ACA228 - Modelos de Regressão e Previsão

Regressão Linear Multipla: variáveis qualitativas e efeito de interação

Prof. Carlos Trucíos
carlos.trucios@facc.ufrj.br
ctruciosm.github.io

Faculdade de Administração e Ciências Contábeis,
Universidade Federal do Rio de Janeiro

Aula 11

Introdução

Introdução

- Até agora temos trabalhando com variaveis explicativas de tipo quantitativo

Introdução

- Até agora temos trabalhando com variaveis explicativas de tipo quantitativo
- No dia a dia, variaveis explicativas de tipo **qualitativo** podem também nos ajudar a explicar a variabilidade de y

Introdução

- Até agora temos trabalhando com variaveis explicativas de tipo quantitativo
- No dia a dia, variaveis explicativas de tipo **qualitativo** podem também nos ajudar a explicar a variabilidade de y
  - sexo

Introdução

- Até agora temos trabalhando com variaveis explicativas de tipo quantitativo
- No dia a dia, variaveis explicativas de tipo **qualitativo** podem também nos ajudar a explicar a variabilidade de y
  - sexo
  - estado civil

Introdução

- Até agora temos trabalhando com variaveis explicativas de tipo quantitativo
- No dia a dia, variaveis explicativas de tipo **qualitativo** podem também nos ajudar a explicar a variabilidade de y
  - sexo
  - estado civil
  - raça

Introdução

- Até agora temos trabalhando com variaveis explicativas de tipo quantitativo
- No dia a dia, variaveis explicativas de tipo **qualitativo** podem também nos ajudar a explicar a variabilidade de y
  - sexo
  - estado civil
  - raça
  - religião

Introdução

- Até agora temos trabalhando com variaveis explicativas de tipo quantitativo
- No dia a dia, variaveis explicativas de tipo **qualitativo** podem também nos ajudar a explicar a variabilidade de y
  - sexo
  - estado civil
  - raça
  - religião
  - estado

Introdução

- Até agora temos trabalhando com variaveis explicativas de tipo quantitativo
- No dia a dia, variaveis explicativas de tipo **qualitativo** podem também nos ajudar a explicar a variabilidade de y
  - sexo
  - estado civil
  - raça
  - religião
  - estado
- Variáveis qualitativas no modelo não afetarão o processo de estimação mas levarão a uma interpretação diferente dos parâmetros do modelo.

Introdução

- Até agora temos trabalhando com variaveis explicativas de tipo quantitativo
- No dia a dia, variaveis explicativas de tipo **qualitativo** podem também nos ajudar a explicar a variabilidade de y
    - sexo
    - estado civil
    - raça
    - religião
    - estado
- Variáveis qualitativas no modelo não afetarão o processo de estimação mas levarão a uma interpretação diferente dos parâmetros do modelo.
- Como definimos a variável qualitativa nos ajudara na interpretação.

Variável independente binária

Variável independente binária (dummy)

Seja

$$wage = \beta_0 + \beta_1 female + \beta_2 educ + u$$

onde:

- female = 1 quando a pessoa for mulher
- female = 0 quando a pessoa for homem

0-1?

Variáveis binárias são recategorizadas como 0 e 1, essa codificação não é obrigatória mas é comum e ajuda na interpretação.

Variável independente binária (dummy)

Seja

$$wage = \beta_0 + \beta_1 female + \beta_2 educ + u$$

onde:

- female = 1 quando a pessoa for mulher
- female = 0 quando a pessoa for homem

0-1?

Variáveis binárias são recategorizadas como 0 e 1, essa codificação não é obrigatória mas é comum e ajuda na interpretação.

- β_1 é a diferença no salário entre mulheres (female = 1) e homens (female = 0) quando todos os otros fatores permanecem fixos.

Variável independente binária (dummy)

Seja

$$wage = \beta_0 + \beta_1 female + \beta_2 educ + u$$

onde:

- female $= 1$ quando a pessoa for mulher
- female $= 0$ quando a pessoa for homem

0-1?

Variáveis binárias são recategorizadas como 0 e 1, essa codificação não é obrigatória mas é comum e ajuda na interpretação.

- β_1 é a diferença no salário entre mulheres (female $= 1$) e homens (female $= 0$) quando todos os otros fatores permanecem fixos.
- β_1 nos dirá, por exemplo, se existe discriminação de genero no salário.

Variável independente binária (dummy)

```r
library(wooldridge)
coef(lm(wage~female + educ, data = wage1))
```

```
## (Intercept)      female        educ
##   0.6228168  -2.2733619   0.5064521
```

$$wage = 0.6228168 - 2.2733619 female + 0.5064521 educ$$

Variável independente binária (dummy)

```r
library(wooldridge)
coef(lm(wage~female + educ, data = wage1))
```

```
## (Intercept)      female        educ
##   0.6228168  -2.2733619   0.5064521
```

$$wage = 0.6228168 - 2.2733619 female + 0.5064521 educ$$

- Quando a pessoa for homem (female $= 0$),

$$\widehat{wage} = 0.6228168 + 0.5064521 \text{ educ}$$

Variável independente binária (dummy)

```
library(wooldridge)
coef(lm(wage~female + educ, data = wage1))
```

```
## (Intercept)      female        educ
##   0.6228168  -2.2733619   0.5064521
```

