UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ

Ministério da Educação
Universidade Tecnológica Federal do Paraná
Campus de Ponta Grossa

Programa de Pós-Graduação em Ensino de Ciência e Tecnologia

SEQUÊNCIA DE ENSINO: UMA PROPOSTA DE ENSINO CONTEXTUALIZADO DE CORRELAÇÃO E REGRESSAO LINEAR PARA UM CURSO DE ENGENHARIA DE COMPUTAÇÃO

Sabrina Anne de Lima
Guataçara dos Santos Junior

PONTA GROSSA
DEZEMBRO, 2014

FICHA CATALOGRÁFICA:Não esquecer de solicitar
ISBN

LISTA DE FIGURAS

LISTA DE QUADROS

LISTA DE GRÁFICOS

LISTA DE TABELAS

SUMÁRIO

1 INTRODUÇÃO

A presente sequência de ensino tem como publico alvo professores de Estatística de cursos de Engenharia de Computação que busquem a aquisição de conhecimento por parte dos alunos através de pressupostos de contextualização.

Visa-se com isso possibilitar aos alunos aprender através de situações especificas de seu curso, mostrando assim a aplicabilidade de conteúdos estatísticos em uma realidade profissional dos estudantes.

Este fato se concretiza ao observar que diversos métodos de engenharia utilizam técnicas estatísticas com a finalidade de tomar decisões que sejam capazes de solucionar problemas ou planejar produtos, serviços ou processos. (LARSON & FARBER, 2010).

Segundo Ara (2000) a prática geral aplicada para o ensino de Estatística em cursos de Engenharia, considera somente a fundamentação teórica seguida da resolução de exercícios que, nem sempre despertam o interesse dos alunos na aprendizagem desta disciplina.

Isso acontece porque, apesar de ser um tema importante para a Engenharia, em muitos casos os alunos não demonstram interesse ou não percebem a relevância e aplicabilidade dos conceitos aprendidos em sala de aula com o seu contexto e cotidiano profissional. (MONTGOMERY & RUNGER, 2009)

Acredita-se que estratégias de ensino contextualizadas possam garantir aos alunos uma melhor visão da Estatística.

Nesta sequencia de ensino, tratar-se-á especificamente de conteúdos de Análise de Correlação e Regressão Linear contextualizados no processo de obtenção do tempo de ordenação de um algoritmo, considerando-se o tamanho do vetor selecionado.

Para isso foram utilizados dois métodos de ordenação: o Método de Bubble Sort (ou Método Bolha) e o Método de Merg Sort.

O trabalho com os alunos foi realizado em duplas, e como se tratava de um curso de Engenharia de Computação, em diversas situações o método manual de alguns cálculos foi comparado (ou substituído) pelo uso do Excel.

Assim sendo, esta sequencia de ensino tem como objetivo auxiliar os professores no ensino de estatística em conteúdos de correlação e regressão linear, buscando favorecer uma aprendizagem contextualizada a partir de uma situação específica de um curso de Engenharia de Computação.

Esta proposta pedagógica será feita em quatro capítulos, sendo o primeiro este apresentado como Introdução onde se apontam os objetivos deste trabalho.

No Capitulo 2, é apresentado um referencial teórico, com uma abordagem geral sobre o ensino contextualizado e, em seguida, conteúdos específicos de Correlação e Regressão Linear que foi utilizada, baseada em dois principais obras literárias: Larson e Farber (2010) e Triola (2005) que foram utilizados como base para o desenvolvimento das atividades propostas.

No Capítulo 3, é mostrada a estrutura da Sequência de Ensino, proposta em quatro etapas: proposição do assunto e escolha do tema, coleta dos dados, desenvolvimento da atividade proposta e síntese dos conteúdos trabalhados.

2 REFERENCIAL TEÓRICO

2.1 CONTEXTUALIZAÇÃO

De acordo com Tafner (2003) quando um professor se dispõe a trabalhar com situações com as quais o estudante é capaz de se identificar, consegue condições de aprendizagem que podem ser consideradas fundamentais. Porém, para que isso ocorra, é preciso que o professor conheça, antecipadamente, a realidade na qual os alunos estão inseridos, quais estratégias pretende usar e o preparo do docente para atingir níveis que sejam compatíveis com a realidade esperada pelos alunos.

Contextualizar é a forma de apresentar, em sala de aula, situações que possibilitem ao aluno encontrar sentido naquilo que está aprendendo e com isso deseje aprender. Com isso é possível ampliar a construção de significados na aprendizagem, por parte dos alunos, tornando a contextualização um fator motivador para a aprendizagem. (VASCONCELOS, 2008).

Na atual conjuntura, a contextualização tem apresentado um papel importante no que diz respeito ao ensino. Para Tufano (2001), a contextualização de uma situação de aprendizagem, favorece o surgimento de ideias, criando um ambiente propício à construção de novos saberes.

Quando se fala de conteúdos matemáticos, especificamente, contextualizar está relacionado à atribuição de significados aos conteúdos matemáticos. (BRASIL, 2010). Porém esta ideia não pode ser reduzida, simplesmente aos aspectos utilitários desta disciplina. É necessário considerar a possibilidade de construção de significados que sejam originados das próprias questões matemáticas, não fosse isso, muitos conteúdos correm o risco de serem simplesmente descartados por não apresentar “utilidade”. (VASCONCELOS & REGO, 2010).

O contexto, de acordo com Brosseau (1996), deve estar associado a uma determinada situação que coloque sentido àquilo que se pretende aprender, de modo a orientar para a aprendizagem matemática, apresentando o conteúdo ao aluno a partir de uma situação problematizadora que seja conjugada com uma situação real na qual o aluno esteja inserido. Isso sugere ao aluno a necessidade de comunicar algo, gerando a necessidade de representar uma situação especifica discutir sobre ela e o que esta circunstância envolve.

Fonseca (1995) defende a necessidade de se contextualizar o conhecimento matemático, buscar suas origens, verificar sua evolução, definir sua finalidade e seu papel na transformação da realidade vivida pelo aluno. Isso não significa desprezar as técnicas matemáticas, mas ampliar os horizontes de significados da aprendizagem para que o conhecimento faça sentido na realidade do estudante.

Assim sendo, pelo exposto até aqui, pode-se entender a contextualização como a prática pedagógica que visa atribuir sentido àquilo que se pretende ensinar, baseado no conhecimento da realidade do aluno, levando em consideração os aspectos específicos do conteúdo a ser ensinado.

Dessa forma, acredita-se que o ensino contextualizado pode gerar efeitos positivos na aprendizagem dos alunos, motivando-os e fazendo com que os conceitos a serem aprendidos, deixem de ser abstratos e passem a ser parte da realidade do discente.

2.2 CORRELAÇÃO E REGRESSÃO LINEAR

2.2.1 Correlação

Dentro da estatística existe um grupo de procedimentos cujo objetivo é estudar e analisar a ligação entre variáveis aleatórias. Neste caso podemos citar o princípio da correlação e da regressão como uma das teorias mais difundidas.(NAGHTETTINI, PINTO,2007).

Larson e Farber (2010) definem correlação da seguinte forma: "Uma **correlação bivariada** é uma relação entre duas variáveis. Os dados podem ser apresentados por pares ordenados (**x,y**), onde **x** é a **variável independente** (ou **explanatória**) e **y** é a **variável dependente** (ou **resposta**)". (grifo do autor) (p. 395).

De acordo com Triola (2005, p.235) diz que: "Existe uma correlação entre duas variáveis quando uma delas está, de alguma forma, relacionada com a outra".

Para melhor entendermos, consideremos o exemplo 1 a seguir:

"Um estudo foi feito para verificar se existe uma relação entre o número de faltas de alunos de uma determinada turma e a nota final em certa disciplina. Os dados obtidos estão dispostos no quadro 1 a seguir:

Quadro 1 - Faltas e Notas dos alunos

Faltas (x)	**Nota Final (y)**
8	78
2	92
5	90
12	58
15	43
9	74
6	81

Fonte - Adaptado de Lima Filho, 2005,

Desta forma, para que se possa verificar se há alguma correlação entre as variáveis x e y, podemos considerar os pressupostos da análise de correlação.

Triola (2005) afirma que a importância de tal determinação está no sentido de que ao concluir a presença de correlação entre as variáveis é possível estimar as notas finais, baseando-se nas faltas dos alunos.

Para tal, pelo fato de se trabalhar com dados amostrais, é necessário formular inferências sobre as populações, supondo:

a) A amostra dos dados em x e y é aleatória;
b) Os pares ordenados (x e y) seguem a distribuição normal bivariada (em forma de sino)

Quando temos um gráfico de pares ordenados (x,y), em estatística chamamos de diagrama de dispersão. Neste caso, cada par ordenado é plotado em um plano cartesiano. A variável x (independente ou explanatória) é medida na direção horizontal e a variável y (dependente ou resposta) é medida na direção vertical. Desta forma é possível perceber se, em um diagrama de dispersão existe uma correlação linear, isto é, se o alinhamento dos pontos se aproximam de uma reta. (LARSON, FARBER, 2010).

