

UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO ACADÊMICO DO AGRESTE
NÚCLEO DE FORMAÇÃO DOCENTE
CURSO DE MATEMÁTICA – LICENCIATURA

ANDERSON MAIKE NASCIMENTO DE MEDEIROS

**CONHECIMENTOS DE FUTUROS PROFESSORES DE MATEMÁTICA SOBRE
ASSOCIAÇÃO DE VARIÁVEIS EM TABELAS DE CONTINGÊNCIA**

CARUARU, 2018

ANDERSON MAIKE NASCIMENTO DE MEDEIROS

**CONHECIMENTOS DE FUTUROS PROFESSORES DE MATEMÁTICA SOBRE
ASSOCIAÇÃO DE VARIÁVEIS EM TABELAS DE CONTINGÊNCIA**

Trabalho de Conclusão de Curso, apresentado(a) ao Curso de Graduação em Matemática - Licenciatura da Universidade Federal de Pernambuco, como requisito parcial para a graduação em Licenciatura em Matemática.

Área de concentração: Ensino/ Matemática

Orientador: Prof.º Dr.º José Ivanildo Felisberto de Carvalho

Catálogo na fonte:
Bibliotecária – Simone Xavier CRB/4-1242

M488c Medeiros, Anderson Maíke Nascimento de.
Conhecimento de futuros professores de matemática sobre associação de variáveis em tabelas de contingência. / Anderson Maíke Nascimento de Medeiros. – 2018.
109f. ; il. : 30 cm.

Orientador: José Ivanildo Felisberto de Carvalho.
Monografia (Trabalho de Conclusão de Curso) – Universidade Federal de Pernambuco, CAA, Licenciatura em Matemática, 2018.
Inclui Referências.

1. Probabilidade. 2. Tabelas de contingência. 3. Matemática – Estudo e ensino. 4. Estatística. I. Carvalho, José Ivanildo Felisberto de (Orientador). II. Título.

371.12 CDD (23. ed.)

UFPE (CAA 2018-206)

ANDERSON MAIKE NASCIMENTO DE MEDEIROS

**CONHECIMENTOS DE FUTUROS PROFESSORES DE MATEMÁTICA SOBRE
ASSOCIAÇÃO DE VARIÁVEIS EM TABELAS DE CONTINGÊNCIA**

Monografia apresentada ao Corpo Docente do Curso de MATEMÁTICA – Licenciatura do Centro Acadêmico do Agreste da Universidade Federal de Pernambuco, como requisito parcial para a obtenção do título de Licenciando em Matemática.

Aprovado em: ___/___/_____.

BANCA EXAMINADORA:

Prof. Dr. José Ivanildo Felisberto de Carvalho (Orientador)
Universidade Federal de Pernambuco

Prof. Dra. Jaqueline Aparecida Foratto Lixandrão Santos (Examinadora Interna)
Universidade Federal de Pernambuco

Prof. Me. César Diogo Bezerra da Silva (Examinador Externo)
Universidade Federal de Pernambuco

AGRADECIMENTOS

Quando paro e penso em toda essa trajetória do ingresso a universidade até a elaboração desta monografia, só tenho de agradecer.

Primeiramente quero agradecer a Deus por ter me dado forças para concluir o meu trabalho de conclusão de curso, pois sempre foi um desafio para mim escrever, mas com ajuda de Deus e de algumas pessoas próximas consegui finalizar minha monografia.

Vivenciei vários momentos difíceis que muitas vezes pensei em desistir, mas Deus me deu forças para continuar. E como recompensa obtive o trabalho aprovado. Em seguida venho agradecer ao meu orientador Prof. Dr. José Ivanildo Felisberto de Carvalho que me aceitou como orientando e me guiou no decorrer dessa caminhada, no qual contribuiu e me apoiou.

Quero agradecer também a Prof. Doutora Cristiane de Arimatéa Rocha que me ajudou bastante com suas contribuições e apoio motivacional.

Também quero agradecer a banca examinadora (Prof. Dra. Jaqueline Lixandrão Santos e o Doutorando César Diogo) que aceitaram participar desta defesa pública, no qual contribuíram bastante com ideias para aprimoração do meu trabalho.

Agradeço também aos demais professores da UFPE-CAA que tiveram papel fundamental na minha formação, especialmente Valdir Bezerra, Elizabeth Lacerda, Simone Queiroz, Marcos Henrique e tanto outros.

Agora quero externar meus agradecimentos a uma pessoa muito especial na minha vida e também neste trabalho Adeliane Mikaely que esteve do meu lado me apoiando e me dando forças nos momentos mais difíceis dessa jornada e aos meus pais Maria José Nascimento de Medeiros (minha mãe) e José Adilson Bezerra de Medeiros (meu pai) e aos meus irmãos David Henrique Nascimento de Medeiros e Averton Mayveson Nascimento de Medeiros e aos demais familiares.

Quero deixar também meus agradecimentos a minha honrosa Avó Severina Medeiros que me deixou cedo, mas foi umas das grandes incentivadoras da carreira de professor.

