INTRODUÇÃO À ECONOMETRIA

Análise de Regressão Múltipla

Aula 5

Escola Nacional de Administração Pública

Mínimos Quadrados Ordinários - Suposições

Modelo de Regressão Linear
Suposições

- Na análise de regressão, o objetivo não é apenas estimar os coeficientes, mas também **realizar inferências** a respeito dos verdadeiros valores de β_0 e β_1 e Y_i.

- Por isso, faz-se necessário fazer algumas suposições a respeito do vetor X_i e do termo de erro u_i.

Modelo de Regressão Linear
Suposições

- Suposição 1:
  - ✓ O modelo de regressão é **linear nos parâmetros**.
  - ✓ $Y_i = \beta_0 + \beta_1 X_i + u_i$

- Observe que o regressando e os regressores podem ser não lineares.
- ***Exs:*** $ln(Y_i) = \beta_0 + \beta_1 ln(X)_i + u_i$
- $Y_i = \beta_0 + \beta_1 X^2_i + u_i$

Modelo de Regressão Linear
Suposições

- Suposição 2:
  - ✓ Os valores de X são fixados em amostragem repetida.

- A análise de regressão é uma análise condicional aos dados valores do(s) regressor(s) *X*.

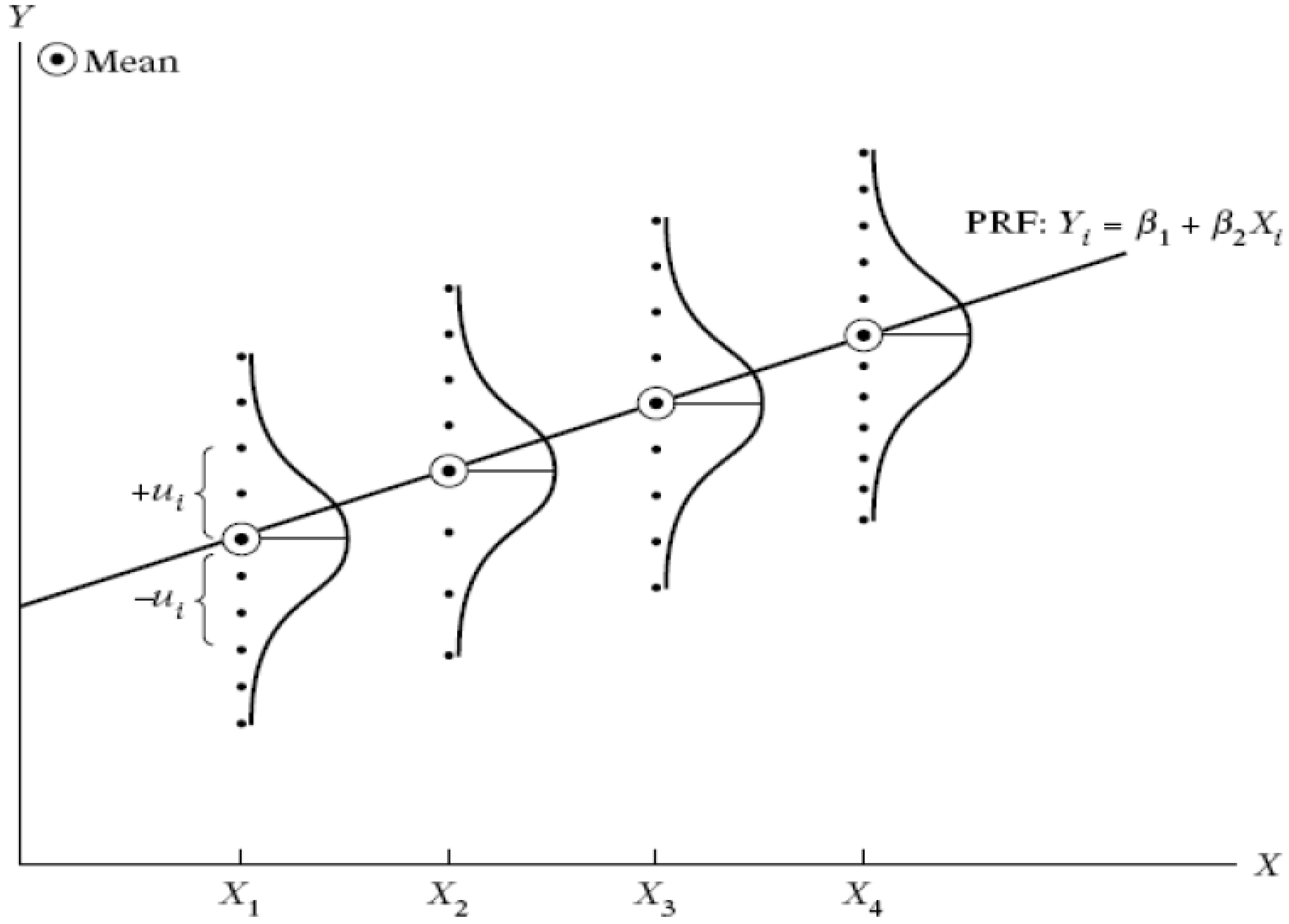
Modelo de Regressão Linear
Suposições

- Suposição 3:
  - ✓ **O valor esperado do erro, condicional a X, é zero.**
  - ✓ $E(u_i/X_i) = 0$

- As variáveis não incluídas em X_i não afetam sistematicamente o valor médio de Y. Os valores positivos de u_i anulam os valores negativos de u_i.

- Note que $E(u_i \mid X_i) = 0$ *implica* $E(Y_i \mid X_i) = \beta_1 + \beta_2 X_i$.

$$E(u_i/X_i) = 0$$



Modelo de Regressão Linear
Suposições

Suposição 4: **Variabilidade de X**

- ✓Os valores de X, em dada amostra, não podem ser todos iguais.
- ✓Var(X) deve ser um número positivo e finito.

