas de trabalho
- Estudos de Utilidades
- Análise de Censo

e muito, muito mais... (<https://bit.ly/2Zp9E5L>)

Baseado nas definições de séries temporais, suas estruturas, metodologias, aplicações e programações para sua análise, este material é baseado principalmente nos seguintes materiais: [Wei \(2019\)](#), [Box et al. \(2015\)](#), [Morettin e Toloi \(2006\)](#), [Hamilton \(1994\)](#), [Shumway e Stoffer \(2017\)](#), [Ferreira \(2018\)](#), [Cryer e Chan \(2008\)](#) e [Perlin \(2018\)](#), entre outros.

Capítulo 2

Séries Temporais

O estudo de séries temporais tem por objetivos principais definir o processo gerador de dados, fazer previsões futuras da série, identificar ciclos, tendências e/ou sazonalidades de forma que a decisão que envolve as variáveis em questão seja a mais acurada possível.

2.1 Séries Temporais: Definição Formal

Neste capítulo vamos descrever os conceitos básicos utilizados dentro da teoria dos modelos de séries temporais. Inicialmente vamos introduzir os conceitos de processos estocásticos, média e função de covariância, processo estacionário, e função de autocorrelação.

2.1.1 Processos Estocásticos

Processos Estocásticos

Definição 2.1. Seja T um conjunto arbitrário de índices. Um *processo estocástico* é uma família de variáveis aleatórias $\{Z_t\}_{t \in T}$ definidas num mesmo espaço de probabilidades, que denotaremos genericamente por (Ω, \mathcal{A}, P) . O conjunto de índices T pode ser o conjunto dos números inteiros $\mathbb{Z} = \{0, \pm 1, \pm 2, \dots\}$, dos naturais $\mathbb{N} = \{1, 2, \dots\}$ ou o conjunto dos números reais \mathbb{R} . Observe ainda que, para cada $t \in T$, Z_t é uma variável aleatória definida sobre Ω , sendo assim de fato uma função de dois argumentos, do índice $t \in T$ e do ponto $\omega \in \Omega$ que determina o valor do processo no tempo t dado por $Z_t(\omega)$.

Uma série temporal, do ponto de vista teórico, nada mais é do que um processo estocástico para o qual o índice T é \mathbb{Z} ou um subconjunto deste. Do ponto de vista prático porém, uma série temporal é um conjunto de dados indexados no tempo. Esta dualidade de nomenclatura será utilizado em todo o trabalho. Invariavelmente, letras maiúsculas, como Z_1, Z_2, \dots denotarão a série temporal do ponto de vista teórico, isto é, como variáveis aleatórias em um processo estocástico indexado pelo tempo, enquanto letras minúsculas como z_1, z_2, \dots denotarão a série temporal do ponto de vista prático, isto é, como uma observação das variáveis aleatórias que compõem o processo estocástico. Assim, se do ponto de vista teórico temos a série temporal $\{Z_t\}_{t \in \mathbb{Z}}$, um processo estocástico indexado pelo tempo, uma série temporal do ponto de vista prático significa uma realização z_1, \dots, z_n do processo $\{Z_t\}$, observados nos tempos $t = 1, \dots, n$. Neste caso, observamos $z_1 = Z_1(\omega), \dots, z_n = Z_n(\omega)$, para um determinado $\omega \in \Omega$ fixo. Embora existam maneiras mais formais e precisas de definir uma série temporal, o ponto de vista aqui adotado, embora aparentemente ambíguo na nomenclatura, servirá bem a nossos propósitos sem causar confusões.

Na Figura 2.1(a), para cada $t \in T$, temos uma variável aleatória $Z(t, \omega)$, com uma distribuição de probabilidade. Na Figura 2.1(b), temos uma família de trajetórias.

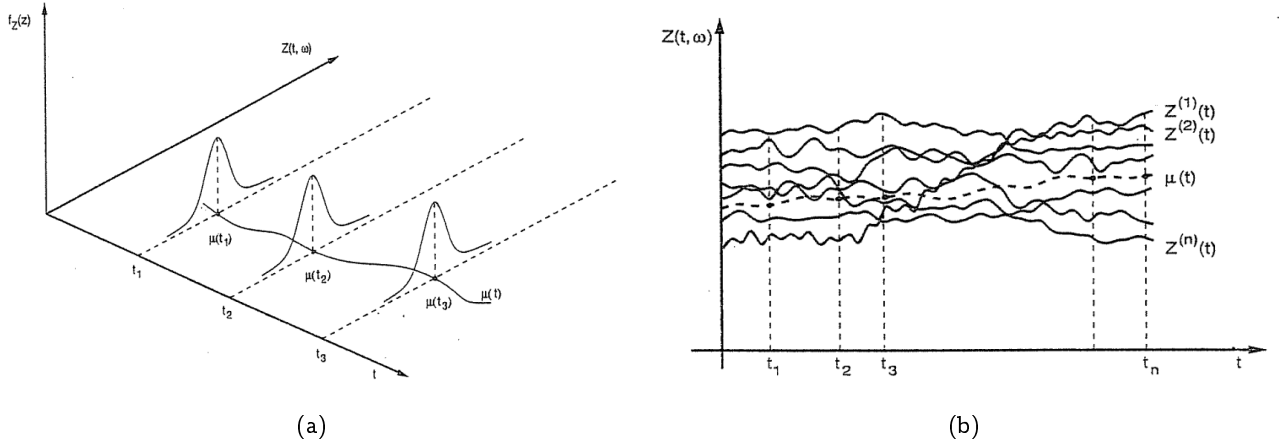


Figura 2.1: Processos Estocásticos: (a) Interpretado como uma família de variáveis aleatórias, (b) Interpretado como uma família de trajetórias. [Fonte: Morettin e Toloi (2006, p.20).]

Exemplo 2.1. Seja $\{X_t\}_{t \in \mathbb{Z}}$ um processo estocástico definido pela equação

$$X_t = A \cos(\theta t) + B \sin(\theta t),$$

onde A e B são duas variáveis aleatórias não correlacionadas com média zero e variância 1 com $\theta \in [-\pi, \pi]$. □

Chamamos atenção ainda que existem condições para que um processo estocástico exista. Estes resultados dependem de uma discussão bastante técnica, bem além das intenções de nossa exposição.

Especificação de um Processo Estocástico

Seja $\{Z_t\}_{t \in T}$ um processo estocástico e para t_1, \dots, t_n elementos quaisquer de T , seja $Z = (Z_{t_1}, \dots, Z_{t_n})$. Denotemos por F_Z a distribuição de Z , isto é, para cada $(z_1, \dots, z_n) \in \mathbb{R}^n$,

$$F_Z(z_1, \dots, z_n) = P(Z_{t_1} \leq z_1, \dots, Z_{t_n} \leq z_n) \quad (2.1)$$

então, o processo estocástico $Z = \{Z(t), t \in T\}$ estará especificado se as distribuições finito-dimensionais de (2.1), são conhecidas para todo $n \geq 1$. Contudo, em termos práticos, não conhecemos todas essas distribuições finito-dimensionais. Estudaremos então certas características associadas a (2.1) e que sejam simples de calcular e interpretar. Uma maneira de especificar o processo Z seria determinar todos os produtos dos momentos, ou seja,

$$\mu(r_1, \dots, r_n; t_1, \dots, t_n) = \mathbb{E}(Z_{t_1}^{r_1} \times \dots \times Z_{t_n}^{r_n}) \quad (2.2)$$

ou

$$\mu(r, t) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} Z_1^{r_1} \dots Z_n^{r_n} f(z_1, \dots, z_n; t_1, \dots, t_n) dz_1 \dots dz_n \quad (2.3)$$

em que $f(Z, t)$ é a função de densidade de $F(Z, t)$. Porém o que vai nos interessar são os momentos de baixa ordem, ou seja, os chamados processos estacionários de 2ª ordem. Consideramos somente os momentos de primeira e segunda ordem, que serão apresentados a seguir.

2.2 Médias e Covariâncias

Naturalmente, quando estamos trabalhando com um processo estocástico, cada variável aleatória que o compõe possui sua própria distribuição, assim como sua própria massa/densidade de probabilidade e sua própria média/variação. Para um processo estocástico $\{Z_t\}_{t \in \mathbb{Z}}$ definimos, para cada $t \in \mathbb{Z}$, a *função média* μ_t e a *função variância* σ_t^2 respectivamente por

$$\mu_t = \mathbb{E}(Z_t) \quad \text{e} \quad \sigma_t^2 = \text{Var}(Z_t), \quad (2.4)$$

desde que as esperanças envolvidas existam. Chamamos a atenção de que embora as esperanças e variâncias de um processo estocástico existam, estas podem ser infinitas. Este fato trás diversos problemas técnicos na análise de séries temporais e requerem técnicas avançadas de análise que estão fora do escopo deste trabalho. Por este motivo, neste trabalho assumiremos tacitamente que todos os processos estocásticos e variáveis aleatórias possuem esperança e variância finitas.

Outra estrutura importante relacionada a um processo estocástico é o que chamamos de estrutura de dependência do processo. Dependência entre variáveis aleatórias pode ser definida de diversas maneiras diferentes. Neste trabalho estamos especialmente interessados na estrutura de dependência relacionadas com a covariância e a correlação entre as variáveis do processo. Observe que num processo estocástico podemos definir a covariância e a correlação entre quaisquer pares Z_i e Z_j de variáveis, para $i, j \in \mathbb{Z}$. No caso de processos, estas funções recebem o prefixo “auto” para enfatizar o fato de que as covariâncias/correlações estão sendo calculadas entre as variáveis do processo. Definimos a *função de autocovariância*, abreviada *FACV*, como

$$\gamma_Z(t, s) = \text{Cov}(Z_t, Z_s) = \mathbb{E}[(Z_t - \mu_t)(Z_s - \mu_s)] = \mathbb{E}(Z_t Z_s) - \mu_t \mu_s, \quad \text{para } t, s \in \mathbb{Z}. \quad (2.5)$$

Analogamente, definimos a *função de autocorrelação*, abreviada *FAC*, por

$$\rho_Z(t, s) = \text{Cor}(Z_t, Z_s) = \frac{\text{Cov}(Z_t, Z_s)}{\sqrt{\text{Var}(Z_t)\text{Var}(Z_s)}} = \frac{\gamma_Z(t, s)}{\sqrt{\gamma_Z(t, t)\gamma_Z(s, s)}}. \quad (2.6)$$

O subscrito “Z” nas definições acima são utilizados para reforçar à qual processo estamos nos referindo. Porém, quando não houver perigo de confusão, podemos eliminar a referência ao processo associado e escrever simplesmente $\gamma_Z(t, s)$ e $\rho_Z(t, s)$.

Observe que, em princípio, as funções $\gamma_Z(t, s)$ e $\rho_Z(t, s)$ dependem tanto de t quanto de s . Nos casos em que isto acontece, qualquer tipo de inferência baseada em autocovariâncias/autocorrelações se torna impossível sem tomarmos medidas para tornar esta estrutura de dependência mais simples. Algumas técnicas relevantes para isso serão estudadas adiante. De qualquer forma, a teoria clássica de séries temporais lida com casos em que essas quantidades possuem uma dependência temporal simplificada, permitindo o seu estudo. Processos com estas características são de grande importância e serão estudados em detalhes mais adiante. Isto por que, do ponto de vista matemático, tal estrutura é conveniente e permite um tratamento rigoroso e aprofundado da teoria enquanto que do ponto de vista prático, é de fácil percepção, permite a modelagem, inferência, previsão e outros aspectos aplicados relevantes de maneira simples e rápida. Tudo isso contribuiu para a difusão de métodos baseado em autocovariâncias/autocorrelações.

Exemplo 2.2. Seja $\{Z_t\}_{t \in \mathbb{Z}}$ um processo estocástico formado por variáveis aleatórias independentes com distribuições alternadas (por exemplo pares e ímpares) entre $N(0, 1)$ e uma distribuição uniforme discreta com $\mathbb{P}(Z_t = -1) = \mathbb{P}(Z_t = 1) = \frac{1}{2}$. Calcule a autocorrelação do processo.

Solução: Temos que $\mathbb{E}(Z_t) = 0$ e $\mathbb{E}(Z_t^2) = 1$ para todo t . Assim,

$$\mathbb{E}(Z_t Z_s) = \begin{cases} 0, & t \neq s, \\ 1, & t = s. \end{cases}$$

$$\rho_Z(t, s) = \frac{\mathbb{E}(Z_t Z_s)}{\sqrt{\mathbb{E}(Z_t^2) \mathbb{E}(Z_s^2)}} = \begin{cases} 0, & t \neq s, \\ 1, & t = s. \end{cases}$$

□

Propriedades Importantes

As seguintes propriedades da função de autocovariância e autocorrelação são análogas às da covariância e correlação ordinárias, mas merecem destaque. Para todo $t, s \in \mathbb{Z}$, com $t \neq s$,

- (1) $\gamma_Z(t, t) = \text{Var}(Z_t)$, $\rho_Z(t, t) = 1$;
- (2) $\gamma_Z(t, s) = \gamma_Z(s, t)$, $\rho_Z(t, s) = \rho_Z(s, t)$.
- (3) $|\gamma_Z(t, s)| \leq \sqrt{\gamma_Z(t, t) \gamma_Z(s, s)}$, $-1 \leq \rho_Z(t, s) \leq 1$.

A propriedade 3 em particular mostra que a covariância entre duas variáveis está bem definida caso estas tenham variância finita.

Como sabemos a correlação é uma medida da dependência linear entre duas variáveis. Se $\text{Cor}(X, Y) = \pm 1$, isto significa que existem constantes β_0 e β_1 tais que $Y = \beta_0 + \beta_1 X$. Ou seja, uma variável é exatamente uma função linear da outra. Valores próximos de ± 1 indicam forte dependência (linear) e valores próximos de 0 indicam fraca dependência (linear). Se $\rho_Z(t, s) = 0$, Z_t e Z_s são ditas não-correlacionadas, mas note que isso não quer dizer que elas são necessariamente independentes. Agora, se Z_t e Z_s são independentes, então $\rho_Z(t, s) = 0$. Por fim, obviamente $\text{Cov}(Z_t, Z_s) = 0$ se, e somente se, Z_t e Z_s são não-correlacionadas.

2.3 Estacionariedade

Nesta seção definiremos o fundamental conceito da estacionariedade de uma série temporal. Existem diversas maneiras de se definir o conceito de estacionariedade, de acordo com as técnicas que se pretendem utilizar na análise das séries temporais. Em poucas palavras, uma série temporal é estacionária quando, com o passar do tempo, a série se desenvolve aleatoriamente em torno de uma média constante, refletindo alguma forma de equilíbrio estável. A ideia é de que uma série temporal estacionária Y tende a “flutuar” aleatoriamente ao redor de uma média constante. A Figura 2.2 apresenta duas séries estacionárias.

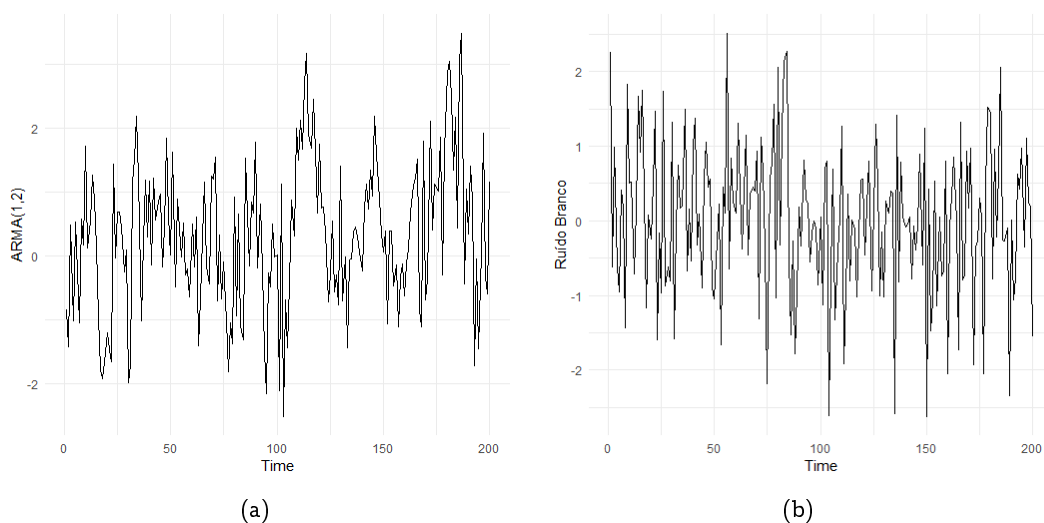


Figura 2.2: Séries estacionárias: (a) ARMA(1, 2), (b) Ruído branco.

```

1 #-----Packages-----
2 suppressWarnings(library(ggplot2))
3 suppressWarnings(library(ggfortify))
4 #
5 set.seed(1087)
6 #Ruído Branco
7 rb<-rnorm(200)
8 rb<-ts(rb)
9 autoplot(rb)+labs(x = "Time", y = "Ruído Branco")+theme_minimal()
10 #ARMA(1,2)
11 arma.sim<-arima.sim(model=list(ar=c(0.9),ma=c(-.7,.1)),n=200)
12 arma.sim<-ts(arma.sim)
13 autoplot(arma.sim)+labs(x = "Time", y = "ARMA(1,2)")+theme_minimal()
14 #

```

Código 2.1: Código fonte em R: Geração da Figura 2.2.

Entretanto, a maior parte das séries que encontramos na prática apresenta alguma forma de não-estacionariedade. A Figura 2.3 apresenta algumas séries que apresentam algum tipo de não-estacionariedade que podem resultar de diversas fontes. Algumas das fontes mais comuns de não-estacionariedade de uma série temporal são:

- (a) a presença de uma tendência determinística (linear, logarítmica, exponencial, etc.) ao redor da qual a série se desenvolve. Geralmente a presença de uma tendência determinística é facilmente reconhecível através do gráfico. Na Figura 2.3(a) apresentamos o gráfico de uma série apresentando uma tendência linear.
- (b) quebra estrutural na série, que pode ser decorrente de uma mudança na média, como representado na Figura 2.3(b), ou uma mudança mais sutil, difícil de ser detectada, como por exemplo mudanças na distribuição da série, na variância, no modelo da série, etc.
- (c) presença do que chamamos de tendência estocástica, como representado na Figura 2.3(c). Neste caso a série parece “vagar” por um caminho que apresenta mudanças aleatórias de trajetória, sendo que fica difícil determinar o seu comportamento.
- (d) presença de sazonalidade. Neste caso a sazonalidade provoca uma mudança de nível local fazendo

com que a média da série se altere nos períodos sazonais. Um exemplo de série sazonal é dado na Figura 2.3(d).

Mais detalhes serão apresentados adiante. A maior parte das séries que encontramos na prática apresenta alguma forma de não-estacionariedade. As séries econômicas apresentam em geral tendências lineares e muito comumente, tendência estocástica. Podemos ter, também, uma forma de não-estacionariedade explosiva, como o crescimento de uma colônia de bactérias.

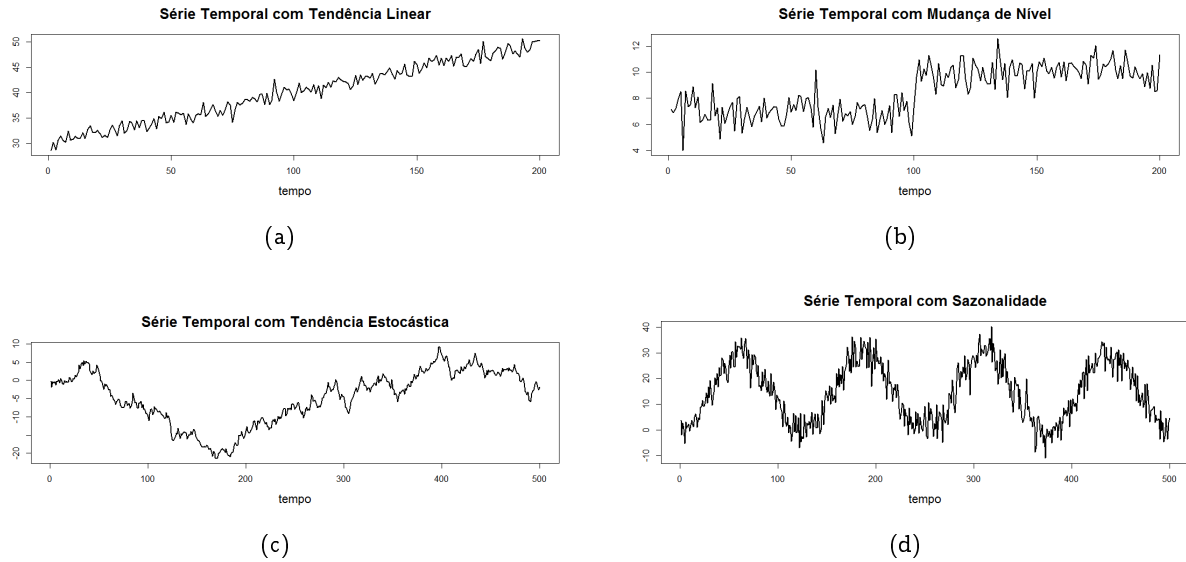


Figura 2.3: Séries não-estacionárias apresentando: (a) Tendência linear, (b) quebra estrutural representada pela mudança de nível da série, (c) tendência estocástica e (d) sazonalidade.

2.3.1 Estacionariedade Forte ou Estrita

Estacionariedade Forte

Definição 2.2. Um processo estocástico $\{Z_t\}_{t \in \mathbb{Z}}$ é dito ser um *processo fortemente (ou estritamente) estacionário* se a distribuição conjunta de Z_{t_1}, \dots, Z_{t_n} é a mesma de $Z_{t_1-k}, Z_{t_2-k}, \dots, Z_{t_n-k}$, para todas as combinações de tempos t_1, \dots, t_n e para todo $k \in \mathbb{Z}$. Observe que este conceito se traduz em dizer que fixados os tempos t_1, \dots, t_n , ao andarmos k passos à frente homogeneamente no tempo, a distribuição das variáveis não se altera.

Quando $n = 1$, a distribuição de Z_t é igual a distribuição de Z_{t-k} para qualquer k , ou seja, os Z_t 's são identicamente distribuídos. Isto implica que num processo fortemente estacionário, as funções média (μ_t) e variância (σ_t^2) são constantes para todo t , isto é, $\sigma^2 = \text{Var}(Z_t) = \text{Var}(Z_{t-k})$ e $\mu = \mathbb{E}(Z_t) = \mathbb{E}(Z_{t-k})$, independentemente de t e k . Quando $n = 2$, a distribuição de (Z_t, Z_s) é a mesma de (Z_{t-k}, Z_{s-k}) , de onde segue que $\text{Cov}(Z_t, Z_s) = \text{Cov}(Z_{t-k}, Z_{s-k})$, para todo t, s e k . Fazendo $k = s$ temos:

$$\gamma_Z(t, s) = \text{Cov}(Z_t, Z_s) = \text{Cov}(Z_{t-k}, Z_{s-k}) = \text{Cov}(Z_{t-s}, Z_{s-s}) = \text{Cov}(Z_{t-s}, Z_0) = \gamma_Z(t-s, 0);$$

e se $k = t$,

$$\gamma_Z(t, s) = \text{Cov}(Z_{t-t}, Z_{s-t}) = \text{Cov}(Z_0, Z_{s-t}) = \text{Cov}(Z_0, Z_{t-s}) = \gamma_Z(0, s-t),$$

de onde podemos concluir que

$$\gamma_Z(t, s) = \gamma_Z(0, |t - s|), \quad \text{lembrando que} \quad |t - s| = \begin{cases} t - s, & \text{para } t > s; \\ s - t, & \text{para } s > t. \end{cases}$$

Analogamente para a função de autocorrelação. Ou seja, num processo fortemente estacionário a covariância entre Z_t e Z_s depende somente da diferença temporal $|t - s|$ e não dos tempos t e s . Ou ainda, podemos dizer que a $\text{Cov}(Z_t, Z_{t+h})$ depende apenas da distância temporal h entre as variáveis (chamada de *defasagem* ou “*lag*” entre as variáveis), e não do tempo t . Isto permite simplificar a notação:

$$\gamma_Z(h) = \text{Cov}(Z_t, Z_{t-h}) = \text{Cov}(Z_t, Z_{t+h}) \quad \rho_Z(h) = \text{Cor}(Z_t, Z_{t-h}) = \text{Cor}(Z_t, Z_{t+h}),$$

para todo $t, h \in \mathbb{Z}$. As propriedades gerais da FAC e FACV para um processo estacionário são:

1. $\gamma_Z(0) = \text{Var}(Z_t), \quad \rho_Z(0) = 1;$
2. $\gamma_Z(h) = \gamma_Z(-h), \quad \rho_Z(h) = \rho_Z(-h);$
3. $|\gamma_Z(h)| \leq \gamma_Z(0), \quad |\rho_Z(h)| \leq 1.$

Se um processo é estritamente estacionário e tem variância finita, então a FACV depende somente do lag h .

2.3.2 Estacionariedade Fraca ou de Segunda Ordem

A estacionariedade forte é um conceito na maioria das vezes difícil de ser identificado na prática, mas muito conveniente do ponto de vista matemático. Uma outra maneira de se definir a estacionariedade de uma série de forma que a teoria é matematicamente tratável e de fácil detecção em problemas práticos é definida a seguir.

Estacionariedade Fraca

Definição 2.3. Um processo estocástico $\{Z_t\}_{t \in \mathbb{Z}}$ é dito ser *fracamente estacionário ou estacionário de segunda-ordem* se satisfaz as seguintes propriedades:

1. $\mathbb{E}(Z_t) = \mu$ para todo tempo t ;
2. $\mathbb{E}|Z_t|^2 < \infty$;
3. $\gamma_Z(t, t - h) = \gamma_Z(0, h) = \gamma_Z(h)$ para todo tempo t e lag h .

A condição $\gamma_Z(t, t - h) = \gamma_Z(h)$ para todo tempo t e lag h é equivalente a $\rho_Z(t, t - h) = \rho_Z(h)$. Além disso, $\text{Var}(Z_t) = \gamma_Z(0)$ não depende de t . Como veremos adiante, em processos fracamente estacionários as funções de autocovariância e autocorrelação desempenham papel central no seu estudo. Neste trabalho, sempre que nos referirmos a um processo estacionário, estaremos nos referindo a processos fracamente estacionários.

Exemplo 2.3. Seja $\{Z_t\}_{t \in \mathbb{Z}}$ um processo estocástico dado por

$$Z_t = A \sin(\omega t + \theta),$$

onde A é a variável aleatória com média zero e variância 1 e θ é uma variável aleatória com distribuição uniforme $[-\pi, \pi]$ independente de A . Encontre a média e a autocovariância do processo estocástico $\{Z_t\}_{t \in \mathbb{Z}}$.

Solução:

$$\mathbb{E}(Z_t) = \mathbb{E}(A)\mathbb{E}[\text{sen}(wt + \theta)] = 0.$$

$$\begin{aligned} \gamma_Z(h) &= \mathbb{E}(Z_t Z_{t+h}) = \mathbb{E}[A^2 \text{sen}(wt + \theta) \text{sen}(w(t+h) + \theta)] \\ &= \mathbb{E}[A^2] \mathbb{E}\left[\frac{1}{2}(\cos(wh) - \cos(w(2t+h) + 2\theta))\right] \end{aligned} \quad (2.7)$$

$$\begin{aligned} &= \frac{1}{2} \cos(wh) - \frac{1}{2} \mathbb{E}[(\cos(w(2t+h) + 2\theta))] \\ &= \frac{1}{2} \cos(wh) - \frac{1}{2} \int_{-\pi}^{\pi} (\cos(w(2t+h) + 2\theta)) \frac{1}{2\pi} d\theta \\ &= \frac{1}{2} \cos(wh) - \frac{1}{8\pi} [\text{sen}(w(2t+h) + 2\theta)] \Big|_{-\pi}^{\pi} \\ &= \frac{1}{2} \cos(wh). \end{aligned} \quad (2.8)$$

A equação (2.7) segue pois $2 \text{sen}(a) \text{sen}(b) = \cos(a-b) - \cos(a+b)$ e a equação (2.8) segue pois $\text{sen}(a) - \text{sen}(b) = 2 \text{sen}\left(\frac{a-b}{2}\right) \cos\left(\frac{a+b}{2}\right)$.

A função de autocovariância depende somente de h .

□

Exemplo 2.4. Seja $\{X_t\}_{t \in \mathbb{Z}}$ um processo estocástico satisfazendo a equação

$$X_t = A \cos(\theta t) + B \text{sen}(\theta t),$$

onde A e B são duas variáveis aleatórias não correlacionadas com média zero e variância um e $\theta \in [-\pi, \pi]$. Verifique se o processo é estacionário.

Solução: Temos que $\mathbb{E}(X_t) = 0$ e

$$\mathbb{E}|X_t|^2 = \mathbb{E}[A \cos(\theta t) + B \text{sen}(\theta t)]^2 = \mathbb{E}[A^2 \cos^2(\theta t) + B^2 \text{sen}^2(\theta t) + 2AB \cos(\theta t) \text{sen}(\theta t)] = 1,$$

pois como A e B são não correlacionadas, $\text{Cov}(A, B) = \mathbb{E}(AB) - \mathbb{E}(A)\mathbb{E}(B) = 0$, portanto $\mathbb{E}(AB) = \mathbb{E}(A)\mathbb{E}(B)$.

$$\begin{aligned} \gamma_Z(h) &= \text{Cov}(X_{t+h}, X_t) \\ &= \text{Cov}[A \cos(\theta(t+h)) + B \text{sen}(\theta(t+h)), A \cos(\theta t) + B \text{sen}(\theta t)] \\ &= \cos(\theta(t+h)) \cos(\theta t) + \text{sen}(\theta(t+h)) \text{sen}(\theta t) \\ &= \cos(\theta h), \end{aligned} \quad (2.9)$$

onde a equação (2.9) segue pois $\cos(a-b) = \cos(a)\cos(b) + \text{sen}(a)\text{sen}(b)$. Portanto o processo $\{X_t\}_{t \in \mathbb{Z}}$ é um processo estacionário (estacionário fraco ou de segunda ordem).

□

Exemplo 2.5. Seja $\{X_t\}_{t \in \mathbb{Z}}$ um processo estocástico satisfazendo a equação

$$X_t = \begin{cases} Y_t, & \text{se } t \text{ é par;} \\ Y_t + 1, & \text{se } t \text{ é ímpar.} \end{cases}$$

Onde $\{Y_t\}_{t \in \mathbb{Z}}$ é um processo estacionário. Verifique se o processo $\{X_t\}_{t \in \mathbb{Z}}$ é estacionário.

Solução: Temos que $\mathbb{E}(X_t) = m$ se t é par e $\mathbb{E}(X_t) = m + 1$ se t é ímpar. Portanto, o processo $\{X_t\}_{t \in \mathbb{Z}}$ não é estacionário pois não tem média constante. □

Relação entre Estacionariedade Forte e Fraca

Observação 2.1. Se um processo estocástico $\{Z_t\}_{t \in \mathbb{Z}}$ é estritamente estacionário segue, tomando $k = 1$, na Definição 2.2, que Z_t possui a mesma distribuição para cada $t \in \mathbb{Z}$. Se $\mathbb{E}|Z_t|^2 < \infty$, implica que $\mathbb{E}(Z_t)$ e $\text{Var}(Z_t)$ são ambas finitas. Além disso, tomando $k = 2$, na Definição 2.2, encontramos que Z_{t+h} e Z_t possuem a mesma distribuição conjunta e portanto a mesma covariância para todo $h \in \mathbb{Z}$. Assim, um processo estritamente estacionário com segundo momento finito é estacionário.

Agora a volta não é verdadeira, ou seja, um processo estacionário não é estritamente estacionário. Por exemplo, se $\{Z_t\}_{t \in \mathbb{Z}}$ é uma sequência de variáveis aleatórias independentes tal que Z_t possui distribuição Exponencial com média 1 para t ímpar e distribuição normal com média 1 e variância 1 se t é par. Então, $\{Z_t\}_{t \in \mathbb{Z}}$ é estacionário com $\gamma_Z(0) = 1$ e $\gamma_Z(h) = 0$ para $h \neq 0$. Além disso, como Z_1 e Z_2 possuem diferentes distribuições, $\{Z_t\}_{t \in \mathbb{Z}}$ não pode ser estritamente estacionário.

2.3.3 Operador de Defasagem ou Operador *Lag*

Em séries temporais é usual trabalhar com operadores que defasam a variável. Definimos então o operador de defasagem B como um operador linear tal que:

Operador Defasagem

$$B^j Z_t = Z_{t-j}$$

As seguintes propriedades do operador B serão úteis no que segue:

1. O operador lag aplicado a uma constante resulta na própria constante, isto é, $Bc = c$;
2. O operador lag segue a propriedade distributiva em relação à soma

$$(B^i + B^j)Z_t = B^i Z_t + B^j Z_t = Z_{t-i} + Z_{t-j};$$

3. É válida a propriedade associativa da multiplicação

$$B^i B^j Z_t = B^i (B^j Z_t) = B^i (Z_{t-j}) = Z_{t-i-j}.$$

4. Potências negativas de B significam um operador de avanço, $B^{-i} Z_t = B^j Z_t$, fazendo $j = -i$. Então $B^{-i} Z_t = B^j Z_t = Z_{t-j} = Z_{t+i}$;

2.3.4 Ruído Branco

Um importante exemplo de processo estacionário é o ruído branco, o qual é definido a seguir.

Ruído Branco

Definição 2.4. Um processo estocástico $\{\varepsilon_t\}_{t \in \mathbb{Z}}$ é dito ser um processo *Ruído Branco* - RB se satisfaz as seguintes propriedades:

1. $\mathbb{E}(\varepsilon_t) = 0$, para todo $t \in \mathbb{Z}$;
2. $\mathbb{E}(\varepsilon_t^2) = \sigma_\varepsilon^2$ para todo $t \in \mathbb{Z}$;
3. $\mathbb{E}(\varepsilon_t \varepsilon_s) = 0$, para todo $t \neq s$, com $t, s \in \mathbb{Z}$.
4. ou $\gamma_\varepsilon(h) = 0$, para todo $h \neq 0$.

Em outras palavras, um ruído branco é uma sequência de variáveis não-correlacionadas com média constante. Denotaremos um processo ruído branco por $\text{RB}(0, \sigma_\varepsilon^2)$.

Muitos processos podem ser construídos a partir de um ruído branco. Pode-se verificar facilmente que se $\{\varepsilon_t\}_{t \in \mathbb{Z}}$ é um $\text{RB}(0, \sigma_\varepsilon^2)$, então é estritamente estacionária, pois

$$\begin{aligned} P(\varepsilon_{t_1} \leq x_1, \varepsilon_{t_2} \leq x_2, \dots, \varepsilon_{t_n} \leq x_n) &= \\ &= P(\varepsilon_{t_1} \leq x_1) P(\varepsilon_{t_2} \leq x_2) \times \dots \times P(\varepsilon_{t_n} \leq x_n) \\ &= P(\varepsilon_{t_1-k} \leq x_1) P(\varepsilon_{t_2-k} \leq x_2) \dots P(\varepsilon_{t_n-k} \leq x_n) \\ &= P(\varepsilon_{t_1-k} \leq x_1, \varepsilon_{t_2-k} \leq x_2, \dots, \varepsilon_{t_n-k} \leq x_n), \end{aligned}$$

onde a primeira igualdade é devido a independência das variáveis e a segunda por serem identicamente distribuídas.

Para um ruído branco $\text{RB}(0, \sigma_\varepsilon^2)$, $\mu_t = \mathbb{E}(\varepsilon_t) = 0$ é constante com *FACV* e *FAC* dadas por

$$\gamma_\varepsilon(h) = \begin{cases} \sigma_\varepsilon^2, & \text{se } h = 0; \\ 0, & \text{se } h \neq 0. \end{cases} \quad \rho_\varepsilon(h) = \begin{cases} 1, & \text{se } h = 0; \\ 0, & \text{se } h \neq 0. \end{cases}$$

```

1 #-----Packages-----
2 suppressWarnings(library(ggplot2))
3 suppressWarnings(library(ggfortify))
4 suppressWarnings(library(forecast))
5 #-----
6 set.seed(1087)
7 #Ruído Branco
8 rb<-rnorm(200)
9 rb<-ts(rb)
10 autoplot(rb)+labs(x = "Time", y = "Ruído Branco Simulado")+theme_minimal()
11 ggAcf(rb, lag.max=100,type = c("correlation"))+labs(y = "FAC Amostral",title="")+theme_
  minimal()
12 ggAcf(rb, lag.max=100,type = c("partial"))+labs(y = "FACP Amostral",title="")+theme_
  minimal()
13 #-----

```

Código 2.2: Código fonte em R: Geração da Figura 2.4.

O termo ruído branco resulta do fato que em uma análise de frequência do modelo, podemos mostrar que todas as frequências são iguais. As características de um processo ruído branco ficam explícitas quando analisamos o seguinte gráfico

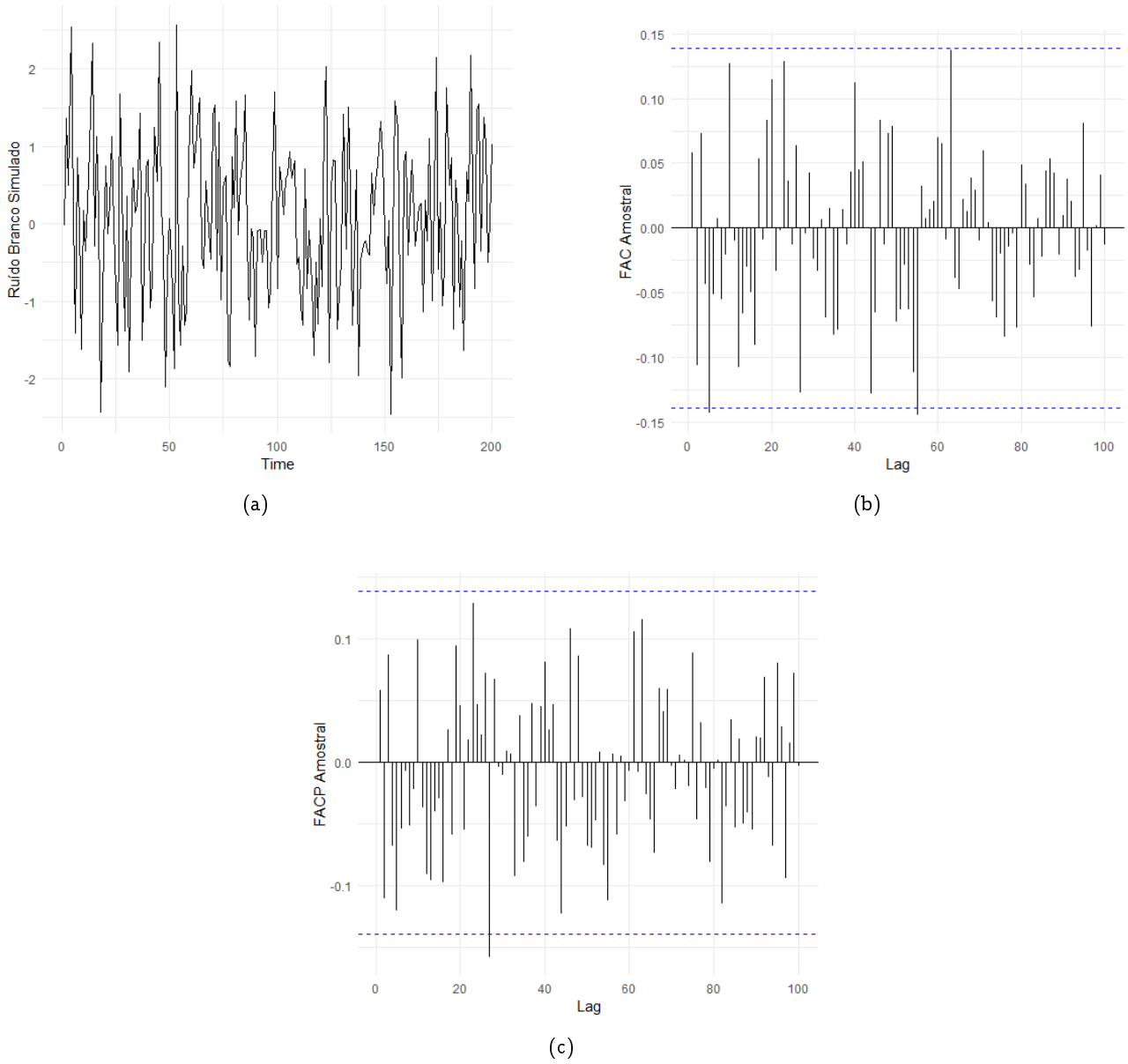


Figura 2.4: (a) Ruído branco gaussiano simulado; (b) FAC amostral; (c) FACP amostral.

O Exemplo 2.6 é um exemplo de um processo estacionário que não é um ruído branco.

Exemplo 2.6. [Média-Móvel de Ordem 1 - MA(1)] Seja $\{Y_t\}_{t \in \mathbb{Z}}$ um processo estocástico satisfazendo a equação

$$Y_t = \varepsilon_t - 0.5\varepsilon_{t-1}, \quad t \in \mathbb{Z},$$

onde $\{\varepsilon_t\}_{t \in \mathbb{Z}}$ é um $\text{RB}(0, \sigma_\varepsilon^2)$.

Solução: Um processo MA(1) possui média e variância dadas por

$$\mu_t = \mathbb{E}(Y_t) = \mathbb{E}(\varepsilon) - 0.5\mathbb{E}(\varepsilon - 1) = 0$$

$$\text{Var}(Y_t) = \text{Var}(\varepsilon_t - 0.5\varepsilon_{t-1}) = \sigma_\varepsilon^2 + 0.5\sigma_\varepsilon^2 = 1.25\sigma_\varepsilon^2.$$

Quanto à estrutura de covariância/correlação de um MA(1), temos

$$\begin{aligned}
 \gamma_Y(h) &= \text{Cov}(Y_t, Y_{t+h}) = \text{Cov}(\varepsilon_t - 0.5\varepsilon_{t-1}, \varepsilon_{t+h} - 0.5\varepsilon_{t+h-1}) \\
 &= \text{Cov}(\varepsilon_t, \varepsilon_{t+h}) - 0.5\text{Cov}(\varepsilon_t, \varepsilon_{t+h-1}) - 0.5\text{Cov}(\varepsilon_{t-1}, \varepsilon_{t+h}) + 0.25\text{Cov}(\varepsilon_{t-1}, \varepsilon_{t+h-1}) \\
 &= \gamma_\varepsilon(h) - 0.5\gamma_\varepsilon(h-1) - 0.5\gamma_\varepsilon(h+1) + 0.25\gamma_\varepsilon(h),
 \end{aligned} \tag{2.10}$$

onde $\gamma_\varepsilon(\cdot)$ denota a função de autocovariância de $\{\varepsilon_t\}_{t \in \mathbb{Z}}$. Da equação (2.10), percebemos que $\text{Cov}(Y_t, Y_{t+h})$ só é diferente de zero quando algum dos argumentos das funções à direita da igualdade em (2.10) é zero. Isto ocorre somente quando $h = 0$ (resultando em $\text{Var}(Y_t) = 1.25\sigma_\varepsilon^2$) e quando $|h| = 1$ (resultando $-0.5\sigma_\varepsilon^2$). Em outras palavras, $\text{Cov}(Y_t, Y_{t+h})$ não depende de t e

$$\gamma(h) = \begin{cases} -0.5\sigma_\varepsilon^2, & \text{se } |h| = 1; \\ 0, & \text{se } |h| > 1. \end{cases} \quad \text{e} \quad \rho(h) = \begin{cases} -0.4, & \text{se } |h| = 1; \\ 0, & \text{se } |h| > 1. \end{cases}$$

Concluimos que um MA(1) é estacionário. □

2.3.5 Passeio Aleatório (*Random Walk*)

Passeio Aleatório

Definição 2.5. Quando $\phi = 1$ no caso anterior, temos o processo chamado passeio aleatório. Seja $\{\varepsilon_t\}_{t \in \mathbb{N}}$ um $\text{WN}(0, \sigma_\varepsilon^2)$. Defina

$$Z_t = Z_{t-1} + \varepsilon_t,$$

que pode ser reescrito de uma maneira bem simples. Defina inicialmente

$$Z_1 = \varepsilon_1, \quad Z_2 = \varepsilon_1 + \varepsilon_2 \leftrightarrow Z_2 = Z_1 + \varepsilon_2$$

e sucessivamente

$$Z_{k-1} = \varepsilon_1 + \cdots + \varepsilon_{k-1}, \quad Z_k = \varepsilon_1 + \cdots + \varepsilon_{k-1} + \varepsilon_k = Z_{k-1} + \varepsilon_k.$$

Com esta representação, o cálculo da média e da variância de Z_t se tornam simples:

$$\mu_t = \mathbb{E}(Z_t) = \mathbb{E}(\varepsilon_1 + \varepsilon_2 + \cdots + \varepsilon_t) = \mathbb{E}(\varepsilon_1) + \mathbb{E}(\varepsilon_2) + \cdots + \mathbb{E}(\varepsilon_t) = 0 + 0 + \cdots + 0 = 0,$$

e

$$\text{Var}(Z_t) = \text{Var}(\varepsilon_1 + \varepsilon_2 + \cdots + \varepsilon_t) = \text{Var}(\varepsilon_1) + \cdots + \text{Var}(\varepsilon_t) = \sigma_\varepsilon^2 + \sigma_\varepsilon^2 + \cdots + \sigma_\varepsilon^2 = t\sigma_\varepsilon^2.$$

Assim concluímos que a variância de um passeio aleatório cresce linearmente com o tempo, sendo portanto um processo não-estacionário. Observe ainda que se $1 \leq t \leq s$, a função de autocovariância de um passeio aleatório é dada por

$$\begin{aligned}
 \gamma_Z(t, s) &= \text{Cov}(Z_t, Z_s) \\
 &= \text{Cov}(\varepsilon_1 + \varepsilon_2 + \cdots + \varepsilon_t, \varepsilon_1 + \varepsilon_2 + \cdots + \varepsilon_s) \\
 &= \sum_{i=1}^t \sum_{j=1}^s \text{Cov}(\varepsilon_i, \varepsilon_j) \\
 &= \text{Cov}(\varepsilon_1, \varepsilon_1) + \text{Cov}(\varepsilon_2, \varepsilon_2) + \cdots + \text{Cov}(\varepsilon_t, \varepsilon_t) \\
 &= \sigma_\varepsilon^2 + \sigma_\varepsilon^2 + \cdots + \sigma_\varepsilon^2 = t\sigma_\varepsilon^2
 \end{aligned}$$

onde $\text{Cov}(\varepsilon_i, \varepsilon_j) = 0$ para $i \neq j$. O mesmo argumento mostra que se $1 \leq s \leq t$, teremos $\gamma_Z(t, s) = s\sigma_\varepsilon^2$, de forma que podemos escrever compactamente $\gamma_Z(s, t) = \min(s, t)\sigma_\varepsilon^2$. A função de autocorrelação de um passeio aleatório é facilmente obtida

$$\begin{aligned}\rho_Z(t, s) &= \frac{\gamma_Z(s, t)}{\sqrt{\text{Var}(X_t)}\sqrt{\text{Var}(X_s)}} = \frac{\min(s, t)\sigma_\varepsilon^2}{\sqrt{t\sigma_\varepsilon^2}\sqrt{s\sigma_\varepsilon^2}} = \frac{\min(s, t)}{\sqrt{t}\sqrt{s}} = \begin{cases} \frac{s}{\sqrt{t}\sqrt{s}} = \sqrt{\frac{s}{t}}, & \text{se } 1 \leq s \leq t; \\ \frac{t}{\sqrt{t}\sqrt{s}} = \sqrt{\frac{t}{s}}, & \text{se } 1 \leq t \leq s \end{cases} \\ &= \sqrt{\frac{\min(s, t)}{\max(s, t)}}.\end{aligned}$$

Em resumo, a FACV e a FAC de um passeio aleatório são dadas por

FACV do passeio aleatório

$$\gamma_Z(t, s) = \min(s, t)\sigma_\varepsilon^2,$$

FAC do passeio aleatório

$$\rho_Z(t, s) = \sqrt{\frac{\min(s, t)}{\max(s, t)}}$$

O passeio aleatório é um exemplo simples que serve de aproximação para diversas situações reais, tais como o movimento comum de preços e títulos e também a posição de pequenas partículas suspensas dentro de um fluido, chamado movimento Browniano.

```

1 #-----Packages-----
2 suppressWarnings(library(ggplot2))
3 suppressWarnings(library(ggfortify))
4 suppressWarnings(library(forecast))
5 #-----
6 set.seed(1087)
7 n<-200
8 ## initialize {x_t} and {w_t}
9 rw <- ww <- rnorm(n = n, mean = 0, sd = 1)
10 ## compute values 2 thru TT
11 for (t in 2:n) {
12   rw[t] <- rw[t - 1] + ww[t]
13 }
14 rw<-ts(rw)
15 autoplot(rw)+labs(x = "Time", y = "Passeio Aleatório Simulado")+theme_minimal()
16 ggAcf(rw, lag.max=100,type = c("correlation"))+labs(y = "FAC Amostral",title="Passeio
17   Aleatório")+theme_minimal()
18 ggAcf(rw, lag.max=100,type = c("partial"))+labs(y = "FACP Amostral",title="Passeio
19   Aleatório")+theme_minimal()
20 #-----

```

Código 2.3: Código fonte em R: Geração da Figura 2.5.

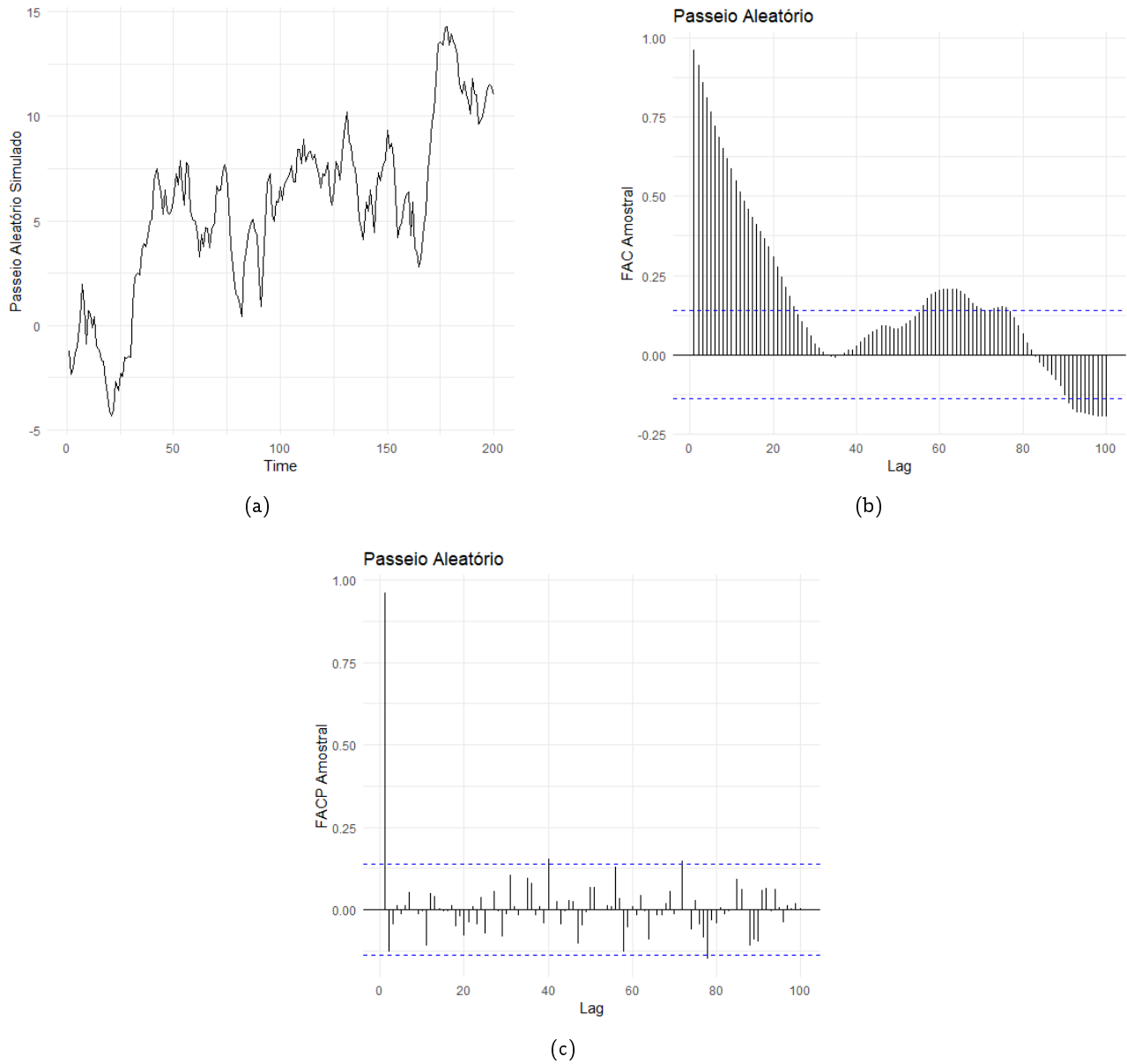


Figura 2.5: (a) Passeio Aleatório simulado; (b) FAC amostral; (c) FACP amostral.

2.4 Média Amostral

Nesta seção estamos trabalhando com uma série temporal estacionária. Uma série temporal é uma amostra de um processo estocástico. Denotamos como série temporal $\{X_t\}_{t=1}^n$. Um estimador natural para a média $\mu = \mathbb{E}(X_t)$ de uma série temporal estacionária é a média amostral

$$\bar{X} = \frac{1}{n} \sum_{t=1}^n X_t$$

A questão agora é se a média amostral é um válido e bom estimador.

$$\mathbb{E}(\bar{X}) = \frac{1}{n} \sum_{t=1}^n \mathbb{E}(X_t) = \frac{1}{n} \cdot n\mu = \mu,$$

o que indica que \bar{X} é um estimador não viesado para μ . Também podemos mostrar que

$$\begin{aligned}
 \text{Var}(\bar{X}) &= \frac{1}{n^2} \sum_{t=1}^n \sum_{s=1}^n \text{Cov}(X_t, X_s) = \frac{\gamma_X(0)}{n^2} \sum_{t=1}^n \sum_{s=1}^n \rho_X(t-s) \\
 &= \frac{\gamma_X(0)}{n^2} \sum_{h=-(n-1)}^{n-1} (n-|h|) \rho_X(h) \\
 &= \frac{\gamma_X(0)}{n} \sum_{h=-(n-1)}^{n-1} \left(1 - \frac{|h|}{n}\right) \rho_X(h),
 \end{aligned} \tag{2.11}$$

onde na equação 2.11 tomamos $h = t - s$. A Tabela 2.1 exemplifica a mudança de variável $h = t - s$ considerando $n = 5$.

Tabela 2.1: Exemplo da mudança de variável $h = t - s$, considerando $n = 5$.

s					
t	1	2	3	4	5
1	$h = 1 - 1 = 0$	$h = 1 - 2 = -1$	$h = 1 - 3 = -2$	$h = 1 - 4 = -3$	$h = 1 - 5 = -4$
2	$h = 2 - 1 = 1$	$h = 2 - 2 = 0$	$h = 2 - 3 = -1$	$h = 2 - 4 = -2$	$h = 2 - 5 = -3$
3	$h = 3 - 1 = 2$	$h = 3 - 2 = 1$	$h = 3 - 3 = 0$	$h = 3 - 4 = -1$	$h = 3 - 5 = -2$
4	$h = 4 - 1 = 3$	$h = 4 - 2 = 2$	$h = 4 - 3 = 1$	$h = 4 - 4 = 0$	$h = 4 - 5 = -1$
5	$h = 5 - 1 = 4$	$h = 5 - 2 = 3$	$h = 5 - 3 = 2$	$h = 5 - 4 = 1$	$h = 5 - 5 = 0$

Para $n = 5$, $h \in \{-4, \dots, -1, 0, 1, \dots, 4\}$, generalizando $h \in \{-(n-1), \dots, -1, 0, 1, \dots, (n-1)\}$.

→ se $h = 0$, temos $(5 - 0) \times \rho_X(0)$,

→ se $h = 1$, temos $(5 - 1) \times \rho_X(1)$

→ se $h = -1$, temos $(5 - 1) \times \rho_X(-1) = (5 - |-1|) \times \rho_X(-1)$

→ se $h = 2$, temos $(5 - 2) \times \rho_X(2)$

→ se $h = -2$, temos $(5 - 2) \times \rho_X(-2) = (5 - |-2|) \times \rho_X(-2)$

Generalizando, para cada $h \in \{-(n-1), \dots, -1, 0, 1, \dots, (n-1)\}$ fixo, temos $(n - |h|) \times \rho_X(h)$. Assim, se

$$\lim_{n \rightarrow \infty} \sum_{h=-(n-1)}^{n-1} \left(1 - \frac{|h|}{n}\right) \rho_X(h),$$

é finito, então $\text{Var}(\bar{X}) \rightarrow 0$, quando $n \rightarrow \infty$ e \bar{X} é um estimador consistente para μ .

Como $\rho_X(h) \rightarrow 0$ quando $h \rightarrow \infty$, temos que para um dado $\epsilon > 0$, podemos escolher um N tal que $|\rho_X(h)| < \epsilon/4$ para todo $h > N$. Então, para $n > (N + 1)$, temos

$$\begin{aligned}
\left| \frac{1}{n} \sum_{h=-(n-1)}^{n-1} \rho_X(h) \right| &\leq \frac{2}{n} \sum_{h=0}^{n-1} |\rho_X(h)| \\
&= \frac{2}{n} \sum_{h=0}^N |\rho_X(h)| + \frac{2}{n} \sum_{h=N+1}^{n-1} |\rho_X(h)| \\
&= \frac{2}{n} \sum_{h=0}^N |\rho_X(h)| + \frac{1}{2} \epsilon \\
&\leq \epsilon,
\end{aligned}$$

onde escolhemos um n grande suficiente para para que o primeiro termo da última desigualdade acima também seja menor que $\epsilon/2$. Assim, para $\rho_X(h) \rightarrow 0$ quando $h \rightarrow \infty$, temos

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{h=-(n-1)}^{n-1} \rho_X(h) = 0,$$

o que implica que

$$\lim_{n \rightarrow \infty} \text{Var}(\bar{X}) = 0.$$

2.5 Função de Autocovariância Amostral

Temos dois estimadores para a função de autocovariância.

$$\hat{\gamma}_X(h) = \frac{1}{n} \sum_{t=1}^{n-h} (X_t - \bar{X})(X_{t+h} - \bar{X}) \quad (2.12)$$

e

$$\hat{\hat{\gamma}}_X(h) = \frac{1}{n-h} \sum_{t=1}^{n-h} (X_t - \bar{X})(X_{t+h} - \bar{X}) \quad (2.13)$$

Calculando o valor esperado de ambas os estimadores temos

$$\mathbb{E}(\hat{\gamma}_X(h)) \simeq \gamma_X(h) - \frac{h}{n} - \left(\frac{n-h}{n} \right) \text{Var}(\bar{X})$$

e

$$\mathbb{E}(\hat{\hat{\gamma}}_X(h)) \simeq \gamma_X(h) - \text{Var}(\bar{X})$$

Podemos perceber que ambos os estimadores são viesados.

1. Quando ignoramos o termo $\text{Var}(\bar{X})$ que é o efeito de estimar μ , $\hat{\hat{\gamma}}_X(h)$ torna-se não viesado, mas $\hat{\gamma}_X(h)$ ainda permanece viesado.
2. Em geral, $\hat{\gamma}_X(h)$ possui maior viés do que $\hat{\hat{\gamma}}_X(h)$, especialmente quando h é grande com respeito a n (se aproxima de n). Logo, para um dado n , é sugerido calcular no máximo $n/4$ autocovariâncias de uma série temporal.
3. Se $\rho_X(h) \rightarrow 0$ quando $h \rightarrow \infty$, então $\lim_{n \rightarrow \infty} \text{Var}(\bar{X}) = 0$ e assim ambos os estimadores $\hat{\gamma}_X(h)$ e $\hat{\hat{\gamma}}_X(h)$ tornam-se não viesados.

4. O estimador $\hat{\gamma}_X(h)$ possui menor erro quadrático médio que o estimador $\widehat{\hat{\gamma}}_X(h)$ (ver [Parzen \(1961\)](#)).
5. Com os resultados a acima, mais o fato do estimador $\hat{\gamma}_X(h)$ sempre ser positivo definido para $\gamma_X(h)$, recomenda-se usar este como estimador da função de autocovariância.
6. Além disso, segundo [Bartlett \(1946\)](#) a variância de $\widehat{\hat{\gamma}}_X(h)$ é maior que a variância de $\hat{\gamma}_X(h)$.

$$\text{Var}(\widehat{\hat{\gamma}}_X(h)) \simeq \frac{1}{n-h} \sum_{i=-\infty}^{\infty} (\gamma_X^2(i) + \gamma_X(i+h)\gamma_X(i-h))$$

$$\text{Var}(\hat{\gamma}_X(h)) \simeq \frac{1}{n} \sum_{i=-\infty}^{\infty} (\gamma_X^2(i) + \gamma_X(i+h)\gamma_X(i-h))$$

Assim,

$$\lim_{n \rightarrow \infty} \hat{\gamma}_X(h) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{t=1}^{n-h} (X_t - \bar{X})(X_{t+h} - \bar{X}) = \gamma(h).$$

2.6 Função de Autocorrelação Amostral

Dada uma série temporal $\{X_t\}_{t=1}^n$ a Função de Autocorrelação amostral é dada por

$$\hat{\rho}_X(h) = \frac{\hat{\gamma}_X(h)}{\hat{\gamma}_X(0)} = \frac{\sum_{t=1}^{n-h} (X_t - \bar{X})(X_{t+h} - \bar{X})}{\sum_{t=1}^{n-h} (X_t - \bar{X})^2}.$$

O gráfico da função $\hat{\rho}(h)$ por h é chamada de correlograma.

Segundo [Bartlett \(1946\)](#), considerando processos gaussianos estacionários, para n grande, $\hat{\rho}_X(h)$ possui distribuição aproximadamente normal com média $\rho_X(h)$ e variância

$$\text{Var}(\hat{\rho}_X(h)) = \frac{1}{n} \sum_{j=-\infty}^{\infty} [\rho_X^2(j) + \rho_X(j+h)\rho_X(j-h) - 4\rho_X(h)\rho_X(j)\rho_X(j-h) + 2\rho_X^2(h)\rho_X^2(j)]$$

Para processos que $\rho_X(h) = 0$ para $h > m$, [Bartlett \(1946\)](#) aproximou a variância acima por

$$\text{Var}(\hat{\rho}_X(h)) = \frac{1}{n} (1 + 2\rho_X^2(1) + 2\rho_X^2(2) + \cdots + 2\rho_X^2(m)).$$

Na prática, as $\rho_X(h)$ são desconhecidas e por este motivo utilizamos a função de autocorrelação amostral e assim temos o desvio padrão de $\hat{\rho}_X(h)$

$$S_{\hat{\rho}_X(h)} = \sqrt{\frac{1}{n} (1 + 2\hat{\rho}_X^2(1) + 2\hat{\rho}_X^2(2) + \cdots + 2\hat{\rho}_X^2(m))}$$

Para um ruído branco $\{\varepsilon_t\}_{t \in \mathbb{Z}}$, $\rho_\varepsilon(h) = 0$, quando $h \neq 0$, neste caso

$$S_{\hat{\rho}_\varepsilon(h)} = \sqrt{\frac{1}{n}}.$$

ou seja, $\hat{\rho}_\varepsilon(h) \sim N\left(0, \frac{1}{n}\right)$.

2.6.1 Teste para significância das autocorrelações

Precisaremos de ferramentas para decidir se uma dada série é correlacionada ou não. Para testar as hipóteses

$$H_0 : \rho(1) = \dots = \rho(m) = 0 \quad \text{vs.} \quad H_1 : \rho(h) \neq 0 \text{ para pelo menos um } h \in \{1, \dots, m\}$$

pode-se usar a estatística Q_{BP} desenvolvida por *Box e Pierce*, ou a estatística Q_{LB} desenvolvida por *Ljung-Box*, definidas, respectivamente, por:

Box e Pierce

$$Q_{BP}(m) = n \sum_{h=1}^m \hat{\rho}(h)$$

em que n é o tamanho da amostra (série) e m é o maior lag considerado na hipótese. A estatística Q_{BP} em grandes amostras tem distribuição qui-quadrado com m graus de liberdade.

Ljung-Box

$$Q_{LB}(m) = n(n+2) \sum_{h=1}^m \frac{\hat{\rho}(h)}{n-h}$$

a qual se distribui como uma qui-quadrado com m graus de liberdade em grandes amostras. A estatística Q_{LB} possui maior poder para amostras pequenas que a estatística Q_{BP} .

Observe que a hipótese nula do teste é que todas as correlações de lag $1, \dots, m$ são nulas, para algum m predeterminado, desta forma a escolha do valor de m é fundamental. Quanto maior o m , caso não seja possível rejeitar a hipótese nula, menor é a evidencia de que a série testada é correlacionada. Porém, se m for muito grande, dois problemas poderão acontecer: primeiro, se m é muito próximo de n haverá poucos pontos amostrais com distância temporal m o que torna a estimação de $\hat{\rho}(h)$ problemática, deteriorando a qualidade do teste; segundo, o poder do teste decresce com o aumento de m . Embora não haja consenso na literatura sobre o valor ideal de m , sugerimos utilizar $m = 20$ para séries com $n \geq 50$. Se a série for curta, na literatura encontra-se a sugestão $m = \min(10, n/5)$.

A estatística do teste terá aproximadamente uma distribuição χ^2 com $m - k$ graus de liberdade, onde m é o número de lags tomados na função de autocorrelação e k número de parâmetros do modelo ajustado. Portanto, rejeitamos a hipótese nula se $Q > \chi^2_{1-\alpha, m-k}$ com um nível de significância α .

Observação 2.2. Testes Box-Pierce ou Ljung-Box. Às vezes, são conhecidos como testes "Portmanteau".

→ `Box.test {stats}`

→ Fonte: <https://bit.ly/2y3CQ6M>

→ Descrição:

Para examinar a hipótese nula de independência em uma determinada série temporal.

→ Hipóteses:

$$\begin{cases} H_0 : \text{A Séries Temporal Não é Autocorrelacionada (Independente)} \\ H_1 : \text{A Séries Temporal é Autocorrelacionada (Dependente).} \end{cases}$$

→ Sintaxe:

`Box.test(x, lag = 1, type = c("Box-Pierce", "Ljung-Box"), fitdf = 0)`

→ Argumentos:

- x : um vetor ou uma séries temporal.

- Lag: número de coeficientes de autocorrelação que serão utilizados para calcular a estatística do teste.

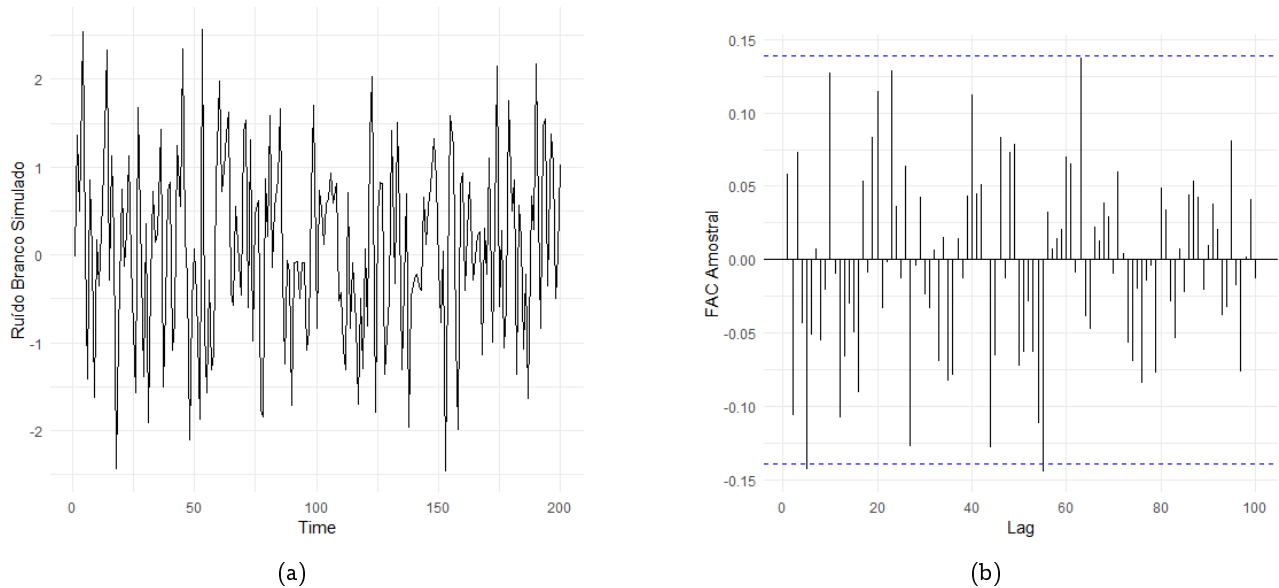


Figura 2.6: (a) Ruído Branco Simulado; (b) FAC amostral.

```

1 #
2 Comandos
3 #
4 set.seed(1087)
5 #Ruído Branco
6 rb<-rnorm(200)
7 rb<-ts(rb)
8 autoplot(rb)+labs(x = "Time", y = "Ruído Branco Simulado")+theme_minimal()
9 ggAcf(rb, lag.max=100,type = c("correlation"))+labs(y = "FAC Amostral",title="")+theme_
  minimal()
10 Box.test(x=rb, lag=10,type="Box-Pierce")
11 #
12   Box-Pierce test
13
14 data:  rb
15 X-squared = 12.929, df = 10, p-value = 0.2277
16 #
17 Box.test(x=rb, lag=10,type="Ljung-Box")
18 #
19   Box-Ljung test
20
21 data:  rb
22 X-squared = 13.432, df = 10, p-value = 0.2005

```

Código 2.4: Código fonte em R: Resultados Testes Box-Pierce ou Ljung-Box.

Observação 2.3. Teste Ljung-Box (Gráfico valor-p versus lag).

→ LBQPlot {FitAR}

→ Fonte: <https://bit.ly/39WfT2y>

→ Descrição:

Elabora o gráfico do valor-p do teste de Ljung-Box versus os lags.

→ Sintaxe:

LBQPlot(res, lag.max = 30, StartLag = k + 1, k = 0, SquaredQ = FALSE)

→ Argumentos:

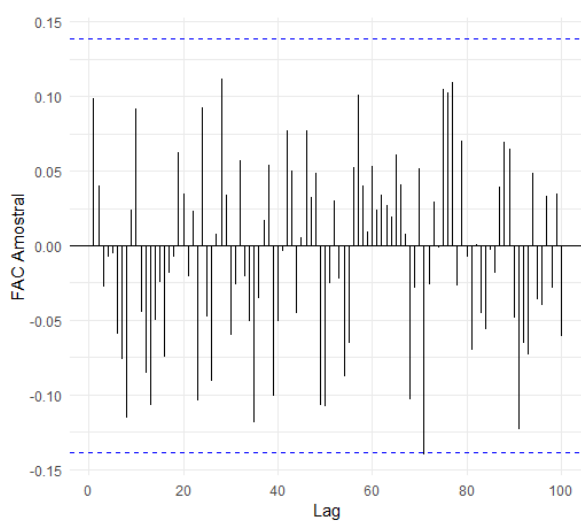
- x : residuals.

```

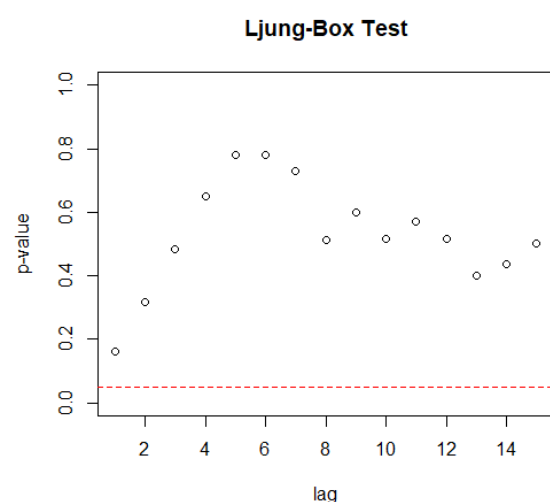
1 #-----
2 Comandos
3 #-----
4 suppressWarnings(library(FitAR))
5 set.seed(1087)
6 #Ruído Branco
7 rb<-rnorm(200)
8 ggAcf(rb, lag.max=100,type = c("correlation"))+labs(y = "FAC Amostral",title="")+theme_
   minimal()
9 LBQPlot(rb, 15)
10 #-----

```

Código 2.5: Código fonte em R: Resultados do Teste de Ljung-Box (valor-p versus lag) (ver Figura 2.7).



(a)



(b)

Figura 2.7: (a) FAC amostral; (b) Teste de Ljung-Box (valor-p versus lag).

Capítulo 3

Análise de Tendência

```
1 library(knitr)
2 library(tseries)
3 library(forecast)
4 library(ggplot2)
5 library(ggfortify)
6 library(plotly)
7 library(magrittr) # para tabulação %>%
8 library(ggseas) #R package aims to help exploratory analysis of time series by making
9 #it easy to do seasonal adjustment and decomposition
10 library(lmtest)
11 library(ggpubr)
12 library(fma)#Data Sets from ‘‘Forecasting: Methods and Applications’’ by
13 #Makridakis, Wheelwright & Hyndman (1998)
14 #devtools::install_github("FinYang/tsdl") #pacote time series data library
15 library(tsdl)
16 library(TTR)
17 library(readxl)
18 library(tidyquant)
19 library(dplyr)
20 library(astsa)#Pacote Livro: Times Series With R.
21 library(L1pack)#lad: Least absolute deviations regression
22 library(webr)
23 library(randtests)
24 library(Kendall)
25 library(trend)
```

Código 3.1: Código fonte em R: Pacotes do R Core Team (2020) utilizados nestes capítulo.

3.1 Tendência Determinística

Considere uma série temporal $\{X_t\}_{t=1}^n$. Temos que um modelo de decomposição consiste em escrever X_t como a soma de três componentes não observáveis

$$X_t = T_t + S_t + \varepsilon_t, \quad t = 1, \dots, n.$$

onde T_t e S_t representam a tendência e a sazonalidade, respectivamente e ε_t a componente aleatória, de média zero e variância constante σ_ε^2 . Se $\{\varepsilon_t\}_{t=1}^n$ são observações de um ruído branco, então $\mathbb{E}(\varepsilon_t \varepsilon_s) = 0$, para $s \neq t$, mas podemos, eventualmente, relaxar esta suposição, tomando $\{\varepsilon_t\}_{t \in \mathbb{Z}}$ um processo estacionário.

Considere $\{X_t\}_{t=1}^n$ uma série não estacionária. O interesse principal em considerar inicialmente um modelo aditivo é o de estimar S_t e construir uma série temporal livre da sazonalidade (ou sazonalmente ajustada). Isto é, se \widehat{S}_t é uma estimativa de S_t , então

$$X_t^s = X_t - \hat{S}_t \quad (3.1)$$

é a séries sazonalmente ajustada. As componentes T_t e S_t são, em geral bastante relacionadas e a influência da tendência sobre a componente sazonal pode ser muito forte Pierce(1979):

- (i) métodos de estimação de S_t podem ser bastante afetados se não levamos em consideração a tendência;
- (ii) a especificação de S_t depende da especificação de T_t ;

Vamos supor que a componente sazonal não esteja presente, ou seja vamos considerar o modelo

$$X_t = T_t + \varepsilon_t, \quad t = 1, \dots, n, \text{ Modelo Aditivo} \quad (3.2)$$

onde $\{\varepsilon_t\}_{t=1}^n$ são observações de um ruído branco.

