

ANÁLISE DE DADOS MULTIVARIADOS I - REGRESSÃO

(AULA 04)

Novembro e dezembro de 2018

Reinaldo Soares de Camargo

Modelos de Regressão

```
> mod1 <- lm(renda_per_capita ~ esperanza_vida_ao_nascer + IDHM_educacao  
+ , data = dados)  
> summary(mod1)
```

Call:
lm(formula = renda_per_capita ~ esperanza_vida_ao_nascer + IDHM_educacao,
data = dados)

Residuals:

| Min | 1Q | Median | 3Q | Max |
|---------|--------|--------|-------|---------|
| -358.85 | -82.99 | -11.48 | 63.59 | 1113.90 |

Coefficients:

| | Estimate | Std. Error | t value | Pr(> t) |
|--------------------------|------------|------------|---------|----------|
| (Intercept) | -3182.2559 | 56.1557 | -56.67 | <2e-16 |
| esperanca_vida_ao_nascer | 40.8346 | 0.8935 | 45.70 | <2e-16 |
| IDHM_educacao | 1236.5080 | 25.6647 | 48.18 | <2e-16 |

Residual standard error: 126.8 on 5561 degrees of freedom
Multiple R-squared: 0.7283, Adjusted R-squared: 0.7282
F-statistic: 7454 on 2 and 5561 DF, p-value: < 2.2e-16

Interpretação: O aumento de uma unidade no idhm_educação aumenta a renda per capita da localidade em R\$ 1.236,50 , mantedo constante a esperança de vida ao nascer.

Modelos de Regressão

- O que representa a elevação de uma unidade no Idmh educação, vai de quanto para quanto dado que o indicador vai de 0 a 1!?

```
> summary(dados$IDHM_educacao)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
0.2070  0.4900  0.5600  0.5591  0.6310  0.8250
```

Modelos de Regressão

- Para responder vamos calcular o \hat{y} para os valores médios das variáveis explicativas; depois calcular qual seria a variação da unidade resolvendo a equação de regressão para o X2 (idmh educação), a partir dos dados da aula 3.

```
x1 <- mean(dados$esperanca_vida_ao_nascer)
x2 <- mean(dados$IDHM_educacao)
```

```
y_hat <- mod1$coefficients[1] + mod1$coefficients[2]*x1 + mod1$coefficients[3]*x2
y_hat2 <- y_hat + mod1$coefficients[3]
```

```
x2_hat <- (y_hat2 - mod1$coefficients[1] - mod1$coefficients[2]*x1)/ mod1$coefficients[3]
```

```
var_x2 <- x2_hat - x2
```

| Global Environment ▾ | |
|----------------------|--------------------------------|
| Data | |
| dados | 5564 obs. of 238 variables |
| mod1 | Large lm (12 elements, 1.5 Mb) |
| values | |
| var_x2 | Named num 1 |
| x1 | 73.0892307692308 |
| x2 | 0.559102444284687 |
| x2_hat | Named num 1.56 |
| y_hat | Named num 494 |
| y_hat2 | Named num 1730 |

A unidade é 1. Porém
essa alteração
extrapola o range do
indicador que é
menor do que 1.

Modelos de Regressão

- Exemplo numérico da interpretação do coeficiente de regressão linear múltipla.

aumentando esperança de vida em uma unidade, renda percapita aumenta em 40,835

| Renda Percapita | B0 | B1 | Esperança de vida ao nascer | B2 | IdHm Educação |
|-----------------|-----------|--------|-----------------------------|----------|---------------|
| 493,64 | -3182,256 | 40,835 | 73,0892 | 1236,508 | 0,5591 |
| 534,48 | -3182,256 | 40,835 | 74,0892 | 1236,508 | 0,5591 |
| 575,31 | -3182,256 | 40,835 | 75,0892 | 1236,508 | 0,5591 |
| 616,15 | -3182,256 | 40,835 | 76,0892 | 1236,508 | 0,5591 |
| 656,98 | -3182,256 | 40,835 | 77,0892 | 1236,508 | 0,5591 |
| 697,82 | -3182,256 | 40,835 | 78,0892 | 1236,508 | 0,5591 |

aumentando idhm educação em uma unidade, renda percapita aumenta em 1.236,508

| Renda Percapita | B0 | B1 | Esperança de vida ao nascer | B2 | IdHm Educação |
|-----------------|-----------|--------|-----------------------------|----------|---------------|
| 493,64 | -3182,256 | 40,835 | 73,0892 | 1236,508 | 0,5591 |
| 1.730,15 | -3182,256 | 40,835 | 73,0892 | 1236,508 | 1,5591 |
| 2.966,66 | -3182,256 | 40,835 | 73,0892 | 1236,508 | 2,5591 |
| 4.203,17 | -3182,256 | 40,835 | 73,0892 | 1236,508 | 3,5591 |
| 5.439,68 | -3182,256 | 40,835 | 73,0892 | 1236,508 | 4,5591 |
| 6.676,18 | -3182,256 | 40,835 | 73,0892 | 1236,508 | 5,5591 |

