

INTRODUÇÃO À ECONOMETRIA

Análise de Regressão Múltipla

Aula 8

Escola Nacional de Administração Pública

FORMAS FUNCIONAIS

Análise de Regressão Múltipla

Formas Funcionais – Modelos em log

Considere o modelo

$$\log(\text{renda}) = \beta_1 + \beta_2 \log(\text{estudo}) + \beta_3 \text{sexo}$$

- Renda – renda habitual do trabalhador
- Estudo - anos de escolaridade
- Sexo – masculino ou feminino

Análise de Regressão Múltipla

Formas Funcionais – Modelos em log

Considere o modelo

$$\log(\text{renda}) = \beta_1 + \beta_2 (\text{estudo}) + \beta_3 \text{ sexo}$$

- β_2 é a elasticidade do tempo de estudo em relação a renda do trabalhador
- O coeficiente β_3 é a mudança em $\log(\text{renda})$ para uma comprar mudança de sexo entre os indivíduos da amostra.

Análise de Regressão Múltipla

Formas Funcionais – Modelos em log

Dados de rendimento

- `setwd("diretorio/diretorio ")`
- `Z<-read.csv("wage1.csv",sep=";", dec=".", head=TRUE)`
- `summary(Z)`

Criando as series em logaritmo

- `Z$lnwage=log(Z$wage)`

Análise de Regressão Múltipla

Formas Funcionais – Modelos em log

Estimando a equação:

$$\log(\text{renda}) = \beta_1 + \beta_2 (\text{estudo}) + \beta_3 \text{ sexo}$$

- `reg_sal=lm(Z$lnwage~Z$educ+Z$female)`
- `summary(reg_sal)`

O que significam os coeficientes?

Análise de Regressão Múltipla

Formas Funcionais – Modelos em log

- Porque tomar o logaritmo das variáveis?
 - Variáveis em log satisfazem as suposições do modelo clássico com mais frequência;
 - Pode reduzir problemas de heteroscedasticidade;
 - Reduz a amplitude dos dados, tornando as estimativas menos sensíveis a observações extremas;

Análise de Regressão Múltipla

Formas Funcionais – Modelos em log

- **Modelos em log com variáveis em forma de taxa.**

Cuidado com a interpretação!!

$$\log(\text{cred}) = \beta_0 + \beta_1 \log(\text{pib}) - \beta_2 \text{taxaj} + u$$

$$\log(\text{cred}) = \beta_0 + \beta_1 \log(\text{pib}) - \beta_2 \log(\text{taxaj}) + u$$

- onde
 - *cred*: demanda por crédito
 - *pib*: nível de renda
 - *taxaj* : taxa de juros

Análise de Regressão Múltipla

Formas Funcionais Quadráticas

- Modelos com formas funcionais quadráticas são geralmente utilizadas para captar efeitos marginais crescentes ou decrescentes.
- No caso mais simples, temos:

$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + u,$$

- Observe que

$$\frac{\partial Y}{\partial X} = \beta_1 + 2\beta_2 X$$

Análise de Regressão Múltipla

Formas Funcionais Quadráticas

- Utilizando dados do arquivo wage1.csv, estime o modelo:

$$\text{salario}_i = \beta_0 + \beta_1 \text{exper}_i + \beta_2 \text{exper}_i^2 + u_i,$$

- Os anos de experiência apresentam retorno crescente ou decrescente sobre o salário?
 - Qual sobre o efeito sobre o salário do primeiro ano de experiência?
 - Qual sobre o efeito sobre o salário do décimo ano de experiência?
 - A partir de quantos anos, a experiência apresenta efeito negativo sobre o salário?
-

Análise de Regressão Múltipla

Formas Funcionais Quadráticas

- estime o modelo utilizando a variável escolaridade e compare os resultados.

$$\log(\text{salario}_h) = \beta_0 + \beta_1 \text{educ} + \beta_2 \text{exper} + \beta_3 \text{exper}^2 + u$$

Análise de Regressão Múltipla

Formas Funcionais – Modelos quadráticos

```
reg_sal1=lm(Z$lnwage~Z$educ+Z$exper)
```

```
summary(reg_sal1)
```

```
Z$exper2=Z$exper*Z$exper
```

```
reg_sal2=lm(Z$lnwage~Z$educ+Z$exper+Z$exper2)
```

```
summary(reg_sal2)
```

O que significam os coeficientes?

Análise de Regressão Múltipla

Formas Funcionais – Interação de variáveis

```
Z$educexper=Z$educ*Z$exper
```

```
reg_sal3=lm(lnwage~Z$educ+Z$exper+Z$exper2+Z$educexper)
```

```
summary(reg_sal3)
```

O que significam os coeficientes?

Variável Binária -
Independente

Variável Binária

Explicação

- Em trabalhos aplicados, muitas variáveis descrevem aspectos qualitativos ou atributos (por exemplo, sexo, raça, cor, religião, nacionalidade, etc.)
- As variáveis dummies ou categóricas são utilizadas para “quantificar” tais atributos.

Variável Binária

Explicação

- Uma variável dummy é uma variável que toma valor 0 ou 1.
- Exemplo:
 - *male* = 1 se a pessoa é do sexo masculino
 - *male* = 0 caso contrário.
- Variáveis dummies também são chamadas variáveis binárias, variáveis categóricas ou variáveis dicotômicas.

