CONTABILIDADE DE CUSTOS – ANÁLISE TEÓRICA PARA A RELAÇÃO VOLUME VERSUS CUSTO DE PRODUÇÃO POR ANÁLISE DE REGRESSÃO

COST ACCOUNTING: THEORETICAL ANALYSIS FOR THE RELATION VOLUME VERSUS COST OF PRODUCTION BY REGRESSION ANALYSIS

Carlos Alberto Stechhahn da Silva[1]

Recebido em: 22/11/2020. Disponibilizado em: 05/12/2020

RESUMO

Neste trabalho serão aplicadas as técnicas da análise de regressão na estimação do custo de produção (em "reais") de posse dos dados da evolução mensal do volume de produção (em unidades). Após uma breve descrição do método, iremos obter as expressões para o cálculo das constantes b_0 e b_1. Cabe consignar que a obtenção de tais constantes, aqui feita passo a passo, envolve o conhecimento de derivada parcial e, portanto, é raramente descrita na literatura. A seguir, é feita uma aplicação de todo um tratamento estatístico com o intuito de se estudar uma situação particular da Contabilidade de Custos, qual seja, na relação volume de produção (em unidades) e custo. A relevância destes cálculos é natural, na medida em que o gestor poderá, por exemplo, calcular o valor do custo quando o volume de produção for maior que os até então apresentados. Ou seja, fazer uma estimativa de custo que a empresa terá no futuro.

Palavras chave: análise de regressão, Contabilidade de Custos, coeficiente de determinação, testes de significância e análise de variância (ANOVA).

ABSTRACT

This work will apply the techniques of regression analysis in the estimation of the cost of production (in "reais") in possession of the data of the monthly evolution of the volume of production (in units). After a brief description of the method, we will obtain the expressions to calculate the b_0 *and afterwards* b_1 *constant. It is up to state that obtaining such constants, here is made step by step, involves the knowledge of partial derivative and therefore is rarely described in the literature. The following is an implementation of an entire statistical treatment with the purpose of studying a particular situation of cost accounting, which is, in respect of volume production (in units) and cost. The relevance of these calculations is natural, to the extent that the Manager can, for example, calculate the value of the cost when the production volume is greater than the so far presented. That is, do a cost estimate that the company will have in the future.*

Keywords*: regression analysis, cost accounting, method of least squares, coefficient of determination, significance tests,* t *test,* F *test, analysis of variance (ANOVA).*

1 Universidade de São Paulo (USP) , Faculdades Integradas Campos Salles

ISSN 1518-9597 | e-ISSN 2316-3852 | doi: 10.22287/ag.v1i24.932

1. INTRODUÇÃO

A Contabilidade de Custos vem ganhando a cada dia elevado destaque e significativa atenção das empresas privadas e estatais. Conhecendo-se os custos de produção ou dos serviços é possível ao gestor estabelecer estratégias no sentido de se aumentar o lucro, reduzir custos e assim, minimizar o preço final de seus produtos ou gastos com a prestação se serviços em seus diversos setores de atuação. A Contabilidade de Custos, portanto, envolve temas bem contemporâneos, como por exemplo, as medidas não financeiras de desempenho e análise de custos, conhecimento do cliente ou de sua área de atuação, enfoque gerencial, entre outros. Com a Revolução Industrial, no final do século XVIII e início do XIX, as empresas, em face de terem uma medida do processo de produção, precisavam avaliar os custos de produção e a razão entre despesas, receitas e custos operacionais. Nesse período ocorreu uma integração entre a contabilidade industrial e a contabilidade geral. Desenvolveram-se maiores detalhes na movimentação e registro das matérias-primas, bem como, no registro e no cálculo dos custos de mão-de-obra. Houve ainda nessa época a ideia de custos fixos, custos variáveis e custos-padrão (STARK, 2013).

A Contabilidade ainda mantém as características desenvolvidas no século XVI, empregadas no método das partidas dobradas. Dessa forma, esse método é cartesiano e linear.

As empresas, desde a Revolução Industrial, sempre tiveram acesso aos dados (entradas e saídas) de sua matéria prima. Contudo, em pleno início do século XX, essas importantes informações não foram devidamente empregadas na gestão de custos. Trabalhava-se apenas o caráter financeiro. Nas décadas de 50 e 60 surgiram conceitos que modificaram este padrão de negócios, quais sejam, custo-padrão, análise de desvios, custo fixo, C_F, custo variável, $C_V(x)$, entre outros. Com tais parâmetros foi possível, por exemplo, se obter o ponto de equilíbrio.

Atualmente, novos métodos e processos de produção foram desenvolvidos. Frente a um imenso aparato tecnológico e de organização, surgiram novas técnicas e métodos de se tratar, por exemplo, o custeio industrial. Novos acrônimos surgiram, tais como, o MRP II, o CAD e o CAM como uma maneira de tratar a relação produto versus custos.

Finalizamos, destacando que a filosofia *Just-in-time, JIT,* foi criada no Japão com o objetivo de diminuir o tempo de produção e o tempo gasto em burocracias que não agregam valor. A gestão pela qualidade total, outro método contemporâneo, tem como foco na qualidade das operações e serviços realizados pela empresa. Exercer de modo melhor possível suas atividades e adotar uma conduta de melhora constante é "*missão*" de muitas empresas. Com esses valores, seus colaboradores devem exercer suas atividades de modo padrão descritos em procedimentos operacionais padrão (*POP´s*). Nesta área ainda temos o sistema de custos *ABC*, *Activity-based costing* indo além dos sistemas de custeio tradicionais (STARK, 2007).

Neste trabalho, iremos abordar o método estatístico da *regressão linear* no intuito de estreitar e aprimorar nosso conhecimento sobre a relação sobre o volume (em unidades) e custos de produção (em reais). Faremos uso, em diversos momentos, do Excel tanto na construção das tabelas e gráficos, bem como, na obtenção dos valores quantitativos dos parâmetros.