Podemos observar alguns de diagrama de dispersão na figura 1 a seguir:

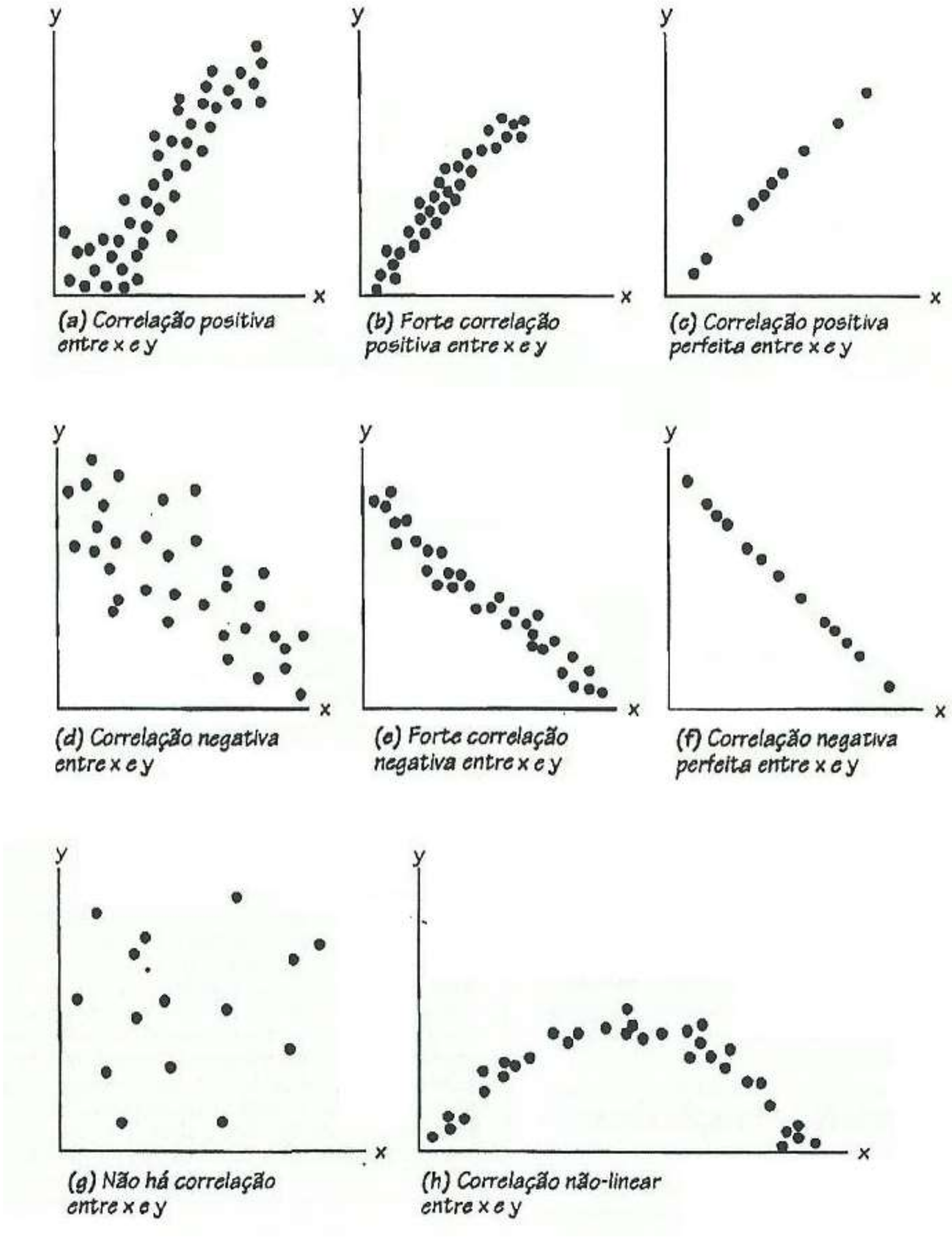
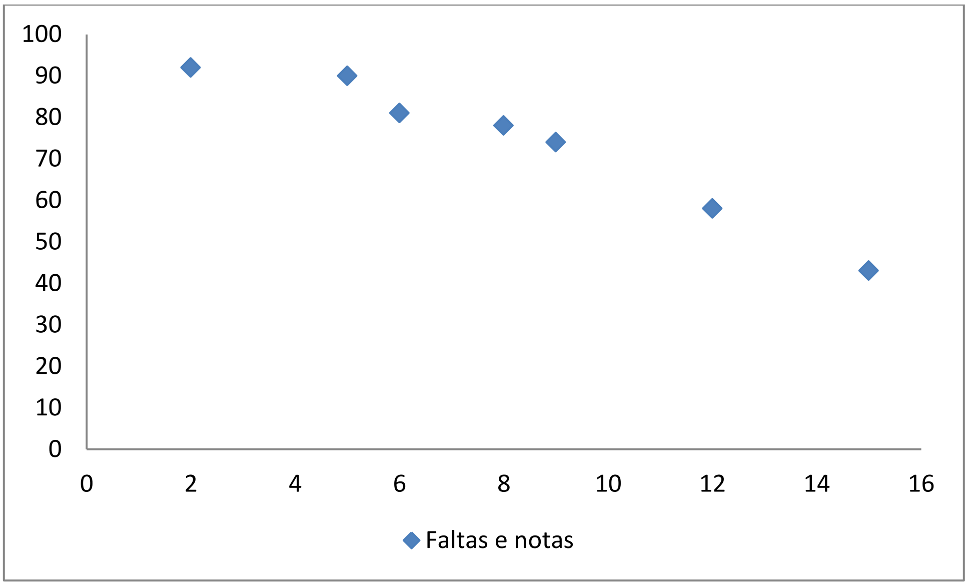


Figura 1- Diagramas de dispersão
Fonte – Triola (2005)

No caso do exemplo dado no quadro 1 temos o seguinte diagrama de dispersão:

Gráfico 1- Faltas e notas dos alunos



Fonte – A própria autora

Porém, analisar somente o diagrama de dispersão pode causar certa subjetividade para que se tire conclusões a respeito de uma suposta correlação. Assim sendo, é necessário um método mais eficiente, preciso e objetivo.

Neste caso, utiliza-se o coeficiente de correlação linear, a fim de verificar padrões de linearidade entre as variáveis.

Triola (2005, p.236) define coeficiente de correlação da seguinte maneira:

> "O **coeficiente de correlação linear** *r* mede o grau de relacionamento linear entre os valores emparelhados *x* e *y* de uma *amostra*. (...) [O coeficiente de correlação linear é chamado às vezes **coeficiente de correlação momento-produto de Pearson**, em homenagem a Karl Pearson (1857-1936) que o estabeleceu]".

O valor numérico do coeficiente de correlação linear (r) é dado pela equação 1 a seguir:

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

Equação 1- Coeficiente de Correlação Linear de Pearson
Fonte: Adaptado de Triola (2005)

O cálculo deste coeficiente é feito com base em dados amostrais, de modo que pode ser utilizado para medir o nível (ou grau) de correlação linear entre duas variáveis x e y. (TRIOLA, 2005).

O quadro 2 a seguir, é a interpretação dos elementos da fórmula do coeficiente de correlação.

Quadro 2 - Interpretação dos elementos da fórmula 1

Elemento	Interpretação
n	Número de pares de dados presentes
$\sum$	Denota a dição dos itens indicados
$\sum x$	Denota a soma de todos os valores de x
$\sum y$	Denota a soma de todos os valores de y
$\sum x^2$	Indica que devemos elevar ao quadrado cada valor de x e somar os resultados
$(\sum x)^2$	Indica que devemos somar os elementos de x e elevar o total ao quadrado. É sumamente importante não confundir $(\sum x)^2$ com $\sum x^2$.
$\sum y^2$	Indica que devemos elevar ao quadrado cada valor de y e somar os resultados
$(\sum y)^2$	Indica que devemos somar os elementos de y e elevar o total ao quadrado. É sumamente importante não confundir $(\sum y)^2$ com $\sum y^2$.
$\sum xy$	Indica que devemos multiplicar cada valor de x pelo valor correspondente de y e somar então todos esses produtos.
r	Representa o coeficiente de correlação linear para uma amostra

Fonte: Adaptado de Triola (2005)

Obs.: o coeficiente de correlação linear para população é representado pela letra grega ρ (rô).

Larson e Farber (2010) afirmam que o coeficiente de correlação linear varia de -1 a 1. Desta forma se as variáveis x e y têm uma correlação linear forte e positiva, o valor de *r* estará próximo de 1. Caso não haja correlação linear ou esta seja muito fraca o valor de *r* ficará próximo de 0 (zero). Quando a correlação linear é forte, porém negativa, o valor de *r* ficará próximo de -1.

Os gráficos da figura 2 a seguir, representam algumas variações de *r*.

