Capítulo 4

- **4. Tópicos sobre formas funcionais**
  - 4.1. Termos quadráticos e interações
  - 4.2. Testes da forma funcional
  - 4.3. Variáveis Dummy
  - 4.4. Exemplos com aplicações empíricas

4.1. Termos quadráticos e interações

- **Termos quadráticos**

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + u$$

- Efeito de uma variação de x

$$\Delta y \approx \left(\frac{dy}{dx}\right)\Delta x \Leftrightarrow \Delta y \approx \Delta x\left(\beta_1 + 2\beta_2 x\right)$$

O efeito depende de x

  - A interpretação de β_1 não é a habitual:

$$x = 0, \Delta x = 1 \Rightarrow \Delta y \approx \beta_1 + 2\beta_2 \times 0 \Leftrightarrow \beta_1 \approx \Delta y \text{ quando } x = 0, \Delta x = 1$$

  - O efeito não é constante e depende de x

$$x = 3, \Delta x = 1 \Rightarrow \Delta y \approx \beta_1 + 2\beta_2 \times 3 = \beta_1 + 6\beta_2$$

$$x = 4, \Delta x = 1 \Rightarrow \Delta y \approx \beta_1 + 2\beta_2 \times 4 = \beta_1 + 8\beta_2$$

  - Ponto de inflexão do efeito parcial: valor de x que maximiza (minimiza) y

$$x^* = \frac{-\beta_1}{2\beta_2}$$

4.1. Termos quadráticos e interações

- **Termos de interação**

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + u$$

- Efeito de uma variação de x_1

$$\Delta x_1 \Rightarrow \Delta y = \Delta x_1 \left(\beta_1 + \beta_3 x_2\right)$$

O efeito depende de x_2

  - A interpretação de β_1 não é a habitual:

$$x_2 = 0, \Delta x_1 = 1 \Rightarrow \Delta y \approx \beta_1 + \beta_3 \times 0 \Leftrightarrow \beta_1 \approx \Delta y \text{ quando } x_2 = 0, \Delta x_1 = 1$$

  - Reparametrização dos efeitos de interação

$$y = \beta_0 + \delta_1 x_1 + \delta_2 x_2 + \beta_3 (x_1 - \mu_1)(x_2 - \mu_2) + u$$

Média de x_1 Média de x_2

Podem ser substituídas pelas médias observadas na amostra

  - interpretação de δ_1

    Efeito de uma variação unitária de x_1 *(ceteris paribus)* quando x_1 e x_2 são iguais às respetivas médias.

4.2. Testes da forma funcional

- **Regression specification error test – RESET**
  - Intuição:

  Se o modelo estiver bem especificado não tem variáveis omitidas relevantes, a forma funcional é correta;

  então polinómios das variáveis explicativas não podem ser relevantes para explicar a variável dependente.

  - Procedimento de teste:

1. OLS do modelo original: $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_3 x_k + u \longrightarrow \hat{y}$
2. Estimação OLS da equação de teste:
   i. Variante 1: $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_3 x_k + \gamma \hat{y}^2 + v_1$

   Testar $\gamma = 0$ contra $\gamma \neq 0$

   ii. Variante 2: $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_3 x_k + \gamma_1 \hat{y}^2 + \gamma_2 \hat{y}^3 + v_2$

   Testar $\gamma_1 = 0, \gamma_2 = 0$ conjuntamente
3. Conclusão:

   Se se rejeitar H_0 então há evidência de que o modelo está mal especificado e necessita ser reformulado.

4.2. Testes da forma funcional

- **Teste J**
  - Serve para avaliar a forma funcional de modelos não encaixados para a mesma v. dependente ➡ permite escolher entre modelos não encaixados
  - Procedimento de teste:

1. Estimar pelo OLS cada um dos modelos

OLS do modelo A: $y = \beta_0 + \beta_1 x + \beta_2 w + u \longrightarrow \hat{y}_A$

OLS do modelo B: $y = \alpha_0 + \alpha_1 x + \alpha_2 z + v \longrightarrow \hat{y}_B$

2. Equações de teste:

OLS de $y = \beta_0 + \beta_1 x + \beta_2 w + \gamma_B \hat{y}_B + u_2 \longrightarrow$ testar $H_0 : \gamma_B = 0 \quad H_1 : \gamma_B \neq 0$

OLS de $y = \alpha_0 + \alpha_1 x + \alpha_2 w + \gamma_A \hat{y}_A + v_2 \longrightarrow$ testar $H_0 : \gamma_A = 0 \quad H_1 : \gamma_A \neq 0$