Modelo de Regressão Linear
Suposições

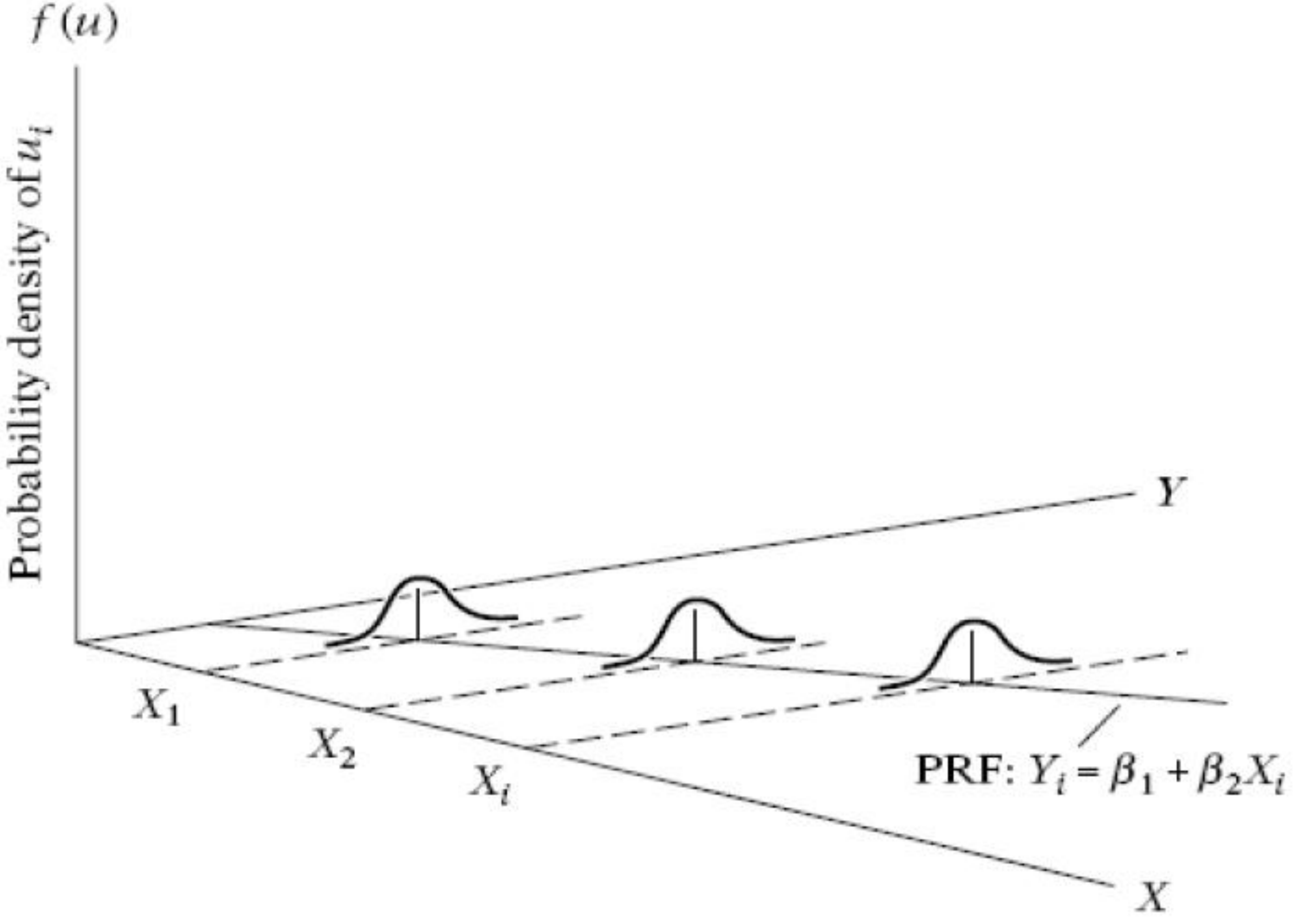
- Suposição 5: **Homoscedasticidade**
  - ✓ A variância de u_i é constante
  - ✓ $Var(u_i/X_i) = \sigma^2$.

- As populações de Y correspondentes aos vários X têm a mesma variância.

- A suposição 4 implica que $var(Y_i/X_i) = \sigma^2$.