Modelos de Regressão

- Inclusão de variáveis qualitativas como variáveis preditoras
- Exemplo, queremos ver como a renda per capita dos municípios é afetada pela região na qual o município se localiza
- Temos cinco regiões: NO, SU, SE, NE, CO
- Precisamos transformar a informação qualitativa em informações quantitativas
- Maneira comumente utilizada:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_k x_{ki} \\ + \delta_1 D_{SU} + \delta_2 D_{NO} + \delta_3 D_{SE} + \delta_4 D_{NE} + \delta_5 D_{CO} + \epsilon_i$$

- Na equação acima, novos parâmetros a serem estimados: $\delta_1, \delta_2, \delta_3, \delta_4, \delta_5$
- As variáveis $D_{SU}, D_{SE}, D_{NO}, D_{NE}, D_{CO}$ são chamadas **variáveis dummy**

Modelos de Regressão

- As variáveis *dummy* $D_{SU}, D_{SE}, D_{NO}, D_{NE}, D_{CO}$ são definidas como:
 - Caso o município da observação i esteja contido na região Sul, então o valor de $D_{SU}=1$, e os valores das demais variáveis *dummy* será zero
 - Caso o município da observação i esteja contido na região Nordeste, então o valor de $D_{NE}=1$, e os valores das demais variáveis *dummy* será zero
 - E assim por diante ...
- Problema: não podemos incluir todas as variáveis *dummy* na regressão ao mesmo tempo, adicionalmente ao intercepto β_0
 - (1) Se mantivermos o intercepto, temos que retirar uma das variáveis *dummy*
 - (2) Se mantivermos todas as variáveis *dummy*, precisamos retirar o intercepto
- Observação:
 - Essas exclusões alternativas são feitas para evitarmos problemas de **multicolinearidade** (perfeita)
 - Dependendo da alternativa (1) ou (2) acima, a interpretação dos parâmetros muda

Modelos de Regressão

- Especificações alternativas:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_k x_{ki} \\ + \delta_1 D_{SU} + \delta_2 D_{NO} + \delta_3 D_{SE} + \delta_4 D_{NE} + \epsilon_i$$

- Ou:

$$y_i = \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_k x_{ki} \\ + \delta_1 D_{SU} + \delta_2 D_{NO} + \delta_3 D_{SE} + \delta_4 D_{NE} + \delta_5 D_{CO} + \epsilon_i$$

- Para fins de previsão, as duas especificações retornam resultados idênticos
- É possível incluir mais de uma variável qualitativa na regressão, sempre atentando para problemas de multicolinearidade
- A primeira especificação é mais utilizada

Modelos de Regressão

- Exemplo:

$$[Salário]_i = \beta_0 + \beta_1 [Experiência]_i + \delta_1 [DummyMulher]_i + \epsilon_i$$

Grupo Base: Masculino

- Para um indivíduo do sexo masculino, a equação se torna:

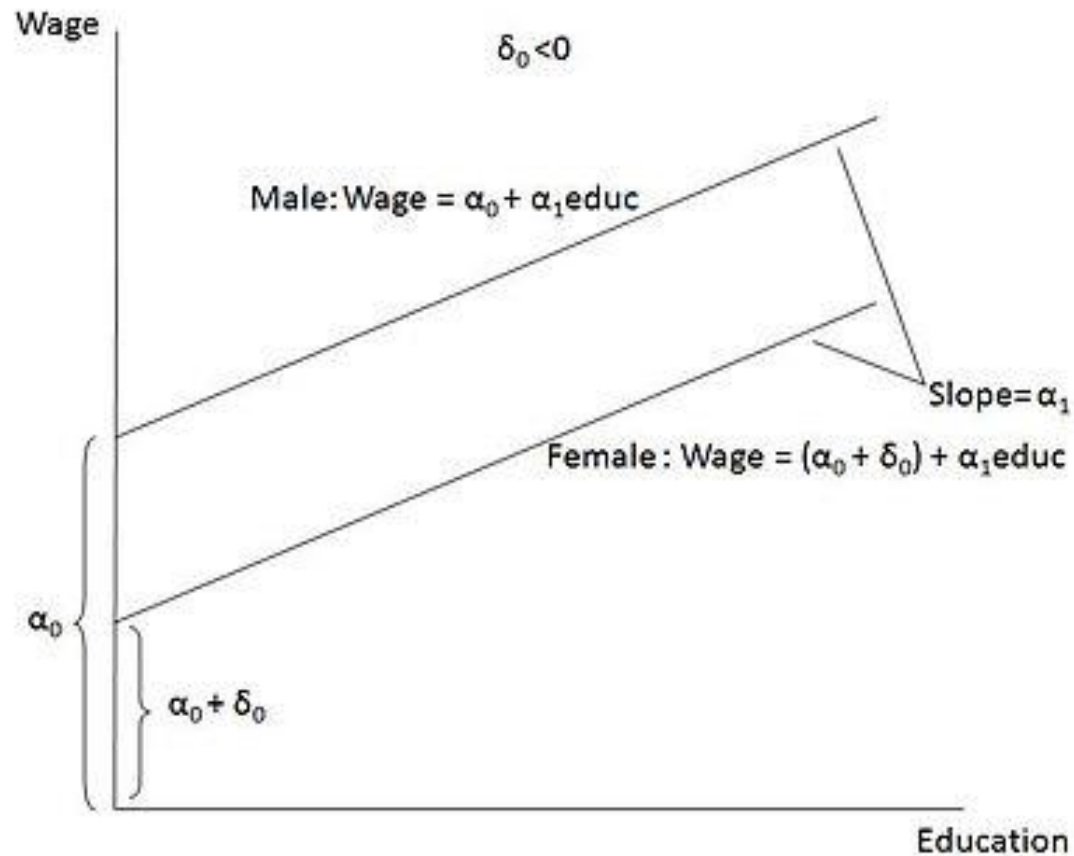
$$[Salário]_i = \beta_0 + \beta_1 [Experiência]_i + \epsilon_i$$

