Modelos Lineares I

Regressão Linear Simples (RLS):

(4ª, 5ª e 6ª Aulas)

Professor: Dr. José Rodrigo de Moraes

Universidade Federal Fluminense (UFF)

Departamento de Estatística (GET)

1

Estimação da variância σ^2 dos erros do modelo:

Introdução:

- A variância σ^2 dos erros do modelo de regressão é de extrema importância para a realização de inferências na análise de regressão.
- É necessário obter alguma informação sobre a variabilidade da distribuição de probabilidade da variável resposta Y do modelo.

2

Estimação da variância σ^2 dos erros do modelo:

Vimos que os resíduos do modelo são definidos por:

$$e_i = Y_i - \hat{Y}_i, \quad \forall \ i = 1,2,...,n$$

Assim a soma dos quadrados dos resíduos (SQRes) é definida por:

$$SQRes = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2$$

Essa soma envolveu em seu cálculo a estimação de dois coeficientes de regressão β_0 e β_1, e portanto 2 gl`s foram perdidos. Desse modo, ao dividir a SQRes por n-2 (graus de liberdade), obtém-se o chamado "*Quadrado Médio dos Resíduos*" (QMRes), dado por:

3

Estimação da variância dos erros, isto é, VAR(ε_i)= σ^2:

$$QMRes = \frac{SQRes}{n-2} = \frac{\sum_{i=1}^{n} e_i^2}{n-2} = \frac{\sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2}{n-2}$$

Será demonstrado mais adiante que o QMRes, também representado, alternativamente, por $\hat{\sigma}^2$, é um estimador não viciado da variância dos erros e, portanto, da variância da variável resposta Y_i do modelo de regressão, já que $VAR(Y_i)=VAR(\varepsilon_i)=\sigma^2$. Ou seja:

$$E(QMRes) = E\left(\frac{\sum_{i=1}^{n} e_i^2}{n-2} \right) = \sigma^2$$

4

Modelo de Regressão Linear Normal:

- As hipóteses gerais do modelo de RLS descritas anteriormente estabeleciam que os erros aleatórios do modelo apresentavam média zero, variância constante σ^2 e $COV(\varepsilon_i,\varepsilon_j)=0 \ \forall \ i \neq j$ (erros não correlacionados).
- Não foi feita nenhuma hipótese sobre a possível *distribuição de probabilidade dos erros*.
- Acrescendo ao conjunto de hipóteses, a hipótese de normalidade dos erros, diremos que $Y_i=\beta_0+\beta_1 X_i + \varepsilon_i$ é um modelo de regressão linear (simples) normal, isto é, um modelo com erros normalmente distribuídos.
- A suposição de que os erros tem distribuição normal simplifica substancialmente a teoria de análise de regressão, e em muitos casos é plenamente justificada na prática.

5

Representação genérica do modelo de RLS normal:

- Modelo de regressão linear normal: $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$
  - $Y_i \rightarrow$ valor observado da variável resposta do i-ésimo elemento da amostra.
  - β_0 e $\beta_1 \rightarrow$ são os parâmetros desconhecidos a serem estimados com base na amostra.
  - $X_i \rightarrow$ valor observado da variável explicativa do i-ésimo elemento da amostra.
  - $\varepsilon_i \rightarrow$ erro aleatório do modelo referente ao i-ésimo elemento da amostra. Os erros ε_i`s são supostamente independentes e normalmente distribuídos com média 0 e variância σ^2, isto é : $\varepsilon_i \sim N(0,\sigma^2)$, i=1,2,...,n

6

Exemplo: Utilizando o método de estimação de máxima verossimilhança (MV) estudado nas disciplinas "Estatística II" e "Inferência Estatística" , pede-se:

a) Obtenha os estimadores dos parâmetros β_0 e β_1 do modelo de regressão linear normal

b) Obtenha o estimador de σ^2.

c) Compare os estimadores obtidos por MV com os obtidos pelo método de Mínimos Quadrados (MQ). Qual a conclusão obtida ?

