Modelos de Regressão Múltipla - Parte IV

Erica Castilho Rodrigues

01 de Fevereiro de 2017

Regressão com Interação

Regressão com Interação

- Nos modelos vistos até agora a mudança de um grupo para o outro estava apenas no intercepto.
- Podemos também incorporar diferenças na forma como as preditoras entram no modelo.
- Ou sejam, a inclinação da reta muda de um grupo para o outro.
- As retas deixam de ser paralelas.

- Suponha que temos dois grupos: A e B.
- Queremos ajustar um modelo de regressão linear simples para cada um deles.
- Essas retas podem ser iguais ou diferentes.

- Temos quatro possibilidades:

- Temos quatro possibilidades:
  - Duas retas distintas (intercetpto e inclinação diferentes)

$$Y_A = \beta_0 + \beta_1 X \quad Y_B = \gamma_0 + \gamma_1 X$$

temos

- Temos quatro possibilidades:
  - Duas retas distintas (intercetpto e inclinação diferentes)

$$Y_A = \beta_0 + \beta_1 X \quad Y_B = \gamma_0 + \gamma_1 X$$

temos 4 parâmetros nesse caso.

- Temos quatro possibilidades:
  - Duas retas distintas (intercetpto e inclinação diferentes)

    $$Y_A = \beta_0 + \beta_1 X \quad Y_B = \gamma_0 + \gamma_1 X$$

    temos 4 parâmetros nesse caso.
  - Duas retas paralelas (interceptos distintos e mesma inclinação)

    $$Y_A = \beta_0 + \beta_1 X \quad Y_B = \gamma_0 + \beta_1 X$$

    temos

- Temos quatro possibilidades:
  - Duas retas distintas (intercetpto e inclinação diferentes)

    $$Y_A = \beta_0 + \beta_1 X \quad Y_B = \gamma_0 + \gamma_1 X$$

    temos 4 parâmetros nesse caso.
  - Duas retas paralelas (interceptos distintos e mesma inclinação)

    $$Y_A = \beta_0 + \beta_1 X \quad Y_B = \gamma_0 + \beta_1 X$$

    temos 3 parâmetros nesse caso.
  - Duas reta com o mesmo intercepto (interceptos iguaise inclinações distintas)

    $$Y_A = \beta_0 + \beta_1 X \quad Y_B = \beta_0 + \gamma_1 X$$

    temos

- Temos quatro possibilidades:
  - Duas retas distintas (intercetpto e inclinação diferentes)

$$Y_A = \beta_0 + \beta_1 X \quad Y_B = \gamma_0 + \gamma_1 X$$

    temos 4 parâmetros nesse caso.
  - Duas retas paralelas (interceptos distintos e mesma inclinação)

$$Y_A = \beta_0 + \beta_1 X \quad Y_B = \gamma_0 + \beta_1 X$$

    temos 3 parâmetros nesse caso.
  - Duas reta com o mesmo intercepto (interceptos iguaise inclinações distintas)

$$Y_A = \beta_0 + \beta_1 X \quad Y_B = \beta_0 + \gamma_1 X$$

    temos 3 parâmetros nesse caso.
  - Duas retas iguais

$$Y_A = \beta_0 + \beta_1 X \quad Y_B = \beta_0 + \beta_1 X$$

    temos

- Temos quatro possibilidades:
  - Duas retas distintas (intercetpto e inclinação diferentes)

$$Y_A = \beta_0 + \beta_1 X \quad Y_B = \gamma_0 + \gamma_1 X$$

    temos 4 parâmetros nesse caso.
  - Duas retas paralelas (interceptos distintos e mesma inclinação)

$$Y_A = \beta_0 + \beta_1 X \quad Y_B = \gamma_0 + \beta_1 X$$

    temos 3 parâmetros nesse caso.
  - Duas reta com o mesmo intercepto (interceptos iguaise inclinações distintas)