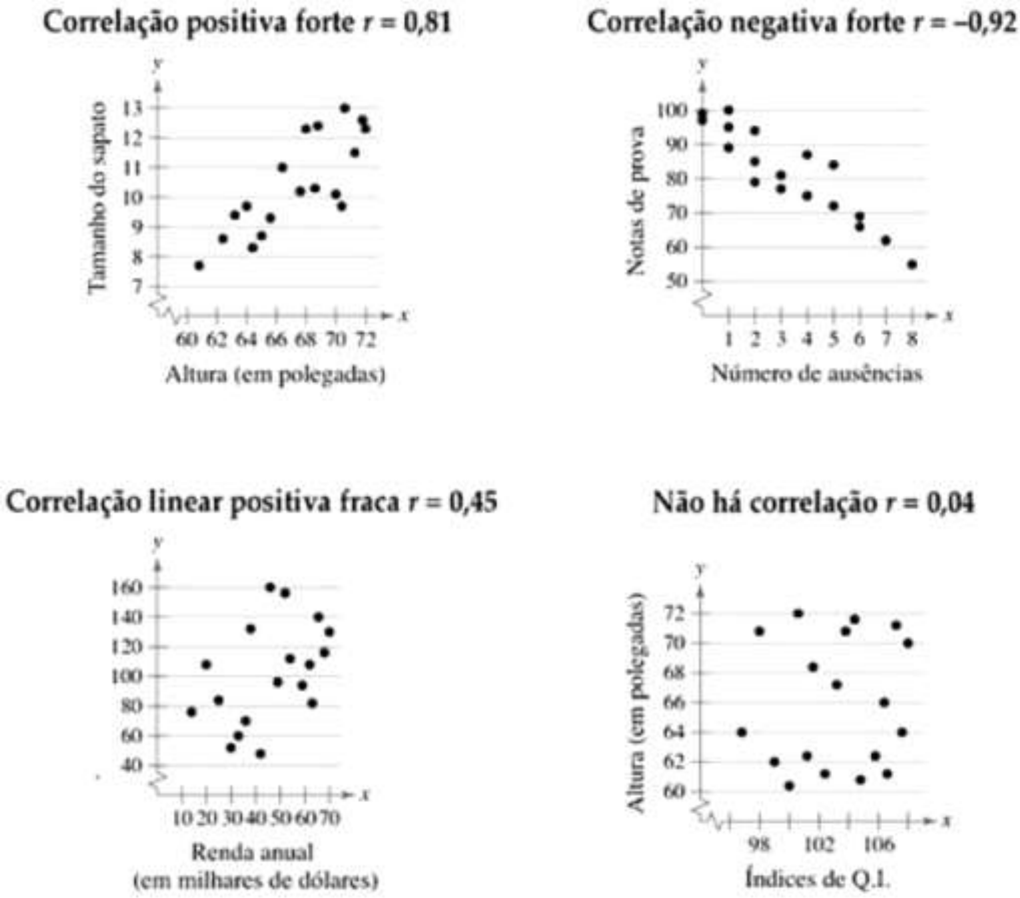


Figura 2- Representação de variação de r
Fonte: Larson e Farber (2010)

Consideremos agora o exemplo dado anteriormente da relação entre as faltas e as notas finais dos alunos e calculemos o coeficiente de correlação linear para os dados.

Para facilitar a confecção da fórmula, sugere-se a construção de uma tabela com seus elementos, conforme pode ser visto na tabela 1 a seguir:

Tabela 1- Dados para cálculo de r

Faltas (x)	***Nota Final (y)***	xy	x^2	y^2
8	78	624	64	6084
2	92	184	4	8464
5	90	450	25	8100
12	58	696	144	3364
15	43	645	225	1849
9	74	666	81	5476
6	81	486	36	6561
$\sum x = 57$	$\sum y = 516$	$\sum xy = 3751$	$\sum x^2 = 579$	$\sum y^2 = 39898$

Fonte: A própria autora

Utilizando os dados da tabela 1 na equação 1 temos:

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

$$r = \frac{7.3751 - (57).(516)}{\sqrt{7.(579 - (57)^2}\sqrt{7.(39898) - (516)^2}}$$

$$r = \frac{-3155}{\sqrt{804}.\sqrt{13030}}$$

$$r = \frac{-3155}{28{,}35\,.\,114{,}15}$$

$$\boxed{r = -0{,}9748}$$

O valor encontrado para *r* através da fórmula, reafirma o exposto no gráfico 1 onde se percebe a tendência de uma correlação linear e negativa (decrescente), uma vez que *r* está próximo de -1.

Desta forma pode-se concluir, para o exemplo citado, que quanto menor o número de faltas dos alunos, maior será a nota final.

2.2.2 Regressão Linear

Larson e Farber (2010, p.409) definem regressão linear como: “Uma **linha de regressão** também chamada de **linha de menor ajuste**, é a linha para a qual a soma dos quadrados dos resíduos é mínimo”.

Para Rezende e Freitas (2006) uma regressão linear consiste em uma tentativa de se estabelecer uma equação matemática de grau um (linear) que seja capaz de descrever a relação que existe entre duas variáveis.

Desta forma, a análise de regressão tem como objetivo, criar uma descrição, através de um modelo matemático, a relação existente entre duas variáveis, tendo como base as observações feitas previamente. (LEITE, 2011).

Correa (2003) afirma que para que se possa avaliar de maneira mais coerente a relação que existe entre as duas variáveis é estabelecer uma reta que seja capaz de prever os comportamentos de valores não contidos incialmente. Chama-se esta reta de reta de regressão.

A equação da reta de regressão é dada por:

$$\hat{y} = ax + b$$

Esta expressão define a relação entre x (variável independente ou preditora) e $\hat{y}$ (variável dependente ou resposta). O coeficiente a é chamado de coeficiente angular e b chamado de intercepto. (TRIOLA, 2005).

Para encontrar os coeficientes a e b, utilizam-se as fórmulas 2 e 3 a seguir:

Para o coeficiente a (coeficiente angular) temos:

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

Equação 2 - Coeficiente “a” (coeficiente angular)
Fonte: Adaptado de Larson e Farber (2010)

Para o coeficiente b temos:

$$b = \bar{y} - a\bar{x}$$

Equação 3 - Coeficiente b (intercepto)
Fonte: Adaptado de Larson e Farber (2010)

Os elementos $\bar{x}$ e $\bar{y}$ representam as médias aritméticas dos dados de x e y, respectivamente.

Consideremos ainda o exemplo 1 da relação entre as notas e as faltas dos alunos. Para se estabelecer a equação de regressão utilizaremos os dados já expostos na tabela 1.

Assim, para os valores de a e b temos:

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

$$a = \frac{7.3751 - (57)(516)}{7.579. (57)^2}$$

$$a = \frac{-3155}{804}$$

$$\boxed{a = -3,92}$$

Em posse de a**,** pode-se encontrar o valor de b.

$$b = \bar{y} - a\bar{x}$$

Temos $\bar{x} = 8,14$ e $\bar{y} = 73,71$.

Logo:

$$b = 73,71 - (-3,92.8,14)$$

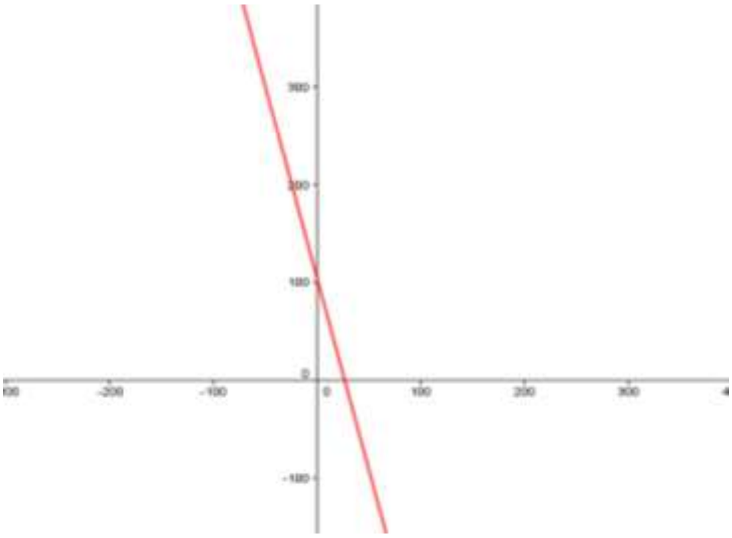
$$\boxed{b = 105,67}$$

Desta forma, a equação da reta de regressão para este caso será:

$$\boxed{\hat{y} = -3,92x + 105,67}$$

O gráfico 2 a seguir, representa a reta de regressão estabelecida pela equação de regressão para o exemplo 1.

Gráfico 2 - Reta de regressão



Fonte: A própria autora

2.2.2.1 Resíduos

De acordo com Triola (2005), um resíduo é a diferença entre um valor observado para y e o valor de y obtido a partir da reta de regressão linear.

2.2.2.2 Coeficiente de determinação

Quando se analisa uma reta de regressão é possível perceber que os pontos x e y estão distribuídos ao longo da reta acima ou abaixo dela, isto é, existe variações nos valores dispostos na reta. (BERTOLO, 2010).

Para uma coleção de dados emparelhados (x,y) consideremos que $\hat{y}$ é o valor predito de y na reta ajustada e $\bar{y},$ a média dos valores amostrais de y.

De acordo com Triola (2005), temos as seguintes definições:

- Desvio total: é a distância vertical entre cada valor de y e sua média $(\bar{y})$.
- Desvio explicado: é a distância vertical entre cada valor de $\hat{y}$ e a média amostral $\bar{y}$.
- Desvio não explicado: é a distância vertical entre cada valor de y e o valor correspondente na reta de regressão $(\hat{y})$.

