

Revisão sobre modelos de regressão normais lineares

Prof. Caio Azevedo

Exemplo 8: Teste de esforço cardiopulmonar

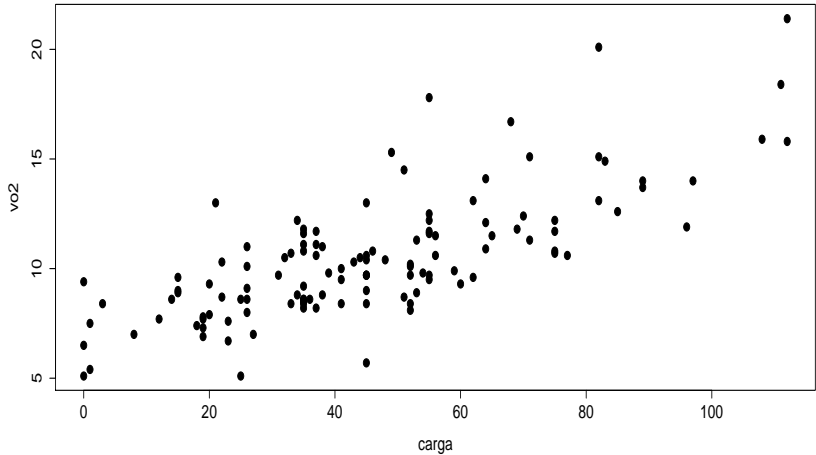
- Considere o estudo sobre teste de esforço cardiopulmonar em pacientes com insuficiência cardíaca realizado no InCor da Faculdade de Medicina da USP pela Dra. Ana Fonseca Braga.
- Um dos objetivos do estudo é comparar os grupos formados pelas diferentes etiologias cardíacas quanto às respostas respiratórias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

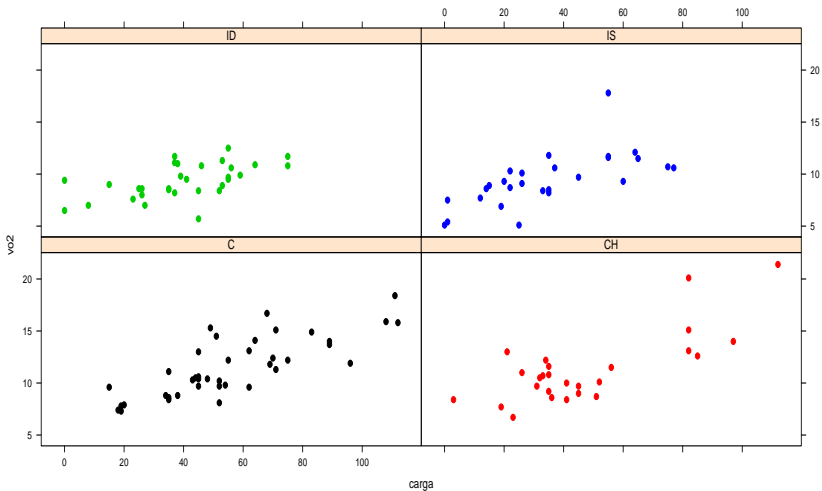
Dados

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 oxigenio em funcao da carga



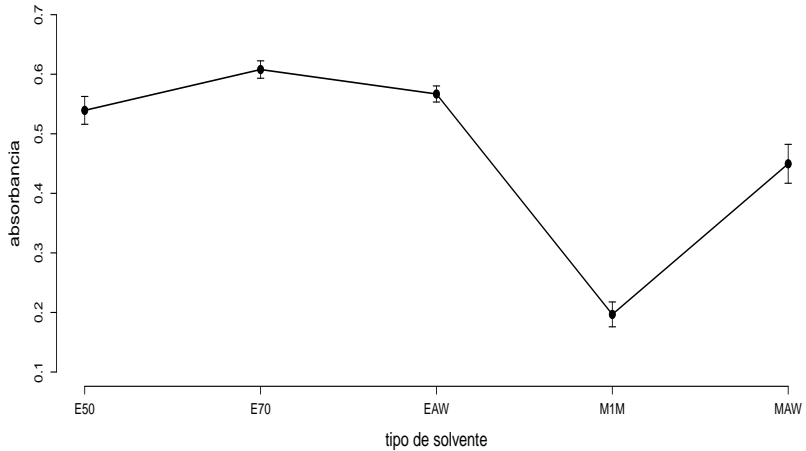
Exemplo 9: Medidas de absorvância

- Exemplo 2: Uma bioquímica (Tecnóloga 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-se o melhor solvente extrator. A escolha do(s) melhor(es) solventes foi realizada através da medida da absorvância de um pigmento natural do fruto de baguaçu. Fator = tipos de solvente; $k=5$ níveis; $n_k=5$ repetições.

- Quanto maior a absorvência, melhor o solvente.
- Unidade experimental: 10 gramas de polpa do fruto de baguaçu.
- 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



Modelagem

- Em todos os exemplos, podemos considerar em algum tipo de modelagem estatística, para responder às perguntas de interesse.
- Uma possibilidade: modelos normais lineares homocedásticos (heterocedásticos).
- A escolha de um modelo deve ser pautada: nos objetivos do experimento, nas características dos dados, em experiências anteriores e na análise descritiva.

Cont.

- Tais modelos (de regressão, de planejamento ou de Análise de Covariância) podem ser decompostos em uma parte sistemática e uma parte aleatória.
- Todos eles podem ser acomodados em uma estrutura geral chamada de modelo linear normal homocedástico (heterocedástico) geral.

Exemplo 8: 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 (conhecido e não aleatório).
- Parte sistemática: $\mathcal{E}(Y_i) = \beta_0 + \beta_1 x_i$.
- Parte aleatória: ξ_i .
- O modelo acima implica que $Y_i \stackrel{ind.}{\sim} N(\beta_0 + \beta_1 x_i, \sigma^2)$

- β_1 : é o aumento 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:

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

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

Exemplo 8: considerando as etiologias cardíacas

$$Y_{ij} : \beta_{0i} + \beta_{1i}x_{ij} + \xi_{ij}, i = 1, \dots, 4; j = 1, \dots, n_i$$

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

Exemplo 9: Modelo (parametrização de médias)

$$Y_{ij} = \mu_i + \xi_{ij},$$

$i = 1, 2, \dots, 5$ (grupos); $j = 1, \dots, 5$ (unidades experimentais)

- Erros (parte aleatória): $\xi_{ij} \stackrel{i.i.d.}{\sim} N(0, \sigma^2)$, $\mu_i, i = 1, 2, 3, 4, 5$ não aleatório.
- $\mathcal{E}_{\xi_{ij}}(Y_{ij}) = \mu_i, \mathcal{V}_{\xi_{ij}}(Y_{ij}) = \sigma^2$.
- Parte sistemática: μ_i - média populacional relacionada ao i -ésimo fator.
- $Y_{ij} \stackrel{ind.}{\sim} N(\mu_i, \sigma^2)$.

