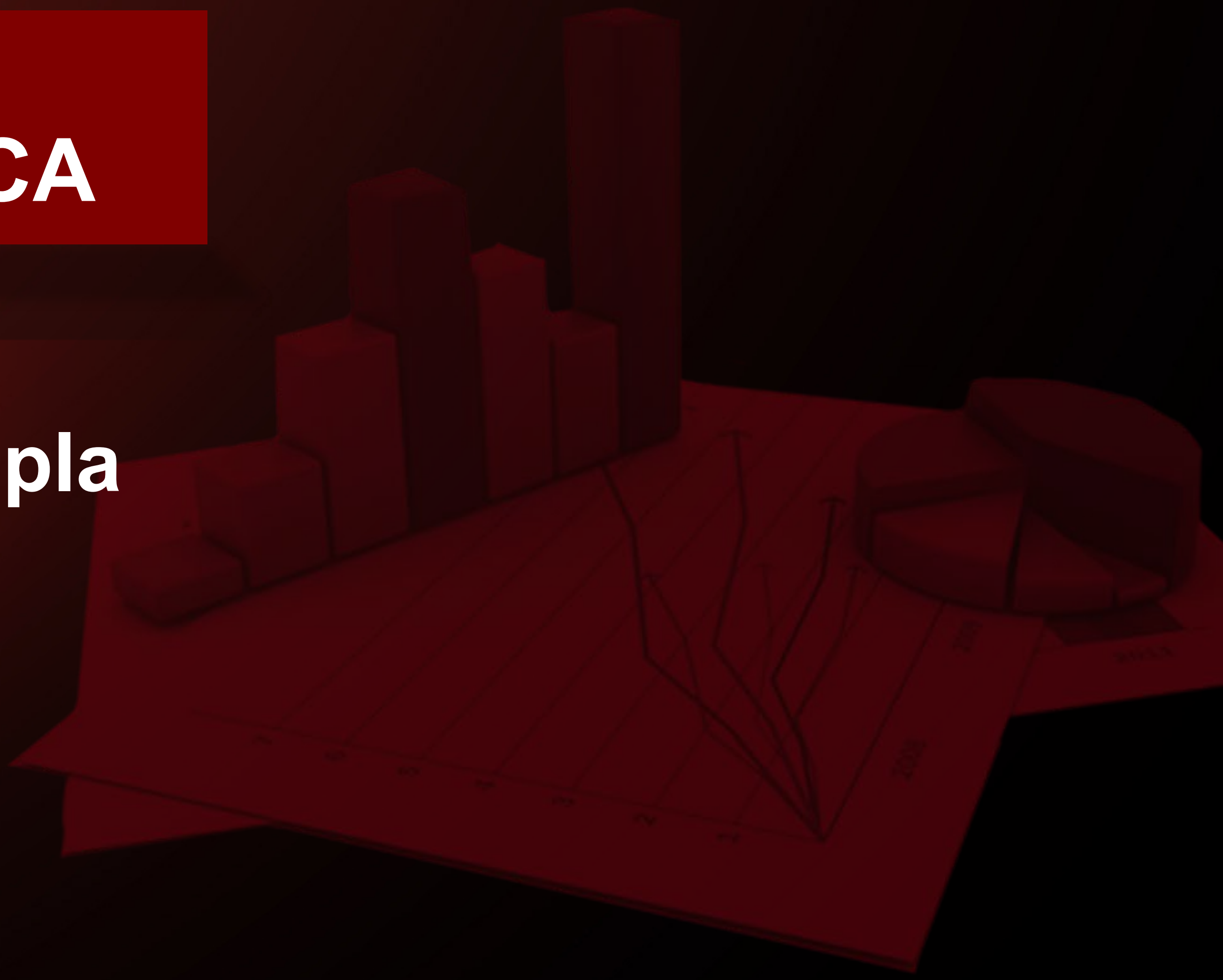
MODELAGEM E INFERÊNCIA ESTATÍSTICA

Modelo de regressão múltipla



O QUE VOU ESTUDAR HOJE?

