CE310 - Modelos de Regressão Linear

Regressão polinomial

Cesar Augusto Taconeli

09 de junho, 2025

Introdução

Introdução

- Principais métodos para a modelagem de relação não lineares entre variáveis:

Introdução

- Principais métodos para a modelagem de relação não lineares entre variáveis:
    - Transformação nas variáveis;

Introdução

- Principais métodos para a modelagem de relação não lineares entre variáveis:
    - Transformação nas variáveis;
    - Regressão polinomial;

Introdução

- Principais métodos para a modelagem de relação não lineares entre variáveis:
    - Transformação nas variáveis;
    - Regressão polinomial;
    - Regressão segmentada;

Introdução

- Principais métodos para a modelagem de relação não lineares entre variáveis:
    - Transformação nas variáveis;
    - Regressão polinomial;
    - Regressão segmentada;
    - Regressão não paramétrica (ou semi paramétrica);

Introdução

- Principais métodos para a modelagem de relação não lineares entre variáveis:
    - Transformação nas variáveis;
    - Regressão polinomial;
    - Regressão segmentada;
    - Regressão não paramétrica (ou semi paramétrica);
    - Regressão não linear.

Introdução

- Principais métodos para a modelagem de relação não lineares entre variáveis:
    - Transformação nas variáveis;
    - Regressão polinomial;
    - Regressão segmentada;
    - Regressão não paramétrica (ou semi paramétrica);
    - Regressão não linear.
- Neste módulo vamos estudar as estratégias para modelagem de relações não lineares no campo da regressão linear, considerando transformação nas variáveis, regressão polinomial e regressão segmentada.

Regressão polinomial

Regressão polinomial

- Num modelo de regressão polinomial, uma mesma variável quantitativa é inserida por meio de dois ou mais termos (potências) no modelo.

Regressão polinomial

- Num modelo de regressão polinomial, uma mesma variável quantitativa é inserida por meio de dois ou mais termos (potências) no modelo.

- Como exemplo, uma variável x pode ser inserida na forma linear (x), quadrática (x^2), cúbica (x^3), ou ainda por meio de termos multiplicativos envolvendo outras variáveis (xy, xz).

Regressão polinomial em uma variável

- O modelo de regressão polinomial de ordem k em uma variável fica especificado por:

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + ... + \beta_k x_i^k + \epsilon_i.$$

- O modelo de regressão polinomial de ordem k em uma variável fica especificado por:

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + ... + \beta_k x_i^k + \epsilon_i.$$

- Particularmente, o modelo polinomial de ordem 2 (quadrático) é definido por:

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \epsilon_i.$$

Regressão polinomial em uma variável

- O modelo de regressão polinomial de ordem k em uma variável fica especificado por:

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + ... + \beta_k x_i^k + \epsilon_i.$$

- Particularmente, o modelo polinomial de ordem 2 (quadrático) é definido por:

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \epsilon_i.$$

- Assumindo que os erros têm média zero, a média de y, condicional a x, fica definida por uma função quadrática de x:

$$E(y|x) = \beta_0 + \beta_1 x + \beta_2 x^2.$$

Regressão polinomial em uma variável

- A aplicação de modelo de regressão polinomial pode ser motivada, por exemplo:

Regressão polinomial em uma variável

- A aplicação de modelo de regressão polinomial pode ser motivada, por exemplo:

1. O analista pode saber, com base na teoria da área, que um modelo polinomial com uma particular ordem descreve a relação entre as variáveis;

Regressão polinomial em uma variável

- A aplicação de modelo de regressão polinomial pode ser motivada, por exemplo:

1. O analista pode saber, com base na teoria da área, que um modelo polinomial com uma particular ordem descreve a relação entre as variáveis;

2. A relação entre as variáveis pode apresentar um ponto ótimo (mínimo ou máximo), o que pode motivar a investigação de um modelo quadrático;

Regressão polinomial em uma variável

- A aplicação de modelo de regressão polinomial pode ser motivada, por exemplo:

1. O analista pode saber, com base na teoria da área, que um modelo polinomial com uma particular ordem descreve a relação entre as variáveis;

2. A relação entre as variáveis pode apresentar um ponto ótimo (mínimo ou máximo), o que pode motivar a investigação de um modelo quadrático;

3. Em boa parte dos casos, tem-se uma relação desconhecida e não linear entre as variáveis, de maneira que polinômios de diferentes ordens podem ser testados buscando um ajuste satisfatório.

Regressão polinomial em uma variável - ordem do modelo

- Caso não se disponha, de antemão, de um particular polinômio a ser aplicado, a ordem do polinômio deve ser escolhida de maneira criteriosa.

Regressão polinomial em uma variável - ordem do modelo

- Caso não se disponha, de antemão, de um particular polinômio a ser aplicado, a ordem do polinômio deve ser escolhida de maneira criteriosa.

- Quanto maior a ordem do polinômio especificado, melhor o ajuste aos dados, mas menor a precisão das estimativas (*tradeoff bias-variance*).

Regressão polinomial em uma variável - ordem do modelo

- Caso não se disponha, de antemão, de um particular polinômio a ser aplicado, a ordem do polinômio deve ser escolhida de maneira criteriosa.

- Quanto maior a ordem do polinômio especificado, melhor o ajuste aos dados, mas menor a precisão das estimativas (*tradeoff bias-variance*).

- A menos que se tenha uma justificativa adequada, polinômios de ordem superior a 3 devem ser evitados.

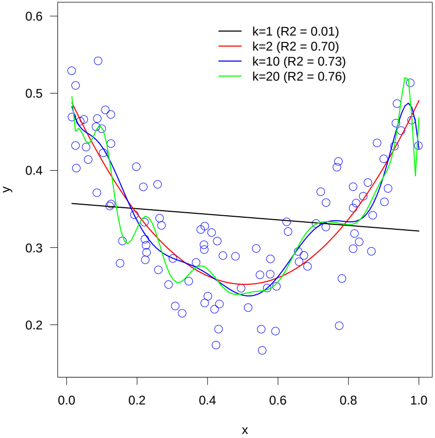


Figura 1: Ajustes de modelos polinomiais de diferentes ordens.

Regressão polinomial em uma variável - ordem do modelo

- A escolha da ordem do polinômio pode ser feita usando estratégias do tipo **forward** ou **backward**.

Regressão polinomial em uma variável - ordem do modelo

- A escolha da ordem do polinômio pode ser feita usando estratégias do tipo **forward** ou **backward**.

- Na estratégia **forward** deve-se começar pelo ajuste sem a variável investigada e incluir, a cada passo, o termo linear (x), depois o quadrático (x^2), o cúbico (x^3)...

Regressão polinomial em uma variável - ordem do modelo

- A escolha da ordem do polinômio pode ser feita usando estratégias do tipo **forward** ou **backward**.

- Na estratégia **forward** deve-se começar pelo ajuste sem a variável investigada e incluir, a cada passo, o termo linear (x), depois o quadrático (x^2), o cúbico (x^3)...

- O processo se encerra quando a adição de um termo de ordem $t + 1$ não for significativa. O modelo escolhido é o de ordem t.

Regressão polinomial em uma variável - ordem do modelo

- Na estratégia **backward** deve-se começar pelo ajuste do modelo polinomial de maior ordem a ser investigado (digamos k) e, a cada passo, excluir o termo de maior ordem (primeiro x^k, depois x^{k-1},...), desde que o termo não tenha efeito significativo.

Regressão polinomial em uma variável - ordem do modelo

- Na estratégia **backward** deve-se começar pelo ajuste do modelo polinomial de maior ordem a ser investigado (digamos k) e, a cada passo, excluir o termo de maior ordem (primeiro x^k, depois x^{k-1},...), desde que o termo não tenha efeito significativo.

- O processo se encerra quando o termo de maior ordem (digamos t) tiver efeito significativo. O modelo escolhido é o de ordem t.

