Modelos de Regressão Múltipla - Parte V

Erica Castilho Rodrigues

3 de Fevereiro de 2017

Mínimos Quadrados Generalizados

Mínimos Quadrados Generalizados

- O modelo de regressão linear supõe que:

- O modelo de regressão linear supõe que:
  - os erros são não correlacionados

- O modelo de regressão linear supõe que:
  - os erros são não correlacionados

$$Cov(\epsilon_i, \epsilon_j) = 0, \ \forall i \neq j$$

- O modelo de regressão linear supõe que:
  - os erros são não correlacionados

$$Cov(\epsilon_i, \epsilon_j) = 0, \ \forall i \neq j$$

  - e possuem a mesma variância

$$Var(\epsilon_i) =$$

- O modelo de regressão linear supõe que:
  - os erros são não correlacionados

$$Cov(\epsilon_i, \epsilon_j) = 0, \ \forall i \neq j$$

  - e possuem a mesma variância

$$Var(\epsilon_i) = \sigma^2, \ \forall i.$$

- Em notação matricial temos que

$$Var(\epsilon) =$$

- O modelo de regressão linear supõe que:
  - os erros são não correlacionados

$$Cov(\epsilon_i, \epsilon_j) = 0, \; \forall i \neq j$$

  - e possuem a mesma variância

$$Var(\epsilon_i) = \sigma^2, \; \forall i.$$

- Em notação matricial temos que

$$Var(\boldsymbol{\epsilon}) = \sigma^2 \mathbf{I} = \sigma^2 \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & & & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix}$$

- Pode acontecer de algumas observações serem menos confiáveis que outras.
- Isso significa que as variância de Y_i e ϵ_i não é constante.
- Dessa maneira não é verdade que

$$Var(\boldsymbol{\epsilon}) = \sigma^2 \mathbf{I}$$

e sim que

$$Var(\boldsymbol{\epsilon}) = \begin{bmatrix} \upsilon_1 & 0 & \dots & 0 \\ 0 & \upsilon_2 & \dots & 0 \\ \vdots & & & \vdots \\ 0 & 0 & \dots & \upsilon_n \end{bmatrix} \Rightarrow Var(\mathbf{Y}) =$$

- Pode acontecer de algumas observações serem menos confiáveis que outras.
- Isso significa que as variância de Y_i e ϵ_i não é constante.
- Dessa maneira não é verdade que

$$Var(\boldsymbol{\epsilon}) = \sigma^2 \mathbf{I}$$

e sim que

$$Var(\boldsymbol{\epsilon}) = \begin{bmatrix} \upsilon_1 & 0 & \dots & 0 \\ 0 & \upsilon_2 & \dots & 0 \\ \vdots & & & \vdots \\ 0 & 0 & \dots & \upsilon_n \end{bmatrix} \Rightarrow Var(\mathbf{Y}) = \begin{bmatrix} \upsilon_1 & 0 & \dots & 0 \\ 0 & \upsilon_2 & \dots & 0 \\ \vdots & & & \vdots \\ 0 & 0 & \dots & \upsilon_n \end{bmatrix}$$

- Pode acontecer também que, mesmo condicionando em X, as variáveis Y_i sejam correlaciondas.
- Por exemplo, se as variáveis Y_i tem estrutura temporal ou espacial.
- Se as variáveis X não refletem essa estrutura,
  - os erros não serão independentes.
- Os elementos fora da diagonal da matriz $Var(\epsilon)$ não serão todos zero.
- Essa matriz fica da forma

$$Var(\epsilon) = \begin{bmatrix} \upsilon_{11} & \upsilon_{12} & \cdots & \upsilon_{1n} \\ \upsilon_{21} & \upsilon_{22} & \cdots & \upsilon_{2n} \\ \vdots & & & \vdots \\ \upsilon_{n1} & \upsilon_{n2} & \cdots & \upsilon_{nn} \end{bmatrix} \Rightarrow Var(\mathbf{Y}) = \begin{bmatrix} \upsilon_{11} & \upsilon_{12} & \cdots & \upsilon_{1n} \\ \upsilon_{21} & \upsilon_{22} & \cdots & \upsilon_{2n} \\ \vdots & & & \vdots \\ \upsilon_{n1} & \upsilon_{n2} & \cdots & \upsilon_{nn} \end{bmatrix}$$

- Em ambos casos o método de Mínimos Quadrados usual

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

  não pode ser aplicado.
- Vamos denotar o método usual por **Mínimos Quadrados Ordinários** (MQO).
- Para o caso de variância não constante aplicaremos o método:
  - **Mínimos Quadrados Ponderados** (MQP) .
- Para o caso de correlação diferente de zero entre os erros e os Y usaremos:
  - **Mínimos Quadrados Generalizados** (MQG) .

