UNIVERSIDADE FEDERAL DE SÃO CARLOS

CENTRO DE CIÊNCIAS E TECNOLOGIA PARA SUSTENTABILIDADE

PROGRAMA DE PÓS GRADUAÇÃO EM ENSINO DE CIÊNCIAS EXATAS

ORLANDO CESAR SANDRE

UMA PROPOSTA CONTEXTUALIZADA PARA O ENSINO MÉDIO - REGRESSÃO LINEAR

Sorocaba

2019

UNIVERSIDADE FEDERAL DE SÃO CARLOS

CENTRO DE CIÊNCIAS E TECNOLOGIA PARA SUSTENTABILIDADE

PROGRAMA DE PÓS GRADUAÇÃO EM ENSINO DE CIÊNCIAS EXATAS

UMA PROPOSTA CONTEXTUALIZADA PARA O ENSINO MÉDIO - REGRESSÃO LINEAR

Orlando Cesar Sandre

Orientador: Prof. Dr. Antônio Luís Venezuela

Sorocaba

2019

Sandre, Orlando Cesar

Uma proposta contextualizada para o Ensino Médio - Regressão Linear / Orlando Cesar Sandre. -- 2019.
134 f. : 30 cm.

Dissertação (mestrado)-Universidade Federal de São Carlos, campus Sorocaba, Sorocaba
Orientador: Prof. Dr. Antônio Luís Venezuela
Banca examinadora: Profa. Dra. Luiza Amália Pinto Cantão, Profa. Dra. Magda da Silva Peixoto
Bibliografia

1. Contextualização Matemática. 2. Estatística. 3. Regressão Linear. I. Orientador. II. Universidade Federal de São Carlos. III. Título.

Ficha catalográfica elaborada pelo Programa de Geração Automática da Secretaria Geral de Informática (SIn).

DADOS FORNECIDOS PELO(A) AUTOR(A)

Bibliotecário(a) Responsável: Maria Aparecida de Lourdes Mariano – CRB/8 6979

UMA PROPOSTA CONTEXTUALIZADA PARA O ENSINO MÉDIO - REGRESSÃO LINEAR

Dissertação apresentada ao Programa de Pós Graduação em Ensino de Ciências Exatas do Centro de Ciências Exatas e Tecnologia para Sustentabilidade da Universidade Federal de São Carlos - Campus Sorocaba, como exigência parcial para a obtenção do título de mestre sob orientação do Professor Doutor Antônio Luís Venezuela.

Sorocaba

2019

UNIVERSIDADE FEDERAL DE SÃO CARLOS

Centro de Ciências Exatas e de Tecnologia
Programa de Pós-Graduação em Ensino de Ciências Exatas

Folha de Aprovação

Assinaturas dos membros da comissão examinadora que avaliou e aprovou a Defesa de Dissertação de Mestrado do candidato Orlando Cesar Sandre, realizada em 25/01/2019:

Prof. Dr. Antonio Luís Venezuela
UFSCar

Profa. Dra. Luiza Amália Pinto Cantão
UNESP

Magda Peixoto

Profa. Dra. Magda da Silva Peixoto
UFSCar

Para Jesus Cristo

Agradecimentos

À Deus, que ao longo da minha vida, vem Se mostrando cada vez mais presente.

Ao meu orientador Prof. Dr. Antônio Luís Venezuela, pela paciência e dedicação, mostrando qual caminho a ser seguido.

Aos meus pais, Milton Sandre (in memorian) e Maria Ap. de Mello Cardia Sandre (in memorian) que foram um exemplo de vida para mim.

À minha esposa Denair de Lima Sandre, que me apoio neste novo desafio, com paciência e sabedoria.

Aos meus filhos Rodrigo, Vinícius, Ana Laura e Giovana.

Aos diretores da Frangoeste Avicultura Ltda em especial ao Sr. Claudio Roberto Foltran, a Sra. Cleide Maria Gaiotto Madureira e Sr. Rogério Fabiano Ferraz, que permitiam a flexibilidade de horário, para eu poder participar do curso de Pós-Graduação da UFSCar-Sorocaba.

Aos professores Antônio Luiz Venezuela, Magda da Silva Peixoto, Paulo César Oliveira, Renato F. Cantão, Sílvia M. Carvalho, Wladimir Seixas, que foram muito importantes para o meu desenvolvimento na área da matemática.

Aos meus amigos da Frangoeste Avicultura Ltda, José Roberto Macruz Pissinatti, Paulo Roberto Belini Junior, Wanderlei Barbosa Peres, que me apoiaram e deram suporte quando eu não estava presente na empresa.

Aos meus colegas de estudo que no decorrer do curso se mostraram solidá-

rios e atenciosos.

Resumo

Este trabalho tem como objetivo principal o ensino da Regressão Linear contextualizado para alunos do Ensino Médio. Para isso, foi elaborado um plano de unidade de ensino que possui quatro planos de aulas. Os planos de aulas são estruturados para que o aluno, através de dados fornecidos pelo professor, os utilize para a análise e criação de tabelas e gráficos, determinação da equação da reta de regressão linear e a determinação da correlação linear. Os conteúdos a serem utilizados são a Estatística do Ensino Médio e Função Afim. Serão apresentados três conjuntos de dados com situações diferentes onde a Regressão Linear pode ser utilizada. A primeira situação é a de um hospital onde os dados são o número de atendimentos no dia e a temperatura média diária. A segunda situação é em uma empresa onde se analisa a produção e o número de horas trabalhadas. A terceira situação é uma analise da pressão sistólica de mulheres e a idade delas. Assim, procuramos mostrar em três situações distintas a utilização da Regressão Linear, que pode ser adaptada por professores de acordo com o cotidiano dos seus alunos.

Palavras-chave: Contextualização, Regressão Linear, Método dos Mínimos Quadrados.

Abstract

This work has as the main objective the teaching of Linear Regression contextualized to High School students. For this, it was elaborated a plan of teaching unity which has four class plans. The class plans are structured so that the student, through the data provided by the teacher, uses it to the analysis and the creation of tables and charts, determination of the linear regression line equation and the determination of linear correlation. The content which will be used is the High Scholl Statistics and Related Function. It will be presented three sets of data with different situations where the Linear Regression can be used. The first one is from a hospital where the data are the number of calls in one day and the average daily temperature. The second situation is in a company where it analyses the production and the number of worked hours. The third situation is an analysis of systolic pressure of women and their age. Like this, we look to show in three different situations the use of the Linear Regression, which can be adapted by the teacher according to the daily life of their students.

Keywords: Contextualization. Linear Regression. The Method of Least Squares.

Sumário

Lista de Figuras

Lista de Tabelas

1 Introdução

A partir de 1997 com a publicação do Parâmetro Curricular Nacional (BRASIL, 1998), um documento que visa nortear os professores para conteúdos a serem ensinados em todas as escolas brasileiras, sendo elas públicas ou particulares, com o objetivo da uniformização do ensino em todo o país, mas de certa forma que ela não seja engessada, podendo adaptar-se às peculiaridades de cada região, o termo "Contextualização" e "Interdisciplinaridade" são muito citadas em todo o documento no sentido de metodologia para o ensino da Matemática. Segundo o dicionário online Michaelis (2008) o significado de "contextualizar" é:

> Dotar de contexto; Inserir ou introduzir em um texto; incorporar em uma narrativa; Incorporar (algo) em determinado contexto; Produzir um texto em que se encontre determinada palavra ou expressão, de modo a ampliar o entendimento de seu uso ou eliminar dúvidas acerca de sua aceitabilidade.

e "contexto" segundo o mesmo dicionário online Michaelis (2008), tem o significado de:

> Conjunto de circunstâncias inter-relacionadas de cuja tessitura se depreende determinado fato ou situação; circunstância(s), conjuntura, situação; Conjunto de circunstâncias que envolvem um fato e são imprescindíveis para o entendimento deste; Encadeamento de ideias ou conjunto de circunstâncias que precedem ou se seguem a determinados elementos e pressupostos de um texto, aprofundando-se o significado quando de sua leitura ou análise;

> Conjunto de circunstâncias que envolvem uma palavra, uma frase ou um segmento de texto, e das quais podem depender seu significado e valor; encadeamento do discurso; composição, entrecho, trama; Aquilo que constitui a totalidade do texto; composição, contextura. Conjunto; todo, totalidade;

Pelas várias definições do dicionário podemos dizer que contextualizar é "dotar de um conjunto de circunstâncias que envolvem um fato e são imprescindíveis para o entendimento deste". Contextualizar um conteúdo matemático então seria a inserção deste dentro de um contexto, ou seja, uma situação onde este conteúdo possa ser melhor compreendido e aceito como algo importante para o aluno que aprende e que já foi de grande importância para o desenvolvimento da humanidade tanto nos aspectos tecnológicos que abrangem quase todas as áreas do conhecimento humano, tais como Medicina, Biologia, Ciências Sociais, entre outras.

Fizemos pesquisas em dissertações e teses para melhor compreender contextualização matemática e os contextos onde ela pode ser inserida, como pode ser aplicada em sala de aula.

A História da Matemática é um exemplo de um contexto onde o conteúdo matemático pode ser explorado, pode descrever a origem e os motivos que levaram os povos em determinado ponto da história da humanidade a desenvolver o conteúdo matemático em questão. Para Vieira (2004), que estudou as estratégias de contextualização em livros didáticos, diz:

> ...os autores tentam situar historicamente o conhecimento matemático para o aluno, procurando fazer com que ele compreenda o desenvolvimento desse conhecimento como uma produção histórica, adaptada às demandas sociais de cada época. Eles apostam em que o reconhecimento da Matemática, como produto cultural, enriquece o processo de ensino-aprendizagem com informações e curiosidades, que são chaves para a compreensão dos modos como os conhecimentos matemáticos se organizam ou como certos procedimentos se estabelecem.

A aplicação da Matemática é um outro contexto que podemos utilizar para a contextualização de conteúdos matemáticos, e está aplicabilidade será articulada com outras áreas do conhecimento humano, ou seja, interdisciplinar, ou com a própria Matemática. Mendes (2010), desenvolveu sua pesquisa introduzindo conceitos teóricos de Matemática e Estatística dentro de um contexto com dados climáticos e meteorológicos do município onde está localizada a escola. Assim, de forma interdisciplinar, procura aplicar os conceitos teóricos, fazendo uma conexão com outras disciplinas, dando um significado contextualizado ao ensino. Sua pesquisa foi desenvolvida em um colégio técnico em agropecuária integrado ao Ensino Médio. Em suas considerações finais, Mendes (2010) delimita a sua pesquisa em não entrar no mérito de problemas complexos e contestar ou autenticar as teorias sobre contextualização e interdisciplinaridade, mas reafirma a possibilidade de construção do conhecimento de uma forma diferente da tradicional:

> Com este trabalho, não se procurou dar respostas prontas e definitivas sobre questões relacionadas a problemas complexos da aprendizagem. Também não teve a pretensão de contestar ou declarar autênticas as teorias sobre contextualização e interdisciplinaridade. Mas, com o seu desenvolvimento, pode-se reafirmar o que foi visualizado ao iniciá-lo, a possibilidade de construir o conhecimento sob uma perspectiva diferente daquela que é adotada pela educação tradicional.

O ensino contextualizado dos conteúdos matemáticos, faz com que os alunos sejam protagonistas da própria aprendizagem, assim afirma Sacco (2015). Em sua dissertação, ela realizou uma pesquisa com duas turmas do 3º ano do Ensino Médio. Ela aplicou um mesmo conteúdo de estatística através de uma sequência didática, com metodologias de ensino diferentes. Para a Turma A, ela utilizou-se da contextualização, e para a Turma B o conteúdo foi transmitido de forma que ela chama de "tradicional", ou seja, aula expositiva e alunos como meros receptores das informações.

Segundo Sacco (2015), os alunos da Turma A se mostraram muito interessados e aprovaram a forma diferenciada de como foi conduzida a aula, os alunos da Turma B, por outro lado, houve um menor entusiasmo em relação a outra turma.

Em relação aos objetivos de aprendizagem, as duas turmas foram satisfatórias, mas a Turma A onde houve a contextualização, teve um rendimento melhor comparado com a Turma B, onde a aula foi somente expositiva. Segundo Segundo Sacco (2015), nem todos os conteúdos matemáticos são fáceis de contextualizar, e que requer um tempo maior ao professor para o preparo, e alerta para uma outra dificuldade em se apresentar aula contextualizada, isto é, a movimentação dos alunos na sala de aula, pois os alunos são separados em grupos, onde existe uma participação ativa de toda a classe, que pode gerar indisciplina e comprometer a aprendizagem.

De maneira geral, existem varias formas onde o conteúdo matemático pode ser inserido e contextualizado, como a História da Matemática, aplicabilidade, Modelagem Matemática, Etnomatemática. Procurando assim dar um sentido aos conteúdos matemáticos, mostrando que a Matemática não é uma ciência isolada e sem conexão com outras áreas do conhecimento humano.

Procuramos neste trabalho, propor um *"Plano de Unidade"*, contendo quatro *Planos de aulas*, utilizando a *Regressão Linear* através do *Método dos Mínimos Quadrados*. Para a determinação da *Regressão Linear*, utilizaremos conteúdos matemáticos a nível de Ensino Médio. Segundo Almeida (2015):

> O Método dos Mínimos Quadrados, bem como qualquer outro procedimento empregado no ajuste de curvas, não faz parte da grade curricular do Ensino Médio, embora os recursos matemáticos necessários para implementá-lo estejam ao alcance do aluno do final (3º ano) desse segmento.

Em seu trabalho Almeida (2015), intitulado como: "O Método dos Mínimos Quadrados: Estudos e Aplicações para o Ensino Médio", ressalta a versatilidade e potencial do Método dos Mínimos Quadrados, como ferramenta de modelagem matemática. Sua proposta busca a motivação que gerou a origem do Método dos Mínimos Quadrados, não somente o tratamento matricial e a execução do método.

Foi versatilidade que nos chamou a atenção para desenvolver este tema, pois é muito grande as possibilidades de utilização da Regressão Linear, em qualquer área do conhecimento humano.

Segundo Jales (2014), "...outras áreas do conhecimento, tais como a Geografia, Biologia, Química, etc, se desempenham com a utilização da linguagem estatísticas."

Buscamos então em contextos reais do cotidiano, correlacionar duas variáveis entre si. Para o plano didático dividido em 4 planos de aulas, foi utilizado um contexto médico-administrativo, na área administrativa hospitalar (publica ou privada). O exercício que será proposto para os alunos desenvolverem extra classe, o contexto será de uma produção industrial. Há uma proposta para avaliação onde o contexto é a pressão arterial, que pertence a medicina, mas muito comum no cotidiano das pessoas.

Através do ensino de Estatística podemos despertar e incentivar nos alunos o espirito de pesquisa como coleta de dados, tabelas, gráficos, analise e discussão. A estatística é uma ferramenta na construção do "estudante-cientista", pela aplicabilidade em varias áreas do conhecimento humano, segundo Jales (2014).

1.1 Motivações e Objetivos

A motivação para o presente trabalho foi o interesse do autor em aprofundar neste novo paradigma da educação matemática contextualizando o ensino através da aplicabilidade, que é sugerida em (BRASIL, 1998) como metodologia para despertar o interesse nos alunos, procurando assim adicionar nossa contribuição aos professores de Matemática.

O objetivo geral desta dissertação é a elaboração de um *plano de unidade de ensino*, para alunos do ensino médio, contextualizando a equação da reta através de análise da regressão linear, correlacionando variáveis de outras áreas do saber.

Os objetivos específicos são:

- Verificar o que os PCN de Matemática sobre contextualização;
- Fazer um estudo sobre planejamento em ensino;
- Revisar os conceitos matemáticos equações e funções do primeiro grau e estatística;
- Entender o que é análise de regressão linear, e como ela é determinada através do Método dos Mínimo Quadrados;
- Elaborar um plano de unidade de ensino contextualizando a correlação entre o valor médio da temperatura diária e o número de atendimentos de pessoas com problemas respiratórios;
- Elaborar quatro planos de aulas para atingir os objetivos do plano de unidade de ensino;

- Propor um exercício contextualizando a correlação entre horas trabalhadas e produção do produto em toneladas, para serem executadas fora de sala de aula;
- Propor um avaliação contextualizando a correlação entre a idade de mulheres e a pressão arterial sistólica
- Propor aos alunos que tragam situações onde possam ser correlacionadas as variáveis.

O primeiro capítulo tem como proposta uma pesquisa sobre o que significa a palavra contextualização e contexto e seu significado nos textos dos Parâmetros Curriculares Nacional (BRASIL, 1998), a quantidade de trabalhos publicados entre os anos de 1987 à 2016 nos anais do Encontro Nacional de Educação (ENEM), traz uma analise de algumas questões contextualizadas publicadas em vestibulares e concursos públicos e um resumo sobre plano e planejamento na área de ensino.

O segundo capítulo tem como objetivo uma revisão teórica sobre os conteúdos de Estatística do Ensino Médio, Equações e funções de primeiro grau, que são utilizados para a a determinação da reta de regressão linear através do método dos mínimos quadrados, e o coeficiente de correlação linear.

No terceiro capítulo temos o plano de unidade de ensino dividido em quatro planos de aula. Esta unidade de ensino, correlacionará duas variáveis: *temperatura média diária de um hospital* e o *número de atendimentos com problemas respiratórios*, através do método dos mínimos quadrados, determinaremos a equação da reta de regressão, o coeficiente de determinação r^2 e correlação r, que representam a qualidade da reta determinada, e a força que existe entre as variáveis respectivamente.

Nos Apêndices E e F temos dois exemplos de correlação resolvidos, para o professor utilizar como exercício e avaliação. O primeiro procura correlacionar as horas trabalhadas em um empresa e a produção em toneladas. A avaliação correlaciona a idade de mulheres com a pressão arterial sistólica. O exercício e a avaliação forma resolvidos passo a passo como no plano de aula do capitulo 3, na forma de perguntas e resposta.

2 Contextualização, LDB/96 e PCN

Neste capítulo investigamos o que é contextualização matemática no ensino, procurando encontrar o seu objetivo como metodologia de aprendizagem, suas citações na Lei de Diretrizes Básicas (LDB/96) e no Parâmetros Curriculares Nacionais (BRASIL, 1998), a quantidade de trabalhos publicados no Encontro Nacional de Educação Matemática (ENEM, 2018), a sua utilização em questões de vestibulares e o que é plano e planejamento e os seus desdobramentos na área da educação.

2.1 Contextualização

Segundo o Dicionário online Dicio (2018) de Português a palavra "Contexto"tem o seguinte significado:

> A relação de dependência entre as situações que estão ligadas a um fato ou circunstância. O que compõe o texto na sua totalidade; a reunião dos elementos do texto que estão relacionados com uma palavra ou frase e contribuem para a modificação ou esclarecimento de seus significados.

Segundo o mesmo Dicionário temos que "contextualização é ato ou efeito de contextualizar", que tem a seguinte definição:

> Mostrar as circunstâncias que estão ao redor de um fato, acontecimento, situação.

> Entender ou interpretar algo tendo em conta as circunstâncias que o rodeiam, colocando num contexto.

Podemos dizer então que o termo *contextualização matemática* é a prática metodológica de inseri-lá em um contexto, dando aos conteúdos abstratos uma aplicabilidade e/ou um contexto histórico. Assim a Matemática será vista como parte de algo maior, que a sua utilização rompe com as barreiras da disciplinaridade, e que em alguns casos ela é insubstituível, como cita Brasil (1998):

> Possivelmente, não existe nenhuma atividade da vida contemporânea, da música à informática, do comércio à meteorologia, da medicina à cartografia, das engenharias às comunicações, em que a Matemática não compareça de maneira insubstituível para codificar, ordenar, quantificar e interpretar compassos, taxas, dosagens, coordenadas, tensões, frequências e quantas outras variáveis houver.

Utilizando-se a contextualização como metodologia de ensino da Matemática pode-se responder a conhecida pergunta feita por alunos: "Onde vou usar isso?", ou "para que serve isto?". Desta forma, o conteúdo matemático ao ser contextualizado além de responder as perguntas acima, também pode despertar o interesse dos aluno, procurando com que ele se motive e que trará benefícios ao seu aprendizado. Lobato (2008), afirma:

> Sabemos que o aluno estará motivado para os estudos somente quando o assunto trabalhado despertar o seu interesse. Assim, ele verá na aprendizagem a satisfação de sua necessidade do conhecimento.

Uma contextualização matemática é também a inserção de um conteúdo matemático em seu contexto histórico, ou seja, onde, quando e porque ele surgiu e como foi resolvido. A aplicabilidade em um contexto atual ou seu uso dentro da própria Matemática. Segundo Fogaça (2011) define contextualização como:

> De forma geral, contextualização é o ato de vincular o conhecimento à sua origem e à sua aplicação. A ideia de contextualização entrou em pauta com a reforma do Ensino Médio, a partir da Lei de Diretrizes e Bases da Educação (LDB nº 9.394/96), que acredita na compreensão dos conhecimentos para uso cotidiano. Além disso, os Parâmetros Curriculares Nacionais (PCN), que são guias que orientam a escola e os professores na aplicação do novo modelo, estão estruturados sobre dois eixos principais: a interdisciplinaridade e a contextualização.

Dessa forma tomemos como exemplo os logaritmos, que surgiram no início do seculo XVII, introduzido por John Napier[1], para a simplificação de cálculos e que atualmente não se utilizam mais para este fim, sendo sua utilização em funções que variam exponencialmente, sendo estas utilizadas na Geografia, Biologia, Química e Matemática Financeira e qualquer outra área que se utilize funções exponenciais ou logarítmicas.

Assim não importa a metodologia que for aplicada, se o conteúdo for isolado, fragmentado, mesmo que de forma completa e aprofundada, mas não houver uma conexão, não é garantia que o aluno estabeleça um significado ao conteúdo (BRASIL, 1998).

Percebemos que a contextualização vem para adicionar ao ensino de Matemática, um novo paradigma no ensino-aprendizado, uma metodologia que busca motivar o aluno, inserindo a Matemática em contexto, tornando a parte de algo maior, com o objetivo principal de um melhor ensino.

2.2 Lei de Diretrizes Básicas - LDB/96

Em 1996 foi criada a Lei de Diretrizes Básicas nº 9.394, com o objetivo de regulamentar o Ensino no país, oferecido gratuitamente a todos no Ensino

[1] John Napier (Edimburgo, 1 de fevereiro de 1550 — Edimburgo, 4 de abril de 1617) foi um matemático, físico, astrônomo, astrólogo e teólogo escocês.

Fundamental (9 anos) e Ensino Médio (3 anos). Segundo Carrocino (2014) , a criação da LDB foi um marco neste novo paradigma da contextualização:

> Nesse sentido, temos como marco no processo da contextualização, a criação de Lei nº 9.394, de 20 de dezembro de 1996, conhecida como a Lei de Diretrizes e Bases da Educação 9394/96(LDB), segundo a qual é condição necessária que a escola ofereça uma aprendizagem significativa.

2.3 Parâmetros Curriculares Nacionais

Os Parâmetros Curriculares Nacionais (PCN) (BRASIL, 1998) são referências para os Ensinos Fundamental e Médio de todo o país. O objetivo dos PCN é garantir a todas as crianças e jovens brasileiros, mesmo em locais com condições socioeconômicas desfavoráveis, o direito de usufruir do conjunto de conhecimentos reconhecidos como necessários para o exercício da cidadania.

Assim o PCN de Matemática define habilidades e competências de ensino a serem atingidas pelos alunos, e também propõe a forma de como atingir-las através da contextualização e interdisciplinariedade.

Dentro das habilidades e competências que os alunos devem atingir, podemos dizer que os Parâmetros Curriculares Nacionais trazem três concepções para o ensino da matemática, são elas formativa, instrumental e a ciência. Segundo Brasil (1998) nos descreve sobre as três concepções: Formativo, Instrumental e Ciência, assim temos:

Formativo:

> Em seu papel formativo, a Matemática contribui para o desenvolvimento de processos de pensamento e a aquisição de atitudes, cuja utilidade e alcance transcendem o âmbito da própria Matemática, podendo formar no aluno a capacidade de resolver problemas genuínos, gerando hábitos de investigação, proporcionando confiança e desprendimento para analisar e enfrentar situações novas,

> propiciando a formação de uma visão ampla e científica da realidade, a percepção da beleza e da harmonia, o desenvolvimento da criatividade e de outras capacidades pessoais.

Instrumental

> No que diz respeito ao caráter instrumental da Matemática no Ensino Médio, ela deve ser vista pelo aluno como um conjunto de técnicas e estratégias para serem aplicadas a outras áreas do conhecimento, assim como para a atividade profissional. Não se trata de os alunos possuírem muitas e sofisticadas estratégias, mas sim de desenvolverem a iniciativa e a segurança para adaptá-las a diferentes contextos, usando-as adequadamente no momento oportuno.

Matemática com Ciência:

> Contudo, a Matemática no Ensino Médio não possui apenas o caráter formativo ou instrumental, mas também deve ser vista como ciência, com suas características estruturais específicas. É importante que o aluno perceba que as definições, demonstrações e encadeamentos conceituais e lógicos têm a função de construir novos conceitos e estruturas a partir de outros e que servem para validar intuições e dar sentido às técnicas aplicadas.

