# R para Economia

Lucas Mendes

24/03/2020

# Modelos Cross - Section

# Regressão Linear Simples

Lembra de sua aula de introdução à microeconomia? Tire seu livro do Mankiw do armário!

Agora pense que você irá analisar o mercado de **bananas**. Representando suas curvas de oferta e demanda

curva de demanda: $Y = \beta_d - \alpha_d X$

curva de oferta: $Y = \beta_o + \alpha_o X$

# Regressão Linear Simples

Se considerarmos que $\beta_d = 80$ e $\beta_o = 10$ sendo que $\alpha_d = 4$ e $\alpha_o = 6$

- curva de demanda: $Y = 80 - 4X$

# Regressão Linear Simples

Se considerarmos que $\beta_d = 80$ e $\beta_o = 10$ sendo que $\alpha_d = 4$ e $\alpha_o = 6$

- curva de demanda: $Y = 80 - 4X$
- curva de oferta: $Y = 10 + 6X$

# Regressão Linear Simples

Temos como agora calcular o equilibrio do mercado igualando a curva de demanda a curva de oferta

$80 - 4X = 10 + 6X$ (1)

$70 = 10X$ (2)

$7 = X$ (3)

Quandidade de equilibrio = 7

Preço de equilibrio = 52

# Regressão Linear Simples

Isso foi o que você provavelmente fez em introdução a micro ou algo do tipo

So que nessa época, o seu professor te dava os valores de $\alpha$ e $\beta$

Agora você mesmo irá calcula - los!

# Disclaimer

$$Y = \beta_1 + \beta_2 X$$

- O $Y$ pode ser chamado de varios nomes, como variavel regressora, variavel dependente, variavel resposta e por ai vai.

# Disclaimer

$$Y = \beta_1 + \beta_2 X$$

- O $Y$ pode ser chamado de varios nomes, como variavel regressora, variavel dependente, variavel resposta e por ai vai.
- Porém eu irei chama - la de variavel **endógena**, ou seja, que é determinada pelo modelo.

# Disclaimer

$$Y = \beta_1 + \beta_2 X$$

- O $Y$ pode ser chamado de varios nomes, como variavel regressora, variavel dependente, variavel resposta e por ai vai.
- Porém eu irei chama - la de variavel **endógena**, ou seja, que é determinada pelo modelo.
- A mesma coisa vale para $X$, que tem varios nomes, mas eu chamarei de varável **exógena**.

# Disclaimer

$$Y = \beta_1 + \beta_2 X$$

- O $Y$ pode ser chamado de varios nomes, como variavel regressora, variavel dependente, variavel resposta e por ai vai.
- Porém eu irei chama - la de variavel **endógena**, ou seja, que é determinada pelo modelo.
- A mesma coisa vale para $X$, que tem varios nomes, mas eu chamarei de varável **exógena**.
- O que estiver no lado esquerdo da equação = **endógena**

# Disclaimer

$$Y = \beta_1 + \beta_2 X$$

- O $Y$ pode ser chamado de varios nomes, como variavel regressora, variavel dependente, variavel resposta e por ai vai.
- Porém eu irei chama - la de variavel **endógena**, ou seja, que é determinada pelo modelo.
- A mesma coisa vale para $X$, que tem varios nomes, mas eu chamarei de varável **exógena**.
- O que estiver no lado esquerdo da equação = **endógena**
- O que estiver no lado direito da equação = **exógena**

# Regressão Linear Simples

Nesse capitulo iremos usar o pacote `AER` (Applied Econometrics with R) e o pacote caret (Machine Learning)

Cole no console e rode

```
# install.packages('AER')
# install.packages('caret')
```

```
library(AER)
library(caret)
library(tidyverse)
```

# Regressão Linear Simples

Iremos analisar agora a base de dados CPS1985, referente a pesquisa de determinação salarial feita em 1985 nos EUA.

