

Eduardo Félix Pita Duarte

**Fatores que Influenciam no Rendimento Académico da disciplina de Matemática dos
alunos da 10^a classe da Escola Secundaria Heróis Moçambicanos de Moatize em 2021:
Um olhar pela Regressão Linear Múltipla**

Licenciatura em Ensino de Matemática com Habilitação em Estatística

Universidade Pungue – Extensão de Tete

Tete, 2022

Eduardo Félix Pita Duarte

**Fatores que Influenciam no Rendimento Académico da disciplina de Matemática dos
alunos da 10^a classe da Escola Secundaria Heróis Moçambicanos de Moatize em 2021:
Um olhar pela Regressão Linear Múltipla**

Licenciatura em Ensino de Matemática com Habilitação em Estatística

Monografia científica do curso licenciatura em ensino de matemática com habilitação em estatística, a ser entregue no departamento de ciências exatas e tecnológicas como requisito final para obtenção do grau de licenciado em ensino de matemática com habilitação em estatística, sob a supervisão do:

Mestre Abrantes Mussafo

Universidade Pungue – Extensão de Tete
Tete, 2022

Índice

Declaração	iv
Dedicatória.....	v
Agradecimentos	vi
Resumo	vii
Abstract.....	viii
Índice de Figuras.....	ix
Lista de Tabelas	x
Lista de Abreviaturas	xi
CAPITULO I.....	11
1.0 Introdução	11
1.1 Justificativa do Estudo	12
1.2 Problematização.....	12
1.3 Objetivos.....	13
1.3.1 Objetivos Geral	13
1.3.2 Objetivos Específicos.....	14
1.4 Hipóteses.....	14
CAPITULO II.....	15
2.0 Metodologia de Pesquisa	15
2.1 Tipo de Pesquisa	15
2.1.1 Quanto a abordagem.....	15
2.1.2 Quanto a natureza.....	16
2.1.3 Quanto aos objetivos	16
2.1.4 Quanto aos procedimentos	16
2.2 Técnica e Instrumentos de Coleta de Dados	16
2.3 População e Amostra	17
2.4 Técnica de Tratamento de dados	18
2.4.1 Descrição das técnicas.....	19
2.4.2 Tabela das Variáveis	22
CAPITULO III	23
3.0 Fundamentação Teórica	23
3.0.1 Modelo de Regressão Linear Múltipla (MRLM)	23
3.0.2 Modelo com Iteração.....	23

3.0.2 Fatores que influenciam no rendimento escolar.....	34
3.0.3 Rendimento Escolar em Matemática.....	38
CAPITULO IV	40
4.0 Análise e Interpretação dos dados	40
4.0.1 Analise Descritiva dos dados	40
4.0.2 Medidas de Assimetria e Curtose.....	41
4.0.3 Análise de Frequências.....	42
4.1 Fatores que influenciam no rendimento da disciplina de matemática	44
4.1.1 Análise da correlação usando o coeficiente de Pearson	44
4.1.2 Modelo de regressão linear múltipla (MRLM)	45
CAPITULO V	49
5.0 Considerações Finais	49
5.1 Sugestões	51
5.2 Barreiras e Limitações	51
5.3 Referencias Bibliográficas	52
CAPITULO VI	54
6.0 Apêndices:	54
6.1 Anexos	59

Declaração

Eu, Eduardo Félix Pita Duarte, declaro por minha honra que esta monografia é o resultado do meu próprio trabalho e esta a ser submetida para a obtenção do grau de licenciado em ensino de matemática com habilitação em estatística na Faculdade de Ciências Exatas e Tecnológicas da Universidade Pungue – Extensão de Tete. Não foi antes submetida para obtenção de nenhum grau ou para avaliação em nenhuma outra universidade.

Tete, Março de 2022

(Eduardo Félix Pita Duarte)

Dedicatória

Aos meus pais Félix Pita Duarte (*em memória*) e Isabel Bernardo pelo exemplo dos bons ensinamentos da vida.

Aos meus irmãos Geraldo Félix, Filda Elias, Palmira Félix (*em memória*) pelo amor e carinho, apoiando-me, aconselhando-me em cada passo e sempre colocando fé em mim.

Ao meu primo, Moisés Duarte, por acreditar em mim e por ter-me dado a oportunidade de me formar e pelo apoio.

A mim mesmo por nunca ter me deixado levar pelas emoções da juventude, por ter focado me nos meus estudos desde o início até chegar a essa etapa de conclusão do meu curso.

Agradecimentos

Agradecer a Deus por cuidar de mim em cada trilha da minha vida e pelo seu amor inabalável por mim, mesmo que quase sempre não o mereça.

A Universidade Pungue, em especial a Faculdade de Ciências Exatas e Tecnológicas, pela oportunidade de realizar o curso de graduação em Licenciatura em Ensino de Matemática com Habilitação em Estatística.

Aos meus pais Félix Pita Duarte (*em memória*) e Isabel Bernardo pelo exemplo dos bons ensinamentos da vida.

Aos meus irmãos Geraldo Félix, Filda Elias, Palmira Félix (*em memória*), minha e minha sobrinha Laura pelo amor e carinho, apoiando-me, aconselhando-me em cada passo e sempre colocando fé em mim.

Ao meu supervisor, Mestre Abrantes João Mussafo, pela confiança e paciência de trabalhar comigo e por todo o incentivo desde o início.

A mim mesmo, por tudo que tenho feito por mim mesmo, por ter me apoiado, me aconselhado, me motivado e por tudo que fiz por mim durante a caminhada acadêmica, minha auto ajuda foi crucial para conseguir chegar ate esta etapa.

A Escola Secundaria Heróis Moçambicanos de Moatize pelo apoio material para a materialização desse trabalho.

Aos meus colegas e amigos, em especial aos do Curso, pelo companheirismo, e acima de tudo pelos ensinamentos dados através da convivência quotidiana.

A todos que, de forma direta e indireta, contribuíram para a realização deste trabalho.

A todos, vai meu muito obrigado!

Resumo

O presente trabalho visa analisar os fatores que influenciam no rendimento acadêmico da disciplina de matemática dos alunos da 10ª classe da escola secundária heróis moçambicanos de moatize. Quanto a metodologia de pesquisa, trata-se abordagem quantitativa, de natureza aplicada quanto aos objetivos ela é explicativa e por último de levantamento estatístico. Dada a vasta extensão da população em estudo, recorreu-se a uma amostragem por conveniência de onde surgiu uma amostra de quatrocentos e vinte e um alunos. A técnica de recolha de dados é a aplicação de um questionário contendo questões fechadas com vista a proporcionar maior sucesso a pesquisa. A principal técnica de análise dos dados é a regressão linear múltipla com auxílio da análise descritiva e de frequências. A classe selecionada para a recolha dos dados é composta na sua totalidade por seis turmas divididas em três sub-turmas por turma, reunindo um universo de quatrocentos e cinquenta e dois alunos. Como resultado desta pesquisa, foi possível construir o modelo de regressão linear múltipla que através dele foi possível concluir que a variável número de refeições (NRef) é a que mais impacta no rendimento acadêmico da disciplina de matemática, com isto o modelo sugere que os alunos devem ter em média três refeições diárias (como sugere também a distribuição da variável NRef) para que tenham maior sucesso na disciplina de matemática.

Palavras - Chave: Fatores, rendimento, Matemática, regressão, linear.

Abstract

The analyzable work seen the class factors of the discipline of academic students of the 0th class of the school of Moatize students. As for the research methodology, as for the research methodology, as for the research methodology, as to the objective nature, it is explanatory and, finally, it deals with a statistical study. Given the vast extent of the population under study, there is an extent where a sample comes from and an extent of four hundred and twenty students. The data collection technique is the application of a more detailed research question. The main data analysis technique is a linear analysis with the aid of frequency analysis. The class selected for a collection of data is the totality of six classes made up of its class and two sub-classes per class, bringing together a universe of four hundred students. As a result of this research, it was possible to build the study model is linear, through which it was possible to use as much as possible from a variable of meals () to a number of impact of mathematics discipline, with this the model suggests that students should take into account media three meals a day (as the distribution of the NRef variable also suggests) for them to be more successful in the mathematics discipline.

Keywords: Factors, yield, mathematic, regression, linear.

Índice de Figuras

Figura 1: Gráfico P-P dos resíduos (Fonte: Rodrigues(2012)).....	25
Figura 2: Confirmação da homoscedasticidade dos resíduos (Fonte: Rodrigues (2012)).....	27
Figura 3: Histograma de distribuição normal (Fonte: autor).....	47
Figura 4: Gráfico dos Resíduos vs Valores Ajustados (Fonte: Autor).....	48

Lista de Tabelas

Tabela 1: Distribuição do número de alunos da 10ª classe em turmas (Fonte: ESHM).....	18
Tabela 2: Tabela das Variáveis (Fonte: Autor)	22
Tabela 3: Tabela de valores críticos de d_L e d_U . (Fonte: Rodrigues (2012))	28
Tabela 4: Tabela de decisão em função de d_L e d_U . (Fonte: Rodrigues (2012)).....	28
Tabela 5: Análise descritiva das variáveis quantitativas (fonte: Autor).....	40
Tabela 6: Medidas de assimetria e curtose (Fonte: Autor).....	42
Tabela 7: Análise de frequências das variáveis categóricas (fonte: autorg).....	43
Tabela 8: Teste de coeficiente de Pearson (Fonte: autor).....	44
Tabela 9: Resumo do Modelo (Fonte: Autor)	45
Tabela 10: ANOVA do Modelo (Fonte: autor)	46
Tabela 11: Coeficientes do Modelo (Fonte: autor).....	46

Lista de Abreviaturas

ESHMM	Escola Secundaria Heróis Moçambicanos de Moatize
IBM	Internacional Business Machine Corporation
SPSS	Statistic Package for Social Sciences (Pacote Estatístico para as Ciências Sociais)
MRLM	Modelo de Regressão Linear Múltipla
MMQ	Método de mínimos Quadrados
SQE	Soma dos quadrados dos Erros
MFM	Media Final de Física
MFP	Media Final de Português
MFF	Media Final de Física
MFQ	Media Final de Química
NRef	Número de Refeições por dia
NPess	Número de Pessoas que vivem em sua residência
CV	Coefficiente de variação
DCE	Distância de casa ate escola
TCE	Tempo que leva ao sair de casa ate escola

CAPITULO I

1.0 Introdução

Com o intuito de detetar o fator que mais impacta no rendimento académico da disciplina de matemática, surge a ideia de estudar os fatores, razões, motivos e aspetos em geral que contribuem para o bom ou mau desempenho dos alunos nessa disciplina específica.

Estudar estes fatores significa procurar soluções para possíveis casos de fraco desempenho em matemática por diversos alunos das escolas nacionais, encontrados os fatores influenciadores, é importante também realçar os fatores que mais impactam no rendimento da disciplina em destaque. Por isso, com o auxílio da ferramenta estatística que é a regressão linear múltipla, pretende-se estudar estes fatores para se avaliar as variáveis que mais impacta no rendimento académico da disciplina de matemática.

No que concerne a organização deste trabalho, dizer que é composto por cinco (5) capítulos na sua totalidade onde o primeiro capítulo enquadra especificamente a fase introdutória deste trabalho, contendo os elementos justificativos da pesquisa e sua contextualização, o capítulo dois (2) do trabalho engloba as metodologias do trabalho e outros aspetos importantes da pesquisa que culminam na descrição geral da população em estudo, o capítulo três (3) do trabalho faz referência a fundamentação teórica geral dos assuntos a serem descritos neste trabalho, o capítulo quatro (4) descreve as análises e interpretações dos dados colhidos desde a análise descritiva, de frequências até a execução da própria regressão linear múltipla, por último, o capítulo cinco (5) inclui os aspetos finais da pesquisa desde a conclusão, limitações e sugestões até os anexos.

1.1 Justificativa do Estudo

A escolha do tema em destaque é resultado das intenções do autor de fazer alguma contribuição na melhoria dos resultados de ensino - aprendizagem e aproveitamento acadêmico dos alunos na disciplina de matemática.

O problema foi observado no momento em que o autor desenvolvia as atividades de estágio pedagógico no ensino de matemática na referida escola, observou-se uma fraca aderência dos alunos nas aulas de matemática, ou por outro lado, em caso de aderência, o foco e dedicação nas aulas era muito reduzido, resultando assim num fraco desempenho e resultado desencorajador.

Desta feita, a identificação dos reais problemas por detrás do fraco empenho dos alunos nas aulas de matemática, como sugestões finais, poderão ser apresentadas algumas propostas que visem reverter o problema e conseqüentemente melhorar o desempenho dos alunos nas aulas de matemática.

As sugestões e contribuições a serem feitas no final são um tanto contribuidor para o melhoramento do processo de ensino e aprendizagem, pois como objetivo de ensino é, de acordo com Rodrigues (2021) “a construção do sujeito apto para agir de acordo com os preceitos estabelecidos como aceitáveis dentro de um contexto social, sujeito com conhecimento”. Com isto, na forma em que o nível de desempenho dos alunos se encontra, é notável que estes objetivos de ensino não serão alcançados pois os fracos resultados observados são proporcionais ao fraco desempenho e dedicação.

1.2 Problematização

No período de estágio pedagógico, o autor observou, com base nas pautas dos trimestres anteriores um percurso não muito convincente dos resultados dos alunos, são resultados cujas médias trimestrais variam no intervalo de seis a oito valores para um número muito significativo de alunos e nenhum deles possuía uma média trimestral correspondente a catorze valores. Na interação com estes mesmos alunos, pouco diziam acerca das reais causas desses resultados.

“Quando nos referimos a fatores de contexto como condicionantes do rendimento acadêmico é possível elencar alguns estudos empíricos onde nesses estudos, fatores como status socioeconômico, etnia/cor, sexo, nível econômico, escolaridades dos pais, infraestrutura e as

variáveis relacionadas a instituição podem ser condicionantes no desempenho escolar” (ALMEIDA, SOUSA, BRAGA e JÚNIOR, 2021, p. 97).