$$wage = 0.6228168 - 2.2733619 female + 0.5064521 educ$$

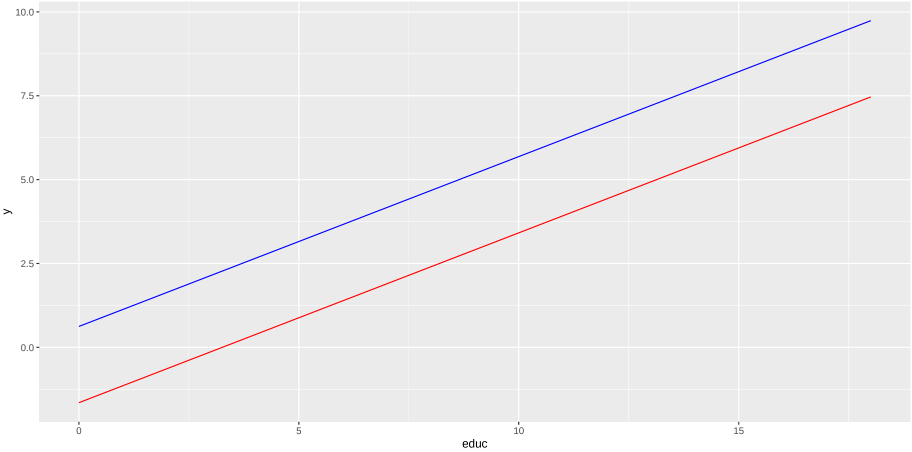
- Quando a pessoa for homem (female $= 0$),

$$\widehat{wage} = 0.6228168 + 0.5064521 \text{ educ}$$

- Quando a pessoa for mulher (female $= 1$),

$$\widehat{wage} = \underbrace{0.6228168 - 2.2733619}_{-1.650545} + 0.5064521 \text{ educ}$$

Variável independente binária (dummy)



Variável independente binária (dummy)

- $\widehat{wage} = -1.650545 + 0.5064521 \times 1 = -1.144093$
- $\widehat{wage} = -1.650545 + 0.5064521 \times 2 = -0.6376408$
- $\widehat{wage} = -1.650545 + 0.5064521 \times 3 = -0.1311887$
- $\widehat{wage} = -1.650545 + 0.5064521 \times 4 = 0.3752634$

Variável independente binária (dummy)

- $\widehat{wage} = -1.650545 + 0.5064521 \times 1 = -1.144093$
- $\widehat{wage} = -1.650545 + 0.5064521 \times 2 = -0.6376408$
- $\widehat{wage} = -1.650545 + 0.5064521 \times 3 = -0.1311887$
- $\widehat{wage} = -1.650545 + 0.5064521 \times 4 = 0.3752634$

Mulheres com menos de 4 anos de educação tem salario negativo?

Variável independente binária (dummy)

- $\widehat{wage} = -1.650545 + 0.5064521 \times 1 = -1.144093$
- $\widehat{wage} = -1.650545 + 0.5064521 \times 2 = -0.6376408$
- $\widehat{wage} = -1.650545 + 0.5064521 \times 3 = -0.1311887$
- $\widehat{wage} = -1.650545 + 0.5064521 \times 4 = 0.3752634$

Mulheres com menos de 4 anos de educação tem salario negativo?

```
prop.table(table(wage1$educ[wage1$female == 1]<4))
```

```
##
##       FALSE        TRUE
## 0.992063492 0.007936508
```

Cuidado!

Resultados contra intuitivos podem ser causados por falta de dados ou estimadores viesados!

Variável independente binária (dummy)

- Temos usado female (0 e 1)

Variável independente binária (dummy)

- Temos usado female (0 e 1)
- E se incluirmos também outra variavel, digamos male (0 e 1)?

Variável independente binária (dummy)

- Temos usado female (0 e 1)
- E se incluirmos também outra variavel, digamos male (0 e 1)?
  - reduntante

Variável independente binária (dummy)

- Temos usado female (0 e 1)
- E se incluirmos também outra variavel, digamos male (0 e 1)?
  - reduntante
  - female + male = 1

Variável independente binária (dummy)

- Temos usado female (0 e 1)
- E se incluirmos também outra variavel, digamos male (0 e 1)?
  - reduntante
  - female + male = 1

Variável independente binária (dummy)

- Temos usado female (0 e 1)
- E se incluirmos também outra variavel, digamos male (0 e 1)?
    - reduntante
    - female + male = 1

```
wage1$male = ifelse(wage1$female == 0, 1, 0)
coef(lm(wage~female + male + educ, data = wage1))
```

```
## (Intercept)      female        male        educ
##   0.6228168  -2.2733619          NA   0.5064521
```

Variável independente binária (dummy)

Explicando a nota média no curso superior

$$colGPA = \beta_0 + \beta_1 PC + \beta_2 hsGPA + \beta_3 ACT + u$$

- colGPA $=$ nota média no curso superior
- PC $= 1$ se a pessoa tem computador próprio e 0 c.c.
- hsGPA: high school GPA
- ACT: nota do teste de avaliação para ingresso no curso superior

Variável independente binária (dummy)

```
summary(gpa1$colGPA)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   2.200   2.800   3.000   3.057   3.300   4.000
```

```
summary(gpa1$hsGPA)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   2.400   3.200   3.400   3.402   3.600   4.000
```

Variável independente binária (dummy)

```r
modelo = lm(colGPA~PC+hsGPA+ACT, data = gpa1)
round(summary(modelo)$coefficients,5)
```

```
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.26352    0.33313 3.79292  0.00022
## PC           0.15731    0.05729 2.74596  0.00684
## hsGPA        0.44724    0.09365 4.77580  0.00000
## ACT          0.00866    0.01053 0.82199  0.41251
```

```r
summary(modelo)$adj.r.squared
```

```
## [1] 0.2022918
```

- Como interpretamos $\hat{\beta}_{hsGPA}$?

