ME720 - Modelos Lineares Generalizados

Parte 3 - Introdução aos modelos normais lineares

Profa. **Larissa Avila Matos**

Exemplo 1: Teste de esforço cardiopulmonar

Considere o estudo sobre teste de esforço cardiopulmonar em pacientes com insufiência cardíaca realizado no InCor da Faculdade de Medicina da USP pela Dra. Ana Fonseca Braga.

Um dos objetivos do estudo é comparar os grupos formados pelas diferentes etiologias cardíacas quanto às respostas respiratórias e metabólicas obtidas do teste de esforço cardiopulmonar.

Outro objetivo do estudo é saber se alguma das características observadas (ou combinação delas) pode ser utilizada como fator prognóstico de óbito.

Os dados podem ser encontrados em

http://www.ime.usp.br/∼jmsinger/doku.php?id=start.

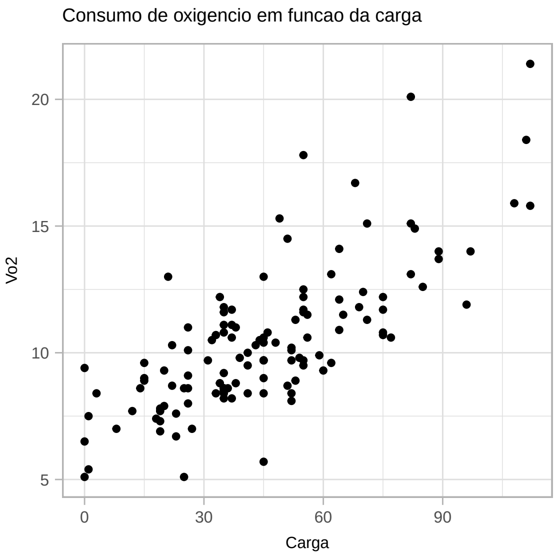
Etiologias = CH: chagásicos, ID: idiopáticos, IS: isquêmicos, C: controle.

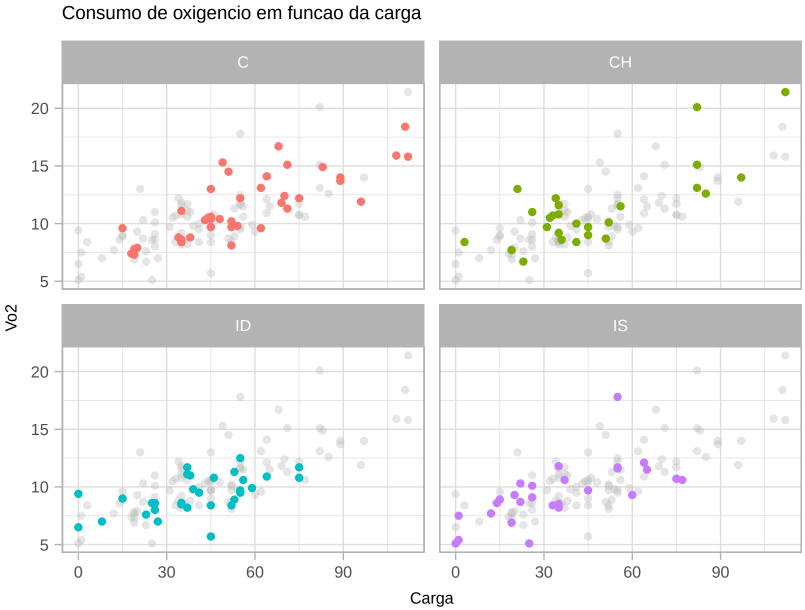
Considere que o objetivo é o de explicar a variação do consumo de oxigênio no limiar anaeróbio ($ml/(kg.min)$) em função da carga utilizada na esteira ergométrica para pacientes com diferentes etiologias (causas) de insuficiência cardíaca.

A grosso modo o Limiar Anaeróbio é um ponto (limite), de divisão entre metabolismo essencialmente aeróbio e metabolismo essencialmente anaeróbio.

Aeróbio (com a utilização de oxigênio); anaeróbio (sem a utilização de oxigênio).

Como responder à pergunta de interesse (ignorando as etiologias cardíacas, num primeiro momento)?


Consumo de oxigencio em funcao da carga
Vo2
20
15
10
5
0
30
60
90
Carga


Consumo de oxigencio em funcao da carga
C
CH
ID
IS
Vo2
Carga
20
15
10
5
0
30
60
90

Exemplo 2: Medidas de absorbância

Uma bioquímica (Tecnolóloga de Alimentos) está interessada em estudar a extração de pigmentos naturais, com aplicação como corante em alimentos. Numa primeira etapa tem-se a necessidade de escolher o melhor solvente extrator. A escolha do(s) melhor(es) solventes foi realizada através da medida da absorbância de um pigmento natural do fruto de baguaçú.

Fator = tipos de solvente; k=5 níveis; n_k=5 repetições.

Quanto maior a absorbância, melhor o solvente.

Unidade experimental: 10 gramas de polpa do fruto de baguaçú.

Casualização: a partir de 1 kg de polpa, foram sendo retiradas amostras de 10 gramas, onde foram aplicados os tratamentos, numa ordem aleatória.

