



Análise de Correlação e Regressão Linear Simples: Contabilometria Aplicada em Indicadores Econômico-Financeiros de 2009 das Empresas de Capital Aberto do Seguimento de Construção Civil Integrantes do Novo Mercado.

Anderson Martins de Melo
Universidade Estadual de Londrina (UEL)
ander_splynter@outlook.com

Resumo

Esta epistemologia objetiva verificar a existência de correlação entre alguns indicadores econômico-financeiros de 2009 e estimar um modelo de regressão que prediga informações para ramo de construção civil, setor que desempenha importante papel na economia nacional. O ambiente altamente globalizado e competitivo vivenciado atualmente pelas organizações traz consigo uma necessidade intrínseca em gerar informações estratégicas e valiosas para a tomada de decisão. Utilizando dos artifícios da Contabilometria, a metodologia aplicada neste artigo foi a pesquisa bibliográfica e documental, empregando a coleta de dados primários e da amostragem semiprobabilística para realizar um estudo descritivo. Desta forma, pode-se constatar que existe uma forte correlação entre os indicadores de rentabilidade do seguimento, dando suporte para, assim, estimar um modelo de regressão que atenda as hipóteses formuladas no decorrer do trabalho e que satisfaça estatisticamente as necessidades para prever valores econômico-financeiros, a certo nível de confiança predeterminado. Corroborando os objetivos pretendidos, esta análise tornou possível identificar as variáveis adequadas à análise de correlação e regressão para a tomada de decisão dentro das entidades, pelos dirigentes e também pelo cientista contábil.

Palavras-chave: Correlação e Regressão. Contabilometria. Construção Civil.



1 INTRODUÇÃO

A presente pesquisa abordará uma análise de correlação e regressão linear simples em alguns indicadores econômico-financeiros de 2009 das empresas de construção civil de capital aberto integrantes do *Novo Mercado*¹.

Segundo Melo (2012, p.8) “Os indicadores econômico-financeiros servem como identificadores que demonstram a potencialidade da empresa”, desta afirmativa não restam dúvidas de quão importante é uma análise mais aprofundada desses indicadores, visando como estes se comportam dependentemente de outras variáveis. Melo (2012, p.8) complementa que “a estatística tem por objetivo a coleta, análise e interpretação de dados numéricos relacionados a um grupo específico, fazendo então previsões com base nesses dados” e é por meio dela que se pode transformá-los em informações úteis para os empreendedores, visando amplificar seus negócios, ou até mesmo extinguir os problemas deste.

A análise de regressão é uma técnica estatística para investigar e modelar a relação entre variáveis, sendo uma das mais utilizadas na análise de dados. Um dos objetivos da análise de regressão é estimar os parâmetros desconhecidos do modelo.

Desta forma propende esta análise solucionar a seguinte questão: **Existe correlação entre alguns indicadores econômico-financeiros de 2009 das empresas de construção civil de capital aberto integrantes do Novo Mercado e qual o modelo de regressão apropriado para realizar previsões?**

Tomando a questão de pesquisa identificada, têm-se os seguintes objetivos a serem trabalhados nesta pesquisa:

Objetivos Gerais:

- Identificar e quantificar a correlação entre os indicadores;
- Encontrar o modelo mais apropriado de regressão para realizar previsões.

Objetivos Específicos:

- Determinar o Coeficiente de Correlação Linear de Pearson;
- Estimação de Mínimos Quadrados para o MRLS;
- Análise de Variância na Regressão;
- Determinar o Coeficiente de Determinação;
- Testar a Falta de Ajuste do MRLS;
- Determinar Intervalos de Confiança;
- Determinar Intervalos de Predição;
- Fazer a Análise de Resíduos.

Algumas hipóteses podem ser elaboradas a fim de esboçar algumas alternativas da solução da questão de pesquisa, são elas:

- Existe uma forte correlação entre as variáveis analisadas;
- O modelo de regressão apropriado é dado pela regressão linear simples.

A análise dos dados tomou por base o ano-calendário de 2009, ano seguinte ao desencadeamento da crise financeira mundial, visto que a construção civil esteve no centro da crise norte-americana que eclodiu no fim de 2008. Foram selecionadas apenas as empresas de capital aberto integrantes do Novo Mercado, que segundo definição da BM&FBOVESPA, é

¹ Segmento especial de negociação de valores mobiliários da BM&FBOVESPA (Bolsa de Valores de São Paulo).

um segmento de listagem destinado à negociação de ações emitidas por companhias que se comprometam, voluntariamente, com a adoção de práticas de *Governança Corporativa*² adicionais em relação ao que é exigido pela legislação. Diversas regulamentações e leis estão sendo criadas e modificadas ao redor do mundo para assegurar as boas práticas de governança corporativa, a mais conhecida mundialmente é a *Lei Sarbanes-Oxley (SOX)*, criada pela SEC (*Securities and Exchange Commission*)³, aprovada no congresso norte-americano e assinada pelo presidente George W. Bush em julho de 2002, esta lei surgiu após inúmeros escândalos contábeis (fraudes) em empresas norte-americanas (Enron, WorldCom, etc.). Esta lei procura restabelecer a ética dos negócios, principalmente no que se refere a preparação das **informações econômico-financeiras**, reforça também os procedimentos de controle interno (SILVA, NASCIMENTO e OTT, 2005). Nesta lei é definido, entre outras definições, que os executivos deverão declarar expressamente que são responsáveis pelo **planejamento**, estabelecimento e manutenção dos controles internos (SILVA, NASCIMENTO e OTT, 2005).

Segundo a Câmara Brasileira da Indústria da Construção, a construção civil constitui um importante setor para a economia nacional e é responsável direta pela parcela significativa do Produto Interno Bruto – PIB (*apud* DUART e LAMOUNIER, 2007, p.12), desta forma é imprescindível uma análise mais estruturada deste seguimento, já que ele será o principal pilar para o país alcançar o desenvolvimento econômico sustentável, dada a sua capacidade de gerar, de forma rápida, emprego e renda (DUART e LAMOUNIER, 2007, p.10-12).

Sendo assim, o motivo pela preferência da análise estatística dos dados das empresas de construção civil de capital aberto integrantes do “Novo Mercado” é pela inovação idealista que essa categoria traz consigo, visando sempre contribuir para o desenvolvimento da entidade e da sociedade, mantendo sua perpetuação.

O estudo possui três seções. Na seção seguinte é apresentada a análise de correlação e de regressão linear simples e na terceira seção são apresentadas as conclusões.

