



Revista Catarinense da Ciência Contábil

ISSN: 1808-3781

revista@crcsc.org.br

Conselho Regional de Contabilidade de  
Santa Catarina  
Brasil

Krauss, Valter Augusto; Wüst Corrêa de Lyra, Ricardo Luiz; da Silva, Tarcísio Pedro  
PROGRAMAÇÃO LINEAR E OS CUSTOS PARA PLANEJAMENTO  
Revista Catarinense da Ciência Contábil, vol. 4, núm. 11, abril-julio, 2005  
Conselho Regional de Contabilidade de Santa Catarina  
Florianópolis, Brasil

Disponível em: <http://www.redalyc.org/articulo.oa?id=477549001002>

- Como citar este artigo
- Número completo
- Mais artigos
- Home da revista no Redalyc

redalyc.org

Sistema de Informação Científica

Rede de Revistas Científicas da América Latina, Caribe, Espanha e Portugal

Projeto acadêmico sem fins lucrativos desenvolvido no âmbito da iniciativa Acesso Aberto

## **PROGRAMAÇÃO LINEAR E OS CUSTOS PARA PLANEJAMENTO**

**Valter Augusto Krauss**

Contador, professor da Furb,  
Mestre em Contabilidade e Controladoria pela USP  
e em Administração de Empresas pela FURB  
vkrauss@furb.br

**Ricardo Luiz Wüst Corrêa de Lyra**

Contador, professor universitário,  
mestre em Ciências Contábeis pela USP  
e especialista em Controladoria pela FURB  
Lyra@furb.br

**Tarcísio Pedro da Silva**

Contador e professor universitário,  
mestrando em Contabilidade e Controladoria pela FURB  
e especialista em Contabilidade Gerencial  
tarcisio@furb.br

### **RESUMO**

A regressão linear simples é uma ferramenta estatística muito útil para a área de ciências sociais, aplicável para analisar e prever o valor de uma variável, sabendo-se o valor de outra. Com base em dados coletados sobre estas duas variáveis, obtém-se uma equação que serve para determinar valores de uma variável, quando conhecidos valores de outra, estimar valores de uma variável e prever valores futuros. As técnicas de regressão e correlação referem-se à estimação de uma relação que possa existir em uma população. A regressão e a correlação baseiam-se em dados amostrais para determinar como duas variáveis são relacionadas entre si. A correlação fornece um número que resume o grau de relacionamento entre duas variáveis e a regressão tem como resultado uma equação matemática que descreve o relacionamento entre as variáveis.

Palavras-chave: Correlação entre variáveis. Definição de regressões para determinação de lucro.

### **1 INTRODUÇÃO**

As inovações em tecnologia e a introdução de diversos cálculos matemáticos e estatísticos em softwares que estão disponíveis a diversos usuários têm permitido a utilização de ferramentas estatísticas com mais facilidade. Profissionais de contabilidade e principalmente da área de custos e orçamentos podem e devem utilizar tais ferramentas para suas projeções e teste de cálculos.

Este trabalho sobre regressão linear simples possui como objetivo principal identificar sua utilidade como ferramenta de trabalho para aplicação em estudos no campo de ciências sociais, principalmente em disciplinas ligadas à contabilidade, e procura apresentar não só um

modelo de aplicação à contabilidade orçamentária, mas também testes úteis à confirmação de tais hipóteses.

Serão explorados os cálculos apresentados por meio da ferramenta Excel, através do comando “Gráfico - adicionar linha de tendência” e “Ferramenta- análise de dados-regressão”. Em seguida são comentados os resultados do gráfico e os resultados da regressão.

## 2 EQUAÇÃO LINEAR

A notação de função surge quando se procura estudar fenômenos e fatos do mundo. Repare quantas vezes criamos ou procuramos relacionar as coisas entre si. Ao estudarmos, por exemplo, os métodos de custeio por absorção, podem querer associar os CIF ao produto através do consumo de MOD.

Na equação matemática da reta, temos o conceito de par ordenado, isto é, para cada elemento de X corresponde um elemento de Y; este par ordenado pode ser representado no plano cartesiano através da relação entre os valores de X e Y.

A relação X e Y é uma função se e somente se, nesta relação, para cada X tivermos um único Y. Dependendo da relação entre X e Y, o gráfico desta função pode ser linear, logarítmico, exponencial e outros.