Portanto, olhando para essa abordagem e fazer uma relação daquilo que é a realidade dos alunos da escola em estudo, podemos reunir os seguintes fatores como sendo os que influenciam no rendimento académico dos alunos da escola secundária heróis moçambicanos de moatize são: a situação académica do aluno que compreende a aprovação ou reprovação no trimestre anterior, o desempenho académico nas disciplinas de português, física e química, a classificação das relações no ambiente escolar, que compreende a relação entre os alunos e os professores, a relação entre os alunos onde destaca – se o tempo de estudo em grupos, a relação entre os alunos e a direção da escola, a situação social do aluno que agrupa fatores inerentes a renda dos seus encarregados, a distancia da escola e casa, o tempo em que este percorre ao sair de casa ate a escola e o grau máximo de escolaridade dos seus encarregados.

Estes fatores foram levantados na fundamentação de que as disciplinas que envolvem cálculos estão relacionadas, ou seja, na maioria das vezes quando se tratam de estudos em grupo, os alunos focam se muito mais em rever essas disciplinas. A diferente relação que o aluno tem com o meio escolar, professores e seus colegas foi dado como fator influenciador, na medida em que viu - se que este também é combustível impulsionador do rendimento académico pois a dedicação deste mesmo aluno parte na medida em que este se sente mais a vontade no recinto escolar e das interações que este tem com os professores na medida em que consegue expor duvidas diversas sobre as aulas.

Por último, a situação social do aluno também é um fator contribuinte para a motivação do mesmo aluno em focar – se nas aulas, constatou – se que muito deles não conseguem separar a escola com os problemas de casa, um aluno mal-humorado porque não almoçou ou porque caminhou uma longa distancia para chegar a escola não estará devidamente concentrado nas aulas.

Com base nas informações apresentadas anteriormente, levanta – se a questão: **quais são as variáveis que mais impactam no rendimento académico na disciplina de matemática dos alunos da 10ª classe da escola secundária heróis moçambicanos de moatize?**

1.3 Objetivos

1.3.1 Objetivos Geral

- Compreender os fatores que Influenciam no Rendimento Académico da disciplina de Matemática dos alunos da 10ª classe da Escola Secundaria Heróis Moçambicanos de Moatize em 2021;

1.3.2 Objetivos Específicos

- Identificar os fatores que influenciam no rendimento académico da disciplina de matemática dos alunos da 10ª classe da escola secundária heróis moçambicanos de moatize;
- Descrever a caracterização dos alunos da 10ª classe da escola secundária heróis moçambicanos de moatize;
- Obter um modelo de regressão linear múltipla que possa explicar os fatores que mais influenciam as notas dos alunos da 10ª classe da escola secundaria heróis moçambicanos de moatize;
- Mostrar qual a variável que tem maior impacto no rendimento académico dos alunos da 10ª classe da ESHMM;

1.4 Hipóteses

H₀: As disciplinas de português, química e física, a relação entre o aluno e o ambiente escolar e a situação social do aluno influenciam no rendimento académico da disciplina de matemática;

H₁: As disciplinas de português, química e física, a relação entre o aluno e o ambiente escolar e a situação social do aluno não influenciam no rendimento académico da disciplina de matemática;

CAPITULO II

2.0 Metodologia de Pesquisa

É essencial entendermos primeiramente o conceito de pesquisa. Para Gil (2007, p. 17) citado por Gerhardt e Silveira (2009, p. 12) a pesquisa é o

“procedimento racional e sistemático que tem como objetivo proporcionar respostas aos problemas que são propostos. A pesquisa desenvolve-se por um processo constituído de várias fases, desde a formulação do problema até a apresentação e discussão dos resultados.”

Com isto, podemos descrever os tipos de pesquisa fazendo assim uma explicação clara quanto a abordagem, quanto a natureza, quanto aos objetivos e quanto aos procedimentos.

No que concerne a metodologia de pesquisa, ela pode ser definida de acordo com a perspectiva de Fonseca (2002) citado por Gerhardt e Silveira (2009, p. 12) que a concebem como sendo

“o estudo da organização, dos caminhos a serem percorridos, para se realizar uma pesquisa ou um estudo, ou para se fazer ciência. Etimologicamente, significa o estudo dos caminhos, dos instrumentos utilizados para fazer uma pesquisa científica.”

Portanto, dentro da metodologia encontram-se os métodos a serem levados a cabo pela pesquisa, Gerhardt e Silveira (2009, p. 13) faz um breve esclarecimento da distinção entre metodologia e método “A metodologia se interessa pela validade do caminho escolhido para se chegar ao fim proposto pela pesquisa; portanto, não deve ser confundida com o conteúdo (teoria) nem com os procedimentos (métodos e técnicas).”

2.1 Tipo de Pesquisa

2.1.1 Quanto a abordagem

Com base nas descrições feitas no tópico anterior, quanto a abordagem, essa pesquisa é de carácter quantitativa pois esta pesquisa visa responder um problema onde viu-se ser necessário trabalhar com um certo público-alvo, público este que é de tamanho muito alargado, daí a necessidade de se definir uma amostra, que de acordo com Fonseca (2002, p. 20) citado por Gerhardt e Silveira (2009, p. 33) é um tipo de pesquisa onde o público-alvo é de tamanho muito maior, face a isto se define uma amostra desse público com o intuito de compreender a realidade com base na análise de dados brutos, recolhidos com o auxílio de instrumentos padronizados e

neutros. E tratando se uma pesquisa cujo modelo de estudo escolheu se uma ferramenta estatística que é a regressão linear, surge a necessidade de esta ser de caracter quantitativa pois trata se de uma ferramenta matemática, “A pesquisa quantitativa recorre à linguagem matemática para descrever as causas de um fenômeno, as relações entre variáveis, etc.” (FONSECA (2002, p. 20) citado por GERHARDT e SILVEIRA (2009, p. 33)).

2.1.2 Quanto a natureza

Quanto a natureza, esta pesquisa é de caracter aplicada, onde Gerhardt e Silveira (2009, p. 35) entendem – a como aquela que “Objetiva gerar conhecimentos para aplicação prática, dirigidos à solução de problemas específicos. Envolve verdades e interesses locais.” Desta feita, com esta pesquisa o autor pretende gerar alguma solução prática do problema referente ao desempenho abaixo do nível dos alunos observados pois para o bem do processo de ensino e aprendizagem, este é um problema que precisa de uma solução urgente.

2.1.3 Quanto aos objetivos

Quanto aos objetivos, esta pesquisa é de caracter explicativa, “Este tipo de pesquisa preocupa – se em identificar os fatores que contribuem para a ocorrência dos fenômenos” (GIL (2007) citado por GERHARDT e SILVEIRA (2009, p. 35)). Com isto, a nossa pesquisa tem como enfoque identificar os fatores que influenciam no rendimento académico da disciplina de matemática dos alunos da 10ª classe da escola secundária heróis moçambicanos de moatize.

2.1.4 Quanto aos procedimentos

Quanto aos procedimentos, esta pesquisa é de levantamento mas concretamente levantamento estatístico, pois o autor pretende fazer o levantamento de uma amostra como descreve Fonseca (2002) citado por Gerhardt e Silevira (2009, p. 38) no que diz respeito aos tipos de pesquisa de levantamento. Esta pesquisa vem a ser de levantamento estatístico na medida em que pretende se adquirir instruções essenciais para a realização da pesquisa, no que diz respeito principalmente a obtenção, organização e processamento e por conseguinte a validação dos dados.

2.2 Técnica e Instrumentos de Coleta de Dados

A coleta de dados foi com base em um questionário que de acordo com Gerhardt e Silveira (2009, p. 69) pode ser definido como sendo “um instrumento de coleta de dados constituído por uma série ordenada de perguntas que devem ser respondidas por escrito pelo informante (...)”. A escolha do questionário como instrumento de coleta de dados surge na medida em que esta pesquisa é de levantamento estatístico e o questionário é uma ferramenta muito prática quando se trata dessa modalidade de pesquisa.

O questionário era de carácter misto, pois possuía cerca de vinte perguntas onde seis eram perguntas abertas de respostas curtas e catorze perguntas fechadas de carácter múltipla escolha das quais duas eram fechadas da modalidade “Sim/Não”. As perguntas do instrumento estavam dífidas em quatro partes nomeadamente: Dados pessoais do aluno, situação académica do aluno, classificação das relações no ambiente escolar e por último a situação social do aluno.

As perguntas do questionário foram desenhadas com base nas principais, motivações que podem levar o aluno a perder o foco nas aulas em geral tal como foi referenciado por Almeida, Sousa, Braga e Júnior (2021, p. 97) onde descrevem as possíveis situações de naturezas variadas onde destaca se principalmente a social e do ambiente escolar como sendo as principais que podem levar o aluno a aplicar esforço abaixo do nível nos estudos.

As respostas do questionário foram obtidas com auxílio da plataforma online de inquéritos (Google forms), o preenchimento das informações foi feito pelo autor na presença de um aluno onde este respondia as perguntas que o autor apresentava, estima – se que em media cada questionário levou cerca de 10 minutos a ser preenchido na sua totalidade.

2.3 População e Amostra

De acordo com Bussab e Moretti (2010, p. 262) “População é o conjunto de todos os elementos ou resultados sob investigação. Amostra é qualquer subconjunto da população.” Desta feita, a população em estudo na presente pesquisa são todos alunos que estão a frequentar a 10ª classe na escola secundária heróis moçambicanos de moatize. O universo é composto por 452 alunos distribuídos da seguinte maneira:

Turmas	Subturmas	Homens	Mulheres	Total/Turma
	A1	10	15	75
	A2	6	19	

A	A3	14	11	
B	B1	10	14	75
	B2	9	16	
	B3	10	16	
C	C1	8	18	75
	C2	11	13	
	C3	10	15	
D	D1	15	10	75
	D2	16	9	
	D3	14	11	
E	E1	12	14	77
	E2	13	13	
	E3	13	12	
F	F1	13	11	75
	F2	11	14	
	F3	20	6	

Tabela 1: Distribuição do número de alunos da 10ª classe em turmas (Fonte: ESHM)

Portanto, na impossibilidade de se inquerir a todos elementos da população devido a diversos fatores, aplicou-se a amostragem não probabilística por conveniência que por sua vez, de acordo com Oliveira (2001) neste tipo de amostragem, o pesquisador seleciona membros da população mais acessíveis, ou seja, aqueles que estão disponíveis para participar na entrevista no momento exato em que o pesquisador está recolhendo os dados. Dada a situação pandémica no país, as escolas não funcionam como anteriormente, adotando um regime de turnos, por isso, viu-se ser necessário aplicar este tipo de amostragem com vista a fazer as entrevistas de uma forma barata e segura assim como explica Oliveira (2001), “A amostra por conveniência é empregada quando se deseja obter informações de maneira rápida e barata.” Portanto, neste tipo de amostragem foi possível reunir uma amostra com dimensão igual a 421 alunos equivalente a 93% da população total.

2.4 Técnica de Tratamento de dados

Neste tópico vamos descrever as técnicas e procedimentos estatísticos levados a cabo no tratamento dos dados assim como descreve Gerhardt e Silveira (2009, p. 80) as técnicas de

análise de dados consistem na “Descrição dos procedimentos adotados para a análise dos dados (quantitativos – análise estatística, tipos de testes estatísticos escolhidos, etc.; qualitativos – análise de conteúdo, análise de discurso, etc.)” Portanto, neste trabalho o tratamento dos dados será com base no software estatístico IBM SPSS Versão 26 que está disponível para “download” grátis na internet.

Os procedimentos escolhidos para a análise e tratamento dos dados incluem a análise exploratória mas concretamente á: verificação de dados atípicos, verificação de erros de lançamento dos dados, verificação de valores ausentes e seu tratamento. Para as variáveis quantitativas proceder – se – a a análise descritiva, análise da correlação de Pearson com vista a avaliar a relação que as variáveis explicativas tem como a variável dependente e aplicação de testes de normalidade. E para as variáveis qualitativas/categóricas vai proceder – se a análise de frequências concretamente no eixo relacionado á dados pessoais. Por último, vai se proceder a regressão linear múltipla e especificação das variáveis de saída assim como as variáveis explicativas.

2.4.1 Descrição das técnicas

2.4.1.1 Verificação de erros de lançamento dos dados

Esta técnica de tratamento dos dados é muito importante para o tratamento dos nossos dados, devido ao erro humano que na perspectiva de Zambon (2010) pode ser definido como sendo “o ato ou efeito de errar, ou o juízo ou julgamento em desacordo com a realidade observada, ou a qualidade daquilo que é inexato, incorreto, ou ainda o desvio do caminho considerado correto, bom, apropriado; um desregramento.”

Portanto, este tipo de imprevisto pode acontecer com qualquer um, e para garantir uma exatidão nos dados lançados na nossa base, esta passará pelo processo de verificação de erros que podem ser troca de valores numéricos por letras ou vice – versa e ainda identificação de valores nos dados que podem ser inconsistentes com os nossos objetivos para uma correção imediata.

2.4.1.2 Verificação de valores ausentes e tratamento

Esta também é uma técnica indispensável no tratamento dos nossos dados, este processo de verificação de valores ausentes permite nos identificar casos em que haja valores

ausentes/omissos em determinada variável, feita a identificação, é possível codificá-los e tratá-los à parte isto com o auxílio do programa escolhido para o processamento dos dados.

2.4.1.3 Análise exploratória

A análise exploratória dos dados vai consistir numa organização geral de modo a resumir o conjunto dos dados, a primeira etapa é organização das variáveis e sua devida classificação, em seguida faremos a apresentação e organização dos dados pois este é procedimento muito importante como descrevem Bussab e Morettin (2006) “A organização dos dados coletados é fundamental para que não hajam erros de processamento e perda de informações. Deve ser feito em um programa apropriado.” Portanto, com o auxílio do SPSS faremos a apresentação dos dados em tabelas e gráficos de acordo com suas variáveis.

