ACA228 - Modelos de Regressão e Previsão

Regressão Linear Multipla: Tópicos adicionais

Prof. Carlos Trucíos
carlos.trucios@facc.ufrj.br
ctruciosm.github.io

Faculdade de Administração e Ciências Contábeis,
Universidade Federal do Rio de Janeiro

Aula 10

Não-Leariedades

Comparação de modelos

Intervalos de Confiança / Previsão

Não-Ineariedades

Não-Leariedades

Até agora, temos trabalhado com modelos da forma

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + u$$

Não-Linealiedades

Até agora, temos trabalhado com modelos da forma

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + u$$

Muitas vezes, para tentar capturar não-lineariedades, podemos estar interessados em incluir, por exemplo x_k^2, x_k^3 $\ldots$,

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_1^2 + \beta_3 x_2 + \beta_4 x_3 + \beta_5 x_4 + u$$

Não-Linearidades

Até agora, temos trabalhado com modelos da forma

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + u$$

Muitas vezes, para tentar capturar não-lineariedades, podemos estar interessados em incluir, por exemplo x_k^2, x_k^3 $\dots$,

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_1^2 + \beta_3 x_2 + \beta_4 x_3 + \beta_5 x_4 + u$$

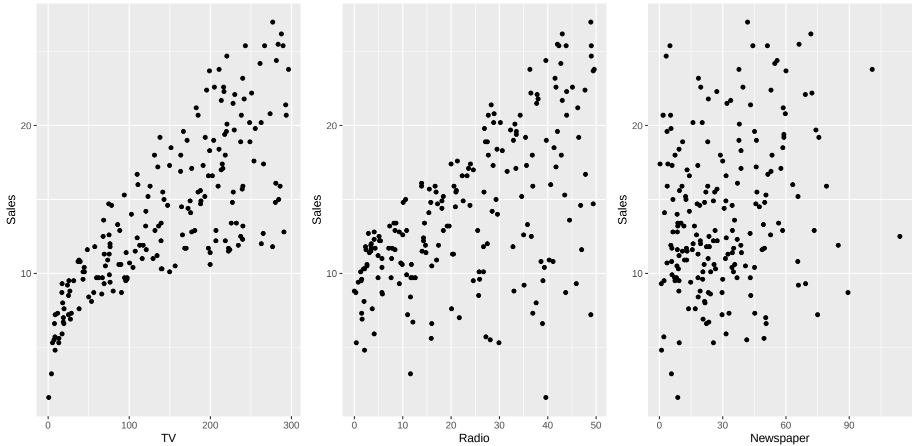
No contexto de MRLM podemos incluir termos polinomiais. O processo de estimação continua sendo o mesmo mas devemos ter muito cuidado na **interpretação**.

Não-Lineariedades

```
uri <- "https://raw.githubusercontent.com/ctruciosm/ISLR/master/c
advertising <-   read.csv(uri)
modelo1 <- lm(Sales ~ TV + Radio + Newspaper,
              data = advertising)
summary(modelo1)$adj.r.squared
```

```
## [1] 0.8956373
```

Não-Lineariedades



Não-Leariedades

```r
modelo2 <- lm(Sales ~ TV + I(TV^2) + Radio + Newspaper,
              data = advertising)
summary(modelo1)$adj.r.squared
```

```
## [1] 0.8956373
```

```r
summary(modelo2)$adj.r.squared
```

```
## [1] 0.9150029
```

Não-Lineariedades

```r
round(summary(modelo2)$coef,5)
```

```
##              Estimate Std. Error  t value Pr(>|t|)
## (Intercept)   1.27016    0.37447  3.39189  0.00084
## TV            0.07847    0.00500 15.68980  0.00000
## I(TV^2)      -0.00011    0.00002 -6.75694  0.00000
## Radio         0.19256    0.00779 24.70605  0.00000
## Newspaper     0.00089    0.00531  0.16784  0.86688
```

Não-Leariedades

```
round(summary(modelo2)$coef,5)
```

```
##              Estimate Std. Error  t value Pr(>|t|)
## (Intercept)   1.27016    0.37447  3.39189  0.00084
## TV            0.07847    0.00500 15.68980  0.00000
## I(TV^2)      -0.00011    0.00002 -6.75694  0.00000
## Radio         0.19256    0.00779 24.70605  0.00000
## Newspaper     0.00089    0.00531  0.16784  0.86688
```

O termo quadrático e significativo, mas será que podemos interpretar ele como estamos acostumados?

Não-Linearidades: Interpretação

Termo Quadrático

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x + \hat{\beta}_2 x^2 \tag{1}$$

- Se estivermos interessados unicamente em predição, podemos usar diretamente (1)

Não-Linearidades: Interpretação

Termo Quadrático

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x + \hat{\beta}_2 x^2 \tag{1}$$

- Se estivermos interessados unicamente em predição, podemos usar diretamente (1)
- Se estivermos interessados em conhecer o efeito de x sobre y precisamos **interpretar** o modelo.

Não-Lineariedades: Interpretação

Termo Quadrático

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x + \hat{\beta}_2 x^2 \tag{1}$$

- Se estivermos interessados unicamente em predição, podemos usar diretamente (1)
- Se estivermos interessados em conhecer o efeito de x sobre y precisamos **interpretar** o modelo.

Não-Lineariedades: Interpretação

Termo Quadrático

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x + \hat{\beta}_2 x^2 \tag{1}$$

- Se estivermos interessados unicamente em predição, podemos usar diretamente (1)
- Se estivermos interessados em conhecer o efeito de x sobre y precisamos **interpretar** o modelo.

