

ANÁLISE BIDIMENSIONAL ESTATÍSTICA APLICADA TWO-DIMENSIONAL STATISTICAL ANALYSIS APPLIED

Sarley de Araújo Silva¹
Lourdiane Gonçalves de Castro²

RESUMO: O estudo e análise bidimensionais proporciona fazer comparativo de variáveis quantitativas em problemas de estatística. Teve como objetivo analisar e explorar múltiplas relações entre variáveis. Na realização do estudo foi utilizado o método estatístico, com foco quantitativo, com objetividade nas análises em variáveis independentes e dependentes. Com o software, que consistiu a participação dos educandos na qual foi desenvolvido o trabalho. Nas ações das atividades, submeteu-se explorar as análises através da: definição, construção de gráficos e análise de dispersão, com propósito de fazer mais significativa e presente a estatística aplicada, valorizando a construção dos conhecimentos. Nesse trabalho foi observado que os alunos entenderam a análise bivariadas com facilidade utilizando o recurso estatístico, pois desenvolveram ideias sobre a proposta colocada em contato direto com o método de estatística aplicada. Os resultados obtidos nos gráficos apresentados no decorrer deste estudo nos permitiram comparar variáveis em problemas de estatística proporcionando soluções claras e precisas para as questões mais complexas. Concluí-sé que a metodologia utilizada teve sucesso ao possibilitar ajustar os modelos multivariados.

Palavras-chaves: Análise multivariada. Regressão Linear Múltipla e Covariância.

ABSTRACT: The two-dimensional study and analysis makes it possible to compare quantitative variables in statistical problems. It aimed to analyze and explore multiple relationships between variables. In carrying out the study, the statistical method was used, with a quantitative focus, with objectivity in the analysis of independent and dependent variables. With the software, which consisted of the participation of the students in which the work was developed. In the actions of the activities, it was submitted to explore the analyzes through: definition, construction of graphs and analysis of dispersion, with the purpose of making applied statistics more meaningful and present, valuing the construction of knowledge. In this work, it was observed that the students easily understood the bivariate analysis using the statistical resource, as they developed ideas about the proposal placed in direct contact with the applied statistics method. The results obtained in the graphs presented during this study allowed us to compare variables in statistical problems, providing clear and precise solutions to the most complex questions. It was concluded that the methodology used was successful in making it possible to adjust the multivariate models.

Keywords: Multivariate analysis. Multiple Linear Regression and Covariance.

¹Mestre em Engenharia de Processos – PPGE/ITEC-UFPA. E-mail: sarley.silva@ifam.edu.br

² Graduada em Licenciatura em Pedagogia pela UNIP.

1 INTRODUÇÃO

As análises bidimensionais correlação, regressão linear múltipla e covariância correspondem análises de dados amostrais quantitativos de um conjunto de variáveis independentes e dependentes relacionadas a várias variáveis. Trabalha-se aplicação de métodos estatísticos com Minitab, sendo o mesmo uma ferramenta poderosa na estatística aplicada e possível mediante comandos opcionais analisar gráficos e verificar seu grau de dispersão. A estatística multivariada corresponde a um conjunto de técnicas utilizadas para estudar diversas variáveis simultaneamente, com o objetivo de sintetizar ou simplificar a estrutura de variabilidade dos dados (Mingoti, 2017). A análise multivariada é um conjunto de técnicas estatísticas que permite a análise e interpretação de conjuntos de dados de natureza quantitativa com grande número de variáveis de forma simplificada (NEISSE; HONGYU, 2016).

As análises bidimensionais são análises que tem como objetivo analisar o comportamento de variáveis. Dentre várias análises analisaremos três análises bidimensional: Correlação, Regressão linear múltipla e covariância.

Através das análises multivariadas com software pelo método estatístico quantitativo a partir da definição, análises de gráfico e análise de dispersão, trabalhou-se a percepção dos alunos do 3º período do curso superior do Instituto Federal Ifam localizado em Manaus. A atividade procura estimular a execução do programa software em análises bidimensionais com o intuito de promover o raciocínio cognitivo e o desenvolvimento de programação do aluno.

