Predição

Tiago M. Magalhães

Departamento de Estatística - ICE-UFJF

Juiz de Fora, 12 de maio de 2025

Roteiro

Roteiro

Modelo de regressão linear

Suponham que $Y_1, Y_2, \ldots, Y_n$ tais que

$$Y_\ell = \boldsymbol{x}_\ell^\top \boldsymbol{\beta} + \varepsilon_\ell, \ \ell = 1, 2, \ldots, n, \tag{1}$$

Modelo de regressão linear

Suponham que $Y_1, Y_2, \ldots, Y_n$ tais que

$$Y_\ell = \boldsymbol{x}_\ell^\top \boldsymbol{\beta} + \varepsilon_\ell, \ \ell = 1, 2, \ldots, n, \tag{1}$$

em que $\boldsymbol{x}_\ell = (x_{\ell 1}, x_{\ell 2}, \ldots, x_{\ell p})^\top$ é conhecido, $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_p)^\top$ é um vetor de parâmetros desconhecidos a serem estimados, $\varepsilon_\ell \sim \mathcal{N}(0, \sigma^2)$, $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$ são variáveis aleatórias independentes e com a mesma variância σ^2, também desconhecida, a ser estimada.

Modelo de regressão linear

Suponham que $Y_1, Y_2, \ldots, Y_n$ tais que

$$Y_\ell = \boldsymbol{x}_\ell^\top \boldsymbol{\beta} + \varepsilon_\ell, \ \ell = 1, 2, \ldots, n, \tag{1}$$

em que $\boldsymbol{x}_\ell = (x_{\ell 1}, x_{\ell 2}, \ldots, x_{\ell p})^\top$ é conhecido, $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_p)^\top$ é um vetor de parâmetros desconhecidos a serem estimados, $\varepsilon_\ell \sim \mathcal{N}(0, \sigma^2)$, $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$ são variáveis aleatórias independentes e com a mesma variância σ^2, também desconhecida, a ser estimada.

Forma matricial

A Equação (1) é o que nós definimos como **modelo de regressão normal linear** (MNL), podendo ser escrita de forma matricial, da seguinte forma:

$$\boldsymbol{Y} = \boldsymbol{X}\beta + \varepsilon, \tag{2}$$

Forma matricial

A Equação (1) é o que nós definimos como **modelo de regressão normal linear** (MNL), podendo ser escrita de forma matricial, da seguinte forma:

$$\boldsymbol{Y} = \boldsymbol{X}\beta + \boldsymbol{\varepsilon}, \tag{2}$$

em que $\boldsymbol{Y} = (Y_1, Y_2, \ldots, Y_n)^\top$ e $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n)^\top$, $\boldsymbol{\varepsilon} \sim \mathcal{N}_n(\mathbf{0}, \sigma^2 \boldsymbol{I}_n)$, sendo $\mathbf{0}$ o vetor nulo de dimensão n, $\boldsymbol{I}_n$ a matriz identidade de ordem n e $\boldsymbol{X} = (\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n)^\top$, a matriz de planejamento.

Forma matricial

A Equação (1) é o que nós definimos como **modelo de regressão normal linear** (MNL), podendo ser escrita de forma matricial, da seguinte forma:

$$\boldsymbol{Y} = \boldsymbol{X}\beta + \boldsymbol{\varepsilon}, \tag{2}$$

em que $\boldsymbol{Y} = (Y_1, Y_2, \ldots, Y_n)^\top$ e $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n)^\top$, $\boldsymbol{\varepsilon} \sim \mathcal{N}_n(\mathbf{0}, \sigma^2 \boldsymbol{I}_n)$, sendo $\mathbf{0}$ o vetor nulo de dimensão n, $\boldsymbol{I}_n$ a matriz identidade de ordem n e $\boldsymbol{X} = (\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n)^\top$, a matriz de planejamento.