2. A REGRESSÃO LINEAR SIMPLES NAS DECISÕES ADMINISTRATIVAS

Frequentemente a contabilidade se depara em suas decisões com a relação entre duas variáveis. Podemos citar, nesta situação, a estimação do custo de produção. Coletando-se dados sobre o volume de produção, em unidades, bem como, do custo para produção dessas unidades em cada etapa do processo. Podemos utilizar, por exemplo, um procedimento estatístico chamado análise de regressão para obtermos uma equação que nos mostra como as duas variáveis se relacionam. Nesta análise de regressão, empregando a variável que é prevista, variável dependente e as variáveis independentes podemos projetar ou prever para o futuro um determinado resultado. Assim, uma vez de posse da equação que relaciona tais variáveis, o volume de produção acima citado, num tempo futuro, poderá fornecer o valor do custo de produção.

Vamos considerar neste trabalho a regressão linear simples, a qual envolve apenas uma variável dependente e outra independente. Veremos situações em que a relação entre ambas pode ser aproximada por uma reta. Uma importante aplicação, embora fora do escopo deste trabalho, é o modelo no qual temos mais de uma variável independente. Tal método é chamado de análise de regressão múltipla.

2. 1 O modelo de regressão linear

Em contabilidade de custos, a condição do aumento do volume de produção, em geral, conduz a um aumento custo de produção dessas unidades. Para cada peça produzida x, a qual é uma variável independente, temos um único valor de y, a variável dependente. Como temos um conjunto de pares ordenados (x_i, y_i), na aproximação por uma reta, surgirá um erro, denotado por ϵ. Para o modelo de regressão linear simples temos a seguinte equação:

$$y = \beta_0 + \beta_1 x + \epsilon \qquad (1)$$

em que β_0 é o coeficiente linear e β_1 o coeficiente linear. A última constante, ϵ, é o erro aleatório ou parcela de erro. Esse erro é devido a variação de y, uma vez que nem todos os pontos estão sobre a reta, e não pode ser explicada pela relação entre x e y.

Temos, neste contexto, a equação de regressão a qual define uma reta que permite fazer a estimativa do valor de uma variável em face do conhecimento da outra. A equação de regressão tem a seguinte forma:

$$E(y) = \beta_0 + \beta_1 x \qquad (2)$$

Se conhecermos os valores populacionais β_i podemos usar a Eq. (2) e obtermos o valor médio de y para um dado valor de x. No entanto, na prática tais valores não são conhecidos. Dessa forma, usamos as estatísticas amostrais (b_0 e b_1) como estimativas dos parâmetros populacionais. Substituindo-se o par b_i no lugar de β_i, $i = 0 \ e \ 1$, na equação de regressão (2) temos a equação de regressão estimada

$$\hat{y} = b_0 + b_1 x' \qquad (3)$$

ISSN 1518-9597 | e-ISSN 2316-3852 | doi: 10.22287/ag.v1i24.932

cujo gráfico da equação de regressão linear simples estimada é chamado de reta de regressão estimada.

Usaremos o método dos mínimos quadrados para estimar os valores amostrais b_i da Eq. (3).

2. 2 O método dos mínimos quadrados – Dedução das equações de inclinação e de intersecção da equação de regressão estimada

No intuito de desenvolver a equação de regressão estimada para prever o custo total para um dado volume de produção (quantidade de unidades produzidas) iremos, a seguir, deduzir as equações da inclinação b_1 e de intersecção com o eixo y, aqui denotada por b_0.

A maioria dos livros de Estatística Aplicada à Administração, Economia ou na área da financeira em geral, expressam as fórmulas da regressão linear prontas, ou seja, não mostram ao leitor como os coeficientes linear e angular, b_0 e b_1, são obtidos.

Vamos inicialmente assumir que os dados amostrais podem ser aproximados por uma função afim. Assim, vamos utilizar o modelo de regressão linear simples para a relação entre (x_i, y_i). Vamos então procurar escrever as equações para os valores b_0 e b_1 da equação de regressão linear simples estimada. Para um elemento da classe i genérico vimos que a Eq. (3) pode ser escrita como:

$$\hat{y}_i = b_0 + b_1 x_i \tag{4}$$

O método dos mínimos quadrados é um procedimento para se calcular os valores b_i, com $i = 1,2$, que minimizam a soma dos quadrados residuais, esta soma é dada por

$$\sum (y_i - \hat{y}_i)^2 . \tag{5}$$

Substituindo-se a Eq. (4) na Eq. (5) temos:

$$\sum (y_i - b_0 - b_1 x_i)^2 \tag{6}$$

a qual deve ser minimizada. Cabe consignar que a expressão acima mede o afastamento (ou desvio) da média, ao quadrado, no intuito de se eliminar o sinal negativo em alguma das diferenças.

Para minimizar a Eq. (6), acima, devemos calcular as derivadas parciais com respeito a b_0, bem como, a b_1 e igualar a zero. Cabe informar que a teoria envolvendo derivadas parciais pode ser encontrada em (STEWART, 2016). Assim, temos:

$$\frac{\partial \sum (y_i - b_0 - b_1 x_i)^2}{\partial b_0} = -2 \sum (y_i - b_0 - b_1 x_i) = 0 \tag{7}$$

e

$$\frac{\partial \sum (y_i - b_0 - b_1 x_i)^2}{\partial b_1} = -2 \sum x_i (y_i - b_0 - b_1 x_i) = 0 \tag{8}$$

Dividindo-se a Eq. (7) por 2 e calculando a soma de cada um dos três termos individualmente, obtemos:

$$-\sum y_i + \sum b_0 + \sum b_1 x_i = 0 \tag{9}$$

i.e.,

$$\sum y_i = nb_0 + (\sum x_i) b_1 \tag{10}$$

em que $\sum b_0 = nb_0$.

Analogamente para a Eq. (8) temos:

$$\sum x_i y_i = (\sum x_i^2) b_1 + (\sum x_i) b_0 \tag{11}$$

Conforme sabemos b_0 e b_1 são constantes que independem do índice de soma, ou seja, podem ser passados para fora do somatório. As equações (10) e (11) são conhecidas como equações Normais de Mínimos Quadrados.