- A idéia básica é transformar o vetor de variáveis **Y** em um novo vetor **Z**.
- Esse novo vetor deve satisfazer as suposições do modelo

$$Var(\mathbf{Z}) =$$

- A idéia básica é transformar o vetor de variáveis **Y** em um novo vetor **Z**.
- Esse novo vetor deve satisfazer as suposições do modelo

$$Var(\mathbf{Z}) = \sigma^2\mathbf{I}\,.$$

- Esse vetor será definido da seguinte maneira

$$\mathbf{Z} = \mathbf{QX} + \mathbf{f}$$

com **f** satisfazendo as seguintes suposições

$$E(\mathbf{f}) = \mathbf{0} \quad Var(\mathbf{f}) = \sigma^2\mathbf{I}\,.$$

- Para que os testes e intervalos sejam válidos precisamos ainda que

$$\mathbf{f} \sim N(\mathbf{0}, \sigma^2\mathbf{I})\,.$$

- Queremos ajustar o seguinte modelo

$$Y = \mathbf{X}\beta + \epsilon$$

onde

$$E(\epsilon) = 0 \quad Var(\epsilon) = \sigma^2\mathbf{V} \quad \epsilon \sim N(\mathbf{0}, \sigma^2\mathbf{V})$$

e

$$\mathbf{V} = \begin{bmatrix} v_{11} & v_{12} & \cdots & v_{1n} \\ v_{21} & v_{22} & \cdots & v_{2n} \\ \vdots & & & \vdots \\ v_{n1} & v_{n2} & \cdots & v_{nn} \end{bmatrix}$$

por ser matriz de covariância, é simétrica e positiva definida.

- Como $\mathbf{V}$ é positiva definida, existe uma matriz $\mathbf{P}$ simétrica, não singular, tal que
$$\mathbf{V} = \mathbf{P}'\mathbf{P} = \mathbf{P}\mathbf{P} = \mathbf{P}^2 \, .$$
- Então para
$$\mathbf{Z} = \mathbf{Q}\mathbf{X} + \mathbf{f}$$
vamos definir
$$\mathbf{f} = \mathbf{P}^{-1}\epsilon$$
e veremos que as suposições do modelo serão válidas.
- Temos que
$$E(\mathbf{f}) =$$

- Como $\mathbf{V}$ é positiva definida, existe uma matriz $\mathbf{P}$ simétrica, não singular, tal que

$$\mathbf{V} = \mathbf{P}'\mathbf{P} = \mathbf{P}\mathbf{P} = \mathbf{P}^2 \,.$$

- Então para

$$\mathbf{Z} = \mathbf{Q}\mathbf{X} + \mathbf{f}$$

vamos definir

$$\mathbf{f} = \mathbf{P}^{-1}\epsilon$$

e veremos que as suposições do modelo serão válidas.
- Temos que

$$E(\mathbf{f}) = E(\mathbf{P}^{-1}\epsilon) =$$

- Como **V** é positiva definida, existe uma matriz **P** simétrica, não singular, tal que

$$\mathbf{V} = \mathbf{P}'\mathbf{P} = \mathbf{P}\mathbf{P} = \mathbf{P}^2 \,.$$

- Então para

$$\mathbf{Z} = \mathbf{Q}\mathbf{X} + \mathbf{f}$$

  vamos definir

$$\mathbf{f} = \mathbf{P}^{-1}\epsilon$$

  e veremos que as suposições do modelo serão válidas.
- Temos que

$$E(\mathbf{f}) = E(\mathbf{P}^{-1}\epsilon) = \mathbf{P}^{-1}E(\epsilon) =$$

- Como $\mathbf{V}$ é positiva definida, existe uma matriz $\mathbf{P}$ simétrica, não singular, tal que

$$\mathbf{V} = \mathbf{P}'\mathbf{P} = \mathbf{P}\mathbf{P} = \mathbf{P}^2 \,.$$

- Então para

$$\mathbf{Z} = \mathbf{Q}\mathbf{X} + \mathbf{f}$$

  vamos definir

$$\mathbf{f} = \mathbf{P}^{-1}\epsilon$$

  e veremos que as suposições do modelo serão válidas.
- Temos que

$$E(\mathbf{f}) = E(\mathbf{P}^{-1}\epsilon) = \mathbf{P}^{-1}E(\epsilon) = \mathbf{0} \,.$$

