

Multicolinearidade

Tiago M. Magalhães

Departamento de Estatística - ICE-UFJF

Juiz de Fora, 25 de maio de 2026



Roteiro

- 1 Introdução
- 2 Identificando multicolinearidade
- 3 Regressão ridge
- 4 Aplicação
- 5 Referências bibliográficas



Roteiro

- 1 Introdução
- 2 Identificando multicolinearidade
- 3 Regressão ridge
- 4 Aplicação
- 5 Referências bibliográficas



Modelo de regressão linear

Suponham que Y_1, Y_2, \dots, Y_n tais que

$$Y_\ell = \mathbf{x}_\ell^\top \boldsymbol{\beta} + \varepsilon_\ell, \ell = 1, 2, \dots, n, \quad (1)$$



Modelo de regressão linear

Suponham que Y_1, Y_2, \dots, Y_n tais que

$$Y_\ell = \mathbf{x}_\ell^\top \boldsymbol{\beta} + \varepsilon_\ell, \quad \ell = 1, 2, \dots, n, \quad (1)$$

em que $\mathbf{x}_\ell = (x_{\ell 1}, x_{\ell 2}, \dots, x_{\ell p})^\top$ é conhecido, $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_p)^\top$ é um vetor de parâmetros desconhecidos a serem estimados, $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$ 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, \dots, Y_n tais que

$$Y_\ell = \mathbf{x}_\ell^\top \boldsymbol{\beta} + \varepsilon_\ell, \ell = 1, 2, \dots, n, \quad (1)$$

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



Forma matricial

A Equação (1) pode ser escrita de forma matricial, da seguinte forma:

$$Y = X\beta + \varepsilon, \quad (2)$$



Forma matricial

A Equação (1) pode ser escrita de forma matricial, da seguinte forma:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad (2)$$

em que $\mathbf{Y} = (Y_1, Y_2, \dots, Y_n)^\top$, $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)^\top$ é a matriz de planejamento e $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n)^\top$, com

$$\mathbb{E}(\boldsymbol{\varepsilon}) = \mathbf{0} \text{ e } \text{Var}(\boldsymbol{\varepsilon}) = \sigma^2 \mathbf{I}_n.$$



Forma matricial

A Equação (1) pode ser escrita de forma matricial, da seguinte forma:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad (2)$$

em que $\mathbf{Y} = (Y_1, Y_2, \dots, Y_n)^\top$, $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)^\top$ é a matriz de planejamento e $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n)^\top$, com

$$\mathbb{E}(\boldsymbol{\varepsilon}) = \mathbf{0} \text{ e } \text{Var}(\boldsymbol{\varepsilon}) = \sigma^2 \mathbf{I}_n.$$



Suposições

Resumindo,

- A relação entre as variáveis resposta e as preditoras é linear;



Suposições

Resumindo,

- A relação entre as variáveis resposta e as preditoras é linear;
- Os erros:



Suposições

Resumindo,

- A relação entre as variáveis resposta e as preditoras é linear;
- Os erros:
 - têm média zero;



Suposições

Resumindo,

- A relação entre as variáveis resposta e as preditoras é linear;
- Os erros:
 - têm média zero;
 - variância constante;



Suposições

Resumindo,

- A relação entre as variáveis resposta e as preditoras é linear;
- Os erros:
 - têm média zero;
 - variância constante;
 - e são não correlacionados.



Suposições

Resumindo,

- A relação entre as variáveis resposta e as preditoras é linear;
- Os erros:
 - têm média zero;
 - variância constante;
 - e são não correlacionados.

Multicolinearidade

A utilidade de um modelo de regressão pode ser impactada pela multicolinearidade, uma dependência (aproximadamente) linear entre as variáveis regressoras,

$$\sum_{m=1}^p t_m x_m = \mathbf{0}. \quad (3)$$

Multicolinearidade

A utilidade de um modelo de regressão pode ser impactada pela multicolinearidade, uma dependência (aproximadamente) linear entre as variáveis regressoras,

$$\sum_{m=1}^p t_m \mathbf{x}_m = \mathbf{0}. \quad (3)$$

Pois nestas situações, a capacidade de estimar os coeficientes de regressão estará comprometida. Uma dependência linear exata resulta em $\mathbf{X}^T \mathbf{X}$ singular, isto é, esta matriz não pode ser invertida.



Multicolinearidade

A utilidade de um modelo de regressão pode ser impactada pela multicolinearidade, uma dependência (aproximadamente) linear entre as variáveis regressoras,

$$\sum_{m=1}^p t_m \mathbf{x}_m = \mathbf{0}. \quad (3)$$

Pois nestas situações, a capacidade de estimar os coeficientes de regressão estará comprometida. Uma dependência linear exata resulta em $\mathbf{X}^T \mathbf{X}$ singular, isto é, esta matriz não pode ser invertida.



Padronização

Para nós lidarmos com a multicolinearidade, primeiramente, nós precisamos padronizar as variáveis, utilizando a **escala de tamanho unitário**, i.e.:

$$Y_{\ell}^* = \frac{Y_{\ell} - \bar{Y}}{S_Y} \text{ e } x_{\ell m}^* = \frac{x_{\ell m} - \bar{x}_m}{s_m},$$



Padronização

Para nós lidarmos com a multicolinearidade, primeiramente, nós precisamos padronizar as variáveis, utilizando a **escala de tamanho unitário**, i.e.:

$$Y_{\ell}^* = \frac{Y_{\ell} - \bar{Y}}{S_Y} \text{ e } x_{\ell m}^* = \frac{x_{\ell m} - \bar{x}_m}{s_m},$$

em que S_Y e s_m são, respectivamente, as raízes quadradas de

$$S_Y^2 = \sum_{\ell=1}^n (Y_{\ell} - \bar{Y})^2 \text{ e } S_m^2 = \sum_{\ell=1}^n (x_{\ell m} - \bar{x}_m)^2,$$

$\ell = 1, 2, \dots, n$ e $m = 1, 2, \dots, p$.



Padronização

Para nós lidarmos com a multicolinearidade, primeiramente, nós precisamos padronizar as variáveis, utilizando a **escala de tamanho unitário**, i.e.:

$$Y_{\ell}^* = \frac{Y_{\ell} - \bar{Y}}{S_Y} \text{ e } x_{\ell m}^* = \frac{x_{\ell m} - \bar{x}_m}{s_m},$$

em que S_Y e s_m são, respectivamente, as raízes quadradas de

$$S_Y^2 = \sum_{\ell=1}^n (Y_{\ell} - \bar{Y})^2 \text{ e } S_m^2 = \sum_{\ell=1}^n (x_{\ell m} - \bar{x}_m)^2,$$

$\ell = 1, 2, \dots, n$ e $m = 1, 2, \dots, p$.