Por fim quero agradecer aos amigos que tiveram papel importante durante a graduação em especial aos meus amigos de turma Monalisa Melo, Karolina Araújo, Marina Carvalho, Thiago Gomes, Mazinho Rocha, Anyla Laíse, Letícia Nascimento, Lazaro Rangel, Leonardo Da Silva Chalegre, enfim a todos que fizeram parte da minha turma. Agradeço também aos amigos que fiz nos corredores da UFPE, de algumas outras disciplinas e aos meus amigos do ônibus, que vivenciaram a minha rotina durante a graduação.

RESUMO

As tabelas de contingência vêm se mostrando uma ferramenta importantíssima no cotidiano, pois elas estão presentes em diversas fontes midiáticas expressando informações que costumam ser relevantes ou não para quem interpreta esses dados. A associação de variáveis é uma das formas mais adequadas para interpretar se as informações expressas na tabela são relevantes ou não. Seguindo esta ideia, foi proposto neste trabalho realizar uma pesquisa que teve como principal objetivo analisar os conhecimentos de futuros professores de matemática sobre associação de variáveis em tabelas de contingência. Para isso, elaboramos um questionário que contempla quatro situações problemas envolvendo associação de variáveis em tabelas de contingência. Essas questões traziam três situações distintas: uma em que os dados da tabela não condiziam com a associação empírica, outra sobre a não existência de associação empírica entre as variáveis e por fim a concordância dos dados da tabela com a associação empírica. O questionário foi aplicado em dois grupos de alunos (os que cursaram a disciplina de Estatística e os que não cursaram). O intuito de aplicamos em dois grupos foi o de confrontar os resultados dos dois e ver se a formação do curso de Matemática-Licenciatura vem contribuindo para o ensino de associação de variáveis. Tivemos resultados satisfatórios no geral, ao qual, a média de acertos foi bastante alta. Concluimos também que pouco influenciou ter cursado a disciplina de Estatística ou não, pois os desempenhos foram bem páreos nos dois grupos de alunos. Quando fizemos uma análise dos dados por item percebemos que no primeiro item, ao qual, os dados não condiziam com a associação empírica, que o grupo dos alunos que não cursou a disciplina teve um desempenho superior aos que cursaram a disciplina. Numa análise das principais estratégias notamos que a maioria dos erros cometidos pelos estudantes foi por olharem apenas a célula de maior valor na tabela. Um outro resultado importante dessa pesquisa é que alguns alunos interpretam os dados da tabela e responderam a este item através de suas crenças. Nos demais itens obtivemos um índice de acertos bastante alto nos dois grupos.

Palavras-chave: Associação de variáveis. Tabelas de contingência. Probabilidade. Matemática. Estatística.

RESUMEN

Las tablas de contingencia se vienen mostrando una herramienta importantísima en el cotidiano, pues ellas están presentes en diversas fuentes mediáticas expresando informaciones que suelen ser relevantes para quien interpreta esos datos. La comprensión de la asociación de variables es una forma importante de interpretar sobre el significado de las asociaciones y la toma de decisiones en el cotidiano. En este trabajo se realizó una investigación que tuvo como principal objetivo analizar los conocimientos de futuros profesores de matemáticas sobre asociación de variables en tablas de contingencia. Para ello, elaboramos un cuestionario que contempla cuatro situaciones problemas involucrando asociación de variables en tablas de contingencia. Estas cuestiones abordaban tres situaciones distintas: una en que los datos de la tabla no concatan con la asociación empírica, otra sobre la no existencia de asociación empírica entre las variables y por fin, dos en que había concordancia de los datos de la tabla con la asociación empírica. El cuestionario fue aplicado en dos grupos de estudiantes (los que cursaron la disciplina de Estadística y los que no cursaron). La intención de aplicar en dos grupos fue el de confrontar los resultados de los dos y ver si la participación en una disciplina de estadística es factor necesario para la comprensión sobre la asociación de variables. Hemos tenido resultados satisfactorios en general, al que el promedio de aciertos fue bastante alto. Concluimos también que poco influenció haber cursado la disciplina de Estadística o no, pues los desempeños fueron bien a los niveles en los dos grupos de estudiantes. En el primer ítem, al que los datos presentados en la tabla no coinciden con la asociación empírica, el grupo de estudiantes que no que no cursó la disciplina obtuvo un desempeño superior a los que cursaron la disciplina. En un análisis de las principales estrategias en todos los cuatro elementos notamos que la mayoría de los errores cometidos por los estudiantes fue tomar la decisión sobre la asociación considerando sólo la celda de mayor valor en la tabla. Otro resultado importante de esta investigación es que algunos alumnos interpretan los datos de la tabla y respondieron sobre las asociaciones teniendo en cuenta sus creencias en lugar de los datos estadísticos presentados.

Palabras clave: Asociación de variables; Tablas de contingencia; Probabilidad: Estadística; Formación de profesores.

