Modelos de Regressão Múltipla - Parte II

Erica Castilho Rodrigues

25 de Outubro de 2016

Modelo de Regressão Linear Múltipla

- Uma variável resposta Y e p variáveis explicativas

$$Y = \beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p + \epsilon$$

onde ϵ é o erro aleatório.

- Com uma amostra de n observações (indivíduos), podemos escrever:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_p X_{ip} \quad \text{para} \quad i = 1, 2, \ldots, n.$$

ou seja

$$\begin{aligned} Y_1 &= \beta_0 + \beta_1 X_{11} + \beta_2 X_{12} + \cdots + \beta_p X_{1p} \\ Y_2 &= \beta_0 + \beta_1 X_{21} + \beta_2 X_{22} + \cdots + \beta_p X_{2p} \\ &\vdots \\ Y_n &= \beta_0 + \beta_1 X_{n1} + \beta_2 X_{n2} + \cdots + \beta_p X_{np} \end{aligned}$$

Interpretação dos Coeficientes

Interpretação do β_j pra $j \geq 1$

Interpretação dos Coeficientes

Interpretação do β_j pra $j \geq 1$

- Matemos constantes todos X_i com exceção do X_j.
- β_j representa a variação no Y quando X_j aumenta em uma unidade.
- Se $\beta_j > 0$, Y aumenta.
- Se $\beta_j < 0$, Y diminui.
- Vejamos porque isso é verdade.

- Para $X_j = x_j$, temos que

$$E(Y|X_j = x_j) =$$

- Para $X_j = x_j$, temos que

  $$E(Y|X_j = x_j) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_j x_j + \cdots + \beta_p x_p\ .$$

- Para $X_j = x_j + 1$, temos

  $$E(Y|X_j = x_j{+}1) =$$

- Para $X_j = x_j$, temos que

$$E(Y|X_j = x_j) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_j x_j + \cdots + \beta_p x_p \, .$$

- Para $X_j = x_j + 1$, temos

$$E(Y|X_j = x_j{+}1) = \beta_0{+}\beta_1 x_1{+}\beta_2 x_2{+}\cdots{+}\beta_j(x_j{+}1){+}\cdots{+}\beta_p x_p.$$

- Assim, a diferença na $E(Y)$ ao se passar de $X_j = x_j$ para $X_j = x_j + 1$ é igual a

$$E(Y|X_j = x_j + 1) - E(Y|X_j = x_j) =$$

- Para $X_j = x_j$, temos que

$$E(Y|X_j = x_j) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_j x_j + \cdots + \beta_p x_p \,.$$

- Para $X_j = x_j + 1$, temos

$$E(Y|X_j = x_j+1) = \beta_0+\beta_1 x_1+\beta_2 x_2+\cdots+\beta_j(x_j+1)+\cdots+\beta_p x_p.$$

- Assim, a diferença na $E(Y)$ ao se passar de $X_j = x_j$ para $X_j = x_j + 1$ é igual a

$$E(Y|X_j = x_j + 1) - E(Y|X_j = x_j) =$$

$$(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_j(x_j + 1) + \cdots + \beta_p x_p) -$$

$$(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_j x_j + \cdots + \beta_p x_p) = \beta_j \,.$$

- Vimos que o modelo de regressão para n observações é dado por

$$
\begin{aligned}
Y_1 &= \beta_0 + \beta_1 X_{11} + \beta_2 X_{12} + \cdots + \beta_p X_{1p} \\
Y_2 &= \beta_0 + \beta_1 X_{21} + \beta_2 X_{22} + \cdots + \beta_p X_{2p} \\
&\vdots \\
Y_n &= \beta_0 + \beta_1 X_{n1} + \beta_2 X_{n2} + \cdots + \beta_p X_{np}
\end{aligned}
$$

- Podemos reescrever em notação matricial

- Vimos que o modelo de regressão para n observações é dado por

$$
\begin{aligned}
Y_1 &= \beta_0 + \beta_1 X_{11} + \beta_2 X_{12} + \cdots + \beta_p X_{1p} \\
Y_2 &= \beta_0 + \beta_1 X_{21} + \beta_2 X_{22} + \cdots + \beta_p X_{2p} \\
&\vdots \\
Y_n &= \beta_0 + \beta_1 X_{n1} + \beta_2 X_{n2} + \cdots + \beta_p X_{np}
\end{aligned}
$$

- Podemos reescrever em notação matricial

$$
\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & X_{11} & X_{12} & \dots & X_{1p} \\ 1 & X_{21} & X_{22} & \dots & X_{2p} \\ \vdots 1 & X_{n1} & X_{n2} & \dots & X_{np} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}
$$

$$
\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}
$$

Onde

- $\mathbf{Y}$ vetor

Onde

- $\mathbf{Y}$ vetor $n \times 1$ da variável resposta.
- $\mathbf{X}$ é uma matrix

Onde

- $\mathbf{Y}$ vetor $n \times 1$ da variável resposta.
- $\mathbf{X}$ é uma matrix $n \times (p+1)$ da variável explicativa.
- β é um vetor

Onde

- $\mathbf{Y}$ vetor $n \times 1$ da variável resposta.
- $\mathbf{X}$ é uma matrix $n \times (p+1)$ da variável explicativa.
- $\boldsymbol{\beta}$ é um vetor $(p+1) \times 1$ dos parâmetros.
- $\boldsymbol{\epsilon}$ é um vetor

Onde

- $\mathbf{Y}$ vetor $n \times 1$ da variável resposta.
- $\mathbf{X}$ é uma matrix $n \times (p+1)$ da variável explicativa.
- $\boldsymbol{\beta}$ é um vetor $(p+1) \times 1$ dos parâmetros.
- $\boldsymbol{\epsilon}$ é um vetor $n \times 1$ dos erros.