Regressão polinomial - extrapolação

- Extrapolar (ou seja, predizer y fora do intervalo observado para x) em geral não é apropriado para modelos polinomiais, podendo produzir resultados incorretos e inconsistentes.

Regressão polinomial - extrapolação

- Extrapolar (ou seja, predizer y fora do intervalo observado para x) em geral não é apropriado para modelos polinomiais, podendo produzir resultados incorretos e inconsistentes.

- Uma situação em que a extrapolação pode ser aplicada com maior segurança é quando o modelo de regressão baseia-se em algum modelo teórico da área que descreva adequadamente o fenômeno.

Regressão polinomial - extrapolação

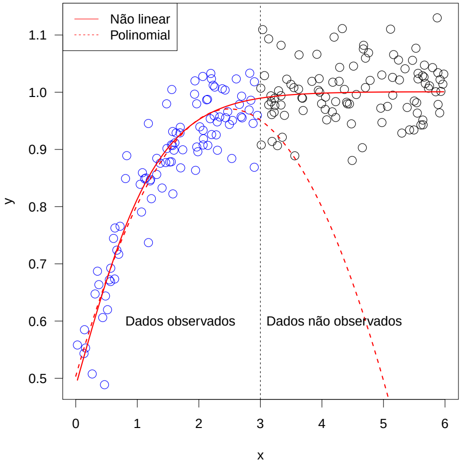


Figura 2: Modelos polinomiais - o problema da extrapolação

Regressão polinomial - multicolinearidade

- Em regressão polinomial, a matriz do modelo é composta por colunas que são potências de uma mesma variável.

Regressão polinomial - multicolinearidade

- Em regressão polinomial, a matriz do modelo é composta por colunas que são potências de uma mesma variável.

- Essa construção implica no mau condicionamento da matriz, ou seja, na correlação entre as colunas da matriz.

Regressão polinomial - multicolinearidade

- Em regressão polinomial, a matriz do modelo é composta por colunas que são potências de uma mesma variável.

- Essa construção implica no mau condicionamento da matriz, ou seja, na correlação entre as colunas da matriz.

- Uma medida corretiva consiste em usar os dados centrados (ou seja, substituir x por $x^* = x - \bar{x}$).

Regressão polinomial - multicolinearidade

- Em regressão polinomial, a matriz do modelo é composta por colunas que são potências de uma mesma variável.
- Essa construção implica no mau condicionamento da matriz, ou seja, na correlação entre as colunas da matriz.
- Uma medida corretiva consiste em usar os dados centrados (ou seja, substituir x por $x^* = x - \bar{x}$).
- Uma alternativa mais eficiente é usar **polinômios ortogonais**.

Exemplo- Valores de imóveis em Boston

Exemplo- Valores de imóveis em Boston

- Nesta aplicação vamos considerar os dados disponíveis na base de dados `Boston` da biblioteca `MASS` do `R`.

Exemplo- Valores de imóveis em Boston

- Nesta aplicação vamos considerar os dados disponíveis na base de dados `Boston` da biblioteca `MASS` do `R`.

- Os dados referem-se aos valores dos imóveis residenciais em 506 subúrvios de Boston.

Exemplo- Valores de imóveis em Boston

- Nesta aplicação vamos considerar os dados disponíveis na base de dados `Boston` da biblioteca `MASS` do `R`.

- Os dados referem-se aos valores dos imóveis residenciais em 506 subúrvios de Boston.

- As variáveis que vamos analisar são as seguintes:

Exemplo- Valores de imóveis em Boston

- Nesta aplicação vamos considerar os dados disponíveis na base de dados `Boston` da biblioteca `MASS` do `R`.

- Os dados referem-se aos valores dos imóveis residenciais em 506 subúrvios de Boston.

- As variáveis que vamos analisar são as seguintes:
  - `medv`: preço mediano dos imóveis ocupados no subúrbio (em $1000s);

Exemplo- Valores de imóveis em Boston

- Nesta aplicação vamos considerar os dados disponíveis na base de dados `Boston` da biblioteca `MASS` do `R`.

- Os dados referem-se aos valores dos imóveis residenciais em 506 subúrvios de Boston.

- As variáveis que vamos analisar são as seguintes:
  - `medv`: preço mediano dos imóveis ocupados no subúrbio (em $1000s);
  - `lstat`: percentual da população do subúrbio na menor faixa de renda.

Exemplo- Valores de imóveis em Boston

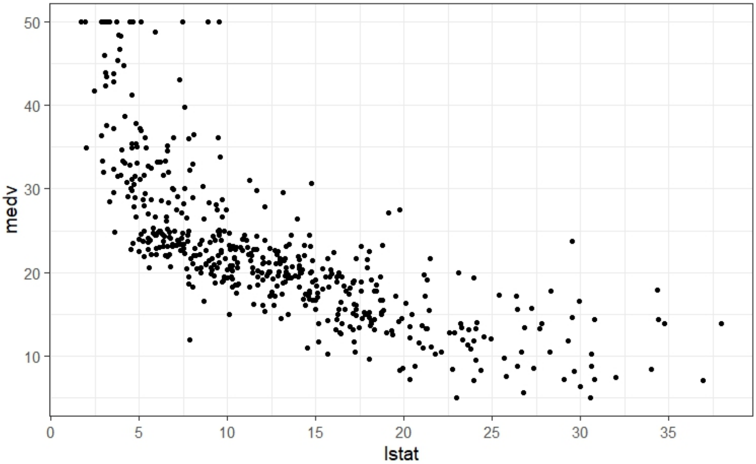


Figura 3: Valores de imóveis e população de menor renda

Exemplo- Valores de imóveis em Boston

- Vamos ajustar modelos polinomiais de primeira a sexta ordem, conforme descrito na sequência:

Exemplo- Valores de imóveis em Boston

- Vamos ajustar modelos polinomiais de primeira a sexta ordem, conforme descrito na sequência:

1. Modelo de regressão polinomial de primeira ordem:

$$\texttt{medv} = \beta_0 + \beta_1 \texttt{lstat} + \epsilon$$

Exemplo- Valores de imóveis em Boston

- Vamos ajustar modelos polinomiais de primeira a sexta ordem, conforme descrito na sequência:

1. Modelo de regressão polinomial de primeira ordem:

$$\texttt{medv} = \beta_0 + \beta_1 \texttt{lstat} + \epsilon$$

2. Modelo de regressão polinomial de segunda ordem:

$$\texttt{medv} = \beta_0 + \beta_1 \texttt{lstat} + \beta_2 \texttt{lstat}^2 + \epsilon$$

Exemplo- Valores de imóveis em Boston

- Vamos ajustar modelos polinomiais de primeira a sexta ordem, conforme descrito na sequência:

1. Modelo de regressão polinomial de primeira ordem:

$$\texttt{medv} = \beta_0 + \beta_1 \texttt{lstat} + \epsilon$$

2. Modelo de regressão polinomial de segunda ordem:

$$\texttt{medv} = \beta_0 + \beta_1 \texttt{lstat} + \beta_2 \texttt{lstat}^2 + \epsilon$$

3. Modelo de regressão polinomial de terceira ordem:

$$\texttt{medv} = \beta_0 + \beta_1 \texttt{lstat} + \beta_2 \texttt{lstat}^2 + \beta_3 \texttt{lstat}^3 + \epsilon$$

Exemplo- Valores de imóveis em Boston

4. Modelo de regressão polinomial de quarta ordem:

$$\texttt{medv} = \beta_0 + \beta_1\texttt{lstat} + \beta_2\texttt{lstat}^2 + \beta_3\texttt{lstat}^3 + \beta_4\texttt{lstat}^4 + \epsilon$$

