(1)

PRELIMINARES:

ESPERANÇA CONDICIONAL:

$$E(Y|X=x) = \sum_{i=1}^{n} y_i \, P(Y=y_i \mid X=x_i)$$

$$\underbrace{E(Y|x)}_{\text{variável dependente (explicada)}} = \underbrace{\mu(x)}_{\text{variável independente (explicativa)}}$$

→ Em forma que $E(Y|X)$ é uma fç de x.

$$\tilde{\Pi} = f(e,\theta) = \mu e + \theta = \mu V + \theta$$

Sempre é importante pensarmos em modelos //

(Queremos saber como nossos concorrentes estão precificando sua matéria-prima)

$$P_\ell = \bar{w} + \beta G$$

(Preço do leite) = Preço mínimo + Qualid (Atributos)

$$P_\ell = \bar{w} + \beta_1 G + \beta_2 C + \beta_3 D + \varepsilon$$

C → contaminação

G → gordura no leite

D → distância

ε → resíduo (erro)

$\beta_1 = 0{,}2$

Vamos olhar, por enquanto, somente G

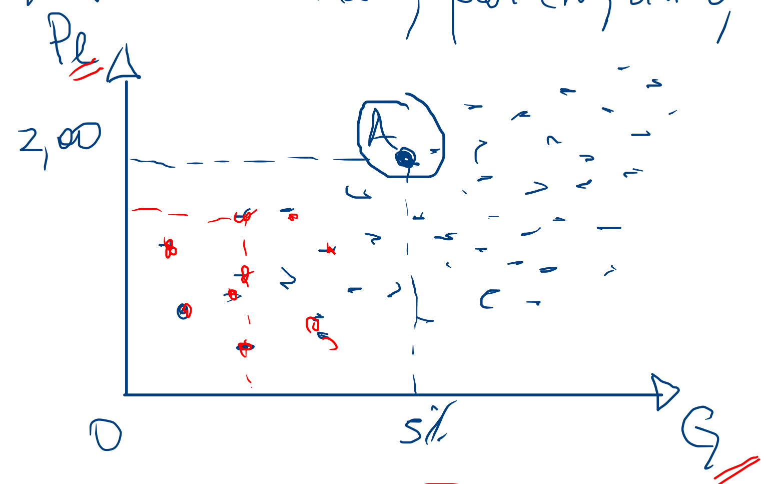


gráfico de dispersão

Produtor A (G, P_ℓ) → Qtd de gordura; Preço recebido pelo leite

→ Retirado de uma planilha:

Produtor	Q de gord (%)	Preço recebido ($)
A	5%	2,00
B		
⋮		

Resumo

$\Pi = f(V, \theta)$

$\Pi = \beta V + \theta$

$\mu_1 = \mu_2$

$\boxed{\hat{\beta} = \beta}$

$\hat{\beta} = 0{,}2$

P $\quad \beta$

Amostra

$\hat{\beta} = 0$

Produtos (leite c/ + G)

$H: \beta = 0$

$W = f(G; G_p, \ldots)$

$\boxed{\hat{W} =} \alpha + \beta_1 G + \beta_2 G_p + \varepsilon$

$(RT - CT)$

Reg. linear (múltipla)

Matrizes

$\hat{\mu}_1$

$\bar{X} \sim \mu$

$\hat{R} = \bar{W} + \beta G$ → (simples)

μ_1	μ_2
μ_3	μ_4

$\hat{\mu}_1 = \hat{\mu}_2 = \hat{\mu}_3 = \hat{\mu}_4$

Parâmetro × Estatística (estimadores)

$P = \bar{W} + \beta G \qquad \hat{P} = \hat{W} + \hat{\beta} G$

→ Média $\{\mu, \bar{X}$

Variância $\{\sigma^2, S^2$

2)

Veja q. na nossa amostra haviam $\underline{n}$ indivíduos onde cada um recebeu P_i, dado G_i.