Variável independente binária (dummy)

```r
modelo = lm(colGPA~PC+hsGPA+ACT, data = gpa1)
round(summary(modelo)$coefficients,5)
```

```
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.26352    0.33313 3.79292  0.00022
## PC           0.15731    0.05729 2.74596  0.00684
## hsGPA        0.44724    0.09365 4.77580  0.00000
## ACT          0.00866    0.01053 0.82199  0.41251
```

```r
summary(modelo)$adj.r.squared
```

```
## [1] 0.2022918
```

- Como interpretamos $\hat{\beta}_{hsGPA}$?
- Como interpretamos $\hat{\beta}_{PC}$?

Variável independente binária (dummy)

Quando nossa variável dependente é $\log(y)$ o β associado à variável *dummy* tem uma interpretação percentual.

Variável independente binária (dummy)

Quando nossa variável dependente é $\log(y)$ o β associado à variável *dummy* tem uma interpretação percentual.

```
modelo = lm(log(price)~ log(lotsize) + log(sqrft) +
              bdrms + colonial, data = hprice1)
round(summary(modelo)$coefficients,5)
```

```
##                 Estimate Std. Error  t value Pr(>|t|)
## (Intercept)   -1.34959    0.65104 -2.07297  0.04128
## log(lotsize)   0.16782    0.03818  4.39539  0.00003
## log(sqrft)     0.70719    0.09280  7.62045  0.00000
## bdrms          0.02683    0.02872  0.93409  0.35297
## colonial       0.05380    0.04477  1.20153  0.23296
```

Variável independente binária (dummy)

$$\widehat{\log(price)} =$$
$$-1.350 + 0.168 log(lotsize) + 0.707 log(sqrft) + 0.027 bdrms + 0.054 colonial$$

- lotsize: tamanho do lote em pés^2
- sqrft: tamanho da casa em pés^2
- bdrms: número de quartos
- colonial:=1 se a casa é de estilo colonial

1 metro = 3.2808 pés

Variável independente binária (dummy)

$\widehat{\log(price)} =$
$-1.350 + 0.168 log(lotsize) + 0.707 log(sqrft) + 0.027 bdrms + 0.054 colonial$

Como interpretamos. . . (ver aula 3)

- $\hat{\beta}_{log(lotsize)}$: Quando o tamanho do lote aumenta em 1%, o preço do imovel aumenta em 0.168%

Variável independente binária (dummy)

$$\widehat{\log(price)} =$$
$$-1.350 + 0.168 log(lotsize) + 0.707 log(sqrft) + 0.027 bdrms + 0.054 colonial$$

Como interpretamos... (ver aula 3)

- $\hat{\beta}_{log(lotsize)}$: Quando o tamanho do lote aumenta em 1%, o preço do imovel aumenta em 0.168%
- $\hat{\beta}_{log(sqrft)}$: Quando o tamanho da casa aumenta em 1%, o preço do imovel aumenta em 0.707%

Variável independente binária (dummy)

$\widehat{\log(price)} =$
$-1.350 + 0.168 log(lotsize) + 0.707 log(sqrft) + 0.027 bdrms + 0.054 colonial$

Como interpretamos... (ver aula 3)

- $\hat{\beta}_{log(lotsize)}$: Quando o tamanho do lote aumenta em 1%, o preço do imovel aumenta em 0.168%
- $\hat{\beta}_{log(sqrft)}$: Quando o tamanho da casa aumenta em 1%, o preço do imovel aumenta em 0.707%
- $\hat{\beta}_{bdrms}$: A cada 1 quarto adicional, espera-se que a casa seja vendida por 2.7%(100×0.027) a mais

Variável independente binária (dummy)

$\widehat{\log(price)} =$
$-1.350 + 0.168 log(lotsize) + 0.707 log(sqrft) + 0.027 bdrms + 0.054 colonial$

Como interpretamos... (ver aula 3)

- $\hat{\beta}_{log(lotsize)}$: Quando o tamanho do lote aumenta em 1%, o preço do imovel aumenta em 0.168%
- $\hat{\beta}_{log(sqrft)}$: Quando o tamanho da casa aumenta em 1%, o preço do imovel aumenta em 0.707%
- $\hat{\beta}_{bdrms}$: A cada 1 quarto adicional, espera-se que a casa seja vendida por 2.7%(100×0.027) a mais
- $\hat{\beta}_{colonial}$: Um imovel colonial espera-se ser vendido por 5.4%(100×0.054) a mais (ou em média casas de tipo colonial são 5.4% mais caras)

