


# A APRENDIZAGEM DA ESTATÍSTICA NA ESP-BIÉ. UMA REFLEXÃO A PARTIR DA EXPERIÊNCIA COM OS ESTUDANTES DO CURSO DE MATEMÁTICA

## Learning Statistics At Esp-Bié. A Reflection From The Experience With The Mathematics Course Students

Rosa Maria de **NASCIMENTO**

Escola Superior Pedagógica do Bié (ESPB), Cuito/Bié-Angola


[rnascimento22@hotmail.com](mailto:rnascimento22@hotmail.com)


<https://orcid.org/0000-0001-7836-7441> 

Ezequias Adolfo Domingas **CASELA**

Escola Superior Pedagógica do Bié (ESPB), Cuito-Bié, Angola

[ezequiasadolfo@hotmail.com](mailto:ezequiasadolfo@hotmail.com)

<https://orcid.org/0000-0001-7703-0097> 

A lista completa com informações dos autores está no final do artigo 

### RESUMO

Este artigo objetiva-se em apresentar os resultados das experiências de aprendizagens adquiridas pelos estudantes do curso de Matemática da Escola Superior Pedagógica do Bié, na disciplina de Estatística. O estudo é quantitativo do tipo experimental, pois, descreve uma experiência com 45 estudantes afetos ao segundo ano do curso de ensino da Matemática, a partir do trabalho com dados de situações do cotidiano. Os resultados do pré-teste e pós-teste evidenciam um aumento do nível de conhecimento dos estudantes após a aplicação da estratégia. Como resultado da experiência destaca-se uma mudança de atitude e de enfoque metodológico no processamento estatístico dos dados, pois, os problemas exigiram a reflexão sobre os tipos de dados, sua estrutura, bem como encontrar padrões que caracterizam o seu comportamento, para posteriormente selecionar o modelo que mais se ajusta aos dados, com vista a correta tomada de decisões.

**Palavras-chave:** Aprendizagem da Estatística, variáveis estatísticas, Medidas estatísticas

### ABSTRACT

This article aims to present the results of the learning experiences acquired by students of the mathematics course at Bié Higher Pedagogical School, in the discipline of Statistics. The study is quantitative of the experimental type, as it describes an experience with 45 students in the second year of the mathematics teaching course, based on the work with data from everyday situations. The results of the pre-test and post-test show an increase in the students' level of knowledge after applying the strategy. As a result of the experience, there is a change in attitude and methodological focus in the statistical processing of data, as the problems required reflection on the types of data, its structure, as well as finding patterns that characterize its behavior, to later select the model that best fits the data, with a view to correct decision making.

**Keywords:** Learning Statistics, Statistical variables, Statistical measures

# 1. INTRODUÇÃO

O acelerado desenvolvimento científico-técnico e a complexa dinâmica social da época contemporânea necessita de homens capazes de manejar um grande volume de informações e utilizá-las na solução adequada de problemas. Muitas destas informações refletem fatos da realidade objetiva representados por modelos estatísticos desde os mais simples como gráficos, tabelas etc. aos mais complexos como as provas de inferência estatística e análise multivariada de dados. Independentemente de como se expressa a informação estatística, faz-se necessário que o cidadão possua um mínimo de conhecimento estatístico para saber interpretar estas informações de maneira crítica e conhecer melhor a natureza dos fenômenos aleatórios da realidade objetiva.

Ao nível internacional muitos pesquisadores desta área de conhecimento indicaram reiteradamente o significado da Estatística na ciência contemporânea, entre eles se destacam: Merino (2003), Martín (2006) e Cadima (2013). De modo geral, estes pesquisadores coincidem em assinalar a incidência da Estatística na análise de dados de problemas derivados da realidade e reiteram a importância desta disciplina para todas as ciências como ferramenta da análise necessária na tomada de decisões em condições de incerteza.

Apesar de ser reconhecida a importância da Estatística na análise de dados em diferentes ramos do saber, são claramente perceptíveis os problemas no seu ensino-aprendizagem dado o carácter abstrato com que se aborda os seus conteúdos.

De acordo com os estudos realizados por Li e Shen (1992) foi possível identificar problemas por parte dos alunos na eleição correta do tipo de gráfico, pois utilizam polígonos de frequências em variáveis qualitativas, ou diagrama de barras horizontal para representar a evolução do índice de produção industrial ao longo dos anos. Por outro lado, Pollatsek, Lima e Well (1981), identificaram problemas nos alunos no cálculo da média aritmética ponderada, pois utilizam a fórmula da média aritmética simples. De acordo com estes autores, a aproximação ao estudo dos estatísticos de posição central, baseada na definição algorítmica e no cálculo das referidas medidas num conjunto de dados “tipo” e descontextualizados, não permite aos alunos uma compreensão integral do conceito de promédio.

Ao nível nacional, Cassinela (2016), identificou problemas nos alunos na compreensão das medidas de tendência central e de posição, pois confundem a média com a mediana, a moda com a maior frequência absoluta. Não ordenam os dados antes de

calcular a mediana, assim como apresentam dificuldades no cálculo da posição dos valores centrais. No âmbito local (contexto da ESP-BIÉ), André (2018), identificou insuficiências pelos estudantes na interpretação de gráficos de frequências simples, na construção de tabelas de frequências, assim como no cálculo e interpretação da média.

Durante as aulas de Estatística, os autores desta pesquisa, enquanto professores desta disciplina, identificaram que os estudantes apresentam dificuldades na análise da incidência que um dado atípico pode exercer sobre a média, na seleção adequada da medida estatística em função do tipo de variável e sua escala de medição, no cálculo e análise das medidas de tendência central, na eleição do tipo de gráfico em função do tipo de variável assim como a sua respectiva análise.

Verificou-se ainda que as atividades desenvolvidas nas aulas não privilegiam o significado do conhecimento estatístico, porquanto enfatizam os procedimentos estatísticos associados à Matemática, sem demonstrar a sua aplicação prática de acordo com as características do profissional que se forma. Nesta perspectiva, considera-se a seguinte questão de pesquisa: como melhorar o ensino-aprendizagem da Estatística, nos alunos do segundo ano do curso de Matemática da Escola Superior Pedagógica do Bié?

De acordo com o anterior, esta pesquisa tem como objetivo: contribuir para a aprendizagem da Estatística a partir do processamento estatístico de dados de problemas do cotidiano. Para o efeito, aplicou-se uma estratégia de aprendizagem baseada na vinculação do conteúdo com situações do cotidiano, em que os alunos determinam o problema a resolver, envolvem-se na recolha de dados e realizam a análise estatística dos dados com base na seleção do tipo de modelo estatístico que melhor representa os dados em função da sua tipologia.