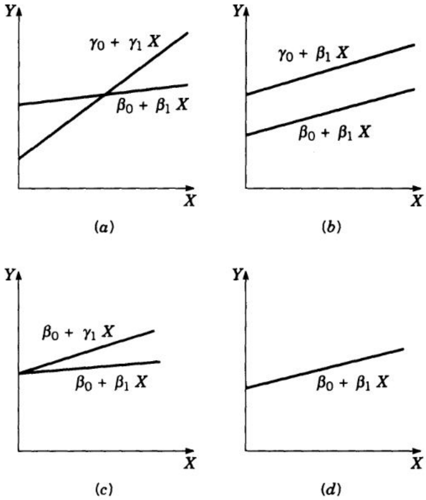
$$Y_A = \beta_0 + \beta_1 X \quad Y_B = \beta_0 + \gamma_1 X$$

    temos 3 parâmetros nesse caso.
  - Duas retas iguais

$$Y_A = \beta_0 + \beta_1 X \quad Y_B = \beta_0 + \beta_1 X$$

    temos 2 parâmetros nesse caso.

- A figura a seguir ilustra esses 4 modelos.



- Podemos escolher uma dentre essas possibilidades usando um único modelo.
- Isso é feito usando variáveis indicadoras.
- Vamos definir as seguintes variáveis

X_0	Z
1	0
1	1

Obsevação: a coluna de 1's referente ao β_0 está sendo contada como indicadora também.

- Ajustamos o modelo

$$Y = X_0(\beta_0 + \beta_1 X) + Z(\alpha_0 + \alpha_1 X) + \epsilon$$

$$= X_0\beta_0 + \beta_1 X_0 X + Z\alpha_0 + \alpha_1 ZX + \epsilon$$

- Como $X_0 = 1$ para qualquer um dos grupos, ficamos com

$$Y = \beta_0 + \beta_1 X + Z\alpha_0 + \alpha_1 ZX + \epsilon$$

- O termo

$$\alpha_1 ZX$$

que é o termo chamado termo de interação.
- Ele permite que a inclinação da reta mude de um grupo para o outro.
- Podemos escrever os modelos para os grupos A e B.
- Para o grupo A, $X_0 = 1$ e $Z = 0$

$$Y = \beta_0 + \beta_1 X \, .$$

- Para o gurpo B, $X_0 = 1$ e $Z = 1$

$$Y = (\beta_0 + \alpha_0) + (\beta_1 + \alpha_1) X \, .$$

- Se quisermos verificar se os interceptos são iguais, devemos testar

- Se quisermos verificar se os interceptos são iguais, devemos testar

$$H_0 : \alpha_0 = 0 \quad H_0 : \alpha_0 \neq 0 \; .$$

- Se quisermos testar se as retas são paralelas equivale a testar

- Se quisermos verificar se os interceptos são iguais, devemos testar

$$H_0 : \alpha_0 = 0 \quad H_0 : \alpha_0 \neq 0 \; .$$

- Se quisermos testar se as retas são paralelas equivale a testar

$$H_0 : \alpha_1 = 0 \quad H_1 : \alpha_1 \neq 0 \; .$$

- Se quisermos verificar se as retas são iguais, devemos testar

- Se quisermos verificar se os interceptos são iguais, devemos testar

$$H_0 : \alpha_0 = 0 \quad H_0 : \alpha_0 \neq 0 \, .$$

- Se quisermos testar se as retas são paralelas equivale a testar

$$H_0 : \alpha_1 = 0 \quad H_1 : \alpha_1 \neq 0 \, .$$

- Se quisermos verificar se as retas são iguais, devemos testar

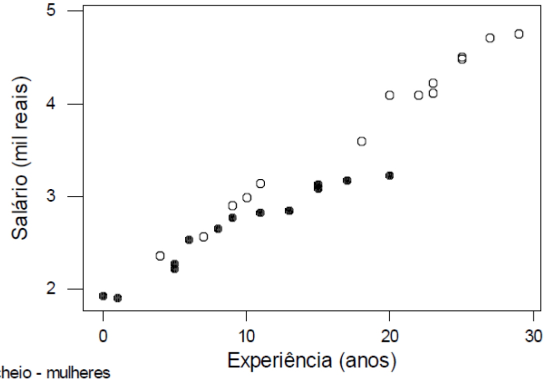
$$H_0 : \alpha_1 = \alpha_0 = 0 \quad H_1 : \alpha_0 = 0 \text{ e/ou } \alpha_1 \neq 0 \, .$$

Exemplo:

- Queremos analisar os salários de gerentes de banco.
- O nosso interesse é verificar como eles variam com a experiência dos profissionais.
- Vamos entrar ainda com uma outra variável: sexo do gerente.
- Sexo é uma variável categórica e deve ser analisado usando indicadoras.