Isolando-se b_0 na Eq. (10) resulta, para o coeficiente linear:

$$b_0 = \frac{\sum y_i}{n} - b_1 \frac{\sum x_i}{n}. \tag{12}$$

Substituindo-se a Eq. (12) acima na Eq. (11) e isolando b_1, teremos:

$$\frac{\sum x_i y_i}{n} - \frac{(\sum x_i)^2}{n} b_1 + (\sum x_i^2) b_1 = \sum x_i y_i \tag{13}$$

$$(\sum x_i^2) b_1 - \frac{(\sum x_i)^2}{n} b_1 = \sum x_i y_i - \frac{\sum x_i \sum y_i}{n} \tag{14}$$

$$b_1 = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{n\sum x_i^2 - (\sum x_i)^2} \tag{15}$$

ou ainda,

$$b_1 = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2}. \tag{16}$$

Devido ao fato de $\bar{y} = \sum y_i / n$ e $\bar{x} = \sum x_i / n$ podemos escrever a Eq. (12) como:

$$b_0 = \bar{y} - b_1 \bar{x} \tag{17}$$

As Eqs. (15), ou ainda a Eq. (16), bem como a Eq. (17) serão aqui utilizadas para o cálculo da equação de regressão estimada; esta usada para prever o custo total de um determinado volume de produção.

3. APLICAÇÃO EM CONTABILIDADE DE CUSTOS

A análise de regressão, teoria acima descrita, pode ser aplicada na estimação de custo y_i. Uma vez que a empresa tenha coletado os dados da evolução mensal do volume de produção (em unidades), para cada elemento x_i teremos o correspondente custo total (em reais). Assim, um profissional da contabilidade pode estimar o valor do custo para um dado volume de fabricação. Poderá, por exemplo, calcular o valor do custo quando o volume de produção for maior que os até então apresentados. Ou seja, fazer uma estimativa de custo que a empresa terá no futuro.

Consideremos uma amostra de seis volumes de produção x_i e os seus respectivos custos para as operações da fábrica, y_i.

Tabela 1 – Volume de Produção (em unidades) versus Custo total (em reais).

Volume da Produção, x_i	Custo total, y_i
80	800
90	1000
110	1080
120	1180
140	1280
150	1400

Fonte: O autor.

Uma vez coletados os dados amostrais (x_i, y_i), teremos um diagrama de dispersão. Nele podemos observar se há uma relação positiva, negativa ou mesmo uma constante (sem relação) entre a variável independente e a dependente, vide Figura 1, a seguir.

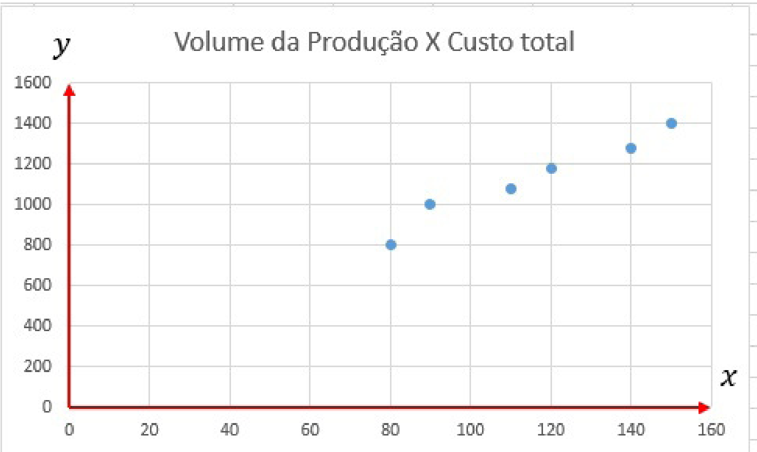
Vejamos, para a Tabela 1, acima, como obter o valor do desvio médio simples, *DMS*. Este desvio pode ser obtido pela Eq. abaixo descrita:

$$DMS = \frac{\sum |x_i - \bar{x}|}{n} = 21{,}67, \tag{18}$$

pois, resulta do quociente 130/6.

Da Tabela 1, acima, obtemos, usando o Excel, o diagrama de dispersão.

Figura 1 – Diagrama de dispersão – Volume (em unidades) versus Custo total (em reais)



Fonte: O autor.

ISSN 1518-9597 | e-ISSN 2316-3852 | doi: 10.22287/ag.v1i24.932

Observando o diagrama de dispersão acima podemos observar uma correlação positiva entre x_i e y_i.

O método de construção do diagrama de dispersão pode ser encontrado em (LIMA, 2016).

Conforme vimos, os valores de y_i representam os custos (valor observado dos custos), em reais, na produção de cada volume de produção (variável independente). Teremos ainda em $\hat{y}_i$ um valor estimado de custos.

No intuito de que a regressão estimada resulte num ajuste eficiente para os nossos dados, vamos impor que as diferenças entre os valores de custos observados e os estimados sejam um mínimo.

Com os nossos dados amostrais vamos calcular os valores de b_0 e b_1 de forma a minimizar a soma dos quadrados dos desvios entre os valores observados (x_i, y_i).

Este método aqui empregado é conhecido na literatura como método dos mínimos quadrados.

As Eqs. (16) e (17) serão empregadas para obtermos os valores de b_0 e b_1. Utilizaremos para tanto uma planilha do Excel. Nela podemos inserir as devidas colunas de dados para obtermos suas respectivas somas.

O valor das somas de cada coluna também pode ser encontrado via manipulação da HP 12C.

Maiores detalhes podem ser encontrados no vídeo (GRINGS, 2013).

A seguir, usamos uma planilha Excel para o cálculo dos termos que envolvem o cálculo dos coeficiente b_0 e b_1.

Tabela 2 – Planilha Excel para o cálculo de b_0 e b_1 $(n = 6)$.

x	y	x.y	x^2	y^2
80	800	64000	6400	640000
90	1000	90000	8100	1000000
110	1080	118800	12100	1166400
120	1180	141600	14400	1392400
140	1280	179200	19600	1638400
150	1400	210000	22500	1960000
690	6740	803600	83100	7797200

Fonte: O autor.