7

Exemplo: Dados sobre a concentração da substância X (mg/L) e ganho de peso Y (kg) de n=30 bois:

Resultados do Ajuste (n=30 bois) usando o SPSS 17.0 – *Statistical Package for the Social Sciences: Analyse /Generalized Linear Models*

Parameter Estimates

Parameter	B	Std. Error	95% Wald Confidence Interval		Hypothesis Test		
			Lower	Upper	Wald Chi-Square	df	Sig.
(Intercept)	10,040	,4783	9,102	10,977	440,530	1	,000
conc_subs_X	,732	,0731	,589	,875	100,236	1	,000
(Scale)	1,248[a]	,3221	,752	2,070			

Dependent Variable: ganho_peso_Y
Model: (Intercept), conc_subs_X

a. Maximum likelihood estimate.

Estimativas dos parâmetros por MV

8

Modelo de Regressão Linear Normal (Inferência sobre os parâmetros do modelo):

- ❑ Agora vamos realizar inferências sobre os parâmetros do modelo β_0 e β_1, através da:
  - ✓ Construção de intervalos de confiança (IC`s)
  - ✓ Realização de testes de hipóteses (TH`s)
- ❑ Definir as distribuições de probabilidade dos estimadores $\hat{\beta}_0$ e $\hat{\beta}_1$, e de suas funções.

9

Distribuição amostral de $\hat{\beta}_1$:

❑ Vimos que o estimador de MQ de β_1 é dado pela seguinte expressão:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(X_i - \bar{X})^2}$$

Desenvolvendo o <u>numerador</u> da expressão:

$$\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y}) = \sum_{i=1}^{n}(X_i - \bar{X})Y_i$$

10

Distribuição amostral de $\hat{\beta}_1$:

❑ Substituindo na expressão do estimador de β_1, temos que:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(X_i - \bar{X})^2} = \frac{\sum_{i=1}^{n}(X_i - \bar{X})Y_i}{\sum_{i=1}^{n}(X_i - \bar{X})^2} = \sum_{i=1}^{n}V_iY_i$$

❑Portanto, temos a seguinte expressão alternativa para o estimador de β_1:

$$\hat{\beta}_1 = \sum_{i=1}^{n}V_iY_i$$

11

<u>**Propriedades importantes:**</u>

$$V_i = \frac{(X_i - \bar{X})}{\sum_{i=1}^{n}(X_i - \bar{X})^2}$$

❑. $\sum_{i=1}^{n}V_i = 0$

❑. $\sum_{i=1}^{n}V_iX_i = \sum_{i=1}^{n}\left[\frac{(X_i - \bar{X})}{\sum_{i=1}^{n}(X_i - \bar{X})^2}\right]X_i = 1$

12

Distribuição amostral de $\hat{\beta}_1$:

O estimador $\hat{\beta}_1$ pode ser expresso por:

$$\hat{\beta}_1 = \beta_1 + \sum_{i=1}^{n} V_i\, \varepsilon_i$$

Fórmula alternativa de $\hat{\beta}_1$

- **Cálculo da média do estimador $\hat{\beta}_1$:**

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

OBS: $\hat{\beta}_1$ é um estimador não viciado para o parâmetro β_1 do modelo.