Variável Binária

Regressão Simples

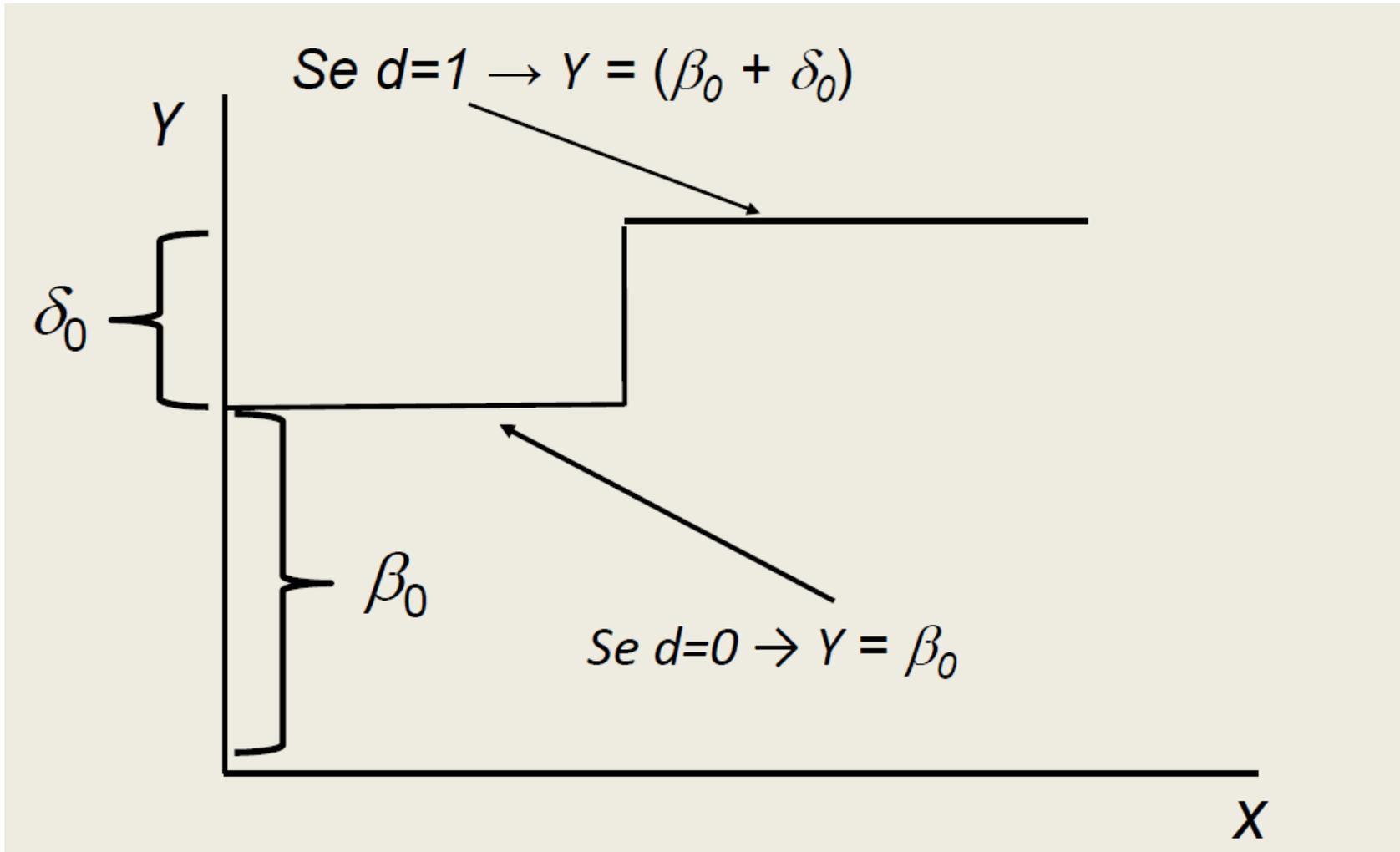
- Considere o seguinte modelo:

$$Y = \beta_0 + \delta_0 d + u$$

- Se $d = 0$, então, $E(Y/d=0) = \beta_0$
- Se $d = 1$, então, $E(Y/d=1) = \beta_0 + \delta_0$
- O caso em que $d = 0$ é considerado o grupo base.

Variável Binária

Regressão Simples



Variável Binária

Regressão Simples

$Z\$male = (Z\$female - 1) * -1$ -> criando o inverso do grupo de mulheres

Renda x Sexo feminino

- `regdum_simples = lm(Z$lnwage ~ Z$male)`
- `summary(regdum_simples)`

Renda x Casados

- `regdum_simples1 = lm(Z$lnwage ~ Z$married)`
- `summary(regdum_simples1)`

Por que casados ganham mais?