Essa análise exploratória dos dados será mista, observado previamente o(s) tipo(s) de variável(s) a ser(em) apresentado(as), poderá optar – se a univariada com gráficos ou sem gráficos e a exploração multivariada com gráficos ou também sem gráficos.

2.4.1.4 Verificação de atípicos

Entender os valores atípicos também chamados de “outliers” é muito crucial para essa pesquisa, na maioria das vezes são esses valores que influenciam de forma negativa nos resultados das pesquisas como defendem Hoppen e Prates (2017):

“Os outliers são dados que se diferenciam drasticamente de todos os outros. Em outras palavras, um outlier é um valor que foge da normalidade e que pode (e provavelmente irá) causar anomalias nos resultados obtidos por meio de algoritmos e sistemas de análise.”

Portanto, para evitarmos situações do gênero, vamos entender e identificar os atípicos nos nossos dados com base em gráficos pois o método tabular não é eficaz para a nossa base devido o elevado número de observações, assim como Hoppen e Prates (2017) defendem:

“(…) quando a quantidade de observações (linhas) passa para a casa dos milhares ou milhões fica impossível de encontrar quais são os dados que destoam do geral. Portanto, a tarefa de identificar anomalias pode se tornar difícil, principalmente quando mais variáveis (mais as colunas) são envolvidas.”

Usando o método gráfico para a identificação dos atípicos, poderá ser fácil o processo da sua identificação, pois no gráfico, um valor atípico estará muito isolado, ou separado dos outros.

Este fato é observado muitas das vezes em gráficos de dispersão ou mesmo histograma, com este gráfico poderemos observar os valores que merecem a nossa atenção.

2.4.1.5 Análise descritiva

A análise descritiva é uma das fases mais importantes do processo de tratamento de dados, é nesta fase onde ocorre maior parte do procedimento estatístico de processamento dos dados, a análise descritiva:

“é a fase inicial deste processo de estudo dos dados coletados. Utilizamos métodos de Estatística Descritiva para organizar, resumir e descrever os aspectos importantes de um conjunto de características observadas ou comparar tais características entre dois ou mais conjuntos.” REIS e REIS (2002, p. 5)

Desta feita, na análise descritiva dos dados vamos tratar e interpretar principalmente: o estudo de análise frequência e distribuição de frequências para variáveis categóricas e quantitativas respectivamente e suas respectivas representações gráficas, determinação de medidas de tendência central, de variabilidade e de posição para variáveis quantitativas e ainda construção de Boxplot (diagrama de caixa) para certas variáveis.

2.4.1.6 Aplicação de testes de Normalidade

Dando uma pequena espreitada no contexto relacionado a regressão, o teste de normalidade é um processo e uma técnica muito importante, este teste permite - nos, na perspectiva de Fundão (2018, p. 52) “obter a distribuição dos estimadores de Mínimos Quadrados, utilizada na estimação dos intervalos de confiança e realização dos testes de hipóteses sobre os parâmetros do modelo.” Existem vários métodos para os testes de normalidade, mas neste trabalho optaremos pelo teste de Kolmogorov-Smirnov (K-S).

“é um teste de aderência que verifica o grau de concordância entre distribuições, considerando um conjunto de valores. Tem como objetivo identificar se os dados seguem uma determinada distribuição. Este teste utiliza a distribuição de frequência acumulada, que ocorre dada a descrição teórica, e compara-a com a distribuição de frequência acumulada observada.” FUNDÃO (2018, p. 54)

A escolha deste teste pelo autor deve – se ao fato de este teste não paramétrico ser de simples e eficaz aplicação, pois este pode ser facilmente no SPSS.

2.4.2 Tabela das Variáveis

Ordem	Variável	Tipo de Variável
1	Idade	Quantitativa discreta
2	Sexo	Qualitativa ordinal
3	Estado Civil	Qualitativa ordinal
4	Resultado obtido no 2º trimestre de 2021?	Quantitativa discreta
5	Média final de Matemática obtida no 2º trimestre de 2021	Quantitativa discreta
6	Média final de Português obtida no 2º trimestre de 2021	Quantitativa discreta
7	Média final de Física obtida no 2º trimestre de 2021	Quantitativa discreta
8	Média final de Química obtida no 2º trimestre de 2021	Quantitativa discreta
9	Relação entre alunos e professores	Qualitativa ordinal
10	Relação entre alunos (Estudos em grupo)	Qualitativa ordinal
11	Relação entre alunos e a direção da escola	Qualitativa ordinal
12	Que tipo de casa vive?	Qualitativa ordinal
13	Seus Pais trabalham?	Qualitativa ordinal
14	Quantas Refeições tem por dia?	Quantitativa discreta
15	Qual a distância da sua casa ate escola?	Quantitativa contínua
16	Vem a escola de que meio?	Qualitativa ordinal
17	Quanto tempo leva ao sair de casa ate a escola?	Quantitativa contínua
18	Teus Pais tem dificuldades na compra de material escolar?	Qualitativa ordinal
19	Quantas pessoas vivem em sua casa?	Quantitativa discreta
20	Qual é o grau máximo de escolaridade dos seus pais?	Qualitativa ordinal

Tabela 2: Tabela das Variáveis (Fonte: Autor)

CAPITULO III

3.0 Fundamentação Teórica

3.0.1 Modelo de Regressão Linear Múltipla (MRLM)

O modelo de regressão linear múltipla é usual quando se estuda fenómenos onde estão disponíveis mais de uma variável independente, e uma variável dependente, desta feita pretende-se estudar o efeito de todas variáveis independentes sobre a variável dependente, formalmente, o MRLM pode ser definido com base na abordagem de Fundão (2018, p.25) onde defende que “Modelo de Regressão Linear Múltipla é o modelo que descreve uma relação linear entre um conjunto de variáveis independentes (variáveis preditores), X_j ; $j = 1; 2; \dots; k$; e uma variável dependente (variável explicada ou resposta) Y .” o modelo de regressão linear múltipla é representado pela seguinte equação:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_k X_{ki} + \epsilon_i, i = 1, 2, 3, \dots, n$$

Onde:

- Y_i – representa a i – ésima observação da variável dependente;
- X_{i1}, \dots, X_{ik} – representa as i – ésimas observações das variáveis independentes;
- $\beta_0, \beta_1, \dots, \beta_k$ – representa os coeficientes ou parâmetros do modelo;
- ϵ_i - variável aleatória residual (erro) que descreve os efeitos em Y_i , não explicadas pelas variáveis independentes;
- n – é o numero de observações;

Por outro lado, Castigo (2021, p. 18) apresenta a equação da regressão linear múltipla mais simplificada que é: $Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n$, onde Y é a variável dependente, X_k é a variável explicativa e n é o numero de observações da amostra.

3.0.2 Modelo com Iteração

O modelo de iteração é um caso particular a regressão linear múltipla em que existem caso onde pode ocorrer que uma variável seja o quadrado da outra, ou ainda produto de duas variáveis, Rodrigues (2012, p. 24) apresenta um exemplo de um modelo de duas variáveis explicativas

onde existe interação entre elas, é apresentada nos seguintes termos: $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2 + \varepsilon$, onde o produto $X_1 X_2$ é que representa a interação.

3.0.1.1 Pressupostos do modelo de regressão linear múltipla

Os pressupostos do modelo de regressão linear múltipla são dados como condições indispensáveis para o uso do modelo, antes de entrar nesse tópico é necessário antes definir o conceito de resíduos, portanto, na perspectiva de Rodrigues (2012, p.33) “resíduos são dados pela diferença entre os valores da variável resposta observada e a variável resposta estimada, isto é, $\varepsilon_i = y_i - \hat{y}_i$, $i = 1, 2, \dots, n$ ”

OS pressupostos são os que determinam o uso correto do modelo, Fundação (2018, p. 27) apresenta os seguintes pressupostos do modelo de regressão linear múltipla:

- Os resíduos devem ter uma distribuição normal com media zero e variância constante ou seja, $\varepsilon_i \sim N(0, \sigma^2)$, $i = 1, 2, \dots, n$;
- Os resíduos devem ser homoscedásticos, o que significa que a sua variância deve ser constante, $\text{var}(\varepsilon_i) = \sigma^2 > 0$, $i = 1, 2, \dots, n$;
- Os resíduos das observações devem ser mutuamente independentes;
- Não devem existir situações de colinearidade e/ou multicolinearidade entre as variáveis independentes do modelo;

3.0.1.2 Validação dos pressupostos e análise dos resíduos

Existem técnicas concebidas exclusivamente para a verificação dos pressupostos apresentados no ponto anterior, com base na abordagem de Rodrigues (2012, p. 33 – p.41) vamos apresentar aqui essas técnicas para cada pressuposto do modelo.

3.0.1.2.1 Diagnóstico da Normalidade

A normalidade dos resíduos pode ser analisada quer através de gráficos, quer usando alguns testes, nomeadamente:

- Gráfico P-P plot dos resíduos;
- Teste de Kolmogorov-Smirnov;
- Teste de Shapiro Wilk;

3.0.1.2.1.1 Gráfico P-P plot dos resíduos

No gráfico P-P, vamos visualizar a distribuição de probabilidades dos valores observados com os valores esperados, representada por uma diagonal, segundo uma distribuição normal. Caso a normalidade se verifique, as observações registadas aproximam-se dessa diagonal, sem nenhum afastamento significativo.

A figura 1 ilustra o gráfico P-P dos resíduos, nesta situação a normalidade é verificada já que os pontos se aproximam da recta.

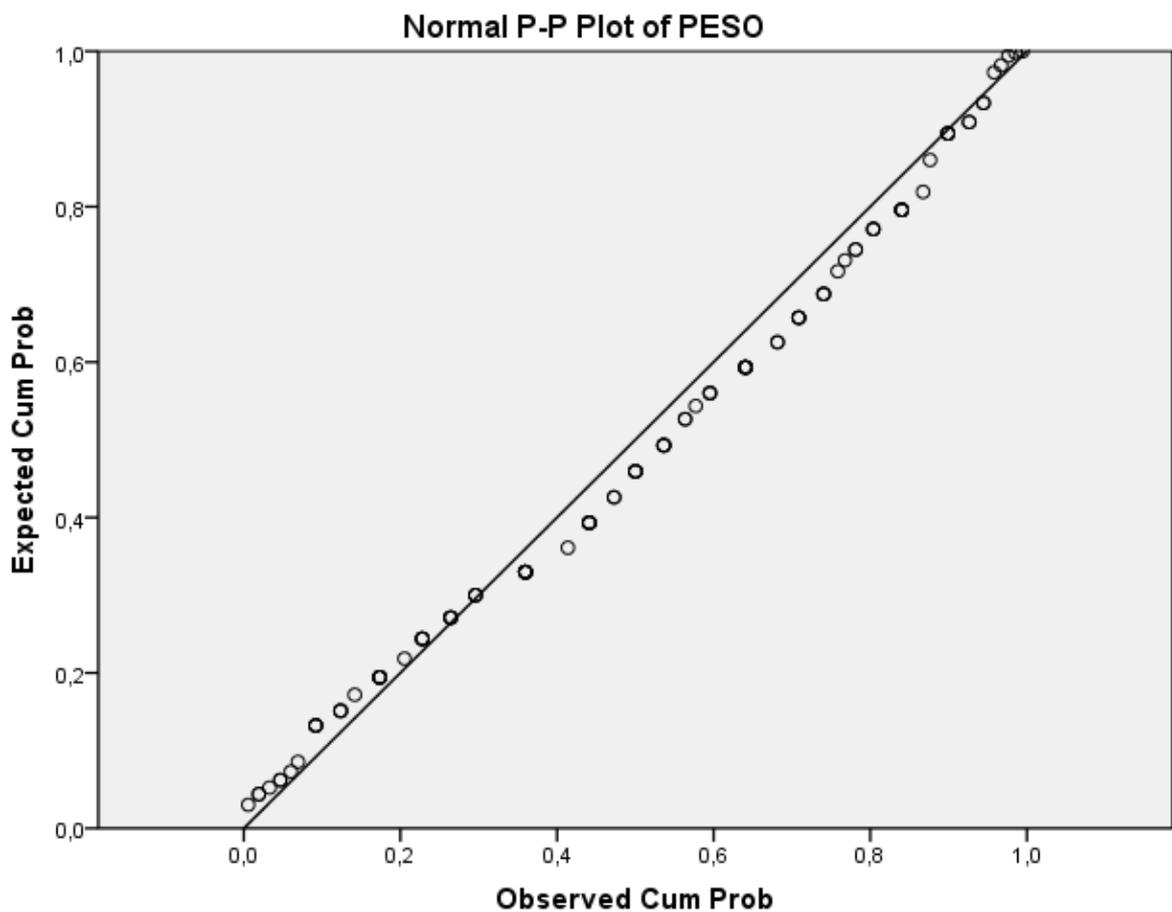


Figura 1: Gráfico P-P dos resíduos (Fonte: Rodrigues(2012))

3.0.1.2.1.2 Teste de Kolmogorov-Smirnov

O teste de K-S é usual para a testagem das seguintes hipóteses:

$$\begin{cases} H_0: \text{A distribuição é normal} \\ H_1: \text{A distribuição não é normal} \end{cases}$$

De acordo com Maroco (2003) citado por Rodrigues (2012, p. 34) a estatística do teste K-S é dada por: $D = \max\{\max(|F(x_i) - F_0(x_i)|); \max(|F(x_i - 1) - F_0(x_i)|)\}$.

Em que $F(x_i) - F_0(x_i)$ representa a diferença entre a frequência acumulada de cada uma das observações e a frequência acumulada que essa observação teria, sendo a sua distribuição normal.