13

Distribuição amostral de $\hat{\beta}_1$:

- **Cálculo da variância do estimador $\hat{\beta}_1$:**

$$VAR(\hat{\beta}_1) = \frac{\sigma^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2}$$

Propriedade:

$$\sum_{i=1}^{n} V_i^2 = \frac{1}{\sum_{i=1}^{n}(X_i - \bar{X})^2}$$

14

Distribuição amostral de $\hat{\beta}_1$:

- Podemos verificar que: $\hat{\beta}_1 = \sum_{i=1}^{n} V_i Y_i$

é uma combinação linear das variáveis aleatórias Y_i\`s, i=1,2,...,n. Baseadas nas hipóteses do **modelo de regressão linear normal**, as v.a\`s Y_i \`s são independentes e tem distribuição normal com média $E(Y_i)=\beta_0+\beta_1 X_i$ e variância $VAR(Y_i)=\sigma^2$. Portanto, $\hat{\beta}_1$ também terá distribuição normal:

$$\hat{\beta}_1 \sim N\left[\beta_1 , \frac{\sigma^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2}\right]$$

15

Distribuição amostral de $\hat{\beta}_0$:

- Vimos que o estimador de MQ de β_0 é dado pela seguinte expressão:

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$$

Reescrevendo:

$$\hat{\beta}_0 = \frac{1}{n}\sum_{i=1}^{n} Y_i - \bar{X}\sum_{i=1}^{n} V_i\, Y_i = \sum_{i=1}^{n}\left(\frac{1}{n} - \bar{X}V_i\right) Y_i$$

$$\hat{\beta}_0 = \beta_0 + \sum_{i=1}^{n}\left(\frac{1}{n} - \bar{X}V_i\right)\varepsilon_i$$

16

Distribuição amostral de $\hat{\beta}_0$:

O estimador $\hat{\beta}_0$ pode ser expresso por:

$$\hat{\beta}_0 = \beta_0 + \sum_{i=1}^{n}\left(\frac{1}{n} - \bar{X}V_i\right)\varepsilon_i$$

Fórmula alternativa de $\hat{\beta}_0$

- **Cálculo da média do estimador $\hat{\beta}_0$:**

$$E(\hat{\beta}_0) = \beta_0$$

OBS: $\hat{\beta}_0$ é um estimador não viciado para o parâmetro β_0 do modelo.

17

Distribuição amostral de $\hat{\beta}_0$:

- **Cálculo da variância do estimador $\hat{\beta}_0$:**

$$VAR(\hat{\beta}_0) = E[\hat{\beta}_0 - E(\hat{\beta}_0)]^2 \rightarrow VAR(\hat{\beta}_0) = E[\hat{\beta}_0 - \beta_0]^2$$

$$(\hat{\beta}_0 - \beta_0)^2 = \sum_{i=1}^{n}\left(\frac{1}{n} - \bar{X}V_i\right)^2 \varepsilon_i^2 + 2\sum_{i<j}^{n}\left(\frac{1}{n} - \bar{X}V_i\right)\left(\frac{1}{n} - \bar{X}V_j\right)\varepsilon_i\varepsilon_j$$

$$VAR(\hat{\beta}_0) = E(\hat{\beta}_0 - \beta_0)^2 = \sum_{i=1}^{n}\left(\frac{1}{n} - \bar{X}V_i\right)^2 \cdot E(\varepsilon_i^2) + 2\sum_{i<j}^{n}\left(\frac{1}{n} - \bar{X}V_i\right)\left(\frac{1}{n} - \bar{X}V_j\right)\cdot E(\varepsilon_i\varepsilon_j)$$

18

Distribuição amostral de $\hat{\beta}_0$:

❑ **Cálculo da variância do estimador $\hat{\beta}_0$ (*continuação*):**

Como $E(\varepsilon_i)=0 \ \forall \ i=1,2,...,n$ e $E(\varepsilon_i \varepsilon_j)=0 \ \forall \ i \neq j$, então:

$VAR(\varepsilon_i)=E(\varepsilon_i^2) = \sigma^2 \ \forall \ i=1,2,...,n$

$$VAR(\hat{\beta}_0) = \sum_{i=1}^{n}\left(\frac{1}{n}-\bar{X}V_i\right)^2\sigma^2 = \sigma^2\left(\frac{1}{n}+\frac{\bar{X}^2}{\sum_{i=1}^{n}(X_i-\bar{X})^2}\right)$$