Método de máxima verossimilhança

O **estimador de máxima verossimilhança** (EMV) de $\boldsymbol{\beta}$ e σ^2 são dados, respectivamente, por:

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{Y}, \tag{3}$$

$$\hat{\sigma}^2_{\mathrm{MV}} = \frac{1}{n} \left(\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}} \right)^\top \left(\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}} \right).$$

Método de máxima verossimilhança

O **estimador de máxima verossimilhança** (EMV) de β e σ^2 são dados, respectivamente, por:

$$\widehat{\beta} = (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{Y}, \tag{3}$$

$$\hat{\sigma}^2_{\mathrm{MV}} = \frac{1}{n} \left(\boldsymbol{Y} - \boldsymbol{X}\widehat{\beta} \right)^\top \left(\boldsymbol{Y} - \boldsymbol{X}\widehat{\beta} \right).$$

Observação: o EMV de σ^2 também pode ser escrito como função do quadrado médio do resíduo (QMRes).

Método de máxima verossimilhança

O **estimador de máxima verossimilhança** (EMV) de $\boldsymbol{\beta}$ e σ^2 são dados, respectivamente, por:

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{Y}, \tag{3}$$

$$\hat{\sigma}^2_{\mathrm{MV}} = \frac{1}{n} \left(\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}} \right)^\top \left(\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}} \right).$$

Observação: o EMV de σ^2 também pode ser escrito como função do quadrado médio do resíduo (QMRes).

Método de máxima verossimilhança

Sob as condições de regularidades, nós temos que:

$$\widehat{\beta} \sim \mathcal{N}_p \left\{ \beta, \sigma^2 \left(\boldsymbol{X}^\top \boldsymbol{X} \right)^{-1} \right\}, \tag{4}$$

$$\hat{\sigma}^2_{\mathrm{MV}} \sim \mathcal{N} \left\{ \sigma^2, \frac{2(\sigma^2)^2}{n} \right\},$$

Método de máxima verossimilhança

Sob as condições de regularidades, nós temos que:

$$\widehat{\boldsymbol{\beta}} \sim \mathcal{N}_p \left\{ \boldsymbol{\beta}, \sigma^2 \left(\boldsymbol{X}^\top \boldsymbol{X} \right)^{-1} \right\}, \tag{4}$$

$$\hat{\sigma}^2_{\mathrm{MV}} \sim \mathcal{N} \left\{ \sigma^2, \frac{2(\sigma^2)^2}{n} \right\},$$

quando o tamanho de amostra é grande, adicionalmente, $\widehat{\boldsymbol{\beta}}$ e $\hat{\sigma}^2_{\mathrm{MV}}$ são ortogonais.

Método de máxima verossimilhança

Sob as condições de regularidades, nós temos que:

$$\widehat{\boldsymbol{\beta}} \sim \mathcal{N}_p \left\{ \boldsymbol{\beta}, \sigma^2 \left(\boldsymbol{X}^\top \boldsymbol{X} \right)^{-1} \right\}, \tag{4}$$

$$\hat{\sigma}^2_{\mathrm{MV}} \sim \mathcal{N} \left\{ \sigma^2, \frac{2(\sigma^2)^2}{n} \right\},$$

quando o tamanho de amostra é grande, adicionalmente, $\widehat{\boldsymbol{\beta}}$ e $\hat{\sigma}^2_{\mathrm{MV}}$ são ortogonais.

Roteiro

Predição de novas observações

Uma importante aplicação dos modelos de regressão é a predição de uma nova observação $Y_{h(\text{nova})}$ para um especifico vetor $\boldsymbol{x}$. Se $\boldsymbol{x}_h$ é o vetor das covariáveis de interesse, então

$$\hat{Y}_{h(\text{nova})} = \boldsymbol{x}_h^\top \widehat{\boldsymbol{\beta}},$$

Predição de novas observações

Uma importante aplicação dos modelos de regressão é a predição de uma nova observação $Y_{h(\text{nova})}$ para um especifico vetor $\boldsymbol{x}$. Se $\boldsymbol{x}_h$ é o vetor das covariáveis de interesse, então

$$\hat{Y}_{h(\text{nova})} = \boldsymbol{x}_h^\top \widehat{\boldsymbol{\beta}},$$

é a estimativa pontual para um novo valor da resposta $Y_{h(\text{nova})}$.