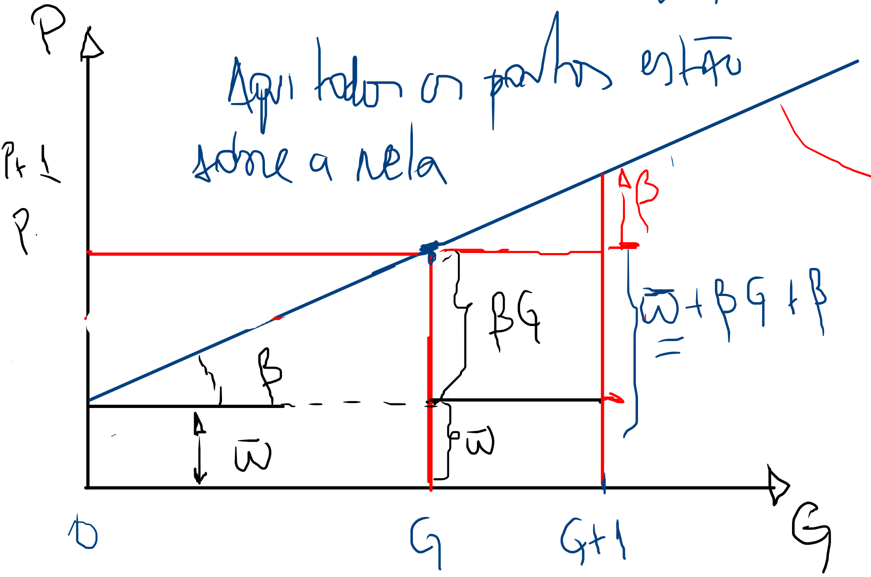
$\Rightarrow \boxed{E(P|G) = \bar{w} + \beta G} + \varepsilon$ → erro ou Resíduo

⇓

Há outros fatores q afetam P.

É uma regressão: Assume que o valor Médio do P é uma f(x) linear de G

↳ Posso considerar uma equação Linear

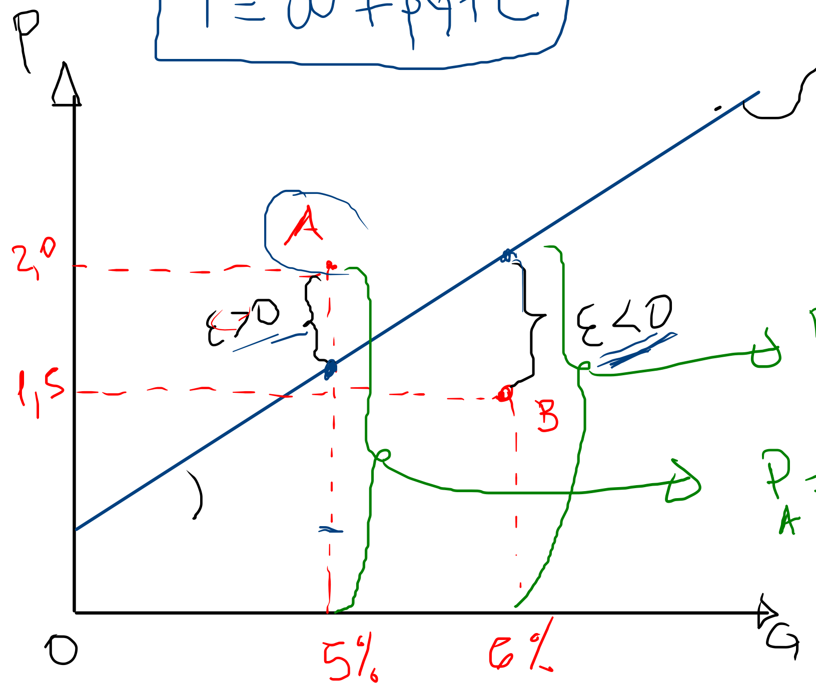


→ $E(P|G)$ s/ ε

$\begin{cases} \bar{w} - \text{Intercepto} \\ \beta - \text{coeficiente angular} \end{cases}$

β – representa o qto da o preço médio (P) varia p/ aumento de uma unidade de garantia

$\boxed{P = \hat{w} + \hat{\beta}G + e}$



Equação da reta (modelo s/ ε)

$P_B = \bar{w} + \beta G_B + \varepsilon$

$P_A = \bar{w} + \beta G_A + \varepsilon$

Representa a população

Veja q. isso só é válido p/ modelo linear.

se $P = \bar{w}e^{\beta G}$ → n linear ∴ n vale

Posso fazer $\ln P = \ln \bar{w} + \beta G$ (log linear)

Q to melhor $\hat{\beta}$ e $\hat{\alpha}$, então menor ε

$\hat{\beta} \to \beta$ / $\hat{\alpha} \to \bar{w}$

3) Nosso objetivo ESTIMAR (β e $\bar{w}$)

↳ estimadores da regressão.

