Regressão Linear Simples

- Modelo linear para explicar a variável Y, denominada variável dependente, explicada ou endógena como função da variável X, denominada variável independente, explicativa ou exógena

$$y = \beta_0 + \beta_1 x + u$$

Propriedades do estimador MQO

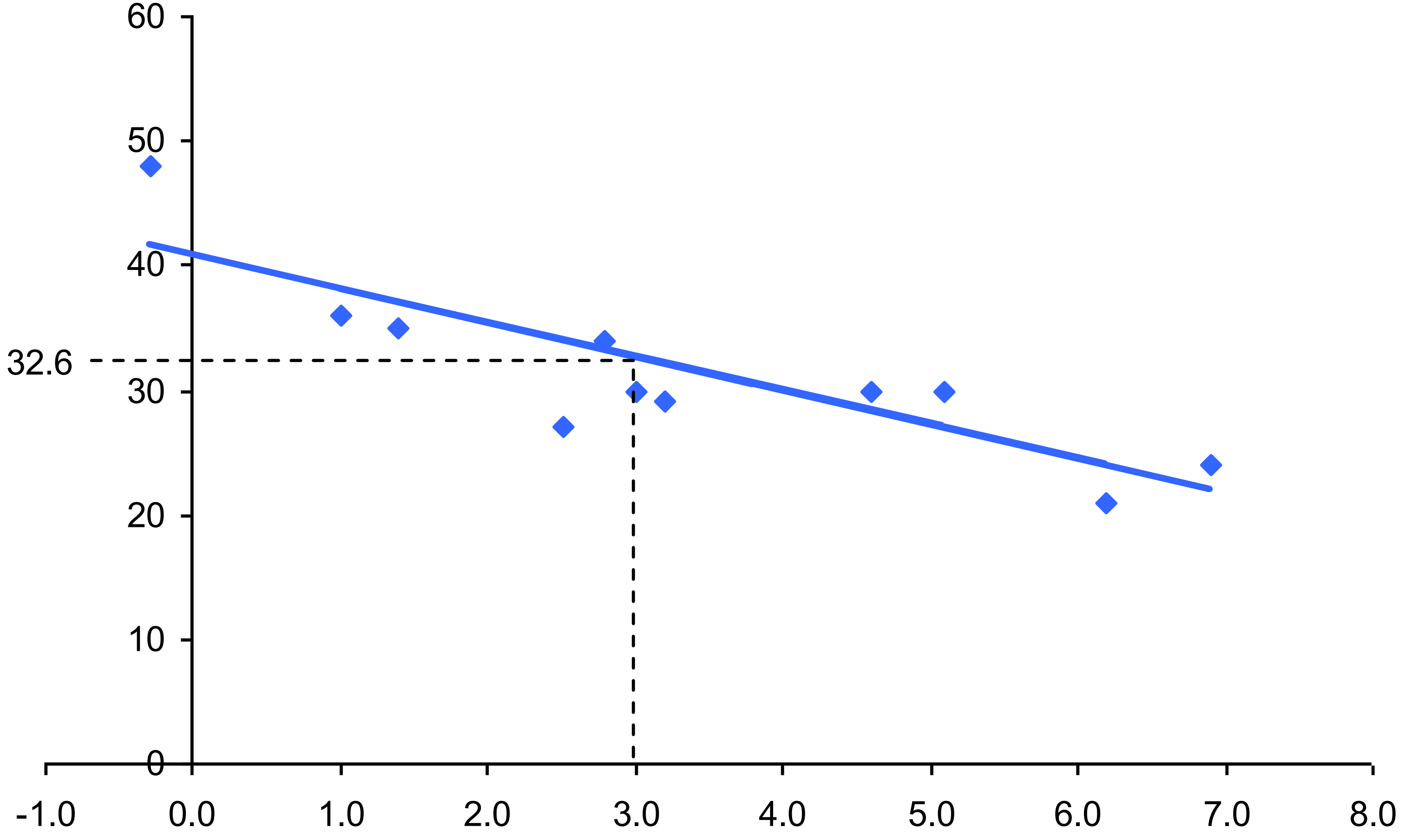
- Pressupostos de MQO:

1. O modelo populacional é linear nos parâmetros.
2. Podemos utilizar uma amostra aleatória de tamanho n.
3. A esperança condicional de u dado x é equivalente a zero.