Variável independente binária (dummy)

Outro exemplo

```
modelo = lm(log(wage)~ female + educ + exper +
         I(exper^2) + tenure + I(tenure^2), data = wage1)
round(summary(modelo)$coefficients,4)
```

```
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.4167     0.0989  4.2121    0.000
## female        -0.2965     0.0358 -8.2812    0.000
## educ           0.0802     0.0068 11.8682    0.000
## exper          0.0294     0.0050  5.9159    0.000
## I(exper^2)    -0.0006     0.0001 -5.4305    0.000
## tenure         0.0317     0.0068  4.6330    0.000
## I(tenure^2)   -0.0006     0.0002 -2.4934    0.013
```

Variável independente binária (dummy)

$$\widehat{\log(wage)} = 0.4167 + -0.2965 female + 0.0802 educ + 0.0294 exper - 0.0006 exper^2 + 0.0317 tenure + -0.0006 tenure^2$$

- Mulheres ganham em média $\approx 29.65\%$ a menos do que homens

Variável independente binária (dummy)

$\widehat{\log(wage)} = 0.4167 + -0.2965 female + 0.0802 educ + 0.0294 exper - 0.0006 exper^2 + 0.0317 tenure + -0.0006 tenure^2$

- Mulheres ganham em média $\approx 29.65\%$ a menos do que homens
- Se quisermos não a percentagem aproximada, mas a exata?

Variável independente binária (dummy)

$\widehat{\log(wage)} = 0.4167 + -0.2965 female + 0.0802 educ + 0.0294 exper - 0.0006 exper^2 + 0.0317 tenure + -0.0006 tenure^2$

- Mulheres ganham em média $\approx 29.65\%$ a menos do que homens
- Se quisermos não a percentagem aproximada, mas a exata?
- $\log(wage_F) - \log(wage_M) = \log(wage_F/wage_M) = -0.2965$

Variável independente binária (dummy)

$\widehat{\log(wage)} = 0.4167 + -0.2965 female + 0.0802 educ + 0.0294 exper - 0.0006 exper^2 + 0.0317 tenure + -0.0006 tenure^2$

- Mulheres ganham em média $\approx 29.65\%$ a menos do que homens
- Se quisermos não a percentagem aproximada, mas a exata?
- $\log(wage_F) - \log(wage_M) = \log(wage_F/wage_M) = -0.2965$
- $wage_F/wage_M = \exp(-0.2965)$

Variável independente binária (dummy)

$\widehat{\log(wage)} = 0.4167 + -0.2965 female + 0.0802 educ + 0.0294 exper - 0.0006 exper^2 + 0.0317 tenure + -0.0006 tenure^2$

- Mulheres ganham em média $\approx 29.65\%$ a menos do que homens
- Se quisermos não a percentagem aproximada, mas a exata?
- $\log(wage_F) - \log(wage_M) = \log(wage_F/wage_M) = -0.2965$
- $wage_F/wage_M = \exp(-0.2965)$
- $\underbrace{wage_F/wage_M - 1}_{\frac{wage_F - wage_M}{wage_M}} = \underbrace{\exp(-0.2965) - 1}_{-0.2565844}$

Variável independente binária (dummy)

$\widehat{\log(wage)} = 0.4167 + -0.2965 female + 0.0802 educ + 0.0294 exper - 0.0006 exper^2 + 0.0317 tenure + -0.0006 tenure^2$

- Mulheres ganham em média $\approx 29.65\%$ a menos do que homens
- Se quisermos não a percentagem aproximada, mas a exata?
- $\log(wage_F) - \log(wage_M) = \log(wage_F/wage_M) = -0.2965$
- $wage_F/wage_M = \exp(-0.2965)$
- $\underbrace{wage_F/wage_M - 1}_{\frac{wage_F - wage_M}{wage_M}} = \underbrace{\exp(-0.2965) - 1}_{-0.2565844}$
- O salário da mulher é, em média, 25.65% menor.

Variável independente com multiplas categorias

```r
modelo = lm(log(wage)~ married + female + educ + exper +
        I(exper^2) + tenure + I(tenure^2), data = wage1)
round(summary(modelo)$coefficients,5)
```

```
##              Estimate Std. Error  t value Pr(>|t|)
## (Intercept)  0.41778    0.09887  4.22575  0.00003
## married      0.05292    0.04076  1.29850  0.19469
## female      -0.29018    0.03611 -8.03565  0.00000
## educ         0.07915    0.00680 11.63989  0.00000
## exper        0.02695    0.00533  5.06095  0.00000
## I(exper^2)  -0.00054    0.00011 -4.81350  0.00000
## tenure       0.03130    0.00685  4.56999  0.00001
## I(tenure^2) -0.00057    0.00023 -2.44753  0.01471
```

Variável independente com multiplas categorias

- *married* não diferencia entre homem e mulher

Variável independente com multiplas categorias

- *married* não diferencia entre homem e mulher
- *female* não diferencia entre casado ou solteiro.