Exemplo 9: Modelo (decomposição das médias)

$$Y_{ij} = \mu + \alpha_i + \xi_{ij},$$

$i = 1, 2, \dots, 5$ (grupos); $j = 1, \dots, 5$ (unidades experimentais)

- Erros (parte aleatória): $\xi_{ij} \stackrel{i.i.d.}{\sim} N(0, \sigma^2)$, $\mu, \alpha_i = 1, 2, 3, 4, 5$ não aleatórios.
- $\mathcal{E}_{\xi_{ij}}(Y_{ij}) = \mu_i, \mathcal{V}_{\xi_{ij}}(Y_{ij}) = \sigma^2$.
- Parte sistemática: $\mu_i = \mu + \alpha_i$ - média populacional relacionada ao i -ésimo fator.
- $Y_{ij} \stackrel{ind.}{\sim} N(\mu + \alpha_i, \sigma^2)$.

- Note que, do modo como o modelo foi definido, temos cinco médias $\theta = (\mu_1, \mu_2, \mu_3, \mu_4, \mu_5)'$ e seis parâmetros $\beta = (\mu, \alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5)'$.
- Ou seja, dizemos que o modelo está superparametrizado.
- Isto implica que o modelo não está identificado, ou seja, diferentes valores para os parâmetros β podem levar à uma mesma verossimilhança.
- Duas opções:
 - Identificar o modelo (reduzindo o número de parâmetros no vetor β) (abordagem escolhida).
 - Trabalhar com o modelo não identificável, concentrando os esforços nas **funções estimáveis** (pesquisar).

- No entanto, se fizermos $\alpha_1 = 0$, teremos que $\mu = \mu_1$, ou seja, ele será a média (populacional) do grupo de referência (grupo 1, neste caso).
- Além disso, $\alpha_i = \mu_i - \mu_1$ é a diferença entre a média do grupo i e a média do grupo de referência.
- Por exemplo: E50 ($i = 1$), E70 ($i = 2$), EAW ($i = 3$), M1M ($i = 4$) e MAW ($i = 5$).

Exemplo 9: Modelo (casela de referência)

$$Y_{ij} = \mu + \alpha_i + \xi_{ij},$$

$i = 1, 2, \dots, 5$ (grupos); $j = 1, \dots, 5$ (unidades experimentais)

- Erros (parte aleatória): $\xi_{ij} \stackrel{i.i.d.}{\sim} N(0, \sigma^2)$, $\mu, \alpha_i, i = 1, 2, 3, 4, 5$ não aleatórios com, $\alpha_1 = 0$.
- $\mathcal{E}_{\xi_{ij}}(Y_{ij}) = \mu_i, \mathcal{V}_{\xi_{ij}}(Y_{ij}) = \sigma^2$.
- Parte sistemática: $\mu_i = \mu + \alpha_i$ que é a média populacional relacionada ao i -ésimo fator.
- $Y_{ij} \stackrel{ind.}{\sim} N(\mu + \alpha_i, \sigma^2)$.
- Interpretações: $\mu = \mu_1$, ou seja é a média do grupo de referência e $\alpha_i = \mu_i - \mu_1$.

Exemplo 9: Modelo (desvios com restrição)

$$Y_{ij} = \mu + \alpha_i + \xi_{ij},$$

$i = 1, 2, \dots, 5$ (grupos); $j = 1, \dots, 5$ (unidades experimentais)

- Erros (parte aleatória): $\xi_{ij} \stackrel{i.i.d.}{\sim} N(0, \sigma^2)$, $\mu, \alpha_i, i = 1, 2, 3, 4, 5$ não aleatórios, com $\sum_{i=1}^5 \alpha_i = 0$ ou $\alpha_5 = -\sum_{i=1}^4 \alpha_i$.
- $\mathcal{E}_{\xi_{ij}}(Y_{ij}) = \mu_i, \mathcal{V}_{\xi_{ij}}(Y_{ij}) = \sigma^2$.
- Parte sistemática: $\mu_i = \mu + \alpha_i$ que é a média populacional relacionada ao i -ésimo fator.
- $Y_{ij} \stackrel{ind.}{\sim} N(\mu + \alpha_i, \sigma^2)$.
- Interpretações: $\mu = \bar{\mu} = \frac{1}{5} \sum_{i=1}^5 \mu_i$, é média das médias e $\alpha_i = \mu_i - \bar{\mu}$.

Exemplo 9: Modelo (desvios sem restrição)

$$Y_{ij} = \mu + \alpha_i + \xi_{ij},$$

$i = 1, 2, \dots, 5$ (grupos); $j = 1, \dots, 5$ (unidades experimentais)

- Erros (parte aleatória): $\xi_{ij} \stackrel{i.i.d}{\sim} N(0, \sigma^2)$, $\mu, \alpha_i, i = 1, 2, 3, 4, 5$ não aleatórios e não se impõe restrições entre os parâmetros.
- $\mathcal{E}_{\xi_{ij}}(Y_{ij}) = \mu_i, \mathcal{V}_{\xi_{ij}}(Y_{ij}) = \sigma^2$.
- Parte sistemática: $\mu_i = \mu + \alpha_i$ que é a média populacional relacionada ao i -ésimo fator .
- $Y_{ij} \stackrel{ind.}{\sim} N(\mu + \alpha_i, \sigma^2)$.
- Os parâmetros μ e $\alpha_i, i = 1, 2, 3, 4, 5$ não possuem interpretação.

Resumindo

- Nas parametrizações: de médias, casela de referência e desvios com restrição o modelo está identificado e os parâmetros μ e α_i possuem interpretações bem definidas.
- Na parametrização desvios sem restrição o modelo não está identificado e os parâmetros μ e α_i não possuem interpretações.
- Trabalharemos quase sempre com a parametrização casela de referência.
- Em qualquer uma das quatro parametrizações temos que $\mu_i = \mu + \alpha_i$. Para um mesmo conjunto de dados e um mesmo método de estimação, as estimativas de μ_i são as mesmas, entre as quatro parametrizações.

Algumas hipóteses de interesse

- Igualdade simultânea das cinco médias:
 - Parametrização de médias: $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$ vs H_1 : há pelo menos uma diferença.
 - Parametrização casela de referência : $H_0 : \alpha_2 = \alpha_3 = \alpha_4 = \alpha_5 = 0$ vs H_1 : há pelo menos uma diferença.
 - Parametrização desvios com restrição: $H_0 : \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 0$ vs H_1 : há pelo menos uma diferença.
 - Parametrização desvios sem restrição:
 $H_0 : \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = \alpha_5 = 0$ vs H_1 : há pelo menos uma diferença.

- Note que, em todos os casos, os modelos estão bem definidos, no sentido de que todas as suposições foram descritas e os parâmetros, interpretados (com exceção da parametrização desvios sem restrição).
- Os modelos anteriores se enquadram na classe dos modelos de regressão normais lineares homocedásticos (MRNL).

Revisão normal multivariada

- Dizemos que $\mathbf{Y} = (Y_1, \dots, Y_p)' \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ se sua fdp é dada por

$$f_{\mathbf{Y}}(\mathbf{y}) = |\boldsymbol{\Sigma}|^{-1/2} (2\pi)^{-p/2} \exp \left\{ -\frac{1}{2} (\mathbf{y} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{y} - \boldsymbol{\mu}) \right\} \mathbb{1}_{\mathcal{R}^p}(\mathbf{y})$$

- $\boldsymbol{\mu}$ é o vetor de médias e $\boldsymbol{\Sigma}$ é a matriz de covariâncias.

