FUCAPE FUNDAÇÃO DE PESQUISA E ENSINO

Elberth Tiago Lopes

A REGRESSÃO LINEAR MÚLTIPLA COMO FERRAMENTA DA CONTROLADORIA

VITÓRIA

2021

ELBERTH TIAGO LOPES

A REGRESSÃO LINEAR MÚLTIPLA COMO FERRAMENTA DA CONTROLADORIA

Artigo tutorial apresentado à Fucape Fundação de Pesquisa e Ensino, como requisito parcial para aprovação na disciplina Métodos Quantitativos Aplicado.

Professor: Adonai José Lacruz

VITÓRIA

2021

A regressão linear múltipla como ferramenta da Controladoria

Resumo

O presente trabalho possui como objetivo demonstrar a viabilidade do uso de instrumentos estatísticos e matemáticos baseados em métodos quantitativos, em especial Regressão Linear, através de suas técnicas e ferramentas no auxilio das atribuições da Controladoria nas organizações, como levanamento de dados, criação de relatórios, estabelecimento de modelos, projeção de cenários, correlação de variáveis e indicadores, assegurando a otimização do resultado econômico da organização na busca pela eficácia institucional, de forma a possibilitar o crescimento dos ganhos corporativos, por meios de decisões assertivas, sempre buscando entender as interações do passado para ser mais assertivo nas decisões do presente em relação aos objetivos futuros.

***Palavras-chave**: Controladoria. Métodos Quantitativos. Regressão Linear Múltipla.*

1. Introdução

Com a crescente necessidade das empresas por informações, controles e projeções que busquem o aumento da produtividade e competitividade, motivada amplamente pela concorrência globalizada que destruiu as fronteiras produtivas, aliada ao contínuo desenvolvimento tecnológico que gera mudanças em velocidades cada vez mais aceleradas, as empresas buscam constantemente refinar seus controles com usos de novas ferramentas e técnicas que contribuam na tomada de decisões em suas variadas camadas hierárquicas.

Com a vocação de satisfazer esta necessidade de aprimoramento surgiu a Controladoria que na visão de Catelli (2001, p. 346) a Controladoria surgiu da necessidade de "assegurar a otimização do resultado econômico da organização na busca pela eficácia institucional, de forma a possibilitar o crescimento dos ganhos corporativos, por meios de decisões assertivas". A Controladoria busca um posicionamento técnico sobre as atividades operacionais das empresas, atuando nos mais diversos setores, dependendo da estrutura e necessidade operacional de cada empresa.

Corroborando com este mesmo entendimento Tambosi, Junior, Hein e Kroenke (2021, p. 4) "uma visão contemporânea, a controladoria é uma área nas organizações que tem como diligência a eficácia nos processos internos e externos de gestão, assegurando que todos os envolvidos (stakeholders) tenham acesso as informações necessárias para o desenvolvimento de seus trabalhos." Contudo, a Controladoria não se limita somente às informações contábeis e revisão de processos internos e externos, sua abrangência vai além desta visão ultrapassada. Na atualidade a Controladoria necessariamente precisa ser estratégica, utilizando de ferramentas técnicas e modernas para extrair do passado dados e experiências que ajudem a entender a realidade do presente e projetar perspectivas e cenários para o futuro, vislumbrando o futuro almejado pela empresa e o mais importante, o que será preciso realizar para conseguir atingir os objetivos almejados, acompanhando os resultados através de relatórios e ferramentas pertinentes em tempo hábil para certificar que os resultados obtidos sejam satisfatórios e caso não sejam, realizar os ajustes necessários para atingí-los.

Com a crescente necessidade de profissionalização da administração das empresas a Controladoria vem conquistando cada vez mais respeito junto aos acionistas e diretores das

organizações, e com isso sendo cada vez mais consultada no planejamento de investimentos das empresas. Reforçando este posicionamento Anagusko, Araki e Moser (2020) entendem que as funções e atividades desenvolvidas pela Controladoria, na atualidade, tornaram-se fatores relevantes para o controle e planejamento a médio e longo prazo de qualquer tipo de organização, com ou sem finalidades lucrativas.

O responsável por gerenciar o Departamento de Controladoria é o Controller, profissional que necessita possuir experiência nas áreas contábil, tributária, financeira, custos, dentre outras áreas afins, para conseguir desempenhar suas atribuições. Neste mesmo sentido, segundo Lopes (2018, p. 1), Controller é o:

> profissional que necessariamente precisa possuir conhecimento multidisciplinar, dominar os processos operacionais da empresa com experiência em diagnósticos preventivos de problemas e capacidade de gerir e solucionar não conformidades de forma respaldada e tempestiva, sempre visando a redução de custo na empresa e otimizando os recursos para obter eficiência operacional. (Lopes, 2018, p. 1).

Corroborando esta mesma visão, Pletsch, Silva e Lavarda (2015, p. 122), consideram que o "controller passou a ser considerado parceiro de negócio, pois sua responsabilidade passa por uma ampla gama de funções, além do papel de núcleo tradicional no fornecimento contínuo de informações para a gestão empresarial." No mesmo entendimento, Souza, De Araújo Wanderley e Horton (2021, p. 5), o Controller além das habilidades técnicas citadas acima também precisa ter habilidades sociais para executar suas atribuições:

> Ao delinear a estrutura de competências requerida dos controllers, o Chartered Global Management Accountant (CGMA) afirma, nos seus Princípios Globais, que tanto as habilidades técnicas (relacionadas a finanças, contabilidade e gerência de negócios) quanto as sociais (relacionadas à liderança e à capacidade de influenciar pessoas) são componentes importantes na formação destes profissionais. (Souza, De Araújo Wanderley e Horton, 2021, p. 5).

Neste mesmo entendimento, segundo Alectoridis, Oyadomari, Carneiro e Antunes (2018, p. 181), "para que os controllers apoiem as áreas de negócios, gerando conhecimento que possa ser compartilhado, é necessário que ocorra a interação com os profissionais dessas áreas, uma vez que o ambiente de trabalho é caracterizado pela interação entre as pessoas."

O Controller através do Departamento de Controladoria pode ajudar os acionistas, diretores e demais usuários, a superar possíveis deficiências gerenciais, pois estão tecnicamente preparados para orientá-los no processo de tomada de decisões assertivas.

2. Sistema de informações

Para a otimização do desempenho das atribuições da Controladoria é fundamental que o Controller, profissional responsável por gerenciar a Controladoria, possua uma visão refinada do sistema de informações existente na organização, extraindo dados de todos os setores da empresa, como contabilidade, produção, financeiro, custos, dentre outros setores.

Contudo, a Controladoria não pode limitar-se a extrair dados internos somente, pois em vários momentos será necessário extrair dados externos, como cotação do dólar, projeção de crescimento da população e do produto interno bruto, novas tecnologias, resultados dos concorrentes, dentre outros.

O fato de apenas extrair dados sem o devido tratamento não traz benefícios para as empresas, assim é atribuído à Controladoria realizar a filtragem e análise dos dados, interpretando e confeccionando relatórios adequados e tempestivos que auxiliem os usuários das informações. Com o aumento contínuo do volume de dados gerados diariamente pelas empresas é essencial e comum o uso de um sistema operacional EPR, sigla em inglês para "*Enterprise Resource Planning*", que é o nome do sistema de informação tecnológica que integra as várias áreas de uma empresa de forma integrada e contínua, funcionando cada área em simbiose, todas imputando informações no sistema que já foi anteriormente parametrizado para recebê-las. As empresas buscam por novas tecnologias que otimizem o processo e que facilitem o trabalho dos usuários, almejando identificar oportunidades e deficiências.

> A perspectiva de uma contínua e dinâmica mudança tecnológica pode oferecer oportunidades de negócios para a empresa, ao mesmo tempo em que sua ausência pode também ameaçar a suas atividades, pois uma tecnologia nova desencadeia forças que provocam mudanças nos valores dos consumidores de bens e serviços que ela produz. (Nascimento, Reginato, 2010, p. 65).

É comum as empresas possuirem em seu sistema de informação além do sistema operacional ERP vários outros subsistemas que atenderão necessidades específicas das empresas e que geralmente se comunicaram com o sistema operacional ERP através de *imputs* de entradas e saídas. Cabe ao Controller a atribuição de conhecer os procedimentos internos da empresa e definir junto a sua alta administração a visão de controle desejado, para isso também padronizar um procedimento contínuo de treinamento para colaboradores que imputam as informações no sistema de informação, garantindo com isso informações corretas.

> Por isso é fundamental que o contador ou o controller conheçam detalhadamente os sistemas de controles internos da empresa. Eles precisam conhecer os parâmetros de contabilização das operações cadastradas em um sistema integrado. Além disso, eles precisam conscientizar os gestores dos outros setores a cumprirem com os procedimentos que visam garantir os objetivos de controle. (Garcia, 2010, p. 43).