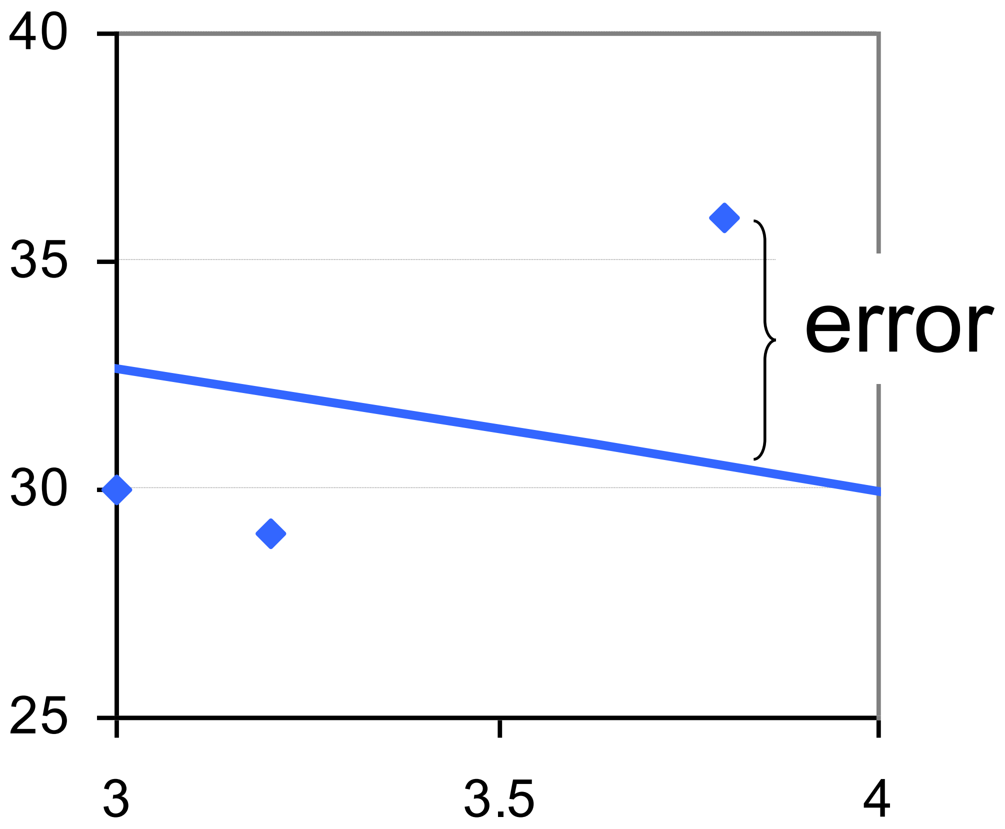
   $E(u/X)=E(u)=0$

4. Há variabilidade em x.

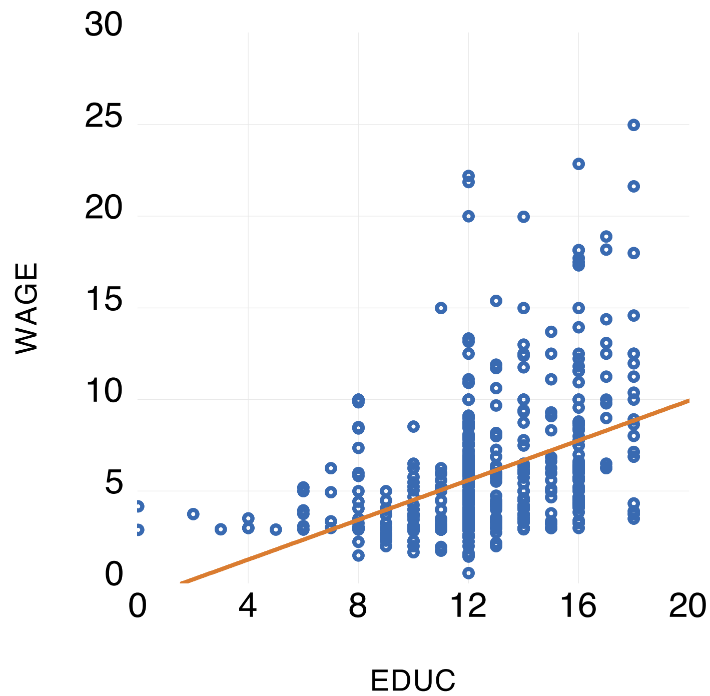
Correlação e Regressão



Regressão Linear Simples



Regressão Linear Simples



Regressão Linear Simples

- Problema: Os parâmetros *beta* são desconhecidos. O objetivo principal de regressão é estimar esses parâmetros a partir de uma amostra de tamanho n $\{(x_i, y_i): i=1, \ldots, n\}$