- Forma quadrática: seja $\mathbf{x} = (x_1, x_2)'$ e $\mathbf{A} = \begin{bmatrix} a_1 & a_2 \\ a_3 & a_4 \end{bmatrix}$, então

$$z = \mathbf{x}' \mathbf{A} \mathbf{x} = \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} a_1 & a_2 \\ a_3 & a_4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \\ a_1 x_1^2 + (a_2 + a_3) x_1 x_2 + a_4 x_2^2.$$

- Se \mathbf{A} for uma matriz simétrica, então $z = a_1 x_1^2 + 2a_2 x_1 x_2 + a_4 x_2^2$.

Parâmetros

$$\blacksquare \boldsymbol{\mu} = \mathcal{E}(\mathbf{Y}) = \begin{bmatrix} \mathcal{E}(Y_1) \\ \mathcal{E}(Y_2) \\ \vdots \\ \mathcal{E}(Y_p) \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{bmatrix}$$

$$\blacksquare \boldsymbol{\Sigma} = \text{Cov}(\mathbf{Y}) = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \dots & \sigma_{1p} \\ \sigma_{12} & \sigma_2^2 & \dots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1p} & \sigma_{2p} & \dots & \sigma_p^2 \end{bmatrix}$$

Propriedades

- Fechada sob marginalização: $Y_i \sim N(\mu_i, \sigma_i^2)$.
- $Y_i \perp Y_j, \forall i \neq j \Leftrightarrow \sigma_{ij} = 0$.
- Se $\mathbf{A}_{(q \times p)}$ e $\mathbf{B}_{(q \times 1)}$ forem matrizes não aleatórias, então $\mathbf{V} = \mathbf{A}\mathbf{Y} + \mathbf{B} \sim N_q(\mathbf{A}\boldsymbol{\mu} + \mathbf{B}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}')$.
- Se $\mathbf{A}_{(p \times p)}$ for uma matriz não aleatória, simétrica e idempotente de rank = p , $\boldsymbol{\mu} = \mathbf{0}$ e $\boldsymbol{\Sigma} = \sigma^2 \mathbf{I}_{(p \times p)}$, então $V = \frac{1}{\sigma^2} \mathbf{Y}'\mathbf{A}\mathbf{Y} \sim \chi_r^2, r = \text{tr}(\mathbf{A})$.
- Se $\mathbf{A}_{(p \times p)}$ for uma matriz não aleatória, então $\mathcal{E}(\mathbf{Y}'\mathbf{A}\mathbf{Y}) = \text{tr}(\mathbf{A}\boldsymbol{\Sigma}) + \boldsymbol{\mu}'\mathbf{A}\boldsymbol{\mu}$.

Propriedades (cont.)

- Se $\mathbf{A}_{(p \times p)}$ for uma matriz não aleatória, simétrica de rank = p e $\boldsymbol{\mu} = \mathbf{0}$, então $V = \mathbf{Y}'\mathbf{A}\mathbf{Y} \sim \chi_r^2, r = \text{tr}(\mathbf{A}) \Leftrightarrow \mathbf{A}\boldsymbol{\Sigma}\mathbf{A} = \mathbf{A}$.
- Sejam $\mathbf{A}_{(q \times p)}$ e $\mathbf{B}_{(s \times p)}$ matrizes não aleatórias e $\mathbf{V} = \mathbf{A}\mathbf{Y}$ e $\mathbf{W} = \mathbf{B}\mathbf{Y}$. Então $\mathbf{V} \perp \mathbf{W} \Leftrightarrow \mathbf{A}\boldsymbol{\Sigma}\mathbf{B}' = \mathbf{0}$.
- Sejam $\mathbf{A}_{(p \times p)}$ e $\mathbf{B}_{(p \times p)}$ matrizes não aleatórias e simétricas e $V = \mathbf{Y}'\mathbf{A}\mathbf{Y}$ e $W = \mathbf{Y}'\mathbf{B}\mathbf{Y}$. Então $\mathbf{V} \perp \mathbf{W} \Leftrightarrow \mathbf{A}\boldsymbol{\Sigma}\mathbf{B} = \mathbf{0}$.

Notação matricial para o MRNL

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

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

- Suposição $\boldsymbol{\xi} \sim N_n(\mathbf{0}, \sigma^2 \mathbf{I}_n)$ (que é o vetor de erros).
- \mathbf{Y} é o vetor das variáveis resposta.
- O índice n da variável resposta é geral e pode representar combinações de índices.
- \mathbf{X} é a matriz de plajenamento (ou delineamento) que define a parte sistemática do modelo.

Exemplo 8

- Para o primeiro modelo, temos que

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

- Naturalmente, depois do experimento ser realizado, teremos um vetor de observações (números): $\mathbf{y} = (y_1, \dots, y_{124})'$.
- Exercício: escrever na forma matricial o segundo modelo (no qual se considera as diferentes etiologias cardíacas).

Exemplo 9 (parametrização de médias)

$$\mathbf{Y} = \begin{bmatrix} Y_{11} \\ Y_{12} \\ Y_{13} \\ Y_{14} \\ Y_{15} \\ Y_{21} \\ \vdots \\ Y_{51} \\ \vdots \\ Y_{55} \end{bmatrix}, \mathbf{X} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \boldsymbol{\beta} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \\ \mu_4 \\ \mu_5 \end{bmatrix}, \boldsymbol{\xi} = \begin{bmatrix} \xi_{11} \\ \xi_{12} \\ \xi_{13} \\ \xi_{14} \\ \xi_{15} \\ \xi_{21} \\ \vdots \\ \xi_{51} \\ \vdots \\ \xi_{55} \end{bmatrix}$$

Naturalmente, depois do experimento ser realizado, teremos um vetor de observações (números)

$$\mathbf{y} = (y_1, \dots, y_{25})'.$$

Exemplo 9 (casela de referência)

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

Naturalmente, depois do experimento ser realizado, teremos um vetor de observações (números)

$$\mathbf{y} = (y_1, \dots, y_{25})'.$$

Exemplo 9 (desvios com restrição)

$$\mathbf{Y} = \begin{bmatrix} Y_{11} \\ Y_{12} \\ Y_{13} \\ Y_{14} \\ Y_{15} \\ Y_{21} \\ \vdots \\ \vdots \\ Y_{51} \\ \vdots \\ \vdots \\ Y_{55} \end{bmatrix}, \mathbf{X} = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 1 & 0 & 0 & 1 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 1 & 0 & 0 & 0 & 1 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 1 & 0 & 0 & 0 & 1 \\ 1 & -1 & -1 & -1 & -1 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 1 & -1 & -1 & -1 & -1 \end{bmatrix}, \boldsymbol{\beta} = \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \end{bmatrix}, \boldsymbol{\xi} = \begin{bmatrix} \xi_{11} \\ \xi_{12} \\ \xi_{13} \\ \xi_{14} \\ \xi_{15} \\ \xi_{21} \\ \vdots \\ \vdots \\ \xi_{51} \\ \vdots \\ \vdots \\ \xi_{55} \end{bmatrix}$$