Há vários métodos de estimar T_t , os mais utilizados consistem em:

- (i) ajustar uma função do tempo, como polinômio, uma exponencial ou outra função suave de t ;
- (ii) suavizar (filtrar) os valores da série ao redor de um ponto (média móvel)

Estimando-se a tendência através de \hat{T}_t , podemos obter a série ajustada para tendência ou livre de tendência fazendo

$$Y_t = X_t - \hat{T}_t.$$

- (iii) outro procedimento que também é utilizado para eliminar a tendência de uma série é diferenciar a série temporal (tendência estocástica). Por exemplo em séries econômicas. Aplicamos diferenças.

Inicialmente definimos o operador defasagem B como um operador linear tal que:

$$B^j X_t = X_{t-j}. \quad (3.3)$$

Diferenciamos uma vez uma série não estacionária aplicando o operador diferença definido como

$$\nabla X_t = (1 - B)X_t = X_t - X_{t-1}. \quad (3.4)$$

Podemos aplicar várias vezes o mesmo operador até tornar a série estacionária.

Existem várias maneiras de eliminarmos tendências determinísticas. Apresentaremos uma metodologia paramétrica de estimação da tendência determinística em uma série. Primeiramente é importante observar que, neste contexto, a forma funcional da tendência determinística deve ser identificada e especificada a priori. Uma maneira muito simples e útil para a remoção da tendência é a inclusão de uma função da variável tempo no modelo, geralmente carregando informações sobre o formato da tendência que se quer remover. Assumiremos que a forma funcional da tendência determinística depende de certos parâmetros de forma linear.

3.1.1 Médias Móveis simples (MMS)

Considere a série temporal $\{Z_t\}_{t=1}^n$, estacionária e localmente constante, composta de seu nível e mais um ruído aleatório da seguinte forma

$$Z_t = \mu_t + \varepsilon_t, \quad t = 1, \dots, n$$

onde $\mathbb{E}(\varepsilon_t) = 0$, $\text{Var}(\varepsilon_t) = \sigma_\varepsilon^2$ e μ_t é um parâmetro desconhecido que varia com o tempo.

A técnica de média móvel consiste em calcular a média aritmética das k observações mais recentes, ou seja

$$\hat{\mu}_t = \frac{Z_t + Z_{t-1} + \dots + Z_{t-k+1}}{k} = \hat{\mu}_{t-1} + \frac{Z_t - Z_{t-k}}{k}$$

. Denotamos por k como sendo o comprimento da média.

Desta forma, $\hat{\mu}_t$ é uma estimativa do nível μ_t que não leva em consideração as observações mais antigas. Note que a cada período a observação mais antiga é substituída pela mais recente, calculando-se uma média nova.

Portanto, a previsão dos h valores futuros é dada pela última média móvel calculada, ou seja

$$\hat{Z}_t(h) = \hat{\mu}_t,$$

ou ainda

$$\hat{Z}_t(h) = \hat{Z}_{t-1}(h+1) + \frac{Z_t - Z_{t-k}}{k}, \quad \forall h > 0.$$

Podemos perceber que a equação acima corrige a previsão de Z_{t+h} a cada instante, ou seja, a cada nova observação na série, Z_{t+h} é atualizado.

Assumindo que $\{\varepsilon_t\}_{t=1}^n$ é um ruído branco com distribuição $N(0, \sigma_\varepsilon^2)$, podemos afirmar que $\hat{Z}_t(h) \sim N\left(\mu_t, \frac{\sigma_\varepsilon^2}{k}\right)$ e construir um intervalo de confiança com coeficiente de confiança $(1 - \alpha)100\%$, dado por

$$\left(\hat{Z}_t(h) - z_{\frac{\alpha}{2}} \frac{\sigma_\varepsilon}{\sqrt{k}}; \hat{Z}_t(h) + z_{\frac{\alpha}{2}} \frac{\sigma_\varepsilon}{\sqrt{k}} \right),$$

onde $z_{\frac{\alpha}{2}}$ é o quantil de uma distribuição normal padrão, com coeficiente de confiança $(1 - \alpha)100\%$.

A Figura 3.1 apresenta os gráficos dos Casos Diários Confirmados de Covid-19 - Santa Maria-RS e suas médias móveis simples de tamanho $k \in \{7, 14, 21\}$ dias. A fonte dos dados de todos os municípios do Brasil foi: *O Brasil em dados libertos*. Um Repositório de dados públicos disponibilizados em formato acessível no link <https://brasil.io/home>.

```
1 covidsm<-read_excel("covid_sm.xlsx", col_types = c("date", "numeric"))
2 t<-1:length(covidsm[,1])
3 n<-length(covidsm[,1])
4 framed<-cbind(rep("d",n),(covidsm))
5 colnames(framed)<-c("cond","data","dados")
6 frame7<-as.data.frame(cbind(rep("k7",n),covidsm[,1],SMA(x=covidsm[,2],n=7)))
7 colnames(frame7)<-c("cond","data","dados")
8 frame14<-as.data.frame(cbind(rep("k14",n),covidsm[,1],SMA(x=covidsm[,2],n=14)))
9 colnames(frame14)<-c("cond","data","dados")
10 frame21<-as.data.frame(cbind(rep("k21",n),covidsm[,1],SMA(x=covidsm[,2],n=21)))
11 colnames(frame21)<-c("cond","data","dados")
12 df<-rbind(framed,frame7,frame14,frame21)
13 ggplot(df, aes(x=data, y=dados, group=cond, colour=cond)) +
14   geom_line()+
15   scale_color_manual(values = c("d" = "black", "k7" = "blue3", "k14" = "red", "k21" = "green"))+
16   labs(y = "Casos Confirmados", x = "Tempo (dias)") + theme_minimal()
```

Código 3.2: Código fonte em R: Casos Diários Confirmados de Covid-19 - Santa Maria-RS. Fonte: Brasil.io.

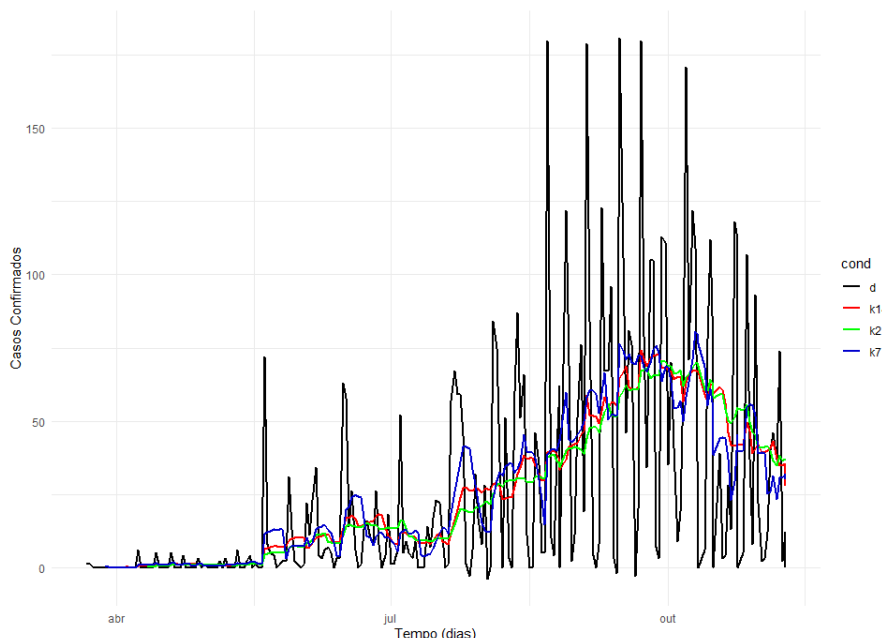


Figura 3.1: Casos Diários Confirmados de Covid-19 - Santa Maria-RS. Fonte: Brasil.io.

A Figura 3.2 apresenta os gráficos dos Casos Diários Confirmados de Covid-19 - Santa Maria-RS e suas médias móveis simples de tamanho $k = 7$ dias. A figura 3.3 apresenta a série dos Casos Diários Confirmados com a tendência retirada por meio da média móvel simples de tamanho $k = 7$ dias.

```

1 covidsm<-read_excel("covid_sm.xlsx",col_types = c("date", "numeric"))
2 t<-1:length(covidsm[,1])
3 n<-length(covidsm[,1])
4 smac<-SMA(x=covidsm[,2],n=7)
5 n1<-length(smac)
6 n2<-n1-6
7 smac<-smac[7:n1]
8 framed<-cbind(rep("d",n2),covidsm[7:n1,])
9 colnames(framed)<-c("cond","data","dados")
10 frame7<-as.data.frame(cbind(rep("k7",n2),covidsm[7:n1,1],smac))
11 colnames(frame7)<-c("cond","data","dados")
12 df<-rbind(framed,frame7)
13 ggplot(df, aes(x=data, y=dados, group=cond, colour=cond)) +
14   geom_line()+
15   scale_color_manual(values = c("d" = "black", "k7" = "blue3"))+
16   labs(y = "Casos Confirmados", x = "Tempo (dias)")+theme_minimal()
17 dframe<-as.data.frame(cbind(rep("detrend",n2),covidsm[7:n1,1],covidsm[7:n1,2]-smac))
18 colnames(dframe)<-c("cond","data","detrend")
19 ggplot(dframe, aes(x=data, y=detrend, group=cond, colour=cond)) +
20   geom_line()+scale_color_manual(values = c("detrend" = "black"))+
21   labs(y = "Casos Confirmados Sem Tendência", x = "Tempo (dias)")+theme_minimal()

```

Código 3.3: Código fonte em R: Casos Diários Confirmados de Covid-19 - Santa Maria-RS - Sem Tendência Por Médias Móveis Simples.

A Figura 3.4 apresenta o gráfico da Função de Autocorrelação Amostral da Série Temporal dos Casos Diários Confirmados de Covid-19 - Santa Maria-RS com a tendência retirada por meio da média

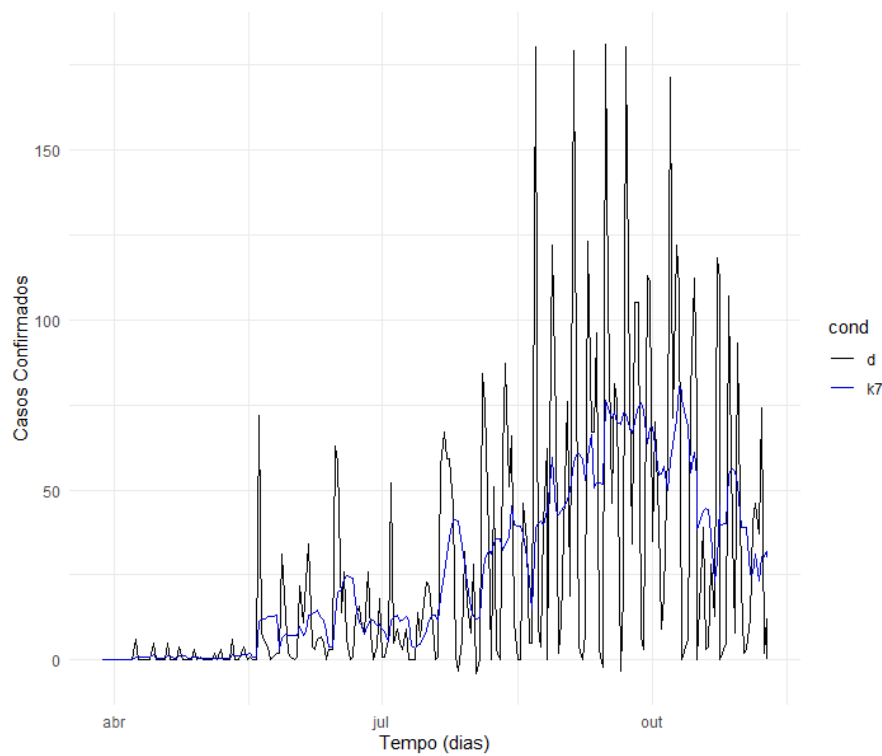


Figura 3.2: Casos Diários Confirmados de Covid-19 - Santa Maria-RS - Tendência Média Móvel Simples de tamanho $k = 7$ dias.

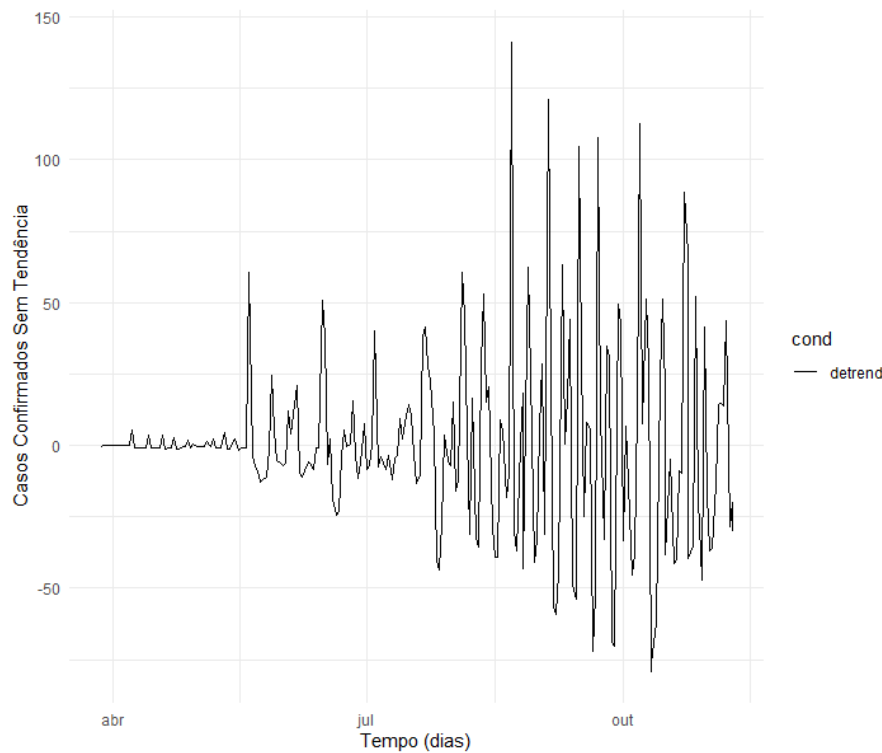


Figura 3.3: Casos Diários Confirmados de Covid-19 - Santa Maria-RS - Sem Tendência Por Médias Móveis Simples de tamanho $k = 7$ dias.

móvel simples de tamanho $k = 7$ dias.

```
1 ggAcf(dframe$detrend, lag.max=20)+labs(y = "FAC Série sem Tendência", title="")+theme_
  minimal()
```

Código 3.4: Código fonte em R: FAC Amostral dos Casos Diários Confirmados de Covid-19 - Santa Maria-RS - Sem Tendência Por Médias Móveis Simples.

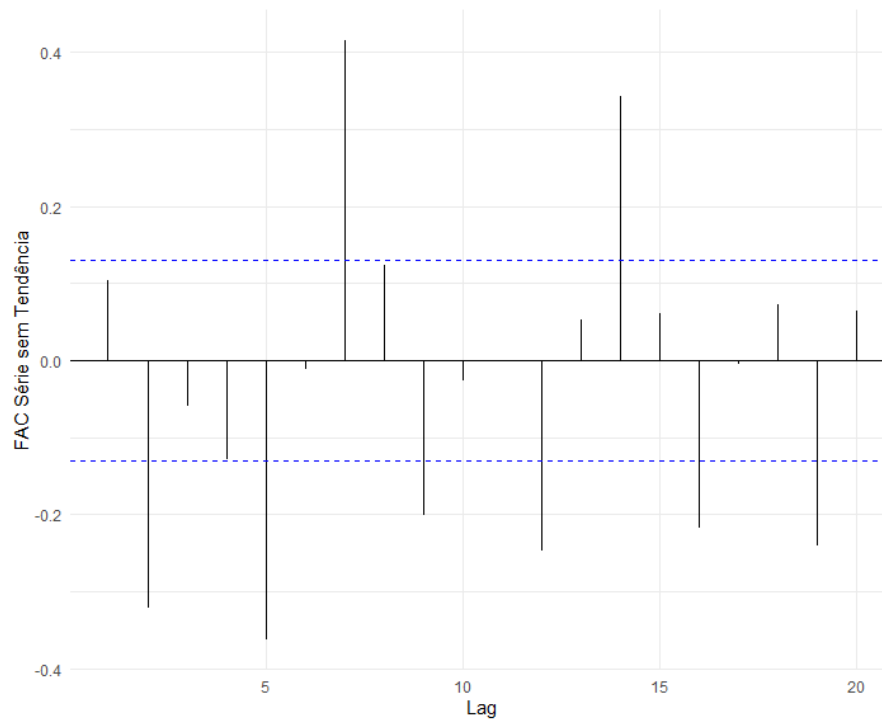


Figura 3.4: Casos Diários Confirmados de Covid-19 - Santa Maria-RS - Sem Tendência Por Médias Móveis Simples de tamanho $k = 7$ dias.

```
1 #
2 #Média Móvel Simples
3 #
4 sma.for<-function(x,k,h,conf,plot=TRUE)
5 {
6   z<-c()
7   z$Dados<-x
8   n<-length(x)
9   y<-ysma<-SMA(x,k)
10  z$fitted.values<-ysma
11  xa<-c(x,rep(NA,h))
12  xa[n+1]<-y[n]
13  for(t in (n+1):(n+h)){
14    y[t]<-mean(xa[(t-1):(t-k)])
15    xa[t+1]<-y[t]
16  }
17  z$residuals<-c(x-ysma)[k:n]
18  ve<-var(z$residuals)
19  q<-qnorm((1-conf)/2,lower.tail = FALSE)
20  ep<-q*sqrt(ve/k)
21  f<-cbind(y[(n+1):(n+h)]-ep,y[(n+1):(n+h)],y[(n+1):(n+h)]+ep)
22  if(plot){
```

```

23   fore<-rbind( matrix( ncol=3,nrow=n) , f)
24   ysmaf<-c( ysmas, rep( NA, h) )
25   dados<-c( x, rep( NA, h) )
26   df<-as.data.frame( cbind( t=1:(n+h), dados=dados, ysmas=ysmaf, fore=fore) )
27   colnames( df)<-c( "t", "Dados", "SMA", "LI", "Prev", "LS" )
28   fig <- plot_ly( df, x = ~t, y = df$Dados, name = 'Data', type = 'scatter', mode = '
      lines' )
29   fig <- fig %>% add_trace( y = ~df$SMA, name = 'SMA', mode = 'lines' )
30   fig <- fig %>% add_trace( y = ~df$LS, name = 'LS', mode = 'lines' )
31   fig <- fig %>% add_trace( y = ~df$LI, name = 'LI', mode = 'lines' )
32   fig <- fig %>% add_trace( y = ~df$Prev, name = 'Forecast', mode = 'lines' )
33   fig <- fig %>% layout( title = paste0( "Médias Móveis Simples - k=", k) ,
34                          xaxis = list( title = 'Tempo' ) ,
35                          yaxis = list( title = 'Dados' ) )
36   #return( fig )
37 }
38 colnames( f)<-c( paste0( "LI:", 100*conf, "%" ) , "Previsão" , paste0( "LS:", 100*conf, "%" ) )
39 z$Forecast<-f
40 return( z )
41 }
42 #
43 #Exemplo
44 #
45 x <-read_excel( "temp.xlsx" )
46 x<-x$Temperatura
47 sma.for( x=x, k=15, h=10, conf=0.95, plot=TRUE)
48 sma.for( x=x, k=15, h=10, conf=0.95, plot=FALSE)

```

Código 3.5: Código fonte em R: Série temporal referente às médias anuais das temperaturas na cidade de Nova York durante os anos de 1912 e 1971.

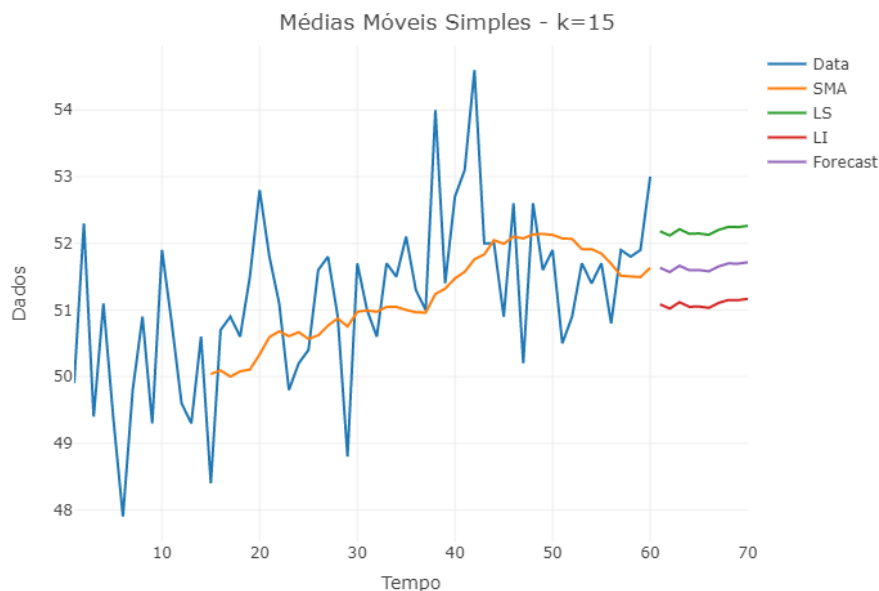


Figura 3.5: Série temporal referente às médias anuais das temperaturas na cidade de Nova York durante os anos de 1912 e 1971. Previsão utilizando Médias Móveis Simples com $k=15$.

A Figura 3.4 apresenta a Série temporal referente às médias anuais das temperaturas na cidade de Nova York durante os anos de 1912 e 1971. Previsão utilizando Médias Móveis Simples com $k=15$.

3.1.2 Tendência Polinomial

Um procedimento utilizado para eliminar tendência polinomial é ajustar uma curva aos valores observados da séries temporal para estimar T_t e fazer previsões. Tradicionalmente são utilizadas funções exponenciais, logarítmicas ou polinômios.

Suponha que

$$T_t = \beta_0 + \beta_1 t + \cdots + \beta_m t^m,$$

onde o grau m do polinômio é bem menor que o número de observações. para estimar os parâmetros β_j utilizamos o método dos mínimos quadrados, ou seja, minimizamos a função

$$\sum_{t=1}^n (X_t - T_t)^2 = \sum_{t=1}^n (X_t - \beta_0 - \beta_1 t - \cdots - \beta_m t^m)^2 \quad (3.5)$$

onde $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_m)$ as estimativas de $\beta = (\beta_0, \beta_1, \dots, \beta_m)$ obtidos pelo método dos mínimos quadrados. As estimativas de $\beta = (\beta_0, \beta_1, \dots, \beta_m)$ podem ser adquiridas também por meio de regressões robustas.

```
1 icv<-c(238,251,256,263,270,275,280,290,298,305,310,318,
2       329,343,359,375,383,393,400,407,415,424,436,449)
3 t<-1:length(icv)
4 df.icv<-as.data.frame(cbind(t,icv))
5 ggplot(df.icv, aes(x=t, y=icv)) +
6   geom_point() +
7   geom_smooth(method = lm, formula = y ~ x, se = TRUE)+
8   labs(y = "Índice de Custo de Vida", x = "Tempo")+theme_minimal()
```

Código 3.6: Código fonte em R: Índice de Custo de Vida no Município de São Paulo. Fonte: Fundação Instituto de Pesquisas Economicas (FIPE) - Função `geom_smooth`.

As Figuras 3.6 (utilizando o comando `geom_smooth()`) e 3.7 (de forma manual) apresentam o gráfico da série temporal do Índice de Custo de Vida no Município de São Paulo. Fonte: Fundação Instituto de Pesquisas Economicas (FIPE) com a reta de tendência $icv = \beta_0 + t\beta_1$, com $t = 1, \dots, n$, onde n é o tamanho da série temporal.

```
1 icv <- read_excel("icv.xlsx", col_types = c("date", "numeric"))
2 n<-length(icv$icv)
3 t<-1:n
4 df.icv<-as.data.frame(cbind(t,icv))
5 reg<-lm(icv~t, data=df.icv)
6 summary(reg)
7
8 Saída summary(reg)
9
10 Call:
11 lm(formula = icv ~ t, data = df.icv)
12
13 Residuals:
14     Min       1Q   Median       3Q      Max
15 -13.518  -5.553   1.160   5.607  11.626
16
17 Coefficients:
18             Estimate Std. Error t value Pr(>|t|)
19 (Intercept)  220.9457     3.0448   72.57  <2e-16 ***
20 t             9.2143     0.2131   43.24  <2e-16 ***
21
22 Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
23
```

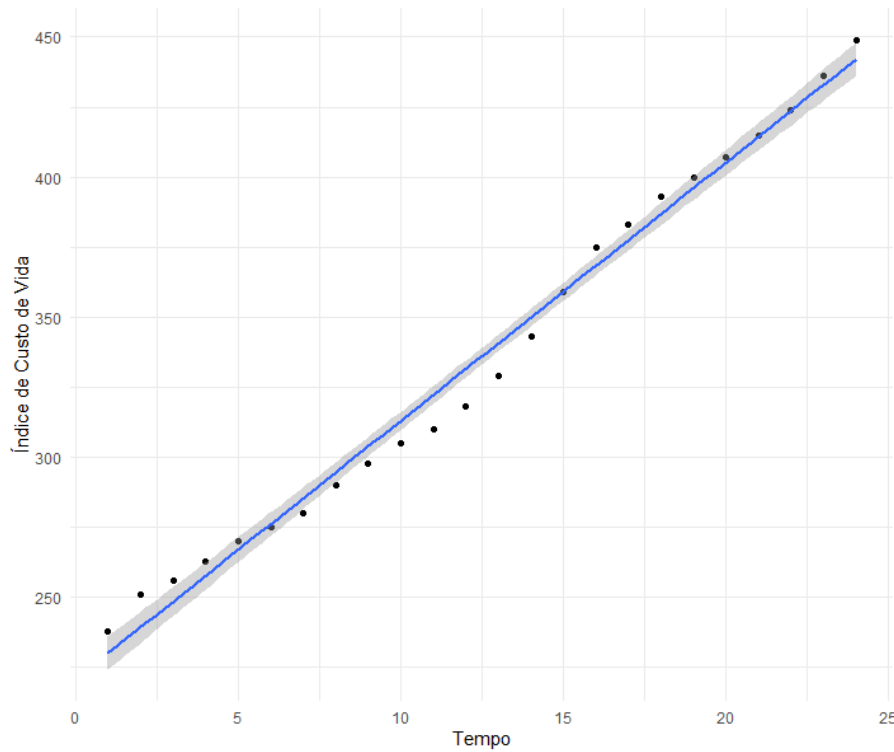


Figura 3.6: Índice de Custo de Vida no Município de São Paulo - Ajuste Tendência Polinomial - Função `geom_smooth`. Fonte: Fundação Instituto de Pesquisas Econômicas (FIPE).

```

24 Residual standard error: 7.226 on 22 degrees of freedom
25 Multiple R-squared: 0.9884, Adjusted R-squared: 0.9878
26 F-statistic: 1870 on 1 and 22 DF, p-value: < 2.2e-16
27
28 y<-reg$coefficients[1]+reg$coefficients[2]*t
29 y<-round(y,1)
30 df1<-as.data.frame(cbind(rep("icv",n),icv))
31 df2<-as.data.frame(cbind(rep("ajustado",n),icv[-2],y))
32 colnames(df1)<-c("cond","t","dados")
33 colnames(df2)<-c("cond","t","dados")
34 df.icv0<-rbind(df1,df2)
35 ggplot(df.icv0, aes(x=t,y=dados,group=cond,colour=cond)) +
36   geom_line(size = 1)

```

Código 3.7: Código fonte em R: Índice de Custo de Vida no Município de São Paulo. Fonte: Fundação Instituto de Pesquisas Econômicas (FIPE).

```

1 d.st<-icv[,2]-y #série sem a tendência
2 d.st<-ts(d.st, start=c(1976,1), frequency=12)
3 autoplot(d.st)+labs(x = "Meses", y = "Série sem Tendência")+theme_minimal()

```

Código 3.8: Código fonte em R: Índice de Custo de Vida no Município de São Paulo sem a Tendência Polinomial.

A Figura 3.8 apresenta o gráfico da série temporal do Índice de Custo de Vida no Município de São Paulo sem a reta de tendência $icv = \beta_0 + t\beta_1$, com $\hat{\beta}_0 = 220,9457$ e $\hat{\beta}_1 = 9,2143$, para $t = 1, \dots, n$, onde n é o tamanho da série temporal. Neste caso, denotamos os resíduos por $\{\hat{\varepsilon}_t\}_{t=1}^n$, onde

$$\hat{\varepsilon}_t = X_t - \hat{T}_t = X_t - [\hat{\beta}_0 + t\hat{\beta}_1],$$

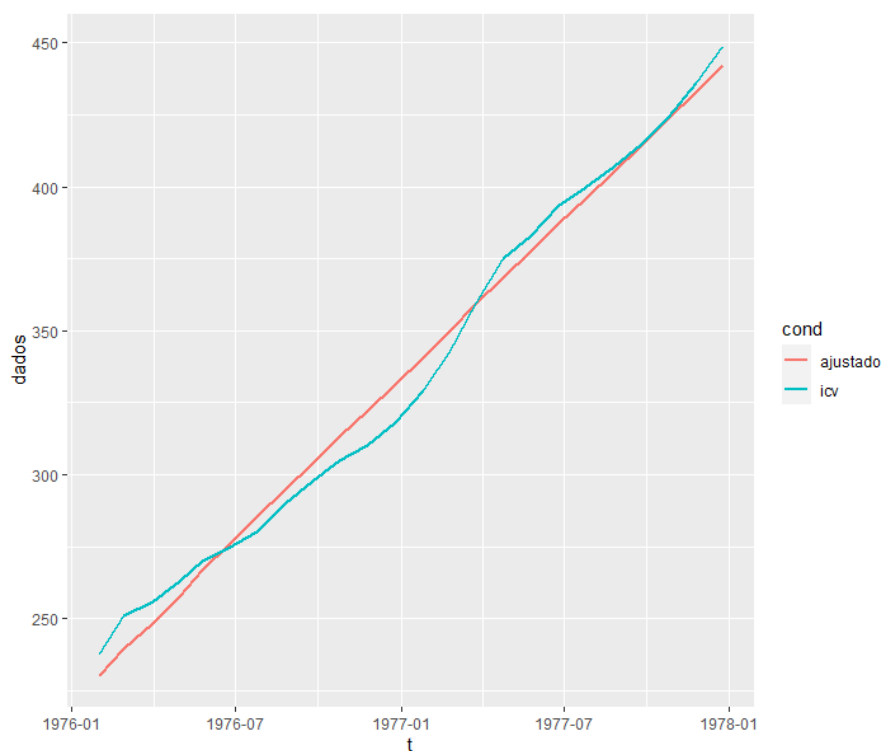


Figura 3.7: Índice de Custo de Vida no Município de São Paulo - Ajuste Tendência Polinomial.
Fonte: Fundação Instituto de Pesquisas Economicas (FIPE) - Ajuste Manual.

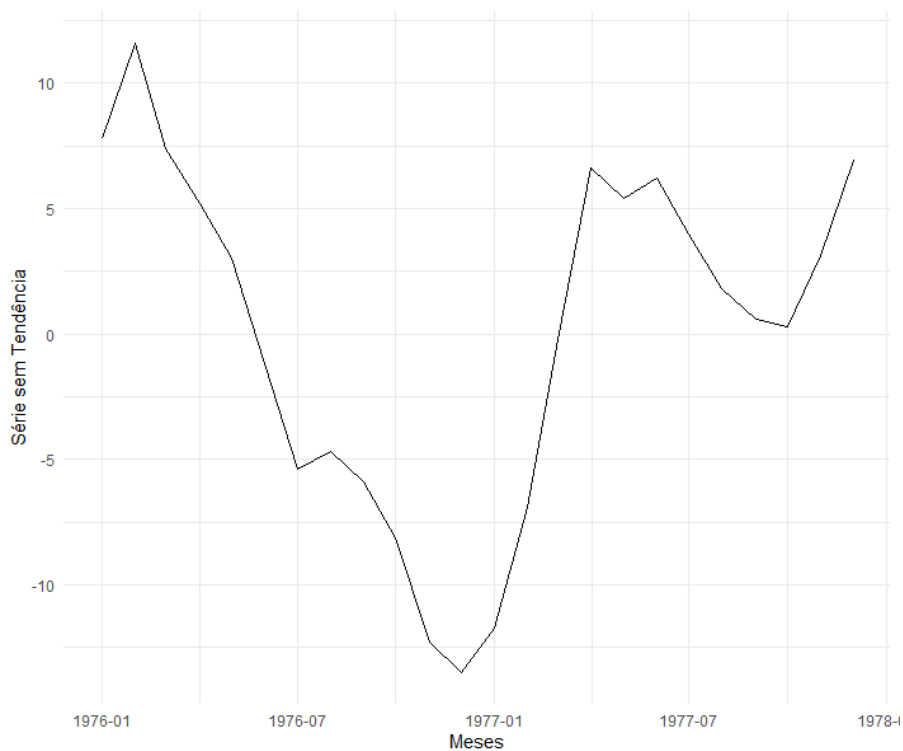


Figura 3.8: Índice de Custo de Vida no Município de São Paulo sem a Tendência Polinomial.

com $\hat{\beta}_0 = 220,9457$ e $\hat{\beta}_1 = 9,2143$, para $t = 1, \dots, n$, onde n é o tamanho da série temporal.

A Figura 3.9 apresenta o gráfico da FAC Amostral dos resíduos do Índice de Custo de Vida no Município de São Paulo após a retirada da tendência $icv = \beta_0 + t\beta_1$, com $\hat{\beta}_0 = 220,9457$ e $\hat{\beta}_1 = 9,2143$, para $t = 1, \dots, n$, onde n é o tamanho da série temporal.

```
1 ggAcf(d.st, lag.max=20)+labs(y = "FAC Série sem Tendência", title="")+theme_minimal()
```

Código 3.9: Código fonte em R: FAC Amostral da Série do Índice de Custo de Vida no Município de São Paulo sem a tendência linear.

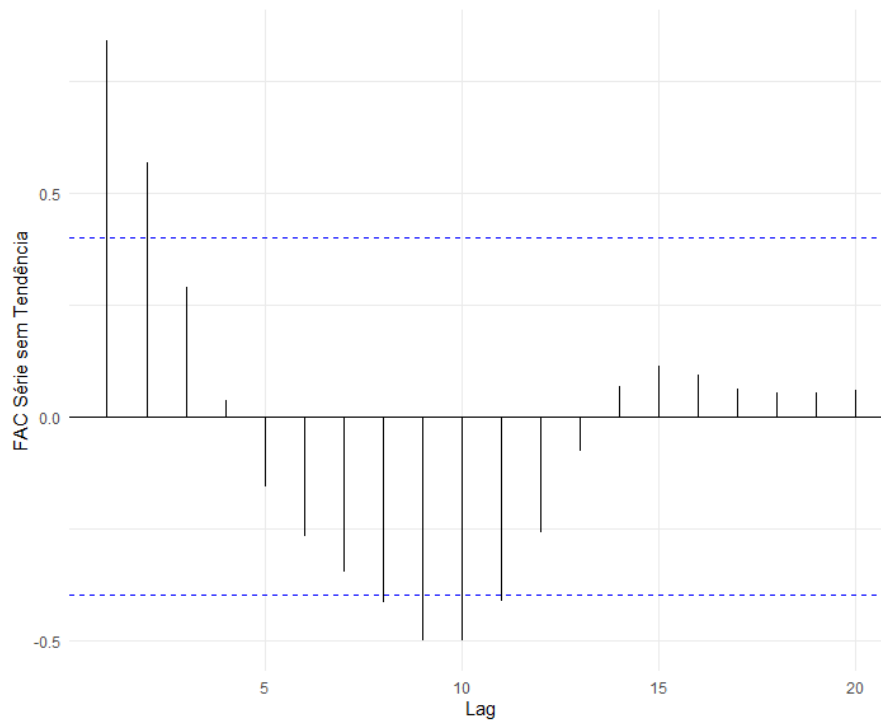


Figura 3.9: FAC Amostral da Série do Índice de Custo de Vida no Município de São Paulo sem a tendência linear.

Agora iremos utilizar a *Least absolute deviations regression* do pacote *L1pack*, ou seja, regressão por desvios absolutos. Neste caso, trata-se uma regressão robusta. A *Least absolute deviations regression* minimiza a função

$$\sum_{t=1}^n |X_t - T_t| = \sum_{t=1}^n |X_t - \beta_0 - \beta_1 t - \dots - \beta_m t^m| \quad (3.6)$$

onde $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_m)$ são as estimativas de $\beta = (\beta_0, \beta_1, \dots, \beta_m)$ obtidos pela *Least absolute deviations regression*.

```
1 icv <- read_excel("icv.xlsx", col_types = c("date", "numeric"))
2 n<-length(icv$icv)
3 t<-1:n
4 df.icv<-as.data.frame(cbind(t, icv))
5 regr<-lad(icv~t, data=df.icv)
6 summary(regr)
7
8 Saída summary(regr)
9
```

```

10 Call:
11 lad(formula = icv ~ t, data = df.icv)
12
13 Residuals:
14     Min       1Q   Median       3Q      Max
15 -13.571  -4.643   0.000   4.536  13.714
16
17 Coefficients:
18             Estimate Std. Error Z value  p-value
19 (Intercept)  218.4286    4.8205   45.3129   0.0000
20 t             9.4286    0.3374   27.9480   0.0000
21
22 Degrees of freedom: 24 total; 22 residual
23 Scale estimate: 8.089638
24 Log-likelihood: -82.49178 on 3 degrees of freedom
25
26 yr<-regr$coefficients[1]+regr$coefficients[2]*t
27 yr<-round(yr,1)
28 df1r<-as.data.frame(cbind(rep("icv",n),icv))
29 df2r<-as.data.frame(cbind(rep("ajustado",n),icv[-2], yr))
30 colnames(df1r)<-c("cond","t","dados")
31 colnames(df2r)<-c("cond","t","dados")
32 df.icvr<-rbind(df1r, df2r)
33 ggplot(df.icvr, aes(x=t,y=dados,group=cond,colour=cond)) +
34   geom_line(size = 1)

```

Código 3.10: Código fonte em R: Índice de Custo de Vida no Município de São Paulo. Fonte: Fundação instituto de Pesquisas Economicas (FIPE) - Ajuste Tendência por regressão robusta

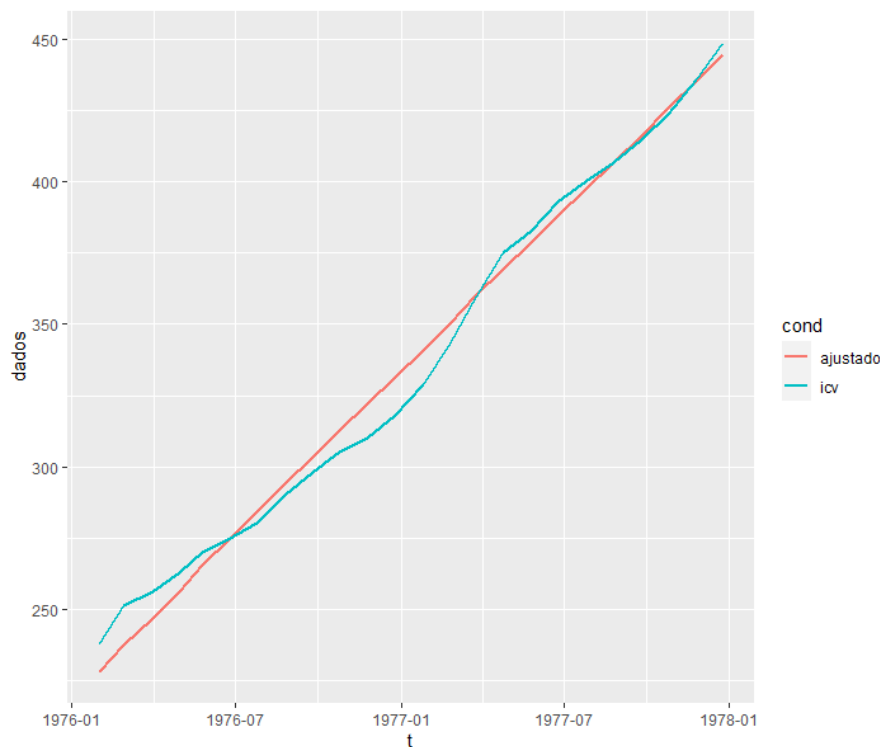


Figura 3.10: Série do Índice de Custo de Vida no Município de São Paulo com ajuste de tendência linear por meio de regressão robusta.

A Figura 3.10 apresenta o gráfico da série temporal do Índice de Custo de Vida no Município de

São Paulo com a reta de tendência robusta $icv = \beta_0 + t\beta_1$, com $\hat{\beta}_0 = 218,4286$ e $\hat{\beta}_1 = 9,4286$, para $t = 1, \dots, n$, onde n é o tamanho da série temporal. Neste caso, denotamos os resíduos por $\{\hat{\varepsilon}_t\}_{t=1}^n$, onde

$$\hat{\varepsilon}_t = X_t - \hat{T}_t = X_t - [\hat{\beta}_0 + t\hat{\beta}_1],$$

com $\hat{\beta}$ e $\hat{\beta}_1$ dados acima.

```
1 d.str<-icv[,2]-yr #série sem a tendência
2 d.str<-ts(d.str, start=c(1976,1), frequency=12)
3 autoplot(d.str)+labs(x="Meses", y="Série sem Tendência")+theme_minimal()
```

Código 3.11: Código fonte em R: Série Temporal do Índice de Custo de Vida no Município de São Paulo sem a tendência linear robusta.

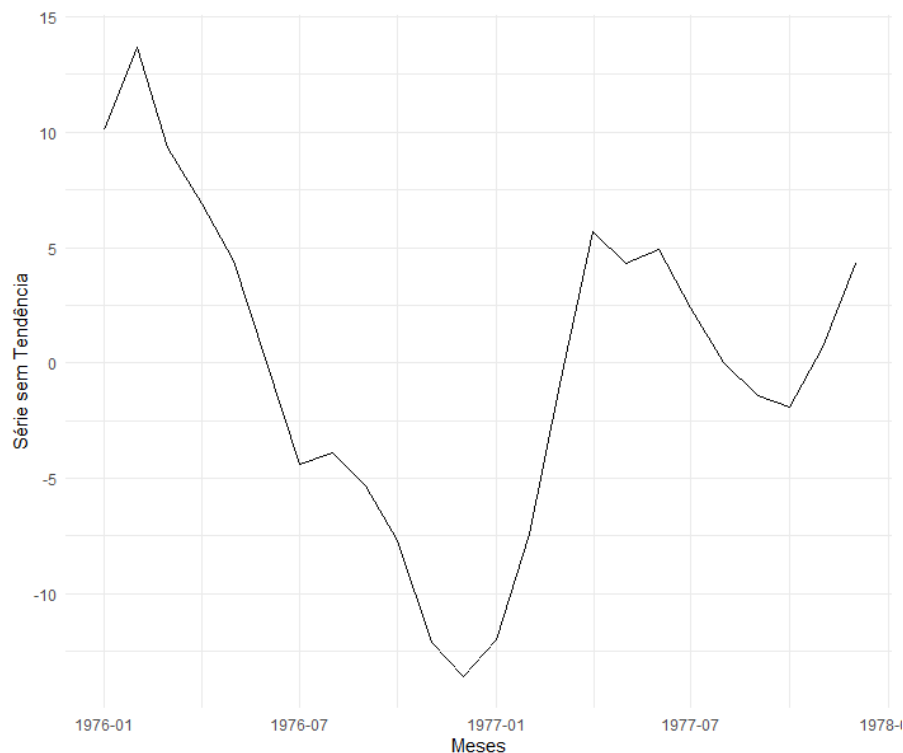


Figura 3.11: Série do Índice de Custo de Vida no Município de São Paulo sem tendência linear ajustada por meio de regressão robusta.

Exemplo 3.1. Análise da série temporal bimestral do Lucro por Ação da Johnson & Johnson do pacote *astsa*.

```
1 #jj (Johnson & Johnson)
2 t<-1:length(jj)
3 df.jj<-as.data.frame(cbind(t1=t, jj=jj))
4 autoplot(jj)+
5   labs(y="Lucro por Ação - Johnson e Johnson", x="Tempo")+theme_minimal()
6 ggplot(df.jj, aes(x=t, y=jj)) +
7   geom_line() +
8   geom_smooth(method = lm, formula=jj~x+I(x^2))+
9   labs(y="Lucro por Ação - Johnson e Johnson", x="Tempo")+theme_minimal()
```

Código 3.12: Código fonte em R: Lucro por Ação - Johnson e Johnson. Fonte: Pacote *astsa* R Core Team (2020).

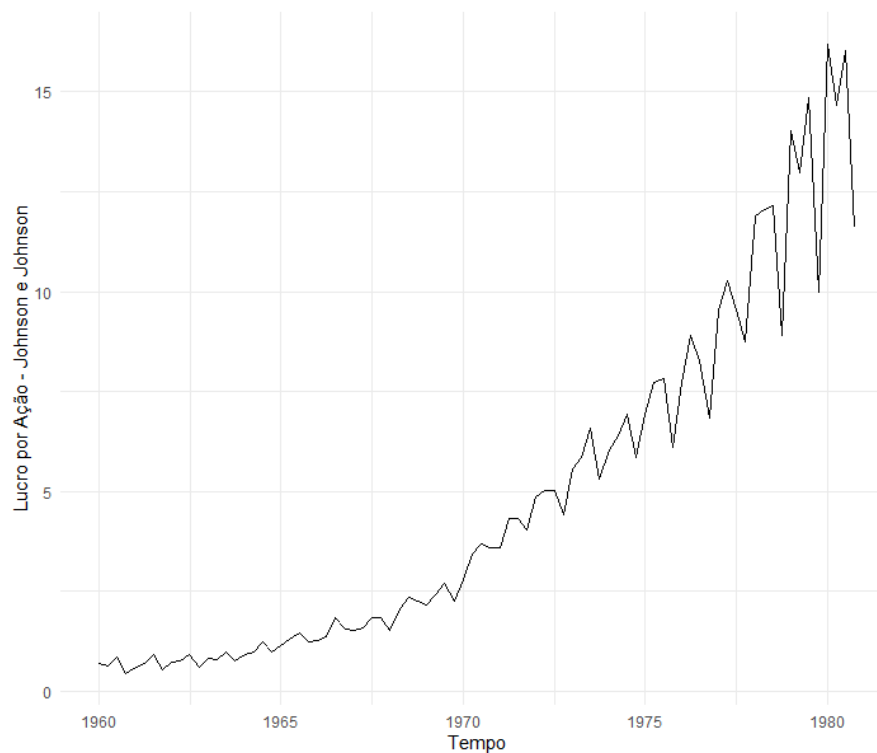


Figura 3.12: Lucro por Ação - Johnson e Johnson. Fonte: Pacote astsa R Core Team (2020).

No caso da série temporal do Lucro por Ação - Johnson & Johnson estamos ajustando uma tendência quadrática, ou seja,

$$T_t = \beta_0 + \beta_1 t + \beta_2 t^2.$$

```

1 #jj (Johnson & Johnson)
2 n<-length(jj)
3 t<-1:n
4 df.jj<-as.data.frame(cbind(t1=t, t2=t^2, jj=jj))
5 reg<-lm(jj~t1+t2, data=df.jj)
6 summary(reg)
7
8 Saída summary(reg)
9
10 Call:
11 lm(formula = jj ~ t1 + t2, data = df.jj)
12
13 Residuals:
14      Min       1Q   Median       3Q      Max
15 -3.07418 -0.21177  0.08893  0.31063  2.77309
16
17 Coefficients:
18             Estimate Std. Error t value Pr(>|t|)
19 (Intercept)  1.0712785   0.3086489   3.471 0.000835 ***
20 t1          -0.0602821   0.0167594  -3.597 0.000552 ***
21 t2           0.0026274   0.0001911  13.752 < 2e-16 ***
22
23 Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1 1
24
25 Residual standard error: 0.9206 on 81 degrees of freedom
26 Multiple R-squared:  0.9555, Adjusted R-squared:  0.9544
27 F-statistic: 869.2 on 2 and 81 DF, p-value: < 2.2e-16

```

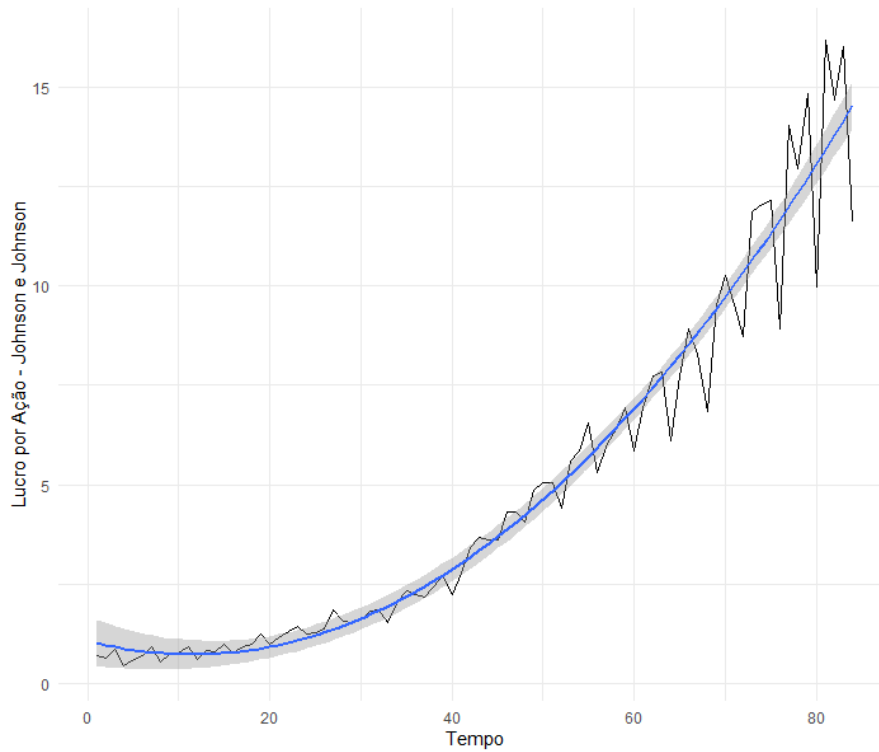


Figura 3.13: Lucro por Ação - Johnson e Johnson. Ajuste realizado utilizando o comando `geom_smooth`.

```

28
29 y<-reg$coefficients[1]+reg$coefficients[2]*t+reg$coefficients[3]*t^2
30 y<-round(y,1)
31 y<-ts(y, start=c(1960,1), frequency = 4)
32 plot.ts(cbind(jj,y), plot.type = c("single"), col=c("black","blue"),
33         main = "Johnson e Johnson", xlab="Tempo", ylab="")
34 d.jj<-jj-y
35 autoplot(d.jj)+
36   labs(y = "Lucro por ação - Johnson & Johnson sem Tendência", x = "Tempo")+theme_
   minimal()

```

Código 3.13: Código fonte em R: Lucro por Ação - Johnson e Johnson. Fonte: Pacote `astsa` R Core Team (2020).

A Figura 3.14 apresenta a série temporal dos Lucro por ação - Johnson & Johnson sem Tendência polinomial, onde $T_t = \beta_0 + \beta_1 t + \beta_2 t^2$, $\hat{\beta}_0 = 1,0712785$, $\hat{\beta}_1 = -0,0602821$ e $\hat{\beta}_2 = 0,0026274$, onde os resíduos são calculados da seguinte forma

$$\hat{\varepsilon}_t = X_t - [T_t = \hat{\beta}_0 + \hat{\beta}_1 t + \hat{\beta}_2 t^2].$$

```

1 ggAcf(d.jj, lag.max=20)+labs(y = "FAC Série sem Tendência", title="")+theme_minimal()

```

Código 3.14: Código fonte em R: FAC Amostral da série temporal do Lucro por Ação - Johnson e Johnson Sem Tendência Quadrática.

A Figura 3.15 apresenta o gráfico da FAC Amostral dos resíduos do Lucro por Ação - Johnson & Johnson após a retirada da tendência $T_t = \beta_0 + t\beta_1 + t^2\beta_2$, com $\hat{\beta}_0 = 1,0712785$, $\hat{\beta}_1 = -0,0602821$ e $\hat{\beta}_2 = 0,0026274$, para $t = 1, \dots, n$, onde n é o tamanho da série temporal.

□

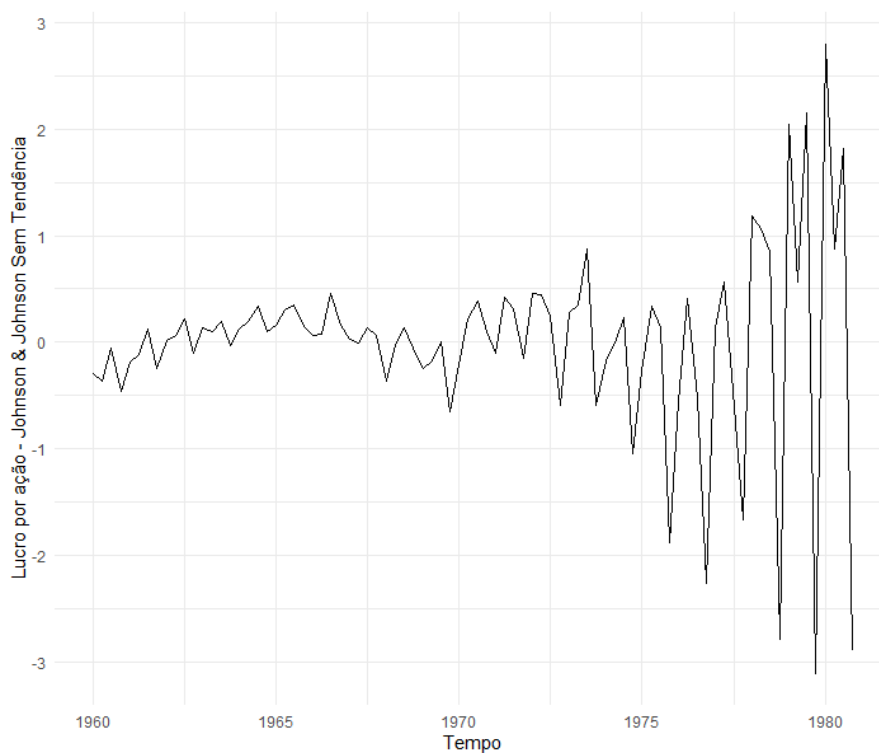


Figura 3.14: Lucro por ação - Johnson & Johnson sem Tendência

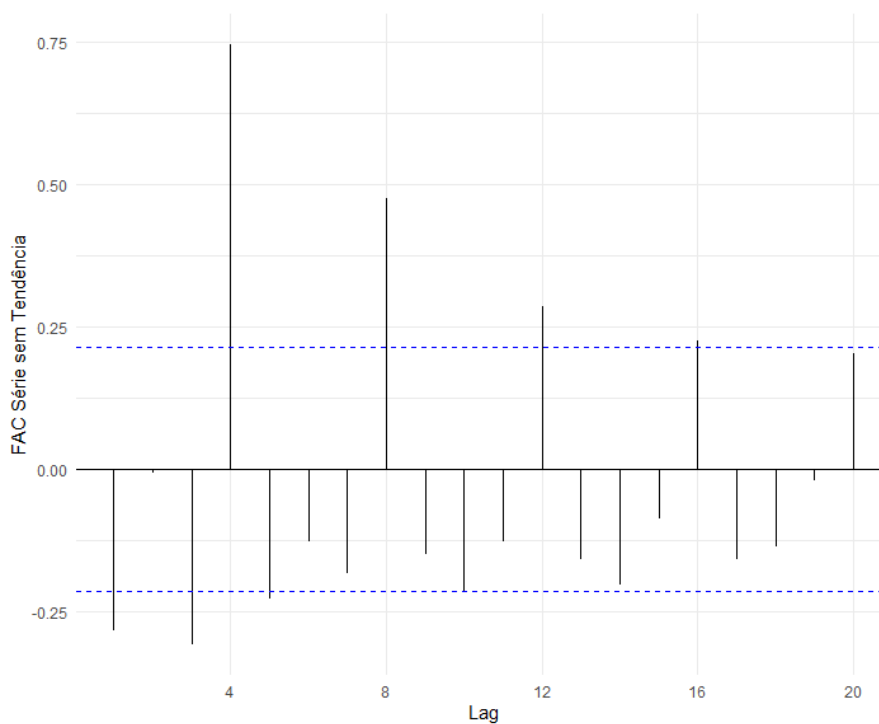


Figura 3.15: FAC Amostral dos resíduos do Lucro por Ação - Johnson & Johnson após a retirada da tendência quadrática.

3.1.3 Tendência Não Linear

A Figura 3.16 apresenta o gráfico de uma série temporal simulada através da função

$$y_t = a(1+t)^b + \varepsilon_t, \quad t = 1, \dots, n,$$

a e b constantes, $\{\varepsilon_t\}_{t \in \mathbb{Z}}$ Ruído Branco Normal- $RB(0, \sigma_\varepsilon^2)$, $\sigma_\varepsilon^2 = 4$. Também podemos observar a função ajustada aos dados.

```

1 #
2 #Série Temporal Simulada
3 #
4 a<-2
5 b<-0.3
6 n<-100
7 t<-1:n
8 e<-rnorm(n, sd=2)
9 y<-a*(1+t)^b+e
10 #
11 y<-ts(y)
12 t<-1:length(y)
13 df<-as.data.frame(cbind(t1=t, y=y))
14 autoplot(y)+
15   labs(y = "Série Temporal Simulada", x = "Tempo")+theme_minimal()
16 ggplot(df, aes(x=t, y=y)) +
17   geom_line() +
18   stat_smooth(method='nls', formula='y~a*(1+x)^b', method.args=list(start =
19     list(a=1, b=1)), se=FALSE)+
20   labs(y = "Série Temporal Simulada - Ajuste Função Não Linear", x = "Tempo")+theme_
    minimal()

```

Código 3.15: Código fonte em R: Série Temporal Simulada - Ajuste Função Não Linear utilizando a função `stat_smooth`.

A Figura 3.17 apresenta o gráfico de uma série temporal simulada com a retirada da tendência através da função

$$y_t = a(1+t)^b + \varepsilon_t, \quad t = 1, \dots, n,$$

$\hat{a} = 2,74670$ e $\hat{b} = 0,23072$.

```

1 t<-1:length(y)
2 mod<-nls(y~a*(1+t)^b, start=list(a=1,b=1))
3 summary(mod)
4 z<-coefficients(mod)[[1]]*(1+t)^coefficients(mod)[[2]]
5 d.y<-y-z
6 autoplot(d.y)+
7   labs(y = "Série Temporal Simulada - Sem Tendência", x = "Tempo")+theme_minimal()

```

Código 3.16: Código fonte em R: Resíduo da Série Temporal Simulada - Ajuste manual Função Não Linear.

A Figura 3.19 apresenta o gráfico da FAC Amostral dos resíduos da série temporal simulada após a retirada da tendência $T_t = a(1+t)^b$, com $\hat{a} = 2,7467$ e $\hat{b} = 0,2307189$, para $t = 1, \dots, n$, onde n é o tamanho da série temporal.

```

1 ggAcf(d.y, lag.max=20)+labs(y = "FAC Série sem Tendência", title="")+theme_minimal()

```

Código 3.17: Código fonte em R: FAC Amostral dos Resíduos da Série Temporal Simulada - Ajuste manual Função Não Linear.

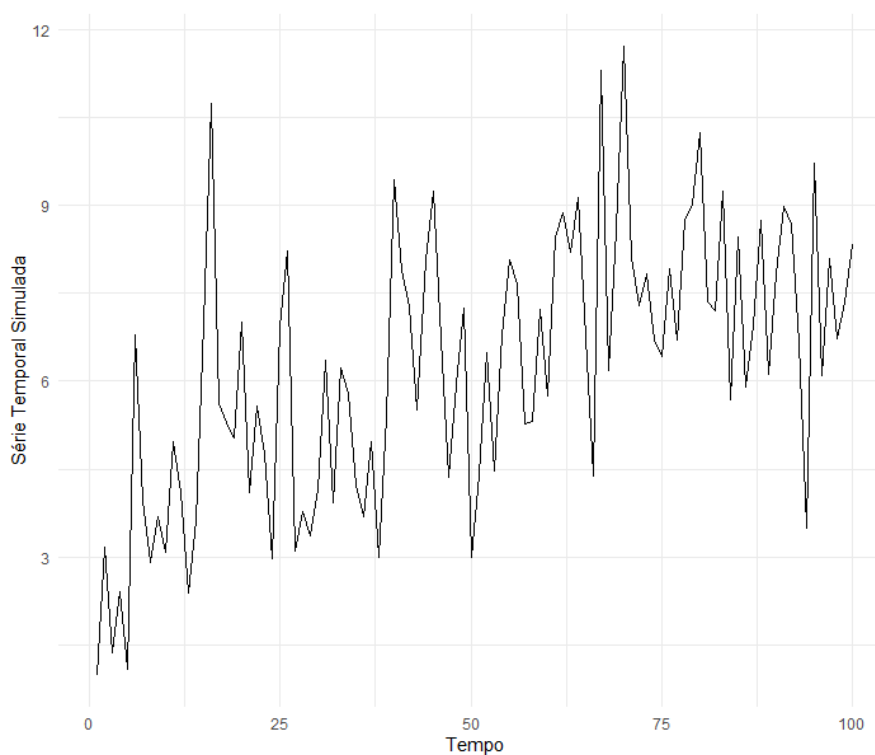


Figura 3.16: Série Temporal Simulada - Função Não Linear.

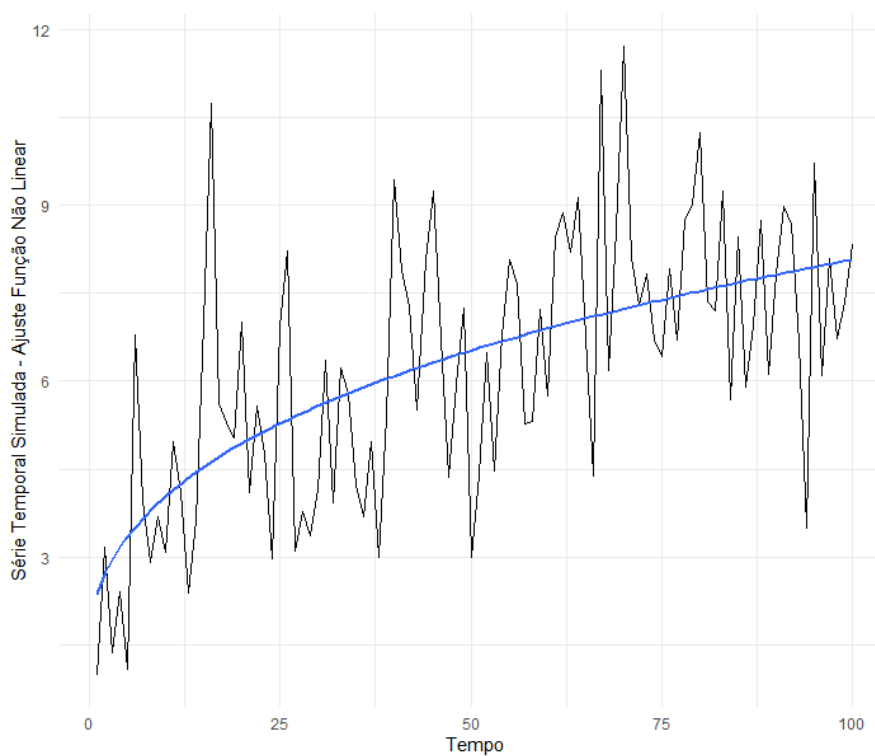


Figura 3.17: Ajuste não linear Série Temporal Simulada - Função Não Linear utilizando a função `stat_smooth`.

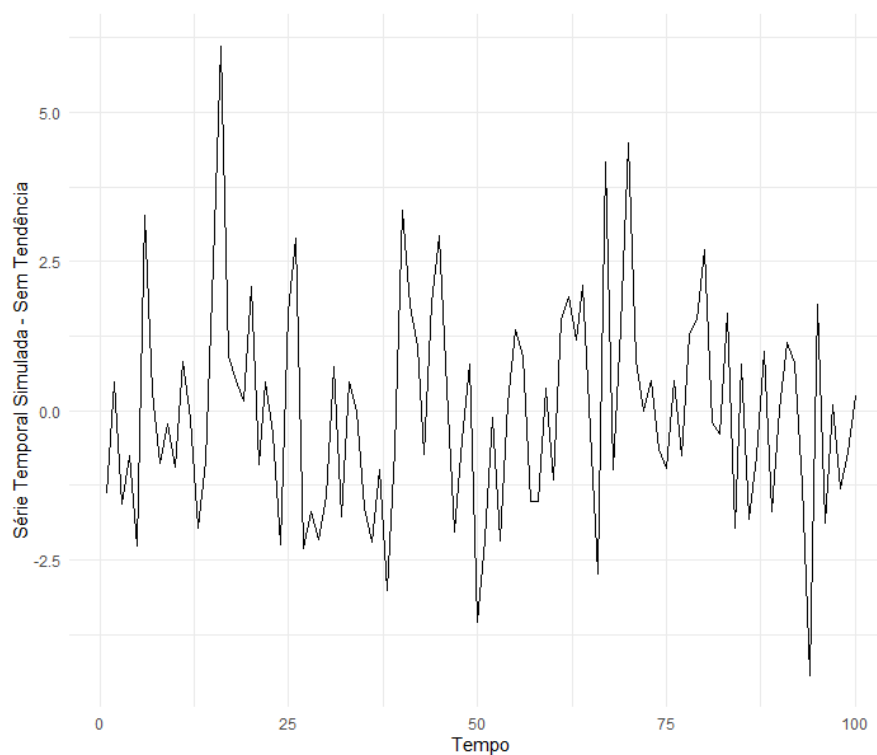


Figura 3.18: Resíduo da Série Temporal Simulada - Ajuste manual Função Não Linear.

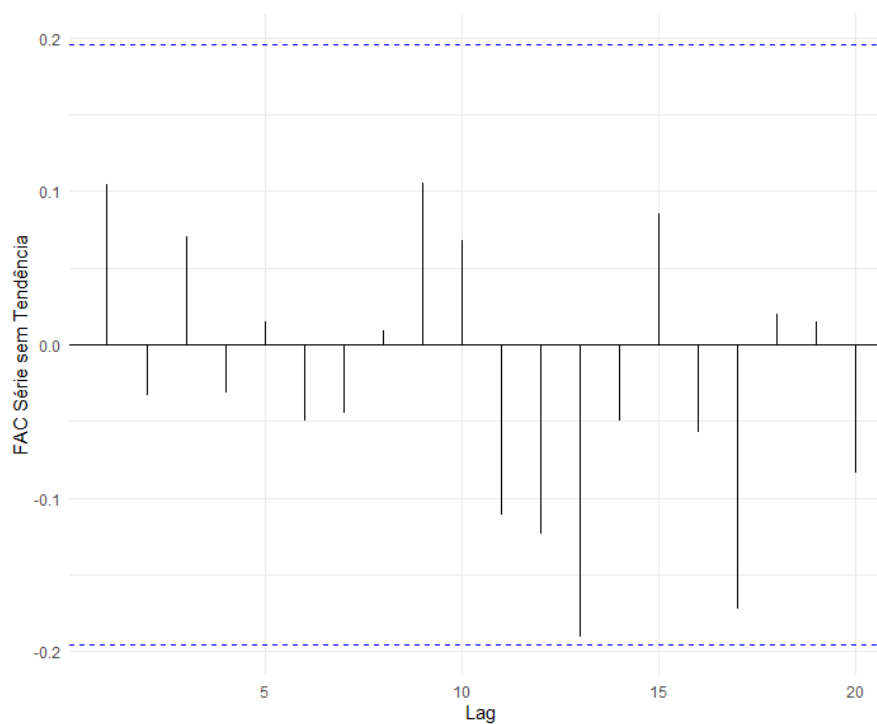


Figura 3.19: FAC Amostral dos Resíduos da Série Temporal Simulada - Ajuste manual Função Não Linear.

Outro exemplo de ajuste de tendência não linear é para os dados a seguir. A Figura 3.20 apresenta os dados e a função

$$T_t = 1 + \theta \left(1 - \exp \left(-\frac{t}{\tau} \right) \right), \quad (3.7)$$

ajustada utilizando a função `geom_smooth`, onde t não são os naturais. Os dados a seguir foram coletados no link <https://bit.ly/39eU1BF>.

```
1 dataset <- data.frame(Exp = c(4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4,
2                               5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5,
3                               6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6),
4                               t = c(0, 0.33, 0.67, 1, 1.33, 1.67, 2, 4,
5                                   6, 8, 10, 0, 33, 0.67, 1, 1.33, 1.67,
6                                   2, 4, 6, 8, 10, 0, 0.33, 0.67, 1, 1.33,
7                                   1.67, 2, 4, 6, 8, 10), fold =
8                                   c(1, 0.957066345654286, 1.24139015724819,
9                                   1.62889151698633, 1.72008539595879,
10                                  1.82725412314402, 1.93164365299958,
11                                  1.9722929538061, 2.15842019312484,
12                                  1.9200507796933, 1.95804730344453, 1,
13                                  0.836176542548747, 1.07077717914707,
14                                  1.45471712491441, 1.61069357875771,
15                                  1.75576377806756, 1.89280913889538,
16                                  2.00219054189937, 1.87795513639311,
17                                  1.85242493827193, 1.7409346372629, 1,
18                                  0.840498729335292, 0.904130905000499,
19                                  1.23116185602517, 1.41897551928886,
20                                  1.60167656534099, 1.72389226836308,
21                                  1.80635095956481, 1.76640786872057,
22                                  1.74327897001172, 1.63581509884482))
23 #
24 df<-subset(dataset, Exp==4)#Subconjunto
25 ggplot(df, aes(x=t, y=fold))+
26   #to make it obvious I use argument names instead of positional matching
27   geom_line()+geom_point()+theme_minimal()
28 ggplot(df, aes(x=t, y=fold))+
29   geom_point()+
30   geom_smooth(method="nls",
31               formula=y~1+theta*(1-exp(-x/tau)),
32               start=c(tau=0.2, theta=2),
33               se=FALSE) +theme_minimal()
```

Código 3.18: Código fonte em R: Série Temporal - Ajuste Função Não Linear

A Figura 3.21 apresenta o gráfico dos dados coletados no link <https://bit.ly/39eU1BF> com o ajuste da função dada pela equação (3.7).

```
1 mod<-nls(fold~1+theta*(1-exp(-t/tau)), start=list(theta=2,tau=0.2), data=df)
2 summary(mod)
3
4 Saída summary(mod)
5
6 Formula: y ~ a * (1 + t)^b
7
8 Parameters:
9   Estimate Std. Error t value Pr(>|t|)
10 a  1.90172    0.36166   5.258 8.53e-07 ***
11 b  0.31338    0.04691   6.681 1.45e-09 ***
12 ---
13 Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1 1
14
15 Residual standard error: 1.795 on 98 degrees of freedom
```

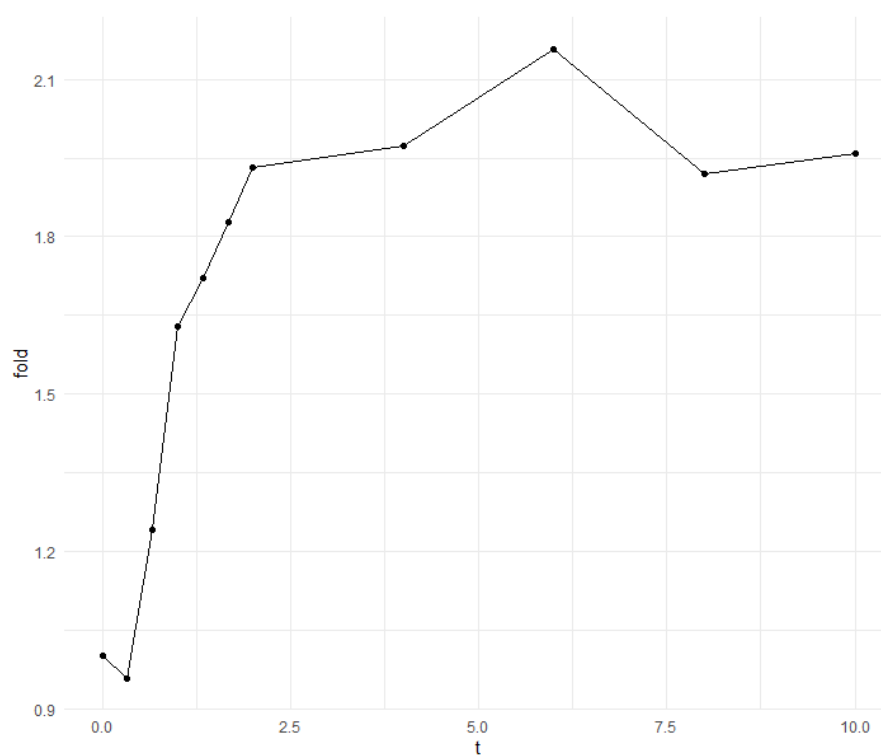



Figura 3.20: Gráfico dos dados coletados. Fonte: <https://bit.ly/39eU1BF>

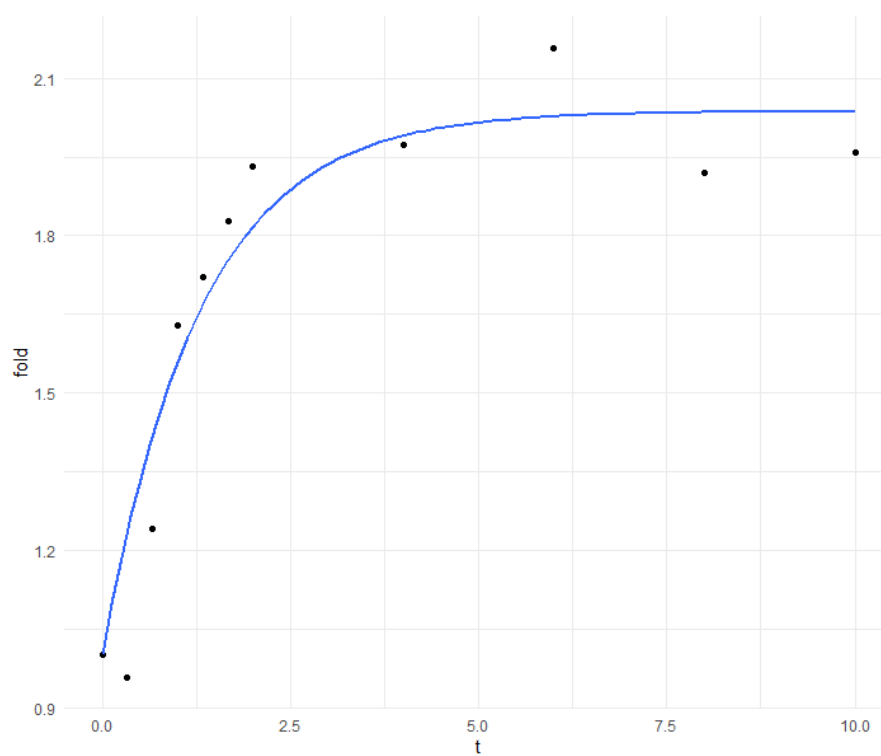


Figura 3.21: Gráfico dos dados coletados - Ajuste utilizando `geom_smooth`. Fonte: <https://bit.ly/39eU1BF>.

```

17 Number of iterations to convergence: 7
18 Achieved convergence tolerance: 4.535e-06
19
20 z<-1+coefficients(mod)[[1]]*(1-exp(-df$t/coefficients(mod)[[2]]))
21 d.y<-data.frame(t=df$t, data=df$fold-z)
22 ggplot(d.y,aes(x=t, y=data))+
23   geom_line()+
24   labs(y = "Série Temporal - Sem Tendência", x = "Tempo")+theme_minimal()

```

Código 3.19: Código fonte em R: Série Temporal - Ajuste Função Não Linear

A Figura 3.22 apresenta o gráfico dos dados coletados no link <https://bit.ly/39eU1BF> após retirada da tendência dada pela equação (3.7), com $\hat{\theta} = 1,03724$ e $\hat{\tau} = 1,29063$, dadas por

$$\hat{\varepsilon}_t = X_t - \hat{T}_t = X_t - \left[1 + \hat{\theta} \left(1 - \exp \left(-\frac{t}{\hat{\tau}} \right) \right) \right]$$

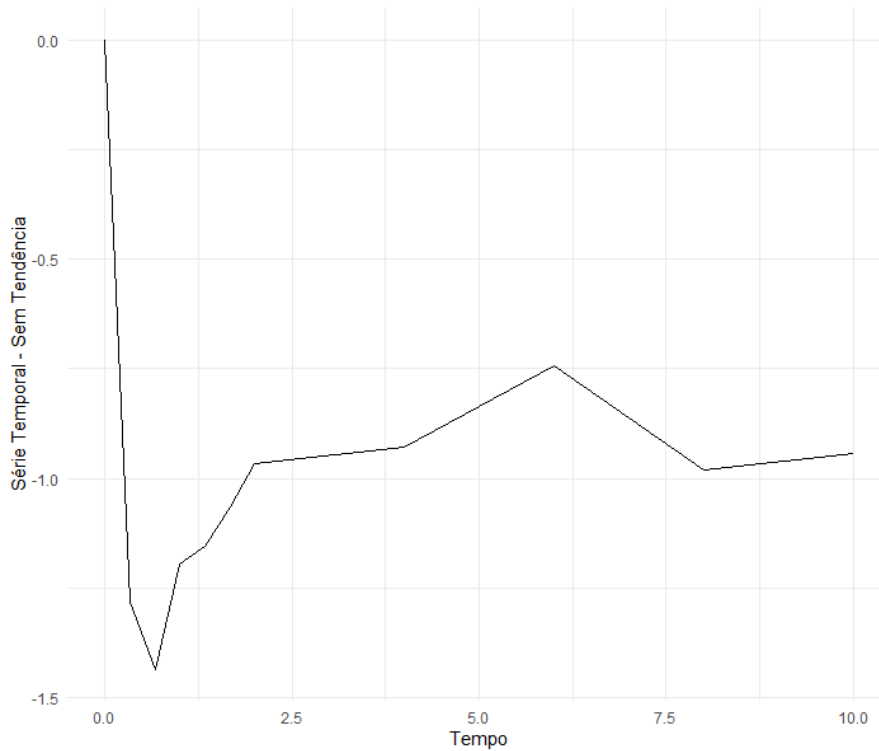


Figura 3.22: Gráfico dos dados coletados - sem tendência não linear.