- Para um indivíduo do sexo feminino, a equação se torna:

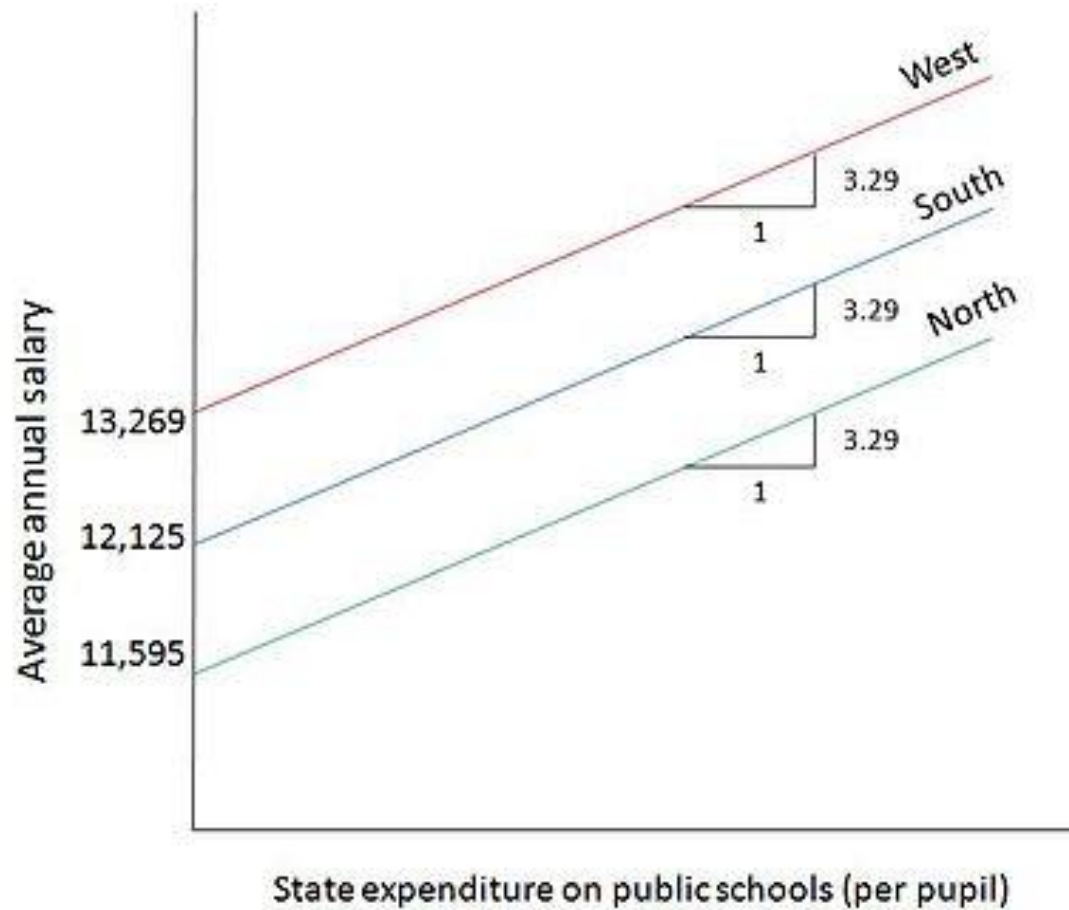
$$[Salário]_i = (\beta_0 + \delta_1) + \beta_1 [Experiência]_i + \epsilon_i$$