Este teste observa a máxima diferença absoluta entre a função de distribuição acumulada assumida pelos dados, neste caso da distribuição normal, e a função de distribuição empírica dos dados.

3.0.1.2.1.3 Teste de Shapiro Wilk

Este teste sugere-nos preferência em relação ao teste de K-S para amostras de pequenas dimensões ($n < 30$). Neste caso, as hipóteses a serem testadas são as definidas anteriormente para o teste de K-S.

A estatística de teste é definida da seguinte forma: $W = \frac{(\sum_{i=1}^n a_i x_i)^2}{\sum_{i=1}^n (x_i - \bar{x})^2}$, onde a_i são constantes gerados a partir da média, variância e covariância de n ordens.

3.0.1.2.2 Diagnóstico De Homoscedasticidade (Variância Constante)

A homoscedasticidade é a condição de que os resíduos devem ter variância constante, A variância ser constante equivale a supor que não existem observações incluídas na variável residual cuja influência seja mais intensa na variável dependente.

Uma das técnicas usadas para verificar a suposição de que os resíduos são homoscedásticos, é a análise do gráfico dos resíduos versus valores ajustados. Este gráfico deve apresentar pontos dispostos aleatoriamente sem nenhum padrão definido, como se pode ver, por exemplo na Figura 1.2.

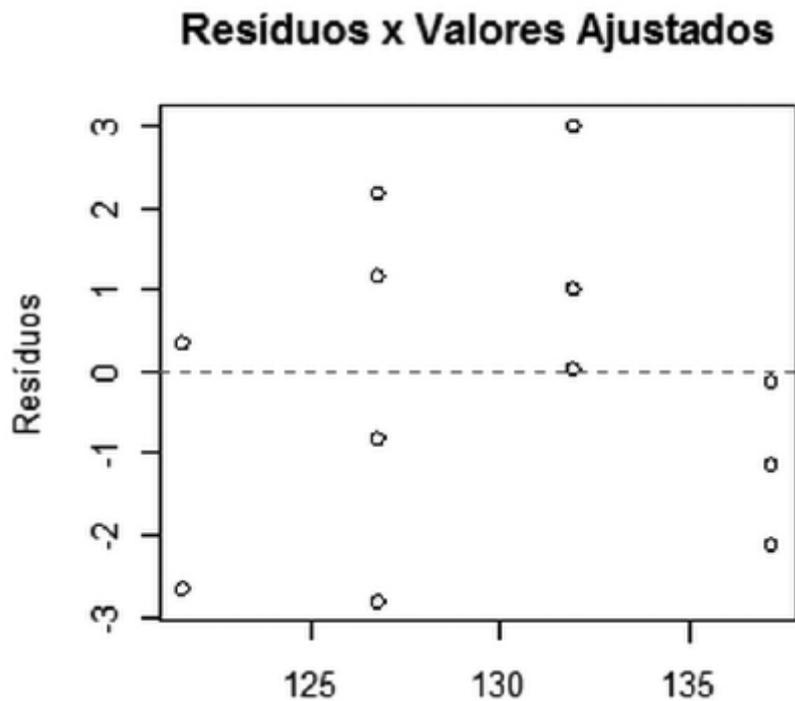


Figura 2: Confirmação da homoscedasticidade dos resíduos (Fonte: Rodrigues (2012))

Os mesmos autores explicam que se os pontos estão aleatoriamente distribuídos em torno da recta $y = 0$, sem nenhum comportamento ou tendência, temos indícios de que a variância dos resíduos é constante. Já a presença, por exemplo, de “funil” é um indicativo da presença de heteroscedasticidade.

3.0.1.2.3 Diagnóstico de independência

Para testar o pressuposto da independência dos resíduos, ou a presença de autocorrelação entre eles, pode utilizar-se o teste de Durbin-Watson (DW). O teste de Durbin-Watson testa as seguintes hipóteses:

$$\begin{cases} H_0: \text{Nao existe autocorrelacao dos residuos} \\ H_1: \text{Existe autocorrelacao positiva dos residuos} \end{cases}$$

A estatística de teste de DW é dada por: $dw = \frac{\sum_{i=2}^n (e_i - e_{i-1})^2}{\sum_{i=1}^n e_i^2}$, onde $0 \leq dw \leq 4$.

Portanto, esta estatística mede a correlação entre cada resíduo e o resíduo correspondente à observação imediatamente anterior. Podemos tomar a decisão comparando o valor de dw com os valores críticos d_L e d_U da tabela de Durbin-Watson a seguir:

		para $\alpha = 0.05$													
n \ p	1		2		3		4		5		10		15		
	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U	
6	0.61	1.40													
10	0.88	1.32	0.70	1.64	0.53	2.02	0.38	2.41	0.24	2.82					
20	1.20	1.41	1.10	1.54	1.00	1.68	0.89	1.83	0.79	1.99	0.34	2.89	0.06	3.68	
30	1.35	1.49	1.28	1.57	1.21	1.65	1.14	1.74	1.07	1.83	0.71	2.36	0.39	2.94	
40	1.44	1.54	1.39	1.60	1.34	1.66	1.29	1.72	1.23	1.79	0.95	2.15	0.68	2.56	
50	1.50	1.59	1.46	1.63	1.42	1.67	1.38	1.72	1.34	1.77	1.11	2.05	0.88	2.35	
60	1.55	1.62	1.51	1.65	1.48	1.69	1.44	1.73	1.41	1.77	1.22	1.98	1.03	2.28	
70	1.58	1.64	1.55	1.67	1.53	1.70	1.49	1.74	1.46	1.77	1.31	1.95	1.14	2.15	
80	1.61	1.66	1.59	1.69	1.56	1.72	1.53	1.74	1.51	1.77	1.37	1.93	1.22	2.09	
90	1.64	1.68	1.61	1.70	1.59	1.73	1.57	1.75	1.54	1.78	1.42	1.91	1.29	2.06	
100	1.65	1.69	1.63	1.72	1.61	1.74	1.59	1.76	1.57	1.78	1.46	1.90	1.35	2.03	
200	1.76	1.78	1.75	1.79	1.74	1.8	1.73	1.81	1.72	1.82	1.67	1.87	1.61	1.93	

Tabela 3: Tabela de valores críticos de d_L e d_U . (Fonte: Rodrigues (2012))

A tabela seguinte dá-nos as decisões a tomar em função dos valores críticos, d_U e d_L .

Zona de Rejeição e de não-rejeição de H_0					
dw	$[0; d_L[$	$[d_L; d_U[$	$[d_U; 4 - d_U[$	$[4 - d_U; 4 - d_L[$	$[4 - d_L; 4]$
Decisão	Rejeitar H_0	Nada se pode concluir	Não rejeitar H_0	Nada se pode concluir	Rejeitar H_0
	Auto-correlação positiva		Os resíduos são independentes		Auto-correlação negativa

Tabela 4: Tabela de decisão em função de d_L e d_U . (Fonte: Rodrigues (2012))

3.0.1.2.4 Colinearidade e Multicolinearidade

Os termos colinearidade e multicolinearidade podem ser definidos segundo Rodrigues (2012, p.39) onde detalha nos seguintes termos:

“O termo colinearidade é utilizado para expressar a existência de correlação elevada entre duas variáveis independentes, enquanto o termo multicolinearidade é utilizado quando se trata de mais do que duas variáveis independentes fortemente correlacionadas. No entanto, existem autores que definem colinearidade como a existência de relação linear entre duas variáveis independentes e multicolinearidade como a existência de relação linear entre uma das variáveis independentes e as restantes.”

Se considerarmos duas quaisquer variáveis independentes, X_1 e X_2 , entre as quais existe uma elevada correlação, a proporção da variação total da variável dependente, explicada por X_1 é idêntica à proporção da variação total da variável dependente, explicada por X_2 .

Quando uma das variáveis independentes já se encontra no modelo de regressão, a inclusão de outra variável independente, não implica, uma explicação adicional significativa da variação total da variável dependente.

A colinearidade pode ser diagnosticada da seguinte maneira:

- Verificando se a matriz de correlações das variáveis independentes demonstra correlações elevadas. Caso a correlação de duas variáveis seja muito próxima de 1, indica de facto um problema;
- Verificando se, ao se realizar a regressão de X_i em função das outras variáveis independentes, o valor de $R \cong 1$.

Um indicador usado com frequência para detectar a multicolinearidade é o Variance Inflation Factor (VIF).

Segundo Maroco (2003) citado por Rodrigues (2012, p.40), caso se obtenham valores de VIF > 5 conclui-se que estamos perante problemas com a estimação de β_i devido à presença de multicolinearidade nas variáveis independentes.

Em acréscimo, Rodrigues (2012, p. 41) acrescenta que Existem vários métodos que permitem, na regressão linear múltipla, fazer uma selecção das variáveis independentes que melhor explicam a variável resposta, nomeadamente:

- FORWARD – o método começa apenas com a constante e adiciona uma variável independente de cada vez. A primeira variável seleccionada é a que apresenta maior correlação com a variável resposta (maior score statistic);
- BACKWARD – o método faz o “contrário” do método Forward. Neste caso todas as variáveis independentes são incorporadas no modelo. Depois, por etapas, cada uma pode ser ou não eliminada.
- STEPWISE - o método Stepwise é uma “modificação” do método Forward que permite resolver problemas de multicolinearidade. Consiste no seguinte: fazemos entrar no modelo a variável explicativa que apresenta maior coeficiente de correlação com a variável dependente.

Em seguida, calculam-se os coeficientes de correlação parcial para todas as variáveis que não fazem parte da primeira equação de regressão, para que, a próxima variável a entrar, seja a que apresenta maior coeficiente de correlação parcial.

Estima-se a nova equação de regressão e analisa-se se uma das duas variáveis independentes deve ser excluída do modelo.

No final, se ambas as variáveis apresentarem valores t significativos, novos coeficientes de correlação parcial são calculados para as variáveis que não entraram.

Este processo finda, assim que se chegue à situação em que nenhuma variável deva ser acrescentada à equação.

3.0.1.1 Estimação dos parâmetros do modelo

Fundão (2018, p.28) explica que para se estimar os parâmetros do MRLM é necessário ter uma amostra aleatória, da qual se vão obter os estimadores não enviesados para a equação ajustada,

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_{i1} + \hat{\beta}_2 X_{i2} + \dots + \hat{\beta}_k X_{ik}.$$

Assim os parâmetros $(\beta_0, \beta_1, \beta_2, \dots, \beta_k)$ do MRLM, são estimados a partir de um conjunto de observações de uma amostra, recorrendo ao Método dos Mínimos Quadrados (MMQ).

3.0.1.1.1 Método dos Mínimos Quadrados

O método dos mínimos quadrados no MRLM permite obter os estimadores não enviesados dos parâmetros do modelo.

O estimador dos mínimos quadrados é aquele que minimiza a soma dos quadrados dos resíduos.

Neste caso se tem: $SQE = \sum_{i=1}^n \varepsilon_i^2 = \sum_{i=1}^n |y_i - \hat{y}_i|^2 = \sum_{i=1}^n [y_i - (\hat{\beta}_0 + \hat{\beta}_1 X_{i1} + \hat{\beta}_2 X_{i2} + \dots + \hat{\beta}_k X_{ik})]^2$.

Os valores de $\widehat{\beta}_0, \widehat{\beta}_1, \widehat{\beta}_2, \dots, \beta_k$, que minimizam este resultado, são aqueles que anulam as suas

$$\text{derivadas parciais em ordem aos respetivos estimadores} \begin{cases} \frac{\partial SQE}{\partial \widehat{\beta}_0} = 0 \\ \frac{\partial SQE}{\partial \widehat{\beta}_1} = 0 \\ \vdots \\ \frac{\partial SQE}{\partial \widehat{\beta}_k} = 0 \end{cases}, \text{ assim tem-se:}$$

$$\frac{\partial SQE}{\partial \widehat{\beta}_0} = \sum_{i=1}^n 2 \times (y_i - \widehat{\beta}_0 - \widehat{\beta}_1 X_{i1} - \widehat{\beta}_2 X_{i2} - \dots - \widehat{\beta}_k X_{ik}) \times (-1) = -2 \sum_{i=1}^n (y_i - \widehat{\beta}_0 - \widehat{\beta}_1 - \widehat{\beta}_2 X_{i2} - \dots - \widehat{\beta}_k X_{ik})$$

$$\frac{\partial SQE}{\partial \widehat{\beta}_1} = -2 \sum_{i=1}^n x_{i1} \times (y_i - \widehat{\beta}_0 - \widehat{\beta}_1 X_{i1} - \widehat{\beta}_2 X_{i2} - \dots - \widehat{\beta}_k X_{ik})$$

(...)

$$\frac{\partial SQE}{\partial \widehat{\beta}_k} = -2 \sum_{i=1}^n x_{ik} \times (y_i - \widehat{\beta}_0 + \widehat{\beta}_1 X_{i1} + \widehat{\beta}_2 X_{i2} + \dots + \widehat{\beta}_k X_{ik}), \text{ desse resultado deduz-se o seguinte sistema de equações:}$$

$$\begin{cases} -2 \sum_{i=1}^n (y_i - \widehat{\beta}_0 - \widehat{\beta}_1 X_{i1} - \widehat{\beta}_2 X_{i2} - \dots - \widehat{\beta}_k X_{ik}) = 0 \\ -2 \sum_{i=1}^n x_{i1} \times (y_i - \widehat{\beta}_0 - \widehat{\beta}_1 X_{i1} - \widehat{\beta}_2 X_{i2} - \dots - \widehat{\beta}_k X_{ik}) = 0 \\ \vdots \\ -2 \sum_{i=1}^n x_{ik} \times (y_i - \widehat{\beta}_0 - \widehat{\beta}_1 X_{i1} - \widehat{\beta}_2 X_{i2} - \dots - \widehat{\beta}_k X_{ik}) = 0 \end{cases} \quad \text{De onde se obtêm as}$$

equações normais dos Mínimos Quadrados para o MRLM que são apresentadas a seguir:

$$\begin{cases} n\widehat{\beta}_0 + \widehat{\beta}_1 \sum_{i=1}^n x_{i1} + \dots + \widehat{\beta}_k \sum_{i=1}^n x_{ik} = \sum_{i=1}^n y_i \\ \widehat{\beta}_0 \sum_{i=1}^n x_{i1} + \widehat{\beta}_1 \sum_{i=1}^n x_{i1}^2 + \dots + \widehat{\beta}_k \sum_{i=1}^n x_{ik} x_{i1} = \sum_{i=1}^n x_{i1} y_i, \text{ isolando } \widehat{\beta}_0 \text{ da primeira} \\ \widehat{\beta}_0 \sum_{i=1}^n x_{ik} + \widehat{\beta}_1 \sum_{i=1}^n x_{ik} x_{i1} + \dots + \widehat{\beta}_k \sum_{i=1}^n x_{ik}^2 = \sum_{i=1}^n x_{ik} y_i \end{cases}$$

equação do sistema obtém-se: $\widehat{\beta}_0 = \frac{1}{n} \sum_{i=1}^n y_i - \widehat{\beta}_1 \frac{1}{n} \sum_{i=1}^n x_{i1} - \dots - \widehat{\beta}_k \frac{1}{n} \sum_{i=1}^n x_{ik}$, tendo em conta que $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$ e $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$, a igualdade de $\widehat{\beta}_0$ pode ser reescrita na seguinte forma: $\widehat{\beta}_0 = \bar{y} - \widehat{\beta}_1 \bar{x}_1 - \dots - \widehat{\beta}_k \bar{x}_k$.