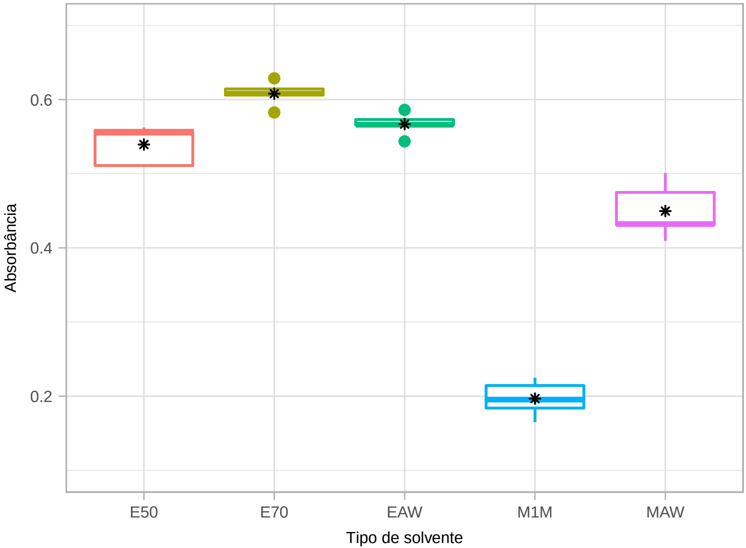
Em princípio, o fator de interesse (solvente) é qualitativo.

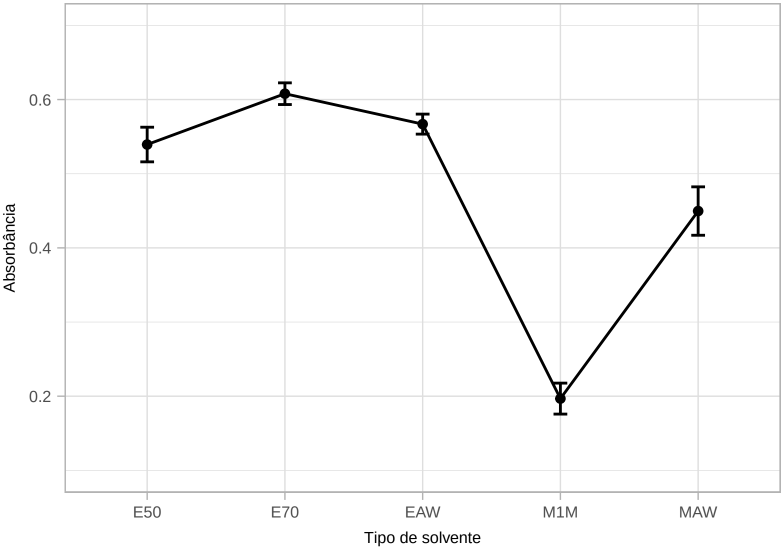
Experimento balanceado : mesmo número de observações (unidades experimentais) por nível do fator.

Possível dependência entre as unidades experimentais?

Dados

Solvente	Absorbância (Observação)				
	1	2	3	4	5
E50	0,5553	0,5623	0,5585	0,5096	0,5110
EAW	0,5436	0,5660	0,5860	0,5731	0,5656
MAW	0,4748	0,4321	0,4309	0,5010	0,4094
E70	0,6286	0,6143	0,5826	0,6079	0,6060
M1M	0,1651	0,1840	0,2144	0,2249	0,1954


0.6
0.4
0.2
Absorbância
E50
E70
EAW
M1M
MAW
Tipo de solvente


0.6
0.4
0.2
Absorbância
E50
E70
EAW
M1M
MAW
Tipo de solvente

Modelagem

Para todas os exemplos, podemos considerar em algum tipo de modelagem estatística, para responder às perguntas de interesse.

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

Tais modelos (de regressão, de planejamento ou de Análise de Covariância) podem ser decompostos em uma parte sistemática e uma parte aleatória.

Exemplo 1: desconsiderando as etiologias cardíacas

O modelo para esse exemplo pode ser dado por

$$Y_i = \beta_0 + \beta_1 x_i + \xi_i, i = 1, ..., 124,$$

onde

- $\xi_i \overset{i.i.d.}{\sim} N(0, \sigma^2)$;
- $(\beta_0, \beta_1, \sigma^2)$ parâmetros desconhecidos;
- x_i: carga à que o paciente i foi submetido (conhecido e não aleatório);
- Parte sistemática: $\mathbb{E}(Y_i) = \beta_0 + \beta_1 x_i$;
- Parte aleatória: ξ_i.

O modelo acima implica que $Y_i \overset{ind.}{\sim} N(\beta_0 + \beta_1 x_i, \sigma^2)$.

β_1 : é o incremento (positivo ou negativo) esperado no consumo de oxigênio para o aumento de uma unidade na carga imposta.

Se for possível observar $x_i = 0$, carga igual à 0, temos que:

β_0 : valor esperado do consumo de oxigênio para pacientes submetidos à uma carga igual à 0.

Caso contrário, podemos considerar o seguinte modelo:

$$Y_i = \beta_0 + \beta_1(x_i - \overline{x}) + \xi_i, i = 1, ..., 124, \overline{x} = \frac{1}{124}\sum_{i=1}^{n} x_i.$$

Neste caso, β_0 é o valor esperado do consumo de oxigênio para pacientes submetidos à uma carga igual à média amostral.