Derivando w.r.t x, temos que $\frac{d\hat{y}}{dx} = (\hat{\beta}_1 + 2\hat{\beta}_2 x)$, então

$$\frac{\Delta \hat{y}}{\Delta x} \approx (\hat{\beta}_1 + 2\hat{\beta}_2 x)$$

Não-Linearidades: Interpretação

...(continuação) Termo Quadrático

- Se $x = 0$, $\dfrac{\Delta \hat{y}}{\Delta x} \approx \hat{\beta}_1$, $\hat{\beta}_1$ pode ser interpretado como a inclinação aproximada na alteração de $x = 0$ para $x = 1$

Não-Linearidades: Interpretação

. . . (continuação) Termo Quadrático

- Se $x = 0$, $\dfrac{\Delta \hat{y}}{\Delta x} \approx \hat{\beta}_1$, $\hat{\beta}_1$ pode ser interpretado como a inclinação aproximada na alteração de $x = 0$ para $x = 1$
- Para $x > 0$, $2\hat{\beta}_2 x$ deve ser considerado na interpretação

Não-Linearidades: Interpretação

. . . (continuação) Termo Quadrático

- Se $x = 0$, $\dfrac{\Delta \hat{y}}{\Delta x} \approx \hat{\beta}_1$, $\hat{\beta}_1$ pode ser interpretado como a inclinação aproximada na alteração de $x = 0$ para $x = 1$
- Para $x > 0$, $2\hat{\beta}_2 x$ deve ser considerado na interpretação

Não-Linearidades: Interpretação

...(continuação) Termo Quadrático

- Se $x = 0$, $\dfrac{\Delta \hat{y}}{\Delta x} \approx \hat{\beta}_1$, $\hat{\beta}_1$ pode ser interpretado como a inclinação aproximada na alteração de $x = 0$ para $x = 1$
- Para $x > 0$, $2\hat{\beta}_2 x$ deve ser considerado na interpretação

```
library(wooldridge)
coef(lm(wage~ exper + I(exper^2), data = wage1))
```

```
##  (Intercept)        exper   I(exper^2)
##  3.725405759  0.298100104 -0.006129887
```

- Como varia *wage* em função de *exper*?

$$\Delta \hat{y} \approx (0.2981 - 2 \times 0.0061x)\Delta x$$

Não-Linearidades: Interpretação

...(continuação) Termo Quadrático

$$\Delta\hat{y} \approx (0.2981 - 2 \times 0.0061x)\Delta x$$

- O primeiro ano a experiência vale 0.2981 USD por hora

Não-Linearidades: Interpretação

...(continuação) Termo Quadrático

$$\Delta\hat{y} \approx (0.2981 - 2 \times 0.0061x)\Delta x$$

- O primeiro ano a experiência vale 0.2981 USD por hora
- Depois do primeiro ano, o efeito de experiência começa a cair:

Não-Linearidades: Interpretação

. . . (continuação) Termo Quadrático

$$\Delta\hat{y} \approx (0.2981 - 2 \times 0.0061x)\Delta x$$

- O primeiro ano a experiência vale 0.2981 USD por hora
- Depois do primeiro ano, o efeito de experiência começa a cair:
  - Para o segundo ano: $0.2981 - 2 \times 0.0061(1) = 0.2859$

Não-Linearidades: Interpretação

...(continuação) Termo Quadrático

$$\Delta\hat{y} \approx (0.2981 - 2 \times 0.0061x)\Delta x$$

- O primeiro ano a experiência vale 0.2981 USD por hora
- Depois do primeiro ano, o efeito de experiência começa a cair:
  - Para o segundo ano: $0.2981 - 2 \times 0.0061(1) = 0.2859$
  - Para o terceiro ano: $0.2981 - 2 \times 0.0061(2) = 0.2737$

Não-Linearidades: Interpretação

...(continuação) Termo Quadrático

$$\Delta \hat{y} \approx (0.2981 - 2 \times 0.0061x)\Delta x$$

- O primeiro ano a experiência vale 0.2981 USD por hora
- Depois do primeiro ano, o efeito de experiência começa a cair:
  - Para o segundo ano: $0.2981 - 2 \times 0.0061(1) = 0.2859$
  - Para o terceiro ano: $0.2981 - 2 \times 0.0061(2) = 0.2737$
  - Para o decimo ano: $0.2981 - 2 \times 0.0061(9) = 0.1883$

Não-Linearidades: Interpretação

. . . (continuação) Termo Quadrático

$$\Delta\hat{y} \approx (0.2981 - 2 \times 0.0061x)\Delta x$$

- O primeiro ano a experiência vale 0.2981 USD por hora
- Depois do primeiro ano, o efeito de experiência começa a cair:
  - Para o segundo ano: $0.2981 - 2 \times 0.0061(1) = 0.2859$
  - Para o terceiro ano: $0.2981 - 2 \times 0.0061(2) = 0.2737$
  - Para o decimo ano: $0.2981 - 2 \times 0.0061(9) = 0.1883$
- Note que o **ponto de infleção** é $x* = -\hat{\beta}_1/2\hat{\beta}_2$

Não-Linearidades: Interpretação

. . . (continuação) Termo Quadrático

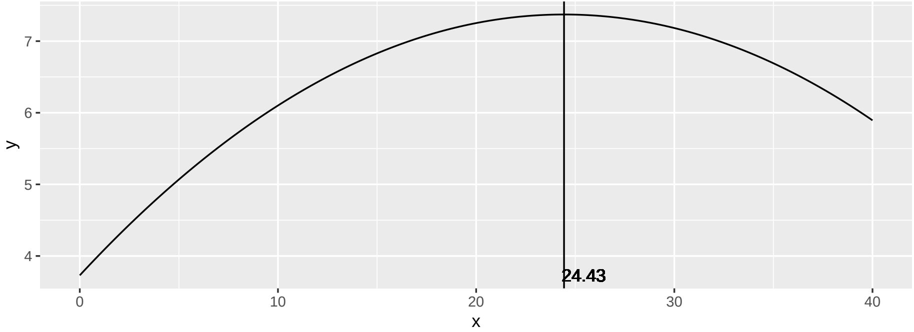
$$\Delta\hat{y} \approx (0.2981 - 2 \times 0.0061x)\Delta x$$