Predição de novas observações

Uma importante aplicação dos modelos de regressão é a predição de uma nova observação $Y_{h(\text{nova})}$ para um especifico vetor $\boldsymbol{x}$. Se $\boldsymbol{x}_h$ é o vetor das covariáveis de interesse, então

$$\hat{Y}_{h(\text{nova})} = \boldsymbol{x}_h^\top \widehat{\boldsymbol{\beta}},$$

é a estimativa pontual para um novo valor da resposta $Y_{h(\text{nova})}$.

Predição de novas observações

Considere a obtenção de uma estimativa intervalar da observação futura $Y_{h(\text{nova})}$. O intervalo de confiança (IC) para a resposta média em $\boldsymbol{x}_h$ não é apropriada para este problema.

Pois ele é intervalo para estimar a média de Y e não uma indicação probabilística sobre as observações futuras dessa distribuição.

Predição de novas observações

Considere a obtenção de uma estimativa intervalar da observação futura $Y_{h(\text{nova})}$. O intervalo de confiança (IC) para a resposta média em $\boldsymbol{x}_h$ não é apropriada para este problema.

Pois ele é intervalo para estimar a média de Y e não uma indicação probabilística sobre as observações futuras dessa distribuição.

Predição de novas observações

Seja a variável aleatória (VA):

$$\begin{aligned}\psi &= Y_h - \hat{Y}_h \\ &= Y_h - \boldsymbol{x}_h^\top \widehat{\boldsymbol{\beta}}.\end{aligned}$$

Predição de novas observações

Seja a variável aleatória (VA):

$$\begin{aligned}\psi &= Y_h - \hat{Y}_h \\ &= Y_h - \boldsymbol{x}_h^\top \widehat{\boldsymbol{\beta}}.\end{aligned}$$

Predição de novas observações

Nós temos que,

$$\mathbb{E}(\psi) = \mathbb{E}(Y_h - \boldsymbol{x}_h^\top \widehat{\beta}) = \mathbb{E}(Y_h) - \boldsymbol{x}_h^\top \mathbb{E}(\widehat{\beta}) = \boldsymbol{x}_h^\top \beta - \boldsymbol{x}_h^\top \beta = 0,$$

$$\begin{aligned}
\mathrm{Var}(\psi) &= \mathrm{Var}(Y_h - \boldsymbol{x}_h^\top \widehat{\beta}) \overset{*}{=} \mathrm{Var}(Y_h) + \boldsymbol{x}_h^\top \mathrm{Var}(\widehat{\beta}) \boldsymbol{x}_h \\
&= \sigma^2 + \sigma^2 \boldsymbol{x}_h^\top \left(\boldsymbol{X}^\top \boldsymbol{X}\right)^{-1} \boldsymbol{x}_h \\
&= \sigma^2 \left[1 + \boldsymbol{x}_h^\top \left(\boldsymbol{X}^\top \boldsymbol{X}\right)^{-1} \boldsymbol{x}_h\right].
\end{aligned}$$

Predição de novas observações

Nós temos que,

$$\mathbb{E}(\psi) = \mathbb{E}(Y_h - \boldsymbol{x}_h^\top \widehat{\boldsymbol{\beta}}) = \mathbb{E}(Y_h) - \boldsymbol{x}_h^\top \mathbb{E}(\widehat{\boldsymbol{\beta}}) = \boldsymbol{x}_h^\top \boldsymbol{\beta} - \boldsymbol{x}_h^\top \boldsymbol{\beta} = 0,$$

$$\begin{aligned}
\mathrm{Var}(\psi) &= \mathrm{Var}(Y_h - \boldsymbol{x}_h^\top \widehat{\boldsymbol{\beta}}) \overset{*}{=} \mathrm{Var}(Y_h) + \boldsymbol{x}_h^\top \mathrm{Var}(\widehat{\boldsymbol{\beta}}) \boldsymbol{x}_h \\
&= \sigma^2 + \sigma^2 \boldsymbol{x}_h^\top \left(\boldsymbol{X}^\top \boldsymbol{X}\right)^{-1} \boldsymbol{x}_h \\
&= \sigma^2 \left[1 + \boldsymbol{x}_h^\top \left(\boldsymbol{X}^\top \boldsymbol{X}\right)^{-1} \boldsymbol{x}_h\right].
\end{aligned}$$