- Note que a diferença está no **intercepto**: β_0 para os homens e $(\beta_0 + \delta_1)$ para as mulheres

Modelos de Regressão

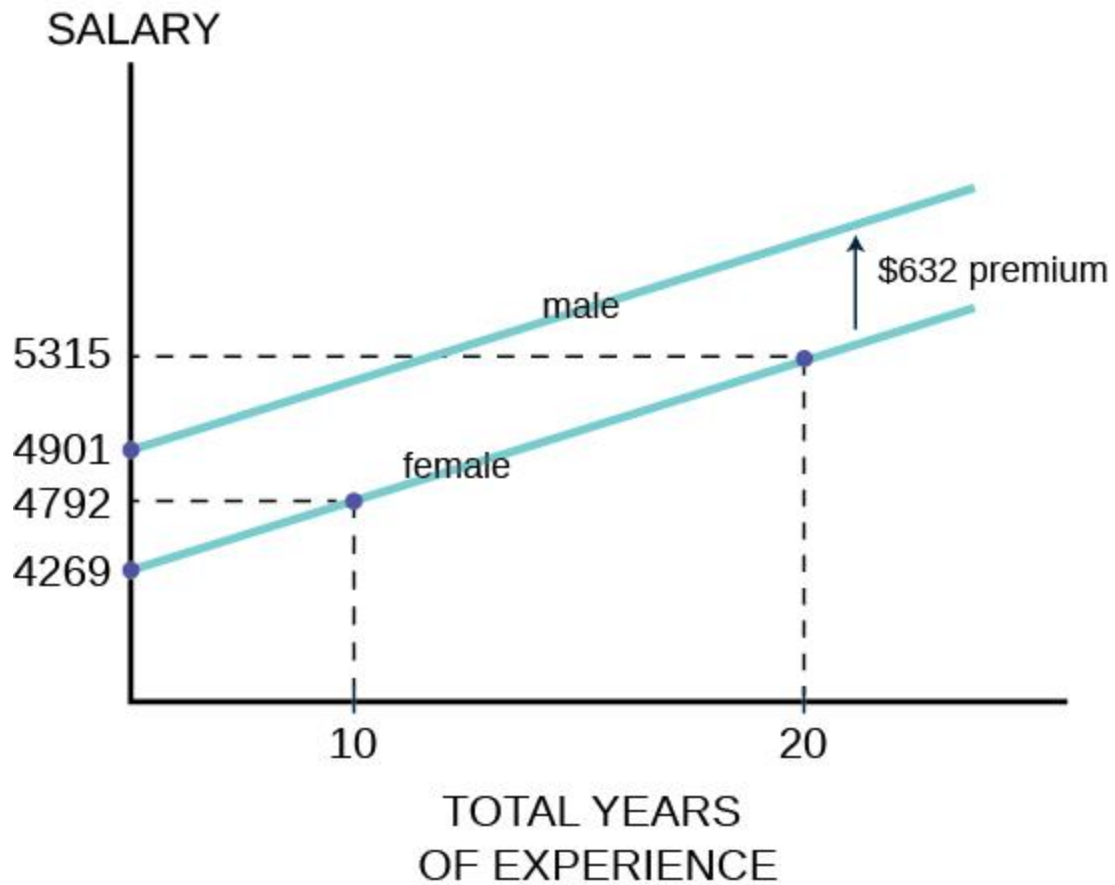


Modelos de Regressão



Modelos de Regressão

TEACHER'S SALARY



Modelos de Regressão

- Variáveis dummies de intercepto.
- Para captar o reflexo da variação do intercepto de acordo com a presença de dada característica. Por exemplo: Para captar o diferencial de preço de uma residência numa vizinhança desejável, incluiríamos a vizinhança como uma *dummy*.

$$\text{Preço} = \beta_0 + \delta_1 \text{Vizinhança} + \beta_1 \text{Tamanho} + \text{outros fatores.}$$

- Se a vizinhança for desejável (Vizinhança =1) o intercepto (ou preço padrão) será ($\beta_0 + \delta_1$).
- Caso contrário o intercepto será β_0 . O parâmetro δ_1 reflete o diferencial de preço entre as vizinhanças.
- Na realidade ao incluir dummies estaremos estimando vários modelos de regressão, um para cada categoria.

$$\text{Preço} = \beta_0 + \delta_1 \text{Vizinhança} + \beta_1 \text{Tamanho} + \text{outros fatores. (Vizinhança = 1)}$$

$$\text{Preço} = \beta_0 + \beta_1 \text{Tamanho} + \text{outros fatores. (Vizinhança = 0)}$$

Modelos de Regressão

- **Aplicação em R**

- Consideremos o modelo:

$$\text{Salar}ioh = \beta_0 + \delta_0 \text{Feminino} + \beta_1 \text{Educação} + \beta_2 \text{Experiência} + \beta_3 \text{Permanencia} + \varepsilon$$

- Importe os dados da planilha: wage1.csv;
- Estime o modelo.
- Qual o grupo de referência (ou grupo base) neste modelo?
- Os coeficientes estimados são estatisticamente significantes com 1% e 5%?
- De acordo com o modelo estimado é possível suspeitar de algum tipo de discriminação salarial?
- Em conjunto os coeficientes estimados são significantes com 5%?

