

ANÁLISE DE CORRELAÇÃO ENTRE OS DOCENTES E MATRÍCULAS NAS ESCOLAS PÚBLICAS E PRIVADAS DO MUNICÍPIO DE MOSSORÓ – RN NO ANO DE 2012.

Annyelly Virginia Brito (UFERSA) britoannyelly@gmail.com

David Edson Macedo Palhares (UFERSA) davidpalharess@gmail.com

Ivanilson Martins da Silva Segundo (UFERSA) segundodc_@hotmail.com

Renata Sayonara Miranda Pimenta (UFERSA) renata_sayonara_61@hotmail.com

Resumo

O artigo tem o objetivo de verificar se existe regressão ao nível de 5%, entre as variáveis docentes e matrículas nas escolas públicas e privadas do município de Mossoró – RN, no ano de 2012. O trabalho mostra a importância da interpretação de dados para a tomada de decisões e com isto, faz uso do Censo, um tipo de pesquisa que informa as características dos habitantes de uma cidade, Estado ou nação por meio de dados estatísticos.

Palavras-chave: Análise de Regressão, Censo, Correlação, Escolas Públicas e Privadas.

1. Introdução

É evidente a importância da interpretação de dados para a tomada de decisões e sua qualidade está diretamente influenciada por uma boa coleta de dados.

O censo é um tipo de pesquisa que através de dados estatísticos que informa diversas características dos habitantes de uma cidade, Estado ou nação. Na educação é realizado o Censo Escolar, que recolhe os dados estatísticos específicos das escolas públicas e privadas em todo país, no caso do presente trabalho trata-se de informações relacionadas a educação da cidade de Mossoró – RN.

No Censo, os pesquisadores do IBGE visitam todos os domicílios do país para aplicar um questionário. Depois de percorrer todos os cantos do Brasil, indo de casa em casa, os pesquisadores organizam e analisam as informações coletadas nos questionários. Em seguida, divulgam os resultados em uma série de publicações sobre os temas estudados (IBGE, 2014).

O modelo matemático proposto será o modelo de regressão por uma ou mais variáveis, de onde temos de acordo com Gujarati (2000) que esse tipo de análise se preocupa em relacionar a variável dependente com variáveis explicativas, sendo:

$$E(Y/X_i) = f(X_i)$$

Sendo que esta equação básica expressa as relações entre a variável dependente Y com as variáveis explicativas X. Ainda que mais especificamente iremos tratar das relações lineares existentes entre as variáveis X e Y.

Em muitos problemas há duas ou mais variáveis que são relacionadas e pode ser importante modelar essa relação, onde a análise de regressão é bastante útil é no caso de experimentos planejados que incluem fatores a níveis contínuos. Assim, a análise de variância é usada para

identificar os fatores significativos, e a seguir a análise de regressão é usada para construir um modelo que incorpore esses fatores (RIBEIRO; CATEN, 2010).

A análise de regressão estuda a relação causal entre uma variável a ser explicada, que possui dependência (número de matrículas - y) e uma variável independente ou explicativa (número de docentes - x). Uma análise de regressão simples indica a dependência funcional entre essas duas variáveis, a cada x corresponde um y . O objetivo é explicar como y varia de acordo com as mudanças em x .

Quando vamos estudar o relacionamento entre duas variáveis, o conceito de correlação é de fundamental importância para o entendimento das variações existentes; segundo Triola (1999) podemos definir correlação como alguma forma de relacionamentos entre duas ou mais variáveis.

Através da análise de regressão pode ser verificada a situação entre alunos e professores das escolas. De maneira que o tema utilizado possui grande importância social, já que pode ter como resposta dessa análise se a quantidade de professores esta sendo adequada a quantidade de alunos para as escolas publicas e privadas do município de Mossoró - RN.

2. Objetivos

2.1 Objetivo Geral

Verificar se existe regressão ao nível de 5%, entre as variáveis docentes e matrículas nas escolas públicas e privadas do município de Mossoró – RN, no ano de 2012.