Em posse dessas definições podemos definir o coeficiente de determinação que, segundo Triola (2005, p. 251) “(...) é o valor da variação de y que é explicado pela reta de regressão”. Seu valor é dado pela seguinte expressão:

$$r^2 = \frac{variação\ explicada}{variação\ total}$$

Equação 4- Coeficiente de determinação
Fonte: Triola (2005)

Para entender essa expressão consideremos a seguinte expressão:

$$Variação\ total = Variação\ explicada + Variação\ não\ explicada$$

Ou ainda:

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

Equação 5 - Variação total
Fonte: Triola (2005)

O valor do coeficiente de correlação pode ser definido pela equação 5 apresentada ou, simplesmente, elevando-se o valor do coeficiente de correlação linear (r) ao quadrado. (TRIOLA, 2005).

Considerando o exemplo 1, o coeficiente de determinação pode ser expresso considerando-se os dados da tabela a seguir:

Tabela 2 - Coeficiente de determinação

Faltas (x)	***Nota Final (y)***	$\hat{y}$ = -3,92x+105,67	$(y-\bar{y})^2$	$(\hat{y}-\bar{y})^2$
8	78	74,31	$(78-73,71)^2$= 18,37	0,35
2	92	97,83	$(92-73,71)^2$= 334,37	581,57
5	90	86,07	$(90-73,71)^2$= 265,22	152,66
12	58	58,63	$(58-73,71)^2$= 246,94	227,54
15	43	46,87	$(43-73,71)^2$= 943,37	72,62
9	74	70,39	$(74-73,71)^2$= 0,081	11,05
6	81	82,15	$(81-73,71)^2$= 53,08	71,16
$\sum x = 57$	$\sum y = 516$		$\sum(y-\bar{y})^2$ = 1861,43	$\sum(\hat{y}-\bar{y})^2$ = 1764,95

Fonte: Elaborado pela autora

Pela formula da equação 5 temos:

$$r^2 = \frac{variação\ explicada}{variação\ total}$$

$$r^2 = \frac{1764,95}{1861,43}$$

$$\boxed{r^2 = 0,95}$$

Se considerarmos o valor de r e elevarmos ao quadrado temos:

$$r = -0,9748$$
$$r^2 = (-0,9748)^2$$
$$\boxed{r^2 = 0,95}$$

Conclui-se, então que 95% da variação total das notas pode ser explicada pela variação das faltas. Os 5% de variação deve ser atribuída a fatores desconhecidos.

2.3C

3 ESTRUTURA DA SEQUÊNCIA DE ENSINO CONTEXTUALIZADA

O objetivo geral desta sequência de ensino é trabalhar conteúdos de estatística, através de pressupostos de contextualização baseados em coleta de dados feita por alunos de um curso de Engenharia de Produção.

Os conteúdos específicos a serem trabalhados nesta Sequência de Ensino serão:

- Correlação linear;
- Coeficiente de correlação linear de Pearson;
- Reta de regressão.

Para que esta Sequência de Ensino seja aplicada, as etapas para sua efetivação, totalizaram XX encontros propostos a partir de um curso de extensão com duração de XX horas aula por encontro.

As etapas para aplicação deste material foram as seguintes:

- 1ª etapa: Proposição do assunto e escolha do tema;
- 2ª etapa: Coleta dos dados;
- 3ª etapa: Desenvolvimento da atividade proposta;
- 4ª etapa: Síntese dos conteúdos trabalhados.

Para cada uma das etapas apresentou-se o tempo estimada para desenvolvimento das atividades propostas, os objetivos a serem alcançados, os recursos físicos a serem utilizados e a proposta de desenvolvimento das atividades. Todo o desenvolvimento foi acompanhado e orientado pela professora que aplicou a Sequência de Ensino por meio de fotografias e gravações. Desta forma ao apresentar cada etapa de desenvolvimento das atividades foi possível apresentar os resultados obtidos em cada atividade, com sugestões para melhorias em trabalhos posteriores.

ATIVIDADE 1

Título: Proposição do assunto e escolha do tema

Duração: 1h/a – 50 minutos

Objetivos: Realizar uma discussão para a escolha do tema a ser trabalhado

Conteúdos trabalhados: Conceito de correlação linear

Materiais utilizados: Quadro e giz

Desenvolvimento da atividade:

A atividade iniciou com uma conversa com os alunos do curso de Engenharia de Computação a partir da seguinte situação problema:

“Supondo que as horas de dedicação aos estudos influenciem na nota de uma determinada prova, ex: quem estudou 1 hora obtenha nota 5,0; quem estudou 2 hora obtenha nota 5,6; quem estudou 2 horas e meia obtenha nota 7,0. Existe alguma ‘proporção’ entre as notas? É possível estimar a nota de uma pessoa que tenha estudado por 3 horas?”.

Baseado nesta situação os alunos foram questionados a sugerir ideias de temas que fossem relacionados ao curso de Engenharia de Computação e que seguissem o mesmo conceito.

Os temas sugeridos foram anotados no quadro pela professora e foram os seguintes:

- *Potência da fonte de um computador em relação ao preço;*
- *Capacidade de armazenamento de um computador em relação ao preço de venda;*
- *Frequência de clock em relação ao preço;*
- *Tamanho (em polegadas) de monitores em relação ao preço;*
- *Desempenho de uma placa de vídeo (em frames por segundo) em relação ao preço;*
- *Densidade de pixels por polegadas (ppi) em relação ao preço;*
- ***Tempo de ordenação de Bubble Sort em relação ao tamanho do vetor;***
- ***Tempo de ordenação de Merg Sort em relação ao tamanho do vetor**.*

Com base nas sugestões feitas pelos alunos e após discussão, chegou-se a conclusão que os temas utilizados seriam os dois últimos citados.

Os alunos foram, então divididos em duplas (oito no total). Quatro destas duplas ficaram responsáveis por trabalhar com a ordenação de Bubble Sort e as outras quatro com a ordenação de Merg Sort.

Sugestão ao professor:

É importante ressaltar que o tema escolhido foi proposto pelos alunos dentre os vários temas apresentados. É possível encontrar outros temas que sigam a mesma ideia apresentada e se utilizem de outros conceitos de Engenharia de Computação ou ainda é possível que se disponibilize um tema para cada aluno (ou dupla) trabalhar.

ATIVIDADE 2

Título: Coleta de dados e cálculo do coeficiente de correlação

Duração: 2h/a – 100 minutos

Objetivos: Coletar dados para a aplicação da sequência de ensino

Conteúdos trabalhados: Diagrama de Dispersão e Coeficiente de correlação linear

Materiais utilizados: Computador, caneta, papel, lápis.

Desenvolvimento da atividade:

Para esta etapa os alunos, divididos em duplas realizaram a coleta dos dados, isto é, dos valores de tempo de ordenação pelo algoritmo a partir do tamanho do vetor considerado.

Para não haver discordância entre os tamanhos de vetores utilizados, foi criada uma padronização, isto é, foram preestabelecidos os tamanhos de vetores que seriam utilizados , independente do tipo de ordenação que fosse considerada.

Os tamanhos destes vetores foram os seguintes: 10000, 15000, 20000, 25000, 30000, 35000, 40000, 45000, 50000 e 55000.

A partir destes tamanhos de vetores os alunos puderam obter o tempo de ordenação do algoritmo.

Os dados obtidos foram anotados por cada uma das duplas, conforme pode ser observado na figura a seguir a partir do registro de uma das duplas:

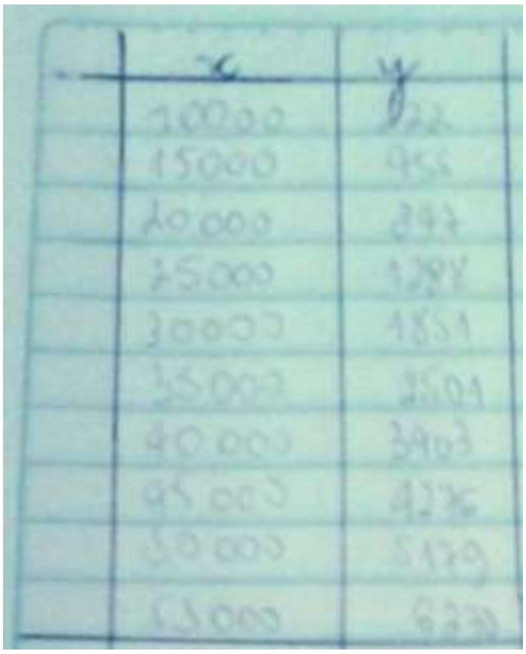


Figura 3 - Tamanho de vetor X tempo de ordenação de Bubble Sort
Fonte: Arquivos da autora

Após a coleta dos dados os alunos foram levados a refletir sobre a existência ou não de uma correlação entre os dados apresentados.

Para isso foi lançado o seguinte questionamento:

“Como identificar, matematicamente a existência ou não de correlação entre os dois conjuntos de dados apresentados”?

Foi então sugerido aos alunos que fizessem uma representação gráfica dos pontos obtidos (diagrama de dispersão).

> ***Sugestão ao professor*****:** Como os alunos tinham domínio do uso do excel, o diagrama de dispersão foi feito com o uso desta ferramenta, facilitando e agilizando a atividade. Porém caso isso não seja possível pode-se pedir que os alunos façam esta atividade usando régua, lápis e caderno.