2.3.1 Metodologia

A Metodologia a ser aplicada em aula de Matemática, diz que o ensino-aprendizagem seja realizado através da contextualização e interdisciplinariedade, ou seja, dentro das concepções formativas e instrumental, segundo Brasil (1998):

> O critério central é o da contextualização e da interdisciplinaridade, ou seja, é o potencial de um tema permitir conexões entre diversos conceitos matemáticos e entre diferentes formas de pensamento matemático, ou, ainda, a relevância cultural do tema, tanto no que diz respeito às suas aplicações dentro ou fora da Matemática, como à sua importância histórica no desenvolvimento da própria ciência.

Segundo (BRASIL, 1998) , exemplifica a citação acima com o estudo das funções, que tem uma ligação interna muito forte com a própria Matemática, assim uma parte da Trigonometria possui as funções Trigonométricas, o gráfico de uma função afim é a representação de uma sequencia de progressão aritmética, o gráfico de uma função exponencial é uma sequência da progressão geométrica. Existe também as ligações externas do estudo de funções que seria a conexão com outras áreas do conhecimento humano, assim define Brasil (1998) as funções e suas conexões internas e externas:

> Além das conexões internas à própria Matemática, o conceito de função desempenha também papel importante para descrever e estudar através da leitura, interpretação e construção de gráficos, o comportamento de certos fenômenos tanto do cotidiano, como de outras áreas do conhecimento, como a Física, Geografia ou Economia.

O estudo da Estatística é rica em exemplos atuais. Grandes corporações utilizam-se da Estatística para a tomada de decisões. Acreditamos que a Estatística é um conteúdo rico em conexões e articulações com a própria matemática e com outras disciplinas. Ela se utiliza de outros conhecimentos da Matemática para ser desenvolvida como a próprio conteúdo de funções e seus gráficos citado acima, potenciação, somatório e geometria analítica que se utiliza de matrizes. Lembrando-se também que a Estatística não é determinista, aproximando-se da realidade. Segundo Brasil (1998):

> Os conceitos matemáticos que dizem respeito a conjuntos finitos de dados ganham também papel de destaque para as Ciências Humanas e para o cidadão comum, que se vê imerso numa enorme quantidade de informações de natureza estatística ou probabilística. No tratamento desses temas, a mídia, as calculadoras e o computadores adquirem importância natural como recursos que permitem a abordagem de problemas com dados reais e requerem habilidades de seleção e análise de informações.

É fato que certos conteúdos matemáticos, estão diariamente nos meios de comunicação atuais, como gráficos e estatística. Assim um gráfico representando o crescimento da população de um país (ou uma cidade), temos elementos para introduzirmos conceitos como plano cartesiano, pares ordenados, domínio, imagem, porém o ensino das funções não se limita apenas a um gráfico. Este contexto inicial procura cativar a atenção do aluno, dando a ele um significado do que vai ser estudado, e posteriormente avançar até a abstração do conteúdo matemático. Assim sugere Druck :

> Um dos pontos de partida para esse processo é tratar, como conteúdo do aprendizado matemático, científico e tecnológico, elementos do domínio vivencial dos educandos, da escola e de sua comunidade imediata. Isso não deve delimitar o alcance do conhecimento tratado, mas sim dar significado ao aprendizado, desde seu início, garantindo um diálogo efetivo. A partir disso, é necessário e possível transcender a prática imediata e desenvolver conhecimentos de alcance mais universal.
>
> Os objetivos do Ensino Médio em cada área do conhecimento devem envolver, de forma combinada, o desenvolvimento de conhecimentos práticos, contextualizados, que respondam às necessidades da vida contemporânea, e o desenvolvimento de conhecimentos mais amplos e abstratos.

Pela análise dos Parâmetros Curriculares Nacionais, no que tange a contextualização e interdisciplinariedade, sugere um início de conteúdo, um contexto, porém não necessariamente todo ele o seja, um ponto de partida para motivar e mostrar ao aluno onde há uma aplicação. Assim continuando até a abstração estando ela contextualizada ou não. E ainda segundo Druck (2003), reforçam a abstração:"Note-se que a interdisciplinaridade do aprendizado científico e matemático não dissolve nem cancela a indiscutível disciplinaridade do conhecimento"

Os Parâmetros Curriculares Nacionais, algumas vezes, estão sendo erroneamente interpretado, causando confusão e desconforto aos professores.

Existe uma necessidade de se contextualizar todo e qualquer conteúdo do ensino de Matemática. Abaixo um trecho do artigo: "O drama do ensino da matemática", de Druck(2003)

> Os Parâmetros Curriculares Nacionais do MEC são erradamente interpretados como se a matemática só pudesse ser tratada no âmbito de situações concretas do dia-a-dia, reduzindo-a a uma sequência desconexa de exemplos o mais das vezes inadequados. Um professor de Ensino Médio relatou que, em sua escola, existe a "matemática junina", enquanto outro contou ter sido obrigado a dar contexto matemático a trechos de um poema religioso. Certamente, esses não são exemplos de uma contextualização criativa e inteligente que pode, em muito, ajudar nossos alunos. Lamentavelmente, esses tipos de exemplo proliferam em nossas escolas.

2.4 Encontro Nacional de Educação Matemática (ENEM)

Um levantamento feito nos trabalhos publicados nos Encontros Nacionais de Educação Matemática (ENEM, 2018) sobre contextualização, encontramos 38 publicações sobre o assunto. Já foram realizados doze edições do ENEM, nos anos 1987, 1988, 1990, 1992, 1995, 1998, 2001, 2004, 2007, 2010, 2013 e 2016. Os trabalhos submetidos ao ENEM, possuem vários eixos, e que são listados abaixo:

* Comunicação Científica;

* Comunicação Oral;

* Relato de Experiência;

* Poster;

* Minicurso;

* Mesa Redonda;

* Palestras.

Porém a nomenclatura dos eixos e a existência de alguns deles sofrem alterações em anos distintos dos encontros. Para selecionar os trabalhos fizemos um recorte, buscando somente nos títulos a palavra CONTEXTUALIZAÇÃO. Pela pesquisa realizada, não encontramos nenhum trabalho anterior ao ano de 1998. O primeiro trabalho apresentado no ENEM, que possuía o termo "contextualização" foi em 2001. Assim encontramos 38 trabalhos apresentados durante os anos de 2001 a 2016. A Tabela 1, mostra os anos e os eixos dos trabalhos.

Tabela 1. Trabalhos Publicados no ENEM

Trabalhos	2001	2004	2007	2010	2013	2016	Total por Eixo
Comunicação Científica	1	2	2	2	2	7	16
MiniCurso	0	5	1	1	0	2	9
Relatos de Experiência	0	1	0	2	5	1	9
Palestras	0	1	0	0	0	0	1
Poster	0	0	2	1	0	0	3
Totais por Ano	1	9	5	6	7	10	38

Fonte: Anais do ENEM

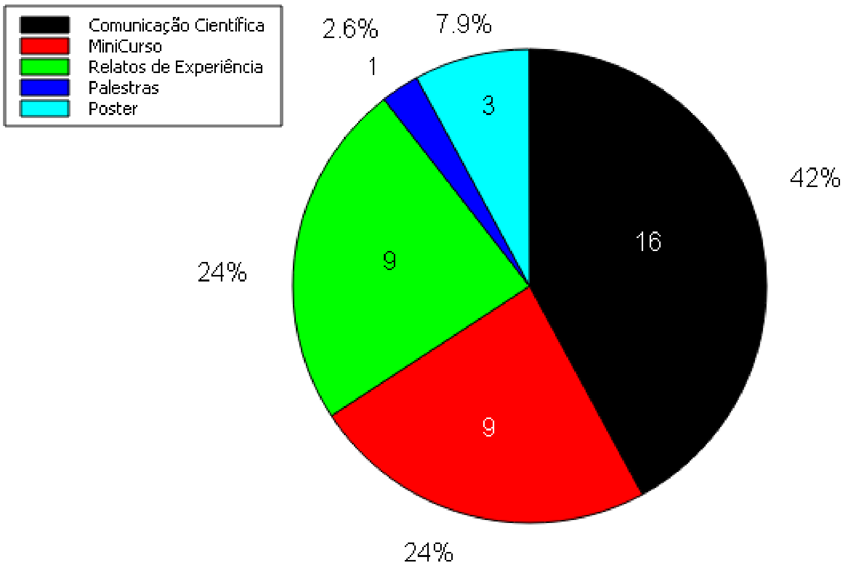
Para uma melhor visualização o gráfico da Figura 1 abaixo representa a quantidade de trabalhos por eixo:

O gráfico de barra da Figura 2, mostra a quantidade de trabalhos apresentados sobre contextualização durante os anos em que foram realizados o ENEM.

2.5 Análise de Questões Contextualizadas

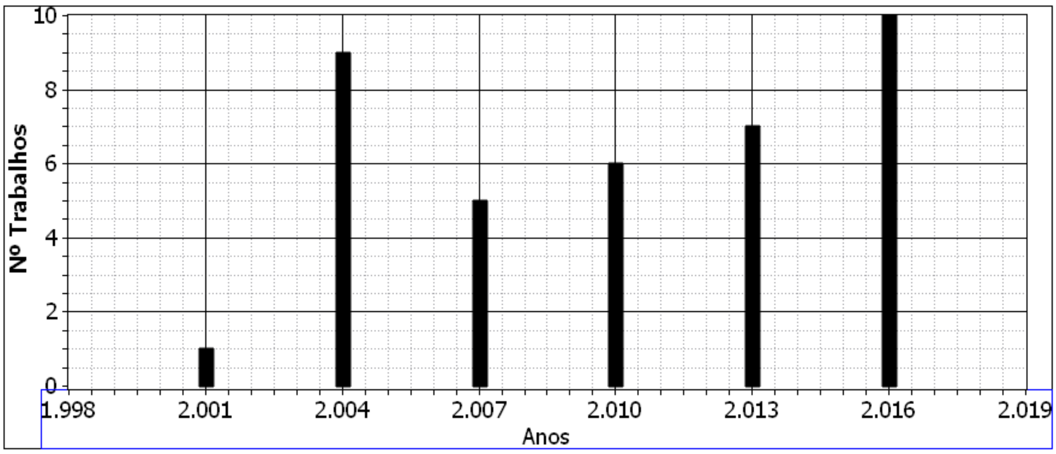
Percebemos que a interpretação errada do PCN, por uma parte de professores, coordenadores e avaliadores pode levar a cometer erros grosseiros na contextualização que pode até prejudicar o aprendizado.

Figura 1. Publicações por Eixo



Fonte: Autoria Própria

Figura 2. Publicações por Ano de Trabalhos no ENEM



Fonte: Autoria Própria

Segundo Santos e Oliveira (2015), os educadores necessitam compreender e saber o que é contextualizar:

> A Contextualização é importante na apropriação do conhecimento e cabe ao professor utilizá-la como uma estratégia do ensino para melhor aprendizagem dos alunos. Para tanto, os educadores necessitam saber o que significa contextualizar e como utilizar o método com objetivo claramente definido, para o qual ele saberá escolher e estabelecer os meios de alcançar.

Analisando algumas questões de vestibulares, concursos públicos entre outros, percebemos a não preocupação da contextualização com o cotidiano, procurando apenas a resolução do problema como observamos em uma questão da OBMEP (Olimpíada Brasileira de Matemática das Escolas Publicas) aplicada em 2014:

(OBMEP/2014, nível III) - Após lançar 2014 vezes uma moeda, Antônio contou 997 caras. Continuando a lançar a moeda, quantas caras seguidas ele deverá obter para que o número de caras fique igual à metade do número total de lançamentos?

a) 10

b) 15

c) 20

d) 30

e) 40

Solução e análise da questão: O detalhe a ser observado nesta questão e perceber que como ele continua a lançar a moeda, e o número total de lançamentos aumenta, logo a metade também alterara o valor. Algebricamente a solução ficaria assim:

$$997 + x = \frac{2014 + x}{2};$$

Logo teremos:

$$1994 + 2x = 2014 + x \Rightarrow 2x - x = 2014 - 1994 \Rightarrow x = 20;;$$

(UNESP/2009) - Um viveiro clandestino com quase trezentos pássaros foi encontrado por autoridades ambientais. Pretende-se soltar esses pássaros seguindo um cronograma, de acordo com uma progressão aritmética, de modo que no primeiro dia sejam soltos cinco pássaros, no segundo dia sete pássaros, no terceiro nove, e assim por diante. Quantos pássaros serão soltos no décimo quinto dia?

a) 55

b) 43

c) 33

d) 32

e) 30

Solução e análise da questão: A questão já diz que é uma Progressão Aritmética, não deixando assim por conta do aluno utilizar uma ferramenta matemática que ele já tenha aprendido, assim podemos resolver desta forma:

$a_1 = 5;$

$a_2 = 7;$

$a_3 = 9;$

Para a determinação da razão r temos:

$$r = a_2 - a_1 \Rightarrow r = 7 - 5 = 2$$

Aplicando se a equação do termo geral da PA temos:

$$
\begin{aligned}
a_n &= a_1 + (n-1)r \\
a_{15} &= a_1 + (15-1)r \\
a_{15} &= 5 + (14)2 \Rightarrow a_{15} = 33
\end{aligned}
$$

Um aluno que não conhece ou não se lembra as equações da PA, mas compreendeu a questão, poderá resolver enumerando os dias com a quantidade de pássaros soltos no respectivo dia fazendo uma listagem com segue abaixo:

1^o dia - 5 aves

2^o dia - 7 aves

3^o dia - 9 aves

4^o dia - 11 aves

...

...

14^o dia - 31 aves

15^o dia - 33 aves

Assim através do raciocínio lógico matemático o aluno poderá chegar ao resultado da questão.

Nada impede também de resolver através da equação da reta, fazendo a quantidade de aves soltas em função dos dias, assim através da equação da reta teremos:

$$y - y_1 = m(x - x_1)$$

Onde:

$$m = \frac{y_2 - y_1}{x_2 - x_1}$$

Logo:

$$m = \frac{7-5}{2-1} \Rightarrow m = \frac{2}{1} \Rightarrow m = 2$$

Assim substituindo m na equação da reta teremos:

$$y - 5 = 2(x-1) \Rightarrow y - 5 = 2x - 2 \Rightarrow y = 2x + 3$$

Assim podemos definir a função $f(x) = 2x + 3$ e para $f(15)$ teremos:

$$f(15) = 2.15 + 3 \Rightarrow f(15) = 33 \tag{2.1}$$

Percebemos que esta questão, também não houve uma preocupação com uma realidade do dia a dia. Essas duas questões mostram que em questões matemáticas contextualizadas, utilizam-se de um contexto que não pertence a realidade. Muitas questões contam uma história irreal, inexistente ou longe de um cotidiano e inserem a questão em meio a história.

O Exame Nacional do Ensino Médio (Enem) é uma prova elaborada pelo Ministério da Educação para verificar o domínio de competências e habilidades dos estudantes que concluíram o Ensino Médio. O Enem é composto por quatro provas de múltipla escolha, com 45 questões cada, e uma redação. Ele tem a proposta de questões contextualizadas e interdisciplinar, mas não podemos esquecer de que outros vestibulares importantes no âmbito nacional, ainda cobra uma matemática tradicional, buscando avaliar os vestibulandos, os conceitos puramente matemáticos. Segundo Brasil (1998):

> Os objetivos do Ensino Médio em cada área do conhecimento devem envolver, de forma combinada, o desenvolvimento de conhecimentos práticos, contextualizados, que respondam às necessidades da vida contemporânea, **e o desenvolvimento de conhecimentos mais amplos e abstratos.**

e também reforça com:

> Note-se que a interdisciplinaridade do aprendizado científico e matemático **não dissolve nem cancela a indiscutível disciplinaridade do conhecimento.**

Assim podemos exemplificar com a questão 24 do vestibular de 2017 do Instituto Tecnológico da Aeronáutica (ITA):

(ITA - 2017) Sejam A e B dois conjuntos com 3 e 5 elementos, respectivamente. Quantas funções sobrejetivas $F : B \to A$ existem?

Análise: É uma questão que avalia a Matemática como ciência, e por ser descritiva, vai exigir também do vestibulando o seu potencial no rigor da escrita matemática. Assim o vestibulando pode ter aprendido os conteúdos matemáticos através das várias formas de contextualização e interdisciplinariedade, mas para resolver questões abstratas como esta, precisa ampliar os seus conhecimentos no conteúdo para poder abstrair os conceitos dos conteúdos matemáticos.

2.6 Planejamento

Segundo o Dicio (2018) planejamento é:

> Ação de preparar um trabalho, ou um objetivo, de forma sistemática; planificação.
> Ação ou efeito de planejar, de elaborar um plano.
> Determinação das etapas, procedimentos ou meios que devem ser usados no desenvolvimento de um trabalho, festa, evento.
> Economia. Política. Desenvolvimento de projetos que buscam sanar os problemas sociais, econômicos, ou atingir certos objetivos de governo.

Segundo Malheiros (2012): "O planejamento de ensino é o detalhamento de todas as atividades que o professor irá realizar dentro da sala de aula", afirma ainda que para a construção do planejamento o professor tenha clareza de quatro questões listadas abaixo:

1 Os objetivos que devem ser atingidos;

2 Os conteúdos que serão trabalhados;

3 Os métodos que serão utilizados;

4 O espaço que estará disponível.

Segundo Haydt (2011) as definições sobre planejamento e plano são:

> **O planejamento** é um processo mental que envolve análise, reflexão e previsão. Nesse sentido, planejar é uma atividade tipicamente humana, e está presente na vida de todos os indivíduos, nos mais variados momentos.
> **O plano** é o resultado, é a culminância do processo mental de planejamento. O plano, sendo um esboço das conclusões resultantes do processo mental de planejar, pode ou não assumir uma forma escrita.

2.6.1 Tipos de Planejamentos

Na esfera da educação e do ensino, há vários níveis de planejamento, que variam em abrangência e complexidade segundo Haydt (2011):

1. planejamento de um sistema educacional;
2. planejamento geral das atividades de uma escola;
3. planejamento de currículo;
4. planejamento didático ou de ensino:
    a) planejamento de curso;
    b) planejamento de unidade didática ou de ensino;
    c) planejamento de aula.

As definições de Planejamento de um sistema educacional, planejamento geral das atividades de uma escola (Planejamento Escolar) e Planejamento de currículo segundo Haydt (2011) são:

O planejamento de um sistema educacional é feito a nível sistêmico, isto é, a nível nacional, estadual e municipal. Consiste no processo de análise e reflexão das várias facetas de um sistema educacional, para delimitar suas dificuldades e prever alternativas de solução. A partir dessas constatações é possível então definir prioridades e metas para o aperfeiçoamento do sistema educacional, estabelecer formas de atuação e calcular os custos necessários à realização das metas. O planejamento de um sistema educacional reflete a política de educação adotada.

O planejamento geral das atividades de uma escola: é o processo de tomada de decisão quanto aos objetivos a serem atingidos e a previsão das ações, tanto pedagógicas como administrativas que devem ser executadas por toda a equipe escolar. É um planejamento participativo envolvendo professores, funcionários, pais e alunos tendo como resultado o *plano escolar*.

O planejamento de currículo: é a previsão dos diversos componentes curriculares que serão desenvolvidos ao longo do curso, com a definição dos objetivos gerais e a previsão dos conteúdos programáticos de cada componente.

2.6.2 Planejamento didático ou de ensino

Segundo Mattos (1954 apud NéRICI, 1986) planejamento de ensino é: "previsão inteligente e bem calculada de todas as etapas do trabalho escolar que envolvem atividades docentes e discentes, de modo a tornar o ensino seguro, econômico e eficiente".

Segundo Turra e outros. (1974 apud NéRICI, 1986) planejamento de ensino é: "processo de tomada de decisões bem informadas que visam à racionalização das atividades do professor e do aluno, na situação ensino-aprendizagem, possibilitando melhores resultados e, em consequência, maior produtividade".

Segundo Nérici (1986): "O planejamento de ensino explicita-se através de três modalidades, percorrendo o caminho do mais geral para o menos geral, que são:"

a) planejamento de curso;

b) planejamento de unidade;

c) planejamento de aula.

Planejamento de curso é a previsão dos conhecimentos a serem desenvolvidos e das atividades a serem realizadas em uma determinada classe, durante um certo período de tempo, geralmente o ano ou semestre letivos. O resultado desse processo é o *plano de curso*, que é a previsão do trabalho docente e discente para o ano ou semestre letivo. O plano de curso é um desdobramento do plano curricular.

Segundo Nérici (1986) "O planejamento de uma unidade deve ser mais objetivo ou mais preciso que um plano de curso, porque está mais próximo da sua execução, da sua efetivação através do ensino". Afirma ainda que uma unidade é um todo significativo, ou seja, nem muito longo quanto um plano de curso, sendo assim é uma previsão, mais detalhada e precisa do que o plano de curso.

Planejamento de aula: segundo Nérici (1986):" é a previsão mais objetiva possível de todas as atividades escolares para a efetivação do processo ensino-aprendizagem que conduza o educando a alcançar objetivos previstos", assim o resultado do planejamento de aula é o *plano de aula*.

Haydt (2011) traz um resumo sobre planejamento, plano na área da educação, como segue:

1. O planejamento é um processo mental que supõe análise, reflexão e previsão. O plano é o resultado do planejamento.
2. O trabalho planejado é importante e necessário porque:
   (a) evita a improvisação;
   (b) ajuda a prever e superar dificuldades;
   (c) contribui para a consecução dos objetivos estabelecidos com economia de tempo e eficiência na ação.
3. As características de um bom plano são:
   (a) coerência e unidade;
   (b) continuidade e sequência;
   (c) flexibilidade;
   (d) objetividade e funcionalidade;
   (e) clareza e precisão.
4. No campo da educação e do ensino há vários tipos de planejamento, que variam de acordo com o seu grau de abrangência:
   (a) planejamento de um sistema educacional;

   (b) planejamento da escola;

   (c) planejamento de currículo;

   (d) planejamento didático ou de ensino:

      i. planejamento de curso;

      ii. planejamento de unidade;

      iii. planejamento de aula.

5. Do ponto de vista didático, planejar é prever os conhecimentos a serem trabalhados e organizar as atividades e experiências de ensino-aprendizagem consideradas mais adequadas para a consecução dos objetivos estabelecidos, levando em conta a realidade dos alunos, suas necessidades e interesses.

O planejamento existe em todas as esferas da educação e do ensino. Dada a importância do planejamento no processo de ensino aprendizagem, utilizaremos em nosso trabalho o planejamento didático de unidade de ensino que será composto por quatro planejamentos de aulas. Estes dois planejamentos estão contidos no planejamento didático ou de ensino.

3 Desenvolvimento Teórico Matemático

Neste capítulo apresentaremos os pré-requisitos necessários que serão utilizados para a compreensão da correlação e regressão linear.

3.1 Funções

3.1.1 Produto Cartesiano e Relação Binária

Dados dois conjuntos A e B, o **produto cartesiano** A x B é o conjunto de todos os pares ordenados (x, y), onde $x \in A$ e $y \in B$. Como exemplo tomemos dois conjuntos $A = \{2, 3, 4\}$ e $B = \{2, 3, 4, 5, 6\}$, o produto cartesiano de A por B é o conjunto de todos os pares ordenados. Assim temos:

$$A \text{ x } B = \{(x, y) | x \in A \text{ e } y \in B\}$$

Dados dois conjuntos quaisquer A e B, uma **relação binária** entre A e B é um subconjunto obtido do produto cartesiano A x B desses conjuntos. Uma **relação binária** de A em B é um conjunto R de pares ordenados, onde o 1º elemento de cada par vem de A e o 2º vem de B, ou seja, R está contido no produto cartesiano de A x B. Assim temos:

R é uma relação binária de A em $B \iff R \subset A \text{ x } B$.

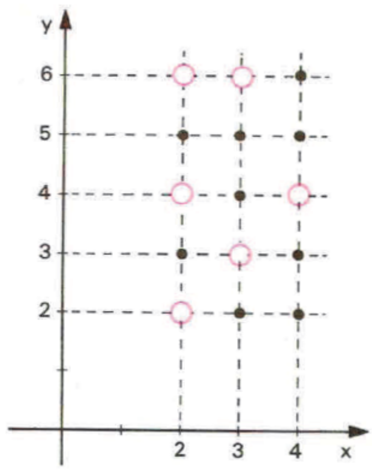
Como exemplo, tomemos os dois conjuntos acima $A = \{2, 3, 4\}$ e $B = \{2, 3, 4, 5, 6\}$, e a partir daí, construímos o conjunto, R_1, dos pares ordenados (x, y), sendo x divisor de y, assim temos o seguinte conjunto:

$$R_1 = \{(x, y) \in A \text{ x } B|x|y\} = \{(2, 2), (2, 4), (2, 6), (3, 3), (3, 6), (4, 4)\}$$

R_1 é uma relação binária de A em $B \iff D \subset A \text{ x } B$.

A Figura 3 representa o produto cartesiano de A x B e o subconjunto R contido em A x B.

