

ME720 - Modelos Lineares Generalizados

Parte 1 - Introdução e Revisão de Álgebra de Matrizes

Profa. **Larissa Avila Matos**

Introdução

Os métodos estatísticos são amplamente usados como parte do processo de aprendizagem do método científico. Ao longo do curso, nós iremos estudar os seguintes modelos:

- 1 Modelos Lineares;
- 2 Modelos Lineares Generalizados;

A escolha de um modelo deve ser pautada: nos objetivos do experimento, nas características dos dados, em experiências anteriores e na análise descritiva.

Modelo de regressão linear simples

Na regressão linear simples, nós nos preocupamos em modelar a relação entre duas variáveis. Exemplos:

- Rendimento salarial e o número de anos de educação;
- Altura e o peso das pessoas;
- Altitude e a temperatura de ebulição da água;

Para uma relação linear, nós usamos um modelo da forma:

$$y = \beta_0 + \beta_1 x + \epsilon, \tag{1}$$

onde y é a variável resposta (dependente), x é a variável preditora (independente), ϵ é o termo aleatório do modelo (o erro no modelo). O erro representa flutuações aleatórias, erros de medidas ou o efeito de fatores não controlados.

Modelos de regressão linear múltipla

Muitas vezes a resposta y é influenciada por mais de uma variável preditora.

Exemplo:

- A produção de uma colheita pode depender das quantidades de nitrogênio, potássio e fosfato;
- As variáveis nitrogênio, potássio e fosfato são controladas pelo experimentador, mas a produção também pode depender de variáveis não controladas como aquelas associadas com o tempo.

Para uma relação linear, nós usamos um modelo da forma:

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + \epsilon \quad (2)$$

onde os parâmetros $\beta_0, \beta_1, \beta_2, \dots, \beta_k$ são chamados coeficientes de regressão, ϵ é a variação aleatória em y não explicada pelos x 's.

Modelos de regressão linear múltipla

Os modelos de regressão tais como em (1) e (2) são usados para vários propósitos, incluindo os seguintes:

- 1 Predição;
- 2 Descrição ou Exploração dos Dados;
- 3 Estimação dos Parâmetros;
- 4 Seleção de variáveis.

Modelos de análise de variância

Em modelos de análise de variância, nós estamos interessados em comparar diversas populações ou comparar diversas condições em um experimento. Modelos de análise de variância podem ser expressos como modelos lineares de valores restritos de x . Frequentemente os x 's são 0's ou 1's.

Exemplo:

Suponha que um pesquisador deseje comparar o rendimento de quatro catalisadores em um processo industrial. Se n observações são obtidas para cada catalisador, um modelo para as $4n$ observações pode ser expresso como:

$$y_{ij} = \mu_i + \epsilon_{ij}, \quad i = 1, 2, 3, 4, \quad j = 1, 2, \dots, n, \quad (3)$$

onde μ_i é a média correspondente ao i -ésimo catalisador. Uma hipótese de interesse é $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$.

Modelos de análise de variância

O modelo em 3 pode ser expresso de uma forma alternativa como

$$y_{ij} = \mu + \alpha_i + \epsilon_{ij}, \quad i = 1, 2, 3, 4, \quad j = 1, 2, \dots, n. \quad (4)$$

Nesta forma, α_i é o efeito do i -ésimo catalisador e a hipótese de interesse pode ser expressa como $H_0 : \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 0$.

Modelos de análise de variância

Suponha agora, que o pesquisador também deseje comparar o efeito de três níveis de temperatura e que n observações são tomadas em cada uma das 12 combinações catalisador-temperatura. Então o modelo pode ser expresso como

$$\begin{aligned} y_{ijk} &= \mu_{ij} + \epsilon_{ijk} \\ &= \mu + \alpha_i + \beta_j + \gamma_{ij} + \epsilon_{ijk}, \quad i = 1, 2, 3, 4, \quad j = 1, 2, 3, \quad k = 1, \dots, n \end{aligned} \quad (5)$$

onde μ_{ij} é a média da (ij) -ésima combinação catalisador-temperatura, α_i é o efeito do i -ésimo catalisador, β_j é o efeito do j -ésimo nível de temperatura, γ_{ij} é a interação ou efeito conjunto do i -ésimo catalisador e j -ésimo nível de temperatura.

Modelos de análise de variância podem ser tratados como um caso especial de modelos de regressão, mas é mais conveniente analisá-los separadamente.

Modelos Lineares Generalizados (MLGs)

Nelder e Wedderburn (1972), propuseram os Modelos Lineares Generalizados (MLGs), que são uma extensão dos modelos normais lineares.

A idéia básica consiste em abrir o leque de opções para a distribuição da variável resposta, permitindo que a mesma pertença à família exponencial de distribuições, bem como dar maior flexibilidade para a relação funcional entre a média da variável resposta (μ) e o preditor linear η .

A ligação entre a média e o preditor linear não é necessariamente a identidade, podendo assumir qualquer forma monótona não-linear.