2 CONTABILOMETRIA APLICADA: ANÁLISE DE CORRELAÇÃO E REGRESSÃO LINEAR SIMPLES

2.1 Fundamentação Teórica

No campo das *Ciências Contábeis*, muitas vezes, o *Cientista Contábil* não consegue tomar decisões usando apenas a análise qualitativa, visto a complexidade dos fatos e as repercutidas mudanças na legislação tributária e contábil, alinhadas ainda a administração competitiva do negócio. Em muitos casos, há a necessidade do profissional ter um olhar mais minucioso e extremamente detalhista para emitir o seu parecer sobre as Demonstrações Contábeis. Diante desta dificuldade, este campo de conhecimento torna-se um local adequado à aplicação de uma metodologia científica, a *Contabilometria*, com o intuito de otimizar a opinião a ser tomada pelo *Cientista Contábil*.

A *Contabilometria* é uma linha de pesquisa científica das *Ciências Contábeis* que busca a aplicação de métodos quantitativos, integrando as Ciências Matemáticas, Estatísticas,

² Segundo o Instituto Brasileiro de Governança Corporativa (IBGC), é um sistema pelo qual as sociedades são dirigidas e monitoradas, envolvendo os acionistas e os cotistas, Conselho de Administração, Diretoria, Auditoria Independente e Conselho Fiscal. As boas práticas de governança corporativa têm a finalidade de aumentar o valor da sociedade, facilitar seu acesso ao capital e contribuir para a sua perenidade.

³ A SEC é o órgão responsável por regular o mercado de capitais dos Estados Unidos da América, equivale à Comissão de Valores Mobiliários (CVM).

Atuárias e de Informática aos conceitos de mensuração contábil e aos instrumentos de gerenciamento existentes, proporcionando informações para os processos de planejamento, controle e tomada de decisão na gestão das entidades.

Algumas características que uma equação contabilométrica deve possuir são ressaltadas por Marion e Silva (1986), os autores enumeram algumas características denominando-as de propriedades, são elas: relevância, simplicidade, capacidade explicativa, capacidade preditiva e plausibilidade teórica. Nossa e Garcia (*apud* MATSUMOTO, PEREIRA e NASCIMENTO, 2004, p.6-7) descrevem alguns instrumentos matemáticos utilizados em Contabilometria: probabilidade, cálculo matricial e programação linear. Corrar e Theóphilo (*apud* MATSUMOTO, PEREIRA e NASCIMENTO, 2004, p.7) também apresentam as principais técnicas quantitativas, bem como suas aplicações a serem utilizadas nas tomadas de decisões administrativas e contábeis, são elas: amostragem, **análise de regressão** e Programação Multiobjetiva ou *Goal Programming*.

Segundo Campos (2010, p.1), no final do século XX e início do século XXI, escândalos envolvendo empresas renomadas mundialmente evidenciaram a importância da solvência para *prevenção* de falência. Entre as organizações envolvidas podem ser destacadas o caso da Enron (2001), da Tyco (2002), da Parmalat (2003) e da WorldCom (2004), neste contexto, os modelos de previsão de falências foram desenvolvidos com a finalidade de *prever* falência ou insolvência. Entre os modelos usados pelas empresas para analisar se uma entidade pode correr o risco de falir, destacam-se o modelo de Kanitz (1978), que elaborou uma equação matemática utilizando a técnica de regressão múltipla e análise discriminante e Elizabethsky (1976) que desenvolveu um modelo matemático baseado na análise discriminante, entre muitos outros modelos matemáticos e estatísticos (Altman, Silva, Fitz Patrick, Winakor e Smith, Merwin, Tamari, Beaver, Backer e Gosman, Letícia Topa, John Moore) (MÁRIO, 2002, p.37-116).

2.2 Metodologia e Sistemática Científica

O presente trabalho contempla uma pesquisa de natureza descritiva, pois ela permite modelar as principais características de uma determinada população ou fenômeno e também o estabelecimento de relação entre variáveis. Baseando-se na técnica de coleta de dados primários, foram levantadas as informações econômico-financeiras de 17 empresas listadas na BM&FBOVESPA, extraídos do artigo elaborado por Melo (2010). As variáveis estudadas são quantitativas contínuas, logo se utilizou do método quantitativo. A coleta dos dados foi feita de forma não aleatória, utilizando-se da amostragem semiprobabilística por cotas, visando captar uma amostra homogênea. A coleta de dados englobou a pesquisa bibliográfica e documental.

Os cálculos e as análises estatísticas foram desenvolvidos com ajuda do *software* multifuncional Excel, um programa da *Microsoft Corporation*⁴. Elaborou-se ainda a disposição dos dados graficamente, pois segundo Barros e Lehfeld (2007, p. 110), auxilia a interpretação da análise e facilita o processo de inter-relação deles e também com as hipóteses de estudo.

⁴ Empresa multinacional de tecnologia informática dos Estados Unidos da América, que desenvolve e fabrica licenças e suporta uma ampla gama de produtos *software* para dispositivos de computador.

2.3 Apresentação dos Dados

No quadro 1 apresentam-se os indicadores econômico-financeiros das empresas estudadas elaborados por Melo (2010, p.8).

Quadro 1 – Quadro de Indicadores Econômico-Financeiros de 2009.

EMPRESAS	LG	LC	LS	LI	CT/PT (%)	PC/CT (%)	CT/PL (%)	PL/CT (%)	GIRO (%)	MG (%)	ROI (%)	ROE (%)
BROOKFIELD	1,65	3,59	1,28	0,42	57,34	37,12	134,43	74,38	33,15	11,14	3,69	8,66
CAMARGO	1,36	6,54	2,98	1,17	68,91	18,28	221,66	45,11	23,75	11,28	2,68	8,62
CR2	1,97	2,28	1,16	0,15	49,94	61,83	99,78	100,21	39,53	5,03	1,98	3,97
CYRELA	1,53	2,10	1,22	0,41	63,65	57,06	175,16	57,08	38,85	17,84	6,93	19,07
DIRECIONAL	2,48	5,77	4,01	1,84	39,69	39,40	65,81	151,93	34,69	20,94	7,26	12,04
EVEN	1,63	4,34	2,99	0,64	61,06	35,08	156,84	63,75	51,43	10,65	5,47	14,07
EZ TEC	5,09	5,88	3,20	0,98	19,47	59,37	24,18	413,54	44,68	32,20	14,38	17,86
GAFISA	1,33	2,62	1,76	0,70	72,51	37,67	263,88	37,89	40,87	7,06	2,88	10,50
HELBOR	1,51	2,13	1,15	0,24	63,79	53,43	176,21	56,74	56,35	12,00	6,76	18,68
INPAR	1,64	3,78	1,19	0,25	59,04	38,49	144,19	69,35	25,17	2,65	0,66	1,63
JHSF	1,32	2,32	1,60	0,86	49,78	51,01	99,15	100,84	27,49	30,04	8,26	16,45
MRV	2,18	2,98	1,86	0,60	45,11	60,75	82,18	121,67	37,80	21,08	7,97	14,52
PDG	1,80	2,52	1,52	0,65	52,04	53,37	108,52	92,14	32,65	17,04	5,56	11,60
RODOBENS	1,67	1,85	1,23	0,16	57,56	78,40	135,66	73,71	31,90	5,95	1,89	4,47
ROSSI	2,02	2,36	1,75	0,77	49,11	59,49	96,53	103,58	36,34	13,87	5,04	9,90
TECNISA	1,60	3,02	1,82	0,71	61,00	43,33	156,46	63,91	31,69	15,55	4,92	12,64
TRISUL	1,71	3,88	2,77	0,56	57,32	39,53	134,34	74,43	51,70	9,19	4,75	11,14

