# Previsão de séries temporais

## Adaptação dos slides do livro:

**M.L. Berenson, D.M. Levine, D.F. Stephan, T.C. Krehbiel, Statistics for Managers Using Microsoft Excel, 6ª Edição, Capítulo 16**



**Prof. Dr. Evandro Marcos Saidel Ribeiro**
**FEA-RP**
**Universidade de São Paulo**

## Neste capítulo:

- Sobre os diferentes modelos de séries temporais de previsão:
  - médias móveis,
  - suavização exponencial,
  - tendência linear,
  - tendência quadrática,
  - tendência exponencial,
  - modelos auto-regressivos e
  - modelos de mínimos quadrados para os dados sazonais
- Escolher o modelo de previsão de séries temporais mais adequado

# A Importância da Previsão

- Previsões governamentais sobre taxas de desemprego, taxas de juros e receitas esperadas dos impostos de renda para fins de política

- Previsões de executivos de marketing sobre demanda, vendas e preferências dos consumidores para o planejamento estratégico

- Para administradores de faculdade na previsão matrículas para planejar instalações e para o recrutamento de professores

- Previsão de lojas de varejo para controlar os níveis de estoque, contratar funcionários e fornecer treinamento

# Abordagem Comum para Previsões

**Abordagem Comum para Previsões**

- **Métodos de Previsão Qualitativos**
  - Usado quando os dados históricos estão indisponíveis
  - Considerado altamente subjetivo
- **Métodos de Previsão Quantitativos**
  - Séries Temporais
  - Causal
  - Use dados passados para prever valores futuros

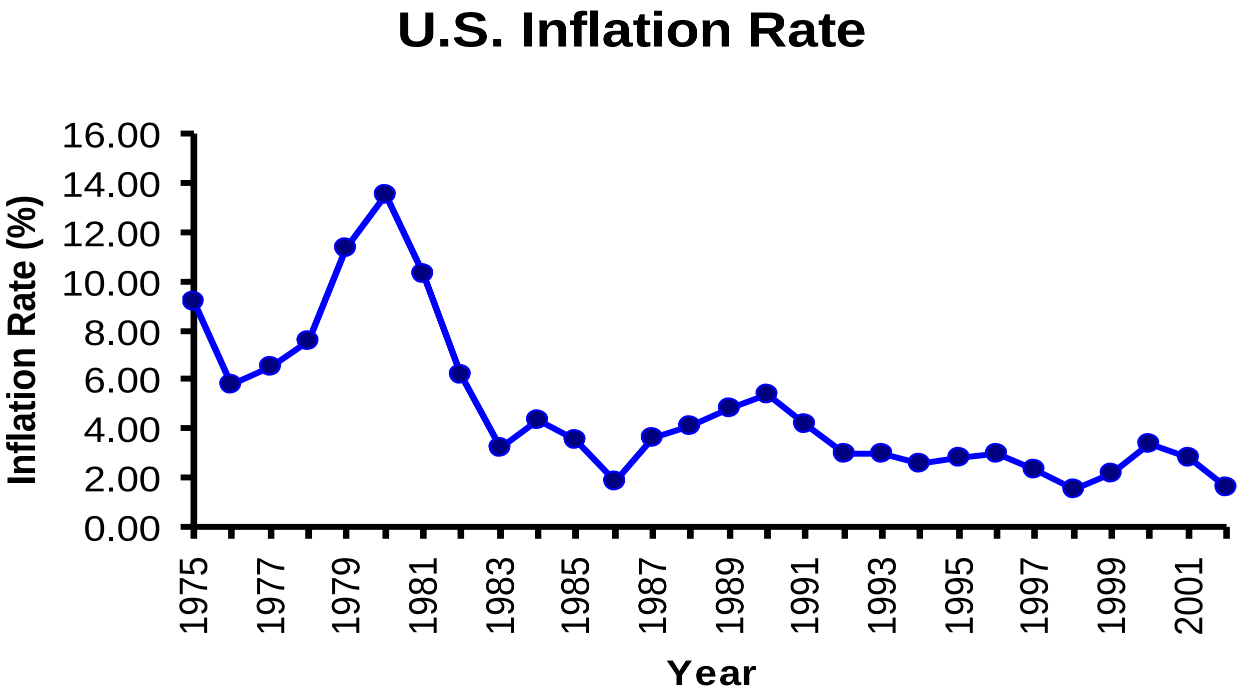
- Dados numéricos obtidos em intervalos de tempo regulares

- Os intervalos de tempo pode ser, anualmente, trimestral, mensal, semanal, diária, por hora, etc.

- Exemplo:

| | | | | | |
|---|---|---|---|---|---|
| Ano: | 2005 | 2006 | 2007 | 2008 | 2009 |
| Vendas: | 75,3 | 74,2 | 78,5 | 79,7 | 80,2 |

Um gráfico de série temporal é um gráfico bidimensional de dados em períodos




- o eixo vertical mede a variável de interesse

- o eixo horizontal corresponde aos períodos

# Componentes das Séries Temporais

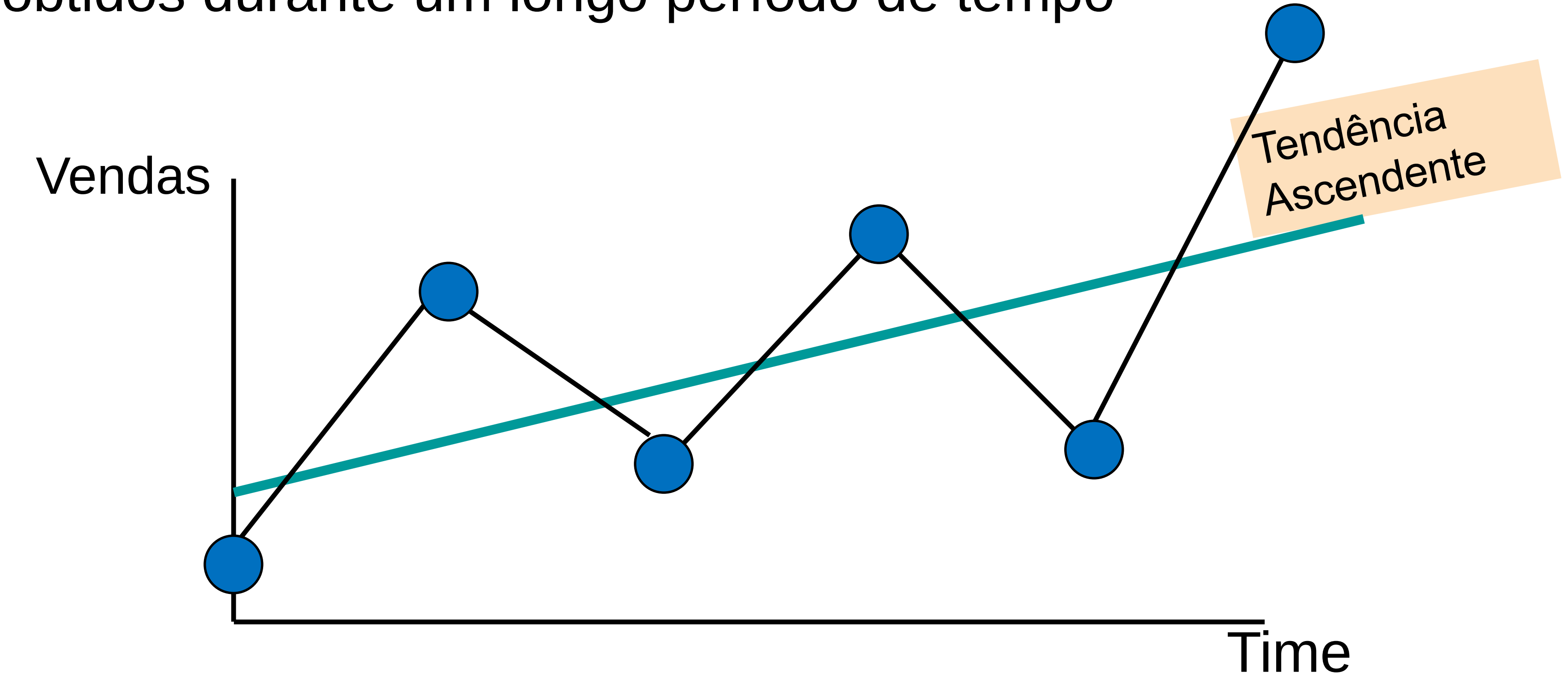
**Séries Temporais**

| Componente de Tendência | Componente Sazonal | Componente Cíclica | Componente Irregular |
|---|---|---|---|
| No geral, persistente, o movimento de longo prazo | Flutuações periódicas regulares, geralmente dentro de um período de 12 meses | Repetição de oscilações ou movimentos ao longo de mais de um ano | Flutuações erráticas ou residuais |

