

# Breve revisão sobre o MRNLH

Prof. Caio Azevedo

## Exemplo 5: Teste de esforço cardiopulmonar

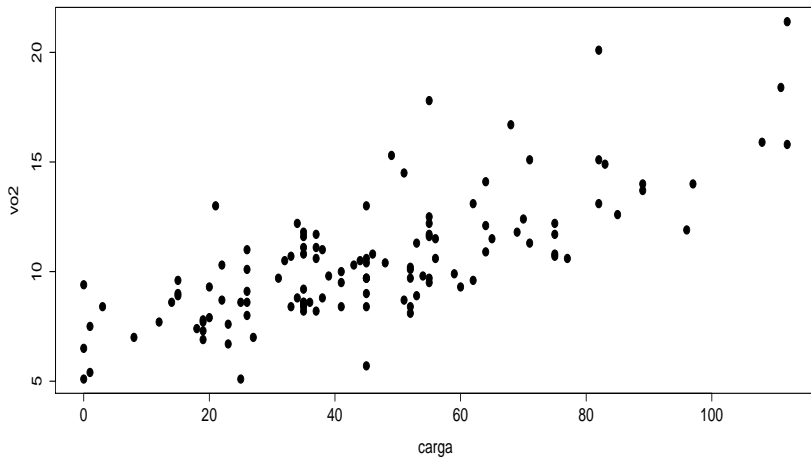
- Considere o estudo sobre teste de esforço cardiopulmonar em pacientes com insuficiência cardíaca realizado no InCor da Faculdade de Medicina da USP pela Dra. Ana Fonseca Braga.
- Um dos objetivos do estudo é comparar os grupos formados pelas diferentes etiologias cardíacas quanto às respostas respiratórias e metabólicas obtidas do teste de esforço cardiopulmonar.
- Outro objetivo do estudo é saber se alguma das características observadas (ou combinação delas) pode(m) ser utilizada(s) 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 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)?.

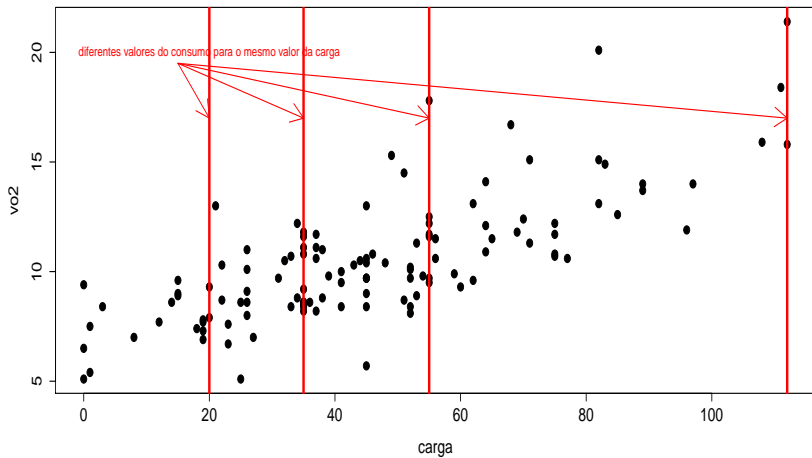
# Dados

ID	Etiologia	Carga	VO2
1	CH	41	10,0
2	CH	56	11,5
3	ID	8	7,0
4	ID	53	8,9
⋮	⋮	⋮	
7	ID	0	6,5
⋮	⋮	⋮	
123	C	64	14,1
124	C	70	12,4