19

Distribuição amostral de $\hat{\beta}_0$:

❑ Analogamente, podemos verificar que: $\hat{\beta}_0 = \sum_{i=1}^{n}\left(\frac{1}{n}-\bar{X}V_i\right)Y_i$

é uma combinação linear das variáveis aleatórias Y_i \`s, i=1,2,...,n. Baseadas nas hipóteses do modelo de regressão linear normal, as v.a\`s Y_i \`s são independentes e tem distribuição normal com média $E(Y_i)=\beta_0+\beta_1 X_i$ e variância $VAR(Y_i)=\sigma^2$. Portanto, $\hat{\beta}_0$ também terá distribuição normal:

$$\hat{\beta}_0 \sim N\left[\beta_0 \, , \, \sigma^2\left(\frac{1}{n}+\frac{\bar{X}^2}{\sum_{i=1}^{n}(X_i-\bar{X})^2}\right)\right]$$

20

Covariância entre os estimadores $\hat{\beta}_0$ e $\hat{\beta}_1$:

❑ A covariância entre os estimadores dos parâmetros β_0 e β_1 do modelo:

$$COV(\hat{\beta}_0,\hat{\beta}_1) = E\left[\left(\hat{\beta}_0 - E(\hat{\beta}_0)\right)\left(\hat{\beta}_1 - E(\hat{\beta}_1)\right)\right]$$

$$COV(\hat{\beta}_0,\hat{\beta}_1) = E\left[(\hat{\beta}_0 - \beta_0)(\hat{\beta}_1 - \beta_1)\right]$$

Sabe-se que:

(1) $\hat{\beta}_0 = \beta_0 + \sum_{i=1}^{n}\left(\frac{1}{n}-\bar{X}V_i\right)\varepsilon_i \quad \rightarrow \quad \hat{\beta}_0 - \beta_0 = \sum_{i=1}^{n}\left(\frac{1}{n}-\bar{X}V_i\right)\varepsilon_i$

(2) $\hat{\beta}_1 = \beta_1 + \sum_{i=1}^{n}V_i\,\varepsilon_i \quad \rightarrow \quad \hat{\beta}_1 - \beta_1 = \sum_{i=1}^{n}V_i\,\varepsilon_i$

21

Covariância entre os estimadores $\hat{\beta}_0$ e $\hat{\beta}_1$:

❑ Fazendo as devidas demonstrações, pode-se provar que a covariância entre os estimadores dos parâmetros β_0 e β_1 é dada por:

$$COV(\hat{\beta}_0,\hat{\beta}_1) = \frac{-\bar{X}\sigma^2}{\sum_{i=1}^{n}(X_i-\bar{X})^2}$$

Pergunta: Qual relação existe entre a média de X e a covariância definida acima ?

22

Intervalo de Confiança para o parâmetro β_1 :

❑ Vimos que:

$$\hat{\beta}_1 \sim N\left[\beta_1 \, , \, \frac{\sigma^2}{\sum_{i=1}^{n}(X_i-\bar{X})^2}\right]$$

Logo: $Z = \dfrac{(\hat{\beta}_1-\beta_1)\sqrt{\sum_{i=1}^{n}(X_i-\bar{X})^2}}{\sigma} \sim N(0,1)$

OBS: A v.a Z depende de σ.

23

Intervalo de Confiança para o parâmetro β_1 :

❑ Aliada ao fato de que:

$$\frac{(n-2)\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{(n-2)}$$

tem distribuição quiquadrada com (n-2) graus de liberdade, obtemos uma nova variável aleatória obtida abaixo:

$$T = \frac{(\hat{\beta}_1-\beta_1)\sqrt{\sum_{i=1}^{n}(X_i-\bar{X})^2}}{\sigma} \Bigg/ \sqrt{\frac{\frac{(n-2)\hat{\sigma}^2}{\sigma^2}}{(n-2)}} = \frac{(\hat{\beta}_1-\beta_1)\sqrt{\sum_{i=1}^{n}(X_i-\bar{X})^2}}{\hat{\sigma}}$$