- Definição e aplicação.
- Modelos com interação e preditores quadráticos.
- Modelos com preditores para variáveis categóricas.
- Estimação de parâmetros e inferências.

MODELO DE REGRESSÃO MÚLTIPLA

A equação do modelo de regressão múltipla aditiva geral é

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \varepsilon$$

em que $E_{(\varepsilon)} = 0$ e $V_{(\varepsilon)} = \sigma^2$. Além disso, para testar hipóteses e calcular IC ou IP, assume-se que ε e tenha distribuição normal.

Valor esperado:

$$\mu_{y.x_1*,\ldots x_2*} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \varepsilon$$

MODELOS COM INTERAÇÃO E PREDITORES QUADRÁTICOS

Para um conjunto de variáveis y, x_1 ex_2:

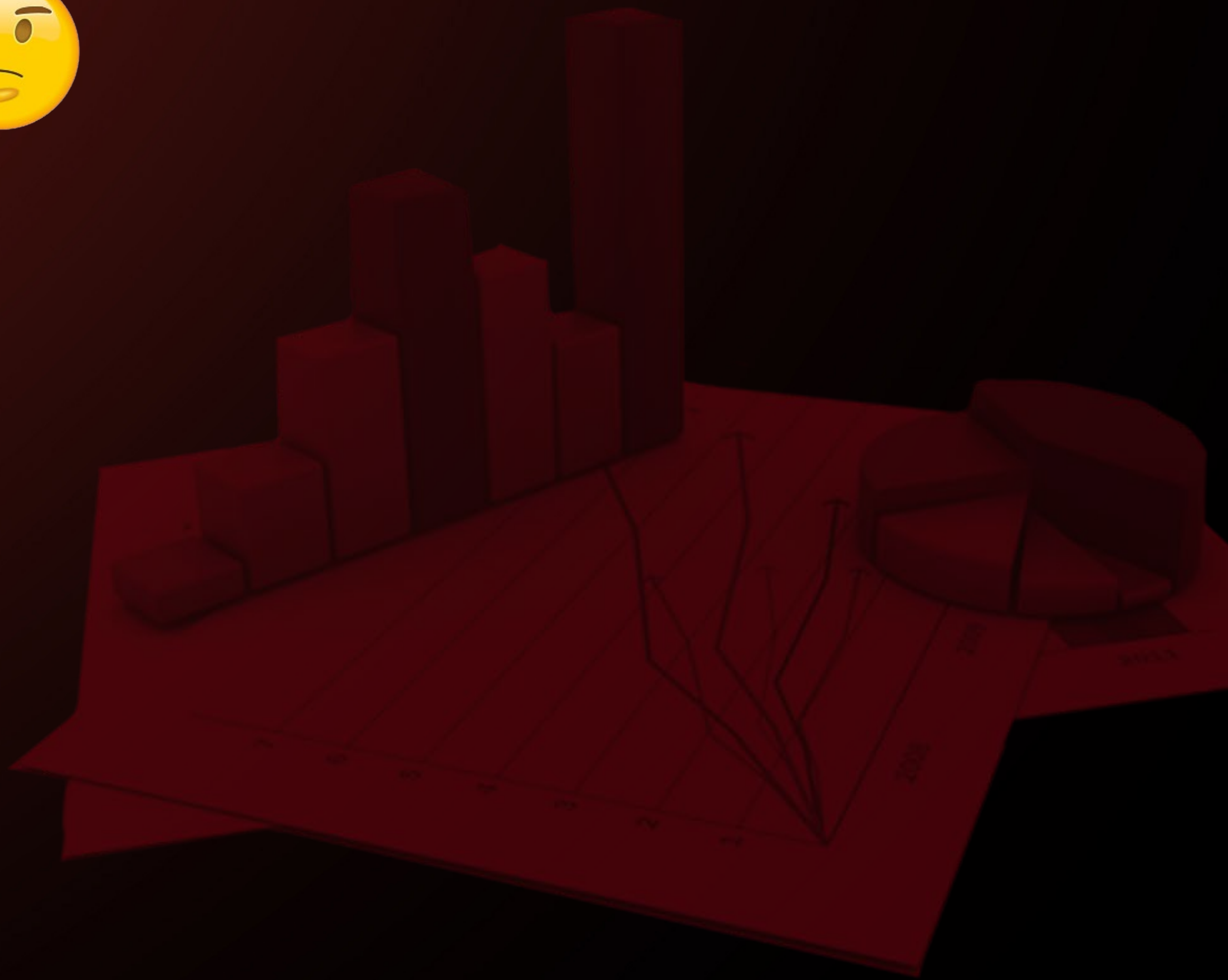
$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$$