Suposições do Modelo

- Os erros
$$\epsilon_1, \epsilon_2, \ldots, \epsilon_n$$
tem média igual a zero e variância igual a σ^2.
- Isso implica que
$$E(Y_i|\mathbf{X}) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_p X_{ip}$$
$$Var(Y_i|\mathbf{X}) = \sigma^2 .$$
- Os erros
$$\epsilon_1, \epsilon_2, \ldots, \epsilon_n$$
são não correlacionados.
- Isso implica que os Y_i são não correlacionados.
- Os erros
$$\epsilon_1, \epsilon_2, \ldots, \epsilon_n$$
têm distribuição normal.
- Isso implica que os Y_i têm distribuição normal.

Resumindo...

- Supõe-se que os erros seguem a seguinte distribuição

$$\epsilon \sim$$

Resumindo...

- Supõe-se que os erros seguem a seguinte distribuição

$$\boldsymbol{\epsilon} \sim N_p(\mathbf{0}, \sigma^2 \mathbf{I})$$

onde

$$\epsilon = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \dots \\ \epsilon_n \end{bmatrix}$$

o vetor de médias é dado por

$$\mathbf{0} = \begin{bmatrix} 0 \\ 0 \\ \dots \\ 0 \end{bmatrix} \text{ e a matriz de covariâncias } \sigma^2 \mathbf{I} = \sigma^2 \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix}$$

Resumindo...

- Isso implica que

$$\mathbf{Y} \sim$$

Resumindo...

- Isso implica que

$$\mathbf{Y} \sim N_p(\mathbf{X}\beta, \sigma^2\mathbf{I})$$

onde

$$\mathbf{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} \quad \mathbf{X}\beta = \begin{bmatrix} 1 & X_{11} & X_{12} & \dots & X_{1p} \\ 1 & X_{21} & X_{22} & \dots & X_{2p} \\ \vdots 1 & X_{n1} & X_{n2} & \dots & X_{np} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix}$$

Estimação de β via mínimos quadrados

- Queremos encontrar β que minimiza a soma dos quadrados dos erros:

$$S = \sum_{i=1}^{n} \epsilon_i^2 = (\epsilon_1^2 + \epsilon_2^2 + \cdots + \epsilon_n^2)$$

$$= \begin{pmatrix} \epsilon_1 & \epsilon_2 & \dots & \epsilon_n \end{pmatrix} \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \dots \\ \epsilon_n \end{pmatrix} = \boldsymbol{\epsilon}'\boldsymbol{\epsilon}$$

onde

$$\boldsymbol{\epsilon} = \mathbf{Y} - \mathbf{X}\beta \, .$$

- Portanto, a soma de quadrados dos erros é dada por

$$\epsilon'\epsilon = (\mathbf{Y} - \mathbf{X}\beta)'(\mathbf{Y} - \mathbf{X}\beta)\,.$$

- Vimos que derivando com relação a β e igualando a zero, obtemos

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{Y})\,.$$

- Vamos usar o seguinte resultado.
- Se $\mathbf{A}$ é uma matriz de constantes e $\mathbf{Y}$ uma matriz de variáveis aleatórias

$$E(\mathbf{AY}) = \mathbf{A}E(\mathbf{Y})\,.$$

- Esse é um estimador não viesado

$E(\hat{\beta}) =$

- Portanto, a soma de quadrados dos erros é dada por

$$\epsilon'\epsilon = (\mathbf{Y} - \mathbf{X}\beta)'(\mathbf{Y} - \mathbf{X}\beta)\,.$$

- Vimos que derivando com relação a β e igualando a zero, obtemos

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{Y})\,.$$

- Vamos usar o seguinte resultado.
- Se $\mathbf{A}$ é uma matriz de constantes e $\mathbf{Y}$ uma matriz de variáveis aleatórias

$$E(\mathbf{AY}) = \mathbf{A}E(\mathbf{Y})\,.$$

- Esse é um estimador não viesado

$$E(\hat{\beta}) = E((\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{Y})) =$$

- Portanto, a soma de quadrados dos erros é dada por

$$\epsilon'\epsilon = (\mathbf{Y} - \mathbf{X}\beta)'(\mathbf{Y} - \mathbf{X}\beta)\,.$$

- Vimos que derivando com relação a β e igualando a zero, obtemos

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{Y})\,.$$

- Vamos usar o seguinte resultado.
- Se $\mathbf{A}$ é uma matriz de constantes e $\mathbf{Y}$ uma matriz de variáveis aleatórias

$$E(\mathbf{AY}) = \mathbf{A}E(\mathbf{Y})\,.$$

- Esse é um estimador não viesado

$$E(\hat{\beta}) = E((\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{Y})) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E(\mathbf{Y}) =$$

