

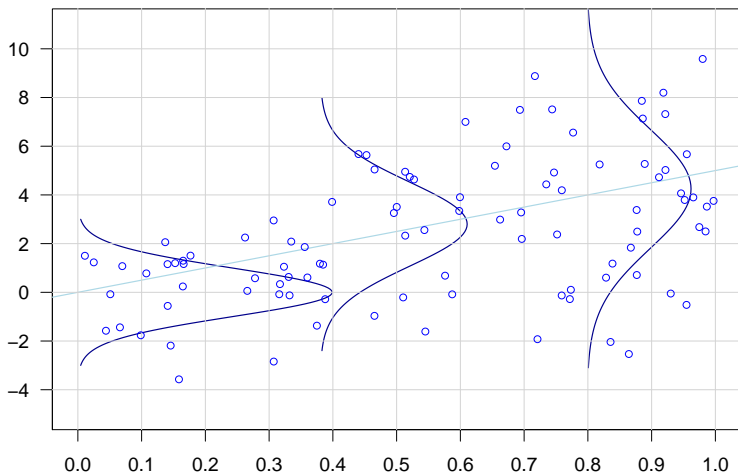
# CE225 - Modelos Lineares Generalizados

Cesar Augusto Taconeli

07 de agosto, 2017

## Aula 3 - Modelo linear com erros heterocedásticos

# Modelo linear com erros heterocedásticos



**Figura 1:** Regressão com erros normais - I

# Modelo linear com erros heterocedásticos

- Uma das pressuposições dos modelos lineares é que  $Var(y_i|\mathbf{x}_i) = \sigma^2$ , ou seja, variância constante quaisquer que sejam os valores das covariáveis.
- Em algumas situações em que essa pressuposição não é verificada, pode ser razoável admitir que as variâncias sejam proporcionais, de forma que:

$$Var(y_i|\mathbf{x}_i) = \sigma_i^2 = \frac{\sigma^2}{\omega_i}, \quad (1)$$

sendo  $\omega_i$ ,  $i = 1, 2, \dots, n$  as constantes de proporcionalidade, particulares a cada observação.

# Método de mínimos quadrados ponderados

- O estimador de mínimos quadrados para os parâmetros de um modelo de regressão linear:

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}, \quad (2)$$

na situação em que os erros são heterocedásticos já não apresenta variância mínima (ineficiente).

- Como alternativa, podemos incorporar pesos no processo de ajuste do modelo, atribuindo maior peso a observações sujeitas a menor variância.

# Método de mínimos quadrados ponderados

- O método de mínimos quadrados ponderados baseia-se na minimização da seguinte soma de quadrados:

$$SQE = \sum_{i=1}^n \omega_i (y_i - \beta_0 - \sum_j \beta_j x_{ij})^2 = (\mathbf{y} - \beta \mathbf{X})' \mathbf{W} (\mathbf{y} - \beta \mathbf{X}), \quad (3)$$

em que

$$\mathbf{W} = \begin{bmatrix} \omega_1 & 0 & \cdots & 0 \\ 0 & \omega_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \omega_n \end{bmatrix}$$

é a matriz diagonal de pesos.

# Método de mínimos quadrados ponderados

- Estimador de mínimos quadrados ponderados:

$$\hat{\beta}_w = (\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}\mathbf{y}. \quad (4)$$

- Estimador de  $\sigma^2$ :

$$\hat{\sigma}^2 = \frac{(\mathbf{y} - \hat{\beta}\mathbf{X})'\mathbf{W}(\mathbf{y} - \hat{\beta}\mathbf{X})}{n - p - 1}. \quad (5)$$

- Matriz de covariâncias de  $\hat{\beta}$ :

$$\text{Var}(\hat{\beta}) = \sigma^2(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}, \quad (6)$$

que é estimada substituindo-se  $\sigma^2$  por  $\hat{\sigma}^2$ .

# Modelo linear com erros heterocedásticos

- Na sequência são apresentadas algumas situações em que há alguma razão para a utilização de mínimos quadrados ponderados, e a forma como o método deveria ser aplicado.
- ① Suponha que as observações sejam, na verdade, médias de amostras de  $n_i$  observações, ou seja:

$$y_i = \bar{u}_i = \sum_{k=1}^{n_i} u_{ik}, \quad i = 1, 2, \dots, n. \quad (7)$$

- Adicionalmente, vamos considerar que as observações individuais ( $u_i$ 's) satisfazem  $\text{Var}(u_i|\mathbf{x}_i) = \sigma^2$ , constante para todo  $\mathbf{x}_i$ .
- Nesse caso:

$$\text{Var}(y_i|\mathbf{x}_i) = \frac{\sigma^2}{n_i}, \quad i = 1, 2, \dots, n. \quad (8)$$

de tal forma que deveríamos adotar  $\omega_i = n_i$ .



# Modelo linear com erros heterocedásticos

- 2 Suponha que o padrão não constante da variância possa ser descrito por alguma função de uma ou mais covariáveis. Como exemplo:

$$\text{Var}(y_i | \mathbf{x}_i) = x_{ij} \sigma^2, \quad (9)$$

ou seja, a variância está linearmente relacionada à variável  $x_j$ .

- Neste caso, os pesos ficam definidos por  $\omega_j = \frac{1}{x_{ij}}$ .
- De maneira semelhante, se tivéssemos  $\text{Var}(y_i | \mathbf{x}_i) = x_{ij}^2 \sigma^2$ , poderíamos definir  $\omega_j = \frac{1}{x_{ij}^2}$ .

# Modelo linear com erros heterocedásticos

- 3 Em muitos estudos as observações estão sujeitas a erros de medida que podem assumir diferentes distribuições para o conjunto de observações.
- Como exemplo, considere um experimento em que cada observação seja medida por um de três equipamentos disponíveis (A, B e C) ;
- Considere que os três equipamentos têm diferentes níveis de precisão, sendo as respectivas variâncias dadas por  $\sigma_A^2$ ,  $\sigma_B^2$  e  $\sigma_C^2$ .
- Neste caso, os pesos poderiam ser determinados pelo inverso das variâncias (estimadas) de cada equipamento.

# Modelo linear com erros heterocedásticos

- **No R:** pesos incorporados por meio do argumento `weights` da função `lm`.
- Para situações mais gerais (envolvendo até mesmo erros correlacionados) pode-se usar a função `gls` do pacote `nlme`.

# Modelo linear com erros heterocedásticos

- Como exemplo de aplicação, vamos considerar a base de dados carsdo 'R'. Os dados correspondem a resultados de testes de frenagem de 50 automóveis. Duas variáveis estão disponíveis:
  - Speed: velocidade do automóvel no momento da frenagem (mph);
  - Dist: distância percorrida do momento da frenagem até o automóvel parar completamente.
- O objetivo é ajustar um modelo para a distância até parada completa em função da velocidade no momento de frenagem.