Assim sendo, no presente artigo, realiza-se uma sistematização das concepções teóricas do ensino-aprendizagem da Estatística, abordam-se as dificuldades na aprendizagem da Estatística, faz-se uma breve descrição da metodologia utilizada assim como a apresentação dos principais resultados obtidos na experiência realizada com os estudantes do segundo ano do curso do ensino da Matemática.

## **2. O PROCESSO DE ENSINO-APRENDIZAGEM DA ESTATÍSTICA**

Apesar de que geralmente a Estatística vem associada a muitos cálculos e operações aritméticas, na sua maior parte os seus fundamentos podem ser dominados sem que o estudante possua conhecimentos avançados de Matemática. De fato, a essência está

na utilização correta dos métodos estatísticos em função do tipo de dados e a interpretação do seu resultado no contexto de uma situação da realidade objetiva que se estuda. Desta forma, a Estatística se relaciona com o método científico complementando-o como ferramenta de análise e mesmo que a investigação científica não requeira necessariamente da Estatística, esta valida muitos dos resultados quantitativos derivados da pesquisa.

Nesta perspectiva, deduz-se que o seu objetivo principal é analisar os fenômenos da realidade objetiva a partir da recolha e processamento estatístico de dados, para a posterior tomada de decisões. Para o efeito, faz-se necessária uma compreensão conceitual e procedimental do seu objeto de conhecimento, ou seja, não basta que o aluno tenha conhecimento das fórmulas e algoritmos associados a um determinado conceito estatístico, mas sim saber quando, como e porquê aplicar um conceito específico em determinada série de dados.

O anteriormente refenciado, implica a realização de um conjunto de análises em torno do tipo de dados, as características de distribuição, ou seja, compreender que o dado não é apenas um número ou atributo mas sim um dado num determinado contexto. Para o cumprimento do anterior, se requiere uma compreensão adequada das medidas estatísticas e a sua aplicação na análise de problemas, o que tem resultado em grandes dificuldades para a maioria dos alunos.

Em função destas dificuldades, pesquisadores de diferentes países como Espanha, Cuba e Angola propõem novas metodologias no ensino-aprendizagem da Estatística, as quais mencionam-se a seguir:

Batanero e Godino (2001), propõem um ensino da Estatística baseado em projetos de análises de dados. Na referida investigação, sugerem a realização de atividades que contextualizam a aprendizagem com vista a permitir aos alunos a aplicação da estatística no processo investigativo.

Cassinela (2016), propõe um modelo interdisciplinar para o ensino-aprendizagem da Estatística que contribui a elevar os níveis motivacionais e o rendimento académico dos estudantes. Deste modo, os estudantes compreendem os conceitos básicos da Estatística e tornam-se capacitados para aplicar corretamente os métodos estatísticos no seu labor profissional assim como na pesquisa científica.

Álvarez (2012), elabora uma concepção teórica-metodológica sustentada no processo de ensino-aprendizagem desenvolvidor, na qual se contextualiza os conteúdos estatísticos com os problemas da prática educativa.

Nas pesquisas anteriormente referenciadas nota-se uma tendência em direcionar o ensino da Estatística nas aplicações, com enfoque interdisciplinar e centrado na análise de dados em problemas do cotidiano. A referida tendência transcende a mera aplicação de procedimentos algoritmos para o cálculo de medidas estatísticas.

Para o cumprimento do anterior, considera-se imperiosa a necessidade do trabalho com problemas que evidenciem o reflexo da realidade objetiva em correspondência com o contexto de atuação dos estudantes. Este aspecto tem uma importância relevante, pois vincula a Estatística com a sociedade em diferentes esferas de atuação e favorece a sua aplicação na interpretação de fenômenos da realidade objetiva. Além do mais, exerce um efeito motivador na medida em que os alunos compreendem o seu significado para a vida.

## **2.1. Dificuldades na aprendizagem da estatística na ESP-BIÉ**

A experiência dos autores desta pesquisa enquanto professores da disciplina de Estatística e o diagnóstico realizado aos alunos do segundo ano dos cursos do ensino da Matemática e da Física permitiu identificar insuficiências relacionadas com a análise da distribuição de frequências, aplicação memorística dos algoritmos, insuficiências na compreensão do significado dos conceitos estatísticos e na identificação dos modelos estatísticos que melhor representam os dados.

As dificuldades de aprendizagem dos conceitos e aplicação dos procedimentos estatísticos estão muitas vezes relacionadas com as concepções que o aluno tem sobre este conceito, associadas aos tipos de práticas no processamento de dados estatísticos. De modo geral, os professores acostumam os alunos a trabalhar com exercícios típicos, ou seja, quando introduzem um conceito, colocam exemplos e exercícios com séries de “dados ideais” para a aplicação do conceito introduzido, sem conduzir o aluno a refletir sobre as características do conceito e em que condições, ou que tipo de dados e de distribuições pode ser aplicado o conceito. Além do mais, após feitos os cálculos, não se conduz o aluno à análise do resultado obtido no contexto da variável em estudo.

Um estudo diagnóstico realizado a partir de uma prova pedagógica em 45 estudantes permitiu identificar as seguintes insuficiências na aprendizagem desta disciplina:

- Distribuição de frequências: Os estudantes, de modo geral calculam corretamente as frequências e compreendem bem as propriedades que se referem aos indivíduos de forma isolada, como por exemplo as notas obtidas numa prova, a altura etc. Entretanto, torna-se mais complexo para eles

compreender a ideia de distribuição desta propriedade no grupo que se estuda para fazer uma análise integral do fenômeno. Portanto, realizam uma “análise” isolada dos valores, sem ter em conta a relação entre os diferentes valores percentuais e destes com relação ao todo;

- Medidas de tendência central: Os estudantes geralmente dominam o algoritmo de cálculo. Entretanto, aplicam-no de forma mecânica sem compreender o significado do conceito. De modo específico verificou-se o seguinte:
  - a) Com relação à média, a maior parte a entende como o valor central, apresentam dificuldades de interpretação quando o resultado do seu cálculo não coincide com nenhum dos valores da série estatística e quando não tende a estar no centro da distribuição, acreditam que ela é sempre representativa de dados quantitativos independentemente das flutuações aleatórias, e ignoram o valor zero no seu cálculo.
  - b) Na aplicação da mediana, não têm em conta a variável em estudo, a ordenação dos dados e a natureza da série estatística (par ou ímpar). Com relação à moda, muitas vezes tomam o valor da frequência relativa em vez do valor da variável correspondente com a maior frequência relativa.
  - c) Possuem dificuldades em entender que o valor de uma medida de tendência central pode não ser o mesmo quando se trabalha de forma isolada e quando se trabalha com dados agrupados, assim como a seleção da medida de tendência central mais adequada numa determinada série de dados, em função do tipo de variável e a sua escala de medição.
- Medidas de dispersão: Os estudantes possuem dificuldades na interpretação do conceito de variabilidade. A sua compreensão baseia-se na falta de semelhança entre os dados, ou seja na diferença ou distância entre os dados e não com relação a uma medida estatística em específico.
- Estudo da associação: Os estudantes apresentam dificuldades na seleção do estatístico de correlação em função dos tipos de dados, na análise da tabela de contingência, na identificação do tipo de relação entre variáveis (direta, inversa ou a não relação), sempre estabelecem uma relação causa-efeito entre as variáveis sem chegar a refletir sobre a possibilidade de uma relação casual.