- Portanto, a soma de quadrados dos erros é dada por
$$\epsilon'\epsilon = (\mathbf{Y} - \mathbf{X}\beta)'(\mathbf{Y} - \mathbf{X}\beta) .$$
- Vimos que derivando com relação a β e igualando a zero, obtemos
$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{Y}) .$$
- Vamos usar o seguinte resultado.
- Se **A** é uma matriz de constantes e **Y** uma matriz de variáveis aleatórias
$$E(\mathbf{AY}) = \mathbf{A}E(\mathbf{Y}) .$$
- Esse é um estimador não viesado
$$E(\hat{\beta}) = E((\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{Y})) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E(\mathbf{Y}) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\beta$$

- Portanto, a soma de quadrados dos erros é dada por

$$\epsilon'\epsilon = (\mathbf{Y} - \mathbf{X}\beta)'(\mathbf{Y} - \mathbf{X}\beta)\,.$$

- Vimos que derivando com relação a β e igualando a zero, obtemos

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{Y})\,.$$

- Vamos usar o seguinte resultado.
- Se **A** é uma matriz de constantes e **Y** uma matriz de variáveis aleatórias

$$E(\mathbf{AY}) = \mathbf{A}E(\mathbf{Y})\,.$$

- Esse é um estimador não viesado

$$E(\hat{\beta}) = E((\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{Y})) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E(\mathbf{Y}) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\beta = \beta$$

pois

$$E(\mathbf{Y}) = \mathbf{X}\beta \quad (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X} = \mathbf{I}\,.$$

- Vimos ainda que

$$
\begin{aligned}
Var(\hat{\beta}) &= \left[\left(\mathbf{X}'\mathbf{X}\right)^{-1} \mathbf{X}' \right] Var(\mathbf{Y}) \left[\left(\mathbf{X}'\mathbf{X}\right)^{-1} \mathbf{X}' \right]' \\
&= \left[\left(\mathbf{X}'\mathbf{X}\right)^{-1} \mathbf{X}' \right] \sigma^2 \mathbf{I}_n \left[\left(\mathbf{X}'\mathbf{X}\right)^{-1} \mathbf{X}' \right]' \\
&= \sigma^2 \left[\left(\mathbf{X}'\mathbf{X}\right)^{-1} \mathbf{X}' \right] \left[\left(\mathbf{X}'\mathbf{X}\right)^{-1} \mathbf{X}' \right]' \\
&= \sigma^2 \left(\mathbf{X}'\mathbf{X}\right)^{-1} \mathbf{X}'\mathbf{X} \left(\mathbf{X}'\mathbf{X}\right)^{-1} \\
&= \sigma^2 \left(\mathbf{X}'\mathbf{X}\right)^{-1} .
\end{aligned}
$$

Tabela ANOVA

- O valor predito é dado por

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \cdots + \hat{\beta}_p X_p \, .$$

- O resíduo (estimativa do erro) é dado por

$$e_i = \hat{Y}_i - Y_i \, .$$

- Vimos que as somas de quadrados podem ser escritas na forma matricial

$$SQT =$$

- O valor predito é dado por

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \cdots + \hat{\beta}_p X_p \,.$$

- O resíduo (estimativa do erro) é dado por

$$e_i = \hat{Y}_i - Y_i \,.$$

- Vimos que as somas de quadrados podem ser escritas na forma matricial

$$SQT = \sum_i (Y_i - \overline{Y})^2 = \mathbf{Y}'\mathbf{Y} - n\overline{Y}^2$$

$$SQR =$$

- O valor predito é dado por

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \cdots + \hat{\beta}_p X_p \,.$$

- O resíduo (estimativa do erro) é dado por

$$e_i = \hat{Y}_i - Y_i \,.$$

- Vimos que as somas de quadrados podem ser escritas na forma matricial

$$SQT = \sum_i (Y_i - \overline{Y})^2 = \mathbf{Y}'\mathbf{Y} - n\overline{Y}^2$$

$$SQR = \sum_i (\hat{Y}_i - \overline{Y})^2 = \sum_i \hat{Y}_i^2 - n\overline{Y}^2 = \hat{\beta}'\mathbf{X}'\mathbf{Y} - n\overline{Y}^2$$

$$SQE =$$

- O valor predito é dado por

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \cdots + \hat{\beta}_p X_p \,.$$

- O resíduo (estimativa do erro) é dado por

$$e_i = \hat{Y}_i - Y_i \,.$$

- Vimos que as somas de quadrados podem ser escritas na forma matricial

$$SQT = \sum_i (Y_i - \overline{Y})^2 = \mathbf{Y}'\mathbf{Y} - n\overline{Y}^2$$

$$SQR = \sum_i (\hat{Y}_i - \overline{Y})^2 = \sum_i \hat{Y}_i^2 - n\overline{Y}^2 = \hat{\beta}'\mathbf{X}'\mathbf{Y} - n\overline{Y}^2$$

$$SQE = \sum_i (\hat{Y}_i - Y_i)^2 = \mathbf{Y}'\mathbf{Y} - \hat{\beta}'\mathbf{X}'\mathbf{Y}$$