Exemplo 1: considerando as etiologias cardíacas

O modelo considerando as etiologias cardíacas é dado por

$$Y_{ij} = \beta_{0i} + \beta_{1i}x_{ij} + \xi_{ij}, i = 1, ..., ; j = 1, ..., n_i,$$

com

- Etiologias = CH ($i = 1$), ID ($i = 2$), IS ($i = 3$), C: ($i = 4$);
- $\xi_{ij} \overset{i.i.d.}{\sim} N(0, \sigma^2)$;
- $(\beta_{01}, \beta_{02}, \beta_{03}, \beta_{04}, \beta_{11}, \beta_{12}, \beta_{13}, \beta_{14}, \sigma^2)$ parâmetros desconhecidos;
- x_{ij}: carga à que o paciente j que apresenta a etiologia cardíaca i foi submetido (conhecido e não aleatório);
- Parte sistemática: $\mathbb{E}(Y_{ij}) = \beta_{0i} + \beta_{1i}x_{ij}$;
- Parte aleatória: ξ_{ij}.

O modelo acima implica que $Y_{ij} \overset{ind.}{\sim} N(\beta_{0i} + \beta_{1i}x_{ij}, \sigma^2)$.

A interpretação dos parâmetros desse modelo é similar ao anterior, por exemplo, na etiologia cardíaca CH ($i = 1$), temos que

- β_{11} : é o incremento (positivo ou negativo) esperado no consumo de oxigênio para o aumento de uma unidade na carga imposta para pacientes com etiologia cardíaca CH.
- Se for possível observar $x_i = 0$, carga igual à 0, temos que:

  β_0 : valor esperado do consumo de oxigênio para pacientes submetidos à uma carga igual à 0 na etiologia cardíaca CH.

Exemplo 2: Modelo (casela de referência)

$$Y_{ij} = \mu + \alpha_i + \xi_{ij},\ i = 1, 2, \cdots, 5 \text{ (grupos)},\ j = 1, \cdots, 5 \text{ (u.e.)};$$

onde u.e. = unidades experimentais; com

- Erros (parte aleatória): $\xi_{ij} \overset{i.i.d}{\sim} N(0, \sigma^2)$;
- μ, α_i não aleatório;
- $\mathbb{E}_{\xi_{ij}}(Y_{ij}) = \mu_i, \text{Var}_{\xi_{ij}}(Y_{ij}) = \sigma^2$;
- Parte sistemática: $\mu + \alpha_i$ que é a média populacional relacionada ao i-ésimo fator, $\alpha_1 = 0$;
- $Y_{ij} \overset{ind.}{\sim} N(\mu + \alpha_i, \sigma^2)$.

Neste caso μ é a média (populacional) do grupo de referência.

$\alpha_i = \mu_i - \mu$ é a diferença entre a média do grupo i e o grupo de referência, $i = 1, .., 4$.

Nesse exemplo: E50 (referência), E70 ($i = 1$), EAW ($i = 2$), M1M ($i = 3$) e MAW ($i = 4$).

Note que, em todos os casos os modelos estão bem definidos, no sentido de que todas as suposições foram descritas e os parâmetros, interpretados.

Os modelos anteriores se enquadram na classe dos modelos normais lineares homocedásticos (de efeitos fixos) (MNL).

Notação matricial para o MNL

Seja,

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\xi},$$

com $\boldsymbol{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}, \boldsymbol{X} = \begin{bmatrix} X_{11} & \dots & X_{1p} \\ X_{21} & \dots & X_{2p} \\ \vdots & \ddots & \vdots \\ X_{n1} & \dots & X_{np} \end{bmatrix}, \boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix}, \boldsymbol{\xi} = \begin{bmatrix} \xi_1 \\ \xi_2 \\ \vdots \\ \xi_n \end{bmatrix}.$

Suposição: $\boldsymbol{\xi} \sim N_n(\mathbf{0}, \sigma^2 \boldsymbol{I}_n)$ (que é o vetor de erros).

$\boldsymbol{Y}$ é o vetor das variáveis resposta.

O índice n da variável resposta é geral e pode representar combinações de índices.

$\boldsymbol{X}$ é a matriz de plajenamento (ou delineamento) que define a parte sistemática do modelo.

Exemplo 1

Para o primeiro modelo, temos que

$$\boldsymbol{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_{124} \end{bmatrix}, \boldsymbol{X} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_{124} \end{bmatrix}, \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}, \boldsymbol{\xi} = \begin{bmatrix} \xi_1 \\ \xi_2 \\ \vdots \\ \xi_{124} \end{bmatrix}.$$

Naturalmente, depois do experimento ser realizado, teremos um vetor de observações (números): $\boldsymbol{y} = (y_1, ..., y_{124})$.