Variável independente com multiplas categorias

- *married* não diferencia entre homem e mulher
- *female* não diferencia entre casado ou solteiro.
- Vamos criar novos grupos *homemcasado*, *mulhercasada*, *homemsolteiro* e *mulhersolteira*

Variável independente com multiplas categorias

- *married* não diferencia entre homem e mulher
- *female* não diferencia entre casado ou solteiro.
- Vamos criar novos grupos *homemcasado*, *mulhercasada*, *homemsolteiro* e *mulhersolteira*

Variável independente com multiplas categorias

- *married* não diferencia entre homem e mulher
- *female* não diferencia entre casado ou solteiro.
- Vamos criar novos grupos *homemcasado*, *mulhercasada*, *homemsolteiro* e *mulhersolteira*

```
wage1 <- wage1 %>%
  mutate(homemcasado = ifelse(female == 0 & married == 1, 1, 0),
         mulhercasada = ifelse(female == 1 & married == 1, 1, 0),
        homemsolteiro = ifelse(female == 0 & married == 0, 1, 0),
       mulhersolteira = ifelse(female == 1 & married == 0, 1, 0))
```

Variável independente com multiplas categorias

$$homemcasado + mulhercasada + homemsolteiro + mulhersolteira = 1$$

Variável independente com multiplas categorias

$$homemcasado + mulhercasada + homemsolteiro + mulhersolteira = 1$$

```
table(wage1$homemcasado + wage1$mulhercasada +
      wage1$homemsolteiro + wage1$mulhersolteira)
```

```
##
##   1
## 526
```

Variável independente com multiplas categorias

$$homemcasado + mulhercasada + homemsolteiro + mulhersolteira = 1$$

```
table(wage1$homemcasado + wage1$mulhercasada +
      wage1$homemsolteiro + wage1$mulhersolteira)
```

```
##
##   1
## 526
```

Se temos p categorias, incluimos $p-1$ variáveis *dummy* no modelo!.

Variável independente com multiplas categorias

```r
modelo = lm(log(wage) ~ homemcasado + mulhercasada + mulhersolte
educ + exper + I(exper^2) +  tenure + I(tenure^2), data = wage1)
round(summary(modelo)$coefficients,5)
```

```
##                 Estimate Std. Error  t value Pr(>|t|)
## (Intercept)      0.32138    0.10001  3.21349  0.00139
## homemcasado      0.21268    0.05536  3.84188  0.00014
## mulhercasada    -0.19827    0.05784 -3.42813  0.00066
## mulhersolteira -0.11035    0.05574 -1.97966  0.04827
## educ             0.07891    0.00669 11.78733  0.00000
## exper            0.02680    0.00524  5.11183  0.00000
## I(exper^2)      -0.00054    0.00011 -4.84710  0.00000
## tenure           0.02909    0.00676  4.30161  0.00002
## I(tenure^2)     -0.00053    0.00023 -2.30555  0.02153
```

Variável independente com multiplas categorias

$$\widehat{\log(wage)} = 0.32138 + 0.21268\text{hc} - 0.19827\text{mc} - 0.11035\text{ms} + \dots$$

Quando temos várias categorias, devemos saber qual é o grupo base

- Grupo base: homem-solteiro
- Homens casados ganham em média 21.27% mais do que homens solteiros
- Mulheres casadas ganham em média 19.83% menos que homens solteiros
- Mulheres solteiras ganham em média 11.03% menos que homens solteiros

Variável independente com multiplas categorias

$$\widehat{\log(wage)} = 0.32138 + 0.21268\text{hc} - 0.19827\text{mc} - 0.11035\text{ms} + \ldots$$

Quando temos várias categorias, devemos saber qual é o grupo base

- Grupo base: homem-solteiro
- Homens casados ganham em média 21.27% mais do que homens solteiros
- Mulheres casadas ganham em média 19.83% menos que homens solteiros
- Mulheres solteiras ganham em média 11.03% menos que homens solteiros

Note que antes *married* não era estatisticamente significativo, mas quando separamos em grupos agora todas as categorias são estatísticamente significativas.

Variável independente ordinal

Variável independente ordinal:

- Se temos poucas categorias podemos criar variáveis dummy

Variável independente ordinal

Variável independente ordinal:

- Se temos poucas categorias podemos criar variáveis dummy
- Quando a variavel ordinal tem muitas categorias podemos trabalhar com ela como se fosse uma variável quantitativa ou recategoriza-la e então utilizar variáveis dummy

Variável independente ordinal

Variável independente ordinal:

- Se temos poucas categorias podemos criar variáveis dummy
- Quando a variavel ordinal tem muitas categorias podemos trabalhar com ela como se fosse uma variável quantitativa ou recategoriza-la e então utilizar variáveis dummy
- Trade-off entre interpretabilidade e proporção da variavel dependente sendo explicada.

Efeito de interação

Efeito de interação

- No exemplo anterior, tivemos que criar as variáveis *homemcasado*, *mulhercasada*, *homemsolteiro* e *mulhersolteira* pois as variáveis *married* e *female* separadas, não consideravam o efeito de alguma **interação**

Efeito de interação

- No exemplo anterior, tivemos que criar as variáveis *homemcasado*, *mulhercasada*, *homemsolteiro* e *mulhersolteira* pois as variáveis *married* e *female* separadas, não consideravam o efeito de alguma **interação**
- Efeitos de interação são interessante quando suspeitamos que o efeito parcial de uma variável explicativa depente da magnitude de **outra** variável explicativa.