Se descobrirmos estes valores conhecemos como G afeta P

(de outra forma: como meu concorrente valoriza a % de gordura no leite.)

P/ descobrir esses estimadores usamos um procedimento: Mínimos Quadrados Ordinários (MQO)

(cap 11 – item 11.4)

⇒ Estamos interessados em descobrir o parâmetro populacional através do estimador da amostra

$\#\ \beta \rightarrow$ p/ população

$\hat{\beta} \rightarrow$ p/ Amostra

$\Rightarrow P = \bar{w} + \beta G$ > População

$\hat{P} = \hat{\alpha} + \hat{\beta} G$ > Amostra

O chamamos $\hat{\beta}$ e dele faremos inferência sobre β.

Hipóteses importantes:

1) $G \rightarrow$ determinística

2) $E(e_i|G) = 0$ (valor médio do erro = 0)

3) $Var(e_i|G) = \sigma_e^2$ (variância constante)

4) Erros são não-correlacionados

↳ $\rho(e_i, e_j) = \dfrac{cov(e_i, e_j)}{\sigma(e_i)\,\sigma(e_j)} = 0$

$P = \overbrace{\hat{\alpha} + \hat{\beta} G_i}^{\hat{P}} + e_i$

$\boxed{e = P - \hat{P}}$

$\hat{P} = \hat{\alpha} + \hat{\beta} G_i + e$

P/ encontrar $\hat{\alpha}$ e $\hat{\beta}$ (estimadores), devemos fazer:

$e_i = P - (\alpha + \beta G_i) \Rightarrow \sum_{i=1}^{n} e_i = \sum_{i=1} (P_i - (\alpha + \beta G_i))$

elevando ao quadrado temos:

$$SQ(\alpha, \beta) = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (P_i - (\alpha + \beta G))^2$$

Ao derivar em relação a α e β temos:

Revisão

Modelo: //

$$\boxed{P_L = \hat{\omega} + \hat{\beta} G + e}$$

$\left.\begin{array}{l}\hat{\omega} - \\ \hat{\beta} - \end{array}\right\}$ estimadores

Quero saber se G afeta o preço

H_0: $G \rightarrow P_L$

afeta

Posso escrever:

$$\left.\begin{array}{l} H_0 : \beta = 0 \\ H_1 : \beta \neq 0 \end{array}\right\} \text{teste}$$

Sei que:

$$\hat{\alpha} = \bar{P} - \hat{\beta} \bar{G}$$

$$\hat{\beta} = \frac{\sum_{i=1}^{n} G_i P_i - n \bar{G} \bar{P}}{\sum_{i=1}^{n} G_i^2 - n \bar{G}^2}$$

P_L = dependente

G = explicativa

estão nos dados

$G(\%) \rightarrow P_L$

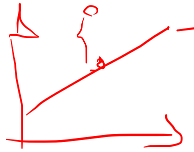
$$Y = \alpha + \beta_1 G + \beta_2 V + \varepsilon$$

erro

$$\hat{Y} = \hat{\alpha} + \hat{\beta}_1 G + \hat{\beta}_2 V + e$$

resíduo

$$Y - \hat{Y} = e$$



4) 1) $n\bar{Y} + \hat{\beta}\sum_{i=1}^{n} G_i = \sum_{i=1}^{n} P_i$

2) $\hat{\alpha}\sum_{i=1}^{n} G_i + \hat{\beta}\sum_{i=1}^{n} G_i^2 = \sum_{i=1}^{n} G_i P_i$

$E(\hat{\alpha}) = \alpha$

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

Reescrevendo (1):

$$\boxed{\hat{\alpha} = \bar{P} - \hat{\beta}\bar{G}}$$

Reescrevendo (2):

$$\boxed{\hat{\beta} = \frac{\sum_{i=1}^{n} G_i P_i - n\bar{G}\bar{P}}{\sum_{i=1}^{n} G_i^2 - n\bar{G}^2}}$$

Nosso modelo p/ a amostra é:

$$\boxed{\hat{P} = \underbrace{\hat{\alpha}}_{\text{representa } \bar{w}} + \underbrace{\hat{\beta}}_{\text{Representa } \beta} G_i}$$