Fonte: Adaptado de Melo (2010, p.8)

2.4 Análise de Correlação e Regressão Linear Simples

• Coeficiente de Correlação Linear de Pearson

Antes de estimar o *modelo de regressão linear simples* (MRLS) é preciso levar em conta outros fatores determinísticos que influenciarão nos parâmetros do modelo. O *Coeficiente de Correlação de Pearson* ($\rho(X; Y)$) tem por objetivo medir o grau de associação entre duas variáveis, ele é obtido por:

$$\rho(\hat{X}; Y) = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sqrt{\left[\sum x^2 - \frac{(\sum x)^2}{n}\right] \left[\sum y^2 - \frac{(\sum y)^2}{n}\right]}} = \frac{S_{XY}}{\sqrt{S_{XX}S_{YY}}}$$

$\rho(\hat{X}; Y)$ é um número adimensional cujo valor se situa entre -1 e 1. Quando X e Y variam no mesmo sentido a correlação é positiva, logo $\rho > 0$, se X e Y variam em sentido contrário a correlação é negativa, logo $\rho < 0$. Se $\rho = 1$, a correlação é positiva perfeita, Se $\rho = -1$, a correlação é negativa perfeita, Se $\rho = 0$, a correlação é nula.

Na tabela 1 se apresentam os *coeficientes de correlação linear de Pearson* amostrais entre todas as variáveis deste estudo:

Tabela 1 – Coeficientes de Correlação Linear de Pearson.

Correlação	LG	LC	LS	LI	CT/PT	PC/CT	CT/PL	PL/CT	GIRO	MG	ROI	ROE
LG	1											
LC	0,4612	1										
LS	0,4545	0,8475	1									
LI	0,3075	0,6964	0,8359	1								
CT/PT	-0,8829	-0,3122	-0,3917	-0,3570	1							
PC/CT	0,2815	-0,6188	-0,4417	-0,4259	-0,4079	1						
CT/PL	-0,6915	-0,1438	-0,2501	-0,2289	0,9353	-0,4798	1					
PL/CT	0,9837	0,4365	0,4359	0,3353	-0,9136	0,2915	-0,7287	1				
GIRO	0,2015	-0,0985	0,1500	-0,1978	-0,0536	0,1318	-0,0031	0,1404	1			
MG	0,5673	0,2326	0,3665	0,5355	-0,6781	0,1719	-0,5799	0,6656	-0,0206	1		
ROI	0,7307	0,2593	0,4154	0,4109	-0,7230	0,2307	-0,5812	0,7807	0,3352	0,9127	1	
ROE	0,2576	0,0133	0,2216	0,2645	-0,2130	0,0675	-0,1165	0,3011	0,4773	0,7286	0,8142	1

Faixas: de 0 (zero) a | 0,5 |: vermelho (fraco); de | 0,5001 | a | 0,7 |: amarelo (razoável); de | 0,7001 | a | 1 |: verde (forte).

Fonte: Melo (2012).

Pela definição do seguimento de construção civil e pelo fato das empresas serem de capital aberto, o interesse deste estudo será estimar o MRLS (Modelo de Regressão Linear Simples) para os índices de rentabilidade, visto os olhos dos investidores e acionistas sobre as respectivas empresas analisadas, mas correlacionando-os com índices patrimoniais (Ativo e Passivo), visto que os índices de rentabilidade possuem praticamente a mesma fonte de derivação, logo não haveria sentido efetuar uma relação desta neste estudo.

A variável resposta selecionada da pesquisa pela sua significância econômico-financeira será o ROI, em contrapartida com o PL/CT, visto que ambos apresentaram a melhor correlação entre índices de resultado e patrimoniais, respectivamente ($\rho = 0,7807$).

• O Modelo de Regressão Linear Simples (MRLS)

Segundo Kazmier (1982, p.299) o objetivo principal da análise de regressão linear é prever o valor de uma variável (a *variável dependente*), dado que seja conhecido o valor de uma variável associada (a *variável independente*). A *equação de regressão* é a fórmula algébrica pela qual se determina o valor previsto da variável dependente.

Seja Y uma variável aleatória de interesse (variável resposta), e seja X uma variável aleatória (variável regressora). O *Modelo de Regressão Linear Simples* descreve a variável Y como uma soma de uma quantidade determinística e uma quantidade aleatória. A parte determinística, uma reta em função de X , representa a informação sobre Y que já se “espera”, apenas com o conhecimento da variável X . A parte aleatória determinada **erro**, representa os inúmeros fatores que, conjuntamente, podem interferir em Y (CHARNET *et al*, 2008, p.11).

O *Modelo de Regressão Linear Simples Amostral* (MRLS) é dado por:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

$$\beta_0, \beta_1 \text{ e } x_i: \text{constantes}$$

$$\text{Cov}[\varepsilon_i; \varepsilon_j] = 0, \quad i \neq j, \quad i, j = 1, 2, \dots, n,$$

em que, n é o número de indivíduos, y_i é a observação da variável dependente (ou resposta) para o i -ésimo indivíduo, $x_i = (x_1, x_2, \dots, x_n)$ é a observação da variável independente (ou regressora) para o i -ésimo indivíduo, β_0 e β_1 são coeficientes de regressão (parâmetros) e ε_i é um componente de erro aleatório. Assume-se que esses erros são independentes e seguem distribuição normal com média zero e variância desconhecida, denotado por: $\varepsilon \sim N(0; \sigma^2)$. A distribuição de probabilidade de Y corresponde ao valor prefixado de X , e é dada por: $Y \sim N(\beta_0 + \beta_1 x; \sigma^2)$.