Exemplo 1: considerando as etiologias cardíacas

Para o primeiro modelo, considerando as etiologias cardíacas, temos que

$$\boldsymbol{Y} = \begin{bmatrix} Y_{11} \\ \vdots \\ Y_{1n_1} \\ Y_{21} \\ \vdots \\ Y_{41} \\ \vdots \\ Y_{4n_4} \end{bmatrix}, \quad \boldsymbol{X} = \begin{bmatrix} 1 & 0 & 0 & 0 & x_{11} & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 0 & 0 & x_{1n_1} & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & x_{21} & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & x_{41} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & x_{4n_4} \end{bmatrix},$$

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_{01} \\ \beta_{02} \\ \beta_{03} \\ \beta_{04} \\ \beta_{11} \\ \beta_{12} \\ \beta_{13} \\ \beta_{14} \end{bmatrix}, \quad \boldsymbol{\xi} = \begin{bmatrix} \xi_{11} \\ \vdots \\ \xi_{1n_1} \\ \xi_{21} \\ \vdots \\ \xi_{41} \\ \vdots \\ \xi_{4n_4} \end{bmatrix}.$$

Naturalmente, depois do experimento ser realizado, teremos um vetor de observações (números): $\boldsymbol{y} = (y_1, \cdots, y_{1n_1}, y_{21}, \cdots, y_{41}, \cdots, y_{4n_4})$.

Exemplo 2

Neste caso, temos que

$$\boldsymbol{Y} = \begin{bmatrix} Y_{11} \\ Y_{12} \\ Y_{13} \\ Y_{14} \\ Y_{15} \\ Y_{21} \\ \vdots \\ Y_{51} \\ \vdots \\ Y_{55} \end{bmatrix}, \boldsymbol{X} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 0 & 0 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 0 & 0 & 1 \end{bmatrix}, \boldsymbol{\beta} = \begin{bmatrix} \mu \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \\ \alpha_5 \end{bmatrix}, \boldsymbol{\xi} = \begin{bmatrix} \xi_{11} \\ \xi_{12} \\ \xi_{13} \\ \xi_{14} \\ \xi_{15} \\ \xi_{21} \\ \vdots \\ \xi_{51} \\ \vdots \\ \xi_{55} \end{bmatrix}.$$

Naturalmente, depois do experimento ser realizado, teremos um vetor de observações (números): $\boldsymbol{y} = (y_{11}, ..., y_{55})$.

Estimação dos parâmetros

Existem vários métodos que podem ser usados para estimar os parâmetros do modelo, por exemplo,

1. mínimos quadrados,
2. mínimos quadrados ordinários,
3. mínimos quadrados generalizados,
4. máxima verossimilhança.

Nós veremos dois casos: mínimos quadrados e máxima verossimilhança.

Estimação dos parâmetros via máxima verossimilhança

Assumindo que $\boldsymbol{\xi} \sim N_p(\mathbf{0}, \sigma^2 \boldsymbol{I}_n)$, então $\boldsymbol{Y} \sim N_p(\boldsymbol{X}\boldsymbol{\beta}, \sigma^2 \boldsymbol{I}_n)$. Portanto, a função de verssomilhança do modelo proposto é dada por

$$L(\boldsymbol{\beta}, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\Big\{ - \frac{1}{2\sigma^2} (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})' (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}) \Big\},$$

e a função de log-verossimilhança por

$$\ell(\boldsymbol{\beta}, \sigma^2) = -\frac{n}{2} \log(2\pi) - \frac{n}{\sigma^2} - \frac{1}{2\sigma^2} (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})' (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}).$$

Os estimadores de máxima verossimilhança (MV) são as soluções das equações:

$$\frac{\partial \ell(\boldsymbol{\beta}, \sigma^2)}{\partial \boldsymbol{\beta}} \Big|_{\boldsymbol{\beta} = \widehat{\boldsymbol{\beta}}} = \mathbf{0} \quad \text{e} \quad \frac{\partial \ell(\boldsymbol{\beta}, \sigma^2)}{\partial \sigma^2} \Big|_{\sigma^2 = \widehat{\sigma}^2} = \mathbf{0}.$$

Temos que,

$$\frac{\partial \ell(\boldsymbol{\beta}, \sigma^2)}{\partial \boldsymbol{\beta}} = \frac{(\boldsymbol{X}'\boldsymbol{Y} - \boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta})}{\sigma^2} \text{ e } \frac{\partial \ell(\boldsymbol{\beta}, \sigma^2)}{\partial \sigma^2} = \frac{(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})'(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})}{2\sigma^4} - \frac{n}{2\sigma^2}.$$

Então,

$$\begin{aligned} \boldsymbol{X}'\boldsymbol{X}\widehat{\boldsymbol{\beta}} &= \boldsymbol{X}'\boldsymbol{Y} \quad \text{(equações normais)} \\ \widehat{\boldsymbol{\beta}}_{MV} &= \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1} \boldsymbol{X}'\boldsymbol{Y} \end{aligned}$$

e

$$\widehat{\sigma}^2_{MV} = \frac{1}{n}\left(\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}\right)'\left(\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}\right),$$

desde que $\boldsymbol{X}'\boldsymbol{X}$ seja inversível. Como $n >>> p$, tal inversibilidade ocorrerá se, o somente se, a matriz $\boldsymbol{X}$ tiver posto coluna completo.

Isto, por sua vez, ocorre quando o modelo está identificado (não está superparametrizado) e/ou quando não há covariáveis que sejam combinações lineares de outras.