- O primeiro ano a experiência vale 0.2981 USD por hora
- Depois do primeiro ano, o efeito de experiência começa a cair:
  - Para o segundo ano: $0.2981 - 2 \times 0.0061(1) = 0.2859$
  - Para o terceiro ano: $0.2981 - 2 \times 0.0061(2) = 0.2737$
  - Para o decimo ano: $0.2981 - 2 \times 0.0061(9) = 0.1883$
- Note que o **ponto de infleção** é $x* = -\hat{\beta}_1/2\hat{\beta}_2$
- No nosso caso: $x* = 0.2981/(2 \times 0.0061) = 24.43$

Não-Linearidades: Interpretação

...(continuação) Termo Quadrático

$$\hat{y} = 3.73 + 0.2981x - 0.0061x^2$$



Não-Linearidades: Interpretação

. . . (continuação) Termo Quadrático

O que significa esse 24.43 anos?

- Se na amostra, apenas uma pequena fração dos dados apresentam idade > 24.43, não devemos nos preocupar (mas. . . .)

Não-Linearidades: Interpretação

. . . (continuação) Termo Quadrático

O que significa esse 24.43 anos?

- Se na amostra, apenas uma pequena fração dos dados apresentam idade $>$ 24.43, não devemos nos preocupar (mas. . . .)

```
prop.table(table(wage1$exper>24.43))
```

```
##
##     FALSE      TRUE
## 0.7205323 0.2794677
```

Não-Lineariedades: Interpretação

. . . (continuação) Termo Quadrático

O que significa esse 24.43 anos?

- Se na amostra, apenas uma pequena fração dos dados apresentam idade > 24.43, não devemos nos preocupar (mas. . . .)

```
prop.table(table(wage1$exper>24.43))
```

```
##
##     FALSE      TRUE
## 0.7205323 0.2794677
```

- É possivel que o retorno de *exper* sobre *wage* seja negativo a partir de algum ponto, mas **cuidado com as variáveis omitidas!** (levam a estimadores viesados)

Não-Leariedades: Interpretação

```r
round(coef(modelo2),5)
```

```
## (Intercept)          TV     I(TV^2)       Radio   Newspaper
##     1.27016     0.07847    -0.00011     0.19256     0.00089
```

Não-Linearidades: Interpretação

```r
round(coef(modelo2),5)
```

```
## (Intercept)          TV     I(TV^2)       Radio   Newspaper
##     1.27016     0.07847    -0.00011     0.19256     0.00089
```

- Como varia *Sales* em função de *TV*?

$$\Delta y \approx (0.07847 - 2 \times 0.00011x)\Delta x$$

Não-Leariedades: Interpretação

$$\Delta\hat{y} \approx (0.07847 - 2 \times 0.00011x)\Delta x$$

Não-Linearidades: Interpretação

$$\Delta \hat{y} \approx (0.07847 - 2 \times 0.00011x)\Delta x$$

- O investimento inicial de 1K USD em TV geram, em média 0.07847 milhares de unidades vendidas (78.47 unidades vendidas).

Não-Linearidades: Interpretação

$$\Delta\hat{y} \approx (0.07847 - 2 \times 0.00011x)\Delta x$$

- O investimento inicial de 1K USD em TV geram, em média 0.07847 milhares de unidades vendidas (78.47 unidades vendidas).
- Depois dos primeiros 1K USD investidos, o efeito começa a diminuir:

Não-Lineariedades: Interpretação

$$\Delta\hat{y} \approx (0.07847 - 2 \times 0.00011x)\Delta x$$

- O investimento inicial de 1K USD em TV geram, em média 0.07847 milhares de unidades vendidas (78.47 unidades vendidas).
- Depois dos primeiros 1K USD investidos, o efeito começa a diminuir:
  - O investimento de 2K em TV geram:
    $0.07847 - 2 \times 0.00011(1) = 0.07825$

Não-Linearidades: Interpretação

$$\Delta\hat{y} \approx (0.07847 - 2 \times 0.00011x)\Delta x$$

- O investimento inicial de 1K USD em TV geram, em média 0.07847 milhares de unidades vendidas (78.47 unidades vendidas).
- Depois dos primeiros 1K USD investidos, o efeito começa a diminuir:
  - O investimento de 2K em TV geram: $0.07847 - 2 \times 0.00011(1) = 0.07825$
  - O investimento de 3K em TV geram: $0.07847 - 2 \times 0.00011(2) = 0.07803$

Não-Lineariedades: Interpretação

$$\Delta\hat{y} \approx (0.07847 - 2 \times 0.00011x)\Delta x$$

- O investimento inicial de 1K USD em TV geram, em média 0.07847 milhares de unidades vendidas (78.47 unidades vendidas).
- Depois dos primeiros 1K USD investidos, o efeito começa a diminuir:
  - O investimento de 2K em TV geram: $0.07847 - 2 \times 0.00011(1) = 0.07825$
  - O investimento de 3K em TV geram: $0.07847 - 2 \times 0.00011(2) = 0.07803$
  - O investimento de 10K em TV geram: $0.07847 - 2 \times 0.00011(9) = 0.07649$