O modelo de regressão linear (modelos lineares) utiliza a linearidade para descrever a relação entre a média da variável resposta e um conjunto de variáveis explicativas, assumindo que a distribuição da variável resposta é normal. Os modelos lineares generalizados (MLGs) estendem os modelos de regressão linear para abranger distribuições de respostas não-normais e possivelmente funções não-lineares da média. Eles têm três componentes.

Componente aleatório: especifica a variável de resposta \mathbf{y} e sua distribuição de probabilidade. As observações $\mathbf{y} = (y_1, \dots, y_n)$ nessa distribuição são tratadas como independentes.

Preditor linear: Para um vetor de parâmetros $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)'$ e uma matriz $\mathbf{X}_{n \times p}$ do modelo que contém valores de p variáveis explicativas para as n observações, o preditor linear é $\mathbf{X}\boldsymbol{\beta} = \boldsymbol{\eta}$.

Função de ligação: Esta é uma função g aplicada a cada componente de $E(\mathbf{y}) = \boldsymbol{\mu}$ que a relaciona com o preditor linear,

$$g(E(\mathbf{y})) = \mathbf{X}\boldsymbol{\beta}.$$

Antes dos MLGs

Os MLGs foram criados com o objetivo de reunir numa mesma família vários modelos estatísticos que eram tratados separadamente.

Em geral, nas análises de regressão, procurava-se algum tipo de transformação que levasse à normalidade, tais como a transformação de Box-Cox (1964) dada abaixo

$$z = \begin{cases} \frac{y^\lambda - 1}{\lambda}, & \text{se } \lambda \neq 0, \\ \log(y), & \text{se } \lambda = 0; \end{cases}$$

em que $y > 0$ e λ é uma constante desconhecida.

Exemplo: Exposição de Bactérias

Vamos considerar um exemplo em que modelos de regressão normal linear são comparados com um modelo log-linear de Poisson para ajustar dados de contagem.

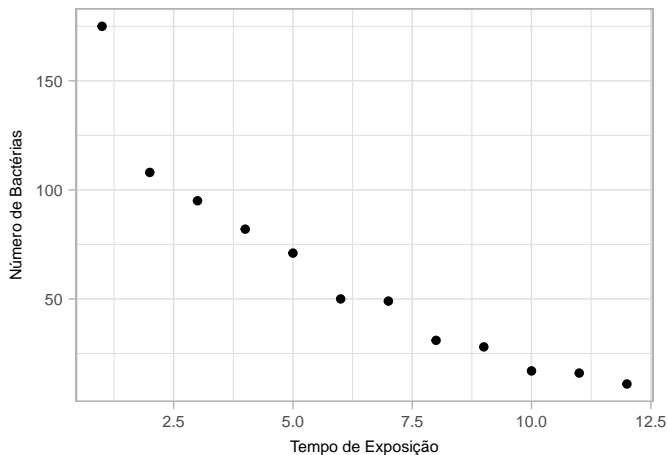
Dados - Variáveis:

- variável resposta: número de bactérias sobreviventes em amostras de um produto alimentício exposto a uma temperatura de 300°F.
- variável explicativa: tempo de exposição do produto (em minutos).

(Montgomery, Peck e Vining, 2001) (Paula, 2013a).

Descrição dos Dados:

Bactérias	175	108	95	82	71	50	49	31	28	17	16	11
Exposição	1	2	3	4	5	6	7	8	9	10	11	12



Exemplo: Exposição de Bactérias

Ajuste Modelos Normais: Com base na aproximação da Poisson pela normal vamos propor inicialmente dois modelos.

1 Modelo Linear

$$\sqrt{y_i} = \alpha + \beta \text{ tempo}_i + \epsilon_i,$$

em que $\epsilon_i \sim N(0, \sigma^2)$.

2 Modelo Quadrático

$$\sqrt{y_i} = \alpha + \beta \text{ tempo}_i + \gamma \text{ tempo}_i^2 + \epsilon_i,$$

em que $\epsilon_i \sim N(0, \sigma^2)$.

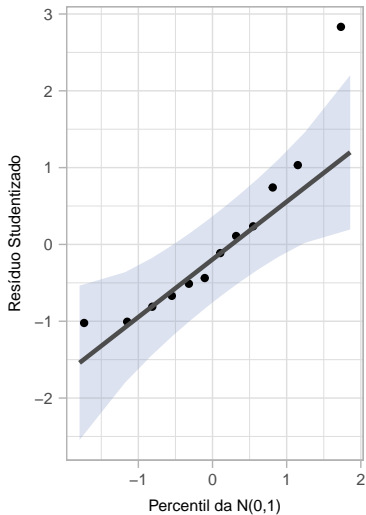
```
fit1<-lm(sqrt(bacterias)~exposicao)
summary(fit1)$coefficients
```

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	12.5727346	0.38090964	33.00713	1.537569e-11
exposicao	-0.8160687	0.05175543	-15.76779	2.162124e-08