Acredita-se que as análises bidimensionais: correlação, regressão linear múltipla e correlação, tornam a aprendizagem significativa entre aluno e programa estatístico, criando situações estimulantes, proporcionando a autonomia intelectual do aluno.

2 FUNDAMENTAÇÃO TEÓRICA

A história da correlação está associada ao nome de Francis Galton na segunda metade do século XIX, foi quem usou pela primeira vez os termos correlação e regressão. Publicou em 1869 o livro *Hereditary Genius*, sobre teoria de regressão (SCHULTZ, 2002,p.439). Galton adotou o termo regressão quando observou que filhos de homens altos não são, em média, tão altos quanto os pais, mais os filhos de homens baixos são em média, mais altos do que os pais. Deve-se a Galton a forma gráfica de representar as

propriedades básicas do coeficiente de correlação. Karl Pearson em contribuições para a teoria matemática da evolução para regressão e no caso de um sistema das variáveis hereditariedades e Panmixia. (PEARSON, 1896, p.187,253-318). Contribuição matemáticas á teoria da evolução, VII sobre a correlação de caracteres não quantitativamente mensuráveis. Pearson, K transações filosóficas da sociedade real de Londres. (PEARSON, 1896, p.1-405). Karl Pearson desenvolveu um método de análise de variância, com a análise de covariância. (PEARSON, 1901, p.559-572). Análises bidimensionais é a análise que analisar o comportamento conjunto de duas ou mais variáveis quantitativas com a finalidade de encontrar possíveis relações entre elas. (BUSSAB, 1987. P.49-63).

Os conceitos das análises bidimensionais exposto a seguir tem a finalidade de informar de quais elas surgiram e como.

Tipos de análises bidimensionais

Os Tipos de análises bidimensionais são: Correlação, Regressão linear múltipla e Covariância.

Correlação: É a medida de relação entre duas variáveis, que determinada numericamente por meio dos coeficientes de correlação que representam o grau de associação entre duas variáveis quantitativas continua.

Regressão linear múltipla: É uma metodologia estatística de previsão de valores de uma ou mais variáveis de resposta dependentes, através de um conjunto de variáveis explicativas independentes.

Covariância: É uma medida que busca avaliar o modo com que duas variáveis quaisquer se inter-relacionam de maneira linear.

METODOLOGIA

O estudo das análises bidimensionais: Correlação, regressão linear múltipla e covariância, foi feita através do método estatístico com os dados quantitativos que permite comprovar as relações entre suas variáveis relacionadas, e obter seu grau de dispersão, as variáveis são as particularidades que podem ser analisadas em uma amostra de uma população, sendo as variáveis um conjunto de resultados possíveis de uma análise estatística bidimensional. Tendo como finalidade estudar as relações entre variáveis e verificar seu grau de distanciamento na reta de dispersão.

O papel do método estatístico é, antes de tudo, fornecer uma descrição quantitativa da sociedade, considerada como um todo organizado. (LAKATUS, 2003,p.108). O método estatístico tem o objetivo coletar, analisar e interpretar variáveis quantitativas contínuas de uma amostra populacional e pela análise bidimensional analisar o comportamento conjunto de duas ou mais variáveis aleatórias.

A construção das análises bidimensional ocorreu através do software e comandos que seguiram regra da definição de análise estatística, essa atividade é voltada para os alunos do 3º período do superior.

Na construção dos gráficos de análises das variáveis independentes e dependentes foram utilizadas as seguintes sequenciais no Minitab: selecionar opção abrir, digitalizar dados na planilha e escolher opções de gráficos. O estudo desenvolveu-se através da construção das análises multivariada com software pelo método estatístico quantitativo tendo como apoio os dados de variáveis quantitativas contínuas. (CRESPO, 2002. P.17). A análise bidimensional na construção de gráficos teve como ferramenta quantitativa a finalidade de estudar, definir, organizar, analisar e interpretar gráficos.

Essas atividades foram direcionadas para os alunos do 3º período do curso superior do Instituto Federal Ifam em Manaus, envolveu-se no trabalho um programador e orientador na construção de análises bidimensional: correlação múltipla, regressão múltipla e covariância.