A figura a seguir mostra o exemplo de um diagrama de dispersão feito por uma das duplas:

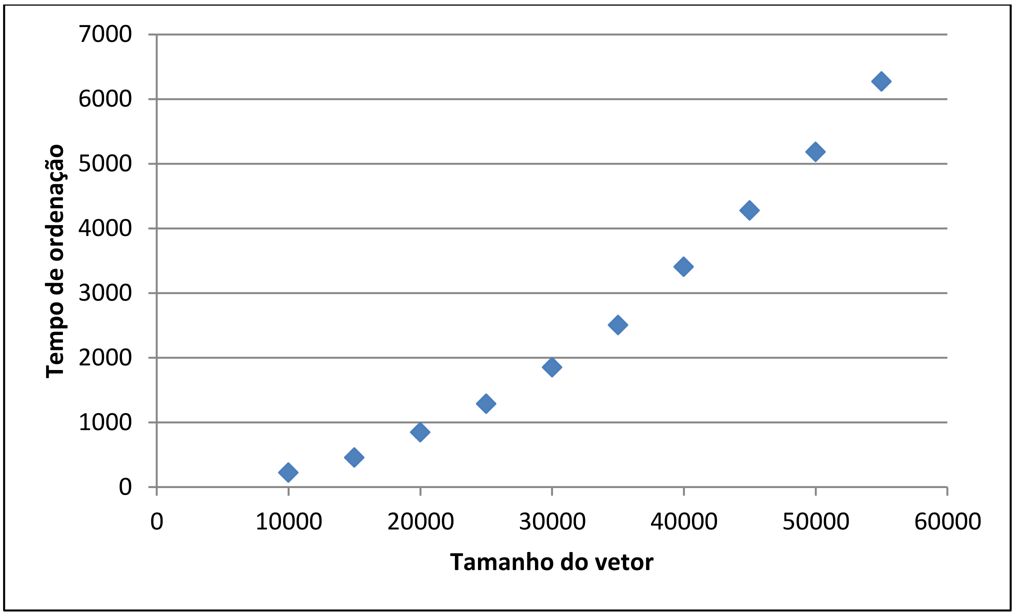
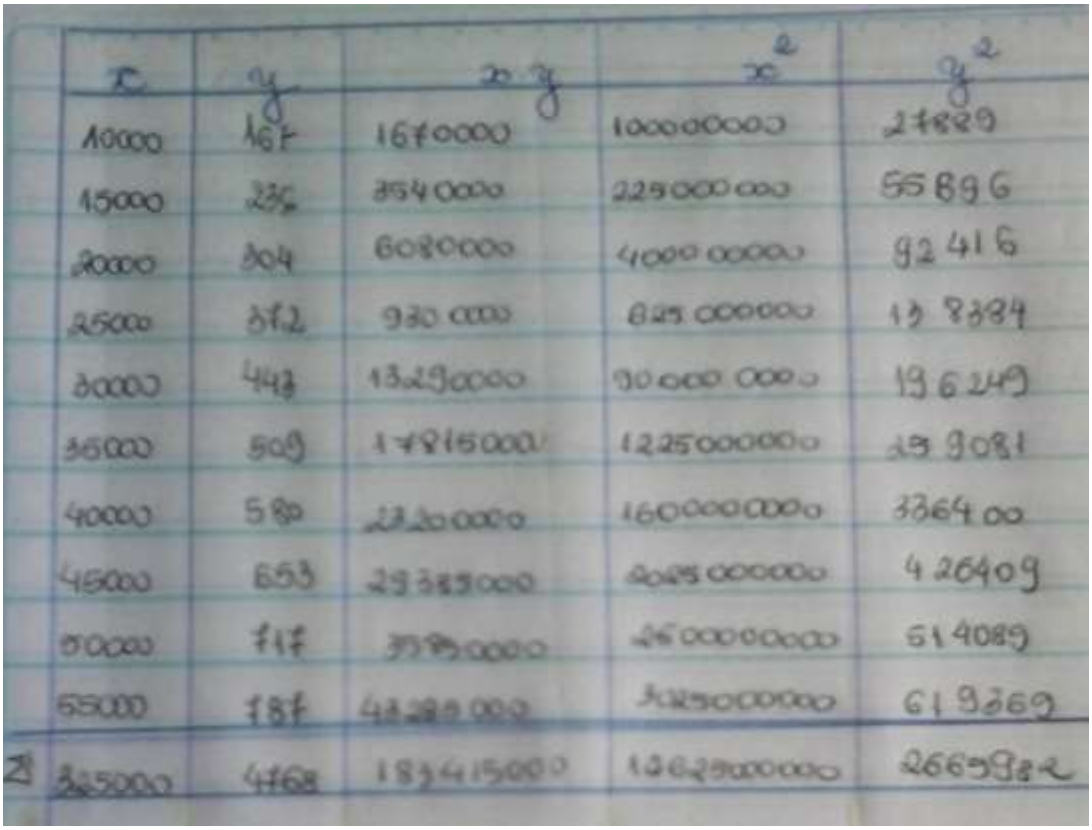


Figura 4 - Diagrama de dispersão feito com auxílio do Excel
Fonte: Arquivos da autora

Depois de feito o diagrama de dispersão foi apresentado aos alunos o conceito de coeficiente de correlação linear.

O cálculo do coeficiente de correlação linear foi calculado a partir da fórmula e cálculo feito manualmente.

Como, porém, o cálculo do coeficiente de correlação sugere vários somatórios, antes de efetuar o cálculo manual, os alunos montaram uma tabela como a que segue na figura:



x	y	xy	x^2	y^2
10000	167	1670000	100000000	27889
15000	236	3540000	225000000	55696
20000	304	6080000	400000000	92416
25000	372	9300000	625000000	138384
30000	443	13290000	900000000	196249
35000	508	17815000	1225000000	259081
40000	580	23200000	1600000000	336400
45000	653	29385000	2025000000	426409
50000	717	35850000	2500000000	514089
55000	787	43285000	3025000000	619369
Σ 325000	4768	183415000	12625000000	2669982

Figura 5 - Tabela para cálculo do coeficiente de correlação linear
Fonte: Arquivos da autora

Após a construção desta tabela, os alunos puderam aplicar a fórmula para cálculo do coeficiente de correlação linear, conforme pode ser observado na figura a seguir:

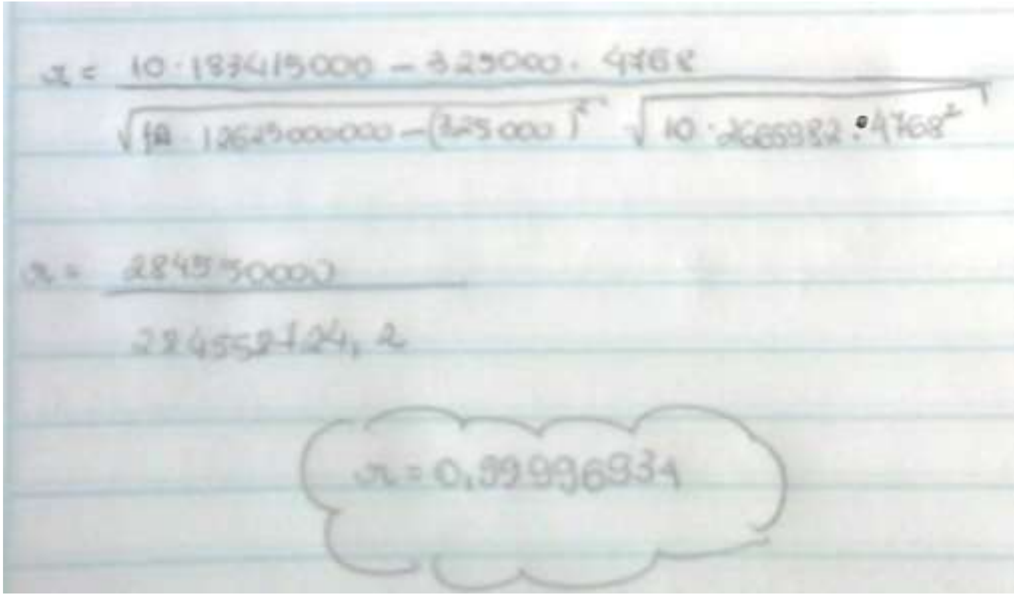


Figura 6 - Cálculo do coeficiente de correlação linear
Fonte: Arquivos da autora

Sugestão ao professor: É possível após mostrar o cálculo manual, apresentar a obtenção do coeficiente de correlação como uso do excel, tanto a partir da função "CORREL" como a partir de configurações do diagrama de dispersão.