Depois de se obter o resultado de $\widehat{\beta}_0$ Substituindo - o nas restantes equações obtém-se um sistema de equações como o seguinte:

$$\begin{cases} \widehat{\beta}_1 S_{x_1 x_1} + \widehat{\beta}_2 S_{x_1 x_2} + \cdots + \widehat{\beta}_k S_{x_1 x_k} = S_{x_1 y} \\ \widehat{\beta}_1 S_{x_2 x_1} + \widehat{\beta}_2 S_{x_2 x_2} + \cdots + \widehat{\beta}_k S_{x_2 x_k} = S_{x_2 y} \\ \vdots \\ \widehat{\beta}_1 S_{x_k x_1} + \widehat{\beta}_2 S_{x_k x_2} + \cdots + \widehat{\beta}_k S_{x_k x_k} = S_{x_k y} \end{cases}$$

Onde $S_{x_j x_j} = \sum_{i=1}^n (x_{ij} - \bar{x}_j)^2$, $j = 1, \dots, k$ e $S_{x_j y} = \sum_{i=1}^n (y_i - \bar{y})(x_{ij} - \bar{x}_j)$, $j = 1, \dots, k$

Substituindo os resultados das somas de quadrados e de produtos cruzados em cada uma das equações que compõem o sistema de equações e resolvendo-o, obtêm-se as equações que nos permitem calcular os estimadores $\widehat{\beta}_1, \widehat{\beta}_2, \dots, \widehat{\beta}_k$, dos parâmetros.

Os estimadores $\widehat{\beta}_1, \widehat{\beta}_2, \dots, \widehat{\beta}_k$, dão as melhores estimativas não enviesadas de $\beta_1, \beta_2, \dots, \beta_k$ e como consequência disso, surge a equação: $\widehat{y}_i = \widehat{\beta}_0 + \widehat{\beta}_1 X_{i1} + \widehat{\beta}_2 X_{i2} + \cdots + \widehat{\beta}_k X_{ik}$, $i=1, 2, \dots, n$, também dá a melhor estimativa linear não enviesada de \widehat{y}_i .

3.0.1.2 Propriedades dos estimadores do modelo

No modelo de regressão linear múltipla, os estimadores $\widehat{\beta}_0, \widehat{\beta}_1, \widehat{\beta}_2, \dots, \widehat{\beta}_k$, Obtidos através do método dos mínimos quadrados, segundo Fundão (2018, p. 34) devem satisfazer as seguintes propriedades:

- Devem ser funções lineares de y_i ;
- Devem ser estimadores centrados ou não enviesados de β como se demonstra a seguir. Dado que, $\widehat{\beta} = (X'X)^{-1}X'Y$ onde $Y = X\beta + \varepsilon$ e sabendo que $(X'X)^{-1}X'X = I_{k+1}$ onde I_{k+1} corresponde a matriz identidade de ordem $k+1$. Deste modo o valor esperado de $\widehat{\beta}$ é dado por:

$$\begin{aligned} E(\widehat{\beta}) &= E((X'X)^{-1}X'Y) \\ &= E((X'X)^{-1}X'(X\beta + \varepsilon)) \\ &= E((X'X)^{-1}X'X\beta + (X'X)^{-1}X'\varepsilon) \\ &= E((X'X)^{-1}X'X\beta) + E((X'X)^{-1}X'\varepsilon) \end{aligned}$$

$$= E(I_{K+1} \times \beta) + E((X'X)^{-1}X'\varepsilon)$$

$$= E(I_{K+1} \times \beta) + (X'X)^{-1}X' \times E(\varepsilon)$$

$= E(I_{K+1} \times \beta) + (X'X)^{-1}X' \times 0_n$ onde 0_n representa o vetor nulo com n componentes. Logo tem – se: $E(\hat{\beta}) = \beta$.

“Este resultado significa que, se fosse possível obter muitas observações particulares do vetor Y , para a mesma matriz X , se obteriam outras tantas estimativas de $\hat{\beta}$, que em média, tenderiam para o verdadeiro valor do vetor dos coeficientes β . Assim, pode se afirmar que o não enviesamento de $\hat{\beta}$ garante que este estimador é centrado.” (MURTELA, RIBEIRO, SILVA, PIMENTA E PIMENTA, 2015) citado por FUNDAO (2018, p. 35)

➤ Os estimadores devem possuir variância mínima, ou seja, deve ser eficientes.

A variância de $\hat{\beta}$ é dada por: $\text{Var}(\hat{\beta}) = \sigma^2(X'X)^{-1}$, tal como se demonstra a seguir:

A matriz da variância e covariância de dimensão $(k + 1) \times (K + 1)$, do vetor $\hat{\beta}$ é dada por:

$$\text{Var}(\hat{\beta}) = \begin{pmatrix} \text{Var}(\hat{\beta}_0) & \text{Cov}(\hat{\beta}_0\hat{\beta}_1) \dots & \text{Cov}(\hat{\beta}_0\hat{\beta}_k) \\ \text{Cov}(\hat{\beta}_1\hat{\beta}_0) & \text{Var}(\hat{\beta}_1) \dots & \text{Cov}(\hat{\beta}_1\hat{\beta}_k) \\ \text{Cov}(\hat{\beta}_k\hat{\beta}_0) & \text{Cov}(\hat{\beta}_k\hat{\beta}_1) \dots & \text{Var}(\hat{\beta}_k) \end{pmatrix}$$

Com base nos resultados desta matriz, e uma vez que $\hat{\beta} = \beta + (X'X)^{-1}X'\varepsilon \Rightarrow \hat{\beta} - \beta = (X'X)^{-1}X'\varepsilon$, tem – se:

$$\text{Var}(\hat{\beta}) = E [\hat{\beta} E(\hat{\beta})](\hat{\beta} - E(\hat{\beta}))]$$

$$= E[(\hat{\beta} - \beta)(\hat{\beta} - \beta)']$$

$$= E[((X'X)^{-1}X'\varepsilon)((X'X)^{-1}X'\varepsilon)']$$

$$= E[(X'X)^{-1}X'\varepsilon\varepsilon'X(X'X)^{-1}]$$

$$= \sigma^2 I_{k+1} (X'X)^{-1} \text{ dado que } E(\varepsilon\varepsilon') = \sigma^2 \text{ e } (X'X)^{-1}X'X = I_{k+1}, \text{ logo } \text{Var}(\hat{\beta}) = \sigma^2(X'X)^{-1}.$$

- Os estimadores devem ser consistentes, ou seja, à medida que o tamanho da amostra aumenta, os seus valores esperados devem convergir para os verdadeiros valores dos parâmetros e as suas variâncias devem tender para o valor nulo.

3.0.2 Fatores que influenciam no rendimento escolar

Apesar dos alunos estudarem e acompanharem o processo educativo de forma conjunta, a capacidade de assimilação dos conteúdos é variante de indivíduo para indivíduo, tal como descreve Freitas (2010) “(...) cada aluno apresenta suas dificuldades e habilidades sociais, técnicas, culturais e educacionais com relação à aprendizagem”. Desta feita percebe – se que cada aluno traz consigo sua bagagem que pode influenciar de forma negativa ou positiva no seu processo de aprendizagem.

Ainda na perspectiva de Freitas (2010), este assume que os fatores que podem vir influenciar no rendimento escolar são vários e de diversas naturezas, e para melhor compreensão e organização, ele agrupa – os em dois grupos nomeadamente: fatores positivos e fatores negativos.

Sobre os fatores positivos, este autor acredita que eles pode destacar - se o interesse do aluno pela aprendizagem, corpo docente capacitado e que saiba orientar, o que contribui para o ensino de qualidade, sobre o segundo fator ele explica que:

Trata-se, portanto, de dotar a instituição escolar de uma estrutura administrativa ágil, que favoreça o bom desempenho do trabalho coletivo e cooperativo, calcado em princípios democráticos que fortaleçam a condição de sujeito (autor) de todos os envolvidos, mas que, ao mesmo tempo (não alternativamente), procure preencher seus postos de trabalho com pessoas identificadas com esses princípios e empenhadas na realização de um ensino de qualidade. (PARO, 2007) citado por FREITAS (2010).

Inerente aos fatores negativos que podem influencia no rendimento escolar, ele sublinha as conversas excessivas em sala de aula, brincadeiras fora de hora e falta de interesse do aluno pela aprendizagem.

Para finalizar, Freitas (2010) ressalta o especto social como sendo um dos contribuintes para o desempenho escolar do aluno, aponta diretamente o papel da família no que diz respeito ao contributo quer esta exerce na formação do seu educado, em sua abordagem, este faz ainda uma

relação da importância da família e do ciclo escolar no desempenho do aluno, explicando nos seguintes termos:

A família é o primeiro grupo social em que o aluno começa a interagir, a aprender e onde busca as primeiras referências no que diz respeito aos valores culturais e emocionais. Assim como a família, a escola é responsável por fazer a mediação entre o indivíduo e a sociedade. Quando os pais são ausentes, ou quando o aluno tem um vínculo familiar ruim, ele pode apresentar autoestima prejudicada e distúrbios na aprendizagem. (PARO, 1997) citado por FREITAS (2010)

Por sua vez, Maia (2010) traz também uma reflexão profunda acerca dos fatores que influenciam no rendimento escolar, essa autora agrupa estes fatores em dois grupos nomeadamente fatores internos e fatores externos.

Maia (2010) começa a sua abordagem explicando acerca dos fatores externos ou seja, locais e acontecimentos que o aluno vivencia fora do ambiente escolar “Os fatores externos ao ambiente escolar têm grande relevância no rendimento dos alunos, pois se torna impossível dissociar as experiências vivenciadas fora da escola com a vida escolar.”, Aponta o meio como sendo um dos principais fatores externos que exercem maior influência no desempenho escolar do aluno.

Portanto, no que concerne ao meio social, esta autora aponta as desigualdades sociais (de entre elas, as más condições de vida e ausência de moradias adequadas) como sendo o fator principal influenciador do rendimento escolar do aluno, de uma forma generalizada, destaca “(...) todo o conjunto de privações com o qual convivem as classes sociais menos privilegiadas surge como o elemento explicativo fundamental para o baixo rendimento escolar.” MAIA (2010)

No que diz respeito aos fatores internos, eles estão relacionados a estrutura e funcionamento da escola, neste sentido, esta se referir ao conjunto de todos aspetos internos da escola onde ocorre toda prática educativa que por sua vez, segundo esta autora pode interferir positiva ou negativamente no rendimento dos alunos.

Desta feita, Maia (2010) destaca primeiramente o ambiente físico da escola como sendo um dos fatores, “O cuidado com o ambiente físico da escola também é de suma importância para o desenvolvimento da aprendizagem.” Nesse sentido, aponta particularmente a higiene e cuidados a ter com as salas de aula.

“Tudo na escola deve ser feito para educar. Tudo. Assim, a sujeira deseduca, o abandono deseduca, a desorganização deseduca. Por outro

lado, a limpeza educa, a organização educa, as paredes educam, os quadros educam, as plantas educam. Por isso a estrutura física para mim é importante para a visualização da seriedade do processo e da concepção que se tem da escola.” (apud LUZ e JESUS, 2006, p. 46) Citado por MAIA (2010)

Assim sendo do ponto de vista pedagógico, a sala de aula não pode ser encarada apenas como local destinado a abrigar alunos e professores durante o trabalho escolar, pois ela é, antes de tudo, um meio educativo, não descartando os recursos físicos e didáticos que a escola dispõe. Em relação aos recursos didáticos é preciso que sejam utilizados pelos professores como material pedagógico e que os alunos tenham acesso a eles.

De nada adianta, por exemplo, um laboratório de informática fechado a maior parte do tempo e que os professores não saibam utilizar em suas aulas. Uma das providências é melhorar a qualificação dos professores e formá - los para poderem elaborar um ensino de qualidade. (LUZ E JESUS, 2006) citado por MAIA (2010)

Dai a necessidade da escola dispor de materiais didáticos eficazes para as aulas e ainda instruir os alunos a manuseia – los com vista a garantir o domínio dos mesmos.

Gomes (2018) através do seu artigo intitulado “fatores que facilitam e dificultam a aprendizagem” faz uma grande contribuição na distinção de diversos fatores que podem de alguma forma ser o fracasso ou sucesso do aluno no processo de ensino e aprendizagem.