Correlação

Definição: É uma medida do grau de associação linear entre duas variáveis. (CRESPO, 2002. P.151).

Coefficiente de correlação de Pearson

$$r = \frac{n \sum X_i Y_i - (\sum X_i) (\sum Y_i)}{\sqrt{[n \sum X_i^2 - (\sum X_i)^2][n \sum Y_i^2 - (\sum Y_i)^2]}} \quad (2.1)$$

O coeficiente de correlação varia entre -1 e +1, inclusive, isto é, $-1 \leq r \leq +1$. Então a correlação se interpreta o valor do coeficiente de correlação:

$r = 1$: correlação perfeita e positiva entre as variáveis.

$r = -1$: correção perfeita e negativa.

$r = 0$: correlação nula.

$0 < r < 1$: correlação positiva

$-1 < r < 0$: correlação negativa

O seguinte quadro fornece um guia de como podemos descrever uma correlação em palavras dado o valor numérico.

De acordo com a Tabela 2.1, são apresentadas regras de correlação.

Se $r = 1$	Correlação positiva perfeita
Se $r = 0,75$	Correlação positiva forte
Se $r = 0,50$	Correlação positiva média
Se $r = 0,25$	Correlação positiva fraca
Se $r = 0$	Correlação linear inexistente
Se $r = - 0,25$	Correlação negativa fraca
Se $r = - 0,50$	Correlação negativa média
Se $r = - 0,75$	Correlação negativa forte
Se $r = -1$	Correlação negativa perfeita

Ajustamento da reta por dispersão da correlação retilínea de duas variáveis, uma variável X independente e outra variável Y dependente.

Ajustamento:

$$Y = a X + b \tag{2.2}$$

Parâmetros a e b.

$$a = \frac{n \sum X_i Y_i - (\sum X_i) (\sum Y_i)}{n \sum X_i^2 - (\sum X_i)^2} \tag{2.3}$$

$$b = \bar{y} - \bar{x} \tag{2.4}$$

n é o número de observações.

\bar{x} é a média dos valores. $X_i \left(\bar{x} = \frac{\sum X_i}{n} \right) \tag{2.5}$

\bar{y} é a média dos valores. $X_i \left(\bar{y} = \frac{\sum \bar{y}_i}{n} \right) \tag{2.6}$

oi) Uma das relações causais mais estudadas na ciência política tenta entender se e como o desempenho da economia afeta a aprovação presidencial em um determinado ano. Considere os seguintes dados referentes ao Brasil para responder aos itens:

Conforme a Tabela 2.2, é possível observar dados referente ao PIB.

Presidente	Ano	Crescimento PIB em %	Aprovação do presidente
Lula	2003	1,3	42,8
Lula	2004	5,7	39,3
Lula	2005	3,2	32,1
Lula	2006	4,0	45,2
Lula	2007	6,1	48,6
Lula	2008	5,2	63,0
Lula	2009	-0,2	68,0
Lula	2010	7,6	78,0
Dilma	2011	3,9	48,0
Dilma	2012	1,8	61,6
Dilma	2013	2,7	40,5
Dilma	2014	-0,1	37,5
Dilma	2015	-3,8	12,3
Temer	2016	-3,6	14,0
Temer	2017	1	6,0

Aprovação do presidente: média anual das pesquisas Data folha com os entrevistados que responderam Ótimo/Bom Crescimento do PIB: IBGE

- a) Considerando o PIB como a variável independente que explica a aprovação presidencial, calcule a covariância e o coeficiente de correlação entre as variáveis X e Y.

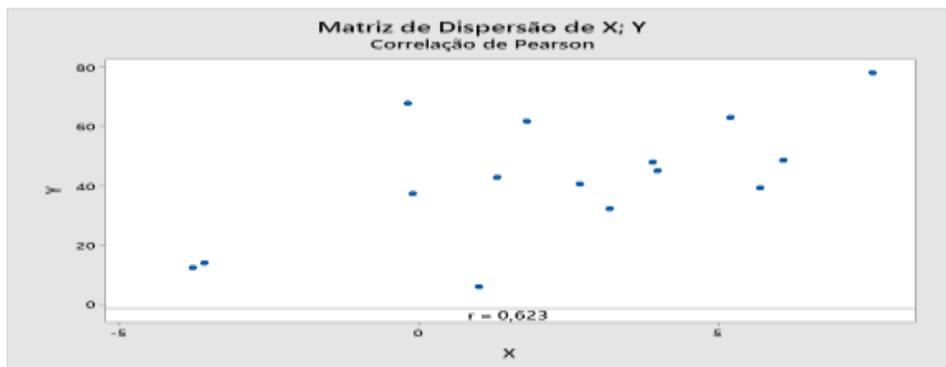


Fig 1- Correlação, SILVA, 2022.