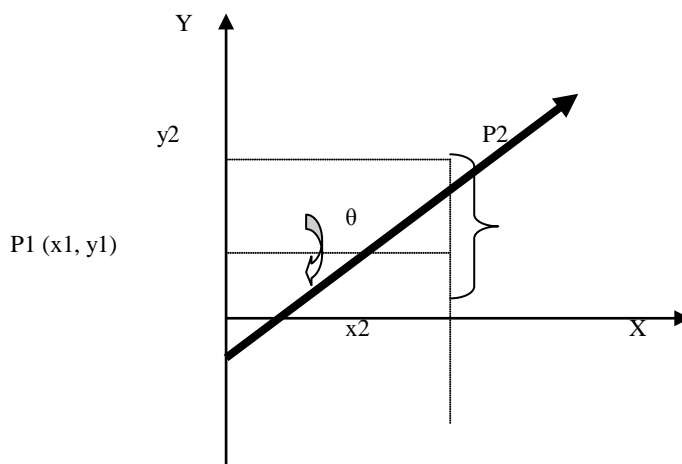
Vamos, neste estudo, nos ater à função linear simples (2 variáveis). Esta função linear pode ser crescente, decrescente ou constante, dependendo da relação entre os valores de X e Y, isto é: se os valores de Y aumentam à medida que os valores de X aumentam, temos uma função crescente; no caso dos valores de Y diminuírem à medida que X aumenta, temos uma função decrescente; e quando os valores de Y não se alteram independentemente dos valores de X, temos uma função constante. Se o sinal do coeficiente angular da reta (b) for “+”, indica ser uma função crescente; se “-“, é uma função decrescente; e se b = zero, a função é constante.

### *Dedução da equação da reta*

- Seja um gráfico cartesiano com os pontos  $P1(x_1; y_1)$  e  $P2(x_2; y_2)$ ;
- A função  $P1P2$  pode ser determinada como sendo a tangente do ângulo  $\theta$ , assim  $\text{tg}\theta = b = y_2 - y_1 / x_2 - x_1$ ;
- Para  $x_1 = 0$  e  $y_1 = a$ , temos a equação:

$$y = a + bx$$

- Onde “a” é o intercepto do eixo y e “b” é o coeficiente angular da reta.



Dizemos então que  $y = f(x)$  ou o valor da variável  $y$  depende do valor da variável  $x$ , sendo, portanto,  $x$  a variável independente e  $y$  a variável dependente da relação.

### 3 REGRESSÃO LINEAR

Regressões lineares são técnicas matemáticas utilizadas para definir fórmulas com base em valores observados, nas quais, com base em valores definidos das variáveis independentes, seja possível definir valores para uma variável dependente.

As regressões lineares que trabalham com duas variáveis, uma independente e outra dependente, são denominadas regressão linear simples e, quando trabalham com mais de duas variáveis, são denominadas regressão múltipla.

Na regressão linear simples trabalha-se com duas variáveis, uma independente ou explicativa e uma dependente. A equação serve para determinar valores de uma variável quando conhecidos valores de outra, estimar valores de uma variável e prever valores futuros.

As técnicas de regressão e correlação referem-se à estimação de uma relação que possa existir em uma população. A regressão e a correlação baseiam-se em dados amostrais para determinar como duas variáveis são relacionadas entre si. A correlação fornece um número que resume o grau de relacionamento entre duas variáveis e a regressão tem como resultado uma equação matemática que descreve o relacionamento entre as variáveis.

É uma tentativa de se estabelecer uma função matemática afim (linha reta) que descreva o relacionamento entre duas variáveis, uma sendo a dependente e a outra a independente, com o propósito de:

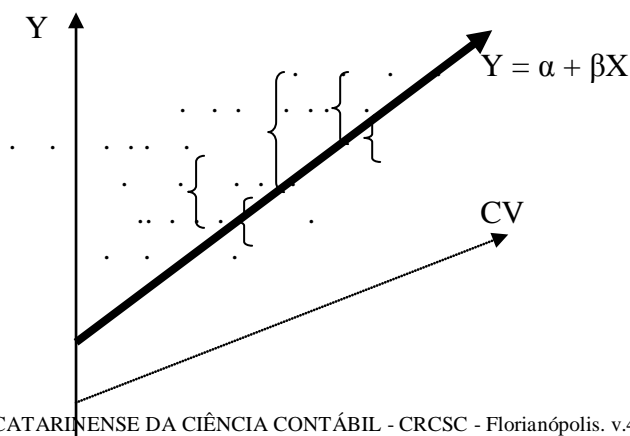
- Explicar os valores de uma variável em termos da outra variável (relação causa e efeito);
- Estimar valores de uma variável, com base em valores conhecidos da outra;
- Predizer valores futuros (estimativa da relação real).

Algumas técnicas para a determinação desta equação matemática serão apresentadas mais adiante. Por hora nos fixaremos em estabelecer alguns conceitos e características desta equação.

Regressão é o efeito da variável explicativa  $x$  sobre a variável dependente  $y$ . Trata-se de estabelecer uma equação que descreva o relacionamento matemático entre as variáveis dependentes e independentes.

Vejamos um gráfico que apresenta, em seu eixo  $X$ , as unidades produzidas e, em seu eixo  $Y$ , os custos totais ( $CF + CV$ ).

Exemplos de dispersões



CF  $\longrightarrow$

$\xrightarrow{\hspace{10em}} X$   
Será que os custos fixos por unidade alocados são iguais para cada unidade produzida dentro de um mesmo lote?

E quanto aos custos variáveis? Eles realmente consomem a mesma quantidade de material, insumos, mão-de-obra, por unidade produzida?

Na realidade, aceitamos o modelo linear como simplificação, porém sabe-se que cada unidade produzida possui um custo diferente, existe uma variação por menor que seja nos custos de cada uma, isto por existirem outras variáveis que influenciam a variável dependente.