Com x_1e x_2 pode ser elaborado outro modelo, Por exemplo, $x_3 = x_1^2$ e $x_4 = x_1 x_4$

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \varepsilon$$

A regressão polinomial, também é um caso especial da regressão múltipla

ONDE SE APLICA?



MODELOS COM INTERAÇÃO E PREDITORES QUADRÁTICOS

Para um conjunto de variáveis y, x_1 e x_2 um conjunto de quatro modelos possíveis:

O modelo de primeira ordem:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$$

O modelo de segunda ordem sem interação:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1^2 + \beta_4 x_2^2 + \varepsilon$$

O modelo com preditores de primeira ordem e interação:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + \varepsilon$$

O modelo completo de segunda ordem ou modelo quadrático completo:

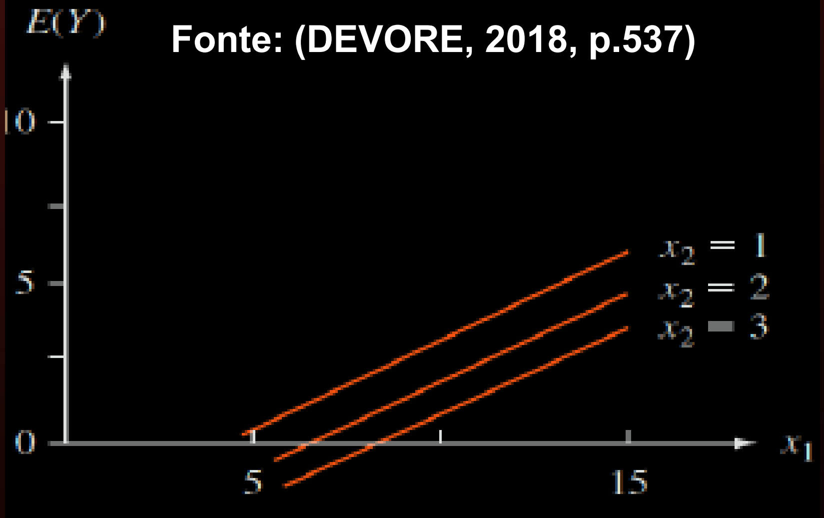
$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1^2 + \beta_4 x_2^2 + \beta_5 x_1 x_2 + \varepsilon$$

MODELOS COM INTERAÇÃO E PREDITORES QUADRÁTICOS

O modelo de primeira ordem: $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$