4.2. Testes da forma funcional

- Discussão dos resultados

  i. $\gamma_A \neq 0$ e $\gamma_B = 0 \longrightarrow$ **Modelo A** (modelo B mal especificado)

  A especificação de A acrescenta informação ao modelo B mas a especificação de B não acrescenta informação a A

  ii. $\gamma_A = 0$ e $\gamma_B \neq 0 \longrightarrow$ **Modelo B** (modelo A mal especificado)

  A especificação de B acrescenta informação ao modelo A mas a especificação de A não acrescenta informação a B

  iii. $\gamma_A = 0$ e $\gamma_B = 0 \longrightarrow$ **Modelo A ou Modelo B**

  A especificação de B não acrescenta informação ao modelo A e a especificação de A não acrescenta informação a B

  iv. $\gamma_A \neq 0$ e $\gamma_B \neq 0 \longrightarrow$ **Rejeitar o Modelo A e o Modelo B**

  A especificação de B acrescenta informação ao modelo A e a especificação de A acrescenta informação a B

4.3. Variáveis Dummy

- **Informação qualitativa**
  - Exemplos:
    - Sexo de uma pessoa
    - Região onde vive o agregado familiar
    - Sector onde trabalha a empresa
    - Nível de satisfação
    - Classe de rendimento
  - Como introduzir no modelo de regressão

variáveis Dummy

Variáveis que assumem apenas o valor 0 ou 1

4.3. Variáveis Dummy

- **Apenas uma variável Dummy nas variáveis explicativas**

$$sal = \beta_0 + \delta_0 mulher + \beta_1 educ + u$$

- Assumindo a hipótese DS.4, $E(u \mid female, educ) = 0$ então:

Regressão para as mulheres

$$E(sal \mid mulher = 1, educ) = \beta_0 + \delta_0 + \beta_1 educ$$

Regressão para os homens

$$E(sal \mid mulher = 0, educ) = \beta_0 + \beta_1 educ$$

- Interpretação do coeficiente da variável Dummy

$$\boxed{\delta_0 = E(sal \mid mulher = 1, educ) - E(sal \mid mulher = 0, educ)}$$

Diferença média ou esperada entre o salário de uma mulher e o salário de um homem com o mesmo nível de escolaridade (mantendo tudo o resto constante)

4.3. Variáveis Dummy

- **Dummy Trap**

$$sal = \beta_0 + \delta_0 mulher + \gamma_0 homem + \beta_1 educ + u$$

Existe multicolienaridade perfeita entre as variáveis Dummy e o termo independente → Não é possível estimar todos os coeficientes do modelo

- **O que fazer?**

- Só introduzir uma variável Dummy na equação

$$sal = \beta_0 + \delta_0 mulher + \beta_1 educ + u$$

ou

$$sal = \beta_0 + \gamma_0 homem + \beta_1 educ + u$$

- Retirar o termo independente

Atenção: a interpretação dos coeficientes das Dummies é diferente

$$sal = \delta_0^* mulher + \gamma_0^* homem + \beta_1 educ + u$$

Inconvenientes:
- Mais difícil de testar e avaliar diretamente a diferença entre os dois grupos
- A fórmula habitual do R^2 não é válida.

4.3. Variáveis Dummy

- **Mudar o grupo base**

$$sal = \beta_0 + \delta_0 mulher + \beta_1 educ + u$$

$$\Downarrow$$

$$sal = \beta_0 + \gamma_0 homem + \beta_1 educ + u$$

então $\boxed{\gamma_0 = -\delta_0}$

- **Exemplo**

Dependent Variable: WAGE ($/hora)
Method: Least Squares
Included observations: 526

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-2.319204	0.738825	-3.139042	0.0018
FEMALE	-2.114035	0.262550	-8.051927	0.0000
EDUC	0.556285	0.050287	11.06210	0.0000
EXPER	0.255128	0.034867	7.317144	0.0000
EXPER^2	-0.004440	0.000776	-5.719805	0.0000

4.3. Variáveis Dummy

- **Variável Dummy para explicar** $\log(y)$

$$\log(y) = \beta_0 + \delta_0 Dummy + \beta_1 x_1 + \ldots + \beta_k x_k + u$$

- **Variação exacta** $\%\Delta y = \left[\exp\left(\delta_0\right) - 1\right] \times 100\%$