Não-Lineariedades: Interpretação

$$\Delta \hat{y} \approx (0.07847 - 2 \times 0.00011x)\Delta x$$

- O investimento inicial de 1K USD em TV geram, em média 0.07847 milhares de unidades vendidas (78.47 unidades vendidas).
- Depois dos primeiros 1K USD investidos, o efeito começa a diminuir:
  - O investimento de 2K em TV geram: $0.07847 - 2 \times 0.00011(1) = 0.07825$
  - O investimento de 3K em TV geram: $0.07847 - 2 \times 0.00011(2) = 0.07803$
  - O investimento de 10K em TV geram: $0.07847 - 2 \times 0.00011(9) = 0.07649$
- Note que o **ponto de infleção** é $x* = -\hat{\beta}_1/2\hat{\beta}_2$

Não-Linearidades: Interpretação

$$\Delta \hat{y} \approx (0.07847 - 2 \times 0.00011x)\Delta x$$

- O investimento inicial de 1K USD em TV geram, em média 0.07847 milhares de unidades vendidas (78.47 unidades vendidas).
- Depois dos primeiros 1K USD investidos, o efeito começa a diminuir:
  - O investimento de 2K em TV geram: $0.07847 - 2 \times 0.00011(1) = 0.07825$
  - O investimento de 3K em TV geram: $0.07847 - 2 \times 0.00011(2) = 0.07803$
  - O investimento de 10K em TV geram: $0.07847 - 2 \times 0.00011(9) = 0.07649$
- Note que o **ponto de infleção** é $x* = -\hat{\beta}_1/2\hat{\beta}_2$
- No nosso caso: $x* = 0.07847/(2 \times 0.00011) = 356.6818$

Não-Linearidades: Interpretação

$$log(price) =$$
$$\beta_0 + \beta_1 log(nox) + \beta_2 log(dist) + \beta_3 rooms + \beta_4 I(rooms^2) + \beta_5 stratio + u$$

Não-Lineariedades: Interpretação

$$log(price) =$$
$$\beta_0 + \beta_1 log(nox) + \beta_2 log(dist) + \beta_3 rooms + \beta_4 I(rooms^2) + \beta_5 stratio + u$$

```
##                Estimate Std. Error   t value     Pr(>|t|)
## (Intercept) 13.38547708 0.56647307 23.629503 1.884304e-83
## log(nox)    -0.90168178 0.11468692 -7.862115 2.340671e-14
## log(dist)   -0.08678134 0.04328071 -2.005081 4.549288e-02
## rooms       -0.54511291 0.16545413 -3.294647 1.055357e-03
## I(rooms^2)   0.06226119 0.01280498  4.862265 1.556648e-06
## stratio     -0.04759019 0.00585419 -8.129254 3.423303e-15
```

Não-Lineariedades: Interpretação

- $\hat{\beta}_{log(nox)}$, $\hat{\beta}_{log(dist)}$ e $\hat{\beta}_{stratio}$ são interpretados como visto nas aulas anteriores.

$$\Delta \widehat{\log(price)} \approx (-0.545 + 2 \times 0.062 rooms)\Delta rooms$$

Não-Lineariedades: Interpretação

- $\hat{\beta}_{log(nox)}$, $\hat{\beta}_{log(dist)}$ e $\hat{\beta}_{stratio}$ são interpretados como visto nas aulas anteriores.
- Como interpretamos *rooms*?

$$\Delta \widehat{\log(price)} \approx (-0.545 + 2 \times 0.062 rooms)\Delta rooms$$

Não-Lineariedades: Interpretação

- $\hat{\beta}_{log(nox)}$, $\hat{\beta}_{log(dist)}$ e $\hat{\beta}_{stratio}$ são interpretados como visto nas aulas anteriores.
- Como interpretamos *rooms*?

$$\Delta \widehat{\log(price)} \approx (-0.545 + 2 \times 0.062 rooms)\Delta rooms$$

Não-Linearidades: Interpretação

- $\hat{\beta}_{log(nox)}$, $\hat{\beta}_{log(dist)}$ e $\hat{\beta}_{stratio}$ são interpretados como visto nas aulas anteriores.
- Como interpretamos *rooms*?

$$\Delta \widehat{\log(price)} \approx (-0.545 + 2 \times 0.062 rooms)\Delta rooms$$

$$\%\Delta \widehat{price} \approx 100(-0.545 + 2 \times 0.062 rooms)\Delta rooms$$

Não-Leariedades: Interpretação

- $\hat{\beta}_{log(nox)}$, $\hat{\beta}_{log(dist)}$ e $\hat{\beta}_{stratio}$ são interpretados como visto nas aulas anteriores.
- Como interpretamos *rooms*?

$$\Delta \widehat{\log(price)} \approx (-0.545 + 2 \times 0.062 rooms)\Delta rooms$$

$$\%\Delta \widehat{price} \approx 100(-0.545 + 2 \times 0.062 rooms)\Delta rooms$$

$$\%\Delta \widehat{price} \approx (-54.5 + 12.4 rooms)\Delta rooms$$

Não-Leneariedades: Interpretação

$$\%\Delta\widehat{price} \approx (-54.5 + 12.4 rooms)\Delta rooms$$

Não-Linearidades: Interpretação

$$\%\Delta\widehat{price} \approx (-54.5 + 12.4 rooms)\Delta rooms$$

- Um aumento em rooms de 5 para 6 aumenta o preço em aprox $-54.5 + 12.4 \times 5 = 7.5\%$

Não-Linearidades: Interpretação

$$\%\Delta\widehat{price} \approx (-54.5 + 12.4 rooms)\Delta rooms$$

- Um aumento em rooms de 5 para 6 aumenta o preço em aprox $-54.5 + 12.4 \times 5 = 7.5\%$
- Um aumento em rooms de 6 para 7 aumenta o preço em aprox $-54.5 + 12.4 \times 6 = 19.9\%$