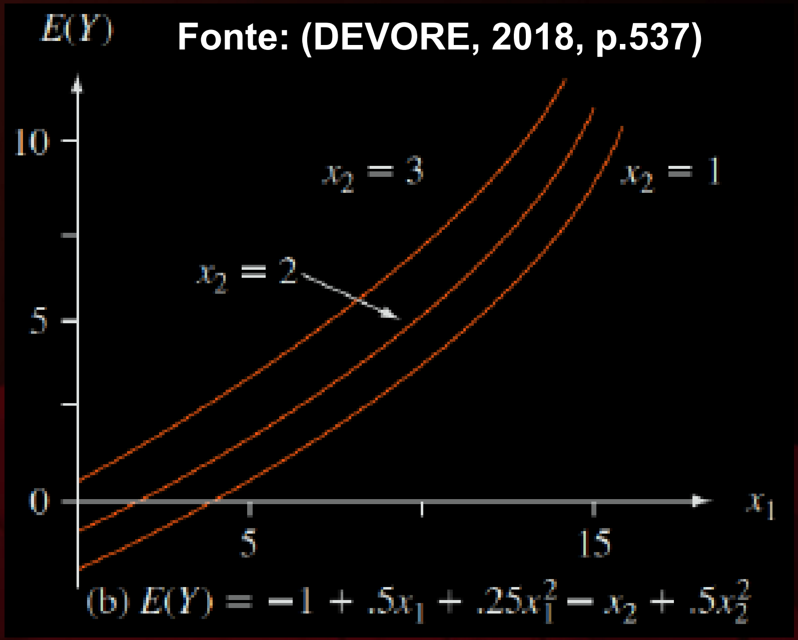
- Valor esperado de y é uma função linear da outra variável.
- Para um valor fixo de x_i: A mudança esperada em y para um aumento de uma unidade em $x_1(x_2)$ é $\beta_1(\beta_2)$ independentemente do nível de x_2 (x_1)
- Diferentes valores de x_2 resultam em um conjunto de retas paralelas.

Fonte: (DEVORE, 2018, p.537)



(a) $E(Y) = -1 + .5x_1 - x_2$

MODELOS COM INTERAÇÃO E PREDITORES QUADRÁTICOS



O modelo de segunda ordem sem interação:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1^2 + \beta_4 x_2^2 + \varepsilon$$

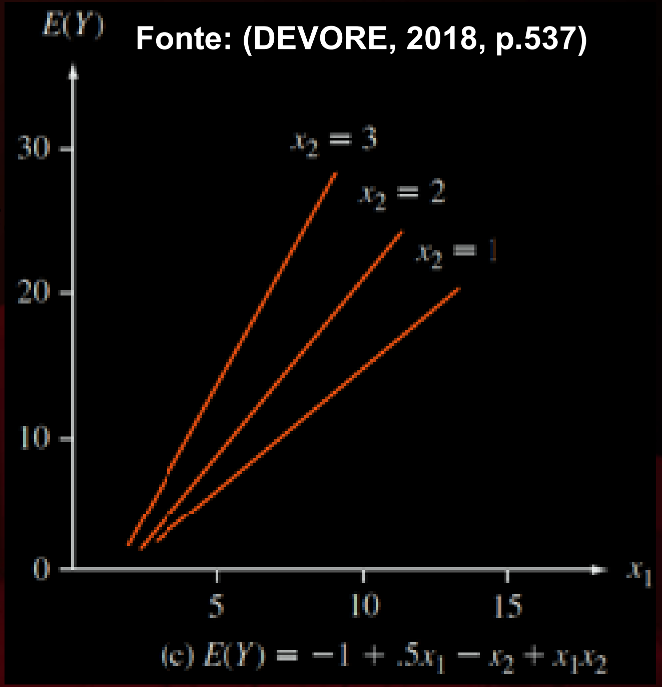
- **A mudança esperada em Y quando x_1 sofre um aumento de 1**

$$\beta_0 + \beta_1(x_1 + 1) + \beta_2 x_2 + \beta_3 (x_1 + 1)^2 + \beta_4 x_2^2$$

$$-(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1^2 + \beta_4 x_2^2) = \boldsymbol{\beta_1 + \beta_3 + 2\beta_3 x_1}$$

- **A dependência da mudança esperada em relação ao valor de x_1 significa que os contornos são agora curvas**

MODELOS COM INTERAÇÃO E PREDITORES QUADRÁTICOS



Fonte: (DEVORE, 2018, p.537)