- Vamos agora usar um resultado que diz que se **Y** é um vetor aleatório então

$$Var(\mathbf{Y}) = E(\mathbf{Y}\mathbf{Y}') - [E(\mathbf{Y})]'[E(\mathbf{Y})] \,.$$

- Temos então que

$$Var(\mathbf{f}) =$$

- Vamos agora usar um resultado que diz que se **Y** é um vetor aleatório então

$$Var(\mathbf{Y}) = E(\mathbf{Y}\mathbf{Y}') - [E(\mathbf{Y})]'[E(\mathbf{Y})] \, .$$

- Temos então que

$$Var(\mathbf{f}) = E(\mathbf{ff}) - [E(\mathbf{f})]'[E(\mathbf{f})]$$

mas vimos que $E(\mathbf{f}) = \mathbf{0}$, logo

$$Var(\mathbf{f}) = E(\mathbf{ff}') = E([\mathbf{P}^{-1}\epsilon][\mathbf{P}^{-1}\epsilon]') = E(\mathbf{P}^{-1}\epsilon\epsilon'\mathbf{P}^{-1})$$
$$= \mathbf{P}^{-1}E(\epsilon\epsilon')\mathbf{P}^{-1}$$

mas

$$Var(\epsilon) = E([\epsilon][\epsilon]') - [E(\epsilon)][E(\epsilon)]'$$

como $E(\epsilon) = \mathbf{0}$

$$Var(\epsilon) = E([\epsilon][\epsilon]') \, . = \sigma^2\mathbf{V} \, .$$

- Substituindo $E([\epsilon][\epsilon]')$ por $\sigma^2\mathbf{V}$ ficamos com

$$Var(\mathbf{f}) = \mathbf{P}^{-1}\sigma^2\mathbf{V}\mathbf{P}^{-1}$$

mas vimos que $\mathbf{V} = \mathbf{P}\mathbf{P}$, logo

$$Var(\mathbf{f}) = \mathbf{P}^{-1}\sigma^2\mathbf{P}\mathbf{P}\mathbf{P}^{-1} = \sigma^2\mathbf{I}\,.$$

- Mostramos assim que se definirmos

$$\mathbf{Z} = \mathbf{Q}\mathbf{X} + \mathbf{f} \quad \text{e} \quad \mathbf{f} = \mathbf{P}^{-1}\epsilon$$

então

$$E(\mathbf{f}) = \mathbf{0} \quad \text{e} \quad Var(\mathbf{f}) = \sigma^2\mathbf{I}$$

o que implica que

$$E(\mathbf{Z}) =$$

- Substituindo $E([\epsilon][\epsilon]')$ por $\sigma^2\mathbf{V}$ ficamos com

$$Var(\mathbf{f}) = \mathbf{P}^{-1}\sigma^2\mathbf{V}\mathbf{P}^{-1}$$

mas vimos que $\mathbf{V} = \mathbf{PP}$, logo

$$Var(\mathbf{f}) = \mathbf{P}^{-1}\sigma^2\mathbf{PPP}^{-1} = \sigma^2\mathbf{I}\,.$$

- Mostramos assim que se definirmos

$$\mathbf{Z} = \mathbf{QX} + \mathbf{f} \quad \text{e} \quad \mathbf{f} = \mathbf{P}^{-1}\epsilon$$

então

$$E(\mathbf{f}) = \mathbf{0} \quad \text{e} \quad Var(\mathbf{f}) = \sigma^2\mathbf{I}$$

o que implica que

$$E(\mathbf{Z}) = \mathbf{QX} \quad Var(\mathbf{Z}) =$$

- Substituindo $E([\epsilon][\epsilon]')$ por $\sigma^2\mathbf{V}$ ficamos com

$$Var(\mathbf{f}) = \mathbf{P}^{-1}\sigma^2\mathbf{V}\mathbf{P}^{-1}$$

mas vimos que $\mathbf{V} = \mathbf{P}\mathbf{P}$, logo

$$Var(\mathbf{f}) = \mathbf{P}^{-1}\sigma^2\mathbf{P}\mathbf{P}\mathbf{P}^{-1} = \sigma^2\mathbf{I}\,.$$

- Mostramos assim que se definirmos

$$\mathbf{Z} = \mathbf{Q}\mathbf{X} + \mathbf{f} \quad \text{e} \quad \mathbf{f} = \mathbf{P}^{-1}\epsilon$$

então

$$E(\mathbf{f}) = \mathbf{0} \quad \text{e} \quad Var(\mathbf{f}) = \sigma^2\mathbf{I}$$

o que implica que

$$E(\mathbf{Z}) = \mathbf{Q}\mathbf{X} \quad Var(\mathbf{Z}) = \sigma^2\mathbf{I}$$

ou seja, o vetor **Z** satisfaz as propriedades do modelo usual.