LISTA DE GRÁFICOS

Gráfico 1 - Número de questões acertadas por estudantes.	57
Gráfico 2 - Distribuição das pessoas com/sem a doença brônquica com relação as pessoas que têm/não o hábito de fumar.	64
Gráfico 3 - Distribuição das pessoas que têm o hábito de fumar em relação ao número de pessoas com doença brônquica.	66
Gráfico 4 - Distribuição dos estudantes aprovados e reprovados no exame em relação com o número de horas dedicadas por cada estudante para estudar para o exame.	98
Gráfico 5 - Distribuição do número de horas dedicadas por cada estudante para estudar para o exame em relação com o número de estudantes aprovados e reprovados no exame.	100

LISTA DE TABELAS

Tabela 1 - Tabela de contingência 2x2	22
Tabela 2 - Tabela de probabilidade conjunta	26
Tabela 3 - Exemplo de comparação de risco por meio da razão de chances	44
Tabela 4 - Variáveis que foram levadas em conta na elaboração do questionário	53
Tabela 5 - Estratégias utilizadas pelos os licenciandos na resolução dos 4 itens.	58
Tabela 6 - Tabela do item 1 do questionário	62
Tabela 7 - Estratégias utilizadas pelos os estudantes para responder o item 1.	68
Tabela 8 - Estratégias utilizadas pelos os estudantes para responder o item 2.	81
Tabela 9 - Estratégias utilizadas pelos os estudantes para responder o item 3.	93
Tabela 10 - Tabela do item 4 das situações problemas do questionário	96
Tabela 11 - Frequência absolutas com suas respectivas porcentagens calculadas em relação a linha marginal do item 4.	99
Tabela 12 - Estratégias utilizadas pelos os estudantes para responder o item 4.	102

LISTA DE FIGURAS

Figura 1 - exemplo de uma tabela de contingência	27
Figura 2 - item 1 do estudo de Cañadas, Batanero, Contreras e Arteaga (2010).	36
Figura 3 - Grau de intensidade do relacionamento de coeficiente de correlação positivos e negativos	39
Figura 4 - item 1 das situações problemas do questionário.	62
Figura 5 - Protocolo do item 1, aluno 18.	70
Figura 6 - Protocolo do item 1, aluno 06.	71
Figura 7 - Protocolo do item 1, aluno 07	72
Figura 8 - Protocolo do item 1, aluno 32.	73
Figura 9 - Protocolo do item 1, aluno 17.	75
Figura 10 - item 2 das situações problemas do questionário.	77
Figura 11 - Protocolo do item 2, aluno 19.	82
Figura 12 - Protocolo do item 2, aluno 22.	83
Figura 13 - Protocolo do item 2, aluno 18.	84
Figura 14 - Protocolo do item 2, aluno 30.	86
Figura 15 - Protocolo do item 2, aluno 31.	87
Figura 16 - Protocolo do item 2, aluno 28.	88
Figura 17 - Protocolo do item 2, aluno 7.	89
Figura 18 - item 3 das situações problemas do questionário	91
Figura 19 - Protocolo do item 3, aluno 19.	94
Figura 20 - Protocolo do item 3, aluno 8.	95
Figura 21- Item 4 das situações problema do questionário	96
Figura 22 - Protocolo do item 4, aluno 6.	103
Figura 23 - Protocolo do item 4, aluno 32.	104

SUMÁRIO

1	INTRODUÇÃO	11
1.1	Objetivo geral	14
1.1.1	<i>Objetivos específicos</i>	14
2	O QUE OS ESTUDOS APONTAM SOBRE O ENSINO DE ESTATÍSTICA E PROBABILIDADE NA FORMAÇÃO DO FUTURO PROFESSOR DE MATEMÁTICA	16
3	UMA BREVE DISCUSSÃO SOBRE INVESTIGAÇÕES COM PROBABILIDADE CONDICIONAL E ASSOCIAÇÃO DE VARIÁVEIS	21
4	O QUE A BASE NACIONAL COMUM CURRICULAR APONTA SOBRE ASSOCIAÇÃO DE VARIÁVEIS EM TABELAS DE CONTINGÊNCIA?	46
5	MÉTODO	49
6	ANÁLISES E DISCUSSÃO DOS RESULTADOS	56
7	CONSIDERAÇÕES FINAIS	106
	REFERÊNCIAS	108

1 INTRODUÇÃO

A probabilidade para muitos é conhecida como a “matemática da incerteza”. Essa fama se deu a partir de sua origem, que se iniciou através de jogos de azar, tais como, jogos de cartas, roletas e dados. Matemáticos famosos como o Girolamo Cardano (1501-1576), Blaise Pascal (1623-1662), Pierre de Fermat (1601-1665), Christiaan Huygens (1629-1695), dentre outros, desenvolveram vários estudos sobre jogos de azar, esses estudos foram o marco do início do que hoje é a Probabilidade.