A partir do diagnóstico aplicado, deduz-se que o conhecimento do algoritmo pelos estudantes não implica necessariamente uma compreensão adequada do conceito a ele subjacente. É necessário que os estudantes antes da aplicação do algoritmo conheçam o significado do conceito a ele subjacente, suas características, vantagens e desvantagens da sua aplicação em cada caso particular.

Nesta perspectiva, considera-se que se deveria trabalhar com os estudantes aproveitando as suas ideias intuitivas e conduzi-los a utilizar novas estratégias de aprendizagem que lhes permitam obter uma melhor compreensão dos conceitos e procedimentos de cálculo afim de utilizá-los de forma consciente e mais acertada.

De acordo com as insuficiências identificadas na aprendizagem dos estudantes, os autores aplicaram uma estratégia de aprendizagem baseada no trabalho com situações da vida real, em cujo processo de resolução utilizou-se o método de busca parcial, no qual o professor coloca sucessivamente determinadas perguntas que paulatinamente vão conduzindo o estudante na obtenção do resultado final. Portanto, vai mais para lá da aplicação memorística de regras ou procedimentos de cálculo. Exige fazer uma integração de conhecimentos e reflexão sobre as possíveis vias de solução.

Desta forma se pode formar no estudante um sistema de conhecimentos, capacidades e habilidades, a fixação de conceitos e procedimentos, em combinação com a exercitação dos modos de atuação que devem aprender, o que lhes permitirá desenvolver um conhecimento conceitual e procedimental da disciplina.

### **3. METODOLOGIA**

Esta pesquisa tem como objetivo apresentar os resultados das experiências da aprendizagem adquirida pelos estudantes do curso de Matemática da Escola Superior Pedagógica do Bié (ESP-Bié), na disciplina de Estatística. A mesma adotou o tipo experimental. Segundo Arias (2012, p.37), a pesquisa experimental “consiste em submeter a um objeto ou grupo de indivíduos a determinadas condições o estímulos (variável independente), para observar os efeitos ou reações que produzem (variável dependente)”.

Para o efeito, selecionaram-se 45 estudantes do segundo ano de licenciatura, no curso do ensino da Matemática, da Escola Superior Pedagógica do Bié, de forma aleatória. Aplicou-se um pré-experimento, no qual os dados foram obtidos a partir da aplicação de dois testes (um pré-teste e um pós-teste).

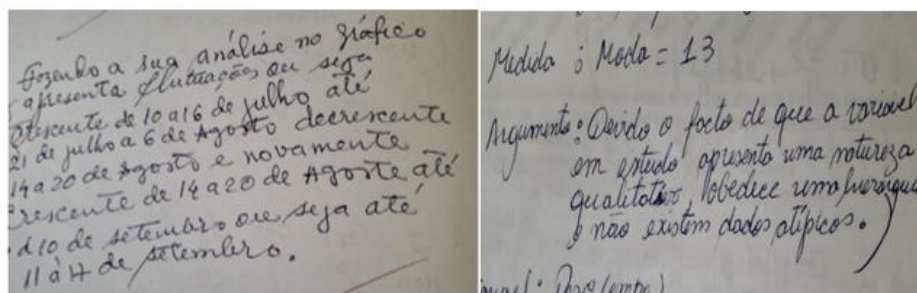
A partir dos resultados obtidos no pré-teste, aplicou-se uma nova estratégia de ensino, na qual o tratamento do novo conteúdo era acompanhado da sua aplicação sistemática no processamento de dados estatísticos em situações do cotidiano. Posteriormente foi aplicado um pós-teste e realizou-se a comparação do antes e depois. Para o efeito utilizaram-se os testes Mc. Nemar e o teste t de Student de comparação de uma amostra.

### 3. 1 Resultados e discussões

Para a medição do estado inicial da aprendizagem dos conteúdos, no pré-teste, colocaram-se 3 perguntas. Na primeira, os estudantes deveriam a partir de um conjunto de dados calcular a média e a mediana, assim como o desvio com respeito a estas medidas. Na segunda questão, apresentou-se um polígono de frequência que representava a tendência do preço do barril de petróleo de 2018 a 2020, e solicitou-se aos estudantes a análise do referido gráfico. Na última pergunta, colocaram-se os dados de duas variáveis estatísticas na qual os estudantes tinham que selecionar correctamente a medida estatística a utilizar, calcular o seu valor e analisar o resultado.

Posteriormente ao levantamento de dados, aplicou-se uma nova estratégia de ensino que teve como essência a aplicação dos conteúdos na análise de situações do cotidiano. Após aplicada a nova estratégia, os estudantes foram submetidos a um pós-teste no qual avaliou-se a análise de dados em tabelas de frequências de uma e duas entradas, a análise de dados pelas medidas de tendência central e de dispersão, a identificação e tratamento de dados outliers assim como a análise correlacional em situações do cotidiano.

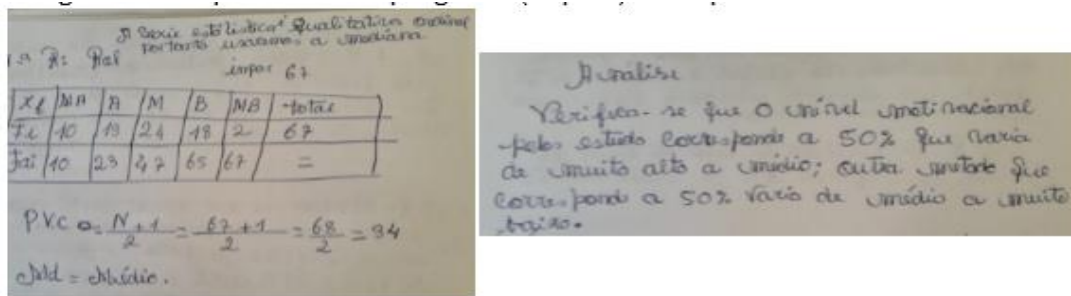
A avaliação do pré-teste e do pós-teste foi realizada na escala de 0 a 20 pontos (escala utilizada na ESP-Bié). Seguidamente apresentam-se alguns excertos das resoluções feitas pelos estudantes antes do experimento:



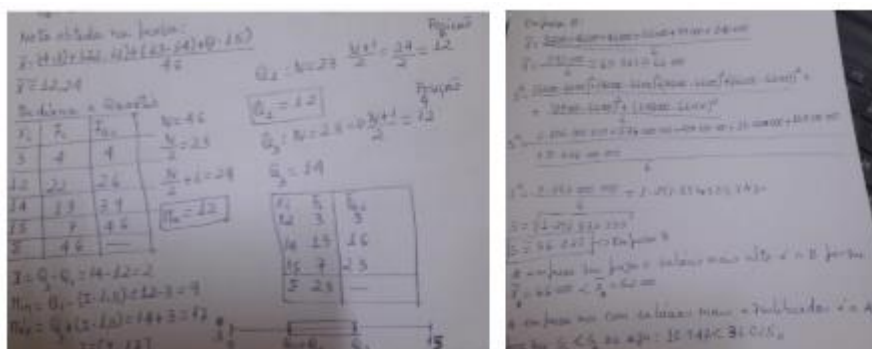
**Figura 1:** Resposta da segunda questão (primeira imagem), resposta da terceira questão (segunda imagem)

Fonte: Elaboração dos estudantes





**Figura 2:** Resposta da terceira questão depois do experimento (a)  
 Fonte: Elaboração dos estudantes



**Figura 3:** Resposta da terceira questão depois do experimento (b)  
 Fonte: Elaboração dos estudantes

Para a análise dos resultados antes e depois do experimento aplicaram-se as provas Mc Nemar e a *t* de Student.

A prova Mc. Nemar segundo Jacinto, Arnaez, Borges e Heredia (2011, p.34), utiliza-se “na pesquisa educativa para analisar o nível de significatividade da mudança de um indicador depois da intervenção na prática”. Nesta perspectiva, para a prova Mc Nemar considerou-se o indicador: utilização correcta dos conhecimentos das variáveis e do tipo de distribuição na análise dos dados, e realizou-se uma valorização das mudanças neste indicador depois do experimento.

Os resultados das observações do comportamento deste indicador antes e depois do experimento foram:

Total de estudantes que realizaram o pré-teste e o pós-teste: 45

Total de estudantes que não aplicavam corretamente os conhecimentos das variáveis e do tipo de distribuição na análise dos dados antes do experimento e o aplicavam corretamente depois: 31 (69%)

Total de estudantes que aplicavam corretamente os conhecimentos das variáveis e do tipo de distribuição na análise dos dados antes do experimento, e não o aplicavam corretamente depois: 2 (4%)

Total de estudantes que não aplicavam corretamente os conhecimentos das variáveis e do tipo de distribuição na análise dos dados nem antes e nem depois do experimento: 7 (16%)

Total de estudantes que aplicavam corretamente os conhecimentos das variáveis e do tipo de distribuição na análise dos dados antes do experimento e continuaram aplicando depois: 5 (11%)

Para analisar se as mudanças no indicador são significativas ou não, considerou-se:

Ho: A estratégia aplicada que não influencia na utilização dos conhecimentos das variáveis e do tipo de distribuição na análise dos dados;

Ha: A estratégia aplicada que influencia na utilização dos conhecimentos das variáveis e do tipo de distribuição na análise dos dados.

Para a definição da região de rejeição optou-se por  $\alpha = 0,01$ . Utilizaram-se os valores críticos de  $\chi^2$  com um grau de liberdade ( $df = 1$ ). Segundo a tabela de valores críticos de  $\chi^2$ , para  $df = 1$  e  $\alpha = 0,01$ ,  $\chi^2 = 6,64$ .

$$\chi^2 = \frac{(|f_D - f_{ND}| - 1)^2}{f_D + f_{ND}} \quad (1)$$

$$\chi^2 = \frac{(|31 - 2| - 1)^2}{31 + 2}$$

$$\chi^2 = 23,8$$

Como o valor de  $\chi^2$  calculado a partir dos registros é maior que  $\chi^2$  crítico, se rejeita a Ho. Logo, se pode inferir que o indicador teve mudanças significativas depois do experimento.

Aplicou-se também a prova de contraste  $t$  de Student para uma amostra. Morales (2010, p.49), considera que a prova  $t$  de Student “é um teste de contraste de hipóteses utilizado para a comparação de médias, e permite analisar se a diferença entre elas é significativa ou não”. Assim sendo, utilizou-se esta prova para a análise do nível de significatividade da diferença entre as médias antes e depois do experimento. Para o efeito, comprovou-se a normalidade das duas distribuições como pré-requisito para a aplicação do referido teste:

Considerou-se como hipóteses:

Ho: Os dados possuem uma distribuição normal.

Ha: Os dados possuem uma distribuição diferente da normal.

**Tabela 1: Testes de Normalidade**

	Kolmogorov-Smirnov <sup>a</sup>			Shapiro-Wilk		
	Estatística	Df	Sig.	Estatística	df	Sig.
Notas obtidas no pré-teste	0,129	45	0,058	0,944	45	0,031

a. Correção de significação de Lilliefors

Fonte: Elaboração própria

**Tabela 2: Testes de Normalidade**

	Kolmogorov-Smirnov <sup>a</sup>			Shapiro-Wilk		
	Estatística	Df	Sig.	Estatística	df	Sig.
Notas obtidas no pós-teste	0,096	45	0,200*	0,956	45	0,085

\*. Este é um limite inferior da significância verdadeira.

a. Correção de significação de Lilliefors

Fonte: Elaboração própria

Assumiu-se a prova de Kolmogorov-Smirnov, entendendo ser como a mais adequada para amostras pequenas ( $n < 100$ ). Como tanto no pré-teste quanto no pós teste,  $p > 0,05$ , logo considera-se que os dados têm uma distribuição normal.

Para a aplicação do teste  $t$  considerou-se as seguintes hipóteses:

Ho: Não existem diferenças significativas nas médias dos estudantes antes e depois do experimento.

Ha: Existem diferenças significativas nas médias dos estudantes antes e depois do experimento.

Para  $\alpha = 0,05$  obteve-se:

**Tabela 3: Resultados do teste  $t$  para uma amostra**

	Valor do teste = 7.9					
	T	Df	Sig. (2 extremidades)	Diferença média	95% Intervalo de confiança da diferença	
					Inferior	Superior
Notas obtidas no pós-teste	13,504	44	,000	4,6222	3,932	5,312

Verifica-se que  $p < 0,05$ . Logo, se dispensa a hipótese nula.

Os resultados da comparação dos dois testes indicam que existem diferenças significativas nas médias antes e depois da realização do pré-experimento. A partir dos resultados das duas provas (Mc Nemar e  $t$  de Student), verificou-se que houve uma melhoria significativa na aprendizagem dos estudantes antes e depois do experimento.