As figuras a seguir representa o cálculo do coeficiente de correlação utilizando a função "CORREL" no Excel **(figura 7)** e a partir do diagrama de dispersão **(figura 8)**:

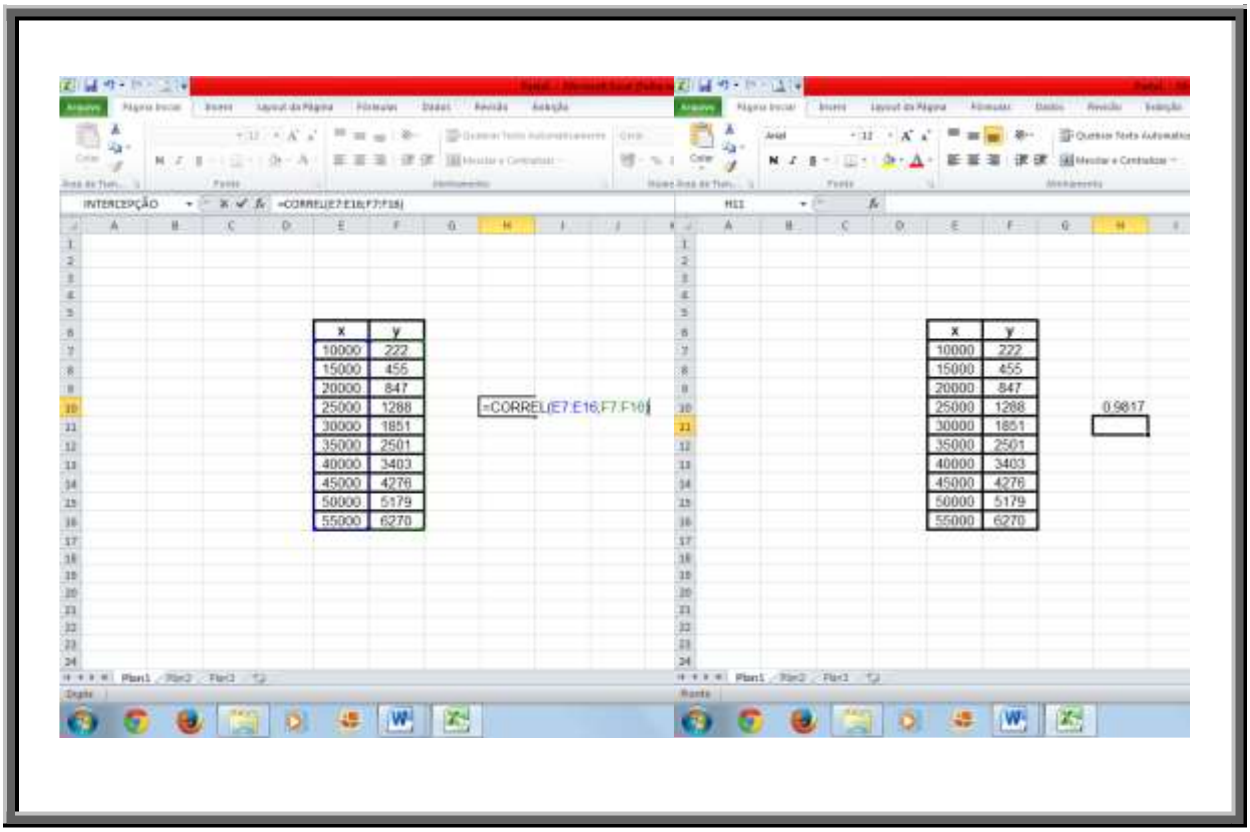


Figura 7 - Cálculo do coeficiente de correlação a partir do Excel
Fonte: A própria autora

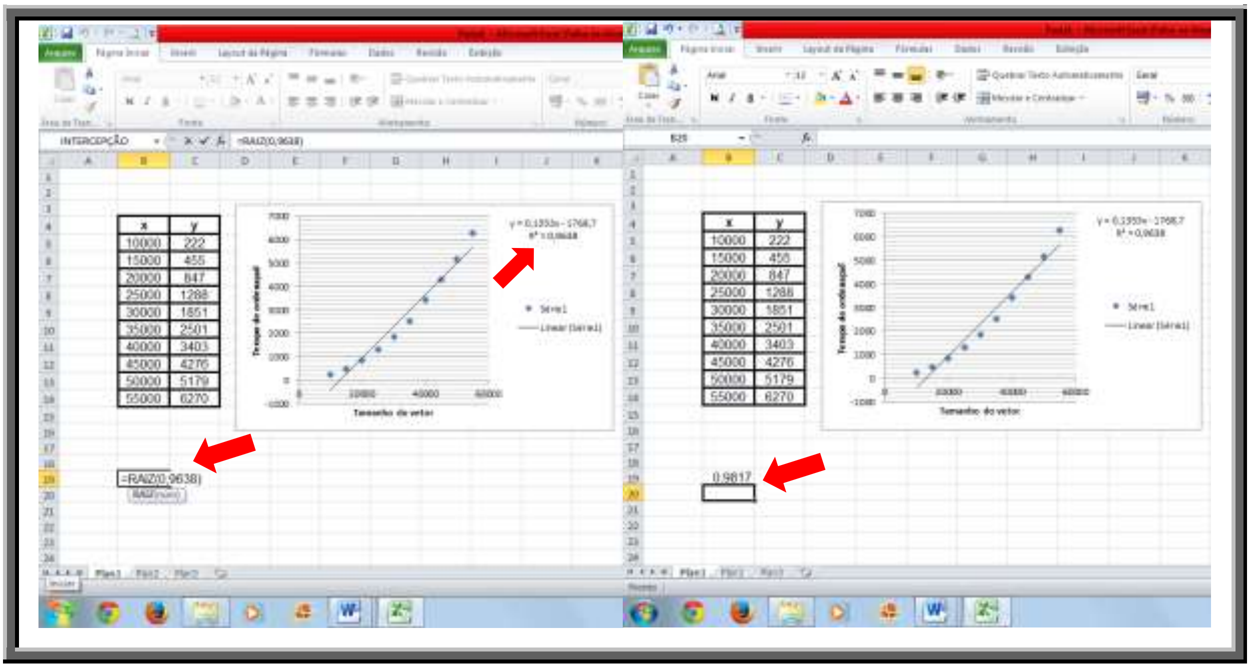


Figura 8 - Coeficiente de correlação a partir do diagrama de dispersão
Fonte: A própria autora

Obs: Como o diagrama de dispersão fornece o valor do coeficiente de determinação (r^2) para determinar o valor do coeficiente de correlação (r) é necessário extrair a raiz quadrada do valor encontrado.

ATIVIDADE 3

Título: Reta de regressão linear, resíduos e coeficiente de determinação

Duração: 2h/a – 100 minutos

Objetivos:

- Identificar os coeficientes da reta de regressão (a e b);
- Encontrar a equação de regressão linear.
- Definir resíduos;
- Encontrar os resíduos.

Conteúdos trabalhados: Reta de regressão linear e resíduos

Materiais utilizados: papel, caneta, planilhas eletrônicas, computador.

Desenvolvimento da atividade:

O conceito de reta de regressão linear foi introduzido a partir da observação do diagrama de dispersão.

É possível fazer com que os próprios alunos percebam que os pontos colocados no diagrama de dispersão sugerem uma reta a partir da disposição destes pontos.

Com os dados coletados, foi apresentado aos alunos as equações que estabelecem os coeficientes “a” e “b” da reta de regressão.

Observação: *Com o cálculo do coeficiente “a” sugere alguns somatórios seria necessário criar uma tabela para facilitar este cálculo. Estes somatórios, porém, são os mesmos utilizados para o cálculo do coeficiente de correlação linear, tornando a construção desta tabela dispensável.*

Sugestão ao professor: A expressão que representa a equação de regressão linear também pode ser observada com auxílio do excel, no diagrama de dispersão.

A figura a seguir representa o cálculo dos coeficientes da reta de regressão linear.