Diante desta “fama” da matemática da incerteza surge a minha vontade pessoal de pesquisar sobre o referido tema, ainda nos primeiros períodos da minha graduação de Matemática-Licenciatura. Essa vontade passa ser mais presente em minha vida acadêmica ao cursar a disciplina de Estatística ofertada no início do curso, para ser mais preciso no segundo período do curso de Matemática-Licenciatura da Universidade Federal de Pernambuco (UFPE). A disciplina teve três grandes professores, a saber: Prof. Dr. Valdir Bezerra dos Santos Junior, Prof. Dr. Edelweis José Tavares Barbosa e Prof. Dr. José Ivanildo Felisberto de Carvalho, no qual, os três conceituaram as aulas explicando o assunto sem se apegar ao simples uso e decoração das fórmulas.

Dentre essa vontade pessoal, fiz algumas pesquisas em estudos realizados em diversas fontes distintas como livros e internet, nos quais, observei que há um consenso que a probabilidade é um ramo da matemática que permite calcular a chance de ocorrência possíveis em um experimento aleatório (fenômeno aleatório).

Perante a diversas dificuldades que perpasssei durante a graduação referente a estatística e probabilidade, me questioneei como estou saindo da academia referente aos conteúdos de probabilidade. E que conhecimentos adquiri durante a graduação? Dessas indagações surge a minha sede de buscar conhecer esse assunto.

É comum que a maioria dos livros didáticos e sites de estudos encontrados na internet abordem a probabilidade com a ideia de equiprobabilidade, que é mais conhecida por probabilidade clássica ou laplaciana. Essa definição em termos matemáticos nos diz que a probabilidade de um evento acontecer pode ser calculada como a razão entre o número de casos favoráveis para o acontecimento desse evento sobre o número total de casos possíveis.

$$P(A) = \frac{n(A)}{n(S)}$$

Onde, $P(A)$ é a probabilidade de acontecer um evento A , $n(A)$ é o número de casos favoráveis ao evento A acontecer e $n(S)$ é o número total de casos possíveis ou espaço amostral.

Então, iniciei esta pesquisa no intuito de entender melhor o conteúdo de probabilidade e por todos os motivos, ao qual, já mencionei anteriormente. Durante a pesquisa me deparei com problemas e situações que a probabilidade clássica por si só não é mais suficiente para resolver e explicar esses problemas do ponto de vista matemático.

No período seguinte ao qual iniciei esta pesquisa, surgiu a oportunidade de aprofundar os conhecimentos sobre essa temática. Foi ofertada na grade curricular do curso de Matemática-Licenciatura a disciplina de “Ensino e Aprendizagem de Estatística, Probabilidade e Risco, Perspectivas Teórico-Práticas” como requisito eletiva, ministrada pelo Prof. Dr. José Ivanildo Felisberto de Carvalho.

Ao cursar a disciplina mencionada, conheci diversos significados probabilísticos, entres esses o laplaciano/clássico, geométrico, frequentista, subjetivo, formal/axiomático, e também discussões sobre a ideia de casualidade e risco. É importante destacar que todos esses significados possuem suas vantagens e suas limitações, como é mencionado e exemplificado nas pesquisas de Amâncio (2012), Silva (2015) e Carvalho (2017). Por exemplo, no significado clássico temos que a probabilidade é uma razão que fornecerá um valor entre 0 e 1, que em outras palavras mede as chances de um evento acontecer. Sua desvantagem é que podemos nos deparar com situações em que o número de casos possíveis seja infinito ou ainda com situações em que o evento não seja equiprovável.

Para salientar ao leitor consideramos que um evento elementar é equiprovável quando todos os casos possíveis são igualmente prováveis, por exemplo, suponha que em uma urna são colocadas 3 bolas pretas e 2 brancas. O evento cor não é equiprovável, no entanto, as bolas P_1, P_2, P_3, B_1 e B_2 possuem a mesma probabilidade.

Quando essas situações ocorrem nos deparamos com problema de probabilidade com enfoques distintos. Por isso defendemos a ideia de Shulman (1986, 1987), sobre o conhecimento especializado do conteúdo e o conhecimento pedagógico do conteúdo. O professor que está em sala de aula deve conhecer a epistemologia do conhecimento que está habilitado; neste sentido a formação inicial do professor de matemática deve contemplar os conhecimentos necessários ao ensino da matemática, em nosso caso, da probabilidade.

Diante a essa vontade pessoal de pesquisar sobre o referido tema, surgem duas justificativas acadêmicas relativas ao campo da pesquisa, uma foi que a probabilidade para muito dos estudantes na Educação Básica e até mesmo no Ensino Superior é tida como de difícil

compreensão fazendo com que os estudantes não demonstrem interesse em resolver problemas que envolva probabilidade ou dados estatísticos.

Costa e Nacarato (2011), nos colocam que essa dificuldade ocorre pois a matemática e a estocástica¹, se diferem. Na estocástica, dados fornecidos em um problema não são apenas números, são números com um contexto inserido. E isso requer que os alunos tenham um pensamento estocástico para resolver esse tipo de problemas.