Exemplo- Valores de imóveis em Boston

4. Modelo de regressão polinomial de quarta ordem:

$$\texttt{medv} = \beta_0 + \beta_1\texttt{lstat} + \beta_2\texttt{lstat}^2 + \beta_3\texttt{lstat}^3 + \beta_4\texttt{lstat}^4 + \epsilon$$

5. Modelo de regressão polinomial de quinta ordem:

$$\texttt{medv} = \beta_0 + \beta_1\texttt{lstat} + \beta_2\texttt{lstat}^2 + \beta_3\texttt{lstat}^3 + \beta_4\texttt{lstat}^4 + \beta_5\texttt{lstat}^5 + \epsilon$$

Exemplo- Valores de imóveis em Boston

4. Modelo de regressão polinomial de quarta ordem:

$$\texttt{medv} = \beta_0 + \beta_1\texttt{lstat} + \beta_2\texttt{lstat}^2 + \beta_3\texttt{lstat}^3 + \beta_4\texttt{lstat}^4 + \epsilon$$

5. Modelo de regressão polinomial de quinta ordem:

$$\texttt{medv} = \beta_0 + \beta_1\texttt{lstat} + \beta_2\texttt{lstat}^2 + \beta_3\texttt{lstat}^3 + \beta_4\texttt{lstat}^4 + \beta_5\texttt{lstat}^5 + \epsilon$$

6. Modelo de regressão polinomial de sexta ordem:

$$\texttt{medv} = \beta_0 + \beta_1\texttt{lstat} + \beta_2\texttt{lstat}^2 + \beta_3\texttt{lstat}^3 + \beta_4\texttt{lstat}^4 + \beta_5\texttt{lstat}^5 + \beta_6\texttt{lstat}^6 + \epsilon$$

Exemplo- Valores de imóveis em Boston

4. Modelo de regressão polinomial de quarta ordem:

$$\texttt{medv} = \beta_0 + \beta_1 \texttt{lstat} + \beta_2 \texttt{lstat}^2 + \beta_3 \texttt{lstat}^3 + \beta_4 \texttt{lstat}^4 + \epsilon$$

5. Modelo de regressão polinomial de quinta ordem:

$$\texttt{medv} = \beta_0 + \beta_1 \texttt{lstat} + \beta_2 \texttt{lstat}^2 + \beta_3 \texttt{lstat}^3 + \beta_4 \texttt{lstat}^4 + \beta_5 \texttt{lstat}^5 + \epsilon$$

6. Modelo de regressão polinomial de sexta ordem:

$$\texttt{medv} = \beta_0 + \beta_1 \texttt{lstat} + \beta_2 \texttt{lstat}^2 + \beta_3 \texttt{lstat}^3 + \beta_4 \texttt{lstat}^4 + \beta_5 \texttt{lstat}^5 + \beta_6 \texttt{lstat}^6 + \epsilon$$

- Em todos os casos os erros são independentes e identicamente distribuídos $\epsilon \sim \mathrm{N}(0, \sigma^2)$.

Exemplo- Valores de imóveis em Boston

- Os modelos ajustados são os seguintes:

Exemplo- Valores de imóveis em Boston

- Os modelos ajustados são os seguintes:

1. Modelo de primeira ordem:

$$\widehat{\texttt{medv}} = 34.55 - 0.95 \times \texttt{lstat}$$

Exemplo- Valores de imóveis em Boston

- Os modelos ajustados são os seguintes:

1. Modelo de primeira ordem:

$$\widehat{\texttt{medv}} = 34.55 - 0.95 \times \texttt{lstat}$$

2. Modelo de segunda ordem:

$$\widehat{\texttt{medv}} = 42.86 - 2.33 \times \texttt{lstat} + 0.043 \times \texttt{lstat}^2$$

Exemplo- Valores de imóveis em Boston

- Os modelos ajustados são os seguintes:

1. Modelo de primeira ordem:

$$\widehat{\texttt{medv}} = 34.55 - 0.95 \times \texttt{lstat}$$

2. Modelo de segunda ordem:

$$\widehat{\texttt{medv}} = 42.86 - 2.33 \times \texttt{lstat} + 0.043 \times \texttt{lstat}^2$$

3. Modelo de terceira ordem:

$$\widehat{\texttt{medv}} = 48.64 - 3.87 \times \texttt{lstat} + 0.15 \times \texttt{lstat}^2 - 0.002 \times \texttt{lstat}^3$$

Exemplo- Valores de imóveis em Boston

4. Modelo de quarta ordem:

$$\widehat{\texttt{medv}} = 57.30 - 7.03 \times \texttt{lstat} + 0.50 \times \texttt{lstat}^2 - 0.016 \times \texttt{lstat}^3 + 0.0002 \times \texttt{lstat}^4$$

Exemplo- Valores de imóveis em Boston

4. Modelo de quarta ordem:

$$\widehat{\texttt{medv}} = 57.30 - 7.03 \times \texttt{lstat} + 0.50 \times \texttt{lstat}^2 - 0.016 \times \texttt{lstat}^3 + 0.0002 \times \texttt{lstat}^4$$

5. Modelo de quinta ordem:

$$\widehat{\texttt{medv}} = 67.70 - 11.99 \times \texttt{lstat} + 1.27 \times \texttt{lstat}^2 - 0.068 \times \texttt{lstat}^3 + 0.0017 \times \texttt{lstat}^4 - 0.000016 \times \texttt{lstat}^5$$

Exemplo- Valores de imóveis em Boston

4. Modelo de quarta ordem:

$$\widehat{\texttt{medv}} = 57.30 - 7.03 \times \texttt{lstat} + 0.50 \times \texttt{lstat}^2 - 0.016 \times \texttt{lstat}^3 + 0.0002 \times \texttt{lstat}^4$$

5. Modelo de quinta ordem:

$$\widehat{\texttt{medv}} = 67.70 - 11.99 \times \texttt{lstat} + 1.27 \times \texttt{lstat}^2 - 0.068 \times \texttt{lstat}^3 + 0.0017 \times \texttt{lstat}^4 - 0.000016 \times \texttt{lstat}^5$$

6. Modelo de sexta ordem ordem:

$$\widehat{\texttt{medv}} = 73.04 - 15.17 \times \texttt{lstat} + 1.93 \times \texttt{lstat}^2 - 0.131 \times \texttt{lstat}^3 + 0.0046 \times \texttt{lstat}^4 - 0.000084 \times \texttt{lstat}^5 + 0.00000059 \times \texttt{lstat}$$

Exemplo- Valores de imóveis em Boston

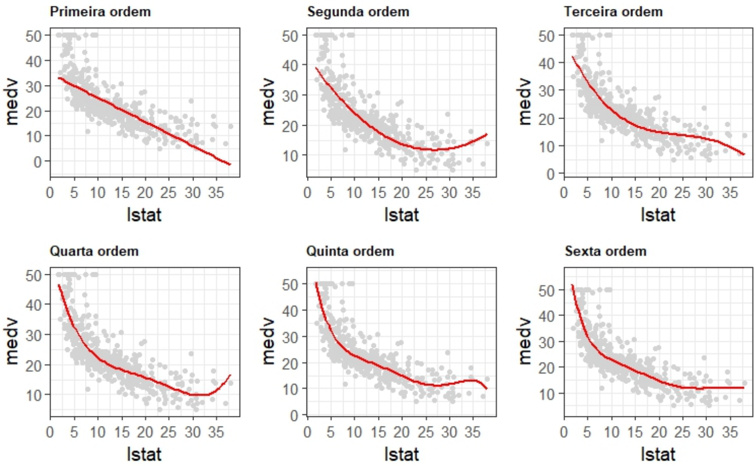


Figura 4: Modelos de regressão polinomial de diferentes ordens ajustados aos dados de valores de imóveis em Boston

Exemplo- Valores de imóveis em Boston

Tabela 1: Tabela de análise de variância para os modelos de regressão polinomial ajustados aos dados de valores de imóveis em Boston

Order	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1	504	19472.38				
2	503	15347.24	1	4125.14	151.86	0.0000
3	502	14615.48	1	731.76	26.94	0.0000
4	501	13967.69	1	647.79	23.85	0.0000
5	500	13597.04	1	370.66	13.65	0.0002
6	499	13554.67	1	42.36	1.56	0.2123

Exemplo- Valores de imóveis em Boston

- A soma de quadrados de resíduos reduz de 19472.38 para 15347.24 do modelo linear para o quadrático. Essa redução de 4125.14 é significativa ($p<0.0001$), de forma que o modelo quadrático é preferível.