2.2 Objetivos Específicos

Interpretação dos parâmetros do modelo, suposições para o modelo;

Estimação dos parâmetros;

Estudo da relação entre as variáveis X e Y , o coeficiente de correlação e análise da regressão;

Coeficiente de determinação;

Análise de variância do modelo;

Análise dos resíduos.

3. Referencial Teórico

3.1 Censo

O Censo é um tipo de pesquisa no qual coleta informações sobre todas as unidades da população. É de caráter obrigatório e as informações coletadas são confidenciais, as quais se destinam apenas para fins estatísticos. Todos os indivíduos civilmente capazes estão obrigados por lei a prestarem informações constantes do questionário da pesquisa (Decreto nº 73.177, de 20 de novembro de 1973).

Segundo o IBGE (2014), o primeiro Censo aconteceu em 1872 e recebeu o nome de Recenseamento da População do Império do Brasil e o mais recente ocorreu em 2010. Os

pesquisadores aplicam os questionários, em seguida organizam e analisam as informações coletadas e divulgam os resultados em uma série de publicações sobre os temas estudados.

Há diversos tipos de censos possíveis de realização, como: Censo demográfico, Censo agropecuário, Censo eleitoral, Censo escolar, entre outros. Através destes censos, é possível obter informações de todos os municípios brasileiros e identificar suas características.

3.2 Censo escolar

O Censo Escolar é um levantamento de dados estatístico-educacionais de âmbito nacional realizado todos os anos e coordenado pelo Inep. Conta com a colaboração das secretarias estaduais e municipais de Educação e com a participação de todas as escolas públicas e privadas do país. É considerado o principal instrumento de coleta de dados da educação básica, abrange da pré-escola até o ensino médio, além dos dados estarem relacionados aos estabelecimentos, docentes, alunos e rendimento escolar (INEP, 2011).

De acordo com o INEP (2011), o preenchimento dos dados é realizado através da Internet, por meio do sistema Educacenso, no endereço eletrônico www.educacenso.inep.gov.br. O sistema é de fácil operação e dispõe de relatórios com informações da escola. É realizado anualmente, nos últimos anos tem sido feito da última quarta-feira de maio até meados de agosto, onde se encerra o período de coletas.

É de fundamental importância para as escolas, participarem deste censo, pois é possível avaliar a situação educacional de cada cidade e os dados coletados servem de base para Programas como o Programa Nacional do Livro Didático (PNLD), o Programa Nacional de Alimentação Escolar (Pnae), o Programa Nacional de Transporte Escolar (Pnate), o Programa Dinheiro Direto na Escola (PDDE) e para a determinação dos coeficientes de distribuição dos recursos do Fundo de Manutenção e Desenvolvimento da Educação Básica e de Valorização dos Profissionais da Educação (Fundeb), entre outras ações (INEP, 2011).

3.3 Análise de regressão

Segundo Freund (2006), em muitas pesquisas estatísticas, o objetivo principal é estabelecer relações causais entre uma variável dependente e uma variável independentes. O ideal seria se pudéssemos prever uma quantidade exatamente em termos de outra, mas isso raramente é possível. Na maioria dos casos, detectamos previsão de médias ou de valores esperados.

- Método dos mínimos quadrados:

O método dos mínimos quadrados é um procedimento que usa dados amostrais para encontrar a equação de regressão estimada. Para obter o cálculo da equação da reta basta aplicar a seguinte equação:

$$y = a + bx$$

As estimativas de a e b, pelo método dos mínimos quadrados são os valores que minimizam a soma dos quadrados. Para se encontrar os valores de a e b, segue a fórmula conforme descreve David e Dennis *et all* (2011):

$$a = \frac{\sum x^2 \sum y^2 - \sum x \sum xy}{n \sum x^2 - (\sum x)^2}$$

$$b = \frac{n \sum xy - \sum x \sum y}{n \sum x^2 - (\sum x)^2}$$

- O coeficiente de correlação e a Análise de regressão

David e Dennis *et all* (2011) diz que para medir o grau de correlação entre duas variáveis, o sinal aritmético associado com o coeficiente de correlação, que é sempre o mesmo sinal associado na equação de regressão, indica a direção da relação X e Y.