### **3.2. Exemplo de atividades realizadas pelos estudantes durante A experiência**

Para a aplicação da estratégia, dividiu-se os 45 estudantes em 3 grupos de 15 indivíduos cada. A cada grupo os professores atribuíram atividades divididas em 3 seções. Na primeira, os estudantes teriam que fazer a análise dos dados pela distribuição de frequências, na segunda pelas medidas de tendência central e de dispersão e na terceira pela análise correlacional.

Na segunda seção, os estudantes realizaram uma análise exploratória dos dados a partir de métodos gráficos (diagrama de caixas e pivôs, histograma com representação da curva normal) e numéricos (coeficiente de assimetria e amplitude interquartil para a determinação de dados outliers) e posteriormente decidiram sobre o modelo mais adequado para a análise dos dados e fizeram a respectiva interpretação.

Na terceira seção, fizeram a recolha de dados a partir de questionários, determinaram o problema de análise correlacional, com base aos dados, decidiram sobre o tipo de estatístico a utilizar e após o cálculo realizaram a análise do resultado obtido do contexto do problema. Identificaram o tipo de correlação e a intensidade da relação entre as variáveis.

Em todas as seções, durante o processo de resolução, foi utilizado o método de busca parcial, que consistiu na colocação de perguntas heurísticas que foram conduzindo os estudantes paulatinamente na busca de solução dos problemas.

Na resolução dos referidos problemas pelos estudantes, as questões colocadas pelos professores foram: Que dados possui?, qual é a natureza dos dados e a sua escala de medição? Os dados são muito diferentes uns dos outros? Que medidas estatísticas aplicar? Porquê aplicar determinada medida? Em que circunstâncias se aplica essa medida? Será esta medida a mais representativa dos dados? Estas questões, conduziram

os estudantes a uma reflexão sobre as características das medidas estatísticas e a sua representatividade ou não no contexto dos dados que o problema apresenta.

Durante a realização das atividades os estudantes utilizaram os meios de cômputo e em alguns casos o software SPSS e o microsoft Excel. Apresentam-se algumas atividades desenvolvidas nas seções 2 e 3 pelo grupo 2:

## Seção II: Medidas de tendência central

*Questão 1:* Nas seguintes séries de dados, identifique o aluno típico com respeito às variáveis idade e média.[...]

Objetivo: analisar os dados a partir dos conhecimentos de medidas de tendência central e de posição.

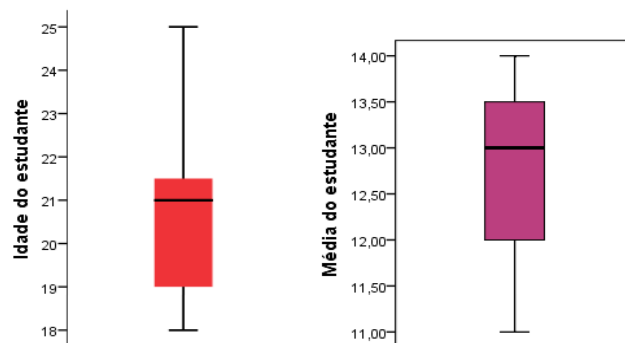
**Tabela 4:** Séries de dados das variáveis: idade e média

Idade				
18	21	22	21	21
21	18	20	18	21
20	23	25	23	18
Média				
13	12	11	14	13
14	13	13	13	14
13	12	12	14	12

Fonte: elaboração própria

Tratando-se de variáveis métricas, provavelmente a média seja o estatístico adequado para a sua análise. Entretanto, atendendo que esta medida é afetada por dados discrepantes (outliers) que podem inflacionar o seu resultado e conduzir a análises distorcidas, incentivou-se os alunos a fazer uma “análise exploratória dos dados”, por via gráfica e numérica para a identificação destes possíveis valores.

Para a via gráfica, os estudantes utilizaram o gráfico de caixas e pivôs, pois este representa a variação dos dados observados numa determinada variável, baseando-se nos quartis, e desta forma, permite facilmente identificar possíveis dados outliers, ou seja, dados atípicos que são mais baixos que o mínimo valor admissível para a série de dados ou mais altos que o máximo valor admissível, e que no caso de existirem, podem inflacionar a média tornando-a numa medida pouco confiável.



**Figura 4:** Gráfico de caixas para a idade e a média dos estudantes  
 Fonte: Elaboração dos estudantes

Na análise dos gráficos, os estudantes chegaram às seguintes conclusões:

Embora haja uma relativa dispersão entre os dados na caixa, em ambos os casos não existem dados outliers. Para a variável idade, o primeiro quartil ( $Q_1$ ) é 19, o segundo quartil ( $Q_2$  ou mediana) é 21 e o terceiro quartil ( $Q_3$ ), 21,5. Os valores extremos (máximo e mínimo), são 18 e 25 respectivamente.

Para a variável média, os valores para  $Q_1$ ,  $Q_2$  e  $Q_3$  são respectivamente: 12, 13 e 13,5. O valor mínimo é 11 e máximo 14. Considerando que a amplitude interquartil é a diferença entre o terceiro e primeiro quartil ( $I = Q_3 - Q_1$ ), para a idade tem-se:  $I = 2,5$ , para a média:  $I = 1,5$ . Calculando o intervalo em torno do qual se admite que os dados estejam dispostos :

Para a idade:

Limite inferior:

$$Q_1 - (1,5 * I) = 19 - (1,5 * 2,5) = 15,25$$

Limite superior:

$$Q_3 + (1,5I) = 21,5 + (1,5 * 2,5) = 25,25$$

Logo, o intervalo de tolerância para a idade é  $[[15,25; 25,25]]$ . Todo o dado fora deste intervalo é considerado um outlier. Como na série de dados os valores extremos são 18 e 25, logo, para esta variável, não existem valores outliers que poderão influenciar o estatístico média.

Analogamente, para a variável média, obteve-se o intervalo:  $[[9,75; 15,75]]$ . Como os valores extremos para esta variável (11 e 14), estão dentro do intervalo de tolerância, igualmente não existem dados outliers nesta série.

Feita esta análise preliminar, os estudantes consideraram que a média é adequada para a representação dos dados. Calculando a idade média obtiveram:

$$\bar{x} = \frac{\sum_{i=1}^n F_i * x_i}{\sum_{i=1}^n F_i} \quad (2)$$
$$\bar{x} = \frac{310}{15} \cong 21$$

Calculando o aproveitamento médio:

$$\bar{x} = \frac{\sum_{i=1}^n F_i * x_i}{\sum_{i=1}^n F_i}$$
$$\bar{x} = \frac{193}{15} \cong 13$$

A partir do estatístico “média”, os estudantes obtiveram a seguinte resposta: “(..) os alunos típicos são os que têm 21 anos e aproveitamento acadêmico (média) 13. Ou seja, os dados com respeito à idade estão ao redor de 21 e com respeito à média, ao redor de 13 valores”.