- **O Método dos Mínimos Quadrados**

Considerando uma reta para representar um conjunto de n pontos $(x_1; y_1), (x_2; y_2), \dots, (x_n; y_n)$, não necessariamente colineares, apresenta-se o *Método dos Mínimos Quadrados* desenvolvida por Carl Gauss (1777-1855)⁵. Para cada reta candidata, o método dos mínimos quadrados analisa as n diferenças entre cada valor y e o valor na reta, correspondente ao respectivo valor x . A reta selecionada é a reta que apresenta a menor soma de quadrados de tais diferenças. Seja $y = a + bx$ a representação de uma reta genérica onde a e b são valores reais. Então se trata de encontrar o mínimo da seguinte função, em a e b , onde $a \in \mathbb{R}$ e $b \in \mathbb{R}$ (CHARNET *et al*, 2008, p.30-31):

$$\sum_{i=1}^n [y_i - (\beta_0 + \beta_1 x_i)]^2$$

Para encontrar esse mínimo é necessário realizar uma série de cálculos matemáticos avançados, cuja notação final é dada por:

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$$

onde:

$$\begin{aligned} \Rightarrow \hat{\beta}_0 &= \frac{\sum y_i - \hat{\beta}_1 \sum x_i}{n} \\ \Rightarrow \hat{\beta}_1 &= \frac{\sum x_i y_i - \frac{\sum x_i \sum y_i}{n}}{\sum x^2 - \frac{(\sum x)^2}{n}} = \frac{S_{XY}}{S_{XX}} \end{aligned}$$

$\hat{\beta}_0$ e $\hat{\beta}_1$, são estimadores de mínimos quadrados do intercepto e inclinação, respectivamente. A expressão dá uma estimativa pontual da média de Y para cada valor de X , as quantidades $\varepsilon_i = y_i - \hat{y}_i$, para $i = 1, 2, \dots, n$, são denominadas **resíduos**, que serão tratadas em um tópico específico.

- **Estimação de Mínimos Quadrados para o MRLS**

A estimação dos parâmetros do modelo de regressão linear simples amostral pelo método de mínimos quadrados calculado, pela análise de regressão entre o ROI e o PL/CT, obteve como resultado os seguintes parâmetros:

$$\hat{\beta}_0 = 0,0241634$$

$$\hat{\beta}_1 = 0,0294086$$

Logo, o modelo de regressão linear simples amostral (**ROI versus PL/CT**) desta pesquisa será dado pela seguinte equação:

$$y_i = 0,0241634 + 0,0294086x_i$$

⁵ Johann Carl Friedrich Gauss (✱ Braunschweig, 30 de Abril de 1777, † Göttingen, 23 de Fevereiro de 1855), foi um matemático, astrônomo e físico alemão que contribuiu muito em diversas áreas da ciência, dentre elas a teoria dos números, estatística, análise matemática, geometria, diferencial, geodésia, geofísica, eletroestática, astronomia e óptica.

- **Análise de Variância na Regressão**

A determinação da equação de regressão deve ser precedida de uma análise de variância, denominada ANOVA, a fim de comprovar estatisticamente se os dados apresentam a suposta relação linear entre as variáveis X e Y . As hipóteses a serem testadas pela análise de variância na regressão são:

$$H_0: \widehat{\beta}_1 = 0 \Leftrightarrow \text{não existe a regressão linear};$$

$$H_1: \widehat{\beta}_1 \neq 0 \Leftrightarrow \text{existe a regressão linear}.$$

Apresentam-se primeiramente as seguintes equações:

$$SQTotal = SQReg + SQRes = S_{YY}, \text{ onde } S_{YY} = \sum y^2 - \frac{(\sum y)^2}{n}$$

$$SQReg = \widehat{\beta}_1 S_{XY}$$

$$SQRes = \sum \varepsilon_i = (y_i - \widehat{y}_i)^2$$

em que $SQTotal$ é a soma de quadrados total, que é a soma de quadrados da regressão ($SQReg$) mais a soma de quadrados do resíduo ($SQRes$). Ao realizar a análise de variância, o procedimento é comparar as variâncias:

$$\sum_{i=1}^n (y_i - \bar{y}_i)^2 = \sum_{i=1}^n (\widehat{y}_i - \bar{y}_i)^2 + \sum_{i=1}^n (y_i - \widehat{y}_i)^2 = SQTotal = SQReg + SQRes$$

$(y_i - \bar{y}_i)^2$: $SQTotal$ é a variação total de Y em torno da média;

$(\widehat{y}_i - \bar{y}_i)^2$: $SQReg$ é a variação das esperanças específicas de Y em torno da média;

$(y_i - \widehat{y}_i)^2$: $SQRes$ é a variação de Y em torno da reta.

A variância residual da amostra é definida por:

$$QMRes = \frac{SQRes}{n - 2}$$

em que $QMRes$ é o quadrado médio do resíduo. A $SQRes$ tem $(n - 2)$ graus de liberdade (gl), pois dois graus de liberdade são associados com as estimativas $\widehat{\beta}_0$ e $\widehat{\beta}_1$. Ainda tem-se:

$$QMReg = \frac{SQReg}{1}$$

em que $QMReg$ é o quadrado médio da regressão, com 1 grau de liberdade.

A seguir a tabela da análise de variância na regressão (ANOVA) pode ser definida conforme tabela 2:

Tabela 2 – ANOVA: Análise de Variância na Regressão.

<i>Fonte da variação</i>	<i>gl</i>	<i>SQ</i>	<i>QM</i>	<i>F₀</i>
<i>Regressão</i>	1	0,010178646	0,010178646	23,40706432
<i>Resíduo</i>	15	0,006522804	0,000434854	
Total	16	0,01670145		

Fonte: Melo (2012).

A estatística teste tem distribuição *F de Snedecor* com 1 e $(n - 2)$ graus de liberdade e com um nível de significância α , logo $F_{1, (n-2), \frac{\alpha}{2}} = 4,543077$, com $n = 17$ e $\alpha = 0,05$, portanto rejeita-se H_0 , pois $F_0 > F_{1, (n-2), \frac{\alpha}{2}}$. Alternativamente, como *p-valor*⁶ neste caso é igual a 0,0002171, rejeita-se H_0 ao nível de significância (α) de 5%, pois *p-valor* < 0,05. Conclui-se que $\widehat{\beta}_1 \neq 0$, logo existe a regressão entre X e Y .