Naturalmente, depois do experimento ser realizado, teremos um vetor de observações (números)

$$\mathbf{y} = (y_1, \dots, y_{25})'.$$

Exemplo 9 (desvios sem restrição)

$$\mathbf{Y} = \begin{bmatrix} Y_{11} \\ Y_{12} \\ Y_{13} \\ Y_{14} \\ Y_{15} \\ Y_{21} \\ \vdots \\ \vdots \\ Y_{51} \\ \vdots \\ \vdots \\ Y_{55} \end{bmatrix}, \mathbf{X} = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 1 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \boldsymbol{\beta} = \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \\ \alpha_5 \end{bmatrix}, \boldsymbol{\xi} = \begin{bmatrix} \xi_{11} \\ \xi_{12} \\ \xi_{13} \\ \xi_{14} \\ \xi_{15} \\ \xi_{21} \\ \vdots \\ \vdots \\ \xi_{51} \\ \vdots \\ \vdots \\ \xi_{55} \end{bmatrix}$$

Naturalmente, depois do experimento ser realizado, teremos um vetor de observações (números)

$$\mathbf{y} = (y_1, \dots, y_{25})'.$$

Estimação dos parâmetros

- Estimador usual para β : mínimos quadrados ordinários (MQO).
- Objetivo: obter β 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 $\frac{\partial Q(\beta)}{\partial \beta} = \mathbf{0}$ (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:

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

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 ($\text{rank}(\mathbf{X}) = p$), a solução única.
 - Se $\mathbf{X}'\mathbf{X}$ for não inversível ($\text{rank}(\mathbf{X}) < p$), 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 que $(\mathbf{X}'\mathbf{X})^{-}$ é uma inversa generalizada de $\mathbf{X}'\mathbf{X}$).
- Em princípio, vamos trabalhar com modelos (parametrizações) para os quais a solução é única.

Propriedades do Estimador de MQO

- 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 associadas 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).
 - $\hat{\beta}_j \sim N_p(\beta_j, \sigma^2 \sigma_{\beta_j}), j = 1, \dots, p$, σ_{β_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).

Estimadores de MQO nos exemplos anteriores

- No Exemplo 8 (modelo 1), temos que

$$\mathbf{X}'\mathbf{Y} = \begin{bmatrix} n\bar{Y} \\ \sum_{i=1}^{124} Y_i X_i \end{bmatrix}; \mathbf{X}'\mathbf{X} = \begin{bmatrix} n & \sum_{i=1}^{124} X_i \\ \sum_{i=1}^{124} X_i & \sum_{i=1}^{124} X_i^2 \end{bmatrix}$$

$$\rightarrow (\mathbf{X}'\mathbf{X})^{-1} = \frac{1}{n \sum_{i=1}^{124} X_i^2 - \left(\sum_{i=1}^{124} X_i\right)^2} \begin{bmatrix} \sum_{i=1}^{124} X_i^2 & -\sum_{i=1}^{124} X_i \\ -\sum_{i=1}^{124} X_i & n \end{bmatrix}$$

- Assim $\hat{\beta} = \begin{bmatrix} \frac{(\sum_{i=1}^{124} X_i^2)(\sum_{i=1}^{124} Y_i) - (\sum_{i=1}^{124} X_i)(\sum_{i=1}^{124} Y_i X_i)}{n \sum_{i=1}^{124} X_i^2 - (\sum_{i=1}^{124} X_i)^2} \\ \frac{n \sum_{i=1}^{124} Y_i X_i - (\sum_{i=1}^{124} X_i)(\sum_{i=1}^{124} Y_i)}{n \sum_{i=1}^{124} X_i^2 - (\sum_{i=1}^{124} X_i)^2} \end{bmatrix}$.

Estimadores de MQO nos exemplos anteriores

- No Exemplo 9 (casela de referência, $\alpha_1 = 0$), temos que

$$\mathbf{X}'\mathbf{Y} = \begin{bmatrix} n\bar{Y} \\ n_2\bar{Y}_2 \\ n_3\bar{Y}_3 \\ n_4\bar{Y}_4 \\ n_5\bar{Y}_5 \end{bmatrix}; \mathbf{X}'\mathbf{X} = \begin{bmatrix} n & n_2 & n_3 & n_4 & n_5 \\ n_2 & n_2 & 0 & 0 & 0 \\ n_3 & 0 & n_3 & 0 & 0 \\ n_4 & 0 & 0 & n_4 & 0 \\ n_5 & 0 & 0 & 0 & n_5 \end{bmatrix}; \hat{\beta} = \begin{bmatrix} \bar{Y}_1 \\ \bar{Y}_2 - \bar{Y}_1 \\ \bar{Y}_3 - \bar{Y}_1 \\ \bar{Y}_4 - \bar{Y}_1 \\ \bar{Y}_5 - \bar{Y}_1 \end{bmatrix}$$

- Exercício: repetir os desenvolvimentos acima para as reparametrizações de médias e desvios com restrição.

Estimadores de MQO nos exemplos anteriores

- No Exemplo 9 (desvios sem restrição), temos que

$$\mathbf{X}'\mathbf{Y} = \begin{bmatrix} n\bar{Y} \\ n_1\bar{Y}_1 \\ n_2\bar{Y}_2 \\ n_3\bar{Y}_3 \\ n_4\bar{Y}_4 \\ n_5\bar{Y}_5 \end{bmatrix}; \mathbf{X}'\mathbf{X} = \begin{bmatrix} n & n_1 & n_2 & n_3 & n_4 & n_5 \\ n_1 & n_1 & 0 & 0 & 0 & 0 \\ n_2 & 0 & n_2 & 0 & 0 & 0 \\ n_3 & 0 & 0 & n_3 & 0 & 0 \\ n_4 & 0 & 0 & 0 & n_4 & 0 \\ n_5 & 0 & 0 & 0 & 0 & n_5 \end{bmatrix}$$

- Neste caso, $\mathbf{X}'\mathbf{X}$ não é inversível.

Estimador de σ^2

- Sob normalidade, o estimador de máxima verosimilhança de σ^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 $\frac{(n-p)\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{(n-p)}$.

Testes de hipóteses simples

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

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

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

$$H_0 : \beta_1 = 0 \text{ vs } H_1 : \beta_1 \neq 0,$$

- No terceiro modelo, podemos ter interesse em testar se as médias dos solventes 1 e 2 são iguais, o que, nas parametrizações casela de referência e desvios com restrição, equivale a testar se:

$$H_0 : \alpha_2 = 0 \text{ vs } H_1 : \alpha_2 \neq 0,$$

- Veremos como testar hipóteses mais gerais em breve. Contudo, hipóteses simples como as apresentadas, podem ser testadas usando-se o fato de que:

$$T = \frac{\hat{\beta}_i - \beta_{0i}}{\sqrt{\hat{\sigma}^2 \sigma_{\beta_j}}} \sim t_{(n-p)}, \text{ sob } H_0. \quad (2)$$

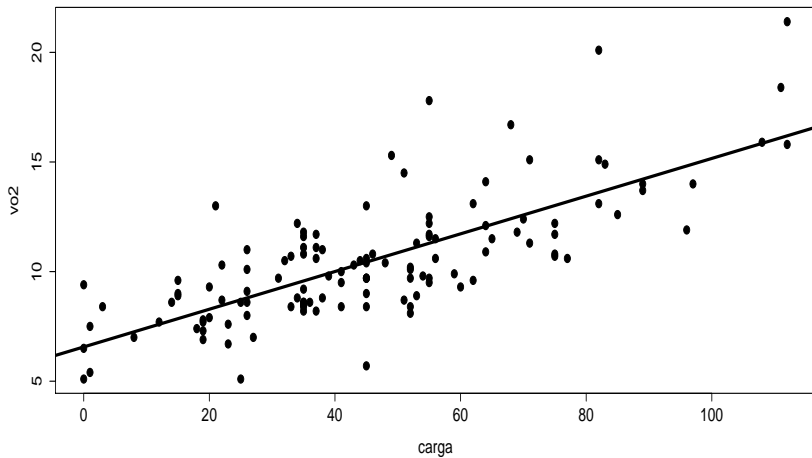
- Lembremos que, no caso de hipóteses bi-laterais ($=$ vs \neq) o p-valor é dado por $p\text{-valor} = 2P(T > |t_c| | H_0)$, em que $T \sim t_{(n-p)}$, sob H_0 e t_c é o valor calculado da estatística definida em (2).
- Demonstração: exercício.