O sistema de equações normais é consistente, ou seja, apresenta pelo menos uma solução.

A justificativa não formal para isso é relativamente simples:

- Se $\boldsymbol{X}'\boldsymbol{X}$ for inversível ($rank(\boldsymbol{X}) = p$), a solução única.
- Se $\boldsymbol{X}'\boldsymbol{X}$ for não inversível ($rank(\boldsymbol{X}) < p$), podemos considerar alguma inversa generalizada de $\boldsymbol{X}'\boldsymbol{X}$. Neste caso, o sistema pode apresentar infinitas soluções e as funções estimáveis passam a ter uma importância maior do que os parâmetros isoladamente.

No último caso, uma solução é dada por $\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}'\boldsymbol{Y}$.

Propriedades dos Estimadores de MV

Uma vez que $\widehat{\boldsymbol{\beta}}_{MV} = \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{Y}$, $\widehat{\sigma}^2_{MV} = \frac{1}{n}(\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}})'(\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}})$, $\boldsymbol{Y} \sim N_n(\boldsymbol{X}\boldsymbol{\beta}, \sigma^2\boldsymbol{I}_n)$ e pelas propriedades associados à vetores aleatórios e a distribuição normal multivariada, temos que:

- $\widehat{\boldsymbol{\beta}}_{MV}$

1. $\mathbb{E}(\widehat{\boldsymbol{\beta}}_{MV}) = \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\mathbb{E}(\boldsymbol{Y}) = \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{\beta}$ (não viciado).
2. $Cov(\widehat{\boldsymbol{\beta}}_{MV}) = \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{X}Cov(\boldsymbol{Y})\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1} = \sigma^2\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}$.
3. $\widehat{\boldsymbol{\beta}}_{MV} \sim N_p(\boldsymbol{\beta}, \sigma^2\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1})$ (normalidade).
4. $\widehat{\beta}_{jMV} \sim N(\beta_j, \sigma^2 c_{jj})$, onde c_{jj} é o j-ésimo elemento da diagonal principal da matriz $\mathbf{C} = (\boldsymbol{X}'\boldsymbol{X}^{-1})$.

Propriedades dos Estimadores de MV

- $\widehat{\sigma}^2_{MV}$

1. $\mathbb{E}(\widehat{\sigma}^2_{MV}) = \sigma^2 \frac{(n-p)}{n}$(viciado).

   Mas, $\widehat{\sigma}^2 = \widehat{\sigma}^2_{MV} \frac{n}{(n-p)} = \frac{1}{(n-p)}(\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}})'(\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}})$ é um estimador não viciado para σ^2.

2. $\frac{(n-p)\widehat{\sigma^2}}{\sigma^2} \sim \chi^2_{(n-p)}$.

3. $\widehat{\sigma}^2 \perp \widehat{\boldsymbol{\beta}}_{MV}$.

Estimadores de MV

Resumo: Os estimadores de MV dados por

$$\widehat{\boldsymbol{\beta}} = \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{Y} \text{ e } \widehat{\sigma}^2 = \frac{(\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}})'(\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}})}{(n-p)} = \frac{\boldsymbol{Y}'(\mathbf{I} - \mathbf{H})\boldsymbol{Y}}{(n-p)},$$

com $\mathbf{H} = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$, são

1. Consistentes,	4.Suficientes,	7. $\widehat{\boldsymbol{\beta}} \sim N_p(\boldsymbol{\beta}, \sigma^2\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1})$,
2. Eficientes,	5. Completo,	8. $\frac{(n-p)\widehat{\sigma^2}}{\sigma^2} \sim \chi^2_{(n-p)}$,
3. Não viciados,	6. Minima variância,	9. $\widehat{\sigma}^2 \perp \widehat{\boldsymbol{\beta}}$.

Estimação dos parâmetros por mínimos quadrados

Estimador usual para $\boldsymbol{\beta}$: mínimos quadrados (MQ).

Objetivo: Encontar $\widehat{\boldsymbol{\beta}}$ (valor de $\boldsymbol{\beta}$) que minimiza a soma de qudrados dos erros, ou seja, obter $\boldsymbol{\beta}$ que minimiza $Q(\boldsymbol{\beta}) = (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})'(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}) = \boldsymbol{\xi}\boldsymbol{\xi}'$. Em geral, $\boldsymbol{\beta} \in \mathcal{R}^p$.

Suposição: $\mathbb{E}(\boldsymbol{\xi}) = \mathbf{0}$ e $Cov(\boldsymbol{\xi}) = \sigma^2\mathbf{I}$. Mas, no nosso curso vamos assumir $\boldsymbol{\xi} \sim N_p(\mathbf{0}, \sigma^2\mathbf{I})$

Assim, para efetuar a minimização, podemos resolver o sistema de equações definido por $\frac{\partial Q(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}$ (chamada de equações normais).