No seu artigo, este autor destaca primeiramente dois fatores: Sociais e económicos como sendo um dos que influenciam positiva e negativamente no desempenho do aluno.

“Quando as condições financeiras ou econômicas das famílias não permitem um maior cuidado ou zelo para com a criança, pode haver baixo rendimento escolar por falta de recursos que lhe proporcionem boa alimentação, boa vestimenta ou melhor qualidade de vida, de saúde, lazer etc.” GOMES (2018)

Dito isso, não descarta o meio social onde a criança ou aluno esta inserido, pois este acredita que o comportamento ou ainda os aspetos da personalidade do individuo em geral são frutos daquilo que ele vivencia no meio onde cresce.

(...) comportamentos inadequados por parte de pais ou responsáveis, principalmente promiscuidade, prostituição, drogas na família, violência doméstica, desemprego e desestruturação familiar são fatores que interferem diretamente no comportamento da criança ou

adolescente, contribuindo para dificultar sua aprendizagem. Idem (2018)

Outro aspecto importante que este descreve nos fatores sociais e económicos, é a situação de emprego dos pais e encarregados de educação da criança ou aluno:

“Fatores como desemprego ou subemprego dos pais ou responsáveis pela criança ou adolescente têm elevado as estatísticas de evasão, desistência, repetência e reprovação escolar, causadas, na maioria das vezes, pelo fato de ele ter que trabalhar para ajudar no aumento da renda familiar, deixando de lado os estudos.” Idem (2018)

Portanto, é visivelmente lógico que um pai ou encarregado de educação desempregado terá mais dificuldades na compra e fornecimento de material escolar ao seu educado em comparação com um pai ou encarregado de educação com a economia estável, tal demora na resposta do pedido da criança pode desencadear reações diversas que por sua vez pode influenciar na maneira de como este aluno se comporta nas aulas.

Gomes (2018) aponta também os fatores físicos e mentais como sendo os influenciadores no rendimento escolar, este explica que alunos deficientes físicos e mentais sofrem limitações diversas, no caso de deficientes físicos ele explica que “alunos com dificuldade de locomoção enfrentam mais dificuldades que alunos que não possuem essas limitações”, é de fato um desafio alunos com essas características frequentarem as escolas adequadamente, principalmente quando as escolas não tem condições criadas para eles, como é o caso de rampas para cadeirantes e banheiros adaptados.

Quando os fatores são mentais, as dificuldades são muito maiores, visto que muitas escolas ainda não estão preparadas para receber essas crianças em suas salas de aulas regulares, Segundo Gomes (2018), é devido principalmente à falta de professores e funcionários qualificados, o que, sem sombra de dúvida, torna-se fator limitante da aprendizagem, principalmente quando eles precisam de cuidados especiais por parte de professores e acompanhantes, tornando esse fator determinante para a aprendizagem de alunos com necessidades especiais, notadamente quando a escola ainda não está inserida na perspectiva inclusiva.

Por último, Gomes (2018) sublinha a atuação dos professores como sendo um dos fatores que influencia no desempenho dos alunos, na perspectiva deste autor, as formas de o professor conduzir a aula que dificultam a aprendizagem do aluno são:

"(...) não ter clareza nas explicações; apresentar aulas monótonas; recusar-se a ensinar e falta de comprometimento; não reconhecer, não permitir a produção e a participação do aluno; não esclarecer as dúvidas dos alunos; não corrigir lições; apresentar uma linguagem inadequada e não ter clareza nas explicações. " GOMES (2018)

É notório que tais práticas pedagógicas são ineficazes, são um grande obstáculo à aprendizagem do aluno. O fato de o professor não responder ou esclarecer as dúvidas levantadas pelos alunos é entendida eles como sinal de descaso para com eles e para com a sua aprendizagem. Tal falta de comprometimento provoca no aluno atitudes como falta de interesse e desestímulo para aprender.

3.0.3 Rendimento Escolar em Matemática

Os fatores que influenciam no rendimento acadêmico mas concretamente no que tange ao ensino da matemática, podem ser analisados com base nas ideias de Oliveira (2015), onde em síntese descreve primeiramente algumas variáveis que podem ser tidas como influenciadoras no rendimento acadêmico em matemática, "As variáveis sexo e estatuto socioeconômico e cultural dos alunos destacam-se entre o conjunto de fatores que poderão influenciar o rendimento escolar em matemática." OLIVEIRA (2015)

De fato, existe uma marcante tendência para os rapazes obterem níveis mais elevados do que as raparigas, sendo as diferenças de sexo maiores nos níveis de proficiência mais baixos.

Em coordenação com os autores que descrevemos anteriormente, Oliveira (2015) repisa também que o contexto familiar e o meio social do aluno também são tios como fatores influenciadores do rendimento escolar.

Sendo a família o primeiro contexto em que a criança aprende e se desenvolve, uma das variáveis que se torna importante na análise do sucesso escolar prende-se à origem sociocultural do aluno: os hábitos de vida, os modelos de interação familiar, as expectativas educativas em relação à aprendizagem dos filhos e a falta ou a baixa qualidade do apoio em relação às tarefas escolares, entre outros, influenciam o desenvolvimento e o rendimento escolar (Miranda et al, 2011) citado por Oliveira (2015)

Por último, realça o nível de escolaridade dos pais influencia o sucesso e o percurso académico dos alunos, sobretudo dos rapazes na medida em que os pais formados reconhecem a importância de estudos em casa e revisão da matéria, estes tem sempre a tendência de motivar os seus filhos a terem o hábito de lerem em casa.

No que diz respeito ao contexto escolar, este autor explora a possibilidade de haver um empurrão ao fracasso por parte dos professores, “A perda de interesse pela escola não decorre apenas de fatores internos ao aluno. Parte dessa perda de interesse deve-se à própria escola, que de alguma forma os afasta”. Desta feita, o meio escolar é também um fator muito contribuidor no desempenho dos alunos.

Uma descrição geral e correlação entre os factores sociais e escolares do aluno é feita nos seguintes termos:

“Existem fatores extrínsecos e intrínsecos às escolas que determinam o desempenho cognitivo, e que podem ser agrupados em três grandes categorias: os associados à estrutura escolar, os associados à família e aqueles relacionados com o próprio aluno.” (Soares, 2004) citado por Oliveira (2015)

Entre estes, os fatores extra-escolares explicam mais das desigualdades observadas no desempenho dos alunos do que os fatores intra-escolares, que de acordo com este autor, são os seguintes: as características socioeconómicas e culturais de proveniência dos alunos.

“Entre os fatores relativos ao próprio aluno estão as experiências prévias de repetência, de expulsões e a idade acima da média dos alunos que frequentam determinado ano. Já do lado da escola encontramos como principal fator o contexto social onde a escola está inserida, capaz de determinar oportunidades e restrições à própria escola e aos alunos que ela recebe.” Idem (2015)

Assim, estudar o rendimento escolar dos alunos implica conhecer o contexto geográfico em que a escola se insere. Entre os fatores internos à escola que determinam a qualidade das aprendizagens e do próprio desenvolvimento do aluno estão os recursos disponíveis (humanos e materiais), o tipo de gestão, a administração, o corpo docente e discente, a relação com a comunidade e com as famílias dos alunos, a cultura de escola, as relações entre os diferentes atores, o projeto pedagógico, a distribuição dos alunos por turmas, os currículos e as avaliações

CAPITULO IV

4.0 Análise e Interpretação dos dados

4.0.1 Analise Descritiva dos dados

Nesta etapa vamos proceder a análise e interpretação dos dados colhidos, começando pelas estatísticas descritivas das variáveis quantitativas. Portanto, esses dados podem ser visualizados na tabela a seguir:

Variável	Máximo	Mínimo	Média	Variância	Mediana	CV
Idade	17	14	15,57	0,302	16	3,52%
MFM	18	0	8,94	12,368	10	39,33%
MFP	17	0	10,74	4,915	11	20,64%
MFF	18	0	10,41	10,585	10	31,25%
MFQ	17	0	8,70	12,576	10	40,76%
NRef	4	2	3,03	0,101	3	10,48%
NPess	10	3	6,50	1,846	6	20,90%

Tabela 5: Análise descritiva das variáveis quantitativas (fonte: Autor)

A tabela acima da análise descritiva univariada agrupa sete variáveis diferentes, os dados são igualmente pertencentes a 421 alunos onde destes, 50% tem 16 anos com uma média equivalente a 16 anos. A variação de 0,302 e o Coeficiente de variação de 3,52% percebe-se que a distribuição das idades é homogénea sem muita variabilidade, ou seja, em termos de idade, maior parte dos alunos estão num nível e classe adequada pois de acordo com o ponto 2 do artigo 6 da lei n.º18/2018 do sistema nacional da educação “a criança deve ser obrigatoriamente matriculada na 1ª classe (...) no ano em que completa 6 anos de idade”, com isto, na 10ª classe ela já seria um jovem de 15 anos, idade esta muito vizinha da idade média dos dados. Este grupo apresenta ainda uma idade mínima de 14 anos e máxima 17 anos. No que diz respeito a média de matemática podemos observar que ela varia no intervalo de 0 a 18 valores, tendo uma media abaixo de dez valores (8,94) o que representa uma preocupação, a

sua variância e coeficiente de variação muito altos indica nos que as médias em destaque estão muito dispersas ou representam uma distribuição heterogénea. Em relação a media final de português, ela varia no intervalo de 0 a 17, tendo uma media muito boa (acima de dez valores) uma variância muito baixa representando assim uma distribuição homogénea e 11 valores separa 50% das médias dessa disciplina. As médias de física e química estão distribuídas no intervalo de 0 a 18 e 0 a 17 respetivamente, a media de física possui uma boa media (acima de dez) em relação a media de química (abaixo de dez), ambas distribuições são heterogéneas pois possuem uma variância muito alta (10,58 e 12,58 respetivamente) por ultimo, metade dessas notas estão separados pela nota 10 valores. Por dia esses alunos tem 2 a 4 refeições, com uma média de 3 refeições (pequeno - almoço, almoço e jantar) a distribuição dessas refeições é homogénea. Por último, na casa desses alunos o número de pessoas varia de 3 a 10 indivíduos com uma media de 7 pessoas por residência, variação muito baixa (1,846) e coeficiente de variação 20,90% dando nos a entender de tratar – se de uma distribuição homogénea.

4.0.2 Medidas de Assimetria e Curtose

No que concerne ao estudo de assimetria e medida de achatamento ou curtose, podemos recorrer as abordagens feitas por Reis (2008, p. 127) onde estabelece uma comparação entre as medidas de tendência central (media, moda e mediana) para concluir acerca da medida de assimetria da distribuição e para a curtose, a conclusão é com base nos valores obtidos no SPSS. Portanto, podemos visualizar a tabela abaixo com o resumo detalhado das medidas para cada variável de interesse:

Variável	Assimetria	Curtose	Distribuição	Atípicos
Idade	Negativa	Platicurtica	Não Normal	Moderado
MFM	Negativa	Leptocurtica	Não Normal	Severo
MFP	Negativa	Leptocurtica	Não Normal	Moderado
MFF	Negativa	Leptocurtica	Não Normal	Moderado
MFQ	Negativa	Leptocurtica	Não Normal	Severo
NRef	Positiva	Leptocurtica	Não Normal	Moderado

NPess	Simétrica	Mesocurtica	Norma	Moderado
-------	-----------	-------------	-------	----------

Tabela 6: Medidas de assimetria e curtose (Fonte: Autor)

Desta feita, pode concluir - se que a variável NPess cuja assimetria é simétrica, seus dados provem de uma distribuição Gaussiana ou normal, por outro lado, a variável NRef com assimetria positiva possui uma grande quantidade de valores pequenos em comparação a existência de valores altos na distribuição e por ultimo, as variáveis assimetricamente negativas possuem uma grande quantidade de valores baixos em relação a valores altos. Por último as medidas de curtose de caracter leptocurtica suas distribuições são menos achatadas do que a distribuição normal, o inverso observa se na variável idade que é platicurtica.

4.0.3 Análise de Frequências

Na análise frequência procede – se o estudo das variáveis categóricas, onde iremos estudar essencialmente a distribuição de frequências da caracterização dos alunos, tal estudo enquadra cada categoria ou divisão que as variáveis apresentam como ilustra a tabela abaixo:

Variável	Categorias	Frequência	%
Sexo	Feminino	229	54,4
	Masculino	192	45,6
Estado Civil	Solteiro	421	100
	Casado	0	0
	VMA	0	0
Tipo de casa	MC	421	100
	MNC	0	0
Resultado obtido no 2º T 2021	Aprovado	301	71,5
	Reprovado	120	28,5
Pais Trabalham	Sim	298	70,8
	Não	123	29,2
Distância de Casa ate Escola	Menos de 1km	13	3,1
	1km a 3km	158	37,5
	3km a 6km	217	51,5

	Mais de 6km	33	7,8
Vem a escola de que meio	A pé	387	91,9
	Transporte Publico	33	7,8
	Viatura Particular	1	0,2
Tempo que leva de casa ate escola	Menos de 30min	74	17,6
	Uma a duas horas	306	72,7
	Duas a quatro horas	41	9,7
Pais tem dificuldade na compra de Material Escolar	Sim	23	5,5
	Não	398	94,5
Grau Máximo de Escolaridade dos Pais	Primário	44	10,5
	Secundário	208	49,4
	Técnico Profissional	108	25,7
	Superior	38	9
	Não estudaram	23	5,5

Tabela 7: Análise de frequências das variáveis categóricas (Fonte: autor)

Os dados da tabela acima ilustram – nos que dos alunos em estudo, maior parte são do sexo feminino, arrecadando cerca de 54,4% contra 45,6% dos homens, devido a idade, 100% dos alunos são solteiros e vivem em casas de material de construção convencional, dos resultados obtidos no trimestre anterior felizmente maior parte deles (71,5%) foram aprovados contra 28,5% de reprovados. 298 Alunos (70,8%) seus pais trabalham tendo assim a facilidade da compra de material escolar e 123 alunos (29,2%) seus pais não trabalham. Por outro lado, 217 alunos (51,5%) percorrem cerca de 3 a 6 km para chegarem a escola, verifica se também que maior parte deles (91,9%) vão a escola a pé e nesse trajeto, 306 deles (72,7%) levam entre uma a duas horas de caminhada para chegarem a escola. Felizmente 94,5% dos alunos seus pais não tem dificuldades na compra de material escolar contra 5,5% de alunos cujos pais enfrentam dificuldades na aquisição de material escolar para seus filhos. Por ultimo, no que concerne ao nível académico máximo dos pais, constata – se que 44 alunos (10,5%) seus pais tem nível primário, 208 alunos (49,4%) pais tem nível secundário, 108 alunos (25,7%) pais tem nível técnico – profissional e por ultimo, menor parte 38 alunos (9%) os pais possuem nível superior completo.