O modelo com preditores de primeira ordem e interação:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + \varepsilon$$

- A mudança esperada em Y quando x_1 sofre um aumento de 1

$$\beta_0 + \beta_1(x_1 + 1) + \beta_2 x_2 + \beta_3 x_2(x_1 + 1)$$
$$-(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2) = \boldsymbol{\beta_1 + \beta_3 x_2}$$

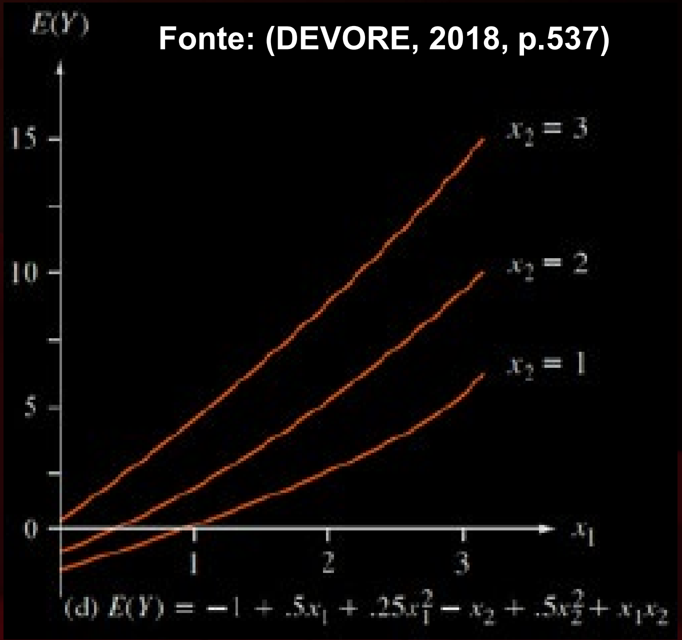
- Linhas retas não paralelas
- A palavra interação reflete o fato de que uma mudança esperada em Y quando uma variável aumenta em valor depende do valor da outra variável.

MODELOS COM INTERAÇÃO E PREDITORES QUADRÁTICOS

- **O modelo completo de segunda ordem ou modelo quadrático completo:**

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1^2 + \beta_4 x_2^2 + \beta_5 x_1 x_2 + \varepsilon$$

- A mudança esperada em Y quando x_1 sofre um aumento de 1: $\beta_1 + \beta_3 + 2\beta_3 x_1 + \beta_5 x_2$
- Os contornos da função de regressão são curvos e não paralelos



Fonte: (DEVORE, 2018, p.537)

MODELOS COM PREDITORES PARA VARIÁVEIS CATEGÓRICAS

- Usar variável *dummy* ou indicadora com 0 e 1

Exemplos de variáveis categóricas:

- Faixa etária: criança e adulto
- Masculino ou feminino
- Dia ensolarado ou dia chuvoso
- Chave philips ou fenda

Exemplos que abrangem uma terceira categoria:

- Pet: gato, cachorro, outro
  - X1= 1 gato
  - X2= 1 cachorro
  - X1=0 e X2 = 0 outro pet
- Localização: aguardando produto, caminho, no local de entrega.

COEFICIENTE DE CORRELAÇÃO R^2 E R_a^2

- **Os valores do coeficiente de determinação multivariado:**

$$R^2 = 1 - \frac{SQE}{SQT}$$

podem ser inflados pela adição de muitos preditores no modelo.