- A Tabela ANOVA fica da seguinte forma

Fonte de Variação	Graus de Liberdade	Soma de Quadrados	Quadrado Médio
Regressão	p	$\hat{\beta}'\mathbf{X}'\mathbf{Y} - n\overline{Y}^2$	QMR=SQR/p
Resíduo	$n-p-1$	$\mathbf{Y}'\mathbf{Y} - \hat{\beta}'\mathbf{X}'\mathbf{Y}$	QME=SQE/(n-p-1)
Total	$n-1$	$\mathbf{Y}'\mathbf{Y} - n\overline{Y}^2$	

Tabela: Tabela ANOVA

Coeficiente de Determinação Ajustado

- O coeficiente de determinação pode ser calculado da mesma forma como na regressão linear simples:

$$R^2 =$$

- O coeficiente de determinação pode ser calculado da mesma forma como na regressão linear simples:

$$R^2 = \frac{SQR}{SQT} = \frac{\sum_i(\hat{Y}_i - \overline{Y})^2}{\sum_i(\hat{Y}_i - Y_i)^2} = \frac{\hat{\beta}'\mathbf{X}'\mathbf{Y} - n\overline{Y}^2}{\mathbf{Y}'\mathbf{Y} - n\overline{Y}^2} .$$

- Sempre que incluirmos mais uma variável no modelo o R^2 irá aumentar.
- Exemplos:
  - se com 2 pontos ajustamos uma reta $\Rightarrow R^2 =$

- O coeficiente de determinação pode ser calculado da mesma forma como na regressão linear simples:

$$R^2 = \frac{SQR}{SQT} = \frac{\sum_i(\hat{Y}_i - \overline{Y})^2}{\sum_i(\hat{Y}_i - Y_i)^2} = \frac{\hat{\beta}'\mathbf{X}'\mathbf{Y} - n\overline{Y}^2}{\mathbf{Y}'\mathbf{Y} - n\overline{Y}^2}.$$

- Sempre que incluirmos mais uma variável no modelo o R^2 irá aumentar.
- Exemplos:
  - se com 2 pontos ajustamos uma reta $\Rightarrow R^2 = 1$;
  - se com 3 pontos ajustamos um polinômio de grau 3 $\Rightarrow R^2 =$

- O coeficiente de determinação pode ser calculado da mesma forma como na regressão linear simples:

$$R^2 = \frac{SQR}{SQT} = \frac{\sum_i(\hat{Y}_i - \overline{Y})^2}{\sum_i(\hat{Y}_i - Y_i)^2} = \frac{\hat{\beta}'\mathbf{X}'\mathbf{Y} - n\overline{Y}^2}{\mathbf{Y}'\mathbf{Y} - n\overline{Y}^2}.$$

- Sempre que incluirmos mais uma variável no modelo o R^2 irá aumentar.
- Exemplos:
  - se com 2 pontos ajustamos uma reta $\Rightarrow R^2 = 1$;
  - se com 3 pontos ajustamos um polinômio de grau 3 $\Rightarrow R^2 = 1$.
- Porém um modelo com muitas variáveis não é o mais adequado.
- O modelo deve ser parcimonioso:
  - ajustar bem aos dados e ser o mais simples possível.

- Solução: penalizar a medida de ajuste pelo número de variáveis.
- Se o modelo ajusta bem aos dados, mas tem muitas variáveis essa medida não fica muito alta.
- Essa nova medida é chamada **Coeficiente de Determinação Ajustado.**
- O coeficiente de determinação é dado por

$$R^2 = \frac{SQR}{SQT} = 1 - \frac{SQE}{SQT}.$$

- O coeficiente de determinação ajustado é dado por

$$R^2_{ajustado} = 1 - \frac{SQE/(n-p-1)}{SQT/(n-1)} =$$

- Solução: penalizar a medida de ajuste pelo número de variáveis.
- Se o modelo ajusta bem aos dados, mas tem muitas variáveis essa medida não fica muito alta.
- Essa nova medida é chamada **Coeficiente de Determinação Ajustado.**
- O coeficiente de determinação é dado por

$$R^2 = \frac{SQR}{SQT} = 1 - \frac{SQE}{SQT}.$$

- O coeficiente de determinação ajustado é dado por

$$R^2_{ajustado} = 1 - \frac{SQE/(n-p-1)}{SQT/(n-1)} = 1 - \left(\frac{n-1}{n-p-1}\right)\left(\frac{SQE}{SQT}\right)$$

$$= 1 - (1 - R^2)\left(\frac{n-1}{n-p-1}\right).$$

Teste F da significância da Regressão

- Queremos testar as seguintes hipóteses:

$$H_0 : \beta_1 = \beta_2 = \cdots = \beta_p = 0, \text{ ou seja, nenhum dos } \beta_j \text{ é significante}$$

$$H_1 : \beta_j \neq 0 \text{ para pelo menos um } j \geq 1, \text{ ou seja, } \beta_j \text{ é significante}$$

- A estatística de teste é dada por

$$F = \frac{QMR}{QME} \sim F_{p,n-p-1} \text{ sob } H_0$$

Observação: não estamos testando o intercepto, queremos verificar quais variáveis são significativas.

- A hipótese nula deveser rejeitada para valores altos ou baixos de F?