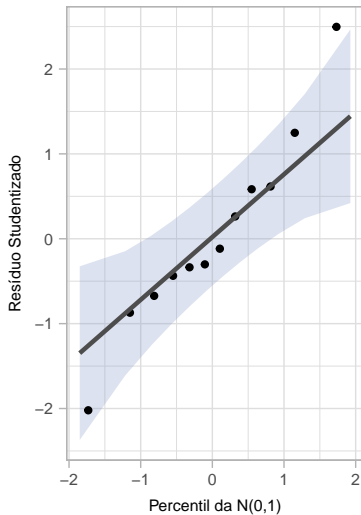
```
exposicao2<-exposicao^2
fit2<-lm(sqrt(bacterias)~exposicao+exposicao2)
summary(fit2)$coefficients
```

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	13.64444308	0.50683805	26.920716	6.518714e-10
exposicao	-1.27537234	0.17925714	-7.114764	5.576294e-05
exposicao2	0.03533105	0.01342333	2.632063	2.726626e-02

Modelo Linear



Modelo Quadrático



Exemplo: Exposição de Bactérias

Ajuste Modelo Log-linear de Poisson: Vamos supor agora o seguinte modelo log-linear de Poisson,

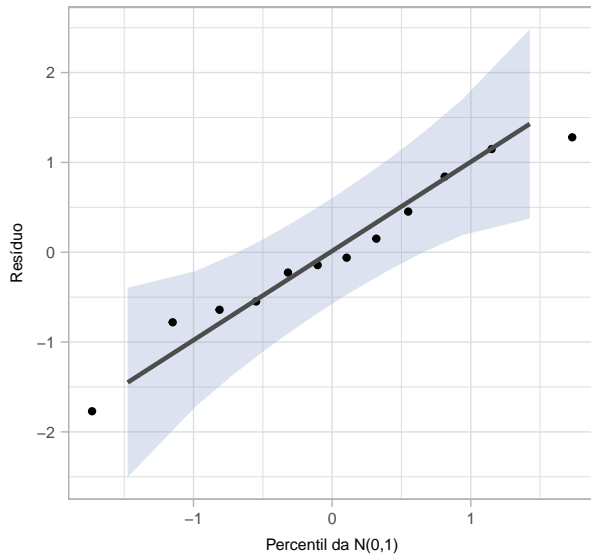
$$\log(\mu_i) = \alpha + \beta \text{ tempo}_i,$$

em que $y_i \sim P(\mu_i)$. As estimativas desse modelo são apresentadas abaixo.

```
fit3<-glm(bacterias~exposicao,family=poisson(link = "log"))  
summary(fit3)$coefficients
```

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	5.3055715	0.06348273	83.57504	0.000000e+00
exposicao	-0.2288956	0.01269994	-18.02336	1.277325e-72

GLM



Exemplo: Exposição de Bactérias

O modelo ajustado fica então dado por

$$\hat{\mu}(x) = e^{5,30-0,23x},$$

em que x denota o tempo de exposição.

Logo, se diminuirmos uma unidade o tempo de exposição a variação no valor esperado fica dada por

$$\frac{\hat{\mu}(x-1)}{\hat{\mu}(x)} = \frac{e^{5,30-0,23(x-1)}}{e^{5,30-0,23x}} = e^{0,23} = 1,259$$

Ou seja, o número esperado de sobreviventes aumenta aproximadamente 25,9%.

Revisão de Álgebra de Matrizes - Matrizes, vetores e escalares

Vetor e Matriz

- Uma matriz é um arranjo retangular de elementos (números, variáveis aleatórias, letras):

$$\mathbf{A}_{(n \times p)} = \begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1p} \\ A_{21} & A_{22} & \cdots & A_{2p} \\ \vdots & & & \\ A_{n1} & A_{n2} & \cdots & A_{np} \end{bmatrix}$$

- Um vetor é uma matriz com somente uma linha (vetor linha) ou somente uma coluna (vetor coluna).

$$\mathbf{a}_{(1 \times p)} = \begin{bmatrix} a_{11} & \cdots & a_{1p} \end{bmatrix}; \quad \mathbf{a}_{(p \times 1)} = \begin{bmatrix} a_{11} \\ \vdots \\ a_{p1} \end{bmatrix}$$

- Se a matriz tiver apenas uma linha e uma coluna teremos um escalar (a).
- Denotaremos uma matriz por uma letra maiúscula em negrito (\mathbf{A}), um vetor por uma letra minúscula em negrito (\mathbf{a}) e um escalar por uma letra minúscula (a).
- Uma matriz é dita ser quadrada se o número de linhas for igual ao número de colunas.

$$\mathbf{A}_{(3 \times 3)} = \begin{bmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{bmatrix}$$

Matriz transposta

Matriz transposta: cada coluna da matriz original é transformada em uma linha.

Se \mathbf{A} é uma matriz $n \times p$ então a sua transposta (\mathbf{A}') é uma matriz $p \times n$.