- **Coefficiente de Determinação**

O *Coefficiente de Determinação*, R^2 , é interpretado como a proporção da variabilidade dos Y 's observados, explicada pelo modelo considerado. O valor de R^2 pertence ao intervalo $[0; 1]$ e quando mais próximo de 1 melhor o ajuste do modelo. Ele é dado por:

$$R^2 = \rho^2 = \frac{SQReg}{SQTotal} = 1 - \frac{SQRes}{SQTotal}$$

O R^2 do modelo de regressão apresentado foi de 0,609446849, isto é, 60,94% da regressão é explicada pelo modelo $y_i = 0,0241634 + 0,0294086x_i$.

- **Teste de Falta de Ajuste do MRLS**

Para testar estatisticamente a falta de ajuste do MRLS é necessário replicações, ou seja, deve ter pelo menos dois valores da variável resposta para alguns valores da variável regressora. Nesta situação tem-se um estimador natural para a $Var[Y|x]$, σ^2 , que não está baseado no MRLS. Um estimador para a $Var[Y|x]$ seria dado pela variância amostral desta amostra:

$$\frac{1}{n_i - 1} \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2,$$

sendo \bar{y}_i a média amostral de todos os valores observados para $X = x_i$. Sempre é conveniente combinar as k variáveis amostrais, utilizando assim toda a amostra. Denotando este estimador por $\widehat{\sigma}_{ep}^2$, no qual o subscrito *ep* significa “erro puro”, por depender da variação pura e não do MRLS.

$$\widehat{\sigma}_{ep}^2 = \frac{\sum_{i=1}^k \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2}{\sum_{i=1}^k (n_i - 1)} = \frac{\sum_{i=1}^k \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2}{n - k}$$

Usando esta notação, a equação do MRLS amostral é expressa da seguinte forma:

$$y_{ij} = \beta_0 + \beta_1 x_i + \varepsilon_{ij}, \quad i = 1, \dots, k, \quad j = 1, \dots, n_i$$

a soma de quadrados de resíduos é dada por:

$$SQE = \sum_{i=1}^k \sum_{j=1}^{n_i} (y_{ij} - \hat{y}_i)^2$$

podendo ser particionada em duas outras somas de quadrados:

$$\sum_{i=1}^k \sum_{j=1}^{n_i} (y_{ij} - \hat{y}_i)^2 = \sum_{i=1}^k \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2 + \sum_{i=1}^k n_i (\hat{y}_i - \bar{y}_i)^2$$

⁶ *p-valor* é a probabilidade de uma variável aleatória com distribuição *F*, com 1 e $(n-2)$ graus de liberdade, ser maior do que o valor observado F_0 .

$$\Rightarrow SQE = SQE_p + SQF_a$$

$(y_{ij} - \hat{y}_i)^2$: SQE é a variação em torno da reta;

$(y_{ij} - \bar{y}_i)^2$: SQE_p é a variação de Y , para X fixo, independente do modelo;

$(\hat{y}_i - \bar{y}_i)^2$: SQF_a é a falta de ajuste do modelo.

Valores relativamente altos de SQF_a indicam evidência contra a adequação do MRLS e pior será o ajuste do modelo, visto que se o ajuste fosse perfeito tem-se que $\hat{y}_i = \bar{y}_i$, portanto a SQF_a seria igual a zero. Considerando estes argumentos, as seguintes hipóteses são construídas:

H_0 : O MRLS é adequado;

H_1 : O MRLS não é adequado.

Para as replicações foram levados em consideração os índices (PL/CT) que não diferiram entre si significativamente (escalas de um centésimo), dos quais foram selecionados apenas 4 escalas de x_i , conforme disposto no quadro 2.

Quadro 2 – Soma de Quadrados do Erro Puro.

EMPRESAS	x_i	Replicações y_{ij}	$\sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2$
HELBOR	0,57	0,0676	0,000001445
CYRELA		0,0693	
EVEN	0,64	0,0547	0,000015125
TECNISA		0,0492	
RODOBENS	0,74	0,0189	0,000418107
BROOKFIELD		0,0369	
TRISUL	1,00	0,0475	0,001971920
CR2		0,0198	
JHSF		0,0826	
SQE_p			0,0024065967

Fonte: Melo (2012).

Em seguida apresenta-se a tabela ANOVA para o teste de falta de ajuste do MRLS:

Tabela 3 – ANOVA: Análise de Variância para Testar a Falta de Ajuste do MRLS.

Fonte da variação	gl	SQ	QM	F_0
Regressão	1	0,010178646	0,010178646	23,40706432
Resíduo	15	0,006522804	0,000434854	
Falta de Ajuste	10	(0,004116207)	0,000411621	0,855192415
Erro Puro	5	(0,002406597)	0,000481319	
Total	16	0,01670145		

Fonte: Melo (2012).

Existem 10 graus de liberdade para a falta de ajuste porque o número k de valores distintos (significativamente) de X é igual a 12. A estatística teste tem distribuição F de *Snedecor* com $(k - 2)$ e $(n - k)$ graus de liberdade (Teorema de Cochran) e com um nível de significância α , logo $F_{1[(k-2),(n-k);\alpha]} = 4,735063$, com $n = 17$ e $\alpha = 0,05$, portanto não descarta-se H_0 , pois $F_0 < F_{1[(k-2),(n-k);\alpha]}$. Concluindo que o MRLS é adequado.