Modelo teórico

$$\boxed{P = \bar{w} + \beta G_i}$$

Vamos a um exemplo:

TABELA c/ DADOS

Produtor	Preço ($)	Gordura (%)	PG	P^2	G^2
1	4	1	4		
2	6	2	12		
3	8	4	32		
4	14	8	112		
5	12	6	72		
6	10	5	50		
7	16	8	128		
8	16	9	144		
9	12	7	84		
$n = 9$	$\sum_{i=1}^{9} = 98$ ✓	$\sum_{i=1}^{9} = 50$ ✓	$\sum G_i P_i$		
	$\sum_{i=1}^{n} P_i$	$\sum_{i=1}^{n} G_i$			

Veja que precisamos gerar $\hat{\alpha}$ e $\hat{\beta}$ p/ chegar a $\hat{P} = \hat{\alpha} + \hat{\beta} G$

5) como
$$\begin{cases} \hat{\alpha} = \bar{P} + \hat{\beta}\bar{G} \\ \hat{\beta} = \dfrac{\sum_{i=1}^{9} G_i P_i - n\bar{G}\bar{P}}{\sum_{i=1}^{9} G_i^2 - n\bar{G}^2} \end{cases}$$

Preciso fazer os cálculos:

$\sum_{i=1}^{9} G_i P_i = 638$ $\bar{G} = \frac{50}{9} = 5,556$

$\sum_{i=1}^{n} G_i^2 = 340$ $\bar{P} = \frac{98}{9} = 10,89$

daí: $\hat{\beta} = \dfrac{638 - 9(5,6)(10,9)}{340 - 9(5,6)^2} = \dfrac{638 - 549,36}{340 - 282,24}$

$\hat{\beta} = \dfrac{88,64}{57,76} \approx 1,5$

Y X

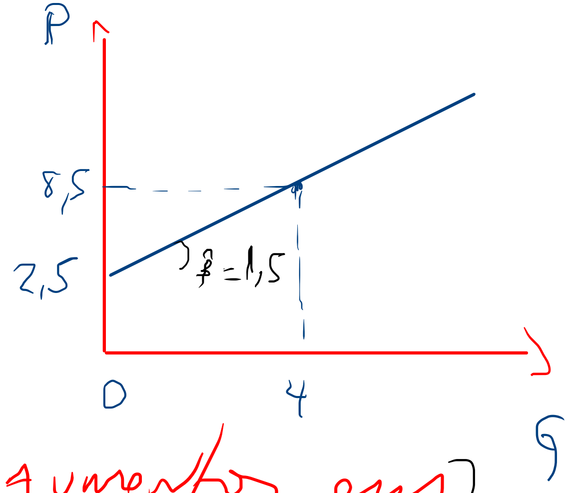
daí: $\hat{\alpha} = \bar{P} - 1,5\bar{G}$

$= 10,9 - 1,5(5,6) = 10,9 - 8,4 \approx 2,5$

Assim:

$$\boxed{\hat{P} = 2,5 + 1,5G}$$

A priori aumentos em 1% em G faz o P aumentar 1,5 vezes



P/ G = 4% $\hat{P}_e = 2,5 + 1,5(4) = 8,5$

G = 5% $\hat{P}_L = 2,5 + 1,5(5) = 10,0$

1,5

6)

Como saber se nosso modelo está adequado?

$\Rightarrow$ Usamos o coeficiente de determinação: R^2

Da aula passada aprendemos a gerar as seguintes somas de quadrados (SQ)

P/ regressão linear podemos fazer:

$$SQ_{Total} = \sum_{i=1}^{n} (P_i - \bar{P})^2$$

$\hookrightarrow$ desvio entre preço observado e o preço médio.

$$SQ_{Res} = \sum_{i=1}^{n} \hat{e}_i^2 = \sum_{i=1}^{n} (\hat{P}_i - \bar{P})^2$$

$\hookrightarrow$ soma dos quadrados do resíduo.