Figura 3. Produto Cartesiano A x B e o subconjunto R contido em A x B

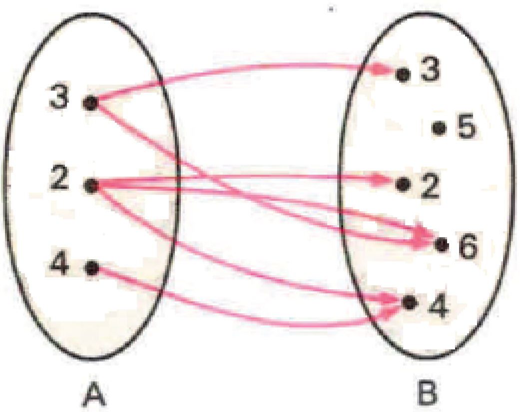


Fonte: Iezzi e Murakami (1977)

A relação binária também pode ser representada graficamente através do diagrama de flechas. A representação, dessa forma, da relação binária de R é mostrada na Figura 4:

Vamos utilizar a seguinte nomenclatura:

Figura 4. Diagrama de flechas da Relação Binária



Fonte: Iezzi e Murakami(1977)

A: conjunto de partida da relação binária de A em B.

B: conjunto de chegada ou contradomínio da relação binária de A em B.

3.1.2 Domínio e imagem

Dados dois conjuntos quaisquer A e B, o subconjunto R, do produto cartesiano A x B é uma relação binária e chamamos de **Domínio** de R, o conjunto, $D(R)$, formado pelos elementos $x \in A$ para os quais existe algum $y \in B$ tal que $(x, y) \in R$. A **Imagem** de R, $Im(R)$, é definida pelos elementos $y \in B$ para os quais existe algum $x \in A$ tal que $(x, y) \in R$. Assim se $R \subset A$xB, então temos:

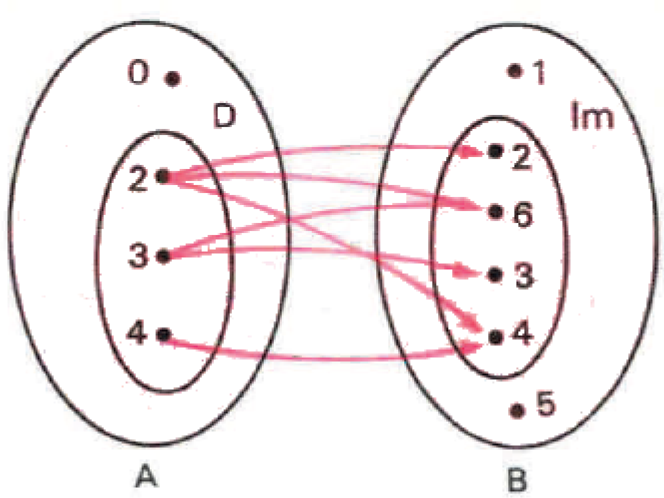
$$D(R) = \{x \in A | \exists y \in B, (x, y) \in R\}$$

$$Im(R) = \{y \in B | \exists x \in A, (x, y) \in R\}$$

Assim na relação binária, R, contida no produto cartesiano, A x B, o Domínio, $D(R)$, está contido em A, ou seja, é um subconjunto de A, e a Imagem, $Im(R)$ está contida no contradomínio B, ou seja, a imagem é um subconjunto do contradomínio.

Tomemos como exemplo dois conjuntos $A = \{0, 2, 3, 4\}$ e $B = \{1, 2, 3, 4, 5, 6\}$ e a relação binária $R_2 = \{x, y) \in A$ x $B | y$ é múltiplo de $x\}$, temos então o domínio e a imagem como mostra o diagrama de flechas da Figura 5:

Figura 5. Diagrama de flechas - Domínio, Contradomínio e Imagem



Fonte: Iezzi e Murakami(1977)

A Figura5, mostra os dois conjuntos A e B, e uma relação binária, R_2, entre A e B. O domínio D está contido no conjunto A e a Imagem Im, está contida no conjunto B. Assim os pares ordenados da relação binária R_2, subconjunto do produto cartesiano A x B, o domínio D que é subconjunto do conjunto A, e a imagem, Im, que é subconjunto do conjunto B são:

$R_2 = \{(2, 2), (2, 4), (2, 6), (3, 3), (3, 6), (4, 4)\}$

$D = \{(2, 3, 4)\}$

$Im = \{(2, 3, 4, 6)\}$

3.1.3 Definição de função

Sejam dois conjuntos A e B, não vazios, uma relação f de A em B recebe o nome de *função definida em A com imagens em B*, se e somente se, para todo $x \in A$ exista um só $y \in B$ tal que $(x, y) \in f$ (IEZZI; MURAKAMI, 1977).

f é uma aplicação de A em $B \Leftrightarrow (\forall\ x \in A, \exists\ y \in B|(x, y) \in f)$

Duas condições fundamentais devem ser satisfeitas para a aplicação f seja uma função:

i) - Estar definida em todos os elementos do domínio.

ii) - Não fazer corresponder mais de um elemento no contradomínio a cada elemento do domínio.

3.1.4 Classificação das funções segundo a sua Imagem

Injetora É toda a função onde cada elemento y do conjunto *Imagem* esta relacionada a um e somente um elemento x do domínio da função. A Figura 6 é um diagrama de setas onde temos os conjuntos A e B que representam o domínio e contradomínio respectivamente da função. Cada elemento do conjunto A, está associado a somente um elemento do conjunto B. Esta função é um exemplo de função injetora.

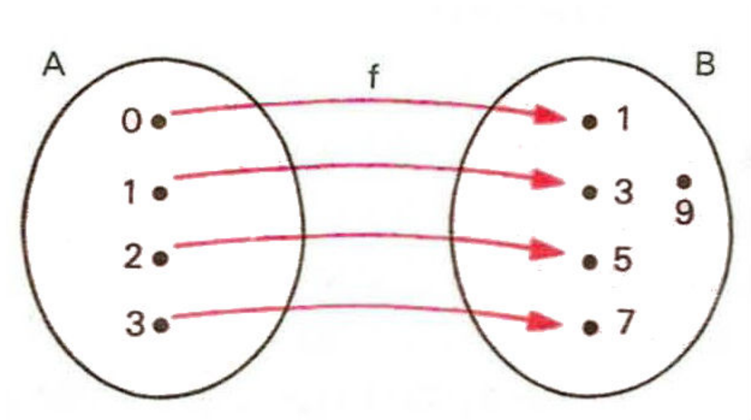
$f : A \to B$

f é injetora $\Leftrightarrow (\forall x_1 \in A, \forall x_2 \in A)(x_1 \neq x_2 \Rightarrow f(x_1) \neq f(x_2))$

ou

$f : A \to B$

Figura 6. Diagrama de flechas - função injetora



Fonte: Iezzi e Murakami(1977)

f é injetora $\Leftrightarrow (\forall x_1 \in A, \forall x_2 \in A)(f(x_1) = f(x_2) \Rightarrow x_1 = x_2)$

Sobrejetora É a função f onde todos os elementos do contradomínio estão relacionados com um elemento x do domínio, ou seja, o conjunto imagem é igual ao contradomínio. Assim tomamos os conjuntos A e B, sendo respectivamente domínio e contradomínio da função temos:

$f : A \to B$

f é sobrejetora $\Leftrightarrow \forall y \in B, \exists x \in A \mid f(x) = y$

ou

$f : A \to B$

f é sobrejetora $\Leftrightarrow Im(f) = B$

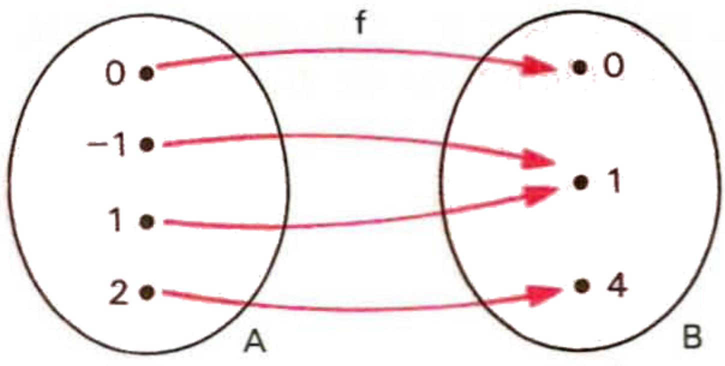
A diagrama de flechas da Figura 7 é um exemplo de função sobrejetora.

Bijetora É a função que é ao mesmo tempo injetora e bijetora;

$f : A \to B$

f é bijetora $\Rightarrow f$ é sobrejetora e injetora

Figura 7. Diagrama de flechas - função sobrejetora



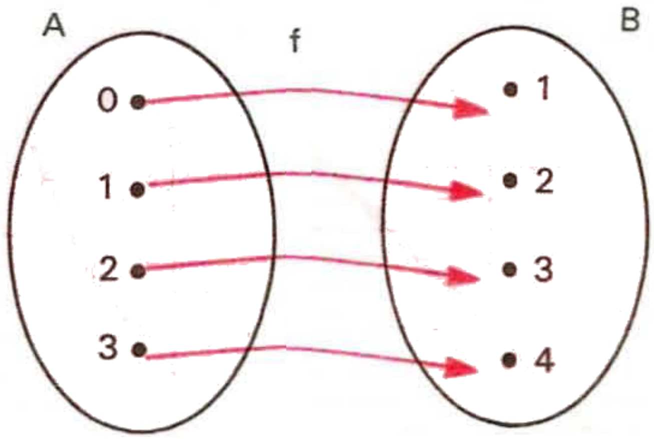
Fonte: FME 01 - Conjuntos e Funções.

ou

$f : A \to B$

f é bijetora $\Leftrightarrow \forall y, y \in B, \exists! x, x \in A \mid f(x) = y$

Figura 8. Diagrama de flechas - função bijetora.



Fonte: FME 01 - Conjuntos e Funções

3.2 Função Afim

A principal característica da função afim é que ela varia a uma taxa constante. Entre as aplicações podemos citar: O Movimento Retilíneo Uniforme, a fórmula para conversão de unidades de medida de temperatura Celsius e Fahrenheit, o modelo do ajuste linear nos problemas de modelagem matemática (ADAMI; FILHO; LORANDIL, 2015).

3.2.1 Definição

Toda função polinomial de primeiro grau é chamada de *função afim.* Segundo Iezzi e Murakami (1977): "Uma aplicação de $\mathbb{R}$ em $\mathbb{R}$ recebe o nome de função afim quando cada $x \in \mathbb{R}$ estiver associado o elemento $(ax+b) \in \mathbb{R}$", com $a, b \in \mathbb{R}$ e $a \neq 0$, definida como:

$$
\begin{aligned}
f : \mathbb{R} &\to \mathbb{R} \\
x &\mapsto ax + b
\end{aligned}
$$

Segundo Adami, Filho e Lorandil (2015)a constante a é denominada de *coeficiente angular* da função e a constante b é denominada *coeficiente linear* e as principais características da função afim são:

(a) O domínio da função afim é o conjunto dos números reais, isto é, $Dom(f) = \mathbb{R}$.
(b) O gráfico de uma função afim é uma reta.
(c) O sinal do coeficiente angular a (positivo ou negativo) indica a inclinação da reta.
(d) O valor do *coeficiente linear* b determina o ponto em que a reta corta o eixo vertical; isto é, $y = b$ é o intercepto vertical.
(e) O zero da função ocorre em $-\frac{b}{a}$; isto é, $x = -\frac{b}{a}$ é o intercepto horizontal.

3.2.2 Determinação do coeficiente angular a da função afim

O coeficiente angular a tem um importante papel na equação da reta: ele determina a sua inclinação. Seja $(x_1, f(x_1))$ e $(x_2, f(x_2))$, dois pontos distintos de uma reta que representa a função f. O coeficiente angular a é dado por:

$$a = \frac{f(x_1) - f(x_2)}{x_1 - x_2} \tag{3.1}$$

O coeficiente angular a de uma função afim determina o crescimento ou decrescimento desta. Segundo Iezzi e Murakami (1977): "A função afim $f(x) = ax + b$ é crescente se, e somente se, coeficiente angular a for positivo."Segue a demonstração:

$$\begin{aligned} f(x) = ax + b \text{ é crescente} \quad &\Longleftrightarrow \quad = \frac{f(x_1) - f(x_2)}{x_1 - x_2} > 0 \Longleftrightarrow \\ \frac{(ax_1 + b) - (ax_2 + b)}{x_1 - x_2} > 0 \quad &\Longleftrightarrow \quad \frac{a(x_1 - x_2)}{x_1 - x_2} > 0 \Longleftrightarrow a > 0 \end{aligned}$$

Segundo Iezzi e Murakami (1977): "A função afim $f(x) = ax+b$ é decrescente se, e somente se, coeficiente angular a for negativo."Segue a demonstração:

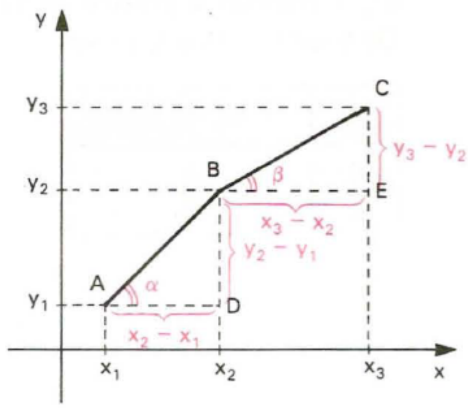
$$\begin{aligned} f(x) = ax + b \text{ é decrescente} \quad &\Longleftrightarrow \quad = \frac{f(x_1) - f(x_2)}{x_1 - x_2} < 0 \Longleftrightarrow \\ \frac{(ax_1 + b) - (ax_2 + b)}{x_1 - x_2} < 0 \quad &\Longleftrightarrow \quad \frac{a(x_1 - x_2)}{x_1 - x_2} < 0 \Longleftrightarrow a < 0 \end{aligned}$$

3.2.3 O gráfico da função afim

Teorema: "O Gráfico cartesiano da função $f(x) = ax + b(a \neq 0)$ é uma reta." Segue a demonstração segundo Iezzi e Murakami (1977):

Sejam A, B, C três pontos quaisquer, distintos dois a dois, do gráfico cartesiano da função $y = ax + b$ $(a \neq 0)$ e $(x_1, y_1), (x_2, y_2)$ e (x_3, y_3), respectivamente, as coordenadas cartesianas desses pontos. Figura 9.

Figura 9. Reta função afim



Fonte: Iezzi e Murakami (1977)

Para provarmos que os pontos A, B e C pertencem a mesma reta, mostremos, inicialmente, que os triângulos retângulos ABD e BCE são semelhantes. De fato:

$$(x_1, y_1) \in f \Rightarrow y_1 = ax_1 + b \tag{3.2}$$

$$(x_2, y_2) \in f \Rightarrow y_2 = ax_2 + b \tag{3.3}$$

$$(x_3, y_3) \in f \Rightarrow y_3 = ax_3 + b \tag{3.4}$$

Subtraindo membro a membro, temos:

$$\left.\begin{array}{l} y_3 - y_2 = a(x_3 - x_2) \\ y_2 - y_1 = a(x_2 - x_1) \end{array}\right\} \Rightarrow \frac{x_3 - x_2}{y_3 - y_2} = \frac{x_2 - x_1}{y_2 - y_1} = a$$

Os triângulos ABD e BCE são retângulos e tem lados proporcionais, então são semelhantes e, portanto, $\alpha = \beta$. Segue-se que os pontos A, B e C estão alinhados.

3.2.4 Função crescente

Segundo: Iezzi e Murakami (1977): "a função: $f : A \to B$ definida por $y = f(x)$ é *crescente* no conjunto $A_1 \subset A$ se, para dois valores quaisquer x_1 e x_2 pertencentes a A_1 com $x_1 < x_2$, tivermos $f(x_1) < f(x_2)$." Assim f é crescente quando:

$$(\forall x_1, x_2)(x_1 < x_2 \Rightarrow f(x_1) < f(x_2))$$

$$(\forall x_1, x_2)(x_1 \neq x_2 \Rightarrow \frac{f(x_1) - f(x_2)}{x_1 - x_2} > 0)$$

3.2.5 Função decrescente

Segundo: Iezzi e Murakami (1977): "a função: $f : A \to B$ definida por $y = f(x)$ é *decrescente* no conjunto $A_1 \subset A$ se, para dois valores quaisquer x_1 e x_2 pertencentes a A_1 com $x_1 < x_2$, tivermos $f(x_1) > f(x_2)$." Assim f é decrescente quando:

$$(\forall x_1, x_2)(x_1 < x_2 \Rightarrow f(x_1) > f(x_2))$$

$$(\forall x_1, x_2)(x_1 \neq x_2 \Rightarrow \frac{f(x_1) - f(x_2)}{x_1 - x_2} < 0)$$

3.3 Somatórios

A letra maiúscula grega $\sum$ (sigma) é utilizada para indicar o somatório , assim temos:

$$\sum_{i=1}^{n} x_i = x_1 + x_2 + x_3 + \cdots + x_n \tag{3.5}$$

Em operações com somas, as seguintes regras devem ser levadas em consideração:

a) O somatório de uma constante é igual ao produto do número de termos pela constante:

$$\begin{aligned} \sum_{i=1}^{n} a &= \underbrace{(a + a + a + a + \cdots + a)}_{n} \\ &= na \\ &= \sum_{i=1}^{n} a \end{aligned}$$

b) Se a é uma constante então pode ser colocada para fora do somatório:

$$\begin{aligned} \sum_{i=1}^{n} ax_i &= ax_1 + ax_2 + ax_3 + \cdots + ax_n \\ &= a(x_1 + x_2 + x_3 + \cdots + x_n) \\ &= a\sum_{i=1}^{n} x_i \end{aligned}$$

c) Se x e y são duas variáveis, então:

$$\begin{aligned} \sum_{i=1}^{n}(x_i + y_i) &= (x_1 + y_1) + (x_2 + y_2) + (x_3 + y_3) + \cdots + (x_n + y_n) \\ &= (x_1 + x_2 + x_3 + \cdots + x_n) + (y_1 + y_2 + y_3 + \cdots + y_n) \\ &= \sum_{i=1}^{n} x_i + \sum_{i=1}^{n} y_i \end{aligned}$$

3.4 Introdução a Estatística

A Estatística é uma ciência que em seu início foi aplicada em aspectos do estado, cresceu e faz parte de praticamente todas as áreas do conhecimento humano como afirma Spiegel (1985):

> A Estatística, ou métodos estatísticos, como é denominada algumas vezes, desempenha papel crescente e importante em quase todas as fases da pesquisa humana. Lidando anteriormente apenas com os negócios de Estado, donde o seu nome, a influência da estatística estendeu se agora à agricultura, biologia, comércio, química, comunicações, economia, educação, eletrônica, medicina, física, ciências políticas, psicologia, sociologia e outros numerosos campos da ciência e engenharia.

Sua origem remota a milhares de anos quando os governos realizavam o censo, para conhecer seus habitantes como afirma Correa (2003):

> A palavra estatística lembra, à maioria das pessoas, recenseamento. Os censos existem há milhares de anos e constituem um esforço imenso e caro feito pelos governos, com o objetivo de conhecer seus habitantes, sua condição socioeconômica, sua cultura, religião, etc. Portanto, associar estatística a censo é perfeitamente correto do ponto de vista histórico, sendo interessante salientar que as palavras estatística e estado têm a mesma origem latina: status.

3.4.1 Definição

Estatística vem do latim e significa status, estado, ou, talvez, situação. Podemos definir *estatística*, com sendo um processo com o objetivo de observar, classificar e analisar fenômenos coletivos, procurando definir as leis que eles obedecem, apresentando os resultados na forma de tabelas e gráficos (RODRIGUES, 1970).

As primeiras utilizações da estatística, foram dados e gráficos que definiam aspectos de um estado ou país. John Graunt[1] publicou dados estatísticos em 1662, e seu trabalho foi secundado como afirma Triola (1998):

> A primitiva utilização da estatística envolvia compilações de dados e gráficos que descreviam vários aspectos de um estado ou país. Em 1662, John Graunt publicou informes estatísticos sobre nascimentos e mortes. O trabalho de Graunt foi secundado por estudos de mortalidade, tamanho de populações, rendas e taxas de desemprego. As famílias, os governos e as empresas se apoiam largamente em dados estatísticos. Assim é que as taxas de desemprego, de inflação, os índices do consumidor, as taxas de natalidade e mortalidade são calculadas cuidadosamente a intervalos regulares, e seus resultados são utilizados por empresários para tomarem decisões que afetam a futura contratação de empregados, níveis de produção e expansão para novos mercados.

Triola (1998) define: "a estatística é uma coleção de métodos para planejar experimentos, obter dados e organizá-los, resumí-los, analisá-los, interpretá-los e deles extrair conclusões.

3.4.2 Estatística descritiva e inferencial

A *Estatística* é dividida em duas áreas: Descritiva e Inferencial descritas como:

A Estatística Descritiva segundo Mann (2015): "consiste em métodos para se organizar, exibir e descrever dados utilizando tabelas, gráficos e medidas resumidas."Quando a coleta de dados é extensa como em pesquisas feitas por órgãos estaduais ou federais, sua analise fica pouco proveitosa para extrair conclusões ou tomar decisões, utilizamos os métodos da estatística descritiva para resumi-los em tabelas e gráficos, reduzindo a um tamanho manejável (MANN, 2015).

[1] John Graunt (Londres, 24 de abril de 1620 – Londres, 18 de abril de 1674) foi um cientista e demógrafo britânico, precursor na construção de Tábuas de Mortalidade.

Sendo chamado na estatística de *população* todos os elementos de interesse, e *amostra* alguns itens desta população, a estatística inferencial são métodos que tomam decisões a partir destas amostras. Mann (2015) define: "Estatística Inferencial consiste em métodos que utilizam resultados de amostras para auxiliar na tomada de decisão ou na realização de prognósticos sobre uma população."

3.4.3 Variável quantitativa e qualitativa

Dizemos que uma variável é um símbolo, como uma letra do alfabeto A, Z, x, y, H que pode assumir qualquer um do conjunto de valores que são atribuídos. Podemos classificar as variáveis como:

Variáveis Quantitativas: elas podem ser medidas em uma escala quantitativa, por isso apresentam valores numéricos que fazem sentido, e podem ser continua ou discreta. Spiegel (1985) define assim: "Uma variável que pode assumir teoricamente qualquer valor entre dois dados chama-se *variável continua*; de outro modo denomina-se *variável discreta*."

Assim como exemplo de variável contínua podemos ter a altura H de um individuo que pode ser 1,75 m, 1,773 m ou 1,7499 m, conforme a precisão da medida.

Como exemplo de variável discreta podemos citar o número N de pessoas inscritas em um curso, que os valores podem ser 0,1,2,3,4,..25,26.. mas não pode ser 25,5 ou 3,5.

Variáveis Qualitativas: elas não possuem valores numéricos quantitativos, e são definidas por categorias, representando uma classificação de indivíduos, podem ser classificadas em nominais e ordinais. Mann (2015) define assim:

> Variável Qualitativa ou Variável Categórica: Uma variável que não pode assumir um valor numérico, mas pode ser classificada em duas ou mais categorias não numéricas, denomina-se variável qualitativa ou variável categórica. Os dados coletados em tal variável são chamados de dados qualitativos.

Segundo Machado (2010) caracteriza variáveis qualitativas nominais como: "Não possui ordenamento nem hierarquia dos atributos analisados." Podemos citar como exemplo: sexo, cor dos olhos, fumante/ não fumante.

"Qualitativas ordinais: São equivalentes às variáveis nominais, porém, incluindo uma ordem ou uma hierarquia" Machado (2010). Exemplos são a escolaridade (1º, 2º, 3º graus), estágio de uma doença (inicial, intermediaria, terminal), mês de observação (janeiro, fevereiro,...., dezembro).

3.4.4 Variáveis dependente e independente

Uma variável é chamada *dependente* e indicamos por Y, quando em uma relação a outra variável denominada *independente* e chamamos de X , é estudada, em Estatística na Regressão Linear (VIEIRA, 2015).

Segundo Marconi e Lakatos (2017), sua definição de variáveis dependente e independente são:

Variável independente (X): é aquela que influencia, determina ou afeta outra variável é fator determinante, condição ou causa para determinado resultado, efeito ou consequência.

Variável dependente (Y): consiste em valores (fenômenos, fatores) a serem explicados ou descobertos, em virtude de serem influenciados, determinados ou afetados pela variável independente.

3.4.5 Medidas de Tendência Central

"Uma *medida de tendência central* é um valor que representa uma entrada típica, ou central, de um conjunto de dados"(LARSON; FARBER, 2004). Fazendo parte da Estatística Descritiva, a média, moda e mediana são as principais medidas de tendência central. Seguem as definições e a forma de calcular as medidas.

A *Média Aritmética* ou simplesmente *Média*, é uma das medidas mais conhecidas "... consiste na somatória de todas as entradas dividida pela quantia total de ocorrências"(BONAFINI, 2011). Pode ser calculada com o total da população (média populacional representado por μ) ou com uma amostra da população (média amostral representado por $\bar{x}$). Assim temos:

$$\text{Média Populacional: } \mu = \frac{\sum x}{N} \tag{3.6}$$

$$\text{Média Amostral: } \bar{x} = \frac{\sum x}{n} \tag{3.7}$$

A *Mediana* é definida por Larson e Farber (2004) como:

> A *Mediana* de um conjunto de dados é o dado que fica no meio quando as entradas são colocadas em ordem crescente ou decrescente. Se o conjunto de dados tiver um número par de entradas, a mediana sera a média entre os dois pontos que estiverem no meio do conjunto.