4.1 Fatores que influenciam no rendimento da disciplina de matemática

4.1.1 Análise da correlação usando o coeficiente de Pearson

Para fazer uma breve análise das possíveis variáveis que estarão correlacionadas com a variável dependente (média da disciplina de matemática), vamos recorreremos ao uso do coeficiente de correlação de Pearson onde este ilustra - nos as variáveis que possuem uma forte correlação com a nossa variável dependente dando a entendes desta forma a possibilidade de esta influenciar no rendimento.

Correlação de Pearson		
Variáveis	MFM	
	r - Pearson	P – valor
Idade	0,040	0,409
MFP	0,218	0,000
MFF	0,263	0,000
MFQ	0,197	0,000
NRef	0,195	0,000
DCE	- 0,091	0,030
NPess	- 0,153	0,001
TCE	- 0,180	0,000

Tabela 8: Teste de coeficiente de Pearson (Fonte: autor)

A tabela 8 apresenta dados de saída do teste de Pearson cuja finalidade é avaliar a potencialidade de correlação entre a variável dependente e as demais variáveis. Portanto, temos o valor de r que é o coeficiente de Pearson acompanhado do p – valor para a correlação entre a variável dependente e a variável independente específica. Portanto, foram identificados os fatores que estão correlacionado com o rendimento pedagógico no nível de confiança de 95%, podemos constatar que MFP, MFF, MFQ e NRef estão correlacionado com a MFM visto que o p-valor <0.05, desta conclui se também que a intensidade da relação é moderada com isso podemos

assumir que as variáveis em causa são preditoras do rendimento da disciplina de matemática de forma proporcional ou seja, quando o estudante tiver positiva nas disciplinas em destaque e o numero de refeições (NRef) aumentar, conseqüentemente este também terá um aumento na sua média da disciplina de matemática

Por outro lado, podemos constatar também que as variáveis DCE, NPess e TCE estão relacionado de forma invertida com a média da disciplina de matemática, isto é, quando aumenta a média da disciplina de matemática, o valor dessas variáveis diminui, em termos práticos, a media de matemática aumenta quando o aluno percorrer curtas distâncias ao sair de casa ate escola, quando o número de pessoas em sua residência diminui e por último quando o tempo que este leva ao sair de casa ate escola também diminui.

4.1.2 Modelo de regressão linear múltipla (MRLM)

Após ser realizada a regressão linear múltipla no SPSS concernente a uma variável dependente (NFM) e três variáveis independentes (NFP, NFF e NRef) através do método Backward, os resultados das saídas são de facto muito interessantes, os mesmos podem ser observados nas tabelas a seguir.

Resumo do Modelo			
Modelo	R quadrado	P - valor	Durbin - Watson
1	0,107	0,000	1,618

Tabela 9: Resumo do Modelo (Fonte: Autor)

A tabela 9 ilustra-nos o resumo do modelo obtido, de acordo com Vieira (2019), “O valor de R^2 é o coeficiente de determinação. É a proporção de variação na variável dependente explicada pelas variáveis independentes. O valor $R^2 = 0,107$ mostra que as variáveis independentes explicam 10,7% da variação da variável dependente MFM. O valor de p valor é menor que 0,05 indicando um boa estimativa do modelo, por ultimo, o valor de Durbin Watson é o principal teste para investigar a independência dos resíduos ou a existência de auto correlação entre eles, Pereira (2019, p. 54) defende que “valores menores do que 1 ou maiores do que 3 devem, definitivamente, ser motivos de preocupação (...). Quanto mais próximo de 2 o valor estiver, melhor.” Para o nosso modelo DW é igual a 1,618 que cita no intervalo de 1 a 3, com isto, conclui – se que os resíduos do modelo são independentes.

ANOVA			
Modelo	Soma dos quadrados		P- valor
1	Regressão	554,093	0,000
	Resíduos	4640,539	
	Total	5194,632	

Tabela 10: ANOVA do Modelo (Fonte: autor)

A segunda tabela de saída é a tabela referente a análise da variância, ou tabela ANOVA (em inglês: Analise of Variance) nesta tabela, observa – se que temos um p – valor correspondente a 0,000, isso remete a ideia de que as variáveis independentes prevê a variável dependente de forma significativa porque p valor é menor que 0,05. Em outras palavras, o modelo de regressão se ajusta bem aos dados.

Coeficientes do Modelo						
Coeficientes (β)	Erros	P - valor	Intervalo de Confiança para β (95%)		VIF	
			L_I	L_S		
Constante	- 0,130	1,659	0,938	-3,391	3,131	
MFP	0,214	0,078	0,006	0,061	0,368	1,131
MFF	0,207	0,054	0,000	0,101	0,312	1,151
NRef	1,524	0,523	0,004	0,496	2,553	1,048

Tabela 11: Coeficientes do Modelo (Fonte: autor)

A tabela 11 é referente aos coeficientes estimados (não padronizados) do modelo, a significância dos coeficientes e o fator de inflação da variância, verifica – se que todos coeficientes (com exceção ao declive de regressão) tem um p – valor menor que 0,05 dando nos a entender que os coeficientes são significativamente diferentes de zero. Variance Inflation Factor (VIF) é um indicador usado com frequência para detectar a multicolinearidade, ele avalia o quanto a variância de um coeficiente de regressão estimado aumenta se as suas preditoras estiverem correlacionadas. Da abordagem de Maroco (2003) citado por Rodrigues (2012, p.40), o valor de VIF ideal deve estar no intervalo de 1 a 10, através da tabela 11 conclui - se que não há multicolinearidade entre as variáveis.

Com base nos dados das Média finais de português (NFP), Média finais de física (NFF) e o Numero de refeições (NRef), foi ajustada uma regressão linear múltipla para prever a nota final de matemática (NFM):

$$NFM = -0,130 + 0,214 \times MFP + 0,207 \times MFF + 1,524 \times NRef + \varepsilon$$

A equação de regressão linear múltipla obtida esclarece – nos que, o aumento de uma unidade que o aluno tiver na sua nota de português, nota de física e o número de refeições, aumenta a sua nota de matemática em 0,214, 0,207 e 1,524 respetivamente. O valor ε corresponde ao resíduo ou erro de estimação do modelo. Desta feita, a variável NRef tem maior impacto no rendimento da disciplina de matemática

4.1.2.1 Normalidade dos Resíduos

A normalidade dos resíduos padronizados do modelo valido podem ser observados com base no histograma com curva de distribuição normal abaixo:

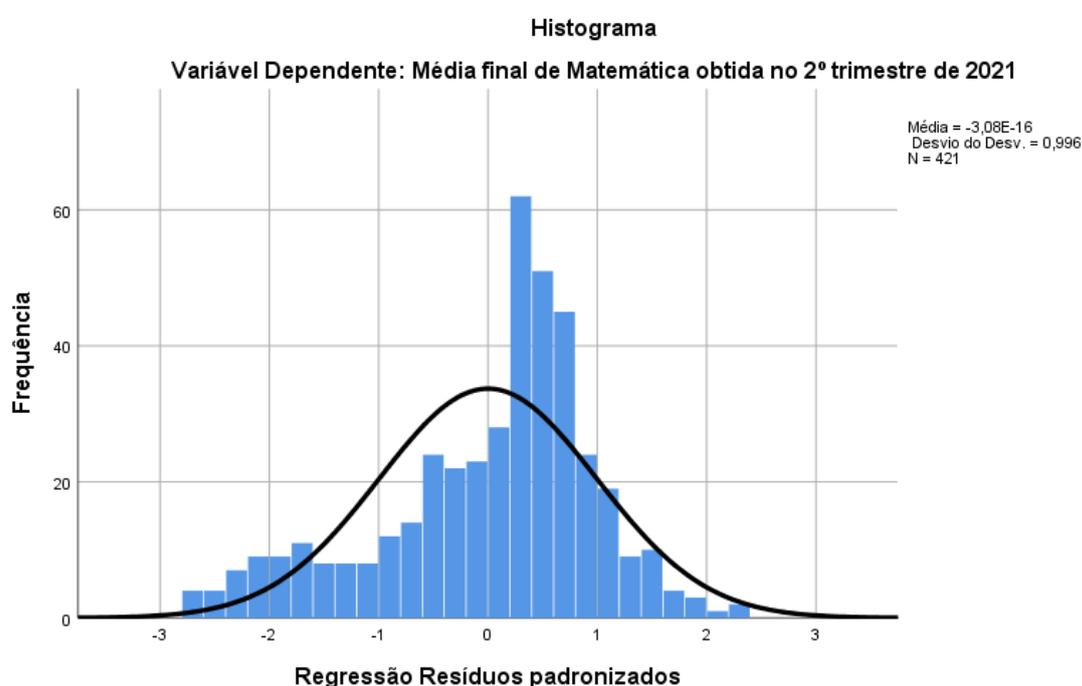


Figura 3: Histograma de distribuição normal (Fonte: autor)

Portanto, com base na figura 3, observa - se que a distribuição dos resíduos se aproxima a distribuição normal ou Gaussiana, contendo alguns valores atípicos de alta frequência.

4.1.2.2 Homoscedasticidade (Variância Constante) dos Resíduos

De acordo com Rodrigues (2012) “Uma das técnicas usadas para verificar a suposição de que os resíduos são homoscedásticos, é a análise do gráfico dos resíduos versus valores ajustados. Este gráfico deve apresentar pontos dispostos aleatoriamente sem nenhum padrão definido”. Dos resíduos do modelo, tem – se o seguinte gráfico:

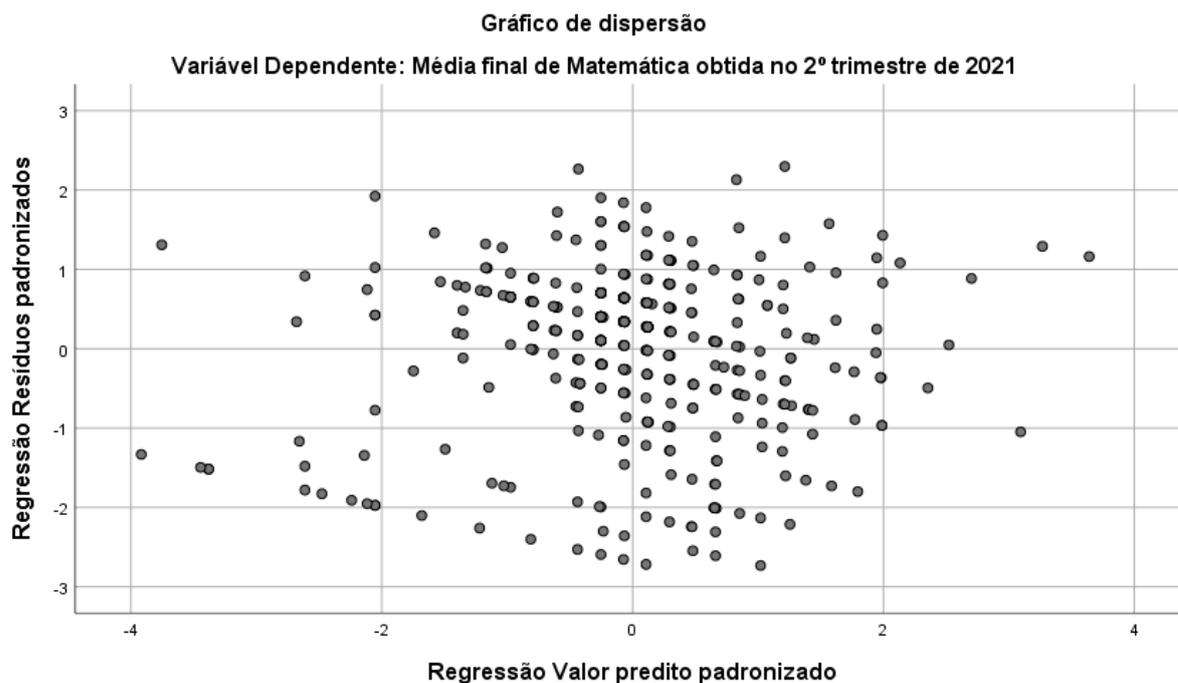


Figura 4: Gráfico dos Resíduos vs Valores Ajustados (Fonte: Autor)

Como podemos observar na figura 4, os pontos estão aleatoriamente distribuídos sem nenhum padrão significativo e a sua maioria em torno do eixo vertical zero (0), face a isto, conclui - se que os resíduos são homoscedásticos.

CAPITULO V

5.0 Considerações Finais

O presente trabalho chegou a reta final de apresentação e elaboração, recordar que este versou acerca de um tópico muito importante no que diz respeito a educação, foi debruçado neste trabalho o estudo prático dos fatores que influenciam no rendimento académico da disciplina de matemática estudo realizado especificamente na escola secundária heróis moçambicanos de moatize província de tete.