Logo, temos que resolver o seguinte sistema:

$$\frac{\partial Q(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}\Big|_{\boldsymbol{\beta}=\widehat{\boldsymbol{\beta}}} = \mathbf{0}$$

Por outro lado, temos que:

$$\begin{aligned}
\frac{\partial}{\partial \boldsymbol{\beta}} Q(\boldsymbol{\beta}) &= \frac{\partial}{\partial \boldsymbol{\beta}} (\boldsymbol{Y}'\boldsymbol{Y} - 2\boldsymbol{Y}'\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta}) = -2\boldsymbol{X}'\boldsymbol{Y} + 2\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta} \\
&\rightarrow \frac{\partial}{\partial \boldsymbol{\beta}} Q(\boldsymbol{\beta})\Big|_{\boldsymbol{\beta}=\widehat{\boldsymbol{\beta}}} = \mathbf{0} \rightarrow -2\boldsymbol{X}'\boldsymbol{Y} + 2\boldsymbol{X}'\boldsymbol{X}\widehat{\boldsymbol{\beta}} = \mathbf{0} \\
&\rightarrow \boldsymbol{X}'\boldsymbol{Y} = \boldsymbol{X}'\boldsymbol{X}\widehat{\boldsymbol{\beta}} \quad \text{(equações normais)} \\
&\rightarrow \widehat{\boldsymbol{\beta}} = \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1} \boldsymbol{X}'\boldsymbol{Y},
\end{aligned}$$

desde que a matriz $\boldsymbol{X}$ tiver posto coluna completo.

Observação: sob a suposição de normalidade, o estimador de MQO coincide com o estimador de MV (máxima verossimilhança).

Estimador de σ^2

Minimizar a soma de quadrados $\boldsymbol{\xi}\boldsymbol{\xi}'$ não fornece um estimador para σ^2. No entanto, um estimador não viaciado de σ^2 baseado nas estimativas de mínimos quadrados é dado por

$$\widehat{\sigma}^2 = \frac{1}{n-p}\left(\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}\right)'\left(\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}\right).$$

Esse estimador é não-viciado. Além disso, pode-se provar que $\widehat{\boldsymbol{\beta}} \perp \widehat{\sigma}^2$ e $\frac{(n-p)\widehat{\sigma^2}}{\sigma^2} \sim \chi^2_{(n-p)}$.

Teorema de Gauss Markov

Se $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\xi}$ é tal que $\mathbb{E}(\boldsymbol{\xi}) = \mathbf{0}$ e $\mathbb{E}(\boldsymbol{\xi}\boldsymbol{\xi}') = \sigma^2\mathbf{I}$, o "melhor" estimador linear não viciado de $\boldsymbol{\beta}$ é dado pelo estimador de mínimos quadrados, ou seja, $\boldsymbol{\beta} = (\boldsymbol{X}\boldsymbol{X}')^{-1}\boldsymbol{X}'\boldsymbol{Y}$ é o melhor estimador linear não viciado de $\boldsymbol{\beta}$ $(BLUE)$.

Prova: ??

Observação: Nesse caso, "melhor" significa mínima variância e linear função linear dos Y's. Em inglês *Best Linear Unbiased Estimator* $(BLUE)$.

Teorema de Gauss Markov

Sob as mesmas condições do teorema anterior, o $BLUE$ de qualquer combinação linear de β_i é a mesma combinação linear do $BLUE$ de β_i, ou seja, o $BLUE$ de $\mathbf{a}'\boldsymbol{\beta}$ é $\mathbf{a}'\widehat{\boldsymbol{\beta}}$, onde $\mathbf{a}$ é um vetor $p \times 1$ conhecido de constantes e $\widehat{\beta}$ é o $BLUE$ de $\boldsymbol{\beta}$.

Prova: ??

Consequências do Teorema de Gauss Markov

1. Se $\mathbf{a}' = (0, \cdots, 0, 1, 0, \cdots, 0)$, 1 na i-ésima posição. Então, a i-ésima componente de $\widehat{\boldsymbol{\beta}}$ ($\widehat{\beta}_i$) é o ENVUM de β_i.

2. Estimação pontual de $\mathbb{E}(\boldsymbol{Y})$.

Para estimar a média de $\boldsymbol{Y}$ (resposta esperada) dado os valores $x_1, \cdots, x_p$. Podemos usar o fato que $\mathbb{E}(\boldsymbol{Y}) = \sum_{i=1}^{p} \beta_i x_i = \mathbf{x}'\boldsymbol{\beta}$ é uma função linear dos $\boldsymbol{\beta}$'s, e tomando $\mathbf{a}' = (x_1, \cdots, x_p)$, temos que

$$\widehat{\mathbb{E}(\boldsymbol{Y})} = \mathbf{a}'\widehat{\boldsymbol{\beta}} = \sum_{i=1}^{p} \widehat{\beta}_i x_i$$

é um ENVUM de $\mathbb{E}(\boldsymbol{Y})$.

Testes de hipóteses simples e IC para os β_i

Em geral, nos modelos descritos acima, tem-se o interesse em testar se

$$H_0 : \beta_i = \beta_{0i} \ vs \ H_1 : \beta_i \neq \beta_{0i}, i = 1, ..., p.$$

Por exemplo, no primeiro modelo, é de interesse testar se a carga não contribui para explicar o consumo de oxigênio, ou seja:

$$H_0 : \beta_1 = 0 \ vs \ H_1 : \beta_1 \neq 0, i = 1, ..., p.$$

Hipóteses simples como as apresentadas, podem ser testadas usando-se o fato de que:

$$\widehat{\boldsymbol{\beta}} \sim N_p(\boldsymbol{\beta}, \sigma^2 \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}) \rightarrow \widehat{\beta}_j \sim N(\beta_j, \sigma^2 c_{jj})$$

$$\frac{(n-p)\widehat{\sigma}^2}{\sigma^2} \sim \chi^2_{(n-p)},$$

e

$$\widehat{\sigma}^2 \perp \widehat{\boldsymbol{\beta}},$$

onde $\mathbf{C} = (\boldsymbol{X}'\boldsymbol{X})^{-1}$.