A Figura 3.23 apresenta a análise dos resíduos da série temporal após a retirada da tendência $T_t = 1 + \theta \left(1 - \exp \left(-\frac{t}{\tau} \right) \right)$, com $\hat{\theta} = 1,03724$ e $\hat{\tau} = 1,29063$.

3.2 Testes de Tendência Determinística

3.2.1 Teste de Cox-Stuart

Considere um conjunto de observações $\{X_t\}_{t=1}^n$. Agrupamos as observações em pares (X_1, X_{1+c}) , (X_2, X_{2+c}) , \dots , (X_{n-c}, X_n) , onde $c = \frac{n}{2}$ se n for par e $c = \frac{n+1}{2}$ se n for ímpar. A cada par de observações associamos o sinal - se $X_i < X_{i+c}$ e o sinal + se $X_i > X_{i+c}$. Caso $X_i = X_{i+c}$ eliminamos esta observação. Considere N o número de pares onde $X_i \neq X_{i+c}$. Assim, queremos testar:

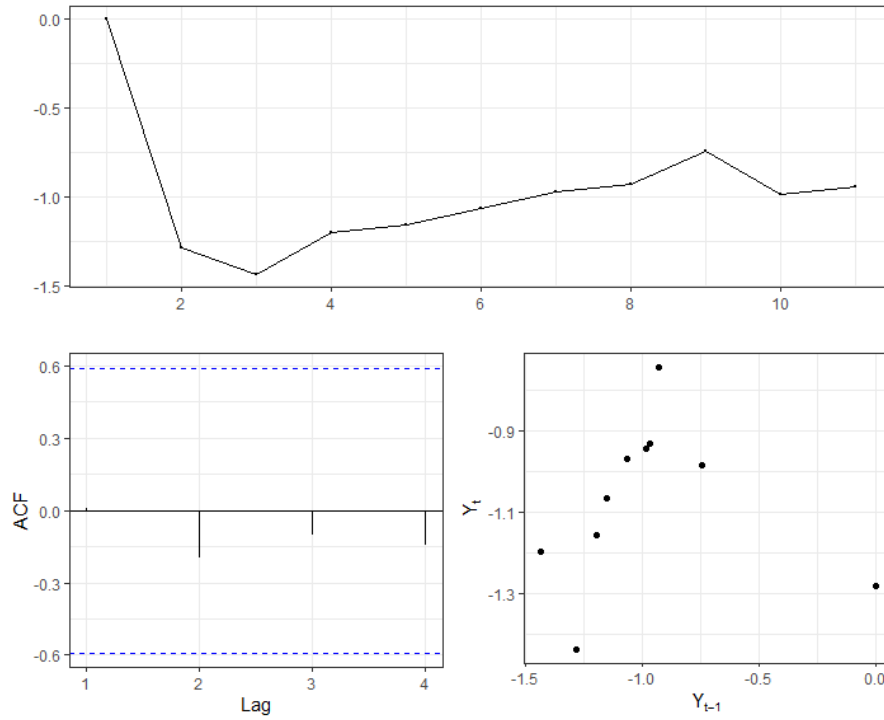


Figura 3.23: Análise dos resíduos da série temporal após a retirada da tendência.

$$\begin{cases} H_0 : \mathbb{P}(X_i < X_{i+c}) = \mathbb{P}(X_i > X_{i+c}), \text{ para todo } i \text{ ou seja, não existe tendência;} \\ H_1 : \mathbb{P}(X_i < X_{i+c}) \neq \mathbb{P}(X_i > X_{i+c}), \text{ para todo } i \text{ ou seja, existe tendência;} \end{cases}$$

Note que trata-se de um teste bilateral, onde a estatística do teste é dada por

Cálculo do p-valor

(i) Na rotina *cox.stuart.test* do pacote *randtests*, a estatística do teste é dada por

$$T = \max\{\text{Número de pares com sinal + ou -}\},$$

onde $T \sim B(k, p)$, com $p = 0,5$ e $k = N$. Como o teste de hipóteses é bilateral, temos que

$$\begin{aligned} \text{p-valor} &= 2 \times \mathbb{P}(T \geq t_{calc}) \\ &= 2 \times (\min(pbinom(q = (t_{calc} - 1), size = N, prob = 0.5, lower.tail = F), \\ &\quad pbinom(q = t_{calc}, size = N, prob = 0.5))) \end{aligned}$$

(ii) Na rotina *cs.test* do pacote *trend*, a estatística do teste é dada por

$$Z = \frac{\frac{|T-n|}{6}}{\sqrt{\frac{n}{12}}},$$

onde

$$T = \max\{\text{Número de pares com sinal + ou -}\},$$

respectivamente. A estatística $Z \sim N(0, 1)$. Neste caso, temos que

$$\text{p-valor} = 2 \times \mathbb{P}(Z \geq z_{calc}) = 2 \times pnorm(z_{calc}, lower.tail = F)$$

Para realizarmos o teste de Cox-Stuart utilizamos a rotina `cox.stuart.test` dos pacotes `randtests` ou `webr`, ou a rotina `cs.test` do pacote `trend`, ambas do software R Core Team (2020). Maiores detalhes sobre a rotina ver <https://bit.ly/3l1xsUj>. Maiores detalhes sobre o teste de Cox-Stuart ver Cox e Stuart (1955). Para o pacote `trend` ver <https://cran.r-project.org/web/packages/trend/trend.pdf>

3.2.2 Teste de Wald-Wolfowitz

Para podermos aplicar este teste, precisamos inicialmente que a série temporal seja independente e identicamente distribuída.

Seja $\{X_t\}_{t=1}^n$ uma série temporal com n observações. Considere M sendo a mediana das n observações da série temporal. Atribuímos a cada valor de X_t o simbolo “a” se ele for maior ou igual a M , e “b” caso ele for menor que M . Então, temos que $n = (n_a \text{ pontos “a”}) + (n_b \text{ pontos “b”})$. Desta forma, teremos grupos de observações marcadas com “a” e grupos de observações marcadas com “b” ao longo da série temporal. O número total de grupos será a estatística do teste, isto é,

$$T = \text{número total de grupos com símbolos iguais.}$$

Consideramos as seguintes hipóteses:

$$H_0 = \text{Não Existe tendência} \quad \text{vs} \quad H_1 = \text{Existe tendência}$$

Rejeitamos a hipótese nula H_0 se tivermos um número pequeno de grupos com símbolos iguais, ou seja, se T for relativamente pequeno.

Cálculo do p-valor

(i) Na rotina `runs.test` dos pacotes `randtests` ou `tseries`, a estatística do teste é dada por

$$T = \text{número total de grupos com símbolos iguais.}$$

onde $T \sim N(\mu, \sigma^2)$, com

$$\mu = \frac{2n_a n_b}{n} + 1, \quad \sigma^2 = \frac{2n_a n_b (2n_a n_b - n)}{n^2 (n - 1)}.$$

Padronizando, temos

$$Z = \frac{T - \mu}{\sigma} \sim N(0, 1)$$

Como o teste de hipóteses é bilateral, temos que

$$\text{p-valor} = 2 \times \mathbb{P}(Z \geq z_{\text{calc}}) = 2 \times \text{pnorm}(z_{\text{calc}}, \text{lower.tail} = F)$$

(ii) Na rotina `ww.test` do pacote `trend`, a estatística do teste é dada por

$$T = \sum_{j=1}^{n-1} X_j X_{j+1} + X_1 X_n.$$

Definindo $\mu_k = \sum_{j=1}^n X_j^k$, para $k = 1, 2, 3, 4$, temos que

$$\mu = \mathbb{E}(T) = \frac{\mu_1^2 - \mu_2}{n - 1}$$

$$\sigma^2 = \text{Var}(T) = \frac{\mu_2^2 - \mu_4}{n-1} - \left[\frac{\mu_1^2 - \mu_2}{n-1} \right]^2 + \frac{\mu_1^4 - 4\mu_1^2\mu_2 + 4\mu_1\mu_3 + \mu_2^2 - 2\mu_4}{(n-1)(n-2)}.$$

Padronizando, temos

$$Z = \frac{T - \mu}{\sigma} \sim N(0, 1)$$

Como o teste de hipóteses é bilateral, temos que

$$\text{p-valor} = 2 \times \mathbb{P}(Z \geq z_{\text{calc}}) = 2 \times \text{pnorm}(z_{\text{calc}}, \text{lower.tail} = F)$$

Para realizarmos o teste de Wald-Wolfowitz utilizamos a rotina *runs.test* dos pacotes *randtests* ou *tseries*, e a rotina *ww.test* do pacote *trend*, ambas do software R Core Team (2020). Maiores detalhes sobre a rotina *runs.test* ver <https://rdrr.io/cran/randtests/man/runs.test.html>. Sobre o Teste de Wald-Wolfowitz ver Wald e Wolfowitz (1940).

3.2.3 Teste de Mann-Kendall

A hipótese nula é que os dados vêm de uma população com realizações independentes e identicamente distribuídas. Para o teste bilateral, a hipótese alternativa é que os dados seguem uma tendência monotônica. Então, queremos testar se as observações da série são independentes e identicamente distribuídas, isto é, queremos testar as hipóteses

$$\begin{cases} H_0 : \text{A série temporal é i.i.d. (Não há tendência)} \\ H_1 : \text{As observações da série possuem tendência monotônica no tempo (Há tendência).} \end{cases}$$

Sendo assim, sob H_0 a estatística do teste é dada por:

$$S = \sum_{k=1}^{n-1} \sum_{j=k+1}^n \text{sign}(x_j - x_k)$$

onde

$$\text{sign}(x) = \begin{cases} 1, & \text{se } x > 0 \\ 0, & \text{se } x = 0 \\ -1, & \text{se } x < 0 \end{cases}$$

É possível mostrar que S é normalmente distribuída, ou seja, $S \sim N(\mu, \sigma^2)$ com

$$\mu = 0 \quad \text{e} \quad \sigma^2 = \frac{n(n-1)(2n+5) - \sum_{j=1}^P t_j(t_j-1)(2t_j+5)}{18}$$

em que n é o número de observações e, considerando o caso em que a série pode ter grupos com observações iguais, P é o número de grupos com observações iguais e t_j é o número de observações iguais no grupo j . Assim, a estatística do teste é dada por

$$Z = \frac{S}{\sigma}.$$

No caso em que o número de observações é superior a 30, utilizando correção de continuidade, a estatística do teste é calculada por

$$Z = \begin{cases} \frac{S-1}{\sigma}, & \text{se } S > 0 \\ 0, & \text{se } S = 0 \\ \frac{S+1}{\sigma}, & \text{se } S < 0 \end{cases} = \text{sign}(S) \left[\frac{|S|-1}{\sigma} \right].$$

Mesmo para um número de observações inferiores a 30, podemos utilizar a estatística Z para realizarmos o teste.

Em um teste bilateral, não rejeitamos a hipótese nula H_0 para um dado nível de significância α , se para o quantil $Z_{\alpha/2}$ de uma distribuição normal padrão temos $|Z| \leq Z_{\alpha/2}$.

Cálculo do p-valor

$$\text{p-valor} = 2 \times \mathbb{P}(Z \geq z_{calc}) = 2 \times \text{pnorm}(z_{calc}, \text{lower.tail} = F)$$

Para realizarmos o teste de Mann-Kendall utilizamos a rotina *MannKendall* do pacote *Kendall* e a rotina *mk.test* do pacote *trend*, ambas do software [R Core Team \(2020\)](https://www.r-project.org/). Maiores detalhes sobre o Teste de Mann-Kendall podem ser encontrados em [Mann \(1945\)](https://doi.org/10.1002/9781118445113.ch10) e <https://up-rs-esp.github.io/mkt/>.

Exemplo 3.2. [<https://bit.ly/3m5uoax>] Considere a série temporal referente aos índices do IBOVESPA no período de 01 de setembro de 2009 a 01 de novembro de 2009 - Figura 3.24.

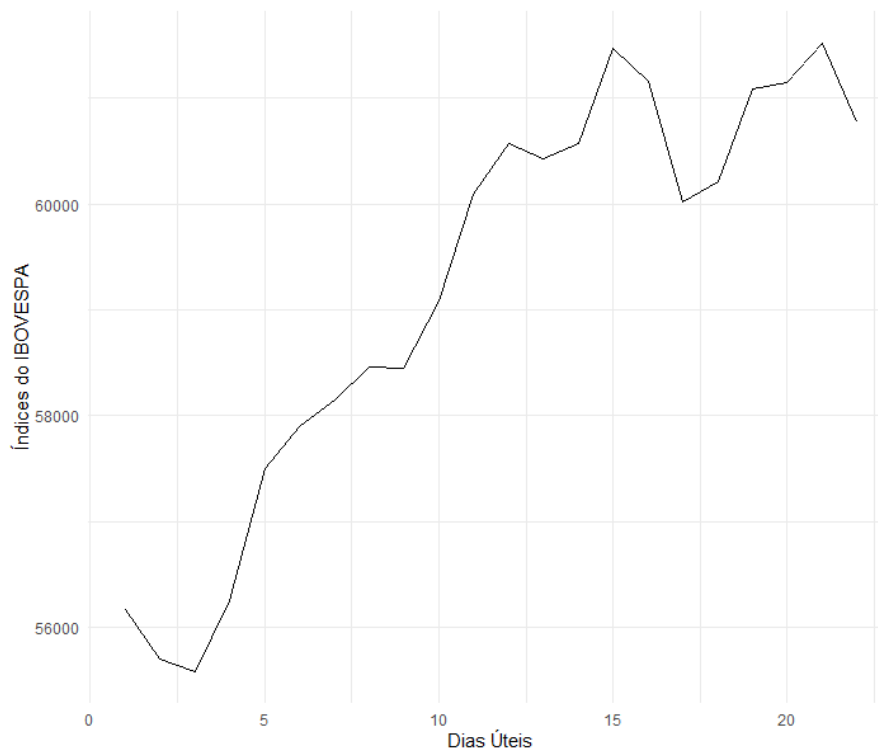


Figura 3.24: Série temporal dos índices do IBOVESPA no período de 01 de setembro de 2009 a 01 de novembro de 2009.

```
1 library(webr)
2 #Teste de Cox-Stuart
3 IBOVESPA<-c(56176,55706,55582,56258,57504,57898,58153,58460,
```

Tabela 3.1: Resultados dos testes de para tendência determinística para a série temporal dos índices do IBOVESPA no período de 01 de setembro de 2009 a 01 de novembro de 2009.

Teste	Rotina	H_0	H_1	P-valor	Conclusão
Wald-Wolfowitz	runs.test	STD	TD	0.000473	TD
	ww.test	STD	TD	0.000027	TD
Cox-Stuart	cox.stuart.test	STD	TD	0.000977	TD
	cs.test	STD	TD	0.004639	TD
Mann-Kendall	mk.test	STD	TD	5.19E-07	TD
	MannKendall	STD	TD	4.77E-07	TD

```

4      58448,59082,60088,60570,60432,60575,61467,61157,
5      60020,60207,61089,61152,61523,60772)
6 IBOVESPA<-ts(IBOVESPA)
7 autoplot(IBOVESPA)+labs(x = "Dias Úteis", y = " Índices do IBOVESPA ")+theme_minimal()
8 IBOVESPA<-c(56176,55706,55582,56258,57504,57898,58153,58460,
9      58448,59082,60088,60570,60432,60575,61467,61157,
10     60020,60207,61089,61152,61523,60772)
11 IBOVESPA<-ts(IBOVESPA)
12 autoplot(IBOVESPA)+labs(x = "Dias Úteis", y = " Índices do IBOVESPA ")+theme_minimal()
13 cox.stuart.test(IBOVESPA)
14 cs.test(IBOVESPA)
15 #-----
16 runs.test(IBOVESPA)
17 ww.test(IBOVESPA)
18 #-----
19 mk.test(IBOVESPA)
20 MannKendall(IBOVESPA)
21 #-----

```

Código 3.20: Código fonte em R: Teste de Tendência para a série temporal dos índices do IBOVESPA no período de 01 de setembro de 2009 a 01 de novembro de 2009.

A Tabela 3.1, apresenta os resultados dos testes de para tendência determinística para a série temporal dos índices do IBOVESPA no período de 01 de setembro de 2009 a 01 de novembro de 2009. Em todos os casos, $p\text{-valor} < 0,05$, assim, rejeitamos H_0 , ou seja a série temporal dos índices do IBOVESPA no período de 01 de setembro de 2009 a 01 de novembro de 2009 apresenta tendência determinística.

□

Exemplo 3.3. [<https://bit.ly/3m5uoax>] Considere a série temporal referente aos valores diários de fechamento das ações da Petrobras no período de 10 de julho de 2007 a 09 de janeiro de 2008, com um total de 123 observações. - Figura 3.25.

```

1 library (webr)
2 #Teste de Cox-Stuart
3 petrobras<-read_excel("petrobras.xlsx", col_types=c("date", "numeric"))
4 ggplot(petrobras, aes(x=Data, y=Dados))+
5   geom_line()+
6   labs(y = "Valores diários de fechamento das ações da Petrobras", x = "Tempo")+theme_
7     minimal()
8 #-----
9 cox.stuart.test(petrobras$Dados)
10 cs.test(petrobras$Dados)
11 #-----
12 runs.test(petrobras$Dados)

```



Figura 3.25: Série temporal referente aos valores diários de fechamento das ações da Petrobras no período de 10 de julho de 2007 a 09 de janeiro de 2008.

```

12 ww.test(petrobras$Dados)
13 #
14 mk.test(petrobras$Dados)
15 MannKendall(petrobras$Dados)
16

```

Código 3.21: Código fonte em R: Teste de Tendência para a série temporal dos valores diários de fechamento das ações da Petrobras de 10 de julho de 2007 a 09 de janeiro de 2008.

A Tabela 3.2, apresenta os resultados dos testes de para tendência determinística para a série temporal dos valores diários de fechamento das ações da Petrobras de 10 de julho de 2007 a 09 de janeiro de 2008. Em todos os casos, $p\text{-valor} < 0,05$, assim, rejeitamos H_0 , ou seja a série temporal dos valores diários de fechamento das ações da Petrobras de 10 de julho de 2007 a 09 de janeiro de 2008 apresenta tendência determinística.

□

Exemplo 3.4.

```

1 #library(randtests)
2
3 # Sweet potato yield per acre, harvested in the United States, between 1868 and 1937.
4 # Data available in this package.
5 data(sweetpotato)
6 sp<-ts(sweetpotato$yield, start=c(1868), frequency = 1)
7 autoplot(sp)+labs(y = "Produção de Batata Doce por Acre", x = "Tempo")+theme_minimal()
8 #
9 cox.stuart.test(sp)
10 cs.test(sp)
11 #
12 runs.test(sp)
13 ww.test(sp)

```


Tabela 3.2: Resultados dos testes de para tendência determinística para a série temporal dos valores diários de fechamento das ações da Petrobras de 10 de julho de 2007 a 09 de janeiro de 2008.

Teste	Rotina	H_0	H_1	P-valor	Conclusão
Wald-Wolfowitz	runs.test	STD	TD	0.0000000	TD
	ww.test	STD	TD	0.0000000	TD
Cox-Stuart	cox.stuart.test	STD	TD	0.0000000	TD
	cs.test	STD	TD	0.0000000	TD
Mann-Kendall	mk.test	STD	TD	0.0000000	TD
	MannKendall	STD	TD	0.0000000	TD

```

14 #
15 mk.test(sp)
16 MannKendall(sp)
17

```

Código 3.22: Código fonte em R: Teste de Tendência para a série temporal Sweet potato yield per acre, harvested in the United States, between 1868 and 1937.

A Figura 3.26 apresenta o gráfico da série temporal da produção de batatas doce por acre nos Estados Unidos entre 1869 e 1937.

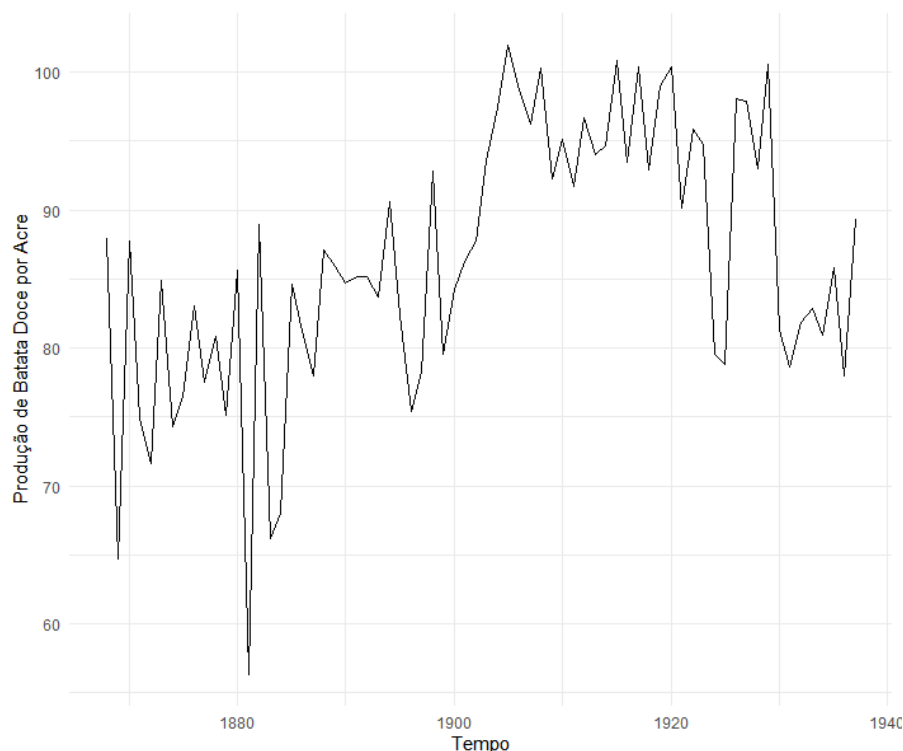


Figura 3.26: Série temporal referente a produção de batatas doce por acre.

A Tabela 3.3, apresenta os resultados dos testes de para tendência determinística para a série temporal da produção de batatas doce por acre nos Estados Unidos entre 1869 e 1937. Em todos os casos, $p\text{-valor} < 0,05$, assim, rejeitamos H_0 , ou seja a série temporal da produção de batatas doce por acre nos Estados Unidos entre 1869 e 1937 apresenta tendência determinística.

□

Tabela 3.3: Resultados dos testes de para tendência determinística para a série temporal da produção de batatas doce por acre nos Estados Unidos entre 1869 e 1937.

Teste	Rotina	H_0	H_1	P-valor	Conclusão
Wald-Wolfowitz	runs.test	STD	TD	0.0000048	TD
	ww.test	STD	TD	0.0000622	TD
Cox-Stuart	cox.stuart.test	STD	TD	0.0000224	TD
	cs.test	STD	TD	0.0023950	TD
Mann-Kendall	mk.test	STD	TD	0.0000689	TD
	MannKendall	STD	TD	0.0000689	TD

3.3 Testes de Raiz Unitária

Primeiramente cabe informar que esta seção baseia-se no trabalho [Mattos \(2018\)](#).

Ao analisarmos séries temporais temos vários aspectos para verificarmos. Um desses aspectos diz respeito aos procedimentos para verificar se uma série temporal é ou não estacionária. Esses procedimentos são conhecidos como *testes de raiz unitária*. Eles recebem essa denominação porque são voltados para verificar se o processo gerador da série apresenta ou não uma raiz unitária na parte AR. Ou seja, se é ou não um processo integrado. Veremos esses testes logo a seguir, mas é válido desde já apontar que, apesar disso, eles também podem ser vistos sob outros ângulos interessantes.

Primeiro, eles também são procedimentos para se detectar a presença ou não de tendências estocásticas no processo gerador das séries. Segundo, como no processo de diferença estacionária a tendência estocástica vem acompanhada de uma tendência determinística, os testes normalmente usam uma representação geral de processo estocástico que permite abarcar as quatro possibilidades apresentadas na seção anterior.

Ao testar restrições particulares impostas à esta representação geral, mais do que verificar a presença ou não de uma raiz unitária (ou de uma tendência estocástica) no processo gerador de uma série temporal, os testes de raiz unitária permitem também distinguir qual dentro os tipos de processos estocásticos considerados está gerando a série. Inicialmente, vamos desenvolver essa representação geral para só depois apresentar os testes de raiz unitária propriamente ditos.

Como vimos nas sessões anteriores, para um processo ARMA ser estacionário, o polinômio característico da parte AR não pode conter raízes de módulo igual a um, chamadas de raízes unitárias. Acontece que a presença de raízes unitárias no polinômio AR resulta na presença de tendência estocástica na série. A identificação de raízes unitárias é de grande importância na análise de séries temporais, e este fato se reflete na literatura relativamente longa tratando do assunto.

3.3.1 Modelo Geral

Considere $\{X_t\}_{t \in \mathbb{Z}}$ o seguinte processo estocástico:

$$X_t = T_t + Y_t, \quad \text{com} \quad T_t = \beta_0 + \beta_1 t \quad (3.8)$$

onde $\beta = (\beta_0, \beta_1)$ são constantes reais e $\{Y_t\}_{t \in \mathbb{Z}}$ um processo AR(1) dado por

$$Y_t = \phi Y_{t-1} + Z_t \quad (3.9)$$

onde $\{Z_t\}_{t \in \mathbb{Z}}$ é um processo estacionário, não necessariamente um processo ruído branco. Note aqui que o processo estocástico para $\{X_t\}_{t \in \mathbb{Z}}$ em (3.8) representa a soma de uma tendência determinística linear mais um processo AR(1) dado por $\{Y_t\}_{t \in \mathbb{Z}}$. Este pode ser estacionário se $|\phi| < 1$ ou não estacionário no sentido de possuir uma raiz unitária se $\phi = 1$. No último caso, $\{Y_t\}_{t \in \mathbb{Z}}$ embute uma

tendência estocástica mais um termo estacionário. Vamos descartar outras possibilidades, em que $\phi \leq -1$ ou $\phi > 1$. Assim, em princípio, a equação (3.8) pode representar um processo tendência estacionária (ou tendência determinística com $\{Y_t\}_{t \in \mathbb{Z}}$ um processo estacionário) ou um processo diferença estacionária ($\{Y_t\}_{t \in \mathbb{Z}}$ segue um processo I(1)), mas ainda há outras possibilidades como veremos a seguir. Para isso, vamos re-escrever a expressão (3.8) como segue:

$$Y_t = X_t - \beta_0 - \beta_1 t. \quad (3.10)$$

Substituindo a equação (3.10) em (3.9) temos

$$X_t - \beta_0 - \beta_1 t = \phi[X_{t-1} - \beta_0 - \beta_1(t-1)] + Z_t$$

ou

$$X_t = \beta_0(1 - \phi) + \beta_1\phi + \beta_1(1 - \phi)t + \phi X_{t-1} + Z_t. \quad (3.11)$$

Tomando $a = \beta_0(1 - \phi) + \beta_1\phi$ e $b = \beta_1(1 - \phi)$ a equação (3.11) pode ser reescrita como

$$X_t = a + bt + \phi X_{t-1} + Z_t. \quad (3.12)$$

A equação (3.12) é uma representação geral para os quatro processos que analisamos antes. Impondo restrições nos valores dos parâmetros $\beta = (\beta_0, \beta_1)$ e ϕ , note que é possível usarmos esta equação para caracterizar os quatro tipos de processos estocásticos que vimos antes.

Muitos livros de Estatística (Séries Temporais) não apresentam a relação entre as expressões (3.11) e (3.12), isto é, que a primeira restringe a segunda. Normalmente, apresentam só a expressão (3.12). A consequência disso é que a formulação (3.11) admite apenas a presença de uma tendência determinística linear, porque não pode acontecer $b \neq 0$ com $\phi = 1$. Se isso fosse possível, haveria uma tendência determinística quadrática junto com a tendência estocástica. Assim, a expressão (3.11) admite apenas os quatro tipos de processos estocásticos considerados até aqui.

Observando a equação (3.12) percebemos que

- (1) Se $\beta_1 = 0$ e $|\phi| < 1$, temos $X_t = a + \phi X_{t-1} + Z_t$ [Processo Estacionário].
- (2) Se $\beta_1 = 0$ e $\phi = 1$, temos $X_t = X_{t-1} + Z_t$ [Processo I(1)].
- (3) Se $|\phi| < 1$, temos $X_t = a + bt + \phi X_{t-1} + Z_t$ [Tendência Estacionária (Determinística)].
- (4) Se $\phi = 1$, temos $X_t = a + X_{t-1} + Z_t$ [Diferencia Estacionária].

Essas possibilidades estão todas reunidas e sintetizadas no Tabela 3.5 a seguir.

Tabela 3.4: Processos estocásticos e tipos de tendências.

Processo Estocástico	Tipo de Tendência	Representação	Restrições equação (3.12)
Processo Estacionário	Sem tendência	$X_t = a + \phi X_{t-1} + Z_t$	$\beta_1 = 0$ e $ \phi < 1$
Processo I(1)	Tendência Estocástica	$X_t = X_{t-1} + Z_t$	$\beta_1 = 0$ e $\phi = 1$
Tendência Estacionária	Tendência Determinística	$X_t = a + bt + \phi X_{t-1} + Z_t$	$ \phi < 1$
Diferencia Estacionária	Tendência Estocástica e Tendência Determinística	$X_t = a + X_{t-1} + Z_t$	$\phi = 1$

Fonte: Mattos (2018), página 23

Várias abordagens para a detecção de raízes unitárias estão a nosso dispor. Um dos testes mais utilizados na literatura é o teste de *Dickey Fuller* que veremos a seguir.

3.3.2 Teste de Dickey Fuller (DF)

O teste de Dickey-Fuller (DF) é o mais antigo e famoso método formal para verificar se uma série temporal é ou não estacionária. Ele foi introduzido em uma versão básica pelos trabalhos de Fuller (1995) e Dickey e Fuller (1979). Posteriormente, foi objeto de uma generalização no trabalho de Dickey e Fuller (1981). A versão generalizada é conhecida como teste de Dickey-Fuller Aumentado, abreviadamente ADF, e constitui a modalidade deste teste que passou a ser usada desde então. Nesta subseção, vamos primeiro falar em detalhe do teste DF, em que se assume que na equação de teste o termo de erro segue um processo ruído branco. Ao final, falaremos de sua versão generalizada, em que o termo de erro segue um processo $I(0)$ autocorrelacionado.

O desenvolvimento do teste DF foi motivado pela necessidade de verificar se uma série precisa ser diferenciada para se tornar estacionária. Essa verificação é a primeira etapa da metodologia proposta por Box et al. (2015) para construção de modelos ARIMA para séries temporais. Esses autores recomendaram que, se o gráfico da série temporal indicar que ela é não estacionária, então devemos diferenciá-la até apresentar um padrão estacionário. Apesar da sofisticação da metodologia de Box e Jenkins, esse procedimento é limitado porque baseia-se simplesmente numa análise gráfica. A preocupação de DF foi propor um método estatístico formal para testar a hipótese de não estacionariedade da série e, assim, ter-se uma indicação mais precisa sobre se é preciso diferenciá-la ou não.

Cabe ressaltar aqui que

- (1) primeiro, DF propuseram na verdade mais de um método para testar a não estacionariedade de uma série. Por exemplo, desenvolveram testes baseados na razão t e em outras estatísticas, como F e de Durbin-Watson. O procedimento mais usado atualmente é baseado na razão t e será apenas este que veremos;
- (2) segundo, a equação de teste inicialmente deixava dúvidas conceituais na interpretação dos parâmetros e essas dúvidas só foram esclarecidas num trabalho posterior, de Dickey et al (1986). Nós já falamos disso, quando apresentamos o desenvolvimento da equação geral para os quatro processos considerados antes;
- (3) terceiro, as hipóteses do teste devem ser interpretadas como se fosse um teste unilateral, mas nosso interesse envolve apenas uma parte da hipótese alternativa.

O primeiro aspecto a destacar sobre o teste DF é que ele procura verificar se uma série é estacionária ou não testando se ela possui uma raiz unitária. A maneira como fazemos isso é assumir que o modelo que gera a série é o mesmo da equação (3.12). Para simplificar o entendimento desse ponto, consideremos uma versão mais simples dessa equação, em que $\beta_0 = \beta_1 = 0$, ou seja $a = b = 0$, então

$$X_t = \phi X_{t-1} + Z_t. \quad (3.13)$$

Observe que, para este modelo, o processo estocástico será estacionário se $|\phi| < 1$ e não estacionário se $|\phi| \geq 1$. Logo o teste de Dickey Fuller (DF) terá como hipóteses

$$H_0 : \phi = 1 \quad \text{versus} \quad H_1 : \phi < 1.$$

A hipótese nula diz que a série é não estacionária, no sentido de que possui uma raiz unitária. De fato, vimos antes que, se $\phi = 1$, então o processo $\{X_t\}_{t \in \mathbb{Z}}$ segue um processo de raiz unitária. Porém, note que, teoricamente, há outras possibilidades para a série ser não estacionária: por exemplo, se $\phi > 1$. O enunciado do teste descarta esta possibilidade porque, neste caso, o processo $\{X_t\}_{t \in \mathbb{Z}}$ seguiria um processo explosivo que cresce ou decresce indefinidamente. Mas, a hipótese nula como definida acima considera uma possibilidade específica de que a série possui uma raiz unitária (e apenas uma). Por sua vez, a hipótese alternativa diz que a série é estacionária, se $|\phi| < 1$, ou não estacionária, se $\phi \leq -1$. Portanto, é preciso cuidado aqui porque a alternativa admite duas possibilidades conflitantes:

a série pode ser estacionária ou não sob essa hipótese. Mas, note que as possibilidades da série ser não estacionária sob H_1 são: a) se $\phi = -1$, a série vai apresentar um ciclo muito curto, com oscilações da série entre o positivo e negativo e com amplitude errática; b) se $\phi < -1$, a série vai exibir o mesmo ciclo curto mas com amplitude explosiva.

Vemos, portanto, que o enunciado acima tanto na hipótese nula quanto na alternativa admite várias formas de não estacionariedade. Porém, é exatamente esse o enunciado do teste DF. Apesar disso, o teste DF nos permite chegar à conclusão de que uma série é não estacionária, no sentido de possuir uma ou mais raízes unitárias, ou estacionária. Na prática, fica mais fácil entendermos o procedimento do teste DF se imaginarmos um enunciado que é um pouco diferente:

$$H_0 : \phi = 1 \quad \text{versus} \quad H_1 : |\phi| < 1.$$

O enunciado re-escrito desta forma, nos coloca exatamente na situação que nos interessa mais. Testaremos a hipótese nula de que a série possui uma raiz unitária, contra a alternativa de interesse de que a série não possui raiz unitária, mas é estacionária.

O teste DF está disponível em três opções, e cada versão se baseia na estimação por Mínimos Quadrados Ordinários (MQO) de uma variante da equação (3.12) para o cálculo das estatísticas de teste. Se H_0 for verdadeira, então o processo estocástico gerador da série possui uma raiz unitária e embute uma tendência estocástica. No entanto, se H_0 for falsa, então o processo é estacionário sem qualquer tendência (opções 1 e 2) ou é uma tendência estacionária (versão 3).

O teste DF, no entanto, não usa diretamente a equação (3.12). Subtraindo X_{t-1} de ambos os lados dessa equação, temos:

$$\Delta X_t = a + bt + (\phi - 1)X_{t-1} + Z_t, \quad (3.14)$$

onde a e b definidos como anteriormente. Na versão original e mais simples do teste, Dickey e Fuller (1979) assumem que o termo $\{Z_t\}_{t \in \mathbb{Z}}$ é um processo ruído branco normalmente distribuído. Na versão aumentada (ADF), Dickey e Fuller (1981) assumem que ele segue um processo $AR(p)$. A equação (3.14), com as mesmas restrições dadas pela equação (3.11), é que é usada no teste DF para cálculo das estatísticas de teste. Embora seja apenas uma representação alternativa do processo gerador do processo $\{X_t\}_{t \in \mathbb{Z}}$ na equação (3.11). É importante observar que, conforme a equação (3.14), as hipóteses nula e de interesse do teste passam a ser escritas como:

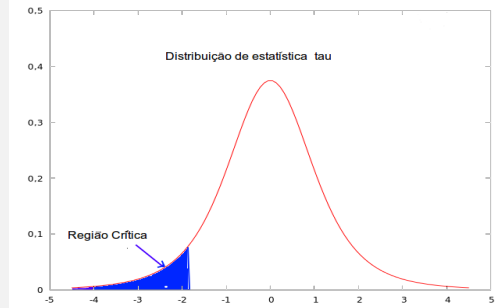
$$H_0 : \phi = 1 \quad \text{versus} \quad H_1 : |\phi| < 1.$$

O teste de Dickey Fuller consiste em fazer um teste t (mas com distribuição de Dickey-Fuller) para a significância do seguinte modelo

Teste de Dickey Fuller

$H_0: \delta = 0$ (Não estacionário) versus $H_1: \delta < 0$ (Estacionário)

$$\Delta X_t = (\phi - 1)X_{t-1} + Z_t = \delta X_{t-1} + Z_t,$$



em que Δ é o operador de diferenciação, dado por $\Delta X_t = X_t - X_{t-1}$. Testar a presença de raiz unitária neste modelo ($\phi = 1$) é equivalente a atestar se $\delta = \phi - 1 = 0$. Como o teste é feito sobre os resíduos, a distribuição de um teste t usual não será usual, nem mesmo assintoticamente. Para isso existe uma estatística de teste específica, τ , cujos valores críticos estão dispostos na tabela de Dickey Fuller. Além disso, o teste de Dickey Fuller é um teste unilateral a esquerda conforme figura acima.

Existem três versões principais do teste:

- Teste para raiz unitária:

$$\Delta X_t = \delta X_{t-1} + Z_t \rightarrow \tau;$$

- Teste para raiz unitária com drift (constante $\mu = a$):

$$\Delta X_t = a + \delta X_{t-1} + Z_t \rightarrow \tau_\mu;$$

- Teste de raiz unitária com drift e tendência temporal determinística:

$$\Delta X_t = a + bt + \delta X_{t-1} + Z_t \rightarrow \tau_\tau$$

A estatística $\hat{\tau}$ para cada um dos modelos pode ser obtida da seguinte forma:

$$\hat{\tau} = \frac{\hat{\delta}}{s(\hat{\delta})} \quad (3.15)$$

em que $s(\hat{\delta})$ é o desvio padrão de

$$\hat{\delta} = \frac{\sum_{t=2}^n Y_{t-1} Y_t}{\sum_{t=2}^n Y_{t-1}^2} - 1,$$

que é a estimativa de mínimos quadráticos de ρ menos 1, para garantir que, sob H_0 , tenhamos $\delta = 0$. O desvio padrão pode ser obtido a partir do cálculo da variância residual, que no caso mais simples se torna

$$S^2 = \frac{1}{n} \sum_{t=1}^n (\Delta Y_t - \hat{\delta} Y_{t-1})^2.$$

Tabela 3.5: Valores críticos para o teste de Dickey Fuller.

Estatística	n	1%	2.5%	5%	10%
τ	25	-2.66	-2.26	-1.95	-1.60
	50	-2.62	-2.25	-1.95	-1.61
	100	-2.60	-2.24	-1.95	-1.61
	250	-2.58	-2.23	-1.95	-1.61
	500	-2.58	-2.23	-1.95	-1.61
	>500	-2.58	-2.23	-1.95	-1.61
τ_μ	25	-3.75	-3.33	-3.00	-2.62
	50	-3.58	-3.22	-2.93	-2.60
	100	-3.51	-3.17	-2.89	-2.58
	250	-3.46	-3.14	-2.88	-2.57
	500	-3.44	-3.13	-2.87	-2.57
	>500	-3.43	-3.12	-2.86	-2.57
τ_τ	25	-4.38	-3.95	-3.60	-3.24
	50	-4.15	-3.80	-3.50	-3.18
	100	-4.04	-3.73	-3.45	-3.15
	250	-3.99	-3.69	-3.43	-3.13
	>500	-3.98	-3.68	-3.42	-3.13

Cada versão do teste (τ , τ_μ e τ_τ) tem sua própria estatística de teste e portanto tem seu próprio valor crítico o qual depende do tamanho amostral. Esses valores foram obtidos a partir de simulações de Monte Carlo.

Em cada caso, a hipótese nula de que *existe raiz unitária* é representada por $\delta = 0$. Para estes testes é conhecido que eles tem *baixo poder* no sentido de que frequentemente não conseguem distinguir entre processos com raiz unitária ($\delta = 0$) de processos com raiz quase-unitária (δ próximo de zero), ou até mesmo com tendências não lineares.

A Tabela 3.5 apresenta alguns valores críticos para o teste de Dickey Fuller.

Compare a estatística τ calculada com o valor crítico tabulado segundo o nível de significância escolhido. O valor crítico pode ser encontrado, por exemplo, na Tabela 3.5. Se a estatística τ for maior ou igual ao valor crítico, não rejeite H_0 , isto é, considere que a série é não estacionária e possui uma raiz unitária. Se a estatística τ for menor, rejeite H_0 e conclua que a série não possui raiz unitária.

3.3.3 Teste Aumentado de Dickey Fuller (ADF)

A versão aumentada do teste DF, que chamamos antes de teste ADF, difere apenas por considerar a existência de alguma estrutura de autocorrelação para os erros da equação de teste. Se essa estrutura não for considerada, há perda de eficiência do estimador de MQO para δ e, o que é mais sério, as estatísticas τ , τ_μ e τ_τ ficam enviesadas. Na prática, isso é considerado usando uma versão aumentada da equação de teste em que se permite termos defasados de ΔX_t como variáveis adicionais no lado direito da expressão (3.14), sendo que para determinar a quantidade geralmente usam-se critérios de Informação do tipo AIC e BIC/SBC:

$$\Delta X_t = a + bt + \delta X_{t-1} + \sum_{j=1}^p \delta_j \Delta X_{t-j} + Z_t, \quad (3.16)$$

onde δ_j , para $j = 1, \dots, p$ são parâmetros. O objetivo desse procedimento é eliminar uma possível existência de autocorrelação serial no termo $\{Z_t\}_{t \in \mathbb{Z}}$, ou seja transformá-lo em um processo ruído

branco $\{\varepsilon_t\}_{t \in \mathbb{Z}}$.

Portanto, as equações do teste (geralmente referenciadas como ADF versão 1, 2 e 3, respectivamente) são:

$$\Delta X_t = \delta X_{t-1} + \sum_{j=1}^p \delta_j \Delta X_{t-j} + \varepsilon_t, \quad (3.17)$$

$$\Delta X_t = a + \delta X_{t-1} + \sum_{j=1}^p \delta_j \Delta X_{t-j} + \varepsilon_t, \quad (3.18)$$

$$\Delta X_t = a + bt + \delta X_{t-1} + \sum_{j=1}^p \delta_j \Delta X_{t-j} + \varepsilon_t, \quad (3.19)$$

Nas três opções, aplica-se o mesmo procedimento de testar $H_0 : \delta = 0$. E em cada uma delas, H_0 continua tendo as mesmas interpretações. Um problema novo que aparece, porém, com esta versão aumentada é a necessidade de se determinar com antecedência o lag máximo p dos termos defasados de $\delta_j \Delta X_{t-j}$. Isto é feito estimando-se várias vezes a equação de teste com números diferentes de termos defasados, isto é, para o valor de p . Escolhe-se o valor de p que minimiza alguma estatística de critério de informação, como a de Schwarz, por exemplo. Uma vez escolhido o lag máximo, implementa-se o teste ADF propriamente dito.

[Mattos \(2018\)](#) apresenta alguns passos para aplicação do teste ADF em séries temporais, apresentados a seguir.

- (1) *Escolha da opção (1, 2, ou 3) do teste:* examine o gráfico da série para verificar a presença ou ausência aparente de uma tendência determinística. Se parecer não haver uma tendência determinística, escolha a opção 1 (sem intercepto nem termo de tendência determinística) ou a opção 2 (só intercepto), que assumem na hipótese H_0 que o processo de raiz unitária possui uma tendência estocástica apenas. No caso de haver forte evidência visual de uma tendência determinística na evolução da série, escolha a opção 3 (intercepto mais termo de tendência). Havendo dúvida, escolha a opção 3, porque é a mais geral e engloba as demais.
- (2) *Lag máximo da equação de teste:* uma vez escolhida a opção do teste, pode-se determinar qual a especificação da equação de teste que será estimada. Isso envolve antes determinar o lag máximo p dos termos defasados da variável dependente ΔX_t que serão usados na estimação da equação de teste. Para tanto, proceda da seguinte forma: Estime a equação de teste sem nenhum termo defasado de ΔX_t e registre o critério de informação (Schwarz, por exemplo). Repita a estimação da equação de teste com um termo defasado ΔX_{t-1} e novamente registre o critério de informação. Compare os dois critérios de informação: se o da última equação estimada for maior do que o da anterior, pare e use a equação anterior para implementar o teste; se for menor, continue. Estime então a equação de teste agora com dois termos defasados, ΔX_{t-1} e ΔX_{t-2} , e proceda à mesma comparação dos critérios de informação da equação atual e da anterior. Pare ou então continue sucessivamente até o momento em que o critério de Schwarz da equação atual aumentar em relação ao da equação anterior. Isso significa que o valor de p da penúltima equação é o lag máximo.
- (3) *Estatística de teste:* Tendo determinado o lag máximo p , estime em definitivo a equação de teste e compute a estatística (τ) correspondente.
- (4) *Decisão Final:* Compare a estatística τ calculada com o valor crítico tabulado segundo o nível de significância escolhido. O valor crítico pode ser encontrado, por exemplo, na Tabela 3.5. Se a estatística τ for maior ou igual ao valor crítico, não rejeite H_0 , isto é, considere que a série é não estacionária e possui uma raiz unitária. Se a estatística τ for menor, rejeite H_0 e conclua

que a série não possui raiz unitária. Refine sua interpretação em termos da presença/ausência de tendências determinística e estocástica em função da opção do teste que você escolheu.

- (5) *Reaplicando o teste*: Se H_0 não for rejeitada no passo (4), significa que o processo gerador da série possui uma raiz unitária. A princípio, isso significa que seu processo gerador é, ou embute, um processo $I(1)$. É possível, no entanto, que o processo gerador da série possua mais raízes unitárias e assim seja, ou possua, um processo integrado de ordem maior. Para verificar isso, diferencie a série uma vez e repita todos os procedimentos anteriores do teste ADF para a série diferenciada. Se H_0 for rejeitada, é porque a série diferenciada não tem raiz unitária e, portanto, a série original é $I(1)$. De outro modo, se H_0 não for rejeitada é porque a série diferenciada possui uma raiz unitária. Neste caso, diferencie novamente a série e aplique de novo o teste ADF. Proceda dessa maneira iterativamente até chegar a um grau de diferenciação da série em que H_0 é finalmente rejeitada. Isso significa que, para esse grau de diferenciação, o processo que gera a série não possui raiz unitária (i.e, é estacionário ou do tipo tendência estacionária)¹.

3.3.4 Teste de Phillips - Perron (PP)

Phillips (1987) e Phillips e Perron (1988) propuseram outro teste de raiz unitária que generaliza o teste ADF para uma ampla classe de modelos em que os erros $\{Z_t\}_{t \in \mathbb{Z}}$ na equação (3.14) são autocorrelacionados e heterogeneamente distribuídos. Ele é conhecido como teste de Phillips-Perron ou, abreviadamente, teste PP. Essencialmente, o procedimento do teste PP é o mesmo que o do teste ADF e envolve a estimação da equação (3.14) para cálculo das estatísticas de teste, introduzindo apenas uma modificação nas últimas. Neste sentido, ele permite as mesmas três opções para testar

$$H_0 : \phi = 1 \quad \text{versus} \quad H_1 : |\phi| < 1.$$

Na primeira opção, considera-se um modelo para $\{X_t\}_{t \in \mathbb{Z}}$ sem constante (i.e., constante nula), logo $\Delta X_t = \delta X_{t-1} + Z_t$; na segunda, considera-se a possibilidade de uma constante não nula, logo $\Delta X_t = a + \delta X_{t-1} + Z_t$ e na terceira admite-se também um termo de tendência $\Delta X_t = a + bt + \delta X_{t-1} + Z_t$.

Para permitir situações mais abrangentes para o termo de erro e também desenvolver uma teoria assintótica de teste conveniente, o teste PP difere do teste ADF em dois aspectos principais. O primeiro é que as fórmulas das estatísticas de teste em cada situação são diferentes das fórmulas da estatística τ do teste ADF. As expressões para as estatísticas do teste PP são:

$$Z(\tau) = \frac{s}{\hat{\sigma}_Z^2} \tau - \frac{n(\hat{\sigma}_Z^2 - S^2)S_{\hat{\delta}}}{2\hat{\sigma}_Z^2 S} \quad (3.20)$$

$$Z(\tau_\mu) = \frac{s}{\hat{\sigma}_Z^2} \tau_\mu - \frac{n(\hat{\sigma}_Z^2 - S^2)S_{\hat{\delta}}}{2\hat{\sigma}_Z^2 S} \quad (3.21)$$

$$Z(\tau_\tau) = \frac{s}{\hat{\sigma}_Z^2} \tau_\tau - \frac{n(\hat{\sigma}_Z^2 - S^2)S_{\hat{\delta}}}{2\hat{\sigma}_Z^2 S} \quad (3.22)$$

De forma análoga ao teste ADF, essas estatísticas são obtidas a partir da estimação por MQO das equações de teste correspondentes a cada opção, mas sem os termos defasados na variável dependente. O termo $S_{\hat{\delta}}$ é o erro-padrão do estimador de MQO para δ . O termo S^2 é a variância residual da regressão de teste e corresponde a um estimador consistente da variância do erro $\{Z_t\}_{t \in \mathbb{Z}}$ sob a hipótese de que este segue um processo ruído branco.

¹ Há um outro procedimento na literatura para se determinar a ordem de integração, ou o número de raízes unitárias, de uma série temporal, proposto por Dickey e Pantula (1987). Esse procedimento é mais rigoroso para isso do que o teste ADF.

O termo $\hat{\sigma}_Z^2$ é um estimador consistente da variância do erro $\{Z_t\}_{t \in \mathbb{Z}}$ sob a hipótese de que este é estacionário fraco mas admitindo condições mais genéricas de autocorrelação e heterogeneidade do processo estocástico que gera esses erros. Este termo é o único componente que não é calculado diretamente das equações de teste estimadas. Phillips e Perron se baseiam no fato de que a variância assintótica neste caso é dada por:

$$\hat{\sigma}_Z^2 = 2\pi f_Z(0),$$

onde $f_Z(\cdot)$ é o valor do espectro (função densidade espectral)² do processo $\{Z_t\}_{t \in \mathbb{Z}}$. Dado este fato, os autores sugerem o uso de procedimentos disponíveis na literatura de análise de séries temporais para estimação consistente do espectro de potência e recomendam, para os três casos, que se use a seguinte expressão:

$$\hat{\sigma}_Z^2 = \frac{1}{n} \left[\sum_{t=1}^n \hat{Z}_t^2 + 2 \sum_{s=1}^l w_l \sum_{t=s+1}^n \hat{Z}_t \hat{Z}_{t-s} \right] \quad (3.23)$$

com \hat{Z}_t para $t = 1, \dots, n$ são os resíduos da regressão correspondente a cada opção. O termo w_l , para $s = 1, \dots, l$, refere-se a um conjunto de pesos que constitui a janela de defasagem (lag window) usada para suavizar as estimativas do espectro. Phillips e Perron apontam que há vários tipos de janelas (métodos para determinar os pesos) que podem ser usadas, como as janelas triangular (ou de Bartlett), de Parzen e de Newey West. Nos softwares computacionais, usualmente são dadas opções ao usuário de escolher a janela de defasagem para estimação do espectro.

Para o ajuste do teste PP no [R Core Team \(2020\)](#), também pode-se utilizar a rotina `pp.test` do pacote `tseries`, e sua sintaxe é

Calcula o Teste PP cuja hipótese nula é que a série temporal possui raiz unitária.

```
pp.test(x, alternative=c("stationary", "explosive"), type=c("Z(alpha)", "Z(t_alpha)"),
lshort=TRUE)
```

onde

- x é a série temporal a ser testada;
- `alternative`: indica a hipótese alternativa do teste (estacionária ou explosiva);
- `type`: indica qual das variantes do teste deve ser calculado "Z(alpha)"(default) ou "Z(t_alpha)".
- `lshort`: quantidade máxima de defasagens para se calcular a estatística do teste (short ou long);

Maiores detalhes ver: <https://bit.ly/2yVfVea>.

3.3.5 Teste de Kwiatkowski, Phillips, Schmidt e Shin (KPSS)

Em econometria, Kwiatkowski-Phillips-Schmidt-Shin (KPSS) ensaios são usados para testar uma hipótese nula de que uma observável série temporal é estacionária em torno de uma tendência determinista (isto é tendência-estacionário) contra a alternativa de uma raiz unitária.

²Veremos a definição de função densidade espectral ou espectro de um processo estocástico mais adiante em processos com longa dependência.

Ao contrário do que a maioria dos testes de raiz unitária, a presença de uma unidade de raiz não é a hipótese nula mas a alternativa. Além disso, no teste KPSS, a ausência de uma raiz unitária não é uma prova de estacionariedade, mas de tendência estacionariedade. Esta é uma distinção importante, pois é possível para uma série de tempo para ser não-estacionária, não têm raiz unitária ainda ser tendência estacionária.

Tais modelos foram propostos em 1982 por Alok Bhargava em seu Ph.D. tese onde vários John von Neumann - ou Durbin-Watson - tipo de testes de amostra finitos para raízes unitárias foram desenvolvidas (ver Bhargava, 1986). Mais tarde, Kwiatkowski et al. (1992) propôs um teste da hipótese nula de que uma série observável é tendência estacionária (estacionária em torno de uma tendência determinística). A série é expressa como a soma de tendência determinística, passeio aleatório, e erro estacionário, e o teste é o teste do multiplicador de Lagrange da hipótese de que o passeio aleatório tem zero variância. Testes KPSS do tipo destinam-se a complementar os testes de raiz unitária, tais como os testes de Dickey-Fuller. Ao testar tanto a hipótese de raiz unidade e a hipótese de estacionariedade, pode-se distinguir série que parece ser estacionário, série que parecem ter uma raiz unidade, e série para a qual os dados (ou os testes) não são suficientemente informativos para ter a certeza se eles são estacionários ou integrado. (Fonte: https://pt.qwe.wiki/wiki/KPSS_test).

Um problema do teste ADF é seu pequeno poder, que o torna quase incapaz de diferenciar uma séries com raiz unitária de uma com raiz “quase” unitária, ou seja, cria um viés para a conclusão da raiz unitária. Isso é um problema sério dado que a correção para uma série não estacionária é sua diferenciação, o que implica que se o teste falha, nos leva a diferenciar uma série estacionária, que elimina muita informação a respeito do processo.

Entre os testes alternativos, o mais conhecido é o teste KPSS apresentado por Kwiatkowski et al. (1992) cujas hipóteses são

$$H_0 = \text{A série é estacionária} \quad \textit{versus} \quad H_1 = \text{A série apresenta raiz unitária}$$

Note que as hipóteses deste teste não são iguais aos testes de Dickey - Fuller e Phillips - Perron para estacionariedade. A hipótese nula deste teste é igual às hipóteses alternativas nos testes anteriores.

Seja $\{X_t\}_{t \in \mathbb{Z}}$ um processo estocástico o qual queremos testar sua estacionariedade. Suponha que podemos decompor o processo em componentes de tendência, passeio aleatório e erro

$$X_t = bt + r_t + \varepsilon_t. \quad (3.24)$$

Neste caso, $\{r_t\}_{t \in \mathbb{Z}}$ é um passeio aleatório

$$r_t = r_{t-1} + u_t$$

onde $\{u_t\}_{t \in \mathbb{Z}}$ é $WN(0, \sigma_u^2)$. O valor inicial r_0 é tratado como fixo e tem o papel de um intercepto. A hipótese nula é de que o processo é *trend-stationary* (tendência-estacionário), ou seja, que a variância do processo que possui raiz unitária é zero ($\sigma_u^2 = 0$). Desde que $\{\varepsilon_t\}_{t \in \mathbb{Z}}$ seja estacionário, sob a hipótese nula $\{X_t\}_{t \in \mathbb{Z}}$ é estacionário-tendência. Também será considerado o caso especial do modelo 3.24 em que definimos $b = 0$, nesse caso, sob a hipótese nula, $\{X_t\}_{t \in \mathbb{Z}}$ é estacionário em torno de um nível r_0 ao invés de em torno de uma tendência.

Nabeya et al. (1988) consideram o modelo de regressão

$$X_t = Y_t \beta_t + Z_t' \gamma + \varepsilon_t,$$

na qual β_t é um passeio aleatório gaussiano ($\beta_t = \beta_{t-1} + u_t$, com $\{u_t\}_{t \in \mathbb{Z}}$ é $WN(0, \sigma_u^2)$). O modelo apresentado por Kwiatkowski et al. (1992) é um caso particular do modelo de Nabeya et al. (1988) na qual $Y_t = 1$, para todo t , $Z_t = t$ e o $\beta_t = r_t$.

Assim a estatística de teste utilizada por Kwiatkowski et al. (1992) é a mesma que Nabeya et al. (1988) que é definida a seguir. Sob as suposições de que $\sigma_u^2 = 0$ e $\{\varepsilon_t\}_{t \in \mathbb{Z}}$ é um ruído branco gaussiano

$N(0, \sigma_\varepsilon^2)$. Então seja e_t , para $t = 1, \dots, n$ os resíduos da regressão entre X_t , para $t = 1, \dots, n$, sobre um intercepto e uma tendência. Seja $\hat{\sigma}_\varepsilon^2$ a estimativa da variância do erro de regressão (soma dos quadrados dividido por n). Defina a soma parcial dos resíduos como

$$S_t = \sum_{j=1}^t e_j, \quad t = 1, \dots, n.$$

A estatística do teste é dada por:

$$LM = \frac{1}{\hat{\sigma}_\varepsilon^2} \sum_{t=1}^n S_t^2.$$

Caso os resíduos da regressão não sejam i.i.d. é necessário a utilização de um estimador para σ_ε^2 o qual depende das autocovariâncias e do lag escolhido na estimação definido na correção de Bartlett. Neste caso, utiliza-se a equação 3.23 para estimar σ_ε^2 .

Kwiatkowski et al. (1992) mostram que a estatística LM tem distribuição que converge assintoticamente para um Movimento Browniano em que seus valores críticos são tabelados.

Para o ajuste do teste KPSS no R Core Team (2020), pode-se utilizar a rotina *kpss.test* do pacote *tseries*, e sua sintaxe é

Calcula o Teste KPSS cuja hipótese nula é que a série temporal é estacionária.

```
kpss.test(x, null=c("Level","Trend"), lshort=TRUE)
```

onde

- x é a série temporal a ser testada;
- `null`: indica a hipótese nula do teste: `level` (default) ou `Trend`;
- `lshort`: quantidade máxima de defasagens para se calcular a estatística do teste (`short` ou `long`);

Maiores detalhes ver: <https://bit.ly/2zGLPeH>.

Exemplo 3.5. Vamos aplicar os testes de raiz unitária a séries temporal do Índice Mensal de Atividade Econômica do Banco Central - IBC-Br, período de 01/2003 a 02/2020. Fonte: Banco Central do Brasil - Departamento Econômico. Endereço dos dados: <https://bit.ly/2xpSqKk>. A Figura 5.2 apresenta o gráfico da Série Temporal do Índice de Atividade Econômica do Banco Central - IBC-Br e sua FAC Amostral. Maiores detalhes sobre o IBC-Br ver <https://bit.ly/3d5LhgN>.

```
1 #
2 #Carregar a série temporal no R Core Team (2020)
3 fig = plot_ly(IBC_Br, x = ~data, y = ~valor, type = 'scatter', mode = 'lines')
4 fig <- fig %>% layout(title = "Índice de Atividade Econômica do Banco Central - IBC-Br")
5 fig
6 ggAcf(IBC_Br$valor, lag.max=100, type=c("correlation"))+labs(y="FAC Amostral", title="IBC-Br")+theme_minimal()
7 IBC_Br <- ts(IBC_Br$valor, start=c(2003,1), frequency = 12)
8 #
9 #Teste ADF
10 adf.test(IBC_Br)
11 #
12 Augmented Dickey-Fuller Test
13
```

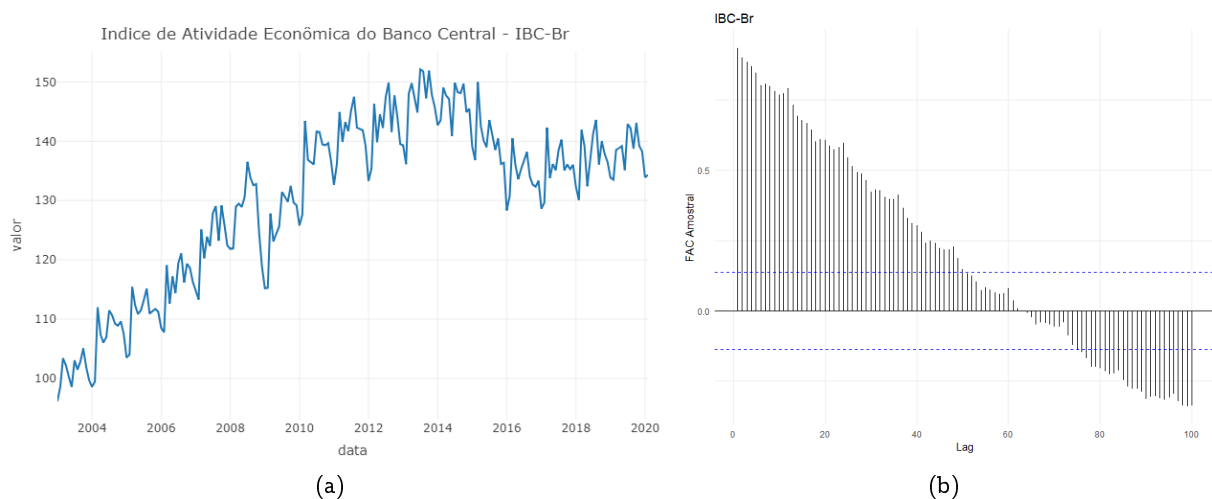


Figura 3.27: Série Temporal do Índice Mensal de Atividade Econômica do Banco Central - IBC-Br, período de 01/2003 a 02/2020: (a) Gráfico Séries Temporal; (b) FAC Amostral.

```

14 data:  ibc_br
15 Dickey-Fuller = -2.0889, Lag order = 5, p-value = 0.5388
16 alternative hypothesis: stationary
17 #
18 #Teste KPSS
19 kpss.test(ibc_br)
20 #
21 KPSS Test for Level Stationarity
22
23 data:  ibc_br
24 KPSS Level = 3.025, Truncation lag parameter = 4, p-value = 0.01
25
26 Warning message:
27 In kpss.test(ibc_br) : p-value smaller than printed p-value
28 #
29 #Teste PP
30 pp.test(ibc_br)
31 #
32 Phillips-Perron Unit Root Test
33
34 data:  ibc_br
35 Dickey-Fuller Z(alpha) = -17.771, Truncation lag parameter = 4, p-value = 0.1045
36 alternative hypothesis: stationary
37 #

```

Código 3.23: Código fonte em R: Comandos para aplicação dos testes ADF, PP e KPSS na séries temporal do Índice Mensal de Atividade Econômica do Banco Central - IBC-Br, período de 01/2003 a 02/2020.

A Tabela 3.3, apresenta os resultados dos testes de para tendência determinística para a série temporal da produção de batatas doce por acre nos Estados Unidos entre 1869 e 1937. Em todos os casos, $p\text{-valor} < 0,05$, assim, rejeitamos H_0 , ou seja a série temporal da produção de batatas doce por acre nos Estados Unidos entre 1869 e 1937 apresenta tendência determinística.

```

1 #
2 diff_ibc_br<-diff(ibc_br, differences = 1)
3 autoplot(diff_ibc_br,ylab = "")+theme_minimal()
4 ggAcf(diff_ibc_br,lag.max=100,type=c("correlation"))+labs(y="FAC Amostral",title="IBC-

```

Tabela 3.6: Resultados dos testes de para tendência (determinística e estocástica - RU) para a série temporal do Índice Mensal de Atividade Econômica do Banco Central - IBC-Br.

Teste	H_0	H_1	P-valor	Conclusão
Wald-Wolfowitz	STD	TD	$< 2.2\text{e-}16$	TD
Cox-Stuart	STD	TD	$< 2.2\text{e-}16$	TD
Mann-Kendall	STD	TD	$< 2.2\text{e-}16$	TD
ADF	RU	Estacionária	0.5388	RU
PP	RU	Estacionária	0.1045	RU
KPSS	Nível Estac.	RU	< 0.01	RU
KPSS	Tend. Estac.	RU	< 0.01	RU

```
Br")+theme_minimal()
#
```

Código 3.24: Código fonte em R: Comandos para diferenciar a séries temporal do Índice Mensal de Atividade Econômica do Banco Central - IBC-Br, período de 01/2003 a 02/2020.

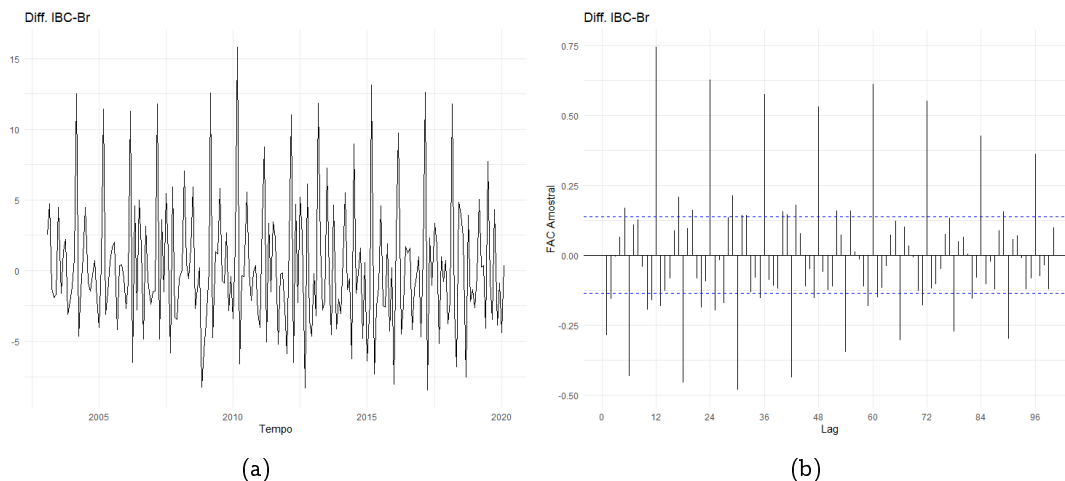


Figura 3.28: Série Temporal do Índice Mensal de Atividade Econômica do Banco Central - IBC-Br, período de 01/2003 a 02/2020 diferenciada com $d = 1$: (a) Gráfico da Séries Temporal Diferenciada; (b) FAC Amostral da Série Temporal Diferenciada.

□

3.4 Regressão Local (LOESS)

[<https://bit.ly/36SzUsx>] Regressão Local (Loess) é um método não paramétrico que estima curvas e superfícies através de suavização (smoothing). Este método ganhou popularidade a partir da década de 70 com o desenvolvimento de computadores e a publicação dos estudos independentes de Stone (1980), Cleveland (1979) e Stone (1977). Sendo que Cleveland (1979) desenvolveu o software *Lowess*, que foi implementado em diversos pacotes estatísticos. As ideias básicas do método podem ser observadas ao considerar-se o mais simples dos modelos de regressão, onde a variável dependente, y , e a independente, x , são relacionadas por:

$$Y_i = f(x_i) + \varepsilon_i, \quad (3.25)$$

onde $f(x)$ é não conhecida e ε_i denota o termo de erro independente e identicamente distribuído média zero e σ_ε^2 constante ou variabilidade de fontes não incluídas em x_i .

Não fazemos suposições globais sobre a função $f(x)$ mas suponha que localmente pode ser bem aproximado com um membro de uma classe simples de função paramétrica, por exemplo uma linha constante ou reta. O Teorema de Taylor diz que qualquer função contínua pode ser aproximado com um polinômio.

Teorema de Taylor

A fórmula de Taylor fornece uma aproximação polinomial local cada vez melhor de uma função (muitas vezes) diferenciável à medida que calculamos as suas derivadas. Ela nos ensina como aproximar funções complicadas por funções mais simples.

Fórmula de Taylor

Teorema 3.1. *Seja $f : \mathbb{R} \rightarrow \mathbb{R}$ uma função e x_0 um número real no qual as derivadas $f^{(1)}(x_0), f^{(2)}(x_0), \dots, f^{(k)}(x_0)$ são contínuas. Então, se definirmos o polinômio (conhecido como polinômio de Taylor) como*

$$P_k(x) = f(x_0) + f^{(1)}(x_0)(x - x_0) + \frac{f^{(2)}(x_0)}{2!}(x - x_0)^2 + \frac{f^{(3)}(x_0)}{3!}(x - x_0)^3 + \dots + \frac{f^{(k)}(x_0)}{k!}(x - x_0)^k,$$

valendo que

$$\lim_{x \rightarrow x_0} \frac{f(x) - P_k(x)}{(x - x_0)^k} = 0.$$

Outra forma de escrever isso é

$$f(x) = P_k(x) + o(|x - x_0|^k)$$

Já a variação desse resultado associada ao teorema do valor médio é apresentado abaixo. Nela é possível ver de forma mais explícita o erro.

Fórmula de Taylor com resto de Lagrange

Teorema 3.2. *Seja $f : \mathbb{R} \rightarrow \mathbb{R}$ uma função e x, y dois números reais. Se as derivadas $f^{(1)}, f^{(2)}, \dots, f^{(k)}, f^{(k+1)}$ estão todas definidas no intervalo $[x, y]$, então existe $z \in (x, y)$ tal que*

$$f(x) = P_k(x) + \frac{f^{(k+1)}(z)}{(k+1)!}(x - x_0)^{k+1}.$$

Ajustando Polinômios Locais

Ao contrário dos métodos paramétricos que estimam a função globalmente, regressão local estima a função $f(x)$ na vizinhança de cada ponto de interesse $x = x_0$. Uma forma simples de estimar uma função localmente é considerar a média ponderada das observações que estão na vizinhança do ponto de interesse, x_0 , a citar, $[x_0 - h(x_0); x_0 + h(x_0)]$, será considerado na estimativa de $f(x_0)$. Duas escolhas devem ser feitas para realizar esta estimativa. Primeiro, deve ser escolhido o tamanho da vizinhança,

h , do ponto $x = x_0$ e, segundo, deve ser escolhida uma função K que pondera o conjunto de pontos vizinhos a x_0 . A função K é denominada de núcleo (Kernel), enquanto que h denominada de banda ou parâmetro de suavização (que em regressão local é chamada de *span* ou *bandwidth*). Com este procedimento, a equação para a média local ponderada por K é dada por:

$$\hat{f}(x_0) = \frac{\sum_{j=1}^n K_h(x_j - x_0) y_j}{\sum_{j=1}^n K_h(x_j - x_0)}.$$

Este estimador de núcleo foi proposto inicialmente por [Nadaraya \(1964\)](#) e [Watson \(1964\)](#). Existem sérias limitações com a estimativa de uma constante localmente, como por exemplo, viés nas regiões de fronteira e no interior se a variável independente não for uniforme e se a função de regressão tiver curvatura. Uma maneira de resolver este problema é através de regressão local linear ponderada, proposta inicialmente por [Cleveland \(1979\)](#) e [Stone \(1977\)](#). Ao estimar (localmente) uma função polinomial ao invés de uma constante, o problema de viés de primeira ordem é eliminado, desta forma, regressão local linear resolve um problema de mínimos quadrados ponderados a cada ponto de interesse, x_0 . Assim, dentro da janela de suavização, $f(x)$ é aproximada por uma função polinomial. Por exemplo, uma equação da reta

$$f(x) \approx \beta_0 + \beta_1(x - x_0) + \epsilon, \quad \text{para } x \in [x_0 - h(x_0); x_0 + h(x_0)]$$

ou seja,

$$\hat{\beta} = \arg \min_{\beta \in \mathbb{R}^2} \sum_{j=1}^n K_h(x_j - x_0) [y_j - \{\beta_0 + \beta_1(x - x_0)\}]^2, \quad (3.26)$$

onde o espaço de parâmetros $\beta = (\beta_0, \beta_1) \in \mathbb{R}^2$.