A v.a T tem distribuição de Student com (n-2) graus de liberdade

Lembre-se que: $\hat{\sigma} = \sqrt{QMRes} = \sqrt{\dfrac{\sum_{i=1}^{n}e_i^2}{n-2}}$

24

Intervalo de Confiança para o parâmetro β_1:

- Para construir um intervalo de confiança (IC) para β_1 ao nível de confiança de 100(1-α)% calcula-se a probabilidade abaixo:

$$P\left[-t_{\alpha/2,n-2} \le \frac{(\hat{\beta}_1-\beta_1)\sqrt{\sum_{i=1}^{n}(X_i-\overline{X})^2}}{\hat{\sigma}} \le t_{\alpha/2,n-2}\right] = 1-\alpha$$

$$P\left[\hat{\beta}_1 - t_{\alpha/2,n-2}\frac{\hat{\sigma}}{\sqrt{\sum_{i=1}^{n}(X_i-\overline{X})^2}} \le \beta_1 \le \hat{\beta}_1 + t_{\alpha/2,n-2}\frac{\hat{\sigma}}{\sqrt{\sum_{i=1}^{n}(X_i-\overline{X})^2}}\right] = 1-\alpha$$

25

Intervalo de Confiança para o parâmetro β_1 :

- Logo o um Intervalo de confiança para o parâmetro β_1 do modelo, ao nível de confiança de 100(1-α)%, é dado por:

$$IC_{\beta_1,100(1-\alpha)\%} = \left[\underbrace{\hat{\beta}_1 - t_{\alpha/2,n-2}\frac{\hat{\sigma}}{\sqrt{\sum_{i=1}^{n}(X_i-\overline{X})^2}}}_{\text{Limite inferior } (L_{inf}) \text{ do intervalo}} \; , \; \underbrace{\hat{\beta}_1 + t_{\alpha/2,n-2}\frac{\hat{\sigma}}{\sqrt{\sum_{i=1}^{n}(X_i-\overline{X})^2}}}_{\text{Limite superior } (L_{sup}) \text{ do intervalo}}\right]$$

26

Exemplo: Considerando os dados dos n=30 bois, construa um intervalo de confiança de 95% para o parâmetro β_1 do modelo de RLS.

Coefficients[a]

Model		Unstandardized Coefficients B	Std. Error	Standardized Coefficients Beta	t	Sig.	95,0% Confidence Interval for B Lower Bound	Upper Bound
1	(Constant)	10,040	,495		20,277	,000	9,025	11,054
	X_conc.subs	,732	,076	,877	9,672	,000	,577	,887

a. Dependent Variable: Y_ganho_peso

IC para β_1 ao nível de 95%

27

Testes de Hipóteses para o parâmetro β_1:

- **Hipóteses a serem testadas:**

$$\begin{cases} H_0: \beta_1 = 0 \\ H_1: \beta_1 \neq 0 \end{cases}$$

- **Estatística de Teste:**

$$T = \frac{\hat{\beta}_1\sqrt{\sum_{i=1}^{n}(X_i-\overline{X})^2}}{\hat{\sigma}} \sim T_{n-2}$$

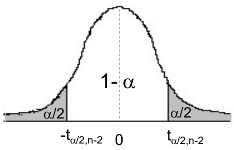
- A Estatística T tem distribuição de Student com (n-2) graus de liberdade (g.l`s).

28

Testes de Hipóteses para o parâmetro β_1:

- **Região crítica:**

$$RC = \left\{ t \in \Re \;/\; t \le -t_{\alpha/2,n-2} \quad \text{ou} \quad t \ge t_{\alpha/2,n-2} \right\}$$



- **Tomada de Decisão:**
  - Se $t_{obs} \in$ RC rejeita-se H_0:β_1=0 ao nível de significância α, e conclui-se que existe relação linear significativa entre X e Y.
  - Se $t_{obs} \notin$ RC não há evidências para rejeitar H_0:β_1=0 ao nível de significância α, e conclui-se que não existe relação linear significativa entre X e Y.