Dispersão é a diferença entre os valores situados na reta de regressão e o valor real. Estes desvios são causados pelos valores que a variável dependente assume, diferentes dos valores estimados para ela por meio da variável independente. O poder de explicação do modelo adotado será tanto melhor quanto menor forem as somas dos desvios.

Se não houvesse dispersões, todas as amostras da população estariam sob a reta e não haveria necessidade de fazer inferências quanto aos verdadeiros valores.

Para se determinar uma reta de regressão a partir de uma amostra de valores supostamente relacionados, precisamos estabelecer algumas premissas básicas, ou hipóteses. Segundo Lapponi (2000, p.391), o modelo de Regressão Linear possui as seguintes premissas:

- Para cada valor de x há um grupo de valores de y e todos os grupos de y têm distribuição normal com o mesmo desvio padrão.
- As médias das distribuições normais de y pertencem à reta de regressão.
- A média dos desvios ou erros é nula, pois a variância é mínima.
- A variância dos desvios é constante e igual à variância da população; isto é, todos os desvios têm a mesma variância.

Os desvios são variáveis aleatórias independentes e têm distribuição normal. Portanto, o coeficiente de correlação entre os desvios, tomado dois a dois, é nulo, e os desvios e a variável independente x não têm qualquer correlação.

Assim podemos concluir que a reta de regressão linear é uma estimativa da relação entre as variáveis X e Y de uma amostra ou população. Sua equação é:

$$\boxed{Y = \alpha + \beta X + e_i} \text{ onde } e_i \text{ é a dispersão da população.}$$

A correlação indica até que ponto os valores de uma variável estão relacionados com os valores da outra variável. O objetivo do estudo da correlação é determinar a força do relacionamento entre duas variáveis que podem se relacionar.

### **Exemplos em que pode existir um relacionamento:**

Produtos melhor elaborados tendem a apresentar melhor aceitação entre os consumidores?

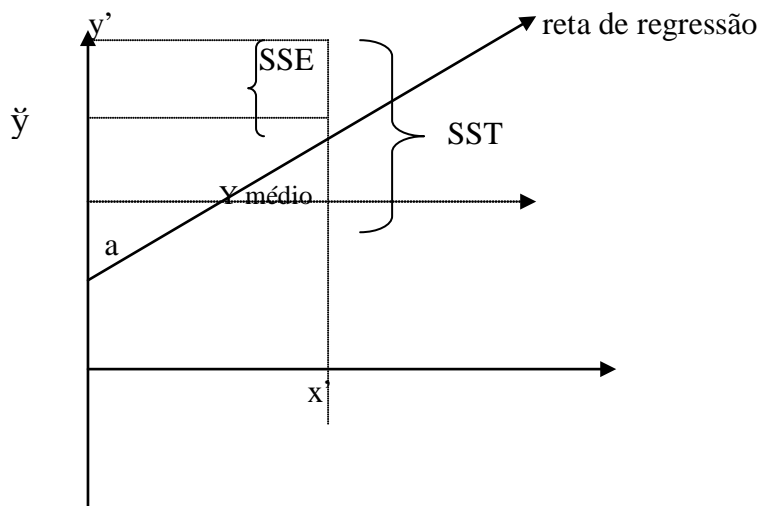
Os custos fixos influenciam a rentabilidade global da empresa?

Funcionários com bons salários tendem a apresentar um melhor rendimento?

Quanto maior o volume de vendas maior deveria ser o lucro?

Dado o gráfico, vamos explicar as variações que podem existir:

Varição em relação à média  $SST = \sum_{i=1}^n (y_i - \bar{y})^2$   
 Variação não explicada  $SSE = \sum_{i=1}^n (y_i - y')^2$



O **Grau de relacionamento (r)** entre duas variáveis é dado por um coeficiente de correlação conhecido como r de Pearson (Karl Pearson, 1857-1936, matemático inglês).

Possui duas propriedades:

- O sinal (+ ou -) é o mesmo que o do coeficiente angular de uma reta imaginária; se positivo, aumentando o valor da variável X, os valores da variável Y também aumentam e, se negativo, os valores da variável Y irão diminuir.
- Sua grandeza ( $-1 \leq r \leq +1$ ), os valores mais próximos de  $\pm 1$  indicam que os valores das observações estão muito próximos ou sobre a reta, enquanto os valores mais próximos de zero sugerem maior dispersão ou que as variáveis são não correlacionadas.

Variável reduzida (z) mede o desvio de determinado valor em relação à média, em unidades de desvio padrão.

$$Z = \frac{x - \mu}{\sigma}$$

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{(\sum (x_i - \bar{x})^2 \cdot \sum (y_i - \bar{y})^2)^{0,5}}$$

$$r = \frac{\sum (Zx \cdot Zy)}{n-1}$$

r = Excel – análise de dados - regressão

r = raiz quadrada de  $r^2$

**Coefficiente de determinação ( $r^2$ )** mede a explicação de reta de regressão, fornece a porcentagem de variação em uma variável que é explicada estatisticamente pela variação da outra. Quanto mais próximo for o seu valor de um (cem por cento), indica que menores são os desvios.

$$r^2 = \frac{SST - SSE}{SST}$$

$$r^2 = \text{Excel RQUAD}(\text{val\_conhecido\_y}; \text{val\_conhecido\_x})$$

$r^2$  = Excel – análise de dados – regressão.