# Componente de Tendência

- Aumento ou diminuição ao longo prazo (em geral movimento ascendente ou descendente)

- Dados obtidos durante um longo período de tempo

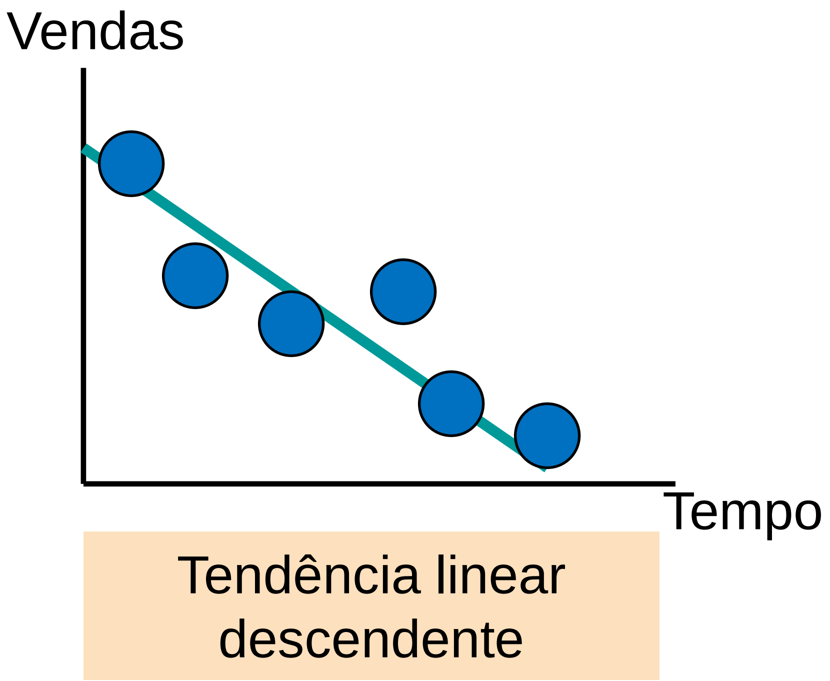


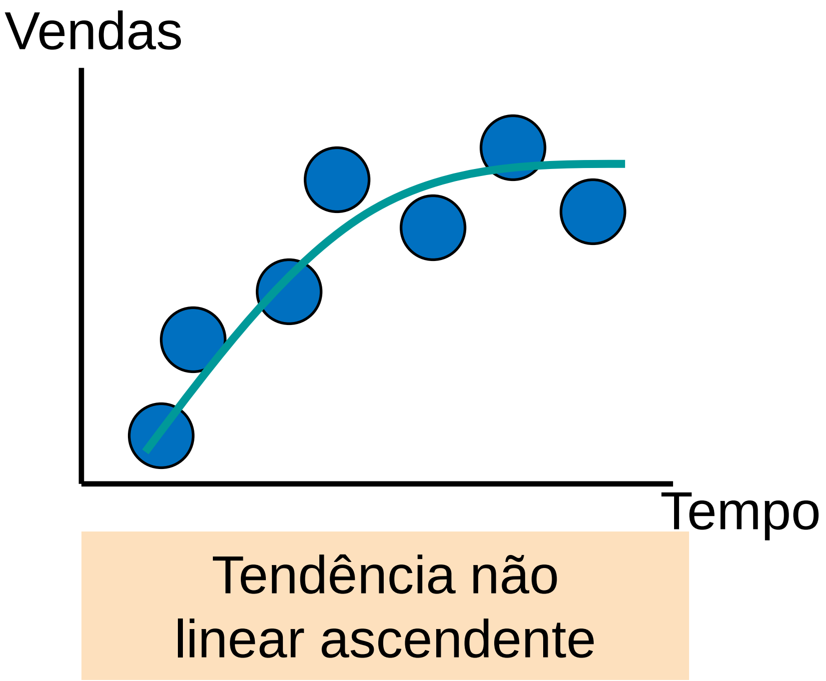
# Componente de Tendência

*(continuação)*

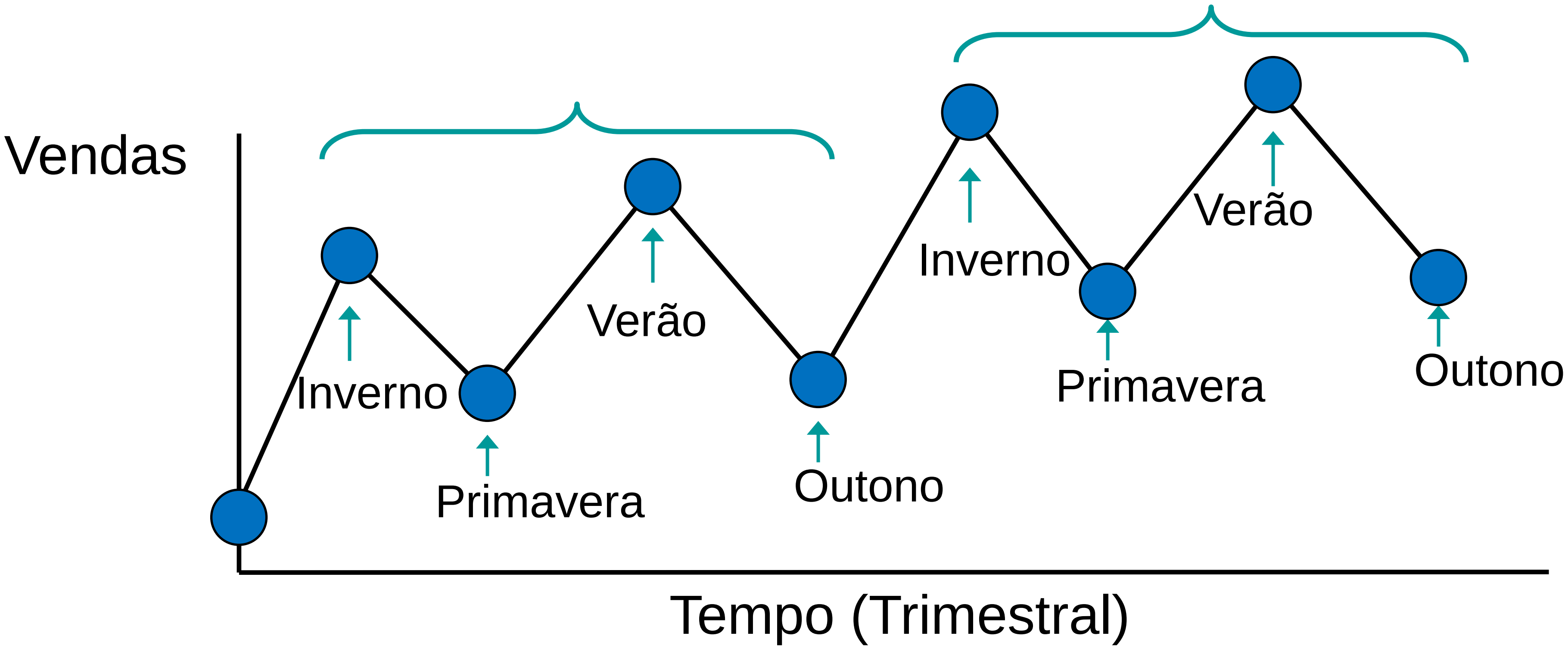
- Tendência pode ser ascendente ou descendente