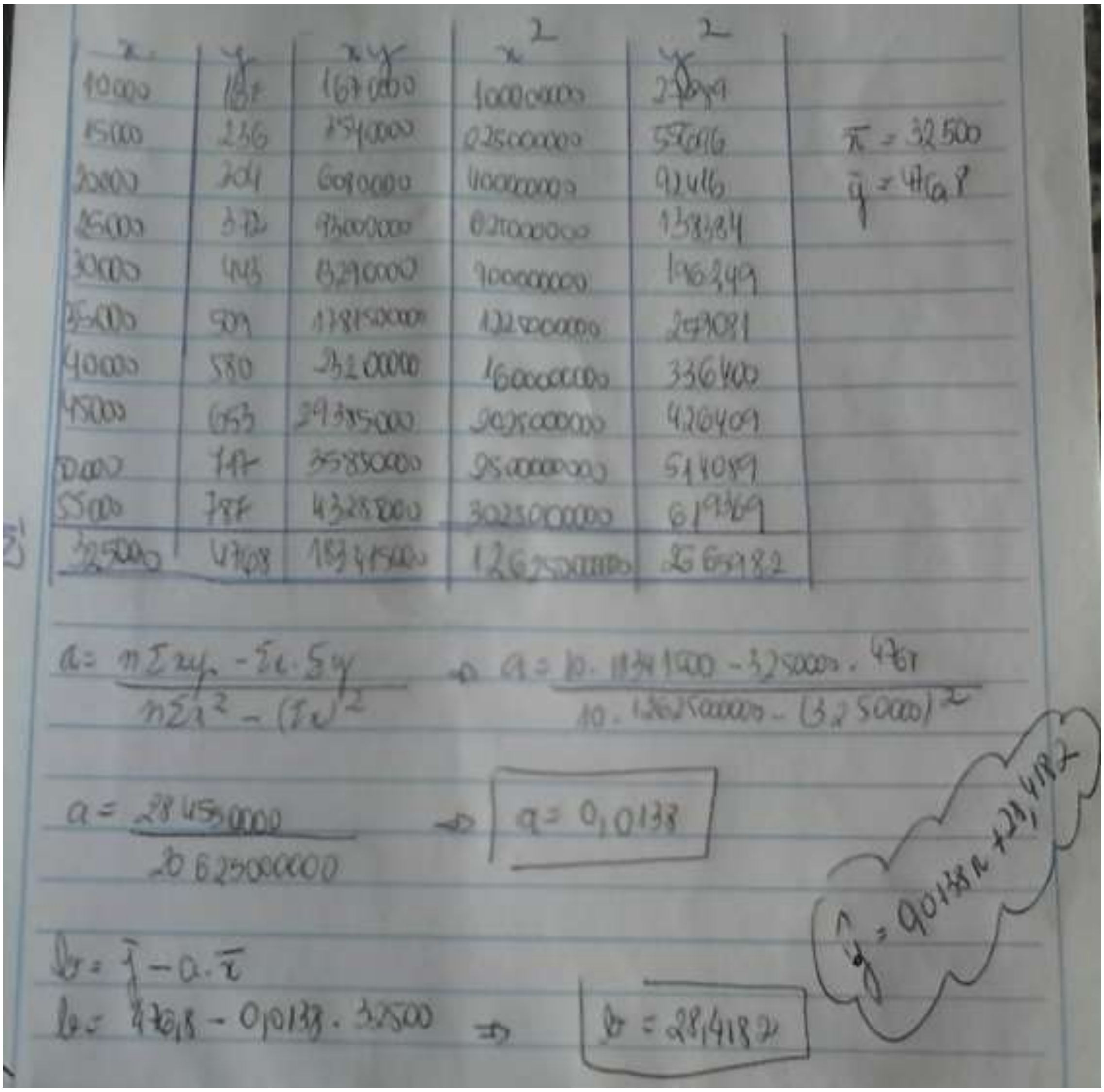


Figura 9 - Determinação da equação de regressão linear
Fonte: Arquivos da autora

A expressão que indica a equação de regressão linear também foi obtida a partir das configurações do diagrama de dispersão, feito com o auxilio do Excel, conforme mostra a figura a seguir:

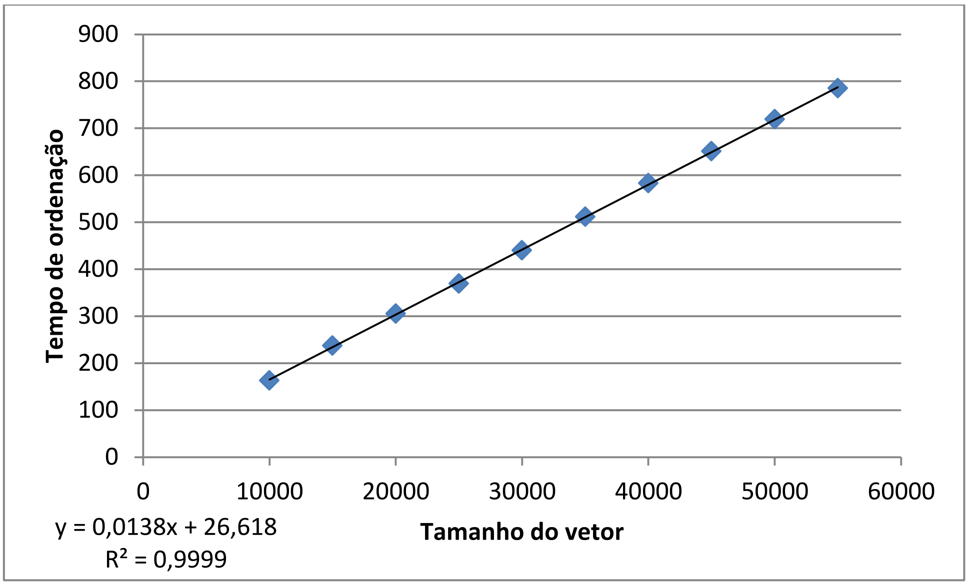
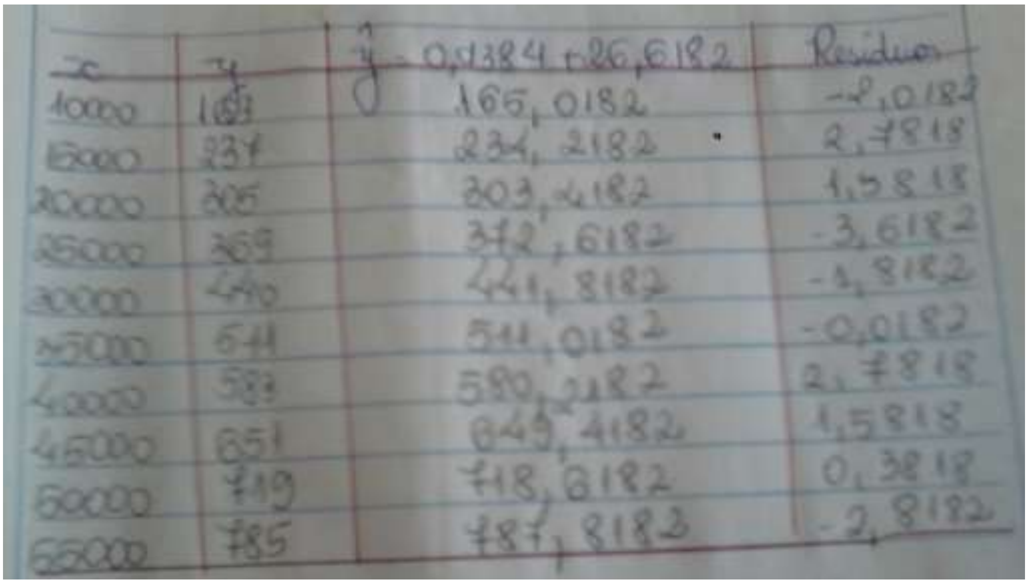


Figura 10 - Equação de regressão linear a partir do diagrama de dispersão
Fonte: Arquivos da autora

Tendo obtido a equação de regressão linear os alunos puderam então determinar os resíduos dos dados obtidos.

Este conceito foi colocado aos alunos a partir da observação, dos próprios alunos de que havia uma diferença entre o verdadeiro valor de y e o valor obtido a partir da reta de regressão.

A figura a seguir representa o cálculo dos resíduos de uma das duplas que fizeram parte deste trabalho:

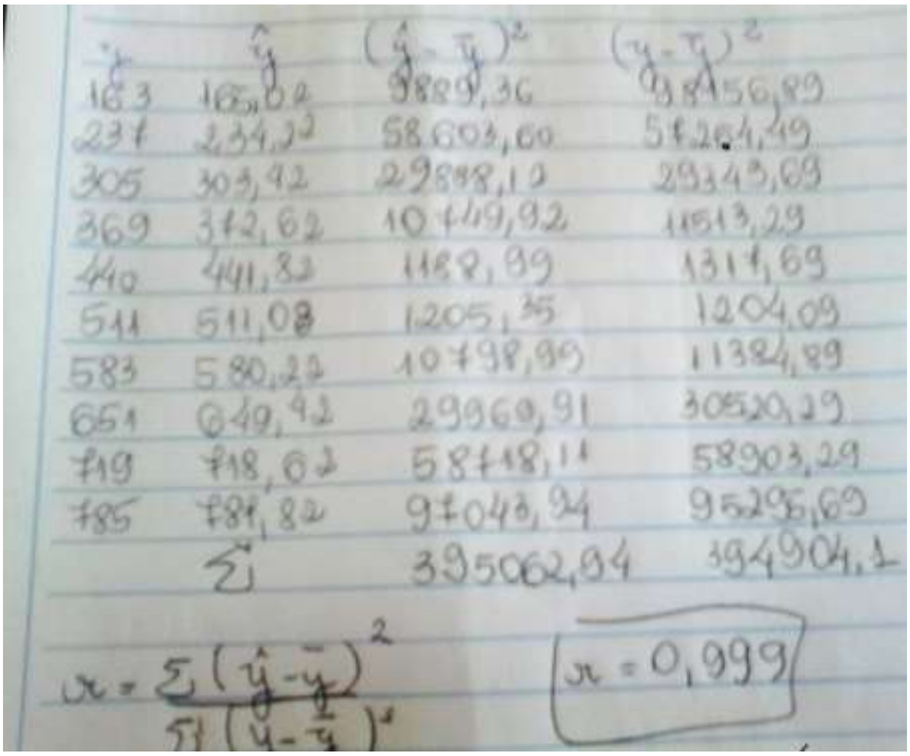


x	y	$\hat{y}$ = 0,1384 +26,6182	Resíduos
10000	163	165,0182	-2,0182
15000	237	234,2182	2,7818
20000	305	303,4182	1,5818
25000	369	372,6182	-3,6182
30000	440	441,8182	-1,8182
35000	511	511,0182	-0,0182
40000	583	580,2182	2,7818
45000	651	649,4182	1,5818
50000	719	718,6182	0,3818
55000	785	787,8182	-2,8182

Figura 11 - Cálculo dos resíduos
Fonte: Arquivos da Autora

Depois do cálculo do resíduo, os alunos ainda fizeram o cálculo do coeficiente de determinação (r^2), primeiramente a partir da equação 5 e depois, fazendo, simplesmente o valor do coeficiente de correlação (r) elevado ao quadrado.