**Erro padrão de estimativa (Se)** mede a dispersão dos valores amostrados, denominados resíduos ao redor da reta de regressão. É equivalente ao desvio padrão que mede a variabilidade dos valores da amostra ao redor da média aritmética desses valores (dispersão entre os valores observados e os valores estimados).

O cálculo do erro padrão se baseia na hipótese de dispersão uniforme, variâncias iguais (HOMOCEDASTICIDADE),

$$\text{EPADYX}(\text{val\_conhecido\_y}; \text{val\_conhecido\_x})$$
$$\text{Se} = (\text{SSE}/(n-2))^{1/2}$$

Por que (n-2)? Porque ao calcularmos as constantes “a” e “b” perdemos dois graus de liberdade.

Se = Excel – análise de dados – regressão.

#### 4 EXEMPLOS

A empresa Agiliza Comércio Eletrônico vem procurando determinar projeções de resultado, com base em dados estudados em períodos anteriores, e o departamento de custos da empresa sugeriu que fosse aplicada a regressão linear simples.

Agiliza, empresa de venda a varejo, que atende a pedidos recebidos via-internet, deseja calcular sua projeção de resultados para períodos futuros com base na projeção de pedidos a serem recebidos e atendidos.

O Gerente de custos, por sua vez, levantou dados dos últimos 15 meses em relação a pedidos atendidos e lucro auferido.

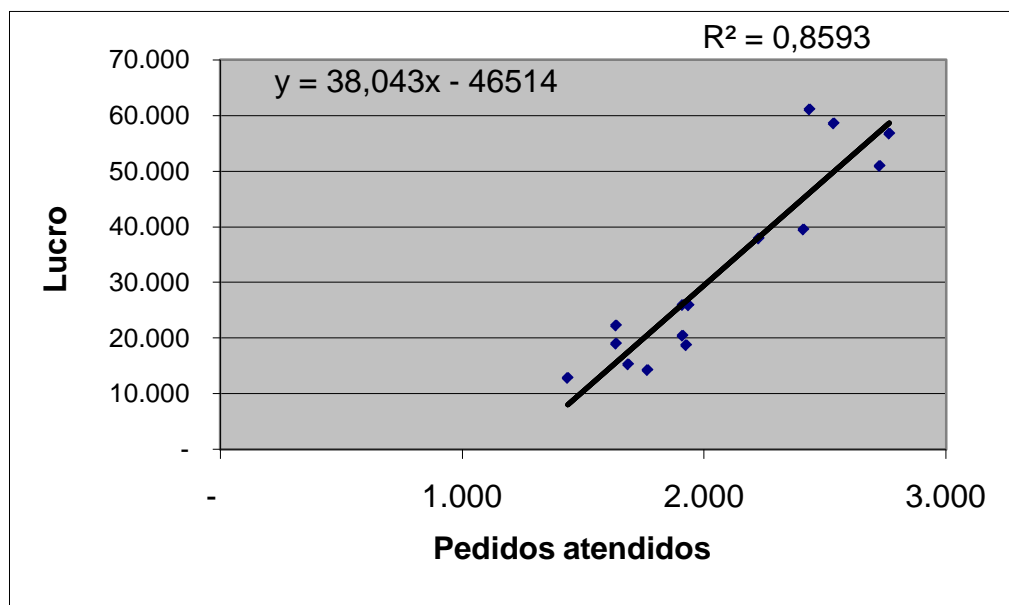
##### Dados para análise

Número de pedidos atendidos	Lucro
1.925	18.683,00
1.435	12.790,00
1.685	15.235,00
1.910	20.390,00
2.435	61.095,00
1.635	18.955,00
2.225	37.875,00
2.765	56.745,00
1.910	25.910,00
1.765	14.203,00
2.725	50.925,00
2.535	58.578,00
1.935	25.855,00
2.410	39.493,00
1.635	22.225,00

A partir dos dados identificados, foram aplicados os conceitos de correlação, elaborou-se o gráfico de dispersão das amostras de pedidos recebidos *versus* lucro alcançado, como verificado a seguir.

Para visualizar a função que identifica uma possível correlação entre a variável dependente (lucro) em função da variável independente (pedidos recebidos e atendidos),

esta foi adicionada no gráfico, utilizando-se o menu Gráfico – Adicionar linha de tendência – Opções, selecionando Exibir equação do gráfico e Exibir valor de R-quadrado no gráfico, apresentando o seguinte resultado:



O gráfico de dispersão mostrou que as variáveis estão correlacionadas de forma positiva próximo da correlação +1. A equação do gráfico apresenta a função  $y = -46.514 + 38,043x$ , identificando que, para cada valor de x (pedidos atendidos), o valor de y (lucro) é diferente. Os pontos acima e abaixo da reta de regressão criada sobre os dados observados indicam que os valores de y obtidos através da fórmula de regressão apresentam dispersões em torno da reta criada.