Exemplo 8: modelo 1

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.

Consumo de oxigênio em função da carga

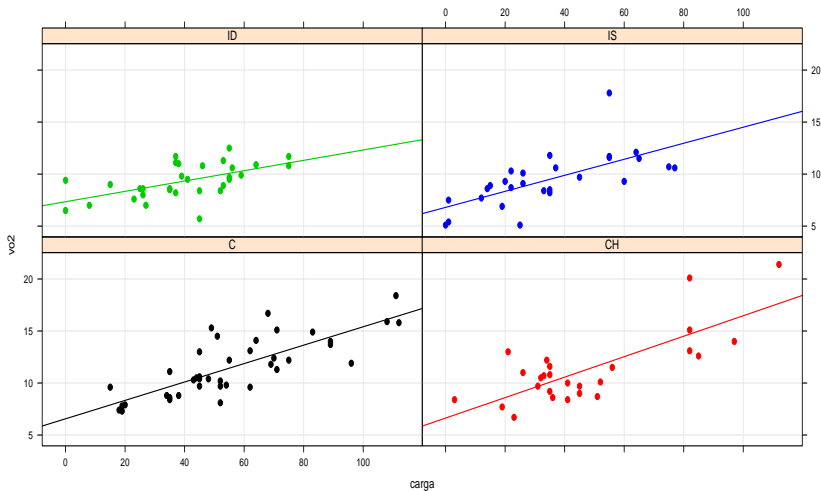


Exemplo 8: modelo 2

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 parecer ser menor que os demais, para pacientes idiopáticos e igual para os outros três tipos.

Consumo de oxigênio em função da carga



Exemplo 9

Parâmetro	Estimativa	EP	IC(95%)	Estat. t	pvalor
μ (E50)	0,539	0,011	[0,517; 0,561]	47,826	< 0,0001
α_2 (E70)	0,069	0,0160	[0,037 ; 0,010]	4,298	0,0003
α_3 (EAW)	0,028	0,0160	[-0,004 ; 0,059]	1,726	0,0998
α_4 (M1M)	-0,343	0,0160	[-0,374; -0,311]	-21,481	< 0,0001
α_5 (MAW)	-0,090	0,0160	[-0,121 ; -0,058]	-5,624	< 0,0001

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

Comentários

- Caso a normalidade não se verifique, podemos utilizar resultados assintóticos.
- Para testar as hipóteses anteriores, podemos utilizar a estatística:

$$Z = \frac{\hat{\beta}_i - \beta_{0i}}{\sqrt{\hat{\sigma}^2 \sigma_{\beta_j}}}$$

a qual, sob H_0 , sob certas condições e para n suficientemente grande, tem distribuição aproximadamente $N(0,1)$, sob H_0 .

- Algumas condições: $\hat{\beta}_i$ tem de ser um estimador com distribuição assintótica $N(\beta_{0i}, \sigma^2 \sigma_{\beta_j})$ e $\hat{\sigma}^2 \xrightarrow[n \rightarrow \infty]{P} \sigma^2$.

Comentários

- A suposição de heterocedasticidade também pode ser relaxada.
- Nesse caso, outro processo de estimação deverá ser usado.
- Também, outras estatísticas de teste terão de ser consideradas.

Mais hipóteses de interesse

- Voltemos ao exemplo 8 (modelo 2). Temos quatro interceptos $(\beta_{01}, \beta_{02}, \beta_{03}, \beta_{04})'$ e quatro coeficientes angulares $(\beta_{11}, \beta_{12}, \beta_{13}, \beta_{14})'$. Ou seja,
$$\beta = (\beta_{01}, \beta_{02}, \beta_{03}, \beta_{04}, \beta_{11}, \beta_{12}, \beta_{13}, \beta_{14})'$$
- Igualdade simultânea dos interceptos:

$$H_0 : \begin{cases} \beta_{01} - \beta_{02} = 0, \\ \beta_{01} - \beta_{03} = 0, \\ \beta_{01} - \beta_{04} = 0. \end{cases} \quad \text{vs } H_1 : \text{há pelo menos uma diferença}$$

Mais hipóteses de interesse

- Igualdade simultânea dos coeficientes angulares:

$$H_0 : \begin{cases} \beta_{11} - \beta_{12} = 0, \\ \beta_{11} - \beta_{13} = 0, \\ \beta_{11} - \beta_{14} = 0. \end{cases} \quad \text{vs } H_1 : \text{há pelo menos uma diferença}$$

Mais hipóteses de interesse

- Igualdade simultânea dos interceptos (entre si) e dos coeficientes angulares (entre si):

$$H_0 : \begin{cases} \beta_{01} - \beta_{02} = 0, \\ \beta_{01} - \beta_{03} = 0, \\ \beta_{01} - \beta_{04} = 0, \\ \beta_{11} - \beta_{12} = 0, \\ \beta_{11} - \beta_{13} = 0, \\ \beta_{11} - \beta_{14} = 0. \end{cases} \quad \text{vs } H_1 : \text{há pelo menos uma diferença}$$

Mais hipóteses de interesse

- Voltando ao exemplo 9 (considerando a parametrização casela de referência). Lembrando que $\beta = (\mu, \alpha_2, \alpha_3, \alpha_4, \alpha_5)'$. Igualdade simultânea entre as médias:

$$H_0 : \alpha_2 = \alpha_3 = \alpha_4 = \alpha_5 = 0 \text{ vs } H_1 : \text{há pelo menos uma diferença}$$

- A média das médias dos três primeiros grupos é igual a média do quinto grupo:

$$H_0 : \frac{\mu_1 + \mu_2 + \mu_3}{3} = \mu_5 \text{ vs } H_1 : \frac{\mu_1 + \mu_2 + \mu_3}{3} \neq \mu_5$$

$$\Leftrightarrow H_0 : \alpha_2 + \alpha_3 - 3\alpha_5 = 0 \text{ vs } H_1 : \alpha_2 + \alpha_3 - 3\alpha_5 \neq 0$$

Hipóteses de interesse

- Todas as hipóteses anteriores (e outras tantas) podem ser reescritas como :

$$H_0 : \mathbf{C}_{(r \times p)} \boldsymbol{\beta}_{(p \times 1)} = \mathbf{M}_{(r \times 1)} \text{ vs } H_1 : \mathbf{C}_{(r \times p)} \boldsymbol{\beta}_{(p \times 1)} \neq \mathbf{M}_{(r \times 1)},$$

em geral, $r \leq p$ e \mathbf{C} é uma matriz de posto linha completo = r e \mathbf{M} um vetor, ambos conhecidos e não aleatórios.