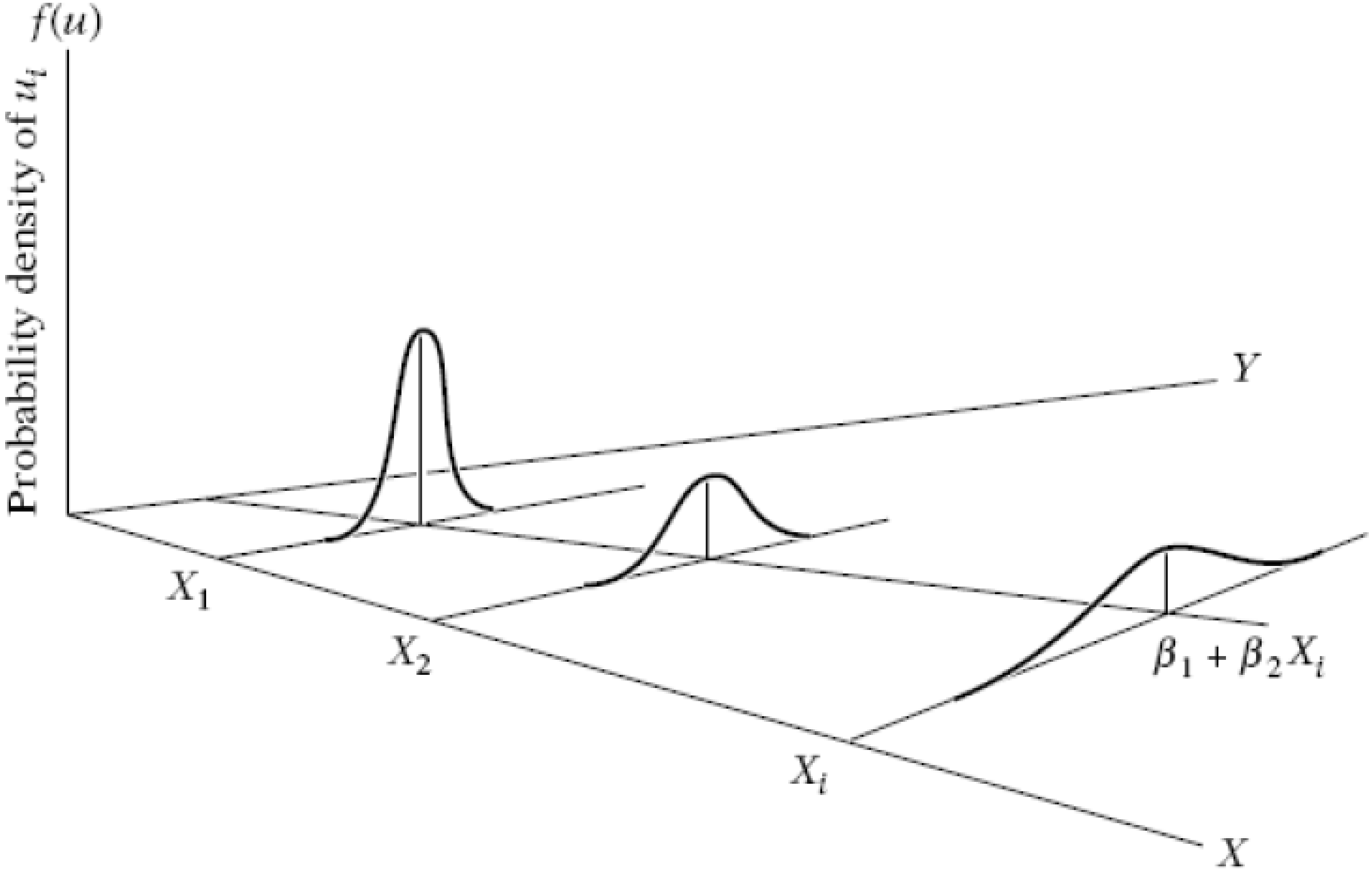
Homoscedasticidade

$var\ (u_i \mid X_i) = \sigma^2$



Heteroscedasticidade

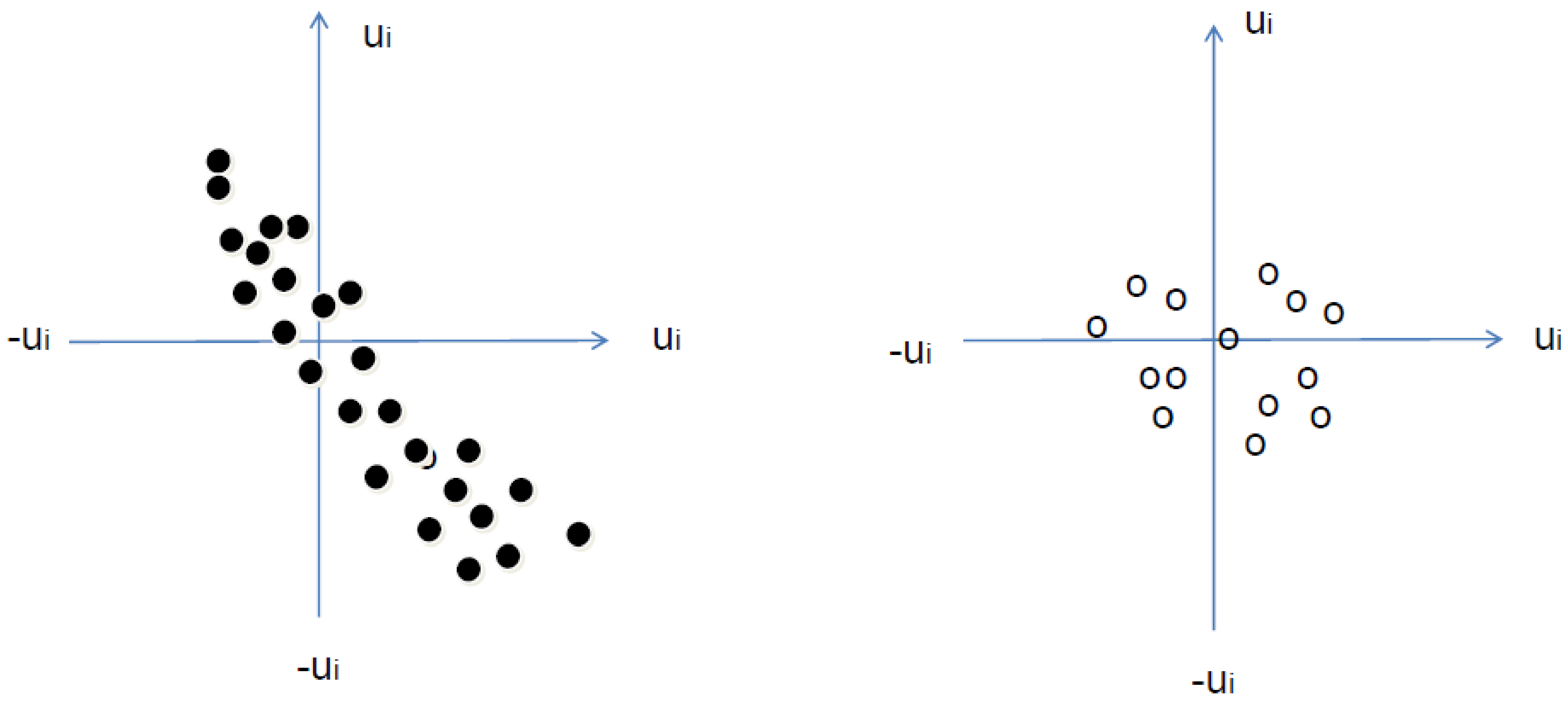
$var\ (u_i \mid X_i) = \sigma^2_i$



Modelo de Regressão Linear
Suposições

- Suposição 6:
  - ✓ **Ausência de autocorrelação** entre os erros.
  - ✓ Cov(u_i ,u_j/X_i ,X_j) = 0, on i e j são duas observações diferentes e Cov significa covariância.