Covariâncias

	X	Y
X	11,1346	
Y	42,9544	426,7369

Os dados calculados de covariância e coeficiente de correlação indicam um grau de correlação entre as variáveis X e Y significativa a correlação positiva entre as variáveis (0,62).

02). Os dados a seguir correspondem à variável renda familiar e gasto com alimentação (em unidades monetárias) para uma amostra de 25 famílias.

De acordo com a Tabela 2.3, podemos informar os dados a seguir correspondem à variável renda familiar e gasto com alimentação.

Renda Familiar (X)	Gasto com Alimentação (Y)
3	1,5
5	2,0
10	6,0
10	7,0
20	10,0
20	12,0
20	15,0
30	8,0
40	10,0
50	20,0
60	20,0
70	25,0
70	30,0
80	25,0
100	40,0
100	35,0
100	40,0
120	30,0
120	40,0
140	40,0
150	50,0
180	40,0
180	50,0
200	60,0
200	50,0

Fonte: USP (2018).

- a) Construa o diagrama de dispersão da variável gasto com alimentação (Y) em função da renda familiar (X).

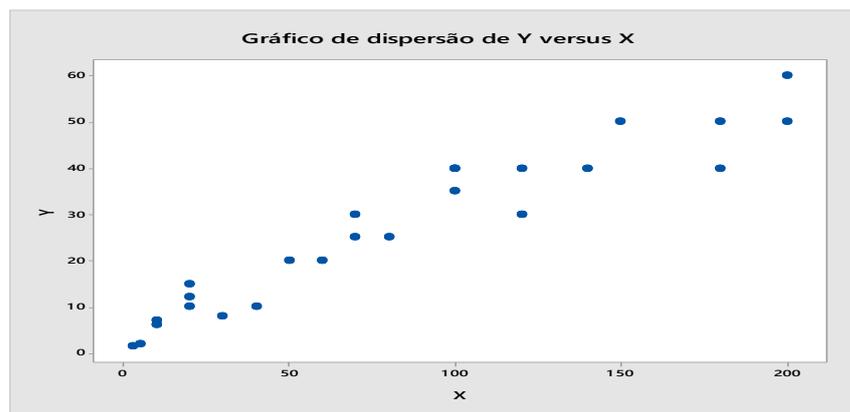


Fig 2 - Correlação, SILVA, 2022.

O gráfico mostra que as variáveis renda familiar x e Gastos com alimentação y apresenta correlação entre si. Quanto maior a renda familiar, maiores são os gastos com alimentação.

b) Calcular o coeficiente de correlação entre essas variáveis.

O coeficiente de correlação de Pearson é igual a 0,954, isso significa que a forte correlação positiva e que a reta está ajustada.

Regressão Linear Múltipla.

Definição: É o estudo de como a variável dependente y se relaciona com duas ou mais variáveis independentes (LARSON, 2002. P.361).

Equação de regressão múltipla.

$$E(Y) = \beta_0 + m_1 X_1 + m_2 X_2 + m_3 X_3 + \dots + m_k X_k \quad (2.7)$$

Práticas de Estatística.

Executar comandos: abrir planilha do Minitab, digitar dados, executar operações em variáveis quantitativas, calculando medidas usuais: Regressão linear múltipla. Selecionar opções comandos, regressão, ajustar o modelo de regressão, resposta em y, selecionar variáveis preditores contínuos C1_y, C2_{X1}, C3_{X2} e ok.

03). Um pesquisador deseja determinar como estão determinados os salários dos empregados de uma determinada campanha com a duração do emprego, a experiência prévia e a educação. O pesquisador então seleciona oito funcionários da empresa e obtém os seguintes dados.