- Além disso, como
$$\epsilon \sim N(\mathbf{0}, \sigma^2 \mathbf{V})$$
e **f** é uma cobinação linear do vetor ϵ então
$$\mathbf{f} \sim N(\mathbf{0}, \sigma^2 \mathbf{I}) \; .$$
- O nosso modelo original é dado por
$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \epsilon \; .$$
- Para que apareça o termpo $\mathbf{P}^{-1}\epsilon$ vamos multiplicar toda equação por $\mathbf{P}^{-1}$
$$\mathbf{P}^{-1}\mathbf{Y} = \mathbf{P}^{-1}\mathbf{X}\boldsymbol{\beta} + \mathbf{P}^{-1}\epsilon \; .$$
- Temos que
$$\mathbf{P}^{-1}\epsilon = \mathbf{f} \; .$$

- Vamos denotar

$$\mathbf{P}^{-1}\mathbf{Y} = \mathbf{Z} \quad \mathbf{P}^{-1}\mathbf{X} = \mathbf{Q}\,.$$

- E assim

$$\mathbf{P}^{-1}\mathbf{Y} = \mathbf{P}^{-1}\mathbf{X}\beta + \mathbf{P}^{-1}\epsilon \quad \Leftrightarrow \quad \mathbf{Z} = \mathbf{Q}\beta + \mathbf{f}\,.$$

- Podemos agora para o novo modelo, em função de **Z**, aplicar o método de mínimos quadrados usual.
- Podemos fazer isso, pois esse modelo satisfaz as suposições de homocedasticidade e independência.

- O erro é dado por
$$\mathbf{f} = \mathbf{Z} - \mathbf{Q}\beta \, .$$
- Portanto, estamos interessados em minimizar
$$\mathbf{f}'\mathbf{f}$$
como $\mathbf{f} = \mathbf{P}^{-1}\epsilon$, ficamos com
$$\mathbf{f}'\mathbf{f} = [\mathbf{P}^{-1}\epsilon]'[\mathbf{P}^{-1}\epsilon] = \epsilon'[\mathbf{P}^{-1}]'\mathbf{P}^{-1}\epsilon$$
como $\mathbf{P}$ é simétrica, $\mathbf{P}^{-1}$ também é, logo
$$[\mathbf{P}^{-1}]' = [\mathbf{P}^{-1}]$$
e ficamos com
$$\mathbf{f}'\mathbf{f} = \epsilon'[\mathbf{P}^{-1}][\mathbf{P}^{-1}]\epsilon \, .$$

- Já vimos que $\mathbf{PP} = \mathbf{V}$ e

$$\mathbf{V}^{-'} = [\mathbf{PP}]^{-1} = \mathbf{P}^{-1}\mathbf{P}^{-1}$$

substituindo ficamos com

$$\mathbf{f}'\mathbf{f} = \epsilon'[\mathbf{P}^{-1}]'\mathbf{P}^{-1}\epsilon = \epsilon'\mathbf{V}^{-1}\epsilon$$

- Mas sabemos que $\epsilon = \mathbf{Y} - \mathbf{X}\beta$ logo

$$\mathbf{f}'\mathbf{f} = (\mathbf{Y} - \mathbf{X}\beta)'\mathbf{V}^{-1}(\mathbf{Y} - \mathbf{X}\beta)$$

$$= (\mathbf{Y}' - \beta'\mathbf{X}')(\mathbf{V}^{-1}\mathbf{Y} - \mathbf{V}^{-1}\mathbf{X}\beta)$$

$$= \mathbf{Y}'\mathbf{V}^{-1}\mathbf{Y} - \mathbf{Y}'\mathbf{V}^{-1}\mathbf{X}\beta - \beta'\mathbf{X}'\mathbf{V}^{-1}\mathbf{Y} + \beta'\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\beta$$

mas, como $\mathbf{Y}'\mathbf{V}^{-1}\mathbf{X}\beta$ é um escalar,

$$[\mathbf{Y}'\mathbf{V}^{-1}\mathbf{X}\beta]' = \beta'\mathbf{X}'\mathbf{V}^{-1}\mathbf{Y}\,.$$