- Em outras palavras, não há nenhum padrão sistemático para os u’s.



ui
-ui
ui
-ui
ui
-ui
ui
-ui

Modelo de Regressão Linear
Suposições

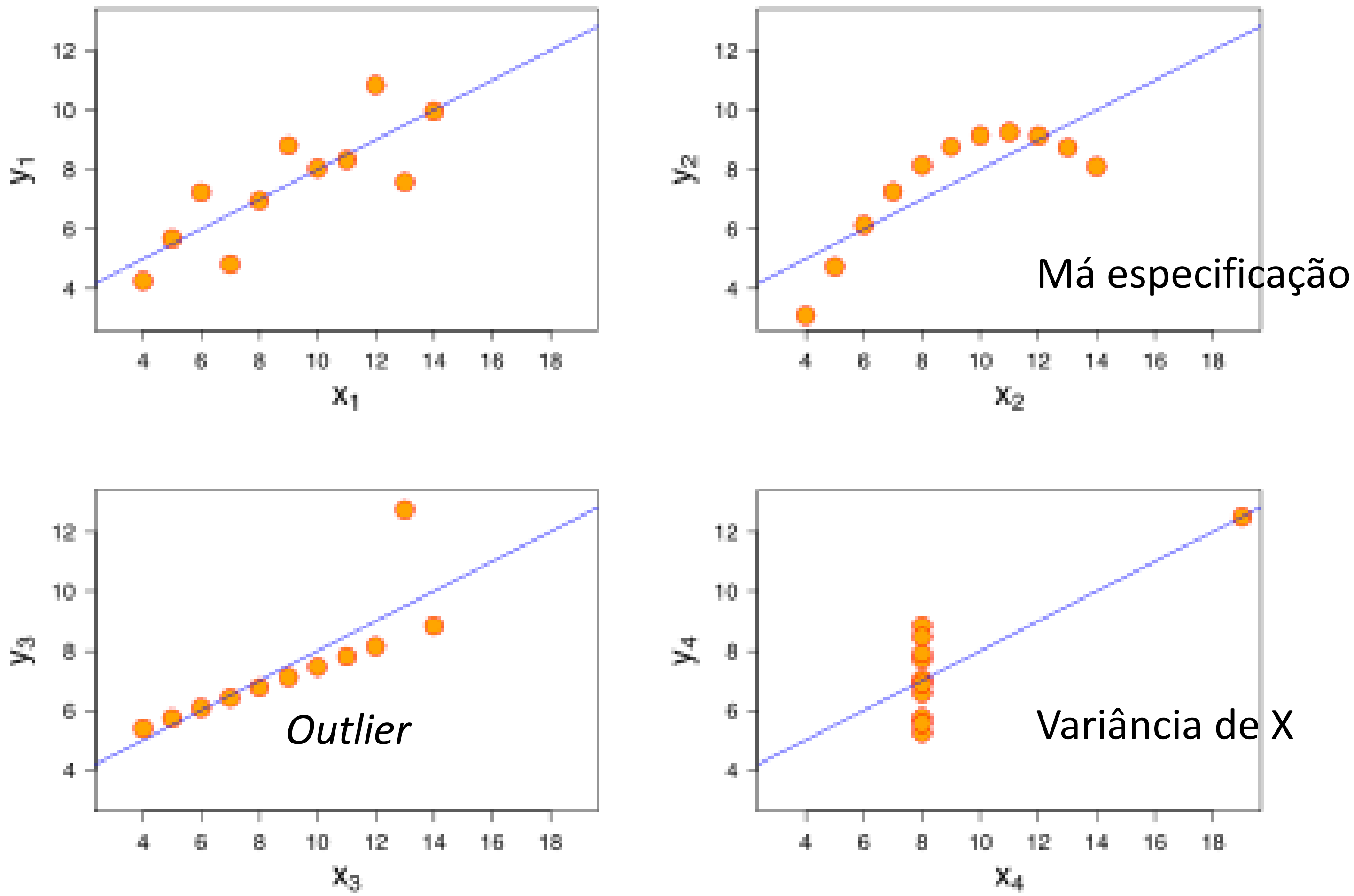
- Suposição 7:
  - ✓ **Ausência de correlação entre $\boldsymbol{u_i}$ e $\boldsymbol{X_i}$.**
  - ✓ $Cov(u_i, X_i) = 0$.

- X e u exercem influência separada em Y.

Modelo de Regressão Linear
Suposições

- Suposição 8:
  - ✓ **Não há multicolinearidade perfeita.** Não existe relação linear perfeita entre as variáveis explicativas.

Diagnosticos gráficos



Erro de especificação

- Problema:
  - Equação verdadeira (FRP):
    - $\boldsymbol{Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t}}$
  - Equação estimada (FRA):
    - $\boldsymbol{Y_t = \widehat{\beta}_0 + \widehat{\beta}_1\, X_{1t}}$
    - Há problema?
      - Sim: corr($\boldsymbol{X_{1t}}$, $\boldsymbol{X_{2t}}$) ≠ 0 -> $\boldsymbol{\widehat{\beta}_1}$ = $\boldsymbol{\beta_1}$+k$\boldsymbol{\beta_2}$
      - Não: corr($\boldsymbol{X_{2t}}$, $\boldsymbol{X_{2t}}$) = 0 -> $\boldsymbol{\widehat{\beta}_1}$ = $\boldsymbol{\beta_1}$
  - Coeficientes viesados!
    - Problemas de má especificação;
    - Variável não observada;
    - Erro de medida.

DIFERENÇA ENTRE ESTIMAÇÃO E INFERÊNCIA

Modelo de Regressão Linear Múltipla *Relembrando a última aula*

setwd("C:/diretorio1/diretorio2") – determinar o diretório: Não se esqueça, **inverter** as barras!!!

list.files() – quais arquivos que estão no seu diretório

X<-read.csv("paises.csv",sep=";", dec=".", head=TRUE) – baixando os dados.

```
reg1_mult_b=lm(X$juros~X$IPC+X$PIB+X$fiscal)
summary(reg1_mult_b)
```

O coeficiente de intercepto é zero? Ele é estatisticamente diferente de zero? O que acontece com os outros coef?

Modelo de Regressão Linear Múltipla Regressão pela Origem

- O termo intercepto β_0 é igual a zero.

$$\widetilde{Y} = \widetilde{\beta}_1 X_1 + \widetilde{\beta}_2 X_2 + \ldots + \widetilde{\beta}_k X_k \qquad (7)$$

- A soma dos resíduos não necessariamente é igual a zero
- O R^2 pode ser negativo
- Se o intercepto no modelo populacional é diferente de zero, as estimativas de (6) serão viesadas.
- O custo de estimar β_0 quando ele é igual a zero é que a variância dos estimadores será maior.