- <u>*Para cada observação na amostra, tem-se:*</u>

$$y_i = \beta_0 + \beta_1 x_i + u_i$$

Método dos Mínimos Quadrados

- A idéia de MQO é encontrar uma linha próxima o suficiente dos dados amostrais tal que a soma dos quadrados dos resíduos seja a menor possível.
- Por definição, o resíduo, û, é uma estimativa do termo de erro u.

Método dos Mínimos Quadrados

- Formalmente, tem-se a minimização de:

$$\sum_{i=1}^{n}\left(\hat{u}_i\right)^2 = \sum_{i=1}^{n}\left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right)^2$$

Método dos Mínimos Quadrados

- A solução do problema de minimização nos fornece as condiçoes de primeira ordem:

$$\sum_{i=1}^{n}\left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right) = 0$$

$$\sum_{i=1}^{n} x_i\left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right) = 0$$

Método dos Mínimos Quadrados

- Portanto, a estimação dos parâmetros é dada por:

$$\bar{y} = \hat{\beta}_0 + \hat{\beta}_1 \bar{x},$$

or

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

Método dos Mínimos Quadrados

- E,

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

$$\text{provided that} \sum_{i=1}^{n}(x_i - \bar{x})^2 > 0$$

Método dos Mínimos Quadrados

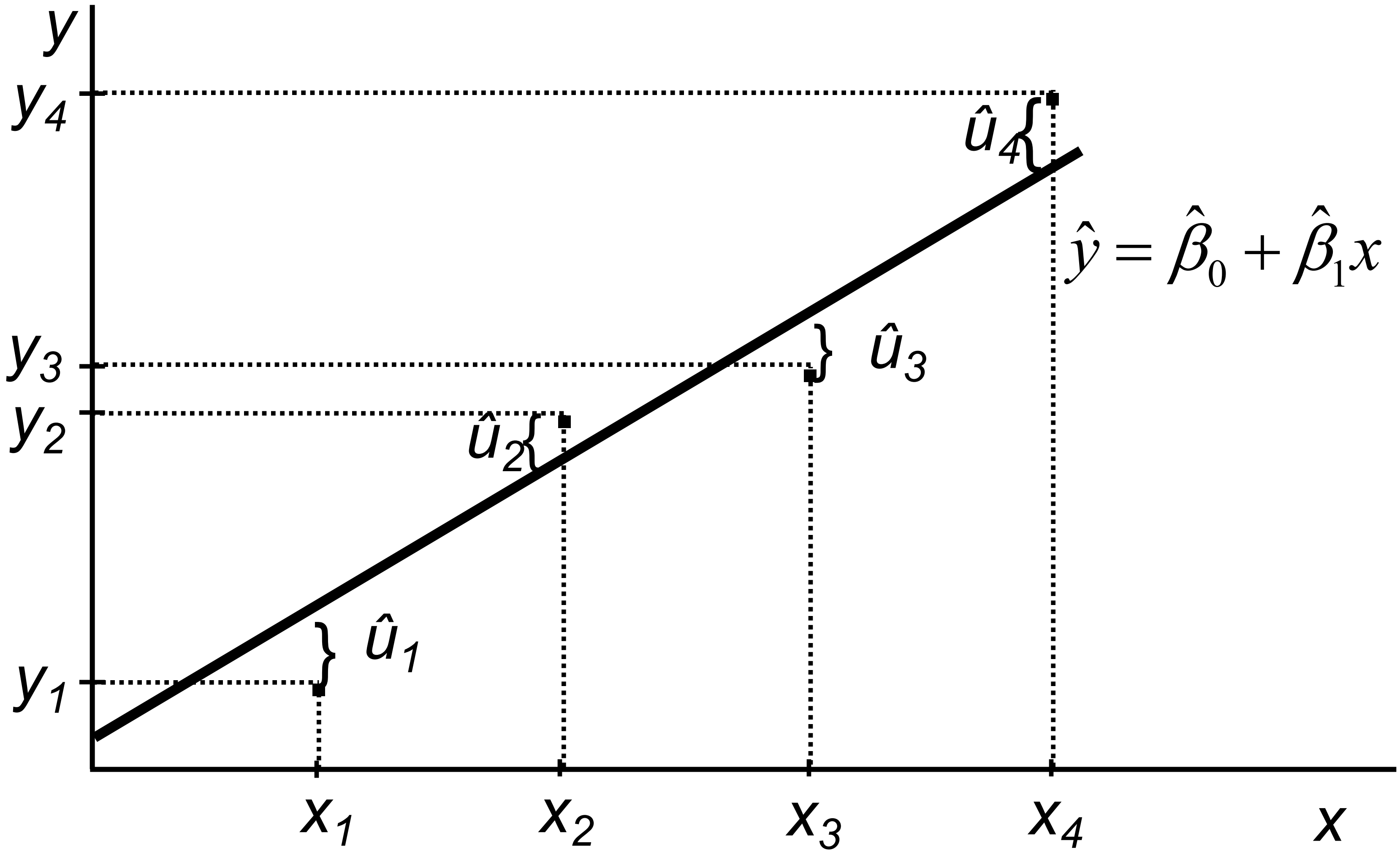
- A inclinação da reta é a covariância amostral entre x e y dividida pela variância amostral de x.