Predição de novas observações

$*$: a observação futura Y_h é independente de $\hat{Y}_h$. Pois, Y_h é "algo" (uma VA) não observado, enquanto $\hat{Y}_h$ depende de $\widehat{\boldsymbol{\beta}}$, uma função de valores já observados.

Como Y_h e $\hat{Y}_h$ têm distribuição normal,

$$Y_h - \hat{Y}_h \sim \mathcal{N}\left\{0, \sigma^2\left[1 + \boldsymbol{x}_h^\top \left(\boldsymbol{X}^\top \boldsymbol{X}\right)^{-1} \boldsymbol{x}_h\right]\right\}. \tag{5}$$

Predição de novas observações

$*$: a observação futura Y_h é independente de $\hat{Y}_h$. Pois, Y_h é "algo" (uma VA) não observado, enquanto $\hat{Y}_h$ depende de $\widehat{\boldsymbol{\beta}}$, uma função de valores já observados.

Como Y_h e $\hat{Y}_h$ têm distribuição normal,

$$Y_h - \hat{Y}_h \sim \mathcal{N}\left\{0,\, \sigma^2 \left[1 + \boldsymbol{x}_h^\top \left(\boldsymbol{X}^\top \boldsymbol{X}\right)^{-1} \boldsymbol{x}_h\right]\right\}. \tag{5}$$

Intervalo de confiança

Por (5), nós temos que,

$$\frac{Y_h - \hat{Y}_h}{\sqrt{\sigma^2 \left[1 + \boldsymbol{x}_h^\top \left(\boldsymbol{X}^\top \boldsymbol{X}\right)^{-1} \boldsymbol{x}_h\right]}} \sim \mathcal{N}(0,1), \tag{6}$$

Intervalo de confiança

Por (5), nós temos que,

$$\frac{Y_h - \hat{Y}_h}{\sqrt{\sigma^2 \left[1 + \boldsymbol{x}_h^\top \left(\boldsymbol{X}^\top \boldsymbol{X}\right)^{-1} \boldsymbol{x}_h\right]}} \sim \mathcal{N}(0,1), \tag{6}$$

Intervalo de confiança

Como a estatística em (6) depende de um parâmetro desconhecido σ^2, nós precisamos substituí-lo por um estimativa não viesada, no caso, pelo QMRes, e, assim, nós temos que

$$\frac{Y_h - \hat{Y}_h}{\sqrt{\text{QMRes}\left[1 + \boldsymbol{x}_h^\top \left(\boldsymbol{X}^\top \boldsymbol{X}\right)^{-1} \boldsymbol{x}_h\right]}} \sim t(n-p), \tag{7}$$

Intervalo de confiança

Como a estatística em (6) depende de um parâmetro desconhecido σ^2, nós precisamos substituí-lo por um estimativa não viesada, no caso, pelo QMRes, e, assim, nós temos que

$$\frac{Y_h - \hat{Y}_h}{\sqrt{\text{QMRes}\left[1 + \boldsymbol{x}_h^\top \left(\boldsymbol{X}^\top \boldsymbol{X}\right)^{-1} \boldsymbol{x}_h\right]}} \sim t(n-p), \tag{7}$$

em que $t(n-p)$ é a distribuição t de Student (Gosset "Student", 1908), com $n-p$ graus de liberdade.