Exemplo:

$$\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 4 & -1 \\ 3 & 2 \end{bmatrix} \rightarrow \mathbf{A}' = \begin{bmatrix} 1 & 4 & 3 \\ 2 & -1 & 2 \end{bmatrix}$$

Tipos especiais de matrizes

- Matriz simétrica: $A_{ij} = A_{ji}, \forall i, j; i \neq j$, ou equivalentemente, $\mathbf{A} = \mathbf{A}'$.

$$\mathbf{A} = \begin{bmatrix} 5 & -2 & 3 \\ -2 & 7 & 1 \\ 3 & 1 & 4 \end{bmatrix}$$

- Matriz diagonal: matriz quadrada em que todos os elementos fora da diagonal principal são iguais à 0.

$$\mathbf{A} = \begin{bmatrix} A_{11} & 0 & 0 \\ 0 & A_{22} & 0 \\ 0 & 0 & A_{33} \end{bmatrix}$$

- Matriz identidade: matriz diagonal cujos elementos da diagonal principal são todos iguais à 1.

$$\mathbf{I} = \mathbf{I}_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}; \quad \mathbf{I}_n = \begin{bmatrix} 1 & \cdots & 0 \\ & \ddots & \\ 0 & \cdots & 1 \end{bmatrix}_{n \times n}$$

- Matriz diagonal inferior: matriz quadrada em que todos os elementos acima da diagonal principal são iguais à 0.

$$\mathbf{A} = \begin{bmatrix} A_{11} & 0 & 0 \\ A_{12} & A_{22} & 0 \\ A_{31} & A_{32} & A_{33} \end{bmatrix}$$

- Matriz diagonal superior: matriz quadrada em que todos os elementos abaixo da diagonal principal são iguais à 0.

$$\mathbf{A} = \begin{bmatrix} A_{11} & A_{12} & A_{13} \\ 0 & A_{22} & A_{23} \\ 0 & 0 & A_{33} \end{bmatrix}$$

Operações com matrizes

- Soma de matrizes: Sejam **A** e **B** matrizes de tamanho $n \times n$ então, $\mathbf{C} = \mathbf{A} + \mathbf{B}$, onde $C_{ij} = A_{ij} + B_{ij}$.

$$\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \\ A_{31} & A_{32} \end{bmatrix} + \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \\ B_{31} & B_{32} \end{bmatrix} = \begin{bmatrix} A_{11} + B_{11} & A_{12} + B_{12} \\ A_{21} + B_{21} & A_{22} + B_{22} \\ A_{31} + B_{31} & A_{32} + B_{32} \end{bmatrix}$$

Obs.: Diferença de matrizes: Sejam **A** e **B** matrizes de tamanho $n \times n$ então, $\mathbf{D} = \mathbf{A} - \mathbf{B}$, onde $D_{ij} = A_{ij} - B_{ij}$.

Multiplicação de matrizes

Para que o produto de duas matrizes (\mathbf{AB}) seja possível, o número de colunas de \mathbf{A} deve ser igual ao número de linhas de \mathbf{B} . Neste caso, dizemos que \mathbf{A} e \mathbf{B} são conformes. Então, $\mathbf{C} = \mathbf{AB}$ é definido como

$$C_{ij} = \sum_k A_{ik} B_{kj}.$$

Exemplo:

$$\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \\ A_{31} & A_{32} \end{bmatrix} \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix} = \begin{bmatrix} A_{11}B_{11} + A_{12}B_{21} & A_{11}B_{12} + A_{12}B_{22} \\ A_{21}B_{11} + A_{22}B_{21} & A_{21}B_{12} + A_{22}B_{22} \\ A_{31}B_{11} + A_{32}B_{21} & A_{31}B_{12} + A_{32}B_{22} \end{bmatrix}$$

Multiplicação de vetores

A multiplicação envolvendo vetores segue as mesmas regras das matrizes.

Suponha $\mathbf{b}_{(p \times 1)}$, então:

$$\mathbf{b} = \begin{bmatrix} b_1 \\ \vdots \\ b_p \end{bmatrix} \Rightarrow \mathbf{b}'\mathbf{b} = \sum_{i=1}^p b_i^2; \quad \mathbf{b}\mathbf{b}' = \begin{bmatrix} b_1^2 & \cdots & b_1 b_p \\ & \ddots & \\ b_p b_1 & \cdots & b_p^2 \end{bmatrix};$$

onde $\mathbf{b}'\mathbf{b}$ é uma soma de quadrados e $\mathbf{b}\mathbf{b}'$ é uma matriz quadrada e simétrica.

Observação:

- Norma euclidiana, ou comprimento do vetor \mathbf{b} :

$$\|\mathbf{b}\| = \sqrt{\mathbf{b}'\mathbf{b}} = \sqrt{\sum_{i=1}^p b_i^2}$$

- Se $\mathbf{A}_{(p \times p)}$ é uma matriz simétrica e \mathbf{y} é um vetor $p \times 1$, o produto $\mathbf{y}'\mathbf{A}\mathbf{y} = \sum_{i=1}^p A_{ii}y_i^2 + 2 \sum_{i \neq j} A_{ij}y_i y_j$ é chamado de forma quadrática.
- Se \mathbf{x} é $n \times 1$, \mathbf{y} é $p \times 1$ e \mathbf{A} é $n \times p$, o produto $\mathbf{x}'\mathbf{A}\mathbf{y} = \sum_{i,j} A_{ij}x_i y_j$ é chamado de forma bilinear.