Modelo de Regressão Linear Múltipla
Regressão pela Origem

Extraindo as médias

- mean(X$juros)
- X$jurosm=X$juros-2.881579
- mean(X$IPC)
- X$IPCm=X$IPC-2.637368
- mean(X$pibm)
- X$PIBm=X$PIB-1.274737
- mean(X$fiscal)
- X$fiscalm=X$fiscal+0.8536842

Modelo de Regressão Linear Múltipla *Regressão pela Origem*

Estimando as equações

- reg_orig=lm(X$jurosm~X$IPCm+X$PIBm+X$fiscalm)
- summary(reg_orig)

- reg_s_interc=lm(X$jurosm~X$IPCm+X$PIBm+X$fiscalm-1)
- summary(reg_s_interc)

- summary(reg1_mult_b)

INFERÊNCIA

Análise de Regressão Múltipla

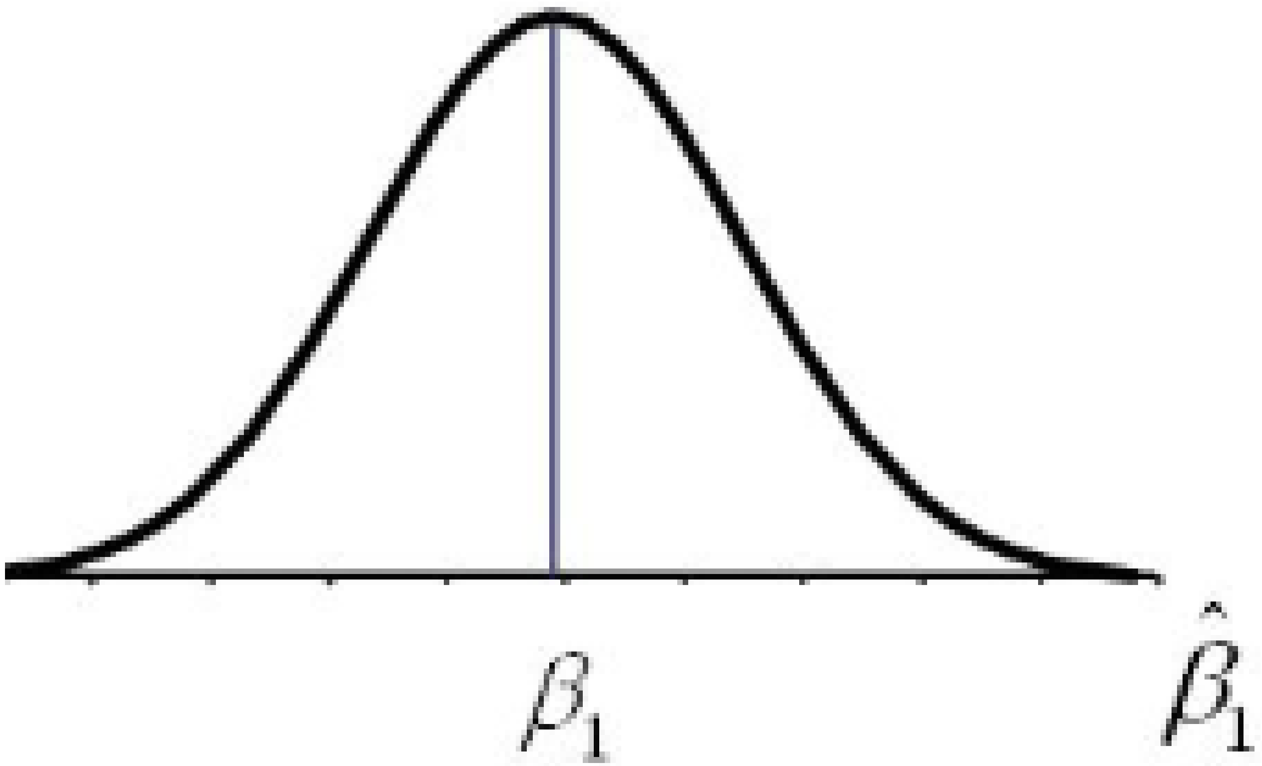
Distribuição Amostral dos Estimadores

- A normalidade do termo erro traduz-se nas distribuições normais amostrais dos estimadores MQO:

$$\hat{\beta}_j \sim \text{Normal}[\beta_j, \text{Var}(\hat{\beta}_j)]$$

$$\frac{\hat{\beta}_j - \beta_j}{\text{dp}(\hat{\beta}_j)} \sim \text{Normal}(0,1)$$

Sampling Distribution



Análise de Regressão Múltipla
Testes de Hipóteses - Teste t

- Dado o modelo populacional abaixo

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots \beta_k X_k + u,$$

- Vamos supor que queremos testar a hipótese nula:
  - H_0: $\beta_1 = 0$.
- Contra uma hipótese alternativa:
  - H_1: $\beta_1 \neq 0$

Modelo de Regressão Linear Múltipla *Diferentes amostras*

Diferenciando os betas

- reg1_mult=lm(X$juros~X$IPC+X$fiscal+ X$PIB)
- summary(reg1_mult)

- reg1_mult1=lm(X$juros~X$IPC+X$fiscal+ X$PIB, subset=c(1:9))
- summary(reg1_mult1)

- reg1_mult2=lm(X$juros~X$IPC+X$fiscal+ X$PIB, subset=c(10:19))
- summary(reg1_mult2)

Testes de Hipóteses
Teste t

- Caso a hipótese nula **não seja rejeitada**, dizemos que X não tem efeito sobre o valor esperado de Y.
- A fim de realizar os testes, precisamos do seguinte resultado:

$$\frac{\hat{\beta}_j - \beta_j}{\mathrm{ep}(\hat{\beta}_j)} \sim t_{n-k}$$

- Onde k é o número de parâmetros desconhecidos (incluindo o intercepto).

Testes de Hipóteses
Teste t

- A estatística usada para testar a hipótese nula é chamada "*estatística t*" ou "*razão t*" :

$$t_{\hat{\beta}_j} \equiv \frac{\hat{\beta}_j}{\mathrm{ep}(\hat{\beta}_j)}$$

  - Um valor amostral de $\hat{\beta}_j$ muito distante de zero fornece evidência contra H_0.
  - Como há um erro amostral em nossa estimativa, o tamanho de $\hat{\beta}_j$ é ponderado pelo seu desvio padrão estimado.