Exemplo- Valores de imóveis em Boston

- A soma de quadrados de resíduos reduz de 19472.38 para 15347.24 do modelo linear para o quadrático. Essa redução de 4125.14 é significativa ($p<0.0001$), de forma que o modelo quadrático é preferível.

- A soma de quadrados de resíduos reduz de 15347.24 para 14615.48 do modelo quadrático para o cúbico. Essa redução de 731.76 é significativa ($p<0.0001$), de forma que o modelo cúbico é preferível.

Exemplo- Valores de imóveis em Boston

- A soma de quadrados de resíduos reduz de 19472.38 para 15347.24 do modelo linear para o quadrático. Essa redução de 4125.14 é significativa ($p<0.0001$), de forma que o modelo quadrático é preferível.

- A soma de quadrados de resíduos reduz de 15347.24 para 14615.48 do modelo quadrático para o cúbico. Essa redução de 731.76 é significativa ($p<0.0001$), de forma que o modelo cúbico é preferível.

- A soma de quadrados de resíduos reduz de 14615.48 para 13967.69 do modelo cúbico para o de quarta ordem. Essa redução de 647.79 é significativa ($p<0.0001$), de forma que o modelo de ordem quatro é preferível.

Exemplo- Valores de imóveis em Boston

- A soma de quadrados de resíduos reduz de 13967.69 para 13597.04 do modelo de quarta para o de quinta ordem. Essa redução de 370.66 é significativa (p=0.0002), de forma que o modelo de ordem cinco é preferível.

Exemplo- Valores de imóveis em Boston

- A soma de quadrados de resíduos reduz de 13967.69 para 13597.04 do modelo de quarta para o de quinta ordem. Essa redução de 370.66 é significativa (p=0.0002), de forma que o modelo de ordem cinco é preferível.

- A soma de quadrados de resíduos reduz de 13597.04 para 13554.67 do modelo de quinta para o de sexta ordem. Essa redução de 42.36 não é significativa (p=0.2123), de forma que o modelo de ordem cinco é preferível.

Exemplo- Valores de imóveis em Boston

- A soma de quadrados de resíduos reduz de 13967.69 para 13597.04 do modelo de quarta para o de quinta ordem. Essa redução de 370.66 é significativa (p=0.0002), de forma que o modelo de ordem cinco é preferível.

- A soma de quadrados de resíduos reduz de 13597.04 para 13554.67 do modelo de quinta para o de sexta ordem. Essa redução de 42.36 não é significativa (p=0.2123), de forma que o modelo de ordem cinco é preferível.

- Desta forma, com base nessa bateria de avaliações de modelos de regressão polinomial de diferentes ordens, o modelo polinomial de ordem cinco deve ser escolhido.

- **Nota:** Embora nesta aplicação um polinômio de ordem elevada (ordem cinco) tenha sido selecionado, vale reforçar o risco de *overfitting* associado a tais polinômios.

Exemplo- Valores de imóveis em Boston

- **Nota:** Embora nesta aplicação um polinômio de ordem elevada (ordem cinco) tenha sido selecionado, vale reforçar o risco de *overfitting* associado a tais polinômios.

- Se a relação não linear entre a resposta e a explicativa não for facilmente descrita por um polinômio mais simples, soluções alternativas como regressão não paramétrica são preferíveis.

Exemplo- Valores de imóveis em Boston

- Para fins de ilustração, vamos calcular alguns valores preditos (ajustados) pelo modelo.

Exemplo- Valores de imóveis em Boston

- Para fins de ilustração, vamos calcular alguns valores preditos (ajustados) pelo modelo.
    - Para `lstat=10`:

$$\widehat{\texttt{medv}} = 67.70 - 11.99 \times 10 + 1.27 \times 10^2 - 0.068 \times 10^3 + 0.0017 \times 10^4 - 0.000016 \times 10^5 = 22.2$$

Exemplo- Valores de imóveis em Boston

- Para fins de ilustração, vamos calcular alguns valores preditos (ajustados) pelo modelo.
    - Para `lstat=10`:

$$\widehat{\texttt{medv}} = 67.70 - 11.99 \times 10 + 1.27 \times 10^2 - 0.068 \times 10^3 + 0.0017 \times 10^4 - 0.000016 \times 10^5 = 22.2$$

* Para `lstat=22`:

$$\widehat{\texttt{medv}} = 67.70 - 11.99 \times 22 + 1.27 \times 22^2 - 0.068 \times 22^3 + 0.0017 \times 22^4 - 0.000016 \times 22^5 = 10.3$$

Polinômios ortogonais

Polinômios ortogonais

- Polinômios ortogonais são uma estratégia eficiente para lidar com o mau condicionamento dos termos em regressão polinomial.

Polinômios ortogonais

- Polinômios ortogonais são uma estratégia eficiente para lidar com o mau condicionamento dos termos em regressão polinomial.

- O modelo de regressão polinomial ortogonal, para uma variável independente x, é reprsentado por:

$$y = \beta_0 P_0(x) + \beta_1 P_1(x) + \beta_2 P_2(x) + \ldots + \beta_k P_k(x) + \epsilon, \quad \epsilon \sim \mathrm{N}(0, \sigma^2),$$

Polinômios ortogonais

- Polinômios ortogonais são uma estratégia eficiente para lidar com o mau condicionamento dos termos em regressão polinomial.
- O modelo de regressão polinomial ortogonal, para uma variável independente x, é reprsentado por:

$$y = \beta_0 P_0(x) + \beta_1 P_1(x) + \beta_2 P_2(x) + ... + \beta_k P_k(x) + \epsilon, \quad \epsilon \sim \mathrm{N}(0, \sigma^2),$$

onde $P_0(x), P_1(x), P_2(x), ..., P_k(x)$ são polinômios ortogonais de ordem $j = 0, 1, 2, ..., k$, ou seja:

$$\sum_{i=1}^{n} P_r(x_i) P_s(x_i) = 0, \quad r \neq s, \quad r, s = 0, 1, 2, ..., k$$

$$P_0(x_i) = 1$$

- Desta forma, a matriz do modelo de regressão linear polinomial ortogonal fica dada por:

$$\boldsymbol{X} = \begin{bmatrix} P_0(x_1) & P_1(x_1) & P_2(x_1) & \dots & P_k(x_1) \\ P_0(x_2) & P_2(x_2) & P_2(x_2) & \dots & P_k(x_2) \\ P_0(x_3) & P_1(x_3) & P_2(x_3) & \dots & P_k(x_3) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ P_0(x_n) & P_1(x_n) & P_2(x_n) & \dots & P_k(x_n) \end{bmatrix}$$

Polinômios ortogonais

- Como a matriz $\boldsymbol{X}$ tem colunas ortogonais, segue que:

$$\boldsymbol{X'X} = \begin{bmatrix} \sum_{i=1}^{n} P_0^2(x_i) & 0 & 0 & \dots & 0 \\ 0 & \sum_{i=1}^{n} P_1^2(x_i) & 0 & \dots & 0 \\ 0 & 0 & \sum_{i=1}^{n} P_2^2(x_i) & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \sum_{i=1}^{n} P_k^2(x_i) \end{bmatrix}$$