Essa difícil compreensão pode também estar atrelada com o que Carvalho (2017, p. 16), colocar:

Temos como pressuposto que o viés determinista da matemática se torna um entrave para o trabalho do professor com a probabilidade. Existe nos professores uma ideia sobre a Matemática alicerçada numa concepção platônica, com uma visão na qual a matemática é estática, a-histórica e portadora de dogmas previamente estabelecidos. No ensino da matemática, o professor, em geral, não considera a probabilidade como um conteúdo “nobre”, digno do ensino tal como o status que é dado, por exemplo, aos conteúdos que envolvem a álgebra e a geometria.

A outra justificativa sobre o referido tema foi de passar a entender o que é uma pesquisa correlacional? O que é tabela de dupla entrada? O que é uma tabela de contingência? O que é associação de variáveis? O que associação de variáveis em uma tabela de dupla entrada e em uma tabela de contingência? E de que forma e como as tabelas de contingência estão relacionadas com o ensino de estatística e probabilidade? No decorrer dessa pesquisa responderemos todas essas indagações levantadas até aqui.

Nossa pesquisa se justifica pela importância do raciocínio correlacional no cenário educacional a cerca do ensino de matemática, e de um modo mais específico no ensino de estatística e probabilidade. Nunes e Bryant (2011, p. 2, tradução nossa²), justificam a importância desse estudo pois “a incerteza de eventos totalmente aleatórios é um mundo imperfeito, mas, no entanto, associações importantes”.

Nós, enquanto professores e até mesmo os alunos, nos deparamos com situações no cotidiano, ao qual, temos que tomar decisões que envolvem um certo risco e que para tomar a decisão mais coerente, temos que analisar se as variáveis estão relacionadas ou não. De acordo com Nunes e Bryant (2011), o raciocínio de correlação entre variáveis é importante pois permite

¹ O trabalho integrado entre Estatística e Probabilidade tem sido denominado, em vários países do mundo, de estocástica.

² uncertainty of totally random events lies a world of imperfect, but nonetheless important associations.

analisar se existe uma associação, entre as variáveis e também indicar que não podemos ter certeza de como a associação afetará a outra variável.

No exemplo dado na pesquisa de Nunes e Bryant (2011), sabemos que existe uma relação entre o quanto as pessoas comem e se o seu peso subiu ou abaixou, mas também sabemos que associação entre essas duas variáveis (peso x quantidade de comidas ingerida por pessoa) não é perfeita, já que o efeito entre essas duas variáveis depende muito do metabolismo de cada indivíduo.

Sabendo da importância do tema, resolvemos realizar uma pesquisa seguindo os princípios de uma pesquisa qualitativa, onde, procuramos responder a seguinte problemática: Quais são os conhecimentos que os futuros professores de matemática (licenciandos) do Centro Acadêmico do Agreste apresentam ou mobilizam referente a associação de variáveis em tabelas de contingência?

1.1 Objetivo geral

Para respondermos à questão central da pesquisa desenvolvemos o seguinte objetivo geral, **analisar os conhecimentos de futuros professores de matemática (licenciandos) sobre associação de variáveis em tabelas de contingência**. Para suprimos o objetivo geral desenvolvemos três objetivos específicos, os quais são apresentados a seguir.

1.1.1 Objetivos específicos

- Identificar as estratégias desenvolvidas pelos futuros professores de matemática nas situações envolvendo associação de variáveis em tabela de contingência.
- Investigar se os tipos de problemas com respeito a dependência das variáveis interferem nas respostas.
- Compreender os conhecimentos mobilizados por estudantes, em dois grupos distintos, um com alunos do 1º período de licenciatura-matemática que ainda não cursaram a disciplina de Estatística e a outra do 5º período que já cursaram a disciplina.

Nos parágrafos seguintes apresentamos como essa pesquisa está estruturada, onde, é exposto de forma clara e breve o que cada capítulo aborda e como estão estruturados.

No capítulo 2 (dois) expomos alguns estudos sobre a estocástica mais especificamente no Ensino Superior, onde, é realizada uma breve discussão com base em pesquisas que apontam lacunas referente a(s) disciplina(s) de Estatística e Probabilidade no curso de formação do

futuro professor de matemática entre elas destacamos (COSTA; NACARATO, 2011; LOPES, 2013).

No capítulo 3 (três) dividimos em duas seções, onde, a primeira foi exposta uma discussão com base em estudos sobre o ensino de probabilidade e de probabilidade condicional (AMÂNCIO, 2012; COUTINHO, 2007; SILVA, 2015; CARVALHO, 2015 e 2017). Na segunda seção apresentamos uma discussão com base em estudos sobre associação estatística entre duas variáveis, também apresentamos algumas estratégias corretas para se realizar uma possível associação (CAÑADAS, BATANERO, CONTRERAS, ARTEAGA, 2010; NUNES, BRYANT, 2011; FERNANDES, MUGABE, CORREIA, 2012).