Consumo de oxigênio em função da carga

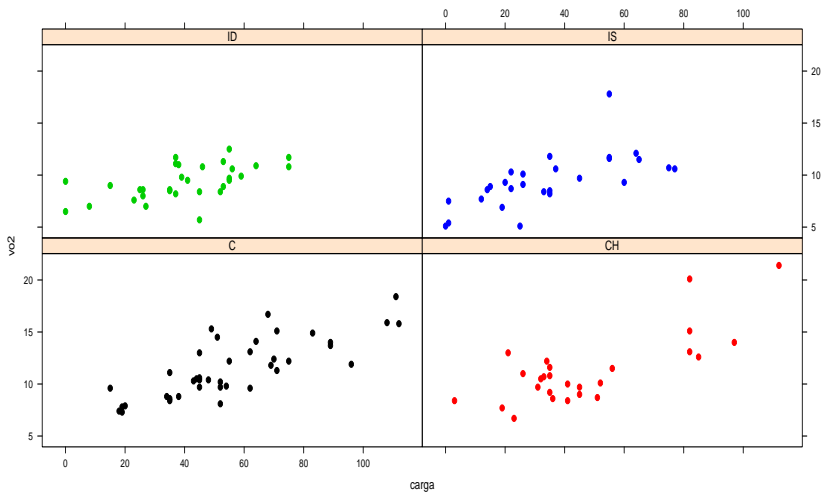


## Consumo de oxigênio em função da carga



- Existe uma relação entre as duas variáveis?
- De que tipo?
- O fato de que quanto maior o valor da carga maior, maior o valor do consumo de oxigênio, implica numa relação de causa e efeito?
- Há outros fatores biológicos (hereditariedade, outras doenças), comportamentais (dieta, prática de exercícios, remédios) e ambientais (poluição, clima), que, verdadeiramente, ditariam os valores dessas duas variáveis para cada indivíduo?
- O que significa dizer: para um dado valor da carga, o comportamento do consumo de oxigênio é aleatório e que pode ser modelado “apropriadamente” por uma estrutura probabilística (paramétrica)?

Consumo de oxigênio em função da carga





- É importante levar em consideração as diferentes etiologias?
- Se sim, como considerá-las na análise?
- Há interesse em comparar a influência da carga no consumo de oxigênio entre as diferentes etiologias cardíacas ?

## Exemplo 5: desconsiderando as etiologias cardíacas

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

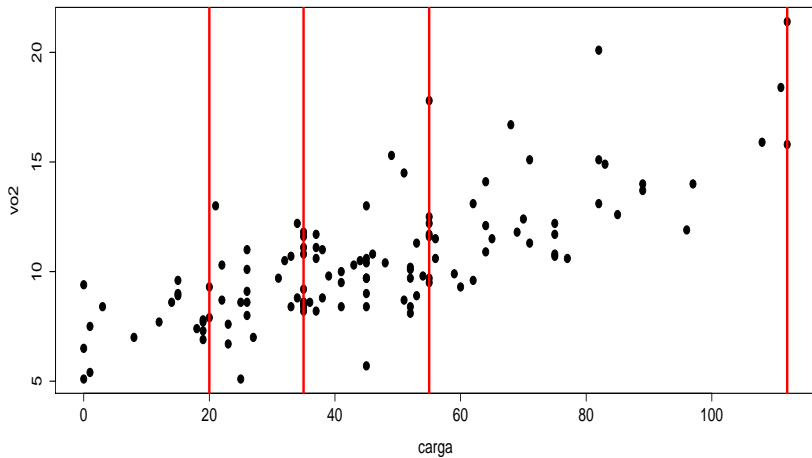
- $\xi_i \stackrel{i.i.d.}{\sim} N(0, \sigma^2)$ .
- $(\beta_0, \beta_1, \sigma^2)'$  : parâmetros desconhecidos.
- $x_i$ : carga à que o paciente  $i$  foi submetido (conhecida e não aleatória).
- Parte sistemática:  $\mathcal{E}(Y_i) = \beta_0 + \beta_1 x_i$ .
- Parte aleatória:  $\xi_i$ .
- O modelo acima implica que  $Y_i \stackrel{ind.}{\sim} N(\beta_0 + \beta_1 x_i, \sigma^2)$ ,  $Y_i$  : valor do consumo de oxigênio do paciente  $i$ .

- $\beta_1$  : 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:
  - $\beta_0$  : valor esperado do consumo de oxigênio para pacientes submetidos à uma carga igual à 0.
- Caso contrário, podemos considerar o seguinte modelo:

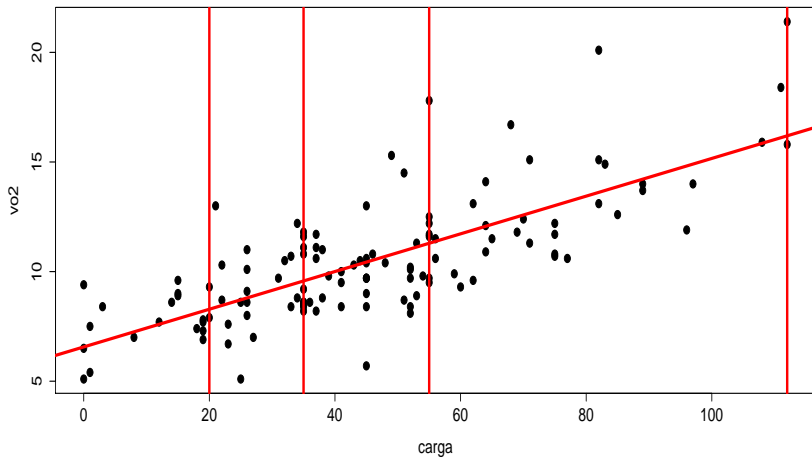
$$Y_i = \beta_0 + \beta_1(x_i - \bar{x}) + \xi_i, i = 1, \dots, 124, \bar{x} = \frac{1}{124} \sum_{i=1}^n x_i.$$