O intercepto Y é igual a -\$ 46.514,00, isto é, a reta inicia-se no ponto Y - 46.514. Neste exemplo o coeficiente angular é positivo, é a relação estabelecida é: quanto maior o número de pedidos X, o lucro Y aumentará.

### Análise da Regressão

Para analisar estas dispersões são apresentados, na sequência, cálculos obtidos através do Excel – Ferramentas – Análise e dados – Regressão, com os seguintes resultados:

#### RESUMO DOS RESULTADOS

##### *Estatística de regressão*

R múltiplo	0,9270
R-Quadrado	0,8593
R-quadrado ajustado	0,8485
Erro padrão	6.769,60
Observações	15

#### ANOVA

	<i>Gl</i>	<i>SQ</i>	<i>MQ</i>	<i>F</i>	<i>F de significação</i>
Regressão	1	3.637.739.749	3637739749	79,37903	0,00000068
Resíduo	13	595.757.059	45.827.466		

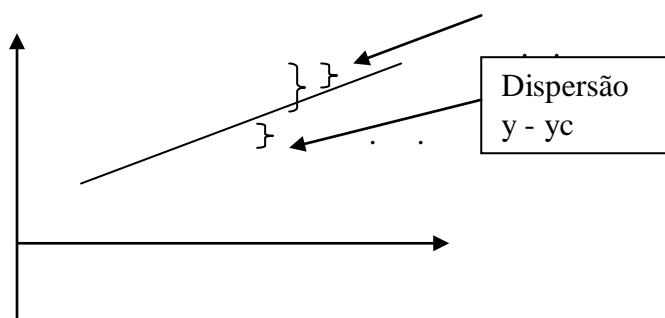


Total	14	4.233.496.808				
	<i>Coefficientes</i>	<i>Erro padrão</i>	<i>Stat t</i>	<i>valor-P</i>	<i>95% inferiores</i>	<i>95% superiores</i>
Interseção	(46.513,87)	8.976,40	(5,182)	0,000177	(65.906,20)	(27.121,53)
Quantidades	38,04	4,27	8,909	6,75E-07	28,82	47,27

### **Análise dos resultados apresentados**

O Grau de relacionamento (R múltiplo) indica que os pontos observados das variáveis lucros e pedidos estão muito próximos da reta de regressão e sendo um valor positivo indica que é uma reta crescente, isto é, aumentando-se os valores dos pedidos, o lucro também aumenta.

O Coeficiente de determinação (R quadrado) indica que 85,93% das variações da variável lucro são explicadas pela variação da variável número de pedidos e quanto mais próximo de 100% menor os desvios entre os valores observados e a reta de regressão.



O  $R^2$  ajustado é uma ferramenta da regressão múltipla que não faz parte da regressão simples.

O erro padrão indica que existe uma dispersão média de 6.769,50 entre os valores observados de y (lucro) e os valores calculados de y (lucro) pela reta de regressão criada sobre os valores observados.

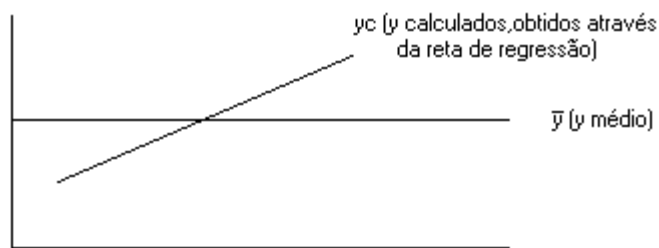
Observações são o número de observações utilizadas para elaborar a reta de regressão.

### **Teste do Relacionamento entre Variáveis**

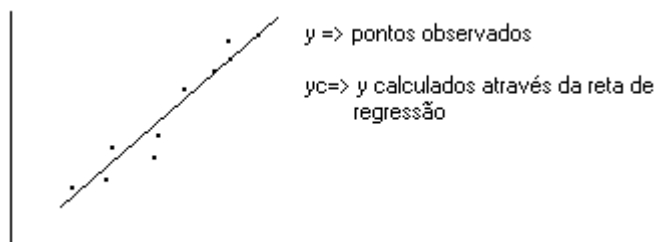
A ferramenta ANOVA serve para verificar se existe ou não relacionamento entre as variáveis, em que o principal índice de análise é obtido pela fórmula do teste F (Fischer):

$$F = \frac{\text{Estimativa "entre" da variância (explicada)}}{\text{Estimativa "dentro" da variância (não explicada)}}$$

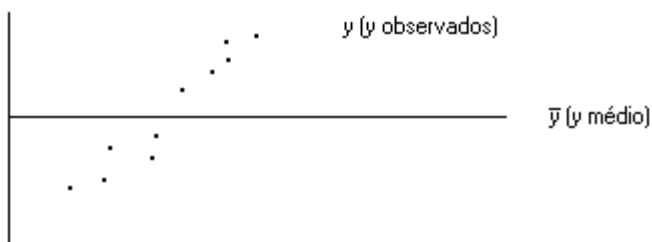
Onde variância “entre” da variância é a soma das diferenças ao quadrado do  $y$  calculado para o  $y$  médio ao quadrado  $\Sigma(y_c - \bar{y})^2$ .