Testes de Hipóteses
Teste t – Regra de Decisão

- Estamos procurando um valor suficientemente grande de $t_{\hat{\beta}_j}$ a fim de rejeitar H_0 em favor de H_1.
- A escolha da regra para rejeitar H_0 depende não somente da hipótese alternativa, mas também do nível de significância (α) ou da probabilidade de rejeitar H_0 quando ela é verdadeira.

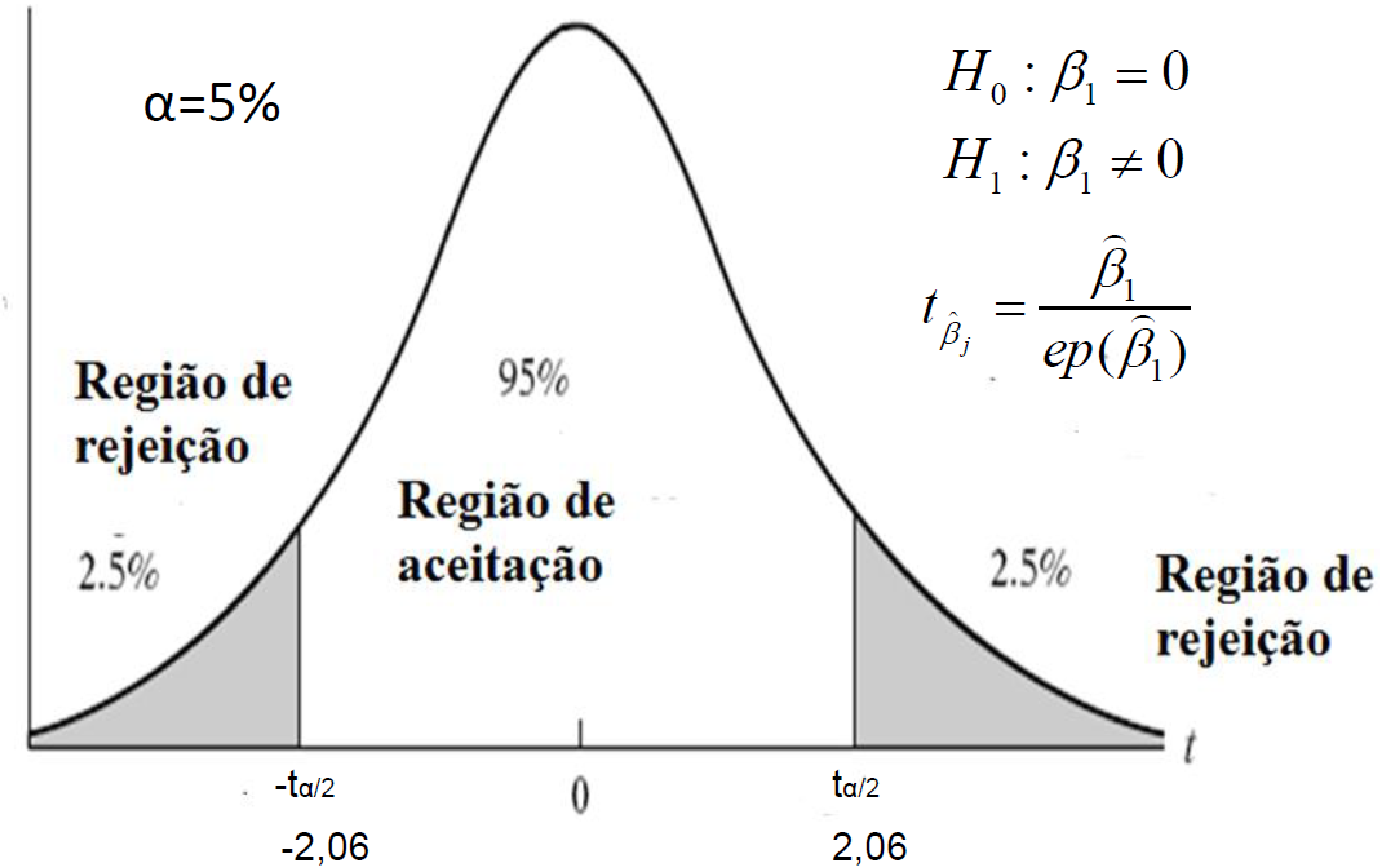
Testes de Hipóteses – Teste t
Alternativas Bilaterais

- Queremos testar:
  - H_0: $\beta_1 = 0$
  - H_1: $\beta_1 \neq 0$
  - Rejeita-se H_0 se:

$$\left| t_{\hat{\beta}_j} = \frac{\hat{\beta}_1}{ep(\hat{\beta}_1)} \right| > t_{\alpha/2}$$

  - onde $t_{\alpha/2}$ é o valor obtido de uma distribuição t com (n – k) graus de liberdade. $t_{\alpha/2}$ é conhecido como t crítico.

Testes de Hipóteses - Teste t
Alternativas Bilaterais



$$H_0 : \beta_1 = 0$$

$$H_1 : \beta_1 \neq 0$$

$$t_{\hat{\beta}_j} = \frac{\hat{\beta}_1}{ep(\hat{\beta}_1)}$$

Modelo de Regressão Linear Múltipla *Baixando os dados – dados em corte*

Diferenciando o nível de significância

- reg1_mult=lm(X$juros~X$IPC+X$fiscal)
- summary(reg1_mult)
- confint(reg1_mult,level=0.95)
- confint(reg1_mult,level=0.99)
- confint(reg1_mult,level=0.90)