A *Moda* é o dado dentro do conjunto de dados que ocorre com maior frequência (LARSON; FARBER, 2004). Quando não há repetições na entrada, então o conjunto de dados não possui moda. São chamados de *Bimodais* em um conjunto de dados, se ocorrerem duas ou mais entradas com a mesma frequência.

3.4.6 Medidas de Dispersão

"As medidas que nos ajudam a saber alguma coisa sobre a dispersão de um conjunto de dados são chamadas de *medidas de dispersão*" (MANN, 2015). Elas são a amplitude, a variância e o desvio-padrão, descritas abaixo:

Segundo (MANN, 2015): "A *amplitude* representa a medida de dispersão mais simples de se calcular. A amplitude é obtida tomando-se a diferença entre o maior valor e o menor valor em um conjunto de dados." Assim sendo x_{max}, o maior valor e x_{min} o menor valor, em um conjunto de dados a amplitude A é determinada por:

$$A = x_{max} - x_{min} \tag{3.8}$$

A amplitude utiliza-se somente de duas medidas do conjunto de dados. A variância e o desvio-padrão, utilizam-se de todas as entradas da população ou amostra. Antes de vermos estas medidas precisamos saber o que é desvio de uma entrada. Segundo (LARSON; FARBER, 2004): "O desvio de uma entrada x em um conjunto de dados de uma população é a diferença entre a entrada e a média μ do conjunto de dados", assim chamamos de Δx o desvio de cada entrada x e é determinado por:

$$\Delta x = x - \mu \tag{3.9}$$

Para o cálculo da *variância populacional* de um conjunto de dados de uma população com N entradas segundo Larson e Farber(2004) é:

$$\text{Variância populacional} : \sigma^2 = \frac{\sum(x-\mu)^2}{N} \tag{3.10}$$

Ou seja, a *variância populacional* σ^2 , é a soma de todos os desvios elevado ao quadrado e dividido pelas N entradas de dados. O *desvio padrão populacional*, considerando um conjunto de dados populacionais com N entradas, é a raiz quadrada da variância populacional (LARSON; FARBER, 2004), e é definida por:

$$\text{Desvio padrão populacional}: \ \sigma = \sqrt{\sigma^2} = \sqrt{\frac{\sum(x-\mu)^2}{N}} \tag{3.11}$$

Analogamente ao conjunto de dados de uma população temos os desvios de cada entrada x e a média $\bar{x}$ em um conjunto de amostras e é representado por Δx e definido como:

$$\Delta x = x - \bar{x} \tag{3.12}$$

Segundo (LARSON; FARBER, 2004) Considerando um conjunto de dados com n entradas. A *variância* e *desvio padrão amostrais* são dados por:

$$\text{Variância amostral}: \ s^2 = \frac{\sum(x-\bar{x})^2}{n-1} \tag{3.13}$$

$$\text{Desvio padrão amostral}: \ s = \sqrt{s^2} = \sqrt{\frac{\sum(x-\bar{x})^2}{n-1}} \tag{3.14}$$

3.4.7 Gráfico de Dispersão

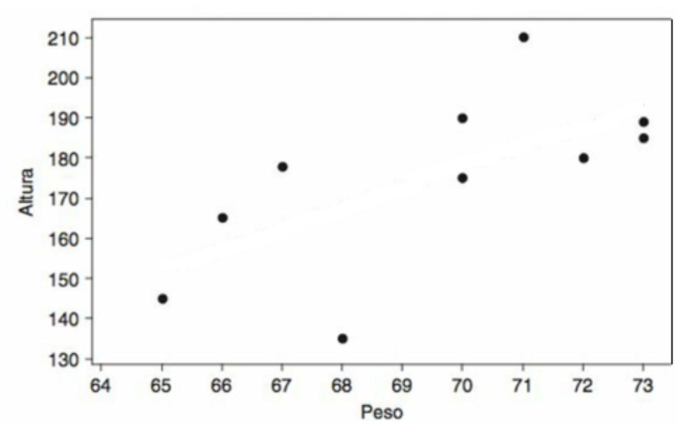
Segundo Levine, Stephan e Szabat (2017):

> Um gráfico de dispersão explora a possível relação entre duas variáveis numéricas, ao inserir no gráfico os valores correspondentes

> a uma variável numérica no eixo horizontal, ou X, e os valores correspondentes a uma segunda variável numérica no eixo vertical, ou Y.

Assim podemos verificar se existe relação por exemplo entre renda e consumo, anos de estudo e salário, gastos em propaganda e venda, ou quaisquer outras duas variáveis que possa ter relação entre elas. Assim denominando as variáveis de X e Y, construímos um gráfico com os pares ordenados. A Figura 10, mostra um gráfico de dispersão com a relação entre o peso em quilogramas e a altura em centímetros.

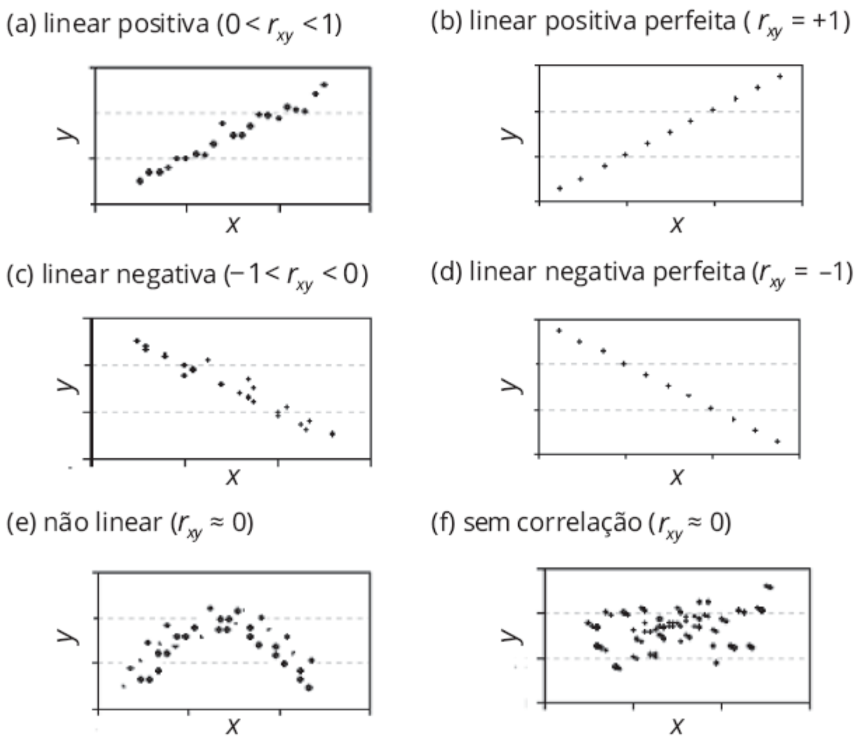
Figura 10. Gráficos de Dispersão: Peso versus Altura



Fonte: Larson e Farber (2004)

Através dos gráficos de dispersão podemos explicar a tendência dessa relação e através do coeficiente de correlação "r", podemos quantificar a força que uma variável exerce sobre a outra ou seja o grau de associação linear. Os gráficos de dispersão a seguir mostram que as correlações podem ser positivas, ou seja, quando as variáveis x e y crescem no mesmo sentido, como está representado nas Figura 11-(a) e 11-(b), podem ser negativas, ou seja, quando as variáveis x e y crescem em sentido opostos, como está representado nas Figura 11-(c) e 11-(d), podem ser não linear representada na Figura 11-(e) e sem correlação como mostra a Figura 11-(f). Nestes casos denominamos de r_{xy} o coeficiente de relação entre x e y. representandos nos gráficos da Figura 11.

Figura 11. Tipos de correlação entre variáveis



Fonte: (MATTOS; KONRATH; AZAMBUJA, 2017)

3.5 Análise de Regressão

Segundo Gurajati (2011) a definição de Analise de Regressão é:

> A analise de regressão diz respeito ao estudo da dependência de uma variável, a *variável dependente*, em relação a uma ou mais variáveis, as *variáveis explanatórias*, visando estimar e/ou prever o valor médio (da população) da primeira em termos dos valores conhecidos ou fixados (em amostragem repetidas) das segundas.

Dessa forma é possível estabelecer uma relação matemática entre essas duas variáveis, sua utilização é largamente usada em muitas áreas do conhecimento e tem por objetivos: Predição, Seleção de Variáveis, Estimação de Parâmetros, Inferência (ACTION, 2015).

3.6 Correlação

Segundo o dicionário Michaelis (2008) o significado de *correlação* na área de Estatística é: "Dependência recíproca entre duas ou mais variáveis. Assim denominamos de "r" o coeficiente que mede o grau de relacionamento entre duas variáveis X e Y (Coeficiente de Correlação de Pearson [2]), sua intensidade e direção , e seu valor fica no intervalo $-1 < r < 1$. Para podermos calcular o valor de r, vamos entender o que é e como calcular a Covariância.

3.6.1 Covariância

Assim como a variância, medida de dispersão da estatística, é utilizado para o cálculo do desvio padrão, a covariância é utilizado para o cálculo do "r" que verifica o grau de relacionamento linear entre duas variáveis. Através dos desvios dos conjuntos de dados calculamos a covariância entre duas variáveis X e Y. Segundo (SARTORIS, 2013) : "A covariância pode ser entendida como uma *variância conjunta* entre duas variáveis. Enquanto a variância sai de quadrados (da variável menos a média), a covariância é definida por meio de produtos". Segue as Equações para a determinação da covariância populacional e amostral:

$$C(X,Y) = \frac{\sum_{i=1}^{n}(x_i - \overline{x}) \cdot (y_i - \overline{y})}{N} \tag{3.15}$$

[2]Matemático Britânico - 1857- 1936

A Formula 3.16, representa o cálculo:

$$C(X,Y) = \frac{\sum_{i=1}^{n}(x_i - \overline{x}) \cdot (y_i - \overline{y})}{n-1} \tag{3.16}$$

onde:

x_i e y_i são dados dos dois conjunto de dados;

$\overline{x}$ é a média aritmética dos valores de X;

$\overline{y}$ é a média aritmética dos valores de Y;

n é o número de amostras

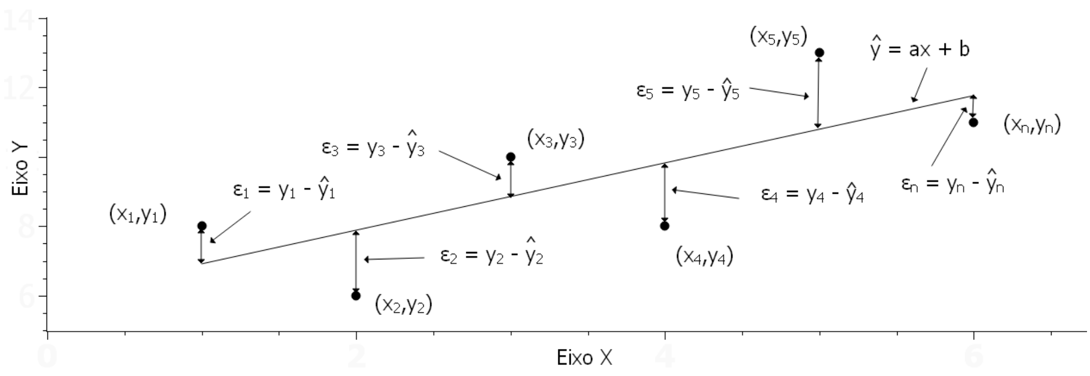
3.7 Método do Mínimos Quadrados

O gráfico da Figura 12, contém os pares ordenados (x_i, y_i), com $x_i \in X$ e $X = \{x_i, ..., x_n\}$ e $y_i \in Y$ e $Y = \{y_i, ..., y_n\}$, e a reta determinada pela equação $\hat{y} = ax_i + b$. Chamaremos de resíduo, ϵ_i, a diferença entre o valor dado $y_i \in Y$ e o valor estimado do $\hat{y}_i$, determinado pela equação $\hat{y}_i = ax_i + b$. A reta que apresenta a menor soma dos resíduos ao quadrado, $(\epsilon_1)^2$, é chamada de reta de regressão e o método dos mínimos quadrados é utilizado para a determinação desta reta. Cada resíduo ϵ_i é determinado por:

$$\epsilon_i = y_i - \hat{y}_i \Rightarrow \epsilon_i = y_i - (ax_i + b) \Rightarrow \epsilon_i = y_i - b - ax_i$$

O Método dos Mínimos Quadrados tem o seguinte princípio segundo Hill, Griffiths e Judge (2010): "Este princípio afirma que, para ajustar uma reta aos valores dados, devemos procurar a reta tal que a soma dos quadrados das distâncias verticais de cada ponto à reta seja a menor possível." Estas

Figura 12. Gráfico de Dispersão



Fonte: Autoria Própria

retas verticais são os resíduos e no Gráfico de Dispersão da Figura 12 são representado pela letra grega "ϵ". São utilizados os quadrados das distâncias para evitar que valores positivos da distância não anule os negativos. Assim a Equação (3.17), representa a soma de todos os erros ou resíduos elevados ao quadrado é dada por:

$$\sum_{i=1}^{n} \epsilon_t^2 \Rightarrow \sum_{i=1}^{n} (y_i - b - ax_i)^2 \tag{3.17}$$

O objetivo e minimizar a soma dos resíduos que denominaremos de EQ (Erros Quadrados). Para minimizar EQ, vamos diferenciar parcialmente em relação a a e b que são os estimadores da reta de regressão. Vamos derivar em relação ao parâmetro a e b, respectivamente:

$$\frac{\partial EQ}{\partial a} = 0 \Rightarrow \frac{\partial}{\partial a}\left(\sum_{i=1}^{n} (y_i - b - ax_i)^2 = 0 \right) \Rightarrow -2 \sum_{i=1}^{n} (y_i - b - ax_i) x_i = 0 \tag{3.18}$$

$$\frac{\partial EQ}{\partial b} = 0 \Rightarrow \frac{\partial}{\partial b}\left(\sum_{i=1}^{n} (y_i - b - ax_i)^2 = 0 \right) \Rightarrow -2 \sum_{i=1}^{n} (y_i - b - ax_i) = 0 \tag{3.19}$$

Utilizando as propriedades dos somatórios nas Equações (3.18) e (3.19) temos:

$$\sum_{i=1}^{n}(y_i - b - ax_i)x_i = 0 \Rightarrow \sum_{i=1}^{n} x_i y_1 = b\sum_{i=1}^{n} x_i + a\sum_{i=1}^{n} x_i^2 \tag{3.20}$$

$$\sum_{i=1}^{n}(y_i - b - ax_i) = 0 \Rightarrow \sum_{i=1}^{n} y_1 = nb + a\sum_{i=1}^{n} x_1 \tag{3.21}$$

Dividindo ambos os lados da Equação 3.21 por n, com $n \neq 0$, temos:

$$\overline{y} = b + a\overline{x} \Rightarrow b = \overline{y} - a\overline{x} \tag{3.22}$$

Substituindo o valor de b determinado na Equação 3.22 na Equação 3.20, temos:

$$\sum_{i=1}^{n} x_i y_1 = (\overline{y} - a\overline{x})\sum_{i=1}^{n} x_i + a\sum_{i=1}^{n} x_1^2 \Rightarrow \sum_{i=1}^{n} x_i y_1 = \overline{y}\sum_{i=1}^{n} x_i - a\overline{x}\sum_{i=1}^{n} x_i + a\sum_{i=1}^{n} x_1^2$$

Isolando o parâmetro a teremos:

$$a = \frac{\displaystyle\sum_{i=1}^{n} x_i y_i - \overline{y}\sum_{i=1}^{n} x_i}{\displaystyle\sum_{i=1}^{n} x_i^2 - \overline{x}\sum_{i=1}^{n} x_i} \tag{3.23}$$

Podemos determinar o parâmetro a pela Equação (3.24), como mostra:

$$a = \frac{s_{xy}}{s_x^2} \tag{3.24}$$

onde:

s_{xy} : Covariância entre x e y.

s_x^2 : Variância de x.

Assim temos:

$$a = \frac{s_{xy}}{s_x^2} \Rightarrow a = \frac{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2} \Rightarrow a = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \tag{3.25}$$

Desenvolvendo o numerador da Equação (3.25) teremos:

$$\begin{aligned}
\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y}) &= \sum_{i=1}^{n}(x_i y_i - \bar{y}x_i - \bar{x}y_i + \bar{x}\bar{y}) \\
&= \sum_{i=1}^{n} x_i y_i - \bar{y}\sum_{i=1}^{n} x_i - \bar{x}\sum_{i=1}^{n} y_i + n\bar{x}\bar{y} \\
&= \sum_{i=1}^{n} x_i y_i - \bar{y}\sum_{i=1}^{n} x_i - n\bar{x}\bar{y} + n\bar{x}\bar{y} \\
&= \sum_{i=1}^{n} x_i y_i - \bar{y}\sum_{i=1}^{n} x_i
\end{aligned}$$

Desenvolvendo o denominador da Equação (3.25) teremos:

$$\begin{aligned}
\sum_{i=1}^{n}(x_i - \bar{x})^2 &= \sum_{i=1}^{n}(x_i^2 - 2x_i\bar{x} + \bar{x}\bar{x}) \\
&= \sum_{i=1}^{n} x_i^2 - 2\bar{x}\sum_{i=1}^{n} x_i + n\bar{x}\bar{x} \\
&= \sum_{i=1}^{n} x_i^2 - 2n\bar{x}^2 - n\bar{x}^2 \\
&= \sum_{i=1}^{n} x_i^2 - \bar{x}\sum_{i=1}^{n} x_i
\end{aligned}$$

Logo:

$$a = \frac{\sum_{i=1}^{n} x_i y_i - \overline{y} \sum_{i=1}^{n} x_i}{\sum_{i=1}^{n} x_i^2 - \overline{x} \sum_{i=1}^{n} x_i} = \frac{\sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y})} = \frac{s_{xy}}{s_x^2}$$

3.7.1 Cálculo do Coeficiente de Correlação Linear r

O Coeficiente de Correlação Linear é uma medida numérica, que tem por objetivo medir a intensidade da relação linear entre duas variáveis (NAVIDI, 2012). Este coeficiente é geralmente representado pela letra r.

O Coeficiente de Correlação Linear se obtêm dividindo a covariância $C(X,Y)$ pelo produto dos desvios padrão de X e Y, assim a Equação (3.26) mostra como o coeficiente de correlação é determinado:

$$r = \frac{C(X,Y)}{\sigma_x \cdot \sigma_y} \tag{3.26}$$

Podendo ser reescrito na forma da Equação (3.27) a seguir:

$$r = \frac{n \sum x_i y_i - (\sum x_i)(\sum y_i)}{(\sqrt{n \sum x_i^2 - (\sum x_i)^2})(\sqrt{n \sum y_i^2 - (\sum y_i)^2})} \tag{3.27}$$

Sobre a análise numérica do coeficiente de correlação linear, (NAVIDI, 2012) diz:

1. O valor de r está sempre entre -1 e 1;
2. Valores positivos de r indica que a reta de mínimos quadrados tem uma inclinação positiva;
3. Valores negativos de r indica que a reta de mínimos quadrados tem uma inclinação negativa;
4. Valores de r próximos a 1 ou -1 indicam relação linear intensa;
5. Valores de r próximos a 0 indicam uma fraca relação.
6. O coeficiente de correlação mede apenas a relação linear.

3.7.2 Cálculo do Coeficiente de Determinação r^2

O coeficiente de determinação indica quantos por cento a variação explicada pela regressão representa da variação total. Em outras palavras ele determina quão bom é o modelo de regressão, para representação dos dados.

Para determinar o Coeficiente de Determinação r^2, utilizamos a Equação (3.28) como segue:

$$r^2 = \frac{SQR}{SQT} = \frac{\text{variação explicada}}{\text{variação total}} \tag{3.28}$$

onde:

$$\begin{aligned}
SQT &= \sum_{i=1}^{n}(y_i - \bar{y})^2 : && \text{Soma de quadrado total.}\\
SQE &= \sum_{i=1}^{n}(y_i - \hat{y}) : && \text{Soma de quadrado do erro.}\\
SQR &= SQT - SQE: && \text{Soma de quadrados da regressão.}
\end{aligned}$$

4 Planejamento de Unidade de Ensino

A proposta deste capítulo é elaborar um plano de unidade de ensino sobre regressão linear e o método dos mínimos quadrados para o Ensino Médio, determinando algebricamente a equação da reta de regressão linear entre duas variáveis quantitativas, determinando também os coeficientes de determinação e correlação.

4.1 Plano de Unidade de Ensino

O plano de unidade de ensino foi realizado com o objetivo de direcionar o aluno do Ensino Médio, por meio de tarefas simples como o preenchimento de tabelas e a confecção de gráficos de dispersão, perceber visualmente se existe uma correlação entre duas variáveis quantitativas e, posteriormente, a determinação algébrica da equação da reta de regressão e, a partir daí, fazer previsões e analisar tendências. A equação será determinada através da teoria dos métodos dos mínimos quadrados e será determinado o coeficiente de determinação "r^2" e correlação "r". Recomendamos ao professor a utilização de uma calculadora simples para a execução dos cálculos. A Tabela 2, mostra um resumo geral do *Plano de Unidade de Ensino*.

Tabela 2. Resumo Plano de Unidade de Ensino

Descrição	Proposta
Conteúdo	Utilizaremos tabelas, gráficos de dispersão, função afim (equação do primeiro grau), medidas de tendência central e dispersão da estatística.
Objetivos Geral	Apresentar o Método dos Mínimos Quadrados e sua abrangência em diversas áreas do conhecimento humano, correlacionando duas variáveis e quantificando a força desse relacionamento. Assim esperamos que os alunos percebam as conexões da Matemática com outras áreas do conhecimento humano, proporcionando também as conexões dentro da própria Matemática.
Número de aulas	O conteúdo a ser apresentando foi separado em 8 horas aulas, que serão numeradas como A1, A2, A3, A4.
Pré-requisitos	Média aritmética, Variância, Covariância , Equação da reta, tabelas e gráficos.
Turma Indicada	Ensino Médio.
Estratégia	Aula expositiva. Preenchimento de tabelas e gráficos. Cálculo de medidas de tendência central e de dispersão.
Materiais utilizados	Quadro, régua, calculadora simples.

Fonte: Autoria Própria

Utilizaremos como contexto um ambiente hospitalar pediátrico, onde será fornecido ao aluno dados das temperaturas médias diárias do mês de julho e o número de atendimentos com problemas respiratórios. Esta contextualização que estamos sugerindo, pode ser visto como um contexto na área de Medicina ou administrativo-hospitalar. O planejamento de unidade de ensino será dividido em 4 planejamentos de aula, resultando em 4 planos de aula. A Tabela 3, mostra o resumo de cada plano de aula elaborado.

Tabela 3. Resumo dos Planos de Aula

Aula	Descrição	Conteúdo
A1	Introdução a Regressão Linear	Introdução a Regressão Linear Pares ordenados Gráficos de Dsipersão Analise e discussão de Gráficos
A2	Determinação da reta	Preenchimento de tabelas Determinação do Parâmetro a Determinação do Parâmetro b Determinação do Equação da Reta.
A3	Resíduos e MMQ	Compreender o que são resíduos. Entender graficamente o MMQ
A4	Coeficientes de Determinação e Correlação	Determinar r^2 e r.

Fonte: Autoria Própria

4.1.1 Plano de Aula A1 - Introdução a Regressão Linear.

Tabela 4. Resumo Plano de Aula 01

Descrição	Proposta
Conteúdo	Pares ordenados e gráfico de dispersão.
Objetivo especifico	Construir por meio de pares ordenados que serão coletados (MARIANO; LAURICELLA; FRUGOLI, 2011), um gráfico de dispersão e analisar com os alunos se há uma tendência e se existe relação entre as variáveis que compõem o par ordenado.
Tempo	2 horas/aula.
Metodologia	Aula expositiva. Utilização do gráfico de modelo "Temperatura" versus "Nº de Atendimentos" que se encontra no apêndice C.
Atividades em Sala de Aula	Plotar os pares ordenados no plano cartesiano com o modelo que se encontra no apêndice C.
Justificativa	Competências e Habilidades a serem desenvolvidas segundo (BRASIL, 1998): Ler, interpretar e utilizar representações matemáticas. (Tabelas, Gráficos e Expressões). Desenvolver a capacidade de utilizar a Matemática na interpretação e intervenção no real. Formular hipóteses e prever resultados.
Avaliação	Analise de gráficos em grupos. Debates. Tarefas fora da sala de aula.
Sugestões de Abordagem	Questionamentos.

Fonte: Autoria Própria.