ISSN 1518-9597 | e-ISSN 2316-3852 | doi: 10.22287/ag.v1i24.932

O uso dos resultados da Tabela 2 leva-nos aos seguintes valores, respectivamente, para a inclinação e intersecção da equação de regressão estimada:

$$b_1 = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{n\sum x_i^2 - (\sum x_i)^2} = 7{,}60 \tag{19}$$

e

$$b_0 = \bar{y} - b_1 \bar{x} = 249{,}33 \tag{20}$$

em que $\bar{y} = 1.123{,}33$.

Uma constatação bastante interessante, para as devidas mudanças que podem ocorrem nos dados da Tabela 2, acima, é que se dobrarmos os valores de cada par (x_i, y_i) o valor de b_0 dobrará e o valor para b_1 permanecerá constante. Isto se deve a própria natureza das fórmulas (18) e (19). Neste caso, teremos um deslocamento da reta para cima, modificando o valor do coeficiente linear, e permanecendo a reta com a mesma inclinação.

Com estes resultados podemos escrever uma equação de regressão estimada que pode ser utilizada para prever o custo total para um determinado volume de produção futuro. Vide Tabela 3, a seguir.

Tabela 3 – Planilha Excel para o custo total no futuro – Cálculo de y_n e y_{n+1}.

Classe	*Volume da produção (unidades)*	*Custo total (reais)*
1	80	800
2	90	1000
3	110	1080
	⋮	⋮
6	150	1400
	⋮	⋮
n	900	$\hat{y}_n$
n + 1	1200	$\hat{y}_{n+1}$

Fonte: O autor.

Das Eqs. (18) e (19) temos que a equação de regressão estimada pode ser escrita como:

$$\hat{y} = 249{,}33 + 7{,}60x \tag{21}$$

Para os valores de volume, 900 e 1.200, devido à Eq. (20), temos, respectivamente, $\hat{y}_n = 7.089{,}33$ e $\hat{y}_{n+1} = 9.369{,}33$.

O custo variável $C(x) = 7{,}60.x$ por unidade produzida será de R$ 7,60.

A seguir faremos uma análise da qualidade do ajuste da equação de regressão estimada.

ISSN 1518-9597 | e-ISSN 2316-3852 | doi: 10.22287/ag.v1i24.932

3. 1 Coeficiente de Determinação

As medidas amostrais produzem, em geral, pontos que estão sobre a reta de regressão estimada, *RRE*, ou fora dela. O afastamento das observações em torno da *RRE* envolve diferenças de ordenadas. Na seção anterior obtivemos, conforme Eq. (20), a equação de regressão estimada. É uma função afim que relaciona o volume produzido com o custo para cada item ou classe. No entanto, podemos, por meio do Coeficiente de Determinação, obter uma ideia da qualidade de ajuste da *RRE*.

Da teoria de Regressão Linear Simples temos as seguintes definições:

I. Soma dos quadrados dos resíduos, *SQRes* (*SSE* para o acrônimo em inglês):

$$SQRes = \sum (y_i - \hat{y})^2$$

II. Soma dos quadrados total, *SQTot* (*SST* para o acrônimo em inglês):

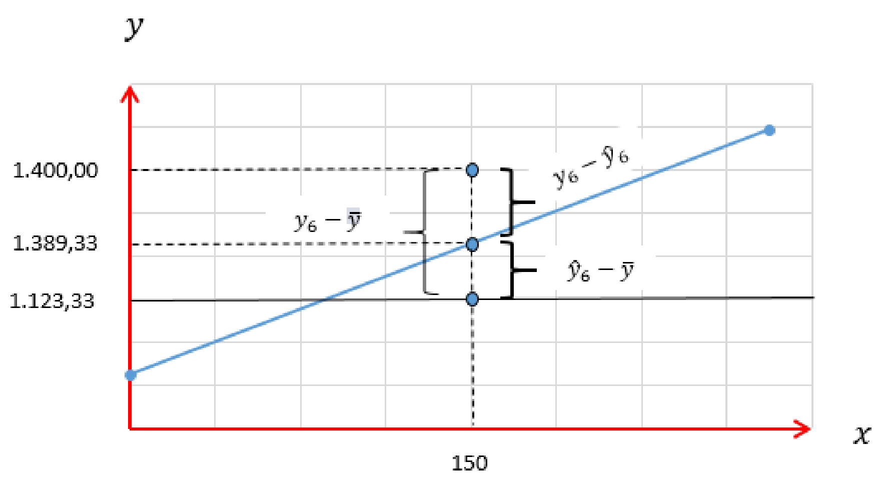
$$SQTot = \sum (y_i - \bar{y})^2$$

III. Soma dos quadrados da Regressão, *SQReg* (*SSR* para o acrônimo em inglês):

$$SQReg = \sum (\hat{y}_i - \bar{y})^2$$

A relação entre estas três grandezas acima descritas pode ser obtida a partir das diferenças de ordenadas do gráfico mostrado na Figura 2, a seguir.

Figura 2 – Reta de Regressão Estimada (em azul) e a reta na cor cinza (horizontal) para $\bar{y}$.



Fonte: O autor.

ISSN 1518-9597 | e-ISSN 2316-3852 | doi: 10.22287/ag.v1i24.932

A partir das três diferenças para as ordenadas ilustradas acima, temos:

$$\hat{y}_6 - \bar{y} = (y_6 - \bar{y}) - (y_6 - \hat{y}_6) \tag{22}$$

Comparando com as relações I. a III., podemos escrever para a soma dos quadrados da regressão a seguinte relação:

$$SQReg = SQTot - SQRes \tag{23}$$

Novamente, uma planilha do Excel nos permitirá calcular o valor para *SQReg.*

Tabela 4 – Planilha Excel para obtenção da *SQReg* da Eq. (22)

x	y	$\lvert y_i - \bar{y}\rvert$	$(y_i - \bar{y})^2$	$\hat{y}_i$	$\lvert y_i - \hat{y}_i\rvert$	$(y_i - \hat{y}_i)^2$	$\lvert \hat{y}_i - \bar{y}\rvert$	$(\hat{y}_i - \bar{y})^2$
80	800	323,33	104542,29	857,33	57,33	3286,73	266	70756
90	1000	123,33	15210,29	933,33	66,67	4444,89	190	36100
110	1080	43,33	1877,49	1085,33	5,33	28,41	38	1444
120	1180	56,67	3211,49	1161,33	18,67	348,57	38	1444
140	1280	156,67	24545,49	1313,33	33,33	1110,89	190	36100
150	1400	276,67	76546,29	1389,33	10,67	113,85	266	70756
690	6740		225933,33			9333,33		216600
			SQTot			SQRes		SQReg
	$\bar{y} = 1.123,33$							

Fonte: O autor.