- A hipótese nula deveser rejeitada para valores altos ou baixos de F? Altos.
- Se F é grande, SQR é grande e SQE é pequena, ou seja, o modelo explica bem os dados.
- Esse é um teste **unilateral**.
- Para um nível de significância α, a região crítica é dada por

$$F > F_\alpha$$

onde α é tal que $P(F_{p,n-p-1} > F_\alpha) = \alpha$.

- Se não rejeitamos H_0:
  - nenhum dos termos $\beta_j X_j$ é significante;
  - nenhuma variável preditora é significativa para explicar a resposta.
- Se rejeitamos H_0:
  - pelo menos um dos termos $\beta_j X_j$ é significante;
  - podemos verificar quais são significativos;
  - isso é feito através de testes t individuais.

Testes t individuais

- Para $j = 1, 2, \ldots, q$, podemos testar a signicância do coeficiente β_j na presença dos demais coeficientes no modelo:

$$H_0 : \beta_j = 0 \text{ na presença dos demais coeficientes };$$

$$H_1 : \beta_j \neq 0 \text{ na presença dos demais coeficientes.}$$

- A estatística de teste é dada por

$$t_j = \frac{\hat{\beta}_j}{\sqrt{Var(\hat{\beta}_j)}} \sim t_{n-p-1} \text{ sob } H_0 \,.$$

- A quantidade $\sqrt{Var(\hat{\beta}_j)}$ pode ser obtida a partir da matriz de variâncias e covariâncias estimada de $\hat{\beta}$.
- Essa matriz é dada por

$$Var(\hat{\beta}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1} .$$

- Não sabemos o valor de σ^2.
- Estimamos esse parâmetro usando

- A quantidade $\sqrt{Var(\hat{\beta}_j)}$ pode ser obtida a partir da matriz de variâncias e covariâncias estimada de $\hat{\boldsymbol{\beta}}$.
- Essa matriz é dada por

$$Var(\hat{\boldsymbol{\beta}}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1} .$$

- Não sabemos o valor de σ^2.
- Estimamos esse parâmetro usando S^2.
- A matriz de variâncias estimada fica da seguinte maneira

- A quantidade $\sqrt{Var(\hat{\beta}_j)}$ pode ser obtida a partir da matriz de variâncias e covariâncias estimada de $\hat{\boldsymbol{\beta}}$.
- Essa matriz é dada por

$$Var(\hat{\boldsymbol{\beta}}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1} \; .$$

- Não sabemos o valor de σ^2.
- Estimamos esse parâmetro usando S^2.
- A matriz de variâncias estimada fica da seguinte maneira

$$\hat{Var}(\hat{\boldsymbol{\beta}}) = S^2(\mathbf{X}'\mathbf{X})^{-1} \; .$$

- O j-ésimo elemento da diagonal nos fornece o valor de $Var(\hat{\beta}_j)$.

- Rejeitamos H_0 para valores altos ou baixos de t.
- Se $\hat{\beta}_j$ está muito distante de 0, H_0 deve ser regeitada.
- O teste é **bilateral**.
- A região crítica do teste é dada por

- Rejeitamos H_0 para valores altos ou baixos de t.
- Se $\hat{\beta}_j$ está muito distante de 0, H_0 deve ser regeitada.
- O teste é **bilateral**.
- A região crítica do teste é dada por

$$t > t_{\alpha/2} \text{ ou } t < -t_{\alpha/2}$$

onde $t_{\alpha/2}$ é tal que $P(t_{n-p-1} > t_{\alpha/2}) = \alpha/2$.

Exemplo:

- Um estudo realisado na Flórida analisa a ocorrência de doenças mentais.
- A variável de interesse é deficiência mental.
- Essa doença apresenta vários sintomas psiquiátricos, como:
  - ansiedade e depressão.
- A variável resposta é um score.
- Valores altos dessa variável indicam altos níveis de deficiência mental.

Exemplo: (continuação)

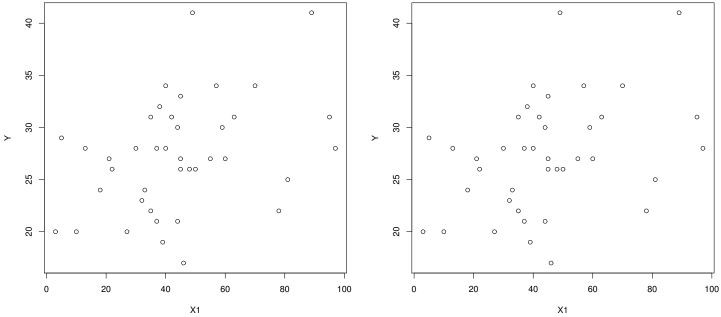
- Duas variáveis resposta são analisadas:
  - X_1= eventos marcantes;
  - X_2= status sócio-econômico (SES).
- A variável X_1 mede:
  - o número e a severidade de eventos marcantes na vida do indivíduo nos últimos três anos.
- Esses eventos podem ser muito severos, como uma morte na família, ou uma condenação criminal.
- Poucos severos, como a mudança para um novo trabalho, uma nova cidade.
- Um score alto para X_1 indica um número alto de eventos e enventos muito severos.

Exemplo: (continuação)

- A variável X_2 é um índice baseado em:
  - ocupação, salário e educação.
- Ela é medida em uma escala padronizada e varia de 0 a 100.
- Quanto maior o valores de X_2, maior o status sócio-econômico.
- Um conjunto de 40 adultos foram entrevistados.