A Tabela 4 apresenta um resumo sobre o *Plano de Aula A1*. Sugerimos ao professor que inicie a aula explicando que um hospital pediátrico é especializado em atendimento à crianças e que podemos considerar as doenças respiratórias como todas as enfermidades que afetam o sistema respiratório humano (nariz, boca, garganta, faringe, laringe, traqueia e pulmões). Assim ele pode descrever a seguinte situação:

"Em um hospital pediátrico, no mês de julho, foram registrados a temperatura média do dia e o número de atendimentos com problemas respiratórios"(MARIANO; LAURICELLA; FRUGOLI, 2011).

O professor então deve apresentar aos alunos os registros das médias da temperatura diária em graus célsius que gerou o seguinte conjunto de dados: $X = \{9, 11, 14, 15, 17, 18, 20, 21, 22\}$, e respectivamente temos o número de atendimentos de casos com problemas respiratórios que gerou o conjunto de dados :$Y = \{28, 26, 22, 22, 22, 16, 12, 6, 6\}$, e com estes dados, faça os seguintes questionamentos:

1. Existe uma relação entre a temperatura média do dia é o número de atendimentos com problemas respiratórios ?
2. Se existe uma relação, há uma tendência quando relacionamos X com Y?, ou seja, se X e Y crescem no mesmo sentido dizemos que existe uma correlação positiva e quando crescem em sentidos opostos dizemos que há uma correlação negativa.
3. Qual a variável dependente e a independente.
4. Existe uma reta que melhor represente estes pontos ?
5. É possível determinar o quanto esta reta representa a tendência dos pontos no gráfico de dispersão?
6. É possível quantificar a força ou o grau de associação linear da relação entre X com Y?

Estes questionamentos são todos respondidos pelo professor no *Plano de Aula A1*, porém o item 4 será determinado no *Plano de Aula A2*, o item 5 e 6 serão determinados no *Plano de Aula 4*, assim Plano de Aula A1, estes questionamentos são também realizados para que o professor possa introduzir os seguintes conceitos:

1. Analise de Regressão;
2. Direção do relacionamento (positivo ou negativo);
3. Variável dependente e independente;
4. Método dos mínimos quadrados;
5. Coeficiente de Determinação;
6. Coeficiente de Correlação;

Para verificar a correlação e o direcionamento da relação entre as variáveis, o professor deve propor/sugerir a confecção de um gráfico de dispersão. Para isso deve-se definir qual a variável dependente que será representada no eixo OX e qual a variável independente que se referirá no eixo OY.

No nosso caso, consideraremos o conjunto com registro das temperaturas médias diárias sendo a variável independente, denominada a partir de agora como "X"e o número de casos com problemas respiratórios sendo a variável dependente, denominada a partir de agora como "Y".

Esta escolha foi feita pelo fato que a variação da temperatura média diária provoca alterações na quantidade de atendimentos. Contudo, o número de atendimento não irá causar uma variação na temperatura diária média. Assim teremos $X = \{x_i, .., x_n\}$ e $Y = \{y_i, .., y_n\}$ e definimos cada par ordenado (x_i, y_i) sendo $x_i \in X$ e $y_i \in Y$ desse relação, como segue:

			X	Y
1.	$(x_1, y_1) = (09, 28)$.		09	28
2.	$(x_2, y_2) = (11, 26)$.		11	26
3.	$(x_3, y_3) = (14, 22)$.		14	22
4.	$(x_4, y_4) = (15, 22)$.		15	22
5.	$(x_5, y_5) = (17, 22)$.	$\Longleftrightarrow$	17	22
6.	$(x_6, y_6) = (18, 16)$.		18	16
7.	$(x_7, y_7) = (20, 12)$.		20	12
8.	$(x_8, y_8) = (21, 06)$.		21	06
9.	$(x_9, y_9) = (22, 06)$.		22	06

O professor pode agora pedir aos alunos para localizarem estes pares ordenados no plano cartesiano. Este exercício pode ser feito em grupo, caso o professor preferir. No Apêndice C temos uma folha modelo com os eixos OX representando as *Temperaturas Médias Diárias* e o eixo OY sendo representado pelo *Número de Casos com Problemas Respiratórios*.

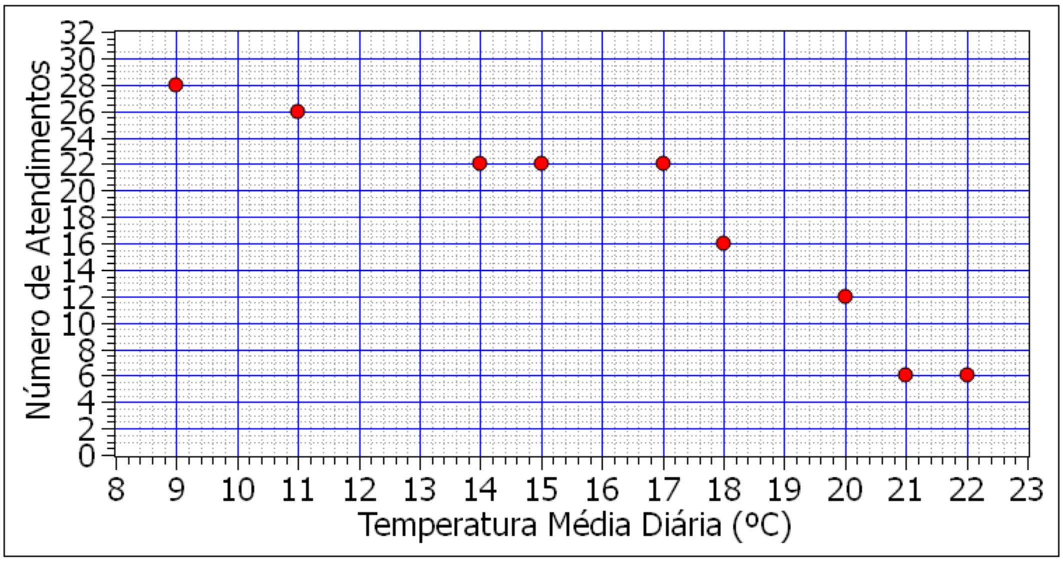
A Figura 13 mostra como deverá ficar estes pontos no gráfico de dispersão com o eixo OX com valores de temperaturas médias diária e o eixo OY com o número de casos com problemas respiratórios.

Analisando o gráfico de dispersão da Figura 13, apresenta uma tendência de que quanto maiores os valores das temperaturas no eixo horizontal OX, menores são os números de atendimentos indicados no eixo vertical OY.

O professor deverá voltar as questões 1, 2, 3 e 4 feitas nos questionamentos da aula, separar os alunos em grupos, para que entre os elementos de cada grupo discutam e respondam os questionamentos.

1. Pergunta: Existe uma relação entre a temperatura média do dia é o número de atendimentos com problemas respiratórios ?

Figura 13. Gráfico de dispersão - Temperatura média diária versus Nº de atendimentos com casos de problemas respiratórios



Fonte: Autoria Própria.

Resposta: Percebemos uma tendência de que quando há uma variação da temperatura o número de atendimentos com problemas respiratórios altera, mas somente após a determinação da regressão linear é que poderemos.

2. Se existe uma relação, há uma tendência quando relacionamos X com Y ?, ou seja, se X e Y crescem no mesmo sentido dizemos que existe uma correlação positiva e quando crescem em sentidos opostos dizemos que há uma correlação negativa.

   Resposta: Percebemos uma tendência de que quanto maior a temperatura média diária, menor é o número de casos com problemas respiratórios. Assim X e Y crescem em sentidos opostos e podemos dizer que há uma **correlação negativa** entre o valor médio diário da temperatura e a quantidade de atendimentos com problemas respiratórios. Somente após a determinação da reta de regressão e analisando o sinal do coeficiente a da equação é que poderemos afirma com certeza.

3. Pergunta:Qual a variável dependente e a independente.?
   Resposta: Já definimos que a Temperatura Média Diária é a variável independente, pois ela causa variações nos Nº Casos com Problemas Respiratórios. A variável dependente é o conjunto com o número de casos com problemas respiratórios.

4. Pergunta: Pode existir uma reta que melhor represente estes pontos ?

Resposta: Sim. Esta reta é chamada de regressão linear. É uma equação em que podemos estimar os valores de Y, em função de X, ou seja, prever aproximadamente o número de casos com problemas respiratórios, em função da temperatura média diária. Veremos nas aulas A2 e A3 os passos para a determinação da equação da reta, que melhor define os pontos dos pares ordenados.

5. É possível determinar o quanto esta reta representa a tendência dos pontos no gráfico de dispersão?
Resposta: Sim. Dá-se o nome de **Coeficiente de Determinação** e após a determinação da reta no Plano de Aula 2, determinaremos o Coeficiente de Determinação no Plano de Aula 4.

6. Pergunta: É possível determinar a força ou o grau de associação linear da relação entre X com Y?
Resposta: Sim. Dá-se o nome de **Coeficiente de Correlação** e será determinado no Plano de Aula 4.

Exercícios: Após a execução do Plano de Aula, os alunos deverão resolver as questões 1, 2 e 3 do Apêndice E, como trabalho extraclasse.

4.1.2 Plano de Aula A2 - Determinação da Reta.

No plano de Aula-A2 será determinado a Equação de Regressão Linear entre as temperaturas médias diárias e o número de atendimentos com problemas respiratórios. A Tabela (5) mostra um resumo desse Plano de Aula.

Tabela 5. Resumo Plano de Aula A2

Descrição	Proposta
Conteúdo	Equação do 1º grau; Valor médio de um conjunto; Variância, covariância, preenchimento de tabelas.
Objetivo especifico	Determinar a equação da reta de regressão linear entre a temperatura média diária e o número de atendimentos com problemas respiratórios;
Tempo	2 horas/aula.
Metodologia	Aula expositiva. Utilização de calculadoras simples.
Atividades em Sala de Aula	Construção e preenchimento de tabela para determinação de medidas de tendências centrais e de dispersão dos conjuntos das temperaturas médias diárias e número de atendimento com problemas respiratórios;

	Determinação da Covariância entre os conjuntos X e Y que representam as temperaturas médias diárias e o número de atendimentos com problemas respiratórios; Determinação da Variância do conjunto X que representa a temperaturas médias diárias; Determinação do coeficiente angular a da equação da reta de regressão; Determinação do coeficiente linear b da equação da reta de regressão; Determinar a Equação de Regressão Linear entre a temperaturas médias diárias e o número de atendimentos com problemas respiratórios.
Justificativa	Competências e Habilidades a serem desenvolvidas segundo (BRASIL, 1998): Ler, interpretar e utilizar representações matemáticas.(Tabelas, Gráficos e Expressões). Desenvolver a capacidade de utilizar a Matemática na interpretação e intervenção no real.
Avaliação	Preenchimento de Tabelas.
Sugestões de Abordagem	Exposição.

Fonte: Autoria Própria.

Para a determinação da Equação da Reta de Regressão, devemos encontrar o valor do coeficiente angular e linear a e b. Através do *Método dos Mínimos Quadrados*, podemos determinar o parâmetro a, dividindo a covariância s_{xy} dos conjuntos X e Y pela variância do conjunto X, assim temos:

$$a = \frac{s_{xy}}{s_x^2}$$

Para facilitar a determinação da covariância s_{xy} e variância s_x^2, o professor deve propor aos alunos a construção e preenchimento de uma tabela com as seguintes colunas:

i :	Número de registros
X :	Temperatura média diária
Y :	Nº de casos com problemas respiratórios
$(x_i - \bar{x})$:	Desvios de x
$(y_i - \bar{y})$:	Desvios de y
$(x_i - \bar{x})^2$:	Desvios de x ao quadrado
$(x_i - \bar{x})(y_i - \bar{y})$:	Multiplicação entre os desvios de x e y

Os alunos devem preencher primeiramente as colunas n^o , a coluna X com os valores das temperaturas média diária e a coluna Y com o número de atendimento com caso de problemas respiratórios, e definiremos os totais e as médias dos valores das colunas X e Y como mostra a Tabela 6:

Tabela 6. Valores das colunas X e Y

i	X	Y	$(x_i - \bar{x})$	$(y_i - \bar{y})$	$(x_i - \bar{x})^2$	$(x_i - \bar{x})(y_i - \bar{y})$
1	9	28				
2	11	26				
3	14	22				
4	15	22				
5	17	22				
6	18	16				
7	20	12				
8	21	6				
9	22	6				
$\sum$	147	160				
Média	16,33	17,78				

Fonte:Autoria Própria.

O Total da coluna X da Tabela 6 que representa os registros da temperaturas média diária determinado com a soma de todos os elementos da coluna

X como segue:

$$\begin{aligned}\sum X &= \sum_{i=1}^{9} x_i \\ &= x_1 + x_2 + x_3 + x_4 + x_5 + x_6 + x_7 + x_8 + x_9 \\ &= 9 + 11 + 14 + 15 + 17 + 18 + 20 + 21 + 22 \\ &= 147\end{aligned}$$

Para a determinação da média $\bar{x}$ da coluna X da Tabela 6, utilizamos o total da coluna X dividido pela quantidade de elementos da coluna. Assim temos:

$$\bar{x} = \frac{\sum X}{n} = \frac{147}{9} = 16,33$$

O Total da coluna Y da Tabela 6 que representa os registros de atendimentos dos casos com problemas respiratórios é determinado com a soma de todos os elementos da coluna Y como segue:

$$\begin{aligned}\sum Y &= \sum_{i=1}^{9} y_i \\ &= y_1 + y_2 + y_3 + y_4 + y_5 + y_6 + y_7 + y_8 + y_9 \\ &= 28 + 26 + 22 + 22 + 22 + 16 + 12 + 6 + 6 \\ &= 160\end{aligned}$$

Para a determinação da média $\bar{y}$ da coluna Y da Tabela 6, utilizamos o total da coluna Y dividido pela quantidade de elementos da coluna. Assim temos:

$$\bar{y} = \frac{\sum Y}{n} = \frac{160}{9} = 17,78$$

Determinado as média $\bar{x}$ e $\bar{y}$ das colunas X e Y respectivamente, da Tabela 6 essas médias serão utilizadas para a determinação dos valores das próximas colunas. Os alunos poderão utilizar uma calculadora para a determinação dos valores das colunas. Segue as colunas da Tabela 6 que devem ser preenchidas, onde os cálculos estão discriminados no Apêndice A:

Coluna $(x_i - \bar{x})$	Desvios de cada x_i em relação a média $\bar{x}$;
Coluna $(y_i - \bar{y})$	Desvios de cada y_i em relação a média $\bar{y}$;
Coluna $(x_i - \bar{x})^2$	Desvios de cada x_i em relação a média $\bar{x}$ elevado ao quadrado;
Coluna $(x_i - \bar{x})(y_i - \bar{y})$	Produto dos desvios de cada x_i em relação a média $\bar{x}$ com cada y_i em relação a média $\bar{y}$;

Precisaremos dos totais das colunas $(x_i - \bar{x})^2$ e $(x_i - \bar{x})(y_i - \bar{y})$ da Tabela 6, onde os cálculos estão discriminados no Apêndice A, segue os resultados:

$$\sum_{i=1}^{9}(x_i - \bar{x})^2 = 160,01$$

$$\sum_{i=1}^{9}(x_i - \bar{x})(y_i - \bar{y}) = -277,31$$

Após os cálculos a Tabela 6, deverá ficar preenchida como a Tabela 7.

Tabela 7. Dados para determinação da equação da reta de regressão

i	X	Y	$(x_i - \bar{x})$	$(y_i - \bar{y})$	$(x_i - \bar{x})^2$	$(x_i - \bar{x})(y_i - \bar{y})$
1	9	28	-7,33	10,22	53,73	-74,91
2	11	26	-5,33	8,22	28,41	-43,81
3	14	22	-2,33	4,22	5,43	-9,83
4	15	22	-1,33	4,22	1,77	-5,61
5	17	22	0,67	4,22	0,45	2,83
6	18	16	1,67	-1,78	2,79	-2,97
7	20	12	3,67	-5,78	13,47	-21,21
8	21	6	4,67	-11,78	21,81	-55,01
9	22	6	5,67	-11,78	32,15	-66,79
Total	147	160	0	0	160,01	-277,33
Média	16,33	17,78				

Com os dados da Tabela 7 determinaremos covariância S_{xy} e a variância, S_x^2 e por fim o parâmetro a.

Na covariância amostral utilizaremos o total da coluna do produto dos desvios de x_i e y_i, denominada $(x_i - \bar{x})(y_i - \bar{y})$, da Tabela 7, dividido pelo número de registros menos um, pois estamos trabalhando com amostras, como segue:

$$s_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{n-1} = \frac{-277,33}{9-1} = \frac{-277,33}{8} = -34,66$$

Para a determinação da variância amostral S_x^2, utilizaremos o total da coluna $(x_i - \bar{x})^2$ da Tabela 7, ou seja, o total dos desvios de x_i elevado ao quadrado dividido pelo numero de registros menos um, pois estamos trabalhando com amostras como segue:

$$s_x^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1} = \frac{160,01}{9-1} = \frac{160,01}{8} = 20$$

Para a determinação do parâmetro a da reta de regressão, devemos dividir a covariância S_{xy} pela variância S_x^2, assim temos

$$a = \frac{s_{xy}}{s_x^2} = \frac{-34,66}{20} = -1,73$$

Para a determinação do parâmetro b utilizaremos as médias de $\bar{x}$, $\bar{y}$ das colunas X e Y da Tabela 7 e o parâmetro a. Asim teremos:

$$b = \bar{y} - a\bar{x} \Rightarrow a = 17,78 - (-1,73 \cdot 16,33) = 46,03$$

Assim a equação da reta é data por:

$$\hat{y} = -1,73x + 46.03$$

Concluímos este Plano de Aula A2, respondendo a questão nº 4: "Existe uma reta que melhor represente estes pontos ?", apresentada no Plano de Aula A1. A reta que melhor representa este pontos é:

$$\hat{y} = -1,73x + 46.03$$

Exercícios: Após a execução do Plano de Aula, os alunos deverão resolver as questões 4, 5 e 6 do Apêndice E, como trabalho extraclasse.

4.1.3 Plano de Aula A3 - Resíduos e o Método dos Mínimos Quadrados

A Tabela 8, mostra um resumo do Plano de Aula A3. Este plano tem a função de expor aos alunos o que são resíduos, para compreenderem graficamente o que é o Método dos Mínimos Quadrado.

Tabela 8. Resumo Plano de Aula A3

Descrição	Proposta
Conteúdo:	Analise de Gráficos.
Objetivo especifico	Traçar a reta de regressão; Compreender o que é Resíduo; Compreender o que representa o Método dos Mínimos Quadrados.
Tempo	1 hora/aula.
Metodologia	Aula expositiva; Utilização de calculadoras simples; Uso da folha Modelo Temperatura versus Nº de Atendimentos do Apêndice C , onde foram localizados os pontos dos pares ordenados da aula A1, ou usar o modelo com os pontos já localizados do Apêndice D.

Atividades em Sala de Aula	Determinar dois pontos da equação da reta de regressão linear; Traçar a reta passando por esses dois pontos; Determinar o resíduo desses dois pontos.
Justificativa	Competências e Habilidades a serem desenvolvidas segundo (BRASIL, 1998): Ler, interpretar e utilizar representações matemáticas (Tabelas, Gráficos e Expressões); Desenvolver a capacidade de utilizar a Matemática na interpretação e intervenção no real.
Avaliação	Acompanhamento em sala de aula.
Sugestão de Abordagem	Expositiva.

Fonte: Autoria Própria.

O aluno deve agora traçar a reta de regressão linear, determinada no Plano de Aula A2. Para isso precisaremos de dois pontos que estejam contidos na reta, assim escolheremos então os pontos $x_5 = 17$ e $x_8 = 21$, para determinar o $\hat{y}_5$ e $\hat{y}_8$, aplicando a equação da reta de regressão determinada por $\hat{y}_i = -1,73x_i + 46,03$. Assim teremos:

$$\hat{y}_5 = -1,73x_5 + 46,03 \Rightarrow \hat{y}_5 = -1,73(17) + 46,03 \Rightarrow \hat{y}_5 = 16,62$$

$$\hat{y}_8 = -1,73x_8 + 46,03 \Rightarrow \hat{y}_8 = -1,73(21) + 46,03 \Rightarrow \hat{y}_8 = 9,7$$

O gráfico de dispersão da Figura 14, mostra como deve ficar a reta de regressão linear, determinada no Plano de Aula A2.

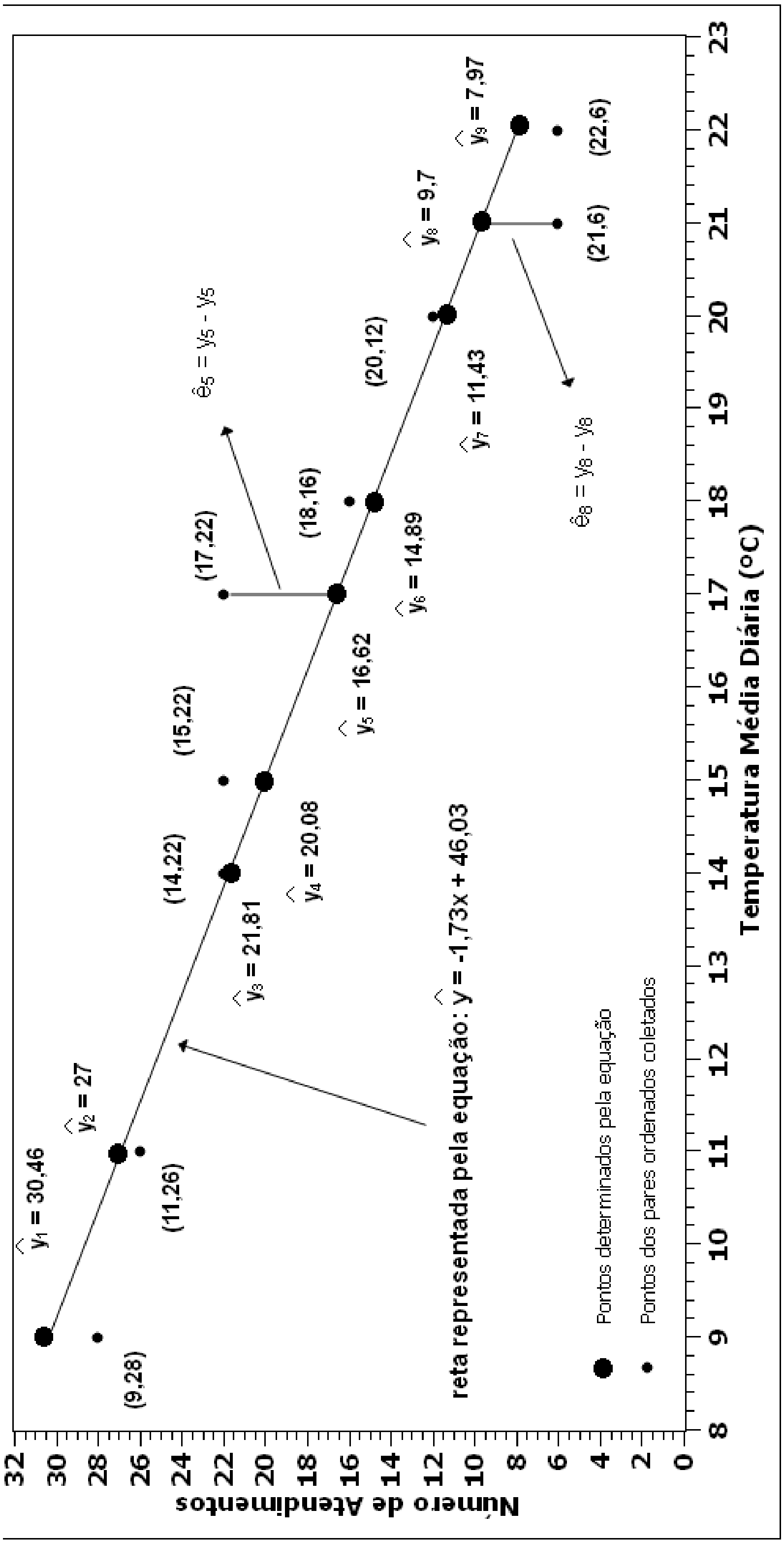
Pode-se verificar na Figura 14, que os pontos coletados não coincidem com a reta, esta diferença é chamada de resíduo $\hat{e}_i$, sendo determinada por $\hat{e}_i = y_i - \hat{y}_i$. Os resíduos, relativos aos pontos (x_5, y_5) e (x_8, y_8) são:

$$\hat{e}_5 = y_5 - \hat{y}_5 \Rightarrow \hat{e}_5 = 22 - 16,62 \Rightarrow \hat{e}_5 = 5,38$$

$$\hat{e}_8 = y_8 - \hat{y}_8 \Rightarrow \hat{e}_8 = 6 - 9,7 \Rightarrow \hat{e}_5 = -3,7$$

Os resíduos $\hat{e}_5$ e $\hat{e}_8$, estão indicados no gráficos de dispersão da Figura 14.

Figura 14. Gráfico Reta de Regressão - Temperatura média diária versus Nº de atendimentos com casos de problemas respiratórios



Fonte: Autoria Própria.

O professor deve explicar aos alunos, que a soma de todos os desvios elevados ao quadrado $\sum_{i=1}^{n}(\hat{e}_i)^2$, tem o menor valor, quando a reta é a determinada pelo *Método dos Mínimos Quadrados* e que neste caso é a Equação $\hat{y} = -1,73x + 46,03$ é a que melhor representa as tendências dos pontos no gráfico da Figura 13. Este é o ponto central da Regressão linear determinada através do Método dos Mínimos Quadrados.