Observações

- ① Cada regressor e a resposta têm média amostral zero e tamanho unitário, ou seja, $\sqrt{\sum_{\ell=1}^n (x_{\ell m}^* - \bar{x}_m^*)^2} = 1$;



Observações

- ① Cada regressor e a resposta têm média amostral zero e tamanho unitário, ou seja, $\sqrt{\sum_{\ell=1}^n (x_{\ell m}^* - \bar{x}_m^*)^2} = 1$;
- ② As matrizes $\mathbf{X}^{*\top} \mathbf{X}^*$ e $\mathbf{X}^{*\top} \mathbf{y}^*$ geram, respectivamente, as correlações entre as covariáveis e das covariáveis com a resposta;

Observações

- ① Cada regressor e a resposta têm média amostral zero e tamanho unitário, ou seja, $\sqrt{\sum_{\ell=1}^n (x_{\ell m}^* - \bar{x}_m^*)^2} = 1$;
- ② As matrizes $\mathbf{X}^{*\top} \mathbf{X}^*$ e $\mathbf{X}^{*\top} \mathbf{y}^*$ geram, respectivamente, as correlações entre as covariáveis e das covariáveis com a resposta;
- ③ Se o modelo tem intercepto, o procedimento de padronização das variáveis o removerá;

Observações

- ① Cada regressor e a resposta têm média amostral zero e tamanho unitário, ou seja, $\sqrt{\sum_{\ell=1}^n (x_{\ell m}^* - \bar{x}_m^*)^2} = 1$;
- ② As matrizes $\mathbf{X}^{*\top} \mathbf{X}^*$ e $\mathbf{X}^{*\top} \mathbf{y}^*$ geram, respectivamente, as correlações entre as covariáveis e das covariáveis com a resposta;
- ③ Se o modelo tem intercepto, o procedimento de padronização das variáveis o removerá;
- ④ A partir das estimativas do modelo padronizado, pode-se chegar nas estimativas dos parâmetros do modelo original, ver Montgomery et al. (2021, p. 119).



Observações

- ① Cada regressor e a resposta têm média amostral zero e tamanho unitário, ou seja, $\sqrt{\sum_{\ell=1}^n (x_{\ell m}^* - \bar{x}_m^*)^2} = 1$;
- ② As matrizes $\mathbf{X}^{*\top} \mathbf{X}^*$ e $\mathbf{X}^{*\top} \mathbf{y}^*$ geram, respectivamente, as correlações entre as covariáveis e das covariáveis com a resposta;
- ③ Se o modelo tem intercepto, o procedimento de padronização das variáveis o removerá;
- ④ A partir das estimativas do modelo padronizado, pode-se chegar nas estimativas dos parâmetros do modelo original, ver Montgomery et al. (2021, p. 119).

Escala normal de unidade

Lembrando que existe uma segunda possibilidade de padronização, denominada de **escala normal de unidade**, feito da seguinte forma:

$$Y_l^* = \frac{Y_l - \bar{Y}}{S_Y} \text{ e } x_{lm}^* = \frac{x_{lm} - \bar{x}_m}{s_m},$$



Escala normal de unidade

Lembrando que existe uma segunda possibilidade de padronização, denominada de **escala normal de unidade**, feito da seguinte forma:

$$Y_{\ell}^* = \frac{Y_{\ell} - \bar{Y}}{S_Y} \text{ e } x_{\ell m}^* = \frac{x_{\ell m} - \bar{x}_m}{s_m},$$

em que S_Y e s_m são, respectivamente, as raízes quadradas de

$$S_Y^2 = \frac{\sum_{\ell=1}^n (Y_{\ell} - \bar{Y})^2}{n-1} \text{ e } s_m^2 = \frac{\sum_{\ell=1}^n (x_{\ell m} - \bar{x}_m)^2}{n-1},$$

$\ell = 1, 2, \dots, n$ e $m = 1, 2, \dots, p$.



Escala normal de unidade

Lembrando que existe uma segunda possibilidade de padronização, denominada de **escala normal de unidade**, feito da seguinte forma:

$$Y_{\ell}^* = \frac{Y_{\ell} - \bar{Y}}{S_Y} \text{ e } x_{\ell m}^* = \frac{x_{\ell m} - \bar{x}_m}{s_m},$$

em que S_Y e s_m são, respectivamente, as raízes quadradas de

$$S_Y^2 = \frac{\sum_{\ell=1}^n (Y_{\ell} - \bar{Y})^2}{n-1} \text{ e } s_m^2 = \frac{\sum_{\ell=1}^n (x_{\ell m} - \bar{x}_m)^2}{n-1},$$

$\ell = 1, 2, \dots, n$ e $m = 1, 2, \dots, p$.



Escala normal de unidade

Observação

Todos os regressores transformados e as respostas transformadas têm média amostral zero e variância amostral um.

Escala normal de unidade

Observação

Todos os regressores transformados e as respostas transformadas têm média amostral zero e variância amostral um.

Modelo de regressão linear

Observação

Nesta apresentação, nós iremos que assumir que (1) ou, equivalentemente, (2) estão na **escala de tamanho unitário**.

Modelo de regressão linear

Observação

Nesta apresentação, nós iremos que assumir que (1) ou, equivalentemente, (2) estão na **escala de tamanho unitário**.



Roteiro

- 1 Introdução
- 2 Identificando multicolinearidade**
- 3 Regressão ridge
- 4 Aplicação
- 5 Referências bibliográficas



Fontes de multilinearidade

Existem quatro fontes de multilinearidade:

- ① A forma como os dados foram coletados. Apenas uma subamostra de um espaço amostral inteiro foi coletada;



Fontes de multicolinearidade

Existem quatro fontes de multicolinearidade:

- ① A forma como os dados foram coletados. Apenas uma subamostra de um espaço amostral inteiro foi coletada;
- ② Restrições do modelo ou da população. Dados composicionais;



Fontes de multicolinearidade

Existem quatro fontes de multicolinearidade:

- ① A forma como os dados foram coletados. Apenas uma subamostra de um espaço amostral inteiro foi coletada;
- ② Restrições do modelo ou da população. Dados composicionais;
- ③ Especificação do modelo. Modelo de regressão polinomial;



Fontes de multicolinearidade

Existem quatro fontes de multicolinearidade:

- ① A forma como os dados foram coletados. Apenas uma subamostra de um espaço amostral inteiro foi coletada;
- ② Restrições do modelo ou da população. Dados composicionais;
- ③ Especificação do modelo. Modelo de regressão polinomial;
- ④ Modelo super definido. Conjunto de dados com mais covariáveis do que observações (*high dimensional data*).



Fontes de multicolinearidade

Existem quatro fontes de multicolinearidade:

- ① A forma como os dados foram coletados. Apenas uma subamostra de um espaço amostral inteiro foi coletada;
- ② Restrições do modelo ou da população. Dados composicionais;
- ③ Especificação do modelo. Modelo de regressão polinomial;
- ④ Modelo super definido. Conjunto de dados com mais covariáveis do que observações (*high dimensional data*).



Identificando multicolinearidade

No caso de variáveis padronizadas, $\mathbf{X}^T \mathbf{X}$ representa uma matriz de correlações. Se algum elemento, em módulo, desta matriz for próximo de 1, haverá um forte indício de multicolinearidade. Porém, o contrário não será verdadeiro.



Identificando multicolinearidade

No caso de variáveis padronizadas, $\mathbf{X}^T \mathbf{X}$ representa uma matriz de correlações. Se algum elemento, em módulo, desta matriz for próximo de 1, haverá um forte indício de multicolinearidade. Porém, o contrário não será verdadeiro.



Identificando multicolinearidade

Uma alta multicolinearidade pode resultar em variâncias e covariâncias altas das estimativas. Além disso, o m -ésimo, $m = 1, 2, \dots, p$, elemento da diagonal principal de $(\mathbf{X}^T \mathbf{X})^{-1}$ é chamado de fator de inflação da variância (*variance inflation factors*, VIF).



Identificando multicolinearidade

Uma alta multicolinearidade pode resultar em variâncias e covariâncias altas das estimativas. Além disso, o m -ésimo, $m = 1, 2, \dots, p$, elemento da diagonal principal de $(\mathbf{X}^T \mathbf{X})^{-1}$ é chamado de fator de inflação da variância (*variance inflation factors*, VIF).