Hipóteses de interesse

- Voltando ao exemplo 8, modelo 2. Lembrando que.

$\beta = (\beta_{01}, \beta_{02}, \beta_{03}, \beta_{04}, \beta_{11}, \beta_{12}, \beta_{13}, \beta_{14})'$. A hipótese de igualdade simultânea dos interceptos é obtida através de:

$$\mathbf{C} = \begin{bmatrix} 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & -1 & 0 & 0 & 0 & 0 \end{bmatrix}; \mathbf{M} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

Hipóteses de interesse

- Em relação à hipótese de igualdade simultânea dos coeficientes angulares, temos que:

$$\mathbf{C} = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & -1 \end{bmatrix}; \mathbf{M} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

Hipóteses de interesse

- Em relação à hipótese de igualdade simultânea dos interceptos (entre si) e dos coeficientes angulares (entre si), temos que:

$$\mathbf{C} = \begin{bmatrix} 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & -1 \end{bmatrix}; \mathbf{M} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

Hipóteses de interesse

- Voltando ao exemplo 9, sob a parametrização casela de referência. Lembrando que. $\beta = (\mu, \alpha_2, \alpha_3, \alpha_4, \alpha_5)'$. A hipótese de igualdade simultânea de igualdade de médias é obtida através de:

$$\mathbf{C} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}; \mathbf{M} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

Hipóteses de interesse

- Em relação à hipótese de igualdade entre a média das médias dos três primeiros grupos e a média do quinto grupo, temos:

$$\mathbf{C} = \begin{bmatrix} 0 & 1 & 1 & 0 & -3 \end{bmatrix}; \mathbf{M} = \begin{bmatrix} 0 \end{bmatrix}$$

Teste de hipóteses $C\beta = M$

- Lembremos que $\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{Y}$, e $\hat{\beta} \sim N_p(\beta, \sigma^2 (\mathbf{X}'\mathbf{X})^{-1})$.
- Assim, $\hat{\theta} = \mathbf{C}\hat{\beta} - \mathbf{M} \sim N_r(\mathbf{C}\hat{\beta} - \mathbf{M}, \sigma^2 \mathbf{C}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{C}')$.
- Logo, $Q_1 = \frac{1}{\sigma^2} \hat{\theta}' [\mathbf{C}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{C}']^{-1} \hat{\theta} \sim \chi_r^2$, sob H_0 .
- Se σ^2 for conhecido, podemos utilizar a estatística Q_1 e rejeitar H_0 se $q_1 > q_c$ (q_1 : valor calculado da estatística do teste),
 $P(\chi_r^2 \geq q_c | H_0)$, ou se se p -valor = $P(\chi_r^2 \geq q_1 | H_0) < \alpha$.
- Esse raciocínio será utilizado em todas as outras estatísticas.

Teste de hipóteses $C\beta = M$

- Contudo, em geral, σ^2 é conhecido.
- Lembremos que $\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{Y}$, e $\hat{\beta} \sim N_p(\beta, \sigma^2 (\mathbf{X}'\mathbf{X})^{-1})$.
- Assim, $\hat{\theta} = C\hat{\beta} - M \sim N_r(C\hat{\beta} - M, \sigma^2 C(\mathbf{X}'\mathbf{X})^{-1}C')$.
- Logo, $\frac{1}{\sigma^2} \hat{\theta}' [C(\mathbf{X}'\mathbf{X})^{-1}C']^{-1} \hat{\theta} \sim \chi_r^2$ sob H_0 .
- Como σ^2 é desconhecido, uma opção seria substituí-lo por um estimador consistente, por exemplo,

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

Teste de hipóteses $C\beta = M$

- Portanto

$$\begin{aligned} Q &= \frac{1}{\hat{\sigma}^2} \hat{\theta}' [C(X'X)^{-1}C']^{-1} \hat{\theta} \\ &= \frac{1}{\hat{\sigma}^2} (C\hat{\beta} - M)' [C(X'X)^{-1}C']^{-1} (C\hat{\beta} - M) \\ &\approx \chi_r^2 \end{aligned}$$

sob H_0 , para n suficientemente grande.

- Contudo, sob as suposições de normalidade, independência e homocedasticidade, podemos utilizar uma outra estatística.
- Lembremos que $\frac{(n-p)\hat{\sigma}^2}{\sigma^2} \sim \chi_{(n-p)}^2$.

Teste de hipóteses $C\beta = M$

- Assim, tem-se que

$$\begin{aligned} Q_0 &= \frac{\left[\frac{1}{\sigma^2} \hat{\theta}' [C(X'X)^{-1}C']^{-1} \hat{\theta} \right] / r}{\left[\frac{(n-p)\hat{\sigma}^2}{\sigma^2} \right] / (n-p)} = \frac{\hat{\theta}' [C(X'X)^{-1}C']^{-1} \hat{\theta}}{r\hat{\sigma}^2} \\ &= \frac{(C\hat{\beta} - M)' [C(X'X)^{-1}C']^{-1} (C\hat{\beta} - M)}{r\hat{\sigma}^2} \sim F_{[r, n-p]} \end{aligned}$$

sob H_0 .

- Na ausência de normalidade, podemos considerar a estatística Q definida anteriormente.

Ausência de homocedasticidade

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

- Suposição: $\boldsymbol{\xi} \sim N_n(\mathbf{0}, \boldsymbol{\Sigma})$, em que

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2 & 0 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & 0 & \dots & 0 \\ 0 & 0 & \sigma_3^2 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \sigma_n^2 \end{bmatrix}$$

Tabela de contingência $r \times 2$: produto de binomiais independentes

		Variável 1 (resposta)		
		C_{11}	C_{12}	Total
Variável 2 (explicativa)	C_{21}	$N_{11}(\theta_{11})$	$N_{12}(\theta_{12})$	$n_{1.}$
	C_{22}	$N_{21}(\theta_{21})$	$N_{22}(\theta_{22})$	$n_{2.}$
	\vdots	\vdots	\vdots	\vdots
	C_{2r}	$N_{r1}(\theta_{r1})$	$N_{r2}(\theta_{r2})$	$n_{r.}$
Total	-	$N_{.1}$	$N_{.2}$	$n_{..}$

Tabela de contingência $r \times 2$: produto de binomiais independentes

- Neste caso, $N_{i1} \stackrel{ind.}{\sim} \text{binomial}(n_i, \theta_{i1}), i = 1, \dots, r$. Definindo

$\mathbf{N} = (N_{11}, N_{21}, \dots, N_{r1})'$, temos que

$$\text{Cov}(\mathbf{N}) = \begin{bmatrix} n_1 \cdot \theta_{11} (1 - \theta_{11}) & 0 & 0 & \dots & 0 \\ 0 & n_2 \cdot \theta_{21} (1 - \theta_{21}) & 0 & \dots & 0 \\ 0 & 0 & n_3 \cdot \theta_{31} (1 - \theta_{31}) & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & n_r \cdot \theta_{r1} (1 - \theta_{r1}) \end{bmatrix}$$

Ausência de homocedasticidade

- Neste caso, é mais apropriado considerar o estimador de mínimos quadrados ponderados (MQP).
- Objetivo: obter β que minimiza $Q(\beta) = (\mathbf{Y} - \mathbf{X}\beta)' \Sigma^{-1} (\mathbf{Y} - \mathbf{X}\beta)$.
Em geral, $\beta \in \mathcal{R}^p$.
- Através de desenvolvimentos semelhantes ao caso do estimador de MQO, obtemos:

$$\hat{\beta}_W = (\mathbf{X}' \Sigma^{-1} \mathbf{X})^{-1} \mathbf{X}' \Sigma^{-1} \mathbf{Y}$$

- Pode-se provar que: $\hat{\beta}_W \sim N_p \left(\beta, (\mathbf{X}' \Sigma^{-1} \mathbf{X})^{-1} \right)$.