Variável Binária

Regressão Múltipla

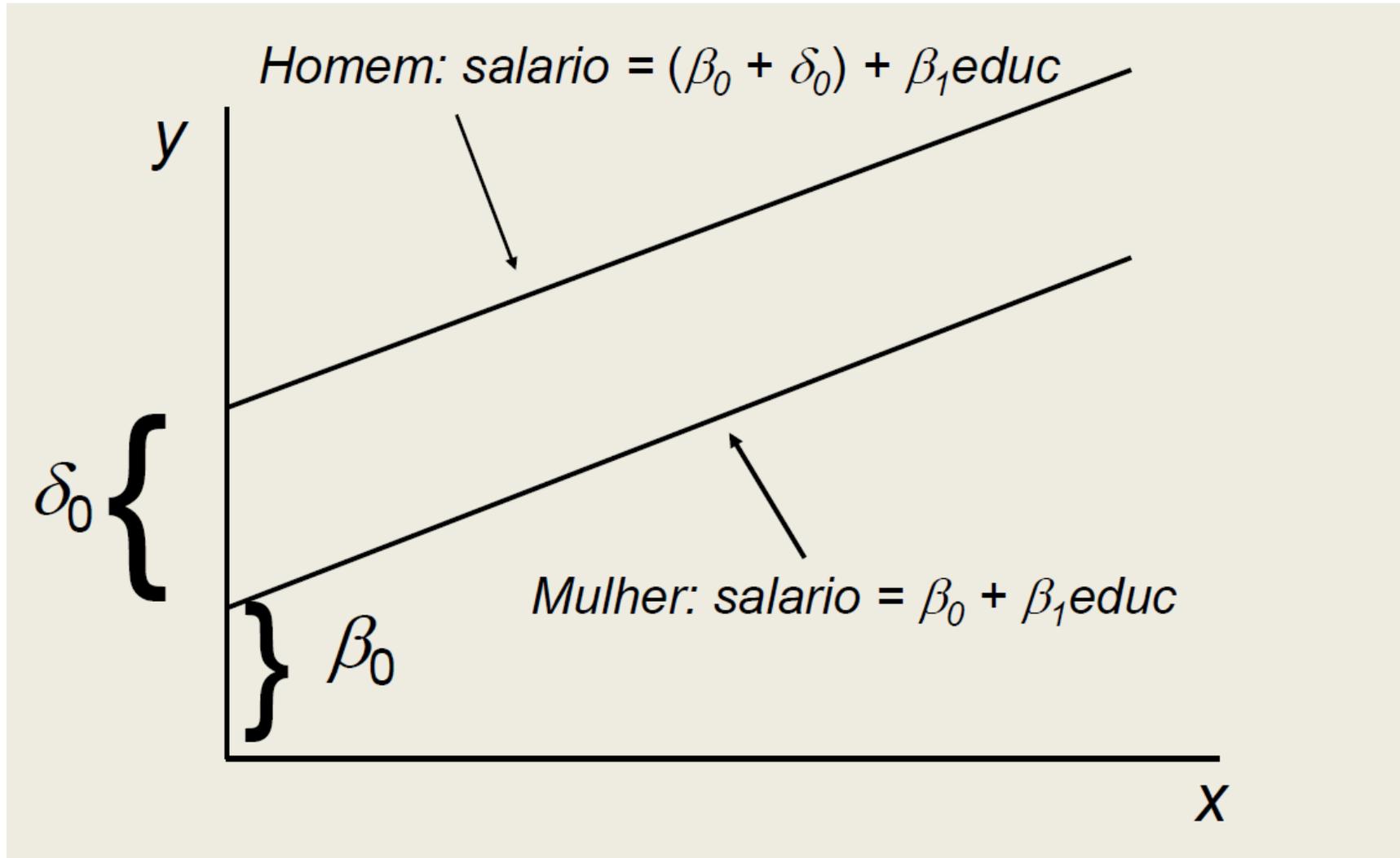
- Exemplo 1. Considere o model abaixo:

$$\text{salario} = \beta_0 + \delta d + \beta_1 \text{educ} + u$$

- Se $d = 1$, se a pessoa é do sexo masculino
- Se $d = 0$, caso contrário
- educ = nível de escolaridade

Variável Binária

Regressão Múltipla



Variável Binária

Regressão Múltipla

- Interpretação

$$\delta_0 = E(\text{salario/homem}, \text{educ}) - E(\text{salario/mulher}, \text{educ})$$

- Para um mesmo nível de escolaridade δ_0 mede a diferença entre o salário médio de homens e mulheres.

Variável Binária

Regressão Múltipla

Renda x Sexo feminino

- `regdum=lm(Z$lnwage~Z$male+Z$educ)`
- `summary(regdum)`

Variável Binária

Regressão Múltipla

- Como interpretar o coeficiente de uma variável dummy quando a variável dependente é $\log(y)$?

$$\log(\text{salario}) = \beta_0 + \delta_1 d_1 + \beta_1 \text{educ} + u$$

- Se $d_1 = 1$, se do sexo masculino e $d_1 = 0$ caso contrário
- O coeficiente δ_1 quando multiplicado por 100 é interpretado como a diferença percentual no salário entre homens e mulheres, mantendo os outros fatores fixos.
- A exata diferença percentual é dada por:

$$100[\exp(\hat{\delta}) - 1]$$

Variável Binária

Regressão Múltipla

- Exercício
- Use a base de dados WAGE1.csv e teste se há diferenças significativas entre os salários médios dos homens e das mulheres.

$$wage = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 tenure + \beta_4 female$$

```
regdum_geral=lm(Z$wage~Z$educ+Z$exper+Z$tenure+Z$female)
```

```
summary(regdum_geral)
```

Variável Binária

Regressão Múltipla

- Se há n categorias, devemos incluir $n-1$ variáveis dummies no modelo.

Por que?