Identificando multicolinearidade

O VIF podem ser escrito como:

$$\text{VIF}_m = \frac{1}{1 - R_m^2},$$



Identificando multicolinearidade

O VIF podem ser escrito como:

$$\text{VIF}_m = \frac{1}{1 - R_m^2},$$

em que R_m^2 é o coeficiente de determinação múltiplo da regressão entre x_m com as demais covariáveis.



Identificando multicolinearidade

O VIF podem ser escrito como:

$$\text{VIF}_m = \frac{1}{1 - R_m^2},$$

em que R_m^2 é o coeficiente de determinação múltiplo da regressão entre x_m com as demais covariáveis. Um $\text{VIF}_m > 5$ indica um forte indício de multicolinearidade. Consequentemente, regressores com valores de VIF, provavelmente, foram mal estimados.



Identificando multicolinearidade

O VIF podem ser escrito como:

$$\text{VIF}_m = \frac{1}{1 - R_m^2},$$

em que R_m^2 é o coeficiente de determinação múltiplo da regressão entre x_m com as demais covariáveis. Um $\text{VIF}_m > 5$ indica um forte indício de multicolinearidade. Conseqüentemente, regressores com valores de VIF, provavelmente, foram mal estimados.



Identificando multicolinearidade

Notem que, o **comprimento** do intervalo de confiança (IC) para β_m , com coeficiente de confiança $(1 - \alpha)$ é dado por

$$L_m = 2t(1 - \alpha/2; n - p)\sqrt{\sigma^2\text{VIF}_m}, \quad (4)$$



Identificando multicolinearidade

Notem que, o **comprimento** do intervalo de confiança (IC) para β_m , com coeficiente de confiança $(1 - \alpha)$ é dado por

$$L_m = 2t(1 - \alpha/2; n - p)\sqrt{\sigma^2\text{VIF}_m}, \quad (4)$$

em que $t(1 - \alpha/2; n - p)$ é o quantil $100(1 - \alpha/2)\%$ da distribuição t de Student (Gosset “Student”, 1908), com $n - p$ graus de liberdade, $m = 1, 2, \dots, p$.



Identificando multicolinearidade

Notem que, o **comprimento** do intervalo de confiança (IC) para β_m , com coeficiente de confiança $(1 - \alpha)$ é dado por

$$L_m = 2t(1 - \alpha/2; n - p)\sqrt{\sigma^2\text{VIF}_m}, \quad (4)$$

em que $t(1 - \alpha/2; n - p)$ é o quantil $100(1 - \alpha/2)\%$ da distribuição t de Student (Gosset “Student”, 1908), com $n - p$ graus de liberdade, $m = 1, 2, \dots, p$. **Observação:** se o modelo original tivesse intercepto, seriam $n - p - 1$ graus de liberdade.



Identificando multicolinearidade

Notem que, o **comprimento** do intervalo de confiança (IC) para β_m , com coeficiente de confiança $(1 - \alpha)$ é dado por

$$L_m = 2t(1 - \alpha/2; n - p)\sqrt{\sigma^2\text{VIF}_m}, \quad (4)$$

em que $t(1 - \alpha/2; n - p)$ é o quantil $100(1 - \alpha/2)\%$ da distribuição t de Student (Gosset “Student”, 1908), com $n - p$ graus de liberdade, $m = 1, 2, \dots, p$. **Observação:** se o modelo original tivesse intercepto, seriam $n - p - 1$ graus de liberdade.



Identificando multicolinearidade

Se as covariáveis forem ortogonais, o comprimento do IC (4) seria dado por

$$L^* = 2t(1 - \alpha/2; n - p)\sqrt{\sigma^2}.$$



Identificando multicolinearidade

Se as covariáveis forem ortogonais, o comprimento do IC (4) seria dado por

$$L^* = 2t(1 - \alpha/2; n - p)\sqrt{\sigma^2}.$$

E a razão $L_m/L^* = \sqrt{\text{VIF}_m}$ é um indicador de quão longe o IC está do cenário ideal, de ortogonalidade.



Identificando multicolinearidade

Se as covariáveis forem ortogonais, o comprimento do IC (4) seria dado por

$$L^* = 2t(1 - \alpha/2; n - p)\sqrt{\sigma^2}.$$

E a razão $L_m/L^* = \sqrt{\text{VIF}_m}$ é um indicador de quão longe o IC está do cenário ideal, de ortogonalidade.



Identificando multicolinearidade

Uma alta multicolinearidade também afeta as estimativa dos parâmetros, resultando em valores altos, em módulo. Seja a distância quadrada entre a estimativa e valor verdadeiro, i.e.,



Identificando multicolinearidade

Uma alta multicolinearidade também afeta as estimativa dos parâmetros, resultando em valores altos, em módulo. Seja a distância quadrada entre a estimativa e valor verdadeiro, i.e.,

$$L_1^2 = (\hat{\beta} - \beta)^\top (\hat{\beta} - \beta). \quad (5)$$



Identificando multicolinearidade

Uma alta multicolinearidade também afeta as estimativa dos parâmetros, resultando em valores altos, em módulo. Seja a distância quadrada entre a estimativa e valor verdadeiro, i.e.,

$$L_1^2 = (\hat{\beta} - \beta)^\top (\hat{\beta} - \beta). \quad (5)$$



Identificando multicolinearidade

De (5), nós temos que:

$$\mathbb{E}(L_1^2) = \mathbb{E} \left\{ (\hat{\beta} - \beta)^\top (\hat{\beta} - \beta) \right\} = \sigma^2 \text{tr} \left\{ (\mathbf{X}^\top \mathbf{X})^{-1} \right\}, \quad (6)$$

$$\text{Var}(L_1^2) = 2\sigma^4 \text{tr} \left\{ (\mathbf{X}^\top \mathbf{X})^{-1} \right\}.$$

Identificando multicolinearidade

De (5), nós temos que:

$$\mathbb{E}(L_1^2) = \mathbb{E} \left\{ (\hat{\beta} - \beta)^\top (\hat{\beta} - \beta) \right\} = \sigma^2 \text{tr} \left\{ (\mathbf{X}^\top \mathbf{X})^{-1} \right\}, \quad (6)$$

$$\text{Var}(L_1^2) = 2\sigma^4 \text{tr} \left\{ (\mathbf{X}^\top \mathbf{X})^{-1} \right\}.$$

Além de,

$$\mathbb{E}(\hat{\beta}^\top \hat{\beta}) = \beta^\top \beta + \sigma^2 \text{tr} \left\{ (\mathbf{X}^\top \mathbf{X})^{-1} \right\}.$$



Identificando multicolinearidade

De (5), nós temos que:

$$\mathbb{E}(L_1^2) = \mathbb{E} \left\{ (\hat{\beta} - \beta)^\top (\hat{\beta} - \beta) \right\} = \sigma^2 \text{tr} \left\{ (\mathbf{X}^\top \mathbf{X})^{-1} \right\}, \quad (6)$$

$$\text{Var}(L_1^2) = 2\sigma^4 \text{tr} \left\{ (\mathbf{X}^\top \mathbf{X})^{-1} \right\}.$$

Além de,

$$\mathbb{E}(\hat{\beta}^\top \hat{\beta}) = \beta^\top \beta + \sigma^2 \text{tr} \left\{ (\mathbf{X}^\top \mathbf{X})^{-1} \right\}.$$



Identificando multilinearidade

Lembrando que $\text{tr}\{\cdot\}$ é o operador traço, a soma dos elementos da diagonal de uma matriz, equivalente a soma dos seus autovalores. Na presença de multilinearidade, alguns autovalores de $\mathbf{X}^T \mathbf{X}$ serão pequenos.