Para a obtenção de r^2 pela equação 5 antes os alunos montaram um quadro como o que sugere a figura a seguir e após obtiveram o valor do coeficiente de determinação:



y	$\hat{y}$	$(\hat{y}-\bar{y})^2$	$(y-\bar{y})^2$
163	165,02	9889,36	98456,89
237	234,22	58603,60	57264,49
305	303,42	29898,12	29343,69
369	372,62	10749,92	11513,29
440	441,82	1188,99	1317,69
511	511,08	1205,35	1204,09
583	580,22	10798,99	11384,89
651	649,42	29969,91	30520,29
719	718,62	58718,11	58903,29
785	787,82	97043,94	95296,69
	Σ	395062,94	394904,1

$$r = \frac{\Sigma(\hat{y}-\bar{y})^2}{\Sigma(y-\bar{y})^2}$$

$\boxed{r = 0{,}999}$

Figura 12 - Cálculo do coeficiente de determinação
Fonte: Arquivos da Autora

> **Observação:** é possível também obter o coeficiente de determinação pelas configurações do diagrama de dispersão.

ATIVIDADE 4

Título: Síntese dos conteúdos trabalhados

Duração: 2h/a – 100 minutos

Objetivos:

- Analisar o entendimento dos conceitos apresentados a partir da perspectiva dos alunos.

- Resolver exercícios de fixação

Conteúdos trabalhados: Correlação e regressão

Materiais utilizados: Papel e caneta

Desenvolvimento da atividade:

Após desenvolver as atividades a partir dos dados coletados por cada uma das duplas, foi sugerido aos alunos que fizessem uma síntese dos conceitos aprendidos durante a aplicação da atividade proposta.

Esta síntese consistia em verificar o entendimento teórico sobre os aspectos relacionados à Correlação e Regressão Linear.

Para isso, cada dupla recebeu uma folha contendo um quadro como o que segue:

Quadro 3 - Conceitos solicitados aos alunos

<table>
<tr><th>Momento</th><th>Definição dos parâmetros</th><th>Definição</th></tr>
<tr><td>1</td><td>Correlação</td><td></td></tr>
<tr><td>2</td><td>Coeficiente de correlação linear (r)</td><td></td></tr>
<tr><td>3</td><td rowspan="2">Reta de regressão ou reta de menor ajuste ou reta dos mínimos quadrados</td><td rowspan="2"></td></tr>
<tr><td>4</td></tr>
<tr><td>5</td><td>Coeficientes da reta (a e b)</td><td></td></tr>
<tr><td>6</td><td>Resíduos</td><td></td></tr>
<tr><td>7</td><td>Coeficiente de determinação (r^2)</td><td></td></tr>
</table>

Fonte: A própria autora

Cada dupla deveria, então, escrever as definições solicitadas a partir de seu entendimento, baseando-se nas atividades que foram propostas.

Para considerar a questão correta ou não, foram considerados três parâmetros: “Crédito Total”, para conceitos que estivessem totalmente corretos, “Crédito Parcial”, para conceitos, parcialmente corretos e “Crédito Nulo” para conceitos incorretos ou não respondidos.

Os conceitos solicitados foram comparados com os contidos nos livros dos autores Triola (2005) e Larson & Farber (2010).

A partir desta comparação foi possível estabelecer a compreensão (ou não) dos conceitos trabalhados, a partir do número de Créditos Totais, Parciais ou Nulos.

Para finalizar a atividade, os alunos puderam, depois, resolver outros exercícios sobre Correlação e Regressão Linear (também dos autores Triola (2005) e Larson & Farber (2010)).

Sugestão ao professor: Os livros adotados para a comparação dos conceitos solicitados e realização de exercícios de fixação podem ser alterados de acordo com a afinidade do professor com estes ou outros autores.

Observação: *as respostas dos alunos a estas questões podem ser observadas nos anexos deste material.*

5 CONSIDERAÇÕES FINAIS

O objetivo deste trabalho foi oferecer ao professor de Matemática (Estatística) do Ensino Superior, especificamente de um curso de Engenharia de Computação um material de apoio (Sequência de Ensino), construída a partir do conceito de correlação para o ensino de Correlação e Regressão Linear.

Ponderando-se a importância do ensino de Estatística e, neste caso, o ensino de Correlação e Regressão Linear, acredita-se que o fato de propiciar ao aluno do Ensino Superior, uma maneira diferenciada para a aprendizagem destes conteúdos, de forma a fazê-lo perceber sua relação com a realidade profissional escolhida, seja uma forma eficaz de construir conhecimentos e aprimorar a aprendizagem.

Com base na literatura apresentada nesta Sequência de Ensino, acredita-se que as atividades propostas são capazes de contribuir significativamente para a aprendizagem de conceitos de Correlação e Regressão.

É importante ressaltar que os conteúdos propostos neste material, foram aplicados a alunos de um curso Superior de Engenharia de Computação, mas que podem ser adaptados a outro curso superior, a partir do conhecimento da realidade vivida por cada profissão. Isso, porém, leva o professor a direcionar mais tempo para a realização da atividade, bem como esforçar-se por compreender o contexto no qual cada tipo de curso superior está inserido.

Além disso, outras situações, mesmo para o curso de Engenharia de Computação podem ser utilizadas para o ensino do mesmo conteúdo.

Desta forma, espera-se que este material possa auxiliar e incentivar outros professores a ensinar Correlação e Regressão de forma diferenciada e que possa gerar significado ao cotidiano dos alunos.

REFERÊNCIAS

ARA, Amilton B. **O Ensino da Estatística no Curso de Engenharia**. In: XXVIII Congresso Brasileiro de Ensino de Engenharia, 2000, Ouro Preto. Anais do XXVIII Congresso Brasileiro de Ensino de Engenharia, 2000.

BERTOLO, Luis A. **Estatística Aplicada no Excel**. IMES, Catanduva. 2010.

BROUSSEAU, Guy. **Os diferentes papéis do professor**. In. Parra, C; C, Saiz, I. et al. Didática da matemática: reflexões pedagógicas. Porto Alegre: Artes Médicas, 1996.

CORREA, Sonia M.B.B. **Probabilidade e estatística**. 2ª edição. Belo Horizonte: PUC Minas, 2003.

FONSECA, Maria C. F. R. **Por que ensinar Matemática**. Presença Pedagógica, Belo Horizonte, v.1, n. 6, mar/abril, 1995.

FURMANN, José . **Desenvolvimento de um modelo para a melhoria do processo de manutenção mediante a análise de desempenho de equipamentos**. 2002. Dissertação (Mestrado em Engenharia de Produção). Universidade Federal de Santa Catarina (UFSC), Florianópolis, 2002.

GUJARATI, Dadomar. N. **Econometria Básica**. São Paulo: MAKRON Books, 2000.

LARSON, Ron; FARBER, Betsy. **Estatística Aplicada**; tradução: Luciane Ferreira Pauleti Vianna. 4ª edição. São Paulo: Pearson Prentice Hall, 2010

LEITE, Osni P. **Apostila de estatística.** Curso Processamento de Dados. Faculdade de Tecnologia de Sorocaba, 2011.

LIMA FILHO, Luiz M.A. **Correlação e regressão**. Universidade Federal da Paraíba. Disponível em < http://www.de.ufpb.br/~luiz/AED/Aula9.pdf>. Acesso em 20 de novembro de 2013.

LIRA, Sachiko A. **Análise de correlação**: abordagem teórica e de construção dos coeficientes com aplicações. Dissertação de Mestrado. Universidade Federal do Paraná. Curitiba, 2004.

MONTGOMERY, D. C.; RUNGER, G. C. **Estatística aplicada e probabilidade para engenheiros**. LTC Livros Técnicos e Científicos, 2009.

NAGHETTINI, Mauro. C. ; PINTO, Eber J. A . **Hidrologia Estatística**. 1. ed. Belo Horizonte: CPRM, 2007. v. 1. 561p .

REZENDE, Oscar L. T. de ; FREITAS, Rony C. O. **Estatística com OpenOffice**. 2006.

VIALI, Lóri. **Estatística Básica**: Correlação e Regressão. Material Didático. 2002.

TAFNER, Elisabeth. P. **A contextualização do ensino como fio condutor do processo de aprendizagem.** Leonardo Pós Revista de Divulgação Técnico Científica do Instituto de Pós Graduação, Blumenau, v. 1, p. 47-51, 2003.

TRIOLA, Mario F. **Introdução à Estatística**. Editora LTC. 9ª edição, 2005.

TUFANO, Wagner. **Contextualização**. In: FAZENDA, Ivani. C. A. (Org). Dicionário em construção: interdisciplinaridade. São Paulo (SP): Cortez, 2001.

VASCONCELOS, Maria. B. F. **A contextualização e o ensino de matemática:** um estudo de caso. 2008. 113 f. Dissertação (Mestrado em Educação) - Universidade Federal da Paraíba, João Pessoa, 2008.

VASCONCELOS, Maria B. F.; RÊGO, Rogéria G. **A contextualização como recurso para o ensino e aprendizagem de matemática**. VI EPBEM – Monteiro, PB – 09, 10 e 11 de novembro de 2010.