Sistema de informações é um dispositovo de controle usado pelo Controller, que será parametrizado pela Controladoria de forma a espelhar a visão de controle da empresa e tornará mais prático o processamento das informações e sua transformação em relatórios de controle, otimizando os resultados da empresa para as espectativas desejadas.

> A integração dos sistemas de informações ao processo de gestão determina a eficácia dos mecanismos de autocontrole e feedback, que constituem requisitos para que o sistema empresa mantenha-se no rumo dos resultados desejados, que são alcançados pelo intermédio de informações gerais. (Schier, 2005, p. 19).

O sistema de informação é um conjunto formado por todas as áreas da empresa imputando informação referentes às operações da mesma contudo, por ser atribuição da Controladoria conhecer os processos operacionais da empresa e muitas vezes ser a responsável por criá-los são necessárias ferramentas de acompanhamento diário dos processos e resultados que permitam verificar se os resultados alcançados estão dentro do que foi projetado, e, caso não estejam, o Controller irá tomar as medidas cabíveis para obter os resultados projetados da forma mais eficaz e satisfatória possível.

3. Caracterização do problema

Uma grande agroindústria de alimentos nacional produtora de proteína animal de aves, especificamente galináceos, pretende aumentar sua produção para os próximos anos com altos investimentos em máquinas, equipamentos, construção de granjas, ampliação de incubatório e abatedouro, isso, prevendo o aumento da demanda de proteínas animal e consequente oportunidades de negócios.

Para tanto, os acionistas da agroindústria solicitaram a criação de uma força tarefa, liderada pela Controladoria, com a finalidade de entender quais são as relações existentes entre várias variáveis que associem ou expliquem o aumento do valor da folha de pagamento.

Dentro desta realidade é comum a Controlaoria participar de projetos de estudo de viabilidade de investimentos nas empresas, realizando estudos de cenários e variáveis, identificando correlações sobre indicadores, tudo isso tentando auxiliar na assertividade das decisões, sendo comum a utilização de técnicas e ferramentes especificas, sendo uma delas a utilização de instrumentos estatísticos e matemáticos, baseados em métodos quantitativos, em especial a Regressão Linear.

Para realizar o presente estudo utilizar-se-á de ferramentas estatísticas e matemáticas, baseadas em métodos quantitativos, em especial a Regressão Linear, através de softwares especializados, neste caso GPower e Stata, para tentar associar ou explicar o aumento da folha de pagamento com várias variáveis estudadas, buscando entender as interações do passado para ser mais assertivo nas decisões do presente em relação aos objetivos futuros.

Dentro da situação prática demonstrada pela presente pesquisa, levantou-se o seguinte problema de pesquisa: é possível através do uso de instrumentos estatísticos e matemáticos, baseados em métodos quantitativos, em especial a Regressão Linear, prever a estimação do comportamento do aumento da folha de pagamento através dos comportamentos de outras variáveis correlacionadas?

4. Objetivos e relevância da pesquisa

O objetivo desta pesquisa é demonstrar e evidenciar a importância do uso de ferramentas estatísticas como elemento de suporte a decisões assertivas, auxiliando a Controladoria em suas atribuições.

Sua relevância fundamenta-se nas contribuições que os instrumentos estatísticos e matemáticos, em especial a Regressão Linear, podem proporcionar no cotidiano da Controladoria como ferramenta útil no auxílio das atribuições, ajudando no entendimento e explicação de problemas empresariais. Sendo um diferencial competitivo para a empresa que

possuir um departamento de controladoria que realmente desenvolvam suas atribuições, que estruture seu sistema de informações de forma efetiva, realizando lenvatamentos, relatórios e previsões que auxiliem os tomadores de decisões a serem assertivos.

5. Metodologia

Nesta seção descreve-se o enquadramento tipológico da pesquisa, os procedimentos para coleta e análise dos dados e utilização do método/técnica com conceitos fundamentais relativos a sua utilização.

5.1. Tipologia da pesquisa

Adotou-se para a presente pesquisa natureza aplicada, gerando conhecimentos úteis visando a solução de problemas específicos, que é conforme a visão de Cervo e Bervian (1996, p. 47), "o pesquisador é movido pela necessidade de contribuir para mais ou fins práticos menos imediatos, buscando uma solução para problemas concretos." Ainda, adotar-se-á abordagem quantitativa, uma vez que serão utilizados números e instrumentos estatísticos com coleta de dados, que segundo Lima Junior, Tenório, Santos e Soares (2007, p. 24), "a pesquisa qualitativa não corresponde simplesmente às técnicas qualitativas para obtenção de respostas, mas sim a natureza da análise requerida para aplica-la. Ela possui índole interpretativa e descritiva."

5.2. Procedimentos para coleta e análise dos dados

Foi realizada pesquisa com seleção de referências bibliográficas pautada de duas formas, a primeira na seleção de artigos científicos e a segunda através de análise de referência bibliográficas.

Dentro destas premissas de pesquisa utilizou-se buscas, por meio do uso de palavras-chave, quais sejam, controladoria, métodos quantitativos e regressão linear múltipla em publicações de revistas com notório reconhecimento científico e classificação junto à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – CAPES, com preferência de notas de A1, A2, B1 e B2, para os artigos. Em relação as referências bibliográficas utilizaram-se obras de autores especialistas no tema pesquisado.

Os dados coleados foram obtidos junto a uma grande agroindústria de alimentos nacional, produtora de proteína animal de aves, especificamente galináceos, que pretende aumentar a sua produção para os próximos anos com altos investimentos em máquinas, equipamentos, construção de granjas, ampliação de incubatório e abatedouro, isso, prevendo o aumento da demanda de proteína animal e consequente oportunidades de negócios. Para tanto, os acionistas da agroindústria solicitaram a criação de uma força tarefa, liderada pela Controladoria, com a finalidade de entender quais são as relações existentes entre várias variáveis que associem ou expliquem o aumento da folha de pagamento.

O controller, após várias pesquisas, identificou duas variáveis independentes que podem influenciar no aumento da folha de pagamento; a primeira chamada de FUN, sendo esta a série histórica da quantidade de funcionários do período de janeiro de 2015 a dezembro de 2020, e a segunda variável independente chamada de PRO que é a série histórica da quantidade produzida no mesmo período da variável independente anterior. Após a

identificação dessas variáveis independentes o Controller decidiu analisá-las por meio da Regressão Linear Múltipla, estudando qual será o comportamento da folha de pagamento, chamada de FOL, em detrimento da correlação existente entre FUN e PRO. Especificou-se que o estudo utilizou dados do período de janeiro de 2015 a dezembro de 2020, considerando o tamanho do efeito convencionado médio de 15% (quinze por cento), a probabilidade de erro ou nível de significância de 5% (cinco por cento), o nível de confiança de 80% (oitenta por cento), com dois preditores (FUN e PRO) para desenvolver um modelo matemático que preveja o valor do FOL a partir dos valores das variáveis FUN e PRO. Seguem os dados obtidos junto a grande agroindústria de alimentos nacional:

PE	PRO	FUN	FOL	PE	PRO	FUN	FOL	PE	PRO	FUN	FOL	PE	PRO	FUN	FOL
1	15.302	2.306	2.725	19	17.674	2.422	3.346	37	20.829	2.703	3.913	55	23.584	2.859	4.358
2	11.950	2.422	2.960	20	18.539	2.449	3.348	38	18.505	2.755	4.047	56	22.411	2.877	4.404
3	15.134	2.411	2.932	21	17.416	2.422	3.413	39	20.505	2.716	4.027	57	20.654	2.861	4.417
4	12.997	2.404	3.052	22	16.494	2.524	3.388	40	21.334	2.745	4.039	58	22.634	2.767	4.429
5	14.457	2.486	3.073	23	15.748	2.541	3.568	41	17.249	2.627	4.034	59	22.599	2.867	4.562
6	13.816	2.505	3.116	24	18.835	2.486	3.471	42	22.214	2.729	4.102	60	18.280	2.796	4.388
7	15.804	2.553	3.087	25	18.451	2.589	3.568	43	21.250	2.787	4.288	61	23.587	2.868	4.356
8	14.456	2.647	3.170	26	17.538	2.663	3.710	44	22.589	2.749	4.051	62	19.450	2.886	4.494
9	13.103	2.638	3.237	27	20.999	2.659	3.744	45	20.308	2.796	4.157	63	19.897	2.877	4.526
10	14.389	2.634	3.169	28	16.126	2.712	3.898	46	21.081	2.722	4.058	64	22.110	2.972	4.980
11	15.578	2.702	3.332	29	19.121	2.689	3.970	47	16.245	2.737	4.402	65	21.675	2.941	4.762
12	16.939	2.686	3.205	30	18.292	2.688	3.984	48	18.489	2.694	4.455	66	20.258	3.002	4.874
13	17.542	2.758	3.425	31	19.617	2.697	3.900	49	21.620	2.766	4.173	67	22.651	3.020	4.895
14	15.347	2.668	3.435	32	21.594	2.688	3.922	50	21.791	2.840	4.500	68	22.301	3.106	5.086
15	18.704	2.583	3.576	33	18.594	2.705	4.002	51	15.046	2.847	4.345	69	20.706	3.030	4.973
16	17.711	2.505	3.549	34	20.325	2.704	4.010	52	20.087	2.877	4.457	70	21.759	3.087	4.871
17	17.929	2.442	3.377	35	19.399	2.732	4.211	53	21.977	2.876	4.417	71	22.450	3.097	4.978
18	17.499	2.447	3.407	36	17.568	2.701	4.074	54	18.144	2.892	4.461	72	23.527	3.027	4.886

Os dados referentes a variável dependente FOL (folha de pagamento) e a variável independente PRO (produção) foram coletados com as unidades em milhões e para melhor adequação ao estudo foram transformadas para unidade de milhares. A variável independente FUN (funcionários) não sofreu alterações.