Posso fazer: $\boxed{SQ_{Reg} = SQ_{Tot} - SQ_{Res}}$

Posso fazer:

$$SQ_{Tot} = SQ_{Reg} + SQ_{Res}$$

Sei que: $(P_i - \bar{P}) = \boxed{P_i - \hat{P}_i} + \boxed{\hat{P}_i - \bar{P}}$

$P_i - \hat{P}_i = e_i$

$$\Rightarrow (P_i - \bar{P}) = \hat{e}_i + (\hat{P}_i - \bar{P})$$

elevando ao quadrado c/ somatórios

$$\sum (P_i - \bar{P})^2 = \sum_{i=1}^{n} e_i^2 + \sum_{i=1}^{n} (\hat{P}_i - \bar{P})^2$$

$$SQ_{Tot} = \sum (\hat{P}_i - \bar{P})^2 + SQ_{Res}$$

$$\therefore SQ_{Reg} = \sum_{i=1}^{n} (\hat{P}_i - \bar{P})^2$$

7)

sei tbm que $\boxed{\hat{P}_i = \hat{\omega} + \hat{\beta} q}$

posso fazer $\hat{P}_i = \bar{P} - \hat{\beta}\bar{q} + \hat{\beta} q_i$

$\hat{P}_i = \bar{P} + \hat{\beta}(\bar{q} - q_i)$

$\Rightarrow \boxed{\hat{P}_i - \bar{P}} = \hat{\beta}(\bar{q} - q_i)$

$$\boxed{\therefore SQReg = \hat{\beta}^2 \sum_{i=1}^{n} (\bar{q} - q_i)^2}$$

Posso fazer agora $R^2 = \frac{SQReg}{SQTOT}$) nunca negativo

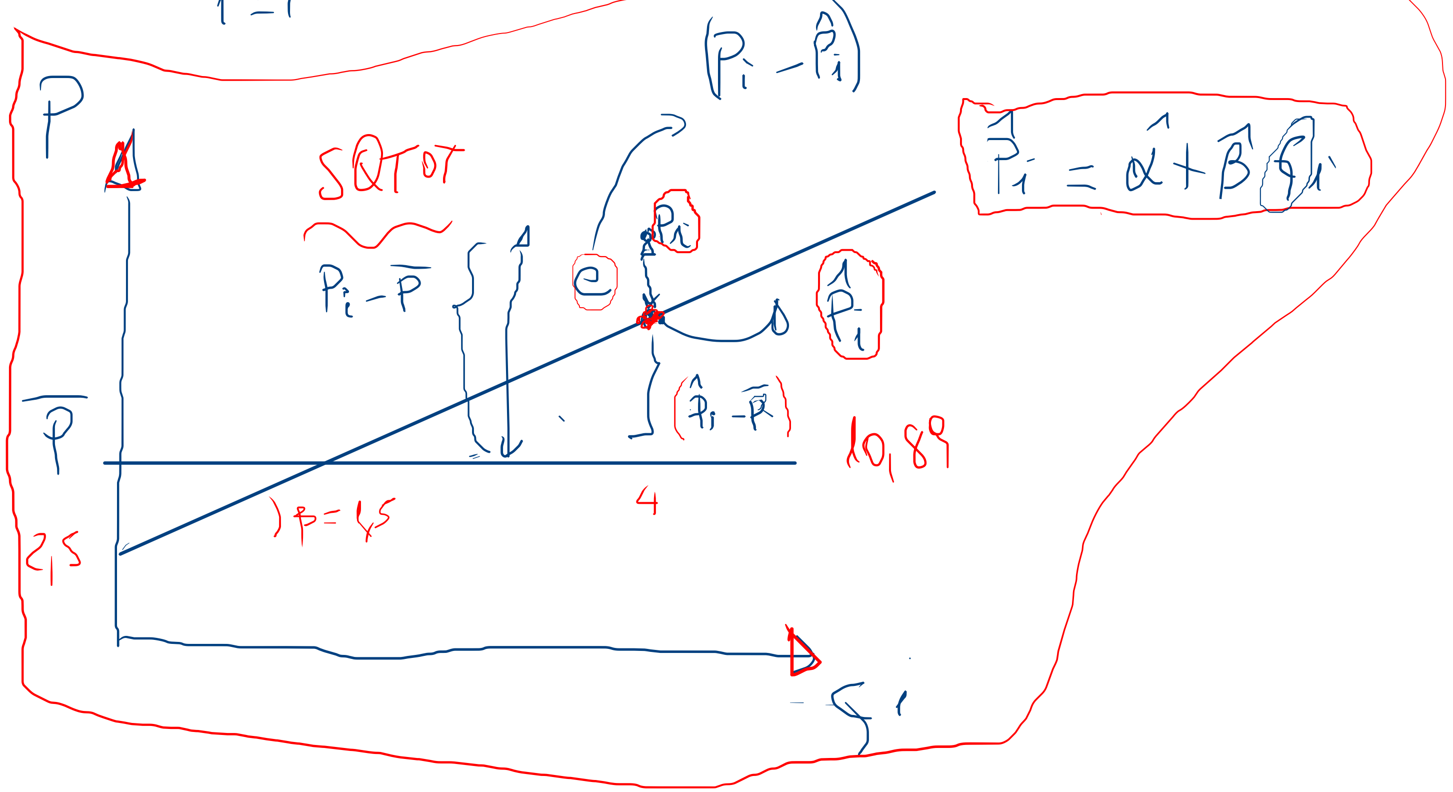
Explica o qto da variação total é explicada pela gordura → Aponta o ajuste à linha de regressão