### **Seção III: Análise correlacional**

Para o desenvolvimento desta seção os estudantes aplicaram dois inquéritos e elaboraram dois problemas com dados que relacionam duas variáveis, com o objetivo de analisar a possível correlação entre elas. Foram advertidos sobre o fato de que a correlação não implica causalidade, ou seja uma relação causa-efeito posto que podem existir variáveis estranhas que estejam a influenciar a variável dependente.

Os estudantes foram orientados a seguir os seguintes passos:

- Definir o problema para a análise de correlação;
- Elegir a amostra de estudo de forma aleatória;
- Selecionar e elaborar as técnicas de recolha de dados: questionários, entrevistas, análise documental etc;
- Recopilar os dados;
- Elegir a medida estatística a ser utilizada;
- Analisar a informação para elaborar conclusões de forma correta.

Assim sendo, os estudantes aplicaram dois questionários, sendo um em 30 pacientes do hospital Walter Stranguay, no Cuito-Bié/Angola, e outro em 179 estudantes da ESP-Bié. Seguidamente formularam dois problemas (um para cada questionário), e analisaram a possível correlação entre as variáveis de cada problema, tal como se apresenta a seguir:

*Problema nº 1:* Recolheu-se os seguintes dados de 30 pacientes de forma aleatória. Determinar se existe uma relação entre o peso dos pacientes e as suas idades. Em caso positivo, classifique-a e represente-a num gráfico de regressão linear.

**Tabela 5:** Dados da idade e do peso de 30 pacientes

<b>Paciente</b>	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
<b>Idade</b>	45	35	38	40	39	31	37	45	42	39	46	38	47	40	31
<b>Peso</b>	67	51	53	68	57	43	49	57	53	49	54	59	69	61	48
<b>Paciente</b>	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
<b>Idade</b>	35	37	48	39	31	30	37	28	33	35	31	35	31	34	31
<b>Peso</b>	51	49	72	58	55	48	49	40	49	50	47	52	48	51	46

Fonte: Elaboração dos estudantes

Para a resolução deste problema, os professores conduziram os alunos a refletir sobre:

Que tipo de variáveis se trata no problema?

Qual é a variável independente e qual a dependente?

Em função do tipo de variável que medida estatística pode ser utilizada para analisar a possível relação entre as variáveis?

Existirá alguma reta que passa pelos pontos que relacionam as variáveis?

Que tipo de relação existe entre as variáveis?

A partir da reflexão sobre as questões colocadas, os estudantes chegaram à seguintes considerações:

*“(...) O problema aborda as variáveis idade e peso, todas elas quantitativas. Como o peso varia em função da idade, logo a idade é a variável independente e o peso a variável dependente. A distribuição apresenta 30 dados pelo que deduz-se que apresenta uma distribuição aproximadamente normal. Uma vez que as variáveis são quantitativas, ambas medidas em escalas métricas, utilizou-se o coeficiente de correlação de Pearson para a análise da associação entre ambas.” (Estudantes do segundo ano da ESP-BIÉ, comunicação pessoal, 2021).*

Seguidamente, calcularam o coeficiente de Pearson e obtiveram:



$$r_{xy} = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{\sqrt{n \sum x_i^2 - (\sum x_i)^2} * \sqrt{n \sum y_i^2 - (\sum y_i)^2}} \quad (3)$$

$$r_{xy} = \frac{(30 * 60236) - (1109 * 1603)}{\sqrt{30 * 41859 - (1109)^2} * \sqrt{30 * 87379 - (1603)^2}} \cong 0,8$$

Para a classificação da força de associação utilizaram o critério determinado por Andrés e Castillo (2004):

$\pm 0,96$ ;  $\pm 1$  correlação linear perfeita

$\pm 0,85$ ;  $\pm 0,95$  correlação linear forte

$\pm 0,7$ ;  $\pm 0,84$  correlação linear significativa

$\pm 0,5$ ;  $\pm 0,69$  correlação linear moderada

$\pm 0,20$ ;  $\pm 0,49$  correlação linear baixa

$\pm 0,1$ ;  $\pm 0,19$  correlação linear muito baixa

$\pm 0,09$ ;  $\pm 0,0$  correlação linear nula.

No âmbito desta classificação, concluíram que:

*“(...) Como  $r_{xy} \cong 0,8$ , a correlação entre as variáveis é positiva forte. Ou seja, supõe-se que à medida em que aumenta a idade aumenta também o peso. Existe uma reta que passa pelos pontos que determinam a sua relação. Como a relação é linear, a equação da reta é dada pela expressão:  $y = ax + b$ , onde  $a$  é a interseção ou seja, o ponto em que a reta corta o eixo  $y$  e  $b$  é o coeficiente angular que determina a inclinação da reta. Se a correlação é positiva, os pontos estarão situados no 1º e 3º quadrantes.” (Estudantes do segundo ano da ESP-BIÉ, comunicação pessoal, 2021).*

Seguidamente, calcularam a equação da recta de regressão linear e representaram graficamente:

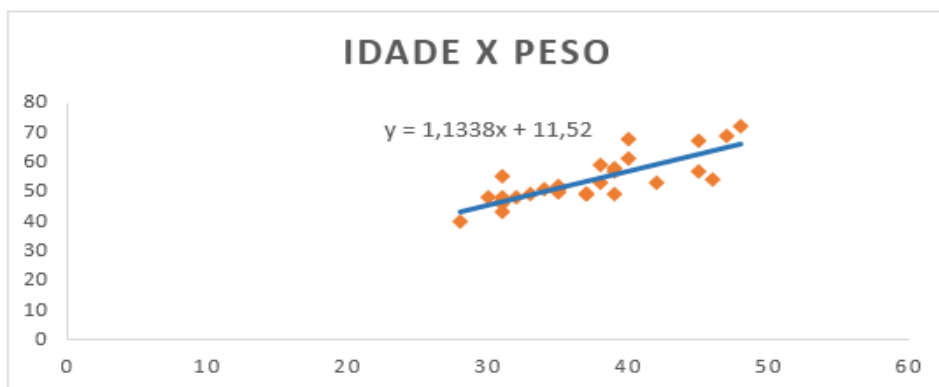
$$A = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{n \sum x_i^2 - (\sum x_i)^2} \quad (4)$$

$$A = \frac{(30 * 60236) - (1109 * 1603)}{(30 * 41859) - (1109)^2} \cong 1,1338$$

$$A = \frac{\sum y_i - A \sum x_i}{n} \quad (5)$$

$$A = \frac{1603 - (1,1338 * 1109)}{30} \cong 11,52$$

Logo, concluíram que a reta de regressão linear será:  $y = 1,13x - 11,52$  cujo gráfico é:



**Figura 5:** Resultado das questões resolvidas pelos estudantes  
Fonte: Elaboração dos estudantes

Os alunos deduziram que:

*“(...) se aumenta a idade, pode aumentar o peso. Entretanto, fez-se reflexões sobre outros fatores que podem provocar o aumento do peso como a dieta alimentar, o sedentarismo ou fatores genéticos. Ou seja, podem haver outras variáveis associadas ao aumento do peso.” (Estudantes do segundo ano da ESP-BIÉ, comunicação pessoal, 2021).*

**Problema nº 2:** A partir da seguinte tabela de contingência, analise a possível relação entre o gênero e o interesse pelo estudo da Matemática.