Efeito de interação

- No exemplo anterior, tivemos que criar as variáveis *homemcasado*, *mulhercasada*, *homemsolteiro* e *mulhersolteira* pois as variáveis *married* e *female* separadas, não consideravam o efeito de alguma **interação**
- Efeitos de interação são interessante quando suspeitamos que o efeito parcial de uma variável explicativa depente da magnitude de **outra** variável explicativa.
- Analisaremos casos de efeitos de interação entre

Efeito de interação

- No exemplo anterior, tivemos que criar as variáveis *homemcasado*, *mulhercasada*, *homemsolteiro* e *mulhersolteira* pois as variáveis *married* e *female* separadas, não consideravam o efeito de alguma **interação**
- Efeitos de interação são interessante quando suspeitamos que o efeito parcial de uma variável explicativa depente da magnitude de **outra** variável explicativa.
- Analisaremos casos de efeitos de interação entre
  - 2 variáveis qualitativas

Efeito de interação

- No exemplo anterior, tivemos que criar as variáveis *homemcasado*, *mulhercasada*, *homemsolteiro* e *mulhersolteira* pois as variáveis *married* e *female* separadas, não consideravam o efeito de alguma **interação**
- Efeitos de interação são interessante quando suspeitamos que o efeito parcial de uma variável explicativa depente da magnitude de **outra** variável explicativa.
- Analisaremos casos de efeitos de interação entre
  - 2 variáveis qualitativas
  - 1 qualitativa e 1 quantitativa

Efeito de interação

- No exemplo anterior, tivemos que criar as variáveis *homemcasado*, *mulhercasada*, *homemsolteiro* e *mulhersolteira* pois as variáveis *married* e *female* separadas, não consideravam o efeito de alguma **interação**
- Efeitos de interação são interessante quando suspeitamos que o efeito parcial de uma variável explicativa depente da magnitude de **outra** variável explicativa.
- Analisaremos casos de efeitos de interação entre
  - 2 variáveis qualitativas
  - 1 qualitativa e 1 quantitativa
  - 2 quantitativas

Efeito de interação

```r
modelo = lm(log(wage)~ married + female + educ + exper +
         I(exper^2) + tenure + I(tenure^2), data = wage1)
round(summary(modelo)$coefficients,5)
```

```
##              Estimate Std. Error  t value Pr(>|t|)
## (Intercept)  0.41778    0.09887  4.22575  0.00003
## married      0.05292    0.04076  1.29850  0.19469
## female      -0.29018    0.03611 -8.03565  0.00000
## educ         0.07915    0.00680 11.63989  0.00000
## exper        0.02695    0.00533  5.06095  0.00000
## I(exper^2)  -0.00054    0.00011 -4.81350  0.00000
## tenure       0.03130    0.00685  4.56999  0.00001
## I(tenure^2) -0.00057    0.00023 -2.44753  0.01471
```

Efeito de interação

```
modelo = lm(log(wage)~ married + female + married*female +
educ + exper +  I(exper^2) + tenure + I(tenure^2), data = wage1)
round(summary(modelo)$coefficients,5)
```

```
##                 Estimate Std. Error  t value Pr(>|t|)
## (Intercept)      0.32138    0.10001  3.21349  0.00139
## married          0.21268    0.05536  3.84188  0.00014
## female          -0.11035    0.05574 -1.97966  0.04827
## educ             0.07891    0.00669 11.78733  0.00000
## exper            0.02680    0.00524  5.11183  0.00000
## I(exper^2)      -0.00054    0.00011 -4.84710  0.00000
## tenure           0.02909    0.00676  4.30161  0.00002
## I(tenure^2)     -0.00053    0.00023 -2.30555  0.02153
## married:female -0.30059    0.07177 -4.18846  0.00003
```

Efeito de interação

```
coef(modelo)
```

```
##    (Intercept)        married         female           educ
##   0.3213780953   0.2126756752  -0.1103502102   0.0789102812
##     I(exper^2)         tenure    I(tenure^2) married:female
##  -0.0005352452   0.0290875220  -0.0005331425  -0.3005930681
```

Efeito de interação

```
coef(modelo)
```

```
##    (Intercept)        married         female           educ
##   0.3213780953   0.2126756752  -0.1103502102   0.0789102812
##      I(exper^2)        tenure   I(tenure^2) married:female
##  -0.0005352452   0.0290875220  -0.0005331425  -0.3005930681
```

- Qual é o *grupo base*? female = 0 e married = 0 (homem solteiro)

Efeito de interação

```
coef(modelo)
```

```
##     (Intercept)       married        female          educ
##    0.3213780953  0.2126756752 -0.1103502102  0.0789102812
##      I(exper^2)        tenure  I(tenure^2) married:female
##   -0.0005352452  0.0290875220 -0.0005331425 -0.3005930681
```

- Qual é o *grupo base*? female $=0$ e married $=0$ (homem solteiro)
- Uma mulher solteira (female $=1$ e married $=0$) ganha 11% menos do que um homem solteiro

Efeito de interação

```
coef(modelo)
```

```
##     (Intercept)       married        female          educ
##    0.3213780953  0.2126756752 -0.1103502102  0.0789102812
##      I(exper^2)        tenure   I(tenure^2) married:female
##   -0.0005352452  0.0290875220 -0.0005331425 -0.3005930681
```