Como podemos constatar, a diferença entre o *SQTot* e o *SQRes* fornece-nos exatamente o valor de *SQReg*, o qual é dado por R$ 216.600,00.

O coeficiente de determinação r^2 mede a qualidade do ajuste da equação de regressão estimada e pode ser escrito como:

$$r^2 = \frac{SQReg}{SQTot} = \frac{216.600}{225.933,33} \approx 0,9587 = 95,87\% \tag{24}$$

Este resultado é a porcentagem da soma dos quadrados total que pode ser explicada pela nossa equação de regressão estimada (20).

É importante aqui considerar que as três somas acima descritas (*SQTot , SQRes e SQReg*), bem como, o valor de r^2, descrito na Eq. (23), podem ser obtidos pela tabela ANOVA do Excel, a qual mostraremos na Seção 3.1 deste trabalho.

Estamos, portanto, cientes que é um excelente ajuste para a Eq. (20). Assim, aproximadamente 96% dos custos de produção (em reais) da empresa podem ser explicados pelo volume da produção em unidades.

ISSN 1518-9597 | e-ISSN 2316-3852 | doi: 10.22287/ag.v1i24.932

4. OS TESTES DE SIGNIFICÂNCIA NA REGRESSÃO LINEAR SIMPLES

Em nossos estudos fizemos uma análise de regressão. Iniciando com uma suposição sobre o modelo ideal para a relação entre as variáveis dependentes e independentes. Vimos que num modelo de regressão devemos nos valer das equações (1) e (3). Temos ainda que o coeficiente de determinação r^2 representa a qualidade de ajuste da equação de regressão estimada. No entanto, mesmo que o valor obtido para r^2 seja muito próximo de 1, devemos fazer uma análise adicional de adequação do modelo suposto. O teste apropriado chama-se teste de significância. Tais testes na análise de regressão se apoiam em algumas suposições sobre a parcela de erro ϵ que é uma variável aleatória com média, ou valor esperado, igual a zero i.e., $E(\epsilon) = 0$. Sendo β_i´s constantes, logo, $E(\beta_i) = \beta_i$ com $i = 1,2$. Portanto,

$$\langle y \rangle = E(y) = \beta_0 + \beta_1 x \tag{25}$$

que é a equação de regressão (2). Temos, neste caso, que têm distribuição normal as distribuições de probabilidade de y.

A variância de ϵ, σ^2 descreve a variância de y_i nas proximidades da reta de regressão. Como, em geral, temos os desvios dos valores de y (resíduos) na vizinhança da reta de regressão, a *SQRes* é uma medida da variação dos resultados observados (custos) em torno da reta estimada.

Vamos definir o quadrado médio dos resíduos (*QMRes*) como uma estimativa de σ^2, i.e.,

$$s^2 = QMRes = \frac{SQRes}{n-2}. \tag{26}$$

Em nosso caso temos que:

$$s^2 = \frac{9.333{,}33}{6-2} = 2.333{,}33 \tag{27}$$

(*MQ Resíduo* da Tabela ANOVA) fornece uma estimativa não viesada (ou não tendenciosa) de σ^2, pois, conforme vimos a variância é a mesma para todos os valores de x. O erro padrão da estimativa é denotado por:

$$s = \sqrt{QMRes} = \sqrt{\frac{SQRes}{n-2}} \approx 48{,}3. \tag{28}$$

O valor de s, descrito acima, Erro padrão da Tabela ANOVA, será usado na próxima seção para o cálculo de uma estimativa para σ_{b_1}.

A seguir iremos tratar dos testes t e F ao nível de significância 0,05. Usaremos o erro padrão da estimativa nestes testes de relação entre as variáveis x e y.

ISSN 1518-9597 | e-ISSN 2316-3852 | doi: 10.22287/ag.v1i24.932

4. 1 O teste t

Vimos que se há uma relação linear entre (x_i, y_i) devemos ter que a constante $\beta_1 \neq 0$.

A motivação para o teste t é para verificarmos se podemos, por meio dos testes de hipóteses, verificar se podemos concluir que

$$H_0: \beta_1 = 0$$

$$H_a: \beta_1 \neq 0$$

Figura 3 – Distribuição Amostral de b_1

$$Valor\ esperado$$

$$E(b_1) = \beta_1$$

$$Desvio\ padrão\ estimado$$

$$\sigma_{b_1} = \frac{\sigma}{\sqrt{\sum (x_i - \bar{x})^2}}$$

$$Forma\ da\ distribuição$$

$$Normal$$

Fonte: O autor.