- Ficamos então com

$$\mathbf{f}'\mathbf{f} = \mathbf{Y}'\mathbf{V}^{-1}\mathbf{Y} - 2\beta'\mathbf{X}'\mathbf{V}^{-1}\mathbf{Y} + \beta'\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\beta\,.$$

- Derivando e relação a β e igualando a zero

$$-2\mathbf{X}'\mathbf{V}^{-1}\mathbf{Y} + 2\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\beta = 0\,.$$

- Isso implica que

$$\mathbf{X}'\mathbf{V}^{-1}\mathbf{Y} = \mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\hat{\beta}$$

e esse é o nosso novo conjunto de equações normais.

- Multiplicando por $[\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}]^{-1}$ dos dois lados ficamos com

$$[\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}]^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\hat{\beta} = [\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}]^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{Y}$$

- O estimador de β é dado por

$$\hat{\beta} = [\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}]^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{Y}\ .$$

- Comparando esse estimador como o estimador de Mínimos Quadrados Ordinários

$$\hat{\beta} = [\mathbf{X}'\mathbf{X}]^{-1}\mathbf{X}'\mathbf{Y}$$

nota-se que a única diferença é a inclusão da matriz $\mathbf{V}^{-1}$.

- A matriz de covariâncias desse estimador é dada por

$$Var(\hat{\beta}) =$$

- A matriz de covariâncias desse estimador é dada por

$$Var(\hat{\beta}) = Var([\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}]^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{Y})$$

$$Var(\mathbf{AY}) = \mathbf{A}\,Var(\mathbf{Y})\mathbf{A}'$$
$$= [\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}]^{-1}\mathbf{X}'\mathbf{V}^{-1}\,Var(\mathbf{Y})\mathbf{V}^{-1}\mathbf{X}[\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}]^{-1}$$

pois $[\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}]^{-1}$ é simétrica.
- Substituindo $Var(\mathbf{Y}) = \sigma^2\mathbf{V}$ ficamos com

$$= [\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}]^{-1}\mathbf{X}'\mathbf{V}^{-1}\sigma^2\mathbf{V}\mathbf{V}^{-1}\mathbf{X}[\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}]^{-1}$$
$$= \sigma^2[\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}]^{-1}$$

ou seja

$$Var(\hat{\beta}) = \sigma^2[\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}]^{-1}$$

antes tínhamos

$$Var(\hat{\beta}) = \sigma^2[\mathbf{X}'\mathbf{X}]^{-1}$$

- A Soma de Quadrados da Regressão é dada por

$$\hat{\beta}'\mathbf{Q}'\mathbf{Z}$$

- A Soma de Quadrados da Regressão é dada por

$$\begin{aligned}\hat{\beta}'\mathbf{Q}'\mathbf{Z} \\ &= ([\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}]^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{Y})'(\mathbf{P}^{-1}\mathbf{X})'\mathbf{P}^{-1}\mathbf{Y} \\ &= \mathbf{Y}'\mathbf{V}^{-1}\mathbf{X}[\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}]^{-1}\mathbf{X}'\mathbf{P}^{-1}\mathbf{P}^{-1}\mathbf{Y}\end{aligned}$$

  pois $\mathbf{V}^{-1}$ e $[\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}]^{-1}$ são simétricas.
- Como $\mathbf{P}^{-1}\mathbf{P}^{-1} = \mathbf{V}^{-1}$ ficamos com

$$\mathbf{Y}'\mathbf{V}^{-1}\mathbf{X}[\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}]^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{Y}$$

- A Soma de Quadrados Totais fica

$$\mathbf{Z}'\mathbf{Z} =$$

- A Soma de Quadrados Totais fica

$$\mathbf{Z}'\mathbf{Z} = (\mathbf{P}^{-1}\mathbf{Y})'\mathbf{P}^{-1}\mathbf{Y} =$$

$$\mathbf{Y}'\mathbf{P}^{-1}\mathbf{P}^{-1}\mathbf{Y} = \mathbf{Y}'\mathbf{V}^{-1}\mathbf{Y}$$

  pois $\mathbf{P}^{-1}$ é simétrica.