Exemplo: (continuação)

- Gráfico de dispersão do salário em função da experiência.



Círculo cheio - mulheres
Círculo vazio - homens

Exemplo: (continuação)

- Ajustamos primeiramente modelos separados para cada um dos grupos.
- Modelo para Mulheres:

$$salario = 1.97 + 0.0722 experiencia$$

- Modelo para Homens:

$$salario = 1.98 + 0.0983 experiencia$$

- Vamos agora ajustar o modelo único.

Exemplo: (continuação)

- A variável sexo vai entrar como variável indicadora

$$sexo = \begin{cases} 1 & \text{se o indivíduo é homem} \\ 0 & \text{se é mulher.} \end{cases}$$

- O modelo ajustado é o seguinte

$$salario = 1,8 + 0,0893(experiencia) + 0,335(sexo)$$

- O que significa o valor 1,8?

Exemplo: (continuação)

- A variável sexo vai entrar como variável indicadora

$$sexo = \begin{cases} 1 & \text{se o indivíduo é homem} \\ 0 & \text{se é mulher.} \end{cases}$$

- O modelo ajustado é o seguinte

$$salario = 1,8 + 0,0893(experiencia) + 0,335(sexo)$$

- O que significa o valor 1,8?
- O salário esperado de uma mulher sem experiência é R$180,00.
- O que significa o valor 0,0893?

Exemplo: (continuação)

- A variável sexo vai entrar como variável indicadora

$$sexo = \begin{cases} 1 & \text{se o indivíduo é homem} \\ 0 & \text{se é mulher.} \end{cases}$$

- O modelo ajustado é o seguinte

$$salario = 1,8 + 0,0893(experiencia) + 0,335(sexo)$$

- O que significa o valor 1,8?
- O salário esperado de uma mulher sem experiência é R$180,00.
- O que significa o valor 0,0893?
- A cada aumento em um ano de experiência, mantendo-se fixo o sexo, espera-se que o salário aumente em R$89,00.

Exemplo: (continuação)

- O que significa o valor 0,335?

Exemplo: (continuação)

- O que significa o valor 0,335?
- Espera-se que um homem ganhe R$335,00 a mais do que uma mulher com a mesma experiência.
- O gráfico de dispersão mostra que:
  - crescimento de salário ocorre de maneira diferente para homens e mulheres.
- As retas não parecem ser paralelas.
- Isso é um indício de que existe interação entre as variáveis sexo e experiência.
- A experiência não influencia no salário da mesma maneira para pessoas de sexos distintos.
- Vamos incorporar essa interação no modelo.

Exemplo: (continuação)

- O modelo fica da forma

$$Y = \beta_0 + \beta_1 X + \alpha_0 Z + \alpha_1 ZX + \epsilon$$

onde

$$Y\text{: salário}$$

$$X\text{: experiência}$$

$$Z = \begin{cases} 1 & \text{se o indivíduo é homem} \\ 0 & \text{se é mulher.} \end{cases}$$

Exemplo: (continuação)

- A figura a seguir mostra o resultado do ajuste do modelo.

```
The regression equation is
salario = 1,97 + 0,0722 experiencia + 0,0091 sexo + 0,0261 sexo*experiência

Predictor        Coef     SE Coef        T       P
Constant      1,96844     0,05369    36,66   0,000
experien     0,072199    0,004750    15,20   0,000
sexo          0,00908     0,08639     0,11   0,917
sexo*exp     0,026062    0,005858     4,45   0,000

S = 0,1018      R-Sq = 98,8%     R-Sq(adj) = 98,6%

Analysis of Variance

Source            DF        SS        MS        F        P
Regression         3   18,9159    6,3053   608,75    0,000
Residual Error    23    0,2382    0,0104
Total             26   19,1541

Source       DF    Seq SS
experien      1   18,1539
sexo          1    0,5570
sexo*exp      1    0,2050
```