Não-Leariedades: Interpretação

$$\%\Delta\widehat{price} \approx (-54.5 + 12.4 rooms)\Delta rooms$$

- Um aumento em rooms de 5 para 6 aumenta o preço em aprox $-54.5 + 12.4 \times 5 = 7.5\%$
- Um aumento em rooms de 6 para 7 aumenta o preço em aprox $-54.5 + 12.4 \times 6 = 19.9\%$
- Um aumento em rooms de 4 para 5 diminui o preço em aprox $-54.5 + 12.4 \times 4 = -4.9\%$

Não-Lineariedades: Interpretação

$$\%\Delta\widehat{price} \approx (-54.5 + 12.4 rooms)\Delta rooms$$

- Um aumento em rooms de 5 para 6 aumenta o preço em aprox $-54.5 + 12.4 \times 5 = 7.5\%$
- Um aumento em rooms de 6 para 7 aumenta o preço em aprox $-54.5 + 12.4 \times 6 = 19.9\%$
- Um aumento em rooms de 4 para 5 diminui o preço em aprox $-54.5 + 12.4 \times 4 = -4.9\%$
- Um aumento em rooms de 3 para 4 diminui o preço em aprox $-54.5 + 12.4 \times 3 = -17.3\%$

Não-Linearidades: Interpretação

$$\%\Delta\widehat{price} \approx (-54.5 + 12.4 rooms)\Delta rooms$$

- Um aumento em rooms de 5 para 6 aumenta o preço em aprox $-54.5 + 12.4 \times 5 = 7.5\%$
- Um aumento em rooms de 6 para 7 aumenta o preço em aprox $-54.5 + 12.4 \times 6 = 19.9\%$
- Um aumento em rooms de 4 para 5 diminui o preço em aprox $-54.5 + 12.4 \times 4 = -4.9\%$
- Um aumento em rooms de 3 para 4 diminui o preço em aprox $-54.5 + 12.4 \times 3 = -17.3\%$

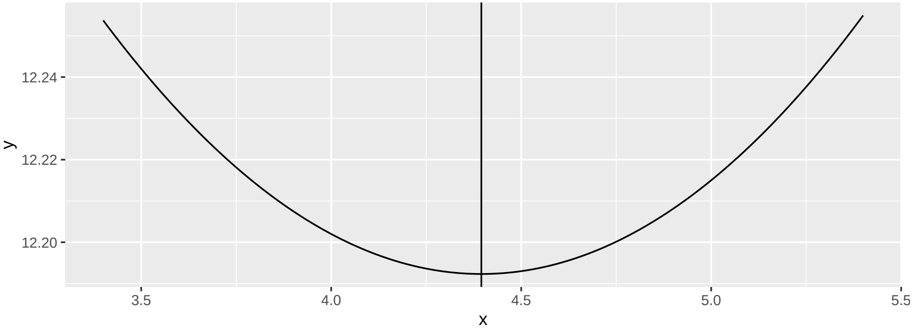
Não-Lineariedades: Interpretação

$$\%\Delta\widehat{price} \approx (-54.5 + 12.4 rooms)\Delta rooms$$

- Um aumento em rooms de 5 para 6 aumenta o preço em aprox $-54.5 + 12.4 \times 5 = 7.5\%$
- Um aumento em rooms de 6 para 7 aumenta o preço em aprox $-54.5 + 12.4 \times 6 = 19.9\%$
- Um aumento em rooms de 4 para 5 diminui o preço em aprox $-54.5 + 12.4 \times 4 = -4.9\%$
- Um aumento em rooms de 3 para 4 diminui o preço em aprox $-54.5 + 12.4 \times 3 = -17.3\%$

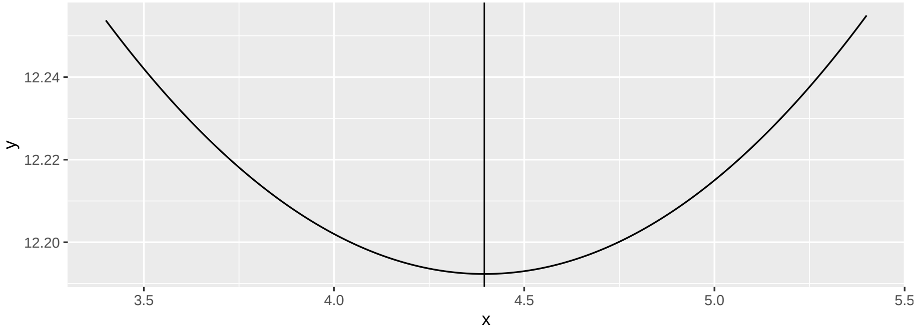
Faz sentido que aumentar o número de quartos cause uma diminuição no preço?

Não-Linearidades: Interpretação



Ponto de infleção 4.3951613.

Não-Lineariedades: Interpretação



Ponto de infleção 4.3951613.

Lembre-se: Se na amostra, apenas uma pequena fração dos dados apresentam idade < 4.3951613, não devemos nos preocupar.

Não-Linearidades: Interpretação

```
summary(hprice2$rooms)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   3.560   5.883   6.210   6.284   6.620   8.780
```

```
prop.table(table(hprice2$rooms<4.4))*100
```

```
##
##      FALSE       TRUE
## 99.0118577  0.9881423
```

Moral da historia

Faça uma boa EDA (exploratory data analysis), alguns resultados contraintuitivos podem desaparecer ao entendermos melhor os dados.

Comparação de modelos

Comparação de modelos

- Vimos que o R^2 nunca diminui quando incluimos mais váriaveis independentes.

Comparação de modelos

- Vimos que o R^2 nunca diminui quando incluimos mais váriaveis independentes.
- O R^2-ajustado, penaliza o número de variáveis e fornece uma ferramenta para compararmos modelos.