- **Intervalos de Confiança**

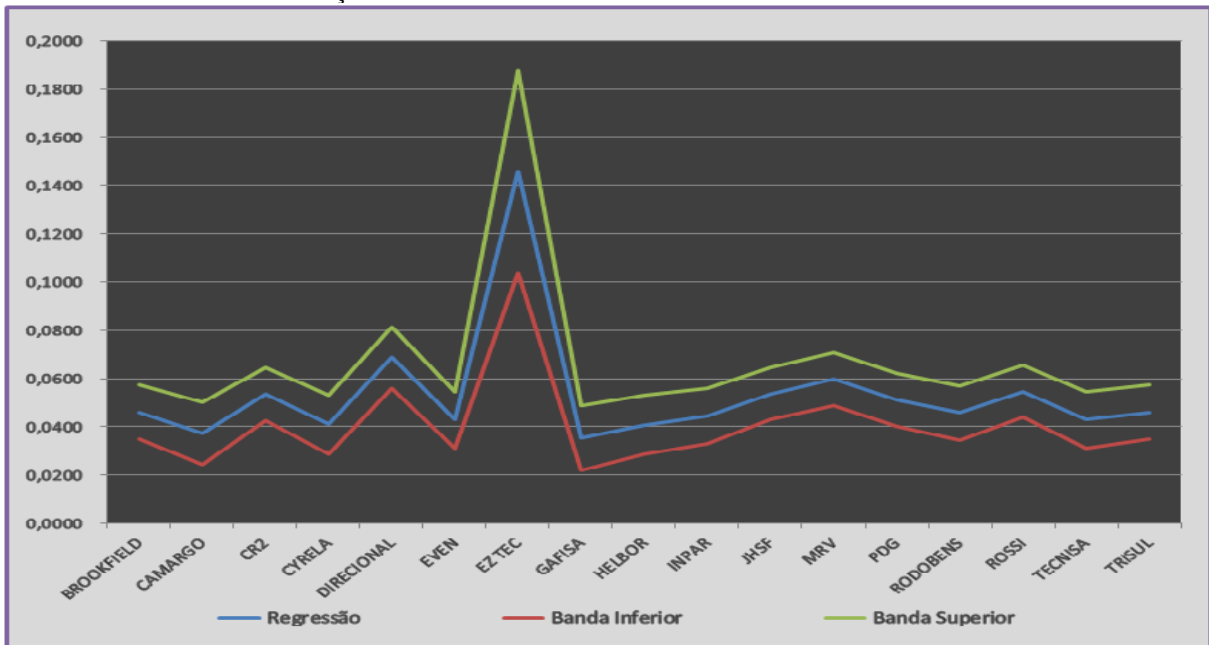
O intervalo de confiança fornece informações sobre a precisão das estimativas, no sentido de que quanto menor a amplitude do intervalo maior a precisão. Calculando intervalos de confiança para alguns valores de x pode-se esboçar uma região em torno da reta estimada, indicando os limites superiores e inferiores desses intervalos, chamadas de **bandas de confiança** (CHARNET *et al*, 2008, p.101).

O intervalo de confiança para a $E(Y|x)$ é dada pela seguinte expressão:

$$\widehat{\beta}_0 + \widehat{\beta}_1 x_i \pm t_{[\frac{\alpha}{2}; (n-2)]} \sqrt{\widehat{\sigma}^2 \left[\frac{1}{n} + \frac{(x - \bar{x})^2}{S_{XX}} \right]}$$

Portanto o intervalo de confiança de 95% para o **ROI** dado o **PL/CT** está disposto conforme gráfico de bandas de confiança abaixo:

Gráfico 1 – Bandas de Confiança de 95% do ROI.



Fonte: Melo (2012).

- **Intervalos de Predição**

O interesse dos intervalos de predição é fazer uma predição para um valor futuro de Y_0 – não observado – correspondente a $X = x_0$. Usando o MRLS ajustado, diz-se que $\widehat{Y}_0 = \widehat{\beta}_0 + \widehat{\beta}_1 x_0$ é uma predição de Y_0 . A predição de Y_0 é igual ao estimador do valor esperado de Y , para $X = x_0$. O erro de predição é definido por $Y_0 - \widehat{Y}_0$, sendo que $Y_0 \sim N(\beta_0 + \beta_1 x_0; \sigma^2)$ (CHARNET *et al*, 2008, p.103-104).

O intervalo de predição $(1 - \alpha)100\%$ de Y , para $X = x_0$ é definido por:

$$\widehat{Y}_0 \pm t_{[\frac{\alpha}{2}; (n-2)]} \sqrt{\widehat{\sigma}^2 \left[1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{XX}} \right]}$$

Portanto o intervalo de predição de 95% de **ROI**, para o **PL/CT** está disposto conforme quadro 3:

Quadro 3 - Intervalos de Predição de Y , para $X = x_0$.

EMPRESAS	x_i	y_i	$\hat{Y}_0 = \hat{\beta}_0 + \hat{\beta}_1 x_0$	$E(Y_0 x_0)$ Inferior	$E(Y_0 x_0)$ Superior
BROOKFIELD	0,7438	0,0369	0,0460	0,0001907	0,0918843
CAMARGO	0,4511	0,0268	0,0374	-0,0088466	0,0837058
CR2	1,0021	0,0198	0,0536	0,0079073	0,0993601
CYRELA	0,5708	0,0693	0,0409	-0,0051136	0,0870132
DIRECIONAL	1,5193	0,0726	0,0688	0,0226256	0,1150621
EVEN	0,6375	0,0547	0,0429	-0,0030557	0,0888784
EZ TEC	4,1354	0,1438	0,1458	0,0846221	0,2069373
GAFISA	0,3789	0,0288	0,0353	-0,0111228	0,0817354
HELBOR	0,5674	0,0676	0,0408	-0,0052189	0,0869185
INPAR	0,6935	0,0066	0,0446	-0,0013403	0,0904568
JHSF	1,0084	0,0826	0,0538	0,0080925	0,0995455
MRV	1,2167	0,0797	0,0599	0,0141325	0,1057571
PDG	0,9214	0,0556	0,0513	0,0055227	0,0969982
RODOBENS	0,7371	0,0189	0,0458	-0,0000127	0,0916936
ROSSI	1,0358	0,0504	0,0546	0,0088961	0,1003535
TECNISA	0,6391	0,0492	0,0430	-0,0030065	0,0889233
TRISUL	0,7443	0,0475	0,0461	0,0002059	0,0918985

Fonte: Melo (2012).

Pela análise do quadro 3 percebe-se que algumas empresas podem até apresentar prejuízo (em vermelho) em um cenário desfavorável na economia.

• Análise de Resíduos

A análise de resíduos desempenha papel fundamental na avaliação do ajuste de um MRLS, investiga a adequação do modelo quanto às suposições básicas do modelo, bem como normalidade, independência dos erros, homocedasticidade⁷, relação linear de X e Y e falta de ajuste do modelo proposto. Além dos testes de significância e adequação, a análise de resíduo vem complementar o elenco de procedimentos que devem ser realizados após o ajuste do qualquer modelo (MESSETI, 2013, p.13-14).

Sob a suposição de normalidade dos erros no modelo $\varepsilon_i \sim N(0; [1 - h_{ii}])$, tem-se que Y_i tem distribuição normal, e visto que a covariância entre dois resíduos distintos é diferente de zero, conclui-se que os resíduos não são independentes. Com intuito de melhor analisar os resíduos levando em conta sua variabilidade, trabalha-se com transformações usuais, definidas a seguir (CHARNET *et al*, 2008, p.119).