Polinômios ortogonais

- Como a matriz $\boldsymbol{X}$ tem colunas ortogonais, segue que:

$$\boldsymbol{X'X} = \begin{bmatrix} \sum_{i=1}^n P_0^2(x_i) & 0 & 0 & \dots & 0 \\ 0 & \sum_{i=1}^n P_1^2(x_i) & 0 & \dots & 0 \\ 0 & 0 & \sum_{i=1}^n P_2^2(x_i) & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \sum_{i=1}^n P_k^2(x_i) \end{bmatrix}$$

- Desta forma, os estimadores de mínimos quadrados para os $\beta's$ ficam dados por:

$$\hat{\beta}_j = \frac{\sum_{i=1}^n P_j(x_i) y_i}{\sum_{i=1}^n P_j^2(x_i)}, \quad j = 0, 1, 2, ..., k$$

Polinômios ortogonais

- No caso onde os níveis (valores amostrados) de x são igualmente espaçados e $P_0(x_i)$ é fixado em 1, $i = 1, 2, ..., n$, os primeiros cinco polinômios ortogonais são:

$$P_0(x_1) = 1$$

$$P_1(x_i) = \lambda_1 \left[\frac{x_i - \bar{x}}{d}\right]$$

$$P_2(x_i) = \lambda_2 \left[\left(\frac{x_i - \bar{x}}{d}\right)^2 - \left(\frac{n^2 - 1}{12}\right)\right]$$

$$P_3(x_i) = \lambda_3 \left[\left(\frac{x_i - \bar{x}}{d}\right)^3 - \left(\frac{x_i - \bar{x}}{d}\right)\left(\frac{3n^2 - 7}{20}\right)\right]$$

Polinômios ortogonais

$$P_4(x_i) = \lambda_4 \left[\left(\frac{x_i - \bar{x}}{d} \right)^4 - \left(\frac{x_i - \bar{x}}{d} \right) \left(\frac{3n^2 - 13}{14} \right) + \frac{3(n^2 - 1)(n^2 - 9)}{560} \right]$$

onde d é a diferença entre os valores sucessivos de x e os λs são constantes para garantir que os polinômios tenham valores inteiros.

Exemplo- Inventário de estoque

Exemplo- Inventário de estoque

- Nesta ilustração, vamos analisar dados de um inventário realizado no estoque de uma empresa.

Exemplo- Inventário de estoque

- Nesta ilustração, vamos analisar dados de um inventário realizado no estoque de uma empresa.

- As variáveis são as seguintes:

Exemplo- Inventário de estoque

- Nesta ilustração, vamos analisar dados de um inventário realizado no estoque de uma empresa.

- As variáveis são as seguintes:
  - `Quantidade`: Número de itens reabastecidos;

Exemplo- Inventário de estoque

- Nesta ilustração, vamos analisar dados de um inventário realizado no estoque de uma empresa.

- As variáveis são as seguintes:
    - `Quantidade`: Número de itens reabastecidos;
    - `Custo`: Custo médio anual de estocagem (resposta).

Exemplo- Inventário de estoque

- Nesta ilustração, vamos analisar dados de um inventário realizado no estoque de uma empresa.

- As variáveis são as seguintes:
  - `Quantidade`: Número de itens reabastecidos;
  - `Custo`: Custo médio anual de estocagem (resposta).

- Os dados apresentados na sequência exibem uma relação claramente quadrática entre o custo médio de estocagem e a quantidade de itens reabastecidos.

Exemplo- Inventário de estoque

Tabela 2: Dados de inventário

Quantidade (x)	Custo (y)
50	335
75	326
100	316
125	313
150	311
175	314
200	318
225	328
250	337
275	345

Exemplo- Inventário de estoque

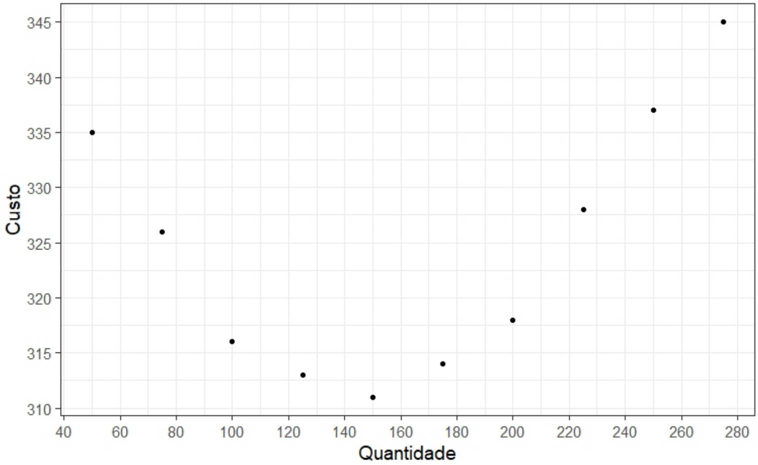


Figura 5: Custo médio de estocagem versus quantidade de itens reabastecidos

Exemplo- Inventário de estoque

- Vamos ajustar um modelo de regressão polinomial de segunda ordem (inicialmente, não ortogonal), definido por:

Exemplo- Inventário de estoque

- Vamos ajustar um modelo de regressão polinomial de segunda ordem (inicialmente, não ortogonal), definido por:

$$\texttt{Custo} = \beta_0 + \beta_1 \texttt{Quant} + \beta_2 \texttt{Quant}^2 + \epsilon, \quad \epsilon \sim \mathrm{N}(0, \sigma^2)$$

Exemplo- Inventário de estoque

- Vamos ajustar um modelo de regressão polinomial de segunda ordem (inicialmente, não ortogonal), definido por:

$$\texttt{Custo} = \beta_0 + \beta_1\texttt{Quant} + \beta_2\texttt{Quant}^2 + \epsilon, \quad \epsilon \sim \mathrm{N}(0, \sigma^2)$$

- Ou, de maneira equivalente:

Exemplo- Inventário de estoque

- Vamos ajustar um modelo de regressão polinomial de segunda ordem (inicialmente, não ortogonal), definido por:

$$\texttt{Custo} = \beta_0 + \beta_1 \texttt{Quant} + \beta_2 \texttt{Quant}^2 + \epsilon, \quad \epsilon \sim \mathrm{N}(0, \sigma^2)$$

- Ou, de maneira equivalente:

$$\texttt{Custo|Quant} \sim \mathrm{N}\left(\mu_{\texttt{Quant}}, \sigma^2\right)$$

$$\mu_{\texttt{Quant}} = \beta_0 + \beta_1 \texttt{Quant} + \beta_2 \texttt{Quant}^2$$

Exemplo- Inventário de estoque

- Matriz do modelo:

Exemplo- Inventário de estoque

- Matriz do modelo:

$$
\boldsymbol{X} = \begin{bmatrix}
1 & 50 & 50^2 = 2500 \\
1 & 75 & 75^2 = 5625 \\
1 & 100 & 100^2 = 10000 \\
1 & 125 & 125^2 = 15625 \\
1 & 150 & 150^2 = 22500 \\
1 & 175 & 175^2 = 30625 \\
1 & 200 & 200^2 = 40000 \\
1 & 225 & 225^2 = 50625 \\
1 & 250 & 250^2 = 62500 \\
1 & 275 & 275^2 = 75625
\end{bmatrix}
$$

Exemplo- Inventário de estoque

- A correlação linear entre `Quantidade` e $\texttt{Quantidade}^2$ é igual a 0.98.

Exemplo- Inventário de estoque

- A correlação linear entre `Quantidade` e `Quantidade`^{2} é igual a 0.98.

- No ajuste do modelo polinomial, o VIF para `Quantidade` e `Quantidade`^{2} é de 27.41.

Exemplo- Inventário de estoque

- A correlação linear entre `Quantidade` e `Quantidade`^{2} é igual a 0.98.

- No ajuste do modelo polinomial, o VIF para `Quantidade` e `Quantidade`^{2} é de 27.41.