Exemplo: (continuação)

- Gráficos de dispersão de Y em função de X_1 e X_2.



- Conclusão:

- Conclusão:
- As duas variáveis parecem ter relação linear com Y.

- Conclusão:
- As duas variáveis parecem ter relação linear com Y.
- A relação com X_1 é positiva.

- Conclusão:
- As duas variáveis parecem ter relação linear com Y.
- A relação com X_1 é positiva.
- A relação com X_2 é negativa.

Exemplo: (continuação)

- Vamos primeiro ajustar modelos de regressão linear simples.
- Y em função de X_1 apenas.
- Primeiro modelo ajustado:

$$Y = \beta_0 + \beta_1 X_1 + \epsilon_i \quad \epsilon_i \sim^{iid} N(0, \sigma^2)$$

```
Call:
lm(formula = Y ~ X1)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 23.30949    1.80675  12.901 1.85e-15 ***
X1           0.08983    0.03633   2.472    0.018 *
---

Residual standard error: 5.133 on 38 degrees of freedom
Multiple R-squared: 0.1385,     Adjusted R-squared: 0.1159
F-statistic: 6.112 on 1 and 38 DF,  p-value: 0.01802
```

Exemplo: (continuação)

- Conclusões:

Exemplo: (continuação)

- Conclusões:
- X_1 é significativa para explicar Y.

Exemplo: (continuação)

- Conclusões:
- X_1 é significativa para explicar Y.
- A relação entre Y e X_1 é positiva.
- Quanto maior o score de eventos marcantes, maior a deficiência mental.

Exemplo: (continuação)

- Conclusões:
- X_1 é significativa para explicar Y.
- A relação entre Y e X_1 é positiva.
- Quanto maior o score de eventos marcantes, maior a deficiência mental.
- O modelo não explica bem a variabilidade dos dados ($R^2 = 0.1385$).
- Devemos incluir mais variáveis explicativas.

Exemplo: (continuação)

- Y em função de X_2 apenas.
- Segundo modelo ajustado:

$$Y = \beta_0 + \beta_1 X_2 + \epsilon_i \quad \epsilon_i \sim^{iid} N(0, \sigma^2)$$

```
Call:
lm(formula = Y ~ X2)


Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 32.17201    1.98765  16.186   <2e-16 ***
X2          -0.08608    0.03213  -2.679   0.0109 *

Residual standard error: 5.072 on 38 degrees of freedom
Multiple R-squared: 0.1589,     Adjusted R-squared: 0.1367
F-statistic: 7.177 on 1 and 38 DF,  p-value: 0.01085
```

Exemplo: (continuação)

- Conclusões:

Exemplo: (continuação)

- Conclusões:
- X_2 é significativa para explicar Y.

Exemplo: (continuação)

- Conclusões:
- X_2 é significativa para explicar Y.
- A relação entre Y e X_2 é negativa.
- Quanto menor o status sócio-econômico, maior a deficiência mental.

Exemplo: (continuação)

- Conclusões:
- X_2 é significativa para explicar Y.
- A relação entre Y e X_2 é negativa.
- Quanto menor o status sócio-econômico, maior a deficiência mental.
- O modelo não explica bem a variabilidade dos dados ($R^2 = 0.1589$).
- Devemos incluir mais variáveis explicativas.

Exemplo: (continuação)

- Vamos ajustar agora o modelo com as duas variáveis.

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon_i \quad \epsilon_i \sim^{iid} N(0, \sigma^2) \; .$$

- Vamos realizar primeiramente o teste F.
- Queremos testar as seguintes hipóteses:

Exemplo: (continuação)

- Vamos ajustar agora o modelo com as duas variáveis.

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon_i \quad \epsilon_i \sim^{iid} N(0, \sigma^2) \ .$$

- Vamos realizar primeiramente o teste F.
- Queremos testar as seguintes hipóteses:

$$H_0 : \beta_1 = \beta_2 = \cdots = \beta_p = 0, \text{ ou seja, nenhum dos } \beta_j \text{ é significante}$$

$$H_1 : \beta_j \neq 0 \text{ para pelo menos um } j \geq 1, \text{ ou seja, } \beta_j \text{ é significante}$$

- A estatística de teste é dada por

$$F = \frac{QMR}{QME} \sim F_{p,n-p-1} \text{ sob } H_0 \ .$$

Exemplo: (continuação)

- Vimos que

$$SQR = \hat{\beta}'\mathbf{X}'\mathbf{Y} - n\overline{Y}^2 = 394.2384$$

logo

$$QMR = \frac{SQR}{p} = \frac{394.2384}{2} = 197.1192$$

- Além disso

$$SQE = \mathbf{Y}'\mathbf{Y} - \hat{\beta}'\mathbf{X}'\mathbf{Y} = 768.1616$$

e

$$QME = \frac{SME}{n-p-1} = \frac{768.1616}{37} = 20.76112\,.$$

Exemplo: (continuação)

- A tabela ANOVA fica da seguinte forma:

Fonte de Variação	Graus de Liberdade	Soma de Quadrados	Quadrado Médio
Regressão	$p = 2$	$SQR = 394.2384$	QMR=394.2384/2
Resíduo	$n - p - 1 = 37$	$SQE = 768.1616$	QME=768.1616/(37)
Total	$n - 1 = 39$	$SQT = 1162.4$	

Tabela: Tabela ANOVA

Exemplo: (continuação)

- A estatística F é dada por

$$F = \frac{QMR}{QME} = \frac{197.1192}{20.76112} = 9.494632\,.$$

- Sob H_0 temos que

$$F \sim F_{2,37}$$

- Fixando $\alpha = 0.05$ temos que $F_{2,37;0.05} \approx F_{2,37;0.05} = 3.232$.
- A região crítica é dada por

$$F > 3.232\,.$$

- Qual a conclusão?