Queremos verificar qual o impacto do total de anos de educação sobre o salario/hora de um indivíduo

Carregando o pacote

```
data('CPS1985')
```

# Regressão Linear Simples

| | wage | education | experience | age | ethnicity |
|---|---|---|---|---|---|
| 1 | 5.10 | 8 | 21 | 35 | hispanic |
| 1100 | 4.95 | 9 | 42 | 57 | cauc |
| 2 | 6.67 | 12 | 1 | 19 | cauc |
| 3 | 4.00 | 12 | 4 | 22 | cauc |
| 4 | 7.50 | 12 | 17 | 35 | cauc |
| 5 | 13.07 | 13 | 9 | 28 | cauc |

# Regressão Linear Simples

$$wage = \beta_1 + \beta_2 educ$$

Essa é nossa especificação da regressão, o código seguinte irá calcular os parâmetros $\beta_1$ e $\beta_2$

# Regressão Linear Simples

Iremos agora treinar um modelo de regressão linear usando a função `train()` do pacote **caret**

```
modelo <- train(wage ~ # Variavel Exógena
                  education, # Variavel endógena
                method = "lm", # Linear Model
                data = CPS1985) # Base de dados
```

# Regressão Linear Simples

Com o modelo criado,podemos observar as estatisticas usando o comando `summary()`.

```
summary(modelo)
```

# Regressão Linear Simples

```
##
## Call:
## lm(formula = .outcome ~ ., data = dat)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -7.911 -3.260 -0.760  2.240 34.740
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.74598    1.04545  -0.714    0.476
## education    0.75046    0.07873   9.532   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 '
##
## Residual standard error: 4.754 on 532 degrees of freedom
```

# Regressão Linear Simples

Eu particularmente não gosto muito do formato que o summary nos retorna.

Como eu sigo a filosofia do tidyverse, eu transformo isso para um dataframe com a função `tidy()` do pacote `broom` (Já instalado com tidyverse)

```r
library(broom)
summary(modelo) %>% tidy()
```

| term | estimate | std.error | statistic | p.value |
|---|---|---|---|---|
| (Intercept) | -0.7459797 | 1.0454541 | -0.7135461 | 0.4758208 |
| education | 0.7504608 | 0.0787337 | 9.5316300 | 0.0000000 |

# Regressão Linear Simples

$$wage = -0.74 + educ0.75$$

O que podemos retirar do modelo e das estatísticas?

Normalmente olhamos para:

- O coefieciente das variaveis

# Regressão Linear Simples

$$wage = -0.74 + educ0.75$$

O que podemos retirar do modelo e das estatísticas?

Normalmente olhamos para:

- O coefieciente das variaveis
- O valor $t$ dessas variaveis

# Regressão Linear Simples

$$wage = -0.74 + educ0.75$$

O que podemos retirar do modelo e das estatísticas?

Normalmente olhamos para:

- O coefieciente das variaveis
- O valor $t$ dessas variaveis
- O $R^2$

# Regressão Linear Simples

## Coeficiente

- Quando analisamos o coeficiente de uma regressão, normalmente nós esperamos o seu sinal devido a uma teoria prévia.

# Regressão Linear Simples

## Coeficiente

- Quando analisamos o coeficiente de uma regressão, normalmente nós esperamos o seu sinal devido a uma teoria prévia.
- No nosso exemplo esperamos que seja positivo ja que é um consenso que mais anos de estudo impactam positivamente no salario.

# Regressão Linear Simples

## Coeficiente

- Quando analisamos o coeficiente de uma regressão, normalmente nós esperamos o seu sinal devido a uma teoria prévia.
- No nosso exemplo esperamos que seja positivo ja que é um consenso que mais anos de estudo impactam positivamente no salario.
- O que normalmente queremos testar é a magnitude do efeito de uma variavel sobre a outra.

# Regressão Linear Simples

## Coeficiente

- Quando analisamos o coeficiente de uma regressão, normalmente nós esperamos o seu sinal devido a uma teoria prévia.
- No nosso exemplo esperamos que seja positivo ja que é um consenso que mais anos de estudo impactam positivamente no salario.
- O que normalmente queremos testar é a magnitude do efeito de uma variavel sobre a outra.
- O nosso modelo nos forneceu que $\beta_2$ era 0.75