5.3. Utilização de métodos quantitativos

O presente estudo busca demonstrar a viabilidade do uso de instrumentos estatísticos e matemáticos, baseados em métodos quantitativos, em especial a Regressão Linear, através de softwares especializados, neste GPower e Stata, para estabelecer modelos, cenários, correlação de variáveis e indicadores, almejando entender as interações do passado para ser mais assertivo nas decisões do presente em relação aos objetivos futuros.

A regressão linear é a equação que define a correlação entre duas variáveis em termos matemáticos. É regressão linear simples quando existe a correlação de duas variáveis, sendo uma variável dependente e outra independente; e regressão linear múltipla quando existe a correlação de mais de duas variáveis, sendo obrigatoriamente uma variável dependente e as demais independentes.

A regressão linear é utilizada para estimar ou explicar os valores de uma variável, no caso a dependente, com base nos valores de outra ou outras variáveis, no caso independentes.

Dentro deste contexto a regressão linear múltipla é a técnica adequada para o estudo de viabilidade de investimentos nas empresas, pois realiza estudos de cenários de suas variáveis, identifica correlações sobre indicadores e prevê os valores para a variável

dependente, tudo isso para auxiliar na assertividade das decisões, sendo comum sua utilização.

A Correlação mede a força de associação entre duas variáveis, podendo ser depenente com independente, ou entre variáveis independentes, podendo variar entre -1 a +1, onde o sinal positivo indica relação positiva, ou seja, quando uma variável aumenta a outra também aumenta, o sinal negativo indica relação negativa, ou seja, quando uma variável aumenta a outra diminue.

A definição de coeficiente de determinação (R^2), segundo Hair, Black, Babin e Anderson, (2009, p. 150), "medida da proporção da variância da variável dependente em torno de sua média que é explicada pelas variáveis independentes ou preditoras. O coeficiente pode variar entre 0 e 1." Ainda segundo Hair, Black, Babin e Anderson, (2009, p. 150), "se o modelo de regressão é propriamente aplicado e estimado, o pesquisador pode assumir que quanto maior o valor de R^2, maior o poder de explicação da equação de regressão e, portanto, melhor a previsão da variável dependente." Ou seja, quanto mais próximo for o valor obtido de 1, melhor a explicação da equação da regressão.

Quanto ao conceito do coeficiente ajustado de determinação (R2 ajustado), segundo Hair, Black, Babin e Anderson, (2009, p. 150), e:

> Coeficiente ajustado de determinação (R2 ajustado) Medida modificada do coeficiente de determinação que considera o número de variáveis independentes incluídas na equação de regressão e o tamanho da amostra. Apesar de a adição de variáveis independentes sempre fazer com que o coeficiente de determinação aumente, o coeficiente ajustado de determinação pode cair se as variáveis independentes acrescentadas tiverem pouco poder de explicação e/ou se os graus de liberdade se tornarem muito pequenos. Essa estatística é muito útil para comparação entre equações com diferentes números de variáveis independentes, diferentes tamanhos de amostras, ou ambos. (Hair, Black, Babin e Anderson, 2009, p. 152).

Considera o mesmo entendimento do coeficiente de determinação (R^2), ou seja, quanto mais próximo for o valor obtido de 1, melhor a explicação da equação da regressão.

Segundo Sell (2005, p. 5), "erro padrão, mede a dispersão da estimativa do parâmetro, indicada pelos dados amostrais, podendo ser obtido pela raiz quadrada da variância residual. É uma espécie de desvio padrão que mede a dispersão em torno da reta de regressão."

Definiu-se no item 5.2 que a presente pesquisa considera uma probabilidade de erro ou nível de significância de 5% (cinco por cento), segundo Hair, Black, Babin e Anderson, (2009, p. 152):

> Nível de significância (alfa) Frequentemente chamado de nível de significância estatística, o nível de significância representa a probabilidade que o pesquisador deseja aceitar de que o coeficiente estimado seja classificado como diferente de zero quando realmente não é. É também chamado de erro Tipo I. O nível de significância mais amplamente usado é 0,05, apesar de pesquisadores utilizarem níveis que variam de 0,01 (mais exigentes) até 0,10 (menos conservador e mais fácil de descobrir significância). (Hair, Black, Babin e Anderson, 2009, p. 152).

O nível de significância geralmente aceito para as ciências sociais é de 5% (cinco por cento), contudo, para os trabalhos das áreas de medicina e psicologia o nível de significância é mais exigente, geralmente 1% (um por cento), e temos também pesquisas menos conservadoras com nível de significância de 10% (dez por cento).

Ainda com conceitos, fala-se da estatística F que segundo Sell (2005, p. 5), e:

> Estatística F, utilizada para testar o efeito conjunto das variáveis independentes sobre a dependente, ou seja, serve para verificar se, pelo menos, um dos X explica a variação do Y. Desse modo, a hipótese nula (Ho) indicará que nenhum dos X afeta Y, enquanto a hipótese alternativa (H1) assegura que, pelo menos uma das variáveis independentes influenciará a variável dependente Y, dentro dos graus de liberdade exigidos e dado um grau de significância. (Sell, 2005, p. 5).

Ainda segundo Sell (2005, p. 5), "considerando um nível de significância igual a 0,05, se F de significação for $< 0,05$, a regressão é significativa, mas se for $\geq 0,05$, a regressão não é significativa." E que a "estatística T para os coeficientes das variáveis independentes, ou valor da prova, tem a finalidade de testar se o efeito de cada uma das variáveis independentes sobre a dependente é ou não estatisticamente significativo."

Intervalo de confiança é o intervalo dentro dos 95% (noventa e cinco por cento) de confiança, isso devido ter-se considerado uma probabilidade de erro ou nível de significância de 5% (cinco por cento), em que há variáveis dependente e independentes que precisam estar contidas, assim como o coeficiente angular Beta. Destarte, estando as observações dentro do intervalo de confiança e constatado que o valor de zero não está contido nestes intervalos, conclui-se pela rejeição da hipótese nula de que o coeficiente angular é = 0, aceita-se a hipótese alternativa de que $b \neq 0$, significando que as variáveis independentes tem influência significativa sobre a variável dependente.

O P-valor informa se o coeficiente angular é igual a zero, comparando a probabilidade de erro ou nível de significância de 5% (cinco por cento), parâmetro definido no problema de pesquisa; se o P-valor das variáveis independentes forem igual a zero, são menores que o nível de significância de 5% (cinco por cento), então rejeita-se a hipótese H=0, e aceita-se a hipótese $H \neq 0$, portanto, as variáveis independentes tem influência significativa sobre a variável dependente. Os testes de hipóteses são segundo Sell (2005, p. 6):

> Teste de hipótese é um processo capaz de afirmar, com base em dados amostrais, se uma hipótese sob prova é correta ou não. É uma afirmação que admite se certo efeito está presente ou não. Por hipótese, entende-se certa afirmação condicionada acerca de uma população, e classificam-se em dois tipos: a) hipótese nula (H=0), quando se admite não haver diferença entre a informação fornecida pela realidade e a afirmação da hipótese; b) hipótese alternativa ($H \neq 0$), quando se admite haver diferença entre a informação fornecida pela realidade e a afirmação da hipótese. Portanto, o processo de teste consiste em aceitar ou rejeitar a hipótese nula (H=0), com base na diferença entre o valor hipotético e seu estimado. (Sell, 2005, p. 6).