$R^2 = 1$ (explica totalmente)

$R^2 = 0$ (ñ explica nada)

$$R^2 = \frac{\sum_{i=1}^{n} (\hat{P}_i - \bar{P})^2}{\sum_{i=1}^{n} (P_i - \bar{P})^2}$$

R^2



Ind. corr → Associação entre duas variáveis

$\boxed{r = \pm\sqrt{R^2}} \Rightarrow R^2 = r^{1/2}$

↳ coeficiente linear de Pearson

7)

$$R^2 = \frac{\left[\sum_{i=1}^{n} G_iP_i - \sum_{i=1}^{n} G_i \sum_{i=1}^{n} P_i\right]^2}{\left[n\sum_{i=1}^{n} G_i^2 - \left(\sum_{i=1}^{n} G_i\right)^2\right]\left[n\sum_{i=1}^{n} P_i^2 - \left(\sum_{i=1}^{n} P_i\right)^2\right]}$$

outra forma

$G \rightarrow P$

Assumindo que $\hat{\alpha}$ e $\hat{\beta}$ são os melhores estimadores p/ α e β $\begin{cases} \text{eficiente} \\ \text{consistente} \\ \text{menor variância} \end{cases}$

Podemos testar, por exemplo, se $H_0: \begin{cases} \hat{\beta} = 0 \\ \hat{\alpha} = 0 \end{cases}$

P/ tanto temos q. conhecer sua distribuição

⇒ Vamos assumir que:

$$\hat{\alpha} \sim N\left(\alpha; \underbrace{\frac{\sigma_e^2 \sum G_i^2}{n\sum (G_i - \bar{G})^2}}_{Var(\hat{\alpha})}\right)$$

$E(\hat{\alpha}) = \alpha$

$$\hat{\beta} \sim N\left(\beta; \underbrace{\frac{\sigma_e^2}{\sum (G_i - \bar{G})^2}}_{Var(\hat{\beta})}\right)$$

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

Podemos fazer:

$$\frac{\hat{\alpha} - \alpha}{\sigma_e}\sqrt{\frac{n\sum (G_i - \bar{G})^2}{\sum G_i^2}} \sim N(0,1)$$

$$\frac{\hat{\beta} - \beta}{\sigma_e}\sqrt{\sum (G_i - \bar{G})^2} \sim N(0,1)$$

Substituindo σ_e por S_e → desvio padrão amostral, então:

Aplicamos um teste t c/ n−2 gl

Posso montar IC e testar hipóteses:

8)

IC p/ α:

chame o coeficiente de confiança de C.

$$\Rightarrow IC(\alpha, C) = \hat{\alpha} \pm t_c(n-2)\, S_e \sqrt{\frac{\sum G_i^2}{n\sum(G_i-\bar{G})^2}}$$

$\boxed{\text{IC p/ } \beta:}$

$$IC(\beta, C) = \hat{\beta} \pm t_c(n-2)\, S_e \sqrt{\frac{1}{\sum(G_i-\bar{G})^2}}$$

P/ testar hipóteses: $(\alpha = 0)$

$$\underbrace{t(\hat{\alpha})}_{t\text{ calculado}} = \frac{\hat{\alpha}}{S_e}\sqrt{\frac{n\sum(G_i-\bar{G})^2}{\sum(G_i^2)}}$$

P/ testar $\beta = 0$

$$\underbrace{t(\hat{\beta})}_{t\text{ calculado}} = \frac{\hat{\beta}}{S_e}\sqrt{\sum(G_i-\bar{G})^2}$$

mais comum