No capítulo 4 (quatro) desta pesquisa que denominamos por “O que os documentos curriculares apontam sobre associação de variáveis em tabelas de contingência? ”. Apresentamos uma análise da nova Base Nacional Curricular Comum que foi homologada recentemente, a análise realizada é referente ao ensino de estatística e probabilidade, dando maior ênfase na temática desta pesquisa que é associações de variáveis em tabelas de contingência.

No capítulo 5 (cinco) expusemos o método adotado por esta pesquisa, onde, apresentamos o objeto de estudo, os sujeitos participantes da pesquisa, as hipóteses levadas em consideração para a realização do instrumento de pesquisa que foi adotado para a coleta de dados. Também são apresentadas as estratégias e procedimentos seguidos para a análise de dados.

No capítulo 6 (seis) apresentamos a análise de dados da pesquisa e também é apresentado uma discussão dos principais resultados dessa pesquisa. Este capítulo está dividido em três momentos. O primeiro momento apresentamos dados da caracterização do perfil dos participantes, o segundo momento apresentamos resultados gerais e por grupos das 4 (quatro) questões problemas do questionário e o terceiro e último momento é apresentado a leitor os resultados gerais e individuais dos grupos por item/questão.

Por fim apresentamos o capítulo intitulado por considerações finais, nele expomos o que propusemos inicialmente para esta pesquisa. Ressaltamos também alguns resultados de estudos atuais sobre a temática desta pesquisa e por fim, apresentamos uma síntese dos resultados dessa pesquisa.

2 O QUE OS ESTUDOS APONTAM SOBRE O ENSINO DE ESTATÍSTICA E PROBABILIDADE NA FORMAÇÃO DO FUTURO PROFESSOR DE MATEMÁTICA

Na pesquisa desenvolvida por Costa e Nacarato (2011), as autoras fazem um recorte de uma dissertação de mestrado que teve como objetivo investigar como os professores de matemática em exercício das escolas básicas percebem a inserção da estocástica em sua formação e na sua prática docente, e como os professores formadores percebem a inserção da mesma, na formação dos professores de Matemática.

As autoras justificam o objetivo defendendo que “não há como desconsiderar que o professor precisa ter um repertório de saberes para desenvolver adequadamente seu trabalho em sala de aula” (COSTA; NACARATO, 2011, p. 368). Em estudos levantados pelas autoras referentes ao campo da formação docente, têm apontado como resultados que, nesse repertório de saberes, o professor precisa ter um domínio conceitual não apenas do conteúdo a ser trabalhado, mas, também, de outros saberes, como o pedagógico do conteúdo e o curricular.

A metodologia utilizada por *ibidem* (2011), foi a aplicação de um questionário onde foram destinados a 50 professores de escolas públicas e particulares da rede básica de ensino, das cidades de Jundiaí-SP e circunvizinhas, onde, obteve um retorno de apenas 30 questionários respondidos. Este questionário indagava os participantes sobre se os mesmos tinham percepções referente à inserção da estocástica na sua formação e na sua prática profissional. E de transcrições de entrevistas com quatro professores formadores (professores do curso de Licenciatura em Matemática que lecionam Estatística). A entrevista abordou quais são as percepções desses professores formadores referente a estocástica na formação docente.

Como resultados obtidos Costa e Nacarato (2011), relatam que a maioria dos professores que responderam o questionário afirmaram que a disciplina de Estatística constou na graduação, mas poucos se referem à probabilidade. Esses professores até lembravam de uma ou duas disciplinas relacionadas à Educação Estatística na grade curricular, porém grande parte dos professores não traziam boas lembranças dessas aulas; pois argumentaram que eram só aplicação de fórmulas e um ensino mecanizado.

O que reforça a análise de Shaughnessy (1992 apud COSTA; NACARATO 2011, p. 375) de que “a maioria dos cursos de Probabilidade e Estatística, em nível universitário, continua a pautar-se em regras e cálculos estatísticos ou em introduções matematizadas desses conteúdos”. Ainda sobre a análise dos questionários temos como resultado que a formação inicial e continuada não tem contribuído para que o professor construa um repertório de saberes

que lhes possibilite atuar com segurança diante do desafio de formar o pensamento estocástico de seus alunos. (COSTA; NACARATO, 2011).

Referente aos resultados das transições da pesquisa de Costa e Nacarato (2011), temos que o ensino na disciplina de Estatística é através de formulas mecanizadas e na maioria das vezes não há nenhuma ligação com Probabilidade. O que corrobora com o que foi respondido no questionário desta mesma pesquisa. E o que faz se tornar meio que um “ciclo vicioso” pouco se faz para trabalhar de forma mais conceitual durante a formação do professor de matemática, por não trabalhar durante sua formação o professor ficar inseguro na Educação Básica e recorre ao livro didático e por sua vez alguns desses livros apresentam falhas conceituais na apresentação desses conteúdos (COSTA; NACARATO, 2011).