Produto de Kronecker:

Sejam $\mathbf{A}_{(n \times p)}$ e $\mathbf{B}_{(m \times q)}$ duas matrizes quaisquer. O produto de Kronecker entre elas é definido por:

$$\mathbf{A} \otimes \mathbf{B} = \begin{bmatrix} A_{11}\mathbf{B} & A_{12}\mathbf{B} & \dots & A_{1p}\mathbf{B} \\ A_{21}\mathbf{B} & A_{22}\mathbf{B} & \dots & A_{2p}\mathbf{B} \\ \vdots & \vdots & \vdots & \vdots \\ A_{n1}\mathbf{B} & A_{n2}\mathbf{B} & \dots & A_{np}\mathbf{B} \end{bmatrix}$$

A resultante do produto de Kronecker será uma matriz $\mathbf{C}_{(nm \times pq)}$.

$\mathbf{A} \otimes \mathbf{B} \neq \mathbf{B} \otimes \mathbf{A}$, em geral.

Se as dimensões são favoráveis,

$$(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = \mathbf{AC} \otimes \mathbf{BD}$$

Potência de matrix quadrada

Seja \mathbf{A} uma matriz quadrada e k um número inteiro e positivo ($k \in \mathbb{Z}$), a k -ésima potência de \mathbf{A} é definida como:

$$\mathbf{A}^k = \underbrace{\mathbf{A} \cdots \mathbf{A}}_{k \text{ vezes}}$$

Em relação à sua segunda potência, uma matriz quadrada \mathbf{A} , será chamada de:

- 1 idempotente, se $\mathbf{A}^2 = \mathbf{A}$.
- 2 nilpotente, se $\mathbf{A}^2 = \mathbf{0}$.
- 3 unipotente, se $\mathbf{A}^2 = \mathbf{I}$.

Posto (rank) de uma matriz

O posto (rank) de qualquer matriz \mathbf{A} (quadrada ou retangular) é definido como o número de colunas (linhas) linearmente independentes de \mathbf{A} .

- Vetores linearmente dependente (*l.d.*) / linearmente independente(*l.i.*)

Um conjunto de vetores $\{\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_p\}$ é dito ser *l.d.* se existe conjunto de escalares c_1, c_2, \dots, c_p (nem todos nulos) que:

$$c_1 \mathbf{a}_1 + c_2 \mathbf{a}_2 + \dots + c_p \mathbf{a}_p = \mathbf{0}.$$

Se $c_1 = c_2 = \dots, c_p = 0$ for a única solução, então o conjunto de vetores $\{\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_p\}$ é dito ser *l.i.*.

Posto (rank) de uma matriz

Observações:

- 1 Se a matriz retangular $\mathbf{A}_{(n \times p)}$ de posto p , onde $p < n$, então \mathbf{A} tem maior posto possível e é dito ter posto coluna completo.
- 2 Em geral, o maior posto possível de uma matrix $\mathbf{A}_{(n \times p)}$ é o $\min(n, p)$. Assim, em uma matrix retangular, as linhas, as colunas ou ambas são linearmente dependentes.

Exemplo: $\mathbf{A} = \begin{bmatrix} 1 & -2 & 3 \\ 5 & 2 & 4 \end{bmatrix}$

Traço de uma matriz

O traço de uma matriz quadrada $\mathbf{A}_{(p \times p)}$ é a soma dos elementos de sua diagonal principal. Então,

$$tr(\mathbf{A}) = \sum_{i=1}^p A_{ii}.$$

$$\text{Exemplo: } tr(\mathbf{A}) = tr \begin{bmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{bmatrix} = A_{11} + A_{22} + A_{33}$$

Operador *vec*

O operador *vec* cria um vetor coluna, a partir de uma matriz, pela concatenação de suas colunas:

$$\text{vec}(\mathbf{A}) = \text{vec} \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \\ A_{31} & A_{32} \end{bmatrix} = \begin{bmatrix} A_{11} \\ A_{21} \\ A_{31} \\ A_{12} \\ A_{22} \\ A_{32} \end{bmatrix}$$

$$\text{vec}(\mathbf{A}) = \text{vec} \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} = \begin{bmatrix} A_{11} \\ A_{21} \\ A_{12} \\ A_{22} \end{bmatrix}$$

Determinantes

O determinante de uma matriz quadrada é uma função que associa um escalar para uma dada matriz quadrada; ela transforma essa matriz em um número real. Existem alguns métodos para se calcular o determinante de uma matriz. O determinante de uma matriz $\mathbf{A}_{(p \times p)}$ é denotado por

$$|\mathbf{A}| = \det(\mathbf{A})$$

Exemplo:

$$|\mathbf{A}| = \det(\mathbf{A}) = \det \begin{bmatrix} 1 & 2 \\ 4 & -1 \end{bmatrix} = -1 - 8 = -9$$