Identificando multicolinearidade

Lembrando que $\text{tr}\{\cdot\}$ é o operador traço, a soma dos elementos da diagonal de uma matriz, equivalente a soma dos seus autovalores. Na presença de multicolinearidade, alguns autovalores de $\mathbf{X}^\top \mathbf{X}$ serão pequenos. Seja $\lambda_m > 0$, o m -ésimo autovalor de $\mathbf{X}^\top \mathbf{X}$, $m = 1, 2, \dots, p$,



Identificando multicolinearidade

Lembrando que $\text{tr}\{\cdot\}$ é o operador traço, a soma dos elementos da diagonal de uma matriz, equivalente a soma dos seus autovalores. Na presença de multicolinearidade, alguns autovalores de $\mathbf{X}^\top \mathbf{X}$ serão pequenos. Seja $\lambda_m > 0$, o m -ésimo autovalor de $\mathbf{X}^\top \mathbf{X}$, $m = 1, 2, \dots, p$, nós podemos escrever a esperança em (6) da seguinte forma:

$$\mathbb{E}(L_1^2) = \sigma^2 \sum_{m=1}^p \frac{1}{\lambda_m}.$$



Identificando multicolinearidade

Lembrando que $\text{tr}\{\cdot\}$ é o operador traço, a soma dos elementos da diagonal de uma matriz, equivalente a soma dos seus autovalores. Na presença de multicolinearidade, alguns autovalores de $\mathbf{X}^\top \mathbf{X}$ serão pequenos. Seja $\lambda_m > 0$, o m -ésimo autovalor de $\mathbf{X}^\top \mathbf{X}$, $m = 1, 2, \dots, p$, nós podemos escrever a esperança em (6) da seguinte forma:

$$\mathbb{E}(L_1^2) = \sigma^2 \sum_{m=1}^p \frac{1}{\lambda_m}.$$



Identificando multicolinearidade

Seja

$$\kappa = \frac{\max \{\lambda_1, \lambda_2, \dots, \lambda_p\}}{\min \{\lambda_1, \lambda_2, \dots, \lambda_p\}} = \frac{\lambda_{\max}}{\lambda_{\min}}.$$

A multicolinearidade se:

$$\left\{ \begin{array}{ll} \kappa < 100, & \text{não é significativa} \\ 100 < \kappa < 1.000, & \text{é moderada} \\ \kappa > 1.000, & \text{é alta} \end{array} \right. .$$



Identificando multicolinearidade

Seja

$$\kappa = \frac{\max \{\lambda_1, \lambda_2, \dots, \lambda_p\}}{\min \{\lambda_1, \lambda_2, \dots, \lambda_p\}} = \frac{\lambda_{\max}}{\lambda_{\min}}.$$

A multicolinearidade se:

$$\left\{ \begin{array}{ll} \kappa < 100, & \text{não é significativa} \\ 100 < \kappa < 1.000, & \text{é moderada} \\ \kappa > 1.000, & \text{é alta} \end{array} \right. .$$



Identificando multicolinearidade

Agora, para identificar as variáveis que estão envolvidas na multicolinearidade, seja

$$\kappa_m = \frac{\lambda_{\max}}{\lambda_m},$$

$$m = 1, 2, \dots, p.$$

Identificando multicolinearidade

Agora, para identificar as variáveis que estão envolvidas na multicolinearidade, seja

$$\kappa_m = \frac{\lambda_{\max}}{\lambda_m},$$

$m = 1, 2, \dots, p$. O número de casos em que $\kappa_m > 1.000$, nos dará uma ideia de quantas colunas de $\mathbf{X}^T \mathbf{X}$ são linearmente dependentes.



Identificando multicolinearidade

Agora, para identificar as variáveis que estão envolvidas na multicolinearidade, seja

$$\kappa_m = \frac{\lambda_{\max}}{\lambda_m},$$

$m = 1, 2, \dots, p$. O número de casos em que $\kappa_m > 1.000$, nos dará uma ideia de quantas colunas de $\mathbf{X}^T \mathbf{X}$ são linearmente dependentes.



Identificando multicolinearidade

A matriz $\mathbf{X}^T \mathbf{X}$ pode ser decomposta da seguinte forma:

$$\mathbf{X}^T \mathbf{X} = \mathbf{T}^T \mathbf{\Lambda} \mathbf{T},$$

em que $\mathbf{\Lambda}$ é uma matriz diagonal $p \times p$, sendo $\lambda_1, \lambda_2, \dots, \lambda_p$, os elementos da diagonal principal e \mathbf{T} é uma matriz ortogonal $p \times p$, com a m -ésima coluna dada pelo m -ésimo autovetor de $\mathbf{X}^T \mathbf{X}$.



Identificando multicolinearidade

A matriz $\mathbf{X}^T \mathbf{X}$ pode ser decomposta da seguinte forma:

$$\mathbf{X}^T \mathbf{X} = \mathbf{T}^T \mathbf{\Lambda} \mathbf{T},$$

em que $\mathbf{\Lambda}$ é uma matriz diagonal $p \times p$, sendo $\lambda_1, \lambda_2, \dots, \lambda_p$, os elementos da diagonal principal e \mathbf{T} é uma matriz ortogonal $p \times p$, com a m -ésima coluna dada pelo m -ésimo autovetor de $\mathbf{X}^T \mathbf{X}$.



Identificando multicolinearidade

Se algum λ_m for próximo de zero, indicando uma relação, ao menos aproximada, de dependência linear dos dados, os elementos da coluna m de \mathbf{T} , podem ser interpretados com os coeficientes que ponderam essa dependência, da forma vista em (3).



Roteiro

- 1 Introdução
- 2 Identificando multicolinearidade
- 3 Regressão ridge**
- 4 Aplicação
- 5 Referências bibliográficas



Métodos para contornar a multicolinearidade

Uma vez que, a presença da multicolinearidade é identificada, existem algumas maneiras de contorná-la:

- Coletando mais dados. Porém, isso não tem viabilidade prática, por conta do custo ou da existência dos dados;



Métodos para contornar a multicolinearidade

Uma vez que, a presença da multicolinearidade é identificada, existem algumas maneiras de contorná-la:

- Coletando mais dados. Porém, isso não tem viabilidade prática, por conta do custo ou da existência dos dados;
- Reespecificando o modelo. Rescrevendo as variáveis independentes ou eliminando uma delas. Porém, isso poderá remover uma variável importante do modelo;



Métodos para contornar a multicolinearidade

Uma vez que, a presença da multicolinearidade é identificada, existem algumas maneiras de contorná-la:

- Coletando mais dados. Porém, isso não tem viabilidade prática, por conta do custo ou da existência dos dados;
- Reespecificando o modelo. Rescrevendo as variáveis independentes ou eliminando uma delas. Porém, isso poderá remover uma variável importante do modelo;
- Utilizando uma outra metodologia, como a regressão *ridge*.



Métodos para contornar a multicolinearidade

Uma vez que, a presença da multicolinearidade é identificada, existem algumas maneiras de contorná-la:

- Coletando mais dados. Porém, isso não tem viabilidade prática, por conta do custo ou da existência dos dados;
- Reespecificando o modelo. Reescrevendo as variáveis independentes ou eliminando uma delas. Porém, isso poderá remover uma variável importante do modelo;
- Utilizando uma outra metodologia, como a regressão *ridge*.



Regressão ridge

A presença de multicolinearidade causa uma inflação das variâncias dos estimadores. Como nós vimos, os estimadores de mínimos quadrados (EMQ), pelo Teorema de Gauss-Markov (Gauss, 1821), têm a menor variância entre todos os estimadores lineares não viesados.