- Tendência pode ser linear ou não-linear

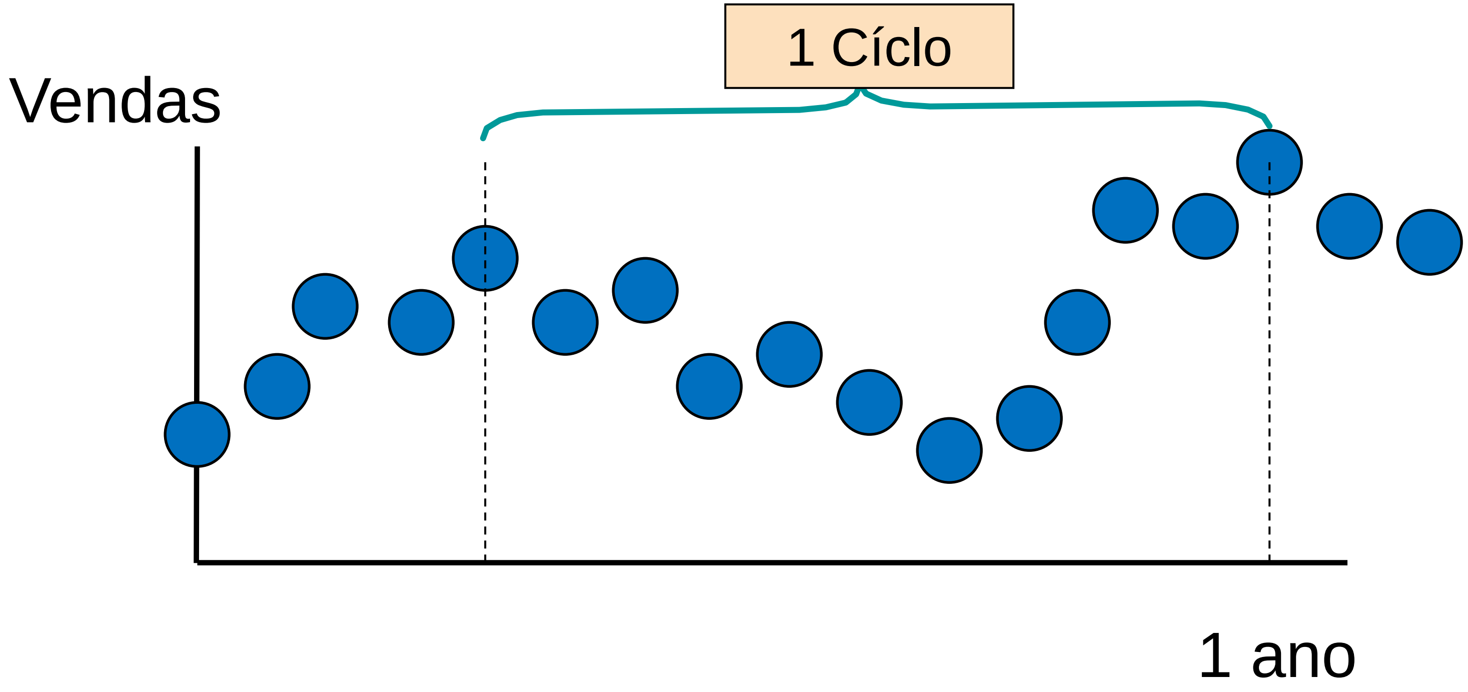


- Padrão oscilatório em períodos pequenos
- Observado dentro de 1 ano
- Muitas vezes, mensal ou trimestral

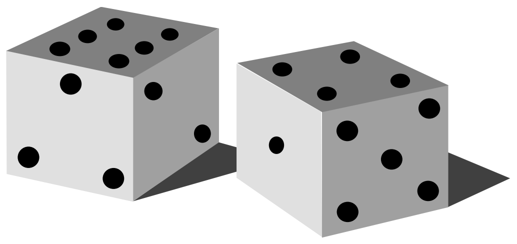


- Padrões de oscilação de longo prazo
- Regularmente ocorrer, mas podem variar em comprimento
- Muitas vezes medida de pico a pico ou através de calha



# Componente Irregular

- Imprevisíveis, aleatório, flutuações "residuais“
- Devido à variação aleatória:
  - Natureza
  - Acidentes ou eventos incomuns
- “Ruído” da série histórica



# Suas séries temporais possuem componente de tendência ?

- Um gráfico de séries temporais deve ajudá-lo a responder à esta pergunta.
- Muitas vezes fazer a suavização dos dados da série temporal ajuda a responder a pergunta.
- Dois métodos de suavização populares são: médias móveis e suavização exponencial.

- **Médias Móveis – MM(L)**
  - Calcular médias móveis para obter uma impressão geral do padrão de movimento ao longo do tempo
  - Médias de valores de séries temporais consecutivas para um período escolhido de comprimento L

- **Suavização exponencial – SE**
  - A média móvel ponderada

# Médias Móveis – MM(L)

- Usado para suavizar
- Uma série de médias aritméticas ao longo do tempo
- O resultado depende da escolha de L (duração do período de meios de computação)
- Última média móvel de comprimento L podem ser extrapolados em um período futuro para uma previsão de curto prazo
- Exemplos:
  - Para uma média móvel de 5 anos, L = 5
  - Para uma média móvel de sete anos, L = 7
  - etc.

*(continuação)*

- **Exemplo:** Média móvel de cinco anos

  – Primeira Média:

$$\boxed{MM(5)_1 = \frac{Y_1 + Y_2 + Y_3 + Y_4 + Y_5}{5}}$$

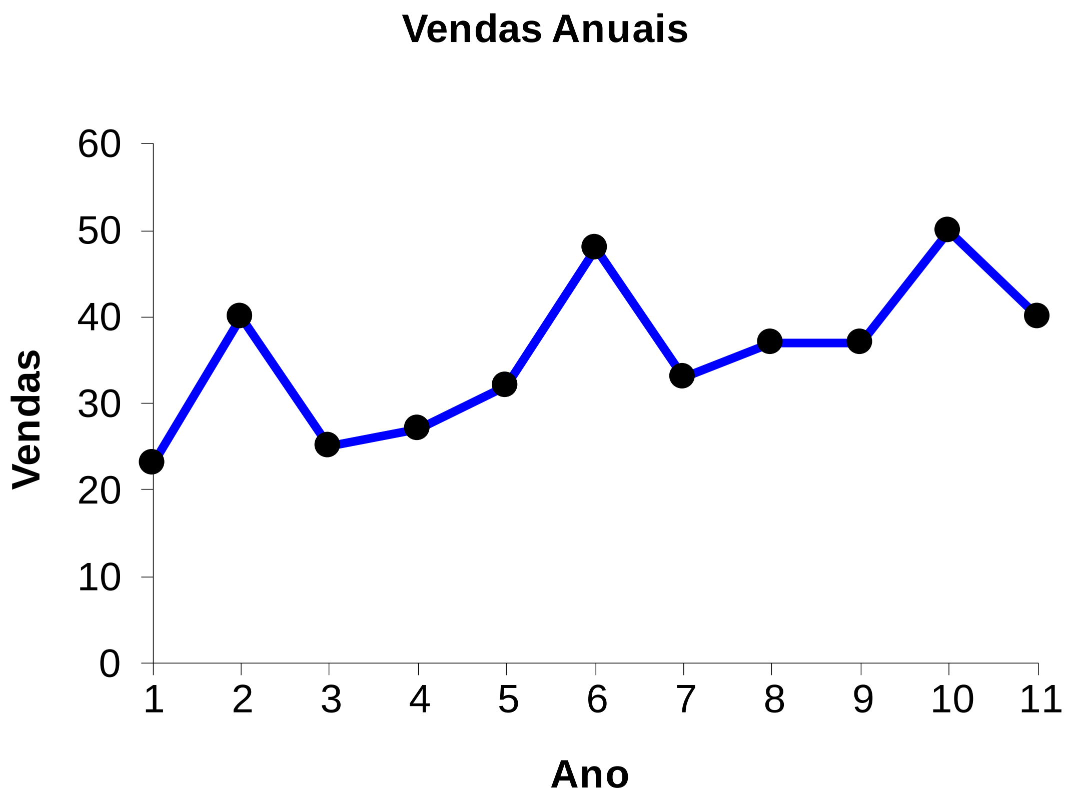
  – Segunda Média:

$$\boxed{MM(5)_2 = \frac{Y_2 + Y_3 + Y_4 + Y_5 + Y_6}{5}}$$

  – etc.