Regressão Linear Simples

- A relação entre valor observado de Y e valor previsto de Y pelo modelo é dada pelo resíduo:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$$

$$y = \hat{y} + \hat{u}$$


y
y_4
$\hat{u}_4$
$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$
y_3
$\hat{u}_3$
y_2
$\hat{u}_2$
$\hat{u}_1$
y_1
x_1
x_2
x_3
x_4
x

Método dos Mínimos Quadrados

Cada observação yi tem uma parte explicada e uma parte inexplicada, o resíduo do modelo

$y_i = \hat{y}_i + \hat{u}_i$. Definimos:

$\sum\left(y_i - \overline{y}\right)^2$ é a soma dos quadrados totais (SQT)

$\sum\left(\hat{y}_i - \overline{y}\right)^2$ é a soma dos quadrados explicados (SQE)

$\sum \hat{u}_i^2$ é a soma dos quadrados dos resíduos (SQR)

Then SQT = SQE + SQR

Regressão Linear Simples

- Como o nosso modelo de regressão se ajusta aos dados?

- Mensuração da qualidade do ajuste:

$$R^2 = SQE/SQT$$
$$= 1 - SQR/SQT$$

Regressão Linear Simples

Dependent Variable: WAGE
Method: Least Squares
Date: 04/13/21 Time: 13:31
Sample: 1 526
Included observations: 526

Variable	Coefficient	Std. Error	t-Statistic	Prob.
EDUC	0.541359	0.053248	10.16675	0.0000
C	-0.904852	0.684968	-1.321013	0.1871

R-squared	0.164758	Mean dependent var	5.896103
Adjusted R-squared	0.163164	S.D. dependent var	3.693086
S.E. of regression	3.378390	Akaike info criterion	5.276470
Sum squared resid	5980.682	Schwarz criterion	5.292688
Log likelihood	-1385.712	Hannan-Quinn criter.	5.282820
F-statistic	103.3627	Durbin-Watson stat	1.823686
Prob(F-statistic)	0.000000		

Propriedades do estimador MQO

- Pressupostos de MQO:

1. O modelo populacional é linear nos parâmetros.
2. Podemos utilizar uma amostra aleatória de tamanho n.
3. A esperança condicional de u dado x é equivalente a zero.
4. Há variabilidade em x.

Propriedades do estimador MQO

- Com base nesses 4 pressupostos, pode-se demonstrar que estimador MQO é não-enviesado:

$$E\left(\hat{\beta}_1\right) = \beta_1$$

- Ou seja, a distribuição amostral está centrada no valor verdadeiro do parâmetro.