Exemplo: (continuação)

- Conclusões:

Exemplo: (continuação)

- Conclusões:
  - A variável experiêcia é significativa para explicar os salário.
  - Qual a conclusão do teste?

$$H_0 = \alpha_0 = 0 \quad H_1 : \alpha_0 \neq 0$$

Exemplo: (continuação)

- Conclusões:
  - A variável experiêcia é significativa para explicar os salário.
  - Qual a conclusão do teste?

$$H_0 = \alpha_0 = 0 \quad H_1 : \alpha_0 \neq 0$$

  - Não rejeitamos a H_0.
  - Isso indica que

Exemplo: (continuação)

- Conclusões:
  - A variável experiêcia é significativa para explicar os salário.
  - Qual a conclusão do teste?

$$H_0 = \alpha_0 = 0 \quad H_1 : \alpha_0 \neq 0$$

  - Não rejeitamos a H_0.
  - Isso indica que as retas têm o mesmo intercepto.
  - Qual a conclusão do teste?

$$H_0 = \alpha_1 = 0 \quad H_1 : \alpha_1 \neq 0$$

Exemplo: (continuação)

- Conclusões:
  - A variável experiêcia é significativa para explicar os salário.
  - Qual a conclusão do teste?

$$H_0 = \alpha_0 = 0 \quad H_1 : \alpha_0 \neq 0$$

  - Não rejeitamos a H_0.
  - Isso indica que as retas têm o mesmo intercepto.
  - Qual a conclusão do teste?

$$H_0 = \alpha_1 = 0 \quad H_1 : \alpha_1 \neq 0$$

  - Rejeitamos a H_0.
  - Isso indica que

Exemplo: (continuação)

- Conclusões:
  - A variável experiêcia é significativa para explicar os salário.
  - Qual a conclusão do teste?

$$H_0 = \alpha_0 = 0 \quad H_1 : \alpha_0 \neq 0$$

  - Não rejeitamos a H_0.
  - Isso indica que as retas têm o mesmo intercepto.
  - Qual a conclusão do teste?

$$H_0 = \alpha_1 = 0 \quad H_1 : \alpha_1 \neq 0$$

  - Rejeitamos a H_0.
  - Isso indica que as retas não têm a mesma inclinação.
  - O salário não cresce da mesma forma para homens e mulheres.

Exemplo: (continuação)

- Eles partem do mesmo intercepto.
- O ganho esperado de um gerente sem experiência é o mesmo ele sendo homem ou mulher.

Exemplo: **(continuação)**

- O modelo ajustado foi o seguinte:

$$Y = 1,971 + 0,0719X + 0,009Z + 0,0266ZX\ .$$

- Qual aumento esperado no salário a cada ano de experiência para as mulheres?

Exemplo: (continuação)

- O modelo ajustado foi o seguinte:

$$Y = 1,971 + 0,0719X + 0,009Z + 0,0266ZX \ .$$

- Qual aumento esperado no salário a cada ano de experiência para as mulheres? 0,07 ou R$70,00
- Qual aumento esperado no salário a cada ano de experiência para os homens?

Exemplo: (continuação)

- O modelo ajustado foi o seguinte:

$$Y = 1,971 + 0,0719X + 0,009Z + 0,0266ZX\ .$$

- Qual aumento esperado no salário a cada ano de experiência para as mulheres? 0,07 ou R$70,00
- Qual aumento esperado no salário a cada ano de experiência para os homens?
- O modelo para os homens é dado por

Exemplo: (continuação)

- O modelo ajustado foi o seguinte:

$$Y = 1,971 + 0,0719X + 0,009Z + 0,0266ZX \; .$$

- Qual aumento esperado no salário a cada ano de experiência para as mulheres? 0,07 ou R$70,00
- Qual aumento esperado no salário a cada ano de experiência para os homens?
- O modelo para os homens é dado por

$$Y = (1,971 + 0,009) + (0,0719 + 0,0266)X$$

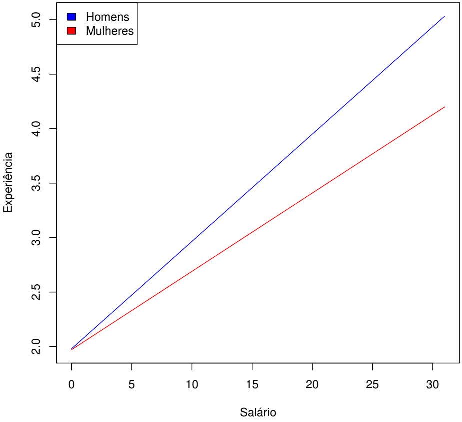
$$= 1,1978 + 0,985X$$

Exemplo: (continuação)