Regressão ridge

A presença de multicolinearidade causa uma inflação das variâncias dos estimadores. Como nós vimos, os estimadores de mínimos quadrados (EMQ), pelo Teorema de Gauss-Markov (Gauss, 1821), têm a menor variância entre todos os estimadores lineares não viesados. Logo, um caminho para contornar a multicolinearidade é encontrar **estimadores viesados**.



Regressão ridge

A presença de multicolinearidade causa uma inflação das variâncias dos estimadores. Como nós vimos, os estimadores de mínimos quadrados (EMQ), pelo Teorema de Gauss-Markov (Gauss, 1821), têm a menor variância entre todos os estimadores lineares não viesados. Logo, um caminho para contornar a multicolinearidade é encontrar **estimadores viesados**.



Regressão ridge

Um método para encontrar estimadores viesados é a regressão *ridge* (rígida, em cristas, Hoerl e Kennard, 1970*b,a*). Esse estimador pode ser escrito como a solução da minimização de:



Regressão ridge

Um método para encontrar estimadores viesados é a regressão *ridge* (rígida, em cristas, Hoerl e Kennard, 1970*b,a*). Esse estimador pode ser escrito como a solução da minimização de:

$$Q = \sum_{\ell=1}^n [Y_{\ell} - (\beta_1 x_{\ell 1} + \beta_2 x_{\ell 2} + \cdots + \beta_p x_{\ell p})]^2 + k \sum_{m=1}^p \beta_m^2, \quad (7)$$



Regressão ridge

Um método para encontrar estimadores viesados é a regressão *ridge* (rígida, em cristas, Hoerl e Kennard, 1970*b,a*). Esse estimador pode ser escrito como a solução da minimização de:

$$Q = \sum_{\ell=1}^n [Y_{\ell} - (\beta_1 x_{\ell 1} + \beta_2 x_{\ell 2} + \cdots + \beta_p x_{\ell p})]^2 + k \sum_{m=1}^p \beta_m^2, \quad (7)$$

em que $k \geq 0$ é uma constante denominada de parâmetro de regularização.

A segunda parte do lado de direito de (7), é um termo de “encolhimento”.



Regressão ridge

Um método para encontrar estimadores viesados é a regressão *ridge* (rígida, em cristas, Hoerl e Kennard, 1970*b,a*). Esse estimador pode ser escrito como a solução da minimização de:

$$Q = \sum_{\ell=1}^n [Y_{\ell} - (\beta_1 x_{\ell 1} + \beta_2 x_{\ell 2} + \cdots + \beta_p x_{\ell p})]^2 + k \sum_{m=1}^p \beta_m^2, \quad (7)$$

em que $k \geq 0$ é uma constante denominada de parâmetro de regularização.

A segunda parte do lado de direito de (7), é um termo de “encolhimento”.



Regressão ridge

Em notação matricial, (7) pode ser escrito como:

$$Q = (\mathbf{Y} - \mathbf{X}\beta)^\top (\mathbf{Y} - \mathbf{X}\beta) + k\beta^\top \beta.$$



Regressão ridge

Em notação matricial, (7) pode ser escrito como:

$$Q = (\mathbf{Y} - \mathbf{X}\beta)^\top (\mathbf{Y} - \mathbf{X}\beta) + k\beta^\top \beta.$$

O estimador *ridge* é dada por:

$$\hat{\beta}_R = (\mathbf{X}^\top \mathbf{X} + kI_p)^{-1} \mathbf{X}^\top \mathbf{Y}, \quad (8)$$



Regressão ridge

Em notação matricial, (7) pode ser escrito como:

$$Q = (\mathbf{Y} - \mathbf{X}\beta)^\top (\mathbf{Y} - \mathbf{X}\beta) + k\beta^\top \beta.$$

O estimador *ridge* é dada por:

$$\hat{\beta}_R = (\mathbf{X}^\top \mathbf{X} + k\mathbf{I}_p)^{-1} \mathbf{X}^\top \mathbf{Y}, \quad (8)$$

em que k é uma constante definida pelo analista, quando $k = 0$, o estimador (8) coincide com o EMQ.



Regressão ridge

Em notação matricial, (7) pode ser escrito como:

$$Q = (\mathbf{Y} - \mathbf{X}\beta)^\top (\mathbf{Y} - \mathbf{X}\beta) + k\beta^\top \beta.$$

O estimador *ridge* é dada por:

$$\hat{\beta}_R = (\mathbf{X}^\top \mathbf{X} + k\mathbf{I}_p)^{-1} \mathbf{X}^\top \mathbf{Y}, \quad (8)$$

em que k é uma constante definida pelo analista, quando $k = 0$, o estimador (8) coincide com o EMQ.



Regressão ridge

Nós podemos reescrever (8) da seguinte forma:

$$\begin{aligned}\hat{\beta}_R &= (\mathbf{X}^\top \mathbf{X} + kI_p)^{-1} \mathbf{X}^\top \mathbf{Y} \\ &= (\mathbf{X}^\top \mathbf{X} + kI_p)^{-1} (\mathbf{X}^\top \mathbf{X}) (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{Y} \\ &= \mathbf{Z}_k \hat{\beta},\end{aligned}\tag{9}$$



Regressão ridge

Nós podemos reescrever (8) da seguinte forma:

$$\begin{aligned}\hat{\beta}_R &= (\mathbf{X}^T \mathbf{X} + kI_p)^{-1} \mathbf{X}^T \mathbf{Y} \\ &= (\mathbf{X}^T \mathbf{X} + kI_p)^{-1} (\mathbf{X}^T \mathbf{X}) (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y} \\ &= \mathbf{Z}_k \hat{\beta},\end{aligned}\tag{9}$$

em que $\mathbf{Z}_k = (\mathbf{X}^T \mathbf{X} + kI_p)^{-1} (\mathbf{X}^T \mathbf{X})$.



Regressão ridge

Nós podemos reescrever (8) da seguinte forma:

$$\begin{aligned}\hat{\beta}_R &= (\mathbf{X}^\top \mathbf{X} + k\mathbf{I}_p)^{-1} \mathbf{X}^\top \mathbf{Y} \\ &= (\mathbf{X}^\top \mathbf{X} + k\mathbf{I}_p)^{-1} (\mathbf{X}^\top \mathbf{X}) (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{Y} \\ &= \mathbf{Z}_k \hat{\beta},\end{aligned}\tag{9}$$

em que $\mathbf{Z}_k = (\mathbf{X}^\top \mathbf{X} + k\mathbf{I}_p)^{-1} (\mathbf{X}^\top \mathbf{X})$.