- Qual é o *grupo base*? female $= 0$ e married $= 0$ (homem solteiro)
- Uma mulher solteira (female $= 1$ e married $= 0$) ganha 11% menos do que um homem solteiro
- Uma mulher casada (female $= 1$ e married $= 1$) ganha 19.8% (0.2126-0.1103-0.3005) menos do que um homem solteiro

Efeito de interação

```
coef(modelo)
```

```
##     (Intercept)          married          female            educ
##    0.3213780953     0.2126756752   -0.1103502102    0.0789102812
##      I(exper^2)           tenure    I(tenure^2)  married:female
##   -0.0005352452     0.0290875220   -0.0005331425   -0.3005930681
```

- Qual é o *grupo base*? female $=0$ e married $=0$ (homem solteiro)
- Uma mulher solteira (female $=1$ e married $=0$) ganha 11% menos do que um homem solteiro
- Uma mulher casada (female $=1$ e married $=1$) ganha 19.8% (0.2126-0.1103-0.3005) menos do que um homem solteiro
- Um homem casado (female $=0$ e married $=1$) ganha 21% a mais do que um homem solteiro

Efeito de interação

Interação também pode acontecer entre uma dummy e uma quantitativa

```
modelo = lm(log(wage)~female + educ + female*educ, data = wage1)
round(coef(modelo),5)
```

```
## (Intercept)      female        educ female:educ
##     0.82595    -0.36006     0.07723    -0.00006
```

Como mudam as retas de regressão?

- Intercepto para o homem (female $= 0$) é 0.82595

Efeito de interação

Interação também pode acontecer entre uma dummy e uma quantitativa

```
modelo = lm(log(wage)~female + educ + female*educ, data = wage1)
round(coef(modelo),5)
```

```
## (Intercept)      female        educ female:educ
##     0.82595    -0.36006     0.07723    -0.00006
```

Como mudam as retas de regressão?

- Intercepto para o homem (female $=0$) é 0.82595
- Intercepto para a mulher é 0.82595-0.36006 $=$ 0.46589

Efeito de interação

Interação também pode acontecer entre uma dummy e uma quantitativa

```
modelo = lm(log(wage)~female + educ + female*educ, data = wage1)
round(coef(modelo),5)
```

```
## (Intercept)      female        educ female:educ
##     0.82595    -0.36006     0.07723    -0.00006
```

Como mudam as retas de regressão?

- Intercepto para o homem (female $=0$) é 0.82595
- Intercepto para a mulher é 0.82595-0.36006 $=$ 0.46589
- Inclinação para o homem (female $=0$) é 0.07723

Efeito de interação

Interação também pode acontecer entre uma dummy e uma quantitativa

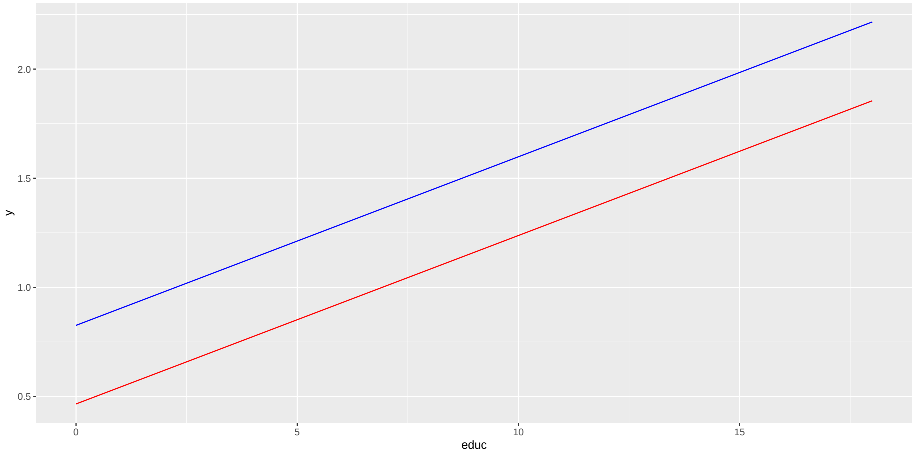
```
modelo = lm(log(wage)~female + educ + female*educ, data = wage1)
round(coef(modelo),5)
```

```
## (Intercept)      female        educ female:educ
##     0.82595    -0.36006     0.07723    -0.00006
```

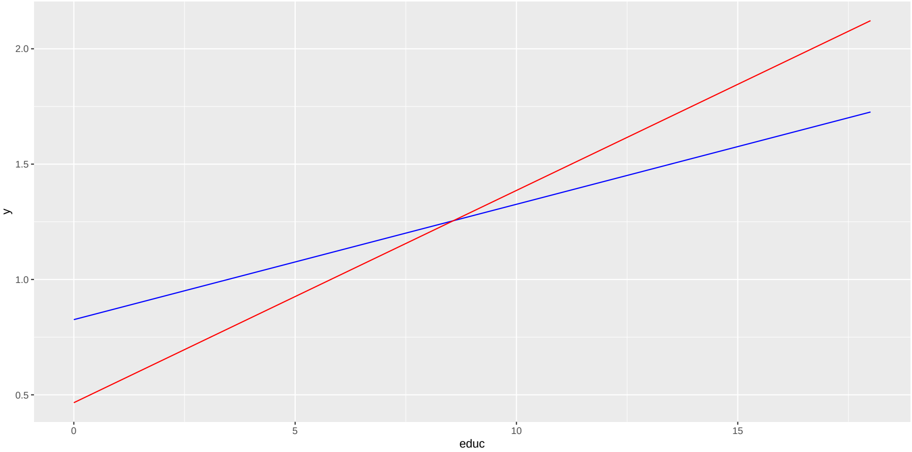
Como mudam as retas de regressão?