- Se o sinal associado for positivo indica que a correlação é direta;
- Se o sinal for negativo a correlação será inversa;
- Se a correlação for nula não haverá padrão de tendência definido.

$$r = \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{\sqrt{n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2}}$$

- Análise da Variância do Modelo

Para Freund (2006), essa análise serve para comparar simultaneamente amostras de variáveis contínuas com distribuição normal e cujas variâncias não diferem significativamente entre si, ou seja, que podem ser consideradas como estimativas da variância populacional. Essa análise é baseada na decomposição da soma de quadrados e nos graus de liberdade associados a variável resposta Y.

- Soma de quadrados total

$$\sum_{i=1}^n (Y_i - \bar{Y})^2 = SQT$$

- Soma de quadrados da regressão

$$\sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2 = SQR$$

- Soma de quadrados dos erros

$$\sum_{i=1}^n (Y_i - \hat{Y}_i)^2 = SQE$$

$$SQT = SQR + SQE.$$

- Medida de Qualidade do Ajuste

Segundo David e Dennis *et all* (2011), essa razão é usada para avaliar a eficiência de ajuste da equação de regressão estimada. Quanto mais alto o valor de R^2 , mais o modelo de regressão linear simples consegue explicar a variação de y .

$$R^2 = \text{SQR} / \text{SQT}$$

- Teste de significância do modelo

Um teste F, baseado na distribuição F de probabilidade, também pode ser usado para testar a significância na regressão, essa lógica baseia-se no desenvolvimento de duas estimativas independentes David e Dennis *et all* (2011).

$$\begin{aligned} \text{QMR} &= \text{SQR}/1 & \text{F}^* &= \text{QMR}/\text{QME} \\ \text{QME} &= \text{SQE}/n-2 \end{aligned}$$

De acordo com o valor de F^* , podemos avaliar Regra de decisão do teste de hipótese:

Rejeita-se H_0 :

$$F^* > F_{\text{tabelado}}$$

Não rejeita-se H_0 :

$$F^* \leq F_{\text{tabelado}}$$

3.4 Análise dos resíduos

- Regressão Simples;

A análise de resíduos é um conjunto de técnicas utilizadas para investigar a adequabilidade de um modelo de regressão com base nos resíduos. Já que na regressão linear e na regressão múltipla, as suposições do modelo ajustado elas precisam ser validadas para que esses resultados sejam confiáveis. Na regressão simples o resíduo (e_i) é dado pela subtração entre a variável resposta observada (Y_i) e a variável resposta estimada (\hat{Y}_i):

$$e_i = Y_i - \hat{Y}_i = Y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{1i} - \dots - \hat{\beta}_p x_{pi} \quad i = 1, \dots, n.$$

Para a análise de resíduos se o modelo for apropriado, os resíduos devem refletir as propriedades impostas pelo termo de erro do modelo, como:

- $Y = X\beta + \varepsilon$,
em que $\varepsilon = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n)'$, com
- ε_i e ε_j são independentes ($i \neq j$);
 - $\text{Var}(\varepsilon_i) = \sigma^2$ (constante);
 - $\varepsilon_i \sim N(0, \sigma^2)$ (normalidade);

iv. Modelo é linear;

v. Não existir outliers (pontos atípicos) influentes.

Existem diversas técnicas usadas para verificação das suposições, elas podem ser informais (gráficos) ou formais (testes). As técnicas informais, por serem visuais, podem ser caracterizadas como subjetivas, e por isso técnicas formais são mais indicadas para uma tomada de decisão. O ideal é que haja uma combinação entre as técnicas disponíveis, tanto formais quanto informais, para o diagnóstico de possíveis problemas nas suposições do modelo. (Action, 2000)

Existem algumas técnicas baseadas em gráficos para a análise de resíduos, como:

- **Gráfico dos resíduos versus valores ajustados:** Uma das principais técnicas utilizadas, ela detecta a heteroscedasticidade e podem indicar que não existe uma relação linear entre as variáveis explicativas, com a variável resposta através de alguma tendência nos pontos.
- **Gráfico dos resíduos versus a ordem de coleta dos dados:** ele vai avaliar a hipótese de independência dos dados.
- **Papel de probabilidade normal:** é utilizada para verificação da adequação de um determinado modelo estatístico aos dados.
- **Gráfico dos Resíduos Studentizados versus valores ajustados:** é verificado se existem outliers em Y.
- **Gráfico dos Resíduos Padronizados versus valores ajustados:** é verificado se existem outliers em Y.
- **Gráfico do Leverage (Diagonal da Matriz H):** é verificado se existem outliers em X.

Já para a análise formal dos resíduos pode ser realizados outros tipos de teste, como:

- **Testes de Normalidade:** no qual os detalhes já estão dentro dos pontos;
- **Teste de Durbin-Watson:** para testar a independência dos resíduos;
- **Teste de Breusch-Pagan e Goldfeld-Quandt:** para testar se os resíduos são homoscedásticos;
- **Teste de Falta de Ajuste:** para verificar se o modelo ajustado é realmente linear.

4. Métodos de pesquisa

4.1 Caracterização da área de estudo

Mossoró é um município brasileiro no interior do estado do Rio Grande do Norte. Pertence à mesorregião do Oeste Potiguar e à microrregião homônima, localizando-se a uma distância de 281 quilômetros a noroeste da capital do estado, Natal. Ocupa uma área de 2 099,333 km², o maior município do estado em área, sendo que 11,5834 km² estão

em perímetro urbano. Em 2013 sua população foi estimada pelo Instituto Brasileiro de Geografia e Estatística em 280 314 habitantes, sendo o segundo mais populoso do Rio Grande do Norte (ficando atrás somente da capital) e o 92º de todo o país (IBGE, 2010).

4.2 Método de Pesquisa

Pesquisa descritiva que “tem como objetivo primordial a descrição das características de determinada população ou fenômeno ou o estabelecimento de relações entre variáveis” (GIL, 1995, p. 45). São levantamentos realizados por meio de coleta de dados. As variáveis não são manipuladas, apenas as observa, registra, analisa e correlaciona.

A coleta de dados foi realizada através de dados estatísticos. Para Gil (1995, p. 158) as fontes escritas na maioria das vezes são muito ricas e ajudam o pesquisador a não perder tanto tempo na hora da busca de material em campo, sabendo que em algumas circunstâncias só é possível a investigação social através de documentos.

Os dados foram coletados através de uma pesquisa bibliográfica referente ao Censo Escolar de 2012, onde é coletado e analisado informações sobre os alunos das escolas públicas e privadas do município de Mossoró/RN. Visando obter resultados de uma análise de regressão linear entre alunos e professores.

Em nosso estudo também fizemos uso da Planilha Eletrônica tipo Excel, nas funções análise de dados e regressão para obtermos os valores que justificaram se havia ou não regressão entre os docentes e matrículas.

5. Aplicação e resultados

Análise de regressão é uma metodologia estatística que utiliza a relação entre duas ou mais variáveis quantitativas (ou qualitativas) de tal forma que uma variável pode ser prognosticada a partir da outra, ou outras. O caso mais simples de regressão é quando temos duas variáveis e a relação entre elas pode ser representada por uma linha reta.

Estamos interessados na relação dessas variáveis, as quais chamaremos de X e Y em cada amostra ou unidade experimental, e vamos usá-las para caracterizar essa relação.