Regressão ridge

Sejam as matrizes quadradas \mathbf{A} e \mathbf{B} , nós sabemos que,

$$(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}.$$

Regressão ridge

Sejam as matrizes quadradas \mathbf{A} e \mathbf{B} , nós sabemos que,

$$(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}.$$

Então, nós podemos reescrever \mathbf{Z}_k da seguinte forma



Regressão ridge

Sejam as matrizes quadradas \mathbf{A} e \mathbf{B} , nós sabemos que,

$$(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}.$$

Então, nós podemos reescrever \mathbf{Z}_k da seguinte forma

$$\begin{aligned}\mathbf{Z}_k &= (\mathbf{X}^\top \mathbf{X} + k\mathbf{I}_p)^{-1} [(\mathbf{X}^\top \mathbf{X})^{-1}]^{-1} \\ &= [(\mathbf{X}^\top \mathbf{X})^{-1}(\mathbf{X}^\top \mathbf{X}) + k(\mathbf{X}^\top \mathbf{X})^{-1}\mathbf{I}_p]^{-1} \\ &= [\mathbf{I}_p + k(\mathbf{X}^\top \mathbf{X})^{-1}]^{-1}.\end{aligned}$$



Regressão ridge

Sejam as matrizes quadradas \mathbf{A} e \mathbf{B} , nós sabemos que,

$$(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}.$$

Então, nós podemos reescrever \mathbf{Z}_k da seguinte forma

$$\begin{aligned}\mathbf{Z}_k &= (\mathbf{X}^\top \mathbf{X} + k\mathbf{I}_p)^{-1} \left[(\mathbf{X}^\top \mathbf{X})^{-1} \right]^{-1} \\ &= \left[(\mathbf{X}^\top \mathbf{X})^{-1} (\mathbf{X}^\top \mathbf{X}) + k(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{I}_p \right]^{-1} \\ &= \left[\mathbf{I}_p + k(\mathbf{X}^\top \mathbf{X})^{-1} \right]^{-1}.\end{aligned}$$



Regressão ridge

De (9), nós podemos ver que,

$$\mathbb{E}(\hat{\beta}_R) = \mathbb{E}(\mathbf{Z}_k \hat{\beta}) = \mathbf{Z}_k \mathbb{E}(\hat{\beta}) = \mathbf{Z}_k \beta$$



Regressão ridge

De (9), nós podemos ver que,

$$\mathbb{E}(\hat{\beta}_R) = \mathbb{E}(\mathbf{Z}_k \hat{\beta}) = \mathbf{Z}_k \mathbb{E}(\hat{\beta}) = \mathbf{Z}_k \beta$$

Logo, $\hat{\beta}_R$ é um estimador viesado.



Regressão ridge

De (9), nós podemos ver que,

$$\mathbb{E}(\hat{\beta}_R) = \mathbb{E}(\mathbf{Z}_k \hat{\beta}) = \mathbf{Z}_k \mathbb{E}(\hat{\beta}) = \mathbf{Z}_k \beta$$

Logo, $\hat{\beta}_R$ é um estimador viesado. Lembrando que, o viés é dado por

$$\begin{aligned} \text{Viés}(\hat{\beta}_R) &= \mathbb{E}(\hat{\beta}_R - \beta) = \mathbb{E}(\hat{\beta}_R) - \beta \\ &= \mathbf{Z}_k \beta - \beta = (\mathbf{Z}_k - \mathbf{I}_p) \beta. \end{aligned}$$



Regressão ridge

De (9), nós podemos ver que,

$$\mathbb{E}(\hat{\beta}_R) = \mathbb{E}(\mathbf{Z}_k \hat{\beta}) = \mathbf{Z}_k \mathbb{E}(\hat{\beta}) = \mathbf{Z}_k \beta$$

Logo, $\hat{\beta}_R$ é um estimador viesado. Lembrando que, o viés é dado por

$$\begin{aligned} \text{Viés}(\hat{\beta}_R) &= \mathbb{E}(\hat{\beta}_R - \beta) = \mathbb{E}(\hat{\beta}_R) - \beta \\ &= \mathbf{Z}_k \beta - \beta = (\mathbf{Z}_k - \mathbf{I}_p) \beta. \end{aligned}$$



Regressão ridge

De (9), nós também temos:

$$\begin{aligned}\text{Var}(\hat{\beta}_R) &= \text{Var}(\mathbf{Z}_k \hat{\beta}) = \mathbf{Z}_k \text{Var}(\hat{\beta}) \mathbf{Z}_k^\top = \sigma^2 \mathbf{Z}_k (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{Z}_k^\top \\ &= \sigma^2 (\mathbf{X}^\top \mathbf{X} + k \mathbf{I}_p)^{-1} (\mathbf{X}^\top \mathbf{X}) (\mathbf{X}^\top \mathbf{X})^{-1} (\mathbf{X}^\top \mathbf{X}) (\mathbf{X}^\top \mathbf{X} + k \mathbf{I}_p)^{-1} \\ &= \sigma^2 (\mathbf{X}^\top \mathbf{X} + k \mathbf{I}_p)^{-1} (\mathbf{X}^\top \mathbf{X}) (\mathbf{X}^\top \mathbf{X} + k \mathbf{I}_p)^{-1}.\end{aligned}$$



Regressão ridge

De (9), nós também temos:

$$\begin{aligned}\text{Var}(\hat{\beta}_R) &= \text{Var}(\mathbf{Z}_k \hat{\beta}) = \mathbf{Z}_k \text{Var}(\hat{\beta}) \mathbf{Z}_k^\top = \sigma^2 \mathbf{Z}_k (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{Z}_k^\top \\ &= \sigma^2 (\mathbf{X}^\top \mathbf{X} + k \mathbf{I}_p)^{-1} (\mathbf{X}^\top \mathbf{X}) (\mathbf{X}^\top \mathbf{X})^{-1} (\mathbf{X}^\top \mathbf{X}) (\mathbf{X}^\top \mathbf{X} + k \mathbf{I}_p)^{-1} \\ &= \sigma^2 (\mathbf{X}^\top \mathbf{X} + k \mathbf{I}_p)^{-1} (\mathbf{X}^\top \mathbf{X}) (\mathbf{X}^\top \mathbf{X} + k \mathbf{I}_p)^{-1}.\end{aligned}$$



Regressão ridge

O erro quadrático médio é dado por:

$$\begin{aligned} \text{EQM}(\hat{\beta}_R) &= \mathbb{E} \left[(\hat{\beta}_R - \beta)^\top (\hat{\beta}_R - \beta) \right] \\ &= \text{tr} \left\{ \sigma^2 (\mathbf{X}^\top \mathbf{X} + k \mathbf{I}_p)^{-1} (\mathbf{X}^\top \mathbf{X}) (\mathbf{X}^\top \mathbf{X} + k \mathbf{I}_p)^{-1} \right\} \quad (10) \\ &\quad + \beta^\top (\mathbf{Z}_k - \mathbf{I}_p)^\top (\mathbf{Z}_k - \mathbf{I}_p) \beta \\ &\stackrel{(?)}{=} \sigma^2 \sum_{m=1}^p \frac{\lambda_m}{(\lambda_m + k)^2} + k^2 \beta^\top (\mathbf{X}^\top \mathbf{X} + k \mathbf{I}_p)^{-2} \beta, \end{aligned}$$



Regressão ridge

O erro quadrático médio é dado por:

$$\begin{aligned} \text{EQM}(\hat{\beta}_R) &= \mathbb{E} \left[(\hat{\beta}_R - \beta)^\top (\hat{\beta}_R - \beta) \right] \\ &= \text{tr} \left\{ \sigma^2 (\mathbf{X}^\top \mathbf{X} + k \mathbf{I}_p)^{-1} (\mathbf{X}^\top \mathbf{X}) (\mathbf{X}^\top \mathbf{X} + k \mathbf{I}_p)^{-1} \right\} \quad (10) \\ &\quad + \beta^\top (\mathbf{Z}_k - \mathbf{I}_p)^\top (\mathbf{Z}_k - \mathbf{I}_p) \beta \\ &\stackrel{(?)}{=} \sigma^2 \sum_{m=1}^p \frac{\lambda_m}{(\lambda_m + k)^2} + k^2 \beta^\top (\mathbf{X}^\top \mathbf{X} + k \mathbf{I}_p)^{-2} \beta, \end{aligned}$$

em que $\lambda_1, \lambda_2, \dots, \lambda_p$ são os autovalores de $\mathbf{X}^\top \mathbf{X}$.