Ausência de homocedasticidade (testes de hipótese)

- Considere Σ conhecido.
- Para testar as hipóteses anteriores, podemos considerar semelhante construção relacionada à estatística Q_O .
- Nesse caso, temos que

$$\hat{\theta}_W = \mathbf{C}\hat{\beta}_W - \mathbf{M} \sim N_r \left(\mathbf{C}\beta - \mathbf{M}, \mathbf{C} (\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1} \mathbf{C}' \right).$$

- Defina

$$\begin{aligned} Q_W &= \hat{\theta}'_W [\mathbf{C}(\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1}\mathbf{C}']^{-1} \hat{\theta}_W \\ &= (\mathbf{C}\hat{\beta}_W - \mathbf{M})' [\mathbf{C}(\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1}\mathbf{C}']^{-1} (\mathbf{C}\hat{\beta}_W - \mathbf{M}) \end{aligned}$$

- Assim, sob H_0 , $Q_W \sim \chi_r^2$.

Ausência de homocedasticidade (testes de hipótese)

- Assuma Σ desconhecido.
- Para testar as hipóteses anteriores, também podemos considerar semelhante construção relacionada à estatística Q_0 .
- Considere $\hat{\Sigma}$ um estimador consistente para Σ , ou seja $\hat{\Sigma} \xrightarrow[n \rightarrow \infty]{P} \Sigma$.

Ausência de homocedasticidade (testes de hipótese)

- Defina $\widehat{\beta}_W^* = \left(\mathbf{X}'\widehat{\Sigma}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'\widehat{\Sigma}^{-1}\mathbf{Y}$. Assim,

$$\widehat{\beta}_W^* \xrightarrow[n \rightarrow \infty]{D} N_p\left(\beta, \left(\mathbf{X}'\Sigma^{-1}\mathbf{X}\right)^{-1}\right).$$

- Defina também, $\widehat{\theta}_W^* = \mathbf{C}\widehat{\beta}_W^* - \mathbf{M}$. Logo,

$$\widehat{\theta}_W^* \xrightarrow[n \rightarrow \infty]{D} N_r\left(\mathbf{C}\beta - \mathbf{M}, \mathbf{C}\left(\mathbf{X}'\Sigma^{-1}\mathbf{X}\right)^{-1}\mathbf{C}'\right).$$

Ausência de homocedasticidade (testes de hipótese)

- Considere

$$\begin{aligned} Q_W^* &= \widehat{\boldsymbol{\theta}}_W^{*'} \left[\mathbf{C}(\mathbf{X}'\widehat{\boldsymbol{\Sigma}}^{-1}\mathbf{X})^{-1}\mathbf{C}' \right]^{-1} \widehat{\boldsymbol{\theta}}_W^* \\ &= \left(\mathbf{C}\widehat{\boldsymbol{\beta}}_W^* - \mathbf{M} \right)' \left[\mathbf{C}(\mathbf{X}'\widehat{\boldsymbol{\Sigma}}^{-1}\mathbf{X})^{-1}\mathbf{C}' \right]^{-1} \left(\mathbf{C}\widehat{\boldsymbol{\beta}}_W^* - \mathbf{M} \right) \end{aligned}$$

- Sob H_0 , para n suficientemente grande, $Q_W^* \approx \chi_r^2$.

Ausência de homocedasticidade e independência

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

- Suposição $\boldsymbol{\xi} \sim N_n(\mathbf{0}, \boldsymbol{\Sigma})$, em que

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} & \cdots & \sigma_{1n} \\ \sigma_{12} & \sigma_2^2 & \sigma_{23} & \cdots & \sigma_{2n} \\ \sigma_{13} & \sigma_{23} & \sigma_3^2 & \cdots & \sigma_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sigma_{1n}^2 & \sigma_{2n}^2 & \sigma_{3n}^2 & \cdots & \sigma_n^2 \end{bmatrix}$$

Tabela de contingência $r \times s$: multinomial

		Variável 1 (resposta)					Total
		C_{11}	C_{12}	...	$C_{1(s-1)}$	C_{1s}	
Variável 2 (resposta)	C_{21}	$N_{11}(\theta_{11})$	$N_{12}(\theta_{12})$...	$N_{1(s-1)}(\theta_{1(s-1)})$	$N_{1s}(\theta_{1s})$	$N_{1.}$
	C_{22}	$N_{21}(\theta_{21})$	$N_{22}(\theta_{22})$...	$N_{2(s-1)}(\theta_{2(s-1)})$	$N_{2s}(\theta_{2s})$	$N_{2.}$
	\vdots	\vdots	\vdots	\ddots	\vdots	\vdots	
	C_{2r}	$N_{r1}(\theta_{r1})$	$N_{r2}(\theta_{r2})$...	$N_{r(s-1)}(\theta_{r(s-1)})$	$N_{rs}(\theta_{rs})$	$N_{r.}$
Total	-	$N_{.1}$	$N_{.2}$...	$N_{.(s-1)}$	$N_{.s}$	$n_{..}$

$$\theta = (\theta_{11}, \theta_{12}, \dots, \theta_{1s}, \dots, \theta_{rs-1}).$$

Tabela de contingência $r \times s$: multinomial

- Neste caso, $\mathbf{N} = (N_{11}, N_{21}, \dots, N_{1s}, \dots, N_{rs-1})' \sim \text{multinomial}(n_{..}, \boldsymbol{\theta})$,

e

$$\text{Cov}(\mathbf{N}) = \begin{bmatrix} n_{..}\theta_{11}(1 - \theta_{11}) & -n_{..}\theta_{11}\theta_{12} & -n_{..}\theta_{11}\theta_{13} & \dots & -n_{..}\theta_{11}\theta_{rs-1} \\ -n_{..}\theta_{11}\theta_{12} & n_{..}\theta_{12}(1 - \theta_{12}) & -n_{..}\theta_{12}\theta_{13} & \dots & -n_{..}\theta_{12}\theta_{rs-1} \\ -n_{..}\theta_{11}\theta_{13} & -n_{..}\theta_{12}\theta_{13} & n_{..}\theta_{13}(1 - \theta_{13}) & \dots & -n_{..}\theta_{13}\theta_{rs-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -n_{..}\theta_{11}\theta_{rs-1} & -n_{..}\theta_{12}\theta_{rs-1} & -n_{..}\theta_{13}\theta_{rs-1} & \dots & n_{..}\theta_{rs-1}(1 - \theta_{rs-1}) \end{bmatrix}$$

Ausência de homocedasticidade e independência (testes de hipótese)

- O desenvolvimento é muito similar à situação relativa à ausência de homocedasticidade (heterocedasticidade).
- Neste caso, é mais apropriado considerar o estimador de mínimos quadrados generalizados (MQG).
- Objetivo: obter β que minimiza $Q(\beta) = (\mathbf{Y} - \mathbf{X}\beta)' \Sigma^{-1} (\mathbf{Y} - \mathbf{X}\beta)$.
Em geral, $\beta \in \mathcal{R}^p$.