E variância “dentro” da variância é a soma das diferenças ao quadrado entre  $y$  e  $\hat{y}$  da regressão  $\Sigma(y - \hat{y})^2$ .



Esta variância também pode ser obtida entre a variância total e a variância “entre” da variância  $\{ (\Sigma(y_c - \bar{y})^2) - (\Sigma(y - \bar{y})^2) \}$ .



Com base nos gráficos acima, apresenta-se, a seguir, os cálculos para apuração das variâncias totais, variâncias de estimativas “entre” e “dentro” da variância.

Pedidos	$(y - \bar{y})^2$	$(y_c - \bar{y})^2$ “entre”	$(y - y_c)^2$ “dentro”
1.925	175.495.373	27.163.577	64.570.832
1.435	366.357.464	568.958.914	22.206.648
1.685	278.738.607	205.697.269	5.538.099
1.910	133.182.371	33.437.438	33.154.048
2.435	850.570.005	201.355.496	224.237.286
1.635	168.362.735	263.877.022	10.685.219
2.225	35.337.477	38.452.186	65.766
2.765	615.761.065	715.247.708	3.723.287

1.910	36.246.019	33.437.438	56.623
1.765	314.263.074	127.661.141	41.328.745
2.725	360.792.297	636.169.555	38.786.346
2.535	710.091.033	323.793.485	74.879.095
1.935	36.911.295	23.342.817	1.547.601
2.410	57.191.910	175.268.679	32.221.227
1.635	94.196.083	263.877.022	42.756.237
Totais	4.233.496.808	3.637.739.749	595.757.059

Para o cálculo do teste F, a soma dos quadrados “dentro” perde dois graus de liberdade da estimativa, sendo dividida por (n-2).

$$F = 3.637.739.749 / (595.757.059 / 13) = 79,38$$

Para verificar se existe relação entre as variáveis, cria-se as seguintes hipóteses:

$H_0$  – Não existe relacionamento entre as variáveis

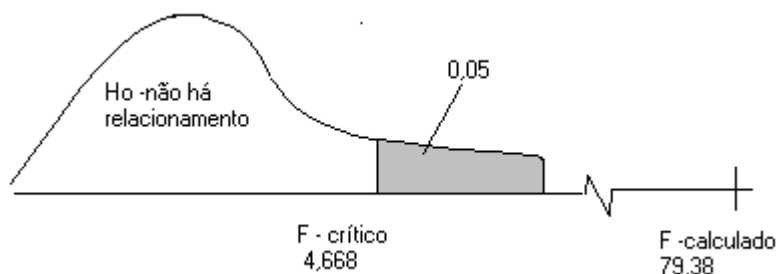
$H_1$  – Existe relacionamento entre as variáveis.

O teste de hipóteses pode ser feito através da comparação do F calculado com o F crítico ou comparação entre o alfa e o F crítico, que obrigatoriamente chegarão ao mesmo resultado, aceitando a hipótese  $H_0$  ou rejeitando-a e aceitando  $H_1$ .

Em termos de análise de regressão, a estimativa “entre” é obtida pela SQ (soma dos quadrados) da regressão, apresentada na tabela ANOVA, dividida pelos graus de liberdade da regressão, resultando na MQ (média dos quadrados) e a estimativa “dentro” é obtida pela SQ (soma dos quadrados) dos resíduos, também apresentados na tabela ANOVA, dividida pelos graus de liberdade.

Desta forma, chegamos ao seguinte F calculado =  $3.637.739.749 / 45.827.466 = 79,38$  e, comparando com o F crítico através da fórmula do excel =INVF(0,05;1;13), ao valor de 4,668.

Assim podemos comparar o F calculado com o F crítico e testar a hipótese  $H_0$  – que não há relacionamento, ou rejeitar  $H_0$  - há relacionamento.



De acordo com o gráfico acima, o F calculado está fora da faixa de teste da hipótese  $H_0$ , portanto rejeita-se  $H_0$  - há relacionamento entre as variáveis.

O F de significância apresentado na tabela ANOVA, 0,00000068 ou calculado pela fórmula =DISTF(79,38;1;13) é menor que o nível de significância 0,05, a hipótese nula deve ser rejeitada e a regressão deve ser aceita.

## Teste dos Coeficientes Angulares

Na seqüência da tabela, na qual são apresentados dados dos coeficientes “a” (interseção) e “b” (quantidades), são analisadas as hipóteses do coeficiente angular “b” ser ou não igual a zero.

Daí define-se as seguintes hipóteses:

$H_0$  – O Coeficiente angular pode ser zero

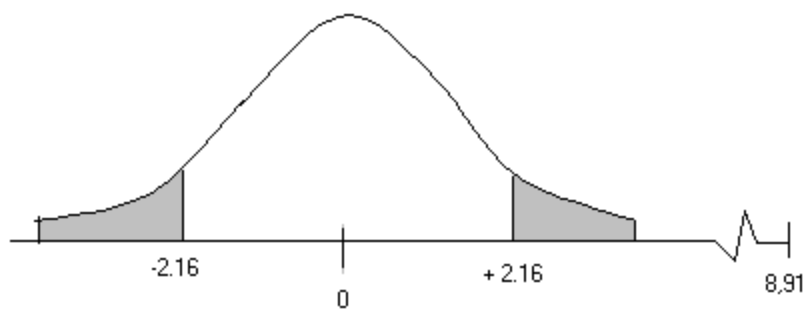
$H_1$  – O Coeficiente angular não pode ser zero.