Intervalo de confiança

Como a estatística em (6) depende de um parâmetro desconhecido σ^2, nós precisamos substituí-lo por um estimativa não viesada, no caso, pelo QMRes, e, assim, nós temos que

$$\frac{Y_h - \hat{Y}_h}{\sqrt{\text{QMRes}\left[1 + \boldsymbol{x}_h^\top \left(\boldsymbol{X}^\top \boldsymbol{X}\right)^{-1} \boldsymbol{x}_h\right]}} \sim t(n-p), \tag{7}$$

em que $t(n-p)$ é a distribuição t de Student (Gosset "Student", 1908), com $n-p$ graus de liberdade.

Intervalo de predição para β_m

Por (7), o intervalo de predição (IP) para $Y_{h(\text{nova})}$, com coeficiente de confiança $(1-\alpha)$ é dado por

$$\text{IC}(Y_{h(\text{nova})}; 1-\alpha) = \hat{Y}_h \mp t(1-\alpha/2; n-p)\sqrt{\text{QMRes}\left[1 + \boldsymbol{x}_h^\top \left(\boldsymbol{X}^\top \boldsymbol{X}\right)^{-1} \boldsymbol{x}_h\right]},$$

Intervalo de predição para β_m

Por (7), o intervalo de predição (IP) para $Y_{h(\text{nova})}$, com coeficiente de confiança $(1-\alpha)$ é dado por

$$\text{IC}(Y_{h(\text{nova})}; 1-\alpha) = \hat{Y}_h \mp t(1-\alpha/2; n-p)\sqrt{\text{QMRes}\left[1 + \boldsymbol{x}_h^\top \left(\boldsymbol{X}^\top \boldsymbol{X}\right)^{-1} \boldsymbol{x}_h\right]},$$

em que $t(1-\alpha/2; n-p)$ é o quantil $100(1-\alpha/2)\%$ da distribuição t de Student, com $n-p$ graus de liberdade.

Intervalo de predição para β_m

Por (7), o intervalo de predição (IP) para $Y_{h(\text{nova})}$, com coeficiente de confiança $(1-\alpha)$ é dado por

$$\text{IC}(Y_{h(\text{nova})}; 1-\alpha) = \hat{Y}_h \mp t(1-\alpha/2; n-p)\sqrt{\text{QMRes}\left[1 + \boldsymbol{x}_h^\top \left(\boldsymbol{X}^\top \boldsymbol{X}\right)^{-1} \boldsymbol{x}_h\right]},$$

em que $t(1-\alpha/2; n-p)$ é o quantil $100(1-\alpha/2)\%$ da distribuição t de Student, com $n-p$ graus de liberdade.

Considerações

Em uma análise de regressão, a predição costuma ser ignorada quando o foco é inferencial, i.e., em que o importante é a interpretação dos coeficientes de regressão.

Há também uma limitação do procedimento, uma boa previsão da variável resposta está condicionada se os novos valores das covariáveis fazem partem do conjunto já observado, ver Seção 3.9 de Montgomery et al. (2021).

Considerações

Em uma análise de regressão, a predição costuma ser ignorada quando o foco é inferencial, i.e., em que o importante é a interpretação dos coeficientes de regressão.

Há também uma limitação do procedimento, uma boa previsão da variável resposta está condicionada se os novos valores das covariáveis fazem partem do conjunto já observado, ver Seção 3.9 de Montgomery et al. (2021).

Considerações

Se o foco principal for a predição, procedimentos adicionais precisam ser utilizados. O primeiro deles é a partição do banco de dados em amostras de **treino** e **validação**. E a obtenção das estimativas dos parâmetros será feita na amostra de treino.

Considerações

A partição do banco de dados e as técnicas subsequentes e alternativas, costumam ser categorizadas como **aprendizado supervisionado**, fazendo parte da área do Aprendizado Estatístico e ainda do Aprendizado de Máquinas.