Figura 1 – Relação entre os docentes e matrículas no município de Mossoró-RN

Modalidades de Ensino	Docentes	Matriculas
Ensino fundamental - escola privada - 2012 (1)	682	11.622
Ensino fundamental - escola pública estadual - 2012 (1)	634	14.521
Ensino fundamental - escola pública federal - 2012 (1)	0	0
Ensino fundamental - escola pública municipal - 2012 (1)	494	12.089
Ensino médio - escola privada - 2012 (1)	215	2.983
Ensino médio - escola pública estadual - 2012 (1)	346	8.780
Ensino médio - escola pública federal - 2012 (1)	64	571
Ensino médio - escola pública municipal - 2012 (1)	0	0
Ensino pré-escolar - escola privada - 2012 (1)	167	2.984
Ensino pré-escolar - escola pública estadual - 2012 (1)	0	0
Ensino pré-escolar - escola pública federal - 2012 (1)	0	0
Ensino pré-escolar - escola pública municipal - 2012 (1)	199	4.488

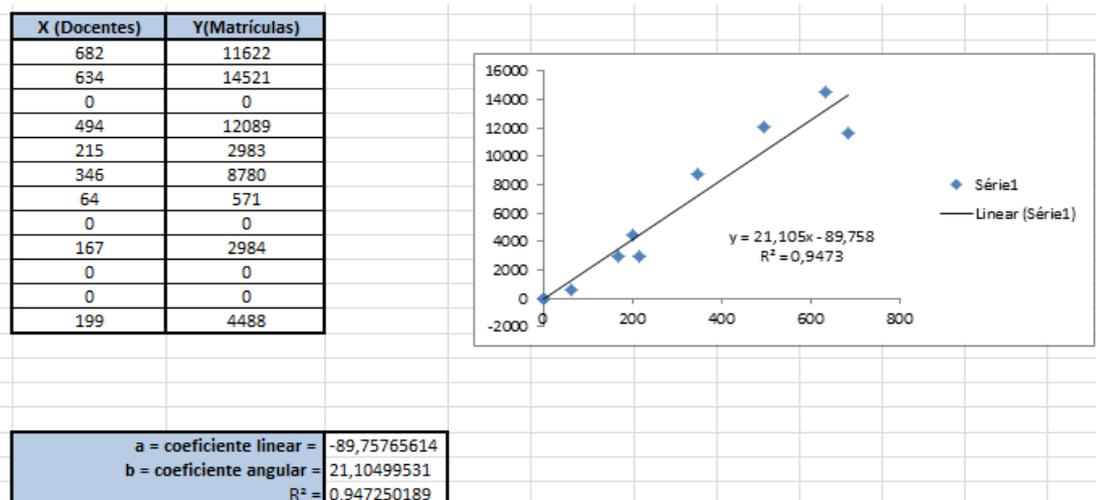
Fonte: IBGE (2012)

A figura acima apresenta o número de matrículas realizadas no município de Mossoró – RN no ano de 2012, como também os seus respectivos docentes que estão a disposição das escolas.

Com base nestes dados primários e nos conceitos discutidos neste artigo, buscaremos as possíveis correlações existentes no âmbito de educação básica no município estudado e verificamos a possibilidade de existir ou não uma regressão entre suas variáveis.

Dando início ao nosso estudo, tivemos o auxílio da planilha eletrônica para obtermos a reta de regressão linear simples. Como podemos observar na figura 2, as variáveis são diretamente proporcionais, ao aumentarmos ou diminuirmos o valor delas um certo número de vezes, o respectivo valor da outra grandeza igualmente aumenta ou diminui a mesma quantidade.

Figura 2: Análise de regressão linear simples.



Fonte: Dados da pesquisa (2014)

Para a obtenção da reta, primeiro calculamos o coeficiente linear “a” através da interceptação que significa o valor da média da distribuição de Y em X=0 e o coeficiente angular “b” através da inclinação, que expressa a taxa de mudança em Y, isto é, a mudança em Y quando ocorre a mudança de uma unidade em X. Assim, foi obtida a função $y = 21.104x - 89,758$.

O R² implica que 94,72% da variabilidade observada nos dados são explicadas pelo modelo de regressão. O quanto a variação de Y é reduzida pela inclusão de X no modelo.

A reta de regressão tem um alto grau de ajustamento; sendo que 94,19% da variação no número de matrículas pode ser explicada pelo número de docentes, medido pela R² ajustado (0,9419).

Figura 3: Tabela de estatística de regressão.