Regressão local linear será igual ao estimador de Nadaraya-Watson expresso pela equação (3.26) se o termo $\beta(x_i - x_0)$ for removido. Neste caso, uma constante será estimada localmente. Apesar de regressão local linear ser utilizado como técnica padrão por muitos autores (ver [Bowman e Azzalini \(1997\)](#)), não há razões para não utilizar polinômios de ordem mais alta, mesmo porque a regressão local pode apresentar viés quando a função a ser estimada possui uma forte curvatura. Nestes casos uma polinomial de grau d pode ser estimada através da seguinte função

$$\hat{\beta} = \arg \min_{\beta \in \mathbb{R}^{d+1}} \left\{ \sum_{j=1}^n K_h(x_j - x_0) \left[y_j - \sum_{k=0}^d \beta_k (x - x_0)^k \right]^2 \right\}, \quad (3.27)$$

onde o espaço de parâmetros $\beta = (\beta_0, \beta_1, \dots, \beta_d) \in \mathbb{R}^{d+1}$ e $d \in \mathbb{N}$.

Portanto para modelar determinado processo por regressão local, deve-se de forma geral fazer três escolhas: a função núcleo, o parâmetro de suavização e o grau da função polinomial. Existe ainda uma outra escolha que deve ser feita, que diz respeito a distribuição assumida para os termos de erro, no presente trabalho assume-se que os erros seguem uma distribuição gaussiana. Para uma discussão de regressão local considerando outras distribuições do erro ver [Loader \(1999\)](#).

Grau do Polinômio

Esta escolha também afeta a relação entre variância e viés, quanto maior o grau da polinomial menor será o viés e maior a variância para um mesmo parâmetro de suavização. De modo geral, o aumento da variância que decorre da utilização de polinomiais de ordem mais elevada pode ser

compensado empregando-se um parâmetro de suavização maior. A utilização de polinômios de baixa ordem é suficiente para produzir estimativas de ótima qualidade, normalmente são utilizados polinômios com graus variando de zero a três. A escolha do grau da polinomial é, em sua maior parte, determinada pelos objetivos do trabalho e pelos dados, na prática a escolha do grau da polinômio pode ser realizada pela inspeção visual do gráfico com os dados originais e a estimativa de regressão local. De forma geral, a presença de “picos” ou “vales” nos dados são um indicativo de que d deve ser igual a dois ou três, enquanto que a presença de um padrão único indicam que d deve ser igual a um.

A função kernel

Esta função é responsável por ponderar as observações na vizinhança de cada ponto de interesse, x_0 . Segundo [Cleveland e Loader \(1996\)](#) e [Loader \(1999\)](#) esta função deve ser contínua, simétrica, com maior peso em torno de x_0 e decrescente a medida em que x se afasta de x_0 . Dentre as escolhas possíveis, destaca-se aqui a função *Tukey's tri-weight* dada por

$$\mathcal{W}(u) = \begin{cases} (1 - |u|^3)^3, & |u| \leq 1 \\ 0, & |u| > 1. \end{cases}$$

A sequência de pesos é facilmente obtida por

$$w_i(x_0) = \mathcal{W}\left(\frac{x_i - x_0}{h(x_0)}\right)$$

Assim, temos que a regressão local (LOESS) é da forma

$$\hat{\beta} = \arg \min_{\beta \in \mathbb{R}^{d+1}} \left\{ \sum_{j=1}^n w_i(x_0) \left[y_j - \sum_{k=0}^d \beta_k (x - x_0)^k \right]^2 \right\},$$

onde o espaço de parâmetros $\beta = (\beta_0, \beta_1, \dots, \beta_d) \in \mathbb{R}^{d+1}$ e $d \in \mathbb{N}$.

Muitas outras funções de Kernel ou pesos, podem ser utilizadas, ela vai depender dos objetivos do trabalho.

Span - Parâmetro de Suavização

Na prática é muito comum termos os x_i irregularmente espaçados. Se tivermos um *span* fixo h , então podemos ter estimativas locais baseadas em muitos pontos e outras são muito poucos pontos. Por esta razão, podemos considerar uma estratégia de vizinho mais próximo para definir um intervalo para cada covariável x_0 .

Defina $\Delta_i(x_0) = |x_i - x_0|$ (distância entre x_i e x_0) e seja $\Delta_{(i)}(x_0)$ os valores de tais distâncias ordenados. Um dos argumentos da função *loess* é *span*, denotado por α , que significa que para cada ajuste local utilizamos $100 \times \alpha\%$ dos dados.

Seja q a parte inteira de $\alpha \times n$. Então definimos o *span* $h(x_0) = \Delta_{(q)}(x_0)$. À medida que α aumenta, a estimativa torna-se mais suave.

Aplicação 1

Dados de economia de combustível de 1999 a 2008 para 38 modelos populares de carros (Dados: mpg do pacote *ggplot2*). Este conjunto de dados contém um subconjunto dos dados de economia de combustível que a EPA disponibiliza em <http://fuelconomy.gov>. Ele contém apenas modelos que tiveram um novo lançamento a cada ano entre 1999 e 2008 - isso foi usado como um proxy para a popularidade do carro. Variáveis no banco de dados: *manufacturer*: nome do fabricante; *model*: nome do modelo; *displ*: cilindrada do motor, em litros; *year*: ano de fabricação; *cyl*: número de

cilindros; *trans*: tipo de transmissão; *drv*: o tipo de transmissão, onde f = tração dianteira, r = tração traseira, 4 = 4wd, *cty*: rendimento na cidade (milhas/galão), *hwy*: rendimento na rodovia (milhas/galão).

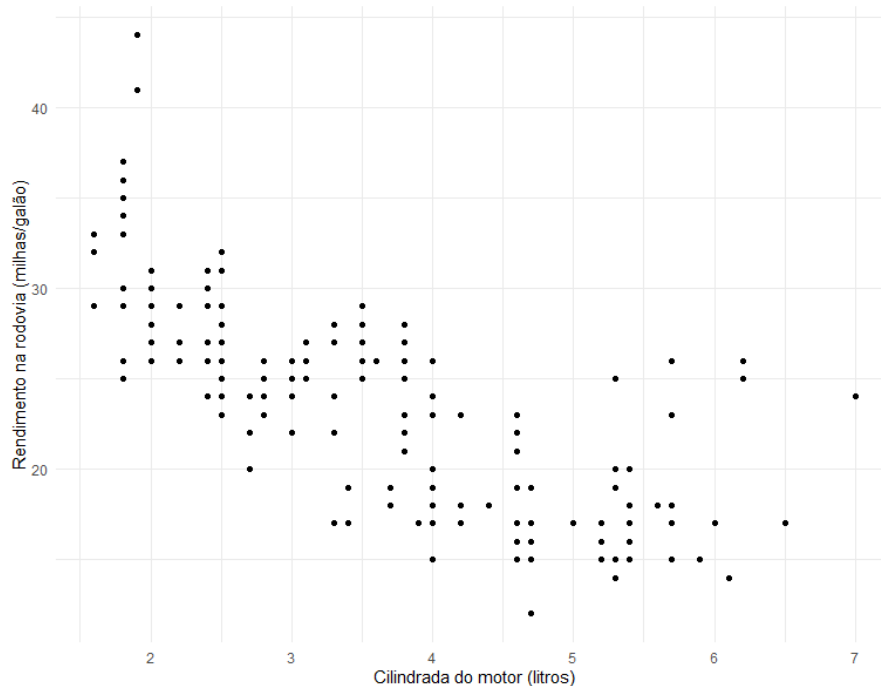


Figura 3.29: Série Temporal dos Dados de economia de combustível de 1999 a 2008 para 38 modelos populares de carros - Rendimento na rodovia (milhas/galão) por Cilindrada do motor (litros).

A Figura 5.4 apresenta a série temporal dos Dados de economia de combustível de 1999 a 2008 para 38 modelos populares de carros - Rendimento na rodovia (milhas/galão) por Cilindrada do motor (litros), enquanto a Figura 5.5 apresenta o ajuste através da função *loess* do software [R Core Team \(2020\)](#) para diferentes valores do parâmetro de suavização *span* $\alpha \in \{0.25, 0.50, 0.75, 0.90\}$.

```

1 #
2 #Regressão Local (LOESS)
3 #
4 #Dados de economia de combustível de 1999 a 2008 para 38
5 #modelos populares de carros
6 #
7 attach(mpg)
8 ggplot(mpg, aes(displ, hwy)) +
9   geom_point()+
10   labs(y = "Rendimento na rodovia (milhas/galão)", x = "Cilindrada do motor (litros)") +
11   theme_minimal()
12 #
13 #LOESS+geom_smooth
14 ggplot(mpg, aes(displ, hwy)) +
15   geom_point() +
16   geom_smooth(method="loess", span = 0.90, formula=y~x) +
17   labs(y = "Rendimento na rodovia (milhas/galão)", x = "Cilindrada do motor (litros)") +
18   theme_minimal()

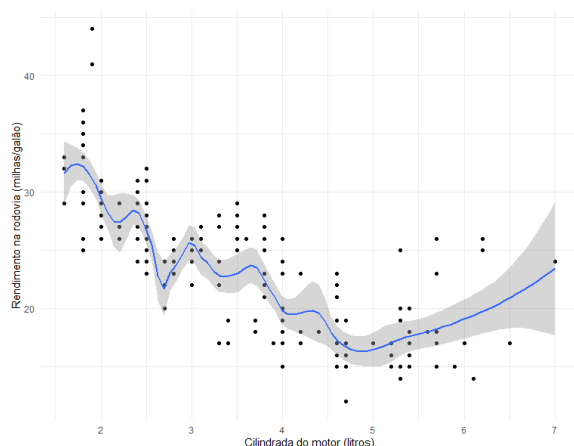
```

Código 3.25: Código fonte em R: Comandos para aplicar a função *loess* de suavização com diferentes parâmetros de suavização a Série Temporal dos Dados de economia de combustível de 1999 a 2008 para 38 modelos populares de carros - Rendimento na rodovia (milhas/galão) por Cilindrada do motor (litros).

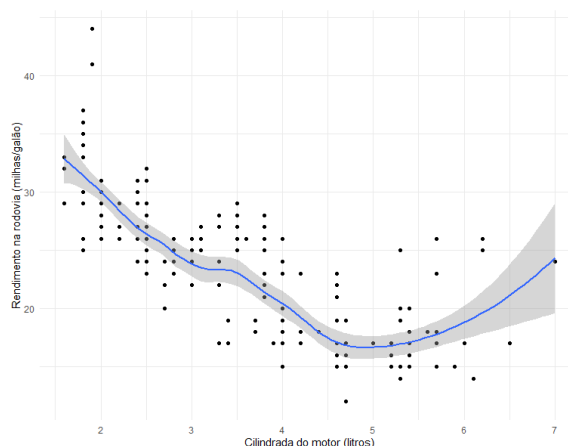
Aplicação 2

Fonte dos Dados: <https://bit.ly/37Jqtee>.

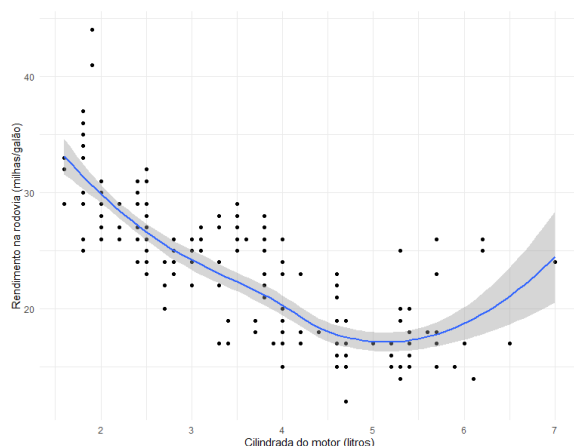
```
1 #
2 mydata <- read.csv("Example-ND.csv", sep=",")
3 attach(mydata)
4 mydata <- data.frame(x=Year, y=EMD)
5 ggplot(mydata, aes(Year, EMD)) +
6   geom_point() +
7   labs(y = "EMD", x = "Ano") + theme_minimal()
```



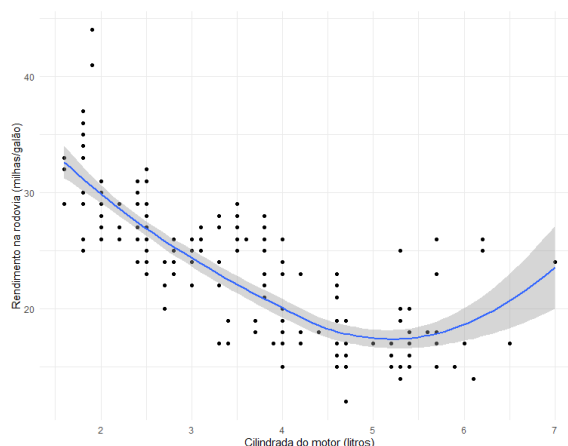
(a) $\alpha = 0.25$



(b) $\alpha = 0.50$



(c) $\alpha = 0.75$



(d) $\alpha = 0.90$

Figura 3.30: Série Temporal dos Dados de economia de combustível de 1999 a 2008 para 38 modelos populares de carros - Rendimento na rodovia (milhas/galão) por Cilindrada do motor (litros) - Aplicada a função *loess* de suavização com diferentes parâmetros de suavização.

```

8 #-----
9 #LOESS
10 y.loess<-loess(y~x,span=0.8,data=data.frame(x=Year,y=EMD),
11             control = loess.control(surface = "direct"))
12 #-----
13 #Cálculo Predição
14 tmp<-predict(y.loess,data.frame(x=Year),se=T)
15 #-----
16 #Cálculo Intervalo de Confiança
17 li<-tmp$fit - qt(0.975,tmp$df)*tmp$se.fit
18 ls<-tmp$fit + qt(0.975,tmp$df)*tmp$se.fit
19 #-----
20 #Gráfico (1)
21 y.predict<-data.frame(x=Year,y=EMD,fit=tmp$fit,li=li,ls=ls)
22 ggplot(y.predict,aes(Year,EMD))+
23   geom_point()+
24   geom_line(aes(Year,fit),colour='red')+
25   geom_line(aes(Year,li),colour='blue')+
26   geom_line(aes(Year,ls),colour='blue')+
27   labs(y="EMD",x="Ano")+theme_minimal()
28 #-----
29 #Gráfico (2)
30 df1<-data.frame(rep("Dados",length(mydata$x)),x=mydata$x,y=mydata$y)
31 colnames(df1)<-c("V1","V2","V3")
32 df2<-data.frame(rep("Fitted",length(mydata$x)),x=mydata$x,y=tmp$fit)
33 colnames(df2)<-c("V1","V2","V3")
34 df3<-data.frame(rep("LI",length(mydata$x)),x=mydata$x,y=li)
35 colnames(df3)<-c("V1","V2","V3")
36 df4<-data.frame(rep("LS",length(mydata$x)),x=mydata$x,y=ls)
37 colnames(df4)<-c("V1","V2","V3")
38 y.predict<-data.frame(rbind(df1,df2,df3,df4))
39 colnames(y.predict)<-c("Legenda","V2","V3")
40 ggplot(y.predict,aes(V2,V3,col=Legenda))+
41   geom_point(size=0.5)+
42   scale_color_manual(values=c("Dados"="black","Fitted"="red","LI"="blue3","LS"="
43     blue3"))+
44   labs(y="EMD",x="Ano",title="Predição-Loess")+theme_minimal()
45 #-----
46 #Cálculo Previsão
47 a<-rep(1/12,12)
48 ano<-c(rep(2005,9)+cumsum(a)[4:12],rep(2006,12)+cumsum(a))
49 tmp2<-predict(y.loess,data.frame(x=c(Year,ano)),se=T)
50 #-----
51 #Cálculo Intervalo de Confiança
52 li<-tmp2$fit - qt(0.975,tmp2$df)*tmp2$se.fit
53 ls<-tmp2$fit + qt(0.975,tmp2$df)*tmp2$se.fit
54 #-----
55 #Gráfico (3)
56 df1<-data.frame(rep("Dados",length(mydata$x)),x=mydata$x,y=mydata$y)
57 colnames(df1)<-c("V1","V2","V3")
58 df2<-data.frame(rep("Fitted",length(c(Year,ano))),x=c(Year,ano),y=tmp2$fit)
59 colnames(df2)<-c("V1","V2","V3")
60 df3<-data.frame(rep("LI",length(c(Year,ano))),x=c(Year,ano),y=li)
61 colnames(df3)<-c("V1","V2","V3")
62 df4<-data.frame(rep("LS",length(c(Year,ano))),x=c(Year,ano),y=ls)
63 colnames(df4)<-c("V1","V2","V3")
64 y.predict<-data.frame(rbind(df1,df2,df3,df4))
65 colnames(y.predict)<-c("Legenda","V2","V3")
66 ggplot(y.predict,aes(V2,V3,col=Legenda))+
67   geom_point(size=0.5)+
68   scale_color_manual(values=c("Dados"="black","Fitted"="red","LI"="blue3","LS"="
69     blue3"))+

```

```
68 | labs(y = "EMD", x = "Ano", title="Predição-Loess")+theme_minimal()
69 | #
```

Código 3.26: Código fonte em R: Comandos para aplicar a função *loess* de suavização com diferentes parâmetros de suavização a Série Temporal <http://doylesdartden.com/Example-ND.csv>.

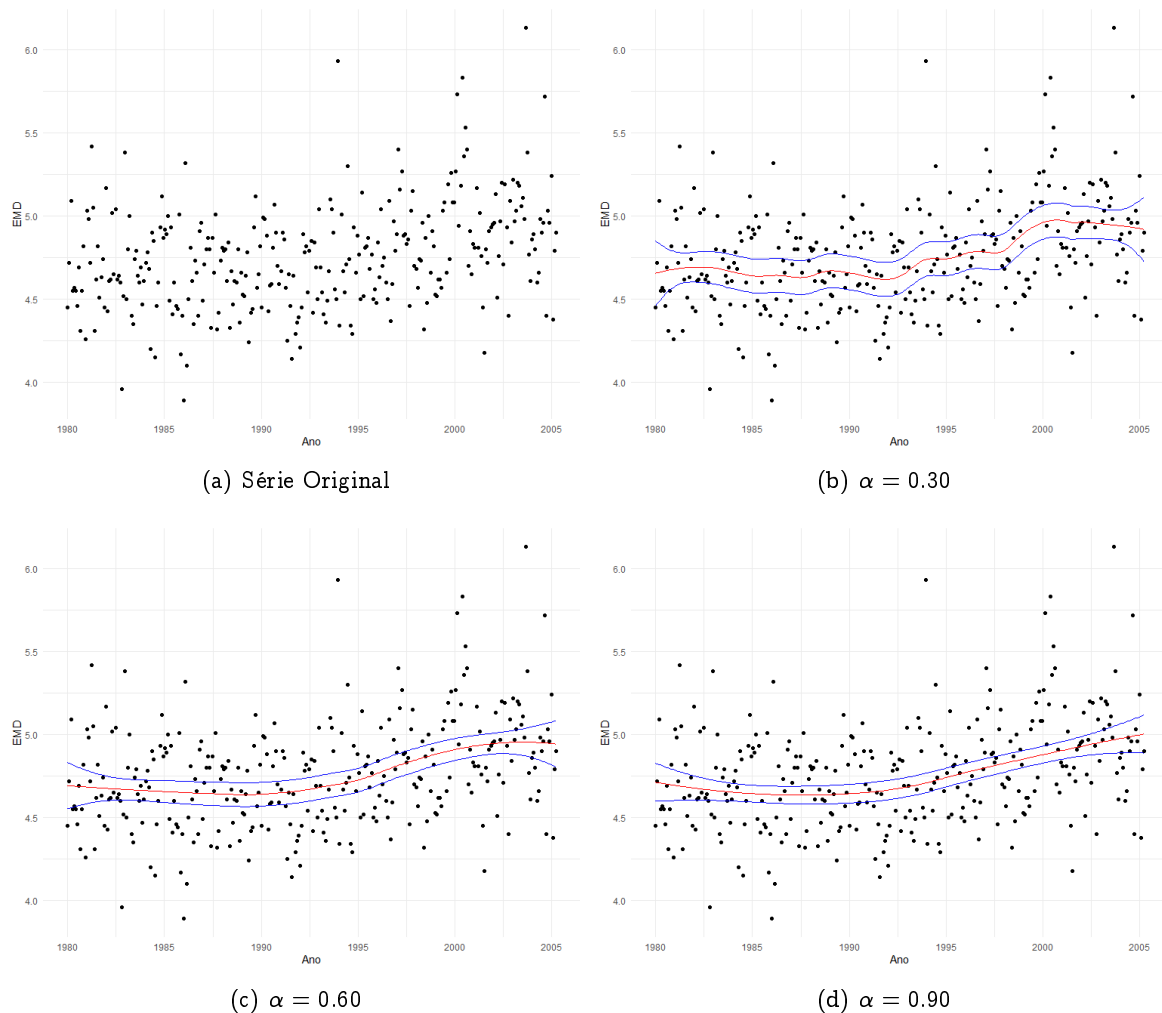


Figura 3.31: Gráfico da Série Temporal original e da predição, utilizando o método *loess* para diferentes valores do parâmetro de suavização $span \alpha \in \{0.30, 0.60, 0.90\}$

A Figura 3.31 apresenta a série temporal dos dados originais e do ajuste através da função *loess* (predição) do software R Core Team (2020) para diferentes valores do parâmetro de suavização $span \alpha \in \{0.30, 0.60, 0.90\}$. Os intervalos de confiança de predição e previsão são calculados utilizando a distribuição *t* com coeficiente de confiança de 95%.

A Figura 3.32 apresenta a predição e previsão da série temporal original utilizando o método *loess* para diferentes valores do parâmetro de suavização $span \alpha \in \{0.30, 0.60, 0.90\}$.

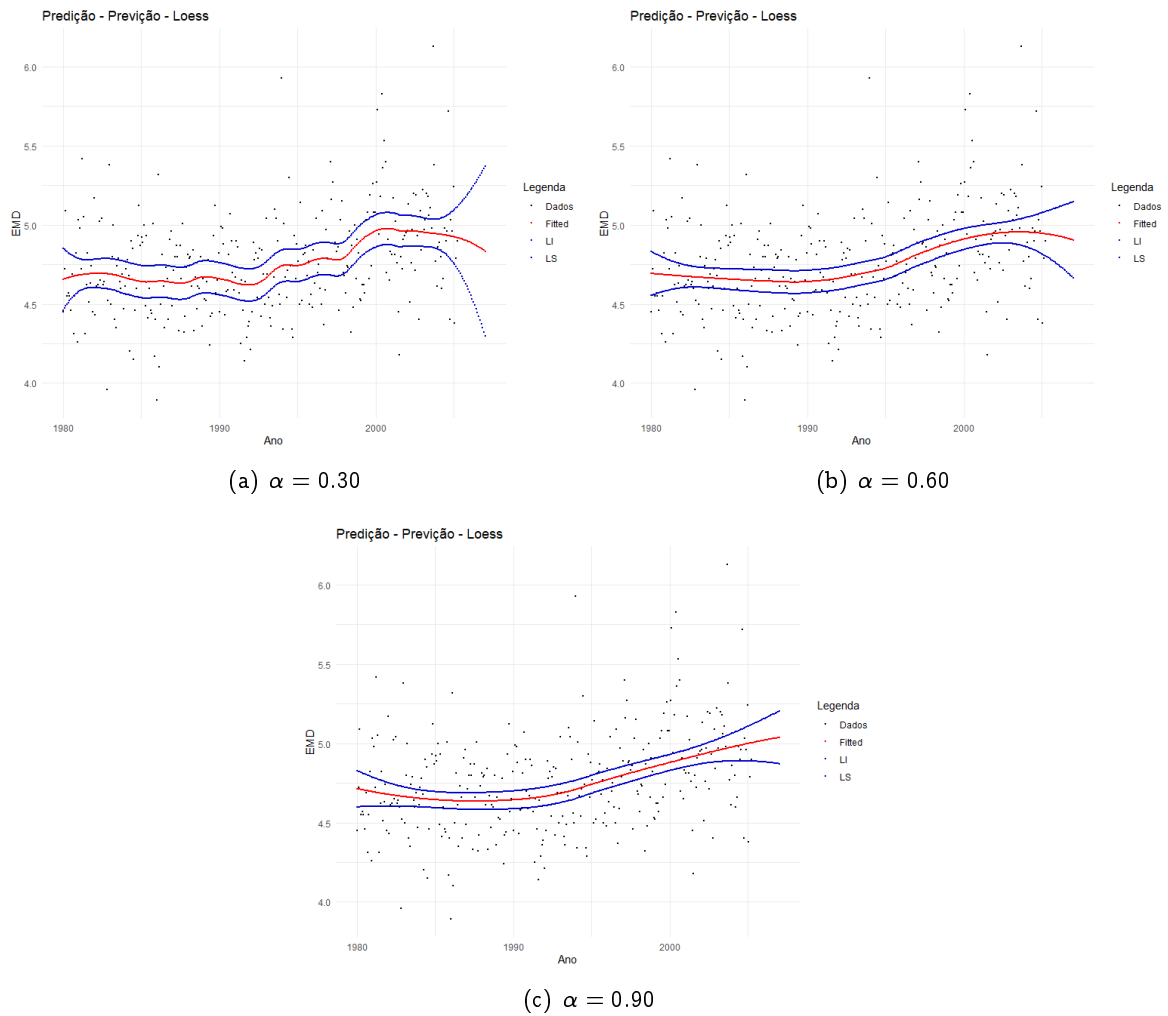


Figura 3.32: Predição e Previsão da série temporal original, utilizando o método *loess* para diferentes valores do parâmetro de suavização $span \alpha \in \{0.30, 0.60, 0.90\}$

Capítulo 4

Análise de Sazonalidade

4.1 Testes para Sazonalidade

[<https://bit.ly/3rB9xzH>] Uma série temporal é sazonal (periódica) quando os fenômenos se repetem a cada período idêntico de tempo - por exemplo, fenômenos que ocorrem diariamente em uma certa hora, todos os dias, ou em um certo mês em todos os anos. A sazonalidade determinística pressupõe um padrão sazonal regular e estável no tempo, o que permite prever o comportamento sazonal perfeitamente a partir de dados anteriores - modelos de regressão como o MMQ são muito aplicados nesse tipo de série. Já a sazonalidade estocástica é quando a componente sazonal da série varia com o tempo. As seções a seguir foram retiradas de [Morettin e Toloi \(2006\)](#).

4.1.1 Teste de Kruskal-Wallis

O teste de Kruskal-Wallis é um método não paramétrico para testar se as amostras são retiradas da mesma população. A significância do efeito do mês (ou trimestre) é testada. O teste de Friedman não requer suposições de distribuição. Ele usa as classificações das observações. Se a hipótese nula de nenhuma sazonalidade estável for rejeitada ao nível de significância, a série é considerada sazonal.

Tabela 4.1: Observações mensais de uma série temporal com k anos

Anos	Meses				
	1	2	3	...	12
1	$X_{1,1}$	$X_{1,2}$	$X_{1,3}$...	$X_{1,12}$
2	$X_{2,1}$	$X_{2,2}$	$X_{2,3}$...	$X_{2,12}$
\vdots	\vdots	\vdots	\vdots		\vdots
p	$X_{k,1}$	$X_{k,2}$	$X_{k,3}$...	$X_{k,12}$

A estatística de teste é construída da seguinte maneira. Considere primeiro a matriz de dados $\{X_{ij}\}_{n_j \times k}$ (ver Tabela 4.1) onde supomos cada coluna uma amostra de uma população, ou seja, temos k (igual a 12) amostras, de tamanho n_j (iguais a p , pela Tabela 4.1), ou seja as observações são denotadas por

$$X_{ij}, \quad j = 1, \dots, k, \quad i = 1, \dots, n_j, \quad n = \sum_{j=1}^k n_j.$$

As observações X_{ij} são substituídas pelos seus postos R_{ij} , obtidos ordenando-se todas as n obser-

vações. Definimos $R_{.j}$ a soma dos postos associados à j -ésima amostra (coluna),

$$R_{.j} = \sum_{i=1}^{n_j} R_{ij}, \quad j = 1, \dots, k.$$

Portanto a estatística do teste é dada por

$$T = \frac{12}{n(n+1)} \sum_{j=1}^k \frac{R_{.j}^2}{n_j} - 3(n+1).$$

As hipóteses do teste de Kruskal-Wallis são

$$H_0 : \text{Série Temporal não Sazonal} \quad \text{vs} \quad H_1 : \text{Série Temporal Sazonal}$$

Rejeitamos H_0 se $\mathbb{P}_{H_0}(T > t_c) = \alpha$, onde α é o nível de significância do teste e t_c é o valor crítico. Para n_j suficientemente grande, ou $k \geq 4$, sob H_0 , a distribuição de T pode ser aproximada por uma χ_{k-1}^2 .

Uma crítica ao teste de Kruskal-Wallis é que uma de suas suposições diz que as variáveis dentro de cada amostra são independentes e que as amostras são independentes entre si, o que não ocorre em séries temporais, pois há dependência entre observações de um mesmo mês, para diferentes anos, e entre observações de vários meses, dentro de um mesmo ano.

Maiores detalhes sobre o Teste de Kruskal-Wallis ver [Kruskal e Wallis \(1952\)](#) e [Hyndman e Khandakar \(2008\)](#).

4.1.2 Teste de Friedman

Outra opção é utilizar o Teste de Friedman para amostras relacionadas. Neste caso, os meses são considerados “tratamentos” e os anos são considerados “blocos”. A ordenação é feita dentro de cada bloco ao invés de ordenar todas as n observações. Mas mesmo aqui, os blocos são considerados independentes, ou seja, as observações de um ano são consideradas independentes das observações de outro ano qualquer. A estatística do Teste de Friedman é dada por

$$T = \frac{12}{pk(k+1)} \sum_{j=1}^k (R_{.j}^{*2} - 3p(k+1)),$$

onde p é o número de blocos (anos), k é o número de tratamentos (12) e $R_{.j}^*$ denota a soma dos postos da j -ésima coluna, isto é,

$$R_{.j}^* = \sum_{i=1}^p R_{ij}^*,$$

com R_{ij}^* o posto X_{ij} dentro do bloco i , para $i = 1, \dots, k$. A distribuição de T pode ser aproximada por uma χ_{k-1}^2 .

Maiores detalhes sobre o Teste de Friedman ver [Friedman \(1937\)](#) e [Hyndman e Khandakar \(2008\)](#).

4.1.3 Teste de Autocorrelação em lags Sazonais

[<http://bit.ly/3aVfC3Q>] O teste QS é uma variante do teste Ljung-Box (ver seção ??) calculado em defasagens sazonais, onde apenas consideramos autocorrelações positivas

$$QS = n(n+2) \sum_{i=1}^k \frac{[\max(0, \hat{\gamma}_{i \cdot l})]^2}{n - i \cdot l}$$

onde $k = 2$, então apenas a primeira e a segunda defasagens sazonais são consideradas. Assim, o teste verificaria a correlação entre a observação real e as observações defasadas em um e dois anos. Observe que $l = 12$ ao lidar com observações mensais, então consideramos as autocovariâncias $\hat{\gamma}_X(12)$ e $\hat{\gamma}_X(24)$ sozinhas. Por sua vez, $k = 4$ no caso de dados trimestrais.

Sob H_0 , que afirma que os dados são distribuídos de forma independente, as estatísticas seguem uma distribuição $\chi^2(k)$. No entanto, a eliminação de correlações negativas torna uma aproximação ruim. Os p-valores seriam dados por $P(\chi^2(k) > Q)$, se $k = 2$. Como $P(\chi^2(2) > 0.05) = 5.99146$ e $P(\chi^2(2) > 0.01) = 9.21034$, $QS > 5.99146$ e $QS > 9.21034$ sugeriria rejeitar a hipótese nula em níveis de significância de 95% e 99%, respectivamente.

Correções

Maravall (2012) propõe aproximar a distribuição correta (p-valores) da estatística QS usando técnicas de simulação. Usando 1000K replicações de tamanho de amostra 240, os valores críticos corretos seriam 3,83 e 7,09 com níveis de confiança de 95% e 99%, respectivamente (inferior a 5,99146 e 9,21034 mostrados acima). Para cada uma das séries simuladas, ele obtém a distribuição assumindo $QS = 0$ quando $\hat{\gamma}_X(12)$, portanto, na prática, este teste detectará sazonalidade apenas quando qualquer uma dessas condições se mantiver:

- (i) Autocorrelação positiva estatisticamente significativa no lag 12
- (ii) Autocorrelação de amostra não negativa no lag 12 e autocorrelação positiva estatisticamente significativa no lag 24

Maiores detalhes sobre o Teste de Autocorrelação em *lags* Sazonais ver Maravall (2011) e Hyndman e Khandakar (2008).

4.1.4 Aplicação

Para aplicarmos os testes para verificarmos se uma série temporal apresenta sazonalidade utilizamos as rotinas do pacote `seastests` do R Core Team (2020) citadas a seguir

- (i) Teste de Kruskal-Wallis - rotina `kw()`
- (ii) Teste de Friedman - rotina `fried()`
- (iii) Teste de Autocorrelação em *lags* Sazonais - rotina `qs()`

No pacote `seastests` [<https://bit.ly/2L4MgoW>] existem outros testes de sazonalidade que podem ser aplicados a séries temporais.

```

1 library(seastests)
2 #
3 #Série temporal: AirPassengers
4 #
5 autoplot(AirPassengers)+theme_minimal()
6 ggAcf(AirPassengers, lag.max=100, type = c("correlation"))+labs(y = "FAC Amostral", title
7   = "")+
8   theme_minimal()
9 kw(AirPassengers, freq = 12, diff = T, residuals = F, autoarima = T)
10 fried(AirPassengers, freq = 12, diff = T, residuals = F, autoarima = T)
11 qs(AirPassengers, freq = NA, diff = T, residuals = F, autoarima = T)
12 #
13 #Série temporal: Australian beer production
14 #

```

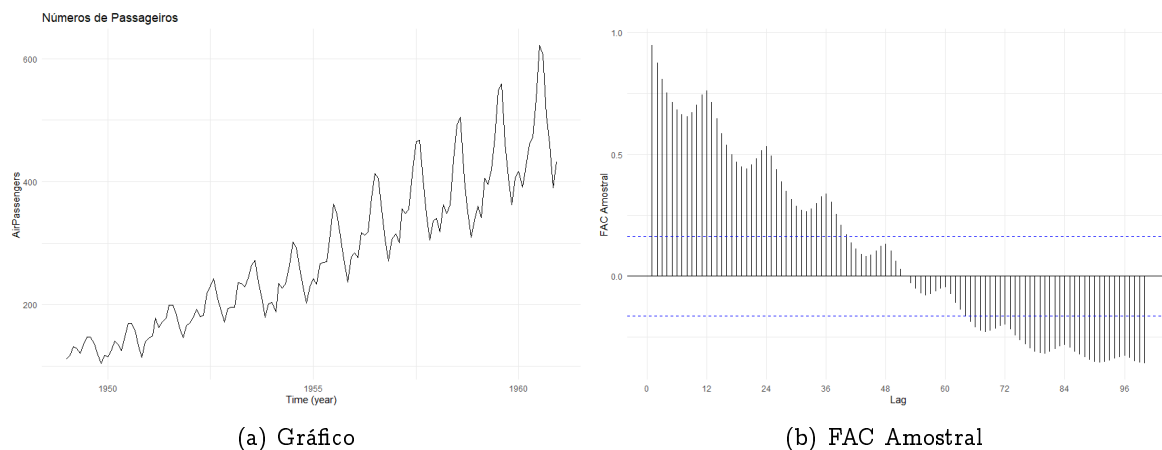


Figura 4.1: Gráficos e Função de Autocorrelação da Série Temporal do número de passageiros internacionais por mês em uma companhia aérea (Pan Am) nos Estados Unidos do período de 1946-1960

```

14 #Total quarterly beer production in Australia (in megalitres) from 1956:Q1 to 2008:Q3.
15 #library(fpp)
16 data(ausbeer)
17 autoplot(ausbeer, ylab = "") + labs(title = "Quarterly Australian Beer Production", x = "
  Tempo") + theme_minimal()
18 ggAcf(ausbeer, lag.max = 50) + labs(title = "Quarterly Australian Beer Production", y = "FAC
  Amostral") + theme_minimal()
19 kw(ausbeer, freq = 12, diff = T, residuals = F, autoarima = T)
20 fried(ausbeer, freq = 12, diff = T, residuals = F, autoarima = T)
21 qs(ausbeer, freq = 12, diff = T, residuals = F, autoarima = T)

```

Código 4.1: Código fonte em R: Funções para verificar sazonalidade utilizando do pacote `seastests`.

4.2 Decomposição

Um modelo para uma série temporal $\{X_t\}_{t=1}^n$ supõe que a mesma possa ser escrita como a soma de 3 componentes: tendência, sazonalidade e termo aleatório, ou seja,

$$X_t = T_t + S_t + \varepsilon_t, \quad t = 1, \dots, n.$$

```

1 #Série Temporal Simulada
2 A<-rnorm(n=1, mean=0, sd=1)
3 B<-rnorm(n=1, mean=0, sd=1)
4 theta<-2
5 n<-50#Tamanho da Série Temporal
6 t<-c(1:n)
7 xt<-A*cos(theta*t)+B*sin(theta*t)
8 y<-0.2*t
9 yt<-xt+y
10 yt<-ts(yt, start = c(2016, 1), frequency=12)
11 autoplot(yt) + labs(x = "Tempo", y = "Série Temporal") + theme_minimal()

```

Código 4.2: Código fonte em R: Série Temporal Simulada com tendência determinística linear e sazonalidade.

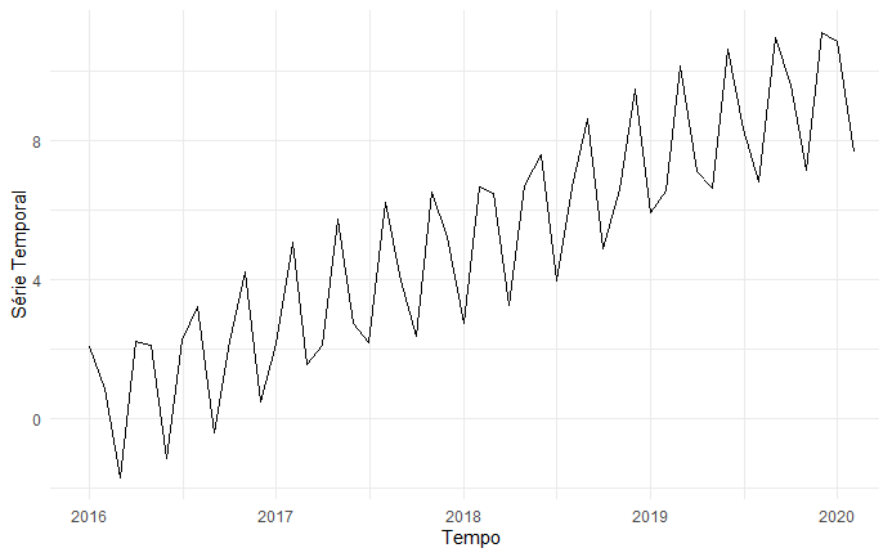


Figura 4.2: Série Temporal Simulada com tendência determinística linear e sazonalidade.

Observando a série temporal da Figura 4.2 notamos que há um aumento gradual das observações. Dizemos que a série apresenta uma tendência (T_t) (no caso, crescente). Pode haver um declínio, em outros casos.

A componente sazonal (S_t) aparece quando as observações são intra-anuais, isto é, registradas mensalmente, trimestralmente ou semanalmente, por exemplo. A série no exemplo apresenta periodicidade de 12 meses.

Comumente, inclui-se no modelo uma componente cíclica, para representar movimentos com longos períodos, geralmente maiores que um ano.

Removendo-se as componentes T_t e S_t o que sobra é a componente aleatória ou resíduo ε_t . A suposição mais clássica é que ε_t decorra de um processo estocástico puramente aleatório ($\{\varepsilon_t\}_{t \in \mathbb{Z}}$ Ruído Branco - $RB(0, \sigma_\varepsilon^2)$), se bem que em alguns casos podemos considerá-lo um processo estacionário, digamos com média zero e variância constante.

O modelo *aditivo* é adequado quando S_t não depende das outras componentes, como T_t . Se as amplitudes sazonais variam com a tendência, um modelo mais adequado é o modelo *multiplicativo*.

$$X_t = T_t \times S_t \times \varepsilon_t, \quad t = 1, \dots, n. \quad (4.1)$$

O modelo multiplicativo pode ser transformado num aditivo tomando-se o logaritmo do modelo multiplicativo, ou seja,

$$\log(X_t) = \log(T_t) + \log(S_t) + \log(\varepsilon_t), \quad t = 1, \dots, n. \quad (4.2)$$

O problema é modelar convenientemente as 3 componentes T_t , S_t e ε_t , a fim de fazer previsão dos futuros valores da série temporal.

```
1 icv<-c(238,251,256,263,270,275,280,290,298,305,310,318,
2       329,343,359,375,383,393,400,407,415,424,436,449)
3 icv<-ts(icv, start=c(1976,1), frequency=12)
4 autoplot(icv)+labs(x = "Tempo", y = "Índice de Custo de Vida")+theme_minimal()
```

Código 4.3: Código fonte em R: Índice de Custo de Vida no Município de São Paulo. Fonte: Fundação instituto de Pesquisas Economicas (FIPE).

A Figura 4.3 apresenta a série temporal do Índice de Custo de Vida no Município de São Paulo. Fonte: Fundação instituto de Pesquisas Econômicas (FIPE).

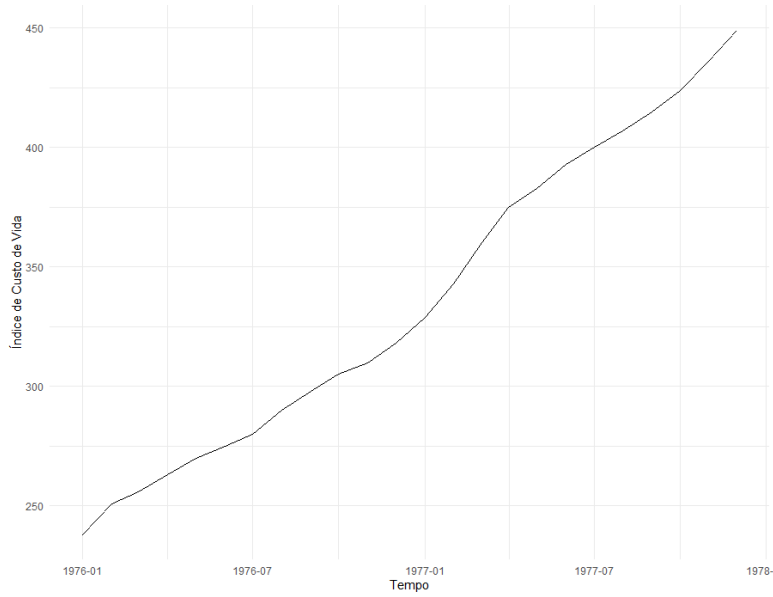


Figura 4.3: Índice de Custo de Vida no Município de São Paulo. Fonte: Fundação Instituto de Pesquisas Econômicas (FIPE).

4.2.1 Modelo Aditivo

Considere uma série temporal $\{X_t\}_{t=1}^n$. Vimos que um modelo de decomposição consiste em escrever X_t como a soma de três componentes não observáveis

$$X_t = T_t + S_t + \varepsilon_t, \quad t = 1, \dots, n.$$

onde T_t e S_t representam a tendência e a sazonalidade, respectivamente e ε_t a componente aleatória, de média zero e variância constante σ_ε^2 . Se $\{\varepsilon_t\}_{t=1}^n$ são observações de um ruído branco, então $\mathbb{E}(\varepsilon_t \varepsilon_s) = 0$, para $s \neq t$, mas podemos, eventualmente, relaxar esta suposição, tomando $\{\varepsilon_t\}_{t \in \mathbb{Z}}$ um processo estacionário.

4.2.2 Modelo Multiplicativo

O modelo *aditivo* é adequado quando S_t não depende das outras componentes, como T_t . Se as amplitudes sazonais variam com a tendência, um modelo mais adequado é o modelo *multiplicativo*.

Considere uma série temporal $\{X_t\}_{t=1}^n$. Vimos que um modelo de decomposição consiste em escrever X_t como a soma de três componentes não observáveis

$$X_t = T_t \times S_t \times \varepsilon_t, \quad t = 1, \dots, n. \quad (4.3)$$

onde T_t e S_t representam a tendência e a sazonalidade, respectivamente e ε_t a componente aleatória, de média zero e variância constante σ_ε^2 . Se $\{\varepsilon_t\}_{t=1}^n$ são observações de um ruído branco, então $\mathbb{E}(\varepsilon_t \varepsilon_s) = 0$, para $s \neq t$, mas podemos, eventualmente, relaxar esta suposição, tomando $\{\varepsilon_t\}_{t \in \mathbb{Z}}$ um processo estacionário.

O modelo multiplicativo pode ser transformado num aditivo tomando-se o logaritmo do modelo multiplicativo, ou seja,

$$\log(X_t) = \log(T_t) + \log(S_t) + \log(\varepsilon_t), \quad t = 1, \dots, n. \quad (4.4)$$

O problema é modelar convenientemente as 3 componentes T_t , S_t e ε_t , a fim de fazer previsão dos futuros valores da série temporal.

Aplicação

Decomposição da série temporal do número de passageiros internacionais por mês em uma companhia aérea (Pan Am) nos Estados Unidos foi obtido da Administração da Aviação Federal para o período 1946-1960, utilizando a função *decompose* do [R Core Team \(2020\)](#).

```

1 #
2 #Gráficos – Decomposição de Séries Temporais
3 #Modelo Aditivo
4 autoplot(decompose(AirPassengers,type = c("additive")))+theme_minimal()
5 #Modelo Multiplicativo
6 autoplot(decompose(AirPassengers,type = c("multiplicative")))+theme_minimal() #
  Additive
7 #
8 #Análise parte aleatória – Decomposição de Séries Temporais
9 mod_ad<-decompose(AirPassengers,type = c("additive"))
10 mod_mt<-decompose(AirPassengers,type = c("multiplicative"))
11 #
12 #Resíduos – Parte Aleatória
13 res_ad<-mod_ad$random[7:138]
14 res_mt<-mod_mt$random[7:138]
15 #
16 #FAC dos Resíduos
17 ggAcf(res_ad, lag.max=100,type = c("correlation"))+labs(y = "FAC Amostral Resíduos
  Aditivos",title="")+
18   theme_minimal()
19 ggAcf(res_mt, lag.max=100,type = c("correlation"))+labs(y = "FAC Amostral Resíduos
  Multiplicativos",title="")+
20   theme_minimal()
21 #
22 #Soma FAC dos Resíduos
23 fac_ad<-sum(abs(acf(res_ad, lag.max=100,type = c("correlation"))$acf))
24 fac_mt<-sum(abs(acf(res_mt, lag.max=100,type = c("correlation"))$acf))
25 #
26 #QQ Plot dos Resíduos
27 ggqqplot(res_ad)+ggtitle("Parte Aleatória – Modelo Aditivo")
28 ggqqplot(res_mt)+ggtitle("Parte Aleatória – Modelo Multiplicativo")
29 #
30 #Densidade dos Resíduos
31 plot(density(res_ad),main="Random Error")
32 plot(density(res_mt),main="Random Error")
33 #
34 #Teste de Normalidade dos Resíduos
35 shapiro.test(res_ad)
36 shapiro.test(res_mt)
37 #

```

Código 4.4: Código fonte em R: Decomposição da Série Temporal do número de passageiros internacionais por mês em uma companhia aérea (Pan Am) nos Estados Unidos - período 1946-1960.

As Figuras 4.4 e 4.5 mostram a decomposição da série temporal em tendência, sazonalidade e parte aleatória (ruído). É claro que a série temporal não é estacionária (possui passeios aleatórios) devido a efeitos sazonais e uma tendência (tendência linear). Cabe ressaltar que a função *decompose* remove a tendência da série temporal através de médias móveis.

Como próximo passo iremos analisar a parte aleatória (resíduos) das decomposições. Seguindo [Werner et al. \(2017\)](#), como ambas as sazonalidade foram calculadas utilizando a rotina *decompose*, do [R Core Team \(2020\)](#), podemos analisar os resíduos de ambos os modelos. Obtemos que a soma dos quadrados das autocorrelações amostrais, para 100 lags, é 19,14231, para o modelo aditivo, e 11,76071, para o modelo multiplicativo. Desta forma percebemos uma pequena vantagem em utilizarmos o modelo multiplicativo.

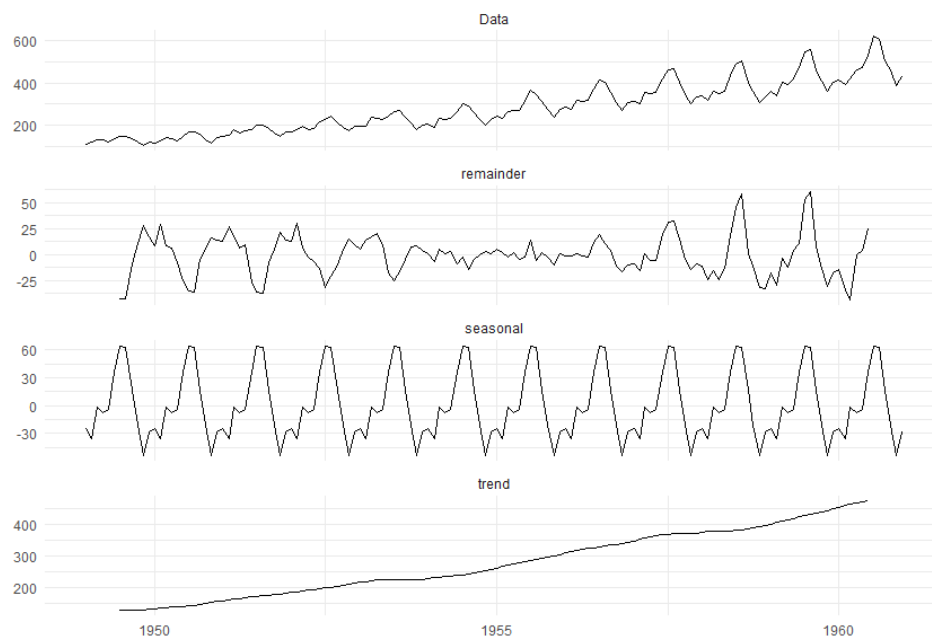


Figura 4.4: Decomposição da Série Temporal do número de passageiros internacionais por mês em uma companhia aérea (Pan Am) nos Estados Unidos do período de 1946-1960 - Modelo Aditivo.

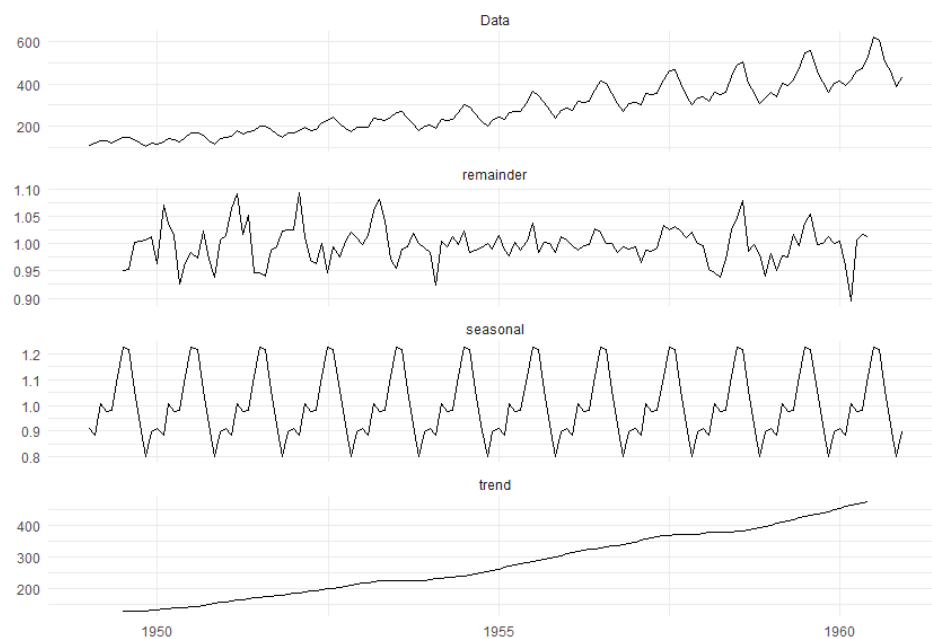


Figura 4.5: Decomposição da Série Temporal do número de passageiros internacionais por mês em uma companhia aérea (Pan Am) nos Estados Unidos do período de 1946-1960 - Modelo Multiplicativo.

A Figura 4.6 apresenta as funções de autocorrelações amostrais dos resíduos das decomposições da série temporal em tendência, sazonalidade e parte aleatória (resíduo).

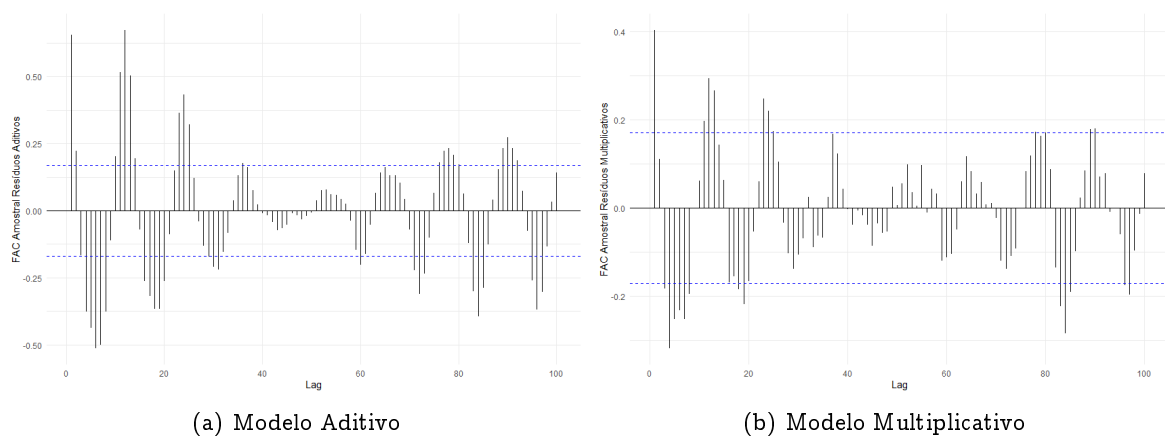


Figura 4.6: Funções de Autocorrelação Amostrais dos resíduos das decomposições da série temporal.

Capítulo 5

Modelos de Suavização

Este capítulo tem como base o livro de Hyndman e Athanasopoulos (2018): <https://otexts.com/fpp2/>

Muitos fenômenos reais possuem a característica de repetir-se durante um certo período de tempo fixo. A este comportamento denominamos *sazonalidade*.

Dizemos que uma série temporal é sazonal quando os fenômenos que ocorrem durante o tempo se repete a cada período idêntico de tempo, ou seja, fenômenos que ocorrem diariamente em uma certa hora, todos os dias, ou em um certo mês em todos os anos. Um exemplo fácil de visualizar seria o aumento das vendas de passagens aéreas todos os finais de ano, ou ainda, o aumento das vendas do comércio no período do natal.

```
1 #  
2 library(readxl)  
3 library(ggplot2)  
4 library(ggfortify)  
5 library(forecast)  
6 library(plotly)  
7 library(magrittr)
```

Código 5.1: Código fonte em R: Pacotes usados neste capítulo.

5.1 Introdução

Para entendermos a componente sazonal de uma série temporal, iniciamos com uma pré análise dos dados clássicos das companhias aéreas, que documentam o total mensal de passageiros de companhias aéreas internacionais de 1949 a 1960.

O número de passageiros internacionais por mês em uma companhia aérea (Pan Am) nos Estados Unidos foi obtido da Administração da Aviação Federal para o período 1946-1960. A empresa usou os dados para prever a demanda futura antes de encomendar novas aeronaves e treinar tripulação. Os dados estão disponíveis como séries temporais no R Core Team (2020) e são denominados *AirPassengers*.

```
1 #  
2 data('AirPassengers')  
3 AirPassengers  
4 autoplot(AirPassengers)+ggtitle("Números de Passageiros") +  
5   xlab("Time (year)") + ylab("AirPassengers")+theme_minimal()  
6 #  
7 ggAcf(AirPassengers, lag.max=100,type = c("correlation"))+labs(y = "FAC Amostral",title  
8   = "")+  
9   theme_minimal()
```

```

9 ggAcf(AirPassengers, lag.max=100, type = c("partial")) + labs(y = "FACP Amostral", title = "")
10   ) +
11   theme_minimal()
12 ggseasonplot(AirPassengers, year.labels=TRUE, continuous=TRUE) + theme_minimal()
13 #
14 ggseasonplot(AirPassengers, year.labels=FALSE, continuous=TRUE, polar = TRUE) + theme_
15   minimal()
16 #
17 ggsubseriesplot(AirPassengers) + theme_minimal()
18 #
19 boxplot(AirPassengers ~ cycle(AirPassengers))
20 #
21 AirPassengers.df <- tsdf(AirPassengers)
22 ggplot(AirPassengers.df, aes(x = x, y = y)) +
23   geom_line(colour = "grey75") +
24   stat_rollapplyr(width = 12, align = "center", geom = "point",
25                   size = 0.5, colour = "steelblue") +
26   labs(x = "", y = "Número de Passageiros - USA\n(média móvel e série original)") +
27   theme_minimal()
28 #

```

Código 5.2: Código fonte em R: Série Temporal do número de passageiros internacionais por mês em uma companhia aérea (Pan Am) nos Estados Unidos - período 1946-1960.

A Figura 5.1 apresenta o gráfico da Série Temporal do número de passageiros internacionais por mês em uma companhia aérea (Pan Am) nos Estados Unidos do período de 1946-1960.

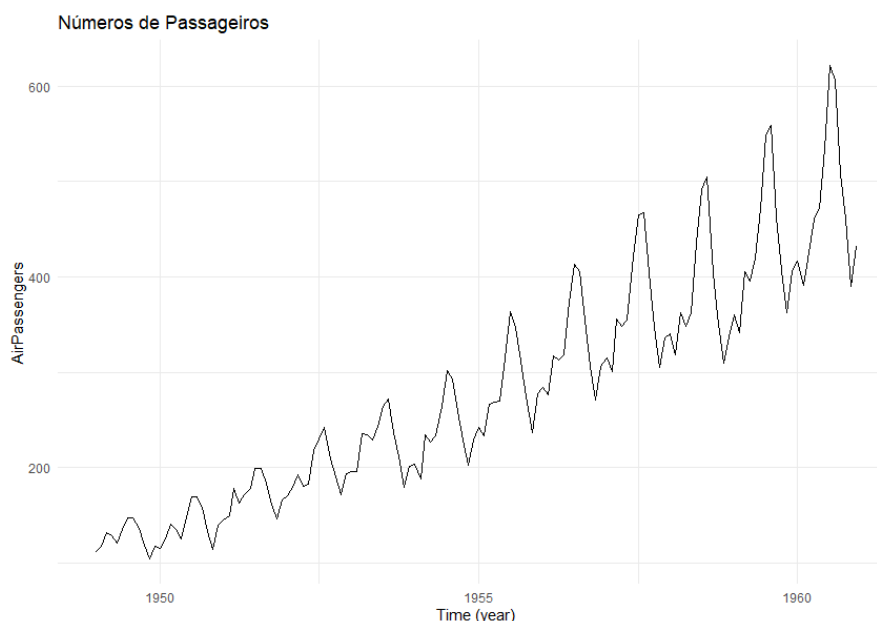


Figura 5.1: Série Temporal do número de passageiros internacionais por mês em uma companhia aérea (Pan Am) nos Estados Unidos do período de 1946-1960

A Figura 5.2 apresenta os gráficos das Funções de Autocorrelação e Autocorrelação Parcial da séries temporal *AirPassengers*.

Existem algumas maneiras úteis de plotar dados para enfatizar padrões sazonais e mostrar alterações nesses padrões ao longo do tempo. Primeiro, um gráfico sazonal é semelhante ao gráfico de tempo, exceto que os dados são plotados em relação às estações individuais em que os dados foram observados. Um gráfico sazonal permite que o padrão sazonal subjacente seja visto com mais clareza

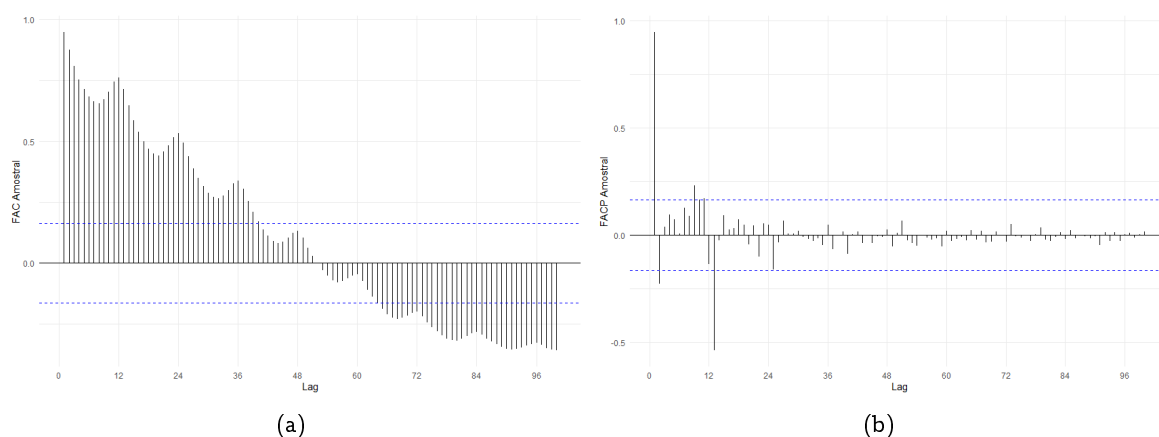


Figura 5.2: Funções de Autocorrelação e Autocorrelação Parcial da séries temporal *AirPassengers*.

e pode ser útil na identificação de anos em que o padrão muda (ver Figura 5.3). Aqui, vemos que o número de passageiros aumentou consistentemente ao longo dos anos, pois as linhas mais baixas (mais escuras) representam anos anteriores e as linhas mais altas (mais claras) representam os anos recentes. Além disso, vemos que ao número de passageiros tende a ser a mais baixa nos meses de janeiro a maio, em geral, atinge o pico nos meses de junho a agosto de se estabilizar ou diminuir levemente nos meses de setembro a dezembro.

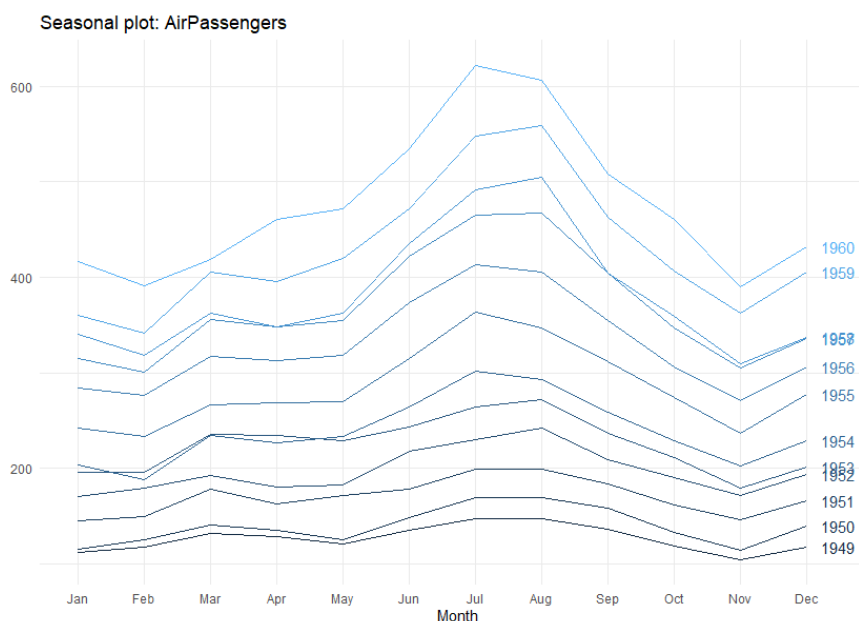


Figura 5.3: Série Temporal do número de passageiros internacionais por mês em uma companhia aérea (Pan Am) nos Estados Unidos do período de 1946-1960 - Gráfico por período anual

Uma variante útil específica de um gráfico de estação usa coordenadas polares (ver Figura 5.4), onde o eixo do tempo é circular e não horizontal. Aqui, plotamos a série temporal *AirPassengers* com o gráfico sazonal convencional (acima) versus uma opção de coordenadas polares para ilustrar essa variante. Ambas as parcelas ilustram o mesmo comportamento apresentado anteriormente.

Um gráfico alternativo que enfatiza os padrões sazonais é onde os dados de cada estação (mensal no nosso exemplo) são coletados em mini gráficos de tempo separados. Um gráfico de sub-séries

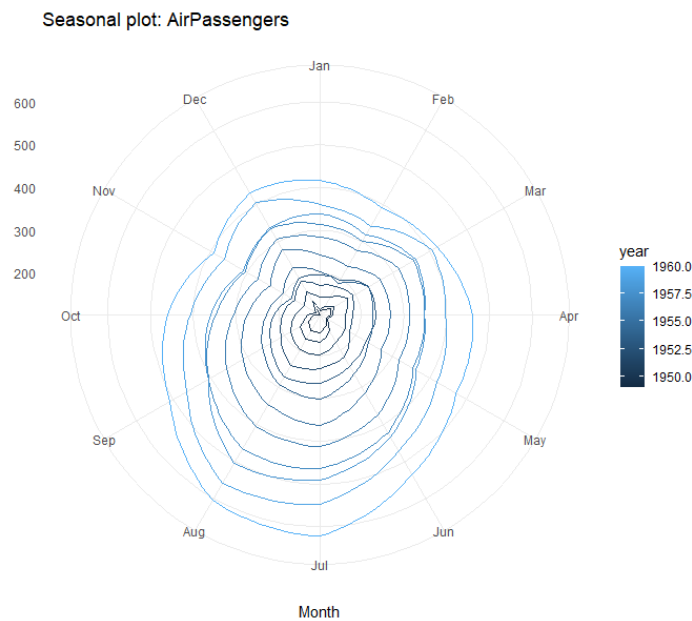


Figura 5.4: Série Temporal do número de passageiros internacionais por mês em uma companhia aérea (Pan Am) nos Estados Unidos do período de 1946-1960 - Gráfico por período anual em coordenadas polares.

produzido por `ggsubseriesplot()` (ver Figura 5.5) cria mini gráficos de tempo para cada mês. Aqui, a média para cada mês é mostrada como uma linha horizontal azul.

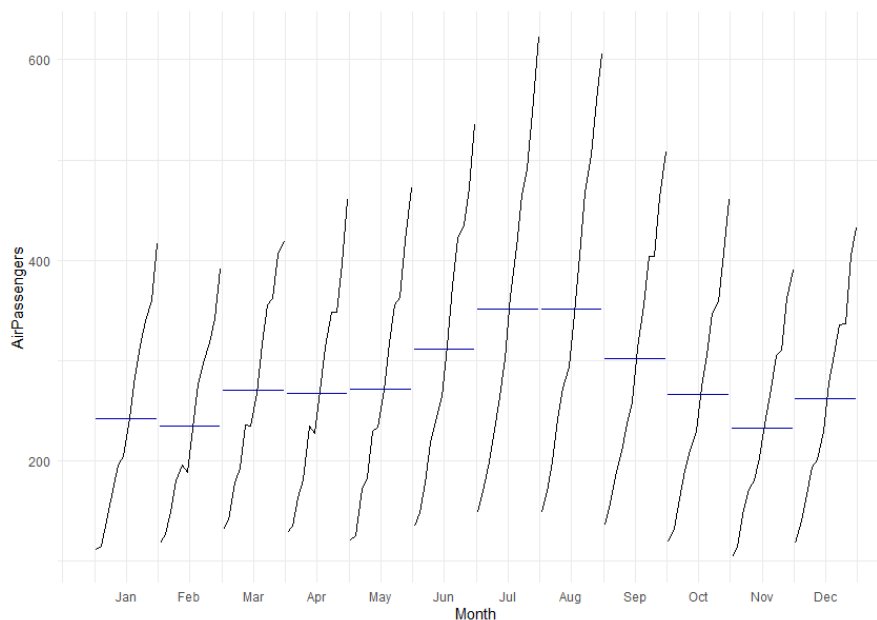


Figura 5.5: Série Temporal do número de passageiros internacionais por mês em uma companhia aérea (Pan Am) nos Estados Unidos do período de 1946-1960 - Gráfico por mês.

Vamos usar a função `boxplot` para verificarmos quaisquer outros efeitos sazonais (ver Figura 5.6).

A função `stat_rollapplyr()` facilita a adição da função média móvel aos dados da série temporal do número de passageiros internacionais por mês em uma companhia aérea (Pan Am) nos Estados Unidos do período de 1946-1960 (ver Figura 5.7).

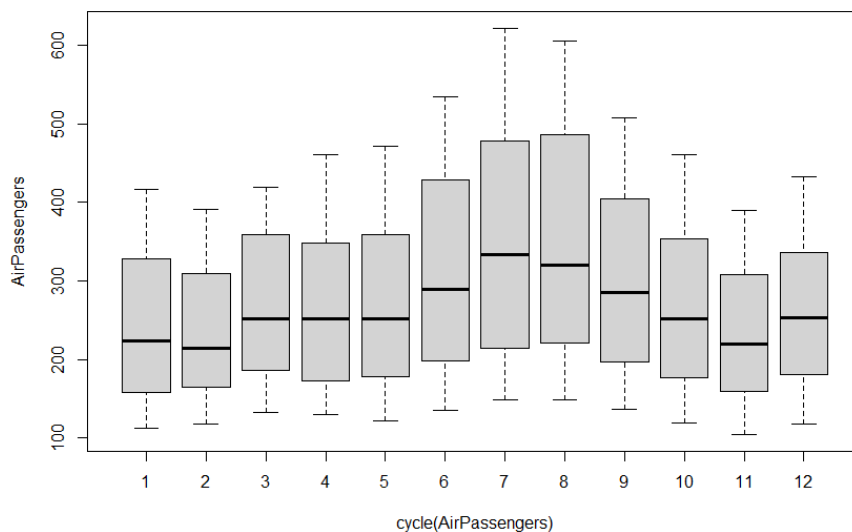


Figura 5.6: Série Temporal do número de passageiros internacionais por mês em uma companhia aérea (Pan Am) nos Estados Unidos do período de 1946-1960 - Gráfico boxplot por mês.

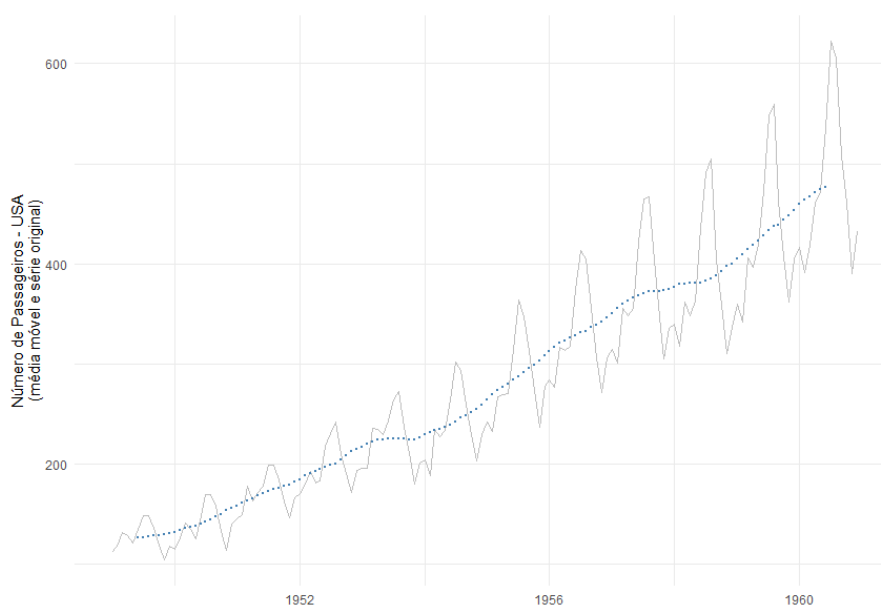


Figura 5.7: Série Temporal do número de passageiros internacionais por mês em uma companhia aérea (Pan Am) nos Estados Unidos do período de 1946-1960 - Ajuste de Média Móvel.

5.2 Avaliação da Precisão das Previsões

É importante avaliar a precisão das previsões usando previsões genuínas. Consequentemente, o tamanho dos resíduos não é uma indicação confiável de quão grandes os erros verdadeiros de previsão podem ser. A precisão das previsões só pode ser determinada considerando o desempenho de um modelo em novos dados que não foram usados durante o ajuste do modelo.

Ao escolher os modelos, é prática comum separar os dados disponíveis em duas porções, dados de treinamento e teste, onde os dados de treinamento são usados para estimar quaisquer parâmetros

de um método de previsão e os dados de teste são usados para avaliar sua precisão. Como os dados de teste não são usados para determinar as previsões, eles devem fornecer uma indicação confiável de quão bem o modelo provavelmente fará as previsões com base nos novos dados.



Fonte: <https://otexts.com/fpp2/accuracy.html#accuracy>

O tamanho do conjunto de teste é normalmente cerca de 20% da amostra total, embora esse valor dependa de quanto tempo a amostra é e quão adiantado você deseja prever. O conjunto de teste deve ser idealmente pelo menos tão grande quanto o horizonte de previsão máximo necessário. Os seguintes pontos devem ser observados.

- Um modelo que se ajusta bem aos dados de treinamento não necessariamente fará uma boa previsão.
- Um ajuste perfeito sempre pode ser obtido usando um modelo com parâmetros suficientes.
- Ajustar excessivamente um modelo aos dados é tão ruim quanto deixar de identificar um padrão sistemático nos dados.

Algumas referências descrevem o conjunto de teste como o “conjunto de validação” porque esses dados são “mantidos fora” dos dados usados para ajuste. Outras referências chamam o conjunto de treinamento de *in-sample data* e o conjunto de teste de *out-of-sample data*. Também podem ser usados “dados de treinamento” e “dados de teste”.

5.2.1 Seleção de Modelos Baseada no Erro de Previsão

Na análise de séries temporais, é bastante comum que existam vários modelos que podem ser utilizados para representar um conjunto de dados. Escolher o melhor modelo nem sempre é fácil. Um modo bastante útil de selecionar um bom modelo para uma série temporal é, utilizar os critérios de seleção de modelos que serão mencionados mais a frente. O principal objetivo em ajustar um modelo à uma serie temporal é podermos prever os seus futuros valores. Com isso, foram criados critérios de seleção de modelos baseados no erro de previsão.

O erro de previsão é dado por

$$e(t) = X_t - \hat{X}_{\mathcal{F}(t-1)}(t),$$

onde $\mathcal{F}(t-1) = \{X_{t-1}, \dots, X_1\}$, $t = 1, \dots, n$.

A seleção do modelo é baseada nas seguintes estatísticas.