De entre os objetivos da pesquisa, tínhamos como meta identificar os fatores que influenciam no rendimento académico da disciplina de matemática dos alunos da 10ª classe da escola secundária heróis moçambicanos de moatize, portanto, a pesquisa permitiu destacar que a media final de português, Física, química e o número de refeições são os que tem maior influência em proporção direta no rendimento da matemática, por outro lado podemos constatou - se também que a distancia de casa ate escola, o numero de pessoas que residem numa casa e o tempo que aluno leva para sair de casa ate a escola estão relacionado de forma invertida com a média da disciplina de matemática, isto é, quando aumenta a média da disciplina de matemática, o valor dessas variáveis diminui.

No que concerne a descrição e caracterização desses alunos, o estudo mostrou que as suas idades variavam no intervalo de 14 a 17 anos, onde todos são solteiros e vivem em casas de material de construção convencional, mais de metade são do sexo feminino, maior números deles não são repetentes, maior parte deles os encarregados de educação trabalham e não tem dificuldades na compra de material escolar, mais de metade percorrem uma distancia de 3km a 6 km para chegar a escola, sendo 306 alunos levam em média uma a duas horas para chegar a escola na sua maioria a pé, por ultimo o estudo mostrou também que quase metade desses alunos os seus encarregados possuem nível secundário como nível de escolaridade máximo.

Do modelo de regressão linear múltipla válido obtido, pode concluir - se que a variável número de refeições é a que tem maior impacto no rendimento académico da disciplina de matemática. Portanto, das hipóteses desenhadas para o estudo, a hipótese nula é aceitável com particular destaque para as médias de português, física e a situação social do aluno mas especificamente ao número de refeições, respondendo a pergunta de pesquisa, a pesquisa mostra que a variável

número de refeições (NRef) é a que mais impacta no rendimento académico da disciplina de matemática dos alunos da 10^a classe da escola secundária heróis moçambicanos de moatize.

5.1 Sugestões

A pesquisa conducente aos fatores que influenciam no rendimento académico da disciplina de matemática dos alunos da 10ª classe da escola secundária heróis moçambicanos de moatize trouxe resultados notáveis e que durante a realização desta mesma notou se ser necessário se ampliar os trabalhos futuros.

Por isso, numa pesquisa a ser levada a cabo no futuro sobre este mesmo assunto, recomenda – se um maior contacto com os alunos ou público-alvo com vista a extrair mais fatores influenciadores com vista a garantir um modelo linearmente regressivo mais amplo.

5.2 Barreiras e Limitações

Durante a pesquisa deparou – se com algumas barreiras notáveis, a que mais se destaca foi a frequente dificuldade de encontrar alunos no recinto para participarem no processo da recolha de dados.

Outro aspecto é inerente a impaciência por parte dos mesmos alunos para fornecer dados precisos e com calma, visto que o autor recolheu os dados com o auxílio da plataforma de formulários do Google, era necessário mais tempo para poder anotar as respostas e nesse período os entrevistados reclamavam de estarem com fome.

5.3 Referencias Bibliográficas

1. Almeida, G. S, et all (2021), *Fatores associados ao rendimento acadêmico na formação inicial de professores*, Fortaleza, Ceara – Brasil disponível em <https://doi.org/10.17979/reipe.2021.8.1.7546> acesso em 16/12/2021;
2. CASTIGO, S., (2021), *Econometria Básica*, Universidade Pungue, Tete;
3. FREITAS, F. M., (2010), *Fatores Que Influenciam O Desempenho Escolar De Adolescentes De Uma Instituição Pública Do Município De Rio Verde*, Rio Verde;
4. FUNDÃO, N. M, (2018), *Modelo de Regressão Linear: Aplicação ao Estudo sobre os Fatores que Influenciam o Rendimento Acadêmico dos Alunos em Angola*, UBI, Covilhã;
5. GERHARDT e SILVEIRA, T. E e D. T, (2009), *Métodos de Pesquisa*, 1ª edição, editora da UFGR, Porto Alegre;
6. GOMES, M. M., (2018), *Fatores que dificultam e facilitam a aprendizagem*, Portal Educação Publica disponível em <https://educacaopublica.cecierj.edu.br/artigos/18/14/fatores-que-facilitam-e-dificultam-a-aprendizagem> acesso em 14/01/2022;
7. HOPPEN e PRATES, J. e H., (2017), *Outliers, o que são e como trata – los em uma análise de dados*, Aquarela, Brasil;
8. Lei n.º18/2018 do Sistema Nacional de Educação disponível em https://www.ilo.org/serial_pdf/28_de_dezembro_de2018/3748/lei_nr_18/2018;
9. MAIA, Z., (2010), *Fatores internos e externos que influenciam no rendimento escolar*, UFP, Corithiba;
10. MORETTI e BUSSAB, P. A. e W. O., (2010), *Estatística Básica*, 6ª edição, Saraiva, São Paulo;
11. MORETTI e BUSSAB, P. A. e W. O., (2006), *Estatística Básica*, 5ª edição, Saraiva, São Paulo;
12. OLIVEIRA, T. M. (2001), *Amostragem não Probabilística: Adequação de Situações para uso e Limitações de amostras por Conveniência, Julgamento e Quotas*, 2ª edição, FECAP;
13. OLIVEIRA, M., (2015), *Rendimento Escolar Em Matemática Vs Atitudes Face À Matemática: Fatores De Contexto E Efeito Escola1*, Portugal;
14. PERREIRA, A. S, et all, (2019), *Regressao linear multipla: como simplificar por meio de Excel e SPSS?*, Brasil;
15. REIS e REIS, E.A e I.A, (2002), *Analise descritiva de dados*, 1ª edição, UFMG, Minas Gerais;

16. REIS, E. (2008), *Estatística Descritiva*, 7ª Edição, Edições Silabo, Lisboa;
17. RODRIGUES, L. O. (2021), *Objetivos da Educação em nossa sociedade*, Brasil Escola. Disponível em: <https://brasilescola.uol.com.br/sociologia/objetivos-educacao.htm>. Acesso em 29 de dezembro de 2021;
18. RODRIGUES, S. C., (2012), *Modelo de regressão linear e suas aplicações*, UBI, Covilhã;
19. VIEIRA, S., (2019), *Regressão Linear Múltipla: Interpretando a saída do SPSS*, disponível em <http://soniavieira.blogspot.com/2019/10/interpretando-saida-do-spss-para.html> acesso em 09/02/2022;
20. ZAMBON, L. S., (2010), *Erro humano – conceitos*, Medicina Net, disponível em https://www.medicinanet.com.br/conteudos/gerenciamento/3097/erro_humano_conceito_s.htm acesso em 07/01/2022;

CAPITULO VI

6.0 Apêndices:

Apêndice 1: Questionário

Este questionário faz parte de uma pesquisa com o título: **Fatores que Influenciam no Rendimento Académico da disciplina de Matemática dos alunos da 10ª classe da Escola Secundaria Heróis Moçambicanos de Moatize em 2021: Um olhar pela Regressão Linear Múltipla.**

Para que este trabalho possa ser concluído, preciso de sua colaboração e máxima sinceridade nas respostas para levantar algumas informações sobre o tema em questão. Os dados desta pesquisa serão analisados de maneira agrupada e com objetivo estritamente acadêmico. A qualidade desta pesquisa depende da precisão de suas respostas. Por favor, leia as instruções com atenção antes de responder às questões e não deixe nenhuma em branco.

Desde já, agradeço sua colaboração
Eduardo Félix Pita Duarte

Estimado aluno do Ensino Secundário! A seguir escolha apenas uma alternativa e marque um “círculo” na opção que melhor caracterize sua situação em relação aos “eixos” das questões abaixo e comenta por suas palavras se fores solicitado.

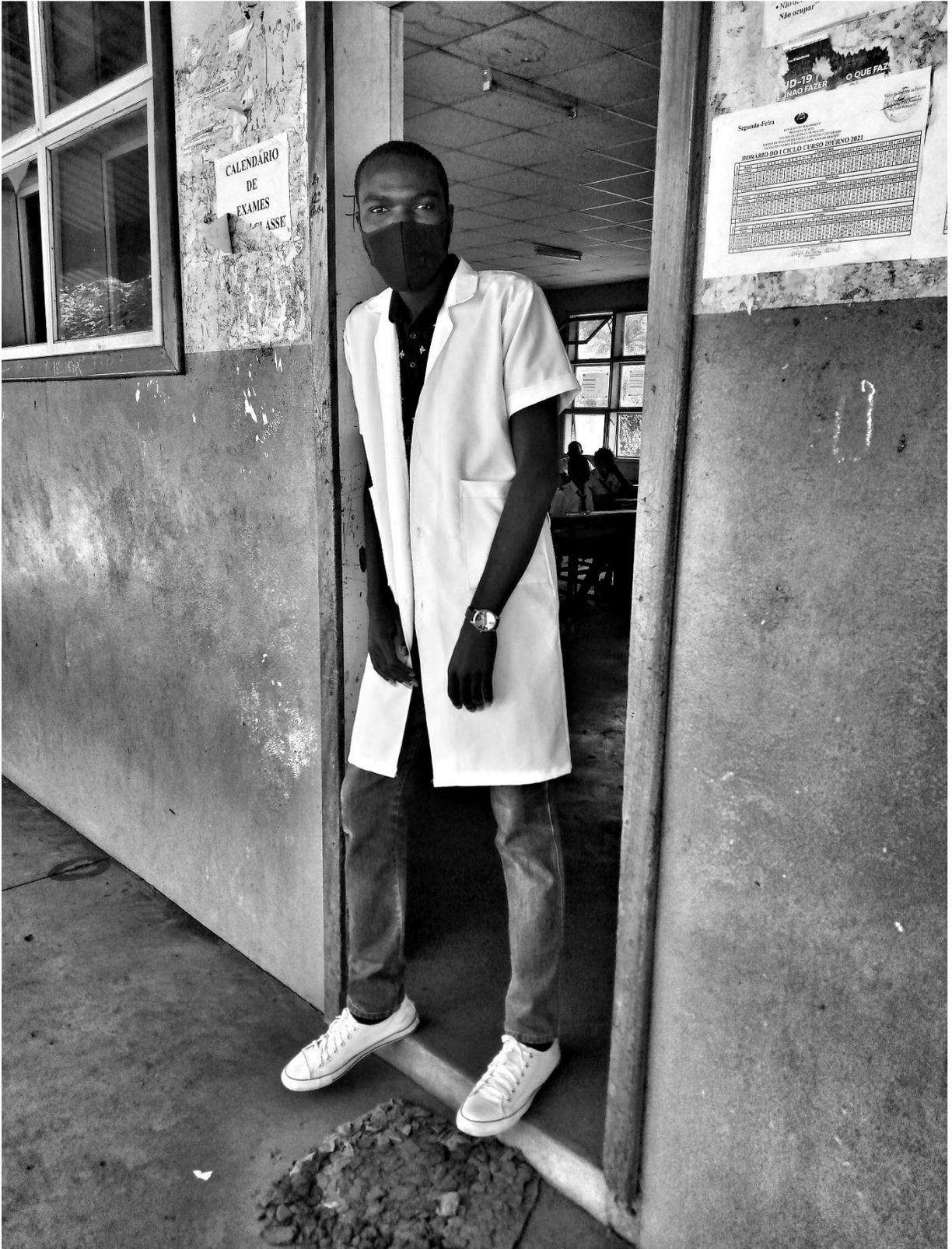
DADOS PESSOAIS	
1.Sexo	1.Masculino 2.Feminino
2.IdadeAnos
3.Estado civil	1.Solteiro 2.Casado 3.Vive maritalmente com alguém
SITUAÇÃO ACADÉMICA DO ALUNO	
4. Resultado obtido no 2º trimestre de 2021?	1.Aprovado 2.Reprovado
5. Média final de Matemática obtida no 2º trimestre de 2021Valores
6. Média final de Português obtida no 2º trimestre de 2021Valores
7. Média final de Física obtida no 2º trimestre de 2021Valores
8. Média final de Química obtida no 2º trimestre de 2021Valores
CLASSIFICAÇÃO DAS RELAÇÕES NO AMBIENTE ESCOLAR	
9. Relação entre alunos e professores	1.Muito insatisfeito 2.Insatisfeito 3.Nem Satisfeito, nem insatisfeito

	4.Satisfeito 5.Muito satisfeito
10. Relação entre Alunos (estudo em grupo)	1.Muito insatisfeito 2.Insatisfeito 3.Nem Satisfeito, nem insatisfeito 4.Satisfeito 5.Muito satisfeito
11. Relação entre alunos e a direção da escola	1.Muito insatisfeito 2.Insatisfeito 3.Nem Satisfeito, nem insatisfeito 4.Satisfeito 5.Muito satisfeito
SITUAÇÃO SOCIAL DO ALUNO	
12. Que tipo de casa vive?	1.Material convencional 2.Material Não convencional
13. Seus Pais trabalham?	1.Sim 2.Não
14. Quantas Refeições tem por dia?	1.Uma 2.Duas 3.Três 4.Quatro
15. Qual a distância de sua casa ate escola?	1.Menos de 1km 2.1km a 3 km 3.3km a 6km 4.Mais de 6km
16. Vem a escola de que meio?	1. A pé 2. Transporte Publico

	3. Viatura Particular
17. Quanto tempo leva ao sair de casa ate a escola?	1.Menos de 30 min 2.Uma hora a Duas horas 3.Duas horas a quatro horas 4.Mais de quatro horas
18. Teus Pais tem dificuldades na compra de material escolar?	1.Sim 2.Nao
19. Quantas pessoas vivem em sua casa? Pessoas
20. Qual é o grau máximo de escolaridade dos seus pais?	1.Primário 2.Secundário 3.Técnico profissional 4.Superior 5.Não Tem

MUITO OBRIGADO PELA ATENÇÃO!

Apêndice 2: Autor na época de estágio pedagógico em ensino de matemática



Apêndice 3: Autor na recolha de dados



Apêndice 4: Alguns alunos da 10^a classe



6.1 Anexos