Comparação de modelos

- Vimos que o R^2 nunca diminui quando incluimos mais váriaveis independentes.
- O R^2-ajustado, penaliza o número de variáveis e fornece uma ferramenta para compararmos modelos.

Comparação de modelos

- Vimos que o R^2 nunca diminui quando incluimos mais váriaveis independentes.
- O R^2-ajustado, penaliza o número de variáveis e fornece uma ferramenta para compararmos modelos.

R^2-ajustado

$$R^2_{Adj} = 1 - (1 - R^2)\frac{n-1}{n-k-1}$$

Comparação de modelos

- Vimos que o R^2 nunca diminui quando incluimos mais váriaveis independentes.
- O R^2-ajustado, penaliza o número de variáveis e fornece uma ferramenta para compararmos modelos.

R^2-ajustado

$$R^2_{Adj} = 1 - (1 - R^2)\frac{n-1}{n-k-1}$$

- Nos permite comparar modelos não aninhados

Comparação de modelos

- Vimos que o R^2 nunca diminui quando incluimos mais váriaveis independentes.
- O R^2-ajustado, penaliza o número de variáveis e fornece uma ferramenta para compararmos modelos.

R^2-ajustado

$$R^2_{Adj} = 1 - (1 - R^2)\frac{n-1}{n-k-1}$$

- Nos permite comparar modelos não aninhados
- **Cuidado com comparar modelos com variavel dependente diferente**

Comparação de modelos

```r
modelo1 = lm(salary~sales+roe, data = ceosal1)
modelo2 = lm(log(salary)~log(sales)+roe, data = ceosal1)
summary(modelo1)$adj.r.squared
```

```
## [1] 0.01974617
```

```r
summary(modelo2)$adj.r.squared
```

```
## [1] 0.2750177
```

Comparação de modelos

```
modelo1 = lm(salary~sales+roe, data = ceosal1)
modelo2 = lm(log(salary)~log(sales)+roe, data = ceosal1)
summary(modelo1)$adj.r.squared
```

```
## [1] 0.01974617
```

```
summary(modelo2)$adj.r.squared
```

```
## [1] 0.2750177
```

Cuidado com comparar modelos com variavel dependente diferente

Comparação de modelos

```r
modelo1 = lm(salary~sales+roe, data = ceosal1)
modelo2 = lm(log(salary)~log(sales)+roe, data = ceosal1)
summary(modelo1)$adj.r.squared
```

```
## [1] 0.01974617
```

```r
summary(modelo2)$adj.r.squared
```

```
## [1] 0.2750177
```

Cuidado com comparar modelos com variavel dependente diferente

R^2-ajustado não pode ser utilizado para comparar modelos que tem diferentes formas funcionais da variável dependente.

Comparação de modelos

Seja

$$\log(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + u$$

Comparação de modelos

Seja

$$\log(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + u$$

- $\widehat{\log(y)} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \hat{\beta}_3 x_3$

Comparação de modelos

Seja

$$\log(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + u$$

- $\widehat{\log(y)} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \hat{\beta}_3 x_3$
- Uma primeira ideia seria fazer $\hat{y} = \exp(\widehat{\log(y)})$ (mas subestima o valor esperado y)

Comparação de modelos

Seja

$$\log(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + u$$

- $\widehat{\log(y)} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \hat{\beta}_3 x_3$
- Uma primeira ideia seria fazer $\hat{y} = \exp(\widehat{\log(y)})$ (mas subestima o valor esperado y)
- Sabemos que $y = \underbrace{\exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + u)}_{\exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3)\exp(u)}$

Comparação de modelos

Seja

$$\log(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + u$$

- $\widehat{\log(y)} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \hat{\beta}_3 x_3$
- Uma primeira ideia seria fazer $\hat{y} = \exp(\widehat{\log(y)})$ (mas subestima o valor esperado y)
- Sabemos que $y = \underbrace{\exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + u)}_{\exp(\beta_0+\beta_1 x_1+\beta_2 x_2+\beta_3 x_3)\exp(u)}$
- Aplicando $\mathbb{E}(\cdot|x)$,

$$\mathbb{E}(y|X) = \exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3)\mathbb{E}(\exp(u)|X)$$

Comparação de modelos

- Se $u \sim N(0, \sigma^2)$, $\mathbb{E}(\exp(u)|X) = \exp(\sigma^2/2)$

$$\hat{y} = \exp(\hat{\sigma}^2/2)\exp(\underbrace{\hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \hat{\beta}_3 x_3}_{\widehat{\log(y)}}) = \exp(\hat{\sigma}^2/2)\exp(\widehat{\log(y)})$$

Comparação de modelos

- Se $u \sim N(0, \sigma^2)$, $\mathbb{E}(\exp(u)|X) = \exp(\sigma^2/2)$

$$\hat{y} = \exp(\hat{\sigma}^2/2)\exp(\underbrace{\hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \hat{\beta}_3 x_3}_{\widehat{\log(y)}}) = \exp(\hat{\sigma}^2/2)\exp(\widehat{\log(y)})$$

- Se não tivermos Normalidade,

$$\hat{y} = \hat{\alpha_0}\exp(\widehat{\log(y)})$$

onde $\hat{\alpha_0}$ é uma estimativa de $\alpha_0 = \mathbb{E}(\exp(u)|X)$

Comparação de modelos

Como estimar α_0?

- $\hat{\alpha}_0 = n^{-1} \sum_{i=1}^{n} \exp(\hat{u}_i) \; (>1)$

Comparação de modelos

Como estimar α_0?