Variável Binária

Várias dummies

$$despesa = \beta_0 + \delta_1 d_1 + \delta_2 d_2 + \beta_1 renda + u$$

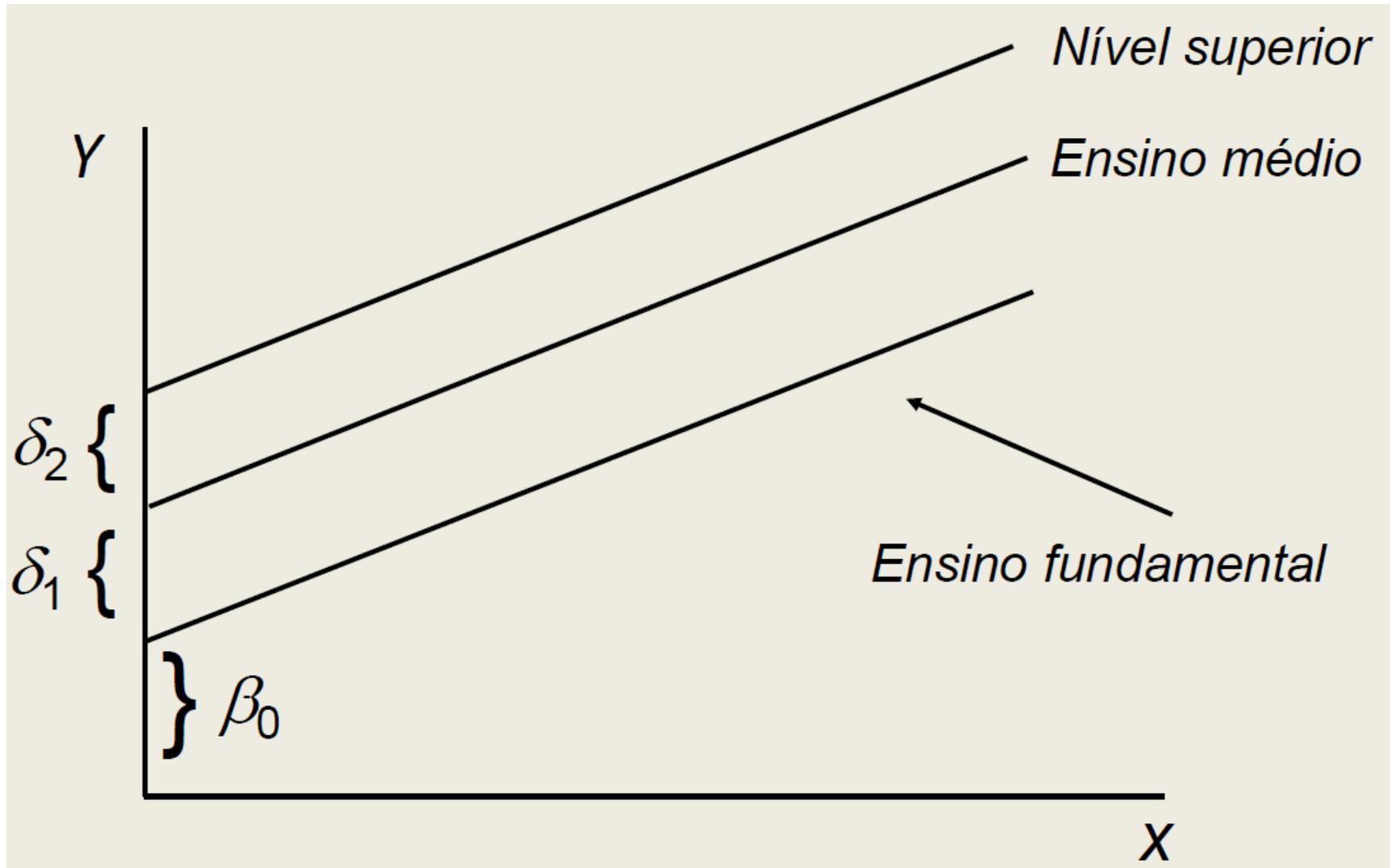
- Se $d_1 = 1$, se tiver formação secundária
- $d_1 = 0$, caso contrário

- Se $d_2 = 1$, se tiver formação superior
- $d_2 = 0$, caso contrário

Aterar o modelo para salário explicado por educação e anos de experiência

Variável Binária

Várias dummies



Variável Binária

Várias dummies

$$\log(wage) = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 tenure + \beta_4 married + \beta_5 black + \beta_6 south + u$$

- *married* = 1 se casado e 0 caso contrário.
- *black* = 1 se o indivíduo é negro e 0 caso contrário.
- *south* = 1 se o indivíduo mora no Sul e 0 caso contrário.
- *urban* = 1 se o indivíduo é morador de área urbana e 0 caso contrário.

Variável Binária

Várias dummies

Equação

- `regdum_geral1=lm(Z$lnwage~Z$educ+Z$exper+Z$tenure+Z$married+Z$nonwhite+Z$south+Z$female)`
- `summary(regdum_geral1)`

Variável Binária

Várias dummies

- Qual o impacto (em percentagem) de um ano a mais de escolaridade, mantidos constantes todos os demais fatores, sobre a renda de um indivíduo casado?
- Qual o impacto (em percentagem) de um ano a mais de escolaridade, mantidos constantes todos os demais fatores, sobre a renda de um indivíduo solteiro?
- Com base nos resultados encontrados, você diria que existe discriminação no mercado de trabalho? Explique.
- Existe diferença significativa entre o salário recebido pelos indivíduos das áreas urbanas e das áreas rurais?