Roteiro

Aplicação

Aplicação 1. Conjunto de dados *Westwood Company* (Neter et al., 1983, p. 36). Após o ajuste, nós temos o seguinte modelo estimado,

$$\hat{Y}_\ell = 10 + 2x_{\ell 2},$$

Aplicação

Aplicação 1. Conjunto de dados *Westwood Company* (Neter et al., 1983, p. 36). Após o ajuste, nós temos o seguinte modelo estimado,

$$\hat{Y}_\ell = 10 + 2x_{\ell 2},$$

em que Y_ℓ: horas trabalhadas, $x_{\ell 2}$: tamanho do lote, $\ell = 1, 2, \ldots, 10$.

Aplicação

Aplicação 1. Conjunto de dados *Westwood Company* (Neter et al., 1983, p. 36). Após o ajuste, nós temos o seguinte modelo estimado,

$$\hat{Y}_\ell = 10 + 2x_{\ell 2},$$

em que Y_ℓ: horas trabalhadas, $x_{\ell 2}$: tamanho do lote, $\ell = 1, 2, \ldots, 10$.

Exemplo

Nós temos também que:

Tabela 1: Estimativas do parâmetros.

Parâmetro	Estimativa	EP	t_c
β_1	10	2,503	3,995
β_2	2	0,047	42,583

Região crítica, para $\alpha = 5\%$: $|t_c| > 2{,}306$, com QMRes $= 7{,}5$.

Exemplo

Nós temos também que:

Tabela 1: Estimativas do parâmetros.

Parâmetro	Estimativa	EP	t_c
β_1	10	2,503	3,995
β_2	2	0,047	42,583

Região crítica, para $\alpha = 5\%$: $|t_c| > 2{,}306$, com QMRes $= 7{,}5$.