# Exemplo: Os dados anuais

| Ano | Vendas |
|---|---|
| 1 | 23 |
| 2 | 40 |
| 3 | 25 |
| 4 | 27 |
| 5 | 32 |
| 6 | 48 |
| 7 | 33 |
| 8 | 37 |
| 9 | 37 |
| 10 | 50 |
| 11 | 40 |
| etc… | etc… |



# Calcular médias móveis

| Ano | Vendas |
|---|---|
| 1 | 23 |
| 2 | 40 |
| 3 | 25 |
| 4 | 27 |
| 5 | 32 |
| 6 | 48 |
| 7 | 33 |
| 8 | 37 |
| 9 | 37 |
| 10 | 50 |
| 11 | 40 |

etc…

| Média dos Anos | Média Móvel de 5 anos |
|---|---|
| 3 | 29,4 |
| 4 | 34,4 |
| 5 | 33,0 |
| 6 | 35,4 |
| 7 | 37,4 |
| 8 | 41,0 |
| 9 | 39,4 |
| … | … |

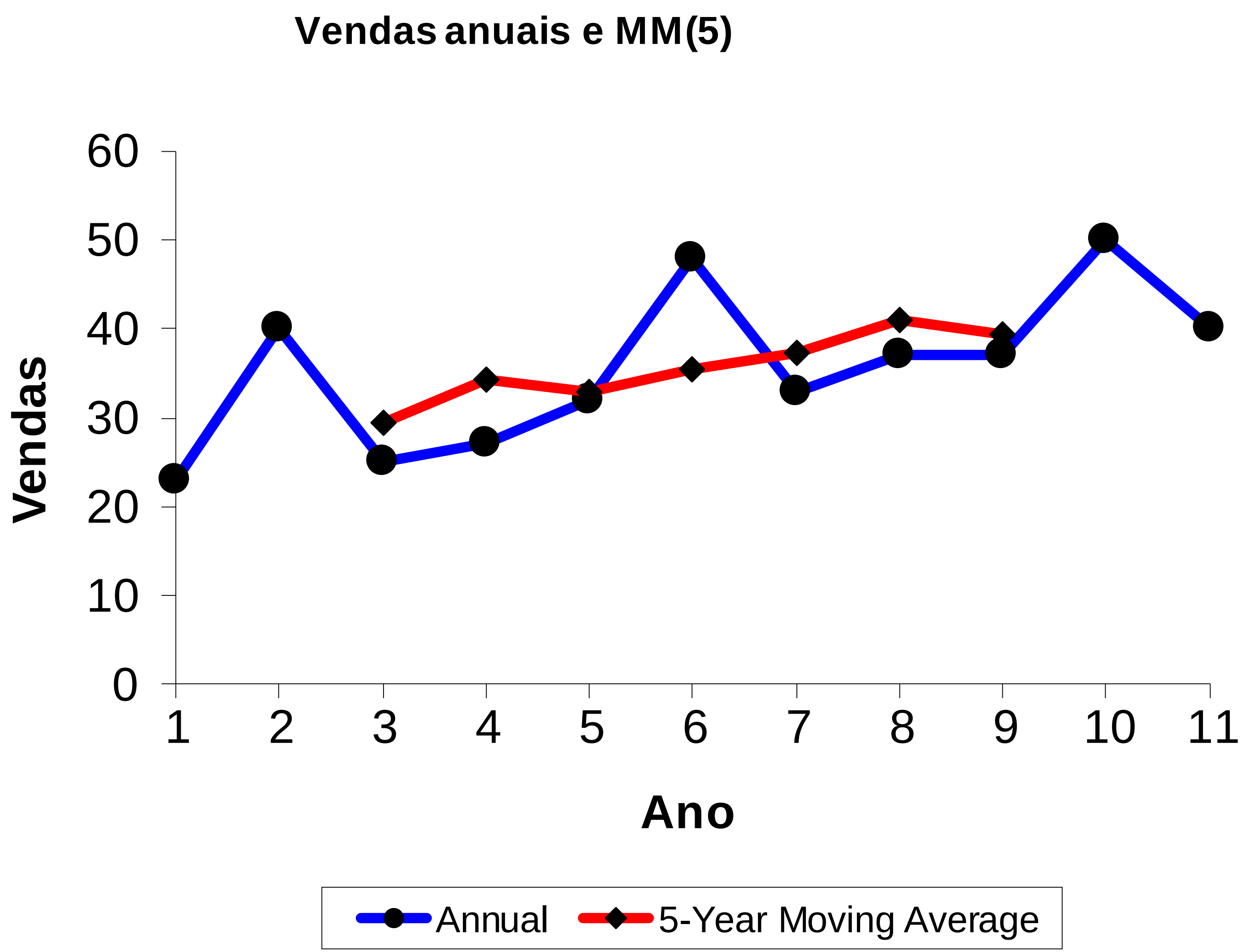
$$3 = \frac{1+2+3+4+5}{5}$$

$$29.4 = \frac{23+40+25+27+32}{5}$$

- Cada média móvel é para um bloco consecutivo de 5 anos

# Anual vs. Média Móvel

A média móvel de 5 anos “suaviza” os dados e torna mais fácil para ver a tendência subjacente



- Usado para suavizamento e previsão de curto prazo (um período para o futuro)
- A média móvel **ponderada**
  - Pesos diminuem exponencialmente
  - Observação mais recente é dado o maior peso
- O peso (coeficiente suavização) é **W**
  - Subjetivamente escolhido
  - Varia de 0 a 1
  - Menor W dá mais suavização, maior W dá menos suavização
- O peso é :
  - Perto de 0 para suavizar componentes cíclicos e irregulares indesejados
  - Perto de 1 para a previsão

- Modelo de Suavização Exponencial

$$\boxed{E_1 = Y_1}$$

$$\boxed{E_i = WY_i + (1 - W)E_{i-1}}$$

Para i = 2, 3, 4, …

onde:
$E_i$ = valor exponencialmente suavizado para um período i
$E_{i-1}$ = valor exponencialmente suavizado já computados para período i – 1
$Y_i$ = valor observado no período i
W = peso (coeficiente suavização), 0 < W < 1

# Exemplo Suavização Exponencial

- Supondo o peso W = **0,2**

| Período de tempo(i) | Vendas ($Y_i$) | Previsão do período anterior($E_{i-1}$) | Valor da Suavização Exponencial neste periodo ($E_i$) |
|---|---|---|---|
| 1 | 23 | -- | 23 |
| 2 | 40 | 23 | (.2)(40)+(.8)(23)=26.4 |
| 3 | 25 | 26.4 | (.2)(25)+(.8)(26.4)=26.12 |
| 4 | 27 | 26.12 | (.2)(27)+(.8)(26.12)=26.296 |
| 5 | 32 | 26.296 | (.2)(32)+(.8)(26.296)=27.437 |
| 6 | 48 | 27.437 | (.2)(48)+(.8)(27.437)=31.549 |
| 7 | 33 | 31.549 | (.2)(48)+(.8)(31.549)=31.840 |
| 8 | 37 | 31.840 | (.2)(33)+(.8)(31.840)=32.872 |
| 9 | 37 | 32.872 | (.2)(37)+(.8)(32.872)=33.697 |
| 10 | 50 | 33.697 | (.2)(50)+(.8)(33.697)=36.958 |
| etc. | etc. | etc. | etc. |