Variável Binária

Diferente Coeficiente de Inclinação

- Considere o modelo abaixo com uma variável dummy d e uma variável contínua x

$$Y = \beta_0 + \delta_1 d + \beta_1 X + \delta_2 d * X + u$$

- Se $d = 0$, $E(Y/X, d=0) = \beta_0 + \beta_1 X$
- Se $d = 1$, $E(Y/X, d=1) = (\beta_0 + \delta_1) + (\beta_1 + \delta_2) X$
- Ocorre uma mudança no intercepto e na inclinação.

Variável Binária

Efeitos das Interação

- Considere o modelo abaixo:

$$Y = \beta_0 + \delta_1 d_1 + \delta_2 d_2 + \delta_3 d_1 d_2 + \beta X + u$$

- Y = gastos com roupa
 - X = renda
 - $d_1=1$ se mulher, $d_1=0$ caso contrário
 - $d_2=1$ se tiver nível superior, $d_2=0$ caso contrário
- A inclusão da interação permite que o efeito diferencial da dummy sexo não seja constante nos dois níveis de instrução e o efeito diferencial da dummy instrução não seja constante nos dois sexos.

Variável Binária

Efeitos das Interação

- Despesas se for mulher e tiver nível superior

$$E(\text{despesa}/d1=1, d2=1, \text{renda}) = (\beta_0 + \delta_1 + \delta_2 + \delta_3) + \beta X$$

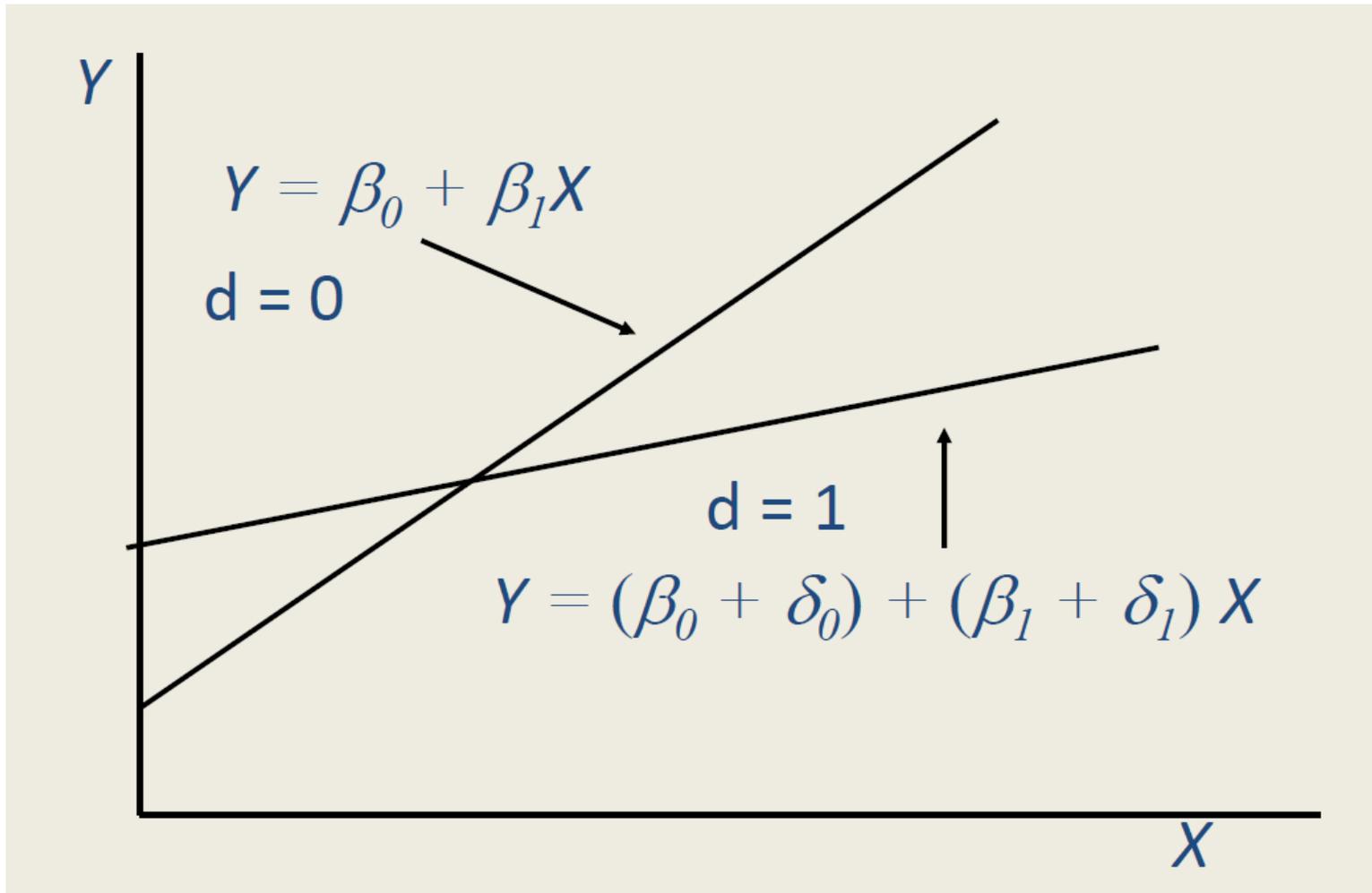
- Despesas se for homem sem nível superior

$$E(\text{despesa}/d1=0, d2=0, \text{renda}) = \beta_0 + \beta_1 X$$

- $\delta_1 =$ efeito diferencial do fato de ser mulher
- $\delta_2 =$ efeito diferencial do fato de ter nível superior
- $\delta_3 =$ efeito diferencial do fato de ser mulher com nível superior