Neste teste, se a hipótese nula for rejeitada podemos concluir que $\beta_1 \neq 0$. Maiores detalhes sobre os Testes de Hipóteses podem ser encontrados, por exemplo, em (LARSON, 2010). No entanto, se H_0 não pode ser rejeitada as evidências são pequenas para concluirmos que existe uma relação significativa. Em nossos cálculos teremos a distribuição amostral de b_i, o estimador de mínimos quadrados de β_i, constituem a base de nosso estudo. Particularmente, para b_1, sua distribuição amostral tem valor esperado $E(b_1) = \beta_1$, ou seja, é um estimador não viesado de β_1 com desvio padrão

$$\sigma_{b_1} = \frac{\sigma}{\sqrt{\sum (x_i - \bar{x})^2}} \tag{29}$$

com distribuição normal. Contudo, em nosso caso, não conhecemos o valor do desvio padrão σ. Assim, iremos estabelecer uma estimativa de σ_{b_1} qual seja:

$$s_{b_1} = \frac{s}{\sqrt{\sum (x_i - \bar{x})^2}} = \frac{48{,}3}{\sqrt{3750}} \approx 0{,}79. \tag{30}$$

ISSN 1518-9597 | e-ISSN 2316-3852 | doi: 10.22287/ag.v1i24.932

Esse resultado, Erro padrão x da tabela ANOVA, descrito na Eq. (29) será aqui considerado como o desvio padrão estimado de b_1. O teste t para uma relação de significância está baseado no fato que o teste estatístico

$$t = \frac{b_1 - \beta_1}{s_{b_1}} \tag{31}$$

tem uma distribuição t com $n-2$ graus de liberdade. Dessa forma, se a hipótese nula H_0, i.e., $\beta_1 = 0$, for verdadeira, então, da equação (31) temos:

$$t = \frac{b_1}{s_{b_1}} = \frac{7{,}60}{0{,}79} = 9{,}63, \tag{32}$$

o qual é apresentado como Stat *t* da Tabela ANOVA.

Usando uma tabela de distribuição t para $n-2=6-2=4$ graus de liberdade, $t = 4{,}604$ representa uma área igual a 0,005 na causa superior. Assim, na cauda superior, a área da distribuição t correspondente à estatística de teste $t = 9{,}63$ deve, portanto, ser menor que 0,005. Sendo o teste bicaudal o valor-p associado ao valor t da Eq. (32) deve ser menor que $0{,}01 = 2.0{,}005$. Usando o Excel, com as seguintes abas em sequência

FÓRMULAS, MAIS FUNÇÕES, ESTATÍSTICA, DIST.T

surgirá na tela: Argumentos da função. Nesta caixa devemos inserir o valor obtido para t, ou seja, $9{,}63$ e, a seguir, 4 para graus de liberdade e a palavra VERDADEIRO. Com os dados supramencionados obtemos para um valor-p muito pequeno, ou seja, $1{,}27.10^{-5}$. Tendo em vista que o valor-p é menor que 0,01, devemos rejeitar H_0 e concluímos que β_1 deve realmente ser diferente de zero. Aceitamos a Hipótese alternativa, H_a. Tal evidência, obtida por meio deste teste, é suficiente para que possamos concluir que existe uma relação linear entre as variáveis volumes e custos.

4. 2 O teste F

Este teste se apoia na distribuição de probabilidade F e será usado para testar a significância na regressão. Testa, portanto, a razão entre duas variâncias. Tendo em vista que temos somente uma variável independente, x, o teste F resultará na mesma conclusão que obtivemos para o teste t. Como obtivemos que $\beta_1 \neq 0$, teremos este mesmo resultado para o teste F. No entanto, se tivermos mais de uma variável independente, o teste F (e somente ele) poderá testar uma relação significativa global.

Vamos iniciar impondo que, se a hipótese nula, dada por $H_0\colon \beta_1 = 0$ for verdadeira, a razão entre *SQReg* e os graus de liberdade da regressão produzirá uma estimativa para a variância σ^2. Tal razão é denominada quadrado médio da regressão e pode ser escrito como:

$$QMReg = \frac{SQReg}{graus\ de\ liberdade\ da\ regressão} \tag{33}$$

ISSN 1518-9597 | e-ISSN 2316-3852 | doi: 10.22287/ag.v1i24.932

É importante destacar que, neste trabalho, os graus de liberdade da regressão são iguais ao número de variáveis independentes do modelo adotado. Dessa forma esse número é igual a 1. Portanto, para os valores da soma, descrito na Eq. (23):

$$QMReg = \frac{SQReg}{núm.\ de\ variáveis\ independentes} = \frac{216.600}{1}. \tag{34}$$

Tal resultado se encontra descrito como *MQ* Regressão na tabela ANOVA. Assim, para os dados da Tabela 4, *C* = *SQReg* = 216.600.

A constante β_1 participa da equação de regressão, e caso assuma o valor zero, a hipótese nula será verdadeira, e consequentemente *SQReg* e *SQRes* são duas estimativas independentes da variância σ^2. Neste caso, a distribuição amostral segue uma distribuição *F*.

O teste f de significância na regressão linear simples é

$$F = \frac{QMReg}{QMRes} = \frac{216.600}{2.333,33} = 92,83 \tag{35}$$

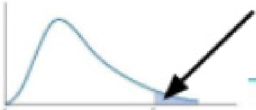
Para 1 grau de liberdade no numerador e 4 graus de liberdade no denominador temos que:

$$F_{0,05} = 7,71. \tag{36}$$

A obtenção deste resultado pode ser visualizada na Tabela 05, a seguir.

ISSN 1518-9597 | e-ISSN 2316-3852 | doi: 10.22287/ag.v1i24.932

Tabela 5 – Os registros dos valores $F_{\sim}$ onde $\alpha = 0{,}05$ é a área (ou probabilidade) na cauda superior da distribuição F.



Graus de liberdade no denominador	Área na cauda superior	Graus de liberdade no denominador: 1	2	3	4	5
1	.10	39.86	49.50	53.59	55.83	57.24
	.05	161.45	199.50	215.71	224.58	230.16
	.025	647.79	799.48	864.15	899.60	921.83
	.01	4052.18	4999.34	5403.53	5624.26	5763.96
2	.10	8.53	9.00	9.16	9.24	9.29
	.05	18.51	19.00	19.16	19.25	19.30
	.025	38.51	39.00	39.17	39.25	39.30
	.01	98.50	99.00	99.16	99.25	99.30
3	.10	5.54	5.46	5.39	5.34	5.31
	.05	10.13	9.55	9.28	9.12	9.01
	.025	17.44	16.04	15.44	15.10	14.88
	.01	34.12	30.82	29.46	28.71	28.24
4	.10	4.54	4.32	4.19	4.11	4.05
	.05	7.71	6.94	6.59	6.39	6.26
	.025	12.22	10.65	9.98	9.60	9.36
	.01	21.20	18.00	16.69	15.98	15.52
5	.10	4.06	3.78	3.62	3.52	3.45
	.05	6.61.	5.79	5.41	5.19	5.05
	.025	10.01	8.43	7.76	7.39	7.15
	.01	16.26	13.27	12.06	11.39	10.97

Fonte: O autor.