- **Exemplo**

Dependent Variable: LOG(WAGE)
Method: Least Squares
Included observations: 526

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.390483	0.102210	3.820413	0.0001
FEMALE	-0.337187	0.036321	-9.283424	0.0000
EDUC	0.084136	0.006957	12.09407	0.0000
EXPER	0.038910	0.004824	8.066683	0.0000
EXPER^2	-0.000686	0.000107	-6.388842	0.0000

4.3. Variáveis Dummy

- **Variáveis Dummy: v. qualitativas <u>com várias categorias</u>**

Introduzir uma variável dummy por categoria, deixando uma de fora que será o grupo base para leitura dos resultados

Dependent Variable: WAGE
Method: Least Squares
Included observations: 526

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-2.655184	0.754921	-3.517167	0.0005
MARRFEM	-1.051078	0.427586	-2.458167	0.0143
MARRMALE	2.079997	0.407073	5.109641	0.0000
SINGFEM	-0.423726	0.424226	-0.998821	0.3183
EDUC	0.578108	0.050528	11.44132	0.0000
EXPER	0.053222	0.010823	4.917597	0.0000

Dependent Variable: WAGE
Method: Least Squares
Included observations: 526

Variable	Coefficient	Std. Error	t-Statistic
C	-3.078910	0.725669	-4.242859
MARRFEM	-0.627352	0.388914	-1.613088
MARRMALE	2.503724	0.366868	6.824585
SINGMALE	0.423726	0.424226	0.998821
EDUC	0.578108	0.050528	11.44132
EXPER	0.053222	0.010823	4.917597

<u>Mudança do grupo base</u>

4.3. Variáveis Dummy

- **Variáveis Dummy: v. qualitativas <u>ordinais</u>**

- As variáveis qualitativas ordinais devem ser transformadas em variáveis dummy para conferir mais flexibilidade à estimação

$$escaleduc = \begin{cases} 1 \text{ se tiver completado 4 anos de escolaridade} \\ 2 \text{ se tiver completado 9 anos de escolaridade} \\ 3 \text{ se tiver completado 12 anos de escolaridade} \\ 4 \text{ se tiver completado 15 anos ou mais de escolaridade} \end{cases}$$

- Com variável ordinal

$$sal = \beta_0 + \beta_1 escaleduc + \text{outros fatores}$$

- Com Dummies

$$sal = \beta_0 + \delta_1 educ_1 + \delta_2 educ_2 + \delta_3 educ_3 + \text{outros fatores}$$

$$educ_j = \begin{cases} 1 \text{ se } escaleduc = j \\ 0 \text{ caso contrário} \end{cases}$$

4.3. Variáveis Dummy

- **Interações envolvendo variáveis Dummy**

$$sal = \beta_0 + \delta_0 mulher + \beta_1 educ + \delta_1 mulher \times educ + u$$

- Assumindo a hipótese DS.4, $E(u \mid female, educ) = 0$ então:

Regressão para as mulheres

$$E(sal \mid mulher = 1, educ) = \beta_0 + \delta_0 + (\beta_1 + \delta_1) educ$$

Regressão para os homens

$$E(sal \mid mulher = 0, educ) = \beta_0 + \beta_1 educ$$

- Interpretação do coeficiente da variável Dummy

$$\boxed{\delta_0 = E(sal \mid mulher = 1, educ = 0) - E(sal \mid mulher = 0, educ = 0)}$$

Diferença média ou esperada entre o salário de uma mulher e o salário de um homem com nível de escolaridade=0 (mantendo tudo o resto constante)

$\delta_1 =$ diferença no retorno do salário de uma mulher face a um homem

4.3. Variáveis Dummy

- **Testar diferenças na forma funcional entre grupos da população**

Variável dummy: $D = \begin{cases} 1 \text{ se a observação pertence ao grupo 1} \\ 0 \text{ caso contrário} \end{cases}$

- Estimar o modelo pelo OLS

$$y = \beta_0 + \delta_0 D + \beta_1 x_1 + \delta_1 x_1 \times D + \ldots + \beta_k x_k + \delta_k x_k \times D + u$$

- Testar

$$H_0 : \delta_0 = \delta_1 = \ldots = \delta_k = 0 \qquad H_1 : H_0 \text{ falsa}$$

Teste F → Rejeitando H_0 há evidência de que a forma funcional é diferente nos dois grupos → Fazer uma regressão separada para cada grupo

- Por vezes tem interesse testar apena diferenças nos coeficientes de declive:

$$H_0 : \delta_1 = \ldots = \delta_k = 0 \qquad H_1 : H_0 \text{ falsa}$$