Para comprovação de que a regressão linear atenda as premissas definidas, que na presente pesquisa é tamanho de efeito convencionado médio de 15% (quinze por cento), probabilidade de erro ou nível de significância de 5% (cinco por cento) e nível de confiança de 80% (oitenta por cento), é necessária a realização dos testes de hipóteses e sua aprovação nos mesmos.

6. Determinação do tamanho da amostra

Inicialmente é necessário determinar se a amostra coletada atende a quantidade mínima sugerida de tamanho de amostra para o presente estudo, para isso, será utilizado o software GPower 3.1.9.7, realizando o teste definido pela opção "Linear multiple regression: Fixed model, R^2 increase", que traduzido seria "Regressão linear múltipla: modelo fixo, aumento de R^2", encontrado pelo caminho "Tests, Correlation and regression, Linear multiple regression: Fixed model, R^2 increase". O referido teste foi escolhido por ser considerado o adequado para o presente estudo e para realizá-lo será considerado o tamanho de efeito convencionado médio de 15% (quinze por cento), probabilidade de erro ou nível de significância de 5% (cinco por cento), nível de confiança de 80% (oitenta por cento) e dois preditores, quais sejam: FUN e PRO.

Após aplicação do teste de "Regressão linear múltipla: modelo fixo, aumento de R^2", obteveram-se os seguintes resultados, a saber:

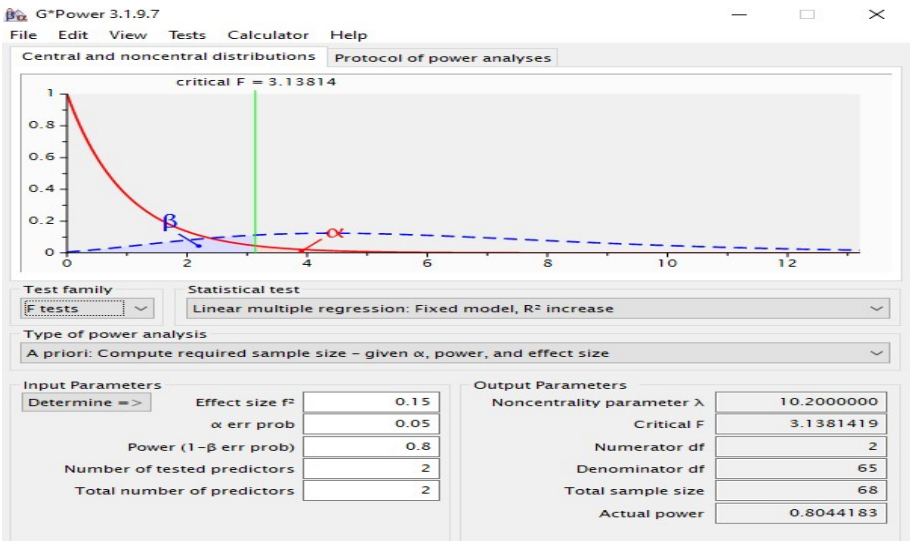

Figura 01: Tamanho mínimo da amostra no software Gpower 3.1.9.7.

No resultado do teste de "Regressão linear múltipla: modelo fixo, aumento de R^2" observa-se que o tamanho mínimo da amostra é de 68 (sessenta e oito), portanto os dados coletados que são de 72 (setenta e duas) amostras estão acima da quantidade mínima sugerida pelo teste, não sendo necessário realizar qualquer tipo de adequação.

7. Processar regressão linear múltipla

Realizou-se com os dados coletados o primeiro processamento da regressão linear múltipla, considerando como nossa variável dependente FOL e como variáveis independentes FUN e PRO, para isso, foi utilizado o software Stata 17.0, através script **"reg FOL FUN PRO"**.

Após processamento da regressão linear múltipla, obtiveram-se os seguintes resultados, a saber:

```
. reg FOL FUN PRO
```

Source	SS	df	MS	Number of obs	=	72
				F(2, 69)	=	253.67
Model	22355599.6	2	11177799.8	Prob > F	=	0.0000
Residual	3040421.23	69	44064.0758	R-squared	=	0.8803
				Adj R-squared	=	0.8768
Total	25396020.8	71	357690.434	Root MSE	=	209.91

FOL	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
FUN	2.361647	.1854769	12.73	0.000	1.991631	2.731663
PRO	.0558591	.0118081	4.73	0.000	.0323025	.0794157
_cons	-3508.13	386.9752	-9.07	0.000	-4280.124	-2736.135

Figura 02: Regressão linear múltipla no software Stata 17.0

Neste primeiro momento não foi analisado o resultado da regressão linear múltipla, devido à necessidade da realização dos testes de pressupostos; não se consegue testar os pressupostos antes de rodar a regressão linear múltipla.

8. Avaliar a presença de valores influentes

Os valores influentes ou Outliers é ou são as presenças de observações atípicas que podem influenciar os valores dos coeficientes de uma equação, podendo ser causada por valores influentes de uma variável ou de várias, tendo a capacidade de modificar a inclinação da reta gráfica. Quando ocorrer deve-se tentar entender a causa, sendo o primeiro passo verificar a existência de dados digitados ou coletados de forma errônea, caso não sejam constatados erros de informações deverá ser relatada a existência de outliers na pesquisa.

Após a confirmação do outlier o pesquisador poderá mantê-los nos dados estudados, poderá retirá-los ou poderá realizar procedimento de equalização das observações atípicas através de procedimento técnico denominado winsorizar.

Através da análise estatística descritiva analisar-se-á os dados, inicialmente com o comando sumarizar, utilizando o software Stata 17.0, através script **"sum FOL FUN PRO"**. Após processamento do comando sumarizar, foram obtidos os seguintes resultados:

```
sum FOL FUN PRO
```

Variable	Obs	Mean	Std. dev.	Min	Max
FOL	72	3965.565	598.0723	2725.413	5085.719
FUN	72	2717.583	185.0242	2306	3106
PRO	72	18899.73	2906.279	11949.76	23586.79

Figura 03: Comando sumarizar no software Stata 17.0.

Como resultado do comando sumarizar foi encontrado na coluna “Obs” o total de observações estuda na amostra de 72 (setenta e duas), na coluna “Mean” a média dos valores das observações FOL, FUN e PRO, na coluna “Std. Dev.” o valor dos desvios padrões de cada variável, “Min” os valores mínimos e “Max” os valores máximos das variáveis.

Quando da análise dos valores encontrados dos desvios padrões com a média dos valores das observações, encontrou-se os seguintes percentuais 15,08%, 6,81% e 15,38% respectivamente para FOL, FUN e PRO, percentuais esses que representam variações moderadas, indicando a princípio, a evidência de não ocorrências de valores influentes ou outliers. Contudo, utilizações de outros testes serão necessárias, sendo o próximo de análise gráfica descritiva, com a construção de um gráfico *box plot* que ilustra através da forma gráfica a ocorrência de valores influentes ou Outliers. Para melhor compreensão do estudo, faremos análise gráfica de cada uma das variáveis de forma separada através de comando, utilizando o software Stata 17.0, através dos scripts **“graph hbox FOL”** para variável dependente FOL, **“graph hbox FUN”**, para variável independente FUN e **“graph hbox PRO”** para variável independente PRO. Após processamento do comando *box plot* para variável dependente FOL, os seguintes resultados foram obtidos, a saber:

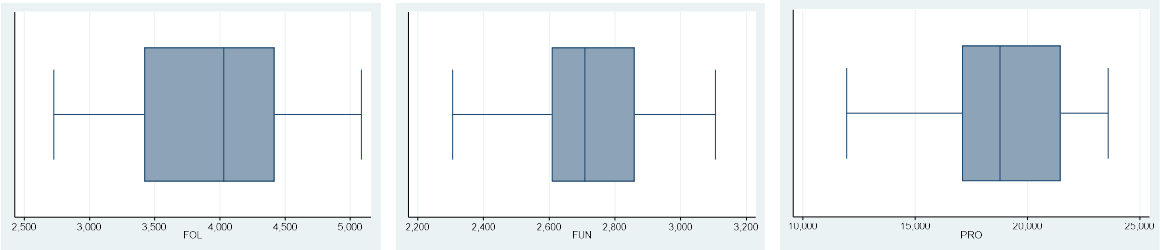

Figura 04: B*ox plot* para a variável FOL, FUN e PRO no software Stata 17.0.

Da análise do gráfico gerado pelo comando *box plot* para variável dependente FOL, e para as variáveis independentes FUN e PRO, ficou demonstrada a inexistência de valores influentes ou Outliers, uma vez que não existe a presença de observações atípicas que podem influenciar os valores dos coeficientes da equação.

Seguindo a pesquisa para confirmação de inexistência de valores influentes ou Outliers foi utilizado o diagrama de dispersão que é a representação gráfica entre duas variáveis, denominadas como independente e dependente, o qual evidência o comportamento conjunto dessas variáveis, permitindo observar se a relação existente entre as mesmas é ou não linear, ou seja, se as projeções dos pontos tendem a aproximar-se ou não de uma reta imaginária.