- $\hat{\alpha}_0 = n^{-1}\sum_{i=1}^{n} \exp(\hat{u}_i)$ (>1)
- $\hat{\alpha}_0 = \Big(\sum_{i=1}^{n} \hat{m}_i y_i\Big) \Big/ \Big(\sum_{i=1}^{n} \hat{m}_i^2\Big)$ onde $\hat{m}_i = \exp(\widehat{\log(y_i)})$

Comparação de modelos

Como estimar α_0?

- $\hat{\alpha}_0 = n^{-1} \sum_{i=1}^{n} \exp(\hat{u}_i)$ (>1)
- $\hat{\alpha}_0 = \left(\sum_{i=1}^{n} \hat{m}_i y_i\right) \Big/ \left(\sum_{i=1}^{n} \hat{m}_i^2\right)$ onde $\hat{m}_i = \exp(\widehat{\log(y_i)})$

Comparação de modelos

Como estimar α_0?

- $\hat{\alpha}_0 = n^{-1} \sum_{i=1}^{n} \exp(\hat{u}_i) \ (>1)$
- $\hat{\alpha}_0 = \Big(\sum_{i=1}^{n} \hat{m}_i y_i\Big) \Big/ \Big(\sum_{i=1}^{n} \hat{m}_i^2\Big)$ onde $\hat{m}_i = \exp(\widehat{\log(y_i)})$

Comparar modelos

- $R^2 = [Cor(y, \hat{y})]^2$
- $R^2 = 1 - \dfrac{\sum_{i=1}^{n}(y_i - \hat{\alpha}_0 \hat{m}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$

Comparação de modelos

$$\hat{\alpha}_0 = n^{-1}\sum_{i=1}^{n}\exp(\hat{u}_i), \quad \hat{\alpha}_0 = \frac{\left(\sum_{i=1}^{n}\hat{m}_i y_i\right)}{\left(\sum_{i=1}^{n}\hat{m}_i^2\right)}, \text{ onde } \hat{m}_i = \exp(\widehat{\log(y_i)})$$

```
modelo1 = lm(salary~sales+roe, data = ceosal1)
modelo2 = lm(log(salary)~log(sales)+roe, data = ceosal1)
uhat = residuals(modelo2)
alpha0_1 = mean(exp(uhat))
m = exp(fitted(modelo2))
alpha0_2 = sum(ceosal1$salary*m)/sum(m^2)
y = ceosal1$salary
```

Comparação de modelos

```
summary(modelo1)$r.squared
```

```
## [1] 0.02917169
```

```
yhat_1 = alpha0_1*m
yhat_2 = alpha0_2*m
c(cor(y, yhat_1)^2,cor(y, yhat_2)^2)
```

```
## [1] 0.04882569 0.04882569
```

```
1- sum((y-yhat_1)^2)/sum((y-mean(y))^2)
```

```
## [1] 0.04124148
```

```
1- sum((y-yhat_2)^2)/sum((y-mean(y))^2)
```

```
## [1] 0.04258757
```

Intervalos de Confiança / Previsão

Intervalos de Confiança / Previsão

- Seja $x_0 = (x_{0,1}, x_{0,2}, \dots, x_{0,k})$ os valores observados de x

Intervalos de Confiança / Previsão

- Seja $x_0 = (x_{0,1}, x_{0,2}, \ldots, x_{0,k})$ os valores observados de x
- O previsão de y dados os valores x_0, são obtidos como

$$\hat{y} = \mathbb{E}(y|x_0) = \hat{\beta}_0 + \hat{\beta}_1 x_{0,1} + \ldots + \hat{\beta}_k x_{0,k}$$

Intervalos de Confiança / Previsão

- Seja $x_0 = (x_{0,1}, x_{0,2}, \ldots, x_{0,k})$ os valores observados de x
- O previsão de y dados os valores x_0, são obtidos como

$$\hat{y} = \mathbb{E}(y|x_0) = \hat{\beta}_0 + \hat{\beta}_1 x_{0,1} + \ldots + \hat{\beta}_k x_{0,k}$$

- E se quisermos medir a incerteza desse valor previsto?

Intervalos de Confiança / Previsão

- Seja $x_0 = (x_{0,1}, x_{0,2}, \ldots, x_{0,k})$ os valores observados de x
- O previsão de y dados os valores x_0, são obtidos como

$$\hat{y} = \mathbb{E}(y|x_0) = \hat{\beta}_0 + \hat{\beta}_1 x_{0,1} + \ldots + \hat{\beta}_k x_{0,k}$$

- E se quisermos medir a incerteza desse valor previsto?
- Vamos construir **intervalos de confiança**

Intervalos de Confiança / Previsão

- Seja $x_0 = (x_{0,1}, x_{0,2}, \dots, x_{0,k})$ os valores observados de x
- O previsão de y dados os valores x_0, são obtidos como

$$\hat{y} = \mathbb{E}(y|x_0) = \hat{\beta}_0 + \hat{\beta}_1 x_{0,1} + \dots + \hat{\beta}_k x_{0,k}$$

- E se quisermos medir a incerteza desse valor previsto?
- Vamos construir **intervalos de confiança**

Intervalos de Confiança / Previsão

- Seja $x_0 = (x_{0,1}, x_{0,2}, \ldots, x_{0,k})$ os valores observados de x
- O previsão de y dados os valores x_0, são obtidos como

$$\hat{y} = \mathbb{E}(y|x_0) = \hat{\beta}_0 + \hat{\beta}_1 x_{0,1} + \ldots + \hat{\beta}_k x_{0,k}$$

- E se quisermos medir a incerteza desse valor previsto?
- Vamos construir **intervalos de confiança**

IC

Um IC assintótico 95% para $\mathbb{E}(y|x_0)$

$$\left(x_0\hat{\beta} - 1.96\sqrt{x_0\widehat{V}_{\hat{\beta}}x_0'} \quad ; \quad x_0\hat{\beta} + 1.96\sqrt{x_0\widehat{V}_{\hat{\beta}}x_0'}\right)$$

onde $\widehat{V}_{\hat{\beta}}$ é a matriz de variância-covariância de $\hat{\beta}$