**Tabela 6:** gênero versus interesse pela Matemática

$i, j$	Interesse pelo estudo da Matemática			Total
	Muito	Pouco	Nenhum	
Feminino	18	34	41	93
Masculino	38	29	19	86
Total	56	63	60	179

Fonte: Elaboração dos estudantes

Para a resolução deste problema, os professores conduziram os estudantes a refletir sobre:

Que tipo de variáveis se trata no problema?

O que descreve a tabela de contingência?

Em função do tipo de variável que medida estatística pode ser utilizada para analisar a possível relação entre as variáveis?

Como determinar a significância da medida calculada?

A que conclusão se pode chegar com respeito à associação das variáveis?

Dado que as variáveis são qualitativas nominais e estão expressas numa tabela de contingência, os estudantes selecionaram o estatístico chi-quadrado ( $\chi^2$ ) para analisar a possível associação entre as variáveis.

$$\chi^2 = \sum_{i=1}^n \frac{(O - E)^2}{E} \quad (6)$$

Este estatístico descreve a soma do quadrado da diferença entre as frequências observadas e as correspondentes frequências esperadas, dividido pela respetiva frequência esperada. Nesta perspectiva, permite que se faça uma distribuição de frequências proporcional à quantidade de estudantes por género. Quanto menor for a diferença entre as frequências observadas e esperadas, mais garantia existirá de que a casualidade não tenha influenciado a frequência observada.

Para o cálculo das frequências esperadas obtiveram os seguintes resultados:

Cálculo das frequências esperadas:

$$E_{i,j} = \frac{n_{1*}n_j}{N} \quad (7)$$

Neste caso:

$$E_{i,j} = \frac{93 * 56}{179} = 29$$

Analogamente para as demais frequências esperadas obtiveram como resultados:

$$E_{12} = 33; E_{13} = 31; E_{21} = 27; E_{22} = 30; E_{23} = 29$$

Os professores chamaram atenção para o fato de a prova ser pouco confiável quando existem frequências esperadas inferiores a 5. Logo, os estudantes observaram que:

*“(...) não há nenhuma frequência esperada menor que 5, pelo que se pode confiar nos resultados da prova.” (Estudantes do segundo ano da ESP-BIÉ, comunicação pessoal, 2021).*

A partir das frequências observadas e esperadas geraram a tabela do quociente entre o quadrado da diferença de frequências e a frequência esperada ( $\frac{(O-E)^2}{E}$ ), e o obtiveram os seguintes resultados:

**Tabela 6:** diferença de frequências observadas e esperadas

$i; j$	Interesse pelo estudo da Matemática		
Género	Muito	Pouco	Nenhum
Feminino	$\frac{(18 - 29)^2}{29} = 4,2$	$\frac{(34 - 33)^2}{33} = 0,03$	$\frac{(41 - 31)^2}{31} = 3,23$
Masculino	$\frac{(38 - 27)^2}{27} = 4,5$	$\frac{(29 - 30)^2}{30} = 0,033$	$\frac{(19 - 29)^2}{29} = 3,45$

Fonte: Elaboração dos estudantes

Calculando o valor de chi-quadrado, obtiveram como resultado:

$$\chi^2 = \sum_{i=1}^n \frac{(O - E)^2}{E}$$

$$\chi^2 = 4,2 + 0,03 + 3,23 + 4,5 + 0,033 + 3,45$$

$$\chi^2 = 15,443$$

Determinação de hipóteses:

Ho: O género e o interesse pelo estudo da matemática são variáveis independentes;

Ha: O género e o interesse pelo estudo da matemática não são variáveis independentes.

Assumiram  $\alpha = 0,05$ , para a determinação do valor crítico de  $\chi^2$ . Como se trata de uma tabela do tipo  $3 * 2$ , para a determinação dos graus de liberdade obtiveram:

$(3 - 1) * (2 - 1) = 2$ , logo, a prova tem 2 graus de liberdade. Para  $\alpha = 0,05$  e com 2 graus de liberdade, a partir da tabela de valores críticos de  $\chi^2$  obtiveram como resultado: 5,9915. Para a tomada da decisão, compararam o valor crítico de  $\chi^2$  com o valor calculado e concluíram que:

Se o valor calculado for menor que o valor crítico de  $\chi^2$  então aceita-se a hipótese nula. Como  $\chi^2$  calculado é maior que o valor crítico ( $15,443 > 5,9915$ ), então, se dispensa a hipótese nula. Logo, há grande probabilidade de que as duas variáveis não sejam independentes.

Entretanto, o fato de o resultado apontar para uma possível relação entre as variáveis, os alunos fizeram uma análise crítica, pois, o género não determina a predisposição para o estudo de uma disciplina, pelo que consireraram que:

*“(...) Apesar de que a amostra foi selecionada aleatoriamente, os resultados podem ter sido influenciados pelas concepções que os alunos da amostra possuem sobre a*

*matemática, a metodologia utilizada pelos professores desses alunos, ou outros fatores. Portanto, não se pode estabelecer uma relação causa-efeito.” (Estudantes do segundo ano da ESP-BIÉ, comunicação pessoal, 2021).*

Os resultados ora apresentados evidenciam a independência cognitiva e criatividade dos estudantes na aplicação dos conhecimentos em situações do cotidiano; o estímulo do pensamento reflexivo e criativo dos estudantes na formulação e resolução de problemas assim como a capacidade de aplicar o aprendido de forma independente e criativa na análise de situações do seu cotidiano;

A integração de diferentes métodos estatísticos na análise de problemas permitiu aos alunos valorar de forma crítica a aplicabilidade de determinados conceitos estatísticos em função dos tipos de variáveis e das características das distribuições

#### **4. CONSIDERAÇÕES FINAIS**

A pesquisa teve como objetivo descrever os resultados das experiências de aprendizagem adquiridas pelos estudantes do curso de Matemática ESP-BIÉ, na disciplina de Estatística. Os resultados obtidos nos testes realizados antes e depois da aplicação da estratégia revelaram diferenças significativas que indicam uma melhor compreensão dos conhecimentos estatísticos e sua aplicação em situações novas.