Resíduos Padronizados: transformação utilizada para reduzir uma variável aleatória a ter esperança igual a zero e desvio padrão igual a um. No caso dos resíduos padronizados, o parâmetro σ^2 é substituído pelo seu estimador $\hat{\sigma}^2$. Sob a hipótese de o MRLS ser adequado, essa nova variável aleatória, z_i , tem distribuição aproximadamente *t de Student*. A distribuição não é exata porque as variáveis aleatórias ε_i e $\hat{\sigma}^2$ no numerador e denominador de z_i , respectivamente, não são independentes. Consequentemente z_i elevado indica a presença de *outliers*⁸. É definido pela expressão:

⁷ Homogeneidade de variâncias.

⁸ Pontos discrepantes, extremos.

$$z_i = \frac{\varepsilon_i}{\sqrt{\widehat{\sigma^2}(1 - h_{ii})}}, \quad i = 1, \dots, n.$$

Resíduos Estudentizados (Resíduos Jackknife): Neste caso, tem-se $\widehat{\sigma^2}_{(i)}$ como a soma de quadrados média dos resíduos do MRLS ajustado sem utilizar a i -ésima observação. Sob a hipótese de o MRLS ser adequado, essa nova variável aleatória, z_i^* , tem distribuição *t de Student* com $(n - 1 - 2)$ graus de liberdade. Aqui as variáveis aleatórias ε_i e $\widehat{\sigma^2}_{(i)}$ são independentes. Difere-se um pouco do resíduo padronizado no sentido de que está mais propício a detectar *pontos influentes*. É definido pela expressão:

$$z_i^* = \frac{\varepsilon_i}{\sqrt{\widehat{\sigma^2}_{(i)}(1 - h_{ii})}}, \quad i = 1, \dots, n.$$

No conjunto dados apresentado, tem-se um ajuste de reta de razoável a bom, pois $R^2 = 0,609446849$. Observa-se também, através do quadro 4 seguinte, que os resíduos padronizados se encontram na faixa $[-2; 2]$, o que vai, *a priori*, de acordo com a suposição de distribuição *t de Student*. Os resíduos padronizados e os estudentizados são bastante parecidos, exceto para a 10ª (INPAR) observação. O valor de $\widehat{\sigma^2}_{(10)} = 0,0003557$ indica que removendo-se a 10ª observação a dispersão em torno da reta ajustada é menor, logo o 10ª resíduo estudentizado de valor $-2,0834$ é mais alto que os demais, visto ainda que se situou fora da faixa $[-2; 2]$.

Através da análise de resíduos estudentizados foi possível perceber que em algumas situações existem observações **influentes**, que não são detectadas pela análise de resíduos padronizados, como no caso descrito acima.

Quadro 4 - Resíduos Padronizados e Estudentizados.

EMPRESAS	x_i	y_i	ε_i	z_i	z_i^*	$\widehat{\sigma^2}_{(i)}$	h_{ii}
BROOKFIELD	0,7438	0,0369	-0,0091	-0,4530	-0,4406	0,0004597	0,0644074
CAMARGO	0,4511	0,0268	-0,0106	-0,5327	-0,5196	0,0004571	0,0844381
CR2	1,0021	0,0198	-0,0338	-1,6724	-1,7912	0,0003791	0,0588239
CYRELA	0,5708	0,0693	0,0284	1,4132	1,4661	0,0004040	0,0744870
DIRECIONAL	1,5193	0,0726	0,0038	0,1880	0,1818	0,0004648	0,0817237
EVEN	0,6375	0,0547	0,0118	0,5862	0,5728	0,0004554	0,0699984
EZ TEC	4,1354	0,1438	-0,0020	-0,2917	-0,2826	0,0004631	0,8940463
GAFISA	0,3789	0,0288	-0,0065	-0,3274	-0,3174	0,0004627	0,0916176
HELBOR	0,5674	0,0676	0,0268	1,3336	1,3720	0,0004109	0,0747360
INPAR	0,6935	0,0066	-0,0380	-1,8843	-2,0834	0,0003557	0,0668136
JHSF	1,0084	0,0826	0,0288	1,4227	1,4774	0,0004032	0,0588293
MRV	1,2167	0,0797	0,0198	0,9786	0,9773	0,0004360	0,0628079
PDG	0,9214	0,0556	0,0043	0,2146	0,2076	0,0004644	0,0593505
RODOBENS	0,7371	0,0189	-0,0269	-1,3359	-1,3750	0,0004105	0,0647031
ROSSI	1,0358	0,0504	-0,0042	-0,2088	-0,2021	0,0004644	0,0589315
TECNISA	0,6391	0,0492	0,0062	0,3104	0,3007	0,0004631	0,0699000
TRISUL	0,7443	0,0475	0,0014	0,0718	0,0694	0,0004657	0,0643856

Fonte: Melo (2012).

Ainda é conveniente realizar um teste que verifique estatisticamente a suposição de normalidade. Dentre os testes mais utilizados destacam-se o de *Shapiro-Wilk* e o de *Kolmogorov*. Utilizando o teste de *Shapiro-Wilk*, definido abaixo:

$$W = \frac{b^2}{\sum (x_i - \bar{x})^2}$$

sendo b uma constante determinada pela fórmula:

$$\begin{cases} \sum_{i=1}^{\frac{n}{2}} a_{(n-i+1)} (x_{(n-i+1)} - x_{(i)}) & \text{se } n \text{ é par} \\ \sum_{i=1}^{\frac{n+1}{2}} a_{(n-i+1)} (x_{(n-i+1)} - x_{(i)}) & \text{se } n \text{ é ímpar} \end{cases}$$

em que $a_{(n-i+1)}$ são constantes geradas pelas médias, variâncias e covariâncias das estatísticas de ordem de uma amostra de tamanho n de uma distribuição Normal, e com as seguintes hipóteses:

H_0 : a amostra provém de uma população normal;

H_1 : a amostra não provém de uma população normal.

ao nível de significância α de 0,05, não rejeita-se a hipótese de que a amostra provém de uma população normal, pois tem-se que $W = 0,94796$, e o valor crítico da estatística W de *Shapiro-Wilk* é 0,981, isto é, $W < W_{(0,05;17)}$, com um nível de significância de 5%, logo não descarta-se a hipótese nula, cientificando-se, desta forma, a suposição de normalidade dos resíduos padronizados.

Algumas Transformações Usuais: Se após a análise dos resíduos é constatado a violação de uma ou mais suposições, algumas vezes é possível amenizar esse problema fazendo alguma transformação pertinente na variável resposta Y . Em seguida algumas transformações possíveis que são normalmente usadas com o objetivo de estabilizar as variâncias dos erros e, em circunstâncias especiais, também para normalizar os dados (CHARNET *et al*, 2008, p.132-133).