Intervalos de Confiança / Previsão

```r
modelo = lm(colgpa~ sat + hsperc + hsize + I(hsize^2),
            data = gpa2)
round(coef(modelo),5)
```

```
## (Intercept)         sat      hsperc       hsize  I(hsize^2)
##     1.49265     0.00149    -0.01386    -0.06088     0.00546
```

```r
x0 = data.frame(sat = 1200, hsperc = 30, hsize = 5)
yhat = predict(modelo,newdata = x0)
V_beta = vcov(modelo)
```

Intervalos de Confiança / Previsão

```
x= matrix(c(1,1200,30,5,25),ncol=5)
x
```

```
##      [,1] [,2] [,3] [,4] [,5]
## [1,]    1 1200   30    5   25
```

```
c(yhat - 1.96*sqrt(x%*%V_beta%*%t(x)),
  yhat + 1.96*sqrt(x%*%V_beta%*%t(x)))
```

```
## [1] 2.661115 2.739036
```

Intervalos de Confiança / Previsão

```
x= matrix(c(1,1200,30,5,25),ncol=5)
x
```

```
##      [,1] [,2] [,3] [,4] [,5]
## [1,]    1 1200   30    5   25
```

```
c(yhat - 1.96*sqrt(x%*%V_beta%*%t(x)),
  yhat + 1.96*sqrt(x%*%V_beta%*%t(x)))
```

```
## [1] 2.661115 2.739036
```

```
predict(modelo,newdata = x0, interval = "confidence")
```

```
##        fit      lwr      upr
## 1 2.700075 2.661104 2.739047
```

Intervalos de Confiança / Previsão

- Um IC da média não é a mesma coisa que um IC de uma unidade em particular

Intervalos de Confiança / Previsão

- Um IC da média não é a mesma coisa que um IC de uma unidade em particular
- IC para uma unidade em particular são conhecidos como **intervalos de previsão** e precisam incluir uma outra fonte de variação, *a variância do erro não observado*.

Intervalos de Confiança / Previsão

- Um IC da média não é a mesma coisa que um IC de uma unidade em particular
- IC para uma unidade em particular são conhecidos como **intervalos de previsão** e precisam incluir uma outra fonte de variação, *a variância do erro não observado*.

Intervalos de Confiança / Previsão

- Um IC da média não é a mesma coisa que um IC de uma unidade em particular
- IC para uma unidade em particular são conhecidos como **intervalos de previsão** e precisam incluir uma outra fonte de variação, *a variância do erro não observado*.

Sabemos que

$$y_{n+1} = \beta_0 + \beta_1 x_{n+1,1} + \cdots + \beta_k x_{n+1,k} + u_{n+1}$$

Então o **erro de previsão** é dado por,

$$\hat{u}_{n+1} = y_{n+1} - \hat{y}_{n+1} = \underbrace{x_{n+1}\beta + u_{n+1}}_{y_{n+1}} - \underbrace{x_{n+1}\hat{\beta}}_{\hat{y}_{n+1}} = x_{n+1}(\beta - \hat{\beta}) + u_{n+1}$$

Intervalos de Confiança / Previsão

$$
\begin{aligned}
V(\hat{u}_{n+1}|x_{n+1}) &= E(\hat{u}_{n+1}^2|x_{n+1}) \\
&= E[(x_{n+1}(\beta - \hat{\beta}) + u_{n+1})^2|x_{n+1}] \\
&= E[x_{n+1}(\beta - \hat{\beta})(\beta - \hat{\beta})'x_{n+1}' + u_{n+1}^2 + 2x_{n+1}(\beta - \hat{\beta})\hat{u}_{n+1}|x_{n+1}] \\
&= x_{n+1}V(\hat{\beta}|x_{n+1})x_{n+1}' + \underbrace{E(u_{n+1}^2|x_{n+1})}_{\sigma^2}
\end{aligned}
\tag{2}
$$

IP

Um IP 95% (assumindo Normalidade) é dado por

$$
\left(x_0\hat{\beta} - 1.96\sqrt{x_0\widehat{V}_{\hat{\beta}}x_0' + \hat{\sigma}^2} \quad ; \quad x_0\hat{\beta} + 1.96\sqrt{x_0\widehat{V}_{\hat{\beta}}x_0' + \hat{\sigma}^2}\right)
$$

Intervalos de Confiança / Previsão

```
sigma2 = summary(modelo)$sigma^2
x= matrix(c(1,1200,30,5,25),ncol=5)
c(yhat - 1.96*sqrt(x%*%V_beta%*%t(x) + sigma2),
  yhat + 1.96*sqrt(x%*%V_beta%*%t(x) + sigma2))
```

```
## [1] 1.602051 3.798100
```

Intervalos de Confiança / Previsão

```r
sigma2 = summary(modelo)$sigma^2
x= matrix(c(1,1200,30,5,25),ncol=5)
c(yhat - 1.96*sqrt(x%*%V_beta%*%t(x) + sigma2),
  yhat + 1.96*sqrt(x%*%V_beta%*%t(x) + sigma2))
```

```
## [1] 1.602051 3.798100
```

```r
predict(modelo,newdata = x0, interval = "prediction")
```

```
##         fit      lwr      upr
## 1 2.700075 1.601749 3.798402
```

Leituras recomendadas

Leituras recomendadas

- Wooldridge, Jeffrey M. *Introdução à Econometria: Uma abordagem moderna*. (2016). Cengage Learning. – **Cap 6**
- Hansen, Bruce. *Econometrics*. (2020). – **Sec 7.14** e **Sec 7.15**