Seguindo a mesma temática apresentada acima, encontramos na pesquisa de Lopes (2013), uma discussão pertinente sobre a disciplina de Estatística no curso de formação inicial do futuro professor de matemática que provavelmente em breve ou futuramente passará à atuar na Educação Básica. A autora aponta que os currículos de matemática no cenário nacional vêm assinalando a necessidade de iniciar a trabalhar uma educação estatística já nos primeiros anos de escolaridade. (LOPES, 2013). E isso faz com que os futuros professores de matemática tenham um domínio do conteúdo, porém isso requer que esses profissionais tenha uma formação adequada e significativa para realizar o que os currículos exigem.

Lopes (2013), coloca que seu estudo tem como objetivo problematizar o desenvolvimento teórico e metodológico da disciplina de Estatística na formação docente, através de experiências vivenciadas durante o estágio da mesma disciplina mencionada anteriormente. O estágio ocorreu no segundo semestre de 2008, na Universidade da Georgia (The University of Georgia).

A autora nos remete que a proposta da disciplina se inicia da concepção de que estatística é uma ciência de análise de dados e que não se pode ser abordada em um curso de formação inicial de professores de matemática como uma parte da matemática. Com isso, a formação matemática, que é centrada em fórmulas, algoritmos e exercícios de fixação, se tornou por se só insuficiente (LOPES, 2013).

Assim, consideramos que a pesquisa realizada por idem (2013), é do tipo bibliográfica e de campo. Bibliográfica pois a autora submete alguns resultados de pesquisas realizadas no cenário nacional, no qual mostra que grande parte dos cursos de formação de professores de matemática ensina estatística através de fórmulas e de ensino mecanizados, e há pouca carga horária durante a sua formação o que contribuem bastante para o fracasso na Educação Básica.

Ainda sobre os levantamentos de estudos bibliográficos, Lopes (2013) analisa estudos norte-americanos e alguns documentos curriculares. Entre estes, a autora destaca o GAISE Project³, onde, recomenda-se como resultado proposto que todos os cursos de introdução estatística sejam capazes de produzir alunos educados estatisticamente, isso significa que os estudantes devem compreender e desenvolver a literacia estatística e ser capaz de pensar estaticamente.

A pesquisa de campo de ibidem (2013) é realizada através de vivencias durante o estágio na disciplina de Estatística e Probabilidade, onde, é destinado para futuros professores de matemática e professores que já atuam na Educação Básica e procuram fazer pós-graduação. A autora afirma que esse curso segue as recomendações do documento curricular citado anteriormente, onde, “sua proposta associa conteúdo estatístico e conteúdo didático, afim de preparar professores para o futuro trabalho docente de qualidade, nos cursos de Estatística a serem ministrados no ensino secundário”. (LOPES, 2013, p. 909).

Segundo Lopes (2013), o desenvolvimento metodológico da disciplina se desenvolveu através de resolução de problemas e com o uso de recursos tecnológicos. Sobre este ensino a autora destaca que:

No ensino de estatística é mais importante se concentrar no entendimento a partir de dados, usando um julgamento e resolução de problemas, ao invés de simplesmente aplicar fórmulas para calcular números, pois a prioridade está sobre o desenvolvimento do pensamento estatístico. O uso de fórmulas deve ser secundário, sua escolha deve ser feita por uma razão específica e ela deve levar a uma conclusão adequada. (LOPES, 2013, p. 910)

Esta análise corrobora com o estudo de Costa e Nacarato (2011), estudo qual, mencionamos anteriormente onde nos dois trabalhos os autores defendem a ideia que o ensino de estocástica através de fórmulas não tem contribuído muito para uma aprendizagem significativa. O que faz com que os cursos de estatística e professores formadores adotem outras estratégias de ensino como resolução de problemas por meio de recursos tecnológicos o que possibilitara ao aluno um melhor entendimento conceitual do conteúdo.

Entre os resultados apresentado por Lopes (2013) destacamos, que embora os dados analisados pela a autora seja de algumas regiões específica do país, esses dados expressam muito a realidade nacional. Ou seja, não somente os futuros professores de matemática se

³ Diretrizes para Avaliação e Ensino em Educação Estatística, que visam em recomendações para a Educação Básica e para o Ensino Superior, coordenados por Christine Franklin (FRANKLIN et al., 2007) e por Joan Garfield (ALIAGA et al., 2010) respectivamente.

sentem inseguros em trabalhar com os conteúdos de estatística nas suas aulas de matemática na Educação Básica, mas também existe a ausência de recursos didático que subsidie um trabalho docente com maior magnitude. Apoiamos que uma formação estatística possa possibilitar ao futuro professor de matemática introduzir a estocástica em suas aulas e não fiquem atrelados a livros didático e paradidáticos.

Lopes (2013), evidencia como resultado de sua pesquisa que as atividades de ensino propostas no curso promovem o desenvolvimento do pensamento estatístico dos futuros professores de matemática e assim auxilia-os na percepção sobre o trabalho a ser desenvolvido com seus futuros alunos.

A seguir enfocamos estudos relacionados ao ensino de Probabilidade.