- A Soma de Quadrados dos Resíduos fica então

$$\mathbf{Y}'\mathbf{V}^{-1}\mathbf{Y} - \mathbf{Y}'\mathbf{V}^{-1}\mathbf{X}[\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}]^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{Y}$$

- Os resíduos a serem analisados são as estimativas de

$$\mathbf{f} = \mathbf{P}^{-1}\epsilon$$

e são dados por

$$\mathbf{f} = \mathbf{Z} - \hat{\mathbf{Z}} = \mathbf{P}^{-1}(\mathbf{Y} - \hat{\mathbf{Y}})$$

Observações

- Nos problemas práticos, é difícil saber a forma da matriz **V**.
- Será necessário, as vezes, primeiro fazer a suposição de que

$$\mathbf{V} \propto I$$

  mesmo sabendo que é errada.
- Em seguida fazer a inferência via mínimos quadrados ordinários.
- E então, examinar os resíduos para se descobrir algo sobre a forma de **V**.

Observações

- Considere um caso em que é necessário usar Mínimos Quadrados Generalizados.
- Suponha que o método dos Mínimos Quadrados Ordinários é usado erroneamente.
- A estimativa de β ainda sim será não- viciada

$$E(\hat{\beta}) = \beta\,.$$

- Porém não terá variância mínima.
- Usar Mínimos Quadrados Ordinários no lugar de Mínimos Quadrados Generalizados produz estimadores com maior variabilidade.

- Vejamos alguns exemplos de forma da matriz **V**.
- Considere o caso em que temos apenas uma parâmetro

$$Y_i = \beta_1 X_i + \epsilon_i \quad \epsilon_i \sim N(0, 1/w_i) \; .$$

- Nesse caso temos que

$$Var(\mathbf{Y}) =$$

- Vejamos alguns exemplos de forma da matriz **V**.
- Considere o caso em que temos apenas uma parâmetro

$$Y_i = \beta_1 X_i + \epsilon_i \quad \epsilon_i \sim N(0, 1/w_i) \; .$$

- Nesse caso temos que

$$Var(\mathbf{Y}) = \sigma^2 \mathbf{V} = \sigma^2 \begin{bmatrix} 1/w_1 & 0 & \dots & 0 \\ 0 & 1/w_2 & \dots & 0 \\ 0 & 0 & \dots & 1/w_n \end{bmatrix}$$

onde os w's são pesos a serem especificados.

- Temos então nesse caso que

$$\mathbf{V}^{-1} =$$

- Temos então nesse caso que

$$\mathbf{V}^{-1} = \begin{bmatrix} w_1 & 0 & \dots & 0 \\ 0 & w_2 & \dots & 0 \\ 0 & 0 & \dots & w_n \end{bmatrix}$$

- Vimos que o estimador de β é dado por

$$\hat{\beta} = [\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}]^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{Y}\,.$$

- Vejamos como fica esse estimador para esse caso específico.

- Temos que

$$\mathbf{X}'\mathbf{V}^{-1} = \begin{bmatrix} X_{11} & X_{21} & \dots & X_{n1} \end{bmatrix} \sigma^2 \begin{bmatrix} w_1 & 0 & \dots & 0 \\ 0 & w_2 & \dots & 0 \\ 0 & 0 & \dots & w_n \end{bmatrix}$$

$$= \sigma^2 \begin{bmatrix} X_{11} w_1 & X_{21} w_2 & \dots & X_{n1} w_n \end{bmatrix}$$

- Logo

$$\mathbf{X}'\mathbf{V}^{-1}\mathbf{X} = \sigma^2 \begin{bmatrix} X_{11} w_1 & X_{21} w_2 & \dots & X_{n1} w_n \end{bmatrix} \begin{bmatrix} X_{11} \\ X_{21} \\ \vdots \\ X_{n1} \end{bmatrix} = \sigma^2 \sum_{i=1}^{n} w_i X_{i1}^2$$

- Logo

$$[\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}]^{-1} = \frac{1}{\sigma^2 \sum_{i=1}^{n} w_i X_{i1}^2}$$

- Além disso

$$\mathbf{X}'\mathbf{V}^{-1}\mathbf{Y} = \sigma^2 \begin{bmatrix} X_{11}w_1 & X_{21}w_2 & \dots & X_{n1}w_n \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_1 \end{bmatrix} = \sum_{i=1}^{n} w_i X_{i1} Y_i$$