Exemplo: (continuação)

- A estatística F é dada por

$$F = \frac{QMR}{QME} = \frac{197.1192}{20.76112} = 9.494632\,.$$

- Sob H_0 temos que

$$F \sim F_{2,37}$$

- Fixando $\alpha = 0.05$ temos que $F_{2,37;0.05} \approx F_{2,37;0.05} = 3.232$.
- A região crítica é dada por

$$F > 3.232\,.$$

- Qual a conclusão? Como $F_{obs} = 9.4946 > 3.232$, rejeitamos H_0.
- O que isso significa?

Exemplo: (continuação)

- A estatística F é dada por

$$F = \frac{QMR}{QME} = \frac{197.1192}{20.76112} = 9.494632\,.$$

- Sob H_0 temos que

$$F \sim F_{2,37}$$

- Fixando $\alpha = 0.05$ temos que $F_{2,37;0.05} \approx F_{2,37;0.05} = 3.232$.
- A região crítica é dada por

$$F > 3.232\,.$$

- Qual a conclusão? Como $F_{obs} = 9.4946 > 3.232$, rejeitamos H_0.
- O que isso significa? Com 5% de significância, podemos dizer que pelo uma das variáveis (X_1 ou X_2) é significativa para explicar Y.

Exemplo: (continuação)

- Vamos agora fazer os testes t individuais.
- Para $j = 1, 2$, queremos testar

$$H_0 : \beta_j = 0 \text{ na presença dos demais coeficientes ;}$$

$$H_1 : \beta_j \neq 0 \text{ na presença dos demais coeficientes.}$$

- A estatística de teste é dada por

$$t_j = \frac{\hat{\beta}_j}{\sqrt{Var(\hat{\beta}_j)}} \sim t_{n-p-1} \text{ sob } H_0 \ .$$

Exemplo: (continuação)

- Precisamos encontrar a variância estimada $\hat{\beta}_1$ e $\hat{\beta}_2$.
- Essa variância é obtida através da matriz

$$Var(\hat{\beta}) = s^2(\mathbf{X}'\mathbf{X})^{-1} = \begin{bmatrix} 4.72721373 & -0.0403235231 & -0.0426999331 \\ -0.04032352 & 0.0010562090 & -0.0001165824 \\ -0.04269993 & -0.0001165824 & 0.0008459206 \end{bmatrix}.$$

- Temos então que

$$Var(\hat{\beta}_0) =$$

Exemplo: (continuação)

- Precisamos encontrar a variância estimada $\hat{\beta}_1$ e $\hat{\beta}_2$.
- Essa variância é obtida através da matriz

$$Var(\hat{\beta}) = s^2(\mathbf{X}'\mathbf{X})^{-1} = \begin{bmatrix} 4.72721373 & -0.0403235231 & -0.0426999331 \\ -0.04032352 & 0.0010562090 & -0.0001165824 \\ -0.04269993 & -0.0001165824 & 0.0008459206 \end{bmatrix}.$$

- Temos então que

$$Var(\hat{\beta}_0) = 4.7272 \quad Var(\hat{\beta}_1) =$$

Exemplo: (continuação)

- Precisamos encontrar a variância estimada $\hat{\beta}_1$ e $\hat{\beta}_2$.
- Essa variância é obtida através da matriz

$$Var(\hat{\boldsymbol{\beta}}) = s^2(\mathbf{X}'\mathbf{X})^{-1} = \begin{bmatrix} 4.72721373 & -0.0403235231 & -0.0426999331 \\ -0.04032352 & 0.0010562090 & -0.0001165824 \\ -0.04269993 & -0.0001165824 & 0.0008459206 \end{bmatrix}.$$

- Temos então que

$$Var(\hat{\beta}_0) = 4.7272 \quad Var(\hat{\beta}_1) = 0.0011 \quad Var(\hat{\beta}_2) = 0.0008$$

Exemplo: (continuação)

- Vamos considerar primeiro a variável X_1.
- Queremos testar

Exemplo: (continuação)

- Vamos considerar primeiro a variável X_1.
- Queremos testar

$$H_0 : \beta_1 = 0 \quad H_1 : \beta_1 \neq 0 \, .$$

- A estatística de teste é dada por

$$t_{obs}^{(1)} = \frac{\hat{\beta}_1}{\sqrt{Var(\hat{\beta}_1)}} = \frac{0.10326}{\sqrt{0.0011}} = 3.177$$

sob H_0 temos que

$$t_{obs}^{(1)} \sim t_{37} \text{ sob } H_0 \, .$$

- Fixando $\alpha = 0.05$ temos que $t_{37;0.025} \approx t_{40;0.025} = 2.021$.