Inversa de uma matriz

A inversa de uma matriz quadrada \mathbf{A} , denotada por \mathbf{A}^{-1} , é tal que:

$$\mathbf{A}\mathbf{A}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}.$$

- Se a matriz \mathbf{A} tem uma inversa, a inversa é única e \mathbf{A} é dita ser *não singular*.
- Se \mathbf{A} é singular $\Rightarrow \det(\mathbf{A}) = 0$.
- Se \mathbf{A} é não singular $\Rightarrow \det(\mathbf{A}) \neq 0$ e $\det(\mathbf{A}^{-1}) = 1/\det(\mathbf{A})$

Existem diversos métodos para se obter (numericamente) a inversa de uma matriz. A solução analítica é dada por:

$$\mathbf{A}^{-1} = \frac{1}{|\mathbf{A}|}\mathbf{C},$$

em que \mathbf{C} é a matriz adjunta de \mathbf{A} , ou seja, a transposta da matriz que se obtém substituindo cada termo A_{ij} pelo determinante da matriz resultante ao se retirar de \mathbf{A} a linha i e a coluna j multiplicado por $(-1)^{i+j}$.

Naturalmente, se $\det(\mathbf{A}) = 0$ sua inversa não existe.

Inversa generalizada

Uma inversa generalizada de uma matriz $\mathbf{A}_{(n \times p)}$ (não necessariamente quadrada), é qualquer matriz \mathbf{A}^- tal que

$$\mathbf{A}\mathbf{A}^-\mathbf{A} = \mathbf{A}.$$

Em princípio, uma matriz pode ter infinitas inversas generalizadas, exceto quando a matriz é não singular ($\mathbf{A}^- = \mathbf{A}^{-1}$).

Mesmo que o $\det(\mathbf{A}) = 0$ sua inversa generalizada pode ser obtida.

Se \mathbf{A} é $n \times p$ então qualquer inversa generalizada \mathbf{A}^- é $p \times n$.

Autovalores e autovetores

Para qualquer matriz quadrada \mathbf{A} , um escalar λ e um vetor não nulo \mathbf{x} podem ser encontrados de tal forma que

$$\mathbf{A}\mathbf{x} = \lambda\mathbf{x},$$

onde λ é chamado um autovalor de \mathbf{A} e \mathbf{x} um autovetor de \mathbf{A} .

Para encontrar λ e \mathbf{x} para uma matriz \mathbf{A} , nós escrevemos a equação acima como:

$$(\mathbf{A} - \lambda\mathbf{I})\mathbf{x} = \mathbf{0}.$$

Temos que $(\mathbf{A} - \lambda\mathbf{I})\mathbf{x}$ é uma combinação das colunas de $\mathbf{A} - \lambda\mathbf{I}$ e essas colunas são linearmente dependentes. Assim a matriz quadrada $\mathbf{A} - \lambda\mathbf{I}$ é singular, e nós podemos encontrar λ usando

$$|\mathbf{A} - \lambda\mathbf{I}| = 0.$$

Os autovalores de uma matriz quadrada de dimensão n são os n números, digamos $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \dots, \lambda_n)'$, que satisfazem

$$|\mathbf{A} - \lambda \mathbf{I}| = 0.$$

Então, os autovetores de uma matriz quadrada de dimensão n são os n vetores que satisfazem a seguinte relação

$$\mathbf{A}\mathbf{x}_i = \lambda_i \mathbf{x}_i; i = 1, \dots, n.$$

Os autovalores e autovetores são de extrema importância na decomposição e na verificação de certas propriedades de matrizes.

Autovetor normalizado: $\mathbf{x}_i \mathbf{x}_i' = 1$.

Tipos de matrizes

Uma matriz \mathbf{A} (quadrada) é dita ser:

- Idempotente se: $\mathbf{A}\mathbf{A} = \mathbf{A}$.
- Ortogonal se: $\mathbf{A}^{-1} = \mathbf{A}'$.
- Positiva definida: se $\lambda_i > 0, \forall i$.
- Positiva semi-definida: se $\lambda_i \geq 0, \forall i$ e \exists pelo menos um $\lambda_i = 0$.
- Negativa semi-definida: se $\lambda_i \leq 0, \forall i$ e \exists pelo menos um $\lambda_i = 0$.
- Negativa definida: se $\lambda_i < 0, \forall i$.

Decomposição Espectral

Seja $\mathbf{A}_{n \times n}$ simétrica e com autovalores $\lambda_1, \dots, \lambda_n$ e autovetores normalizados $\mathbf{x}_1, \dots, \mathbf{x}_n$. Então, \mathbf{A} pode ser expressa como

$$\mathbf{A} = \mathbf{E} \mathbf{\Lambda} \mathbf{E}',$$

em que $\mathbf{\Lambda} = \text{diag}(\lambda_1, \dots, \lambda_n)$ (autovalores) e as colunas da matriz \mathbf{E} são formadas pelos respectivos autovetores ortonormalizados, matrix ortonormal, $\mathbf{E} = [\mathbf{x}_1, \dots, \mathbf{x}_n]$.