- Neste caso,  $\beta_0$  é o valor esperado do consumo de oxigênio para pacientes submetidos à uma carga igual à média amostral.

Consumo de oxigênio em função da carga



Consumo de oxigênio em função da carga



# Notação matricial para o MNL

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

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

- Suposição:  $\boldsymbol{\xi} \sim N_n(\mathbf{0}, \sigma^2 \mathbf{I}_n)$  (vetor de erros).
- O índice  $n$  da variável resposta é geral e pode representar combinações de índices.

# Continuação

- $\mathbf{X}$  é a matriz de planejamento (ou delineamento) que define a parte sistemática do modelo (conhecida e não aleatória).
- $\mathbf{Y}$  é o vetor associado à variável resposta. Assim, temos que
$$\mathbf{Y} \sim N_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I}_n)$$
- Depois dos dados coletados teremos um conjunto  $\mathbf{y} = (y_1, \dots, y_n)'$  de observações.

## Exemplo 5

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



# Estimação dos parâmetros

- Estimador usual para  $\beta$ : mínimos quadrados ordinários (MQO).
- Objetivo: obter  $\beta$  que minimiza  $Q(\beta) = (\mathbf{Y} - \mathbf{X}\beta)'(\mathbf{Y} - \mathbf{X}\beta)$ . Em geral,  $\beta \in \mathcal{R}^p$ . Assim, para efetuar a minimização, podemos resolver o sistema de equações definido por  $\frac{\partial Q(\beta)}{\partial \beta}$  (chamada de equações normais).
- Logo, temos que resolver o seguinte sistema:

$$\left. \frac{\partial Q(\beta)}{\partial \beta} \right|_{\beta=\hat{\beta}} = \mathbf{0}$$

- Por outro lado, temos que:

$$\begin{aligned}\frac{\partial Q(\beta)}{\partial \beta} &= \frac{\partial}{\partial \beta} (\mathbf{Y}'\mathbf{Y} - 2\mathbf{Y}'\mathbf{X}\beta + \beta'\mathbf{X}'\mathbf{X}\beta) = -2\mathbf{X}'\mathbf{Y} + 2\mathbf{X}'\mathbf{X}\beta \\ \rightarrow \frac{\partial Q(\beta)}{\partial \beta} \Big|_{\beta=\hat{\beta}} &= \mathbf{0} \rightarrow -2\mathbf{X}'\mathbf{Y} + 2\mathbf{X}'\mathbf{X}\hat{\beta} = \mathbf{0} \quad (1) \\ \rightarrow \hat{\beta} &= (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{Y}\end{aligned}$$

desde que  $\mathbf{X}'\mathbf{X}$  seja inversível. Como  $n \gg \gg p$ , tal inversibilidade ocorrerá se, e somente se, a matriz  $\mathbf{X}$  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 definido por (1) é consistente, ou seja, apresenta pelo menos uma solução.
- A justificativa não formal para isso é relativamente simples:
  - Se  $\mathbf{X}'\mathbf{X}$  for inversível ( $rank(\mathbf{X}) = p$ ), a solução única.
  - Se  $\mathbf{X}'\mathbf{X}$  for não inversível ( $rank(\mathbf{X}) < p$ ), podemos considerar alguma inversa generalizada de  $\mathbf{X}'\mathbf{X}$ . 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  $\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-} \mathbf{X}'\mathbf{Y}$ .
- Em geral, vamos trabalhar com modelos em que a solução é única.