Quando existir relação entre as duas variáveis positiva, ou seja, se os valores de uma das variáveis aumentar os valores da outra variável também tendem a aumentar, neste caso ocorrerá uma relação positiva entre elas. Porém, quando a relação existente entre as duas for inversa, ou seja, se os valores de uma das variáveis aumentar enquanto os valores da outra variável diminuir, neste caso ocorrerá uma relação negativa.

Pode-se afirmar que existe relação linear entre as variáveis quando os dados de ambas se aproximarem de uma linha imaginária reta no diagrama de dispersão, e quando os dados de ambas estiverem dispersos em uma linha imaginária reta não existirá relação linear. Podendo ocorrer também valores influentes ou Outliers que ocorre quando tem-se a existência de

valores ou muito pequenos ou muito grandes, atípicos, que se dispersem dos valores das demais variáveis.

Utilizou-se o software Stata 17.0, para gerar o diagrama de dispersão entre as variáveis FOL e FUN, através do script **"scatter FOL FUN"**, em seguida para gerar o diagrama de dispersão entre as variáveis FOL e PRO, utilizou-se o script **"scatter FOL PRO"**. Após processamento do comando de diagrama de dispersão, obtiveram-se os seguintes resultados:

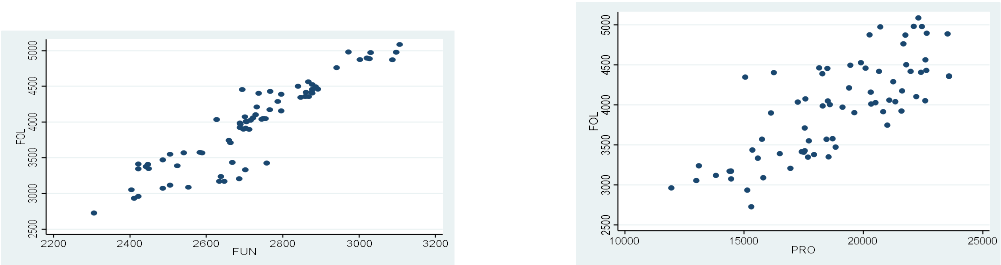

Figura 05: Diagrama de dispersão entre variáveis FOL e FUN, e FOL e PRO no software Stata 17.0.

Como resultado dos diagramas de dispersão das variáveis FOL e FUN, e FOL e PRO, verifica-se que os dados estão próximos de uma linha imaginária reta no gráfico, ficando caracterizado a existência de relação linear entre as variáveis e a não existência de valores influentes ou Outliers.

Dando continuidade à pesquisa para confirmação de inexistência de valores influentes ou Outliers, utilizou-se da análise de resíduos que é a representação gráfica de todas as variáveis e suas respectivas observações, que devem estar próximas de uma linha reta gerada a partir do ponto de dispersão zero do gráfico, sendo o ideal que as observações estejam o mais próximo da reta gerada a partir do ponto de dispersão zero do gráfico. Antes de se gerar a análise de resíduos é necessário realizar o processamento da regressão linear múltipla para geração dos resíduos utilizando o software Stata 17.0, através script **"reg FOL FUN PRO"**, para, em seguida, processar-se a análise dos resíduos, no mesmo software, através do script **"rvfplot, yline(0)"**. Após processamento do comando análise dos resíduos, obtiveram-se os seguintes resultados, a saber:

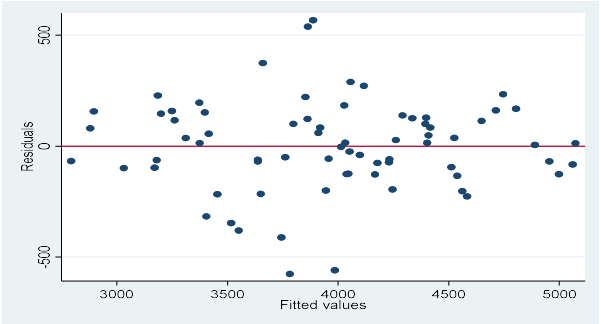

Figura 06: Análise dos resíduos no software Stata 17.0.

Da análise do gráfico gerado percebeu-se que as observações estão todas próximas da linha reta gerada a partir do ponto de dispersão zero do gráfico, não existindo a caracterização da existência de valores influentes ou Outliers.

Agora, como última ferramenta a ser utilizada para verificação da inexistência de valores influentes ou Outliers, será usado a distância de Leverage que é a representação gráfica de todas as variáveis e suas respectivas observações, que devem estar próximas de duas linhas retas perpendiculares na representação gráfica, sendo o ideal que as observações estejam o mais próximo das duas retas perpendiculares. Para gerar o comando distância de Leverage utilizaremos o software Stata 17.0, através script **"lvr2plot, mlabel(FOL FUN PRO)"**. Após processamento do comando distância de Leverage, obtivemos os seguintes resultados, a saber:

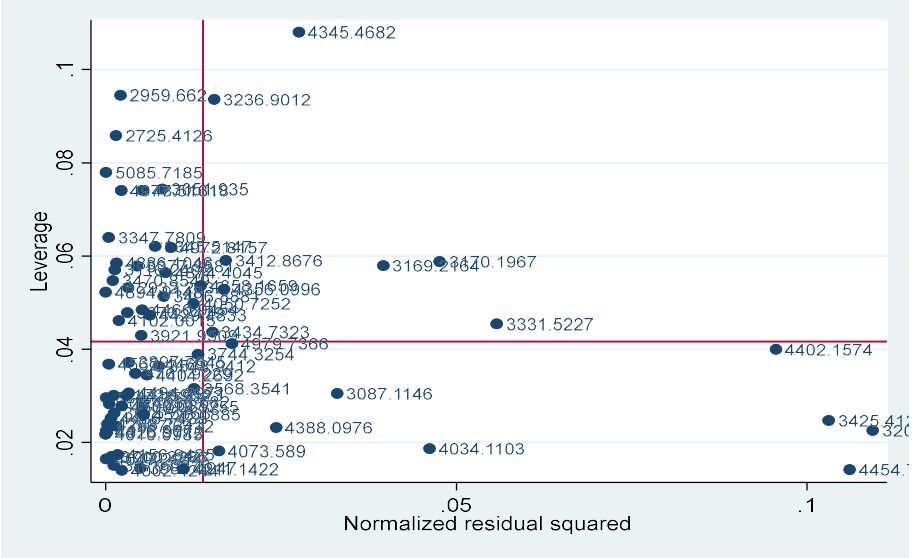

Figura 07: Distância de Leverage no software Stata 17.0.

Ao analisar as observações da distância de Leverage no gráfico acima fica demonstrado que as observações estão todas próximas das duas linhas retas perpendiculares na representação gráfica, não existindo a caracterização da existência de valores influentes ou Outliers.

Após realizadas as verificações através da sumarização **"sum FOL FUN PRO"**, *box plot* **"graph hbox FOL"/"graph hbox FUN"/"graph hbox PRO"**, diagrama de dispersão **"scatter FOL FUN"/ "scatter FOL PRO"**, análise de resíduos **"rvfplot, yline(0)"** e distância de Leverage **"lvr2plot, mlabel(FOL FUN PRO)"**, todos realizados por meio do software Stata 17.0, evidenciou-se a não existência de valores influentes ou Outliers.

9. Avaliar pressupostos

Segundo Levine, Stephan e Szabat (2017), pressupostos são testes de hipóteses usados para validar quaisquer conclusões a respeito de uma regressão linear.

9.1. Multicolinearidade

Segundo Hair, Black, Babin e Anderson (2009), ocorre a multicolinearidade quando as variáveis independentes possuem correlação entre si, quando apenas duas variáveis independentes possuem correlação entre si chamamos de colinearidade. Considerado

problema de dados e não de modelo, sendo a situação ideal quando termos diversas variáveis independentes altamente correlacionadas com a variável dependente, no entanto, pouco correlacionadas entre as próprias variáveis independentes.

Entre quase todos os conjuntos existem algum grau de correlação entre os preditores, o grau desta correlação que irá determinar a existência de multicolinearidade e para estimar este grau utilizar-se o teste correlação entre as variáveis no software Stata 17.0, através script **"pwcorr FOL FUN PRO, star(0.05)"**. Após processamento do comando de correlação entre as variáveis, obtivemos os seguintes resultados, a saber:

```
. pwcorr FOL FUN PRO, star(0.05)

             |      FOL      FUN      PRO
-------------+---------------------------
         FOL |   1.0000
         FUN |   0.9173*  1.0000
         PRO |   0.7739*  0.6878*  1.0000
```

Figura 08: Correlação entre as variáveis no software Stata 17.0.