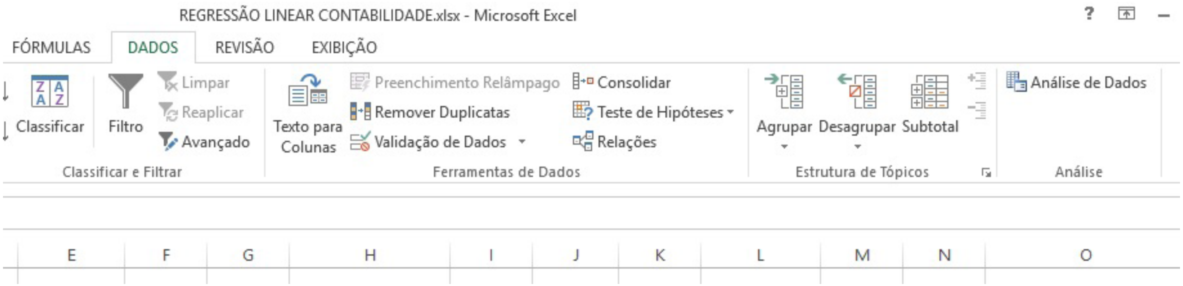
Com o resultado da Eq. (35) temos que, $F > F_{0,05}$, ou seja, $92{,}83 > 7{,}71$ rejeitamos H_0 e concluímos que o volume de produção, em unidades, e o custo total estão relacionados.

Conforme expomos, a maioria dos valores acima descritos podem ser obtidos da tabela ANOVA a qual iremos, passo a passo, descrever como obtê-la.

Selecionando-se apenas as colunas de x_i e de y_i clicamos na aba DADOS e ANÁLISE DE DADOS.

ISSN 1518-9597 | e-ISSN 2316-3852 | doi: 10.22287/ag.v1i24.932

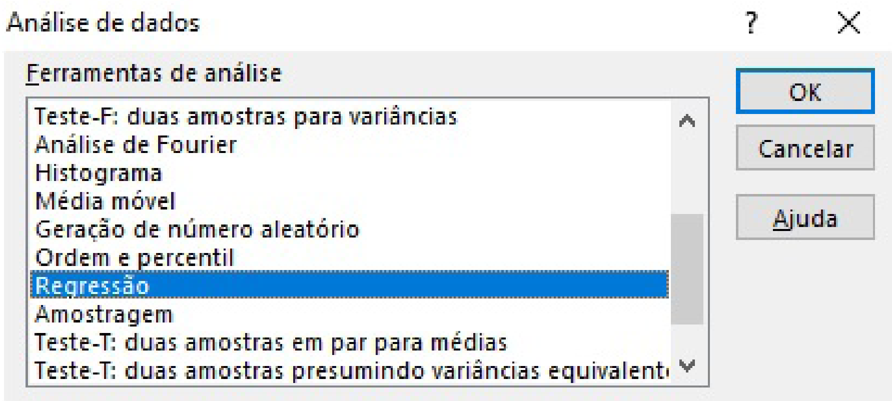
Figura 4 – Passos iniciais para obtenção da tabela ANOVA.



Fonte: O autor.

Em seguida, abrir-se-á uma nova janela. Nela, devemos selecionar REGRESSÃO.

Figura 5 – Passo seguinte para obtenção da tabela ANOVA



Fonte: O autor.

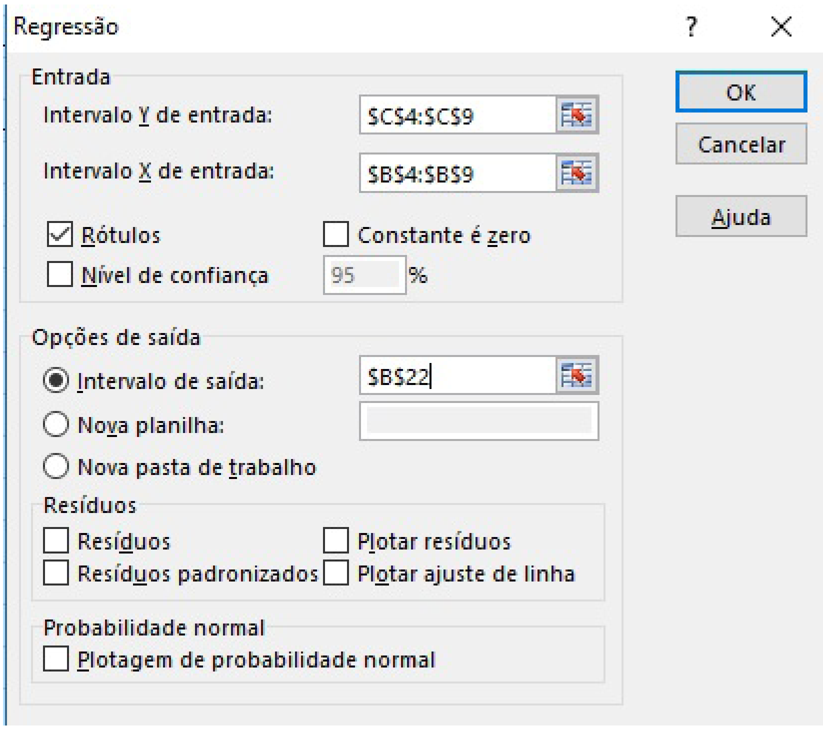
Finalizamos com o preenchimento da última janela. Devemos selecionar primeiro os dados (sem o rótulo) da coluna y_i. A seguir, selecionamos os dados da coluna x_i, novamente sem o rótulo superior. Clicamos em rótulos e, por último em intervalo de saída. A inserção da tabela ANOVA será feita na célula que determinarmos após o preenchimento do retângulo destinado ao intervalo de saída.