Regressão ridge

O erro quadrático médio é dado por:

$$\begin{aligned} \text{EQM}(\hat{\beta}_R) &= \mathbb{E} \left[(\hat{\beta}_R - \beta)^\top (\hat{\beta}_R - \beta) \right] \\ &= \text{tr} \left\{ \sigma^2 (\mathbf{X}^\top \mathbf{X} + k \mathbf{I}_p)^{-1} (\mathbf{X}^\top \mathbf{X}) (\mathbf{X}^\top \mathbf{X} + k \mathbf{I}_p)^{-1} \right\} \quad (10) \\ &\quad + \beta^\top (\mathbf{Z}_k - \mathbf{I}_p)^\top (\mathbf{Z}_k - \mathbf{I}_p) \beta \\ &\stackrel{(?)}{=} \sigma^2 \sum_{m=1}^p \frac{\lambda_m}{(\lambda_m + k)^2} + k^2 \beta^\top (\mathbf{X}^\top \mathbf{X} + k \mathbf{I}_p)^{-2} \beta, \end{aligned}$$

em que $\lambda_1, \lambda_2, \dots, \lambda_p$ são os autovalores de $\mathbf{X}^\top \mathbf{X}$.



Regressão ridge

O primeiro e o segundo termos do EQM são, respectivamente, a variância e o viés ao quadrado de $\hat{\beta}_R$. Se $k > 0$, notem que, o viés de $\hat{\beta}_R$ aumenta, enquanto a variância diminui.



Regressão ridge

O primeiro e o segundo termos do EQM são, respectivamente, a variância e o viés ao quadrado de $\hat{\beta}_R$. Se $k > 0$, notem que, o viés de $\hat{\beta}_R$ aumenta, enquanto a variância diminui.



Métodos para selecionar k

Hoerl e Kennard (1970b) sugeriram selecionar o valor de k baseado no **traço ridge**, um gráfico com as estimativas de $\hat{\beta}_R$ (no eixo das ordenadas) baseada em k (no eixo das abscissas).

Métodos para selecionar k

Hoerl e Kennard (1970b) sugeriram selecionar o valor de k baseado no **traço ridge**, um gráfico com as estimativas de $\hat{\beta}_R$ (no eixo das ordenadas) baseada em k (no eixo das abscissas).

Enquanto, Hoerl et al. (1975) sugeriram o seguinte método analítico:

$$k = \frac{p\hat{\sigma}^2}{\hat{\beta}^T \hat{\beta}}.$$



Métodos para selecionar k

Hoerl e Kennard (1970*b*) sugeriram selecionar o valor de k baseado no **traço ridge**, um gráfico com as estimativas de $\hat{\beta}_R$ (no eixo das ordenadas) baseada em k (no eixo das abscissas).

Enquanto, Hoerl et al. (1975) sugeriram o seguinte método analítico:

$$k = \frac{p\hat{\sigma}^2}{\hat{\beta}^\top \hat{\beta}}.$$



Métodos para contornar a multicolinearidade

Uma segunda metodologia para contornar a multicolinearidade é a **regressão por componentes principais**, ver, por exemplo, a Seção 9.5.4 de Montgomery et al. (2021).



Roteiro

- 1 Introdução
- 2 Identificando multicolinearidade
- 3 Regressão ridge
- 4 Aplicação**
- 5 Referências bibliográficas



Aplicação

(Montgomery et al., 2021, p. 298) Um conjunto de dados, com 16 observações,



Aplicação

(Montgomery et al., 2021, p. 298) Um conjunto de dados, com 16 observações, relacionado a porcentagem de heptano convertido em acetileno (Y),



Aplicação

(Montgomery et al., 2021, p. 298) Um conjunto de dados, com 16 observações, relacionado a porcentagem de heptano convertido em acetileno (Y), com três variáveis explicativas:



Aplicação

(Montgomery et al., 2021, p. 298) Um conjunto de dados, com 16 observações, relacionado a porcentagem de heptano convertido em acetileno (Y), com três variáveis explicativas: temperatura do reator (x_2 , em $^{\circ}\text{C}$),



Aplicação

(Montgomery et al., 2021, p. 298) Um conjunto de dados, com 16 observações, relacionado a porcentagem de heptano convertido em acetileno (Y), com três variáveis explicativas: temperatura do reator (x_2 , em $^{\circ}\text{C}$), proporção de H_2 (x_3 , em mols)



Aplicação

(Montgomery et al., 2021, p. 298) Um conjunto de dados, com 16 observações, relacionado a porcentagem de heptano convertido em acetileno (Y), com três variáveis explicativas: temperatura do reator (x_2 , em $^{\circ}\text{C}$), proporção de H_2 (x_3 , em mols) e tempo de contato (x_4 , em segundos).



Aplicação

(Montgomery et al., 2021, p. 298) Um conjunto de dados, com 16 observações, relacionado a porcentagem de heptano convertido em acetileno (Y), com três variáveis explicativas: temperatura do reator (x_2 , em $^{\circ}\text{C}$), proporção de H_2 (x_3 , em mols) e tempo de contato (x_4 , em segundos).

Na Figura 1, nós apresentamos o gráfico de dispersão entre as covariáveis x_2 e x_4 .



Aplicação

(Montgomery et al., 2021, p. 298) Um conjunto de dados, com 16 observações, relacionado a porcentagem de heptano convertido em acetileno (Y), com três variáveis explicativas: temperatura do reator (x_2 , em $^{\circ}\text{C}$), proporção de H_2 (x_3 , em mols) e tempo de contato (x_4 , em segundos).

Na Figura 1, nós apresentamos o gráfico de dispersão entre as covariáveis x_2 e x_4 .



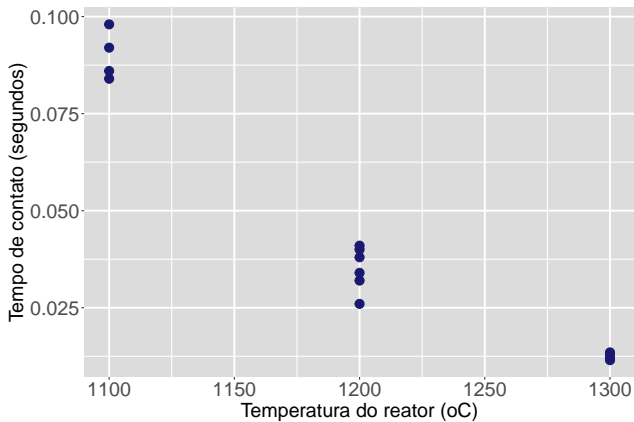


Figura 1: Gráfico de dispersão.