Como parâmetro de observação para o teste de correlação realizado acima considera-se correlação fraca quando possuir resultado de correlação abaixo de 30% (trinta por cento), correlação moderada quando possuir correlação entre 30% (trinta por cento) e 50% (cinquenta por cento) e por fim correlação forte quando possuir correlação acima de 50% (cinquenta por cento), dessa forma, demonstra-se acima que todas correlações entre as variáveis são fortes. Ter correlação forte somente entre a variável dependente e as independentes é o ideal segundo Hair, Black, Babin e Anderson (2009). Contudo, ter correlação forte entres as variáveis independente não é o ideal, e para testar se está correlação forte será considerada multicolinearidade é usado o teste de **Fator de inflação de variância - VIF** que é um indicador do efeito que as outras variáveis independentes tem sobre o erro padrão de um coeficiente de regressão, altos valores de VIF indicam um alto grau de colinearidade ou multicolinearidade entre as variáveis independentes, será realizado através do software Stata 17.0, através script **"vif"**. Antes de gerar o teste *vif* é necessário realizar o processamento da regressão linear múltipla para geração de dados utilizando o software Stata 17.0, através script **"reg FOL FUN PRO"**, logo em seguida processa-se comando *vif*, obtiveram-se os seguintes resultados:

```
. vif

    Variable |      VIF       1/VIF
-------------+----------------------
         FUN |     1.90    0.526974
         PRO |     1.90    0.526974
-------------+----------------------
    Mean VIF |     1.90
```

Figura 09: Comando VIF no software Stata 17.0.

Como parâmetro de observação para o teste *vif* realizado acima considera-se existência de multicolinearidade quando o resultado do teste for acima de 10 (dez), ou seja, demonstrou-se acima que não existe multicolinearidade na correlação das observações analisadas uma vez que o resultado do teste *vif* foi de 1,90 (um virgula noventa), sendo descartada a multicolinearidade.

Contudo, a título de conhecimento quando ocorrer multicolinearidade (quando o *vif* for maior que 10) o procedimento adequado será verificar a matriz de correlação para identificar e remover as variáveis altamente correlacionadas da regressão linear múltipla ou a partir destas variáveis criar uma variável composta e incluir na regressão linear múltipla.

9.2. Omissão de variável

Ocorre omissão de variável quando alguma variável relevante para o modelo não foi incluída na amostra estudada, enviesando os coeficientes da equação e distorcendo os resultados. Para testar a existência de variáveis omitidas é necessário realizar o teste *reset* no software Stata 17.0, através script **"ovtest"**. Antes de aplicar o teste *reset* é necessário realizar o processamento da regressão linear múltipla para geração de dados, pelo software Stata 17.0, através do script **"reg FOL FUN PRO"**, logo em seguida processa-se o comando *reset*, obtiveram-se os seguintes resultados:

```
. ovtest

Ramsey RESET test for omitted variables
Omitted: Powers of fitted values of FOL

H0: Model has no omitted variables

F(3, 66) =    0.32
Prob > F = 0.8121
```

Figura 10: Comando *reset* no software Stata 17.0.

Como parâmetro de observação para o teste *reset* realizado acima e considerado a existência de variáveis omitidas quando o resultado do teste for menor de 5 (cinco), ou seja, demonstrou-se acima que não existe variáveis omitidas na correlação das observações analisadas uma vez que o resultado do teste *reset* foi de 81,21 (oitenta e um virgula vinte e um), sendo descartada a existência de variáveis omitidas. No caso de ocorrência de variáveis omitidas a literatura orienta procurar novas variáveis de controle e incluí-las no modelo pesquisado.

9.3. Heterocedasticidade

Ocorre heterocedasticidade quando a variância de erro não e constate acarretando possíveis subgrupos dentro de um modelo. Sendo a situação ideal de pesquisa quando os dados apresentam Homocedasticidade, e quando não existi ocorrências de subgrupos dentro do modelo. Para testar a existência de heterocedasticidade realizar-se o teste de heterocedasticidade no software Stata 17.0, através script **"hettest"**. Após processamento do teste de heterocedasticidade, obtiveram-se os seguintes resultados:

```
. hettest

Breusch-Pagan/Cook-Weisberg test for heteroskedasticity
Assumption: Normal error terms
Variable: Fitted values of FOL

H0: Constant variance

    chi2(1) =   1.37
Prob > chi2 = 0.2416
```

Figura 11: Teste de heterocedasticidade no software Stata 17.0.

Como parâmetro de observação para o teste de heterocedasticidade realizado acima considera-se existência de heterocedasticidade quando possuir resultado menor de 5 (cinco), ou seja, demonstrou-se acima que não existe heterocedasticidade na correlação das observações analisadas, uma vez que o resultado do teste de heterocedasticidade foi de 24,16 (vinte e quatro virgula dezesseis), sendo descartada a existência de heterocedasticidade. No caso de ocorrência de heterocedasticidade a literatura nos orienta a realizar a regressão com estimador robusto de White, que é realizado através do script **"reg FOL FUN PRO, rob"**.

9.4. Erros não normais

Ocorre erros não normais quando existir observações espalhadas além dos intervalos de confiança do modelo, sendo a situação ideal de pesquisa quando as observações se apresentam dentro do intervalo de confiança do modelo. Pequenos desvios de normalidades não afetam a equação de forma significativa, sendo assim, não iremos realizar o teste de erros não normais.

Contudo, a título de conhecimento o teste adequado para testar erros não normais, que no caso, é o teste de normalidade Shapiro-Francia no software Stata 17.0, através script **"sfrancia residuals"**. Antes de aplicar o teste Shapiro-Francia é necessário realizar o processamento da regressão linear múltipla para geração de dados através do script **"reg FOL FUN PRO"**, em seguida é necessário salvar os erros gerados através do comando do script "**predict residuals, residuals**", para então, processar o teste de normalidade Shapiro-Francia, todos os testes no software Stata 17.0.

9.5. Não linearidade

Ocorre não linearidade quando a relação entre a variável dependente e a variável independente seja uma relação quadrática e não a forma funcional adequada linear, quando se usa mais de uma variável independente deve-se certificar que cada relação entres as variáveis independentes sejam lineares para garantir adequada representação na equação. Sendo a situação ideal de pesquisa quando as variáveis sejam todas lineares. Para testar a existência de não linearidade realiza-se o teste *linktest* no software Stata 17.0, através script **"linktest"**. Após processamento do teste de não linearidade, obtiveram-se os seguintes resultados:

```
. linktest
```

Source	SS	df	MS	Number of obs	=	72
				F(2, 69)	=	254.01
Model	22359157.2	2	11179578.6	Prob > F	=	0.0000
Residual	3036863.63	69	44012.5164	R-squared	=	0.8804
				Adj R-squared	=	0.8770
Total	25396020.8	71	357690.434	Root MSE	=	209.79

FOL	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
_hat	.8451701	.546406	1.55	0.126	-.24488	1.93522
_hatsq	.0000196	.0000688	0.28	0.777	-.0001177	.0001568
_cons	300.3242	1071.205	0.28	0.780	-1836.672	2437.32

Figura 12: Teste de não linearidade no software Stata 17.0.

Como parâmetro de observação para o teste *linktest* realizado acima e considerado a existência de não linearidade quando possuir resultado menor de 5 (cinco) no item *hatsq*, ou seja, demonstrou-se acima falta de existência de não linearidade na correlação das observações analisadas, uma vez que o resultado do teste *linktest* foi de 77,70 (setenta e sente virgula sete), sendo descartada não linearidade.

Contudo, a título de conhecimento quando ocorrer não linearidade (quando o *hatsq* for maior que 5) o procedimento adequado será alterar a forma funcional do modelo estimando um novo modelo correto.

10. Analisar regressão linear múltipla

Após todos os testes de pressupostos serem realizados de forma satisfatória, processa-se a regressão linear múltipla, considerando como variável dependente FOL e como variáveis independentes FUN e PRO, utilizando o software Stata 17.0, através script **"reg FOL FUN PRO"**. Após processamento da regressão linear múltipla, obtiveram-se os seguintes resultados, a saber:

```
. reg FOL FUN PRO
```

Source	SS	df	MS		
				Number of obs	= 72
				F(2, 69)	= 253.67
Model	22355599.6	2	11177799.8	Prob > F	= 0.0000
Residual	3040421.23	69	44064.0758	R-squared	= 0.8803
				Adj R-squared	= 0.8768
Total	25396020.8	71	357690.434	Root MSE	= 209.91

FOL	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
FUN	2.361647	.1854769	12.73	0.000	1.991631	2.731663
PRO	.0558591	.0118081	4.73	0.000	.0323025	.0794157
_cons	-3508.13	386.9752	-9.07	0.000	-4280.124	-2736.135

Figura 13: Regressão linear múltipla no software Stata 17.0.