Estatística de regressão	
R múltiplo	0,973267789
R-Quadrado	0,947250189
R-quadrado aju	0,941975208
Erro padrão	1312,824523
Observações	12

Fonte: Dados da pesquisa (2014)

Ainda é importante constatar que os valores d P, F e t student obtidos nos indicam que o ajustamento da reta tem significância estatística com 95% de aceitação,

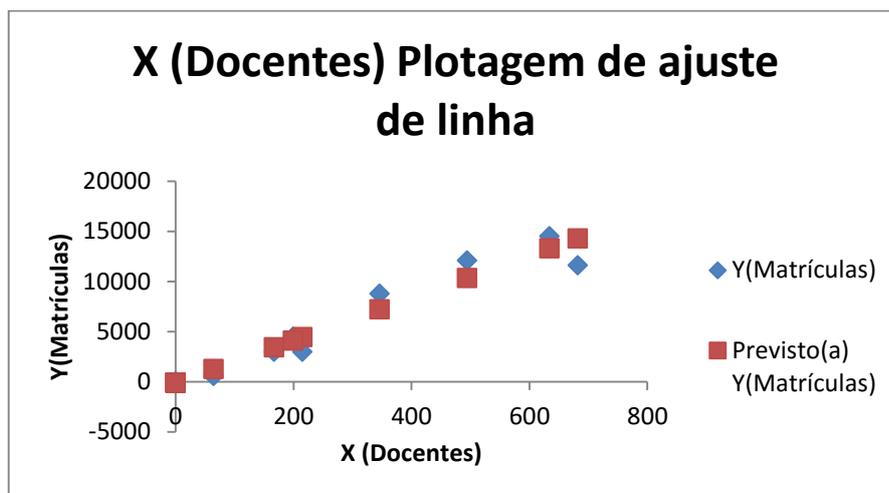
Figura 4: Tabela de análise

	Coeficientes	Erro padrão	Stat t	valor-P
Interseção	-89,75765614	527,9845832	-0,17	0,8684
X (Docentes)	21,10499531	1,574937643	13,40053	1,03E-07

Fonte: Dados da pesquisa (2014)

De onde podemos visualizar no seguinte gráfico e dispersão:

Figura 5: Gráfico de dispersão de ajuste de linha



Fonte: Dados da pesquisa (2014)

Ao nível de significância ($\alpha = 5\%$), encontramos $F^* = 179,57$ através da análise de variância do modelo – ANOVA, e na tabela F buscamos os valor do $F_{tabelado}$ que implica $F(0,05;1;10) = 4,96$. Com, isso rejeitamos a hipótese nula ao nível de 5% , já que $F^* > F_{tabelado}$, portanto existe regressão do número de matriculados sobre o número de docentes, ou seja, existe uma relação funcional linear entre Y e X.

Há diferenças significativas entre os grupos. Observa-se que MQG é muito superior a MQR, indicando uma forte variância entre os grupos.

Figura 6: Análise de variância do modelo - ANOVA

ANOVA					
	<i>gl</i>	<i>SQ</i>	<i>MQ</i>	<i>F</i>	<i>F de significação</i>
Regressão	1	309497506,7	3,09E+08	179,5741	1,02796E-07
Resíduo	10	17235082,29	1723508		
Total	11	326732589			

Fonte: Dados da pesquisa (2014)

Figura 7: Tabela com percentis 95% da distribuição F

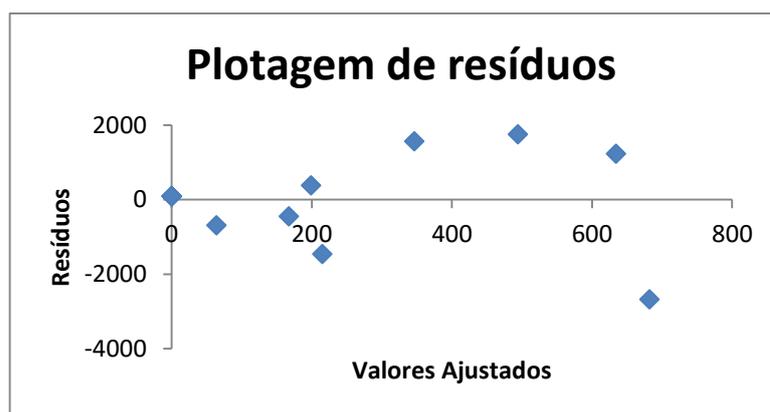
v1 →	1	2
v2 ↓		
1	161.45	199
2	18.51	19
3	10.13	9
4	7.71	6
5	6.61	5
6	5.99	5
7	5.59	4
8	5.32	4
9	5.12	4
10	4.96	4
11	4.84	3
12	4.75	3