Segundo a Tabela 2.4, é possível verificar dados de salário, emprego, experiência e educação.

Empregado Y	Salário	Emprego	Experiência	Educação
A	37.310	10	2	16
B	37.380	5	6	16
C	34.135	3	1	12
D	36.985	6	5	14
E	38.715	8	8	16
F	40.620	20	0	12
G	39.200	8	4	18
H	40.320	14	6	17

Fonte: UNICAMP (2018).

(a) Encontre o valor dos coeficientes das variáveis e a equação da regressão.

Coeficientes

Equação de Regressão

$$w = 35153 + 316,8 k$$

256

Coeficientes

Termo	EP de		Valor-T	Valor-P	VIF
	Coef	Coef			
Constante	35153	940	37,38	0,000	
k	316,8	89,0	3,56	0,012	1,00

Análise de Variância

Fonte	GL	SQ (Aj.)	QM (Aj.)	Valor F	Valor-P
Regressão	1	21022183	21022183	12,68	0,012
k	1	21022183	21022183	12,68	0,012
Erro	6	9949514	1658252		
Falta de ajuste	5	9831901	1966380	16,72	0,183
Erro puro	1	117613	117613	*	*
Total	7	30971697			

O gráfico de linha ajustada mostra os resultados de regressão.

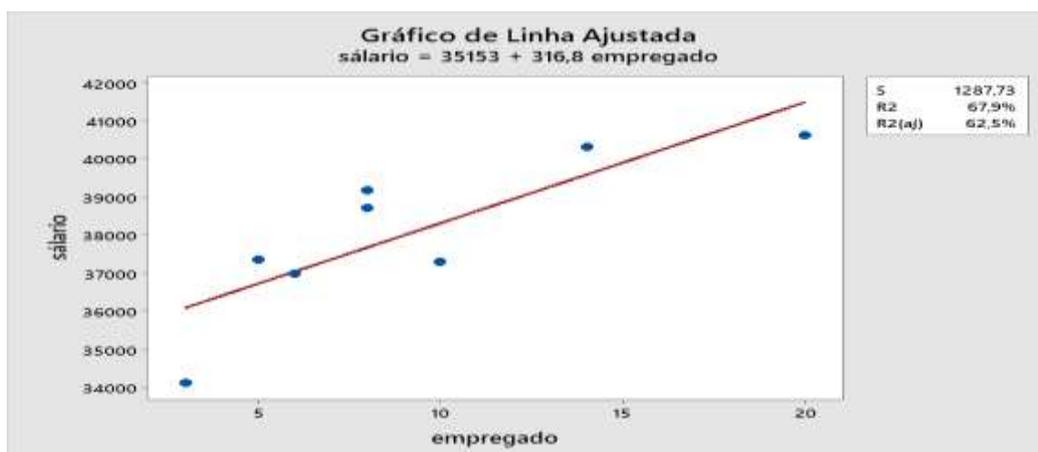


Fig 3- Gráfico de Regressão Linear, SILVA, 2022.

Conclui-se que os resultados indicam que os pontos estão ajustados entre a reta e que os valores para p valor são menores do que o nível de significância de 0,05. Indicando efeito estatisticamente significativo.

Covariância

Definição: É a covariância como medidas descritivas da relação entre duas variáveis (DENNIS, J. Sweeney, 2015. p.123).

Formula de covariância amostral.

$$\sigma_{xy} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{n - 1} \quad (2.8)$$

A fórmula de covariância emparelhada cada x_i com um y_i soma-se os produtos obtidos ao multiplicarmos o desvio de cada x_i tem em sua média amostral \bar{x} pelo desvio que o y_i correspondente em sua média amostral \bar{y} essa soma é então dividida por $n - 1$.

Práticas de Estatística.

Seguir sequencia de manuseio: Abrir planilha do Minitab, digitar dados, executar operações em variáveis quantitativas, calculando medidas usuais: Covariância. Selecionar opções stat, estatística básica, covariância, selecionar variáveis X e Y, opção ok, gráfico de dispersão simples, simples e ok.

o4) Dada uma população e extraídas as seguintes amostras de: $A = (2,5,1,3,4,1,5,3,4,2)$ e $B = (50,57,41,54,54,38,63,48,59,46)$. Verifique sua covariância.