Modelos de Regressão – Variáveis dummy

```
call:
lm(formula = salarioh ~ feminino + educacao + experiencia + permanencia,
    data = dados)
```

Residuals:

| | Min | 1Q | Median | 3Q | Max |
|--|---------|---------|---------|--------|---------|
| | -7.7675 | -1.8080 | -0.4229 | 1.0467 | 14.0075 |

Coefficients:

| | Estimate | Std. Error | t value | Pr(> t) | |
|-------------|----------|------------|---------|----------|-----|
| (Intercept) | -1.56794 | 0.72455 | -2.164 | 0.0309 | * |
| feminino | -1.81085 | 0.26483 | -6.838 | 2.26e-11 | *** |
| educacao | 0.57150 | 0.04934 | 11.584 | < 2e-16 | *** |
| experiencia | 0.02540 | 0.01157 | 2.195 | 0.0286 | * |
| permanencia | 0.14101 | 0.02116 | 6.663 | 6.83e-11 | *** |

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.958 on 521 degrees of freedom

Multiple R-squared: 0.3635, Adjusted R-squared: 0.3587

F-statistic: 74.4 on 4 and 521 DF, p-value: < 2.2e-16

Os sinais dos coeficientes estão de acordo com o esperado?

Modelos de Regressão – Variáveis dummy

```
call:
lm(formula = salarioh ~ feminino + educacao + experiencia + permanencia,
    data = dados)

Residuals:
    Min       1Q   Median       3Q      Max
-7.7675 -1.8080 -0.4229  1.0467 14.0075

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)  -1.56794     0.72455   -2.164   0.0309 *
feminino      -1.81085     0.26483  -6.838 2.26e-11 ***
educacao       0.57150     0.04934  11.584 < 2e-16 ***
experiencia    0.02540     0.01157   2.195   0.0286 *
permanencia    0.14101     0.02116   6.663 6.83e-11 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.958 on 521 degrees of freedom
Multiple R-squared:  0.3635,    Adjusted R-squared:  0.3587
F-statistic: 74.4 on 4 and 521 DF,  p-value: < 2.2e-16
```

Grupo de referência ou grupo base: Masculino.

Modelos de Regressão – Variáveis dummy

```
call:  
lm(formula = salarioh ~ feminino + educacao + experiencia + permanencia,  
    data = dados)
```

Residuals:

| | Min | 1Q | Median | 3Q | Max |
|--|---------|---------|---------|--------|---------|
| | -7.7675 | -1.8080 | -0.4229 | 1.0467 | 14.0075 |

Coefficients:

| | Estimate | Std. Error | t value | Pr(> t) | |
|-------------|----------|------------|---------|----------|-----|
| (Intercept) | -1.56794 | 0.72455 | -2.164 | 0.0309 | * |
| feminino | -1.81085 | 0.26483 | -6.838 | 2.26e-11 | *** |
| educacao | 0.57150 | 0.04934 | 11.584 | < 2e-16 | *** |
| experiencia | 0.02540 | 0.01157 | 2.195 | 0.0286 | * |
| permanencia | 0.14101 | 0.02116 | 6.663 | 6.83e-11 | *** |

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.958 on 521 degrees of freedom

Multiple R-squared: 0.3635, Adjusted R-squared: 0.3587

F-statistic: 74.4 on 4 and 521 DF, p-value: < 2.2e-16

Coeficientes estimados estatisticamente significantes com 1%
(feminino, educação, permanência); com 5% todos.

Modelos de Regressão – Variáveis dummy

```
call:
lm(formula = salarioh ~ feminino + educacao + experiencia + permanencia,
    data = dados)

Residuals:
    Min       1Q   Median       3Q      Max
-7.7675 -1.8080 -0.4229  1.0467 14.0075

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -1.56794    0.72455  -2.164   0.0309 *
feminino     -1.81085    0.26483  -6.838 2.26e-11 ***
educacao     0.57130    0.04934  11.584 < 2e-16 ***
experiencia  0.02540    0.01157   2.195  0.0286 *
permanencia  0.14101    0.02116   6.663 6.83e-11 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.958 on 521 degrees of freedom
Multiple R-squared:  0.3635,    Adjusted R-squared:  0.3587
F-statistic: 74.4 on 4 and 521 DF,  p-value: < 2.2e-16
```

Salário pago para o gênero feminino inferior ao pago ao masculino em média R\$ 1,81 por hora, pode indicar discriminação!!??