Fonte: Dados da pesquisa (2014)

Os resíduos podem ser considerados como erros observados, e não são variáveis aleatórias independentes pois eles envolvem os valores Y, os quais são baseados na mesma equação de regressão.

A verificação de que a função de regressão é adequada aos dados pode ser feita através do gráfico dos resíduos versus valores ajustados. Com esse gráfico temos indícios sobre o comportamento da variância dos resíduos com relação aos valores ajustados. Na figura 8 temos um protótipo da situação em que um modelo de regressão linear é adequado. Observe que os resíduos se distribuem aleatoriamente em torno da média zero.

Figura 8: Gráfico de Resíduos versus Valores Ajustados.



Fonte: Dados da pesquisa (2014)

6. Considerações finais

O presente artigo teve objetivo exploratório, no sentido de fazer uma discussão conceitual sobre regressão e da correlação linear, e com base nos dados primários buscar identificar os

mecanismos de correlação existente no âmbito da educação básica do município de Mossoró – RN, especificamente, alunos e professores.

Podemos deste modo concluir que o número de docentes é diretamente proporcional ao número de alunos, pelo ajustamento linear da regressão; visto que houve uma regressão ao nível de 5%, porém existe uma diferença significativa entre os docentes, o que implica um professor para vários alunos.

Não foi objetivo do nosso estudo, identificar a questão social quanto a educação, mas podemos observar e é bem interessante para a análise dos governantes, a sobrecarga dos professores na pré escola, como consequência da falta de escolas para este nível. Ainda que, talvez seja o nível que o aluno precise de um acompanhamento mais próximo devido estar iniciando sua vida escolar e a sua pouca idade.

Referencial Bibliográfico

ACTION, Portal. **Análise de Resíduos**. 2000. Disponível em: <<http://www.portalaction.com.br/content/15-análise-de-resíduos>>. Acesso em: 02 jul. 2014

ANDERSON, David R.; SWEENEY, Dennis J.; et all. **Estatística Aplicada à Administração e Economia**. São Paulo: Cengage Learning, 2011.

Decreto nº 73177, de 20 de novembro de 1973. **Plano Nacional de Estatísticas Básicas e Ao Plano Geral de Informações Estatísticas e Geográficas**. Brasília, 1973.

FREUND, John E. **Estatística aplicada: economia, administração e contabilidade**. Porto Alegre: Bookman, 2006. 536p.

GIL, Antônio Carlos. **Métodos e técnicas de pesquisa social**. 4. ed. São Paulo: Atlas, 1995.

GUJARATI, D. N. **Econometria básica**. São Paulo: Makron Books, 2000.

IBGE. **O que é censo?** 2014. Disponível em: <<http://7a12.ibge.gov.br/sobre-o-ibge/o-que-e-censo>>. Acesso em: 26 maio 2014.

INEP. **Descrição do censo escolar**. 2011. Brasília. Disponível em: <<http://portal.inep.gov.br/basica-censo>>. Acesso em: 02 jun. 2014.

INEP. **O que é censo escolar?** 2011. Brasília. Disponível em: <<http://portal.inep.gov.br/basica-censo>>. Acesso em: 02 jun. 2014.

RIBEIRO, José Luis Duarte; CATEN, Carta ten. **Estatística Industrial**. Porto Alegre: UFRGS, 2010. 143 p.

TRIOLA, M. F. **Introdução à estatística**. Rio de Janeiro: Livros Técnicos Científicos, 1999.