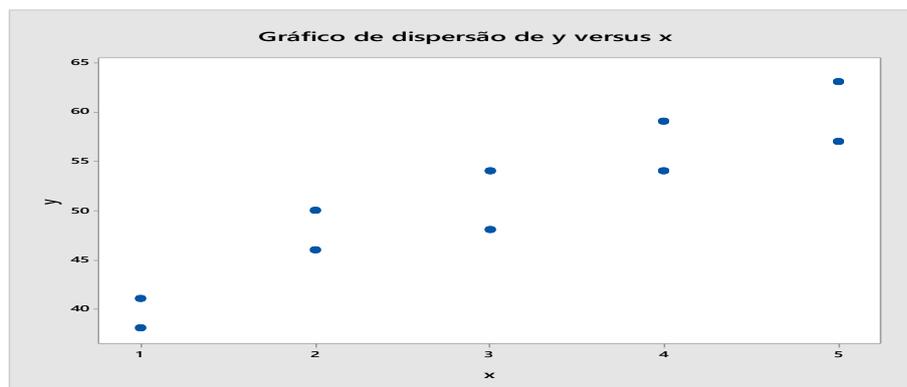


Fig 2- Covariância, SILVA, 2022.

A covariância mostra a tendência na relação linear entre as variáveis, sendo igual a 11,00 e sua reta está ajustada em seus pontos e suas variáveis estão correlacionadas.

3 RESULTADOS E DISCUSSÕES

Através das análises bidimensionais com Minitab e construção de gráficos, foi trabalhada a maneira de perceber acompanhada de recursos cognitivos (conceito, coleta, organização, descrição, análise, interpretação de dados e aprendizagem em grupo). Participaram do estudo alunos do 3º período do curso superior do Instituto Federal Ifam em Manaus. A atividade foi desenvolvida em duas etapas a primeira etapa execução de comandos estatísticos, execução, construção de gráficos, exploração do conteúdo e da metodologia a ser utilizada.

A segunda etapa a turma foi distribuída no laboratório de estatística cada aluno de posse do computador com o programa Minitab, seguindo as sequências seguintes: Abrir programa, digitar dados na planilha, escolher opções estatísticas e construção de gráficos. Foi possível os alunos através da execução do programa estatístico e análise de gráficos, cada uma aplicando seus conhecimentos, e trabalhando a definição para análise bidimensional, verificou-se que pela definição e utilizando comandos do programa, estabelecendo opção que seguem a definição torna-se simples a construção de gráficos, houve a inter-relação professor-aluno e um interesse maior com relação a disciplina de estatística aplicada. Antes da aplicação do projeto foram feitas as seguintes perguntas para os alunos: O que são análise bidimensional? Conceito de correlação e a

história das análises bidimensionais e seus gráficos estão associado a qual nome? Quem mais desenvolveu o estudo? Como verificação do conhecimento dos alunos em relação ao conteúdo já ministrado pelo professor responsável pela turma, verificou-se que dos 40 alunos, somente 20 responderam alguma informação sendo que com imprecisão, depois do trabalho executado todos responderam, passaram a ter melhor compreensão trabalhando com as análises bidimensionais. Tiveram melhor esclarecimento em relação aos seus gráficos, verificaram a importância do conceito em relação às análises. Segundo o professor responsável a metodologia utilizada contribuiu muito com a turma, passaram a ter melhor desempenho. Foi confirmado que é possível construir gráficos de análises bidimensionais, a partir da definição e comandos estatísticos do Minitab mediada pela aprendizagem significativa.

Trabalhando conceito, execução e análise de dispersão em gráficos constataram que realmente é possível construir gráficos e fazer análise pelo método estatístico quantitativo a partir da definição, atuaram de forma ativa e coletiva na construção do conhecimento. As análises auxiliam na compreensão de conceitos bidimensionais, bem como desenvolver habilidades cognitivas para executar procedimentos estatísticos em programa, estimulam a criatividade do aluno, quando é dada a ele a oportunidade de executar programa e criar as análises bidimensionais estatísticas.