- Desta forma, o modelo ajustado está fortemente suscetível aos problemas da multicolinearidade.

Exemplo- Inventário de estoque

- O modelo de regressão polinomial (não ortogonal) de segunda ordem, ajustado pelo método de mínimos quadrados, é o seguinte:

Exemplo- Inventário de estoque

- O modelo de regressão polinomial (não ortogonal) de segunda ordem, ajustado pelo método de mínimos quadrados, é o seguinte:

$$\widehat{\texttt{Custo}} = 362.17 - 0.67 \times \texttt{Quant} + 0.0022 \times \texttt{Quant}^2$$

Exemplo- Inventário de estoque

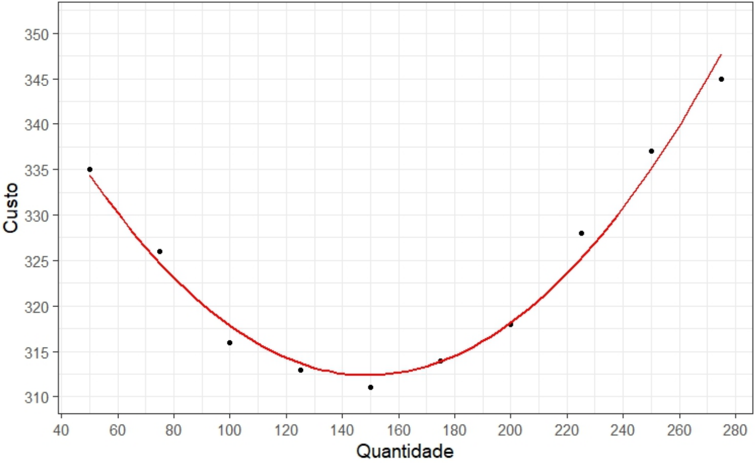


Figura 6: Custo médio de estocagem versus quantidade de itens reabastecidos com regressão polinomial ajustada

Exemplo- Inventário de estoque

- Vamos agora considerar o modelo de regressão polinomial ortogonal de segunda ordem:

Exemplo- Inventário de estoque

- Vamos agora considerar o modelo de regressão polinomial ortogonal de segunda ordem:

$$\texttt{Custo} = \beta_0 P_0(\texttt{Quant}) + \beta_1 P_1(\texttt{Quant}) + \beta_2 P_2(\texttt{Quant}) + \epsilon, \quad \epsilon \sim \mathrm{N}(0, \sigma^2)$$

Exemplo- Inventário de estoque

Tabela 3: Coeficientes dos polinômios ortogonais para os dados do inventário

i	$P_0(x_i)$	$P_1(x_i)$	$P_2(x_i)$
1	1	-9	6
2	1	-7	2
3	1	-5	-1
4	1	-3	-3
5	1	-1	-4
6	1	1	-4
7	1	3	-3
8	1	5	-1
9	1	7	2
10	1	9	6
	$\sum_{i=1}^{10} P_0^2(x_i) = 10$	$\sum_{i=1}^{10} P_1^2(x_i) = 330$	$\sum_{i=1}^{10} P_2^2(x_i) = 132$
		$\lambda_1 = 2$	$\lambda_2 = 1/2$

Exemplo- Inventário de estoque

- Neste caso,

$$\boldsymbol{X'X} = \begin{bmatrix} \sum_{i=1}^{10} P_0^2(x_i) & 0 & 0 \\ 0 & \sum_{i=1}^{10} P_1^2(x_i) & 0 \\ 0 & 0 & \sum_{i=1}^{10} P_2^2(x_i) \end{bmatrix} = \begin{bmatrix} 10 & 0 & 0 \\ 0 & 330 & 0 \\ 0 & 0 & 132 \end{bmatrix}$$

Exemplo- Inventário de estoque

- Neste caso,

$$\boldsymbol{X'X} = \begin{bmatrix} \sum_{i=1}^{10} P_0^2(x_i) & 0 & 0 \\ 0 & \sum_{i=1}^{10} P_1^2(x_i) & 0 \\ 0 & 0 & \sum_{i=1}^{10} P_2^2(x_i) \end{bmatrix} = \begin{bmatrix} 10 & 0 & 0 \\ 0 & 330 & 0 \\ 0 & 0 & 132 \end{bmatrix}$$

$$\boldsymbol{X'y} = \begin{bmatrix} \sum_{i=1}^{n} P_0(x_i)y_i \\ \sum_{i=1}^{n} P_1(x_i)y_i \\ \sum_{i=1}^{n} P_2(x_i)y_i \end{bmatrix} = \begin{bmatrix} 3243 \\ 245 \\ 369 \end{bmatrix}$$

Exemplo- Inventário de estoque

$$\hat{\beta} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y} = \begin{bmatrix} \frac{1}{10} & 0 & 0 \\ 0 & \frac{1}{330} & 0 \\ 0 & 0 & \frac{1}{132} \end{bmatrix} \times \begin{bmatrix} 3243 \\ 245 \\ 369 \end{bmatrix} = \begin{bmatrix} 324.30 \\ 0.74 \\ 2.79 \end{bmatrix}$$

Exemplo- Inventário de estoque

$$\hat{\beta} = (\boldsymbol{X'X})^{-1}\boldsymbol{X'y} = \begin{bmatrix} \frac{1}{10} & 0 & 0 \\ 0 & \frac{1}{330} & 0 \\ 0 & 0 & \frac{1}{132} \end{bmatrix} \times \begin{bmatrix} 3243 \\ 245 \\ 369 \end{bmatrix} = \begin{bmatrix} 324.30 \\ 0.74 \\ 2.79 \end{bmatrix}$$

- Logo, o modelo ajustado fica dado por:

$$\widehat{\texttt{Custo}} = 324.30 + 0.7424 \times P_1(\texttt{Quant}) + 2.7955 \times P_2(\texttt{Quant})$$

Modelos polinomiais em duas ou mais variáveis

Modelos polinomiais em duas ou mais variáveis

- Modelos polinomiais podem ser estendidos para o caso de duas ou mais variáveis explicativas.

Modelos polinomiais em duas ou mais variáveis

- Modelos polinomiais podem ser estendidos para o caso de duas ou mais variáveis explicativas.

- Por exemplo, um modelo polinomial de segunda ordem definido em duas variáveis é dado por:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1^2 + \beta_4 x_2^2 + \beta_5 x_1 x_2 + \epsilon,$$

com $\epsilon \sim \mathrm{N}(0, \sigma^2)$.

Modelos polinomiais em duas ou mais variáveis

- Modelos polinomiais podem ser estendidos para o caso de duas ou mais variáveis explicativas.

- Por exemplo, um modelo polinomial de segunda ordem definido em duas variáveis é dado por:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1^2 + \beta_4 x_2^2 + \beta_5 x_1 x_2 + \epsilon,$$

com $\epsilon \sim \mathrm{N}(0, \sigma^2)$.

- Neste caso, é comum chamar a representação gráfica de $\mathrm{E}(y|\boldsymbol{x})$ de **superfície de resposta**:

$$\mathrm{E}(y|\boldsymbol{x}) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1^2 + \beta_4 x_2^2 + \beta_5 x_1 x_2$$

Exemplo- Gastos em educação e desempenho escolar

Exemplo- Gastos em educação e desempenho escolar

- Nesta aplicação vamos considerar a base de dados `sat` disponível na biblioteca `faraway` do `R`.

Exemplo- Gastos em educação e desempenho escolar

- Nesta aplicação vamos considerar a base de dados `sat` disponível na biblioteca `faraway` do `R`.

- O objetivo é ilustrar uma análise de regressão polinomial envolvendo duas variáveis explicativas.

Exemplo- Gastos em educação e desempenho escolar

- Nesta aplicação vamos considerar a base de dados `sat` disponível na biblioteca `faraway` do `R`.
- O objetivo é ilustrar uma análise de regressão polinomial envolvendo duas variáveis explicativas.
- Os dados referem-se a indicadores de gastos com educação pública e resultados de testes estudantis nos 50 estados norte-americanos no período 1994-95.