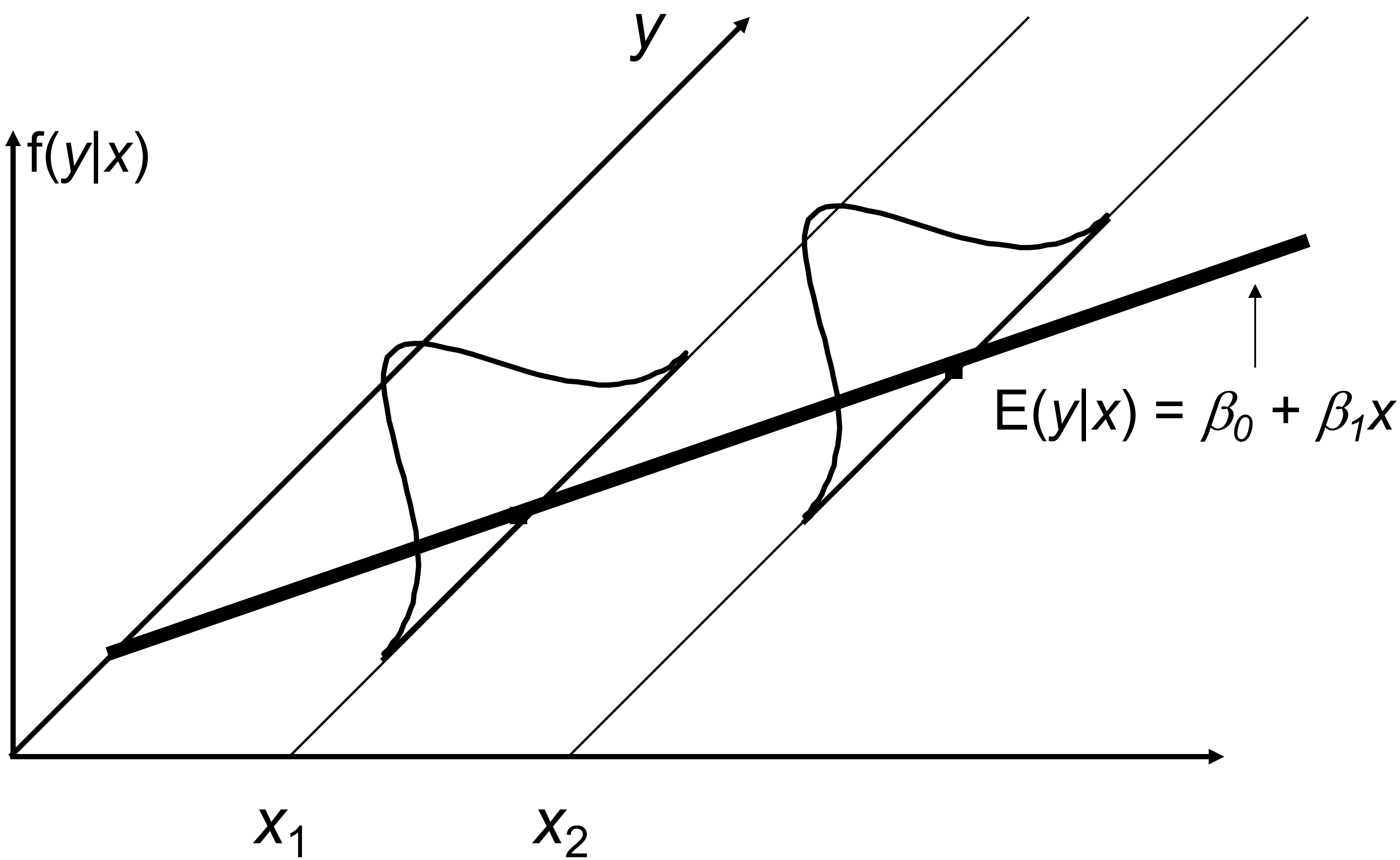
- E como a distribuição amostral está espalhada?