Então, a estatística

$$T = \frac{\widehat{\beta}_j - \beta_{0j}}{\sqrt{\widehat{\sigma}^2 c_{jj}}} \sim t_{(n-p)}, \text{ sob } H_0. \tag{1}$$

Lembremos que, no caso de hipóteses bi-laterais ($= vs \neq$) o pvalor é dado por $p - valor = 2P(T > |t_c| | H_0)$, em que $T \sim t_{(n-p)}$, sob H_0 e t_c é o valor calculado da estatística definida em (1).

Demonstração:

- IC para β_j: $\widehat{\beta}_j \sim N(\beta_j, \sigma^2 c_{jj})$

Se $\beta_j \neq 0$, podemos escrever que

$$\mathbb{P}\left(-t_{\alpha/2,n-p} \leq \frac{\widehat{\beta}_j - \beta_j}{\sqrt{\widehat{\sigma}^2 c_{jj}}} \leq t_{\alpha/2,n-p}\right) = 1 - \alpha$$

$$\rightarrow \mathbb{P}\left(\widehat{\beta}_j - t_{\alpha/2,n-p}\sqrt{\widehat{\sigma}^2 c_{jj}} \leq \beta_j \leq \widehat{\beta}_j + t_{\alpha/2,n-p}\sqrt{\widehat{\sigma}^2 c_{jj}}\right) = 1 - \alpha$$

Então,

$$IC(100(1-\alpha)\%, \beta_j) = \widehat{\beta}_j \pm t_{\alpha/2,n-p}\sqrt{\widehat{\sigma}^2 c_{jj}}$$

Exemplo 1: Modelo 1

Os resultados do modelo ajustado são dados a seguir.

Parâmetro	Estimativa	EP	Estat. t	IC(95%)	p-valor
β_0	6,56	0,36	18,43	[5,87 ; 7,26]	<0,0001
β_1	0,09	0,01	12,52	[0,07 ; 0,10]	<0,0001

Os dois parâmetros são diferentes de 0. A carga influencia positivamente o consumo de oxigênio. O consumo de oxigênio para pacientes submetidos à carga 0 tende a se apresentar entre 5,87 e 7,26.

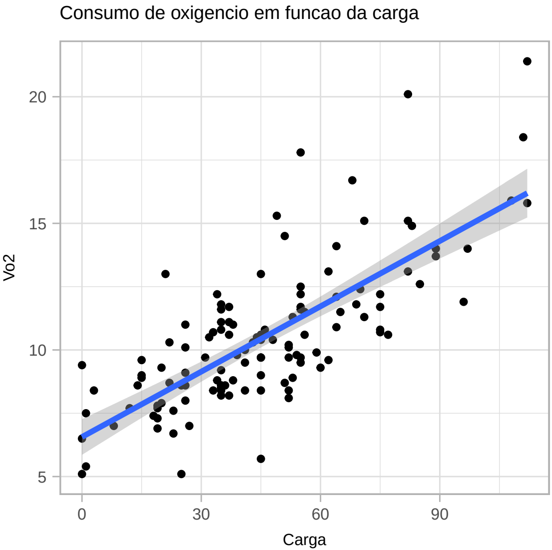
```r
fit.model<-lm(vo2~carga)
summary(fit.model)
```

```
Call:
lm(formula = vo2 ~ carga)

Residuals:
    Min      1Q  Median      3Q     Max
-4.7327 -1.1680 -0.3317  1.1524  6.5075

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 6.563493   0.356037   18.43   <2e-16 ***
carga       0.085982   0.006869   12.52   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.892 on 122 degrees of freedom
Multiple R-squared:  0.5622,    Adjusted R-squared:  0.5586
F-statistic: 156.7 on 1 and 122 DF,  p-value: < 2.2e-16
```


Consumo de oxigencio em funcao da carga
Vo2
20
15
10
5
0
30
60
90
Carga

Exemplo 1: Modelo 2

Os resultados do modelo ajustado considerando os grupos de etiologias cardíacas são dados a seguir.