Ausência de homocedasticidade e independência (testes de hipótese)

- Considere Σ conhecido.
- Para testar as hipóteses anteriores, podemos considerar semelhante construção relacionada à estatística Q_G .
- Nesse caso, temos que

$$\hat{\theta}_G = \mathbf{C}\hat{\beta}_G - \mathbf{M} \sim N_r \left(\mathbf{C}\beta - \mathbf{M}, \mathbf{C} (\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1} \mathbf{C}' \right).$$

Ausência de homocedasticidade e independência (testes de hipótese)

- Defina

$$\begin{aligned} Q_G &= \hat{\theta}'_G [\mathbf{C}(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{C}']^{-1} \hat{\theta}_G \\ &= (\mathbf{C}\hat{\beta}_G - \mathbf{M})' [\mathbf{C}(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{C}']^{-1} (\mathbf{C}\hat{\beta}_G - \mathbf{M}) \end{aligned}$$

- Assim, sob H_0 , $Q_G \sim \chi_r^2$.

Ausência de homocedasticidade e independência (testes de hipótese)

- Assuma Σ desconhecido.
- Para testar as hipóteses anteriores, também podemos considerar semelhante construção relacionada à estatística Q_G .
- Seja $\hat{\Sigma}$ um estimador consistente para Σ , ou seja $\hat{\Sigma} \xrightarrow[n \rightarrow \infty]{P} \Sigma$.

Ausência de homocedasticidade e independência (testes de hipótese)

- Defina $\hat{\beta}_G^* = \left(\mathbf{X}' \hat{\Sigma}^{-1} \mathbf{X} \right)^{-1} \mathbf{X}' \hat{\Sigma}^{-1} \mathbf{Y}$. Assim,

$$\hat{\beta}_G^* \xrightarrow[n \rightarrow \infty]{D} N_p \left(\beta, \left(\mathbf{X}' \Sigma^{-1} \mathbf{X} \right)^{-1} \right).$$

- Defina também, $\hat{\theta}_G^* = \mathbf{C} \hat{\beta}_G^* - \mathbf{M}$. Logo,

$$\hat{\theta}_G^* \xrightarrow[n \rightarrow \infty]{D} N_r \left(\mathbf{C} \beta - \mathbf{M}, \mathbf{C} \left(\mathbf{X}' \Sigma^{-1} \mathbf{X} \right)^{-1} \mathbf{C}' \right);$$

Ausência de homocedasticidade e independência (testes de hipótese)

- Considere

$$\begin{aligned} Q_G^* &= \hat{\theta}_G^{*'} \left[\mathbf{C}(\mathbf{X}'\hat{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{C}' \right]^{-1} \hat{\theta}_G^* \\ &= \left(\mathbf{C}\hat{\beta}_G^* - \mathbf{M} \right)' \left[\mathbf{C}(\mathbf{X}'\hat{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{C}' \right]^{-1} \left(\mathbf{C}\hat{\beta}_G^* - \mathbf{M} \right) \end{aligned}$$

- Sob H_0 , para n suficientemente grande, $Q_G^* \approx \chi_r^2$.

Resumo

- Relembrando que estamos considerando a estrutura $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\xi}$ e nosso interesse consiste em testar:

$$H_0 : \mathbf{C}_{(r \times p)}\boldsymbol{\beta}_{(p \times 1)} = \mathbf{M}_{(r \times 1)} \text{ vs } H_1 : \mathbf{C}_{(r \times p)}\boldsymbol{\beta}_{(p \times 1)} \neq \mathbf{M}_{(r \times 1)}$$

- Sob normalidade, homocedasticidade e independência: pode-se considerar o estimador de mínimos quadrados ordinários e utilizar a estatística Q_0 a qual, sob H_0 , tem **distribuição exata** $F_{[r, n-p]}$.
- Sob **não-normalidade**, homocedasticidade e independência: pode-se considerar o estimador de mínimos quadrados ordinários e utilizar a estatística Q a qual, sob H_0 , tem distribuição aproximada χ_r^2 , para n suficientemente grande.

- Sob normalidade, **heterocedasticidade** e independência:
 - Σ conhecido: pode-se considerar o estimador de mínimos quadrados ponderados ($\hat{\beta}_W$) e utilizar a estatística Q_W a qual, sob H_0 , tem distribuição exata χ_r^2 .
 - Σ desconhecido: pode-se considerar o estimador de mínimos quadrados ponderados ($\hat{\beta}_W^*$) e utilizar a estatística Q_W^* a qual, sob H_0 , tem distribuição aproximada χ_r^2 , para n suficientemente grande.

- Sob normalidade, heterocedasticidade e **dependência**:
 - Σ conhecido: pode-se considerar o estimador de mínimos quadrados generalizados ($\hat{\beta}_G$) e utilizar a estatística Q_G a qual, sob H_0 , tem distribuição exata χ_r^2 .
 - Σ desconhecido: pode-se considerar o estimador de mínimos quadrados generalizados ($\hat{\beta}_G^*$) e utilizar a estatística Q_G^* a qual, sob H_0 , tem distribuição aproximada χ_r^2 , para n suficientemente grande.

Resumo

- Na segunda parte do curso veremos outras estruturas de regressão em situações nas quais mais de uma das suposições anteriores não se verifica.
- Observação: nos resultados relativos à convergência em distribuição dos estimadores $\hat{\beta}_W^*$, $\hat{\theta}_W^*$, $\hat{\beta}_G^*$, $\hat{\beta}_G^*$ cometeu-se um abuso de notação, já que a matriz de planejamento depende do tamanho da amostra n .
- Uma alternativa (mais apropriada) é dizer que $\hat{\beta}_W^* \approx N_p \left(\beta, (\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1} \right)$, para n suficientemente grande (valendo o mesmo para os outros três estimadores).

Voltando ao exemplo 8

- Vamos utilizar a estatística Q_0 .
- Teste para igualdade simultânea dos interceptos (entre si) e dos coeficientes angulares (entre si): $q_0 = 2,17$, $p - valor = 0,0509$.
- Teste para igualdade simultânea dos interceptos : $q_0 = 0,22$, $p - valor = 0,8842$.
- Teste para igualdade simultânea dos coeficientes angulares (entre si): $q_0 = 1,72$, $p - valor = 0,1666$.
- Conclusão: o modelo 1, que considera interceptos iguais entre si e coeficientes angulares iguais entre si é mais adequado.

Voltando ao exemplo 9

- Vamos utilizar a estatística Q_0 .
- Teste para igualdade entre a média das médias dos três primeiros grupos com a média do quinto grupo: $q_0 = 87,37$,
 $p - \text{valor} < 0,0001$.