Os testes das hipóteses podem ser feitos de três maneiras:

- Comparação entre t crítico com t calculado;
- Verificando se o zero se encontra na faixa determinada para os 95% superiores e 95% inferiores dos valores determinados para Y;
- Comparando valor p com alfa.

Os valores dos coeficientes “a” e “b” são os mesmos apresentados no gráfico de dispersão. Os erros padrão do coeficiente “b” apresentados na tabela correspondem aos erros padrões dos coeficientes e o Stat t é o valor dos coeficientes dividido pelo erro padrão correspondente. Comparando o T calculado do coeficiente  $b = 8,91$  com o T crítico calculado pela fórmula  $=\text{INVT}(0,05;13) \Rightarrow 2,16$ , faz-se a seguinte análise:

Se t calculado for maior que (-t crítico) e menor que (+t crítico) então aceita-se a hipótese do coeficiente angular ser igual a zero.



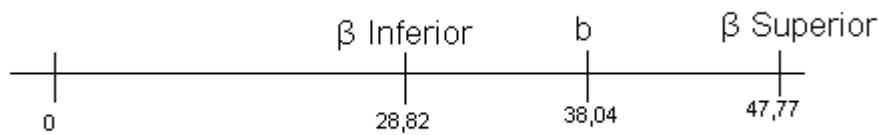
Estando o Stat t calculado 8,91 fora da área de t teste para coeficiente angular igual a zero, conclui-se que o coeficiente angular não é igual a zero.

Os 95% inferiores e 95% superiores determinam a amplitude dos coeficientes  $\alpha$  e  $\beta$ , considerando o erro padrão dos coeficientes e o t crítico calculado.

95% inferiores do coeficiente  $\alpha = a - t^*Sa$  (coeficiente a deduzido do t crítico vezes erro padrão do coeficiente)  $= -46.513,87 - 8976,40 \cdot 2,16036 \Rightarrow -65.906,20$ .

95% superiores do coeficiente  $\alpha = a + t^*Sa$  (coeficiente  $\alpha$  somado ao t crítico vezes erro padrão do coeficiente)  $= -46.513,87 + 8976,40 \cdot 2,16036 \Rightarrow 27.121,53$ .

95% superiores ou inferiores do coeficiente  $\beta = b (+/-) t^*S_b$  (coeficiente  $\beta$  somado ou deduzido do t crítico vezes erro padrão do coeficiente) =  $38,04 - 4,27 * 2,16036 = 28,82$  ou superiores =  $38,04 + 4,27 * 2,16036 = 47,27$ .



Uma outra forma de análise do coeficiente angular  $\beta$  é por meio da análise do Valor\_p.

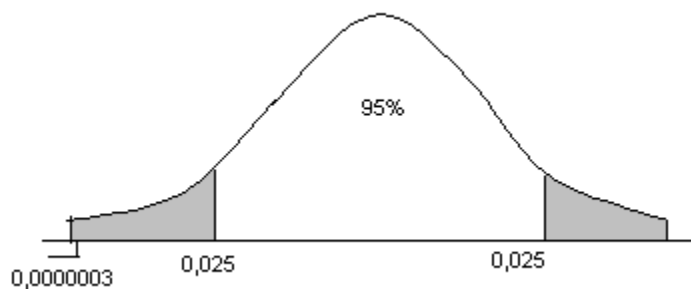
A condição básica é a comparação do valor\_p com o nível de significância ( $\alpha$ ), através das seguintes hipóteses:  $H_0: B = 0$  e  $H_1: B \neq 0$

Se  $\begin{cases} \text{Valor\_p} > \alpha \text{ implica que } H_0 \text{ é aceita} \\ \text{Valor\_p} < \alpha \text{ implica que } H_0 \text{ é rejeitada, isto é, } B \neq 0. \end{cases}$

Em nosso exemplo, temos:

Valor\_p = 0,00000067  
 $\alpha = 0,05$ .

Conclui-se que  $B \neq 0$ , logo, rejeita-se  $H_0$  e podemos tratar a equação como uma regressão linear.



### Intervalo de Predição para análise de Regressão

O valor predito de y, através da reta de regressão para valores específicos de x, pode ser encarado de duas maneiras:

- Valor médio de y (lucro) para dada quantidade x (quantidades);
- Valor individual de y.

A Administração pode querer saber o lucro médio para uma quantidade de 2.000 pedidos e o Diretor querer saber o lucro para sua quantidade de pedidos a ser atendida no mês em referência.