(1) Percentual médio de erro - MPE

O Percentual médio de erro também pode ser referenciado como vício já que ele mede o vício de previsão,

$$\text{MPE} = \left(\frac{1}{n} \sum_{t=1}^n \frac{e(t)}{X_t} \right) \times 100\%.$$

(2) Erro médio absoluto de previsão - MAE

$$\text{MAE} = \frac{1}{n} \sum_{t=1}^n |e(t)|.$$

(3) *Percentual médio absoluto de erro - MAPE*

A média dos erros percentuais absolutos (MAPE) é uma das medidas de acurácia mais utilizada, devido as vantagens de interpretação e independência na escala. Essa medida está expressa na equação (5.1).

$$\text{MAPE} = \left(\frac{1}{n} \sum_{t=1}^n \left| \frac{e(t)}{X_t} \right| \right) \times 100\%. \quad (5.1)$$

(4) *Raiz quadrada da média dos erros quadráticos - RMSE*

Se a acurácia de técnicas de previsão é avaliada em séries temporais que apresentam a mesma escala, uma das medidas sugeridas é a raiz quadrada da média dos erros quadráticos (RMSE). Além disto, esta medida tem sido usada como métrica padrão para medir o desempenho de técnicas de previsão em estudos de meteorologia, qualidade do ar e pesquisa climática.

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{t=1}^n [e(t)]^2}$$

(5) *U de Theil*

O coeficiente de desigualdade de Theil é uma medida de acurácia frequentemente citada na literatura. No entanto, existe confusão sobre esse coeficiente, que pode ser resultado do fato de que o próprio Theil propôs duas fórmulas diferenciadas em momentos distintos, porém com o mesmo nome. Ambos são chamados de "coeficiente de desigualdade" e rotulados com o símbolo "U." A primeira proposta do U de Theil é delimitada entre 0 e 1. Já na segunda proposta o limite superior é infinito. Neste estudo será utilizada a primeira proposta do U de Theil e sua expressão é dada conforme a equação (5.2).

$$\text{U-Theil} = \frac{\sqrt{\frac{1}{n} \sum_{t=1}^n [X_t - \hat{X}_{\mathcal{F}(t-1)}(t)]^2}}{\sqrt{\frac{1}{n} \sum_{t=1}^n X_t^2} \sqrt{\frac{1}{n} \sum_{t=1}^n \hat{X}_{\mathcal{F}(t-1)}^2(t)}} \quad (5.2)$$

O coeficiente U de Theil analisa a qualidade das previsões, sendo que, quanto mais próximo de zero significa que o erro de previsão, gerada por determinado modelo ou combinação, é menor que da previsão ingênua, ou seja, que uma previsão é melhor que a previsão trivial (Makridakis et al., 1998).

(6) *Erro Escalonado Absoluto Médio - MASE*

Erros escalonados foram propostos por Hyndman e Koehler (2006) como uma alternativa ao uso de erros percentuais ao comparar a precisão da previsão em séries com unidades diferentes. Eles propuseram escalonar os erros com base no treinamento MAE a partir de um método simples de previsão.

Para uma série temporal não sazonal, uma maneira útil de definir um erro escalonado usa previsões ingênuas:

$$q_t = \frac{e_t}{\frac{1}{n-1} \sum_{t=2}^n |X_t - X_{t-1}|}.$$

Como o numerador e o denominador envolvem valores na escala dos dados originais, q_t é independente da escala dos dados. Um erro escalado é menor que um se resultar de uma previsão melhor do que a previsão ingênua média calculada nos dados de treinamento. Por outro lado, é maior que um se a previsão for pior do que a previsão ingênua média calculada nos dados de treinamento.

Para séries temporais sazonais, um erro escalonado pode ser definido usando previsões ingênuas sazonais:

$$q_t = \frac{e_t}{\frac{1}{n-s} \sum_{t=2}^n |X_t - X_{t-s}|},$$

onde s é a sazonalidade.

Assim, O Erro Médio Absoluto Escalonado (ou Erro Escalonado Absoluto Médio) é definido por

$$\text{MASE} = \frac{1}{n} \sum_{t=1}^n |q_t|.$$

Observação 5.1. Estas medidas são calculadas pela rotina *accuracy* do pacote *forecast* do R Core Team (2020).

5.3 Critérios de Seleção de Modelos

Nesta seção descrevemos os critérios de seleção de modelos baseados na soma mínima dos erros quadráticos (SSE), dado por

$$\text{SSE} = \sum_{t=1}^n e_t^2.$$

5.3.1 Critério da Informação de Akaike

O Critério da Informação de Akaike é definido como

$$\text{AIC}_c = n \log \left(\frac{\text{SSE}}{n} \right) + 2(k+2),$$

onde n é o número de observações usadas para estimativa e k é o número de preditores no modelo. Pacotes computacionais diferentes usam definições ligeiramente diferentes para o AIC, embora todos devam levar à seleção do mesmo modelo. O $k+2$ parte da equação ocorre porque há $k+2$ parâmetros no modelo: o k coeficientes para os preditores, o intercepto e a variância dos resíduos. A ideia aqui é penalizar o ajuste do modelo (SSE) com o número de parâmetros que precisam ser estimados, ou seja, penalizar os modelos com maior número de parâmetros. O modelo com o valor mínimo do AIC é frequentemente o melhor modelo para previsão.

5.3.2 Critério da Informação de Akaike Corrigido

O Critério da Informação de Akaike Corrigido é definido como

$$\text{AIC}_c = \text{AIC} + \frac{2(k+2)(k+3)}{n-k-3} = n \log \left(\frac{\text{SSE}}{n} \right) + 2(k+2) + \frac{2(k+2)(k+3)}{n-k-3}.$$

Tal como acontece com o AIC, o AICc deve ser minimizado.

5.3.3 Critério da Informação Bayesiana de Schwarz

Outra medida relacionada é o Critério de Informação Bayesiana de Schwarz (geralmente abreviado para BIC, SBIC ou SC):

$$\text{BIC} = n \log \left(\frac{\text{SSE}}{n} \right) + (k + 2) \log(n).$$

Tal como acontece com o AIC, minimizar o BIC tem como objetivo fornecer o melhor modelo. O modelo escolhido pelo BIC é igual ao escolhido pelo AIC ou com menos termos. Isso ocorre porque o BIC penaliza o número de parâmetros mais fortemente do que o AIC.

5.4 Modelos de Suavização Exponencial

Modelos de suavização é uma grande classe de métodos de previsão que se baseiam na ideia de que observações passadas contêm informações sobre o padrão da série temporal. O propósito dos métodos é distinguir um padrão de comportamento de qualquer outro ruído que possa estar contido nas observações da série e então usar esse padrão para prever valores futuros da série.

Observação 5.2. Notação:

$$\widehat{X}_{\mathcal{F}(t)}(t + h) = \widehat{X}_t(t + h)$$

onde $\mathcal{F}(t) = \{X_t, \dots, X_1\}$, para $t = 1, \dots, n$ e $h = 1, 2, 3, \dots$

Usando o método ingênuo, todas as previsões para o futuro são iguais ao último valor observado da série,

$$\widehat{X}_n(n + h) = X_n,$$

para $h = 1, 2, 3, \dots$. Portanto, o método ingênuo assume que a observação mais recente é a única importante e todas as observações anteriores não fornecem informações para o futuro. Isso pode ser considerado como uma média ponderada em que todo o peso é atribuído à última observação.

Usando o método da média, todas as previsões futuras são iguais a uma média simples dos dados observados,

$$\widehat{X}_n(n + h) = \frac{1}{n} \sum_{t=1}^n X_t,$$

para $h = 1, 2, 3, \dots$. Portanto, o método da média assume que todas as observações são de igual importância e atribui a elas pesos iguais ao gerar previsões.

Frequentemente, queremos algo entre esses dois extremos. Por exemplo, pode ser sensato atribuir pesos maiores às observações mais recentes do que às observações do passado distante. Este é exatamente o conceito por trás da Suavização Exponencial Simples. As previsões são calculadas usando médias ponderadas, onde os pesos diminuem exponencialmente conforme as observações vêm de mais longe no passado - os menores pesos estão associados às observações mais antigas.

5.4.1 Suavização Exponencial Simples

[<https://otexts.com/fpp2/ses.html>] Seja $\{X_t\}_{t=1}^n$ uma série temporal estacionária. Podemos descrever uma SES por

$$\widehat{X}_t(t + 1) = \alpha \sum_{j \geq 0} (1 - \alpha)^j X_{t-j} = \alpha X_t + \alpha(1 - \alpha)X_{t-1} + \alpha(1 - \alpha)^2 X_{t-2} \cdots \quad (5.3)$$

onde $0 \leq \alpha \leq 1$. A previsão um passo à frente para o tempo $t + 1$ é uma média ponderada de todas as observações da série X_1, \dots, X_t . A taxa na qual os pesos diminuem é controlada pelo parâmetro α .

Desta forma, SES é uma média ponderada que dá pesos maiores às observações mais recentes, eliminando uma das desvantagens do método de MMS.

A previsão no momento $t + 1$ é igual a uma média ponderada entre a observação mais recente e a previsão anterior $\hat{X}_{t-1}(t)$

$$\hat{X}_t(t+1) = \alpha X_t + (1 - \alpha) \hat{X}_{t-1}(t),$$

onde $0 \leq \alpha \leq 1$ é o parâmetro de suavização para $t = 1, \dots, n$. Lembre-se de que os valores preditos (ajustados) são simplesmente previsões de um passo a frente dos dados de treinamento.

O processo tem que começar em algum lugar, então deixamos o primeiro valor ajustado no tempo 1 ser denotado por ℓ_0 (que teremos que estimar). Então

$$\begin{aligned} \hat{X}_1(2) &= \alpha X_1 + (1 - \alpha) \ell_0 \\ \hat{X}_2(3) &= \alpha X_2 + (1 - \alpha) \hat{X}_1(2) \\ \hat{X}_3(4) &= \alpha X_3 + (1 - \alpha) \hat{X}_2(3) \\ \hat{X}_4(5) &= \alpha X_4 + (1 - \alpha) \hat{X}_3(4) \\ &\vdots \\ \hat{X}_{n-1}(n) &= \alpha X_{n-1} + (1 - \alpha) \hat{X}_{n-2}(n-1) \\ \hat{X}_n(n+1) &= \alpha X_n + (1 - \alpha) \hat{X}_{n-1}(n) \end{aligned}$$

Substituindo cada equação na seguinte equação, obtemos

$$\begin{aligned} \hat{X}_2(3) &= \alpha X_2 + (1 - \alpha) \hat{X}_1(2) [\alpha X_1 + (1 - \alpha) \ell_0] \\ &= \alpha X_2 + \alpha(1 - \alpha) X_1 + (1 - \alpha)^2 \ell_0 \\ \hat{X}_3(4) &= \alpha X_3 + (1 - \alpha) [\alpha X_2 + \alpha(1 - \alpha) X_1 + (1 - \alpha)^2 \ell_0] \\ &= \alpha X_3 + \alpha(1 - \alpha) X_2 + \alpha(1 - \alpha)^2 X_1 + (1 - \alpha)^3 \ell_0 \\ &\vdots \\ \hat{X}_t(t+1) &= \sum_{j=0}^{t-1} \alpha(1 - \alpha)^j X_{t-j} + (1 - \alpha)^t \ell_0 \end{aligned}$$

O último termo se torna minúsculo para t grande. Assim, a forma da média ponderada é a mesma previsão da equação (5.3).

5.4.2 Estimação do Parâmetro

A aplicação de todo método de suavização exponencial requer que os parâmetros de suavização e os valores iniciais sejam escolhidos. Em particular, para a suavização exponencial simples, precisamos selecionar os valores de α e ℓ_0 . Todas as previsões podem ser calculadas a partir dos dados, uma vez que conhecemos esses valores. Para os métodos a seguir, geralmente há mais de um parâmetro de suavização e mais de um componente inicial a ser escolhido.

Em alguns casos, os parâmetros de suavização podem ser escolhidos de maneira subjetiva - o previsor especifica o valor dos parâmetros de suavização com base na experiência anterior. No entanto, uma forma mais confiável e objetiva de obter valores para os parâmetros desconhecidos é estimá-los a partir dos dados observados.

Os coeficientes de um modelo de regressão minimizando a soma dos resíduos quadrados (geralmente conhecido como SSE ou “soma dos erros quadrados”). Da mesma forma, os parâmetros desconhecidos e

os valores iniciais para qualquer método de suavização exponencial podem ser estimados minimizando o SSE. Os resíduos são especificados como $e_t = X_t - \hat{X}_{t-1}(t)$, para $t = 1, \dots, n$. Portanto, encontramos os valores dos parâmetros desconhecidos e os valores iniciais que minimizam

$$\text{SSE} = \sum_{t=1}^n \left(X_t - \hat{X}_{t-1}(t) \right)^2 = \sum_{t=1}^n e_t^2. \quad (5.4)$$

Ao contrário do caso de regressão (onde temos fórmulas que retornam os valores dos coeficientes de regressão que minimizam o SSE), isso envolve um problema de minimização não linear, e precisamos usar uma ferramenta de otimização para resolvê-lo.

[<https://bit.ly/2JS0XKd>] Supondo $\{e_t\}_{t=1}^n$ é um ruído branco com distribuição $N(0, \sigma_e^2)$, podemos construir um intervalo de confiança assintótico para X_{t+h} , dado por

$$\left(\hat{X}_t(h) - z_{\frac{\alpha}{2}} \sigma_e \sqrt{\frac{\alpha}{2-\alpha}} ; \hat{X}_t(h) + z_{\frac{\alpha}{2}} \sigma_e \sqrt{\frac{\alpha}{2-\alpha}} \right)$$

onde $z_{\frac{\alpha}{2}}$ é o ponto crítico da distribuição normal padrão com coeficiente de confiança $(1 - \alpha)100\%$.

5.4.3 Aplicação

Série temporal referente às médias anuais das temperaturas na cidade de Nova York durante os anos de 1912 e 1971. Vamos ajustar um modelo SES a está serie temporal e tentar prever a temperatura nos próximos 12 anos.

```

1 #
2 #Suavização Exponencial Simples
3 #
4 #Série temporal referente às médias anuais das temperaturas na cidade
5 #de Nova York durante os anos de 1912 e 1971. Vamos ajustar um modelo
6 #SES a está serie temporal e tentar prever a temperatura nos próximos
7 #12 anos.
8 #
9 temperatura <- read_excel("temperatura.xls")
10 dados<-temperatura$Temperatura
11 dados<-ts(temperatura$Temperatura, start=c(1912), frequency = 1)
12 #
13 tmp<-ses(dados, h=12)
14 #
15 names(tmp)
16 summary(tmp)
17 #
18 autoplot(tmp)+theme_minimal()
19 new.year<-seq(from=1972, to=1983, by=1)
20 df1<-data.frame(rep("Dados", length(dados)), x=temperatura$Ano, y=as.vector(dados))
21 colnames(df1)<-c("V1", "V2", "V3")
22 df2<-data.frame(rep("Fitted", length(dados)), x=temperatura$Ano, y=as.vector(tmp$fitted))
23 colnames(df2)<-c("V1", "V2", "V3")
24 df3<-data.frame(rep("LI", length(tmp$mean)), x=new.year, y=as.vector(tmp$lower[, 2]))
25 colnames(df3)<-c("V1", "V2", "V3")
26 df4<-data.frame(rep("LS", length(tmp$mean)), x=new.year, y=as.vector(tmp$upper[, 2]))
27 colnames(df4)<-c("V1", "V2", "V3")
28 df5<-data.frame(rep("Forecast", length(tmp$mean)), x=new.year, y=as.vector(tmp$mean))
29 colnames(df5)<-c("V1", "V2", "V3")
30 y.predict<-data.frame(rbind(df1, df2, df3, df4, df5))
31 colnames(y.predict)<-c("Legenda", "V2", "V3")
32 ggplot(y.predict, aes(V2, V3, col=Legenda)) +
33   geom_line(size = 0.5)+

```

```

34   scale_color_manual(values = c("Dados" = "black", "Fitted" = "blue", "LI"="blue3", "LS"=
      "blue3",
35                               "Forecast"="red"))+
36   labs(y = "Temperatura", x = "Ano", title="Predição - Previsão - SES")+theme_minimal()
37 #-----
38 #Resíduo
39 res_ses<-tmp$residuals
40 #-----
41 #Análise dos resíduos
42 ggtsdisplay(res_ses, plot.type="scatter", theme=theme_bw())
43 #-----
44 #FAC dos Resíduos
45 ggAcf(res_ses, lag.max=100, type = c("correlation"))+labs(y = "FAC Amostral Resíduos SES
      ", title="")+
46   theme_minimal()
47 #-----
48 #QQ Plot dos Resíduos
49 ggqqplot(res_ses)+ggtitle("Resíduo Modelo SES")
50 #-----
51 #Densidade dos Resíduos
52 plot(density(res_ses), main="Random Error")
53 #-----
54 #Teste de Normalidade dos Resíduos
55 shapiro.test(res_ses)
56 #-----

```

Código 5.3: Código fonte em R: Série temporal referente às médias anuais das temperaturas na cidade de Nova York durante os anos de 1912 e 1971 - Modelo SES

A Figura 5.8 refere-se a Predição da Série temporal referente às médias anuais das temperaturas na cidade de Nova York durante os anos de 1912 e 1971 e a previsão para os 12 anos seguintes, utilizando os modelos de Suavização Exponencial Simples ($\hat{\alpha} = 0,182$ e $\ell_0 = 50,2755$).

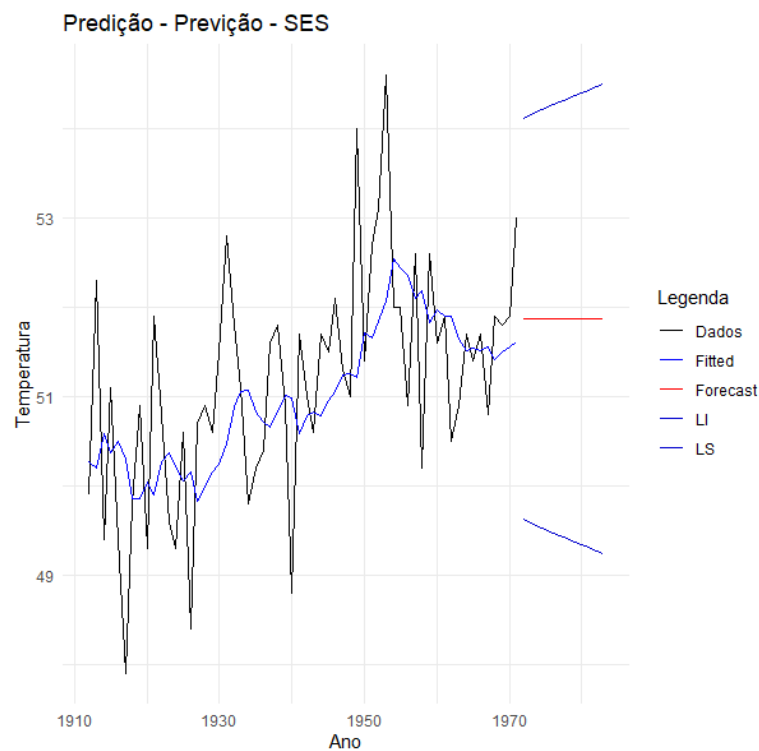


Figura 5.8: Modelos SES. Predição da Série temporal referente às médias anuais das temperaturas na cidade de Nova York durante os anos de 1912 e 1971. Previsão para os 12 anos seguintes.

A Figura 5.9 apresenta o gráfico da função de autocorrelação amostral do resíduo do Modelo SES aplicado a Série temporal referente às médias anuais das temperaturas na cidade de Nova York durante os anos de 1912 e 1971.

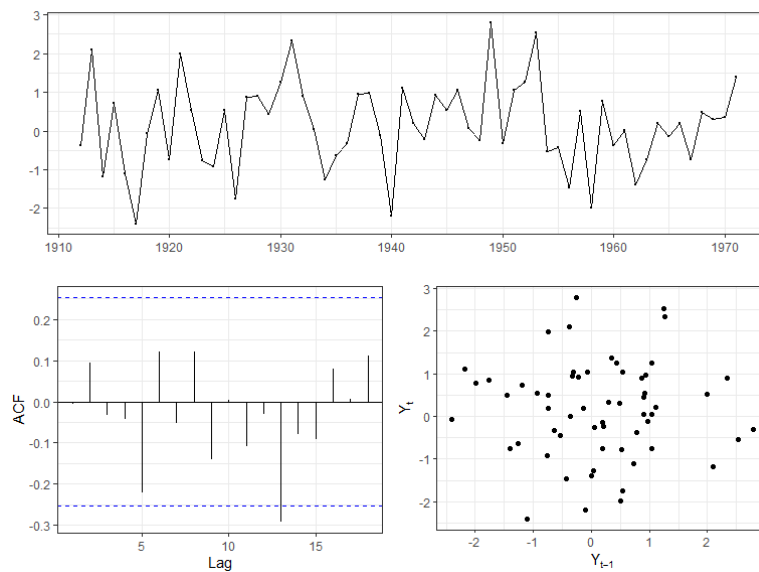


Figura 5.9: Análise dos resíduo do Modelo SES aplicado a Série temporal referente às médias anuais das temperaturas na cidade de Nova York durante os anos de 1912 e 1971.

5.5 Suavização Exponencial de Holt

5.5.1 Suavização Exponencial de Holt Clássico

Esta técnica de suavização é recomendada para séries que apresentam tendência. Considere o caso de uma série temporal $\{X_t\}_{t=1}^n$ que é composta localmente da soma de nível, tendência e resíduo com média zero e variância constante σ_ε^2 , isto é

$$X_t = \mu_t + T_t + \varepsilon_t, \quad t = 1, \dots, n.$$

Holt (2004) [*Reprinted Version of the 1957*] estendeu a Suavização Exponencial Simples para permitir a previsão de dados com uma tendência. Este método envolve uma equação de previsão e duas equações de suavização (uma para o nível e outra para a tendência):

Equação de Previsão	$\hat{X}_t(t+h) = \ell_t + hb_t$
Equação do Nível	$\ell_t = \alpha X_t + (1-\alpha)(\ell_{t-1} + b_{t-1})$
Equação da Tendência	$b_t = \beta(\ell_t - \ell_{t-1}) + (1-\beta)b_{t-1},$

onde ℓ_t denota uma estimativa do nível da série no momento t , b_t denota uma estimativa da tendência (*slope*) da série no tempo t , α é o parâmetro de suavização para o nível $0 \leq \alpha \leq 1$ e β é o parâmetro de suavização para a tendência $0 \leq \beta \leq 1$.

Tal como acontece com a Suavização Exponencial Simples, a equação de nível aqui mostra que ℓ_t é uma média ponderada de observação X_t e a previsão de treinamento de um passo à frente para o tempo t , aqui dado por $\ell_{t-1} + b_{t-1}$. A equação de tendência mostra que b_t é uma média ponderada da tendência estimada no momento t baseado em $\ell_t - \ell_{t-1}$ e b_{t-1} , a estimativa anterior da tendência.

A função de previsão não é mais constante. O previsão h passos a frente é igual ao último nível estimado mais h vezes o último valor de tendência estimado. Portanto, as previsões são uma função linear de h .

5.5.2 Suavização Exponencial de Holt Amortecido

As previsões geradas pelo método linear de Holt exibem uma tendência constante (aumentando ou diminuindo) indefinidamente no futuro. A evidência empírica indica que esses métodos tendem a superestimar, especialmente para horizontes de previsão mais longos. Motivados por esta observação, Gardner e McKenzie (1985) introduziram um parâmetro que “amortece” a tendência para uma linha plana em algum momento no futuro. Os métodos que incluem uma tendência atenuada provaram ser muito bem-sucedidos e são indiscutivelmente os métodos individuais mais populares quando as previsões são exigidas automaticamente para muitas séries.

Em conjunto com os parâmetros de suavização α e β (com $0 \leq \alpha, \beta \leq 1$ como no Método de Holt), este método também inclui um parâmetro de amortecimento $0 < \phi < 1$:

Equação de Previsão	$\hat{X}_t(t+h) = \ell_t + (\phi + \phi^2 + \dots + \phi^h)b_t$
Equação do Nível	$\ell_t = \alpha X_t + (1-\alpha)(\ell_{t-1} + \phi b_{t-1})$
Equação da Tendência	$b_t = \beta(\ell_t - \ell_{t-1}) + (1-\beta)\phi b_{t-1},$

Se $\phi = 1$ o método é idêntico ao método linear de Holt. Para valores entre 0 e 1, ϕ amortece a tendência para que se aproxime de uma constante em algum momento no futuro. Na verdade, as previsões convergem para $\frac{\ell_t + \phi b_t}{(1-\phi)}$ quando $h \rightarrow \infty$ para qualquer valor de $0 < \phi < 1$. Isso significa que as previsões de curto prazo têm tendência, enquanto as previsões de longo prazo são constantes.

Na prática, ϕ raramente é menor que 0,8, pois o amortecimento tem um efeito muito forte para valores menores. Valores de ϕ perto de 1 significa que um modelo com amortecimento não pode ser distinguido de um modelo sem amortecimento. Por esses motivos, geralmente restringimos ϕ para um mínimo de 0,8 e um máximo de 0,98.

5.5.3 Estimação dos Parâmetros

Os parâmetros de suavização, α e β , e os valores iniciais ℓ_0 e b_0 , e o parâmetro de amortecimento ϕ são estimados minimizando a soma dos erros de previsão ao quadrados (geralmente conhecido como SSE ou “soma dos erros quadrados”). Da mesma forma, os parâmetros desconhecidos e os valores iniciais para qualquer método de suavização exponencial podem ser estimados minimizando o SSE. Os resíduos são especificados como $e_t = X_t - \hat{X}_{t-1}(t)$, para $t = 1, \dots, n$. Portanto, encontramos os valores dos parâmetros desconhecidos e os valores iniciais que minimizam

$$\text{SSE} = \sum_{t=1}^n \left(X_t - \hat{X}_{t-1}(t) \right)^2 = \sum_{t=1}^n e_t^2. \quad (5.5)$$

5.5.4 Aplicação 1

A Figura 5.10 apresenta o gráfico e a função de autocorrelação amostral da Série Temporal do Número anual de passageiros anuais de transportadoras aéreas registradas na Austrália (milhões de passageiros, 1990-2016).

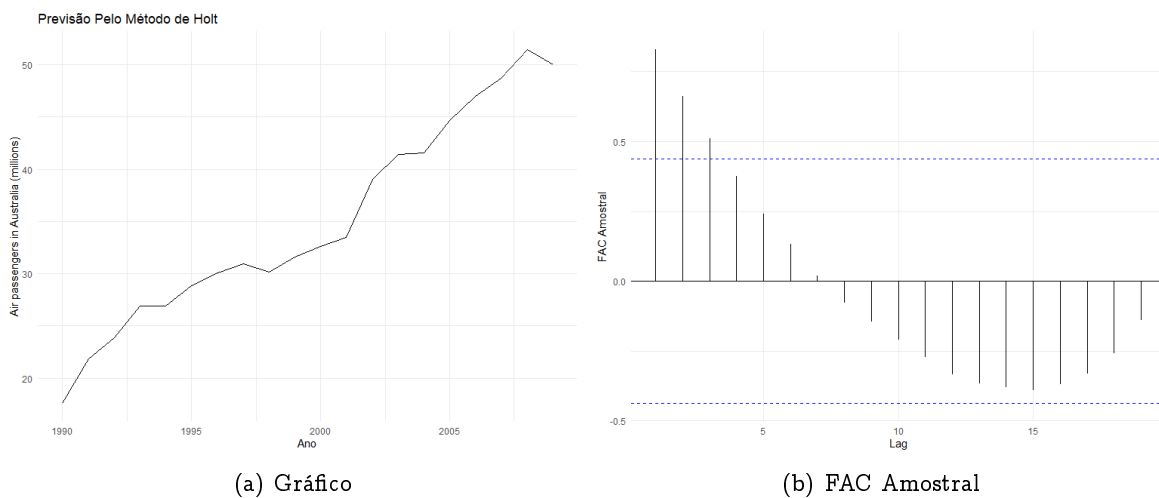


Figura 5.10: Gráfico e FAC Amostral da Série Temporal do Número anual de passageiros anuais de transportadoras aéreas registradas na Austrália (milhões de passageiros, 1990-2016).

A Figura 5.11 apresenta o gráfico da previsão da Série Temporal do Número anual de passageiros anuais de transportadoras aéreas registradas na Austrália (milhões de passageiros, 1990-2016), utilizando o Modelo de Suavização de Holt Amortecido cujas estimativas dos parâmetros são $\hat{\alpha} = 0,8792$ e $\hat{\beta} = 2e - 04$, e os valores iniciais $\hat{\ell}_0 = 18,6576$, $\hat{b}_0 = 1,9867$ e $\hat{\phi} = 0,98$. A Figura 5.12 apresenta a análise de resíduo do referido modelo.

A Figura 5.13 apresenta o gráfico comparando a previsão da Série Temporal do Número anual de passageiros anuais de transportadoras aéreas registradas na Austrália (milhões de passageiros, 1990-2016), utilizando o Modelo de Suavização de Holt (clássico e amortecido). No caso do modelo clássico as estimativas dos parâmetros são $\hat{\alpha} = 0,796$ e $\hat{\beta} = 1e - 04$, e os valores iniciais $\hat{\ell}_0 = 18,1302$ e

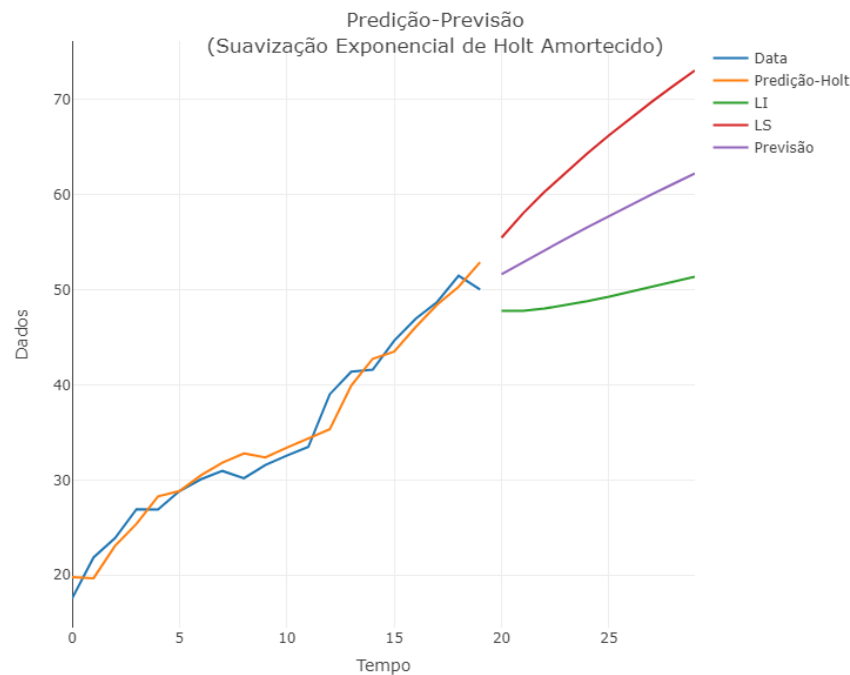


Figura 5.11: Previsão do total de passageiros anuais de transportadoras aéreas registradas na Austrália (milhões de passageiros, 1990-2016), utilizando o Modelo de Suavização de Holt Amortecido cujas estimativas dos parâmetros são $\hat{\alpha} = 0,8792$ e $\hat{\beta} = 2e - 04$, e os valores iniciais $\hat{\ell}_0 = 18,6576$, $\hat{b}_0 = 1,9867$ e $\hat{\phi} = 0,98$.

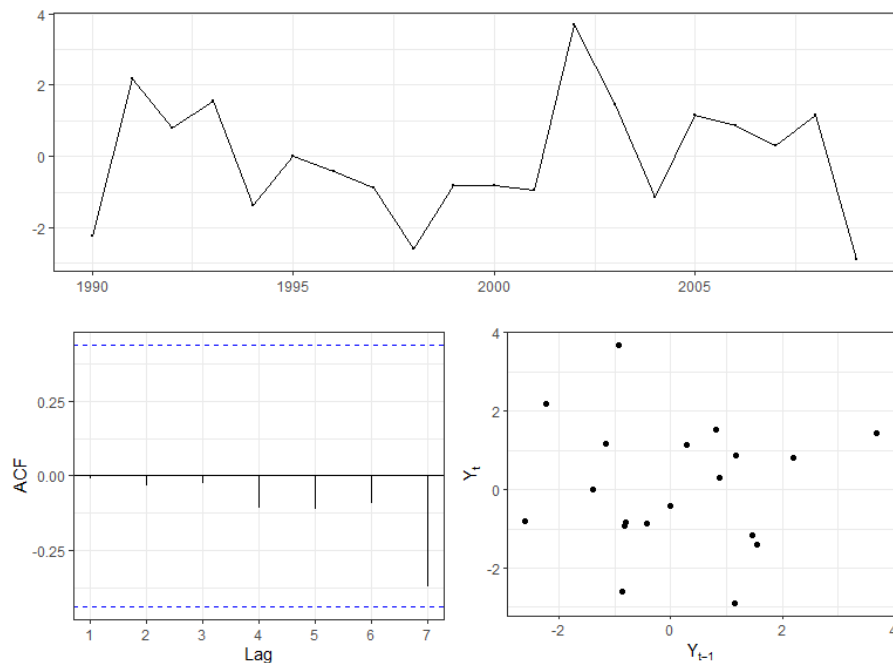


Figura 5.12: Análise de resíduo Modelo de Suavização de Holt Amortecido aplicado a série temporal do total de passageiros anuais de transportadoras aéreas registradas na Austrália (milhões de passageiros, 1990-2016).

$\hat{b}_0 = 1,6647$. Para o modelo suavizado as estimativas dos parâmetros são $\hat{\alpha} = 0,8792$ e $\hat{\beta} = 2e - 04$, e os valores iniciais $\hat{\ell}_0 = 18,6576$, $\hat{b}_0 = 1,9867$ e $\hat{\phi} = 0,98$.

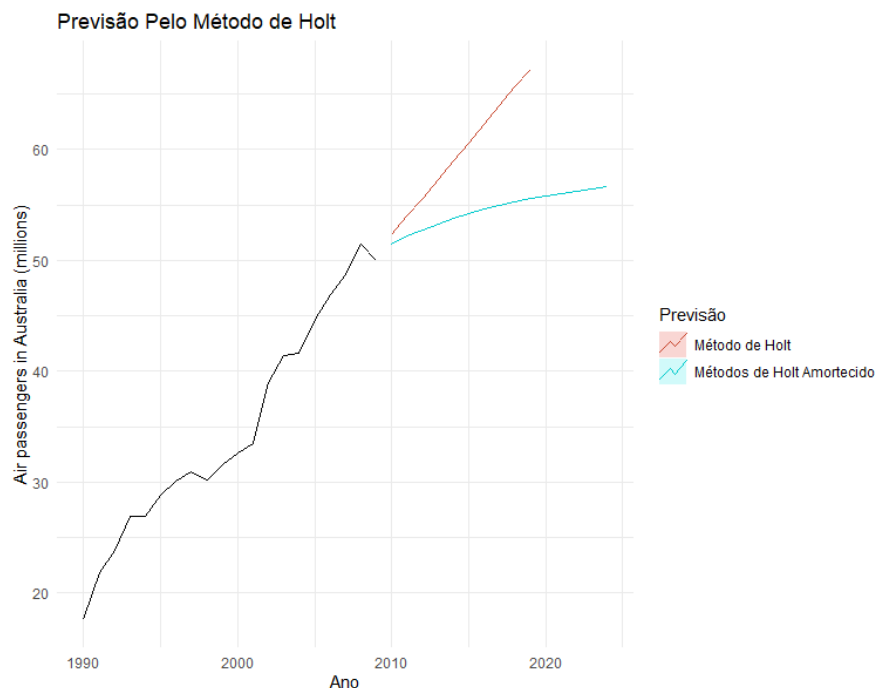


Figura 5.13: Previsão do total de passageiros anuais de transportadoras aéreas registradas na Austrália (milhões de passageiros, 1990-2016).

```

1 #
2 #Suavização Exponencial de Holt – SEH
3 #
4 #Número anual de passageiros das companhias aéreas australianas.
5 #
6 air <- window(ausair, start=1990)
7 autoplot(air) +
8   ggtitle("Previsão Pelo Método de Holt") + xlab("Ano") +
9   ylab("Air passengers in Australia (millions)") + theme_minimal()
10 ggAcf(air, lag.max=20, type = c("correlation")) + labs(y = "FAC Amostral", title = "") +
11   theme_minimal()
12 #
13 mod1 <- holt(air, h=10, damped=FALSE) #Clássico
14 mod2 <- holt(air, h=10, damped=TRUE, phi = NULL) #Amortecido
15 #
16 summary(mod1)
17 summary(mod2)
18 #
19 #Análise de Resíduos para cada modelo
20 mod <- mod2
21 names(mod)
22 summary(mod)
23 #
24 #Resíduo
25 res_holt <- mod$residuals
26 #
27 #Análise dos resíduos
28 ggtsdisplay(res_holt, plot.type = "scatter", theme = theme_bw())
29 #
30 #FAC dos Resíduos

```

```

31 ggAcf(res_holt, lag.max=100,type = c("correlation"))+labs(y = "FAC Amostral Resíduos
    SEH",title="")+
32   theme_minimal()
33 #
34 #QQ Plot dos Resíduos
35 ggqqplot(res_holt)+ggtitle("Resíduo Modelo SEH")
36 #
37 #Densidade dos Resíduos
38 plot(density(res_holt),main="Random Error")
39 #
40 #Teste de Normalidade dos Resíduos
41 shapiro.test(res_holt)
42 #
43 #Predição-Previsão Suavização Exponencial de Holt
44 cd<-rep("NA",length(air))
45 fig <- plot_ly(y = air, name = 'Data', type = 'scatter', mode = 'lines')
46 fig <- fig %>% add_trace(y = mod1$fitted, name = 'Predição-Holt', mode = 'lines')
47 fig <- fig %>% add_trace(y = c(cd,mod$lower[,2]), name = 'LI', mode = 'lines')
48 fig <- fig %>% add_trace(y = c(cd,mod$upper[,2]), name = 'LS', mode = 'lines')
49 fig <- fig %>% add_trace(y = c(cd,mod$mean), name = 'Previsão', mode = 'lines')
50 fig <- fig %>% layout(title =paste0("Predição-Previsão\n(Suavização Exponencial de Holt
    Amortecido)"),
51                        xaxis = list(title = 'Tempo'),
52                        yaxis = list(title = 'Dados'))
53 fig
54 #
55 #Previsão dos dois modelos no mesmo gráfico
56 autoplot(air) +
57   autolayer(mod1, series="Método de Holt", PI=FALSE) +
58   autolayer(mod2, series="Métodos de Holt Amortecido", PI=FALSE) +
59   ggtitle("Previsão Pelo Método de Holt") + xlab("Ano") +
60   ylab("Air passengers in Australia (millions)") +
61   guides(colour=guide_legend(title="Previsão"))+theme_minimal()
62 #

```

Código 5.4: Código fonte em R: Análise da Série Temporal do Número anual de passageiros anuais de transportadoras aéreas registradas na Austrália (milhões de passageiros, 1990-2016) - Suavização Exponencial de Holt Clássico

5.5.5 Aplicação 2

Nesta aplicação, comparamos o desempenho de previsão dos três métodos de suavização exponencial que consideramos até agora na previsão da população de gado ovino na Ásia (em milhões de cabeças). Os dados abrangem o período de 1961-2007 e são mostrados na Figura 5.14.

Tabela 5.1: Critérios de Seleção de Modelos para a Série Temporal da população de gado ovino na Ásia no período de 1961-2007 (em milhões de cabeças)

	AIC	BIC	AICC
SES	427.5522	433.1026	428.1103
SEH Clássico	424.5944	433.8451	426.0578
SEH Amortecido	427.637	438.7379	429.737

Pelos Critérios de Seleção de Modelos (Tabela 5.1), o Modelo SEH Clássico possui menor AIC e AICC, enquanto o Modelo SES possui menor BIC. Quanto as medidas de acurácia (Tabela 5.2), o

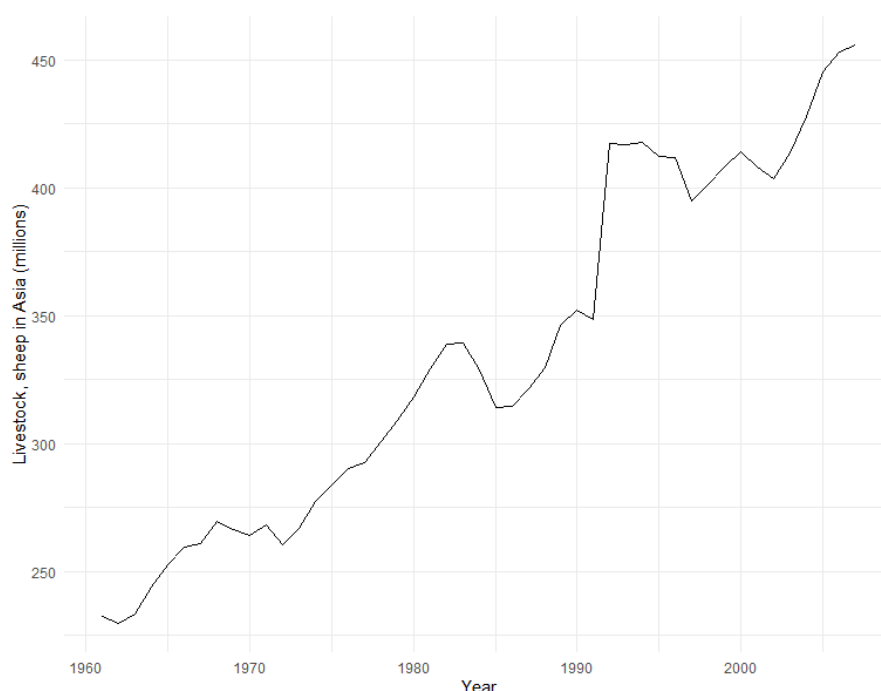


Figura 5.14: Série Temporal da população de gado ovino na Ásia no período de 1961-2007 (em milhões de cabeças).

Tabela 5.2: Medidas de Acurácia dos Modelos para a Série Temporal da população de gado ovino na Ásia no período de 1961-2007 (em milhões de cabeças)

	ME	RMSE	MAE	MPE	MAPE
SES	4.754439	12.92966	8.135768	1.371099	2.384596
SEH Clássico	-0.05148	12.00718	7.107686	-0.10743	2.062476
SEH Amortecido	0.555277	12.1411	7.213785	0.011527	2.095781

modelo SEH Clássico possui menores ME, RMSE, MAE e MAPE, enquanto o modelo SEH Amortecido possui menor MPE.

A Figura 5.15 apresenta o gráfico da previsão da Série Temporal da população de gado ovino na Ásia no período de 1961-2007 (em milhões de cabeças) utilizando o modelo SEH Clássico cujas estimativas dos parâmetros são $\hat{\alpha} = 0,999$ e $\hat{\beta} = 1e - 04$, e os valores iniciais $\hat{\ell}_0 = 225,4909$ e $\hat{b}_0 = 4,9513$.

5.6 Modelo de Holt-Winters

Holt (2004) [Reprinted Version of the 1957] e Winters (1960) estenderam o método de Holt para capturar a sazonalidade. O método sazonal de Holt-Winters compreende a equação de previsão e três equações de suavização - uma para o nível ℓ_t , um para a tendência b_t e um para o componente sazonal s_t , com os correspondentes parâmetros de suavização α , β e γ . Para denotar a frequência da sazonalidade utilizamos s , ou seja, o número de temporadas em um ano. Por exemplo, para dados trimestrais $s = 4$ e para dados mensais $s = 12$.

Existem duas variações desse método que diferem na natureza do componente sazonal. O método aditivo é preferido quando as variações sazonais são aproximadamente constantes ao longo da série,

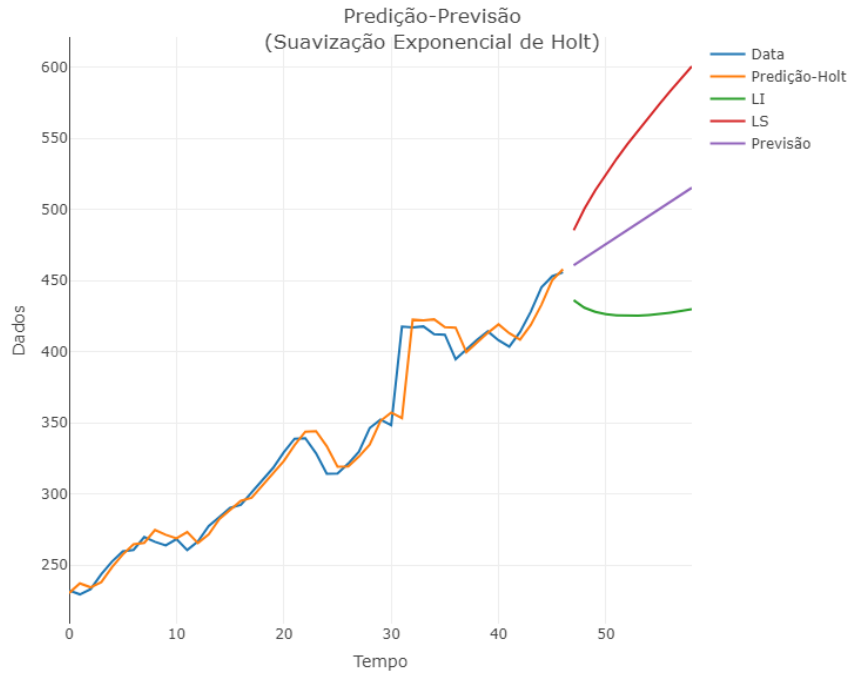


Figura 5.15: Previsão da Série Temporal da população de gado ovino na Ásia no período de 1961-2007 (em milhões de cabeças) utilizando o modelo SEH Clássico.

enquanto o método multiplicativo é preferido quando as variações sazonais estão mudando proporcionalmente ao nível da série. Com o método aditivo, a componente sazonal é expressa em termos absolutos na escala das séries observadas e na equação de nível a série é ajustada sazonalmente subtraindo a componente sazonal. Dentro de cada ano, o componente sazonal somará aproximadamente zero. Com o método multiplicativo, a componente sazonal é expressa em termos relativos (percentagens) e a série é ajustada sazonalmente pela divisão pela componente sazonal. Dentro de cada ano, o componente sazonal somará aproximadamente s .

5.6.1 Holt-Winters Aditivo

Considere uma série temporal $\{X_t\}_{t=1}^n$, com período s e n o tamanho da série temporal. No caso dos modelos de Holt-Winters para efeitos sazonais aditivos, consideramos que tal série é formada pela soma do nível, tendência e um fator sazonal. O modelo de Holt-Winters para efeitos sazonais aditivos satisfaz as equações

$$\begin{aligned}\hat{X}_t(t+h) &= \ell_t + hb_t + s_{t+h-s(k+1)} \\ \ell_t &= \alpha(X_t - s_{t-s}) + (1-\alpha)(\ell_{t-1} + b_{t-1}) \\ b_t &= \beta(\ell_t - \ell_{t-1}) + (1-\beta)b_{t-1} \\ s_t &= \gamma(X_t - \ell_t) + (1-\gamma)s_{t-s},\end{aligned}$$

onde α é o coeficiente de ponderação exponencial do nível ($0 \leq \alpha \leq 1$), β é o coeficiente de ponderação exponencial da tendência ($0 \leq \beta \leq 1$) e γ é o coeficiente de ponderação exponencial do fator sazonal ($0 \leq \gamma \leq 1$) e k é a parte inteira de $\frac{(h-1)}{s}$ o que garante que as estimativas dos índices sazonais utilizados para a previsão venham do último ano da amostra.

5.6.2 Holt-Winters Multiplicativo

Considere uma série temporal $\{X_t\}_{t=1}^n$, com período s e n o tamanho da série temporal. No caso dos modelos de Holt-Winters para efeitos sazonais multiplicativos, consideramos que tal série é formada pelo produto do nível, tendência e um fator sazonal. O modelo de Holt-Winters para efeitos sazonais multiplicativos satisfaz as equações

$$\begin{aligned}\widehat{X}_t(t+h) &= (\ell_t + hb_t)s_{t+h-s(k+1)} \\ \ell_t &= \alpha \frac{X_t}{s_{t-s}} + (1-\alpha)(\ell_{t-1} + b_{t-1}) \\ b_t &= \beta(\ell_t - \ell_{t-1}) + (1-\beta)b_{t-1} \\ s_t &= \gamma \frac{X_t}{\ell_t} + (1-\gamma)s_{t-s},\end{aligned}$$

onde α é o coeficiente de ponderação exponencial do nível ($0 \leq \alpha \leq 1$), β é o coeficiente de ponderação exponencial da tendência ($0 \leq \beta \leq 1$) e γ é o coeficiente de ponderação exponencial do fator sazonal ($0 \leq \gamma \leq 1$) e k é a parte inteira de $\frac{(h-1)}{s}$ o que garante que as estimativas dos índices sazonais utilizados para a previsão venham do último ano da amostra.

5.6.3 Holt-Winters Multiplicativo com fator de amortecimento

Para obter uma versão amortecida do modelo Holt-Winters Multiplicativo, seguimos o procedimento desenvolvido em [Gardner e McKenzie \(1985\)](#) e multiplicar o componente de tendência por ϕ em todos os lugares que aparece. O novo sistema de previsão é denominado Holt-Winters Multiplicativo com fator de amortecimento e satisfaz as seguintes equações

$$\begin{aligned}\widehat{X}_t(t+h) &= (\ell_t + (\phi + \phi^2 + \dots + \phi^h)b_t)s_{t+h-s(k+1)} \\ \ell_t &= \alpha \frac{X_t}{s_{t-s}} + (1-\alpha)(\ell_{t-1} + \phi b_{t-1}) \\ b_t &= \beta(\ell_t - \ell_{t-1}) + (1-\beta)\phi b_{t-1} \\ s_t &= \gamma \frac{X_t}{\ell_t} + (1-\gamma)s_{t-s},\end{aligned}$$

5.6.4 Holt-Winters Aditivo com fator de amortecimento

O desenvolvimento da versão amortecida do modelo Holt-Winters aditiva, para modelar uma tendência amortecida com sazonalidade aditiva, é semelhante ao caso multiplicativo. O sistema Holt-Winters Aditivo com fator de amortecimento satisfaz as seguintes equações

$$\begin{aligned}\widehat{X}_t(t+h) &= \ell_t + (\phi + \phi^2 + \dots + \phi^h)b_t + s_{t+h-s(k+1)} \\ \ell_t &= \alpha(X_t - s_{t-s}) + (1-\alpha)(\ell_{t-1} + \phi b_{t-1}) \\ b_t &= \beta(\ell_t - \ell_{t-1}) + (1-\beta)\phi b_{t-1} \\ s_t &= \gamma(X_t - \ell_t) + (1-\gamma)s_{t-s},\end{aligned}$$

5.6.5 Estimação dos Parâmetros

Os parâmetros de suavização, α , β e γ , e os valores iniciais ℓ_0 , b_0 e $s_0(1), s_0(2), \dots, s_0(s)$, e o parâmetro de amortecimento ϕ são estimados minimizando a soma dos erros de previsão ao quadrados

(geralmente conhecido como SSE ou “soma dos erros quadrados”). Da mesma forma, os parâmetros desconhecidos e os valores iniciais para qualquer método de suavização exponencial podem ser estimados minimizando o SSE. Os resíduos são especificados como $e_t = X_t - \hat{X}_{t-1}(t)$, para $t = 1, \dots, n$. Portanto, encontramos os valores dos parâmetros desconhecidos e os valores iniciais que minimizam

$$\text{SSE} = \sum_{t=1}^n \left(X_t - \hat{X}_{t-1}(t) \right)^2 = \sum_{t=1}^n e_t^2. \quad (5.6)$$

No caso dos modelos de Holt-Winters Aditivo e Multiplicativos podemos utilizar as funções *hw* do pacote *forecast* ou a rotina *HoltWinters* do pacote *stats*, ambas do software [R Core Team \(2020\)](#). Para os modelos de Holt-Winters Aditivo e Multiplicativos com fator de amortecimento podemos utilizar as funções *hw* do pacote *forecast*.

5.6.6 Aplicação 1

Análise da Série temporal do número de turistas internacionais na Austrália. Aplicamos o método de Holt-Winters com sazonalidade aditiva e multiplicativa para prever os números trimestrais de turistas internacionais na Austrália. A Figura 5.16 mostra o gráfico dos dados de 1999-2010 e sua respectiva função de autocorrelação amostral. Os dados mostram um padrão sazonal óbvio, com picos observados no trimestre de março de cada ano, correspondendo ao verão australiano.

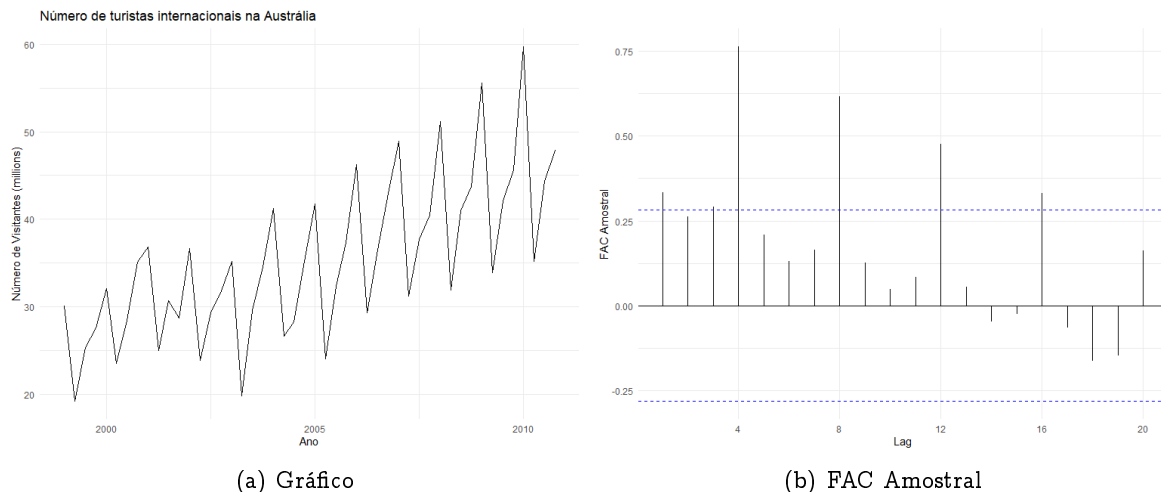


Figura 5.16: Gráfico e FAC Amostral da Série Temporal do número de turistas internacionais na Austrália (milhões, 1999-2010).

Tabela 5.3: Critérios de Seleção de Modelos para a Série Temporal do número de turistas internacionais na Austrália (milhões, 1999-2010)

	AIC	BIC	AICC
HW-A	284.5460	301.3868	289.2828
HW-M	280.5030	297.3438	285.2398
HW-AA	288.4291	307.1411	294.3750
HW-MA	278.9244	297.6364	284.8703

A Figura 5.17 apresenta os gráficos da previsão, utilizando os modelos de Holt-Winters Aditivo e Multiplicativo (com e sem amortecimento), para a Série Temporal do número de turistas internacionais

na Austrália (milhões, 1999-2010)

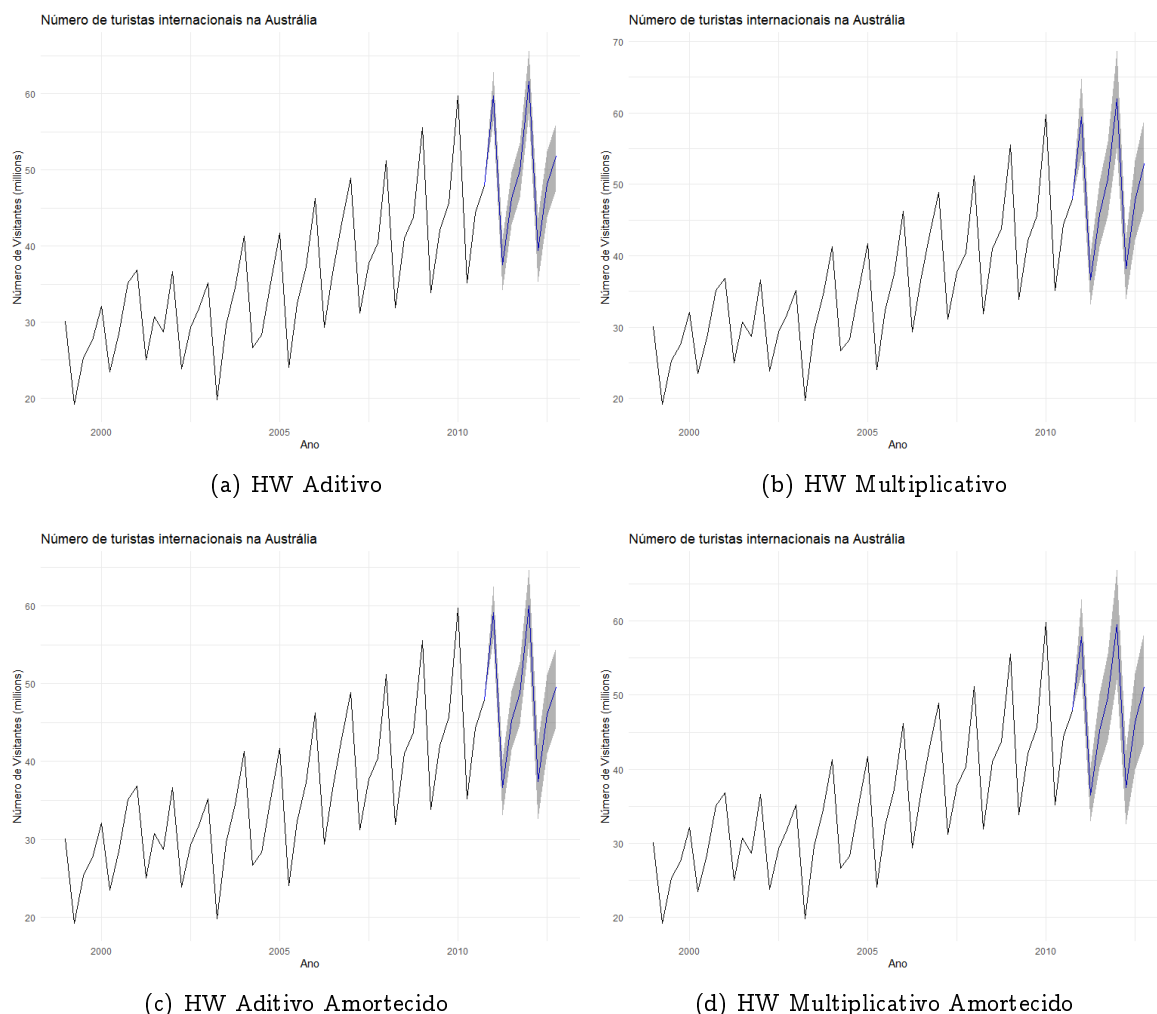


Figura 5.17: Previsão utilizando os modelos de Holt-Winters Aditivo e Multiplicativo (com e sem amortecimento) para a Série Temporal do número de turistas internacionais na Austrália (milhões, 1999-2010).

Tabela 5.4: Medidas de Acurácia dos Modelos para a Série Temporal do número de turistas internacionais na Austrália (milhões, 1999-2010)

	ME	RMSE	MAE	MPE	MAPE	MASE
HW-A	0.016031	2.3185	1.70989	-0.59421	5.345153	0.58992
HW-AA	0.175228	2.364426	1.747508	-0.23057	5.391091	0.602898
HW-M	-0.02873	2.123305	1.568994	-0.54228	4.771798	0.54131
HW-MA	-0.17222	2.120519	1.571684	-1.02042	4.705302	0.542238

A Figura 5.18 apresenta o gráfico da predição da Série Temporal do número de turistas internacionais na Austrália (milhões, 1999-2010) utilizando todos os Modelos de Holt-Winters.

Analisando a Tabela 5.3 com os Critérios de Seleção de Modelos, percebemos que o Modelo de Holt-Winters Multiplicativo com amortecimento apresenta menor AIC e AICc, enquanto o modelo de Holt-Winters Multiplicativo apresenta menor BIC. Analisando a Tabela 5.4 com as Medidas de

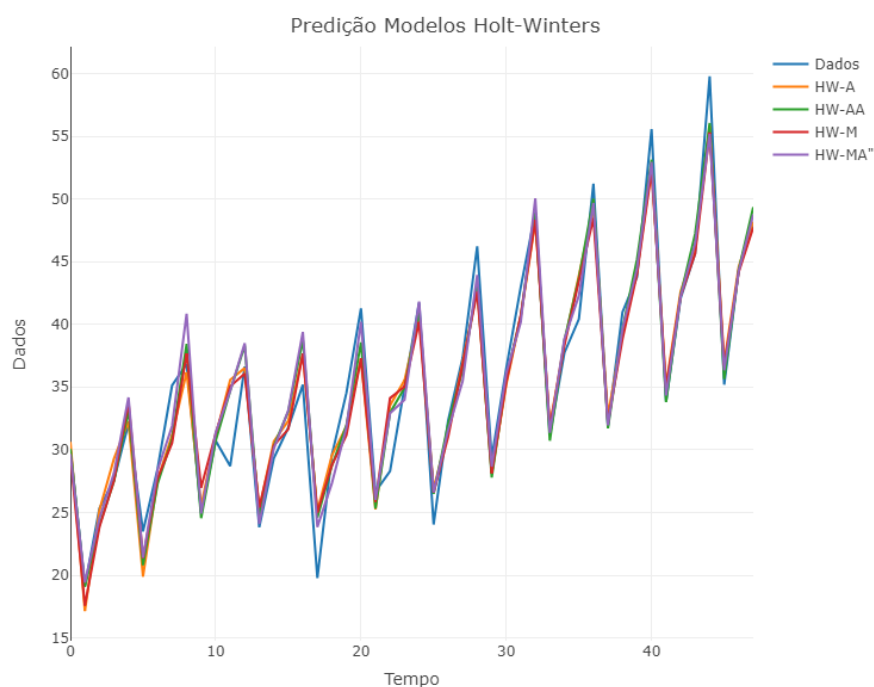


Figura 5.18: Predição da Série Temporal do número de turistas internacionais na Austrália (milhões, 1999-2010) utilizando os Modelos de Holt-Winters.

Acurácia o modelo de Holt-Winters Aditivo apresenta menor ME, modelo de Holt-Winters Aditivo com amortecimento apresenta menor MPE, modelo de Holt-Winters Multiplicativo apresenta menor MAE e MASE e o modelo de Holt-Winters Multiplicativo com amortecimento menor RMSE e MAPE. Pelos Critérios de Seleção de Modelos e Medidas de Acurácia escolhemos o modelo de Holt-Winters Multiplicativo com amortecimento cuja predição e previsão estão na Figura 5.19.

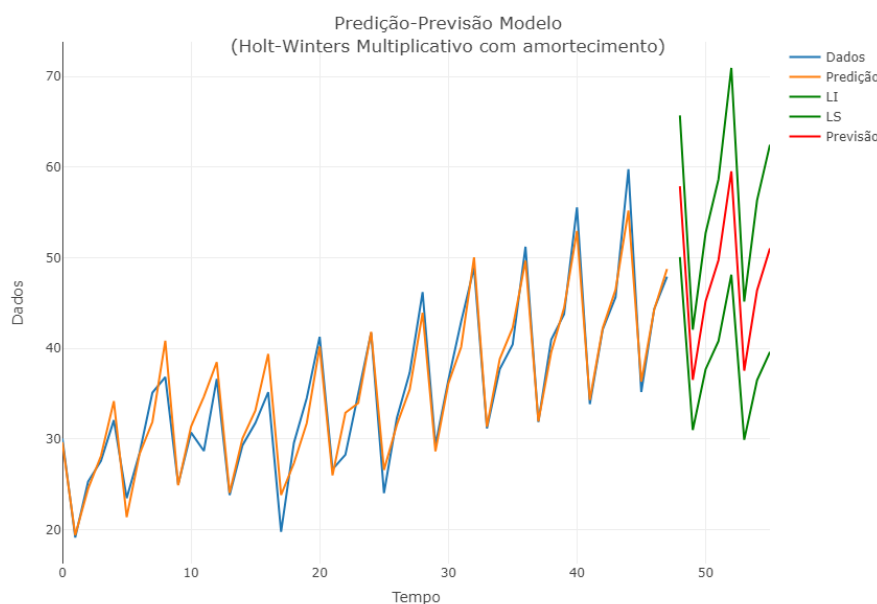


Figura 5.19: Predição-Previsão da Série Temporal do número de turistas internacionais na Austrália (milhões, 1999-2010) utilizando o Modelo de Holt-Winters Multiplicativo com amortecimento.


```

1 #
2 #Análise da Série temporal do número de turistas internacionais na Austrália.
3 #
4 #aust <- window(austourists, start=2005)
5 aust <- austourists
6 autoplot(aust) +
7   xlab("Ano") +
8   ylab("Número de Visitantes (millions)") +
9   ggtitle("Número de turistas internacionais na Austrália")+theme_minimal()
10 ggAcf(aust, lag.max=20,type = c("correlation"))+labs(y = "FAC Amostral",title="")+
11   theme_minimal()
12 #
13 #Modelos de Holt-Winters
14 mod1 <- hw(aust, seasonal="additive")
15 mod2 <- hw(aust, seasonal="multiplicative")
16 mod3 <- hw(aust, seasonal="additive", damped =TRUE)
17 mod4 <- hw(aust, seasonal="multiplicative", damped =TRUE)
18 autoplot(mod1) +
19   xlab("Ano") +
20   ylab("Número de Visitantes (millions)") +
21   ggtitle("Número de turistas internacionais na Austrália") +
22   guides(colour=guide_legend(title="Forecast"))+theme_minimal()
23 #
24 #Comparando as Predições
25 fig <- plot_ly(y = aust, name = 'Dados', type = 'scatter', mode = 'lines')
26 fig <- fig %>% add_trace(y = mod1$fitted, name = 'HW-A', mode = 'lines')
27 fig <- fig %>% add_trace(y = mod2$fitted, name = 'HW-AA', mode = 'lines')
28 fig <- fig %>% add_trace(y = mod3$fitted, name = 'HW-M', mode = 'lines')
29 fig <- fig %>% add_trace(y = mod4$fitted, name = 'HW-MA', mode = 'lines')
30 fig <- fig %>% layout(title = paste0("Predição Modelos Holt-Winters"),
31   xaxis = list(title = 'Tempo'),
32   yaxis = list(title = 'Dados'))
33
34 fig
35 #
36 #Critérios de Seleção de Modelos
37 csm.mod1<-c(mod1$model$aic, mod1$model$bic, mod1$model$aicc)
38 csm.mod2<-c(mod2$model$aic, mod2$model$bic, mod2$model$aicc)
39 csm.mod3<-c(mod3$model$aic, mod3$model$bic, mod3$model$aicc)
40 csm.mod4<-c(mod4$model$aic, mod4$model$bic, mod4$model$aicc)
41 scm.mod<-rbind(csm.mod1, csm.mod2, csm.mod3, csm.mod4)
42 colnames(scm.mod)<-c("AIC", "BIC", "AICC")
43 rownames(scm.mod)<-c("HW-A", "HW-M", "HW-AA", "HW-MA")
44 kable(scm.mod)
45 #
46 #Medidas de Acurácia
47 ac.mod1<-accuracy(mod1)
48 ac.mod2<-accuracy(mod2)
49 ac.mod3<-accuracy(mod3)
50 ac.mod4<-accuracy(mod4)
51 ac.mod<-rbind(ac.mod1, ac.mod2, ac.mod3, ac.mod4)
52 rownames(ac.mod)<-c("HW-A", "HW-M", "HW-AA", "HW-MA")
53 kable(ac.mod[, 1:6])
54 #
55 #Gráfico Predição-Previsão Modelo 4
56 cd<-rep("NA", length(aust))
57 fig <- plot_ly(y = aust, name = 'Dados', type = 'scatter', mode = 'lines')
58 fig <- fig %>% add_trace(y = mod4$fitted, name = 'Predição', mode = 'lines')
59 fig <- fig %>% add_trace(y = c(cd, mod4$lower[,2]), name = 'LI', mode = 'lines', line =
60   list(color = 'green'))
61 fig <- fig %>% add_trace(y = c(cd, mod4$upper[,2]), name = 'LS', mode = 'lines', line =

```

```

list(color = 'green'))
61 fig <- fig %>% add_trace(y = c(cd, mod4$mean), name = 'Previsão', mode = 'lines', line =
list(color = 'red'))
62 fig <- fig %>% layout(title = paste0("Predição—Previsão Modelo\n(Holt—Winters
Multiplicativo com amortecimento)"),
63 xaxis = list(title = 'Tempo'),
64 yaxis = list(title = 'Dados'))
65 fig
66
67 #

```

Código 5.5: Código fonte em R: Análise da Série Temporal do número de turistas internacionais na Austrália (milhões, 1999-2010) utilizando os Modelos de Holt-Winters

5.6.7 Aplicação 2

Análise da Série temporal de Produção de Bens Intermediários, período janeiro de 2002 a outubro de 2020.

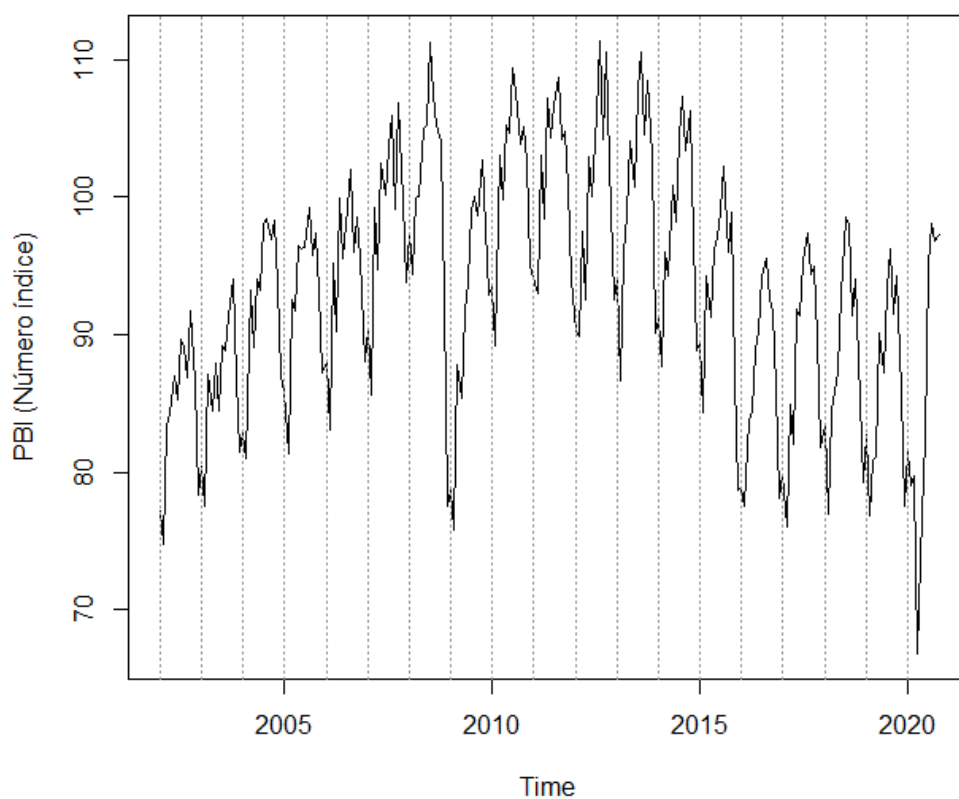


Figura 5.20: Gráfico da Série temporal de Produção de Bens Intermediários, período janeiro de 2002 a outubro de 2020.

Analisando a Tabela 5.5 percebemos que o Modelo de Holt-Winters Multiplicativo com tendência amortecida possui menores AIC e AICc, enquanto O Modelo de Holt-Winters Aditivo possui menor BIC.

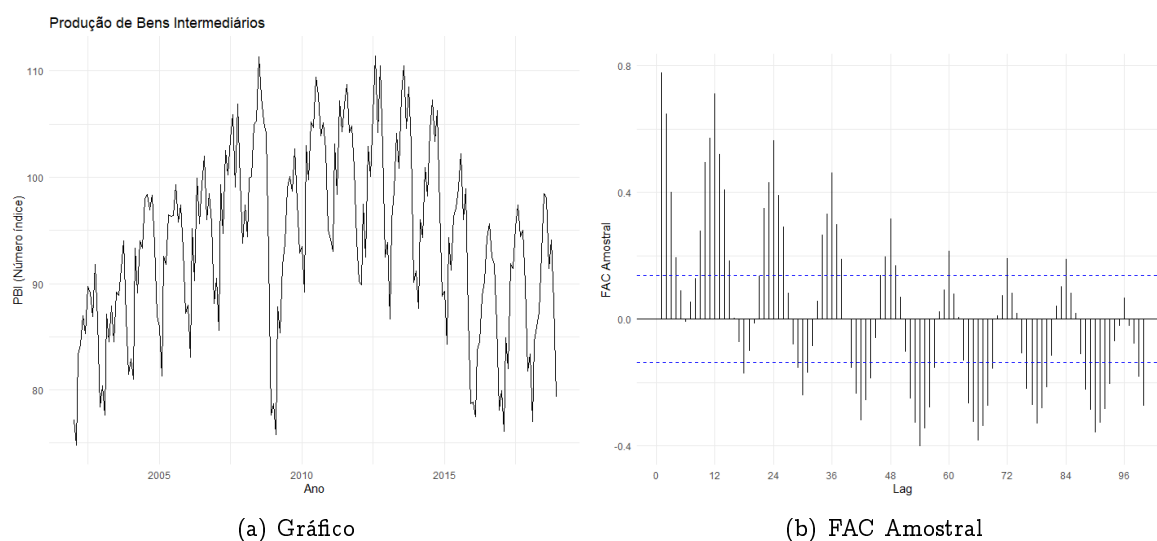


Figura 5.21: Gráficos e Função de Autocorrelação da Série Temporal da Produção de Bens Intermediários, período janeiro de 2002 a dezembro de 2018 - Conjunto Treino.

Tabela 5.5: Critérios de Seleção de Modelos para a Série Temporal de Produção de Bens Intermediários. Período:01/2002 - 12/2018

	AIC	BIC	AICc
HW-A	1436.629	1493.037	1439.919
HW-M	1522.513	1578.921	1525.803
HW-AA	1438.294	1498.020	1441.991
HW-MA	1435.206	1494.932	1438.903

Tabela 5.6: Medidas de Acurácia dos Modelos para a Série Temporal de Produção de Bens Intermediários. Período:01/2002 - 12/2018

Conjunto Treino - Período:01/2002 - 12/2018						
	ME	RMSE	MAE	MPE	MAPE	MASE
HW-A	-0.0263	2.1788	1.7169	-0.0562	1.8369	0.4099
HW-M	-0.2173	2.6840	2.0098	-0.2995	2.1691	0.4798
HW-AA	0.0171	2.1770	1.7209	-0.0098	1.8412	0.4108
HW-MA	-0.0329	2.1587	1.6874	-0.0675	1.8037	0.4028
Conjunto Teste - Período:01/2019 - 10/2020						
	ME	RMSE	MAE	MPE	MAPE	MASE
HW-A	-2.1697	5.7736	3.6294	-3.0198	4.6143	0.6343
HW-M	-1.8494	5.5256	3.6699	-2.5924	4.6074	0.5855
HW-AA	-1.6874	5.6251	3.6942	-2.4578	4.6534	0.6336
HW-MA	-1.2805	5.5817	3.8037	-2.0172	4.7374	0.6429

Analisando as medidas de acurácia de predição do conjunto treino (ver Tabela 5.6), o modelo de Modelo de Holt-Winters Multiplicativo com tendência amortecida possui menores RMSE, MAE, MAPE e MASE, enquanto o modelo de Modelo de Holt-Winters Aditivo com tendência amortecida possui menores ME e MPE. Analisando as medidas de acurácia de predição do conjunto teste (ver

Tabela 5.6), o modelo de Modelo de Holt-Winters Multiplicativo com tendência amortecida possui menores NE e MPE, o modelo de Modelo de Holt-Winters Multiplicativo possui menores RMSE, MAPE e MASE e por fim o modelo de Holt-Winters Aditivo possui menor MAE.