- Intercepto para o homem (female = 0) é 0.82595
- Intercepto para a mulher é 0.82595-0.36006 = 0.46589
- Inclinação para o homem (female = 0) é 0.07723
- Inclinação para a mulher é 0.07723-0.00006 = 0.07717

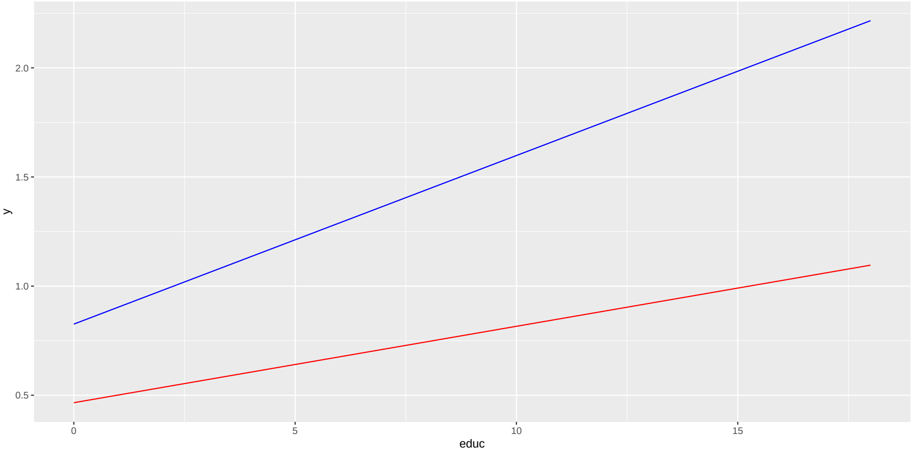
Efeito de interação



Efeito de interação



Efeito de interação



Efeito de interação

Duas variáveis quantitativas

```
modelo = lm(log(wage)~female + educ + exper+ tenure + exper*tenur
round(summary(modelo)$coefficients,5)
```

```
##                Estimate Std. Error  t value Pr(>|t|)
## (Intercept)     0.48172    0.10019  4.80820    0e+00
## female         -0.29638    0.03660 -8.09802    0e+00
## educ            0.08293    0.00689 12.03537    0e+00
## exper           0.00765    0.00173  4.40802    1e-05
## tenure          0.04772    0.00738  6.46207    0e+00
## exper:tenure   -0.00097    0.00022 -4.47592    1e-05
```

```
summary(modelo)$adj.r.squared
```

```
## [1] 0.4091884
```

Efeito de interação

$$\widehat{\log(wage)} = 0.4817 - 0.2964 female + 0.0829 educ + 0.0076 exper + 0.0477 tenure - 0.0010 exper \times tenure$$

- $\Delta \log(wage) = 0.0076 \Delta exper - 0.0010 tenure \Delta exper$

Efeito de interação

$$\widehat{\log(wage)} = 0.4817 - 0.2964 female + 0.0829 educ + 0.0076 exper + 0.0477 tenure - 0.0010 exper \times tenure$$

- $\Delta \log(wage) = 0.0076 \Delta exper - 0.0010 tenure \Delta exper$
- $\dfrac{\%\Delta wage}{\Delta exper} \approx \dfrac{100 \Delta log(wage)}{\Delta exper} = 0.76 - 0.1 tenure$

Efeito de interação

$\widehat{\log(wage)} = 0.4817 - 0.2964 female + 0.0829 educ + 0.0076 exper + 0.0477 tenure - 0.0010 exper \times tenure$

- $\Delta \log(wage) = 0.0076 \Delta exper - 0.0010 tenure \Delta exper$
- $\dfrac{\%\Delta wage}{\Delta exper} \approx \dfrac{100\Delta log(wage)}{\Delta exper} = 0.76 - 0.1 tenure$
- $\%\Delta wage = (0.76 - 0.1 tenure)\Delta exper$

Resumo

- Variáveis qualitativas podem ser incluimos no modelo
- O processo de estimação não muda
- A interpretação dos betas muda
- Quando temos uma variável categórica com p categorias, utilizamos $p-1$ variáveis *dummy*
- Quando a variável dependente é $\log(y)$ a variável *dummy* tem uma interpretação percentual
- Efeitos de interação precisam ser analisados cuidadosamente

Leituras recomendadas

Leituras recomendadas

- Wooldridge, Jeffrey M. *Introdução à Econometria: Uma abordagem moderna*. (2016). Cengage Learning. – **Cap 6.2** e **Cap 7**
- Hansen, Bruce. *Econometrics*. (2020). – **Sec 7.14** e **Sec 7.15**