Exemplo- Gastos em educação e desempenho escolar

- As variáveis consideradas na análise são as seguintes:

Exemplo- Gastos em educação e desempenho escolar

- As variáveis consideradas na análise são as seguintes:
    - `total`: Escore médio dos estudantes no SAT (*Scholastic Aptitude Test*);

Exemplo- Gastos em educação e desempenho escolar

- As variáveis consideradas na análise são as seguintes:
  - `total`: Escore médio dos estudantes no SAT (*Scholastic Aptitude Test*);
  - `takers`: Porcentagem de estudantes elegíveis que realizaram o SAT;

Exemplo- Gastos em educação e desempenho escolar

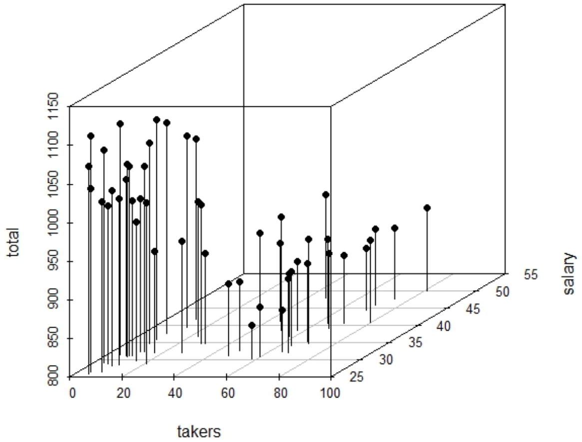
- As variáveis consideradas na análise são as seguintes:
  - `total`: Escore médio dos estudantes no SAT (*Scholastic Aptitude Test*);
  - `takers`: Porcentagem de estudantes elegíveis que realizaram o SAT;
  - `salary`: Salário médio estimado dos professores do ensino básico.

Exemplo- Gastos em educação e desempenho escolar

- As variáveis consideradas na análise são as seguintes:
  - `total`: Escore médio dos estudantes no SAT (*Scholastic Aptitude Test*);
  - `takers`: Porcentagem de estudantes elegíveis que realizaram o SAT;
  - `salary`: Salário médio estimado dos professores do ensino básico.

- O objetivo é ajustar um modelo de regressão que explique o desempenho dos estudantes (`total`) em função das outras duas variáveis (`takers` e `salary`).

Exemplo- Gastos em educação e desempenho escolar



Exemplo- Gastos em educação e desempenho escolar

- Diferentes modelos de regressão linear serão ajustados, considerando (ou não):

Exemplo- Gastos em educação e desempenho escolar

- Diferentes modelos de regressão linear serão ajustados, considerando (ou não):
  - Efeito quadrático de `takers`;

Exemplo- Gastos em educação e desempenho escolar

- Diferentes modelos de regressão linear serão ajustados, considerando (ou não):
  - Efeito quadrático de `takers`;
  - Efeito quadrático de `salary`;

Exemplo- Gastos em educação e desempenho escolar

- Diferentes modelos de regressão linear serão ajustados, considerando (ou não):
  - Efeito quadrático de `takers`;
  - Efeito quadrático de `salary`;
  - Efeito de interação entre `takers` e `salary`.

Exemplo- Gastos em educação e desempenho escolar

- Diferentes modelos de regressão linear serão ajustados, considerando (ou não):
    - Efeito quadrático de `takers`;
    - Efeito quadrático de `salary`;
    - Efeito de interação entre `takers` e `salary`.

1. Modelo com efeitos quadráticos de `takers` e `salary` e com efeito de interação:

Exemplo- Gastos em educação e desempenho escolar

- Diferentes modelos de regressão linear serão ajustados, considerando (ou não):
  - Efeito quadrático de `takers`;
  - Efeito quadrático de `salary`;
  - Efeito de interação entre `takers` e `salary`.

1. Modelo com efeitos quadráticos de `takers` e `salary` e com efeito de interação:

$$\texttt{total} = \beta_0 + \beta_1\texttt{takers} + \beta_2\texttt{salary} + \beta_3\texttt{takers}^2 + \beta_4\texttt{salary}^2 + \beta_5\texttt{takers} \times \texttt{salary} + \epsilon$$

Exemplo- Gastos em educação e desempenho escolar

2. Modelo com efeitos quadráticos de `takers` e `salary`:

Exemplo- Gastos em educação e desempenho escolar

2. Modelo com efeitos quadráticos de `takers` e `salary`:

$$\texttt{total} = \beta_0 + \beta_1 \texttt{takers} + \beta_2 \texttt{salary} + \beta_3 \texttt{takers}^2 + \beta_4 \texttt{salary}^2 + \epsilon$$

Exemplo- Gastos em educação e desempenho escolar

2. Modelo com efeitos quadráticos de `takers` e `salary`:

$$\texttt{total} = \beta_0 + \beta_1\texttt{takers} + \beta_2\texttt{salary} + \beta_3\texttt{takers}^2 + \beta_4\texttt{salary}^2 + \epsilon$$

3. Modelo com efeito quadrático de `takers`:

Exemplo- Gastos em educação e desempenho escolar

2. Modelo com efeitos quadráticos de `takers` e `salary`:

$$\texttt{total} = \beta_0 + \beta_1 \texttt{takers} + \beta_2 \texttt{salary} + \beta_3 \texttt{takers}^2 + \beta_4 \texttt{salary}^2 + \epsilon$$

3. Modelo com efeito quadrático de `takers`:

$$\texttt{total} = \beta_0 + \beta_1 \texttt{takers} + \beta_2 \texttt{salary} + \beta_3 \texttt{takers}^2 + \epsilon$$

Exemplo- Gastos em educação e desempenho escolar

4. Modelo com efeitos lineares de `takers` e `salary`:

Exemplo- Gastos em educação e desempenho escolar

4. Modelo com efeitos lineares de `takers` e `salary`:

$$\texttt{total} = \beta_0 + \beta_1 \texttt{takers} + \beta_2 \texttt{salary} + \epsilon$$

Exemplo- Gastos em educação e desempenho escolar

4. Modelo com efeitos lineares de **takers** e **salary**:

$$\texttt{total} = \beta_0 + \beta_1 \texttt{takers} + \beta_2 \texttt{salary} + \epsilon$$

5. Modelo com efeitos lineares de **takers** e **salary** e com efeito de interação:

Exemplo- Gastos em educação e desempenho escolar

4. Modelo com efeitos lineares de **takers** e **salary**:

$$\texttt{total} = \beta_0 + \beta_1\texttt{takers} + \beta_2\texttt{salary} + \epsilon$$

5. Modelo com efeitos lineares de **takers** e **salary** e com efeito de interação:

$$\texttt{total} = \beta_0 + \beta_1\texttt{takers} + \beta_2\texttt{salary} + \beta_3\texttt{takers} \times \texttt{salary} + \epsilon$$

Exemplo- Gastos em educação e desempenho escolar

4. Modelo com efeitos lineares de **takers** e **salary**:

$$\texttt{total} = \beta_0 + \beta_1\texttt{takers} + \beta_2\texttt{salary} + \epsilon$$

5. Modelo com efeitos lineares de **takers** e **salary** e com efeito de interação:

$$\texttt{total} = \beta_0 + \beta_1\texttt{takers} + \beta_2\texttt{salary} + \beta_3\texttt{takers} \times \texttt{salary} + \epsilon$$

- Em todos os casos os erros são independentes e identicamente distribuídos $\epsilon \sim \mathrm{N}(0, \sigma^2)$.