Coma base nas análise acima e como estamos interessados em prever passos a frente iremos utilizar o Modelo de Holt-Winters Multiplicativo para realizar as previsões passos a frente. A Figura 5.22 apresenta o gráficos da Predição-Previsão (in-sample) da Série Temporal da Produção de Bens Intermediários, período janeiro de 2002 a outubro de 2020 utilizando o Modelo de Holt-Winters Multiplicativo.

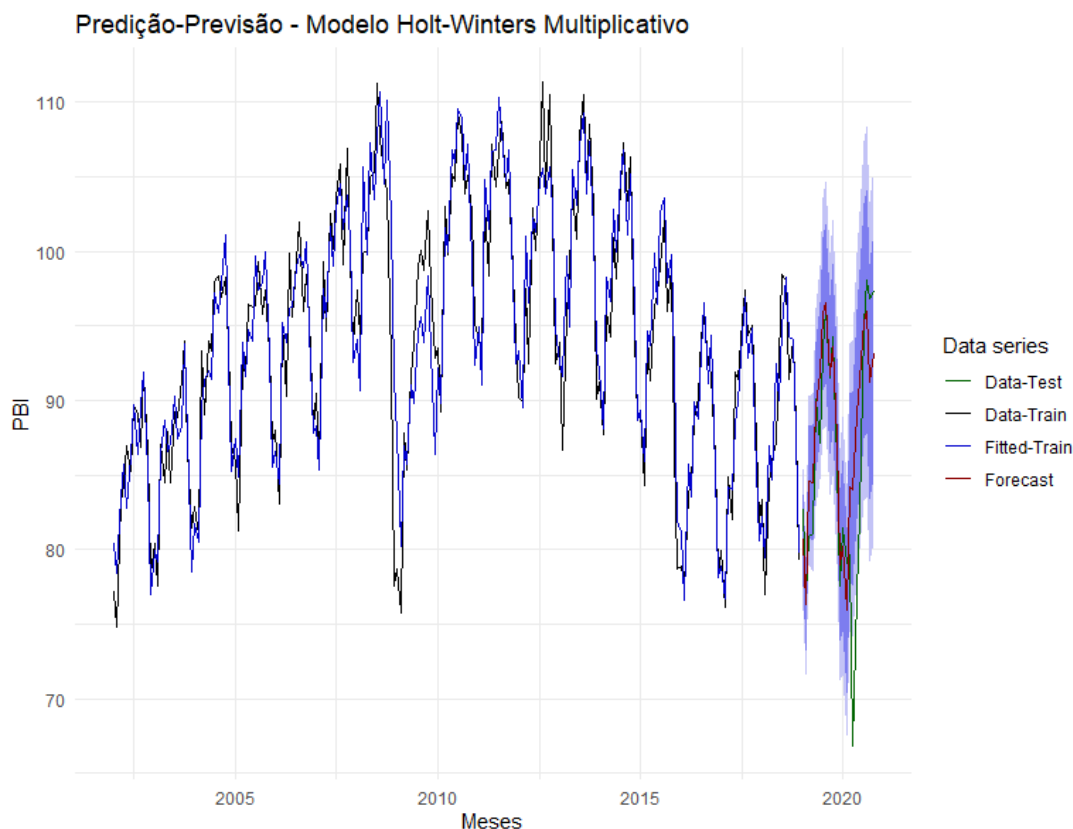


Figura 5.22: Gráficos da Predição-Previsão (in-sample) da Série Temporal da Produção de Bens Intermediários, período janeiro de 2002 a outubro de 2020 utilizando o Modelo de Holt-Winters Multiplicativo.

A Tabela 5.7 apresenta as estimativas dos parâmetros do Modelo de Holt-Winters Multiplicativo.

A Figura 5.23 apresenta o gráficos da Predição-Previsão (out-of-sample) da Série Temporal da Produção de Bens Intermediários, para o período novembro de 2020 a outubro de 2022 utilizando o Modelo de Holt-Winters Multiplicativo.

```

1 #
2 #Análise da Série temporal de Produção de Bens Intermediários.
3 #
4 #library(BETS)
5 data <- BETSget(21864)
6 #save(data, file = "SERIE_PBI.RData")
7 #load("SERIE_PBI.RData")
8 plot(data, ylab = "PBI (Número índice)", main = "")
9 abline(v = seq(2002,2020,1), col = "gray60", lty = 3)
10 #
11 #Conjuntos de Treino e Teste

```

Tabela 5.7: Estimativas dos parâmetros do Modelo de Holt-Winters Multiplicativo para a Série Temporal de Produção de Bens Intermediários. Período:01/2002 - 10/2020

α	β	γ	ℓ_0	b_0
0.3797	0.0046	0.3442	85.7762	0.1699
$s_0(1)$	$s_0(2)$	$s_0(3)$	$s_0(4)$	$s_0(5)$
0.8969	1.0003	1.0798	1.0425	1.019
$s_0(6)$	$s_0(7)$	$s_0(8)$	$s_0(9)$	$s_0(10)$
1.0597	1.0163	0.9967	1.0267	1.0016
$s_0(11)$	$s_0(12)$	AIC	BIC	AICc
0.9251	0.9354	1522.513	1525.803	1578.921

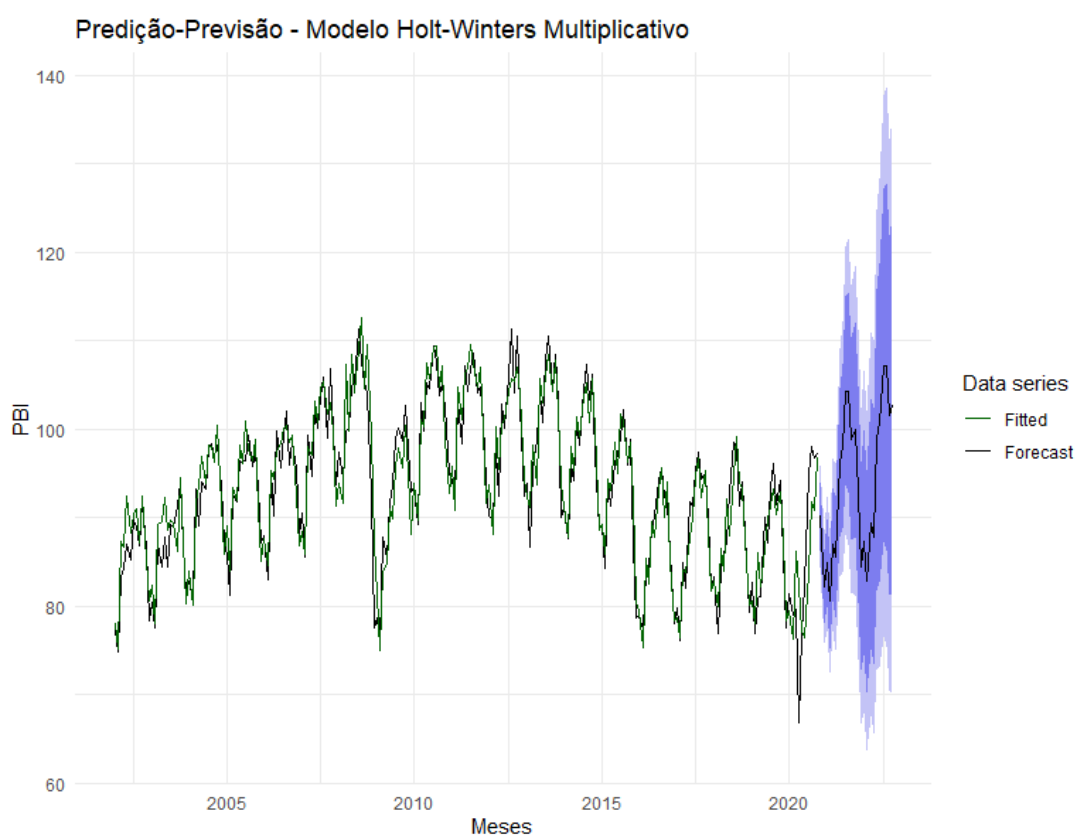


Figura 5.23: Gráficos da Previsão (out-of-sample) da Série Temporal da Produção de Bens Intermediários, período novembro de 2020 a outubro de 2022 utilizando o Modelo de Holt-Winters Multiplicativo.

```

12 data.tr <- window(data, start=2002, end=c(2018,12))
13 data.te <- window(data, start=2019)
14 #-----
15 autoplot(data.tr) +
16   xlab("Ano") +
17   ylab("PIB (Número índice)") +
18   ggtitle("Produção de Bens Intermediários")+theme_minimal()
19 ggAcf(data.tr, lag.max=100,type = c("correlation"))+labs(y = "FAC Amostral", title="")+
20   theme_minimal()
21 #-----

```

```

22 #Modelos de Holt-Winters
23 mod1 <- hw(data.tr, seasonal="additive", h=length(data.te))
24 mod2 <- hw(data.tr, seasonal="multiplicative", h=length(data.te))
25 mod3 <- hw(data.tr, seasonal="additive", damped =TRUE, h=length(data.te))
26 mod4 <- hw(data.tr, seasonal="multiplicative", damped =TRUE, h=length(data.te))
27 #
28 #Critérios de Seleção de Modelos
29 csm.mod1<-c(mod1$model$aic, mod1$model$bic, mod1$model$aicc)
30 csm.mod2<-c(mod2$model$aic, mod2$model$bic, mod2$model$aicc)
31 csm.mod3<-c(mod3$model$aic, mod3$model$bic, mod3$model$aicc)
32 csm.mod4<-c(mod4$model$aic, mod4$model$bic, mod4$model$aicc)
33 scm.mod<-rbind(csm.mod1, csm.mod2, csm.mod3, csm.mod4)
34 colnames(scm.mod)<-c("AIC", "BIC", "AICC")
35 rownames(scm.mod)<-c("HW-A", "HW-M", "HW-AA", "HW-MA")
36 kable(scm.mod)
37 #
38 #Medidas de Acurácia – Conjunto Treino
39 ac.mod1<-accuracy(mod1)
40 ac.mod2<-accuracy(mod2)
41 ac.mod3<-accuracy(mod3)
42 ac.mod4<-accuracy(mod4)
43 ac.mod<-rbind(ac.mod1, ac.mod2, ac.mod3, ac.mod4)
44 rownames(ac.mod)<-c("HW-A", "HW-M", "HW-AA", "HW-MA")
45 kable(ac.mod[, 1:6])
46 #
47 mod1.for<-forecast(mod1)
48 mod2.for<-forecast(mod2)
49 mod3.for<-forecast(mod3)
50 mod4.for<-forecast(mod4)
51 ac.mod1.for<-accuracy(mod1.for$mean, data.te)
52 ac.mod2.for<-accuracy(mod2.for$mean, data.te)
53 ac.mod3.for<-accuracy(mod3.for$mean, data.te)
54 ac.mod4.for<-accuracy(mod4.for$mean, data.te)
55 ac.mod.for<-rbind(ac.mod1.for, ac.mod2.for, ac.mod3.for, ac.mod4.for)
56 rownames(ac.mod.for)<-c("HW-A", "HW-M", "HW-AA", "HW-MA")
57 kable(ac.mod.for[, 1:6])
58 #
59 #Melhor Modelo Treino/Teste
60 clrs <- c("darkgreen", "black", "blue3", "darkred")
61 autoplot(mod2) +
62   autolayer(data.tr, series='Data-Train') +
63   autolayer(data.te, series='Data-Test') +
64   autolayer(fitted(mod2), series='Fitted-Train') +
65   autolayer(mod2$mean, series='Forecast') +
66   xlab("Meses") +
67   ylab("PBI") +
68   ggtitle("Predição-Previsão – Modelo Holt-Winters Multiplicativo")+
69   guides(colour=guide_legend(title="Data series"),
70          fill=guide_legend(title="Prediction interval"))+
71   scale_color_manual(values=clrs)+theme_minimal()
72 #
73 #Gráfico Predição-Previsão – Melhor Modelo
74 bestmod<-hw(data, seasonal="multiplicative", h=24)
75 bestmod.for<-forecast(bestmod, h=24)
76 cd<-rep("NA", length(data))
77 fig <- plot_ly(y = data, name = 'Dados', type = 'scatter', mode = 'lines')
78 fig <- fig %>% add_trace(y = bestmod$fitted, name = 'Predição', mode = 'lines')
79 fig <- fig %>% add_trace(y = c(cd, bestmod$lower[, 2]), name = 'LI', mode = 'lines', line
80   = list(color = 'green'))
81 fig <- fig %>% add_trace(y = c(cd, bestmod$upper[, 2]), name = 'LS', mode = 'lines', line
82   = list(color = 'green'))
83 fig <- fig %>% add_trace(y = c(cd, bestmod$mean), name = 'Previsão', mode = 'lines', line

```

```

      = list(color = 'red'))
82 fig <- fig %>% layout(title = paste0("Predição-Previsão Modelo\n(Holt-Winters
      Multiplicativo)"),
83                        xaxis = list(title = 'Tempo'),
84                        yaxis = list(title = 'Dados'))
85 fig
86 #
87 clrs <- c("darkgreen", "black", "blue3", "darkred")
88 autoplot(bestmod.for) +
89   autolayer(bestmod$mean, series="Forecast") +
90   autolayer(fitted(bestmod), series='Fitted') +
91   xlab("Meses") +
92   ylab("PBI") +
93   ggtitle("Predição-Previsão - Modelo Holt-Winters Multiplicativo")+
94   guides(colour=guide_legend(title="Data series"),
95          fill=guide_legend(title="Prediction interval"))+
96   scale_color_manual(values=clrs)+theme_minimal()
97 #

```

Código 5.6: Código fonte em R: Análise da Série Temporal da Produção de Bens Intermediários, período janeiro de 2002 a outubro de 2020 e previsão para o período de novembro de 2020 a outubro de 2022

5.7 Taxonomia de Modelos de Suavização Exponencial

Taxonomia do grego antigo táxis, “arranjo” e nomia, “método”. Os métodos de suavização exponencial não se restringem àqueles que apresentamos até o momento. Hyndman et al. (2008) propuseram a taxonomia para os modelos de suavização exponencial. Ao considerar as variações nas combinações dos componentes de tendência e sazonalidade, nove métodos de suavização exponencial são possíveis, listados na Tabela 5.8. Cada método é rotulado por um par de letras (T, S) que definem o tipo de componente *Tendência* e *Sazonalidade*. Por exemplo, (A, M) é o método com tendência aditiva e sazonalidade multiplicativa; (A_d , N) é o método com tendência amortecida e sem sazonalidade; e assim por diante.

Tabela 5.8: Reclassificação dos Modelos de Suavização Exponencia

Tendência		Sazonalidade		
		N	A	M
		(Sem)	(Aditiva)	(Multiplicativa)
N	(Sem)	(N,N)	(N,A)	(N,M)
A	(Aditiva)	(A,N)	(A,A)	(A,M)
A_d	(Aditiva Amortecida)	(A_d ,N)	(A_d ,A)	(A_d ,M)
M	(Multiplicativa)	(M,N)	(M,A)	(M,M)
M_d	(Multiplicativa Amortecida)	(M_d ,N)	(M_d ,A)	(M_d ,M)

Alguns desses métodos já vimos usando outros nomes:

Abreviatura	Modelo
(N,N)	Suavização Exponencial Simples
(A,N)	Suavização Exponencial de Holt
(A_d ,N)	Suavização Exponencial de Holt Amortecido
(A,A)	Holt-Winters Aditivo
(A,M)	Holt-Winters Multiplicativo
(A_d ,M)	Holt-Winters Aditivo Amortecido

Esse tipo de classificação foi proposta pela primeira vez por [Pegels \(1969\)](#), que também incluiu um método com tendência multiplicativa. Posteriormente, foi estendido por [Gardner \(1985\)](#) para incluir métodos com tendência amortecida aditiva e por [Taylor \(2003a\)](#) para incluir métodos com tendência amortecida multiplicativa. Não consideramos os métodos de tendência multiplicativos neste livro, pois eles tendem a produzir previsões ruins. Veja [Athanasopoulos e Hyndman \(2008\)](#) para uma discussão mais completa de todos os Modelos de Suavização Exponencial.

A Tabela 5.9 fornece as fórmulas recursivas para aplicar os nove métodos de suavização exponencial da Tabela 5.8. Cada célula inclui a equação de previsão para gerar previsões h passo a frente e as equações de suavização para a aplicação do método.

Tabela 5.9: Equações dos modelos e suas respectivas previsões. Em cada caso, ℓ_t denota o nível da série no momento t , b_t denota a tendência no tempo t , s_t denota o componente sazonal da série no momento t , e s denota o número de estações em um ano; α , β , γ e ϕ denotam os parâmetros de suavização $\phi_h = \phi + \phi^2 + \dots + \phi^h$ e k é a parte inteira de $\frac{h-1}{s}$

Tendência	Sazonalidade		
	N	A	M
N	$\ell_t = \alpha X_t + (1 - \alpha)\ell_{t-1}$ $\widehat{X}_t(t+h) = \ell_t$	$\ell_t = \alpha(X_t - s_{t-s}) + (1 - \alpha)\ell_{t-1}$ $s_t = \gamma(X_t - \ell_{t-1}) + (1 - \gamma)s_{t-s}$ $\widehat{X}_t(t+h) = \ell_t + s_{t-s+(k+1)}$	$\ell_t = \alpha(X_t/s_{t-s}) + (1 - \alpha)\ell_{t-1}$ $s_t = \gamma(X_t/\ell_{t-1}) + (1 - \gamma)s_{t-s}$ $\widehat{X}_t(t+h) = \ell_t s_{t-s+(k+1)}$
A	$\ell_t = \alpha X_t + (1 - \alpha)(\ell_{t-1} + b_{t-1})$ $b_t = \beta(\ell_t - \ell_{t-1}) + (1 - \beta)b_{t-1}$ $\widehat{X}_t(t+h) = \ell_t + hb_t$	$\ell_t = \alpha(X_t - s_{t-s}) + (1 - \alpha)(\ell_{t-1} + b_{t-1})$ $b_t = \beta(\ell_t - \ell_{t-1}) + (1 - \beta)b_{t-1}$ $s_t = \gamma(X_t - \ell_{t-1} - b_{t-1}) + (1 - \gamma)s_{t-s}$ $\widehat{X}_t(t+h) = \ell_t + hb_t + s_{t-s+(k+1)}$	$\ell_t = \alpha(X_t/s_{t-s}) + (1 - \alpha)(\ell_{t-1} + b_{t-1})$ $b_t = \beta(\ell_t - \ell_{t-1}) + (1 - \beta)b_{t-1}$ $s_t = \gamma(X_t/(\ell_{t-1} - b_{t-1})) + (1 - \gamma)s_{t-s}$ $\widehat{X}_t(t+h) = (\ell_t + hb_t)s_{t-s+(k+1)}$
A_d	$\ell_t = \alpha X_t + (1 - \alpha)(\ell_{t-1} + \phi b_{t-1})$ $b_t = \beta(\ell_t - \ell_{t-1}) + (1 - \beta)\phi b_{t-1}$ $\widehat{X}_t(t+h) = \ell_t + \phi_h b_t$	$\ell_t = \alpha(X_t - s_{t-s}) + (1 - \alpha)(\ell_{t-1} + \phi b_{t-1})$ $b_t = \beta(\ell_t - \ell_{t-1}) + (1 - \beta)\phi b_{t-1}$ $s_t = \gamma(X_t - \ell_{t-1} - \phi b_{t-1}) + (1 - \gamma)s_{t-s}$ $\widehat{X}_t(t+h) = \ell_t + \phi_h b_t + s_{t-s+(k+1)}$	$\ell_t = \alpha(X_t/s_{t-s}) + (1 - \alpha)(\ell_{t-1} + \phi b_{t-1})$ $b_t = \beta(\ell_t - \ell_{t-1}) + (1 - \beta)\phi b_{t-1}$ $s_t = \gamma(X_t/(\ell_{t-1} - \phi b_{t-1})) + (1 - \gamma)s_{t-s}$ $\widehat{X}_t(t+h) = (\ell_t + \phi_h b_t)s_{t-s+(k+1)}$
M	$\ell_t = \alpha X_t + (1 - \alpha)(\ell_{t-1} b_{t-1})$ $b_t = \beta(\ell_t/\ell_{t-1}) + (1 - \beta)b_{t-1}$ $\widehat{X}_t(t+h) = \ell_t + b_t^h$	$\ell_t = \alpha(X_t - s_{t-s}) + (1 - \alpha)(\ell_{t-1} b_{t-1})$ $b_t = \beta(\ell_t/\ell_{t-1}) + (1 - \beta)b_{t-1}$ $s_t = \gamma(X_t - \ell_{t-1} b_{t-1}) + (1 - \gamma)s_{t-s}$ $\widehat{X}_t(t+h) = \ell_t + b_t^h + s_{t-s+(k+1)}$	$\ell_t = \alpha(X_t/s_{t-s}) + (1 - \alpha)(\ell_{t-1} b_{t-1})$ $b_t = \beta(\ell_t/\ell_{t-1}) + (1 - \beta)b_{t-1}$ $s_t = \gamma(X_t/(\ell_{t-1} - b_{t-1})) + (1 - \gamma)s_{t-s}$ $\widehat{X}_t(t+h) = (\ell_t + b_t^h)s_{t-s+(k+1)}$
M_d	$\ell_t = \alpha X_t + (1 - \alpha)(\ell_{t-1} b_{t-1}^\phi)$ $b_t = \beta(\ell_t - \ell_{t-1}) + (1 - \beta)b_{t-1}^\phi$ $\widehat{X}_t(t+h) = \ell_t + b_t^{\phi_h}$	$\ell_t = \alpha(X_t - s_{t-s}) + (1 - \alpha)(\ell_{t-1} b_{t-1}^\phi)$ $b_t = \beta(\ell_t/\ell_{t-1}) + (1 - \beta)b_{t-1}^\phi$ $s_t = \gamma(X_t - \ell_{t-1} b_{t-1}^\phi) + (1 - \gamma)s_{t-s}$ $\widehat{X}_t(t+h) = \ell_t b_t^{\phi_h} + s_{t-s+(k+1)}$	$\ell_t = \alpha(X_t/s_{t-s}) + (1 - \alpha)(\ell_{t-1} b_{t-1}^\phi)$ $b_t = \beta(\ell_t/\ell_{t-1}) + (1 - \beta)b_{t-1}^\phi$ $s_t = \gamma(X_t/(\ell_{t-1} b_{t-1}^\phi)) + (1 - \gamma)s_{t-s}$ $\widehat{X}_t(t+h) = (\ell_t b_t^{\phi_h})s_{t-s+(k+1)}$

5.8 Modelos de Espaço de Estado de Inovações para Suavização Exponencial

No restante deste capítulo, estudaremos os modelos estatísticos que fundamentam os métodos de suavização exponencial que consideramos até agora. Os métodos de suavização exponencial apresentados na Tabela 5.9 são algoritmos que geram previsões pontuais. Os modelos estatísticos nesta seção geram as mesmas previsões de pontos, mas também podem gerar intervalos de predição (ou previsão). Um modelo estatístico é um processo de geração de dados estocásticos (ou aleatórios) que pode produzir uma distribuição de previsão inteira. Também descreveremos como usar os critérios de seleção de modelo introduzidos na Seção 5.3 para escolher o modelo de maneira objetiva.

Cada modelo consiste em uma equação de medição que descreve os dados observados e algumas equações de estado que descrevem como as componentes ou estados não observados (nível, tendência, sazonalidade) mudam ao longo do tempo. Portanto, eles são chamados de modelos de espaço de estado.

Para cada método existem dois modelos: um com erros aditivos e outro com erros multiplicativos. As previsões de pontos produzidas pelos modelos são idênticas se usarem os mesmos valores de parâmetro de suavização. Eles irão, no entanto, gerar diferentes intervalos de previsão.

Para distinguir entre um modelo com erros aditivos e outro com erros multiplicativos (e também para distinguir os modelos dos métodos), adicionamos uma terceira letra à classificação da Tabela 5.9. Rotulamos cada modelo de espaço de estado como ETS(\cdot, \cdot, \cdot) para (erro, tendência, sazonalidade). Este rótulo também pode ser denominado Suavização Exponencial (*Exponential Smoothing*). Usando a mesma notação da Tabela 5.9, as possibilidades para cada componente são: Erro = $\{A, M\}$, Tendência = $\{N, A, A_d, M, M_d\}$ e Sazonalidade = $\{N, A, M\}$.

5.8.1 ETS(A,N,N): Suavização Exponencial Simples com Erros Aditivos

Considere uma série temporal $\{X_t\}_{t=1}^n$, com período s e n o tamanho da série temporal. Então o modelo de Suavização Exponencial Simples satisfaz as equações

$$\begin{array}{ll} \text{Equação de Previsão} & \widehat{X}_t(t+1) = \ell_t \\ \text{Equação de Suavização} & \ell_t = \alpha X_t + (1 - \alpha)\ell_{t-1}, \end{array}$$

Reorganizarmos a equação de suavização para o nível, obteremos o formulário de “correção de erro”

$$\begin{aligned} \ell_t &= \ell_{t-1} + \alpha(X_t - \ell_{t-1}) \\ &= \ell_{t-1} + \alpha e_t \end{aligned}$$

onde $e_t = X_t - \ell_{t-1} = X_t - \widehat{X}_{t-1}(t)$ é o resíduo no tempo t , $t = 1, \dots, n$.

Podemos escrever $X_t = \ell_{t-1} + e_t$ de modo que cada observação pode ser representada pelo nível anterior mais um erro. Para transformar isso em um modelo de espaço de estado de inovações, tudo o que precisamos fazer é especificar a distribuição de probabilidade para $\{e_t\}_{t=1}^n$. Para um modelo com erros aditivos, assumimos que os resíduos (os erros de treinamento) $\{e_t\}_{t=1}^n$ é um ruído branco normalmente distribuídos com média 0 e variância σ_e^2 . Notação: $e_t = \varepsilon_t \sim \text{NID}(0, \sigma_e^2)$. Desta forma as equações do modelo podem se escritas como

$$X_t = \ell_{t-1} + \varepsilon_t \tag{5.7}$$

$$\ell_t = \ell_{t-1} + \alpha \varepsilon_t. \tag{5.8}$$

Referimo-nos a (5.8) como a equação de medição (ou observação) e (5.8) como a equação de estado (ou transição). Essas duas equações, junto com a distribuição estatística dos erros, formam um modelo estatístico totalmente especificado. Especificamente, eles constituem um modelo de espaço de estado de inovações subjacente à suavização exponencial simples.

A equação de estado mostra a evolução do estado ao longo do tempo. A influência do parâmetro de suavização α é o mesmo que para os métodos discutidos anteriormente. Por exemplo, α rege a quantidade de mudança em níveis sucessivos: altos valores de α permitir mudanças rápidas no nível; valores baixos de α levam a mudanças suaves. E se $\alpha = 0$, o nível da série não muda com o tempo; E se $\alpha = 1$, o modelo se reduz a um modelo de passeio aleatório, $X_t = X_{t-1} + \varepsilon_t$.

5.8.2 ETS(M,N,N): Suavização Exponencial Simples com Erros Multiplicativos

De maneira semelhante, podemos especificar modelos com erros multiplicativos, escrevendo os erros de treinamento um passo à frente como erros relativos

$$\varepsilon_t = \frac{X_t - \hat{X}_{t-1}(t)}{\hat{X}_{t-1}(t)}$$

onde $\varepsilon_t \sim \text{NID}(0, \sigma_\varepsilon^2)$. Substituindo $\hat{X}_{t-1}(t) = \ell_{t-1}$ obtemos $X_t = \ell_{t-1} + \ell_{t-1}\varepsilon_t$ e $e_t = X_t - \hat{X}_{t-1}(t) = \ell_{t-1}\varepsilon_t$. Então podemos escrever a forma multiplicativa do modelo de espaço de estado como

$$\begin{aligned} X_t &= \ell_{t-1}(1 + \varepsilon_t) \\ \ell_t &= \ell_{t-1}(1 + \alpha\varepsilon_t). \end{aligned}$$

5.8.3 ETS(A,A,N): Suavização Exponencial de Holt com Erros Aditivos

Para este modelo, assumimos que os erros de treinamento um passo à frente são dados por $\varepsilon_t = X_t - \ell_{t-1} - b_{t-1} \sim \text{NID}(0, \sigma_\varepsilon^2)$. Substituindo ε_t nas equações de correção de erro para o modelo de Suavização Exponencial de Holt, obtemos

$$\begin{aligned} X_t &= \ell_{t-1} + b_{t-1} + \varepsilon_t \\ \ell_t &= \ell_{t-1} + b_{t-1} + \alpha\varepsilon_t \\ b_t &= b_{t-1} + \alpha\beta\varepsilon_t, \end{aligned}$$

5.8.4 ETS(M,A,N): Suavização Exponencial de Holt com Erros Multiplicativos

Especificar erros de treinamento um passo à frente como erros relativos, como

$$\varepsilon_t = \frac{X_t - (\ell_{t-1} + b_{t-1})}{(\ell_{t-1} + b_{t-1})}$$

e seguindo uma abordagem semelhante à usada acima, o modelo de espaço de estado de inovações subjacente ao modelo de Suavização Exponencial de Holt com erros multiplicativos é especificado como

$$\begin{aligned} X_t &= (\ell_{t-1} + b_{t-1})(1 + \varepsilon_t) \\ \ell_t &= (\ell_{t-1} + b_{t-1})(1 + \alpha\varepsilon_t) \\ b_t &= b_{t-1} + \alpha\beta(\ell_{t-1} + b_{t-1})\varepsilon_t \end{aligned}$$

onde $\varepsilon_t \sim \text{NID}(0, \sigma_\varepsilon^2)$.

5.9 Modelos de Espaço de Estados para todos os ETS

Agora fornecemos os modelos de espaço de estado para todas as 30 suavizações exponenciais variações. O modelo geral envolve um vetor de estado $\mathbf{Y}_t = (\ell_t, b_t, s_t, s_{t-1}, \dots, s_{t-s+1})^\top$ e as equações de espaço de estados

$$X_t = w(\mathbf{Y}_{t-1}) + r(\mathbf{Y}_{t-1})\varepsilon_t, \quad (5.9)$$

$$\mathbf{Y}_t = f(\mathbf{Y}_{t-1}) + g(\mathbf{Y}_{t-1})\varepsilon_t, \quad (5.10)$$

$$(5.11)$$

onde $\{\varepsilon_t\}_{t \in \mathbb{Z}}$ é um ruído branco gaussiano com variância σ_ε^2 , e $\mu_t = w(\mathbf{Y}_{t-1})$. O modelo com erros aditivos tem $r(\mathbf{Y}_{t-1}) = 1$, de modo que $X_t = \mu_t + \varepsilon_t$. O modelo com erros multiplicativos tem $r(\mathbf{Y}_{t-1}) = \mu_t$, de modo que $X_t = \mu_t(1 + \varepsilon_t)$. Assim, $\varepsilon_t = (X_t - \mu_t)/\mu_t$ são os erros relativos do modelo multiplicativos. E os erros aditivos não são únicos. Qualquer valor de $r(\mathbf{Y}_{t-1})$ levará a idênticos previsões pontuais para X_t .

Cada modelo da Tabela 5.9, pode ser escrito nas formas dadas pelas equações (5.9) e (5.10). As Figuras 5.24 e 5.25 apresentam os modelo de espaço de estado de inovações aditivas e multiplicativas, respectivamente. Os modelos de erro multiplicativo são obtidos substituindo ε_t por $\mu_t \varepsilon_t$ nas equações na tabela da Figura 5.24. As equações de erro multiplicativos resultantes são fornecidas na tabela da Figura 5.25.

Algumas das combinações de tendência, sazonalidade e erro podem ocasionalmente levar a dificuldades numéricas; especificamente, qualquer modelo de equação que requer a divisão por um componente de estado pode envolver divisão por zero. Isto é um problema para modelos com erros aditivos e tendência multiplicativa ou sazonalidade multiplicativa, bem como o modelo com erros multiplicativos, tendência multiplicativa e sazonalidade aditiva. Portanto, esses modelos devem ser usados com cautela.

Os modelos de erros multiplicativos são úteis quando os dados são estritamente positivos, mas não são numericamente estáveis quando os dados contêm zeros ou valores negativos. Então, quando a série temporal não é estritamente positiva, apenas os seis modelos aditivos podem ser aplicados.

As previsões de pontos fornecidas anteriormente são obtidas a partir desses modelos iterando as equações (5.9) e (5.10) para $t = n + 1, \dots, n + h$, e setando $\varepsilon_{n+j} = 0$, para $j = 1, \dots, h$. Na maioria dos casos (exceções notáveis dos modelos com sazonalidade multiplicativa ou tendência multiplicativa para $h \leq 2$), as previsões pontuais podem ser mostradas ser iguais a

$$\mu_{t+h|t} = \mu_t(t+h) = \mathbb{E}(X_{\mathbf{Y}_t}(t+h))$$

a esperança condicional do respectivo modelo de espaço de estado.

Os modelos também fornecem um meio de obter intervalos de predição. No caso dos modelos lineares, onde as distribuições de predição são gaussianas, podemos derivar a variância condicional

$$v_{t+h|t} = \text{Var}(X_{\mathbf{Y}_t}(t+h))$$

e obtendo assim os intervalos de predição.

Uma abordagem mais direta que funciona para todos os modelos é simplesmente simular muitos caminhos de amostra futuros, condicional à última estimativa do vetor de estado, \mathbf{Y}_t . Então, os intervalos de predição podem ser obtidos a partir dos percentis dos caminhos de amostra simulados. As previsões de pontos também podem ser obtidas desta forma tomando a média dos valores simulados em cada período de tempo futuro. Uma vantagem dessa abordagem é que geramos uma estimativa da distribuição preditiva completa, que é especialmente útil em aplicações como planejamento de estoque, onde os custos esperados dependem de toda a distribuição

Trend	Seasonal		
	N	A	M
N	$\mu_t = \ell_{t-1}$ $\ell_t = \ell_{t-1} + \alpha \varepsilon_t$	$\mu_t = \ell_{t-1} + s_{t-m}$ $\ell_t = \ell_{t-1} + \alpha \varepsilon_t$ $s_t = s_{t-m} + \gamma \varepsilon_t$	$\mu_t = \ell_{t-1} s_{t-m}$ $\ell_t = \ell_{t-1} + \alpha \varepsilon_t / s_{t-m}$ $s_t = s_{t-m} + \gamma \varepsilon_t / \ell_{t-1}$
A	$\mu_t = \ell_{t-1} + b_{t-1}$ $\ell_t = \ell_{t-1} + b_{t-1} + \alpha \varepsilon_t$ $b_t = b_{t-1} + \beta \varepsilon_t$	$\mu_t = \ell_{t-1} + b_{t-1} + s_{t-m}$ $\ell_t = \ell_{t-1} + b_{t-1} + \alpha \varepsilon_t$ $b_t = b_{t-1} + \beta \varepsilon_t$ $s_t = s_{t-m} + \gamma \varepsilon_t$	$\mu_t = (\ell_{t-1} + b_{t-1}) s_{t-m}$ $\ell_t = \ell_{t-1} + b_{t-1} + \alpha \varepsilon_t / s_{t-m}$ $b_t = b_{t-1} + \beta \varepsilon_t / s_{t-m}$ $s_t = s_{t-m} + \gamma \varepsilon_t / (\ell_{t-1} + b_{t-1})$
A _d	$\mu_t = \ell_{t-1} + \phi b_{t-1}$ $\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha \varepsilon_t$ $b_t = \phi b_{t-1} + \beta \varepsilon_t$	$\mu_t = \ell_{t-1} + \phi b_{t-1} + s_{t-m}$ $\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha \varepsilon_t$ $b_t = \phi b_{t-1} + \beta \varepsilon_t$ $s_t = s_{t-m} + \gamma \varepsilon_t$	$\mu_t = (\ell_{t-1} + \phi b_{t-1}) s_{t-m}$ $\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha \varepsilon_t / s_{t-m}$ $b_t = \phi b_{t-1} + \beta \varepsilon_t / s_{t-m}$ $s_t = s_{t-m} + \gamma \varepsilon_t / (\ell_{t-1} + \phi b_{t-1})$
M	$\mu_t = \ell_{t-1} b_{t-1}$ $\ell_t = \ell_{t-1} b_{t-1} + \alpha \varepsilon_t$ $b_t = b_{t-1} + \beta \varepsilon_t / \ell_{t-1}$	$\mu_t = \ell_{t-1} b_{t-1} + s_{t-m}$ $\ell_t = \ell_{t-1} b_{t-1} + \alpha \varepsilon_t$ $b_t = b_{t-1} + \beta \varepsilon_t / \ell_{t-1}$ $s_t = s_{t-m} + \gamma \varepsilon_t$	$\mu_t = \ell_{t-1} b_{t-1} s_{t-m}$ $\ell_t = \ell_{t-1} b_{t-1} + \alpha \varepsilon_t / s_{t-m}$ $b_t = b_{t-1} + \beta \varepsilon_t / (s_{t-m} \ell_{t-1})$ $s_t = s_{t-m} + \gamma \varepsilon_t / (\ell_{t-1} b_{t-1})$
M _d	$\mu_t = \ell_{t-1} b_{t-1}^\phi$ $\ell_t = \ell_{t-1} b_{t-1}^\phi + \alpha \varepsilon_t$ $b_t = b_{t-1}^\phi + \beta \varepsilon_t / \ell_{t-1}$	$\mu_t = \ell_{t-1} b_{t-1}^\phi + s_{t-m}$ $\ell_t = \ell_{t-1} b_{t-1}^\phi + \alpha \varepsilon_t$ $b_t = b_{t-1}^\phi + \beta \varepsilon_t / \ell_{t-1}$ $s_t = s_{t-m} + \gamma \varepsilon_t$	$\mu_t = \ell_{t-1} b_{t-1}^\phi s_{t-m}$ $\ell_t = \ell_{t-1} b_{t-1}^\phi + \alpha \varepsilon_t / s_{t-m}$ $b_t = b_{t-1}^\phi + \beta \varepsilon_t / (s_{t-m} \ell_{t-1})$ $s_t = s_{t-m} + \gamma \varepsilon_t / (\ell_{t-1} b_{t-1}^\phi)$

Figura 5.24: Equações dos modelos de espaços de estados, com erros aditivos. Notação: $m = s$ sazonalidade. Fonte: Hyndman et al. (2008), página 21.

5.10 Estimação dos Parâmetros dos Modelos ETS

Uma alternativa para estimar os parâmetros minimizando a soma dos erros quadrados é maximizar a função de verossimilhança. A verossimilhança é a probabilidade dos dados surgirem do modelo especificado. Assim, uma alta verossimilhança está associada a um bom modelo. Para um modelo de erro aditivo, maximizar a verossimilhança (assumindo erros normalmente distribuídos) dá os mesmos resultados que minimizar a soma dos erros quadrados. No entanto, resultados diferentes serão obtidos para modelos de erros multiplicativos. Nesta seção, iremos estimar os parâmetros de suavização α , β , γ e ϕ , e os valores iniciais ℓ_0 , b_0 e $s_0(1), s_0(2), \dots, s_0(s)$, maximizando a função de verossimilhança.

Os valores possíveis que os parâmetros de suavização podem assumir são restritos. Tradicionalmente, os parâmetros foram restringidos para ficar entre 0 e 1 para que as equações possam ser interpretadas como médias ponderadas. Assim, $0 < \alpha, \beta, \gamma, \phi < 1$.

Trend	Seasonal		
	N	A	M
N	$\mu_t = \ell_{t-1}$ $\ell_t = \ell_{t-1}(1 + \alpha\varepsilon_t)$	$\mu_t = \ell_{t-1} + s_{t-m}$ $\ell_t = \ell_{t-1} + \alpha(\ell_{t-1} + s_{t-m})\varepsilon_t$ $s_t = s_{t-m} + \gamma(\ell_{t-1} + s_{t-m})\varepsilon_t$	$\mu_t = \ell_{t-1}s_{t-m}$ $\ell_t = \ell_{t-1}(1 + \alpha\varepsilon_t)$ $s_t = s_{t-m}(1 + \gamma\varepsilon_t)$
A	$\mu_t = \ell_{t-1} + b_{t-1}$ $\ell_t = (\ell_{t-1} + b_{t-1})(1 + \alpha\varepsilon_t)$ $b_t = b_{t-1} + \beta(\ell_{t-1} + b_{t-1})\varepsilon_t$	$\mu_t = \ell_{t-1} + b_{t-1} + s_{t-m}$ $\ell_t = \ell_{t-1} + b_{t-1} + \alpha(\ell_{t-1} + b_{t-1} + s_{t-m})\varepsilon_t$ $b_t = b_{t-1} + \beta(\ell_{t-1} + b_{t-1} + s_{t-m})\varepsilon_t$ $s_t = s_{t-m} + \gamma(\ell_{t-1} + b_{t-1} + s_{t-m})\varepsilon_t$	$\mu_t = (\ell_{t-1} + b_{t-1})s_{t-m}$ $\ell_t = (\ell_{t-1} + b_{t-1})(1 + \alpha\varepsilon_t)$ $b_t = b_{t-1} + \beta(\ell_{t-1} + b_{t-1})\varepsilon_t$ $s_t = s_{t-m}(1 + \gamma\varepsilon_t)$
A _d	$\mu_t = \ell_{t-1} + \phi b_{t-1}$ $\ell_t = (\ell_{t-1} + \phi b_{t-1})(1 + \alpha\varepsilon_t)$ $b_t = \phi b_{t-1} + \beta(\ell_{t-1} + \phi b_{t-1})\varepsilon_t$	$\mu_t = \ell_{t-1} + \phi b_{t-1} + s_{t-m}$ $\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha(\ell_{t-1} + \phi b_{t-1} + s_{t-m})\varepsilon_t$ $b_t = \phi b_{t-1} + \beta(\ell_{t-1} + \phi b_{t-1} + s_{t-m})\varepsilon_t$ $s_t = s_{t-m} + \gamma(\ell_{t-1} + \phi b_{t-1} + s_{t-m})\varepsilon_t$	$\mu_t = (\ell_{t-1} + \phi b_{t-1})s_{t-m}$ $\ell_t = (\ell_{t-1} + \phi b_{t-1})(1 + \alpha\varepsilon_t)$ $b_t = \phi b_{t-1} + \beta(\ell_{t-1} + \phi b_{t-1})\varepsilon_t$ $s_t = s_{t-m}(1 + \gamma\varepsilon_t)$
M	$\mu_t = \ell_{t-1}b_{t-1}$ $\ell_t = \ell_{t-1}b_{t-1}(1 + \alpha\varepsilon_t)$ $b_t = b_{t-1}(1 + \beta\varepsilon_t)$	$\mu_t = \ell_{t-1}b_{t-1} + s_{t-m}$ $\ell_t = \ell_{t-1}b_{t-1} + \alpha(\ell_{t-1}b_{t-1} + s_{t-m})\varepsilon_t$ $b_t = b_{t-1} + \beta(\ell_{t-1}b_{t-1} + s_{t-m})\varepsilon_t / \ell_{t-1}$ $s_t = s_{t-m} + \gamma(\ell_{t-1}b_{t-1} + s_{t-m})\varepsilon_t$	$\mu_t = \ell_{t-1}b_{t-1}s_{t-m}$ $\ell_t = \ell_{t-1}b_{t-1}(1 + \alpha\varepsilon_t)$ $b_t = b_{t-1}(1 + \beta\varepsilon_t)$ $s_t = s_{t-m}(1 + \gamma\varepsilon_t)$
M _d	$\mu_t = \ell_{t-1}b_{t-1}^\phi$ $\ell_t = \ell_{t-1}b_{t-1}^\phi(1 + \alpha\varepsilon_t)$ $b_t = b_{t-1}^\phi(1 + \beta\varepsilon_t)$	$\mu_t = \ell_{t-1}b_{t-1}^\phi + s_{t-m}$ $\ell_t = \ell_{t-1}b_{t-1}^\phi + \alpha(\ell_{t-1}b_{t-1}^\phi + s_{t-m})\varepsilon_t$ $b_t = b_{t-1}^\phi + \beta(\ell_{t-1}b_{t-1}^\phi + s_{t-m})\varepsilon_t / \ell_{t-1}$ $s_t = s_{t-m} + \gamma(\ell_{t-1}b_{t-1}^\phi + s_{t-m})\varepsilon_t$	$\mu_t = \ell_{t-1}b_{t-1}^\phi s_{t-m}$ $\ell_t = \ell_{t-1}b_{t-1}^\phi(1 + \alpha\varepsilon_t)$ $b_t = b_{t-1}^\phi(1 + \beta\varepsilon_t)$ $s_t = s_{t-m}(1 + \gamma\varepsilon_t)$

Figura 5.25: Equações dos modelos de espaços de estados, com erros multiplicativos. Notação: $m = s$ sazonalidade. Fonte: Hyndman et al. (2008), página 22.

5.11 Critérios de Seleção de Modelos

Uma grande vantagem da estrutura estatística do ETS é que os critérios de informação podem ser usados para a seleção do modelo. AIC, AICc e o BIC, apresentado na Seção 5.3, pode ser usado aqui para determinar qual dos modelos ETS é o mais apropriado para uma determinada série temporal.

5.11.1 Critério da Informação de Akaike

Para modelos ETS, o Critério de Informação de Akaike (AIC) é definido como

$$\text{AIC} = -2 \log(L) + 2k,$$

onde L é a verossimilhança do modelo e k é o número total de parâmetros e estados iniciais que foram estimados (incluindo a variância residual).

5.11.2 Critério da Informação de Akaike Corrigido

O AIC corrigido para o viés da pequena amostra (AICc) é definido como

$$\text{AIC}_c = \text{AIC} + \frac{2k(k+1)}{n-k-1}.$$

5.11.3 Critério da Informação Bayesiana de Schwarz

O Critério de Informação Bayesiana de Schwarz é definido como

$$\text{BIC} = \text{AIC} + k[\log(n) - 2].$$

Tal como acontece com o AIC, minimizar o BIC tem como objetivo fornecer o melhor modelo.

Três das combinações de (erro, tendência, sazonalidade) podem levar a dificuldades numéricas. Especificamente, os modelos que podem causar tais instabilidades são ETS(A,N,M), ETS(A,A,M) e ETS(A,A_d,M), devido à divisão por valores potencialmente próximos de zero nas equações de estado. Normalmente não consideramos essas combinações particulares ao selecionar um modelo.

Os modelos com erros multiplicativos são úteis quando os dados são estritamente positivos, mas não são numericamente estáveis quando os dados contêm zeros ou valores negativos. Portanto, modelos de erro multiplicativos não serão considerados se a série temporal não for estritamente positiva. Nesse caso, apenas os seis modelos totalmente aditivos serão aplicados.

5.12 Previsão com os Modelos ETS

As previsões de pontos são obtidas a partir dos modelos, iterando as equações para $t = n+1, \dots, n+h$, e definindo $\varepsilon_t = 0$, para $t > n$. Por exemplo, para o modelo ETS(M,A,N), $X_{n+1} = (\ell_n + b_n)(1 + \varepsilon_{n+1})$. Portanto, $\hat{X}_n(n+1) = \ell_n + b_n$. De forma similar

$$\begin{aligned} X_{n+2} &= (\ell_{n+1} + b_{n+1})(1 + \varepsilon_{n+2}) \\ &= [(\ell_n + b_n)(1 + \alpha\varepsilon_{n+1}) + b_n + \beta(\ell_n + b_n)\varepsilon_{n+1}](1 + \varepsilon_{n+2}). \end{aligned}$$

Portanto, $\hat{X}_n(n+2) = \ell_n + 2b_n$, e assim por diante. Essas previsões são idênticas às previsões do modelo de Suavização Exponencial de Holt e também às do modelo ETS(A, A, N). Assim, as previsões de pontos obtidas com o método e com os dois modelos que o embasam são idênticas (assumindo que são usados os mesmos valores de parâmetro).

As previsões do ponto ETS são iguais às medianas das distribuições das previsões. Para modelos com apenas componentes aditivos, as distribuições de previsão são normais, portanto, as medianas e as médias são iguais. Para modelos ETS com erros multiplicativos, ou com sazonalidade multiplicativa, as previsões pontuais não serão iguais às médias das distribuições das previsões. Para obter previsões de um modelo ETS, usamos a função `forecast()`.

5.12.1 Intervalos de Confiança de Previsão

Uma grande vantagem dos modelos é que intervalos de predição também podem ser gerados - algo que não pode ser feito usando os métodos. Os intervalos de predição serão diferentes entre os modelos com métodos aditivos e multiplicativos.

Para a maioria dos modelos ETS, um intervalo de predição pode ser escrito como

$$\hat{X}_n(n+h) \pm c\sigma_h,$$

onde $c = z_{\frac{\alpha}{2}}$ depende da probabilidade de cobertura (coeficiente de confiança) $100(1 - \alpha)\%$ e σ_h^2 é a variação da previsão. Para modelos ETS, fórmulas para σ_h^2 pode ser complicado; os detalhes são fornecidos no Capítulo 6 de [Hyndman et al. \(2008\)](#). Na Tabela 5.10, pode ser encontrado algumas fórmulas para os modelos ETS aditivos, que são os mais simples.

5.13 A função `ets()` no R Core Team (2020)

Os modelos podem ser estimados em R usando a função `ets()` no pacote de `forecast`. Ao contrário das funções `ses()`, `holt()` e `hw()`, a função `ets()` não produz previsões. Em vez disso, estima os

Tabela 5.10: Expressões da variância de previsão para cada modelo de espaço de estado aditivo, onde σ^2 é a variância do resíduo, s é a sazonalidade, e k é a parte inteira de $\frac{h-1}{s}$ (ou seja, o número de anos completos no período de previsão antes do tempo $n + h$)

Modelo	Variância de Previsão σ_h^2
(A,N,N)	$\sigma_h^2 = \sigma^2 [1 + \alpha^2(h-1)]$
(A,A,N)	$\sigma_h^2 = \sigma^2 [1 + (h-1) \{ \alpha^2 + \alpha\beta h + \frac{1}{6}\beta^2 h(2h-1) \}]$
(A,A _d ,N)	$\sigma_h^2 = \sigma^2 \left[1 + \alpha^2(h-1) + \frac{\beta\phi h}{(1-\phi)^2} \{2\alpha(1-\phi) + \beta\phi\} - \frac{\beta\phi(1-\phi^h)}{(1-\phi)^2(1-\phi^2)} \{2\alpha(1-\phi^2) + \beta\phi(1+2\phi-\phi^h)\} \right]$
(A,N,A)	$\sigma_h^2 = \sigma^2 [1 + \alpha^2(h-1) + \gamma k(2\alpha + \gamma)]$
(A,A,A)	$\sigma_h^2 = \sigma^2 [1 + (h-1) \{ \alpha^2 + \alpha\beta h + \frac{1}{6}\beta^2 h(2h-1) \} + \gamma k \{ 2\alpha + \gamma + \beta m(k+1) \}]$
(A,A _d ,A)	$\sigma_h^2 = \sigma^2 \left[1 + \alpha^2(h-1) + \gamma k(2\alpha + \gamma) + \frac{\beta\phi h}{(1-\phi)^2} \{2\alpha(1-\phi) + \beta\phi\} - \frac{\beta\phi(1-\phi^h)}{(1-\phi)^2(1-\phi^2)} \{2\alpha(1-\phi^2) + \beta\phi(1+2\phi-\phi^h)\} + \frac{2\beta\gamma\phi}{(1-\phi)(1-\phi^m)} \{k(1-\phi^m) - \phi^m(1-\phi^{mk})\} \right]$

parâmetros do modelo e retorna informações sobre o modelo ajustado. Por padrão, ele usa o AICc para selecionar um modelo apropriado, embora outros critérios de informação possam ser selecionados.

O código R abaixo mostra os argumentos mais importantes que essa função pode assumir e seus valores padrão. Se apenas a série temporal for especificada e todos os outros argumentos forem deixados em seus valores padrão, um modelo apropriado será selecionado automaticamente. Explicamos os argumentos abaixo. Veja o arquivo de ajuda para uma descrição completa.

```

1 ets(y, model="ZZZ", damped=NULL, alpha=NULL, beta=NULL,
2   gamma=NULL, phi=NULL, lambda=NULL, biasadj=FALSE,
3   additive.only=FALSE, restrict=TRUE,
4   allow.multiplicative.trend=FALSE)

```

Parâmetros da função `ets()` do pacote `forecast` do [R Core Team \(2020\)](#).

y

A série temporal a ser analisada.

model

Um código de três letras que indica o modelo a ser estimado usando a classificação e notação ETS. As entradas possíveis são “N” para nenhum, “A” para aditivo, “M” para multiplicativo ou “Z” para seleção automática. Se qualquer uma das entradas for deixada como “Z”, então este componente é

selecionado de acordo com o critério de informação. O valor padrão de “ZZZ” garante que todos os componentes sejam selecionados usando o critério de informação.

`damped`

Se `damped=TRUE`, uma tendência amortecida será usada (A ou M). Se `damped=FALSE`, então uma tendência não amortecida será usada. Se `damped=NULL` (o padrão), então uma tendência amortecida ou não amortecida será selecionada, dependendo de qual modelo tem o menor valor para o critério de informação.

`alpha`, `beta`, `gamma`, `phi`

Os valores dos parâmetros de suavização podem ser especificados usando esses argumentos. Se eles forem definidos como `NULL` (a configuração padrão para cada um deles), os parâmetros são estimados.

`lambda`

Parâmetro de transformação Box-Cox. Ele será ignorado se `lambda = NULL` (o valor padrão). Caso contrário, a série temporal será transformada antes que o modelo seja estimado. Quando `lambda` não é `NULL`, `additive.only` é definido como `TRUE`.

`biasadj`

Se `TRUE` e `lambda` não for `NULL`, os valores ajustados e as previsões transformados de volta serão ajustados pelo viés.

`additive.only`

Somente modelos com componentes aditivos serão considerados se `additive.only=TRUE`. Caso contrário, todos os modelos serão considerados.

`restrict`

Se `restrict=TRUE` (o padrão), os modelos que causam dificuldades numéricas não são considerados na seleção do modelo.

`allow.multiplicative.trend`

Modelos de tendência multiplicativos também estão disponíveis. Defina este argumento como `TRUE` para permitir que esses modelos sejam considerados.

Trabalhando com objetos da classe `ets`

A função `ets()` retornará um objeto de classe `ets`. Existem muitas funções do [R Core Team \(2020\)](#) projetadas para facilitar o trabalho com objetos `ets`. Alguns deles são descritos abaixo.

`coef()`

Retorna todos os parâmetros ajustados.

`accuracy()`

Retorna as medidas de acurácia calculadas utilizando os dados de treinamento.

`summary()`

Imprime algumas informações resumidas sobre o modelo ajustado.

`autoplot()` e `plot()`

Produz gráficos dos componentes.

`residuals()`

Retorna os resíduos do modelo estimado.

`fitted()`

Retorna previsões de um passo para os dados de treinamento.

`simulate()`

Simula uma amostra utilizando o modelo ajustado.

`forecast()`

Calcula previsões pontuais e intervalos de previsão.

5.13.1 Aplicação 1

Agora empregamos a estrutura estatística do ETS para prever o número de turistas internacionais na Austrália no período 1999-2010. Deixamos a função `ets()` selecionar o modelo minimizando o AICc.

```

1 #
2 #Modelos ETS
3 #
4 #Análise da Série temporal do número de turistas internacionais na Austrália.
5 #
6 #library(fpp)
7 aust <- window(austourists, start=1999)
8 #
9 #Modelo
10 mod <- ets(aust)
11 summary(mod)
12 #
13 #Gráfico previsão +IC
14 autoplot(mod)+theme_minimal()
15 mod %>% forecast(h=8) %>%
16   autoplot() +
17   ylab("Número de turistas internacionais na Austrália (millions)") +
18   theme_minimal()
19 #
20 #Comparação resíduos e erros de previsão
21 cbind('Residuals' = residuals(mod),
22       'Forecast errors' = residuals(mod,type='response')) %>%
23   autoplot(facet=TRUE) + xlab("Year") + ylab("")+theme_minimal()
24 #
25 #Resíduo
26 res_ets<-residuals(mod)
27 #
28 #Análise dos resíduos
29 ggtsdisplay(res_ets,plot.type="scatter", theme=theme_bw())
30 #
31 #FAC dos Resíduos
32 ggAcf(res_ets, lag.max=100,type = c("correlation"))+labs(y = "FAC Amostral Resíduos SEH",
33   title="")+
34   theme_minimal()
35 #
36 #QQ Plot dos Resíduos
37 ggqqplot(res_holt)+ggtitle("Resíduo Modelo SEH")
38 #
39 #Densidade dos Resíduos
40 plot(density(res_ets),main="Random Error")

```

```

40 #-----
41 #Teste de Normalidade dos Resíduos
42 shapiro.test(res_ets)
43 #-----
44 #Teste de Correções
45 Box.test(res_ets, lag=10)
46 #-----

```

Código 5.7: Código fonte em R: Função `ets()` para selecionar o modelo ETS minimizando o AICc.

O modelo selecionado foi ETS(M,A,M):

$$\begin{aligned}
 X_t &= (\ell_{t-1} + b_{t-1})s_{t-s}(1 + \varepsilon_t) \\
 \ell_t &= (\ell_{t-1} + b_{t-1})(1 + \alpha\varepsilon_t) \\
 b_t &= b_{t-1} + \beta(\ell_{t-1} + b_{t-1})\varepsilon_t \\
 s_t &= s_{t-s}(1 + \gamma\varepsilon_t).
 \end{aligned}$$

Tabela 5.11: Estimativas dos parâmetros do Modelo de Holt-Winters Multiplicativo para a Série Temporal do número de turistas internacionais na Austrália no período 1999-2010

α	β	γ
0.5153	1.00E-04	5.00E-04
ℓ_0	b_0	$s_0(1)$
24.5569	0.4981	1.0555
$s_0(2)$	$s_0(3)$	$s_0(4)$
0.9471	0.7665	1.2309
AIC	AICc	BIC
276.3074	281.0442	293.1482

Comparando os valores dos coeficientes obtidos na Tabela 5.11 com os coeficientes do Modelo de Holt-Winters equivalente com sazonalidade multiplicativa apresentada na Seção 5.6.6. O modelo ETS (M, A, M) dará diferentes previsões de pontos para o método de Holt-Winters multiplicativo, porque os parâmetros foram estimados de forma diferente. Com a função `ets()`, o método de estimativa padrão é o da máxima verossimilhança, em vez da soma mínima dos quadrados.

A Figura 5.26 mostra os estados ao longo do tempo, a Figura 5.27 mostra previsões de pontos e intervalos de predição gerados a partir do modelo. Os pequenos valores de β e γ significa que a inclinação e os componentes sazonais mudam muito pouco ao longo do tempo. Os estreitos intervalos de previsão indicam que a série é relativamente fácil de prever devido à forte tendência e sazonalidade. A Figura 5.28

Como esse modelo tem erros multiplicativos, os resíduos não são equivalentes aos erros de previsão da etapa de treinamento. Os resíduos são dados por $\hat{\varepsilon}_t$, enquanto os erros de treinamento de uma etapa são definidos como $X_t - \hat{X}_{t-1}(t)$. Podemos obter ambos usando a função `residuals()`. O argumento de type é usado na função `residuals()` para distinguir entre residuais e erros de previsão. O padrão é type = 'innovation' que dá resíduos regulares.

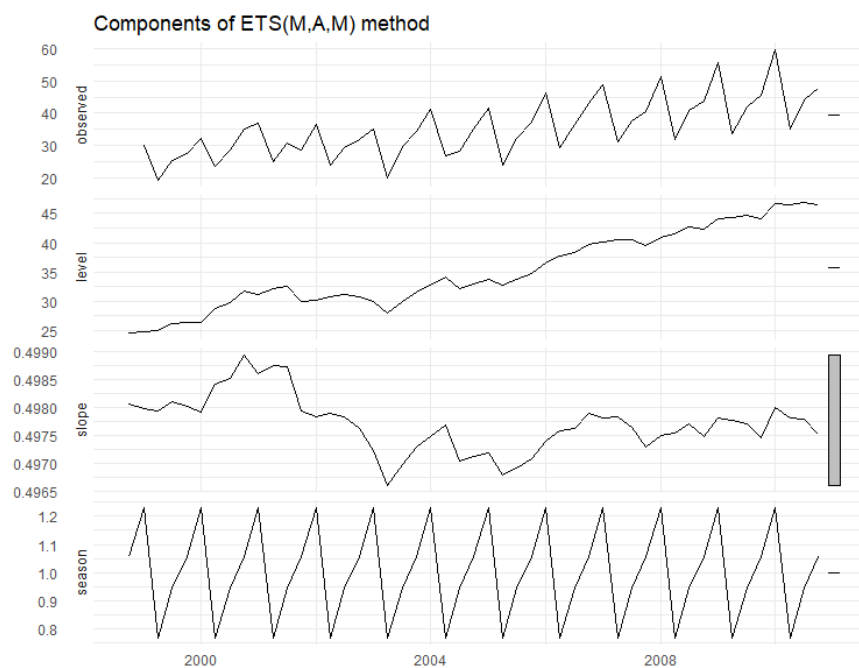


Figura 5.26: Representação gráfica dos estados estimados ao longo do tempo.

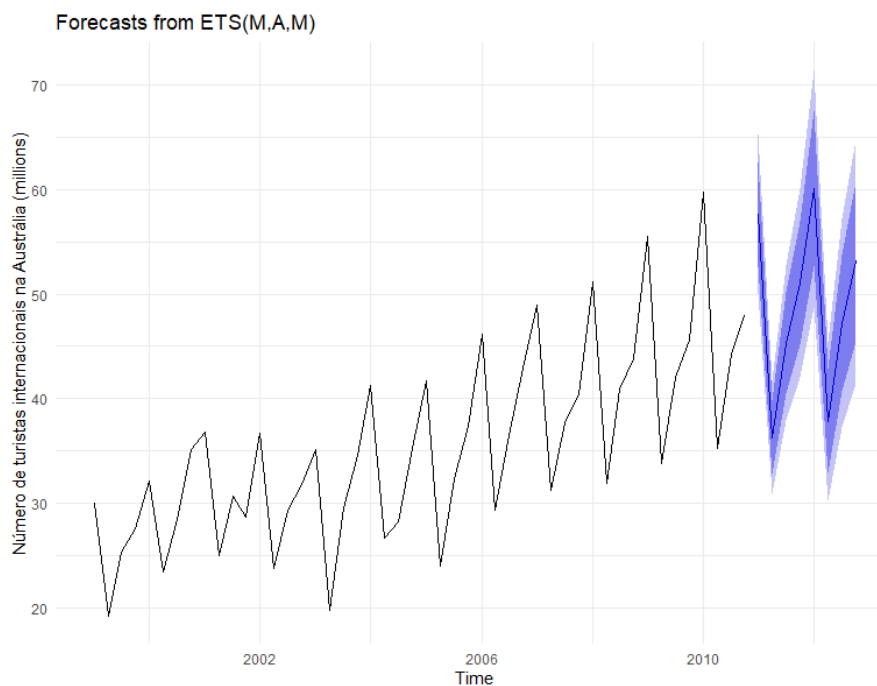


Figura 5.27: Gráficos da Previsão (out-of-sample) da Série Temporal do número de turistas internacionais na Austrália no período 1999-2010 utilizando o Modelo ETS (M, A, M).

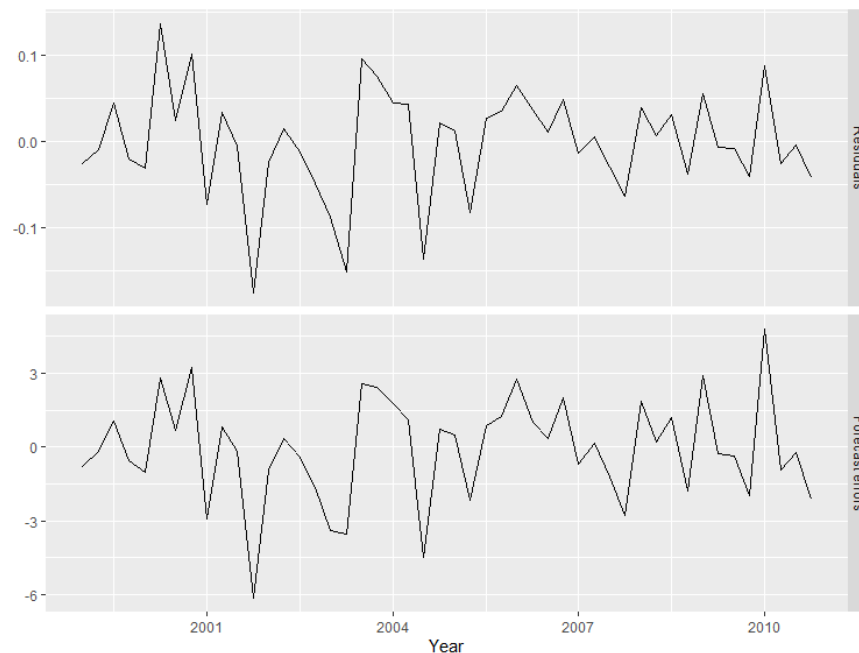


Figura 5.28: Comparação dos resíduos e dos erros de previsão.

5.14 Séries Temporais com Dupla Sazonalidade

5.14.1 Decomposição `mstl()`

A função `mstl()` [<http://bit.ly/2KFL9wg>] do pacote `forecast` pode decompor uma série temporal em componentes sazonais (mais de uma sazonalidade), de tendência e aleatório. Os componentes sazonais são estimados iterativamente usando a função `stl()` (Cleveland et al., 1990) [<http://bit.ly/3hsBa9k>]. Vários períodos sazonais são permitidos. O componente de tendência é calculado para a última iteração de STL. As séries temporais não sazonais são decompostas em tendência e restante apenas. Opcionalmente, a série temporal pode ser transformada em Box-Cox antes da decomposição. Ao contrário da função `stl()`, a função `mstl()` é totalmente automatizado.

Ambas as funções `stl()` e `mstl()` são baseadas no modelo aditivo ($X_t = b_t + s_t + \varepsilon_t$) e utilizam o modelo `loess()` apresentado na seção 3.4.

```
1 taylor %>% autoplot()+theme_minimal()
2 ts(taylor[1:800]) %>% autoplot()+theme_minimal()
3 mstl(taylor) %>% autoplot()+theme_minimal()
```

Código 5.8: Código fonte em R: Função `mstl()` do pacote `forecast`.

A Figura 5.29 apresenta o gráfico da Série Temporal *taylor* que apresenta duas sazonalidades $s_1 = 48$ e $s_2 = 336$ e a Figura 5.30 apresenta o gráfico da sua decomposição utilizando a a função `mstl()`.

Na seção a seguir apresentados o modelo de Holt-Winters para dupla sazonalidade multiplicativa.

5.14.2 Modelo de Holt-Winters com Dupla Sazonalidade Multiplicativa

Embora Holt-Winters padrão seja amplamente utilizado para prever séries temporais sazonais, o método só é capaz de acomodar um padrão sazonal. Uma formulação que pode acomodar mais de um

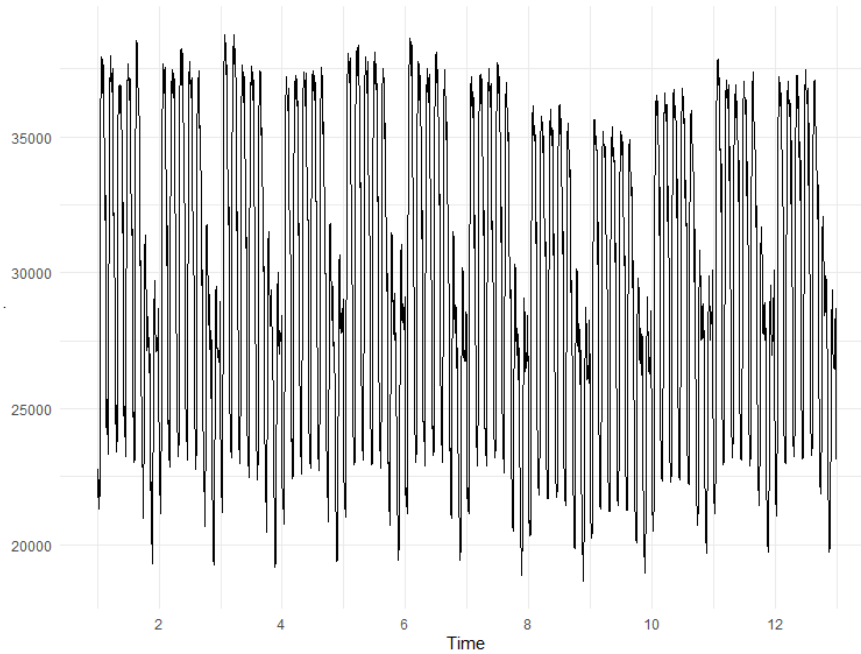


Figura 5.29: Gráfico da Série Temporal *taylor* que apresenta duas sazonalidades $s_1 = 48$ e $s_2 = 336$.

padrão sazonal não foi considerado na literatura de suavização exponencial. Isto é evidente a partir das taxonomias recentes de [Hyndman et al. \(2002\)](#) e [Taylor \(2003a\)](#).

Método de suavização exponencial com dupla sazonalidade (DS) de Taylor ([Taylor, 2003b](#)) foi desenvolvido para prever séries temporais com dois ciclos sazonais: um curto que se repete muitas vezes em um mais longo. Não deve ser confundido com suavização exponencial dupla ([Brown, 1959](#)), cujo foco principal está em uma tendência linear local. [Taylor \(2003b\)](#) desenvolveu um método para sazonalidade multiplicativa (ou seja, maior variação sazonal em valores mais altos de X_t), que foi adaptado para sazonalidade aditiva (ou seja, tamanho da variação sazonal não afetada por o nível de X_t) por [Gould et al. \(2008\)](#).

Denotamos s_1 e s_2 as duas sazonalidades. Então o modelo de Holt-Winters para dupla sazonalidade multiplicativa satisfaz as expressões a seguir

$$\begin{aligned}
 \text{Nível} \quad \ell_t &= \alpha \frac{X_t}{s_{t-s_1}^{(1)} s_{t-s_2}^{(2)}} + (1 - \alpha)(\ell_{t-1} + b_{t-1}) \\
 \text{Tendência} \quad b_t &= \beta(\ell_t - \ell_{t-1}) + (1 - \beta)b_{t-1} \\
 \text{Sazonalidade 1} \quad s_t^{(1)} &= \gamma_1 \frac{X_t}{\ell_t s_{t-s_2}^{(2)}} + (1 - \gamma_1)s_{t-s_1}^{(1)}, \\
 \text{Sazonalidade 2} \quad s_t^{(2)} &= \gamma_2 \frac{X_t}{\ell_t s_{t-s_1}^{(1)}} + (1 - \gamma_2)s_{t-s_2}^{(2)}, \\
 \text{Previsão} \quad \hat{X}_t(k) &= (\ell_t + kb_t)s_{t-s_1+k}^{(1)}s_{t-s_2+k}^{(2)}
 \end{aligned}$$

O modelo é adequado quando há dois padrões sazonais em a série temporal.

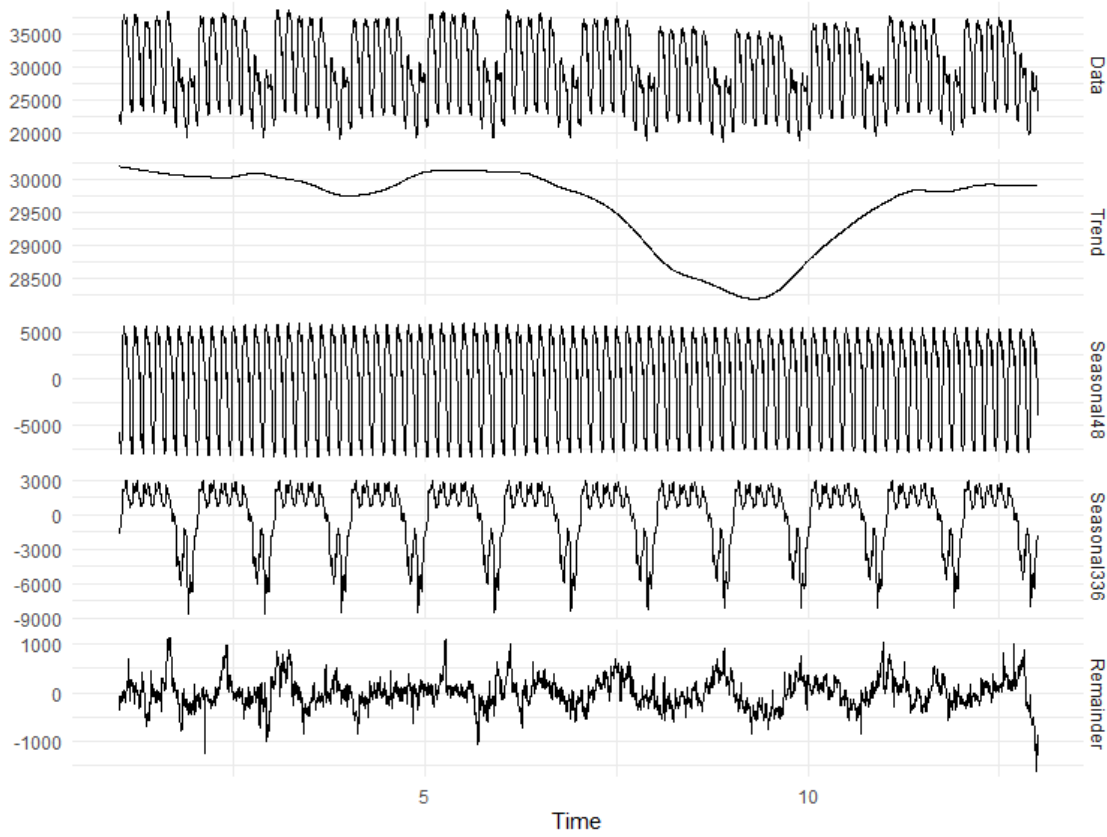


Figura 5.30: Gráfico da decomposição da Série Temporal *taylor* que apresenta duas sazonalidades $s_1 = 48$ e $s_2 = 336$ utilizando a função `mstl()` do pacote `forecast`.

5.14.3 Estimação dos Parâmetros

O vetor de parâmetros do modelo é dado por $\theta = \{\alpha, \beta, \gamma_1, \gamma_2, \ell_0, b_0, s_0^{(1)}, s_0^{(2)}\}$, onde

$$s_0^{(1)} = (s_t^{(1)}, s_1^{(1)}, s_2^{(1)}, \dots, s_{s_1}^{(1)}) \quad \text{e} \quad s_0^{(2)} = (s_t^{(2)}, s_1^{(2)}, s_2^{(2)}, \dots, s_{s_2}^{(2)})$$

são estimados minimizando a soma dos erros de previsão ao quadrados (geralmente conhecido como SSE ou “soma dos erros quadrados”). Da mesma forma, os parâmetros desconhecidos e os valores iniciais para o método de suavização exponencial podem ser estimados minimizando o SSE. Os resíduos são especificados como $e_t = X_t - \hat{X}_{t-1}(t)$, para $t = 1, \dots, n$. Portanto, encontramos os valores dos parâmetros desconhecidos e os valores iniciais que minimizam

$$\text{SSE} = \sum_{t=1}^n \left(X_t - \hat{X}_{t-1}(t) \right)^2 = \sum_{t=1}^n e_t^2. \quad (5.12)$$

Em [Franse \(2017\)](#) o Modelo de Holt-Winters com Dupla Sazonalidade Multiplicativa é aplicado para prever a Demanda de Eletricidade no País de Gales e Inglaterra. Os dados disponíveis para pesquisa consistem em 12 semanas de demanda de eletricidade de meia hora no País de Gales e Inglaterra e foi coletado durante 84 dias entre 5/06/2000 e 27/08/2000.