$E_1 = Y_1$

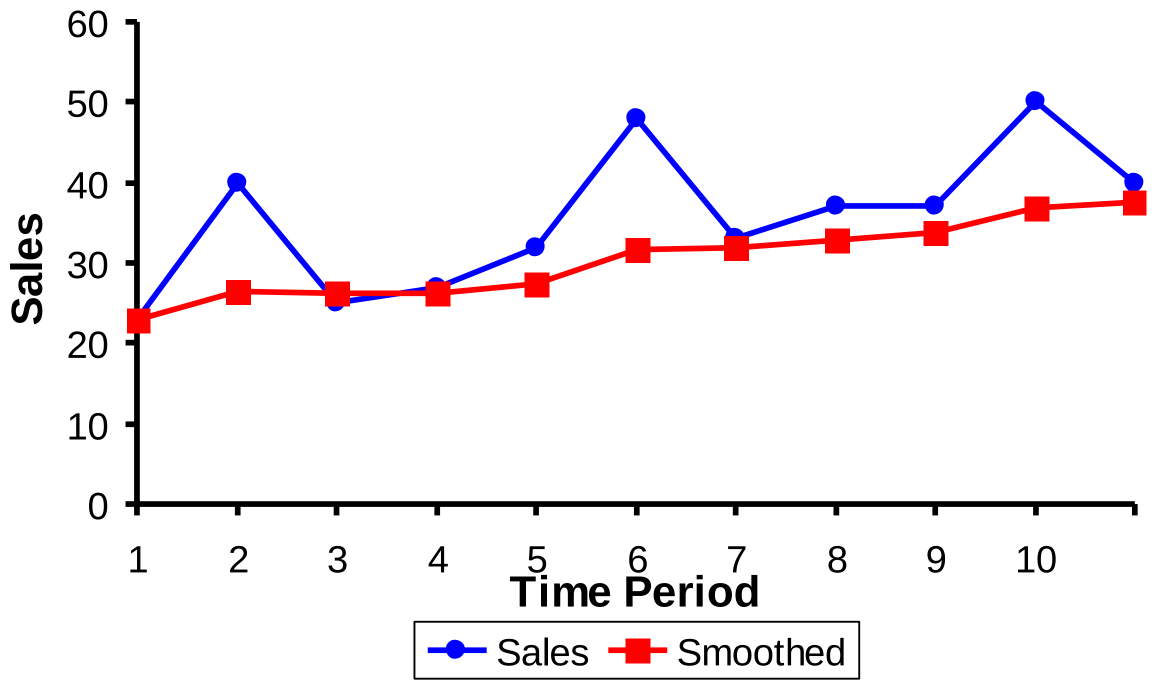
já que não existe informação prévia

$$E_i = WY_i + (1 - W)E_{i-1}$$

# Vendas vs. Vendas Suavizadas

- As flutuações foram suavizadas

- **NOTA:** o valor suavizado, neste caso, é geralmente um pouco mais baixo, uma vez que a tendência é ascendente e o fator de ponderação é de apenas **0.2**



# Previsão Período i + 1

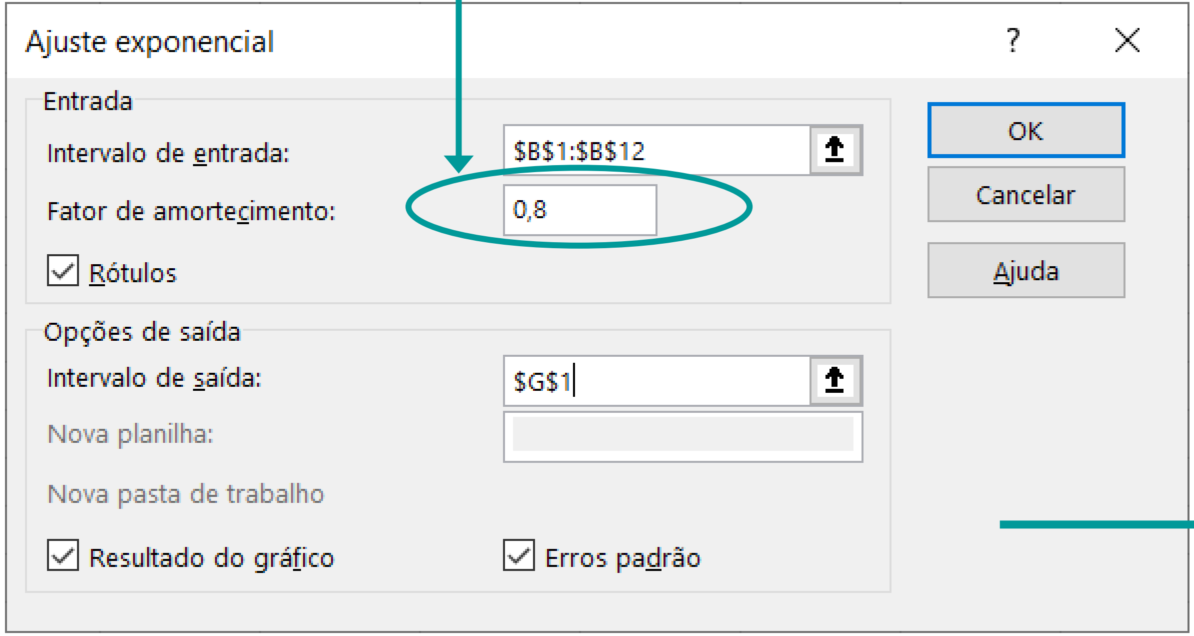
O valor suavizado no período atual (n) é usado como o valor de previsão para o próximo período (n + 1) :

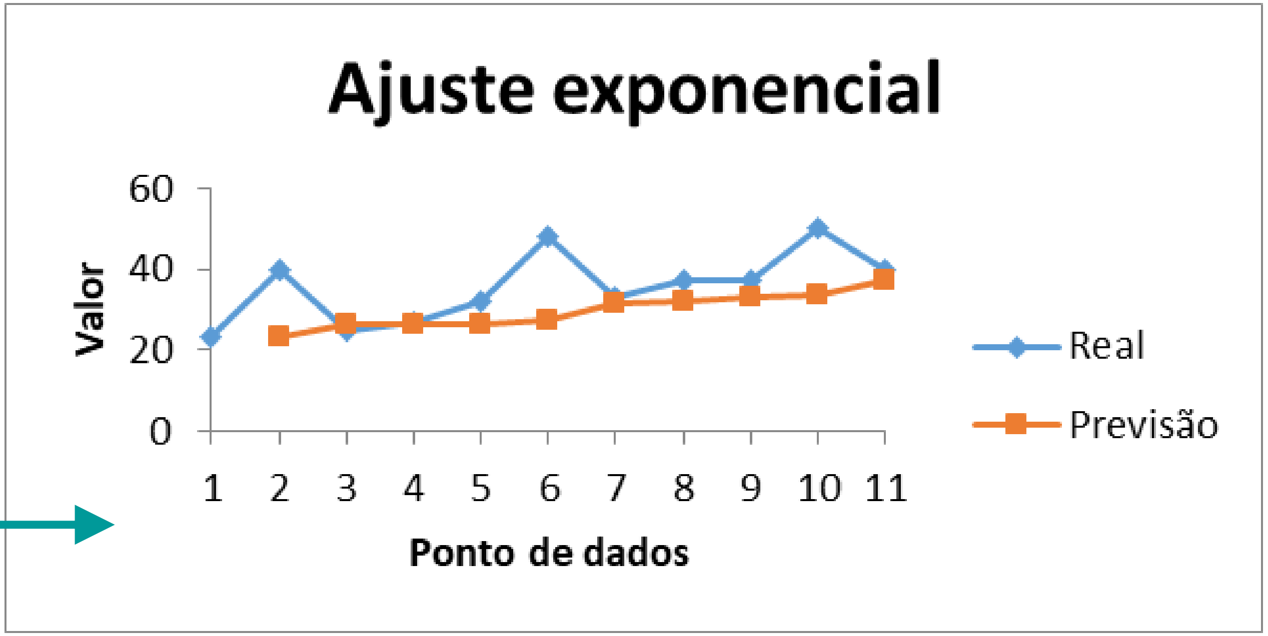
$$\boxed{\hat{Y}_{n+1} = E_n}$$

No Excel:

- Dados → Análise → Análise de dados → Ajuste exponencial

O "fator de amortecimento" é (1 - W)



# Existem três métodos mais populares para a previsão baseados em tendência

- Previsão de tendência linear
- Previsão de tendência não-linear
- Previsão tendência exponencial

## Estimar uma linha de tendência usando análise de regressão

| Ano | Período (X) | Vendas (Y) |
|---|---|---|
| 2004 | 0 | 20 |
| 2005 | 1 | 40 |
| 2006 | 2 | 30 |
| 2007 | 3 | 50 |
| 2008 | 4 | 70 |
| 2009 | 5 | 65 |

- Use **tempo (X)** como uma variável independente:

$$\hat{Y} = b_0 + b_1 X$$

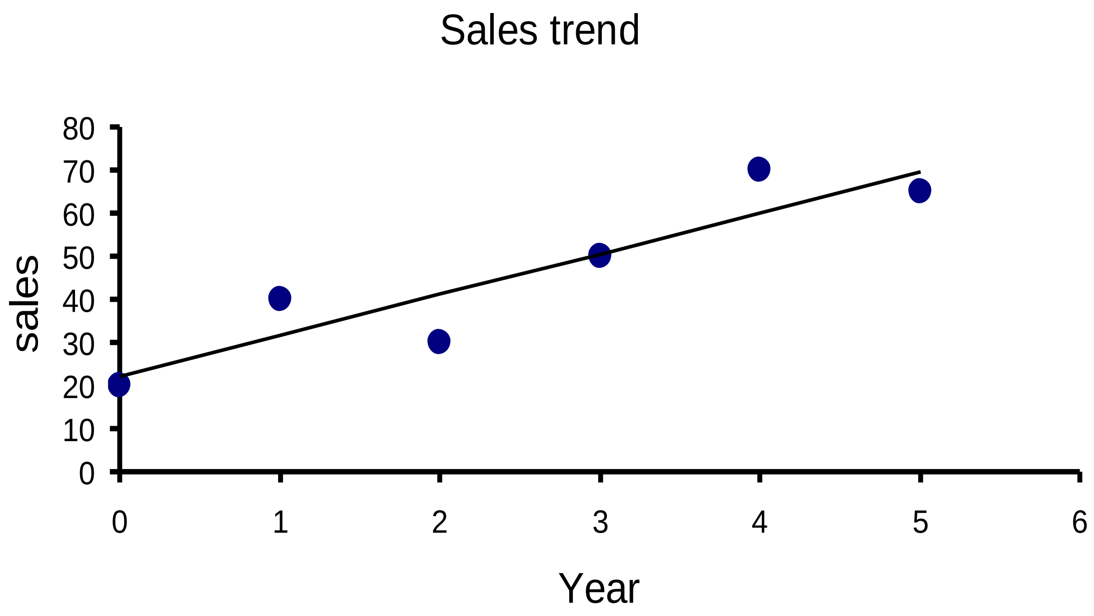
No quadrado linear, não linear, e modelagem exponencial, os períodos de tempo são numeradas começando com 0 e aumentando de 1 para cada período de tempo.

*(continuação)*

A equação de previsão de tendência linear é :

$$\hat{Y}_i = 21.905 + 9.5714\,X_i$$

| Ano | Período (X) | Vendas (Y) |
|---|---|---|
| 2004 | 0 | 20 |
| 2005 | 1 | 40 |
| 2006 | 2 | 30 |
| 2007 | 3 | 50 |
| 2008 | 4 | 70 |
| 2009 | 5 | 65 |



# Previsão de Tendência Linear

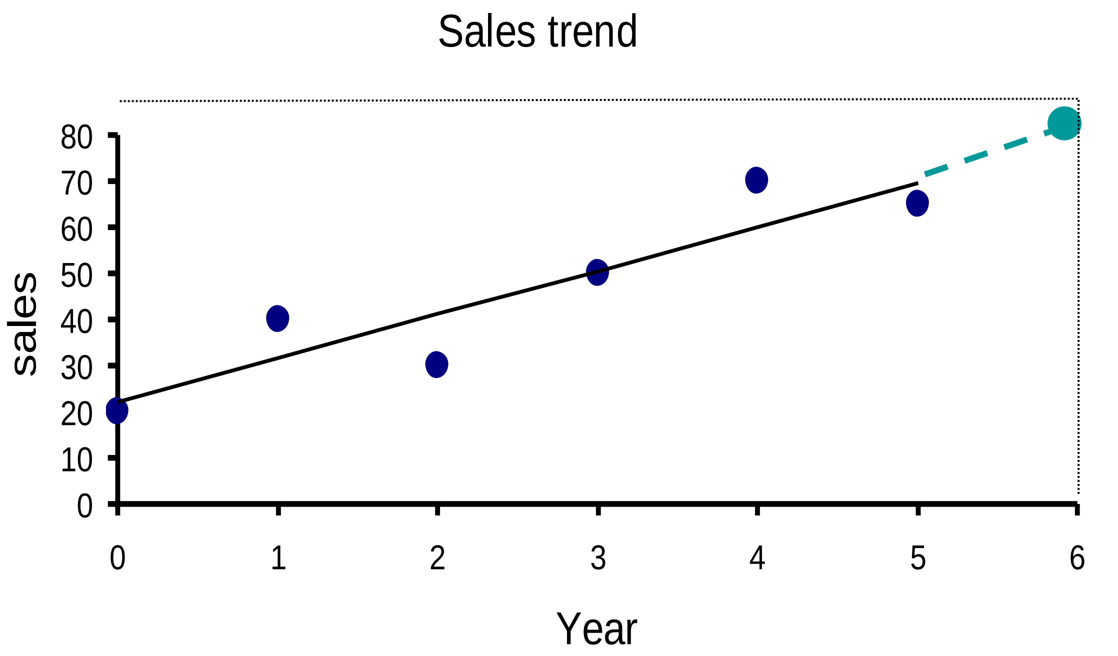
*(continuação)*

- Previsão para o período de tempo 6 (2010):

$$\hat{Y} = 21.905 + 9.5714\,(6)$$

$$\boxed{= 79.33}$$

| Ano | Período (X) | Vendas (Y) |
|---|---|---|
| 2004 | 0 | 20 |
| 2005 | 1 | 40 |
| 2006 | 2 | 30 |
| 2007 | 3 | 50 |
| 2008 | 4 | 70 |
| 2009 | 5 | 65 |



# Previsão de tendência não-linear

- Um modelo de regressão não-linear pode ser usado quando a série temporal apresenta uma tendência não linear
- **Fórmula Quadratica** é um tipo de um modelo não-linear:

$$\boxed{Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \varepsilon_i}$$

- Compare o ajuste $R^2$ e o erro padrão ao do modelo linear para ver se esta é uma melhoria
- Pode tentar outras formas funcionais para obter melhor ajuste