CONSIDERAÇÕES

A partir das aplicações nas análises multivariada foi possível obter equações para a estimativas das variáveis independentes e dependentes relacionadas a aplicações estatísticas. Sabe-se que o ensino-aprendizado de estatística e probabilidade não pode ser restringido por análises bidimensionais com software, todavia sempre que possível, devemos apropriar-se da poderosa ferramenta estatística, para tirar o aluno da condição de espectador passivo e provocar aprendizagem significativa que mobilize o aluno e estabeleça entre ele e o objeto do conhecimento uma relação interativa. Os resultados obtidos permitiu concluir que as análises multivariadas são eficazes no estudo das relação entre as variáveis em aplicações práticas de estatísticas.

O trabalho foi de suma importância, pois, os alunos aprenderam a utilizar os conceitos de análises em estatística aplicada e a partir de opções construir gráficos e

analisá-lo, trabalharam em equipe e participaram na prática. Sendo assim o educador em estatística não deve colocar obstáculo em aceitar o desafio de procurar trabalhar análise bidimensionais com software e desenvolver atividades que possam despertar o interesse dos alunos e aumentar a participação durante as aulas de estatística aplicada. Portanto, as principais contribuições do presente trabalho foram a proposta de análise multivariadas, assim como a comparação das variáveis independentes e dependentes pelos cálculos de regressão múltipla, correlação e regressão, os quais possibilitaram analisar e comparar famílias de variáveis.

REFERÊNCIAS BIBLIOGRÁFICAS

BUSSAB, Wilton. **Estatística Básica**. 4^o ed. São Paulo: Atual, 1987.

CRESPO, Arnot. **Estatística Fácil**. 17^o ed. São Paulo: Saraiva, 2002.

DENNIS, J. Sweeney. **Estatística Aplicada á Administração e Economia**. 3^o ed. São Paulo: Cengage learning, 2015.

SCHULTZ, Duane P.; SCHULTZ, Sydney Ellen. **História da psicologia moderna**. 16. ed. São Paulo: Cultrix, 1992. 439 p.

LAKATUS, E. M.; MARCONI, M. A. **Fundamento de metodologia científica**. 5^o ed. São Paulo: Atlas, 2003.

MINGOTI, S. A. (2017). **Análise de dados através de métodos de estatística multivariada – uma abordagem aplicada**. In *Análise de dados através de métodos estatística multivariada: uma abordagem aplicada*.

NEISSE, A. C.; HONGYU, K. Aplicação de componentes principais e análise fatorial a dados criminais de 26 estados dos EUA. *E&S Engineering and Science*, v.6, n.2, 2016.

PEARSON, K. **Regression, heredity and panmixi**, *Philosophical Transactions of the Royal Society of London*, A, London, v. 187, p. 253-318, 1896.

PEARSON, K *Philosophical Transactions of the Royal Society of London. Mathematical Contributions to the Theory of Evolution*. VII. On the Correlation of Characters Not Quantitatively Measurable. Series A, Containing Papers of a Math. or Phys. Character (1896-1934). 1900-01-01. 195:1-405.

PERASON, Karl (1901). **On Lines and Planes of Closest Fit to Systems of Points in Space**. *Philosophical Magazine*, 6th Series, Vol. II, pp. 559-572.

LARSON, Faber. **Estatística Aplicada**. 2^o ed. São Paulo: Pearson Prentice Hall, 2004.

SCHULTZ, Duane P.; SCHULTZ, Sydney Ellen. **História da psicologia moderna**. 16. ed. São Paulo: Cultrix, 1992. 439 p.

https://edisciplinas.usp.br/pluginfile.php/5642228/mod_resource/content/1/Aula_03.pdf

https://www4.eco.unicamp.br/docentes/gori/images/arquivos/EconometriaI/Econometria_RevisaoRegressaoLinear.pdf

https://edisciplinas.usp.br/pluginfile.php/4336593/mod_resource/content/1/Lista_10_2018.pdf

[1] Mestre em Engenharia de Processos – PPGEP/ITEC-UFPA.