- **log(Y)** Transformação logarítmica na base e ;
- \sqrt{Y} Transformação raiz-quadrada;
- $\frac{1}{Y}$ Transformação recíproca;
- Y^2 Transformação quadrática;
- **arcsen** \sqrt{Y} Transformação arco-seno.

Como não foram constatadas evidências de violação de uma ou mais suposições pela análise gráfica e por meio de um dos testes estatísticos, aqui não se fez necessário a transformação dos dados.

3 CONSIDERAÇÕES FINAIS

Com o intuito de identificar e quantificar a correlação entre os indicadores deste estudo, conclui-se que o ROI e o PL/CT possuem uma forte correlação e servirão como variáveis para a tomada de decisão dentro da entidade. Logo, foi proposto o modelo de regressão linear simples para prever valores de resultado (ROI) e outros derivados da mesma equação para a tomada de decisões dentro das entidades do setor estudado.

Pela definição de Pearson foi possível determinar o Coeficiente de Correlação Linear e então quantificar como forte essa correlação. Através da estimação de Mínimos Quadrados, formulada por Carl Gauss, foi possível definir o modelo de regressão linear simples amostral deste estudo, de onde surgiram todas as demais análises seguintes.

Dentre estas análises, destaca-se a Análise de Variância na Regressão, onde foi plausível testar a hipótese da existência de regressão linear entre as variáveis do estudo, chegando à conclusão de que existe a regressão linear entre as variáveis ROI e PL/CT, pela tabela ANOVA, ao nível de significância de 5%.

Através do Coeficiente de Determinação explicou-se que uma boa proporção da variabilidade dos valores preditos de y_i é explicada pelo modelo de regressão linear simples amostral.

Por meio do Teste de Falta de Ajuste do MRLS, e pela tabela ANOVA, foi identificado de que o MRLS é adequado, ao nível de significância de 5%.

Quanto à precisão das estimativas foi possível, através de intervalos de confiança e ao nível de significância de 5%, definir valores mínimos e máximos para a variável resposta y_i , criando bandas de confiança em torno da reta de regressão.

Com interesse em fazer previsões para um valor futuro de Y_0 , ainda não observado correspondente a um $X = x_0$, foi possível produzir intervalos de predição, conforme disposto no quadro 4, logo, tem-se, também, valores mínimos e máximos para a esperança de Y_0 dado x_0 .

Através da Análise de Resíduos confirmaram-se todas as suposições básicas de um modelo, além dos testes de significâncias e adequação do modelo. Isto se deu por meio da análise de resíduos padronizados e estudatizados e pelo teste de *Shapiro-Wilk*.

Ao final do estudo constatou-se que não foi necessário utilizar-se das transformações usuais, visto que não foram constatadas quaisquer violações de uma ou mais suposições básicas de um modelo, um ótimo resultado para a pesquisa, pois a torna confiável para aplicações pelas empresas do seguimento na tomada de decisão.

Se finda este estudo com o propósito de novas análises no campo da Contabilometria, ciência ainda muito carente de estudos mais aprofundados e práticos indutivos, pois a única fonte fidedigna de conhecimento é o empirismo, que nasce da observação e da experimentação para decidir o que é verdadeiro. Suscitar novos temas de estudo e traçar objetivos será sempre uma meta contínua das Ciências Contábeis, para elucidar as temáticas e necessidades econômico-financeiras.

4 REFERÊNCIAS

BARROS, Aidil Jesus da Silveira; LEHFELD, Neide Aparecida de Souza. **Fundamentos de metodologia científica**. 3 ed. São Paulo: Pearson Prentice Hall, 2007.

BM&FBOVESPA, Bolsa de Valores de São Paulo. **Governança Corporativa**. Disponível em: <<http://www.bmfbovespa.com.br/cias-listadas/consultas/governanca-corporativa/governanca-corporativa.aspx?Idioma=pt-br>>. Acesso em: 14 ago. 2013.

CAMPOS, Isabelle Carlos. **Aplicação dos modelos de Elizabetsky e Kanitz na previsão de falência**: um estudo descritivo das melhores e maiores empresas por setor listadas na Revista Exame em 2010. V Seminário UFPE de Ciências Contábeis. 2010.

CHARNET, Reinaldo *et al.* **Análises de modelos de regressão linear**: com aplicações. 2 ed. Campinas: Unicamp, 2008.

KAZMIER, Leonard J. **Estatística aplicada à economia e a administração**. 1 ed. São Paulo: McGraw-Hill do Brasil, 1982.

MÁRIO, Poueri do Carmo. **Contribuição ao estudo da solvência empresarial**: uma análise de modelos de previsão – estudo exploratório aplicado em empresas mineiras. 2002. Dissertação (Mestrado em Controladoria e Contabilidade) – Universidade de São Paulo, São Paulo.

MARION, José Carlos; SILVA, Laércio Baptista. **Contabilometria**: novo campo de estudos para a contabilidade. Revista Brasileira de Contabilidade, Rio de Janeiro, ano 16, n. 59, p.34-41, outubro/dezembro, 1986.

MATARAZZO, Dante Carmine. **Análise financeira de balanços**: abordagem básica e gerencial. 6 ed. São Paulo: Atlas, 2008.

MATSUMOTO, Alberto Shigueru; PEREIRA, Sebastião Eustáquio; NASCIMENTO, Gilmar de Souza do. **A utilização da Contabilometria e a agregação de valor à informação contábil**. 3º Congresso USP Iniciação Científica em Contabilidade 27 e 28 de julho de 2006. Disponível em: <<http://www.congressousp.fipecafi.org/artigos32006/255.pdf>>. Acesso em: 26 abr. 2013.

MELO, Anderson Martins de. **Índices-padrão de indicadores econômico-financeiros das empresas de capital aberto do seguimento de construção civil integrantes do Novo Mercado**. 2012. Monografia (Especialização em Contabilidade e Controladoria Empresarial) – Universidade Estadual de Londrina, Londrina.

MESSETI, Ana Verginia Libos. **Análise de correlação e regressão**. 2013. Curso de especialização “*Lato Sensu*” em estatística. Universidade Estadual de Londrina, Londrina.

SILVA, Letícia Medeiros; NASCIMENTO, Auster Moreira; OTT, Ernani. **A influência da Lei Sarbanes-Osley e do Código Civil brasileiro nos controles internos de empresas localizadas no Brasil**. 2005. Universidade do Vale do Rio dos Sinos.