4.1.4 Plano de Aula A4 - Coeficientes de Determinação r^2 e Correlação r

Tabela 9. Resumo Plano de Aula A4

Descrição	Proposta
Conteúdo	Preenchimento de tabelas e analise de gráficos, cálculos.
Objetivo especifico	Determinar a Soma de Quadrados (SQT); Determinar a Soma de Quadrados do Erro (SQE); Determinar a Soma de Quadrados da Regressão SQR; Determinar o Coeficiente de Determinação; Determinar o Coeficiente de Correlação.
Tempo	2 horas/aula.
Metodologia	Aula expositiva; Construção de Tabelas; Utilização de calculadoras simples.
Atividades em Sala de Aula	Preenchimento de tabelas; Cálculos; Determinar o Coeficiente de Determinação r^2; Determinar o Coeficiente de Correlação r.
Justificativa	Competências e Habilidades a serem desenvolvidas segundo (BRASIL, 1998): Ler, interpretar e utilizar representações matemáticas (Tabelas, Gráficos e Expressões); Desenvolver a capacidade de utilizar a Matemática na interpretação e intervenção no real.
Avaliação	Acompanhamento em sala de aula.
Sugestão de Abordagem	Expositiva.

Este Plano de Aula, irá responder a questões nº 5: "*É possível determinar o quanto esta reta representa a tendência dos pontos no gráfico de dispersão?*" e a questão nº6: "*É possível a força ou o grau de associação linear* *da relação entre X com Y?"*, questionadas no Plano de Aula 01. O resumo deste Plano de Aula 04 está na Tabela 9.

Para responder a questão nº 5 "*É possível determinar o quanto esta reta representa a tendência dos pontos no gráfico de dispersão?"*, devemos determinar o Coeficiente de Determinação denominado como r^2 que varia entre 0 e 1, este coeficiente é que ira mostrar o quanto a reta pode explicar os valores observados, sua medida é em porcentagem. A Equação para a determinação é:

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

Determinaremos então os itens abaixo para posteriormente determinar o Coeficiente de Determinação r^2:

$SQT =$ Soma de quadrado total.

$SQE =$ Soma de quadrado do erro.

$SQR =$ Soma de quadrados da regressão $(SQT - SQE)$.

$$\textit{Soma de quadrado total}: SQT = \sum_{i=1}^{n} (y_i - \bar{y})^2$$

$$\textit{Soma de quadrados do erro}: SQE = \sum_{i=1}^{n} (y_i - \hat{y})$$

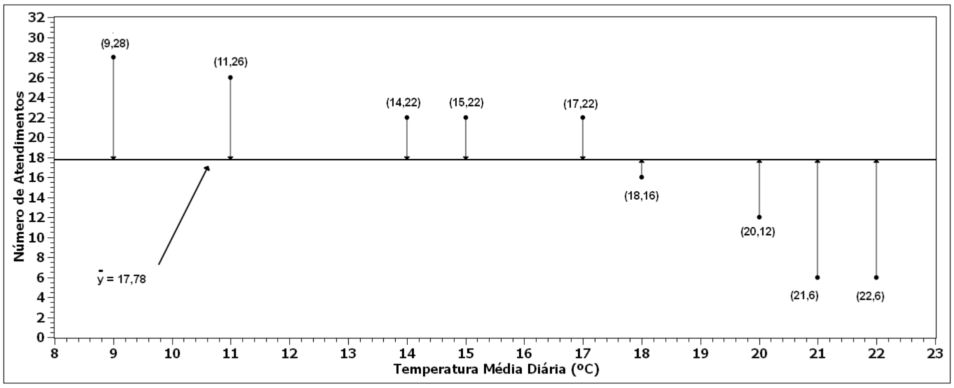
$$\textit{Soma de quadrados da regressão}: SQR = SQT - SQE$$

Assim temos que:

$$r^2 = \frac{SQR}{SQT} = \frac{\text{variação explicada}}{\text{variação total}}, \text{ sendo:}$$

Soma de quadrado total: é a soma das diferenças entre os valores coletados da variável dependente y e a sua média $\bar{y}$, elevado ao quadrado, ou seja, $(y_i - \bar{y})^2$, que em nosso caso a variável dependente é o número de casos com problemas respiratórios. A Figura 15, mostra a linha média dos valores da coluna Y da Tabela 11 de valor $\bar{y} = 17.78$ e os pontos dos pares ordenados.

Figura 15. Gráfico de Desvios entre cada y_i e a média $\bar{y}$ da coluna Y



Fonte: Autoria Própria.

Para facilitar a determinação da *Soma de Quadrado Total SQT*, utilizaremos uma tabela que contará com seis colunas sendo:

$n.^o$: número do registro;

X : temperatura média diária;

Y : número de casos com problemas respiratórios;

$\bar{y}$: valor médio da coluna Y da Tabela 11;

$(y_i - \bar{y})$: desvios entre os valores coletados de Y e o valor médio $\bar{y}$;

$(y_i - \bar{y})$: desvios entre os valores coletados de Y e o valor médio $\bar{y}$ elevados ao quadrado.

Podemos visualizar através o gráfico da Figura 15 os desvios de cada y_i

com a média $\bar{y}$ da coluna Y da Tabela 11. A Tabela 11 mostra como deverá ficar o preenchimento da tabela para a determinação dos cálculos da Soma de Quadrado Total.

Tabela 11. Determinação da Soma de Quadrado Total

i	X	Y	$\bar{y}$	$(y_i - \bar{y})$	$(y_i - \bar{y})^2$
1	9	28	17,78	10,22	104,45
2	11	26	17,78	8,22	67,57
3	14	22	17,78	4,22	17,81
4	15	22	17,78	4,22	17,81
5	17	22	17,78	4,22	17,81
6	18	16	17,78	-1,78	3,17
7	20	12	17,78	-5,78	33,41
8	21	6	17,78	-11,78	138,77
9	22	6	17,78	-11,78	138,77
$\sum$		160			539,56
Média		17,78			

Assim a determinação da Soma de Quadrados Totais SQT é o valor da soma da coluna $(y_i - \bar{y})^2$ da Tabela 11.

$$SQT = \sum_{i=1}^{9}(y_i - \bar{y})^2 = 539,56$$

A *Soma de Quadrado do Erro*, SQE, é a soma de todos os resíduos elevados ao quadrado e_i^2. Assim para a sua determinação construiremos outra tabela contendo as seguintes colunas:

$n.^o$: número do registro;

X : temperatura média diária;

Y : número de casos com problemas respiratórios;

$\hat{y}_i$: valor determinado por:$\hat{y}_i = -1,73x + 46,03$;

e_i : resíduos $y_i - \hat{y}$;

e_i^2 : resíduos elevado ao quadrado.

A Tabela 13 deverá ficar preenchida como demonstrado.

Tabela 13. Resíduos da reta de regressão

i	X	Y	$\hat{y}$	$\hat{e}$	$(\hat{e})^2$
1	9	28	30,46	-2,46	6,05
2	11	26	27,00	-1,00	1,00
3	14	22	21,81	0,19	0,04
4	15	22	20,08	1,92	3,69
5	17	22	16,62	5,38	28,94
6	18	16	14,89	1,11	1,23
7	20	12	11,43	0,57	0,32
8	21	6	9,70	-3,7	13,69
9	22	6	7,97	-1,97	3,88
$\sum$	147	160			58,85
Média	16,33	17,78			

A soma da coluna $\hat{e}^2$ da Tabela 13 é a Soma de Quadrados dos Erros SQE:

$$SQE = \sum_{i=1}^{n}(y_i - \hat{y})^2 = \sum_{i=1}^{n}\hat{e}^2 = 58,85$$

Para a determinação da Soma de Quadrados da Regressão SQR temos:

$$SQR = SQT - SQE \Rightarrow SQR = 539,56 - 58,85 = 480,71$$

Para a determinação do r^2 temos:

$$r^2 = \frac{SQR}{SQT} = \frac{\text{variação explicada}}{\text{variação total}} \Rightarrow r^2 = \frac{480,71}{539,56} = 0,89 \text{ ou } 89\%$$

A Resposta para a questão n^o 5 "*É possível determinar o quanto esta reta representa a tendência dos pontos no gráfico de dispersão?*" é:

Podemos afirmar que a reta de regressão linear determinada no Plano de Aula A2 por $\hat{y} = -1,73x + 46,03$, representa em 89% os pontos dos pares ordenados.

A determinação do **Coeficiente de Correlação** r, é dada pela divisão entre a covariância dos dois conjuntos de dados X e Y dividido pelo produto dos desvios padrão dos conjuntos X e Y, representado pela Equação a seguir:

$$r = \frac{Cov(X,Y)}{S_x S_y}$$

A covariância amostral s_{xy} e a variância s_x^2 já foi determina no Plano de Aula A2 para a determinação do parâmetro a da equação da reta de regressão, assim utilizaremos, esse dados para a determinação do coeficiente de correlação. Segue a equação de determinação da covariância s_{xy}:

$$s_{xy} = \frac{\displaystyle\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{n-1} = \frac{-277,33}{9-1} = \frac{-277,33}{8} = -34,66$$

A Equação para determinação do coeficiente de correlação r, pede os desvios padrão de s_x e s_y. Temos a variância de s_x^2 que foi utilizada no Plano de Aula A2 para determinar o parâmetro a da equação da reta de regressão. Para encontra o desvios padrão, basta extrair a raiz quadrada da variância, então segue:

$$s_x = \sqrt{s_x^2} = \sqrt{\frac{\displaystyle\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{160,01}{8}} = \sqrt{20} \Rightarrow s_x = 4,47$$

Para determinar o desvio padrão s_y, utilizaremos o total da coluna $(y_i - \bar{y})^2$ da Tabela 11, assim temos:

$$s_y = \sqrt{s_y^2} = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \bar{y})^2}{n-1}} = \sqrt{\frac{539,56}{8}} = \sqrt{67,45} \Rightarrow s_y = 8,21$$

Com a covariância s_xy e as variâncias s_x e s_y determinadas, aplicamos na equação como segue:

$$r = \frac{Cov(X,Y)}{S_x S_y} = \frac{-34,66}{4,47 \cdot 8,21} = \frac{-34,66}{36,70} \Rightarrow r = -0,94$$

O coeficiente de correlação r, varia entre $-1 < r < 1$, e quanto mais próximo a unidade, positiva ou negativa, maior é a força entre as variáveis.

A resposta para questão nº 6: *"É possível a força da relação entre X com Y?"*, questionadas no Plano de Aula 01 é:

Podemos afirmar que a força de correlação entre os conjuntos de dados X e Y é muito forte, pois o Coeficiente de Correlação $r = -0,94$ é muito próximo de -1. Assim é forte a dependência do número de atendimentos em função da temperatura média diária.

Exercícios: Após a execução do Plano de Aula, os alunos deverão resolver as questões 7 e 8 do Apêndice E, como trabalho extraclasse.

Considerações Finais

A realização do estudo sobre Regressão Linear através do Método dos Mínimos Quadrados teve por objetivo o planejamento de ensino contextualizado e interdisciplinar para alunos do Ensino Médio, procurando abranger áreas distintas do conhecimento humano.

Durante nossas pesquisas sobre contextualização, percebemos que esta é muito ampla e que pode ser observada em outras metodologias do ensino da Matemática, como a *História da Matemática*, *Modelagem Matemática*, *Etnomatemática* e *Aplicabilidade da Matemática*.

O Método dos Mínimos Quadrados utilizados para analise de Regressão Linear, utiliza-se de pares ordenados, construções e análises de gráficos de dispersão, tabelas, equações da reta e medidas de tendência central e de dispersão, que são habilidades e competências dos Parâmetros Curriculares Nacionais. Dessa forma o planejamento que no inicio tinha um objetivo de contextualizar o ensino da Matemática se tornou também a contextualização da própria Matemática, onde equações de primeiro grau e Estatística são utilizados no contexto matemático da Regressão Linear.

O planejamento da unidade de ensino, procura apresentar dados, não sobre um problema a ser resolvido, mas sim uma previsão do que há por vir, dados que serão investigados e previstos. Procuramos dessa forma instigar a curiosidade dos alunos. Divido em planos de aulas, procuramos responder as investigações, de forma que o aluno gradualmente absorva e compreenda o que é o ensino de regressão linear, exercitando os conceitos básicos da equação da reta e estatística.

De forma interdisciplinar, o plano de unidade traz dados de um hospital pediátrico, para que os alunos definam um padrão, uma lógica entre as variáveis *temperatura diária* e o *número de atendimentos com problemas respiratórios* que pode ser simplesmente com a analise do gráfico de dispersão. Posteriormente a utilização da Matemática para prever e quantificar a força da relação entre as variáveis, através da reta de regressão.

A regressão linear, é muito ampla e podemos tentar correlacionar qualquer duas variáveis, mas o professor deve ficar atendo, pois há casos onde existe uma correlação, e não se escolhe com critério as variáveis. Exemplo disso: se fizermos uma correlação entre o número de sorvetes vendidos em certa praia e o número de ocorrência de afogamentos, provavelmente haverá uma correlação positiva entre as duas variáveis, chegando a conclusão de que quanto mais vender sorvete maior é o número de afogamentos, porém as variáveis corretas neste caso seriam a temperatura média da praia e o número de afogamentos, pois em dias mais quente há um número maior de pessoas na praia consequentemente haverá um consumo maior de sorvete e uma probabilidade maior de pessoas se afogarem.

Há tantas possibilidades com a regressão linear, que em sala de aula mesmo, o professor pode através das características dos alunos, por exemplo, correlacionar tamanho dos pés com altura.

Assim procuramos com esse trabalho composto de um plano de unidade dividido em quatro plano de aulas, sobre regressão linear, contextualizando diferentes áreas do saber e utilizando-se de conceitos de estatísticas e funções de primeiro graus, proporcionar ao professor do Ensino Médio, um plano de unidade de ensino que pode ser alterada conforme a característica da região onde leciona. Assim as variáveis que ele pode relacionar podem estar ligadas diretamente com o cotidiano dos alunos.

Referências Bibliográficas

ACTION, P. *Analise de Regressão*. 2015. Acessado em 10/09/2017. Disponível em: <http://www.portalaction.com.br/analise-de-regressao>.

ADAMI, A. M.; FILHO, A. A. D.; LORANDIL, M. M. *Pré-Cálculo*. Porto Alegre: BOOKMAN EDITORA LTDA., 2015.

ALMEIDA, R. N. de. *O Método dos Mínimos Quadrados: Estudos e Aplicações para o Ensino Médio*. Campos dos GoyTacazes - RJ: Universidade Estadual do Norte Fluminense - Darcy Ribeiro - UENF, 2015.

BONAFINI, F. C. *Estatística*. São Paulo: Pearson Education do Brasil, 2011.

BRASIL. *Parâmetros Curriculares Nacionais*. [S.l.]: MEC, 1998.

CARROCINO, C. H. G. *Questões Contextualizadas nas Provas de Matemática*. Rio Janeiro RJ: IMPA-Instituto de Matemática Pura e Aplicada - RJ, 2014.

CORREA, S. M. B. B. *Probabilidade e Estatística*. [S.l.]: Belo Horizonte: Puc Minas Virtual, 2003.

DICIO. *Dicionário online de Português*. 2018. Acessado em 09/07/2018. Disponível em: <https://www.dicio.com.br/>.

DRUCK, S. *O drama do ensino da matemática*. Acessado em 31/08/2017. Disponível em: <http://www1.folha.uol.com.br/folha/sinapse/ult1063u343.shtml>.

ENEM. *Encontro Nacional de Educação Matemática*. 2018. Acessado em 09/07/2018. Disponível em: <http://www.sbembrasil.org.br/sbembrasil/index.php/anais/enem>.

FOGAçA, J. *Contextualização*. 2011. Acessado em 31/08/2017. Disponível em: <http://educador.brasilescola.uol.com.br/trabalho-docente/contextualizacao.htm>.

GURAJATI, D. N. *Econometria Básica*. [S.l.]: AMGH Editora Ltda, 2011.

HAYDT, R. C. C. *Curso de Didática Geral*. São Paulo: Editora Ática — Livros Técnicos e Científicos Editora Ltda., 2011.

HILL, R. C.; GRIFFITHS, W. E.; JUDGE, G. G. *Econometria*. São Paulo SP: Editora Saraiva, 2010.

IEZZI, G.; MURAKAMI, C. *Fundamentos de Matemática Elementar 1 - Funções e Conjuntos*. [S.l.]: Atual Editora, 1977.

JALES, P. C. *A importância do Ensino de Regressão Linear Simples no Ensino Médio: um estudo com os alunos do 3º Ano do Ensino Médio IFMA - Imperatriz*. Teresina - PI: Universidade Federal do Piauí, 2014.

LARSON, R.; FARBER, B. *Estatística Aplicada*. São Paulo: Pearson Education do Brasil, 2004.

LEVINE, D. M.; STEPHAN, D. F.; SZABAT, K. A. *Estatística: Teoria e Aplicações*. Rio de Janeiro, RJ: LTC — Livros Técnicos e Científicos Editora Ltda, 2017.

LOBATO, A. C. *Contextualização: Um conceito em debate*. 2008. Acessado em: 31/08/2017. Disponível em: <http://www.educacaopublica.rj.gov.br/biblioteca/educacao/0173.html>.

MACHADO, J. F. *Método Estatístico:Gestão da qualidade para melhoria continua*. São Paulo - SP: Editora Saraiva, 2010.

MALHEIROS, B. T. *Didática Geral*. Rio de Janeiro: LTC — Livros Técnicos e Científicos Editora Ltda., 2012.

MANN, P. S. *Introdução à Estatística*. Rio Janeiro RJ: LTC — Livros Técnicos e Científicos Editora Ltda, 2015.

MARCONI, M. de A.; LAKATOS, E. M. *Fundamentos de Metodologia Científica*. [S.l.]: Editora Atlas S.A., 2017.

MARIANO, M. V.; LAURICELLA, C. M.; FRUGOLI, A. D. *Estatísca Indutiva*. [S.l.: s.n.], 2011.

MATTOS, L. A. de. *Sumário de didática geral*. São Paulo: Editora Aurora, 1954.

MATTOS, V. L. D. de; KONRATH, A. C.; AZAMBUJA, A. M. V. de. *Introdução à Estatística - Aplicações em Ciências Exatas*. Rio Janeiro RJ: LTC — Livros Técnicos e Científicos Editora Ltda, 2017.

MENDES, A. F. *Contextualização e Interdisciplinaridade na utilização da Matemática no Estudo de Fenômenos climáticos e Meteorológicos*. Soropédica - RJ: UFRR -Universidade Federal RURAL DO RIO DE JANEIRO, 2010.

MICHAELIS. *Dicionário online Michaelis*. 2008. Acessado em: 27/03/2018. Disponível em: <http://michaelis.uol.com.br>.

NAVIDI, W. *Probabilidade e Estatística para Ciências Exatas*. [S.l.]: Porto Alegre - RS AMGH - Editora, 2012.

NéRICI, I. G. *Didática uma Introdução*. São Paulo: Editora Atlas., 1986.

RODRIGUES, M. da S. *Dicionário Brasileiro de Estatística*. Rio de Janeiro: Fundação IBGE, 1970.

SACCO, T. P. V. *Analise de duas metodologias distintas para o ensino de estatística nos anos finais do ensino fundamental: Metodologia Tradicional e Contextualizada*. Três Lagoas - MS: UFMS -Universidade Federal de Mato Grosso do Sul - MS, 2015.

SANTOS, A. O.; OLIVEIRA, G. S. de. Contextualização no ensino-aprendizagem da matemática: Princípios e práticas. Educação em Rede:Formação e Prática Docente, http://ojs.cesuca.edu.br/index.php/educacaoemrede/article/view/819, v. 4, n. 5, 2015.

SARTORIS, A. *Estatística e introdução a econometria 2 ed*. São Paulo: Editora Saraiva, 2013.

SPIEGEL, M. R. *Estatística*. [S.l.]: MgGRAW-Hill do Brasil, 1985.

TRIOLA, M. F. *Introdução à Estatística*. [S.l.]: LTC - Livros Técnicos e Científicos Editora S.A., 1998.

TURRA, E.; OUTROS. *Planejamento de ensino e avaliação*. Porto Alegre: PUC-EMMA, 1974.

VIEIRA, G. M. *Estratégias de Contextualização nos Livros Didáticos de Matemática dos ciclos iniciais do Ensino Fundamental*. Belo Horizonte - MG: UFMG -Universidade Federal de Minas Gerais, 2004.

VIEIRA, S. *Estatística Básica*. [S.l.]: Cengage Learning, 2015.

A Cálculos parâmetros a e b da equação da reta

Demonstração dos cálculos dos desvios de cada x_i em relação a média $\bar{x}$ da coluna X da Tabela 6:

$(x_1 - \bar{x}) = 09 - 16,33 = -7,33$

$(x_2 - \bar{x}) = 11 - 16,33 = -5,33$

$(x_3 - \bar{x}) = 14 - 16,33 = -2,33$

$(x_4 - \bar{x}) = 15 - 16,33 = -1,33$

$(x_5 - \bar{x}) = 17 - 16,33 = 0,67$

$(x_6 - \bar{x}) = 18 - 16,33 = 1,67$

$(x_7 - \bar{x}) = 20 - 16,33 = 3,67$

$(x_8 - \bar{x}) = 21 - 16,33 = 4,67$

$(x_9 - \bar{x}) = 22 - 16,33 = 5,67$

Demonstração dos cálculos dos desvios de cada y_i em relação a média $\bar{y}$ da coluna Y da Tabela 6:

$(y_1 - \bar{y}) = 28 - 17,78 = 10,22$

$(y_2 - \bar{y}) = 26 - 17,78 = 8,22$

$(y_3 - \bar{y}) = 22 - 17,78 = 4,22$

$(y_4 - \bar{y}) = 22 - 17,78 = 4,22$

$(y_5 - \bar{y}) = 22 - 17,78 = 4,22$

$(y_6 - \bar{y}) = 16 - 17,78 = -1,78$

$(y_7 - \bar{y}) = 12 - 17,78 = -5,78$

$(y_8 - \bar{y}) = 06 - 17,78 = -11,78$

$(y_9 - \bar{y}) = 06 - 17,78 = -11,78$

Demonstração dos cálculos do Quadrado dos desvios de cada x_i em relação a média $\bar{x}$ da coluna X da Tabela 6:

$(x_1 - \bar{x})^2 = (-7,33)^2 = 53,73$

$(x_2 - \bar{x})^2 = (-5,33)^2 = 28,41$

$(x_3 - \bar{x})^2 = (-2,33)^2 = 5,43$

$(x_4 - \bar{x})^2 = (-1,33)^2 = 1,77$

$(x_5 - \bar{x})^2 = (0,67)^2 = 0,45$

$(x_6 - \bar{x})^2 = (1,67)^2 = 2,79$

$(x_7 - \bar{x})^2 = (3,67)^2 = 13,47$

$(x_8 - \bar{x})^2 = (4,67)^2 = 21,81$

$(x_9 - \bar{x})^2 = (5,67)^2 = 32,15$

Demonstração dos cálculos do produto entre desvios de cada x_i em relação a média $\bar{x}$ da coluna X da Tabela 6 com os desvios de cada y_i em relação a média $\bar{y}$ da coluna Y da Tabela 6:

$(x_1 - \bar{x})(y_1 - \bar{y}) = (-7,33)(10,22) = -74,91$

$(x_2 - \bar{x})(y_2 - \bar{y}) = (-5,33)(8,22) = -43,81$

$(x_3 - \bar{x})(y_3 - \bar{y}) = (-2,33)(4,22) = -9,83$

$(x_4 - \bar{x})(y_4 - \bar{y}) = (-1,33)(4,22) = -5,61$

$(x_5 - \bar{x})(y_5 - \bar{y}) = (0,67)(4,22) = 2,83$

$(x_6 - \bar{x})(y_6 - \bar{y}) = (1,67)(-1,78) = -2,97$

$$(x_7 - \bar{x})(y_7 - \bar{y}) = (3,67)(-5,78) = -21,21$$

$$(x_8 - \bar{x})(y_8 - \bar{y}) = (4,67)(-11,78) = -55,01$$

$$(x_9 - \bar{x})(y_9 - \bar{y}) = (5,67)(-11,78) = -66,79$$

B Cálculos para a determinação de r^2

Demonstração dos cálculos para o cálculo de todos os $\hat{y_i}$ preditos para o preenchimento da Tabela 13.

$$\hat{y_1} = -1,73x + 46.03 \Rightarrow \hat{y_1} = -1,73(9) + 46.03 \Rightarrow \hat{y_1} = 30,46$$

$$\hat{y_2} = -1,73x + 46.03 \Rightarrow \hat{y_2} = -1,73(11) + 46.03 \Rightarrow \hat{y_2} = 27,00$$

$$\hat{y_3} = -1,73x + 46.03 \Rightarrow \hat{y_3} = -1,73(14) + 46.03 \Rightarrow \hat{y_3} = 21,81$$

$$\hat{y_4} = -1,73x + 46.03 \Rightarrow \hat{y_4} = -1,73(15) + 46.03 \Rightarrow \hat{y_4} = 20,08$$

$$\hat{y_5} = -1,73x + 46.03 \Rightarrow \hat{y_5} = -1,73(17) + 46.03 \Rightarrow \hat{y_5} = 16,62$$

$$\hat{y_6} = -1,73x + 46.03 \Rightarrow \hat{y_6} = -1,73(18) + 46.03 \Rightarrow \hat{y_6} = 14,89$$

$$\hat{y_7} = -1,73x + 46.03 \Rightarrow \hat{y_7} = -1,73(20) + 46.03 \Rightarrow \hat{y_7} = 11,43$$

$$\hat{y_8} = -1,73x + 46.03 \Rightarrow \hat{y_5} = -1,73(21) + 46.03 \Rightarrow \hat{y_8} = 9,70$$

$$\hat{y_9} = -1,73x + 46.03 \Rightarrow \hat{y_9} = -1,73(18) + 46.03 \Rightarrow \hat{y_9} = 7,97$$

Demonstração dos cálculos de todos os desvios $\hat{e}$ entre o valor y_i predito e o valor previsto determinado pela equação da reta $\hat{y_i} = -1,73x + 46.03$, para o preenchimento da Tabela 13.