Propriedades do estimador MQO

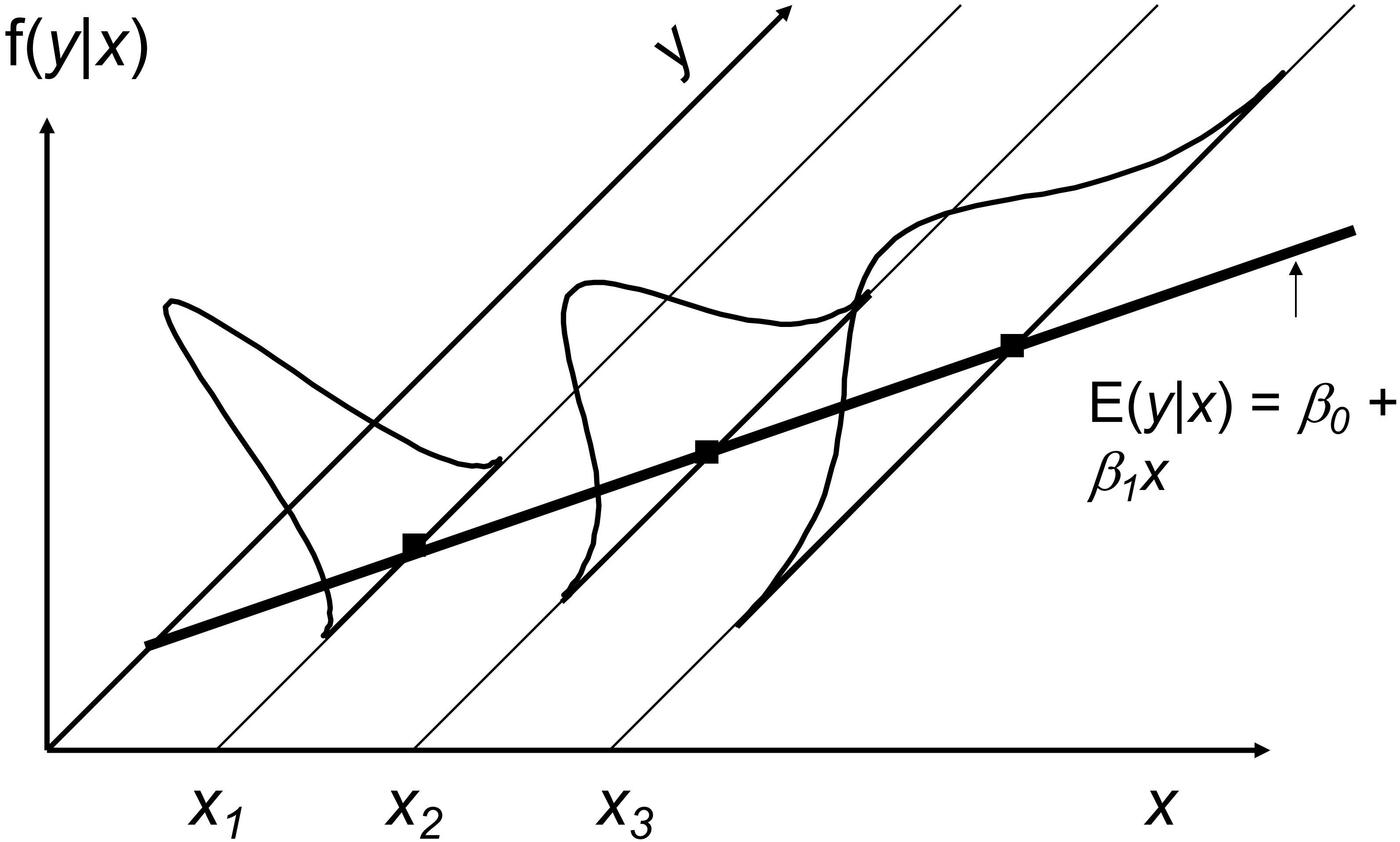
- Assumimos inicialmente homocedasticidade:

$$Var\left(u / x\right) = Var(u) = \sigma^2$$

Caso 1: Homocedasticidade



Caso 2: Heterocedasticidade



Variância de MQO

$$Var\left(\hat{\beta}_1\right) = \sigma^2 \big/ s_x^2$$

Variância de MQO

- Quanto maior a variância do erro, maior a variância do estimador beta
- Quanto maior a variabilidade de x, menor essa variância.
- Problema: a variância do erro é desconhecida!