Exemplo- Gastos em educação e desempenho escolar

4. Modelo com efeitos lineares de `takers` e `salary`:

$$\texttt{total} = \beta_0 + \beta_1 \texttt{takers} + \beta_2 \texttt{salary} + \epsilon$$

5. Modelo com efeitos lineares de `takers` e `salary` e com efeito de interação:

$$\texttt{total} = \beta_0 + \beta_1 \texttt{takers} + \beta_2 \texttt{salary} + \beta_3 \texttt{takers} \times \texttt{salary} + \epsilon$$

- Em todos os casos os erros são independentes e identicamente distribuídos $\epsilon \sim \mathrm{N}(0, \sigma^2)$.
- Na sequência são apresentados os modelos ajustados.

Exemplo- Gastos em educação e desempenho escolar

1. Modelo com efeitos quadráticos de **takers** e **salary** e com efeito de interação:

$$\widehat{\texttt{total}} = 909.66 - 6.612 \times \texttt{takers} + 8.678 \times \texttt{salary} + 0.058 \times \texttt{takers}^2 - 0.082 \times \texttt{salary}^2 - 0.018 \times \texttt{takers} \times \texttt{salary}$$

Exemplo- Gastos em educação e desempenho escolar

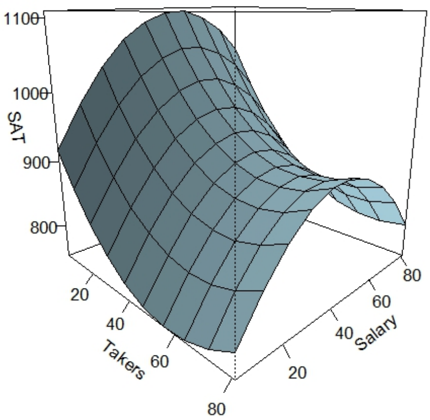
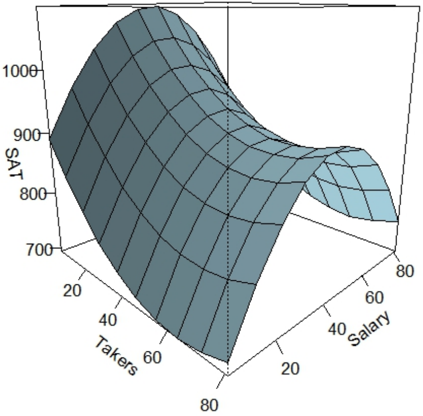


Figura 8: Superfície de resposta para o modelo com efeitos quadráticos de `takers` e `salary` e com

2. Modelo com efeitos quadráticos de `takers` e `salary`:

$$\widehat{\texttt{total}} = 875.72 - 7.081 \times \texttt{takers} + 11.114 \times \texttt{salary} + 0.056 \times \texttt{takers}^2 - 0.125 \times \texttt{salary}^2$$

Exemplo- Gastos em educação e desempenho escolar



Exemplo- Gastos em educação e desempenho escolar

3. Modelo com efeito quadrático de `takers`:

$$\widehat{\texttt{total}} = 1039 - 6.646 \times \texttt{takers} + 1.803 \times \texttt{salary} + 0.051 \times \texttt{takers}^2$$

Exemplo- Gastos em educação e desempenho escolar

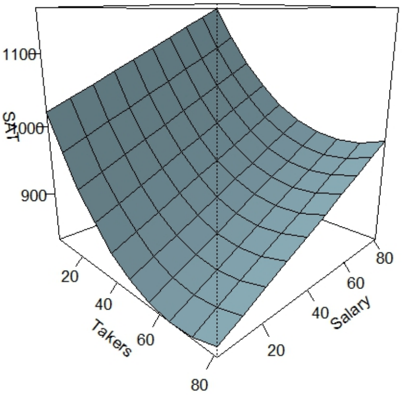


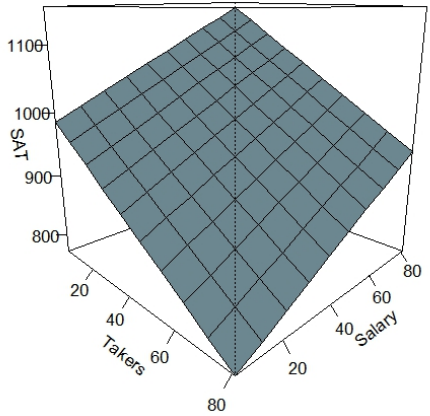
Figura 10: Superfície de resposta para o modelo com efeito quadrático de `takers`

Exemplo- Gastos em educação e desempenho escolar

4. Modelo com efeitos lineares de `takers` e `salary`:

$$\widehat{\texttt{total}} = 987.9 - 2.778 \times \texttt{takers} + 2.180 \times \texttt{salary}$$

Exemplo- Gastos em educação e desempenho escolar



Exemplo- Gastos em educação e desempenho escolar

5. Modelo com efeitos lineares de **takers** e **salary** e com efeito de interação:

$$\widehat{\texttt{total}} = 1082 - 5.026 \times \texttt{takers} - 0.719 \times \texttt{salary} + 0.064 \times \texttt{takers} \times \texttt{salary}$$

Exemplo- Gastos em educação e desempenho escolar

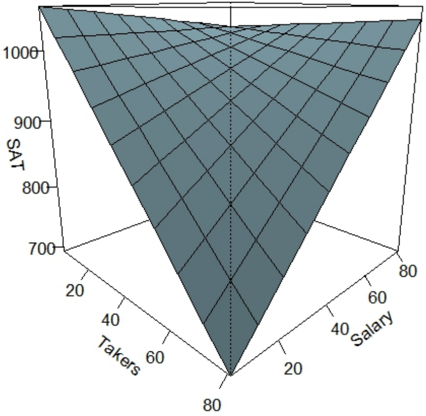


Figura 12: Superfície de resposta para o modelo com efeitos lineares de `takers` e `salary` e com

Exemplo- Gastos em educação e desempenho escolar

Tabela 4: Resumo do modelo ajustado com os efeitos quadráticos de `takers` e `salary` e com efeito de interação

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	909.6693	161.5353	5.63	0.0000
takers	-6.6127	1.4411	-4.59	0.0000
salary	8.6787	9.8935	0.88	0.3851
$\texttt{takers}^2$	0.0583	0.0115	5.06	0.0000
$\texttt{salary}^2$	-0.0820	0.1513	-0.54	0.5907
takers × salary	-0.0182	0.0467	-0.39	0.6991

Exemplo- Gastos em educação e desempenho escolar

- Com base no resumo do modelo ajustado com efeitos quadráticos de `takers` e `salary` e efeito de interação, concluímos que:

Exemplo- Gastos em educação e desempenho escolar

- Com base no resumo do modelo ajustado com efeitos quadráticos de `takers` e `salary` e efeito de interação, concluímos que:

1. O termo de interação não é estatisticamente significativo ($p = 0.6991$), de forma que pode ser removido do modelo;

Exemplo- Gastos em educação e desempenho escolar

- Com base no resumo do modelo ajustado com efeitos quadráticos de `takers` e `salary` e efeito de interação, concluímos que:

1. O termo de interação não é estatisticamente significativo ($p = 0.6991$), de forma que pode ser removido do modelo;
2. O efeito quadrático de `salary` também não é significativo ($p = 0.5907$), ao contrário do efeito quadrático de `takers` ($p < 0.0001$);

Exemplo- Gastos em educação e desempenho escolar

- Com base no resumo do modelo ajustado com efeitos quadráticos de `takers` e `salary` e efeito de interação, concluímos que:

1. O termo de interação não é estatisticamente significativo ($p = 0.6991$), de forma que pode ser removido do modelo;
2. O efeito quadrático de `salary` também não é significativo ($p = 0.5907$), ao contrário do efeito quadrático de `takers` ($p < 0.0001$);
3. Desta forma, baseado nos resultados apresentados, o modelo poderia apenas conter o efeito quadrático de `takers`. Adicionalmente, poderia se avaliar a remoção do efeito de `salary` (exercício).