Modelos Lineares
Regressão Linear Simples (continuação)

Susana Faria

O modelo de regressão simples pode ser utilizado para analisar a relação entre uma variável dependente quantitativa e uma variável independente qualitativa com 2 níveis.

$$Y_i = \beta_0 + \beta_1 x_i + \epsilon_i \quad i = 1, \ldots, n$$

- $E(Y_i|x_i = 0) = \beta_0$ e $V(Y_i|x_i = 0) = \sigma^2$
- $E(Y_i|x_i = 1) = \beta_0 + \beta_1$ e $V(Y_i|x_i = 1) = \sigma^2$

Para as estimativas obtém-se:

$$\hat{\beta}_0 = \overline{y}_{x=0}$$

e

$$\hat{\beta}_1 = \overline{y}_{x=1} - \overline{y}_{x=0}$$

Interpretação de β_1:

Não é difícil mostrar que o teste-t para testar

$$H_0 : \beta_1 = 0$$

vs

$$H_1 : \beta_1 \neq 0$$

é equivalente ao teste-t para comparação das médias de dois grupos (com variâncias iguais)

$$H_0 : \mu_1 = \mu_2$$

vs

$$H_1 : \mu_1 \neq \mu_2$$

Exemplo

Homens:

38.7	35.5	32.4	31.1	32.8	27.4	27.7	34.0	28.2	35.3
33.6	29.8	30.7	30.1	34.4	29.4	32.0	30.2	33.7	32.5
35.8	29.8	34.7	27.4						

Mulheres:

28.8	29.4	28.1	27.9	31.8	30.0	31.3	28.2	28.7	30.7
30.3	29.4	31.1	29.3	27.3	31.8	31.9	28.5	30.6	25.4
29.9	30.6	27.7	34.4						

E se a variável explicativa (chamemos-lhe z) tivesse três níveis? Faria sentido usar, por exemplo, a codificação 0, 1, 2 ?

$$Y_i = \beta_0 + \beta_1 z_i + \epsilon_i, i = 1, \ldots, n$$

- se $z_i = 0$
- se $z_i = 1$
- se $z_i = 2$

É equivalente a dizer que o nível 2 tem um efeito (em termos da diferença para o nível zero) igual a 2 vezes o efeito do nível 1.
É uma restrição muito forte e irrealista na maior parte dos casos!

Solução: Criar duas variáveis binárias fictícias (dummy variables)

Nota: Há muitas codificações possíveis. No *bom software* basta indicar que a variável é qualitativa para que seja usada a codificação correta. No entanto, para interpretar corretamente os resultados, é importante saber concretamente qual a codificação que está a ser usada.