A aplicação da estratégia permitiu identificar resultados qualitativamente superiores na aprendizagem dos alunos, dos quais se destacam: o trabalho com diferentes tipos de dados que permitiu a valoração crítica de diferentes métodos estatísticos para explorar o comportamento dos dados do problema; a interpretação e valoração dos resultados obtidos com a utilização das diferentes medidas estatísticas; a fixação dos conceitos estatísticos a partir da interpretação do seu significado e das vantagens e desvantagens do seu uso.

#### **REFERÊNCIAS**

- Álvarez, J. C. (2012). *Concepción teórico-metodológica del proceso de enseñanza-aprendizaje de la unidad curricular análisis del dato estadístico*. Tese de doutoramento. Instituto Pedagógico Latinoamericano y Caribeño. La Habana, Cuba.
- André, A. L. (2018). *Representação e interpretação de dados e medidas estatísticas por alunos do 3º ano do curso de matemática da Escola Superior Pedagógica do Bié*. Dissertação de mestrado. Universidade de Lisboa. Lisboa.

- Andrés, M. e Castillo, L. (2004). *Bioestadística*. Barcelona, Espanha. Editorial norma.
- Arias, F. G. (2012). *El proyecto de investigación. Introducción a la metodología científica. (6ta edición)*. Caracas, Venezuela. Editorial Episteme.
- Cadima, S. (2013). *Tratamento de dados no 7.º ano de escolaridade: Medidas estatísticas. (Relatório de mestrado)*. Universidade de Lisboa. Portugal
- Batanero, C., e Godino, J. D. (2001). *Análisis de datos y su didáctica*. Departamento de Didáctica de la Matemática, Universidad de Granada. Recuperado de [https://www.researchgate.net/publication/273453581 Analisis de datos y su didactica](https://www.researchgate.net/publication/273453581_Analisis_de_datos_y_su_didactica).
- Cassinela, E. N. (2016). *Modelo didáctico para el proceso de enseñanza y aprendizaje de la Estadística Descriptiva en la formación de profesores de Geografía en el ISCED-Huambo*. (Tese de doutoramento). Universidad de ciencias pedagógicas Enrique José Varona. La Habana, Cuba.
- Jacinto, O. P.; Borges, T. C; Arnaez, I.; Heredia, R. H. (2011). *Los diseños estadísticos en la investigación pedagógica*. Editorial educación cubana. La Habana, Cuba.
- Li, K. Y. e Shen, S. M. (1992). Students' weaknesses in statistical projects. *Teaching Statistics*, 14 (1), 2-8.
- Martín, P. J. (2006). *Apuntes para el tratamiento estadístico de datos*. Madrid: Editorial ISCI.Espanha
- Merino, B. C. (2003). *Significado de las medidas de posición central para los estudiantes de la secundaria*. (Tese de doutoramento). Universidade de Granada. Espanha.
- Morales, E. E. (2010). *La estadística. Herramienta fundamental en la investigación pedagógica*. (2da edición). La Habana, Cuba. Editorial Pueblo y Educación.
- Pollasek, A., Lina, S. y Well, A. D. (1981). *Concept or computation: Students' understanding of the mean*. *Educational Studies in Mathematics*, 12, 191-204.

## NOTAS

### TÍTULO DA OBRA

A aprendizagem da Estatística na ESP-BIÉ. Uma reflexão a partir da experiência com os estudantes do curso de Matemática

#### Rosa Maria de Nascimento

Doutora em Educação Matemática na especialidade do Ensino da Estatística (UCPJV Havana-Cuba) Professora do Curso de Licenciatura em Matemática e em Física da Escola Superior Pedagógica do Bié (ESPB), Cuito/Bié-Angola

[rnascimento22@hotmail.com](mailto:rnascimento22@hotmail.com)

<https://orcid.org/0000-0001-7836-7441>

#### Ezequias Adolfo Domingas Cassela

Mestre em Matemática para professores pela Universidade da Beira Interior Portugal, Doutorando em Educação Matemática pela Pontifícia Universidade Católica de São Paulo, docente da Escola Superior Pedagógica do Bié, Departamento de ciências Exatas, sector de Matemática, Cuito-Bié, Angola

[ezequiasadolfo@hotmail.com](mailto:ezequiasadolfo@hotmail.com)

<https://orcid.org/0000-0001-7703-0097>



**Endereço de correspondência do principal autor**

Centralidade Horizonte do Cuito, Quadra 31, prédio 9, apt. 2.2., Angola.

**AGRADECIMENTOS**

A Deus pai todo poderoso pelas imensuráveis bençãos, a Direção da Escola Superior Pedagógica do Bié, na pessoa do Diretor Geral, PhD. Alfredo Maria de Jesus Paulo, a PhD. Rosa Maria de Nascimento pelo apoio e contributo.

**CONTRIBUIÇÃO DE AUTORIA**

**Concepção e elaboração do manuscrito:** R. M. De Nascimento ,E. A. D. Cassela

**Coleta de dados:** R. M. De Nascimento

**Análise de dados:** R. M. De Nascimento, E.A.D. Cassela

**Revisão bibliográfica:** R. M. De Nascimento, R. M. De Nascimento

**Revisão e aprovação:** R. M. De Nascimento

**CONJUNTO DE DADOS DE PESQUISA**

O conjunto de dados que dá suporte aos resultados deste estudo não está disponível publicamente.

**FINANCIAMENTO**

Não se aplica.

**CONSENTIMENTO DE USO DE IMAGEM**

Não se aplica.

**APROVAÇÃO DE COMITÊ DE ÉTICA EM PESQUISA**

Não se aplica.

**CONFLITO DE INTERESSES**

Não se aplica.

**LICENÇA DE USO** – uso exclusivo da revista

Os autores cedem à **Revemat** os direitos exclusivos de primeira publicação, com o trabalho simultaneamente licenciado sob a [Licença Creative Commons Attribution](#) (CC BY) 4.0 International. Esta licença permite que **terceiros** remixem, adaptem e criem a partir do trabalho publicado, atribuindo o devido crédito de autoria e publicação inicial neste periódico. Os **autores** têm autorização para assumir contratos adicionais separadamente, para distribuição não exclusiva da versão do trabalho publicada neste periódico (ex.: publicar em repositório institucional, em site pessoal, publicar uma tradução, ou como capítulo de livro), com reconhecimento de autoria e publicação inicial neste periódico.

**PUBLISHER** – uso exclusivo da revista

Universidade Federal de Santa Catarina. Grupo de Pesquisa em Epistemologia e Ensino de Matemática (GPEEM). Publicação no [Portal de Periódicos UFSC](#). As ideias expressadas neste artigo são de responsabilidade de seus autores, não representando, necessariamente, a opinião dos editores ou da universidade.

**EDITOR** – uso exclusivo da revista

Mérciles Thadeu Moretti e Rosilene Beatriz Machado.

**HISTÓRICO** – uso exclusivo da revista

Recebido em: 12-02-2021 – Aprovado em: 22-06-2021