A decomposição acima também vale se autovetores não estiverem ortonormalizados, neste caso, ela é dada por:

$$\mathbf{A} = \mathbf{E} \mathbf{\Lambda} \mathbf{E}^{-1}$$

Decomposição de Cholesky

Seja $\mathbf{A}_{p \times p}$, simétrica e positiva definida, ou seja, todos os seus autovalores são positivos (exemplo, matriz de covariâncias e matriz de correlações).

Então, \mathbf{A} pode ser expressa como

$$\mathbf{A} = \mathbf{L}\mathbf{L}',$$

onde \mathbf{L} é uma matriz diagonal inferior com os valores da diagonal estritamente positivos.

Tal decomposição é única.

Exemplo:

$$\Sigma = \begin{bmatrix} 3 & 1 & 0.5 \\ 1 & 2 & 0.8 \\ 0.5 & 0.8 & 4 \end{bmatrix}$$

$$\mathbf{L} \approx \begin{bmatrix} 1,732 & 0 & 0 \\ 0,577 & 1,291 & 0 \\ 0,289 & 0,491 & 1,917 \end{bmatrix}$$

Existem alguns algoritmos que possibilitam a obtenção de \mathbf{L} .

Propriedades

Algumas propriedades importantes: Se todas as operações estiverem bem definidas, teremos:

- $(\mathbf{A}')^{-1} = (\mathbf{A}^{-1})'$.
- $(\mathbf{AB})' = \mathbf{B}'\mathbf{A}'$.
- $(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$.
- $tr(\mathbf{A} + \mathbf{B}) = tr(\mathbf{A}) + tr(\mathbf{B})$.
- $tr(a\mathbf{A}) = a tr(\mathbf{A})$, a um escalar.
- $tr(\mathbf{AB}) = tr(\mathbf{BA})$.
- $det(\mathbf{A}) = det(\mathbf{A}')$.
- $det(\mathbf{A}^{-1}) = [det(\mathbf{A})]^{-1}$.

- $\det(\mathbf{AB}) = \det(\mathbf{A})\det(\mathbf{B})$ se \mathbf{A} e \mathbf{B} forem ambas quadradas.
- $\mathbf{A} \otimes (\mathbf{B} \otimes \mathbf{C}) = (\mathbf{A} \otimes \mathbf{B}) \otimes \mathbf{C}$.
- $a \otimes \mathbf{A} = \mathbf{A} \otimes a = a\mathbf{A}$, a escalar.
- $a\mathbf{A} \otimes b\mathbf{B} = ab(\mathbf{A} \otimes \mathbf{B})$.
- $(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = \mathbf{AC} \otimes \mathbf{BD}$.
- $(\mathbf{A} \otimes \mathbf{B})' = (\mathbf{A}' \otimes \mathbf{B}')$.
- $(\mathbf{A} \otimes \mathbf{B})^{-1} = (\mathbf{A}^{-1} \otimes \mathbf{B}^{-1})$.
- $\text{tr}(\mathbf{A} \otimes \mathbf{B}) = \text{tr}(\mathbf{A})\text{tr}(\mathbf{B})$.
- Para $\mathbf{A}_{(m \times m)}$ e $\mathbf{B}_{(n \times n)}$, temos que $|\mathbf{A} \otimes \mathbf{B}| = |\mathbf{A}|^m |\mathbf{B}|^n$.

Derivadas de funções lineares e formas quadráticas

Seja $u = f(\mathbf{x})$ uma função das variáveis x_1, x_2, \dots, x_p em que $\mathbf{x} = (x_1, x_2, \dots, x_p)'$ e sejam $\frac{\partial u}{\partial x_1}, \frac{\partial u}{\partial x_2}, \dots, \frac{\partial u}{\partial x_p}$ as derivadas parciais.

Então, nós definimos $\frac{\partial u}{\partial \mathbf{x}}$ como

$$\frac{\partial u}{\partial \mathbf{x}} = \begin{bmatrix} \frac{\partial u}{\partial x_1} \\ \vdots \\ \frac{\partial u}{\partial x_p} \end{bmatrix}.$$

Em alguns casos nós podemos encontrar o máximo ou mínimo de u resolvendo $\frac{\partial u}{\partial \mathbf{x}} = 0$.