Exemplo

Como as covariáveis estão em escalas muito diferentes, foi feita uma transformação para deixá-las em escala normal de unidade. Em seguida, foi definido o seguinte modelo a ser ajustado:



Exemplo

Como as covariáveis estão em escalas muito diferentes, foi feita uma transformação para deixá-las em escala normal de unidade. Em seguida, foi definido o seguinte modelo a ser ajustado:

$$\begin{aligned} Y_\ell = & \beta_1 + \beta_2 x_{\ell 2} + \beta_3 x_{\ell 3} + \beta_4 x_{\ell 4} \\ & + \beta_5 x_{\ell 2} x_{\ell 3} + \beta_6 x_{\ell 2} x_{\ell 4} + \beta_7 x_{\ell 3} x_{\ell 4} \\ & + \beta_8 x_{\ell 2}^2 + \beta_9 x_{\ell 3}^2 + \beta_{10} x_{\ell 4}^2 + \varepsilon_\ell, \end{aligned}$$

$$\ell = 1, 2, \dots, 16.$$



Exemplo

Como as covariáveis estão em escalas muito diferentes, foi feita uma transformação para deixá-las em escala normal de unidade. Em seguida, foi definido o seguinte modelo a ser ajustado:

$$\begin{aligned} Y_\ell = & \beta_1 + \beta_2 x_{\ell 2} + \beta_3 x_{\ell 3} + \beta_4 x_{\ell 4} \\ & + \beta_5 x_{\ell 2} x_{\ell 3} + \beta_6 x_{\ell 2} x_{\ell 4} + \beta_7 x_{\ell 3} x_{\ell 4} \\ & + \beta_8 x_{\ell 2}^2 + \beta_9 x_{\ell 3}^2 + \beta_{10} x_{\ell 4}^2 + \varepsilon_\ell, \end{aligned}$$

$$\ell = 1, 2, \dots, 16.$$



Exemplo

Pela Figura 1, há um forte indício de uma relação entre as variáveis x_2 e x_4 . Para aplicação das técnicas aprendidas, nós iremos transformar o banco de dados para a escala de tamanho unitário.



Exemplo

Pela Figura 1, há um forte indício de uma relação entre as variáveis x_2 e x_4 . Para aplicação das técnicas aprendidas, nós iremos transformar o banco de dados para a escala de tamanho unitário. Em seguida, nós definimos o seguinte modelo a ser ajustado:

$$\begin{aligned} Y_l^* &= \beta_2^* x_{l2}^* + \beta_3^* x_{l3}^* + \beta_4^* x_{l4}^* \\ &+ \beta_5^* x_{l2}^* x_{l3}^* + \beta_6^* x_{l2}^* x_{l4}^* + \beta_7^* x_{l3}^* x_{l4}^* \\ &+ \beta_8^* x_{l2}^{2*} + \beta_9^* x_{l3}^{2*} + \beta_{10}^* x_{l4}^{2*} + \varepsilon_l, \end{aligned}$$

$$l = 1, 2, \dots, 16.$$



Exemplo

Pela Figura 1, há um forte indício de uma relação entre as variáveis x_2 e x_4 . Para aplicação das técnicas aprendidas, nós iremos transformar o banco de dados para a escala de tamanho unitário. Em seguida, nós definimos o seguinte modelo a ser ajustado:

$$\begin{aligned} Y_l^* &= \beta_2^* x_{l2}^* + \beta_3^* x_{l3}^* + \beta_4^* x_{l4}^* \\ &+ \beta_5^* x_{l2}^* x_{l3}^* + \beta_6^* x_{l2}^* x_{l4}^* + \beta_7^* x_{l3}^* x_{l4}^* \\ &+ \beta_8^* x_{l2}^{2*} + \beta_9^* x_{l3}^{2*} + \beta_{10}^* x_{l4}^{2*} + \varepsilon_l, \end{aligned}$$

$$l = 1, 2, \dots, 16.$$



Exemplo

Tabela 1: Estimativas do parâmetros.

Parâmetro	Estimativa	EP	t_c
β_2^*	0,336	0,351	0,959
β_3^*	0,233	0,024	9,773
β_4^*	-0,676	0,472	-1,431
β_5^*	-0,480	0,101	-4,757
β_6^*	-2,034	1,467	-1,386
β_7^*	-0,266	0,108	-2,459
β_8^*	-0,835	0,760	-1,098
β_9^*	-0,090	0,032	-2,805
β_{10}^*	-1,001	0,616	-1,625

Região crítica, para $\alpha = 5\%$: $|t_c| > 2,365$, com $QMRes = 0,00033$.

Exemplo

Tabela 2: Matriz de correlações.

1,00	0,22	-0,96	-0,13	0,44	0,21	-0,27	0,03	-0,58
0,22	1,00	-0,24	0,04	0,19	-0,02	-0,15	0,50	-0,22
-0,96	-0,24	1,00	0,19	-0,66	-0,27	0,50	-0,02	0,77
-0,13	0,04	0,19	1,00	-0,26	-0,97	0,25	0,40	0,27
0,44	0,19	-0,66	-0,26	1,00	0,32	-0,97	0,13	-0,97
0,21	-0,02	-0,27	-0,97	0,32	1,00	-0,28	-0,37	-0,36
-0,27	-0,15	0,50	0,25	-0,97	-0,28	1,00	-0,12	0,89
0,03	0,50	-0,02	0,40	0,13	-0,37	-0,12	1,00	-0,16
-0,58	-0,22	0,77	0,27	-0,97	-0,36	0,89	-0,16	1,00



Exemplo

Tabela 3: Medidas de multicolinearidade.

	x_2	x_3	x_4	x_2x_3	x_2x_4	x_3x_4	x_2^2	x_3^2	x_4^2
VIF	375,25	1,74	680,28	31,04	6563,35	35,61	1762,58	3,16	1156,77
AV	4,21	2,16	1,14	1,04	0,39	0,05	0,01	0,01	0,00
κ	1,00	1,95	3,69	4,04	10,92	84,89	308,63	820,08	43381,31

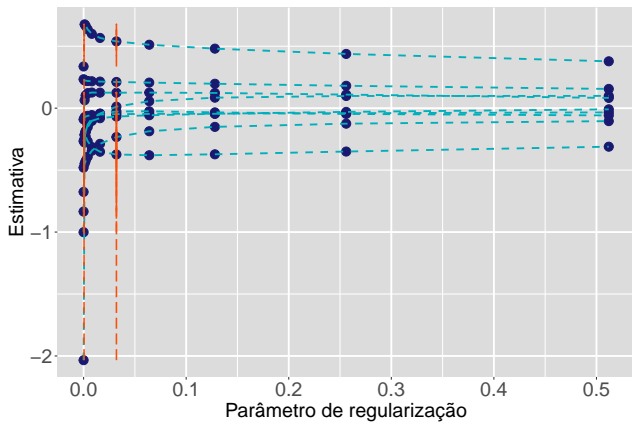


Figura 2: Traço *ridge*.

Tabela 4: Estimativas dos coeficientes.

	EMQ (0,000)	Ridge (0,032)
β_2	0,336	0,539
β_3	0,233	0,212
β_4	-0,676	-0,374
β_5	-0,480	-0,233
β_6	-2,034	-0,068
β_7	-0,266	0,012
β_8	-0,835	0,125
β_9	-0,090	-0,048
β_{10}	-1,001	-0,027

Roteiro

- 1 Introdução
- 2 Identificando multicolinearidade
- 3 Regressão ridge
- 4 Aplicação
- 5 Referências bibliográficas



Referências bibliográficas I

Gauss, C. F. (1821), 'Theoria combinationis observationum erroribus minimis obnoxiae - part 1', *Werke* **4**, 1–108.

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

Hoerl, A. E. e Kennard, R. W. (1970a), 'Ridge regression: Applications to nonorthogonal problems', *Technometrics* **12**(1), 69–82.

Hoerl, A. E. e Kennard, R. W. (1970b), 'Ridge regression: biased estimation for nonorthogonal problems', *Technometrics* **12**(1), 55–67.



Referências bibliográficas II

Hoerl, A. E., Kennard, R. W. e Baldwin, K. F. (1975), 'Ridge regression: some simulations', *Communications in Statistics - Theory and Methods* **4**(2), 105–123.

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



Obrigado!

✉ tiago.magalhaes@ufjf.br

🌐 ufjf.br/tiago_magalhaes

🌐 Departamento de Estatística, Sala 319