Modelos de Regressão – Variáveis dummy

```
call:
lm(formula = salarioh ~ feminino + educacao + experiencia + permanencia,
    data = dados)

Residuals:
    Min       1Q   Median       3Q      Max
-7.7675 -1.8080 -0.4229  1.0467 14.0075

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -1.56794    0.72455  -2.164   0.0309 *
feminino     -1.81085    0.26483  -6.838 2.26e-11 ***
educacao     0.57130    0.04934  11.584 < 2e-16 ***
experiencia  0.02540    0.01157   2.195   0.0286 *
permanencia  0.14101    0.02116   6.663 6.83e-11 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.958 on 521 degrees of freedom
Multiple R-squared:  0.3635,    Adjusted R-squared:  0.3587
F-statistic: 74.4 on 4 and 521 DF,  p-value: < 2.2e-16
```

Interpretação dos coeficientes: Genero femino recebe em média – R\$ 1,81 por h, ceteris paribus.

Modelos de Regressão – Variáveis dummy

```
call:
lm(formula = salarioh ~ feminino + educacao + experiencia + permanencia,
    data = dados)

Residuals:
    Min       1Q   Median       3Q      Max
-7.7675 -1.8080 -0.4229  1.0467 14.0075

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  -1.56794    0.72455  -2.164   0.0309 *
feminino      -1.81085    0.26483  -6.838 2.26e-11 ***
educacao       0.57150    0.04934  11.584 < 2e-16 ***
experiencia   0.02540    0.01157   2.195  0.0286 *
permanencia   0.14101    0.02116   6.663 6.83e-11 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.958 on 521 degrees of freedom
Multiple R-squared:  0.3635,    Adjusted R-squared:  0.3587
F-statistic: 74.4 on 4 and 521 DF,  p-value: < 2.2e-16
```

Interpretação dos coeficientes: Aumento de um ano na educação aumenta o salário horário em R\$ 0,57 , ceteris paribus.

Modelos de Regressão – Variáveis dummy

```
call:
lm(formula = salarioh ~ feminino + educacao + experiencia + permanencia,
    data = dados)

Residuals:
    Min       1Q   Median       3Q      Max
-7.7675 -1.8080 -0.4229  1.0467 14.0075

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  -1.56794    0.72455   -2.164   0.0309 *
feminino      -1.81085    0.26483  -6.838 2.26e-11 ***
educacao       0.57150    0.04934  11.584 < 2e-16 ***
experiencia   0.02540    0.01157   2.195   0.0286 *
permanencia   -0.14101    0.02116   6.663 6.83e-11 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.958 on 521 degrees of freedom
Multiple R-squared:  0.3635,    Adjusted R-squared:  0.3587
F-statistic: 74.4 on 4 and 521 DF,  p-value: < 2.2e-16
```

Interpretação dos coeficientes: Aumento de um ano na experiencia aumenta o salario horário em R\$ 0,52 , ceteris paribus.

Modelos de Regressão – Variáveis dummy

```
call:  
lm(formula = salarioh ~ feminino + educacao + experiencia + permanencia,  
    data = dados)
```

Residuals:

| | Min | 1Q | Median | 3Q | Max |
|--|---------|---------|---------|--------|---------|
| | -7.7675 | -1.8080 | -0.4229 | 1.0467 | 14.0075 |

Coefficients:

| | Estimate | Std. Error | t value | Pr(> t) | |
|-------------|----------|------------|---------|----------|-----|
| (Intercept) | -1.56794 | 0.72455 | -2.164 | 0.0309 | * |
| feminino | -1.81085 | 0.26483 | -6.838 | 2.26e-11 | *** |
| educacao | 0.57150 | 0.04934 | 11.584 | < 2e-16 | *** |
| experiencia | 0.02346 | 0.01157 | 2.195 | 0.0286 | * |
| permanencia | 0.14101 | 0.02116 | 6.663 | 6.83e-11 | *** |