Teorema: Seja $u = \mathbf{a}'\mathbf{x} = \mathbf{x}'\mathbf{a}$, onde $\mathbf{a}' = (a_1, a_2, \dots, a_p)'$ é um vetor de constantes, então

$$\frac{\partial u}{\partial \mathbf{x}} = \frac{\partial \mathbf{a}'\mathbf{x}}{\partial \mathbf{x}} = \frac{\partial \mathbf{x}'\mathbf{a}}{\partial \mathbf{x}} = \mathbf{a}.$$

Teorema: Seja $u = \mathbf{x}'\mathbf{A}\mathbf{x}$, onde \mathbf{A} é uma matriz simétrica de constantes, então

$$\frac{\partial u}{\partial \mathbf{x}} = \frac{\partial \mathbf{x}'\mathbf{A}\mathbf{x}}{\partial \mathbf{x}} = 2\mathbf{A}\mathbf{x}.$$

- **Exercícios:** pesquisar sobre as propriedades do *vec* e também sobre diferenciação matricial.
- **Observação:** basicamente **todas** as operações e decomposições matriciais apresentadas estão implementadas no programa R.

Comandos no R

```
require(MASS)
require(Matrix)
A=matrix(c(3,2,6,2,10,-7,6,-7,9),3,3)
A
```

```
      [,1] [,2] [,3]
[1,]     3     2     6
[2,]     2    10    -7
[3,]     6    -7     9
```

```
isSymmetric(A)
```

```
[1] TRUE
```

```
t(A) #Transposta
```

```
      [,1] [,2] [,3]
[1,]     3     2     6
[2,]     2    10    -7
[3,]     6    -7     9
```

```
B=matrix(c(3,1,1,1,2,8,5,8,4),3,3)
```

```
B
```

	[,1]	[,2]	[,3]
[1,]	3	1	5
[2,]	1	2	8
[3,]	1	8	4

```
isSymmetric(B)
```

```
[1] FALSE
```

```
c=5
```

```
str(c)
```

```
num 5
```

```
diag(5) #Matriz identidade 5x5
```

	[,1]	[,2]	[,3]	[,4]	[,5]
[1,]	1	0	0	0	0
[2,]	0	1	0	0	0
[3,]	0	0	1	0	0
[4,]	0	0	0	1	0
[5,]	0	0	0	0	1

```
solve(A) #Inversa
```

```
      [,1]      [,2]      [,3]  
[1,] -0.09297052  0.13605442  0.16780045  
[2,]  0.13605442  0.02040816 -0.07482993  
[3,]  0.16780045 -0.07482993 -0.05895692
```

```
ginv(A) #Inversa generalizada
```

```
      [,1]      [,2]      [,3]  
[1,] -0.09297052  0.13605442  0.16780045  
[2,]  0.13605442  0.02040816 -0.07482993  
[3,]  0.16780045 -0.07482993 -0.05895692
```

```
det(A) #Determinante
```

```
[1] -441
```

```
rankMatrix(A)[1] # Posto de A
```

```
[1] 3
```

`A*%*%B` *#Produto de matrizes*

	[,1]	[,2]	[,3]
[1,]	17	55	55
[2,]	9	-34	62
[3,]	20	64	10

`A+B` *#Soma de matrizes*

	[,1]	[,2]	[,3]
[1,]	6	3	11
[2,]	3	12	1
[3,]	7	1	13

`c*A` *#Produto de matriz por escalar*

	[,1]	[,2]	[,3]
[1,]	15	10	30
[2,]	10	50	-35
[3,]	30	-35	45

```
# Autovalores e autovetores  
A=matrix(c(4,2,2,7),2,2)  
eigen(A)
```

eigen() decomposition

\$values

[1] 8 3

\$vectors

 [,1] [,2]

[1,] 0.4472136 -0.8944272

[2,] 0.8944272 0.4472136

Comandos no R

```
#Decomposição espectral
```

```
C=matrix(c(1/sqrt(5),2/sqrt(5),-2/sqrt(5),1/sqrt(5)),2,2)
```

```
D=matrix(c(8,0,0,3),2,2)
```

```
C%*%D%*%t(C)
```

```
      [,1] [,2]  
[1,]    4    2  
[2,]    2    7
```

```
t(C)%*%A%*%C
```

```
      [,1] [,2]  
[1,]    8    0  
[2,]    0    3
```

```
C1=matrix(c(1,2,-2,1),2,2)
```

```
C1%*%D%*%solve(C1)
```

```
      [,1] [,2]  
[1,]    4    2  
[2,]    2    7
```

Comandos no R

```
#Decomposição de Cholesky
```

```
Sigma=matrix(c(3,1,0.5,1,2,0.8,0.5,0.8,4),3,3)
```

```
chol(Sigma)
```

```
      [,1]      [,2]      [,3]  
[1,] 1.732051 0.5773503 0.2886751  
[2,] 0.000000 1.2909944 0.4905779  
[3,] 0.000000 0.0000000 1.9172898
```

```
t(chol(Sigma))%*%(chol(Sigma))
```

```
      [,1] [,2] [,3]  
[1,] 3.0  1.0  0.5  
[2,] 1.0  2.0  0.8  
[3,] 0.5  0.8  4.0
```

Referência

- Notas de aula do Prof. Caio Azevedo.
- [Notas](#) de aula do Prof. Gilberto de Paula.
- Agresti, A. (2015). *Foundations of Linear and Generalized Linear Models*. Wiley series in probability and statistics.