Analisando o resultado da regressão linear múltipla, podemos constatar que:

A variável dependente é FOL enquanto as variáveis independentes são FUN e PRO.

A interseção que representa o valor da folha de pagamento é -3.508,13.

O coeficiente angular Beta que representa a variável independente FUN é 2,361647, enquanto a que representa a variável independente PRO é 0,0558591. O que significa que FOL aumentará 2,361647 unidades para cada aumento em FUN e que FOL aumentará 0,0558591 unidades para cada aumento em PRO.

A equação que expressa a relação matemática entre as variáveis é dada por FOL = -3.508,13 + 2,361647.FUN + 0,0558591.PRO.

O coeficiente de determinação R^2 indica que o poder de explicação das variações de FOL são explicadas em 88,03% pelas variáveis FUN e PRO, ao mesmo tempo. Sendo que quanto maior o valor de R^2, maior o poder de explicação da equação de regressão e, portanto, melhor a previsão da variável dependente.

Já o coeficiente de determinação R^2 ajustado indica que o poder de explicação das variações de FOL são explicadas em 87,68% pelas variáveis FUN e PRO, ao mesmo tempo.

Sendo que, similar ao coeficiente de determinação R^2, quanto maior o valor de R^2 ajustado, maior o poder de explicação da equação de regressão e, portanto, melhor a previsão da variável dependente.

O erro padrão encontrado é de 209,91, o que significa que as observações estão dispersas em torno da reta de regressão 209,91 para ambos os sentidos.

Para responder as hipóteses levantadas no enunciado do problema, também e necessário testar o modelo:

Teste de intervalo de confiança - o intervalo com 95% (noventa e cinco por cento) de confiança, está para FUN, entre 1,991631 e 2,731663, sendo o coeficiente angular 2,361647 está neste intervalo, para PRO entre 0,0323025 e 0,0794157, sendo o coeficiente angular 0,0558591 está neste intervalo e para o coeficiente angular Beta, entre -4.280,124 e – 2.736,135, sendo o coeficiente angular Beta – 3.508,13. Assim sendo, com base nestes valores verifico que o zero não está contido nestes intervalos e concluo que rejeito a hipótese nula de que o coeficiente angular é = 0, aceito a hipótese alternativa de que b≠0, significando que FUN e PRO, tem influência significativa sobre FOL.

Teste Stat "*t*" - significa que o coeficiente angular da amostra para a variável FUN está 12,73 e para a variável PRO está 4,73, desvios padrões distantes em relação ao coeficiente da população. A região de aceitação apresenta um "t" crítico entre –1,99 e +1,99 (FUN: 2,361647-1,991631/0,1854769 e PRO: 0,0558591-0,0323025/0,0118081), tanto para FUN quanto para PRO, então, rejeito a hipótese H=0, com um nível de confiança de 95%, tendo em vista que o coeficiente é 12,73 para FUN e 4,73 para PRO, fora da região de aceitação. Portanto, as variáveis FUN e PRO tem influência significativa sobre FOL.

Teste P-valor - dá como resposta que o coeficiente angular é igual a zero, mas comparando com o alfa=0,05, que é o nível de significância definido no problema, concluímos que o P-valor das variáveis FUN=0,000 e PRO=0,000 são menores que o alfa=0,05, então rejeita-se a hipótese H=0, e aceita-se a hipótese H≠0, portanto, as variáveis FUN e PRO tem influência significativa sobre FOL.

Portanto os três testes realizados comprovam influência significativa das variáveis independentes FUN e PRO, sobre a variável dependente FOL.

11. Determinar tamanho do efeito

Para determinar o tamanho do efeito deve-se utilizar o software GPower 3.1.9.7, realizando o teste definido pela opção "Linear multiple regression: Fixed model, R^2 increase", que traduzido seria "Regressão linear múltipla: modelo fixo, aumento de R^2", encontrado pelo caminho "Tests, Correlation and regression, Linear multiple regression: Fixed model, R^2 increase", aberta a tela indicada pelo caminho informado deve-se alterar o campo de "Statistical test" de "Linear multiple regression: Fixed model, R^2 increase" para "Linear multiple regression: Fixed model, R^2 deviation from zero", seguindo com o teste deve-se acionar o campo "determine", que abrirá uma aba lateral no software GPower 3.1.9.7, e nesta aba lateral deve-se alterar o campo "squared multiple correlation p^2" para o valor do "R-squared" gerado pela regressão linear múltipla no software Stata 17.0, que na presente pesquisa é de 88,03% (oitenta e oito virgula zero três por cento), mantendo todos os demais dados definidos no item "Determinação do tamanho da amostra", quais sejam, considerar o tamanho de efeito convencionado médio de 15% (quinze por cento), probabilidade de erro ou

nível de significância de 5% (cinco por cento), nível de confiança de 80% (oitenta por cento). Após aplicação do teste, obtiveram-se os seguintes resultados, a saber:

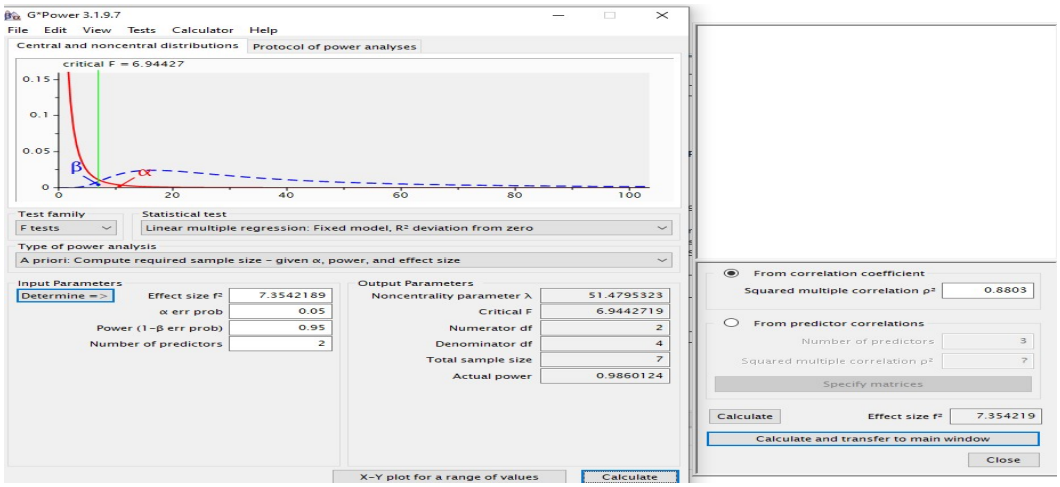

Figura 14: Tamanho do efeito no software Gpower 3.1.9.7.

Como resultado do teste observa-se que o valor encontrado foi de 7,3542189, portanto maior que o tamanho de efeito convencionado médio de 15% (quinze por cento).

12. Determinar poder de estatística

Para determinar o poder de estatística deve-se utilizar o software GPower 3.1.9.7, continuando onde finalizou-se no item anterior, assim sendo, altera-se o campo “Type of power analysis” de “A priori: Compute required sample size – given α, power, and effect size” para “Post hoc: Compute achieved power – given α, sample size, and effect size”, seguindo com o teste altera-se a informação do campo “Total sample size”, para o total de nossa amostra, que nesta pesquisa são 72 (setenta e duas) observações. Após processamento do teste de poder de estatística, obtiveram-se os seguintes resultados:

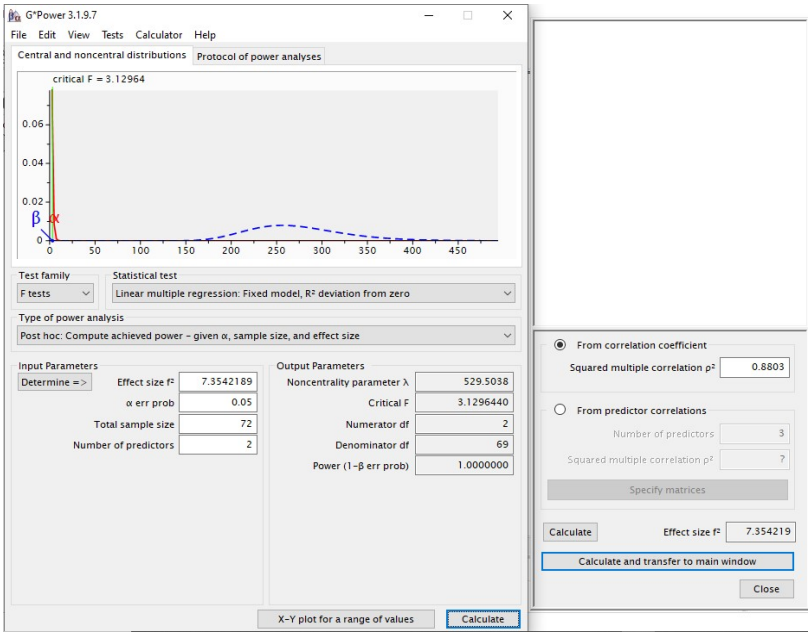

Figura 15: Poder de estatística no software Gpower 3.1.9.7.