- **Portanto deve-se calcular o R^2-ajustado:**

$$R_a^2 = 1 - \frac{n-1}{n-(k+1)} \frac{SQE}{SQT}$$

- $\mathrm{R_a^2}$ **pode ser positivo ou negativo**
- $0 < \mathrm{R^2} < 1$
- $\mathrm{R_a^2} < \mathrm{R^2}$
- **(+) R é o coeficiente de correlaçãomúltiplo** $(\hat{y}_i, y_i)$

VARIÂNCIA ESTIMADA s²

A variância estimada:

$$\hat{\sigma}^2 = s^2 = \frac{SQE}{n-(k+1)}$$

Exemplo para calcular gl=n-(k+1)

$$\mathbf{y} = \boldsymbol{\beta}_0 + \boldsymbol{\beta}_1\mathbf{x}_1 + \boldsymbol{\beta}_2\mathbf{x}_2 + \boldsymbol{\beta}_3\mathbf{x}_1\mathbf{x}_2 + \boldsymbol{\varepsilon}$$

- **Para uma amostra de 29 dados**
- **Na equação exemplo k=3**
- **Assim os graus de liberdade gl=29-(3+1) → gl=25**

TESTE DE UTILIDADE DO MODELO

Hipótese nula: H_0: $\beta_1 = \beta_2 = \cdots = \beta_k = 0$

Hipótese alternativa: H_a: pelo menos um $\beta_i \neq 0$ $(i = 1, \ldots, k)$

Valor da estatística de teste: $$f = \frac{R^2/k}{(1 - R^2)/[n - (k+1)]} \tag{13.19}$$

$$= \frac{\text{SQR}/k}{\text{SQE}/[n - (k+1)]} = \frac{\text{QMR}}{\text{QME}}$$

onde SQR = soma de quadrados de regressão = SQT – SQE

Quando H_0 é verdadeira, a estatística de teste F tem uma distribuição F com k gl no numerador e $n - (k+1)$ gl no denominador. O teste é unilateral à direita, de modo que o valor-*p é a área sob a curva* $F_{k,\, n-(k+1)}$ à direita de f.

Fonte: (DEVORE, 2018, p.544)

TESTE DE UTILIDADE DO MODELO

Exemplo:

- Para uma amostra de 29 dados
- Na equação exemplo k=3
- Assim os graus de liberdade gl= 29-(3+1) → gl= 25
- Estatística de teste de f= 15,6

http://www.socr.ucla.edu/Applets.dir/F_Table.html

Para $\alpha = 0.001$ $F_{crit} = 6{,}49$		ν_1 = numerator df				
	α	1	2	3	4	5
25	.100	2.92	2.53	2.32	2.18	2.09
	.050	4.24	3.39	2.99	2.76	2.60
	.010	7.77	5.57	4.68	4.18	3.85
	.001	13.88	9.22	7.45	6.49	5.89
26	.100	2.91	2.52	2.31	2.17	2.08
	.050	4.23	3.37	2.98	2.74	2.59
	.010	7.72	5.53	4.64	4.14	3.82
	.001	13.74	9.12	7.36	6.41	5.80

$$f \geq F_{crit} \text{ rejeitar } H_0$$

TESTE DE UTILIDADE DO MODELO

- Para uma amostra de 29 dados
- Na equação exemplo k=3
- Assim os graus de liberdade gl=29-(3+1) → gl=25
- Estatística de teste de f=15,6
  http://www.socr.ucla.edu/Applets.dir/F_Table.html

RESOLVENDO:

- Analisar se $\mathrm{f} \geq \mathrm{F}_{\mathrm{crit}}$ rejeitar H_0
- f=15,6 e $\mathrm{F}_{\mathrm{crit}} = 6,49$
- 15,6$\geq 6,49$ SIM, portanto rejeitar a hipótese nula H_0 isto é os parâmetros são diferentes de zero, e o modelo multivariado pode ser usado.