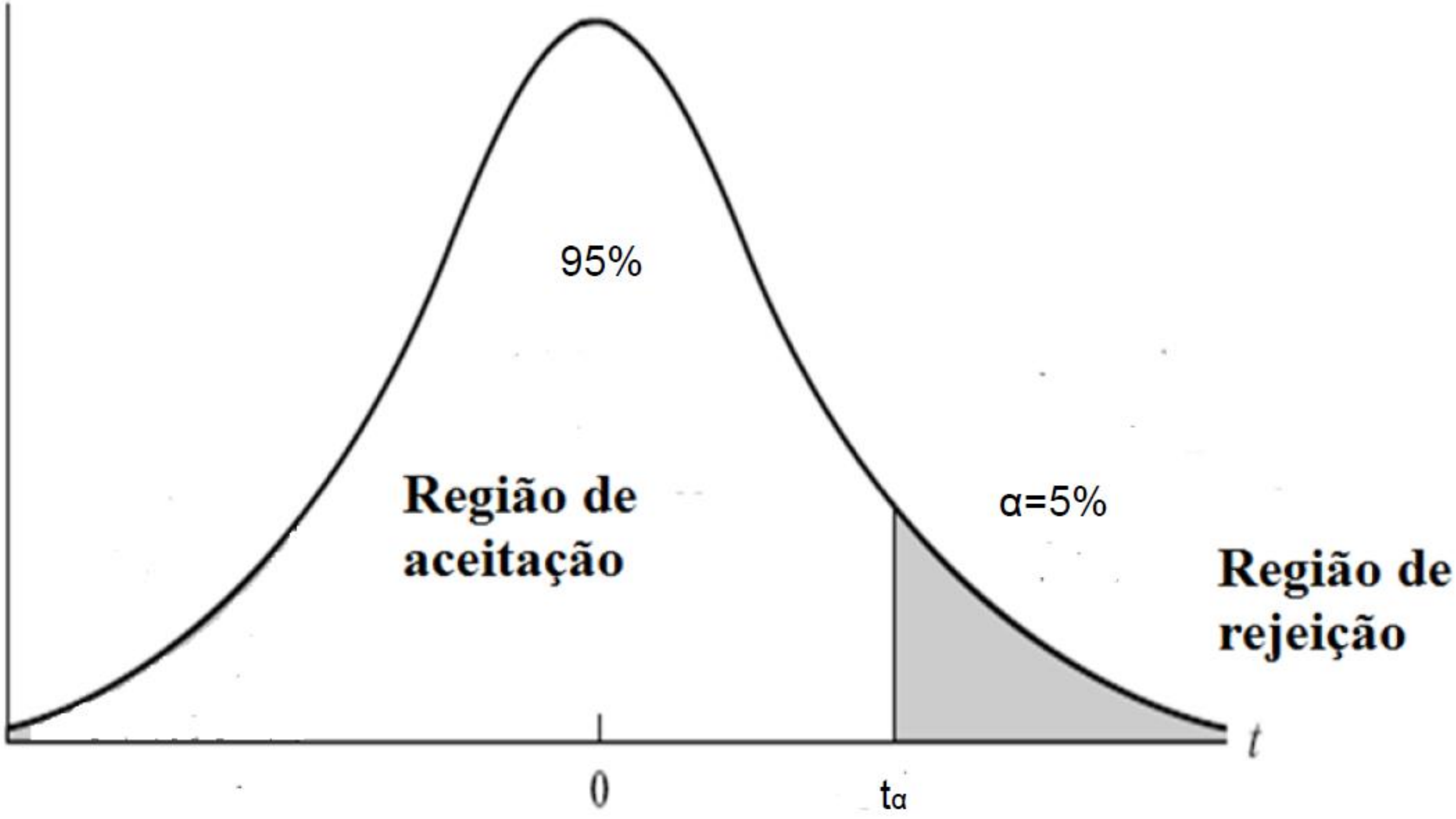
Testes de Hipóteses - Teste t
Alternativas Unilaterais

- **O objetivo é testar:**
  - $H_0: \beta_1 = 0$
  - $H_1: \beta_1 > 0$
  - Rejeita-se H_0 se:

$$t_{\hat{\beta}_j} = \frac{\widehat{\beta}_1}{ep(\widehat{\beta}_1)} > t_\alpha$$

  - onde t_α é o valor obtido de uma distribuição *t* com (*n* – *k*) graus de liberdade. t_α é conhecido como *t* crítico.

Testes de Hipóteses - Teste t
Alternativas Unilaterais



Testes de Hipóteses - Teste t Alternativas Unilaterais

- **O objetivo é testar:**
  - H_0: $\beta_1 = 0$
  - H_1: $\beta_1 < 0$
  - Rejeita-se H_0 se:

$$t_{\hat{\beta}_j} = \frac{\widehat{\beta}_1}{ep(\widehat{\beta}_1)} < -t_\alpha$$

  - onde t_α é o valor obtido de uma distribuição *t* com ($n - k$) graus de liberdade. t_α é conhecido como *t* crítico.

Testes de Hipóteses - Teste t Alternativas Bilaterais

- Como você formularia a hipótese nula de que não há diferença nas taxas de empréstimos entre os bairros devido à composição racial e étnica?
- Como você formularia a hipótese alternativa de que há discriminação nas taxas de empréstimos?

Testes de Hipóteses
Teste contra Hipóteses Alternativas

- Supondo que $\alpha = 0{,}05$ e queremos testar:
  - H_0: $\beta_1 = \theta_j$
  - H_1: $\beta_1 \neq \theta_j$
  - Onde θ_j é o valor hipotético de β_1 .
  - A estatística apropriada para esse teste é:

$$t_{\hat{\beta}_j} = \frac{\widehat{\beta}_1 - \theta_j}{ep(\widehat{\beta}_1)}$$

Modelo de Regressão Linear Múltipla *Baixando os dados – dados em corte*

Fazendo o exercício bicaudal e unicaudal:

```
reg1_mult_b=lm(X$juros~X$IPC+X$PIB+X$fiscal)
```

Bicaudal (p-valor):

```
2*pt(-abs(coef(res)[, 3]), reg1_mult_b$df)
```

Unicaudal (p-valor) para beta <0:

```
pt(coef(res)[,3],reg1_mult_b$df,lower=TRUE)
```

Unicaudal (p-valor) para beta >0:

```
pt(coef(res)[,3],reg1_mult_b$df,lower=FALSE)
```

Testes de Hipóteses
Nível Exato de Significância – *p* valor

- Qual valor de α escolher?
  - Dado o valor da estatística t, qual o mais baixo nível de significância (α) com o qual a hipótese nula pode ser rejeitada?
- Este nível é conhecido com p-valor.
- O **p-valor** pode ser interpretado como sendo a probabilidade de rejeitar H_0 quando ela é verdadeira.
- Ou seja, é a probabilidade de mantermos equivocadamente uma variável explicativa no modelo.