29

Exemplo: Considerando os dados dos n=30 bois, realize um teste estatístico de hipóteses para o parâmetro β_1 ao nível de significância α de 5%.

Coefficients[a]

Model		Unstandardized Coefficients B	Std. Error	Standardized Coefficients Beta	t	Sig.	95,0% Confidence Interval for B Lower Bound	Upper Bound
1	(Constant)	10,040	,495		20,277	,000	9,025	11,054
	X_conc.subs	,732	,076	,877	9,672	,000	,577	,887

a. Dependent Variable: Y_ganho_peso

TH para β_1 ao nível de 5%

30

Testes de Hipóteses para o parâmetro β_1:

No caso do analista desejar testar se o parâmetro β_1 do modelo é igual a algum valor de interesse (β_1*), realiza-se o seguinte teste de hipóteses:

- **Hipóteses a serem testadas:**

$$\begin{cases} H_0: \beta_1 = \beta_1^* \\ H_1: \beta_1 \neq \beta_1^* \end{cases}$$

- **Estatística de Teste:**

$$T = \frac{\left(\hat{\beta}_1 - \beta_1^*\right)\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2}}{\hat{\sigma}} \sim T_{n-2}$$

31

Testes de Hipóteses para o parâmetro β_1 :

- **Região crítica:**

$$RC = \left\{ t \in \Re \ / \ t \le -t_{\alpha/2,n-2} \quad \text{ou} \quad t \ge t_{\alpha/2,n-2} \right\}$$

ou alternativamente:

$$RC = \left\{ t \in \Re \ / \ |t| \ge t_{\alpha/2,n-2} \right\}$$

- **Tomada de Decisão:**
  - Se $t_{obs} \in$ RC rejeita-se H_0:β_1=0 ao nível de significância α, e conclui-se que β_1 é significativamente diferente de β_1^*.
  - Se $t_{obs} \notin$ RC não há evidências para rejeitar H_0:β_1=0 ao nível de significância α, e conclui-se que β_1 não é significativamente diferente de β_1^*.

32

Intervalo de Confiança para o parâmetro β_0 :

- Vimos que:

$$\hat{\beta}_0 \sim N\left[\beta_0\ ,\ \sigma^2\left(\frac{1}{n} + \frac{\bar{X}^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2}\right)\right]$$

Logo: $$Z = \frac{\hat{\beta}_0 - \beta_0}{\sigma\sqrt{\frac{1}{n} + \frac{\bar{X}^2}{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2}}} \sim N(0,1)$$

OBS: A v.a Z depende de σ.

33

Intervalo de Confiança para o parâmetro β_0 :

- Aliada ao fato de que: $\dfrac{(n-2)\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{(n-2)}$

tem distribuição quiquadrada com (n-2) graus de liberdade, obtemos uma nova variável aleatória obtida abaixo:

$$T = \frac{(\hat{\beta}_0 - \beta_0)}{\sigma\sqrt{\frac{1}{n} + \frac{\bar{X}^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2}}} \Bigg/ \sqrt{\frac{\frac{(n-2)\hat{\sigma}^2}{\sigma^2}}{(n-2)}} = \frac{\hat{\beta}_0 - \beta_0}{\hat{\sigma}\sqrt{\frac{1}{n} + \frac{\bar{X}^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2}}}$$

A v.a T tem distribuição de Student com (n-2) graus de liberdade

Lembre-se que: $\hat{\sigma} = \sqrt{QMRes} = \sqrt{\dfrac{\sum_{i=1}^{n} e_i^2}{n-2}}$

34

Intervalo de Confiança para o parâmetro β_0:

- Para construir um intervalo de confiança (IC) para β_0 ao nível de confiança de 100(1-α)% calcula-se a probabilidade abaixo:

$$P\left[-t_{\alpha/2,n-2} \le \frac{\hat{\beta}_0 - \beta_0}{\hat{\sigma}\sqrt{\frac{1}{n} + \frac{\bar{X}^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2}}} \le t_{\alpha/2,n-2}\right] = 1-\alpha$$

$$P\left[\hat{\beta}_0 - t_{\alpha/2,n-2}\cdot\hat{\sigma}\sqrt{\frac{1}{n} + \frac{\bar{X}^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2}} \ \le\ \beta_0 \ \le\ \hat{\beta}_0 - t_{\alpha/2,n-2}\cdot\hat{\sigma}\sqrt{\frac{1}{n} + \frac{\bar{X}^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2}}\right] = 1-\alpha$$

35

Intervalo de Confiança para o parâmetro β_0 :

- Logo um intervalo de confiança para o parâmetro β_0 do modelo, ao nível de confiança de 100(1-α)%, é dado por:

$$IC_{\beta_0,100(1-\alpha)\%} = \left[\underbrace{\hat{\beta}_0 - t_{\alpha/2,n-2}\ \hat{\sigma}\sqrt{\frac{1}{n} + \frac{\bar{X}^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2}}}_{\text{Limite inferior } (L_{inf}) \text{ do intervalo}}\ ,\ \underbrace{\hat{\beta}_0 - t_{\alpha/2,n-2}\ \hat{\sigma}\sqrt{\frac{1}{n} + \frac{\bar{X}^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2}}}_{\text{Limite superior } (L_{sup}) \text{ do intervalo}}\right]$$

36

Exemplo: Considerando os dados dos n=30 bois, construa agora um intervalo de confiança de 95% para o parâmetro β_0 do modelo de RLS.

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95,0% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	10,040	,495		20,277	,000	9,025	11,054
	X_conc.subs	,732	,076	,877	9,672	,000	,577	,887

a. Dependent Variable: Y_ganho_peso

IC para β_0 ao nível de 95%

37

Testes de Hipóteses para o parâmetro β_0:

- **Hipóteses a serem testadas:**

$$\begin{cases} H_0: \beta_0 = 0 \\ H_1: \beta_0 \neq 0 \end{cases}$$

- **Estatística de Teste:**

$$T = \frac{\hat{\beta}_0}{\hat{\sigma}\sqrt{\frac{1}{n} + \frac{\bar{X}^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2}}} \sim T_{n-2}$$

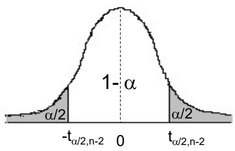
- A Estatística T tem distribuição de Student com (n-2) graus de liberdade.

38

Testes de Hipóteses para o parâmetro β_0 :

- **Região crítica:**

$$RC = \left\{ t \in \Re \ / \ t \le -t_{\alpha/2,n-2} \quad \text{ou} \quad t \ge t_{\alpha/2,n-2} \right\}$$



- **Tomada de Decisão:**
  - Se $t_{obs} \in$ RC rejeita-se H_0:β_0=0 ao nível de significância α, e conclui-se que β_0 é significativamente diferente de zero.
  - Se $t_{obs} \notin$ RC não há evidências para rejeitar H_0:β_0=0 ao nível de significância α, e conclui-se que β_0 não é significativamente diferente de zero.

39

Exemplo: Considerando os dados dos n=30 bois, realize um teste estatístico de hipóteses para o parâmetro β_0 ao nível de significância α de 5%. Qual a conclusão ?

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95,0% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	10,040	,495		20,277	,000	9,025	11,054
	X_conc.subs	,732	,076	,877	9,672	,000	,577	,887

a. Dependent Variable: Y_ganho_peso

TH para β_0 ao nível de 5%

40