INTERVALOS DE CONFIANÇA E PREVISÃO

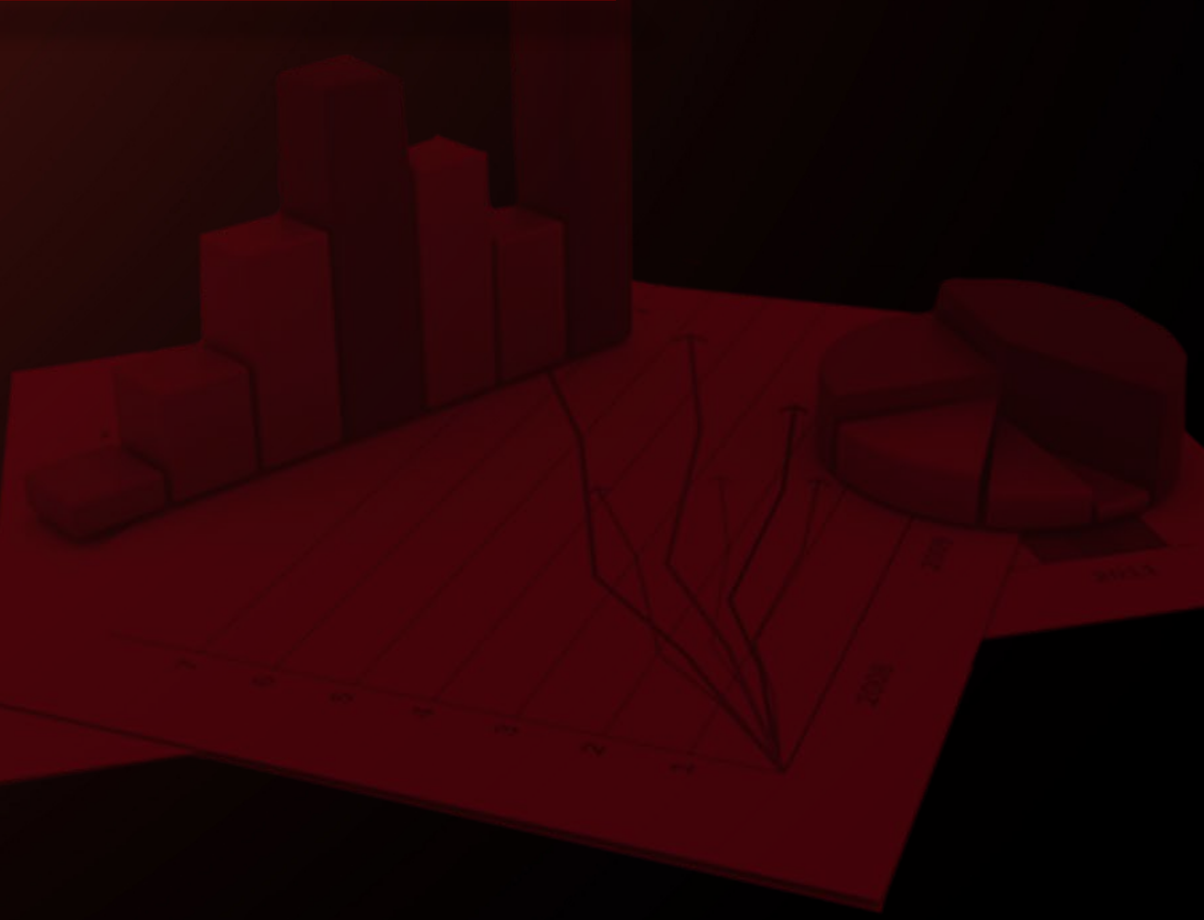
IC deve ser calculado mediante:

- Para um parâmetro: $\hat{\beta}_i \pm t_{\frac{\alpha}{2},(n-(k+1))} s_{\hat{\beta}_i}$
- Para o valor esperado $\mu_{y \cdot x_1^*, \ldots, x_k^*}$

$$\hat{y} \pm t_{\frac{\alpha}{2},(n-(k+1))} s_{\hat{y}}$$

IP deve ser calculado mediante:

$$\hat{y} \pm t_{\frac{\alpha}{2},(n-(k+1))} \sqrt{s_{\hat{y}}^2 + s^2}$$



MODELAGEM E INFERÊNCIA ESTATÍSTICA

Modelo de regressão múltipla