- Outro modelo de tendência não linear:

$$\boxed{Y_i = \beta_0 \beta_1^{X_i} \varepsilon_i}$$

- Transformar a forma linear:

$$\boxed{\log(Y_i) = \log(\beta_0) + X_i \log(\beta_1) + \log(\varepsilon_i)}$$

*(continuação)*

- Equação do Modelo do Tendência Exponencial:

$$\boxed{\log(\hat{Y}_i) = b_0 + b_1 X_i}$$

Onde: $b_0$ = estimativa do $\log(\beta_0)$

$b_1$ = estimativa do $\log(\beta_1)$

Interpretação:

$(b_1 - 1)\times 100\%$ é a taxa composta de crescimento no período (em %)

## Seleção do Modelo de Tendência usando Diferenças

- Usar um modelo de tendência linear se as primeiras diferenças são aproximadamente constante

$$(Y_2 - Y_1) = (Y_3 - Y_2) = \cdots = (Y_n - Y_{n-1})$$

- Use um modelo de tendência quadrática se as segundas diferenças são aproximadamente constante

$$[(Y_3 - Y_2) - (Y_2 - Y_1)] = [(Y_4 - Y_3) - (Y_3 - Y_2)] = \cdots = [(Y_n - Y_{n-1}) - (Y_{n-1} - Y_{n-2})]$$

- Use um modelo de tendência exponencial se as diferenças percentuais são aproximadamente constantes

$$\frac{(Y_2 - Y_1)}{Y_1} \times 100\% = \frac{(Y_3 - Y_2)}{Y_2} \times 100\% = \cdots = \frac{(Y_n - Y_{n-1})}{Y_{n-1}} \times 100\%$$

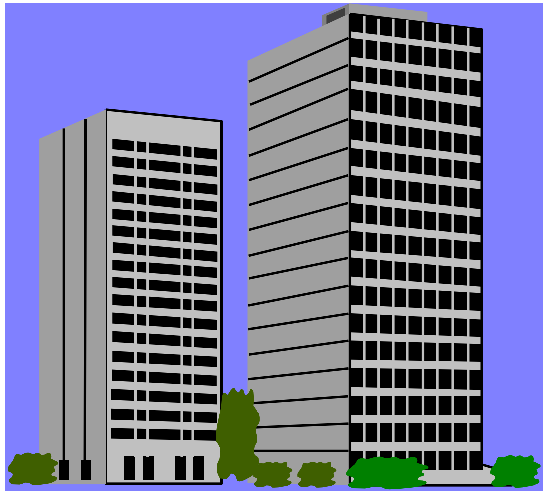
- Usado para a previsão
- Tira proveito de autocorrelação
  - 1ª ordem - correlação entre os valores consecutivos
  - 2ª ordem - correlação entre os valores de dois períodos separados
- **$p^{th}$ order** Modelo Autoregressivo:

$$\boxed{Y_i = A_0 + A_1 Y_{i-1} + A_2 Y_{i-2} + \cdots + A_p Y_{i-p} + \delta_i}$$

Erro aleatório

O Office Concept Corp. adquiriu um número de unidades de escritório (em milhares de pés quadrados) ao longo dos últimos oito anos. Desenvolver a segunda ordem do modelo Autoregressivo.

| Ano | Unid |
|---|---|
| 02 | 4 |
| 03 | 3 |
| 04 | 2 |
| 05 | 3 |
| 06 | 2 |
| 07 | 2 |
| 08 | 4 |
| 09 | 6 |



- Desenvolver a tabela de 2ª ordem
- Usar o Excel para estimar um modelo de regressão

| Ano | $Y_i$ | $Y_{i-1}$ | $Y_{i-2}$ |
|---|---|---|---|
| 02 | 4 | -- | -- |
| 03 | 3 | 4 | -- |
| 04 | 2 | 3 | 4 |
| 05 | 3 | 2 | 3 |
| 06 | 2 | 3 | 2 |
| 07 | 2 | 2 | 3 |
| 08 | 4 | 2 | 2 |
| 09 | 6 | 4 | 2 |

**Saída Excel**

| | *Coefficients* |
|---|---|
| Intercept | 3.5 |
| X Variable 1 | 0.8125 |
| X Variable 2 | -0.9375 |

$$\hat{Y}_i = 3.5 + 0.8125 Y_{i-1} - 0.9375 Y_{i-2}$$

Use a equação de segunda ordem para prever o número de unidades para 2010:

$$\hat{Y}_i = 3.5 + 0.8125 Y_{i-1} - 0.9375 Y_{i-2}$$

$$\begin{aligned}\hat{Y}_{2010} &= 3.5 + 0.8125(Y_{2009}) - 0.9375(Y_{2008}) \\ &= 3.5 + 0.8125(6) - 0.9375(4) \\ &= 4.625\end{aligned}$$

1. Escolha p (Observe que df = n – 2p – 1)

2. Forme uma série de variáveis "de previsões defasadas" $Y_{i-1}$ , $Y_{i-2}$ , … ,$Y_{i-p}$

3. Use Excel or R para executar modelo de regressão utilizando todas as defasagens p

4. Teste de significância de $A_p$
   – Se hipótese nula rejeitada, este modelo é selecionado
   – Se a hipótese nula não é rejeitada, diminuir p por 1 e repita

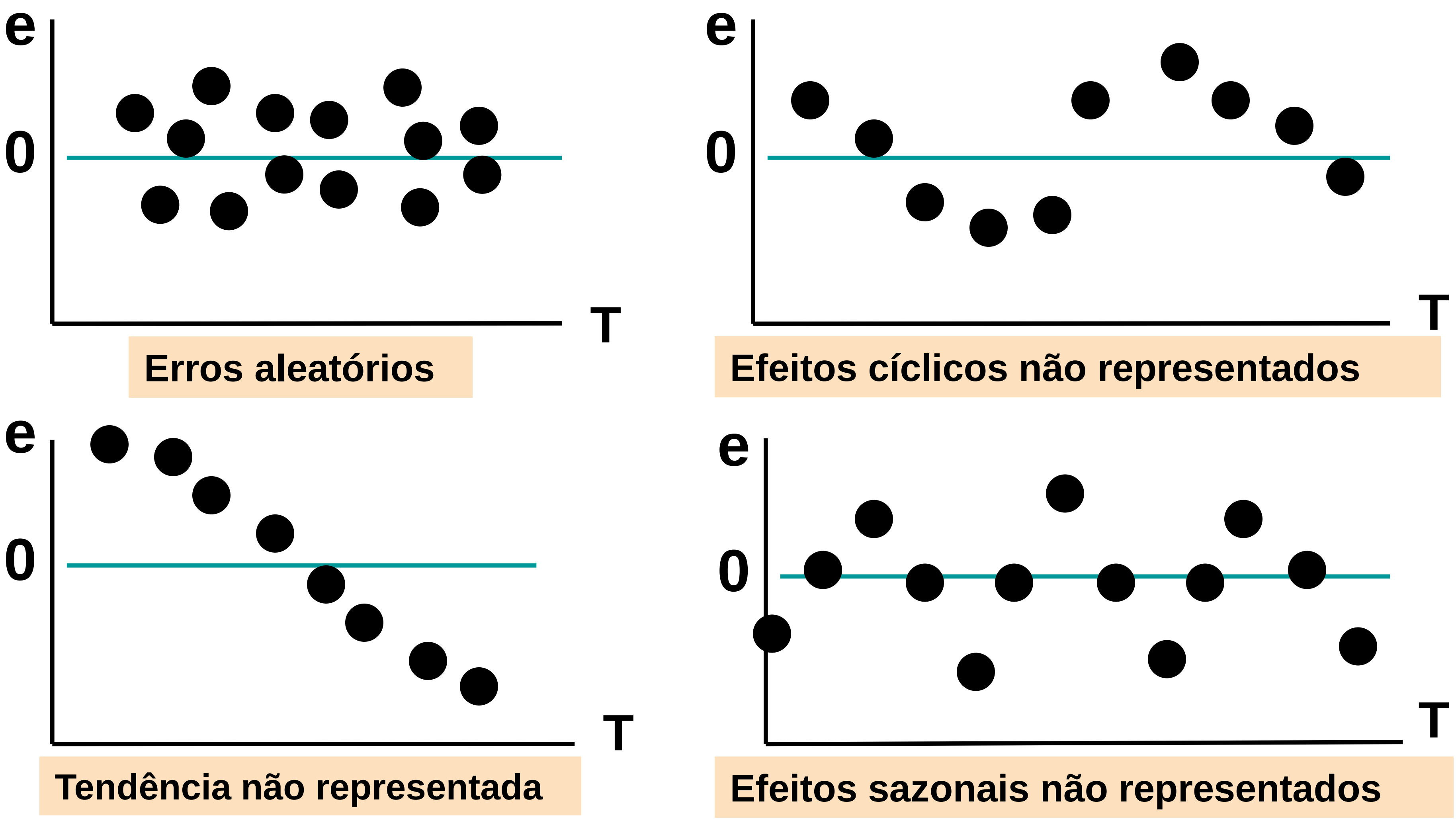
# Escolhendo um Modelo de Previsão

- Realize uma análise de resíduos
  - Elimine um modelo que mostra um padrão ou tendência