Como resultado do teste observa-se que o valor encontrado foi de 1,0000000, poder considerado adequado para nossa pesquisa.

13. Determinar o tamanho mínimo da amostra

Para determinar o tamanho mínimo da amostra deve-se utilizar o software GPower 3.1.9.7, continuando onde finalizou-se no item anterior, assim sendo, altera-se novamente o campo de “Type of power analysis” de “Post hoc: Compute achieved power – given α, sample size, and effect size” para “A priori: Compute required sample size – given α, power, and effect size”, seguindo com o teste altera-se a informação do campo “Power (1-Beta err prob)”, para o poder encontrado no teste poder estatístico do item anterior de 1,0000000, contudo o software GPower 3.1.9.7 não consegue processar utilizando o poder de 1,0000000, assim sendo, e necessário utilizarmos o poder de 0,9999999. Após processamento do teste do tamanho mínimo da amostra, obtiveram-se os seguintes resultados:

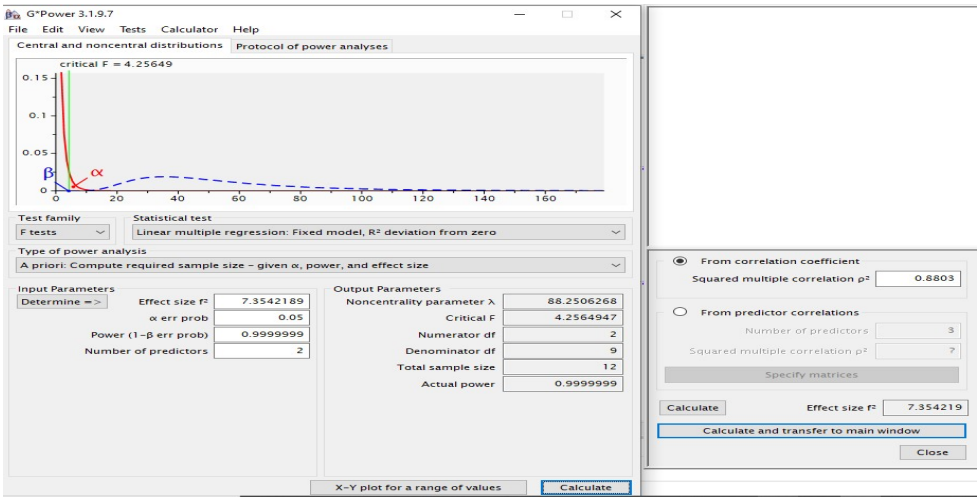

Figura 16: Tamanho mínimo da amostra no software Gpower 3.1.9.7.

Como resultado do teste observa-se que o valor encontrado foi de 12 (doze) amostras, considerando que a amostra da presente pesquisa são de 72 (setenta e duas) observações, a amostra e considerado adequado para nossa pesquisa.

14. Conclusão

Através do presente estudo conclui-se que existe possibilidade e viabilidade da utilização de instrumentos estatísticos e matemáticos, baseados em métodos quantitativos, em especial a Regressão Linear Múltipla, como ferramenta útil no auxílio das atribuições da Controladoria no seu cotidiano, ajudando no entendimento e explicação de problemas empresariais. Sendo um diferencial competitivo para a empresa que possuir um departamento de controladoria que realmente desenvolvam suas atribuições, que estruture seu sistema de informações de forma efetiva, realizando lenvatamentos, relatórios e previsões que auxiliem os tomadores de decisões a serem assertivos.

Destacou-se a importância da utilização de softwares estatísticos especializados para facilitar o trabalho operacional de cálculos, estabelecer modelos, cenários, definir correlação

de variáveis e entendimento das interações entre as variáveis estudadas, nesta pesquisa utilizamos os softwares Gpower 3.1.9.7 e Stata 17.0.

Nesta pesquisa ficou evidenciado que é possível através do uso de instrumentos estatísticos e matemáticos, baseados em métodos quantitativos, em especial a Regressão Linear Múltipla, prever a estimação do comportamento do aumento da folha de pagamento através dos comportamentos das variáveis independentes quantidade de funcionarios e quantidade de produção.

Para a presente pesquisa a utilização do método de Regressão Linear Múltipla satisfez as necessidades apresentadas, contudo, não satisfaz a resolução de todos os problemas administrativos enfrentados pelas empresas, assim sendo, precisa ser estudado sua aplicação a cada caso prático.

Referências

Alectoridis, D., Oyadomari, J. C. T., Carneiro, W. N., & Antunes, M. T. P. (2018). **Criação e socialização do conhecimento: estudo com profissionais de controladoria e operações.** Revista Contemporânea de Contabilidade, 15(35), 179-203.

American Psychological Association. (2010). **Publication manual of the American Psychological Association** (6th ed.). Washington, DC: Author.

Américo, B. L., & Lacruz, A. J. (2017). **Contexto e desempenho escolar: análise das notas na Prova Brasil das escolas capixabas por meio de regressão linear múltipla.** Revista de Administração Pública, 51, 854-878.

Anagusko, L. H., Araki, M. E., & Moser, E. M. (2020). **Implantação da controladoria em micro e pequenas empresas no Brasil: O que considerar?.** Revista da Micro e Pequena Empresa, 14(1), 59-78.

Catelli, Armando. **Controladoria: Uma Abordagem da Gestão Econômica**. 2° ed. São Paulo-SP: Atlas, 2001.

Cervo, A., & Bervian, P. (1996). **Metodologia científica.** São Paulo: Makron Books.

da Cunha, L. S., & Carlos, A. (2007). **Regressão linear múltipla. Análise multivariada: para os cursos de administração, ciências contábeis e economia.** São Paulo: Atlas.

de Oliveira, C. R. I. (2011). **Utilização da análise de regressão na estimação do comportamento dos custos: um estudo em oficina multimarcas.** REVISTA HUM@ NAE, 5(1).

Garcia, Alexandre Sanches. **Introdução à controladoria: instrumentos básicos de controle** de **gestão das empresas.** São Paulo-SP: Atlas, 2010.

Gujarati, D. N., & Porter, D. C. (2011). **Econometria básica-5**. Amgh Editora.

Hair, J. F. J.; Black, W. C.; Babin, B. J.; Anderson, R. E. **Análise multivariada de dados.** 6 ed. Porto Alegre: Bookman, 2009. [Minha Biblioteca]. Retirado de https://integrada.minhabiblioteca.com.br/#/books/9788577805341/

Levine, D.M., Stephan, D.F. & Szabat, K.A., (2017). **Estatística - Teoria e Aplicações usando MS Excel em Português,** 7ª edição. Grupo GEN. https://integrada.minhabiblioteca.com.br/books/9788521631972.

Lima Junior, C. A. C., Tenório, J. N. B., Santos, S. M., Soares, F. A. (2007). **A controladoria como ferramenta de apoio as estratégias competitivas: estudo de caso numa empresa de telefonia móvel.**

Lopes, E.T., (2018). **A importância da Controladoria e do Controller para as empresas.** Artigo de Conclusão MBA Gestão de Negócios, Controladoria e Finanças Corporativas. Instituto de Pós-Graduação – IPOG.

Nascimento, A. M., Reginato, L., (2010). **Controladoria: um enfoque na eficácia organizacional.** São Paulo-SP: Atlas.

Pletsch, C. S., Silva, A., & Lavarda, C. E. F. (2015). **Conteúdos da disciplina de Controladoria e as funções do Controller no mercado de trabalho.** Revista Pretexto.

Schier, Carlos Ubiratan da Costa. **Controladoria: como instrumento de gestão.** 2° ed. Curitiba-PR: Juruá, 2005.

Sell, I. (2005). **Utilização da regressão linear como ferramenta de decisão na gestão de custos.** In Anais do Congresso Brasileiro de Custos-ABC.

Souza, G. H. C., de Araújo Wanderley, C., & Horton, K. (2021). **Perfis dos Controllers: Autonomia e envolvimento dos profissionais de controladoria.** Advances in Scientific and Applied Accounting, 003-022.

Tambosi, S. S. V., Junior, J. T., Hein, N., & Kroenke, A. (2021). **Dimensões da importância da controladoria nas organizações: um estudo multivariado e multicritério.** Revista Contemporânea de Contabilidade, 18(46).