# Regressão Linear Simples

## Coeficiente

- Quando analisamos o coeficiente de uma regressão, normalmente nós esperamos o seu sinal devido a uma teoria prévia.
- No nosso exemplo esperamos que seja positivo ja que é um consenso que mais anos de estudo impactam positivamente no salario.
- O que normalmente queremos testar é a magnitude do efeito de uma variavel sobre a outra.
- O nosso modelo nos forneceu que $\beta_2$ era 0.75
- A interpretação portanto é: Se um indivíduo estuda 1 ano a mais, ele ganha em média 0.75 centavos/hora a mais de salário

# Regressão Linear Simples

## Coeficiente

- Supondo que um indivíduo A estudou 10 anos

# Regressão Linear Simples

## Coeficiente

- Supondo que um indivíduo A estudou 10 anos
- $\text{Salario}/\text{hora} = -0.74 + 10 * 0.75 = 6.76$

# Regressão Linear Simples

## Coeficiente

- Supondo que um indivíduo A estudou 10 anos
- $\text{Salario/hora} = -0.74 + 10 * 0.75 = 6.76$
- Supondo que um individuo B estudou 11 anos

# Regressão Linear Simples

## Coeficiente

- Supondo que um indivíduo A estudou 10 anos
- $\text{Salario/hora} = -0.74 + 10 * 0.75 = 6.76$
- Supondo que um individuo B estudou 11 anos
- $\text{Salario/hora} = -0.74 + 11 * 0.75 = 7.51$

# Regressão Linear Simples

## Valor T

- O valor $t$ é um valor que vem da formula $t = \frac{\beta}{EP(\beta)}$

# Regressão Linear Simples

## Valor T

- O valor $t$ é um valor que vem da formula $t = \frac{\beta}{EP(\beta)}$
- Essa pequena conta é um teste estatistico que avalia se o nosso coeficiente $\beta_i$ é diferente de zero.

# Regressão Linear Simples

## Valor T

- O valor $t$ é um valor que vem da formula $t = \frac{\beta}{EP(\beta)}$
- Essa pequena conta é um teste estatistico que avalia se o nosso coeficiente $\beta_i$ é diferente de zero.
- A regra de bolso que levamos é que se $t > |2|$, podemos rejeitar que o coeficiente é igual a zero.

# Regressão Linear Simples

## Valor T

- O valor $t$ é um valor que vem da formula $t = \frac{\beta}{EP(\beta)}$
- Essa pequena conta é um teste estatistico que avalia se o nosso coeficiente $\beta_i$ é diferente de zero.
- A regra de bolso que levamos é que se $t > |2|$, podemos rejeitar que o coeficiente é igual a zero.
- Vamos fazer a conta

# Regressão Linear Simples

## Valor T

| term | estimate | std.error | statistic | p.value |
|---|---|---|---|---|
| (Intercept) | -0.7459797 | 1.0454541 | -0.7135461 | 0.4758208 |
| education | 0.7504608 | 0.0787337 | 9.5316300 | 0.0000000 |

# Regressão Linear Simples

## $R^2$

- O $R^2$ mede o poder de explicação de uma regressão

# Regressão Linear Simples

### $R^2$

- O $R^2$ mede o poder de explicação de uma regressão
- Seus valores variam de 0 a 1.

# Regressão Linear Simples

## $R^2$

- O $R^2$ mede o poder de explicação de uma regressão
- Seus valores variam de 0 a 1.
- No nosso exemplo ele é 0.14, ou 14%

# Regressão Linear Simples

## $R^2$

- O $R^2$ mede o poder de explicação de uma regressão
- Seus valores variam de 0 a 1.
- No nosso exemplo ele é 0.14, ou 14%
- Muitos podem se enganar olhando apenas esse indicador, use o com cuidado.

# Regressão Linear Simples

**$R^2$**

$$R\check{s} = \frac{SQE}{SQT}$$

$$SQR = SQT - SQE$$

SQT = Soma dos quadrados Totais SQE = Soma dos quadrados Explicados
SQR = Soma dos quadrados dos Resíduos

# Regressão Linear Simples

## $R^2$

```r
summary(modelo) %>% glance()
```

| r.squared | adj.r.squared | sigma | statistic | p.value | df |
|---|---|---|---|---|---|
| 0.1458645 | 0.1442589 | 4.753987 | 90.85197 | 0 | 2 |

# Exercicios

# Elasticidades

# Regressão Linear Simples (Elasticidades)

Talvez você ja tenha ouvido falar sobre elasticidades, talvez até calculado na forma discreta.