- Verifique a magnitude do erro residual usando diferenças ao quadrado e selecione o modelo com o menor valor

- Verifique a magnitude do erro residual usando diferenças absolutas e selecione o modelo com o menor valor

- Use o modelo mais simples
  - Princípio da parcimônia


e
0
T
Erros aleatórios
e
0
T
Efeitos cíclicos não representados
e
0
T
Tendência não representada
e
0
T
Efeitos sazonais não representados

- Escolha o modelo que dá os erros de medição menores

- Soma Quadrática dos Erros

$$\boxed{\mathrm{SQE} = \sum_{i=1}^{n} (\mathrm{Y_i} - \widehat{\mathrm{Y}}_\mathrm{i})^2}$$

  - Sensiveis a discrepâncias

- Desvio Médio Absoluto

$$\boxed{\mathrm{DMA} = \frac{\sum_{i=1}^{n} \left|\mathrm{Y_i} - \widehat{\mathrm{Y}}_\mathrm{i}\right|}{n}}$$

  - Menos sensível às observações extremas

# Princípio da Parcimônia

- Suponha que dois ou mais modelos fornecem um bom ajuste para os dados
- Escolha o modelo mais simples
  - Modelos do tipo mais simples:
    - Mínimos quadrados linear
    - Mínimos quadrados quadrática
    - 1º ordem autoregressiva
  - Tipos mais complexos:
    - 2ª e 3ª ordem autoregressiva
    - Mínimos quadrados exponencial

- Séries temporais são frequentemente coletadas mensalmente ou trimestralmente
- Estas séries de tempo, muitas vezes contêm uma componente de tendência, uma componente sazonal, e uma componente irregular
- Suponha que a sazonalidade é trimestral
  - Definir “três novas variáveis” dummy para trimestres:
    - $T_1$ = 1 se primeiro trimestre, 0 caso contrário
    - $T_2$ = 1 se segundo trimestre, 0 caso contrário
    - $T_3$ = 1 se terceiro trimestre, 0 caso contrário
  - Trimestre 4 é o padrão se $T_1 = T_2 = T_3 = 0$)

- Forma Exponencial

$$Y_i = \beta_0 \beta_1^{X_i} \beta_2^{T_1} \beta_3^{T_2} \beta_4^{T_3} \varepsilon_i$$

$(\beta_1 - 1)$x100% é a taxa composta de crescimento trimestral

$\beta_i$ fornece o multiplicador para o trimestre i-1 em relação ao 4º trimestre (i = 2, 3, 4)

- Transformar em forma linear :

$$\log(Y_i) = \log(\beta_0) + X_i \log(\beta_1) + T_1 \log(\beta_2) + T_2 \log(\beta_3) + T_3 \log(\beta_4) + \log(\varepsilon_i)$$

- Equação de previsão exponencial com sazonalidade trimestral:

$$\log(\hat{Y}_i) = b_0 + b_1 X_i + b_2 T_1 + b_3 T_2 + b_4 T_3$$

Onde: $b_0$ = estimativa do $\log(\beta_0)$, então: $10^{b_0} = \hat{\beta}_0$

$b_1$ = estimativa do $\log(\beta_1)$, então: $10^{b_1} = \hat{\beta}_1$

etc...

**Interpretação**:

$(\hat{\beta}_1 - 1) \times 100\%$ = taxa de crescimento estimada composto trimestral (em%)

$\hat{\beta}_2$ = Multiplicador estimado para o primeiro trimestre em relação ao quarto trimestre

$\hat{\beta}_3$ = Multiplicador estimado para o segundo trimestre em relação ao quarto trimestre

$\hat{\beta}_4$ = Multiplicador estimado para o terceiro trimestre em relação ao quarto trimestre

- Suponha que a equação de previsão é :

$$\log(\hat{Y}_i) = 3{,}43 + 0{,}017X_i - 0{,}082T_1 - 0{,}073T_2 + 0{,}022T_3$$

$b_0 = 3{,}43$, assim $10^{b_0} = \hat{\beta}_0 = 2691.53$

$b_1 = 0{,}017$, assim $10^{b_1} = \hat{\beta}_1 = 1.040$

$b_2 = -0{,}082$, assim $10^{b_2} = \hat{\beta}_2 = 0.827$

$b_3 = -0{,}073$, assim $10^{b_3} = \hat{\beta}_3 = 0.845$

$b_4 = 0{,}022$, assim $10^{b_4} = \hat{\beta}_4 = 1.052$

# Exemplo de Modelo Trimestral

*(continuação)*

| Valor: | Interpretação: |
|---|---|
| $\hat{\beta}_0 = 2691.53$ | valor de tendência não ajustado para o primeiro trimestre de primeiro ano |
| $\hat{\beta}_1 = 1.040$ | $Beta_1$ -1 = 0,04: taxa composta de crescimento trimestral de 4,0% |
| $\hat{\beta}_2 = 0.827$ | média de vendas em T1 são 82,7% das vendas médias do 4º trimestre, após o ajuste para a taxa de crescimento trimestral de 4% |
| $\hat{\beta}_3 = 0.845$ | média de vendas em T2 são 84,5% das vendas médias do 4ºtrimestre, após o ajuste para a taxa de crescimento trimestral de 4% |
| $\hat{\beta}_4 = 1.052$ | média de vendas em T3 são 105,2% das vendas médias do 4ºtrimestre, após o ajuste para a taxa de crescimento trimestral de 4% |

- **Assumir** que o mecanismo que governa o comportamento de séries temporais no passado ainda irá governar no futuro.

- Usar **extrapolação automática** da tendência para prever o futuro sem considerar julgamentos pessoais, experiências de negócios, mudanças tecnológicas e hábitos, etc.

## Resumo do capítulo

- Discussão da importância da previsão
- Abordagem dos fatores componentes do modelo de séries temporais
- Realização de suavização de séries de dados
  - Médias móveis
  - Suavização exponencial
- Descrição dos mínimos quadrados e suas tendências de montagem e previsão
  - Linear, quadrática e modelos exponenciais
- Abordagem de modelos autorregressivos
- Descrição de procedimentos para escolha de modelos adequados
- Abordagem previsão de séries temporais de dados mensais ou trimestrais (uso de variáveis dummy)
- Discussão de armadilhas em matéria de análise de séries temporais

## Bibliografia

**BERENSON, M.L.; LEVINE, D.M.; STEPHAN, D.F.; KREHBIEL, T.C.;** *Statistics for Managers Using Microsoft Excel*, **6ª ed. Pearson Education: Prentice Hall, 2011,**

**LEVINE, David M.; STEPHAN, David F.; KREHBIEL, Thimothy C.;  BERENSON, Mark L.** *Estatística: Teoria e aplicações usando Microsoft® Excel em português*, **6ª ed. Rio de Janeiro: LTC, 2012.**