Exemplo: (continuação)

- A região crítica é dada por

Exemplo: (continuação)

- A região crítica é dada por

$$t_{obs}^{(1)} > 2.021 \quad \text{ou} \quad t_{obs}^{(1)} < -2.021 \,.$$

- Rejeitamos ou não H_0?

Exemplo: (continuação)

- A região crítica é dada por

$$t_{obs}^{(1)} > 2.021 \quad \text{ou} \quad t_{obs}^{(1)} < -2.021\ .$$

- Rejeitamos ou não H_0? Rejeitamos, pois $t_{obs}^{(1)} = 3.177 > 2.021$.
- Conclusão:

Exemplo: (continuação)

- A região crítica é dada por

$$t_{obs}^{(1)} > 2.021 \quad \text{ou} \quad t_{obs}^{(1)} < -2.021\,.$$

- Rejeitamos ou não H_0? Rejeitamos, pois $t_{obs}^{(1)} = 3.177 > 2.021$.
- Conclusão: Com 5% de significância, há evidências de que X_1 é significa para explicar Y.

Exemplo: (continuação)

- Vamos considerar primeiro a variável X_2.
- Queremos testar

Exemplo: (continuação)

- Vamos considerar primeiro a variável X_2.
- Queremos testar

$$H_0 : \beta_2 = 0 \quad H_1 : \beta_2 \neq 0 \,.$$

- A estatística de teste é dada por

$$t_{obs}^{(2)} = \frac{\hat{\beta}_2}{\sqrt{Var(\hat{\beta}_2)}} = \frac{-0.09748}{\sqrt{0.0008}} = -3.351$$

sob H_0 temos que

$$t_{obs}^{(2)} \sim t_{37} \text{ sob } H_0 \,.$$

- Fixando $\alpha = 0.05$ temos que $t_{37;0.025} \approx t_{40;0.025} = 2.021$.

Exemplo: (continuação)

- A região crítica é dada por

Exemplo: (continuação)

- A região crítica é dada por

$$t_{obs}^{(1)} > 2.021 \quad \text{ou} \quad t_{obs}^{(1)} < -2.021\,.$$

- Rejeitamos ou não H_0?

Exemplo: (continuação)

- A região crítica é dada por

$$t_{obs}^{(1)} > 2.021 \quad \text{ou} \quad t_{obs}^{(1)} < -2.021 \, .$$

- Rejeitamos ou não H_0? Rejeitamos, pois $t_{obs}^{(1)} = -3.351 < -2.021$.
- Conclusão:

Exemplo: (continuação)

- A região crítica é dada por

$$t_{obs}^{(1)} > 2.021 \quad \text{ou} \quad t_{obs}^{(1)} < -2.021 \,.$$

- Rejeitamos ou não H_0? Rejeitamos, pois $t_{obs}^{(1)} = -3.351 < -2.021$.
- Conclusão: Com 5% de significância, há evidências de que X_2 é significa para explicar Y

Exemplo: (continuação)

- Vamos agora encontrar o Coeficiente de Determinação ajustado.
- O Coeficiente de Determinação é dado por

$$R^2 = \frac{SQR}{SQT} = \frac{394.2384}{1162.4} = 0.3392\,.$$

- O coeficiente de determinação ajustado é dado por

$$R^2_{ajustado} = 1 - (1 - R^2)\left(\frac{n-1}{n-p-1}\right)$$

$$= 1 - (1 - 0.3392)\left(\frac{40-1}{40-2-1}\right) = 0.3034\,.$$

- Interpretação:

Exemplo: (continuação)

- Vamos agora encontrar o Coeficiente de Determinação ajustado.
- O Coeficiente de Determinação é dado por

$$R^2 = \frac{SQR}{SQT} = \frac{394.2384}{1162.4} = 0.3392\,.$$

- O coeficiente de determinação ajustado é dado por

$$R^2_{ajustado} = 1 - (1 - R^2)\left(\frac{n-1}{n-p-1}\right)$$

$$= 1 - (1 - 0.3392)\left(\frac{40-1}{40-2-1}\right) = 0.3034\,.$$

- Interpretação: 30.34% da variabilidade dos dados pode ser explicada pelo modelo de regressão.
- Esse modelo não explica bem os dados.
- Talvez devemos incluir mais variáveis.

Exemplo: (continuação)

- O saída do R nos fornece todos os resultado obtidos até aqui:

```
Call:
lm(formula = Y ~ X1 + X2)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 28.22981    2.17422  12.984 2.38e-15 ***
X1           0.10326    0.03250   3.177  0.00300 **
X2          -0.09748    0.02908  -3.351  0.00186 **
---

Residual standard error: 4.556 on 37 degrees of freedom
Multiple R-squared: 0.3392,     Adjusted R-squared: 0.3034
F-statistic: 9.495 on 2 and 37 DF,  p-value: 0.0004697
```

Exemplo: (continuação)

```
Analysis of Variance Table

Response: Y
          Df Sum Sq Mean Sq F value   Pr(>F)
X1         1 161.05 161.048  7.7572 0.008385 **
X2         1 233.19 233.190 11.2320 0.001863 **
Residuals 37 768.16  20.761
---
```