- Portanto o aumento de salário esperado para os homens para cada ano de experiência é de R$98,50 .
- Um ano de experiência a mais para os homens tem um efeito maior do que para as mulheres.
- Se eles não tem experiência, o ganho esperado é o mesmo para os dois sexos.

Exemplo: (continuação)

- A figura mostra a reta para cada um dos grupos.



Exemplo:

- Vamos analisar o salário anual (em mil USD) de uma amostra aleatória de 220 executivos.
- 145 homens e 75 mulheres.
- O salário será relacionado com as seguintes variáveis explicativas:
  1. sexo (1:masculino; 0: feminino);
  2. posição na empresa (varia de 1 a 9;
  3. anos de experiência no cargo.

Exemplo:

- Vamos ajustar o modelo mais simples primeiro

$$Y_i = \beta_0 + \beta_1 Sexo + \beta_2 Experiencia + \beta_3 Posic + \epsilon_i$$

em que Y_i denota o salário do executivo.

Exemplo:

- A tabela a seguir mostra os resultados do modelo.

Efeito	Estimativa	Erro padrão	valor-t	valor-P
Constante	115,262	1,401	82,25	0,000
Experiência	-0,472	0,113	-4,17	0,000
SexoM	-2,201	1,080	-2,04	0,043
Posição	6,710	0,313	21,46	0,000
R^2	0,712			
R^2-ajustado	0,708			
s	6,770			
F	177,90	(3 e 216 g.l.)		0,000

Exemplo:

- Conclusões:

Exemplo:

- Conclusões:
- Todas as variáveis são significativas.
- Mantendo as demais variáveis fixas, o aumento em um ano na experiência reduz o salário esperado em -0,472.

Exemplo:

- Conclusões:
- Todas as variáveis são significativas.
- Mantendo as demais variáveis fixas, o aumento em um ano na experiência reduz o salário esperado em -0,472.
- O salário esperado para os homens é menor do que o das mulheres em 2,201 unidades.

Exemplo:

- Conclusões:
- Todas as variáveis são significativas.
- Mantendo as demais variáveis fixas, o aumento em um ano na experiência reduz o salário esperado em -0,472.
- O salário esperado para os homens é menor do que o das mulheres em 2,201 unidades.
- Para cada aumento na posição, mantendo as demais variáveis fixas, o salário esperado aumenta em 6,710.
- Vamos incluir a interação.

Exemplo:

- Vamos testar a interação com as três variáveis.

Interação	valor-F	valor-P
sexo*exper	1,615	0,20
sexo*posicão	0,001	0,97
exper*posição	7,594	0,00

- Conclusão:

Exemplo:

- Vamos testar a interação com as três variáveis.

Interação	valor-F	valor-P
sexo*exper	1,615	0,20
sexo*posicão	0,001	0,97
exper*posição	7,594	0,00

- Conclusão: apenas o segundo termo é singificativo.

Exemplo:

- O modelo final é dado por

Efeito	Estimativa	Erro padrão	valor-t	valor-P
Constante	108,042	2,961	36,48	0,000
Experiência	0,336	0,314	1,07	0,285
SexoM	-2,811	1,087	-2,59	0,010
Posição	8,096	0,590	13,73	0,000
Posic*Exper	-0,135	0.049	-2,76	0.006
R^2	0,722			
R^2-ajustado	0,716			
s	6,670			
F	139,40	(4 e 215 g.l.)		0,000

Exemplo:

- Conclusões:

Exemplo:

- Conclusões:
- Experiência deixa de ser significativa.
- Se o executivo é mais experiente, o efeito no salário de aumentar a posição é menor.