$\hat{e_1} = y_1 - \hat{y_1} \Rightarrow \hat{e_1} = 28 - 30,46 \Rightarrow \hat{e_1} = -2,46$

$\hat{e_2} = y_2 - \hat{y_2} \Rightarrow \hat{e_2} = 26 - 27 \Rightarrow \hat{e_2} = -1,00$

$\hat{e_3} = y_3 - \hat{y_3} \Rightarrow \hat{e_3} = 22 - 21,81 \Rightarrow \hat{e_3} = 0,19$

$\hat{e_4} = y_4 - \hat{y_4} \Rightarrow \hat{e_4} = 22 - 20,08 \Rightarrow \hat{e_4} = 1,92$

$\hat{e_5} = y_5 - \hat{y_5} \Rightarrow \hat{e_5} = 22 - 16,62 \Rightarrow \hat{e_5} = 5,38$

$\hat{e_6} = y_6 - \hat{y_6} \Rightarrow \hat{e_6} = 16 - 14,89 \Rightarrow \hat{e_6} = 1,11$

$\hat{e_7} = y_7 - \hat{y_7} \Rightarrow \hat{e_7} = 12 - 11,43 \Rightarrow \hat{e_7} = 0,57$

$\hat{e_8} = y_8 - \hat{y_8} \Rightarrow \hat{e_8} = 6 - 9,7 \Rightarrow \hat{e_8} = -3,7$

$\hat{e_9} = y_9 - \hat{y_8} \Rightarrow \hat{e_9} = 6 - 7,97 \Rightarrow \hat{e_9} = -1,97$

Demonstração da soma da coluna $(\hat{e})^2$ da Tabela 13.

$$\sum_{i=1}^{9} (\hat{e_i})^2 = (\hat{e_1})^2 + (\hat{e_2})^2 + (\hat{e_3})^2 + (\hat{e_4})^2 + (\hat{e_5})^2 + (\hat{e_6})^2 + (\hat{e_7})^2 + (\hat{e_8})^2 + (\hat{e_9})^2$$

$$\sum_{i=1}^{9} (\hat{e_i})^2 = (-2,46)^2 + (-1)^2 + (0,19)^2 + (1,92)^2 + (5,38)^2 + (1,11)^2 + (0,57)^2 + (-3.7)^2 + (-1,97)^2$$

$$\sum_{i=1}^{9}(\hat{e}_i)^2 = 6,05+1,00+0,04+3,69+28,94+1,23+0,32+13,69+3,88$$

$$\sum_{i=1}^{9}(\hat{e}_i)^2 = 58,85$$

Demonstração dos cálculos dos desvios entre o valor predito y_i e a média $\bar{y}$ da coluna Y da Tabela 11.

$$(y_1 - \bar{y}) = 28 - 17,78 = 10,22$$

$$(y_2 - \bar{y}) = 26 - 17,78 = 8,22$$

$$(y_3 - \bar{y}) = 22 - 17,78 = 4,22$$

$$(y_4 - \bar{y}) = 22 - 17,78 = 4,22$$

$$(y_5 - \bar{y}) = 22 - 17,78 = 4,22$$

$$(y_6 - \bar{y}) = 16 - 17,78 = -1,78$$

$$(y_7 - \bar{y}) = 12 - 17,78 = -5,78$$

$$(y_8 - \bar{y}) = 6 - 17,78 = -11,78$$

$$(y_9 - \bar{y}) = 6 - 17,78 = -11,78$$

Demonstração dos cálculos dos desvios entre o valor predito y_i e a média $\bar{y}$ da coluna Y da Tabela 11 elevado ao quadrado.

$$(y_1 - \bar{y})^2 = (10,22)^2 = 104,45$$

$$(y_2 - \bar{y})^2 = (8,22)^2 = 67,57$$

$$(y_3 - \bar{y})^2 = (4,22)^2 = 17,81$$

$$(y_4 - \bar{y})^2 = (4,22)^2 = 17,81$$

$$(y_5 - \bar{y})^2 = (4,22)^2 = 17,81$$

$$(y_6 - \bar{y})^2 = (-1,78)^2 = 3,17$$

$$(y_7 - \bar{y})^2 = (-5,78)^2 = 33,41$$

$$(y_8 - \bar{y})^2 = (-11,78)^2 = 138,77$$

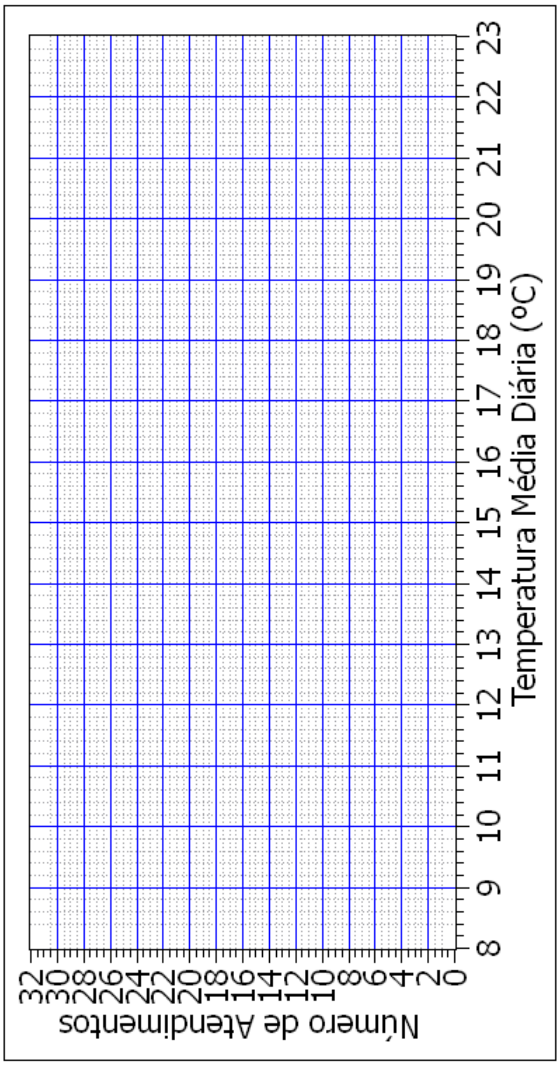
$$(y_9 - \bar{y})^2 = (-11,78)^2 = 138,77$$

SQT =104,45 + 67,57 + 17,81 + 17,81 +17,81 +3,17 + 33,41 + 138,77 + 138,77 = 539,56.

C | Modelo Gráfico Dispersão

Folha Modelo Temperatura versus Nº de Atendimentos

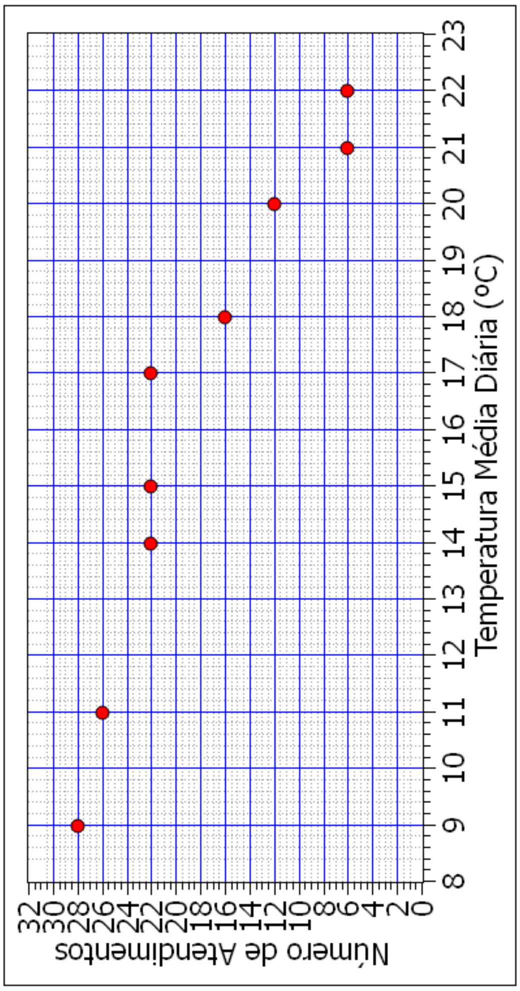
Figura 16. Gráfico de dispersão - Temperatura média diária (Eixo OX) versus Nº de atendimentos com casos de problemas respiratórios (Eixo OY)



D | Gráfico de Dispersão

Gráfico de Dispersão - Temperatura versus Nº de Atendimentos

Figura 17. Gráfico de dispersão - Temperatura média diária versus Nº de atendimentos com casos de problemas respiratórios



E | Exercício

Exercício do Plano de Aula

Em determinada empresa industrial a relação entre horas trabalhadas e a produção de um produto em toneladas é mostrada na Tabela 14.

Tabela 14. Horas Trabalhadas versus Produção

Horas Trabalhadas	Produção
03	24
05	32
10	42
12	48
10	46
02	15
06	35
08	38

Questão 1

Definindo-se as horas trabalhadas como a variável independente e denominada como o conjunto $X = \{3, 5, 10, 12, 10, 2, 6, 8\}$ e a produção como a variável dependente denominada como o conjunto $Y = \{24, 32, 42, 48, 46, 15, 35, 38\}$, determine os pares ordenados desta relação.

Resposta:

1. $(x_1, y_1) = (03, 24)$.
2. $(x_2, y_2) = (05, 32)$.
3. $(x_3, y_3) = (10, 42)$.

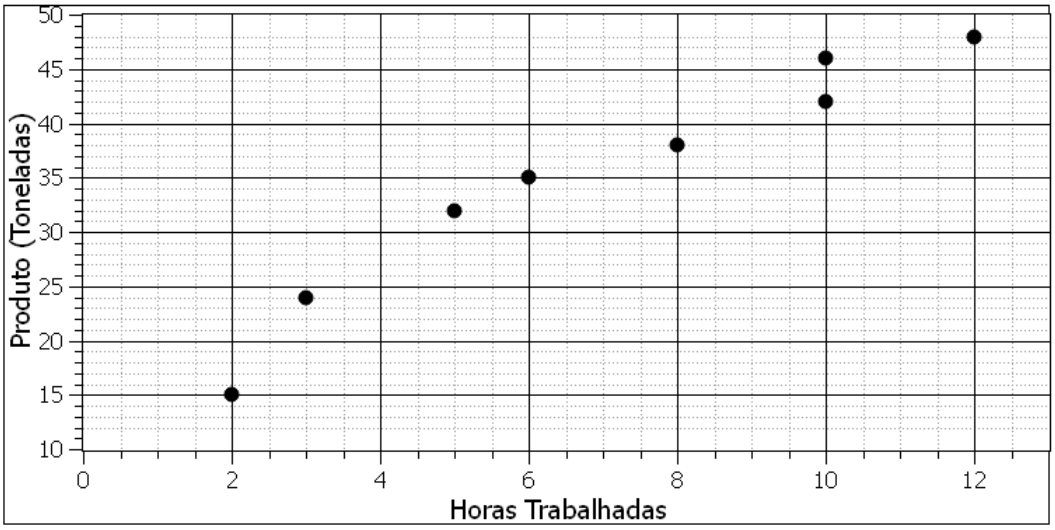
4. $(x_4, y_4) = (12, 48)$.
5. $(x_5, y_5) = (10, 46)$.
6. $(x_6, y_6) = (02, 15)$.
7. $(x_7, y_7) = (06, 35)$.
8. $(x_8, y_8) = (08, 38)$.

Questão 2

Construa um gráfico de dispersão e localize os pares ordenados?

Resposta: Gráfico da Figura 18.

Figura 18. Gráfico de dispersão - Horas Trabalhadas versus Produção



Fonte: Autoria Própria.

Questão 3

Depois de construído o gráfico, o que você diria sobre a tendência da disposição dos pontos?

Respostas: Os pontos tem uma tendência positiva, ou seja, quando aumentamos as horas trabalhadas, a produção também aumenta.

Questão 4

Determine a equação da reta de regressão desta industria, entre as horas trabalhadas e a produção do produto em toneladas.

Resposta: A equação é: $\hat{y} = 3,02x + 13,86$

Segue os cálculos para a determinação dos parâmetros a e b da equação da reta de regressão linear. Começaremos construindo uma tabela com as colunas descritas abaixo:

i :	Número de registros
X :	Nº de horas trabalhadas
Y :	Produção em Toneladas
$(x_i - \bar{x})$:	Desvios de x
$(y_i - \bar{y})$:	Desvios de y
$(x_i - \bar{x})^2$:	Desvios de x ao quadrado
$(x_i - \bar{x})(y_i - \bar{y})$:	Multiplicação entre os desvios de x e y

A Tabela 15 mostra como deve ficar preenchida. Para a determinação de todas as colunas devemos primeiramente determinar o total da coluna X da Tabela 15, que representa as horas trabalhadas e sua média $\bar{x}$. Determinar o total da coluna Y da Tabela 15, que representa a produção do produto em toneladas e sua média $\bar{y}$ como segue:

Total da coluna X:

$$\begin{aligned}
\sum X &= \sum_{i=1}^{8} x_i \\
\sum_{i=1}^{8} x_i &= 3 + 5 + 10 + 12 + 10 + 2 + 6 + 8 = 56
\end{aligned}$$

Média da coluna X denominada por $\bar{x}$:

$$\bar{x} = \frac{\sum X}{n} = \frac{56}{8} = 7$$

Total da coluna Y:

$$\sum Y = \sum_{i=1}^{8} y_i$$
$$\sum_{i=1}^{8} y_i = 24 + 32 + 42 + 48 + 46 + 15 + 35 + 38 = 280$$

Média da coluna Y denominada por $\bar{y}$:

$$\bar{y} = \frac{\sum Y}{n} = \frac{280}{8} = 35$$

Tabela 15. Dados para a determinação da covariância s_{xy} e a variância s_x^2

i	X	Y	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x})^2$	$(x_i - \bar{x})(y_i - \bar{y})$
1	3	24	-4	-11	16	44
2	5	32	-2	-3	4	6
3	10	42	3	7	9	21
4	12	48	5	13	25	65
5	10	46	3	11	9	33
6	2	15	-5	-20	25	100
7	6	35	-1	0	1	0
8	8	38	1	3	1	3
$\sum$	56	280	0	0	90	272
Média	7	35				

Com os valores de $\bar{x}$ e $\bar{y}$ determinados, preencheremos as colunas $x_i - \bar{x}$ e $y_i - \bar{y}$ da Tabela 15, que representam respectivamente os desvios de cada x_i em relação a média $\bar{x}$ e desvios de cada y_i em relação a média $\bar{y}$, assim temos:

Coluna $x_i - \bar{x}$ da Tabela 15:

$(x_1 - \bar{x}) =$ 03 - 7 = -04

$(x_2 - \bar{x}) =$ 05 - 7 = -02

$(x_3 - \bar{x}) =$ 10 - 7 = 03

$(x_4 - \bar{x}) =$ 12 - 7 = 05

$(x_5 - \bar{x}) =$ 10 - 7 = 03

$(x_6 - \bar{x}) =$ 02 - 7 = -05

$(x_7 - \bar{x}) =$ 06 - 7 = -01

$(x_8 - \bar{x}) =$ 08 - 7 = 01

Coluna $y_i - \bar{y}$ da Tabela 15:

$(y_1 - \bar{y}) =$ 24 - 35 = -11

$(y_2 - \bar{y}) =$ 32 - 35 = -03

$(y_3 - \bar{y}) =$ 42 - 35= 07

$(y_4 - \bar{y}) =$ 48 - 35 = 13

$(y_5 - \bar{y}) =$ 46 - 35 = 11

$(y_6 - \bar{y}) =$ 15 - 35 = -20

$(y_7 - \bar{y}) =$ 35 - 35 = 0

$(y_8 - \bar{y}) =$ 38 - 35 = 3

Com os valores da coluna $x_i - \bar{x}$ da Tabela 15, determinaremos os valores da coluna $(x_i - \bar{x})^2$, como segue:

$(x_1 - \bar{x})^2 = \quad (-04)^2 \quad = 16$

$(x_2 - \bar{x})^2 = \quad (-02)^2 \quad = 04$

$(x_3 - \bar{x})^2 = \quad (03)^2 \quad = 09$

$(x_4 - \bar{x})^2 = \quad (05)^2 \quad = 25$

$$
\begin{aligned}
(x_5 - \bar{x})^2 &= \quad (03)^2 &&= 03\\
(x_6 - \bar{x})^2 &= \quad (-05)^2 &&= 25\\
(x_7 - \bar{x})^2 &= \quad (-01)^2 &&= 01\\
(x_8 - \bar{x})^2 &= \quad (01)^2 &&= 01
\end{aligned}
$$

Preenchermos a ultima coluna $(x_i - \bar{x})(y_i - \bar{y})$ Tabela 15, que é o produto de $x_i - \bar{x}$ e $y_i - \bar{y}$ da mesma tabela como segue:

$$
\begin{aligned}
(x_1 - \bar{x})(y_1 - \bar{y}) &= \text{(-04)(-11)} &&= 44\\
(x_2 - \bar{x})(y_2 - \bar{y}) &= \text{(-02)(-03)} &&= 06\\
(x_3 - \bar{x})(y_3 - \bar{y}) &= (03)(07) &&= 21\\
(x_4 - \bar{x})(y_4 - \bar{y}) &= (05)(13) &&= 65\\
(x_5 - \bar{x})(y_5 - \bar{y}) &= (03)(11) &&= 33\\
(x_6 - \bar{x})(y_6 - \bar{y}) &= \text{(-05)(-20)} &&= 100\\
(x_7 - \bar{x})(y_7 - \bar{y}) &= (01)(00) &&= 0\\
(x_8 - \bar{x})(y_8 - \bar{y}) &= (01)(03) &&= 3
\end{aligned}
$$

Determinaremos os totais das colunas $(x_i - \bar{x})^2$ e $(x_i - \bar{x})(y_i - \bar{y})$ da Tabela 15, como segue:

$$\sum_{i=1}^{8} (x_i - \bar{x})^2 = 90$$

$$\sum_{i=1}^{8} (x_i - \bar{x})(y_i - \bar{y}) = 272$$

Determinação da covariância amostral s_{xy}, utilizando o total da coluna $(x_i - \bar{x})(y_i - \bar{y})$ da Tabela 15, dividido pelo número de registros menos um, pois estamos trabalhando com amostras, como segue:

$$s_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{n-1} = \frac{272}{8-1} = \frac{272}{7} = 38,86$$

Determinação da variância amostral S_x^2, utilizando o total da coluna $(x_i - \bar{x})^2$ da Tabela 15, ou seja, o total dos desvios de x_i elevado ao quadrado dividido pelo numero de registros menos um, pois estamos trabalhando com amostras como segue:

$$s_x^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1} = \frac{90}{8-1} = \frac{90}{7} = 12,86$$

Determinação do parâmetro a da reta de regressão, basta dividir a covariância S_{xy} pela variância S_x^2, ambas determinadas anteriormente, assim temos:

$$a = \frac{s_{xy}}{s_x^2} = \frac{38,86}{12,86} = 3,02$$

Determinação do parâmetro b utilizaremos as médias de $\bar{x}$, $\bar{y}$ da Tabela 15, e o parâmetro a determinado anteriormente. Asim teremos:

$$b = \bar{y} - a\bar{x} \Rightarrow a = 35 - (3,02 \cdot 7) = 13,86$$

Assim a equação da reta é data por:

$$\hat{y} = 3,02x + 13,86$$

Questão 5

Determine através da equação da reta determinada na questão nº 4 a previsão de produção em toneladas para quatro e nove horas trabalhadas.

Resposta: Para quatro e nove horas trabalhadas temos:

$$\hat{y} = 3,02x + 13,86 \Rightarrow \hat{y} = 3,02(4) + 13,86 \Rightarrow \hat{y} = 25,94 \text{ toneladas.}$$

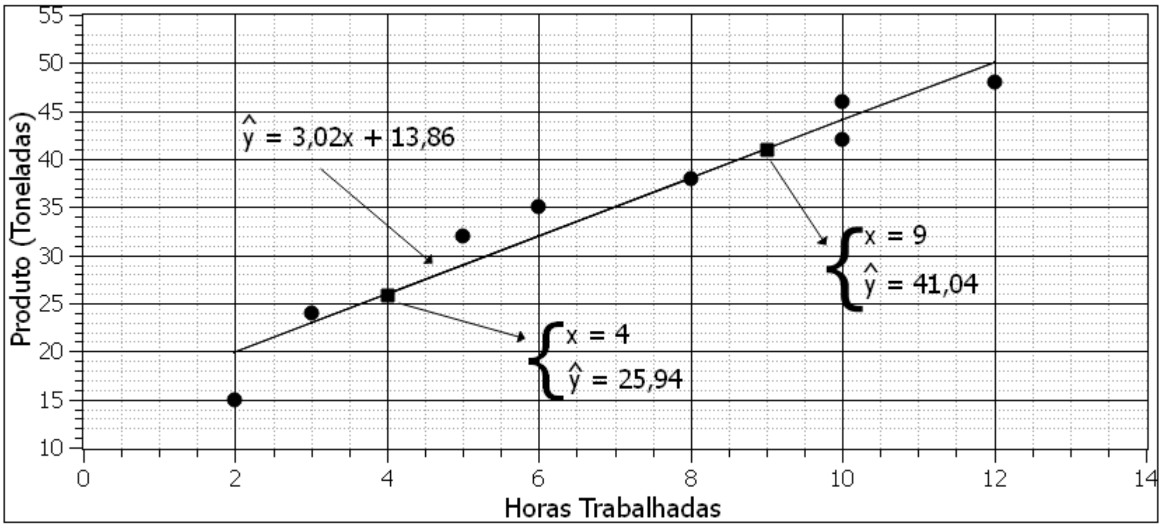
$$\hat{y} = 3,02x + 13,86 \Rightarrow \hat{y} = 3,02(9) + 13,86 \Rightarrow \hat{y} = 41,04 \text{ toneladas.}$$

Questão 6

Com o gráfico da questão nº 2 e os dois pontos determinados na questão 5, localize os dois pontos e trace a reta de regressão.

Resposta: Gráfico da Figura 19.

Figura 19. Gráfico de dispersão - Horas Trabalhadas versus Produção



Fonte: Autoria Própria.

Questão 7

Determine o Coeficiente de Determinação.

Resposta: $r^2 = 94\%$

Para determinar o coeficiente de determinação "r^2", utilizaremos duas tabelas. A primeira é mostrada pela Tabela 16.

Tabela 16. Desvios de y em relação a média aritmética ($\bar{y}$) da coluna Y

i	X	Y	$\bar{y}$	$(y_i - \bar{y})$	$(y_i - \bar{y})^2$
1	3	24	35	-11	121
2	05	32	35	-03	09
3	10	42	35	07	49
4	12	48	35	13	169
5	10	46	35	11	121
6	02	15	35	-20	400
7	06	35	35	0	0
8	08	38	35	03	09
$\sum$		280			878
Média		35			

Fonte: Autoria Própria.

As colunas X e Y da Tabela 16 é a mesma da Tabela 15. O total da coluna Y, $\sum Y = 280$ e sua média é: $\bar{y} = 35$, podendo ser copiados também da Tabela 15. A coluna $\bar{y}$ da Tabela 16, deve ser toda preenchida com o valor da média aritmética da coluna Y que é $\bar{y} = 35$.