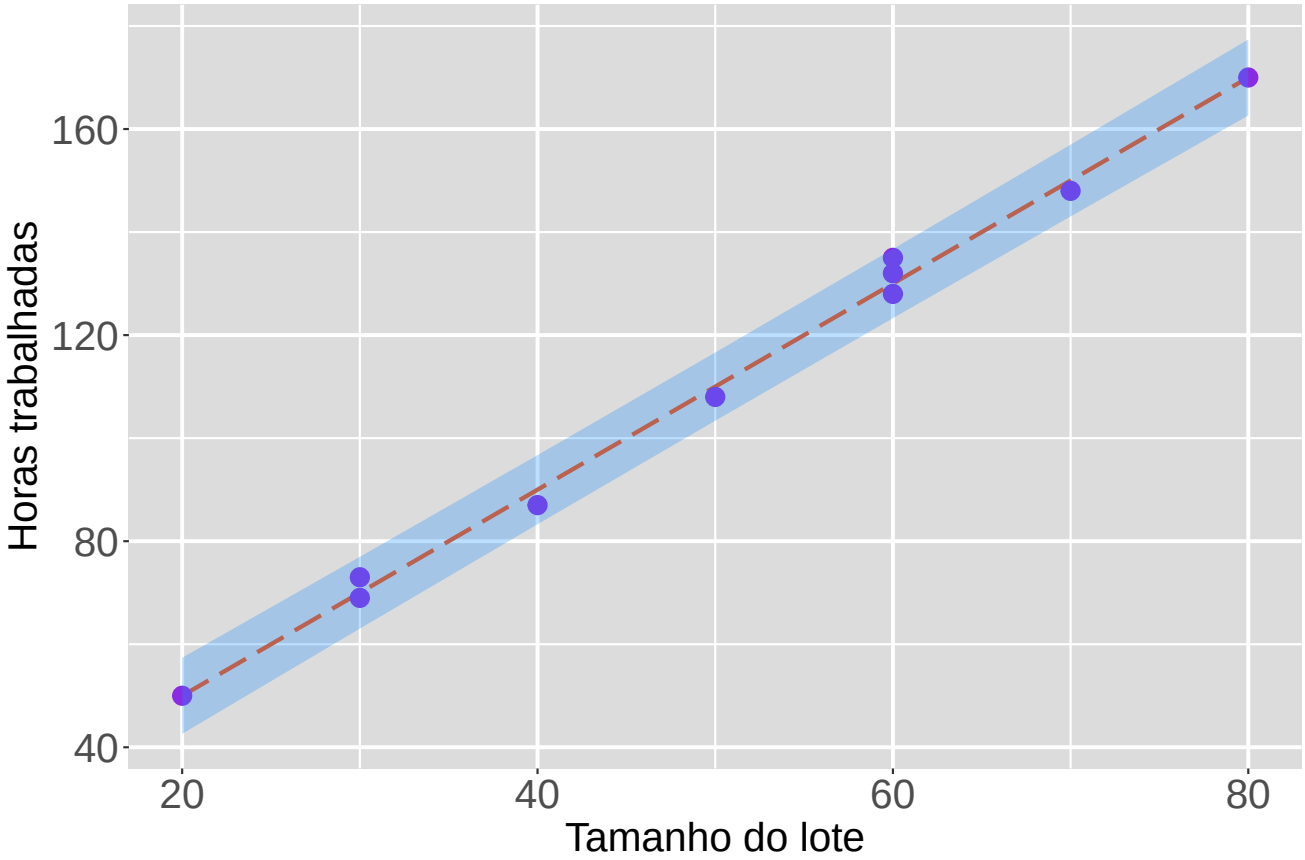

Figura 1: GD, reta ajustada e IP para os dados *Westwood Company*.

Aplicação

Aplicação 2. Um conjunto de dados de 1920, disponível em Ezekiel (1930), que relaciona a velocidade de um carro (variável explicativa) e a distância percorrida entre o momento da freagem e a parada total (variável resposta).

Após o ajuste, nós temos o seguinte modelo estimado,

$$\hat{Y}_\ell = -17{,}58 + 3{,}93 x_{\ell 2},$$

Aplicação

Aplicação 2. Um conjunto de dados de 1920, disponível em Ezekiel (1930), que relaciona a velocidade de um carro (variável explicativa) e a distância percorrida entre o momento da freagem e a parada total (variável resposta). Após o ajuste, nós temos o seguinte modelo estimado,

$$\hat{Y}_\ell = -17{,}58 + 3{,}93 x_{\ell 2},$$

em que Y_ℓ: distância percorrida (pés), $x_{\ell 2}$: velocidade (milhas por hora), $\ell = 1, 2, \ldots, 50$.

Aplicação

Aplicação 2. Um conjunto de dados de 1920, disponível em Ezekiel (1930), que relaciona a velocidade de um carro (variável explicativa) e a distância percorrida entre o momento da freagem e a parada total (variável resposta). Após o ajuste, nós temos o seguinte modelo estimado,

$$\hat{Y}_\ell = -17{,}58 + 3{,}93 x_{\ell 2},$$

em que Y_ℓ: distância percorrida (pés), $x_{\ell 2}$: velocidade (milhas por hora), $\ell = 1, 2, \ldots, 50$.

Exemplo

Nós temos também que:

Tabela 2: Estimativas do parâmetros.

Parâmetro	Estimativa	EP	t_c
β_1	-17,58	6,76	-2,60
β_2	3,93	0,42	9,46

Região crítica, para $\alpha = 5\%$: $|t_c| > 2{,}011$, com QMRes $= 236{,}53$.

Exemplo

Nós temos também que:

Tabela 2: Estimativas do parâmetros.

Parâmetro	Estimativa	EP	t_c
β_1	-17,58	6,76	-2,60
β_2	3,93	0,42	9,46

Região crítica, para $\alpha = 5\%$: $|t_c| > 2{,}011$, com QMRes $= 236{,}53$.