Para calcular elasticidades, precisamos deixar as variaveis logarizadas usando a função `log()`

No nosso exemplo sobre educação, ficaria da seguinte maneira

```
modelo <- train(log(wage) ~ log(education),
      method = "lm",
      data = CPS1985)
```

# Regressão Linear Simples (Elasticidades)

Observando as estatisticas

```
summary(modelo) %>% broom::tidy()
```

| term | estimate | std.error | statistic | p.value |
|---|---|---|---|---|
| (Intercept) | 0.0701300 | 0.2371804 | 0.2956821 | 0.7675883 |
| log(education) | 0.7822175 | 0.0928906 | 8.4208482 | 0.0000000 |

# Regressão Linear Simples (Elasticidades)

$$log(wage) = 0.07 + 0.78 log(educ)$$

- Agora a interpretação dos coeficientes muda um pouco.

# Regressão Linear Simples (Elasticidades)

$$log(wage) = 0.07 + 0.78 log(educ)$$

- Agora a interpretação dos coeficientes muda um pouco.
- Nós lemos da seguinte maneira:

# Regressão Linear Simples (Elasticidades)

$$log(wage) = 0.07 + 0.78log(educ)$$

- Agora a interpretação dos coeficientes muda um pouco.
- Nós lemos da seguinte maneira:
- Se eu aumentar meus anos de estudo em 1%

# Regressão Linear Simples (Elasticidades)

$$log(wage) = 0.07 + 0.78 log(educ)$$

- Agora a interpretação dos coeficientes muda um pouco.
- Nós lemos da seguinte maneira:
- Se eu aumentar meus anos de estudo em 1%
- Meu salario/hora irá aumentar em média 0.78%

# Regressão Linear Simples (Elasticidades)

$$log(wage) = 0.07 + 0.78 log(educ)$$

- Agora a interpretação dos coeficientes muda um pouco.
- Nós lemos da seguinte maneira:
- Se eu aumentar meus anos de estudo em 1%
- Meu salario/hora irá aumentar em média 0.78%
- Todas as estatisticas seguem o mesmo procedimento de análise

# Regressão Linear Simples (Elasticidades)

# Logs

Logarizar os também serve para:

- Deixar relações exponenciais em lineares

$$y = x_1^{\beta_1} x_2^{\beta_2}$$

Se aplicarmos log

$$log(y) = \beta_1 log(x_1) + \beta_2 log(x_2)$$

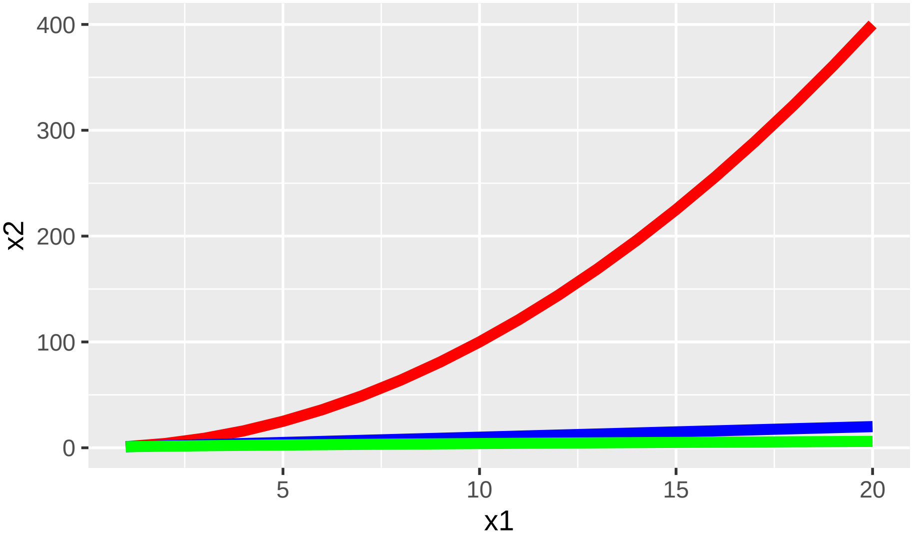
# Cobb Douglas

Testando no R retornos de escala

```
cobb_douglas <- function(x1,x2,beta1 = 0.5,beta2 = 0.5){
  x1^beta1 * x2^beta2
}

crescente <- cobb_douglas(seq(1:20),seq(1:20),beta1 = 1,beta2
constante <- cobb_douglas(seq(1:20),seq(1:20))
decrescente <- cobb_douglas(seq(1:20),seq(1:20),beta1 = 0.3,be
```