Teremos, portanto, uma última janela pronta com o formato do tipo apresentado a seguir.

Este tipo de tarefa, conforme podemos perceber, exige um pouco de prática no Excel.

O acrônimo ANOVA representa a abreviação em inglês de "análise de variância" (*analysis of variance*).

Figura 6 – Último passo para obtenção da tabela ANOVA.



Fonte: O autor.

A tabela ANOVA apresentada pelo Excel terá o seguinte mostrado na Tabela 6, a seguir.

Tabela 6 – Tabela ANOVA – valor da estatística de teste F – Graus de liberdade, entre outros.

ANOVA					
	gl	*SQ*	*MQ*	*F*	*F de significação*
Regressão	1	216600	216600	92,82857	0,00064897
Resíduo	4	9333,333333	2333,333		
Total	5	225933,3333			

RESUMO DOS RESULTADOS	
Estatística de regressão	
R múltiplo	0,979127101
R-Quadrado	0,958689879
R-quadrado ajustado	0,948362349
Erro padrão	48,30458915
Observações	6

Fonte: O autor.

ISSN 1518-9597 | e-ISSN 2316-3852 | doi: 10.22287/ag.v1i24.932

Conforme vimos a tabela ANOVA apresenta o valor da estatística de teste *F*, qual seja, $F = 92{,}83 > 7{,}71$ o que nos levou a rejeita H_0 e concluir que o volume de produção e o custo total estão relacionados.

5. CONSIDERAÇÕES FINAIS

Vimos, neste trabalho, uma importante e avançada aplicação de uma ferramenta da Estatística, a análise de regressão, aplicada à Contabilidade de Custos. Atualmente, este ramo da Contabilidade Geral envolve conceitos, princípios e práticas contábeis aplicadas aos custos. Seus sistemas e métodos de custeio envolvem o custeio *ABC – Activity-based costing*, Custo-ativo – *kaizen*, método de custeio por processo, método da unidade esforço de produção (UEP) entre outros. Torna-se, portanto, uma área bastante vasta e de interessantes aplicações.

Aqui, consideramos apenas uma amostra de seis volumes de produção x_i e os seus respectivos custos para as operações da fábrica, y_i. Usando o método estatístico da regressão linear, para a relação volume (em unidades) e custos de produção (em reais), fizemos uso em diversos momentos do Excel, tanto na construção das tabelas e gráficos apresentados, bem como, na obtenção dos valores quantitativos dos parâmetros.

Iniciamos com o modelo de regressão linear. Tendo em vista que, a maioria dos livros de Estatística, principalmente os com foco na Administração, Economia ou área financeira em geral, expressam as fórmulas da regressão linear prontas, mostramos ao leitor como os coeficientes linear e angular, b_0 e b_1, são obtidos. E, com método dos mínimos quadrados fizemos a dedução da Eq. (20), a equação de regressão estimada. Esta permite prever o custo total de um determinado volume de produção em tempos futuros.

A seguir, a teoria se desenvolve e o coeficiente de determinação r^2, o qual mede a qualidade do ajuste da equação de regressão estimada, foi calculado pela relação entre a soma dos quadrados da Regressão pela soma dos quadrados total. O resultado obtido para $r^2 = 95{,}87\%$ mostra que este 95,87% da variabilidade dos custos podem ser explicados por meio da relação linear existente entre o tamanho do volume da produção (em unidades) e o custo (em reais). O ajuste, portanto, foi muito bom para a nossa equação de regressão estimada (20).

Finalizamos com um capítulo sobre os Testes de Significância na Regressão Linear Simples. Parâmetros como a variância amostral e número de graus de liberdade permitiram, por meio do teste t, usando testes de hipóteses da teoria bicaudal, e do teste F concluir nossos resultados. Tais testes, foram suficientes para concluirmos que existe uma relação linear entre as variáveis volumes e custos, com $\alpha = 0{,}05$.

Cabe finalmente consignar que, os resultados aqui obtidos por meio do Excel podem também ser encontrados por meio da tabela ANOVA. Todos os detalhes para a obtenção dos dados apresentados por meio desta tabela foram detalhados passo a passo neste trabalho.

Outras importantes aplicações envolvem este tema, embora se encontrem fora do escopo deste trabalho. Numa situação real, os modelos não envolvem somente um parâmetro x_i (volume

ISSN 1518-9597 | e-ISSN 2316-3852 | doi: 10.22287/ag.v1i24.932

de produção) para a obtenção dos custos y_i. Tais modelos envolvem mais de uma variável independente, ou seja, usam a análise de regressão múltipla.

REFERÊNCIAS

GRINGS - Correlação e Regressão com uso da HP 12c - aula 23 - (2013).

Disponível em:

< **https://www.youtube.com/watch?v=tvxpdr1SIAI** > Acesso em 19 abril 2019.

LARSON, R.; FARBER, B., ESTATÍSTICA APLICADA – 4ª. Edição (2010).

LIMA, R. H. P. - Diagrama de Dispersão (Ferramenta da Qualidade): Teoria + Exemplo no Excel (2016).

Disponível em:

< **https://www.youtube.com/watch?v=rTcp161k8q4** > Acesso em 19 abril 2019.

STEWART, J – CÁLCULO – Volume 2 – 6ª. edição (2016)

SÁ, S. Manual de Microsoft Excel, (2015).

Disponível em: < **http://www.ispa.pt/ficheiros/documentos/microsoft_excel.pdf** >. Acesso em: 11 abr. 2019.

STARK, J. A. – CONTABILIDADE CUSTOS – São Paulo: Pearson Prentice Hall (2007).

SWEENEY, D. J. et. al. – ESTATÍSTICA APLICADA à administração e economia – 3ª. Edição (2015).

INFORMAÇÕES DO AUTOR

Carlos Alberto Stechhahn da Silva é mestre e doutor em física pela Universidade de São Paulo (USP) e professor da Faculdades Integradas Campos Salles. E-mail: **carlos.silva@cs.edu.br**

ISSN 1518-9597 | e-ISSN 2316-3852 | doi: 10.22287/ag.v1i24.932