Para ajuste do Modelo de Holt-Winters com Dupla Sazonalidade Multiplicativa utilizamos a rotina `dshw()` do pacote `forecast` ([R Core Team, 2020](#)).

Modelo de Holt-Winters com Dupla Sazonalidade Multiplicativa - `dshw {forecast}`

Retorna previsões e intervalos de predição usando o modelo Double-Seasonal Holt-Winters de [Taylor \(2003b\)](#).

Comando:

```
dshw(y, period1, period2, h=2*max(period1,period2), alpha=NULL, beta=NULL, gamma=NULL,
omega=NULL, phi=NULL, lambda=NULL, armethod=TRUE)
```

Argumentos:

`y` um vetor numérico ou uma série temporal.

`period1` período sazonal mais curto.

`period2` período sazonal mais longo.

`h` Número de passos para previsão.

`alpha` Parâmetro de suavização para o nível.

`beta` Parâmetro de suavização da tendência

`gamma` Parâmetro de suavização para o primeiro período sazonal.

`omega` Parâmetro de suavização para o segundo período sazonal.

`phi` Parâmetro autoregressivo.

`lambda` Parâmetro de transformação Box-Cox. Ignorado se for NULL. Caso contrário, os dados transformados antes da estimativa do modelo.

`armethod` Se TRUE (padrão), as previsões são ajustadas usando um modelo AR(1) para os erros.

Detalhes:

O método de dupla sazonalidade de [Taylor \(2003b\)](#) de Holt-Winters usa tendência aditiva e sazonalidade multiplicativa, onde existem dois componentes sazonais que são multiplicados juntos. Por exemplo, com uma série de dados de meia hora, definir-se-ia `periodo1 = 48` para o período diário e `periodo2 = 336` para o período semanal. A notação de parâmetro de suavização usada aqui é diferente daquela em [Taylor \(2003b\)](#); em vez disso, corresponde ao usado em [Hyndman et al. \(2008\)](#) e ao usado para a função `ets()`.

Valores:

Um objeto da classe *forecast*.

A função `summary()` é usada para obter e imprimir um resumo dos resultados, enquanto a função `plot()` produz um gráfico das previsões e intervalos de predição.

As funções `fitted.values()` e `residuals()` extraem características úteis do valor retornado por `meanf`.

Um objeto da classe *forecast* é uma lista contendo pelo menos os seguintes elementos:

`model` Uma lista contendo informações sobre o modelo ajustado.

method O nome do método de previsão como uma string de caracteres.

mean previsões como uma série temporal.

lower limite inferior do intervalo de confiança de previsão.

upper limite superior do intervalo de confiança de previsão.

level coeficiente de confiança do intervalo de confiança de previsão.

× A série temporal original (o próprio objeto ou a série temporal usada para criar o modelo armazenado como objeto)

residuals Resíduos do modelo ajustado. Isso é x menos os valores ajustados.

fitted Valores ajustados (previsões de um passo a frente)

Maiores detalhes: <http://bit.ly/3rCpp4J> e <http://bit.ly/38Rxtpe>.

5.14.4 Aplicação

Nesta seção apresentamos a aplicação dos modelos de Holt-Winters com Dupla Sazonalidade Multiplicativa utilizamos a rotina `dshw()` do pacote `forecast` (R Core Team, 2020). A Figura 5.29 apresenta o gráfico da Série Temporal *taylor* que apresenta duas sazonalidades $s_1 = 48$ e $s_2 = 336$. A Figura 5.31 apresenta o gráfico da Predição/Previsão da Série Temporal *taylor*.

A Tabela 5.12 apresenta as estimativas dos parâmetros e medidas de acurácia de predição da Série Temporal *taylor* que apresenta duas sazonalidades $s_1 = 48$ e $s_2 = 336$ utilizando a função `dshw()` do pacote `forecast`. As estimativas de $s_0^{(1)}$ e $s_0^{(2)}$ não foram apresentadas pois teríamos que incluir duas tabelas grandes, uma com 48 entradas e outra com 336 entradas.

Tabela 5.12: Estimativas dos parâmetros e medidas de acurácia de predição da Série Temporal *taylor* que apresenta duas sazonalidades $s_1 = 48$ e $s_2 = 336$ utilizando a função `dshw()` do pacote `forecast`

α	β	γ_1	γ_2	ℓ_0	b_0
0.433461	5.76E-09	0.4181925	0.5556816	29296.17	2.258012
ϕ_1	RMSE	MAE	MPE	MAPE	MASE
0.586139	159.6836	113.6605	-0.00988172	0.389437	0.200419

```

1 #Modelo DSHW
2 mod<-dshw(taylor)
3 mod %>% autoplot()+theme_minimal()
4 summary(mod)
5 #
6 clrs <- c("darkgreen","black", "blue3", "darkred")
7 autoplot(mod) +
8   autolayer(taylor, series="Data") +
9   autolayer(fitted(mod), series="Fitted")+
10  autolayer(mod$mean, series="Forecast") +
11  xlab("Tempo") +
12  ylab("Taylor") +
13  ggtitle("Predição—Previsão — Modelo DS Holt—Winters")+
14  guides(colour=guide_legend(title="Data Series"))+

```

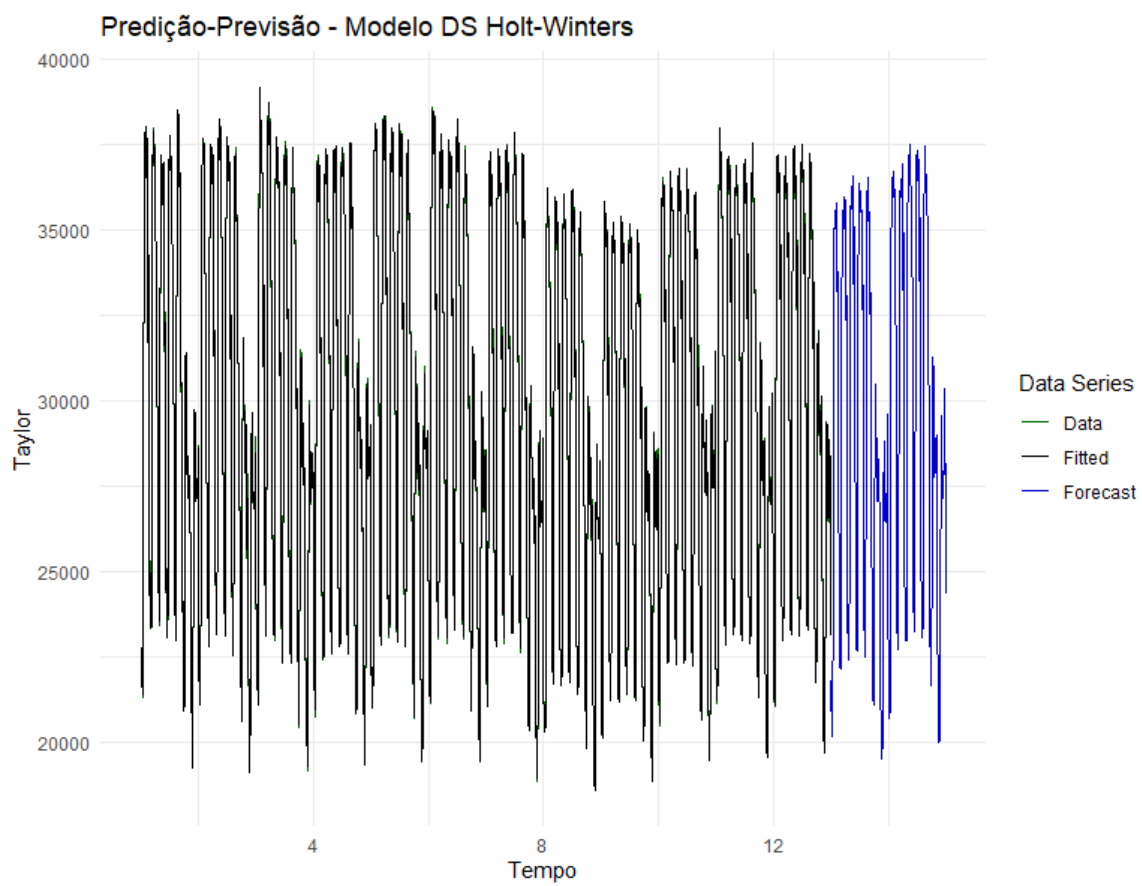


Figura 5.31: Gráfico da Predição/Previsão da Série Temporal *taylor* que apresenta duas sazonalidades $s_1 = 48$ e $s_2 = 336$ utilizando a função `dshw()` do pacote `forecast`.

```
15 | scale_color_manual(values=clrs)+theme_minimal()
```

Código 5.9: Código fonte em R: Função `dshw()` do pacote `forecast`.

Capítulo 6

Outliers

6.1 Introdução

Este capítulo é baseado no artigo *Outliers detection in R* [<http://bit.ly/2WIOwql>] de Antoine Soetewey do blog Stats and R [<https://statsandr.com/>].

Um *outlier* é um valor ou uma observação que está distante de outras observações, ou seja, um ponto de dados que difere significativamente de outros pontos de dados. Uma observação deve sempre ser comparada a outras observações feitas no mesmo fenômeno antes de realmente chamá-lo de *outlier*. Na verdade, alguém com 200 cm de altura provavelmente será considerado um *outlier* em comparação com a população em geral, mas essa mesma pessoa pode não ser considerada um *outlier* se medirmos a altura de jogadores de basquete.

Um *outlier* pode ser devido à variabilidade inerente ao fenômeno observado. Por exemplo, muitas vezes ocorrem discrepâncias ao coletar dados sobre salários, pois algumas pessoas ganham muito mais dinheiro do que as outras. *Outlier* também podem surgir devido a um erro experimental, de medição ou codificação. Por exemplo, um peso humano de 786 kg (1733 libras) é claramente um erro ao codificar o peso do sujeito. Seu peso é provavelmente 78,6 kg (173 libras) ou 7,86 kg (17 libras), dependendo se o peso de adultos ou bebês foi medido.

Neste capítulo, são apresentadas várias abordagens para detectar *outliers* no R Core Team (2020), desde técnicas simples, como estatística descritiva [<http://bit.ly/3pBw9yj>] (incluindo mínimo, máximo, histograma, boxplot e percentis) até técnicas mais formais como o filtro de Hampel, o Grubbs, o Dixon e o Rosner testa para *outliers*.

Embora não haja uma regra estrita ou exclusiva sobre se os valores discrepantes devem ser removidos ou não do conjunto de dados antes de fazer análises estatísticas, é bastante comum, pelo menos, remover valores discrepantes que são devido a um erro experimental ou de medição (como o peso de 786 kg (1.733 libras) para um humano). Alguns testes estatísticos exigem a ausência de *outliers* para tirar conclusões sólidas, mas a remoção de *outliers* não é recomendada em todos os casos e deve ser feita com cautela.

Este artigo não dirá se você deve remover *outliers* ou não (nem se deve imputá-los à mediana, média, modo ou qualquer outro valor), mas ajudará você a detectá-los para, como uma primeira etapa, verifique-os. Após sua verificação, é sua escolha excluí-los ou incluí-los em suas análises. Remover ou manter *outliers* depende principalmente de três fatores

- (1) O domínio e o contexto de suas análises. Em alguns domínios, é comum remover valores discrepantes, pois eles ocorrem frequentemente devido a um processo com defeito. Em outros campos, os valores discrepantes são mantidos porque contêm informações valiosas. Também ocorre que as análises são realizadas duas vezes, uma com e outra sem *outliers* para avaliar seu impacto nas conclusões.

- (2) Se os testes que você vai aplicar são robustos à presença de *outliers* ou não. Por exemplo, a inclinação de uma regressão linear simples pode variar significativamente com apenas um *outliers*, enquanto testes não paramétricos como o teste de Wilcoxon são geralmente robustos para *outliers*.
- (3) Quão distantes estão os *outliers* de outras observações. Algumas observações consideradas outliers (de acordo com as técnicas apresentadas abaixo) não são realmente extremas em comparação com todas as outras observações, enquanto outros *outliers* potenciais podem estar realmente distantes do resto das observações.

O conjunto de dados `mpg` do pacote `ggplot2` será utilizado para ilustrar diferentes técnicas para detecção de *outliers* no [R Core Team \(2020\)](#), em particular usaremos a variável `hwy` (highway miles per gallon).

6.2 Estatística Descritiva

6.2.1 Máximo e Mínimo

O primeiro passo para detectar *outliers* em R é começar com algumas estatísticas descritivas e, em particular, com o mínimo e o máximo.

No [R Core Team \(2020\)](#), isso pode ser feito facilmente com a função `summary()` onde o mínimo e o máximo são, respectivamente, o primeiro e o último valores na saída acima. Alternativamente, eles também podem ser calculados com as funções `min()` e `max()`.

```

1 dat <- ggplot2::mpg
2 summary(dat$hwy)
3 ##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
4 ##      12.00   18.00   24.00   23.44   27.00   44.00
5 #-----
6 min(dat$hwy)
7 ## [1] 12
8 max(dat$hwy)
9 ## [1] 44

```

Código 6.1: Código fonte em R: Estatística Descritiva.

Algun erro de codificação claro, como um peso de 786 kg (1733 libras) para um humano, já será facilmente detectado por esta técnica muito simples.

6.2.2 Histograma

Outra maneira básica de detectar *outliers* é desenhar um histograma dos dados.

Um Histograma dos dados pode ser elaborado usando o pacote `stats` base [R Core Team \(2020\)](#) (com o número de classes correspondendo à raiz quadrada do número de observações, a fim de ter mais classes do que a opção padrão) ou utilizando o pacote `ggplot2`.

```

1 hist(dat$hwy,
2     xlab = "hwy",
3     main = "Histogram of hwy",
4     breaks = sqrt(nrow(dat))
5 ) # set number of bins
6 #-----
7 ggplot(dat) +
8   aes(x = hwy) +
9   geom_histogram(bins = 30L, fill = "#0c4c8a") +
10  theme_minimal()

```

Código 6.2: Código fonte em R: Histograma.

A partir do histograma, parece haver algumas observações mais altas do que todas as outras observações (veja a barra do lado direito do gráfico), ver Figura 6.1.

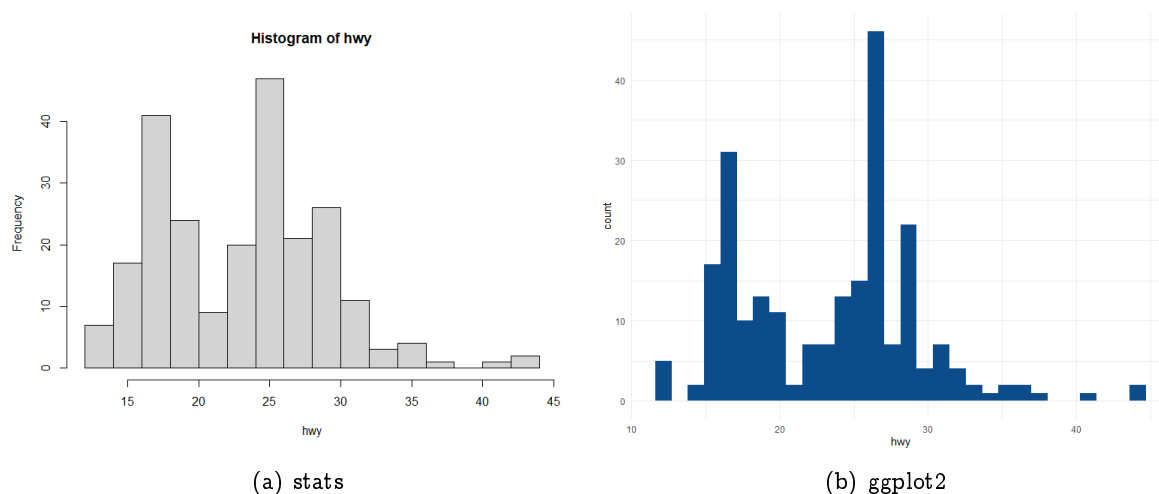


Figura 6.1: Histogramas elaborados pelos pacotes *stats* e *ggplot2*, respectivamente, da variável *hwy*.

6.2.3 Boxplot

Além dos histogramas, os boxplots também são úteis para detectar possíveis *outliers*.

A Figura 6.2 apresenta os boxplots elaborados utilizando os pacotes *stats* e *ggplot2*, respectivamente, da variável *hwy*.

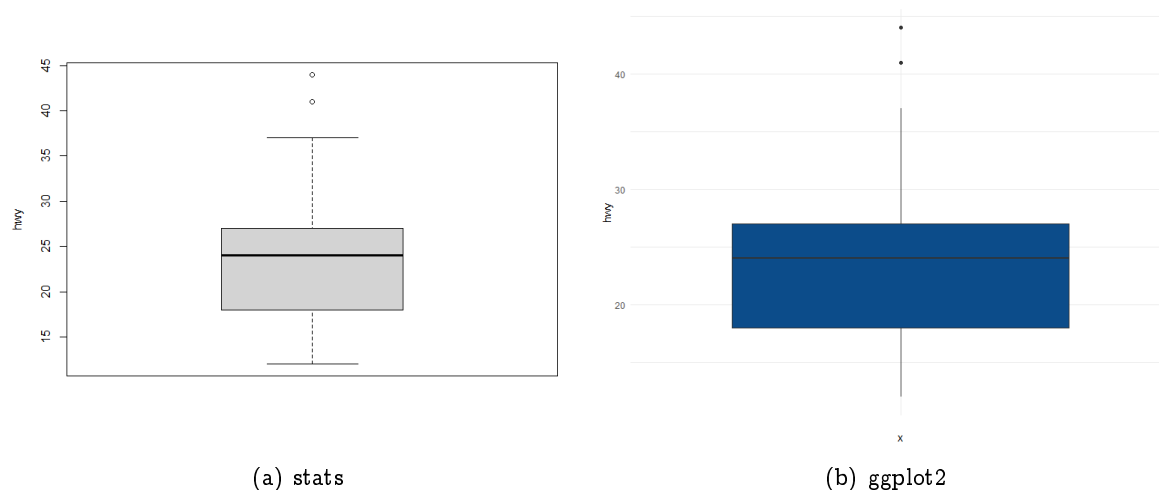


Figura 6.2: Boxplots elaborados utilizando os pacotes *stats* e *ggplot2*, respectivamente, da variável *hwy*.

Um boxplot ajuda a visualizar uma variável quantitativa exibindo cinco resumos de localização comuns (mínimo, mediana, primeiro e terceiro quartis e máximo) e qualquer observação que foi classificada como um *outlier* suspeito usando o critério de intervalo interquartil (IQR). O critério IQR significa que todas as observações acima de $q_{0,75} + 1,5\text{IQR}$ ou abaixo de $q_{0,25} - 1,5\text{IQR}$ (onde $q_{0,25}$ e $q_{0,75}$ correspondem ao primeiro e terceiro quartil respectivamente, e IQR é a diferença entre o terceiro e primeiro quartil) são considerados potenciais *outliers* por R. Em outras palavras, todas as

observações fora do intervalo seguinte serão consideradas potenciais *outliers*:

$$I = [q_{0.25} - 1.5 \cdot \text{IQR}; q_{0.75} + 1.5 \cdot \text{IQR}]$$

As observações consideradas como *outliers* potenciais pelo critério IQR são exibidas como pontos no boxplot. Com base neste critério, existem 2 *outliers* potenciais (consulte os 2 pontos acima da linha vertical, no topo do boxplot).

Lembre-se de que não é porque uma observação é considerada um outlier potencial pelo critério IQR que você deve removê-la. Remover ou manter um *outliers* depende

- (i) do contexto de sua análise,
- (ii) se os testes que você vai realizar no conjunto de dados são robustos para *outliers* ou não,
- (iii) longe está o *outlier* de outras observações.

Também é possível extrair os valores dos *outliers* potenciais com base no critério IQR, graças à função `boxplot.stats()$out`.

Como você pode ver, existem na verdade 3 pontos considerados como *outliers* potenciais: 2 observações com um valor de 44 e 1 observação com um valor de 41.

Graças à função `which()`, é possível extrair o número da linha correspondente a esses *outliers*. Com essas informações, agora você pode facilmente voltar às linhas específicas no conjunto de dados para verificá-las ou imprimir todas as variáveis para esses *outliers*. Também é possível imprimir os valores dos *outliers* diretamente no boxplot com a função `mtext()`.

A Figura 6.3 apresenta o boxplot com os possíveis *outliers* da variável *hwy* impressos no sub título do gráfico.

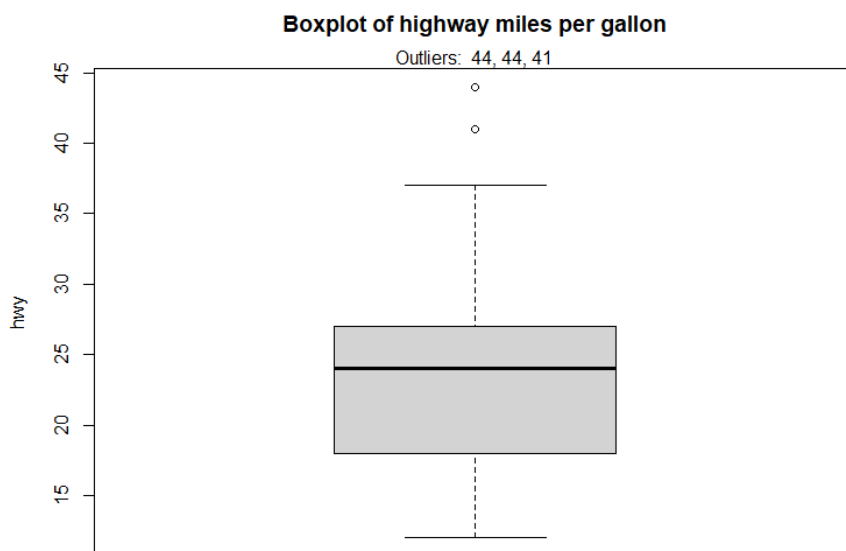


Figura 6.3: Boxplot com os possíveis *outliers* da variável *hwy*.

```
1 boxplot(dat$hwy,
2         ylab = "hwy"
3 )
```

```

4 #
5 ggplot(dat) +
6   aes(x = "", y = hwy) +
7   geom_boxplot(fill = "#0c4c8a") +
8   theme_minimal()
9 #
10 boxplot.stats(dat$hwy)$out
11 #
12 out <- boxplot.stats(dat$hwy)$out
13 out_ind <- which(dat$hwy %in% c(out))
14 out_ind
15 #
16 boxplot(dat$hwy,
17         ylab = "hwy",
18         main = "Boxplot of highway miles per gallon"
19 )
20 mtext(paste("Outliers: ", paste(out, collapse = ", ")))

```

Código 6.3: Código fonte em R: Boxplot.

6.2.4 Percentis

Este método de detecção de *outliers* é baseado em percentis. Com o método dos percentis, todas as observações que estiverem fora do intervalo formado pelos percentis 2,5 e 97,5 serão consideradas potenciais *outliers*. Outros percentis, como 1 e 99, ou 5 e 95, também podem ser considerados para construir o intervalo.

Os valores dos percentis inferior e superior (e, portanto, dos limites inferior e superior do intervalo) podem ser calculados com a função `quantil()`.

De acordo com este método, todas as observações abaixo de 14 e acima de 35,175 serão consideradas potenciais *outliers*. Os números das linhas das observações fora do intervalo podem então ser extraídos com a função `which()`.

Existem 11 outliers potenciais de acordo com o método dos percentis. Para reduzir esse número, você pode definir os percentis para 1 e 99. Definir os percentis como 1 e 99 dá os mesmos valores discrepantes potenciais do critério IQR.

```

1 lower_bound <- quantile(dat$hwy, 0.025)
2 lower_bound
3 upper_bound <- quantile(dat$hwy, 0.975)
4 upper_bound
5 #
6 outlier_ind <- which(dat$hwy < lower_bound | dat$hwy > upper_bound)
7 outlier_ind
8 #
9 dat[outlier_ind, "hwy"]
10 #
11 dat[outlier_ind, ]
12 #
13 lower_bound <- quantile(dat$hwy, 0.01)
14 upper_bound <- quantile(dat$hwy, 0.99)
15 outlier_ind <- which(dat$hwy < lower_bound | dat$hwy > upper_bound)
16 dat[outlier_ind, ]

```

Código 6.4: Código fonte em R: Percentil.

6.2.5 Filtro de Hampel

Outro método, conhecido como filtro de Hampel, consiste em considerar como *outliers* os valores fora do intervalo (I) formado pela mediana, mais ou menos 3 desvios absolutos medianos (MAD).

$$I = [\text{median} - 3 \times \text{MAD}; \text{median} + 3 \times \text{MAD}]$$

onde MAD é o desvio absoluto médio e é definido como a mediana dos desvios absolutos da mediana dos dados $\tilde{X} = \text{median}(X)$ dado por

$$\text{MAD} = \text{median}(|X_i - \tilde{X}|)$$

Para este método, primeiro definimos os limites do intervalo graças às funções `median()` e `mad()`. De acordo com este método, todas as observações abaixo de 1.761 e acima de 46.239 serão consideradas potenciais *outliers*. Os números das linhas das observações fora do intervalo podem então ser extraídos com a função `which()`. De acordo com o filtro de Hampel, não há *outlier* potencial para a variável `hwy`.

```
1 lower_bound <- median(dat$hwy) - 3 * mad(dat$hwy)
2 lower_bound
3 upper_bound <- median(dat$hwy) + 3 * mad(dat$hwy)
4 upper_bound
5 #
6 outlier_ind <- which(dat$hwy < lower_bound | dat$hwy > upper_bound)
7 outlier_ind
```

Código 6.5: Código fonte em R: Filtro de Hampel.

6.3 Testes Estatísticos

Vários testes formais de *outliers* foram propostos na literatura. Eles podem ser agrupados pelas seguintes características:

- (i) Qual é o modelo de distribuição dos dados? Restringimos nossa discussão a testes que assumem que os dados seguem uma distribuição aproximadamente normal.
- (ii) O teste foi projetado para um único valor discrepante ou para vários valores discrepantes?
- (iii) Se o teste for projetado para vários *outliers*, o número de *outliers* precisa ser especificado exatamente ou podemos especificar um limite superior para o número de *outliers*?

A seguir estão alguns dos testes de *outliers* mais comumente usados para dados normalmente distribuídos. Esta lista não é exaustiva (um grande número de testes *outliers* foi proposto na literatura). Os testes aqui apresentados baseiam-se essencialmente no critério de “distância da média”. Este não é o único critério que pode ser usado. Por exemplo, o teste de Dixon, que não é discutido aqui, baseia-se em um valor muito grande (ou pequeno) em comparação com seu vizinho mais próximo.

Nesta seção, apresentamos algumas técnicas mais formais para detectar *outliers*:

- (i) Teste de Grubbs
- (ii) Teste de Dixon
- (iii) Teste de Rosner

Esses 3 testes estatísticos fazem parte de técnicas mais formais de detecção de *outliers*, pois todos envolvem o cálculo de uma estatística de teste que é comparada aos valores críticos tabulados (que são baseados no tamanho da amostra e no nível de confiança desejado).

Observe que os 3 testes são apropriados apenas quando os dados (sem quaisquer *outliers*) são distribuídos de forma aproximadamente normal. A suposição de normalidade deve, portanto, ser verificada antes de aplicar esses testes para *outliers* (veja como testar a suposição de normalidade no R Core Team (2020) - <http://bit.ly/3hjEtQ1>).

6.3.1 Teste de Normalidade de Doornik-Hansen

[<http://bit.ly/3mW6fU6>] Nesta seção apresentamos o Teste de Normalidade de Doornik e Hansen (2008) que leva em consideração que as variáveis são fracamente dependentes em vez de independentes. O teste foi sugerido implicitamente por Lobato e Velasco (2004).

É mais poderoso do que o teste de Shapiro-Wilk para a maioria das distribuições multivariadas testadas. A estatística de teste Doornik-Hansen deriva de Shenton e Bowman (1977) e usa versões transformadas de assimetria e curtose.

A transformação para o coeficiente de assimetria s em z_1 é como em D'Agostino (1970)

$$\beta = \frac{3(n^2 + 27n - 70)(n + 1)(n + 3)}{(n - 2)(n + 5)(n + 7)(n + 9)}, \quad \omega^2 = -1 + \sqrt{2(\beta - 1)},$$

$$\delta = \frac{1}{\sqrt{\log(\omega^2)}}, \quad y = s \sqrt{\frac{(\omega^2 - 1)(n + 1)(n + 3)}{12(n - 2)}},$$

$$z_1 = \delta \log(y + \sqrt{y^2 - 1}).$$

A curtose k é transformada de uma distribuição Gama para χ^2 , que é então transformado em z_2 normal padrão usando a transformação de raiz cúbica de Wilson-Hilferty

$$\delta = (n - 3)(n + 1)(n^2 + 15n - 4), \quad a = \frac{(n - 2)(n + 5)(n + 7)(n^2 + 27n - 70)}{6\delta},$$

$$c = \frac{(n - 7)(n + 5)(n + 7)(n^2 + 2n - 5)}{6\delta}, \quad l = \frac{(n + 5)(n + 7)(n^3 + 37n^2 + 11n - 313)}{12\delta},$$

$$\alpha = a + c \times s^2, \quad \chi = 2l(k - 1 - s^2),$$

$$z_2 = \sqrt{9\alpha} \left(\frac{1}{9\alpha} - 1 + \sqrt[3]{\frac{\chi}{2\alpha}} \right).$$

Por fim, a estatística de teste de Doornik-Hansen é definida como a soma das transformações quadradas da assimetria e curtose. Aproximadamente, a estatística de teste segue uma distribuição χ^2 , ou seja:

$$DH = z_1^2 + z_2^2 \sim \chi^2(2).$$

Tal teste é realizado pela função `normality.test2` do pacote `normwhn.test` [<https://bit.ly/38G7Vf4>] no R Core Team (2020).

Teste de Normalidade de Omnibus - `normality.test2 {normwhn.test}`

Comando:

normality.test2(x)

Argumentos:

1. `x` Matriz de entrada por linha n (observações) e coluna p (variáveis). No caso univariado, a matriz de entrada é a linha n (observações) por 1.

Valores: Uma lista com a classe `hstest` contendo os seguintes componentes:

1. `sk` Coeficiente de Assimetria
2. `k` Coeficiente de Curtose
3. `rtb1` Coeficiente de Assimetria das variáveis padronizadas
4. `b2` Coeficiente de Curtose das variáveis padronizadas
5. `z1` Coeficiente de Assimetria das variáveis transformadas
6. `z2` Coeficiente de Curtose das variáveis transformadas
7. `pvalsk` p-valor sob H_0 : a série temporal é não simétrica
8. `pskneg` p-valor sob H_0 : a série temporal não apresenta simetria negativa
9. `pskpos` p-valor sob H_0 : a série temporal não apresenta simetria positiva
10. `pvalk` p-valor sob H_0 : a série temporal não apresenta curtose
11. `pkneg` p-valor sob H_0 : a série temporal não apresenta curtose negativa
12. `pkpos` p-valor sob H_0 : a série temporal não apresenta curtose positiva
13. `Ep` valor da estatística de teste
14. `dof` graus de liberdade
15. `Sig.Ep` significância da estatística de teste

```
1 ts.sim <- arima.sim(list(order = c(1,1,0), ar = 0.7), n = 200)
2 autoplot(ts.sim)
3 ggAcf(ts.sim, lag.max=100, type = c("correlation")) + labs(y = "FAC Amostral", title="") +
4   theme_minimal()
5 shapiro.test(ts.sim)
6 jarque.test(as.vector(ts.sim))
7 normality.test2(as.matrix(ts.sim))
```

Código 6.6: Código fonte em R: Teste de Normalidade para Séries Temporais.

6.3.2 Teste de Normalidade de Anderson-Darling

O teste de Anderson-Darling ([Stephens, 1974](#)) é usado para testar se uma amostra de dados veio de uma população com uma distribuição específica. É uma modificação do teste de Kolmogorov-Smirnov (K-S) e dá mais peso às caudas do que o teste K-S. O teste K-S é livre de distribuição no sentido de que os valores críticos não dependem da distribuição específica que está sendo testada (observe

que isso é verdadeiro apenas para uma distribuição totalmente especificada, ou seja, os parâmetros são conhecidos). O teste de Anderson-Darling faz uso da distribuição específica no cálculo de valores críticos. Isso tem a vantagem de permitir um teste mais sensível e a desvantagem de que valores críticos devem ser calculados para cada distribuição. Atualmente, as tabelas de valores críticos estão disponíveis para as distribuições Normal, Uniforme, Lognormal, Exponencial, Weibull, Valor Extremo Tipo I, Pareto Generalizado e distribuição Logística. Para as tabelas de valores críticos ver Stephens (1974), 1976, 1977 e 1979.

As hipóteses do teste de Anderson-Darling são definidas como:

H_0 : Os dados seguem uma dist. especificada vs H_1 : Os dados não seguem a dist. especificada

A estatística do teste de Anderson-Darling é definida como

$$A = -n - \frac{1}{n} \sum_{i=1}^n [2i - 1] [\ln(p_{(i)}) + \ln(1 - p_{(n-i+1)})],$$

onde $p_{(i)} = F_Z([x_{(i)} - \bar{x}]/s)$, onde $F_Z(\cdot)$ é a função de distribuição cumulativa da distribuição normal padrão, e \bar{x} e s são a média e o desvio padrão dos valores dos dados. O p-valor é calculado a partir da estatística modificada $Z = A(1.0 + \frac{0.75}{n} + \frac{2.25}{n^2})$.

Teste de Anderson-Darling - `ad.test {nortest}`

Executa o teste de Anderson-Darling para a hipótese composta de normalidade.

Comando:

`ad.test(x)`

Argumentos:

x um vetor numérico de valores de dados, cujo número deve ser maior que 7. Valores ausentes são permitidos.

Maiores detalhes: <http://bit.ly/3nXuoL8>.

O pacote `nortest` [<http://bit.ly/3nY9awJ> ou <https://bit.ly/37VqrB3>] do R Core Team (2020) contém outros testes de normalidade, como o Teste de Lilliefors (Kolmogorov-Smirnov) [<http://bit.ly/3pA2cyx>], o Teste qui-quadrado de Pearson [<http://bit.ly/2WTL2j3>], o Teste de Cramer-von Mises [<http://bit.ly/3aPG8Mg>] e o Teste de Shapiro-Francia [<http://bit.ly/2KH17w6>].

6.3.3 Teste de Grubbs

[<https://bit.ly/2JpbmOU>] O teste de Grubbs (Grubbs (1969) e Stefansky (1972)) é usado para detectar um único valor discrepante em um conjunto de dados univariado que segue uma distribuição aproximadamente normal.

Se você suspeitar que mais de um *outlier* pode estar presente, é recomendado que você use o teste de Tietjen-Moore ou o teste de desvio estudentizado extremo generalizado em vez do teste de Grubbs.

O teste de Grubbs também é conhecido como teste residual máximo normatizado.

As hipóteses do teste de Grubbs são definidas como:

H_0 : Não existe *outliers* nos dados versus H_1 : Existe exatamente um *outlier* nos dados

A estatística do teste de Grubbs é definida como:

$$G = \frac{\max_i |X_i - \bar{X}|}{s} \quad (6.1)$$

com \bar{X} e s denotando a média e o desvio padrão da amostra, respectivamente. A estatística de teste de Grubbs é o maior desvio absoluto da média da amostra em unidades do desvio padrão da amostra. Esta é a versão bilateral do teste. O teste de Grubbs também pode ser definido como um dos seguintes testes unilaterais:

- (1) para testar se o valor mínimo é um *outlier*

$$G = \frac{\bar{X} - X_{(1)}}{s}$$

onde $X_{(1)}$ denota o mínimo dos dados.

- (2) para testar se o valor máximo é um *outlier*

$$G = \frac{X_{(n)} - \bar{X}}{s}$$

onde $X_{(n)}$ denota o máximo dos dados.

Para um teste bilateral, a hipótese de não existir *outliers* é rejeitada de

$$G > \frac{(n-1)}{\sqrt{n}} \sqrt{\frac{(t_{\alpha/(2n), n-2})^2}{n-2 + (t_{\alpha/(2n), n-2})^2}}$$

onde $t_{\alpha/(2n), n-2}$ denota o valor crítico da distribuição t_{n-2} e nível de significância de $\frac{\alpha}{2n}$.

Como para qualquer teste estatístico, se o p-valor for menor que o limite de significância escolhido (geralmente $\alpha = 0,05$), então a hipótese nula é rejeitada e concluiremos que o valor mais baixo / mais alto é um *outlier*. Pelo contrário, se o p-valor for maior ou igual ao nível de significância, a hipótese nula não é rejeitada, e concluiremos que, com base nos dados, não rejeitamos a hipótese de que o menor / maior valor não é um *outlier*.

Observe que o teste de Grubbs não é apropriado para tamanho de amostra de 6 ou menos ($n \leq 6$).

Questões: O teste de Grubbs pode ser usado para responder às seguintes perguntas: O valor máximo é um *outlier*? O valor mínimo é um *outlier*?

Importância: Muitas técnicas estatísticas são sensíveis à presença de *outliers*. Por exemplo, cálculos simples da média e do desvio padrão podem ser distorcidos por um único ponto de dados totalmente impreciso.

A verificação de *outliers* deve ser uma parte rotineira de qualquer análise de dados. Os valores discrepantes em potencial devem ser examinados para verificar se estão possivelmente errados. Se o ponto de dados estiver com erro, ele deve ser corrigido se possível e excluído se não for possível. Se não houver razão para acreditar que o ponto periférico esteja errado, ele não deve ser excluído sem uma consideração cuidadosa. No entanto, o uso de técnicas mais robustas pode ser justificado. Técnicas robustas geralmente reduzem o efeito de pontos remotos sem excluí-los.

Técnicas Relacionadas: Várias técnicas gráficas podem e devem ser usadas para ajudar a detectar *outliers*. Um gráfico de probabilidade normal simples, gráfico de sequência de corrida, gráfico de caixa ou um histograma deve mostrar quaisquer pontos obviamente periféricos. Além de mostrar possíveis *outliers*, vários desses gráficos também ajudam a avaliar se os dados seguem uma distribuição aproximadamente normal.

Para realizar o teste de Grubbs no [R Core Team \(2020\)](#), usamos a função `grubbs.test()` do pacote `outliers`.

O p-valor é 0,056. No nível de significância de 5%, não rejeitamos a hipótese de que o valor mais alto 44 não é um *outlier*.

Por padrão, o teste é executado no valor mais alto (conforme mostrado na saída [R Core Team \(2020\)](#) hipótese alternativa: o valor mais alto 44 é um *outlier*). Podemos fazer o teste para o valor mais baixo, basta adicionar o argumento `opposite = TRUE` na função `grubbs.test()`.

A saída [R Core Team \(2020\)](#) indica que o teste agora é executado no valor mais baixo (consulte a hipótese alternativa: o valor mais baixo 12 é um *outlier*).

O p-valor é 1. No nível de significância de 5%, não rejeitamos a hipótese de que o menor valor 12 não é um *outlier*.

```

1 library(outliers)
2 test <- grubbs.test(dat$hwy)
3 test
4
5 Grubbs test for one outlier
6
7 data: dat$hwy
8 G = 3.45274, U = 0.94862, p-value = 0.05555
9 alternative hypothesis: highest value 44 is an outlier
10
11 test <- grubbs.test(dat$hwy, opposite = TRUE)
12 test
13
14 Grubbs test for one outlier
15
16 data: dat$hwy
17 G = 1.92122, U = 0.98409, p-value = 1
18 alternative hypothesis: lowest value 12 is an outlier
19
20 dat[34, "hwy"] <- 212
21
22 test <- grubbs.test(dat$hwy)
23 test
24
25 Grubbs test for one outlier
26
27 data: dat$hwy
28 G = 13.72240, U = 0.18836, p-value < 2.2e-16
29 alternative hypothesis: highest value 212 is an outlier

```

Código 6.7: Código fonte em R: Teste de Grubbs.

Para fins de ilustração, iremos agora substituir uma observação por um valor mais extremo e realizar o teste de Grubbs neste novo conjunto de dados. Vamos substituir a 34ª linha por um valor de 212.

O p-valor é $< 0,001$. No nível de significância de 5%, concluímos que o valor mais alto 212 é um *outlier*.

Teste de Grubbs - `grubbs.test {outliers}`

Executa o teste de Grubbs para um *outlier*, dois *outliers* em uma cauda ou dois *outliers* em caudas opostas, em uma amostra pequena.

Comando:

```
grubbs.test(x, type = 10, opposite = FALSE, two.sided = FALSE)
```

Argumentos:

`x` um vetor numérico para valores de dados.

`opposite` uma lógica que indica se você deseja verificar não o valor com a maior diferença da

média, mas o oposto (o mais baixo, se o mais suspeito for o mais alto etc.)

`type` Valor inteiro indicando variante de teste. 10 é um teste para um *outlier* (o lado é detectado automaticamente e pode ser revertido pelo parâmetro oposto). 11 é um teste para dois *outliers* em caudas opostas, 20 é um teste para dois *outliers* em uma cauda.

`two.sided` Valor lógico que indica se há necessidade de tratar este teste como bilateral.

Detalhes:

A função pode realizar três testes fornecidos e discutidos por Grubbs (1950).

O primeiro teste (10) é usado para detectar se o conjunto de dados de amostra contém um *outlier*, estatisticamente diferente dos outros valores. O teste é baseado no cálculo da pontuação desse valor discrepante G (valor discrepante menos média e dividido por sd) e na comparação com valores críticos apropriados. O método alternativo é calcular a proporção de variâncias de dois conjuntos de dados - conjunto de dados completo e conjunto de dados sem *outlier*. O valor obtido, denominado U , é vinculado a G por uma fórmula simples.

O segundo teste (11) é usado para verificar se os valores mais baixo e mais alto são dois *outliers* em caudas opostas da amostra. Baseia-se no cálculo da proporção da faixa para o desvio padrão da amostra.

O terceiro teste (20) calcula a razão de variância da amostra completa e da amostra sem duas observações extremas. É usado para detectar se o conjunto de dados contém dois *outliers* na mesma cauda.

Valores:

`statistic` o valor da estatística do teste. Para o tipo 10, é a diferença entre o *outlier* e a média dividida pelo desvio padrão, e para o tipo 20 é o intervalo da amostra dividido pelo desvio padrão. O valor adicional U é a proporção das variâncias da amostra com e sem nosso *outlier* suspeito. De acordo com Grubbs (1950), esses valores para o tipo 10 são limitados por fórmula simples e apenas um deles pode ser usado, mas a função fornece os dois. Para o tipo 20, o G é o mesmo que U .

`p.value` p-valor do teste

`alternative` uma sequência de caracteres que descreve a hipótese alternativa.

`method` uma sequência de caracteres que indica o tipo de teste realizado.

Maiores detalhes: <http://bit.ly/3pjFZoe>.

6.3.4 Teste de Dixon

[<http://bit.ly/34IchRZ>] Semelhante ao teste de Grubbs, o teste de Dixon é usado para testar se um único valor baixo ou alto é um *outlier*. Portanto, se houver suspeita de mais de um *outliers*, o teste deve ser executado nesses *outliers* suspeitos individualmente.

Observe que o teste de Dixon é mais útil para amostras de tamanho pequeno (geralmente $n \leq 25$).

O teste Q de Dixon, muitas vezes referido simplesmente como o Teste Q , é um teste estatístico usado para detectar outliers em um conjunto de dados.

A estatística de teste para o teste Q é a seguinte:

$$Q = \frac{|X_a - X_b|}{R},$$

onde X_a é o valor discrepante suspeito, X_b é o ponto de dados mais próximo de X_a e R é o intervalo do conjunto de dados. Na maioria dos casos, X_a é o valor máximo no conjunto de dados, mas também pode ser o valor mínimo.

É importante notar que o teste Q é normalmente realizado em pequenos conjuntos de dados e o teste assume que os dados são normalmente distribuídos. Também é importante observar que o teste Q deve ser realizado apenas uma vez para um determinado conjunto de dados.

As hipóteses do teste Q de Dixon são

$$H_0 : \text{O máximo não é um outlier} \quad \text{versus} \quad H_1 : \text{O máximo é um outlier}$$

Para realizar o teste de Dixon no [R Core Team \(2020\)](#), usamos a função `dixon.test()` do pacote `outliers`.

Teste de Dixon - `dixon.test {outliers}`

Executa várias variantes do teste de Dixon para detectar valores discrepantes na amostra de dados.

Comando:

```
dixon.test(x, type = 0, opposite = FALSE, two.sided = TRUE)
```

Argumentos:

`x` um vetor numérico para valores de dados.

`opposite` uma lógica que indica se você deseja verificar não o valor com a maior diferença da média, mas o oposto (o mais baixo, se o mais suspeito for o mais alto etc.).

`type` um inteiro especificando a variante do teste a ser executado. Os valores possíveis são compatíveis com os dados por [Dixon \(1950\)](#) 10, 11, 12, 20, 21. Se este valor for definido como zero, uma variante do teste é escolhida de acordo com o tamanho da amostra (10 para 3-7, 11 para 8-10, 21 para 11-13, 22 para 14 e mais). O valor mais baixo ou mais alto é selecionado automaticamente e pode ser revertido usando o parâmetro oposto.

`two.sided` Valor lógico que indica se há necessidade de tratar este teste como bilateral.

Detalhes:

O p-valor é calculado por interpolação usando `qdixon` e `qtable`. De acordo com as conclusões de [Dixon \(1951\)](#), os valores críticos podem ser obtidos numericamente apenas para $n = 3$. Outros valores críticos são obtidos por simulações, retiradas do artigo original de Dixon, e referentes às correções fornecidas por [Rorabacher \(1991\)](#).

Valores:

`statistic` o valor da estatística do teste.

`p.value` p-valor do teste

`alternative` uma sequência de caracteres que descreve a hipótese alternativa.

`method` uma sequência de caracteres que indica o tipo de teste realizado.

Maiores detalhes: <http://bit.ly/37PnGBe> e <http://bit.ly/37Rqgqw>

O cálculo dos valores críticos da Estatística Q do teste de Dixon, pode ser encontrado em [McBane \(2006\)](https://bit.ly/3aMMhZA) [<https://bit.ly/3aMMhZA>].

No entanto, restringimos nosso conjunto de dados às 20 primeiras observações, pois o teste de Dixon só pode ser feito em um tamanho de amostra pequeno ([R Core Team \(2020\)](#) lançará um erro informando que aceita apenas conjunto de dados de 3 a 30 observações):

Os resultados mostram que o menor valor 15 é um *outlier* (p-valor = 0,007). Para testar o valor mais alto, basta adicionar o argumento `opposite = TRUE` à `dixon.test()`.

Os resultados mostram que o maior valor 31 não é um *outlier* (p-valor = 0,858).

É uma boa prática sempre verificar os resultados do teste estatístico para valores discrepantes em relação ao boxplot para garantir que testamos todos os valores discrepantes em potencial.

No boxplot, vemos que também poderíamos aplicar o teste de Dixon no valor 20 além do valor 15 feito anteriormente. Isso pode ser feito encontrando o número da linha do valor mínimo, excluindo este número da linha do conjunto de dados e, finalmente, aplicar o teste de Dixon neste novo conjunto de dados. Os resultados mostram que o segundo menor valor 20 não é um *outlier* (p-valor = 0,13).

```

1 subdat <- dat[1:20, ]
2 test <- dixon.test(subdat$hwy)
3 test
4
5 Dixon test for outliers
6
7 data: subdat$hwy
8 Q = 0.57143, p-value = 0.006508
9 alternative hypothesis: lowest value 15 is an outlier
10
11 test <- dixon.test(subdat$hwy, opposite = TRUE)
12 test
13
14 Dixon test for outliers
15
16 data: subdat$hwy
17 Q = 0.25, p-value = 0.8582
18 alternative hypothesis: highest value 31 is an outlier
19
20 out <- boxplot.stats(subdat$hwy)$out
21 boxplot(subdat$hwy,
22         ylab = "hwy"
23 )
24 mtext(paste("Outliers: ", paste(out, collapse = ", ")))
25
26 remove_ind <- which.min(subdat$hwy)
27 subsubdat <- subdat[-remove_ind, ]
28 # Dixon test on dataset without the minimum
29 test <- dixon.test(subsubdat$hwy)
30 test
31
32 Dixon test for outliers
33
34 data: subsubdat$hwy
35 Q = 0.44444, p-value = 0.1297
36 alternative hypothesis: lowest value 20 is an outlier

```

Código 6.8: Código fonte em R: Teste de Dixon.

6.3.5 Teste de Rosner

O teste de Rosner para outliers tem as vantagens de:

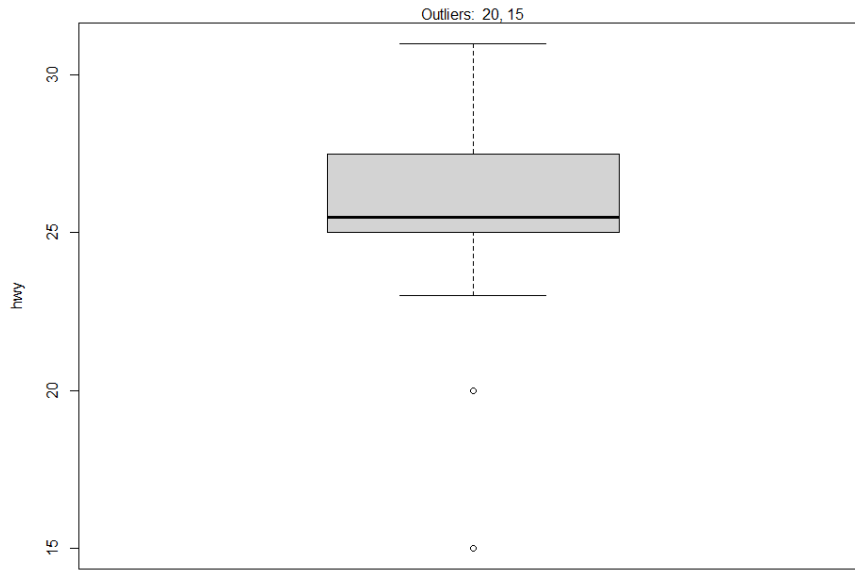


Figura 6.4: Boxplot com os possíveis *outliers* dos primeiros 20 valores da variável *hwy*.

- (i) É usado para detectar vários *outliers* de uma vez (ao contrário do teste de Grubbs e Dixon, que deve ser realizado iterativamente para rastrear múltiplos *outliers*), e ele é projetado para evitar o problema de mascaramento, onde um valor discrepante próximo a outro valor discrepante pode passar despercebido.
- (ii) Ao contrário do teste de Dixon, observe que o teste de Rosner é mais apropriado quando o tamanho da amostra é grande ($n \geq 20$).

O teste ESD generalizado (desvio estudentizado extremo) (Rosner, 1983) é usado para detectar um ou mais *outliers* em um conjunto de dados univariado que segue uma distribuição aproximadamente normal. A principal limitação do teste de Grubbs e do teste de Tietjen-Moore é que o número suspeito de *outliers*, k , deve ser especificado exatamente. Se k não for especificado corretamente, isso pode distorcer as conclusões desses testes. Por outro lado, o teste ESD generalizado (Rosner, 1983) requer apenas que um limite superior para o número suspeito de *outliers* seja especificado.

Definição do Teste: Dado o limite superior, k , o teste ESD generalizado realiza essencialmente k testes separados: um teste para um *outlier*, um teste para dois *outliers* e assim por diante até k *outliers*.

As hipóteses do teste teste ESD generalizado são

H_0 : Não existem *outliers* nos dados versus H_1 : Existem até r *outlier* nos dados

Estatística do teste:

$$R_i = \frac{\max_i |X_i - \bar{X}|}{s}$$

com \bar{X} e s denotando a média da amostra e o desvio padrão da amostra, respectivamente.

Remova a observação que maximiza $|X_i - \bar{X}|$ e então recalcule a estatística acima com $n - 1$ observações. Repita este processo até que r observações tenham sido removidas. Isso resulta nas k estatísticas de teste R_1, R_2, \dots, R_k .

O teste com nível de significância α e correspondendo às r estatísticas de teste, calcule os seguintes k valores críticos

$$\lambda_i = \frac{(n-i)t_{p,n-i-1}}{\sqrt{(n-i-1+t_{p,n-i-1}^2)(n-i+1)}} \quad i = 1, 2, \dots, k$$

onde, $t_{p,\nu}$ é o ponto percentual de $100p$ da distribuição t com ν graus de liberdade e

$$p = 1 - \frac{\alpha}{2(n-i+1)}$$

O número de *outliers* é determinado encontrando o maior i tal que $R_i > \lambda_i$.

Estudos de simulação por Rosner (1983) indicam que esta aproximação de valor crítico é muito precisa para $n \geq 25$ e razoavelmente precisa para $n \geq 15$.

Observe que, embora o ESD generalizado seja essencialmente um teste de Grubbs aplicado sequencialmente, existem algumas distinções importantes:

- (i) O teste ESD generalizado faz ajustes apropriados para os valores críticos com base no número de outliers sendo testados para o que a aplicação sequencial do teste de Grubbs não faz.
- (ii) Se houver mascaramento significativo, a aplicação do teste Grubbs sequencialmente pode parar muito cedo.

Para realizar o teste de Rosner, usamos a função `rosnerTest()` do pacote `EnvStats`. Esta função requer pelo menos 2 argumentos: os dados e o número de *outliers* suspeitos k (com $k = 3$ como o número padrão de *outliers* suspeitos).

Teste de Rosner - `rosnerTest` {*EnvStats*}

Execute o teste de desvio estudantizado extremo generalizado de Rosner para até k *outliers* potenciais em um conjunto de dados, assumindo que os dados sem *outliers* vêm de uma distribuição normal (gaussiana).

Comando:

```
rosnerTest(x, k = 3, alpha = 0.05, warn = TRUE)
```

Argumentos:

`x` vetor numérico de observações. Valores ausentes (NA), indefinidos (NaN) e infinitos (Inf, -Inf) são permitidos, mas serão removidos. Deve haver pelo menos 10 observações finitas em `x`.

`k` inteiro positivo indicando o número de *outliers* suspeitos. O argumento `k` deve estar entre 1 e $n - 2$, onde n denota o número de valores finitos não ausentes no argumento `x`. O valor padrão é `k = 3`.

`alpha` escalar numérico entre 0 e 1 indicando o Erro Tipo I associado ao teste de hipótese. O valor padrão é `alpha = 0,05`.

`warn` escalar lógico que indica se deve ser emitido um aviso (`warn = TRUE`; o padrão) quando o número de valores finitos não ausentes em `x` e o valor de `k` são tais que o nível de Erro Tipo I assumido pode não ser mantido.

Maiores detalhes: <http://bit.ly/3h1G1b6>.

O teste ESD generalizado pode ser usado para responder à seguinte pergunta: Quantos *outliers* o conjunto de dados contém?

Muitas técnicas estatísticas são sensíveis à presença de *outliers*. Por exemplo, cálculos simples da média e do desvio padrão podem ser distorcidos por um único ponto de dados totalmente impreciso.

A verificação de *outliers* deve ser uma parte rotineira de qualquer análise de dados. Os valores discrepantes em potencial devem ser examinados para verificar se estão possivelmente errados. Se o ponto de dados estiver com erro, ele deve ser corrigido se possível e excluído se não for possível. Se não houver razão para acreditar que o ponto periférico esteja errado, ele não deve ser excluído sem uma consideração cuidadosa. No entanto, o uso de técnicas mais robustas pode ser justificado. Técnicas robustas geralmente reduzem o efeito de pontos remotos sem excluí-los.

Várias técnicas gráficas podem e devem ser usadas para ajudar a detectar *outliers*. Um gráfico de probabilidade normal simples, gráfico de sequência de corrida, gráfico de caixa ou um histograma deve mostrar quaisquer pontos obviamente periféricos. Além de mostrar possíveis *outliers*, vários desses gráficos também ajudam a avaliar se os dados seguem uma distribuição aproximadamente normal.

Para este exemplo, definimos o número de *outliers* suspeitos igual a 3, conforme sugerido pelo número de *outliers* potenciais descritos no boxplot.

```
1 library (EnvStats)
2 test <- rosnerTest (dat$hwy, k = 3)
3 test
```

Código 6.9: Código fonte em R: Teste de Dixon.

Com base no teste de Rosner, vemos que há apenas um outlier (veja a coluna Outlier), e que é a observação 34 (veja Obs.Num) com um valor de 212 (veja Value).

Iglewicz e Hoaglin (1993) fornecem uma extensa discussão sobre os testes *outliers* dados acima (bem como alguns não fornecidos acima) e também fornecem um bom tutorial sobre o assunto de *outliers*. Rothamsted et al. (1996) fornecem um tratamento do assunto em forma de livro. Além de discutir testes adicionais para dados que seguem uma distribuição aproximadamente normal, essas fontes também discutem o caso em que os dados não são normalmente distribuídos.

Para o caso de *outliers* multivariados podem ser usadas as técnicas apresentadas em <https://rpubs.com/melinatarituba/356726>.

6.3.6 Observações Adicionais

Podemos utilizar outros métodos para encontrar *outliers*.

- (1) No pacote `outliers` do R Core Team (2020).
 - (2) por meio da função `lofactor()` do pacote `DMwR`: Local *Outlier* Factor (LOF) é um algoritmo usado para identificar *outliers* comparando a densidade local de um ponto com a de seus vizinhos,
 - (3) a função `outlierTest()` do pacote `car` fornece a observação mais extrema com base no modelo fornecido e permite testar se é um *outlier*, e
 - (4) no pacote `OutlierDetection`.
 - (5) O teste de Tietjen-Moore (Tietjen e Moore, 1972) é usado para detectar múltiplos *outliers* em um conjunto de dados univariado que segue uma distribuição aproximadamente normal. O teste Tietjen-Moore é uma generalização do teste de Grubbs para o caso de múltiplos *outliers*. Se estiver testando um único *outlier*, o teste de Tietjen-Moore é equivalente ao teste de Grubbs.
- É importante observar que o teste de Tietjen-Moore requer que o número suspeito de *outliers* seja especificado com exatidão. Se isso não for conhecido, é recomendado que o teste ESD generalizado seja usado em vez disso (este teste requer apenas um limite superior no número de discrepâncias suspeitas). Maiores detalhes ver <https://bit.ly/3ru7bm2>.

Observe também que algumas transformações podem eliminar “naturalmente” *outliers*. O logaritmo natural ou raiz quadrada de um valor reduz a variação causada por valores extremos, portanto, em alguns casos, a aplicação dessas transformações eliminará os *outliers*, ou outras transformações aplicadas a séries temporais para que a sua distribuição passe a ser a normal. Por exemplo, a transformação de Box Cox apresentada na Seção 6.4.

Outliers não são necessariamente valores incorretos

Se uma observação é um *outlier* depende do modelo estatístico assumido subjacente. McBean e Rovers (1992) declaram: “Pode ser possível ignorar o *outlier* se uma justificativa física estiver disponível, mas, na falta disso, o valor deve ser incluído ... Observe que o uso de estatísticas não interpreta os fatos, simplesmente torna os fatos mais fáceis de ver. Portanto, é responsabilidade do analista identificar se o valor alto ... é ou não verdadeiramente representativo do produto químico que está sendo monitorado ou, em vez disso, é um *outlier* por razões como resultado de amostragem ou erro de laboratório.”

EPA (2006), página 51, afirma: “Se o raciocínio científico não explica o *outlier*, ele não deve ser descartado do conjunto de dados.”

Por fim, um editorial do Editor-chefe da revista Science trata desse tema em McNutt (2014).

6.4 Transformação de Box-Cox

Dado que os testes para detecção de *outliers* apresentados na Seção 6.3 exigem que as séries temporais apresentem distribuição aproximadamente normal. Como comentado anteriormente, aplicando o logaritmo natural ou raiz quadrada reduz a variação causada por valores extremos, portanto, em alguns casos, a aplicação dessas transformações eliminará os *outliers*. Outra razão para efetuar transformações é obter uma distribuição para a série temporal mais simétrica e próxima da normal.

Uma forma alternativa, para dados não normais, as vezes é aplicar uma transformação tal como função logarítmica pode produzir dados normais. Neste caso, esta transformação é apropriada se o desvio padrão da série temporal (ou outra medida de dispersão) for proporcional à média.

No caso de uma variável positiva Y_t , Box e Cox (1964) apresentam outra transformação geral

$$Y_t^{(\lambda)} = \begin{cases} \frac{Y_t^\lambda - c}{\lambda}, & \text{para } \lambda \neq 0 \\ \log(Y_t), & \text{para } \lambda = 0 \end{cases}$$

denominada de *transformação de Box-Cox*. Neste caso, λ necessita ser estimado. Box et al. (2015) (página 384) sugerem uma forma de encontrar os valores de λ : aplicar a transformação de Box-Cox na série temporal original para vários valores de λ , e escolher o valor de λ que resulta na menor soma dos quadrados dos resíduos

$$S(\eta) = \sum_{t=2}^n e_t^2(\eta).$$

Para estimar o parâmetro λ pode-se utilizar também a rotina `BoxCox.lambda()` do pacote `forecast` do R Core Team (2020).

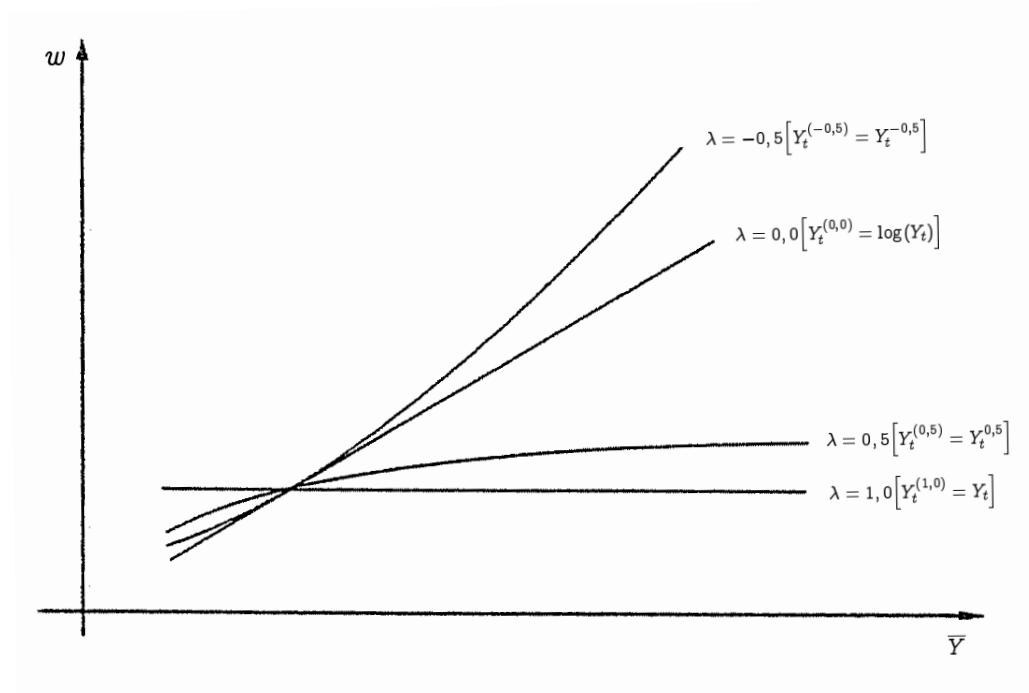


Figura 6.5: Gráfico: Amplitude \times Média, para ilustrar alguns valores de λ .

Para se ter uma ideia de que tipo de transformação é mais adequada, podemos utilizar um gráfico que traz no eixo das abcissas médias de subconjuntos de observações da série temporal original e no eixo das ordenadas a amplitude de cada um desses subconjuntos. Por exemplo, se $Y = \{Y_1, \dots, Y_k\}$ for um subconjunto com k observações, calculamos

$$\bar{Y} = \frac{1}{k} \sum_{j=1}^k X_{t_j} \quad \text{e} \quad w = \max(Y) - \min(Y).$$

que são medidas de posição e variabilidade, respectivamente. O para (\bar{Y}, w) será um ponto no gráfico.

Se w independe de \bar{Y} , obteremos pontos espalhados ao redor de uma reta paralela ao eixo das abcissas, e neste caso não haverá necessidade de transformação. Se w for diretamente proporcional a \bar{Y} , a transformação logarítmica é apropriada. A Figura 6.5, foi extraída de Jenkins (1979), nos fornece uma ideia dos tipos de gráficos que podem ocorrer e os respectivos valores de λ .

Hinkley (1977) sugere que se calcule a média, mediana e um estimador de escala (desvio padrão, ou outro estimador robusto) e, então, para a transformação $Y_t^{(\lambda)}$ e λ tomamos os valores $\{\dots, -3, -2, -1, -\frac{1}{2}, -\frac{1}{4}, 0, \frac{1}{4}, \frac{1}{2}, 1, 2, 3, \dots\}$, escolha-se o valor de λ que minimize

$$d_\lambda = \frac{|\text{média} - \text{mediana}|}{\text{medida de escala}},$$

que pode ser vista como uma medida de assimetria, em uma distribuição simétrica $d_\lambda = 0$.

Para mais alguns detalhes sobre a transformação de Box-Cox ver <https://bit.ly/2RZQ77g>.

6.5 Aplicação Transformação Box-Cox

Nesta seção apresentamos uma aplicação da transformação de Box-Cox cuja série temporal é não gaussiana. Serão utilizadas as rotinas `BoxCox.lambda()` do pacote `forecast`.

Maiores detalhes sobre a estimação dos parâmetros da transformação Box-Cox utilizando a rotina `BoxCox.lambda()` do pacote `forecast` ver [Box e Cox \(1964\)](#) e [Guerrero \(1993\)](#).

Capítulo 7

Modelo Vetorial Autorregressivo

Capítulo baseado nos materiais: <http://bit.ly/3536qa1>, <https://bit.ly/385oSRh>, <https://bit.ly/3pMe0NY>, <https://bit.ly/3b1WwsR>, Ferreira (2018), Hamilton (1994)

7.1 Introdução

Os modelos econômicos em geral são expressos por meio de diversas variáveis. Portanto, o uso de modelos univariados, como visto até o momento, é limitado para expressar modelos econômicos. Um exemplo básico são os modelos macroeconômicos de curto prazo onde PIB, consumo, investimento e gastos governamentais são determinados simultaneamente.

A partir do artigo de Sims (1980), o uso do *Modelo Vetorial Autorregressivo* (VAR) se difundiu rapidamente e estão entre os instrumentos mais usados para investigar empiricamente a macroeconomia.

Os modelos VAR são sistemas de equações simultâneas que capturam a existência de relações de interdependência entre variáveis, e que permitem avaliar o impacto de choques estocásticos sobre determinada variável do sistema. O Banco Central, assim como a grande maioria de seus pares internacionais, utiliza modelos VAR como instrumento de análise e, principalmente, de previsão de inflação desde a implementação do regime de metas para a inflação, em junho de 1999.

7.2 Regressão Espúria

Na seção 2.3.5, definimos um passeio aleatório e suas propriedades as quais apresentamos a seguir.

Passeio Aleatório

Definição 7.1. Seja $\{\varepsilon_t\}_{t \in \mathbb{N}}$ um $WN(0, \sigma_\varepsilon^2)$. Defina

$$Z_t = Z_{t-1} + \varepsilon_t,$$

que pode ser reescrito de uma maneira bem simples. Defina inicialmente

$$Z_1 = \varepsilon_1, \quad Z_2 = \varepsilon_1 + \varepsilon_2 \leftrightarrow Z_2 = Z_1 + \varepsilon_2$$

e sucessivamente

$$Z_{k-1} = \varepsilon_1 + \cdots + \varepsilon_{k-1}, \quad Z_k = \varepsilon_1 + \cdots + \varepsilon_{k-1} + \varepsilon_k = Z_{k-1} + \varepsilon_k.$$

Com esta representação, o cálculo da média e da variância de Z_t se tornam simples:

$$\mu_t = \mathbb{E}(Z_t) = \mathbb{E}(\varepsilon_1 + \varepsilon_2 + \cdots + \varepsilon_t) = \mathbb{E}(\varepsilon_1) + \mathbb{E}(\varepsilon_2) + \cdots + \mathbb{E}(\varepsilon_t) = 0 + 0 + \cdots + 0 = 0,$$

e

$$\text{Var}(Z_t) = \text{Var}(\varepsilon_1 + \varepsilon_2 + \cdots + \varepsilon_t) = \text{Var}(\varepsilon_1) + \cdots + \text{Var}(\varepsilon_t) = \sigma_\varepsilon^2 + \sigma_\varepsilon^2 + \cdots + \sigma_\varepsilon^2 = t\sigma_\varepsilon^2.$$

Assim concluímos que a variância de um passeio aleatório cresce linearmente com o tempo, sendo portanto um processo não-estacionário. Observe ainda que se $1 \leq t \leq s$, a função de autocovariância de um passeio aleatório é dada por

$$\begin{aligned} \gamma_Z(t, s) &= \text{Cov}(Z_t, Z_s) \\ &= \text{Cov}(\varepsilon_1 + \varepsilon_2 + \cdots + \varepsilon_t, \varepsilon_1 + \varepsilon_2 + \cdots + \varepsilon_s) \\ &= \sum_{i=1}^t \sum_{j=1}^s \text{Cov}(\varepsilon_i, \varepsilon_j) \\ &= \text{Cov}(\varepsilon_1, \varepsilon_1) + \text{Cov}(\varepsilon_2, \varepsilon_2) + \cdots + \text{Cov}(\varepsilon_t, \varepsilon_t) \\ &= \sigma_\varepsilon^2 + \sigma_\varepsilon^2 + \cdots + \sigma_\varepsilon^2 = t\sigma_\varepsilon^2 \end{aligned}$$

onde $\text{Cov}(\varepsilon_i, \varepsilon_j) = 0$ para $i \neq j$. O mesmo argumento mostra que se $1 \leq s \leq t$, teremos $\gamma_Z(t, s) = s\sigma_\varepsilon^2$, de forma que podemos escrever compactamente $\gamma_Z(s, t) = \min(s, t)\sigma_\varepsilon^2$. A função de autocorrelação de um passeio aleatório é facilmente obtida

$$\begin{aligned} \rho_Z(t, s) &= \frac{\gamma_Z(s, t)}{\sqrt{\text{Var}(X_t)}\sqrt{\text{Var}(X_s)}} = \frac{\min(s, t)\sigma_\varepsilon^2}{\sqrt{t\sigma_\varepsilon^2}\sqrt{s\sigma_\varepsilon^2}} = \frac{\min(s, t)}{\sqrt{t}\sqrt{s}} = \begin{cases} \frac{s}{\sqrt{t}\sqrt{s}} = \sqrt{\frac{s}{t}}, & \text{se } 1 \leq s \leq t; \\ \frac{t}{\sqrt{t}\sqrt{s}} = \sqrt{\frac{t}{s}}, & \text{se } 1 \leq t \leq s \end{cases} \\ &= \sqrt{\frac{\min(s, t)}{\max(s, t)}}. \end{aligned}$$

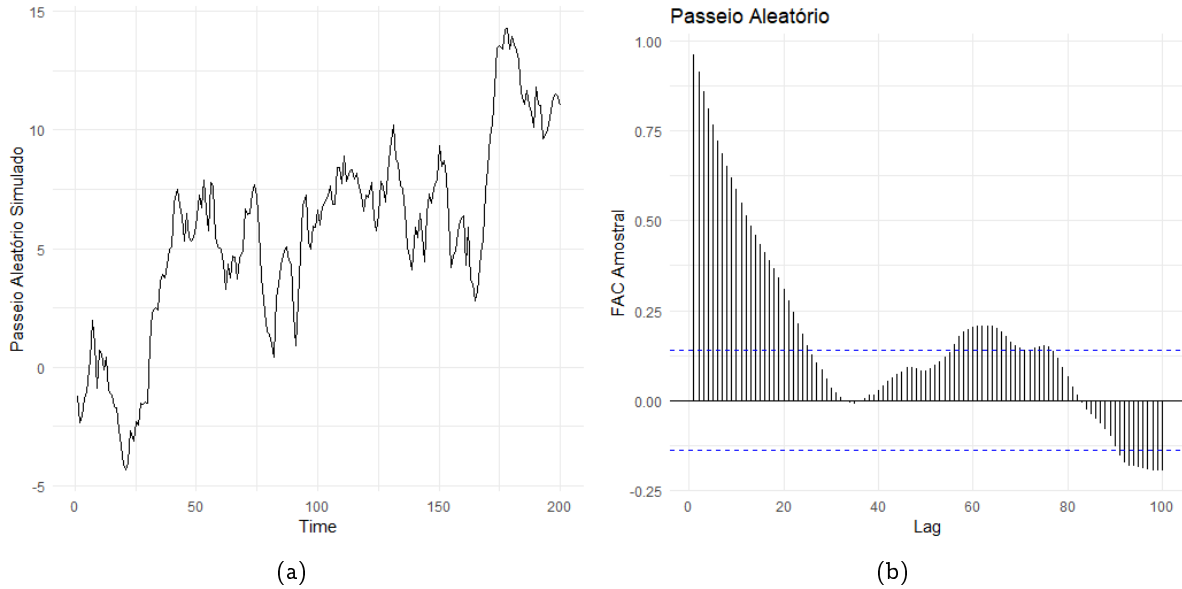


Figura 7.1: (a) Passeio Aleatório simulado; (b) FAC amostral.

Em resumo, a FACH e a FAC de um passeio aleatório são dadas por

FACHV do passeio aleatório

$$\gamma_Z(t, s) = \min(s, t)\sigma_\varepsilon^2,$$

FAC do passeio aleatório

$$\rho_Z(t, s) = \sqrt{\frac{\min(s, t)}{\max(s, t)}}$$

Existe a possibilidade de encontrarmos uma relação espúria entre séries temporais que contenham tendência. Quando as séries são dependentes de forma determinística dessa tendência, ou seja, a série é de tendência estacionária (determinística), o problema é solucionado adicionado uma tendência temporal no modelo de regressão. No entanto, quando tratamos de processos integrados de ordem um $I(1)$ existe outro agravante. Quando tentamos fazer a regressão de uma série não estacionária contra outra série que tem a mesma característica, encontramos com frequência um resultado com significância estatística mesmo que tal resultado não tenha nenhum significado estatístico ou econômico.

Para ilustrar melhor vamos supor dois passeios aleatórios $\{X_t\}_{t \in \mathbb{N}}$ e $\{Y_t\}_{t \in \mathbb{N}}$ onde

$$X_t = X_{t-1} + \varepsilon_{1t},$$

$$Y_t = Y_{t-1} + \varepsilon_{2t},$$

com $\{\varepsilon_{1t}\}_{t \in \mathbb{N}}$ e $\{\varepsilon_{2t}\}_{t \in \mathbb{N}}$ dois processos ruídos brancos definidos respectivamente por $RB(0, \sigma_{\varepsilon_1}^2)$ e $RB(0, \sigma_{\varepsilon_2}^2)$. Temos que $\{X_t\}_{t \in \mathbb{N}}$ e $\{Y_t\}_{t \in \mathbb{N}}$ são independentes $\{X_t\}_{t \in \mathbb{N}}$ e $\{Y_t\}_{t \in \mathbb{N}}$ e ambos valores iniciais das séries são zero. Se for feita a regressão

$$\hat{Y}_t = \beta_0 + \beta_1 X_t + \varepsilon_{1t}$$

e encontrarmos a estatística de teste t para β_1 é esperado que o teste $H_0 : \beta_1 = 0$ versus $H_1 : \beta_1 \neq 0$, cujo nível de significância é 5%(por exemplo), nos forneça uma estatística t para β_1 cujos valores são não significativos em 95% das vezes.