# Cobb Douglas



# Exercicios

# Regressão Linear Multipla

# Regressão Linear Multipla

A regressão linear multipla é quando estamos usando mais de uma varivel endógena.

Exemplo

$$wage = \beta_1 + \beta_2 educ + \beta_3 experience$$

# Regressão Linear Multipla

Nós usamos a mesma função no R

```
modelo <- train(wage ~ education + experience,
      method = "lm",
      data = CPS1985)
```

# Regressão Linear Multipla

```
summary(modelo) %>% broom::tidy()
```

| term | estimate | std.error | statistic | p.value |
|---|---|---|---|---|
| (Intercept) | -4.9044823 | 1.2189240 | -4.023616 | 6.56e-05 |
| education | 0.9259646 | 0.0814035 | 11.374999 | 0.00e+00 |
| experience | 0.1051316 | 0.0171975 | 6.113181 | 0.00e+00 |

# Regressão Linear Multipla

$$wage = -4.9 + 0.92educ + 0.1exp$$

# Regressão Linear Multipla

## F-statistic, Valor - P e $R^2$ ajustado

- Além de todas as estatisticas que estudamos, agora temos mais três para analisar

# Regressão Linear Multipla

## F-statistic, Valor - P e $R^2$ ajustado

- Além de todas as estatisticas que estudamos, agora temos mais três para analisar
- A estatistica F é uma continha que testa se conjuntamente, há pelo menos um coeficiente diferente de zero.

# Regressão Linear Multipla

## F-statistic, Valor - P e $R^2$ ajustado

- Além de todas as estatisticas que estudamos, agora temos mais três para analisar
- A estatistica F é uma continha que testa se conjuntamente, há pelo menos um coeficiente diferente de zero.
- Porém não há uma regra de bolso pois ele depende do graus de liberdade da regressão, então olhamos o valor - p por facilidade.

# Regressão Linear Multipla

## F-statistic, Valor - P e $R^2$ ajustado

- Além de todas as estatisticas que estudamos, agora temos mais três para analisar
- A estatistica F é uma continha que testa se conjuntamente, há pelo menos um coeficiente diferente de zero.
- Porém não há uma regra de bolso pois ele depende do graus de liberdade da regressão, então olhamos o valor - p por facilidade.
- A regra de bolso do valor - p é, caso seja menor que 5%(0.05), sua regressão tem pelo menos um coeficiente diferente de zero.

# Regressão Linear Multipla

## F-statistic, Valor - P e $R^2$ ajustado

- Além de todas as estatisticas que estudamos, agora temos mais três para analisar
- A estatistica F é uma continha que testa se conjuntamente, há pelo menos um coeficiente diferente de zero.
- Porém não há uma regra de bolso pois ele depende do graus de liberdade da regressão, então olhamos o valor - p por facilidade.
- A regra de bolso do valor - p é, caso seja menor que 5%(0.05), sua regressão tem pelo menos um coeficiente diferente de zero.
- O $R^2$ de uma regressão sempre irá crescer ou pelo menos ficar constante caso você acrescente uma variavel endógena

# Regressão Linear Multipla

## F-statistic, Valor - P e $R^2$ ajustado

- Além de todas as estatisticas que estudamos, agora temos mais três para analisar
- A estatistica F é uma continha que testa se conjuntamente, há pelo menos um coeficiente diferente de zero.
- Porém não há uma regra de bolso pois ele depende do graus de liberdade da regressão, então olhamos o valor - p por facilidade.
- A regra de bolso do valor - p é, caso seja menor que 5%(0.05), sua regressão tem pelo menos um coeficiente diferente de zero.
- O $R^2$ de uma regressão sempre irá crescer ou pelo menos ficar constante caso você acrescente uma variavel endógena
- Por isso, para compararmos regressões multiplas, usamos o $R^2$ ajustado, que penaliza o incremento de variáveis que não ajudem o modelo a explicar melhor