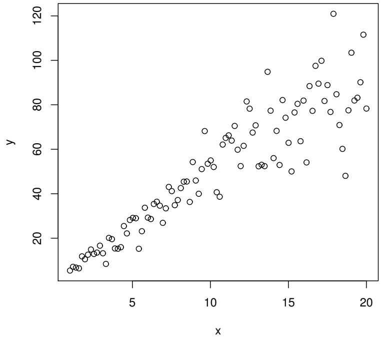
- Portanto

$$\hat{\beta} = [\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}]^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{Y} = \frac{\sum_{i=1}^{n} w_i X_{i1} Y_i}{\sum_{i=1}^{n} w_i X_{i1}^2}.$$

- Vejamos agora alguns exemplos de pesos possíveis.

Exemplo

- Vamos considerar o caso em que a variância é proporcional a X_i.
- Quano maior o valor da variável explicativa, maior a variabilidade.



Exemplo (continuação)

- Temos então que

$$w_i = \frac{1}{kX_i} \Rightarrow \hat{\beta}_1 = \frac{\sum_{i=1}^n w_i X_{i1} Y_i}{\sum_{i=1}^n w_i X_{i1}^2} =$$

Exemplo (continuação)

- Temos então que

$$w_i = \frac{1}{kX_i} \Rightarrow \hat{\beta}_1 = \frac{\sum_{i=1}^n w_i X_{i1} Y_i}{\sum_{i=1}^n w_i X_{i1}^2} = \frac{\sum_{i=1}^n X_{i1} Y_i/(kX_i)}{\sum_{i=1}^n X_{i1}^2/(kX_i)}$$

$$= \frac{\sum_{i=1}^n Y_i}{\sum_{i=1}^n X_{i1}} =$$

Exemplo (continuação)

- Temos então que

$$w_i = \frac{1}{kX_i} \Rightarrow \hat{\beta}_1 = \frac{\sum_{i=1}^n w_i X_{i1} Y_i}{\sum_{i=1}^n w_i X_{i1}^2} = \frac{\sum_{i=1}^n X_{i1} Y_i/(kX_i)}{\sum_{i=1}^n X_{i1}^2/(kX_i)}$$

$$= \frac{\sum_{i=1}^n Y_i}{\sum_{i=1}^n X_{i1}} = \ = \frac{\sum_{i=1}^n Y_i/n}{\sum_{i=1}^n X_{i1}/n} = \frac{\overline{Y}}{\overline{X}}.$$

- Além disso, sabemos que

$$Var(\hat{\beta}) = \sigma^2[\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}]^{-1}$$

mas vimos que

$$[\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}]^{-1} =$$

Exemplo (continuação)

- Temos então que

$$w_i = \frac{1}{kX_i} \Rightarrow \hat{\beta}_1 = \frac{\sum_{i=1}^n w_i X_{i1} Y_i}{\sum_{i=1}^n w_i X_{i1}^2} = \frac{\sum_{i=1}^n X_{i1} Y_i/(kX_i)}{\sum_{i=1}^n X_{i1}^2/(kX_i)}$$

$$= \frac{\sum_{i=1}^n Y_i}{\sum_{i=1}^n X_{i1}} = = \frac{\sum_{i=1}^n Y_i/n}{\sum_{i=1}^n X_{i1}/n} = \frac{\overline{Y}}{\overline{X}} .$$

- Além disso, sabemos que

$$Var(\hat{\beta}) = \sigma^2[\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}]^{-1}$$

mas vimos que

$$[\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}]^{-1} = \sum_{i=1}^n w_i X_{i1}^2 .$$

Exemplo (continuação)

- Portanto

$$Var(\hat{\beta}_1) = \frac{\sigma^2}{\sum_{i=1}^{n} w_i X_{i1}^2} = \frac{\sigma^2}{\sum_{i=1}^{n} X_{i1}^2/(kX_{i1})} =$$

Exemplo (continuação)

- Portanto

$$Var(\hat{\beta}_1) = \frac{\sigma^2}{\sum_{i=1}^{n} w_i X_{i1}^2} = \frac{\sigma^2}{\sum_{i=1}^{n} X_{i1}^2/(kX_{i1})} = \sigma^2 \frac{k}{\sum_{i=1}^{n} X_{i1}}$$

Exemplo

- Vejamos agora um exemplo numérico.
- Um modelo linear simples é ajustado usando Mínimos Quadrados Ordinários.
- O gráfico a seguir mostra os resíduos em função dos valores ajustados e variável explicativa.