Variância de MQO

Um estimador não-enviesado de σ^2 é:

$$\hat{\sigma}^2 = \frac{1}{(n-2)} \sum \hat{u}_i^2 = SQR/(n-2)$$

Regressão Linear Múltipla

Regressão Linear Múltipla

- Se a variável dependente for uma função de k variáveis explicativas, tem-se:

  - $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots \beta_k x_k + u$

- Como anteriormente, os valores dos parâmetros são estimados através da minimização da soma dos quadrados dos resíduos.

Regressão Linear Múltipla

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \ldots + \hat{\beta}_k x_k \,, então$$

$$\Delta\hat{y} = \hat{\beta}_1 \Delta x_1 + \hat{\beta}_2 \Delta x_2 + \ldots + \hat{\beta}_k \Delta x_k \,,$$

Se mantivermos $x_2, \ldots, x_k$ fixos, tem-se:

$$\Delta\hat{y} = \hat{\beta}_1 \Delta x_1$$

Pressupostos para ausência de viés

- Assuma que o modelo populacional é linear nos parâmetros: $y = b_0 + b_1x_1 + b_2x_2 + \ldots + b_kx_k + u$
- Podemos utilizar uma amostra aleatória $\{(x_{i1}, x_{i2}, \ldots, x_{ik}, y_i): i=1, 2, \ldots, n\}$,
- $E(u|x_1, x_2, \ldots x_k) = 0$,
- Nenhum das variavéis explicativas é constante ou uma combinação linear das demais.

Variance of OLS (cont)

- Seja $\boldsymbol{x} = (x_1, x_2, \ldots x_k)$

- Assuma homocedasticidade, como no modelo de regressão linear simples,

  $\text{Var}(u|\boldsymbol{x}) = s^2$

- Os 4 pressupostos para ausência de viés, juntamente com o pressuposto de homocedasticidade, são os pressupostos de Gauss-Markov.

Variance of OLS (cont)

Dados os pressupostos de Gauss-Markov, tem-se:

$$Var\left(\hat{\beta}_j\right) = \frac{\sigma^2}{SST_j\left(1 - R_j^2\right)}, \text{ onde}$$

$$SST_j = \sum\left(x_{ij} - \overline{x}_j\right)^2 \text{ e } R_j^2 \text{ é o } R^2 da$$

$$regress\tilde{a}ode\ x_j \text{ nos demais } x\text{'s}$$

Variance of OLS (cont)

Dados os pressupostos de Gauss-Markov, MQO é o melhor estimador linear não-enviesado dos parâmetros populacionais!

"OLS is BLUE"

Inferência

Regressão Linear Simples - Inferência

- Considere:
  - $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots \beta_k x_k + u$

- Utilizamos dados amostrais para calcular as estimativas dos parâmetros.

Regressão Linear Simples - Inferência

- Para realizar testes de hipóteses, assumimos que u tem distribuição normal com média 0 e variância constante.

- Nesse caso,

$$\left(\hat{\beta}_{\mathrm{j}} - \beta_j\right) \Big/ se\left(\hat{\beta}_j\right) \sim t_{n-k-1}$$

Regressão Linear Simples - Inferência

- Podemos testar hipóteses com relação aos parâmetros populacionais

- Para isso, precisamos calcular o desvio-padrão associado aos estimadores.

$$Var\left(\hat{\beta}_j\right) = \frac{\sigma^2}{SST_j\left(1 - R_j^2\right)}, \text{ onde}$$

Regressão Linear Simples - Inferência

- Estamos prontos para testar hipóteses com relação aos valores dos parâmetros populacionais

- Exemplo:

$$H_0: \beta = 0$$
$$H_1: \beta \neq 0$$

Regressão Linear Simples - Inferência

- A estatística do teste é:

$$t = {}^{\left(\hat{\beta}_{\mathrm{j}} - 0\right)}\!\Big/\!_{se\left(\hat{\beta}_j\right)}$$

Teste de Hipótese

- As etapas do teste são:

1. Escrever as hipóteses alternativas e nulas
2. Escolher o nível de significância do teste α
3. Calcular a estatística t, conhecida como a **estatística do teste**
4. Encontrar o **valor crítico** do teste t^*,
5. Decidir: Se o valor absoluto de t for maior do que o de t^*, rejeitar H_0 com um nível de confiança de $1-\alpha$