Testes de Hipóteses

Significância Estatística vs. Significância Econômica

- A significância estatística de uma variável é determinada pelo valor da estatística t.
- A significância econômica de uma variável está relacionada à magnitude e ao sinal do coeficiente estimado.
- **Exemplo 1**:

$$\widehat{prate} = 80.29 + 5.44 mrate + 0.269 age + 0.00013\ totemp$$

$$(0.78) \quad (0.52) \qquad\qquad (0.045) \quad (0.00004)$$

$$n = 1{,}534,\ R2 = .100$$

Testes de Hipóteses

Significância Estatística vs. Significância Econômica

- **Exemplo 2:**
  - Variável estatisticamente não significativo aos níveis padrões (10%, 5%, 1%)
  - Variável com efeito esperado e elevado sobre Y.

- **Exemplo 3:**
  - Variável estatisticamente significativa
  - Sinal diferente do esperado.

Testes de Hipóteses – Teste F Restrições Lineares Múltiplas

- Suponha que o objetivo é testar: é:
  - H_0: $\beta_1 = \beta_2 = \beta_3 = ... = \beta_k = 0$
  - H_1: pelo menos um β_j diferente de 0

- A estatística usado para o teste é:

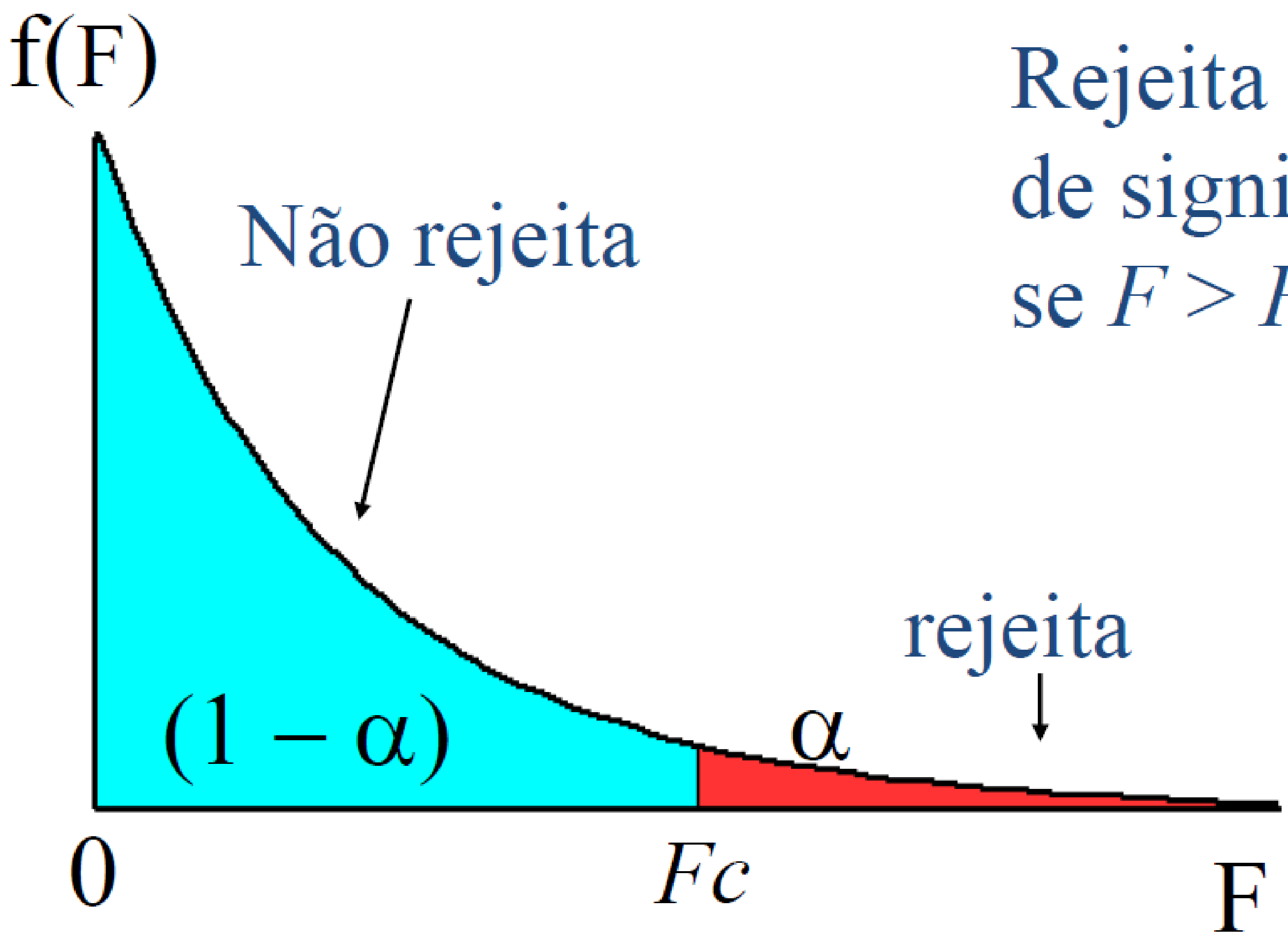
$$F = \frac{(SQR_R - SQR_{NR})/q}{SQR_{NR}/n-k}$$

- q: número de restrições

Testes de Hipóteses – Teste F
Restrições Lineares Múltiplas

- SQR_R representa a soma dos quadrados dos resíduos do modelo restrito
- SQR_{NR} representa a soma dos quadrados dos resíduos do modelo não restrito.
- A estatística F segue uma distribuição *F* com q e *n-k* graus de liberdade.

Teste de Hipótese
Estatística *F*



Rejeita H_0 ao nível de significância α se $F > Fc$

Modelo de Regressão Linear Múltipla *Baixando os dados – dados em corte*

Test F

- summary(reg1_mult) – final da regressão

Test F para dois coeficientes

- install.packages("car")
- library(car)
- linearHypothesis(reg1_mult_b, c("X$PIB=0", "X$fiscal=0"))

O que o teste F nos mostra?

EXERCÍCIO

Modelo de Regressão Linear Múltipla *Exercício*

1. Estime uma regressão múltipla com a base “serie_temporal”
2. Calcule a equação passando pela origem
3. Calcule o intervalo de confiança para algum coeficiente a significância de 95%