Exemplo (continuação)

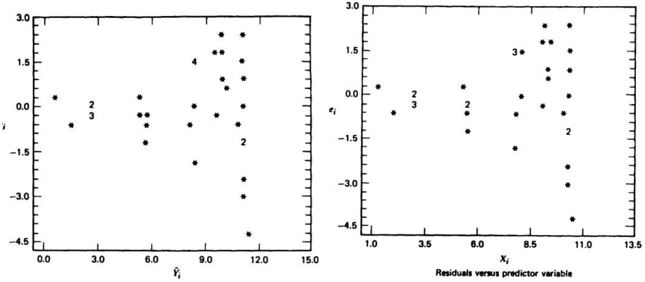


Figura: Gráfico de Resíduos.

Exemplo (continuação)

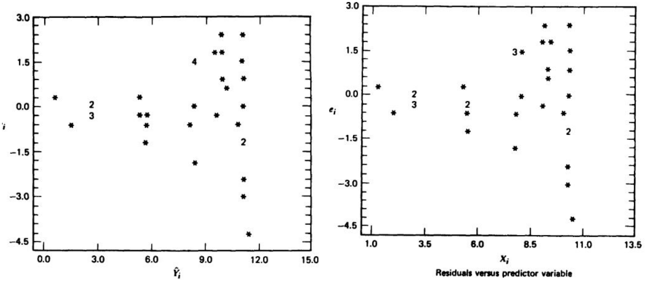


Figura: Gráfico de Resíduos.

- Notamos claramente que a variância não é constante.

Exemplo (continuação)

- Esses gráficos mostram que usar Mínimos Quadrados Ordinários não é o mais adequado nesse caso.
- Devemos usar mínimos quadrados generalizados.
- Vamos assumir que os Y_i são independentes, ou seja, a matriz de covariâncias é diagonal.
- Iremos permitir, porém, que as variâncias mudem de uma observação para a outra.
- A base de dados possui alguns valores repetidos de X ou muito próximos.
- Esses valores serão usados para estimar os pesos w's.

Exemplo (continuação)

- Foram calculadas as médias e o quadrado médio do erro puro para cada conjunto de valores repetido de X

$\overline{X}_j$	3.0	5.4	7.8	9.1	10.2
s^2_{ej}	0.0072	0.3440	1.7404	0.8683	3.8964

- Como alguns valores de X foram arredondados, precisamos estimar s^2_{ej} para esses valores.
- Ajustamos um polinômio de segundo grau

$$\hat{s}^2_{ej} = 1.5329 - 0.7334\overline{X} + .0883\overline{X}^2 .$$

Exemplo (continuação)

- Substituindo os valores de X_i na equação, obtemos os valores de $\hat{s}^2_{ej}$.
- Os pesos são dados por

$$w_i = \frac{1}{\hat{s}^2_{ej}} \,.$$

- A Figura a seguir mostra os gráficos de resíduos obtidos com esse novo modelo ajustado.

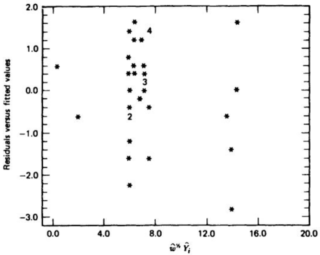


Figura: Gráfico de Resíduos usando Mínimos Quadrados Generalizados.

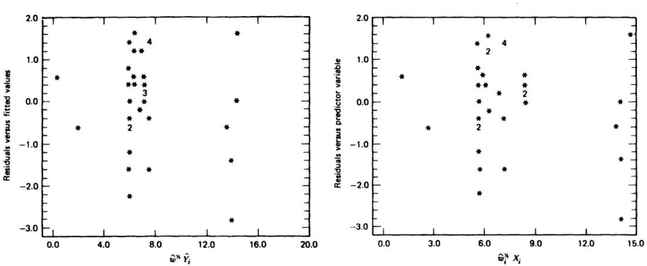


Figura: Gráfico de Resíduos usando Mínimos Quadrados Generalizados.

- Notamos claramente que o problema de heterocedasticidade foi resolvido.

- Veremos como usar esse método no R.

```
lm.gls(formula, data, W, subset, na.action,
 inverse = FALSE,
       method = "qr", model = FALSE,
 x = FALSE, y = FALSE,
       contrasts = NULL, ...)
```