Granger e Newbold (1974), através de simulações, demonstraram que se $\{X_t\}_{t \in \mathbb{N}}$ e $\{Y_t\}_{t \in \mathbb{N}}$ são $I(1)$ e passeios aleatórios, a regressão de uma variável contra a outra gera resultados inconsistentes e, além disso, os testes de significâncias usuais mostram que as variáveis são relacionadas. Regressões espúrias são caracterizadas por alto R^2 e um baixo valores da estatística de Durbin-Watson (ver Granger e Newbold, 1974). Segundo da Silveira Bueno (2011), em resumo, temos:

7.3 Modelo Vetorial Autorregressivo de Ordem 1 - VAR(1)

Uma série temporal multivariada $\mathbf{r}_t = (r_{1t}, r_{2t}, \dots, r_{kt})'$ composta por k componentes no tempo t é um processo VAR de ordem 1, ou $VAR(1)$, se segue o modelo:

$$\mathbf{r}_t = \phi_0 + \Phi \mathbf{r}_{t-1} + \mathbf{a}_t$$

onde ϕ_0 é um vetor de dimensão k , Φ é uma matriz $k \times k$ e \mathbf{a}_t é um ruído branco formado por uma sequência de vetores aleatórios independentes e identicamente distribuídos com média 0 e matriz de covariância Σ_a . Na literatura, é frequentemente assumido que \mathbf{a}_t segue uma distribuição Normal multivariada e que a matriz Σ_a é positiva definida.

Considere o caso bivariado (isto é, $k = 2$, $\mathbf{r}_t = (r_{1t}, r_{2t})'$ e $\mathbf{a}_t = (a_{1t}, a_{2t})'$). O modelo VAR(1) pode ser escrito como:

$$\begin{bmatrix} r_{1t} \\ r_{2t} \end{bmatrix} = \begin{bmatrix} \phi_{10} \\ \phi_{20} \end{bmatrix} + \begin{bmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{bmatrix} \begin{bmatrix} r_{1,t-1} \\ r_{2,t-1} \end{bmatrix} + \begin{bmatrix} a_{1t} \\ a_{2t} \end{bmatrix}, \Sigma_a = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{bmatrix}$$

onde ϕ_{ij} é o (i, j) -ésimo elemento de Φ , ϕ_{i0} é o i -ésimo elemento de ϕ_0 , $var(a_{1t}) = \sigma_1^2$, $var(a_{2t}) = \sigma_2^2$ e $cov(a_{1t}, a_{2t}) = \sigma_{12} = \sigma_{21}$.

Equivalentemente, podemos escrever o modelo como:

$$\begin{aligned} r_{1t} &= \phi_{10} + \phi_{11} r_{1,t-1} + \phi_{12} r_{2,t-1} + a_{1t} \\ r_{2t} &= \phi_{20} + \phi_{21} r_{1,t-1} + \phi_{22} r_{2,t-1} + a_{2t} \end{aligned}$$

Tal forma de especificar o modelo (tanto a matricial quanto a última) é chamada de *forma reduzida* em função da relação contemporânea entre r_{1t} e r_{2t} não ser mostrada explicitamente. Observe que trata-se de um modelo composto por duas equações (em função de termos duas variáveis) e cada equação é função de seu valor defasado e da primeira defasagem da outra variável.

Temos duas maneiras de avaliar as equações acima:

(1) Separadamente

- (i) ϕ_{12} denota a dependência linear de r_{1t} em relação a $r_{2,t-1}$ na presença de $r_{1,t-1}$. Portanto, ϕ_{12} é o efeito condicional de $r_{2,t-1}$ sobre r_{1t} dado $r_{1,t-1}$.
- (ii) Se $\phi_{12} = 0$, então r_{1t} não depende de $r_{2,t-1}$ e o modelo mostra que r_{1t} depende apenas de seus valores passados.
- (iii) Similarmente, se $\phi_{21} = 0$, então a segunda equação mostra que r_{2t} não depende de $r_{1,t-1}$ quando $r_{2,t-1}$ é dado

(2) Em conjunto

- (i) Se $\phi_{12} = 0$ e $\phi_{21} \neq 0$ o que mostra que r_{1t} não depende de r_{2t} , mas r_{2t} depende de r_{1t} . Consequentemente, temos uma relação unidirecional de r_{1t} para r_{2t} . Na literatura de econometria, o modelo implica na existência de causalidade de Granger entre as duas séries com r_{1t} causando r_{2t} , mas não sendo causada por r_{2t} .
- (ii) Se $\phi_{12} = \phi_{21} = 0$, então r_{1t} e r_{2t} não são dinamicamente correlacionados. Neste caso, cada série segue um modelo AR(1) univariado. Dizemos que as duas séries são desacopladas.
- (iii) Se $\phi_{12} \neq 0$ e $\phi_{21} \neq 0$, então existe uma relação entre as duas séries, ou seja, elas se impactam.

Além disso, um choque em r_{1t} por meio de a_{1t} tem efeito contemporâneo sobre r_{1t} . No período seguinte, este efeito se torna $r_{1,t-1}$ e afeta r_{2t} que no período seguinte se torna $r_{2,t-1}$ e afeta r_{1t} e assim sucessivamente.

Em um modelo estacionário este efeito do choque desaparece após alguns períodos. Como definimos anteriormente, uma série temporal multivariada r_t é fracamente estacionária se tem média constante ($\mu = E(r_t)$) e matriz de covariância constante ($\Gamma_{t,l} = Cov(r_t, r_{t-l}) = \Gamma_l$), onde l é uma defasagem qualquer. Assim, precisamos confirmar se o modelo VAR(1) respeita as hipóteses de estacionariedade fraca, como segue:

7.3.1 Média Constante

Fazendo uso das hipóteses $E(r_t) = E(r_{t-1}) = \mu$ e que $E(a_t) = 0$, temos:

$$\begin{aligned} E(r_t) &= \phi_0 + \Phi E(r_{t-1}) \\ \mu &= \phi_0 + \Phi \mu \\ \mu &= (I - \Phi)^{-1} \phi_0 \end{aligned}$$

onde I é uma matriz identidade $k \times k$ e a matriz $I - \Phi$ deve ser não singular (determinante diferente de zero) para que a inversa exista e, consequentemente, o vetor de médias.

7.3.2 Matriz de Covariância Constante

Reescrevendo a média do processo como $\phi_0 = (I - \Phi)\mu$, podemos escrever o modelo VAR(1) como:

$$\begin{aligned}
r_t &= (I - \Phi)\mu + \Phi r_{t-1} + a_t \\
r_t &= \mu - \Phi\mu + \Phi r_{t-1} + a_t \\
(r_t - \mu) &= \Phi(r_{t-1} - \mu) + a_t
\end{aligned}$$

Fazendo $\tilde{r}_t = r_t - \mu$, temos:

$$\tilde{r}_t = \Phi\tilde{r}_{t-1} + a_t$$

Esta formulação é apenas uma forma diferente de escrever o modelo VAR(1) definido no início do texto. Substituindo \tilde{r}_{t-1} recursivamente, produz:

$$\begin{aligned}
\tilde{r}_t &= \Phi(\Phi\tilde{r}_{t-2} + a_{t-1}) + a_t \\
\tilde{r}_t &= \Phi^2\tilde{r}_{t-2} + \Phi a_{t-1} + a_t \\
\tilde{r}_t &= \Phi^2(\Phi\tilde{r}_{t-3} + a_{t-2}) + \Phi a_{t-1} + a_t \\
\tilde{r}_t &= \Phi^3\tilde{r}_{t-3} + \Phi^2 a_{t-2} + \Phi a_{t-1} + a_t \\
\tilde{r}_t &= a_t + \Phi a_{t-1} + \Phi^2 a_{t-2} + \Phi^3 a_{t-3} + \dots
\end{aligned}$$

Esta formulação mostra algumas características de um processo VAR(1). Elas são:

- (1) Uma vez que a_t é serialmente não correlacionado (em função da hipótese de ser um ruído branco), segue que $Cov(a_t, r_{t-l}) = 0$.
- (2) Multiplicando a expressão por a_t' , calculando o valor esperado e usando a hipótese de a_t ser serialmente não correlacionado, obtemos $Cov(r_t, a_t) = \sum_a$.
- (3) Para um VAR(1), \tilde{r}_t e consequentemente r_t dependerá dos erros passados com matriz de coeficientes igual a Φ^j . Para esta dependência ser significativa Φ^j deve convergir para zero na medida que $j \rightarrow \infty$. Isto significa que os k autovalores de Φ devem ser menores que 1, caso contrário Φ^j vai explodir quando $j \rightarrow \infty$.

Usado as hipóteses acima e $Cov(a_t) = \sum_a$ para qualquer t , temos que a covariância de r_t será:

$$Cov(r_t) = Cov(\tilde{r}_t) = \Gamma_0 = \sum_a + \Phi \sum_a \Phi' + \Phi^2 \sum_a (\Phi^2)' + \dots = \sum_{i=0}^{\infty} \Phi^i \sum_a (\Phi^i)'$$

onde $\Phi^0 = I$ é uma matriz identidade $k \times k$. Note que a covariância do processo não depende de qualquer t , sendo invariante do tempo.

7.3.3 Matrizes de Covariância e Correlação Cruzada Dependente apenas de 1

Sabemos que o modelo reescrito usando $\Phi_0 = (I - \Phi)\mu$ e $\tilde{r}_t = r_t - \mu$ pode ser definido como:

$$\tilde{r}_t = \Phi\tilde{r}_{t-1} + a_t$$

Multiplicando ambos os lados por $\tilde{r}_{t-l}' = (r_{t-l} - \mu)'$, usando a hipótese de que $E(a_t, \tilde{r}_{t-l}') = 0$ em função de a_t ser um ruído branco e calculando o valor esperado, teremos a matriz de covariância cruzada para qualquer defasagem $l > 0$:

$$\begin{aligned}
E(\tilde{r}_t, \tilde{r}_{t-l}') &= \Phi E(\tilde{r}_{t-1}, \tilde{r}_{t-l}') \\
E[(r_t - \mu)(r_{t-l} - \mu)'] &= \Phi E[(r_{t-1} - \mu)(r_{t-l} - \mu)']
\end{aligned}$$

que pode ser escrita usando Γ_l

$$\Gamma_l = \Phi \Gamma_{l-1}$$

onde Γ_l é a matriz de covariância cruzada de r_t para a defasagem l .

Sabemos que a divisão da matriz de covariância cruzada pela matriz de desvios-padrão proporcionará a correlação cruzada. Assim, teremos:

$$\rho_l = D^{-1} \Phi \Gamma_{l-1} D^{-1} = \Upsilon \rho_{l-1}$$

onde $\Upsilon = D^{-1} \Phi D^{-1}$. Consequentemente, a matriz de correlação cruzada de um modelo VAR(1) será:

$$\rho_l = \Upsilon^l \rho_0$$

para $l > 0$.

7.4 Modelo Vetorial Autorregressivo de Ordem p - VAR(p)

Uma série temporal multivariada $r_t = (r_{1t}, r_{2t}, \dots, r_{kt})'$ composta por k componentes no tempo t é um processo VAR de ordem p , ou VAR(p), se para $p > 0$ segue o modelo:

$$r_t = \phi_0 + \Phi_1 r_{t-1} + \dots + \Phi_p r_{t-p} + a_t$$

onde ϕ_0 é um vetor de dimensão k , Φ_j são matrizes $k \times k$ para $j = 1, \dots, p$ e a_t é um ruído branco formado por uma sequência de vetores aleatórios independentes e identicamente distribuídos com média 0 e matriz de covariância Σ_a . Na literatura, é frequentemente assumido que a_t segue uma distribuição Normal multivariada e que a matriz Σ_a é positiva definida.

Utilizando um operador de defasagem B , podemos reescrever o modelo como:

$$(I - \Phi_1 B - \dots - \Phi_p B^p) r_t = \phi_0 + a_t$$

onde I é uma matriz identidade $k \times k$. Esta representação também pode ser reescrita como:

$$\Phi(B) r_t = \phi_0 + a_t$$

onde $\Phi(B) = I - \Phi_1 B - \dots - \Phi_p B^p$ é uma matriz polinomial. Se r_t é fracamente estacionário, então temos:

$$\mu = (I - \Phi_1 - \dots - \Phi_p)^{-1} \phi_0$$

onde $(I - \Phi_1 - \dots - \Phi_p)^{-1}$ deve ser não singular (determinante diferente de zero) para que o vetor de médias exista. Fazendo $\tilde{r}_t = r_t - \mu$ o VAR(p) se torna:

$$\tilde{r}_t = \Phi_1 \tilde{r}_{t-1} + \dots + \Phi_p \tilde{r}_{t-p} + a_t$$

Usando esta equação e o mesmo processo aplicado nas propriedades de estacionariedade do VAR(1), obtemos:

1. $Cov(r_t, a_t) = \Sigma_a$ que é a matriz de covariância de a_t 2. $Cov(a_t, r_{t-l}) = 0$ para $l > 0$ 3. $\Gamma_l = \Phi_1 \Gamma_{l-1} + \dots + \Phi_p \Gamma_{l-p}$ para $l > 0$

7.5 Estimação

Um modelo VAR(p) pode ser estimado por mínimos quadrados ordinários (MQO), método da máxima verossimilhança (MV) ou pelo procedimento bayesiano de estimação. Aqui, vamos mostrar brevemente como fazer uso de MQO e MV para obter os parâmetros do modelo VAR(p).

7.5.1 Mínimos Quadráticos Ordinários

Suponha que temos a amostra $\{r_t | t = 1, \dots, T\}$ para estimar um modelo VAR(p). Os parâmetros de interesse são $\{\phi_0, \Phi_1, \dots, \Phi_p\}$ e \sum_a . Para a estimação via MQO, os dados disponíveis são $t = p + 1, \dots, T$ e a equação a ser estimada é:

$$r_t = \phi_0 + \Phi_1 r_{t-1} + \dots + \Phi_p r_{t-p} + a_t$$

onde a matriz de covariância de a_t é \sum_a . Considere o caso bivariado (isto é, $k = 2$, $r_t = (r_{1t}, r_{2t})'$ e $a_t = (a_{1t}, a_{2t})'$) e o modelo VAR(p) pode ser escrito como:

$$\begin{aligned} r_{1t} &= \phi_{10} + \phi_{11}^1 r_{1,t-1} + \phi_{12}^1 r_{2,t-1} + \dots + \phi_{11}^p r_{1,t-p} + \phi_{12}^p r_{2,t-p} + a_{1t} \\ r_{2t} &= \phi_{20} + \phi_{21}^1 r_{1,t-1} + \phi_{22}^1 r_{2,t-1} + \dots + \phi_{21}^p r_{1,t-p} + \phi_{22}^p r_{2,t-p} + a_{2t} \end{aligned}$$

Que em formato matricial pode ser expresso como abaixo:

$$\begin{bmatrix} r_1 \\ r_2 \end{bmatrix} = \begin{bmatrix} X_1 & 0 \\ 0 & X_2 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} A_1 \\ A_2 \end{bmatrix}$$

onde para cada série temporal (no nosso caso $k = 1, 2$) teremos:

$$X_1 = X_2 = \begin{bmatrix} 1 & r_{1,p} & r_{2,p} & \dots & r_{1,1} & r_{2,1} \\ 1 & r_{1,p+1} & r_{2,p+1} & \dots & r_{1,2} & r_{2,2} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & r_{1,T-1} & r_{2,T-1} & \dots & r_{1,T-1} & r_{2,T-1} \end{bmatrix}$$

e $r_k = [r_{k,p+1}, \dots, r_{k,T}]'$, $\beta_k = [\phi_{k0} \ \phi_{k1}^1 \ \phi_{k2}^1 \ \dots \ \phi_{k1}^p \ \phi_{k2}^p]'$ e $A_k = [a_{k,p+1}, \dots, a_{k,T}]'$.

Note que aplicar o método de mínimos quadrados ordinários (MQO) sobre cada uma das equações do modelo VAR(p) gerará para cada equação (usando a notação matricial anterior):

$$\hat{\beta}_k = (X_k' X_k)^{-1} X_k' r_k$$

tal estimador é consistente e assintoticamente eficiente. Além disso, é igual ao estimador obtido por meio do método da máxima verossimilhança condicional. Em função de estarmos com as mesmas variáveis explicativas em cada equação (o retorno de cada ativo como função de suas defasagens e das defasagens dos outros ativos), o estimador obtido ao aplicar o método de mínimos quadrados em cada uma das equações separadamente é equivalente ao estimador obtido pelo método multivariado (todas as equações ao mesmo tempo).

O estimador de mínimos quadrados para a matriz de covariância dos erros será:

$$\widehat{\sum}_a = \frac{1}{T - p - (kp + 1)} \sum_{t=p+1}^T \hat{a}_t \hat{a}_t'$$

onde o denominador é a amostra efetiva (disponível para estimação) menos o número de parâmetros na equação para cada série temporal que compõe r_t .

7.5.2 Verossimilhança Condicional

Para estimação do modelo VAR(p) via máxima verossimilhança assumimos que a_t segue uma distribuição Normal multivariada. Deixe $r_{h:q}$ denotar as observações de $t = h$ até $t = q$ (inclusive). A função de verossimilhança condicional dos dados pode ser escrita como:

$$\begin{aligned}
L \left(\mathbf{r}_{(p+1):T} | \mathbf{r}_{1:p}, \boldsymbol{\beta}, \boldsymbol{\Sigma}_a \right) &= \prod_{t=p+1}^T p \left(\mathbf{r}_t | \mathbf{r}_{1:(t-1)}, \boldsymbol{\beta}, \boldsymbol{\Sigma}_a \right) \\
&= \prod_{t=p+1}^T p \left(\mathbf{a}_t | \mathbf{r}_{1:(t-1)}, \boldsymbol{\beta}, \boldsymbol{\Sigma}_a \right) \\
&= \prod_{t=p+1}^T p \left(\mathbf{a}_t | \boldsymbol{\beta}, \boldsymbol{\Sigma}_a \right) \\
&= \frac{1}{(2\pi)^{\frac{1}{2}} |\boldsymbol{\Sigma}_a|^{\frac{1}{2}}} \exp \left[-\frac{1}{2} \mathbf{a}_t' \boldsymbol{\Sigma}_a^{-1} \mathbf{a}_t \right]
\end{aligned}$$

e o logaritmo da verossimilhança condicional se torna:

$$\ln L \left(\boldsymbol{\beta}, \boldsymbol{\Sigma}_a \right) = c - \frac{T-p}{2} \log \left(\left| \boldsymbol{\Sigma}_a \right| \right) - \frac{1}{2} \sum_{t=p+1}^T \text{tr} \left(\mathbf{a}_t' \boldsymbol{\Sigma}_a^{-1} \mathbf{a}_t \right)$$

Assim, aplicando tal função de verossimilhança condicional aos dados do problema encontremos as estimativas para o vetor de parâmetros $\hat{\boldsymbol{\beta}}$ e $\hat{\boldsymbol{\Sigma}}_a$ que maximizam o logaritmo da verossimilhança condicional.

7.6 Definindo a Ordem do Modelo

Já sabemos da efetividade dos critérios de informação para selecionar um modelo estatístico. Basicamente, todos os critérios de informação são baseados em verossimilhança e consistem de dois componentes.

O primeiro está relacionado com a qualidade do ajuste do modelo enquanto o segundo penaliza modelos complicados (muitos parâmetros a serem estimados). A qualidade do ajuste de um modelo é frequentemente medido pela máxima verossimilhança.

Três critérios são comumente usados para determinar a ordem de um modelo VAR(p). Assumindo que \mathbf{a}_t segue uma distribuição Normal multivariada, esses três critérios para um VAR(1) são:

$$\begin{aligned}
AIC(l) &= \ln \left| \widehat{\boldsymbol{\Sigma}}_{a,l} \right| + \frac{2}{T} lk^2 \\
BIC(l) &= \ln \left| \widehat{\boldsymbol{\Sigma}}_{a,l} \right| + \frac{\ln(T)}{T} lk^2 \\
HQ(l) &= \ln \left| \widehat{\boldsymbol{\Sigma}}_{a,l} \right| + \frac{2 \ln[\ln(T)]}{T} lk^2
\end{aligned}$$

onde T é o tamanho da amostra, l a defasagem do modelo e $\widehat{\boldsymbol{\Sigma}}_{a,l}$ é a estimativa de máxima verossimilhança para $\boldsymbol{\Sigma}_a$ que é a matriz de variâncias e covariâncias dos resíduos. Enquanto o critério AIC penaliza cada parâmetro por um fator 2, BIC e HQ aplicam uma penalidade que depende do tamanho da amostra. Podemos seguir o seguinte processo:

1. Estimar diversos modelos VAR com diferentes valores para p por meio do método de máxima verossimilhança condicional;
2. Escolher dentre os modelos aquele que apresente menor valor para algum critério de interesse.

7.7 Diagnóstico do Modelo Estimado

Como já vimos no caso univariado, verificar o modelo também conhecido como analisar os resíduos tem um papel importante na construção de um modelo econométrico. Os principais objetivos desta análise são garantir que o modelo ajustado é adequado.

Tipicamente, um modelo ajustado é dito ser adequado se:

- (1) Todos os parâmetros estimados são estatisticamente significantes
 - (i) O objetivo é eliminar do modelo parâmetros que não são estatisticamente significantes. Se alguma variável não Granger-causa as outras variáveis, podemos encontrar coeficientes nulos. Além disso, os dados não podem ser ricos o suficiente para fornecer estimativas precisas.
 - (ii) Assim, simplificamos o modelo VAR estimado usando, por exemplo, o critério de manter apenas parâmetros que são estatisticamente significantes ao nível de 5%.
- (2) Os resíduos não apresentam autocorrelação serial, correlação cruzada ou heterocedasticidade condicional
 - (i) Os resíduos de um modelo adequado deveriam se comportar como um ruído branco. Para verificar a autocorrelação serial podemos fazer uso dos gráficos da FAC e FACP dos resíduos de cada equação do modelo estimado. O ideal é não ter defasagens significativas.
 - (ii) Já a correlação cruzada pode ser visualizada por meio de gráfico ou via teste estatístico (conforme explicado neste [link](https://rpubs.com/hudsonchavs/multiseries)(<https://rpubs.com/hudsonchavs/multiseries>)).
 - (iii) A presença de heterocedasticidade condicional vies a análise dado que há o pressuposto que a variância do termo de erro de cada equação do modelo seja constante. A verificação da heterocedasticidade condicional pode ser feita pelo gráfico da FAC do quadrado dos resíduos de cada equação.
- (3) Analisar a estabilidade do modelo estimado através dos autovalores associados ao mesmo:
 - (i) Em módulo os autovalores do modelo precisam ser menores que 1 para que o modelo seja estacionário
 - (ii) Não podemos confundir a estacionariedade do modelo com a estacionariedade individual das séries
- (4) Os resíduos não violam a hipótese de distribuição, por exemplo, Normal.

7.8 Causalidade de Granger

[Granger \(1969\)](#) introduziu o conceito de causalidade, que é fácil de inserir a partir de um modelo VAR. Considere o caso de uma série temporal bivariada e a previsão h passos à frente. Neste caso, podemos usar o modelo VAR e o modelo univariado para cada componente e produzir as previsões.

Dizemos que r_{1t} causa r_{2t} se a previsão bivariada de r_{2t} tem maior acurácia do que a previsão univariada. Em outras palavras, sobre a abordagem de [Granger \(1969\)](#), dizemos que r_{1t} causa r_{2t} se as informações passadas de r_{1t} aumentam a acurácia da previsão de r_{2t} em relação a previsão obtida por meio de um modelo univariado para r_{2t} .

Porém, o teste não informa nada a respeito de causalidade em termos literais, mas oferece evidências estatísticas de que oscilações passadas de uma variável estão correlacionadas com as de outra variável.

Em um sistema bivariado, testar se r_{2t} Granger-causa r_{1t} equivale a estimar se para a equação abaixo, $\beta_{2,t-i} = 0$ para qualquer $i = 1, \dots, k$.

$$r_{1t} = \mu_1 + \beta_{1,t-1}r_{1,t-1} + \dots + \beta_{1,t-p}r_{1,t-p} + \beta_{2,t-1}r_{2,t-1} + \dots + \beta_{2,t-k}r_{2,t-k} + \varepsilon_{1,t}$$

Assim, temos o seguinte teste F convencional (para qualquer $i = 1, \dots, k$):

$$\begin{aligned} H_0 : \beta_{2,t-i} &= 0 & r_{2t} \text{ não Granger-causa } r_{1t} \\ H_1 : \beta_{2,t-i} &\neq 0 & r_{2t} \text{ Granger-causa } r_{1t} \end{aligned}$$

Basicamente, verificamos o p-valor da estatística de teste e decidimos se rejeitamos ou aceitamos a hipótese nula. Tal teste é realizado após a estimação de um modelo VAR. Além do objetivo proposto pelo teste, podemos utilizá-lo após estimar um modelo VAR e verificar que há defasagens não estatisticamente significantes e assim, decidir por retirá-las ou não de alguma equação do modelo VAR.

7.9 Função Impulso-Resposta

Esta função mostra os efeitos de choques nas variáveis do sistema e possibilita calcular o impacto (sinal e magnitude) dinâmico de mudança em uma variável sobre ela e as demais variáveis do modelo ao longo do tempo. Assim, procura-se medir o efeito de um choque unitário em uma variável no período t sobre todas as variáveis em períodos subsequentes.

Os coeficientes da função impulso-resposta são os coeficientes obtidos por meio da inversão do VAR(p) em um *Modelo Vetorial de Médias Móveis*, VMA(∞). Seja um VAR(1) com variáveis medidas em termos de desvios das médias, ou seja, $\tilde{r}_t = r_t - \mu$:

$$\tilde{r}_t = \Phi \tilde{r}_{t-1} + a_t$$

Esta formulação é apenas uma forma diferente de escrever o modelo VAR(1) definido no início do texto. Substituindo \tilde{r}_{t-1} recursivamente, produz:

$$\begin{aligned} \tilde{r}_t &= \Phi(\Phi \tilde{r}_{t-2} + a_{t-1}) + a_t \\ \tilde{r}_t &= \Phi^2 \tilde{r}_{t-2} + \Phi a_{t-1} + a_t \\ \tilde{r}_t &= \Phi^2(\Phi \tilde{r}_{t-3} + a_{t-2}) + \Phi a_{t-1} + a_t \\ \tilde{r}_t &= \Phi^3 \tilde{r}_{t-3} + \Phi^2 a_{t-2} + \Phi a_{t-1} + a_t \\ \tilde{r}_t &= a_t + \Phi a_{t-1} + \Phi^2 a_{t-2} + \Phi^3 a_{t-3} + \dots \\ \tilde{r}_t &= \sum_{i=0}^{t-1} \Phi^i a_{t-i} \text{ sendo } \Phi^0 = I \end{aligned}$$

que é a inversão do VAR(1) em um VMA(∞). A interpretação da função de impulso-resposta é dada por:

$$\frac{\partial r_t}{\partial a_{t-i}} = \Phi^i$$

em que o elemento (m, n) de Φ^i mostra o efeito de uma unidade de aumento no erro da variável n no tempo t sobre a variável m no tempo $t + i$, mantendo-se constante todos os outros erros. Se as variáveis estão medidas nos logaritmos os valores da função impulso-resposta podem ser interpretados como elasticidades. Normalmente a análise da função impulso-resposta se prende ao valor da resposta ao choque unitário, ao sinal e quantos períodos leva para o efeito do choque desaparecer. Para um VAR estável esta função eventualmente decresce até atingir zero.

Por exemplo, para um sistema consistindo de taxa de inflação (r_{1t}) e taxa de juros (r_{2t}), o efeito de um aumento na taxa de inflação pode ser de interesse. No mundo real, tal incremento ou choque pode

ser exogeneamente originado fora do sistema por eventos como a greve dos caminhoneiros. Assim, suponha o sistema:

$$\mathbf{r}_t = \begin{bmatrix} r_{1t} \\ r_{2t} \end{bmatrix}$$

e que ocorreu um choque na inflação. Para isolar tal efeito, suponha que os juros bem como a inflação assumem seu valor médio antes de $t = 0$, ou seja, que $\mathbf{r}_t = \boldsymbol{\mu}$ para $t < 0$. O aumento na inflação por uma unidade no período $t = 0$ será $a_{10} = 1$.

Agora, podemos mapear o que acontece com o sistema durante os períodos $t = 1, 2, \dots$ se nenhum outro choque ocorrer, ou seja, $\mathbf{a}_2 = \dots = \mathbf{a}_t = \mathbf{0}$. Além disso, se assumirmos que as duas variáveis tem média igual a zero ($\phi_0 = 0$) e que são modeladas por um $VAR(1)$ que pode ser escrito como:

$$\mathbf{r}_t = \Phi \mathbf{r}_{t-1} + \mathbf{a}_t$$

Supondo que já conhecemos a matriz Φ , o modelo é reescrito como:

$$\mathbf{r}_t = \begin{bmatrix} r_{1t} \\ r_{2t} \end{bmatrix} = \begin{bmatrix} 0.5 & 0 \\ 0.10 & 1 \end{bmatrix} \begin{bmatrix} r_{1,t-1} \\ r_{2,t-1} \end{bmatrix} + \begin{bmatrix} a_{1t} \\ a_{2t} \end{bmatrix}$$

Adicionando no sistema um choque unitário na primeira variável (inflação) em $t = 0$, temos:

$$\mathbf{r}_0 = \begin{bmatrix} r_{10} \\ r_{20} \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$$\mathbf{r}_1 = \Phi \mathbf{r}_0 \Rightarrow \begin{bmatrix} r_{11} \\ r_{21} \end{bmatrix} = \begin{bmatrix} 0.5 & 0 \\ 0.10 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0.5 \\ 0.1 \end{bmatrix}$$

$$\mathbf{r}_2 = \Phi \mathbf{r}_1 = \Phi \Phi \mathbf{r}_0 = \Phi^2 \mathbf{r}_0 \Rightarrow \begin{bmatrix} r_{12} \\ r_{22} \end{bmatrix} = \begin{bmatrix} 0.5 & 0 \\ 0.10 & 1 \end{bmatrix} \begin{bmatrix} 0.5 & 0 \\ 0.10 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0.25 \\ 0.06 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0.25 \\ 0.06 \end{bmatrix}$$

Perceba que $\mathbf{r}_i = (r_{1i}, r_{2i})'$ é a primeira coluna de Φ^i que é obtida de:

$$\mathbf{r}_t = \Phi^i \mathbf{a}_i$$

onde $\mathbf{a}_i' = (0, 0, \dots, 1, \dots, 0)$ e os elementos de Φ^i representam o efeito de um choque unitário nas variáveis do sistema depois de i períodos.

7.10 Função Impulso-Resposta Ortogonalizada

Uma hipótese problemática no tipo de resposta ao impulso mostrada anteriormente é que este choque ocorre apenas em uma variável em um tempo. Tal hipótese é razoável se as variáveis são independentes.

Em geral, a matriz de variância e covariância do termo de erro \mathbf{a}_t não é diagonal, ou seja, existe correlação contemporânea entre os erros das diferentes equações do modelo VAR. Assim, os choques são contemporaneamente correlacionados e isso impede que se tenha o efeito puro/líquido de cada choque.

Quando ocorre choque em uma variável do sistema seu efeito se confunde com o efeito advindo da correlação dos erros de outras variáveis. Não há como garantir que o choque ocorre em cada variável isoladamente. Para contornar este problema é necessário transformar os erros em erros ortogonais que terão matriz de variância e covariâncias diagonal. Esta operação pode ser denominada de ortogonalização dos erros ou diagonalização da matriz de variâncias e covariâncias.

7.11 Decomposição da Variância do Erro de Previsão

Uma análise complementar à função de impulso-resposta é a decomposição da variância do erro de previsão que procura determinar qual a percentagem da variância do erro de previsão de uma variável que é devido a ela e qual percentagem é devido a cada uma das outras variáveis do modelo ao longo do horizonte de previsão. Esta análise fornece elementos para se inferir sobre causalidade ao longo de um período temporal e sobre qual variável é mais exógena no sistema.

As previsões obtidas com o modelo VAR contém dois elementos: o valor esperado para a variável e o choque inesperado em cada equação. A decomposição de variância inicia com o cálculo da variância do erro de previsão do VAR na forma de médias móveis, isto é, na forma invertida VMA(∞). Assim, o erro de previsão h períodos à frente no modelo VAR(p) na forma VMA(∞) é dado por:

$$\tilde{r}_{t+h} - E[\tilde{r}_{t+h}|T] = a_{t+h} + \Phi a_{t+h-1} + \Phi^2 a_{t+h-2} + \Phi^3 a_{t+h-3} + \dots$$

O lado esquerdo desta equação é a diferença entre o valor observado do vetor de variáveis endógenas no tempo $t + h$ e o valor previsto pelo VAR. O lado direito é a representação VMA(∞) dos erros de previsão. Observe que o erro de previsão corrente depende dos termos de erros passados. A análise precisa ser feita com os erros ortogonais.

A decomposição da variância do erro de previsão é, normalmente, apresentada em forma de tabela que indica a percentagem do erro de previsão de uma variável que pode ser atribuída a ela e a cada uma das outras do sistema h períodos à frente depois do choque. Uma variável tipicamente exógena terá alta percentagem explicada por ela própria por um longo período. Se uma variável é importante para a dinâmica temporal de outra variável, um erro de previsão na primeira variável terá efeito sobre o erro de previsão na segunda.

Basicamente, a análise concentra no percentual da variância do erro de previsão de uma variável que decorre dela e de cada uma das outras, se o percentual varia ao longo do horizonte e por quantos períodos uma variável permanece importante para explicar a outra.

7.12 Modelo Vetorial Autorregressivo Estrutural

Nesta parte vamos apresentar o modelo VAR estrutural (SVAR) como uma extensão do modelo VAR mostrado anteriormente. Apesar do modelo VAR conseguir captar as características dinâmicas de séries temporais multivariadas, identificamos a necessidade de transformar a matriz de variância e covariâncias dos erros para se ter choques ortogonais, isto é, não correlacionados, para a análise correta das funções de impulso-resposta e decomposição do erro de previsão.

Contudo, mesmo tendo erros ortogonais surge a dificuldade de como interpretar as funções impulso-resposta sem referência à teoria econômica. A incorporação da teoria para definir as relações contemporâneas entre as variáveis conduz ao VAR na forma estrutural (SVAR).

7.12.1 Do VAR para o SVAR

Uma série temporal multivariada $r_t = (r_{1t}, r_{2t}, \dots, r_{kt})'$ composta por k componentes no tempo t é um processo VAR de ordem p , ou VAR(p), se para $p > 0$ segue o modelo:

$$r_t = \phi_0 + \Phi_1 r_{t-1} + \dots + \Phi_p r_{t-p} + a_t$$

onde ϕ_0 é um vetor de dimensão k , Φ_j são matrizes $k \times k$ para $j = 1, \dots, p$ e a_t é um ruído branco formado por uma sequência de vetores aleatórios independentes e identicamente distribuídos com média 0 e matriz de covariância Σ_a .

Assumindo $\phi_0 = 0$ para simplificar a notação e assumindo a representação de médias móveis, VMA(∞), dada por:

$$r_t = a_t + \Phi a_{t-1} + \Phi^2 a_{t-2} + \Phi^3 a_{t-3} + \dots$$

em que $\tilde{r}_t = \sum_{i=0}^{t-1} \Phi^i a_{t-i}$ sendo $\Phi^0 = I$.

Como vimos, os coeficientes das matrizes Φ^i fornecem as respostas das variáveis a choques nos sistema, constituindo as funções impulso-resposta. Porém, quando os erros a_t são contemporaneamente correlacionados esses coeficientes não refletem de forma adequada os efeitos dos choques nas variáveis do sistema. Assim, os erros devem ser ortogonalizados de modo a apresentarem matriz de variâncias e covariâncias diagonal (ausência de covariância). A ortogonalização consiste em uma transformação dos erros gerando outro vetor de erros com matriz de variâncias e covariâncias diagonal. No entanto, esta transformação envolve modificar todo o modelo o que traz consequências para o relacionamento contemporâneo entre as variáveis.

Dado \sum_a simétrica e positiva definida, existe uma matriz P não singular (determinante diferente de zero) tal que $P \sum_a P'$ é uma matriz diagonal. Multiplicando o modelo por P , tem-se:

$$Pr_t = P\Phi_1 r_{t-1} + \dots + P\Phi_p r_{t-p} + Pa_t$$

$$Pr_t = \Phi_1^* r_{t-1} + \dots + \Phi_p^* r_{t-p} + u_t$$

que constitui um novo modelo, com novo erro $u_t = Pa_t$ e com o termo Pr_t do lado esquerdo que incorpora relações contemporâneas entre as variáveis do sistema.

Tal modelo é conhecido como VAR Estrutural que permite a análise de funções impulso-resposta ortogonais, decomposição da variância do erro de previsão e estimativas de coeficientes de relações contemporâneas entre as variáveis. A matriz de variâncias e covariâncias dos erros estruturais é dada por:

$$Var(u_t) = \sum_a = E[u_t u_t'] = E[Pa_t a_t' P'] = PE[a_t a_t'] P' = P \sum_a P'$$

7.12.2 do SVAR para o VAR

De forma alternativa, podemos definir um VAR a partir de um SVAR. Considere o seguinte modelo SVAR:

$$Ar_t = A_1^* r_{t-1} + \dots + A_p^* r_{t-p} + Bu_t$$

Pré-multiplicando por A^{-1} , temos:

$$A^{-1}Ar_t = A^{-1}A_1^* r_{t-1} + \dots + A^{-1}A_p^* r_{t-p} + A^{-1}Bu_t$$

$$r_t = \Phi_1 r_{t-1} + \dots + \Phi_p r_{t-p} + a_t$$

que é um VAR padrão e $a_t = A^{-1}Bu_t$, ou, $Aa_t = Bu_t$.

Observe que a_t são os erros do modelo VAR na sua forma reduzida (correlacionados) enquanto que u_t são os erros da forma estrutural (não correlacionados).

7.12.3 Identificação

Como é possível observar, o modelo estrutural não é observável, mas a forma reduzida pode ser estimada. A estratégia é especificar e estimar o VAR e depois estimar o SVAR e concentrar na análise de funções impulso-resposta, decomposição de variância e relações contemporâneas entre as variáveis.

Para estimar o SVAR temos que resolver primeiro o problema de *identificação*. Será que a partir da estimativa de \sum_a é possível obter estimativas dos coeficientes de A e B do modelo estrutural?

Para isso, é preciso impor restrições nas matrizes A e B . O número de restrições necessário para alcançar identificação depende da relação entre o número de coeficientes estimados na forma reduzida e o número de coeficientes a serem obtidos na forma estrutural. É um problema de número de incógnitas e de número de equações (relações) semelhante à condição de ordem em equações simultâneas.

O número de parâmetros do VAR que servem para identificar os elementos das matrizes A e B é o número de coeficientes não redundantes da matriz de variâncias e covariâncias Σ_a . Os coeficientes das variáveis defasadas não contam. Como Σ_a é simétrica, temos $\frac{k(k+1)}{2} = \frac{k^2+k}{2}$ coeficientes que é o número máximo de elementos identificáveis em A e B . Se nosso modelo tem 3 variáveis, teremos $\frac{3(3+1)}{2} = 6$ elementos em Σ_a e só podemos identificar 6 elementos em A e B . Por outro lado, se nosso modelo tem 4 variáveis, teremos $\frac{4(4+1)}{2} = 10$ elementos em Σ_a e só podemos identificar 10 elementos em A e B .

Resumindo, temos k elementos em A e k elementos em B totalizando $2k^2$ elementos a serem identificados em A e B (por exemplo, para $k = 2$ as matrizes A e B são 2×2 e teremos 4 elementos em cada, totalizando $2k^2 = 8$ elementos a serem identificados). Porém, como mostramos anteriormente, apenas $\frac{k(k+1)}{2} = \frac{k^2+k}{2}$ destes elementos são observados/estimados pela matriz Σ_a .

Para identificar $2k^2$ coeficientes desconhecidos a partir de $\frac{k(k+1)}{2}$ valores de Σ_a , será necessário impor $2k^2 - \frac{k(k+1)}{2} = k^2 + \frac{k(k-1)}{2}$ restrições nas matrizes A e B .

Com base em $Aa_t = Bu_t$ e dependendo das restrições impostas, três tipos de modelos SVAR podem ser definidos:

7.12.4 Modelo A

Neste modelo a ideia é modelar as relações contemporâneas entre as variáveis diretamente pela matriz A considerando $B = I_k$, isto é:

$$Aa_t = u_t$$

o que reduz o número de valores desconhecidos de $2k^2$ para k^2 , pois só temos a matriz A a ser identificada. Assim, o número mínimo de restrições para identificação será $\frac{k(k-1)}{2}$.

Por exemplo, em um sistema de 4 variáveis temos 16 elementos a serem identificados. Os elementos estimados são 10 da matriz Σ_a . Logo, temos que impor, no mínimo 6 restrições. A matriz A seria:

$$A = \begin{bmatrix} a_{11} & 0 & 0 & 0 \\ a_{21} & a_{22} & 0 & 0 \\ a_{31} & a_{32} & a_{33} & 0 \\ a_{41} & a_{42} & a_{43} & a_{44} \end{bmatrix}$$

Assim, o sistema é exatamente identificado. Se colocarmos 1 na diagonal principal o sistema fica super identificado. Se a teoria indica um relacionamento contemporâneo não recursivo, os zeros podem aparecer em posições diferentes fora da diagonal principal.

7.12.5 Modelo B

Neste modelo, ao invés de modelar diretamente as relações contemporâneas entre as variáveis, especifica-se as relações entre os erros identificando-se os choques estruturais diretamente pelos choques da forma reduzida. A matriz A é considerada identidade, isto é, $A = I_k$. Logo,

$$a_t = Bu_t$$

Ou seja, os erros da forma reduzida são funções lineares dos erros estruturais. O número mínimo de restrições para identificação também é $\frac{k(k-1)}{2}$.

7.12.6 Modelo AB

É o modelo que considera os dois tipos de restrições simultaneamente, ou seja:

$$Aa_t = Bu_t$$

Como A e B têm k^2 elementos cada, temos que impor no mínimo $k^2 + \frac{k(k-1)}{2}$ restrições para identificação.

Para o exemplo do modelo VAR do BACEN apresentado em sala de aula, poderíamos estimar um SVAR do tipo AB (ignorando as defasagens de r_t e ϕ_0), como segue:

$$Ar_t = Bu_t$$

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ a_{21} & 1 & 0 & 0 \\ a_{31} & a_{32} & 1 & 0 \\ a_{41} & a_{42} & a_{43} & 1 \end{bmatrix} \begin{bmatrix} \Delta juros \\ \Delta cambio \\ \Delta adm \\ \Delta livres \end{bmatrix} = \begin{bmatrix} b_{11} & 0 & 0 & 0 \\ 0 & b_{22} & 0 & 0 \\ 0 & 0 & b_{33} & 0 \\ 0 & 0 & 0 & b_{44} \end{bmatrix} \begin{bmatrix} u_{juros} \\ u_{cambio} \\ u_{adm} \\ u_{livres} \end{bmatrix}$$

Como $k = 4$, é necessário um mínimo de $2k^2 - \frac{k(k+1)}{2} = k^2 + \frac{k(k-1)}{2} = 22$ restrições em A e B para identificação do modelo. Observe que exatamente 22 restrições foram impostas e o sistema é *exatamente identificado*.

Se chamarmos r_{t-1}, \dots, r_{t-p} de "termos defasados", o modelo pode ser escrito como:

$$\Delta juros_t = \phi_{10} + \text{termos defasados} + b_{11}u_{juros}$$

$$\Delta cambio_t = \phi_{20} - a_{21}\Delta juros_t + \text{termos defasados} + b_{22}u_{cambio}$$

$$\Delta adm_t = \phi_{30} - a_{31}\Delta juros_t - a_{32}\Delta cambio_t + \text{termos defasados} + b_{33}u_{adm}$$

$$\Delta livres_t = \phi_{40} - a_{41}\Delta juros_t - a_{42}\Delta cambio_t - a_{43}\Delta adm_t + \text{termos defasados} + b_{44}u_{livres}$$

Observe que a forma como escrevemos as matrizes A e B definiram que a variável mais endógena é $\Delta livres$ e a mais exógena é $\Delta juros$. Isso é de suma importância quando temos embasamento teórico capaz de sustentar esta escolha.

Perceba que a forma de especificação de A altera a caracterização das variáveis como mais endógenas e mais exógenas. Por exemplo, caso optássemos por:

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ a_{31} & 0 & 1 & 0 \\ a_{41} & 0 & a_{43} & 1 \end{bmatrix} \begin{bmatrix} \Delta juros \\ \Delta cambio \\ \Delta adm \\ \Delta livres \end{bmatrix} = \begin{bmatrix} b_{11} & 0 & 0 & 0 \\ 0 & b_{22} & 0 & 0 \\ 0 & 0 & b_{33} & 0 \\ 0 & 0 & 0 & b_{44} \end{bmatrix} \begin{bmatrix} u_{juros} \\ u_{cambio} \\ u_{adm} \\ u_{livres} \end{bmatrix}$$

Como $k = 4$, é necessário um mínimo de $2k^2 - \frac{k(k+1)}{2} = k^2 + \frac{k(k-1)}{2} = 22$ restrições em A e B para identificação do modelo. Observe que 25 restrições foram impostas e o sistema é *super identificado*.

Se chamarmos r_{t-1}, \dots, r_{t-p} de "termos defasados", o modelo pode ser escrito como:

$$\Delta j_{uros_t} = \phi_{10} + \text{termos defasados} + b_{11}u_{j_{uros}}$$

$$\Delta cambio_t = \phi_{20} + \text{termos defasados} + b_{22}u_{cambio}$$

$$\Delta adm_t = \phi_{30} - a_{31}\Delta j_{uros_t} + \text{termos defasados} + b_{33}u_{adm}$$

$$\Delta livres_t = \phi_{40} - a_{41}\Delta j_{uros_t} - a_{42}\Delta cambio_t + \text{termos defasados} + b_{44}u_{livres}$$

Agora, temos duas variáveis mais exógenas no sistema que são Δj_{uros} e $\Delta cambio$ enquanto que a variável mais endógena continua sendo $\Delta livres$.

Como outro exemplo, suponha o modelo apresentado por [Lütkepohl \(2005\)](#) envolvendo as variáveis produto (q_t), taxa de juros (i_t) e moeda (m_t). Os erros da forma reduzida são denotados por:

$$\mathbf{a}_t = (a_t^q, a_t^i, a_t^m)'$$

e os erros estruturais por:

$$\mathbf{u}_t = (u_t^{IS}, u_t^{LM}, u_t^m)'$$

De acordo com uma visão keynesiana a relação entre os erros da forma reduzida e os erros estruturais é dada por:

$$a_t^q = -a_{12}a_t^i + b_{11}u_t^{IS} \quad \text{que representa a curva IS}$$

$$a_t^i = -a_{21}a_t^q - a_{23}a_t^m + b_{22}u_t^{LM} \quad \text{que representa a curva LM}$$

$$a_t^m = b_{33}u_t^m \quad \text{que representa a regra de oferta de moeda}$$

A primeira equação representa uma curva IS com um parâmetro negativo para o termo de erro da taxa de juros e um choque estrutural na própria IS. A segunda equação é melhor entendida resolvendo a demanda por moeda em função do termo de erro na taxa de juros, ou seja, $a_t^m = \beta_1 a_t^q + \beta_2 a_t^i + u_t^{LM}$ de onde se tira que β_1 deve ser positivo porque mais moeda é necessário para um maior volume de transação e que β_2 deve ser negativo porque quanto mais alta a taxa de juros maior o custo de se reter moeda e, conseqüentemente, menor a quantidade de moeda será demandada. Por fim, a terceira equação postula que o termo de erro da base monetária é determinado por choques exógenos na oferta monetária. As três equações forma um modelo AB que pode ser escrito na forma $\mathbf{Aa}_t = \mathbf{Bu}_t$:

$$\begin{bmatrix} 1 & a_{12} & 0 \\ a_{21} & 1 & a_{23} \\ 0 & 0 & 1 \end{bmatrix} \mathbf{a}_t = \begin{bmatrix} b_{11} & 0 & 0 \\ 0 & b_{22} & 0 \\ 0 & 0 & b_{33} \end{bmatrix} \mathbf{u}_t$$

Como $k = 3$, é necessário um mínimo de $2k^2 - \frac{k(k+1)}{2} = k^2 + \frac{k(k-1)}{2} = 12$ restrições em \mathbf{A} e \mathbf{B} para identificação do modelo. Com três 1 e três 0 em \mathbf{A} e seis 0 em \mathbf{B} a condição é satisfeita e o modelo é exatamente identificado.

7.13 Processo de Estimação

Abaixo, os passos para estimação e avaliação dos modelos VAR, SVAR e VEC. Apesar de ainda não termos estudado os modelos VEC (próxima etapa da disciplina), o processo aqui considera tal

modelo. Porém, perceba que dependendo das decisões nas etapas, seguimos com a estimação dos modelos VAR e SVAR.

- (1) Visualizar os dados e identificar observações fora do padrão (outliers, sazonalidade, tendência)
- (2) Se necessário, transformar os dados para estabilizar a variância (logaritmo ou retirar sazonalidade, por exemplo)
- (3) Avaliar a função de correlação cruzada para confirmar a possibilidade de modelagem multivariada.
- (4) Testar se os dados são estacionários ou cointegrados:
 - (i) Caso não tenha raiz unitária (estacionários), estimar VAR com as séries em nível
 - (ii) Caso tenha raiz unitária, mas sem cointegração é preciso diferenciar os dados até se tornarem estacionários e estimar VAR com as séries diferenciadas
 - (iii) Caso tenha raiz unitária, mas com cointegração devemos estimar o VEC com as séries em nível
- (5) Definir a ordem p para os dados em análise por meio de critérios de informação (escolher modelo com menor AIC, por exemplo)
- (6) Estimar o modelo escolhido no passo 4
 - (i) Se VAR (forma reduzida):
 - Verificar significância estatística do modelo estimado e, caso seja necessário, eliminar parâmetros não significantes.
 - Analisar a causalidade de Granger (variáveis que não granger causa as demais podem ser retiradas do modelo)
 - (ii) Se SVAR (forma estrutural):
 - Definir a estrutura para as matrizes A e B e o modelo de interesse (A, B ou AB)
 - Verificar significância estatística do modelo estimado e, caso seja necessário, eliminar parâmetros não significantes.
 - Analisar a causalidade de Granger (variáveis que não granger causa as demais podem ser retiradas do modelo)
 - (iii) Se VEC (Modelo Vetorial de Correção de Erros)
 - Usar a quantidade de vetores de cointegração obtidos no teste de cointegração para estimar o modelo VEC
- (7) Examinar se os resíduos se comportam como ruído branco e condições de estacionariedade do modelo. Caso contrário, retornar ao passo 3 ou 4.
 - (i) Verificar a autocorrelação serial por meio da FAC e FACP dos resíduos de cada equação do modelo estimado. O ideal é não ter defasagens significativas.
 - (ii) Verificar correlação cruzada por meio da FCC dos resíduos.
 - (iii) Analisar a estabilidade do modelo estimado através dos autovalores associados ao mesmo.
 - (iv) Verificar a distribuição de probabilidade (Normal) para os resíduos de cada equação do modelo.
 - (v) Analisar heterocedasticidade condicional (resíduos devem ser homocedásticos, ou seja, variância condicional constante)
- (8) Uma vez que os resíduos são ruído branco e o modelo é estável:

- (i) Analisar funções de resposta ao impulso
- (ii) Analisar a importância das variáveis para explicar a variância do erro de previsão de cada variável
- (iii) Fazer previsões para as variáveis do modelo

Referências Bibliográficas

- Athanasopoulos, G. e Hyndman, R. J. (2008). Modelling and forecasting australian domestic tourism. *Tourism Management*, 29(1):19–31.
- Bartlett, M. S. (1946). On the theoretical specification and sampling properties of autocorrelated time-series. *Supplement to the Journal of the Royal Statistical Society*, 8(1):27.
- Bhargava, A. (1986). On the theory of testing for unit roots in observed time series. *The Review of Economic Studies*, 53(3):369.
- Bowman, A. W. e Azzalini, A. (1997). *Applied smoothing techniques for data analysis: the kernel approach with S-Plus illustrations*. Oxford: Oxford University Press.
- Box, G. E. e Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society: Series B (Methodological)*, 26(2):211–243.
- Box, G. E. P., Reinsel, G. C., e Jenkins, G. M. (2015). *Time Series Analysis: Forecasting and Control*. John Wiley & Sons.
- Brown, R. G. (1959). *Statistical forecasting for inventory control*. McGraw/Hill.
- Cleveland, R. B., Cleveland, W. S., McRae, J. E., e Terpenning, I. (1990). Stl: A seasonal-trend decomposition procedure based on loess (with discussion). *Journal of Official Statistics*, 6:3–73.
- Cleveland, W. S. (1979). Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association*, 74(368):829–836.
- Cleveland, W. S. e Loader, C. (1996). Smoothing by local regression: Principles and methods. In *Statistical theory and computational aspects of smoothing*, pages 10–49. Springer.
- Cox, D. R. e Stuart, A. (1955). Some quick sign tests for trend in location and dispersion. *Biometrika*, 42(1/2):80.
- Cryer, J. D. e Chan, K.-S. (2008). *Time series analysis : with applications in R*. Springer, New York.
- da Silveira Bueno, R. D. L. (2011). *Econometria de Séries Temporais*.
- D'Agostino, R. B. (1970). Transformation to normality of the null distribution of g_1 . *Biometrika*, 57(3):679.
- Dickey, D. A. e Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, 74(366a):427–431.
- Dickey, D. A. e Fuller, W. A. (1981). Likelihood ratio statistics for autoregressive time series with a unit root. *Econometrica*, 49(4):1057.

- Dixon, W. J. (1950). Analysis of extreme values. *The Annals of Mathematical Statistics*, 21(4):488–506.
- Dixon, W. J. (1951). Ratios involving extreme values. *The Annals of Mathematical Statistics*, pages 68–78.
- Doornik, J. A. e Hansen, H. (2008). An omnibus test for univariate and multivariate normality. *Oxford Bulletin of Economics and Statistics*, 70:927–939.
- EPA, D. Q. A. (2006). A reviewer's guide. Technical report, EPA QA/G-9R 2000 & EPA/240/B-06/002.
- Ferreira, P. G. C. (2018). *Análise de Séries Temporais em R: curso introdutório*. Rio de Janeiro: Elsevier.
- Franse, A. (2017). Using the multiplicative double seasonal holt-winters method to forecast short-term electricity demand.
- Friedman, M. (1937). The use of ranks to avoid the assumption of normality implicit in the analysis of variance. *Journal of the American Statistical Association*, 32(200):675–701.
- Fuller, W. A. (1995). *ntroduction to Statistical Time Series*. John Wiley & Sons.
- Gardner, E. S. (1985). Exponential smoothing: The state of the art. *Journal of Forecasting*, 4(1):1–28.
- Gardner, E. S. e Mckenzie, E. (1985). Forecasting trends in time series. *Management Science*, 31(10):1237–1246.
- Gould, P. G., Koehler, A. B., Ord, J. K., Snyder, R. D., Hyndman, R. J., e Vahid-Araghi, F. (2008). Forecasting time series with multiple seasonal patterns. *European Journal of Operational Research*, 191(1):207–222.
- Granger, C. e Newbold, P. (1974). Spurious regressions in econometrics. *Journal of Econometrics*, 2(2):111–120.
- Granger, C. W. J. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, 37(3):424.
- Grubbs, F. E. (1969). Procedures for detecting outlying observations in samples. *Technometrics*, 11(1):1–21.
- Guerrero, V. M. (1993). Time-series analysis supported by power transformations. *Journal of Forecasting*, 12(1):37–48.
- Hamilton, J. D. (1994). *Time series analysis*, volume 2. Princeton university press Princeton.
- Hinkley, D. (1977). On quick choice of power transformation. *Applied Statistics*, 26(1):67.
- Holt, C. C. (2004). Forecasting seasonals and trends by exponentially weighted moving averages. *International Journal of Forecasting*, 20(1):5–10.
- Hyndman, R. e Athanasopoulos, G. (2018). *Forecasting : principles and practice*. OTexts, Melbourne, 2 edition.
- Hyndman, R., Koehler, A. B., Ord, J. K., e Snyder, R. D. (2008). *Forecasting with Exponential Smoothing*. Springer Berlin Heidelberg.

- Hyndman, R. J. e Khandakar, Y. (2008). Automatic time series forecasting: TheforecastPackage for R. *Journal of Statistical Software*, 27(3).
- Hyndman, R. J. e Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting*, 22(4):679–688.
- Hyndman, R. J., Koehler, A. B., Snyder, R. D., e Grose, S. (2002). A state space framework for automatic forecasting using exponential smoothing methods. *International Journal of Forecasting*, 18(3):439–454.
- Iglewicz, B. e Hoaglin, D. (1993). *How to detect and handle outliers*. ASQC Quality Press, Milwaukee, Wis.
- Jenkins, G. (1979). *Practical experiences with modelling and forecasting time series*. Gwilym Jenkins and Partners (Overseas) Ltd, St. Helier.
- Kruskal, W. H. e Wallis, W. A. (1952). Use of ranks in one-criterion variance analysis. *Journal of the American Statistical Association*, 47(260):583–621.
- Kwiatkowski, D., Phillips, P. C., Schmidt, P., e Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root. *Journal of Econometrics*, 54(1-3):159–178.
- Loader, C. (1999). *Local Regression and Likelihood*. Springer New York.
- Lobato, I. N. e Velasco, C. (2004). A simple test of normality for time series. *Econometric Theory*, pages 671–689.
- Lucambio, F. (2008). Diferentes testes para verificar normalidade de uma amostra aleatória. [<https://bit.ly/2WSuKa8>]. Technical report, Departamento de Estatística da Universidade Federal do Paraná - UFPR.
- Lütkepohl, H. (2005). *New Introduction to Multiple Time Series Analysis*. Springer-Verlag GmbH.
- Makridakis, S., Wheelwright, S. C., e Hyndman, R. J. (1998). *Forecasting methods and applications*. John Wiley and Sons.
- Mann, H. B. (1945). Nonparametric tests against trend. *Econometrica*, 13(3):245.
- Maravall, A. (2011). Seasonality tests and automatic model identification in tramo-seats. *Bank of Spain*.
- Maravall, A. (2012). Update of seasonality tests and automatic model identification in tramo-seats. *Bank of Spain*.
- Mattos, R. (2018). Tendências e raízes unitárias. texto didático. [<https://bit.ly/2Y0l96d>]. Technical report, Universidade Federal de Juiz de Fora.
- McBane, G. C. (2006). Programs to compute distribution functions and critical values for extreme value ratios for outlier detection. *Journal of Statistical Software*, 16(3).
- McBean, E. A. e Rovers, F. A. (1992). Estimation of the probability of exceedance of contaminant concentrations. *Groundwater Monitoring & Remediation*, 12(1):115–119.
- McNutt, M. (2014). Raising the bar.
- Morettin, P. A. e Toloi, C. (2006). *Análise de séries temporais*. Blucher.

- Nabeya, S., Tanaka, K., et al. (1988). Asymptotic theory of a test for the constancy of regression coefficients against the random walk alternative. *The Annals of Statistics*, 16(1):218–235.
- Nadaraya, E. A. (1964). On estimating regression. *Theory of Probability & Its Applications*, 9(1):141–142.
- Parzen, E. (1961). Mathematical considerations in the estimation of spectra. *Technometrics*, 3(2):167–190.
- Pegels, C. C. (1969). Exponential forecasting: some new variations. *Management Science*, pages 311–315.
- Perlin, M. (2018). *Processamento e Análise de Dados Financeiros e Econômicos com o R* [<https://www.msperlin.com/padfeR>]. Self Published.
- Phillips, P. C. B. (1987). Time series regression with a unit root. *Econometrica*, 55(2):277.
- Phillips, P. C. B. e Perron, P. (1988). Testing for a unit root in time series regression. *Biometrika*, 75(2):335–346.
- R Core Team (2020). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Rorabacher, D. B. (1991). Statistical treatment for rejection of deviant values: critical values of Dixon's "q" parameter and related subrange ratios at the 95% confidence level. *Analytical Chemistry*, 63(2):139–146.
- Rosner, B. (1983). Percentage points for a generalized esd many-outlier procedure. *Technometrics*, 25(2):165–172.
- Rothamsted, V., Lewis, T., e Barnett, V. (1996). Outliers in statistical data.
- Shenton, L. R. e Bowman, K. O. (1977). A bivariate model for the distribution of $\sqrt{b_1}$ and $\sqrt{b_2}$. *Journal of the American Statistical Association*, 72(357):206–211.
- Shumway, R. H. e Stoffer, D. S. (2017). *Time Series Analysis and Its Applications*. Springer International Publishing.
- Sims, C. A. (1980). Macroeconomics and reality. *Econometrica*, 48(1):1.
- Stefansky, W. (1972). Rejecting outliers in factorial designs. *Technometrics*, 14(2):469–479.
- Stephens, M. A. (1974). EDF statistics for goodness of fit and some comparisons. *Journal of the American Statistical Association*, 69(347):730–737.
- Stephens, M. A. (1976). Asymptotic results for goodness-of-fit statistics with unknown parameters. *The Annals of Statistics*, 4(2):357–369.
- Stephens, M. A. (1977). Goodness of fit for the extreme value distribution. *Biometrika*, 64(3):583–588.
- Stephens, M. A. (1979). Tests of fit for the logistic distribution based on the empirical distribution function. *Biometrika*, 66(3):591–595.
- Stone, C. J. (1977). Consistent nonparametric regression. *The Annals of Statistics*, 5(4):595–620.
- Stone, C. J. (1980). Optimal rates of convergence for nonparametric estimators. *The Annals of Statistics*, 8(6):1348–1360.

- Taylor, J. W. (2003a). Exponential smoothing with a damped multiplicative trend. *International Journal of Forecasting*, 19(4):715–725.
- Taylor, J. W. (2003b). Short-term electricity demand forecasting using double seasonal exponential smoothing. *Journal of the Operational Research Society*, 54(8):799–805.
- Tietjen, G. L. e Moore, R. H. (1972). Some grubbs-type statistics for the detection of several outliers. *Technometrics*, 14(3):583–597.
- Wald, A. e Wolfowitz, J. (1940). On a test whether two samples are from the same population. *The Annals of Mathematical Statistics*, 11(2):147–162.
- Watson, G. S. (1964). Smooth regression analysis. *Sankhyā: The Indian Journal of Statistics, Series A*, pages 359–372.
- Wei, W. (2019). *Time series analysis univariate and multivariate methods*. Pearson, Boston.
- Werner, L., Bisognin, C., e Menegotto, L. (2017). Previsão da umidade relativa do ar diária de porto alegre. *Cadernos do IME: Série Estatística. Rio de Janeiro, RJ. Vol. 43 (dez. 2017), p. 39-59*.
- Winters, P. R. (1960). Forecasting sales by exponentially weighted moving averages. *Management Science*, 6(3):324–342.

Capítulo 8

Adendo A

Neste adendo encontra-se o artigo *Diferentes testes para verificar normalidade de uma amostra aleatória* de Fernando Lucambio do Departamento de Estatística da Universidade Federal do Paraná ([Lucambio, 2008](#)).

Diferentes testes para verificar normalidade de uma amostra aleatória

Fernando Lucambio

Departamento de Estatística
Universidade Federal do Paraná
Curitiba/PR, 81531-990, Brasil

email: `lucambio@ufpr.br`

Maio de 2008

Mostraremos os testes para normalidade mais utilizados: teste Jarque-Beta, teste D'Agostino, teste χ^2 de Pearson de bondade de ajuste, teste Shapiro-Wilk, teste Lilliefors, teste Anderson-Darling e teste de Cramer-von Mises. Este tema é extremamente importante e por isso existem muitas outras propostas de testes para verificar normalidade de amostras em diferentes contextos, uma excelente referência é Stephens (1986).

1 Teste Jarque-Beta

Proposto por Bera & Jarque (1980), baseia-se na diferença entre os coeficientes de skewness e kurtosis dos dados y_1, y_2, \dots, y_n e àqueles da distribuição assumida normal.

As hipóteses nula e alternativa no teste Jarque-Bera são:

$$H_0 : y_1, y_2, \dots, y_n \sim N(\mu, \sigma^2) \quad vs \quad H_1 : \text{não } H_0$$

A estatística de teste é

$$JB = n \left(\frac{\alpha_3^2}{6} + \frac{(\alpha_4 - 3)^2}{24} \right),$$

onde

$$\begin{aligned} \alpha_3 &= \frac{\sum_{i=1}^n (y_i - \bar{y})^3}{ns^3}, \\ \alpha_4 &= \frac{\sum_{i=1}^n (y_i - \bar{y})^4}{ns^4}, \\ s^2 &= \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n}. \end{aligned}$$

Aqui, \bar{y} é a média amostral e s^2 , α_3 e α_4 o segundo, terceiro e quarto momentos centrais, respectivamente. A estatística JB tem distribuição assintótica $\chi^2(2)$ sob a hipóteses nula.

O teste de Jarque-Beta é conhecido por ter boas propriedades para verificar normalidade, é claro e simples de calcular e é muito utilizado no contexto de regressão em econometria. Uma limitação é que somente pode ser utilizado na verificação de normalidade.

No R (R Development Core Team, 2006) a função `rjb.test` no pacote *lawstat* fornece a estatística de teste JB acima assim como o p-valor e também *NormalityTests* no pacote *fBasics*.

2 Teste D'Agostino-Pearson

Também conhecido como teste D foi proposto por D'Agostino (1970) e têm sido muito utilizado para verificar normalidade. Suponha que y_1, y_2, \dots, y_n é a amostra aleatória e que $y_{(1)}, y_{(2)}, \dots, y_{(n)}$ é a amostra ordenada, isto é, $y_{(1)} \leq y_{(2)} \leq \dots \leq y_{(n)}$.

A estatística D de teste é

$$D = \frac{T}{n^2 s},$$

onde s é o desvio padrão amostral, o qual é calculado como a raiz quadrado positiva de s^2 segundo definido no contexto do teste Jarque-Bera e

$$T = \sum_{i=1}^n \left(i - \frac{n+1}{2} \right) y_{(i)}.$$

Se a amostra é da distribuição normal, temos que

$$E\{D\} = \frac{(n-1)\Gamma(\frac{n}{2} - \frac{1}{2})}{2\sqrt{2n\pi}\Gamma(\frac{n}{2})} \approx \frac{1}{2\sqrt{\pi}} \approx 0.28209479,$$

o desvio padrão assintótico da estatística D é

$$s\{D\} = \sqrt{\frac{12\sqrt{3} - 37 + 2\pi}{24n\pi}} \approx \frac{0.02998598}{\sqrt{n}}.$$

Utiliza-se a estatística D padronizada como

$$D^* = \frac{D - E\{D\}}{s\{D\}},$$

a qual têm distribuição normal aproximada sob hipótese nula. No R a função *NormalityTests* no pacote *fBasics* fornece a estatística de teste D acima assim como o p-valor.

3 Teste χ^2 de Pearson de bondade de ajuste

A estatística do teste Pearson é

$$P = \sum_{i=1}^n \frac{(C_i - E_i)^2}{E_i},$$

onde C_i é o número de observações e E_i é o número de observações esperadas (sob a hipótese nula) na i -ésima classe. Estas classes são escolhidas de maneira sejam equiprováveis sob a hipótese de normalidade.

O p-valor é calculado da distribuição χ^2 com graus de liberdade iguais ao número de classes menos 3 ou ao número de classes menos 1 se decide-se pela correção ou não (a correção recomenda-se somente em pequenos tamanhos de amostras). Em ambos os casos este não é o correto valor do p-valor, ficando na maioria das vezes entre estes dois valores (Moore, 1986).

Este teste não é recomendado para testar normalidade devido às inferiores propriedades comparado aos outros testes disponíveis. Em situações práticas calcula-se o p-valor da distribuição χ^2 com graus de liberdade iguais ao número de classes menos 3, devido à estimação dos dois parâmetros da densidade normal. As funções R *pchiTest* no pacote *fBasics* e *pearson.test* no pacote *nortest* calculam este teste. Devido a que, como mencionado, o verdadeiro valor do p-valor na maioria das vezes fica entre dois valores recomenda-se utilizar primeiro a função *pearson.test(y, adjust=TRUE)* e depois *pearson.test(y, adjust=FALSE)*, também sugere-se alterar o número de classes e observar o efeito disso no p-valor.

4 Teste Shapiro-Wilk

Proposto por Shapiro & Wilk (1965) utiliza a estatística

$$W = \frac{(\sum_{i=1}^n a_i y_{(i)})^2}{\sum_{i=1}^n (y_i - \bar{y})^2},$$

onde as constantes a_1, a_2, \dots, a_n são calculadas como a solução de

$$(a_1, a_2, \dots, a_n) = \frac{m^\top V^{-1}}{(m^\top V^{-1} V^{-1} m)^{1/2}},$$

sendo $m = (m_1, m_2, \dots, m_n)^\top$ o vetor dos valores esperados das estatísticas de ordem da amostra e V a matriz de covariâncias dessas estatísticas. O p-valor deste teste é calculado exatamente para $n = 3$, em outras situações utilizam-se aproximações diferentes para $4 \leq n \leq 11$ e para $n \geq 12$, (Shapiro & Francia, 1972).

No R as funções *shapiro.test* no pacote *stats* fornece a estatística de teste acima assim como o p-valor e a função *sfTest* no pacote *fBasics*.

5 Teste Lilliefors

O teste Lilliefors, uma modificação do teste Kolmogorov-Smirnov, utiliza a estatística D de Kolmogorov-Smirnov que mede a diferença máxima absoluta entre a função de distribuição acumulada empírica e teórica.

A estatística de teste é

$$D = \max\{D^+, D^-\},$$

onde $D^+ = \max_{i=1, \dots, n} \{\frac{i}{n} - p_{(i)}\}$, $D^- = \max_{i=1, \dots, n} \{p_{(i)} - \frac{i-1}{n}\}$, e $p_{(i)} = \Phi([x_{(i)} - \bar{x}]/s)$. Nesta expressão, Φ é a função de distribuição acumulada normal padrão e \bar{x} and s são a média e desvio padrão amostrais.

O p-valor é calculado pela fórmula em Dallal & Wilkinson (1986), a qual é confiável quando o p-valor é menor do que 0.1. Se o p-valor Dallal-Wilkinson fornece um valor maior do que 0.1 o verdadeiro p-valor deve ser calculado da distribuição da estatística de teste modificada

$$Z = D \left(\sqrt{n} - 0.01 + \frac{0.85}{\sqrt{n}} \right)$$

(Stephens, 1986), o p-valor é obtido então via simulações e aproximações.

As funções R *ksnormTest* no pacote *fBasics* e *ks.test* no pacote *nortest* calculam o teste Kolmogorov-Smirnov e as funções *lillieTest* no pacote *fBasics* e *lillie.test* no pacote *nortest* calculam o teste Lilliefors.

O teste Lilliefors (Kolmogorov-Smirnov) é o mais famoso teste para verificar normalidade. Comparado aos testes Anderson-Darling e Cramer-von Mises é conhecido que sua performance é pior. Embora as estatísticas obtidas das funções R *lillie.test(y)* e *ks.test(y, "pnorm", mean(y), sd(y))* sejam iguais, isto não acontece com o p-valor devido a que a distribuição das estatísticas de teste diferem quando estimados os parâmetros.

6 Teste Anderson-Darling

Proposto por Anderson & Darling (1952) é mais utilizado quando o tamanho da amostra não é superior a 25. Este teste é baseado na função de distribuição empírica, a idéia é que dada a função de distribuição sob hipóteses nula, os dados podem ser transformados à distribuição uniforme. Os dados transformados podem então serem testados para uniformidade.

A estatística de teste é

$$A = -n - \frac{1}{n} \sum_{i=1}^n [2i - 1] [\ln(p_{(i)}) + \ln(1 - p_{(n-i+1)})],$$

onde $p_{(i)} = \Phi([y_{(i)} - \bar{y}]/s)$ são os percentis ordenados da distribuição normal padrão e Φ representa a função de distribuição acumulada normal padrão.

O p-valor é calculado da estatística modificada

$$Z = \left(1.0 + \frac{0.75}{n} + \frac{2.25}{n^2}\right) A,$$

como $1 - e^{-13.436+101.14Z-223.73Z^2}$ se $Z < 0.2$, $1 - e^{-8.318+42.796Z-59.938Z^2}$ se $0.2 \leq Z < 0.34$, $e^{0.9177-4.279Z-1.38Z^2}$ se $0.34 \leq Z < 0.6$ e $e^{1.2937-5.709Z+0.0186Z^2}$ se $Z \geq 0.6$. No R as funções *ad.test* no pacote *nortest* fornece a estatística de teste acima assim como o p-valor e a função *adTest* no pacote *fBasics*.

7 Teste Cramer-von Mises

Este teste é também baseado na distribuição acumulada, foi proposto por Darling (1957).

A estatística de teste é

$$W = \frac{1}{12n} + \sum_{i=1}^n \left(p_{(i)} - \frac{2i-1}{2n}\right),$$

onde $p_{(i)} = \Phi([x_{(i)} - \bar{x}]/s)$ são os percentis ordenados da distribuição normal padrão e Φ representa a função de distribuição acumulada normal padrão.

O p-valor é calculado da estatística modificada

$$Z = \left(1.0 + \frac{1}{2n}\right) W,$$

como $1 - e^{-13.953+775.5Z-12542.61Z^2}$ se $Z < 0.0275$, $1 - e^{-5.903+179.546Z-1515.29Z^2}$ se $0.0275 \leq Z < 0.051$, $e^{0.886-31.62Z+10.897Z^2}$ se $0.051 \leq Z < 0.092$ e $e^{1.111-34.242Z+12.832Z^2}$ se $Z \geq 0.092$. No R as funções *cvm.test* no pacote *nortest* fornece a estatística de teste acima assim como o p-valor e a função *cvmTest* no pacote *fBasics*.

Referências

- Anderson, T.W. & Darling, D.A. (1952). Asymptotic theory of certain goodness-of-fit criteria based on stochastic processes. *Annals of Mathematical Statistics*, **23**, 193–212.
- Bera, A. & Jarque, C. (1980). Efficient test for normality, heterocedasticity and serial independence of regression residuals. *Econometrics Letters*, **6**, 255–259.

- D'Agostino, R.B. (1970). Transformation to normality of the null distribution of g_1 . *Biometrika*, **57**(3), 679–681.
- Dallal, G.E. & Wilkinson, L. (1986). An analytic approximation to the distribution of Lilliefors's test for normality. *The American Statistician*, **40**, 291–296.
- Darling, D.A. (1957). The Kolmogorov-Smirnov, Cramer-von Mises Tests. *Annals of Mathematical Statistics*, **28**(4 (Dec.)), 823–838.
- Moore, D.S. (1986). *Tests of the chi-squared type*. D'Agostino, R.B. and Stevens, M.A., eds.: Goodness-of-Fit Techniques. New York: Marcel Dekker.
- R Development Core Team (2006). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.
- Shapiro, S.S. & Francia, R.S. (1972). An Approximate Analysis of Variance Test for Normality. *Journal of the American Statistical Association*, **67**, 215–216.
- Shapiro, S.S. & Wilk, M. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, **52**, 591–611.
- Stephens, M.A. (1986). *Tests based on EDF statistics*. In *Goodness-of-Fit Techniques*. D'Agostino, R.B. and Stevens, M.A., eds.: Goodness-of-Fit Techniques. New York: Marcel Dekker.