Variável Binária

Efeitos da Interação



Variável Binária

Efeitos da Interação

- Usando a base de dados wage1.csv, estime o modelo abaixo e analise se um ano a mais de escolaridade tem efeito diferente para homens e mulheres.

$$\log(wage) = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 exper^2 + \beta_4 tenure + \beta_5 tenure^2 + \beta_6 female * educ$$

Variável Binária

Efeitos da Interação

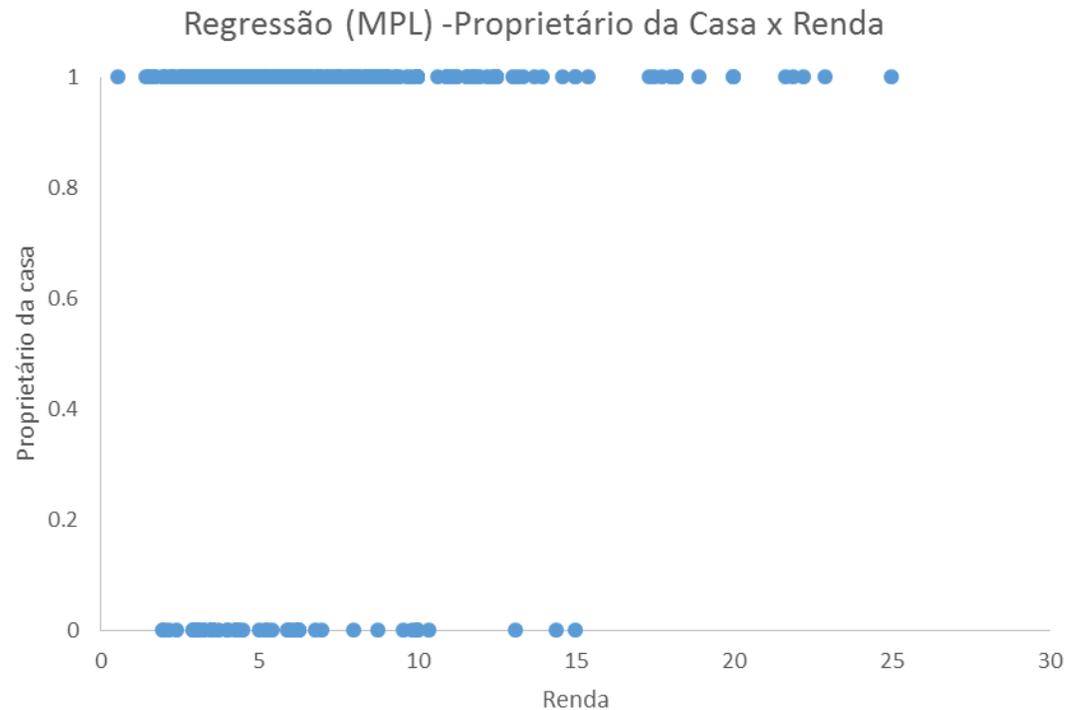
Equação

- $Z\$tenure2 = Z\$tenure * Z\$tenure$
- $Z\$fem_educ = Z\$female * Z\$educ$
- $regdum_geral2 = \text{lm}(Z\$lnwage \sim Z\$educ + Z\$exper + Z\$exper2 + Z\$tenure + Z\$tenure2 + Z\$fem_educ)$
- `summary(regdum_geral2)`