Os intervalos de confiança para essas predições baseiam-se no desvio padrão das respectivas quantidades. Para o valor médio de y, o desvio padrão de yc é

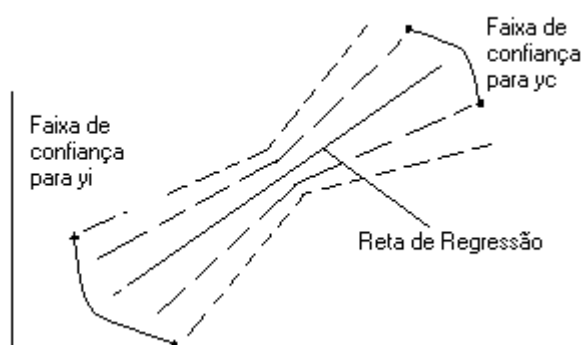
$$S_{yc} = Se \sqrt{\frac{1}{N} + \frac{(X_g - \bar{X})^2}{\sum x^2 - \{(\sum x)^2/n\}}}$$

Onde Xg é a quantidade de pedidos e X é a média de pedidos das observações. Para os valores individuais de y, soma-se um único termo 1 à expressão sob o radical:

$$S_{yi} = Se \sqrt{1 + \frac{1}{N} + \frac{(X_g - \bar{X})^2}{\sum x^2 - \{(\sum x)^2/n\}}}$$

A explicação destes dois intervalos de confiança é que o intervalo de valores individuais de y (lucro) através de um valor determinado de x (quantidades) é superior ao valor médio de y (lucro).

Dentro destes intervalos de predição, quanto mais distante da média de x os valores dos intervalos de yc e yi são maiores, conforme demonstrado a seguir.



Conforme gráfico acima, fica evidente que, quanto mais distante da média, maior a faixa de confiança para yi e yc. Vejamos a faixa de confiança para yi (pedidos individuais) e yc (pedidos médios) considerando quantidade média de (2062) pedidos atendidos. Quanto seriam os desvios padrões para y médio e y individual se considerarmos a quantidade individual de 1.435 pedidos e a quantidade média de 1.435 pedidos ?

Apresenta-se abaixo os cálculos dos desvios padrões, considerando a quantidade individual de 1.435 pedidos.

$$S_{yi}(1.435) = 6.769,60 \times \sqrt{1 + \frac{1}{15} + \frac{(1.435 - 2.062)^2}{66.291.200 - (956.664.900 / 15)}}$$

$$S_{yi}(1.435) = 6.769,6 \times 1,106 = 7.487,18$$

Considerando os desvios padrões para a quantidade individual de 1.435 pedidos, a faixa de confiança correspondente é:

$$Y_c = -46.514 + 38,043(1.435) \pm 7.487,18$$

$$Y_c = 8.077,71 \pm 7.487,18 \quad \text{ou} \quad 590,52 \leq y_i \leq 15.564,89.$$

Portanto o lucro para uma quantidade de 1.435 pedidos corresponde entre \$590,52 e \$15.564,89.

Abaixo se apresentam cálculos para os desvios padrões valores de  $y$ , considerando uma quantidade média de 1.435 pedidos:

$$S_{yc}(1435) = 6.769,60 * \sqrt{\frac{1}{15} + \frac{(1.435 - 2.062)^2}{66.291.200 - (956.664.900 / 15)}}$$

$$S_{yc}(1435) = 6.769,6 * 0,472 = 3.195,25$$

Considerando os desvios padrões para a quantidade média de 1.435 pedidos, a faixa de confiança correspondente é:

$$Y_c = -46.514 + 38,043(1.435) \pm 3.195,25$$

$$Y_c = 8.077,71 \pm 3.195,25 \quad \text{ou} \quad 4.882,46 \leq y_i \leq 11.272,96.$$

## 5 CONCLUSÃO

A regressão e a correlação são ferramentas usadas para determinar uma tendência, sendo importante definir bem as variáveis dependentes e independentes, pois a regressão linear é uma estimativa da relação entre estas duas variáveis.

Embora se saiba que podem existir mais de duas variáveis, alguns cálculos da análise de dados fornecem índices que podem medir graus de correlação e significância.

Para obter bons resultados na aplicação de uma regressão é imprescindível identificar primeiramente as variáveis dependentes e independentes e, em seguida, verificar se corresponde a uma regressão simples ou múltipla, linear ou outra, que pode ser avaliada primeiramente pelo  $r^2$  e depois pelos testes dos coeficientes angulares.

## REFERÊNCIAS BIBLIOGRÁFICAS

KMENTA, Jan. **Elementos de Econometria**. Tradução de Carlos Vieira Araújo. São Paulo: Atlas, 1978.

LAPPONI, Juan Carlos. **Estatística usando Excel Versões 5.0 e 7.0**. São Paulo: Laponi Treinamentos e Editora Ltda., 2000.

LEVINE, David M., BERENSON, Mark L., STEPHAN, David. **Estatística: Teoria e Aplicações**. Rio de Janeiro: LTC, 2000.

MATOS, O. C. **Econometria Básica**. São Paulo: Atlas, 1997.

STEVENSON, William J. **Estatística Aplicada à Administração**. São Paulo: Harper & Row, 1981.

VASCONCELLOS, Marco A. S., ALVES, Denisard (editores). **Manual de Econometria: Nível Intermediário**. São Paulo: Atlas, 2000.

WONNACOTT, Thomas, WONNACOTT, Ronald J. **Estatística Aplicada à Economia e à Administração**. Rio de Janeiro: LTC, 1981.