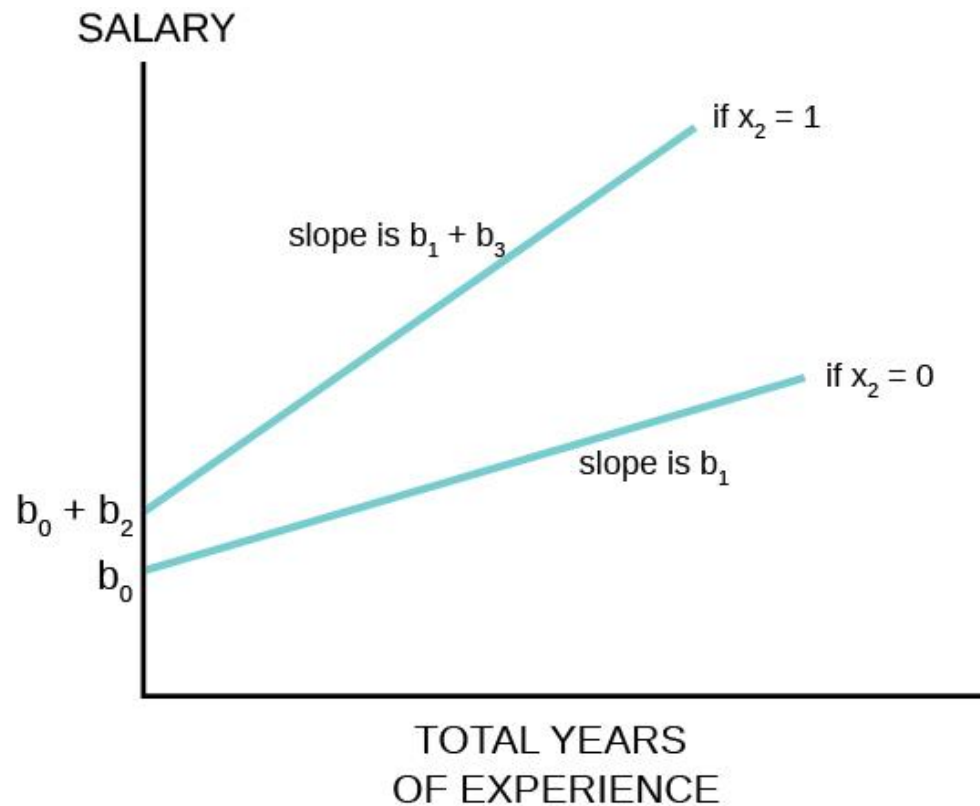
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.958 on 521 degrees of freedom
Multiple R-squared: 0.3635, Adjusted R-squared: 0.3587
F-statistic: 74.4 on 4 and 521 DF, p-value: < 2.2e-16

Interpretação dos coeficientes: cada ano de permanência no emprego aumenta o salário horário em R\$ 0,14 , ceteris paribus.

Modelos de Regressão

E se quisermos que o coeficiente dos anos de experiência variem entre homens e mulheres?



$$\hat{y} = b_0 + b_2x_2 + b_1x_1 + b_3x_2x_1$$

Modelos de Regressão

- Alteração no modelo com dummy para mulheres:

$$[Salário]_i = \beta_0 + \beta_1 [Experiência]_i + \delta_1 [DummyMulher]_i + \gamma_1 [DummyMulher]_i \times [Experiência]_i + \epsilon_i$$

- Para um indivíduo do sexo masculino, a equação se torna:

$$[Salário]_i = \beta_0 + \beta_1 [Experiência]_i + \epsilon_i$$

- Para um indivíduo do sexo feminino, a equação se torna:

$$[Salário]_i = (\beta_0 + \delta_1) + (\beta_1 + \gamma_1) [Experiência]_i + \epsilon_i$$

- Diferença no **intercepto**: β_0 para os homens e $(\beta_0 + \delta_1)$ para as mulheres
- Diferença no **coeficiente** da variável anos de experiência: β_1 versus $(\beta_1 + \gamma_1)$

Modelos de Regressão

- **Aplicação em R**

- Consideremos o modelo:

$$\text{Salario}_i = \beta_0 + \delta_0 \text{Feminino}_i + \beta_1 \text{Educação}_i + \beta_2 \text{Experiência}_i + \beta_3 \text{Permanencia}_i + \gamma_1 (\text{Experiência}_i * \text{Feminino}_i) + \varepsilon_i$$

- Estime o modelo.
- Os coeficientes estimados são estatisticamente significantes com 1% e 5%?
 - Em conjunto os coeficientes estimados são significantes com 5%?
 - Com base nos resultados dessa nova equação, como o efeito da experiência, sobre salário, se altera de acordo com o gênero?
 - Houve melhora do R^2 ajustado em relação ao modelo anterior?

Modelos de Regressão – Variáveis dummy

```
call:
lm(formula = salarioh ~ feminino + educacao + experiencia + experiencia *
    feminino + permanencia, data = dados)
```

Residuals:

| Min | 1Q | Median | 3Q | Max |
|---------|---------|---------|--------|---------|
| -8.3215 | -1.6447 | -0.4678 | 1.0431 | 13.8889 |

Coefficients:

| | Estimate | Std. Error | t value | Pr(> t) | |
|----------------------|----------|------------|---------|----------|-----|
| (Intercept) | -2.27347 | 0.75952 | -2.993 | 0.002891 | ** |
| feminino | -0.87766 | 0.41570 | -2.111 | 0.035223 | * |
| educacao | 0.58954 | 0.04938 | 11.938 | < 2e-16 | *** |
| experiencia | 0.05720 | 0.01589 | 3.601 | 0.000348 | *** |
| permanencia | 0.12810 | 0.02148 | 5.964 | 4.56e-09 | *** |
| feminino:experiencia | -0.05635 | 0.01944 | -2.898 | 0.003908 | ** |

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.937 on 520 degrees of freedom
Multiple R-squared: 0.3737, Adjusted R-squared: 0.3676
F-statistic: 62.04 on 5 and 520 DF, p-value: < 2.2e-16

Os coeficientes estimados são estatisticamente significantes a 5%.