Variável Binária –
Dependente

Variável Binária

Modelo de Probabilidade Linear

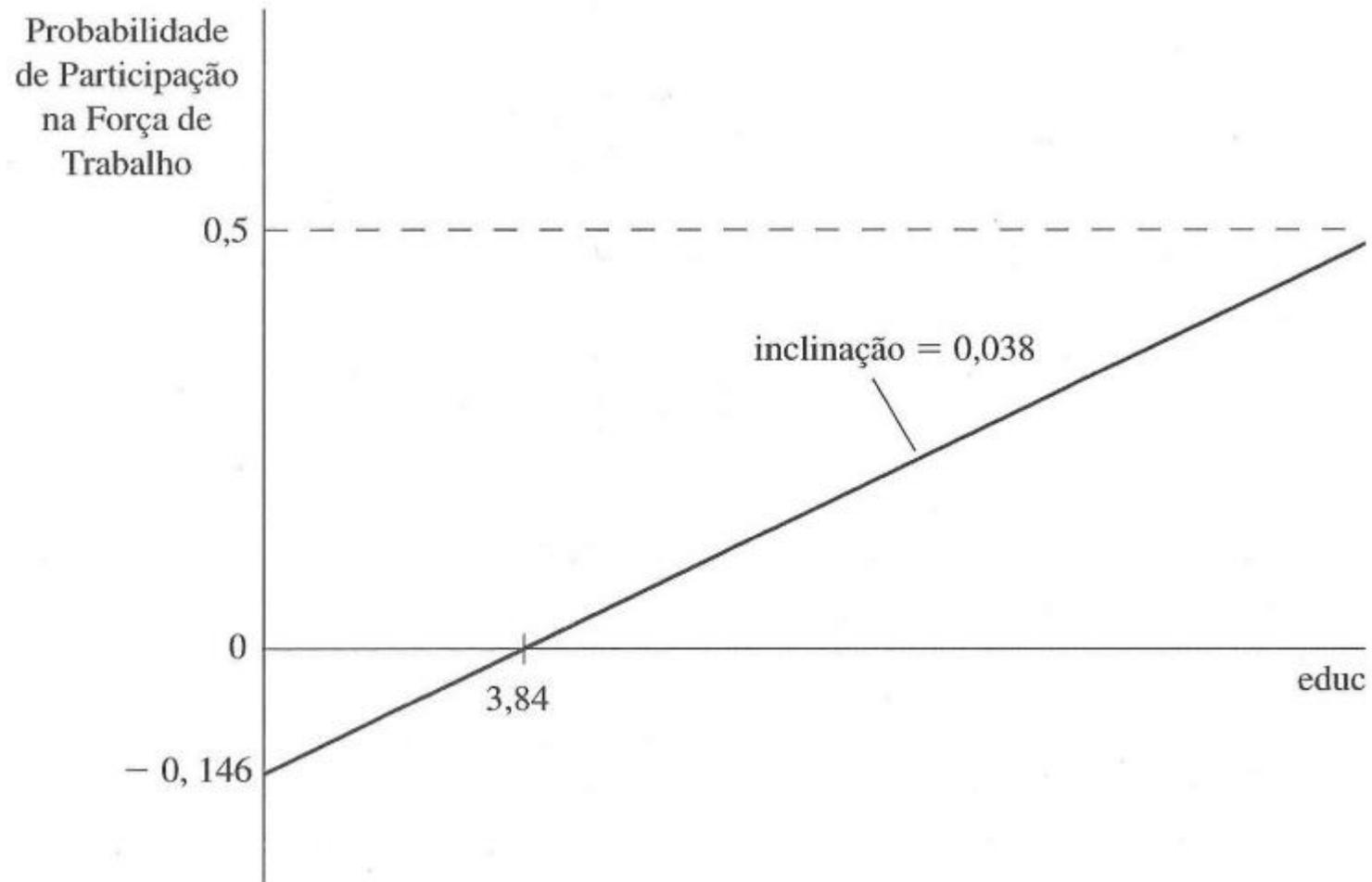


$$\text{Proprietário da casa}_i = \beta_0 + \beta_1 \text{Renda familiar}_i$$

$$\text{pr}(\text{Proprietário da casa}_i = 1 | \text{Renda familiar}_i)$$

Variável Binária

Modelo de Probabilidade Linear



Variável Binária

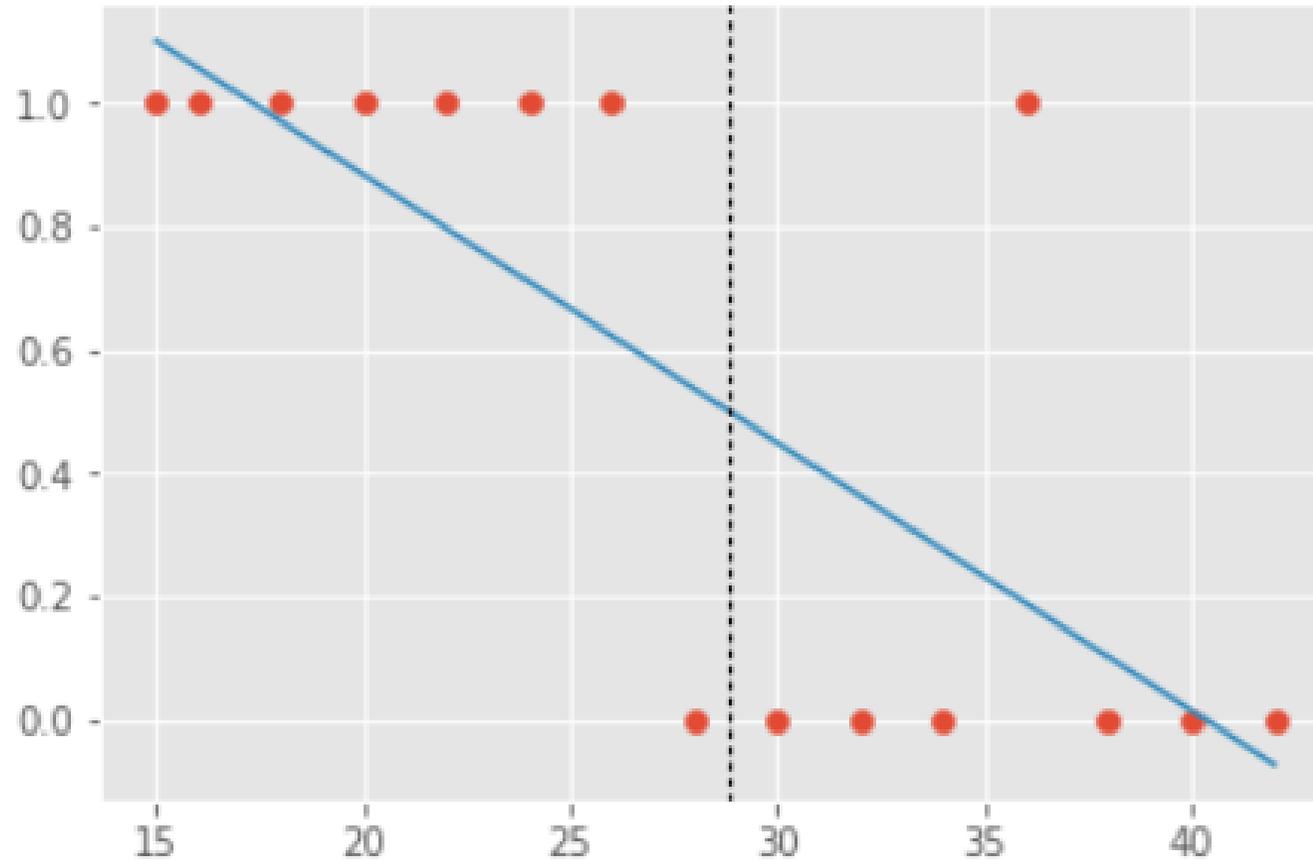
Modelo de Probabilidade Linear

Problemas com o MPL

- Resíduos não tem distribuição Normal (binomial) – há problema da inferência;
- Pode assumir valores negativos. Isso faz sentido?
 - Não satisfação de $0 \leq pr(Y_i | X_i) \leq 1$
- Heterocedasticidade;

Variável Binária

MPL (problemas)



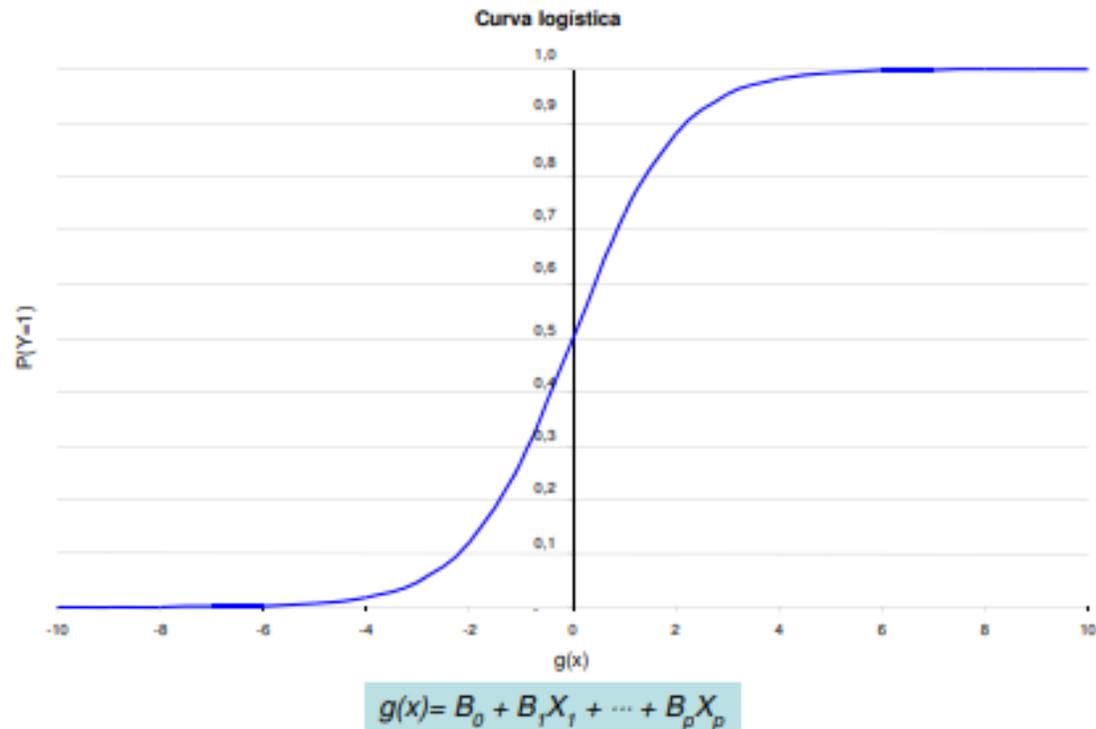
Variável Binária

Probit

- Emerge de uma função normal acumulada;
- Embora linear nos parâmetros, a variável independente não é linear;
- A variável dependente está limitada $[0,1]$;

Variável Binária

Probit



$$P(Y = 1) = \frac{1}{1 + e^{-g(x)}}$$

$g(x) \rightarrow +\infty$, então $P(Y = 1) \rightarrow 1$

$g(x) \rightarrow -\infty$, então $P(Y = 1) \rightarrow 0$

Variável Binária

Exemplo de MPL e Probit

Equação MPL

- `eua<-read.csv("EUA.csv",sep=";", dec=".", head=TRUE)`
- `summary(eua)`
- `reg_recss=lm(eua$recessao~eua$juro_real+eua$Prod_industrial)`
- `summary(reg_recss)`

O que significam os coeficientes?

Variável Binária

Exemplo de MPL e Probit

Equação PROBIT

- `install.packages("aod")`
- `require(aod)`
- `probit <- glm(eua$recessao ~ eua$Prod_industrial + eua$juro_real, family = binomial(link = "probit"), data=eua)`
- `summary(probit)`
- `confint(probit)`

EXERCÍCIO

Modelo de Regressão

Exercício

1. Estime a regressão múltipla e adicione alguma variável binária
2. Faça a interação da variável binária e o sexo do trabalhador
3. Estime a regressão de recessão nos EUA usando a taxa de inflação