# Regressão Linear Multipla

### F-statistic, Valor - P e $R^2$ ajustado

```
summary(modelo) %>% broom::glance()
```

| r.squared | adj.r.squared | sigma | statistic | p.value | df |
|---|---|---|---|---|---|
| 0.2020248 | 0.1990192 | 4.599365 | 67.21709 | 0 | 3 |

# Regressão com variáveis categóricas

# Regressão com variáveis categóricas

- Até agora vimos regressões somente com variáveis endógenas contínuas e/ou discretas.

# Regressão com variáveis categóricas

- Até agora vimos regressões somente com variáveis endógenas contínuas e/ou discretas.
- Agora iremos ver como aplicar regressões com variaveis exógenas categóricas, na qual representam classes.

# Regressão com variáveis categóricas

- Até agora vimos regressões somente com variáveis endógenas contínuas e/ou discretas.
- Agora iremos ver como aplicar regressões com variaveis exógenas categóricas, na qual representam classes.
- Para quem não sabe o que são as categóricas aqui em baixo vão dois exemplos:

# Regressão com variáveis categóricas

- Até agora vimos regressões somente com variáveis endógenas contínuas e/ou discretas.
- Agora iremos ver como aplicar regressões com variaveis exógenas categóricas, na qual representam classes.
- Para quem não sabe o que são as categóricas aqui em baixo vão dois exemplos:
- Categóricas cardinais: Quando não há um ordenamento. Sexo (H,M)

# Regressão com variáveis categóricas

- Até agora vimos regressões somente com variáveis endógenas contínuas e/ou discretas.
- Agora iremos ver como aplicar regressões com variaveis exógenas categóricas, na qual representam classes.
- Para quem não sabe o que são as categóricas aqui em baixo vão dois exemplos:
- Categóricas cardinais: Quando não há um ordenamento. Sexo (H,M)
- Categóricas ordinais: Quando há um ordenamento. Educação (Doutorado > Mestrado > Graduação)

# Regressão com variáveis categóricas

- Até agora vimos regressões somente com variáveis endógenas contínuas e/ou discretas.
- Agora iremos ver como aplicar regressões com variaveis exógenas categóricas, na qual representam classes.
- Para quem não sabe o que são as categóricas aqui em baixo vão dois exemplos:
- Categóricas cardinais: Quando não há um ordenamento. Sexo (H,M)
- Categóricas ordinais: Quando há um ordenamento. Educação (Doutorado $>$ Mestrado $>$ Graduação)
- No R essas variáveis são da classe `factor`

# Exercicios

# Regressão com variáveis categóricas

```r
# Gender é uma variavel categorica
modelo <- train(wage ~ education + experience + gender,
      method = "lm",
      data = CPS1985)
```

# Regressão com variáveis categóricas

```r
summary(modelo) %>% broom::tidy()
```

| term | estimate | std.error | statistic | p.value |
|---|---|---|---|---|
| (Intercept) | -4.1668750 | 1.1866703 | -3.511401 | 0.0004838 |
| education | 0.9405066 | 0.0788634 | 11.925765 | 0.0000000 |
| experience | 0.1133003 | 0.0167082 | 6.781112 | 0.0000000 |
| genderfemale | -2.3376324 | 0.3880624 | -6.023857 | 0.0000000 |

# Regressão com variáveis categóricas

$$wage = -4.16 + educ0.94 + exp0.11 - genderfemale2.33$$

# Modelo Logístico

# Modelo Logístico

O modelo de regressão logística também usa variáveis categóricas, so que agora endógenamente.

Ou seja, não queremos agora prever um possível número médio, mas sim uma classe como **sim** ou **não**.

Vamos pensar o contrário na nossa base de dados agora. Dado o salário/hora,anos de educação e experiencia conseguimos descobrir se a pessoa é do sexo masculino ou feminino?