Modelos de Regressão – Variáveis dummy

```
call:
lm(formula = salarioh ~ feminino + educacao + experiencia + experiencia *
    feminino + permanencia, data = dados)
```

Residuals:

| Min | 1Q | Median | 3Q | Max |
|---------|---------|---------|--------|---------|
| -8.3215 | -1.6447 | -0.4678 | 1.0431 | 13.8889 |

Coefficients:

| | Estimate | Std. Error | t value | Pr(> t) | |
|----------------------|----------|------------|---------|----------|-----|
| (Intercept) | -2.27347 | 0.75952 | -2.993 | 0.002891 | ** |
| feminino | -0.87766 | 0.41570 | -2.111 | 0.035223 | * |
| educacao | 0.58954 | 0.04938 | 11.938 | < 2e-16 | *** |
| experiencia | 0.05720 | 0.01589 | 3.601 | 0.000348 | *** |
| permanencia | 0.12810 | 0.02148 | 5.964 | 4.56e-09 | *** |
| feminino:experiencia | -0.05635 | 0.01944 | -2.898 | 0.003908 | ** |

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.937 on 520 degrees of freedom
Multiple R-squared: 0.3737, Adjusted R-squared: 0.3676
F-statistic: 62.04 on 5 and 520 DF, p-value: < 2.2e-16

Em conjunto os coeficientes estimados são estatisticamente significantes a 5%.

Modelos de Regressão – Variáveis dummy

```
call:
lm(formula = salarioh ~ feminino + educacao + experiencia + experiencia *
  feminino + permanencia, data = dados)
```

Residuals:

| Min | 1Q | Median | 3Q | Max |
|---------|---------|---------|--------|---------|
| -8.3215 | -1.6447 | -0.4678 | 1.0431 | 13.8889 |

Coefficients:

| | Estimate | Std. Error | t value | Pr(> t) | |
|----------------------|----------|------------|---------|----------|-----|
| (Intercept) | -2.27347 | 0.75952 | -2.993 | 0.002891 | ** |
| feminino | -0.87766 | 0.41570 | -2.111 | 0.035223 | * |
| educacao | 0.58954 | 0.04938 | 11.938 | < 2e-16 | *** |
| experiencia | 0.05720 | 0.01589 | 3.601 | 0.000348 | *** |
| permanencia | 0.12810 | 0.02148 | 5.964 | 4.56e-09 | *** |
| feminino:experiencia | -0.05635 | 0.01944 | -2.898 | 0.003908 | ** |

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.937 on 520 degrees of freedom
Multiple R-squared: 0.3737, Adjusted R-squared: 0.3676
F-statistic: 62.04 on 5 and 520 DF, p-value: < 2.2e-16

O salário horário de indivíduos do sexo feminino de mesma experiência é menor em R\$ -0,05, ceteris paribus.

Modelos de Regressão – Variáveis dummy

```
call:
lm(formula = salarioh ~ feminino + educacao + experiencia + experiencia *
    feminino + permanencia, data = dados)
```

Residuals:

| Min | 1Q | Median | 3Q | Max |
|---------|---------|---------|--------|---------|
| -8.3215 | -1.6447 | -0.4678 | 1.0431 | 13.8889 |

Coefficients:

| | Estimate | Std. Error | t value | Pr(> t) | |
|----------------------|----------|------------|---------|----------|-----|
| (Intercept) | -2.27347 | 0.75952 | -2.993 | 0.002891 | ** |
| feminino | -0.87766 | 0.41570 | -2.111 | 0.035223 | * |
| educacao | 0.58954 | 0.04938 | 11.938 | < 2e-16 | *** |
| experiencia | 0.05720 | 0.01589 | 3.601 | 0.000348 | *** |
| permanencia | 0.12810 | 0.02148 | 5.964 | 4.56e-09 | *** |
| feminino:experiencia | -0.05635 | 0.01944 | -2.898 | 0.003908 | ** |

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.937 on 520 degrees of freedom
Multiple R-squared: 0.3737, Adjusted R-squared: 0.3676
F-statistic: 62.04 on 5 and 520 DF, p-value: < 2.2e-16

O R² ajustado passou de 35,7% para 36,76%. Este modelo nos parece melhor do que o anterior..

Obrigado.