Parâmetro	Estimativa	EP	Estat. t	IC(95%)	p-valor
$\beta_{01}(C)$	6,56	0,71	9,18	[5,16 ; 7,96]	<0,0001
$\beta_{02}(CH)$	6,63	0,75	8,88	[5,17 ; 8,10]	<0,0001
$\beta_{03}(ID)$	7,35	0,78	9,45	[5,82 ; 8,87]	<0,0001
$\beta_{04}(IS)$	6,80	0,66	10,33	[5,51 ; 8,09]	<0,0001
$\beta_{11}(C)$	0,09	0,01	7,62	[0,07 ; 0,11]	<0,0001
$\beta_{12}(CH)$	0,10	0,01	7,14	[0,07 ; 0,13]	<0,0001
$\beta_{13}(ID)$	0,05	0,02	2,82	[0,02 ; 0,08]	0,0056
$\beta_{14}(IS)$	0,08	0,02	4,78	[0,05 ; 0,11]	<0,0001

O consumo de oxigênio dos pacientes para carga 0 parecem ser semelhantes entre os grupos. O aumento no consumo parece ser menor que os demais para pacientes idiopáticos e igual para os outros três tipos.

```r
fit.model<-lm(vo2~-1+etiofac+carga:etiofac)
summary(fit.model)
```

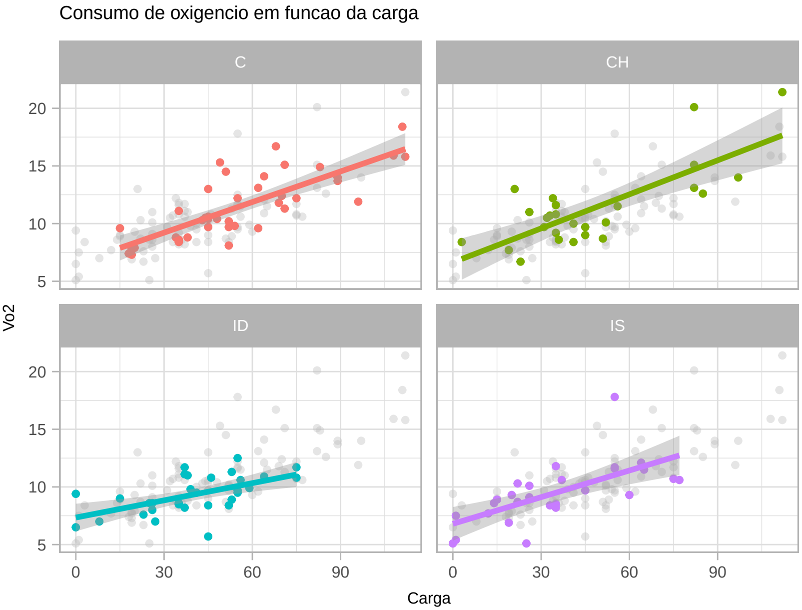
```

Call:
lm(formula = vo2 ~ -1 + etiofac + carga:etiofac)

Residuals:
    Min      1Q  Median      3Q     Max 
-3.8824 -1.0629 -0.3659  0.9445  6.7618 

Coefficients:
                Estimate Std. Error t value Pr(>|t|)    
etiofacC         6.56104    0.71441   9.184 1.94e-15 ***
etiofacCH        6.63213    0.74645   8.885 9.64e-15 ***
etiofacID        7.34504    0.77709   9.452 4.57e-16 ***
etiofacIS        6.80127    0.65814  10.334  < 2e-16 ***
etiofacC:carga   0.08846    0.01161   7.619 7.59e-12 ***
etiofacCH:carga  0.09835    0.01377   7.143 8.62e-11 ***
etiofacID:carga  0.04972    0.01763   2.821  0.00564 ** 
etiofacIS:carga  0.07704    0.01612   4.778 5.24e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.84 on 116 degrees of freedom
Multiple R-squared:  0.9731,    Adjusted R-squared:  0.9713 
F-statistic: 525.3 on 8 and 116 DF,  p-value: < 2.2e-16
```



Consumo de oxigencio em funcao da carga
C
CH
ID
IS
20
15
10
5
0
30
60
90
Vo2
Carga

Exemplo 2

Os resultados do modelo ajustado são dados a seguir.

Parâmetro	Estimativa	EP	IC(95%)	Estat. t	pvalor
μ (E50)	0,539	0,011	[0,516; 0,563]	47,826	$<$ 0,0001
α_2 (E70)	0,069	0,0160	[0,035 ; 0,102]	4,298	0,0003
α_3 (EAW)	0,028	0,0160	[-0,006 ; 0,061]	1,726	0,0998
α_4 (M1M)	-0,343	0,0160	[-0,376; -0,309]	-21,481	$<$ 0,0001
α_5 (MAW)	-0,090	0,0160	[-0,123 ;-0,056]	-5,624	$<$ 0,0001

Parâmetro α_3 não significativo. Isto sugere uma possível equivalência entre os solventes E50 e EAW.

```r
solvfac <- C(as.factor(solvente))
fit.model <- lm(mabsor~solvfac)
summary(fit.model)
```

```
Call:
lm(formula = mabsor ~ solvfac)

Residuals:
     Min       1Q   Median       3Q      Max
-0.04024 -0.01874 -0.00086  0.01914  0.05136

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.53934    0.01128  47.826  < 2e-16 ***
solvfacE70   0.06854    0.01595   4.298 0.000351 ***
solvfacEAW   0.02752    0.01595   1.726 0.099843 .
solvfacM1M  -0.34258    0.01595 -21.481 2.75e-15 ***
solvfacMAW  -0.08970    0.01595  -5.624 1.67e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.02522 on 20 degrees of freedom
Multiple R-squared:  0.977, Adjusted R-squared:  0.9725
F-statistic: 212.8 on 4 and 20 DF,  p-value: 4.378e-16
```

Referência

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