Os cálculos para o preenchimento da coluna $y_i - \bar{y}$ da Tabela 16 são determinados como segue:

$$
\begin{aligned}
(y_1 - \bar{y}) &= \text{24 - 35} = -11 \\
(y_2 - \bar{y}) &= \text{32 - 35} = -03 \\
(y_3 - \bar{y}) &= \text{42 - 35} = 07 \\
(y_4 - \bar{y}) &= \text{48 - 35} = 13 \\
(y_5 - \bar{y}) &= \text{46 - 35} = 11 \\
(y_6 - \bar{y}) &= \text{15 - 35} = -20 \\
(y_7 - \bar{y}) &= \text{35 - 35} = 0 \\
(y_8 - \bar{y}) &= \text{38 - 35} = 03
\end{aligned}
$$

A coluna $(y_i - \bar{y})^2$ da Tabela 16 é determinada pelos cálculos a seguir:

$$(y_1 - \bar{y})^2 = \quad (-11)^2 = \quad 121$$
$$(y_2 - \bar{y})^2 = \quad (-03)^2 = \quad 09$$
$$(y_3 - \bar{y})^2 = \quad (07)^2 = \quad 49$$
$$(y_4 - \bar{y})^2 = \quad (13)^2 = \quad 169$$
$$(y_5 - \bar{y})^2 = \quad (11)^2 = \quad 121$$
$$(y_6 - \bar{y})^2 = \quad (-20)^2 = \quad 400$$
$$(y_7 - \bar{y})^2 = \quad (0)^2 = \quad 0$$
$$(y_8 - \bar{y})^2 = \quad (03)^2 = \quad 09$$

A outra tabela a ser utilizada possui as colunas i, X, Y, com os totais e médias aritméticas das colunas X, Y como mostra a Tabela 17. Estes valores são os mesmos da Tabela 16.

Tabela 17. Dados para determinação da Soma de Quadrado Total

i	X	Y	$\hat{y}_i$	$\hat{e}_i$	$(\hat{e}_i)^2$
1	3	24	22,92	-1,08	1,17
2	5	32	28,96	-3,04	9,24
3	10	42	44,06	2,06	4,24
4	12	48	50,10	2,10	4,41
5	10	46	44,06	-1,94	3,76
6	2	15	19,90	4,90	24,01
7	6	35	31,98	-3,02	9,12
8	8	38	38,02	0,02	0,00
$\sum$	56	280			55,96
Média	7	35			

Fonte: Autoria Própria.

A coluna $\hat{y}_i$ da Tabela 17, deve ser preenchida com todos os valores da coluna X, aplicados na equação da reta $\hat{y} = 3,02x + 13,86$, determinada na questão 4, deste apêndice. Segue os cálculos:

$$\hat{y}_1 = 3,02x + 13,86 \Rightarrow \hat{y}_1 = 3,02(3) + 13,86 \Rightarrow \hat{y}_1 = 22,92$$

$\hat{y_2} = 3,02x + 13,86 \Rightarrow \hat{y_2} = 3,02(5) + 13,86 \Rightarrow \hat{y_2} = 28,96$

$\hat{y_3} = 3,02x + 13,86 \Rightarrow \hat{y_3} = 3,02(14) + 13,86 \Rightarrow \hat{y_3} = 44,06$

$\hat{y_4} = 3,02x + 13,86 \Rightarrow \hat{y_4} = 3,02(15) + 13,86 \Rightarrow \hat{y_4} = 50,10$

$\hat{y_5} = 3,02x + 13,86 \Rightarrow \hat{y_5} = 3,02(17) + 13,86 \Rightarrow \hat{y_5} = 44,06$

$\hat{y_6} = 3,02x + 13,86 \Rightarrow \hat{y_6} = 3,02(18) + 13,86 \Rightarrow \hat{y_6} = 19,90$

$\hat{y_7} = 3,02x + 13,86 \Rightarrow \hat{y_7} = 3,02(20) + 13,86 \Rightarrow \hat{y_7} = 31,98$

$\hat{y_8} = 3,02x + 13,86 \Rightarrow \hat{y_5} = 3,02(21) + 13,86 \Rightarrow \hat{y_8} = 38,02$

A coluna $\hat{e_i}$ da Tabela 17, deve ser preenchida pela diferença entre cada y_i da coluna Y e o seu correspondente $\hat{y_i}$. Assim a determinação dos cálculos são:

$\hat{e_1} = y_1 - \hat{y_1} \Rightarrow \hat{e_1} = 24 - 22,92 \Rightarrow \hat{e_1} = -1,08$

$\hat{e_2} = y_2 - \hat{y_2} \Rightarrow \hat{e_2} = 32 - 28,96 \Rightarrow \hat{e_2} = -3,04$

$\hat{e_3} = y_3 - \hat{y_3} \Rightarrow \hat{e_3} = 42 - 44,06 \Rightarrow \hat{e_3} = 2,06$

$\hat{e_4} = y_4 - \hat{y_4} \Rightarrow \hat{e_4} = 48 - 50,10 \Rightarrow \hat{e_4} = 2,10$

$\hat{e_5} = y_5 - \hat{y_5} \Rightarrow \hat{e_5} = 46 - 44,06 \Rightarrow \hat{e_5} = -1,94$

$\hat{e_6} = y_6 - \hat{y_6} \Rightarrow \hat{e_6} = 15 - 19,90 \Rightarrow \hat{e_6} = 4,90$

$\hat{e_7} = y_7 - \hat{y_7} \Rightarrow \hat{e_7} = 35 - 31,98 \Rightarrow \hat{e_7} = -3,02$

$\hat{e_8} = y_8 - \hat{y_8} \Rightarrow \hat{e_8} = 38 - 38,02 \Rightarrow \hat{e_8} = 0,02$

A coluna $\hat{e}^2$ e determinada como segue:

$\hat{e_1}^2 = (-1,08)^2 \Rightarrow \hat{e_1}^2 = 1,17$

$\hat{e_2}^2 = (-3,04)^2 \Rightarrow \hat{e_2}^2 = 9,24$

$\hat{e_3}^2 = (2,06)^2 \Rightarrow \hat{e_3}^2 = 4,24$

$\hat{e_4}^2 = (2,10)^2 \Rightarrow \hat{e_4}^2 = 4,41$

$\hat{e_5}^2 = (-1,94)^2 \Rightarrow \hat{e_5} = 3,76$

$\hat{e_6}^2 = (4,90)^2 \Rightarrow \hat{e_6} = 24,01$

$\hat{e_7}^2 = (-3,02)^2 \Rightarrow \hat{e_7} = -9,12$

$\hat{e_8}^2 = (0,02)^2 \Rightarrow \hat{e_8} = 0$

O valor da Soma dos quadrados dos erros SQE é obtido do total da coluna $\hat{e}_i$ da tabela 17, assim temos:

$$SQE = \sum_{i=1}^{n}(y_i - \hat{y}) = \sum_{i=1}^{n}\hat{e}_i = 55,96$$

Para a determinação da *soma de quadrados da regressão* SQR temos a diferença entre a *soma dos quadrados totais* SQT, que é a soma da coluna $(y_i - \bar{y})^2$ da Tabela 16 e a soma dos quadrados dos erros que é o somatório da coluna $\hat{e}_i$ da tabela 17, assim temos:

$SQR = SQT - SQE \Rightarrow SQR = 878 - 55,96 = 822,04$

Para a determinação do r^2 temos:

$$r^2 = \frac{SQR}{SQT} = \frac{\text{variação explicada}}{\text{variação total}} \Rightarrow R^2 = \frac{822,04}{878} = 0,94 \text{ ou } 94\%$$

Questão 8

Determine o Coeficiente de Correlação r.

Resposta: $r = 0,97$

A equação para a determinação do Coeficiente de Correlação é mostrado a seguir:

$$r = \frac{Cov(X,Y)}{S_x S_y}$$

Utilizaremos a covariância $Cov(X,Y) = s_{xy} = 38,86$, determinada na questão nº 4 e a variância s_x^2 determinada na questão nº 4, para determinar o desvio padrão s_x:

$$s_x^2 = 12,86 \Rightarrow s_x = \sqrt{s_x^2} \Rightarrow s_x = \sqrt{12,86} \Rightarrow s_x = 3,59$$

Para a determinação do desvio padrão s_y utilizaremos o total da coluna $(y_i - \bar{y})^2$ da Tabela 16 na questão nº 7. Assim temos:

$$s_y = \sqrt{s_y^2} = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \bar{y})^2}{n-1}} = \sqrt{\frac{878}{7}} = \sqrt{125,43} \Rightarrow s_y = 11,20$$

Determinados a covariância s_{xy} e os desvios padrão s_x e s_y, segue a determinação do Coeficiente de Correlação:

$$r = \frac{Cov(X,Y)}{S_x S_y} \Rightarrow r = \frac{38,86}{3,59 \cdot 11,20} \Rightarrow r = \frac{38,86}{3,59 \cdot 11,20} \Rightarrow r = \frac{38,86}{40,21} \Rightarrow r = 0,97$$

Pelo valor do Coeficiente de Correlação $r = 0,97$, podemos afirmar que é forte a força de dependência entre as variáveis.

F Avaliação

Avaliação para o professor aplicar em sala de aula, podendo ser individual ou em grupo.

Pressão arterial sistólica (PAS) é o maior valor verificado durante a aferição de pressão arterial. Exemplo: 120x80 significa: 120 refere-se à pressão arterial sistólica e 80 à pressão arterial diastólica, ambas medidas em milímetros de mercúrio (mmHg).

A Tabela 18, mostra as idades e a pressão sistólica de doze mulheres.

Tabela 18. Idade versus pressão arterial sistólica

Idade	56	42	72	36	63	47	55	49	38	42	68	60
Pressão	147	125	160	118	149	128	150	145	115	140	152	155

Questão 1

Definindo-se a idade como variável independente e denominada como o conjunto:

$X = \{56, 42, 72, 36, 63, 47, 55, 49, 38, 42, 68, 60\}$

e a pressão arterial sistólica como a variável dependente denominada como o conjunto:

$Y = \{147, 125, 160, 118, 149, 128, 150, 145, 115, 140, 152, 155\}$,

determine os pares ordenados desta relação.

Resposta:

01. $(x_1, y_1) = (56, 147)$.

02. $(x_2, y_2) = (42, 125)$.

03. $(x_3, y_3) = (72, 160)$.

04. $(x_4, y_4) = (36, 118)$.

05. $(x_5, y_5) = (63, 149)$.

06. $(x_6, y_6) = (47, 128)$.

07. $(x_7, y_7) = (55, 150)$.

08. $(x_8, y_8) = (49, 145)$.

09. $(x_8, y_8) = (38, 115)$.

10. $(x_8, y_8) = (42, 140)$.

11. $(x_8, y_8) = (68, 152)$.

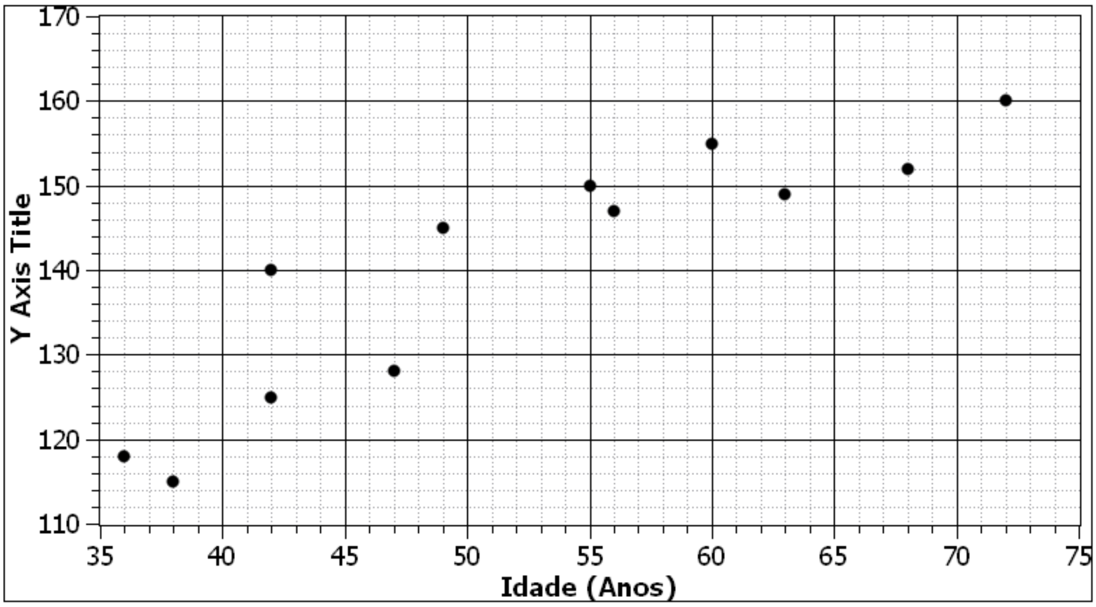
12. $(x_8, y_8) = (60, 155)$.

Questão 2

Construa um gráfico de dispersão e localize os pares ordenados?

Resposta:Gráfico da Figura 20.

Figura 20. Gráfico de dispersão - Idade versus Pressão



Fonte: Autoria Própria.

Questão 3

Depois de construído o gráfico, o que você diria sobre a tendência da disposição dos pontos?

Respostas: Os pontos tem uma tendência positiva, ou seja, quando a idade da mulheres aumentam a tendência de um aumento na pressão arterial sistólica.

Questão 4

Determine a equação da reta de regressão linear entre a idade e a pressão arterial sistólica (mmHg).

Resposta: A equação é: $\hat{y} = 1,14x + 80,67$

Para a determinação dos parâmetros a e b da equação da reta de regressão linear, construiremos uma tabela e definiremos como X e Y as colunas contendo dos dados da *Idade* e *Pressão Arterial Sistólica*, respectivamente. Os campos da tabela estão discriminados abaixo:

i :	Número de registros
X :	Idade
Y :	Pressão em mmHg
$(x_i - \bar{x})$:	Desvios de x
$(y_i - \bar{y})$:	Desvios de y
$(x_i - \bar{x})^2$:	Desvios de x ao quadrado
$(x_i - \bar{x})(y_i - \bar{y})$:	Multiplicação entre os desvios de x e y

A Tabela 19 mostra os valores de cada coluna assim como o Total da colunas X, Y, $(x_i - \bar{x})^2$ e $(x_i - \bar{x})(y_i - \bar{y})$ assim como as médias $\bar{x}$ e $\bar{y}$ das colunas X e Y respectivamente e que serão utilizadas para a determinação dos parametros a e b da equação da reta de regressão linear.

Tabela 19. Idade versus pressão arterial sistólica

i	X	Y	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$	$(x_i - \bar{x})(y_i - \bar{y})$
1	56	147	3,67	6,67	13,47	44,49	24,48
2	42	125	- 10,33	-15,33	106,71	235,01	158,36
3	72	160	19,67	19,67	386,91	386,91	386,91
4	36	118	- 16,33	-22,33	266,67	498,63	364,65
5	63	149	10,67	8,67	113,85	75,17	92,51
6	47	128	- 5,33	-12,33	28,41	152,03	65,72
7	55	150	2,67	9,67	7,13	93,51	25,82
8	49	145	- 3,33	4,67	11,09	21,81	- 15,55
9	38	115	- 14,33	-25,33	205,35	641,61	362,98
10	42	140	- 10,33	-0,33	106,71	0,11	3,41
11	68	152	15,67	11,67	245,55	136,19	182,87
12	60	155	7,67	14,67	58,83	215,21	112,52
$\sum$	628	1.684			1.550,68	2500,68	1.764,67
Média	52,33	140,33					

Determinação da covariância amostral s_{xy}, utilizando o total da coluna $(x_i - \bar{x})(y_i - \bar{y})$, da Tabela 19, dividido pelo número de registros menos um, pois estamos trabalhando com amostras, como segue:

$$s_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{n-1} = \frac{1764,67}{12-1} = \frac{1764,67}{11} = 160,42$$

Determinação da variância amostral S_x^2, utilizando o total da coluna $(x_i - \bar{x})^2$ da Tabela 19, ou seja, o total dos desvios de x_i elevado ao quadrado dividido pelo número de registros menos um, pois estamos trabalhando com amostras como segue:

$$s_x^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1} = \frac{1550,68}{12-1} = \frac{1550,68}{11} = 140,97$$

Determinado a covariância s_{xy} e a variância s_x^2, para determinar o parâmetro a da reta de regressão, basta dividir a covariância S_{xy} pela variância S_x^2, ambas determinadas anteriormente, assim temos:

$$a = \frac{s_{xy}}{s_x^2} = \frac{160,42}{140,97} = 1,1379 \Rightarrow a = 1,138$$

Para a determinação do parâmetro b utilizaremos as médias de $\bar{x}$, $\bar{y}$ da Tabela 19, e o parâmetro a determinado anteriormente. Asim teremos:

$$b = \bar{y} - a\bar{x} \Rightarrow a = 140,33 - (1,14 \cdot 52,33) = 80,67$$

Assim a equação da reta é data por:

$$\hat{y} = 1,14x + 80,67$$

Questão 5

Determine através da equação determinada na questão nº 4 a previsão de pressão sistólica de duas mulheres, com idade de 50 e 65 anos.

Resposta: Para quatro e nove horas trabalhadas temos:

$$\hat{y} = 1,14x + 80,67 \Rightarrow \hat{y} = 1,14(50) + 80,67 \Rightarrow \hat{y} = 137,67 \text{ mmHg.}$$

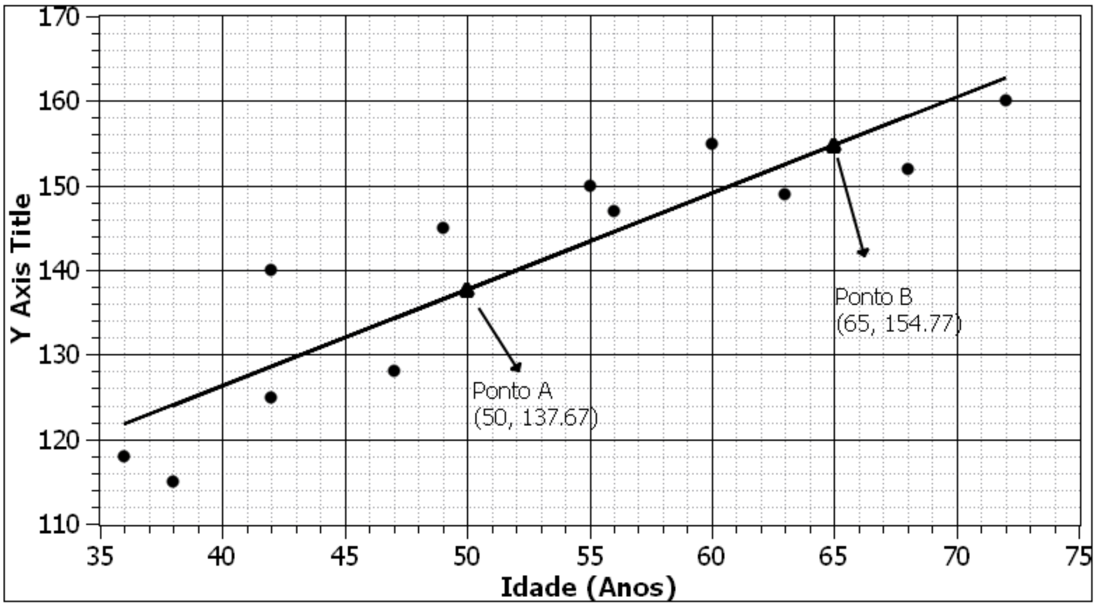
$$\hat{y} = 1,14x + 80,67 \Rightarrow \hat{y} = 1,14(65) + 80,67 \Rightarrow \hat{y} = 154,77 \text{ mmHg.}$$

Questão nº 6

Com o gráfico da questão nº 2 e os dois pontos determinados na questão nº 5, trace a reta de regressão passando por esses dois pontos.

Resposta: Gráfico da Figura 21.

Figura 21. Reta de Regressão: Idade versus Pressão Sistólica



Fonte: Autoria Própria.

Questão 7

Determine o Coeficiente de Determinação.

Resposta: $r^2 = 80\%$, isto significa que a reta de regressão linear $\hat{y} = 1,14x + 80,67$, determinada na questão nº 4 representa em 80 % os pares ordenados plotados no gráfico de dispersão da questão nº 2.

Para a determinação do coeficiente de correlação "r^2", utilizaremos duas tabelas. A primeira é mostrada pela Tabela 20. Ela contem 6 colunas onde:

i :	Número de registros
X :	Idade
Y :	Pressão Arterial Sistólica
$\bar{y}$:	Média de Y
$y_i - \bar{y}$:	Desvios de y
$(y_i - \bar{y})^2$:	Desvios de y ao quadrado

Tabela 20. Desvios de y em relação a média aritmética $\bar{y}$ da coluna Y

i	X	Y	$\bar{y}$	$(y_i - \bar{y})$	$(y_i - \bar{y})^2$
1	56	147	140,33	6,67	44,49
2	42	125	140,33	-15,33	235,00
3	72	160	140,33	19,67	386,91
4	36	118	140,33	-22,33	498,63
5	63	149	140,33	8,67	75,17
6	47	128	140,33	-12,33	152,03
7	55	150	140,33	9,67	93,51
8	49	145	140,33	4,67	21,81
9	38	115	140,33	-25,33	641,61
10	42	140	140,33	-0,33	0,11
11	68	152	140,33	11,67	136,19
12	60	155	140,33	14,67	215,21
$\sum$		1.684			2.500,68
Média		140,33			

Fonte: Autoria Própria.

A outra tabela a ser utilizada possui as colunas i, X, Y, $\hat{y}_i$, $\hat{e}_i$ e $(\hat{e}_i)^2$. As colunas i, X, Y são idênticas aos da Tabela 21. A coluna $\hat{y}_i$ são todos os valores do conjunto definido pela equação da reta $\hat{y} = 1.14x + 80,67$, com as entradas sendo os valores da coluna X da Tabela 20. A coluna $\hat{e}_i$ são o valores dos resíduos, que são determinados entre a diferença do valor y_i e o determinado através da equação $\hat{y} = 1.14x + 80,67$, ou seja, $y_i - \hat{y}$. A coluna $(\hat{e}_i)^2$ é o resíduo elevado ao quadrado. A Tabela 21, mostra como deve ficar o preenchimento. Com as Tabela preenchidas podemos determinar a *soma dos quadrados dos erros* SQE, a *soma dos quadrados de regressão* SQR e a *soma dos quadrados totais* SQT, que serão utilizados para a determinação do *Coeficiente de Determinação* r^2.

Tabela 21. Determinação da Soma de Quadrado Total

i	X	Y	$\hat{y}_i$	$\hat{e}_i$	$(\hat{e}_i)^2$
1	56	147	144,51	-2,49	6,2
2	42	125	128,55	3,55	12,6
3	72	160	162,75	2,75	7,56
4	36	118	121,71	3,71	13,76
5	63	149	152,49	3,49	12,18
6	47	128	134,25	6,25	39,06
7	55	150	143,37	-6,63	43,96
8	49	145	136,53	-8,47	71,74
9	38	115	123,99	8,99	80,82
10	42	140	128,55	-11,45	131,1
11	68	152	158,19	6,19	38,32
12	60	155	149,07	-5,93	35,16
$\sum$	628	1684			492,46
Média	52,33	140,33			

Fonte: Autoria Própria.

Soma dos quadrados dos erros (SQE): O valor da Soma dos quadrados dos erros SQE é obtido do total da coluna $\hat{e}_i$ da tabela 21, assim temos:

$$SQE = \sum_{i=1}^{n}(y_i - \hat{y}) = \sum_{i=1}^{12} \hat{e}_i = 492,46$$

Para a determinação da **soma de quadrados da regressão** SQR temos a diferença entre a **soma dos quadrados totais** SQT, que é a soma da coluna $(y_i - \bar{y})^2$ da Tabela 20 e a **soma dos quadrados dos erros** SQE que é o somatório da coluna $\hat{e}_i$ da tabela 21, assim temos:

$$SQR = SQT - SQE \Rightarrow SQR = 2500,68 - 492,46 = 2.008,22$$

Para a determinação do R^2 temos:

$$r^2 = \frac{SQR}{SQT} = \frac{\text{variação explicada}}{\text{variação total}} \Rightarrow R^2 = \frac{2.008,22}{2.500,68} = 0,80 \text{ ou } 80\%$$

Questão 8

Determine o Coeficiente de Correlação.

A equação para a determinação do Coeficiente de Correlação é mostrado a seguir:

$$r = \frac{Cov(X,Y)}{S_x S_y}$$

Utilizaremos a covariância $Cov(X,Y) = s_{xy} = 160,42$, determinada na questão nº 4 e a variância s_x^2 determinada na questão nº 4, para determinar o desvio padrão s_x:

$$s_x^2 = 140,97 \Rightarrow s_x = \sqrt{s_x^2} \Rightarrow s_x = \sqrt{140,97} \Rightarrow s_x = 11,87$$

Para a determinação do desvio padrão s_y utilizaremos o total da coluna $(y_i - \bar{y})^2$ da Tabela 16 na questão nº 7. Assim temos:

$$s_y = \sqrt{s_y^2} = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \bar{y})^2}{n-1}} = \sqrt{\frac{2.500,68}{11}} = \sqrt{227,34} \Rightarrow s_y = 15,08$$

Determinados a covariância s_{xy} e os desvios padrão s_x e s_y, segue a determinação do Coeficiente de Correlação:

$$r = \frac{Cov(X,Y)}{S_x S_y} \Rightarrow r = \frac{160,42}{11,87 \cdot 15,08} \Rightarrow r = \frac{160,42}{179} \Rightarrow r = 0,90$$

Pelo valor do Coeficiente de Correlação $r = 0,90$, podemos afirmar que é forte a força de dependência entre as variáveis.