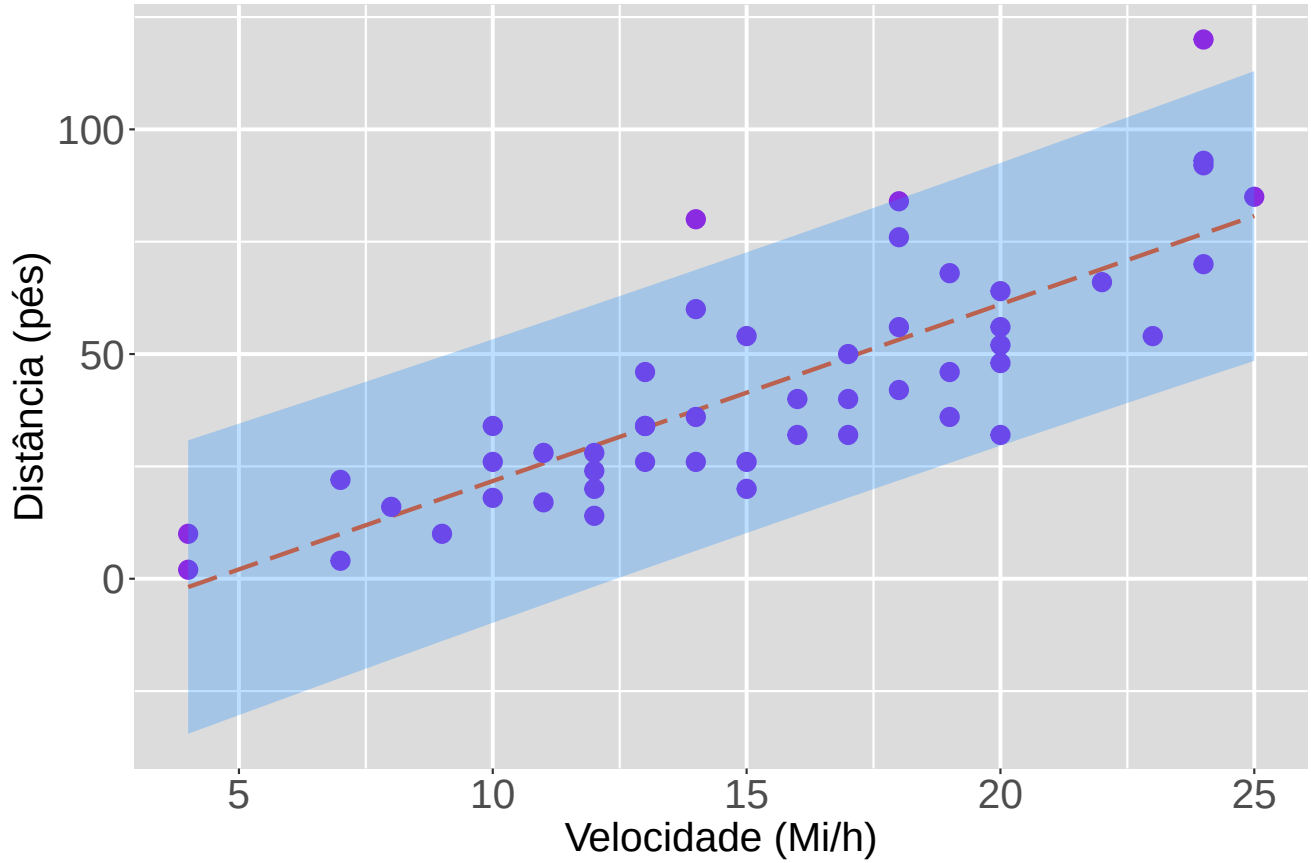

Figura 2: GD, reta ajustada e IP para os dados.

Roteiro

Referências bibliográficas I

Ezekiel, M. (1930), *Methods of Correlation Analysis*, Wiley, New York.

Gosset "Student", W. S. (1908), 'The probable error of a mean', *Biometrika* **6**(1), 1–25.

Montgomery, D. C., Peck, E. A. e Vining, G. G. (2021), *Introduction to linear regression analysis*, 6th edn, Wiley, New York.

Neter, J., Wasserman, W. e Kutner, M. H. (1983), *Applied linear regression models*, Richard D. Irwin Inc, Homewood, Illinois.

Obrigado!

✉ tiago.magalhaes@ufjf.br

ufjf.br/tiago_magalhaes

Departamento de Estatística, Sala 319