## Propriedades do Estimador de MQO (exercício)

- Uma vez que  $\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{Y}$ ,  $\mathbf{Y} \sim N_n(\mathbf{X}\beta, \sigma^2\mathbf{I}_n)$  e pelas propriedades associados à vetores aleatórios e a distribuição normal multivariada, temos que:
  - $\mathcal{E}(\hat{\beta}) = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathcal{E}(\mathbf{Y}) = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{X}\beta = \beta$ . (não viciado).
  - $Cov(\hat{\beta}) = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'Cov(\mathbf{Y})\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} = \sigma^2 (\mathbf{X}'\mathbf{X})^{-1}$ .
  - $\hat{\beta} \sim N_p(\beta, \sigma^2 (\mathbf{X}'\mathbf{X})^{-1})$  (normalidade).
  - $\hat{\beta}_j \sim N_1(\beta_j, \sigma^2\psi_j)$ ,  $\psi_j$  é o  $j$ -ésimo elemento da diagonal principal da matriz  $(\mathbf{X}'\mathbf{X})^{-1}$ .
  - Observação: sob a suposição de normalidade, o estimador de MQO coincide com o estimador de MV (máxima verossimilhança).

## Estimador de $\sigma^2$

- Sob normalidade, o estimador de máxima verosimilhança de  $\sigma^2$  é dado por

$$\hat{\sigma}_{MV}^2 = \frac{1}{n} (\mathbf{Y} - \mathbf{X}\hat{\beta})' (\mathbf{Y} - \mathbf{X}\hat{\beta})$$

o qual é viciado.

- Na prática considera-se o seguinte estimador:

$$\hat{\sigma}^2 = \frac{1}{n-p} (\mathbf{Y} - \mathbf{X}\hat{\beta})' (\mathbf{Y} - \mathbf{X}\hat{\beta})$$

- O qual é não-viciado. Além disso, pode-se provar que  $\hat{\beta} \perp \hat{\sigma}^2$  e  $\frac{(n-p)\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{(n-p)}$  (exercício).

# Inferência

- Temos que

$$\hat{\beta}_j \sim N(\beta_j, \sigma^2 \psi_j); \frac{(n-p)\hat{\sigma}^2}{\sigma^2} \sim \chi_{(n-p)}^2; \hat{\beta}_j \perp \frac{(n-p)\hat{\sigma}^2}{\sigma^2}.$$

- Logo,  $\frac{\hat{\beta}_j - \beta_j}{\sqrt{\hat{\sigma}^2 \psi_j}} \sim t_{(n-p)}$ , portanto (considerando-se  $P(X \leq t_{\frac{1+\gamma}{2}}) = \frac{1+\gamma}{2}$ ,  $X \sim t_{(n-p)}$ ), temos que

$$IC(\beta_j, \gamma) = \left[ \hat{\beta}_j - t_{\frac{1+\gamma}{2}} \sqrt{\hat{\sigma}^2 \psi_j}; \hat{\beta}_j + t_{\frac{1+\gamma}{2}} \sqrt{\hat{\sigma}^2 \psi_j} \right]$$

# Testes de hipóteses

- Suponha que queremos testar  $H_0 : \beta_j = \beta_{j0}$  vs  $H_1 : \beta_j \neq \beta_{j0}$ , para algum  $j$ , em que  $\beta_{j0}$  é um valor fixado.
- Estatística do teste  $T_t = \frac{\hat{\beta}_j - \beta_{j0}}{\sqrt{\hat{\sigma}^2 \psi_j}}$ , em que  $\hat{\beta}_j$  é o estimador de MQO de  $\beta_j$  e  $\hat{\sigma}^2 = \frac{1}{n-p} (\mathbf{Y} - \mathbf{X}\hat{\beta})' (\mathbf{Y} - \mathbf{X}\hat{\beta})$ .

# Testes de hipóteses

- Sob  $H_0$ ,  $T_t \sim t_{(n-p)}$ . Assim, rejeita-se  $H_0$  se  $|t_t| \geq t_c$ , em que  $t_t = \frac{\tilde{\beta}_j - \beta_{j0}}{\sqrt{\tilde{\sigma}^2 \psi_j}}$  e  $P(X \geq t_c | H_0) = \alpha/2$ ,  $X \sim t_{(n-p)}$ ,  $\tilde{\beta}_j$  é o  $j$ -ésimo elemento de  $\tilde{\beta} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}\mathbf{y}$  e  $\tilde{\sigma}^2 = \frac{1}{n-p} (\mathbf{y} - \mathbf{X}\tilde{\beta})' (\mathbf{y} - \mathbf{X}\tilde{\beta})$
- De modo equivalente, rejeita-se  $H_0$  se p-valor  $\leq \alpha$ , em que p-valor =  $2P(X \geq |t_t| | H_0)$ ,  $X \sim t_{(n-p)}$ .
- Para o MRNLH, a grande maioria dos resultados inferenciais são exatos.