# Modelo Logístico

```r
modelo <- train(gender ~ wage + education + experience,
      method = "glm",
      family = "binomial",
      data = CPS1985)
```

# Modelo Logístico

```r
summary(modelo)
```

```
##
## Call:
## NULL
##
## Deviance Residuals:
##    Min      1Q  Median      3Q     Max
## -1.742  -1.085  -0.673   1.143   3.264
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.440471   0.572072  -2.518 0.011803 *
## wage        -0.134674   0.023724  -5.677 1.37e-08 ***
## education    0.148640   0.043049   3.453 0.000555 ***
## experience   0.029629   0.008334   3.555 0.000378 ***
## 
```

# Modelo Logístico

$$gender = -1.44 - 0.13wage + 0.14educ + 0.02exp$$

# Modelo Logístico

Considerações sobre o modelo logístico

- A interpretação dos coeficientes são feitas em forma de probablidade, e temos que passar a fórmula $e^{\beta}$ para calcula-los

# Modelo Logístico

Considerações sobre o modelo logístico

- A interpretação dos coeficientes são feitas em forma de probablidade, e temos que passar a fórmula $e^{\beta}$ para calcula-los
- O mesmo ocorre com a variavel `gender`, que temos que passar a função $\frac{1}{1+e^{y}}$

# Modelo Logístico

Considerações sobre o modelo logístico

- A interpretação dos coeficientes são feitas em forma de probablidade, e temos que passar a fórmula $e^{\beta}$ para calcula-los
- O mesmo ocorre com a variavel `gender`, que temos que passar a função $\frac{1}{1+e^{y}}$
- Vamos calcular um exemplo

# Modelo Logístico

Caso tívessemos uma observação com as seguintes varíaveis

wage = 5.1, educ = 8, exp = 21, qual seria a probabilidade dessa pessoa ser do gênero feminino

Jogando na fórmula

$-1.44 - 0.13 * 5.1 + 0.14 * 8 + 0.02 * 21 = 0.76$

Jogando agora na formula $\frac{1}{1+e^{(}0.76)} = 0.31$

A chance de ser do gênenro feminino seria de 31%

# Modelo Logístico

Agora para isso voltar como uma variavel categórica nós precisamos definir um valor de decisão que varia entre 0 a 1.

- Na maioria dos casos esse valor é 0.5, ou seja.

# Modelo Logístico

Agora para isso voltar como uma variavel categórica nós precisamos definir um valor de decisão que varia entre 0 a 1.

- Na maioria dos casos esse valor é 0.5, ou seja.
- Se gender $> 0.5$, o indivíduo é do genero feminino

# Modelo Logístico

Agora para isso voltar como uma variavel categórica nós precisamos definir um valor de decisão que varia entre 0 a 1.

- Na maioria dos casos esse valor é 0.5, ou seja.
- Se gender $> 0.5$, o indivíduo é do genero feminino
- Se gender $< 0.5$, o indivíduo é do genero masculino

# Modelo Logístico

Podemos automatizar todo esse processo no R com a função `predict`, na qual nos retorna um vetor com a previsão de classificação do nosso data frame.

```
previsao <-predict(modelo,newdata = CPS1985)
previsao
```

```
##   [1] male   female male   male   male   male   female male
##  [11] male   male   male   male   male   female male   male
##  [21] female male   female female male   female male   fema
##  [31] male   female male   female male   male   male   fema
##  [41] male   male   male   male   female male   male   male
##  [51] male   male   male   female male   male   male   male
##  [61] male   male   male   male   male   male   male   male
##  [71] male   male   male   male   female female female male
##  [81] male   male   female female male   male   male   male
##  [91] male   male   male   female male   male   male   male
```

# Modelo Logístico

Verificando a acurácia do modelo, usando uma matriz de confusão

```
table(previsao,CPS1985[,"gender"])
```

```
##
## previsao male female
##   male     218    115
##   female    71    130
```

Essa matriz de confusão nos retorna diversos indicadores de acurácia do nosso modelo.

Para calcular a acurácia geral fazemos o seguinte: $(218 + 130)/\ 534 = 65\%$

Ou seja, nosso modelo acertou no geral 65% das classificações.

Para saber mais sobre essa matriz, clique aqui