Coutinho (2007) apresenta uma discussão sobre o que a história aponta sobre o conceito de probabilidade, e para isto a autora analisa como e de que forma surgiu a ideia de “acaso” no decorrer da história, para depois chegar nos conceitos dos enfoques probabilísticos. A autora defende que o papel da história desses conceitos é fundamental para a escolha de contextos para apresentação desses primeiros conceitos probabilísticos no Ensino Fundamental.

Tendo em vista a importância do conceito probabilístico, como foi identificado nas pesquisas de Costa e Nacarato (2011) e Lopes (2013) também julgamos como importante que o professor de matemática tenha domínio desses conceitos e concordamos com Coutinho (2007), onde, defende como necessário que o mesmo profissional saiba dos contextos, para assim melhor inserir os conteúdos no Ensino Fundamental.

Em seu trabalho, a autora visa “propiciar ao professor uma diversidade de contextos possíveis e apreensões probabilísticas para o trabalho com a *ideia de acaso e as noções de probabilidade e de modelo probabilista*, sobre o ponto de vista de sua gênese histórica”. (COUTINHO, 2007, p. 51). Assim ela traz fatos e curiosidades da história do surgimento do acaso e da probabilidade.

Na pesquisa de Amâncio (2012) que tem como título “Planejamento e aplicação de uma sequência didática para o ensino de probabilidade no âmbito do PIBID” estabelece alguns objetivos como:

- Criar uma proposta para iniciar o conteúdo de Probabilidade para o Ensino Médio que possa ser aplicada pelos licenciandos vinculados ao PIBID. Para isto, nos valem dos vários conceitos que permeiam a Probabilidade, dos conhecimentos do conteúdo e pedagógico do conteúdo e da análise do conteúdo de Probabilidade apresentado no livro didático do Ensino Médio adotado na escola em que se dará o acompanhamento da pesquisa.

- Identificar, via a proposta criada, algumas das contribuições do PIBID na formação profissional dos licenciados em Matemática da Universidade Federal do Rio de Janeiro. (AMÂNCIO, 2012, p. 2)

Percebemos com o estudo de Amâncio (2012) a preocupação de propostas metodológicas para a abordagem da probabilidade na Educação Básica. A autora conclui que a probabilidade deve ser trabalhada com os licenciandos e que o PIBID foi um importante espaço para a inserção desta temática na formação inicial dos professores.

A pesquisa realizada por Silva (2015), buscou compreender a abordagem dos significados de probabilidade em livros didáticos aos anos finais do Ensino Fundamental.

Silva (2015), levanta pontos positivos e algumas limitações de todos os significados de probabilidade (Intuitivo, Clássico, Geométrico, Frequentista, Subjetivo e Formal), onde utiliza como fio condutor para a análise da sua pesquisa os significados de probabilidade sistematizados por Batanero (2005). Não será detalhado todos os significados de probabilidade pois não é o foco desta pesquisa. Como resultado Silva (2015) aponta lacunas com o conceito de probabilidade nas coleções didáticas investigadas, e que por ser um dos recursos mais utilizados pelo professor em sala de aula seria necessário melhorar a abordagem da probabilidade em tais recursos.

3 UMA BREVE DISCUSSÃO SOBRE INVESTIGAÇÕES COM PROBABILIDADE CONDICIONAL E ASSOCIAÇÃO DE VARIÁVEIS

No estudo realizado por Carvalho (2015), o autor investiga e analisa conceitos de probabilidade condicional de 25 futuros professores de matemática da Universidade Federal de Pernambuco (UFPE), por meio do jogo das “três fichas”. Como resultado desse estudo o autor aponta lacunas no conhecimento comum e especializado do conteúdo dos participantes da pesquisa sobre o conceito de probabilidade condicional.

O autor discute em seu estudo os conhecimentos fundamentais para a compreensão do conceito de probabilidade condicional na formação inicial do professor matemática. Este estudo tem como objetivo o mesmo apresentado na tese de doutoramento de Contreras (2011) de acordo com Carvalho (2015), fazer com que os professores vivenciassem a experimentação de uma situação didática baseada em um paradoxo clássico da teoria probabilística (o jogo das três fichas) com intuito de aflorar alguns conhecimentos matemáticos e didáticos deste grupo concernente a probabilidade condicional.

O método utilizado por Carvalho (2015) para obtenção do seu objetivo foi a vivência de situações ocorridas por meio do jogo das três fichas e através de um protocolo com perguntas referentes às situações ocorridas durante o jogo. Participaram ao todo 25 estudantes do curso de Matemática-Licenciatura da Universidade Federal de Pernambuco - Centro Acadêmico do Agreste (UFPE-CAA).

Este jogo foi sistematizado com base no Paradoxo das Caixas de Bertrand, assim conhecido por ter sido estudado pelo matemático francês do século XIX Joseph Bertand. O jogo tem o seguinte enunciado:

Se tomam 3 fichas da mesma forma e tamanho, das quais uma é vermelha em ambas as faces; outra é azul por uma face e vermelho na outra e a terceira é azul nas duas faces. O professor coloca as três fichas em uma caixa, que agita convenientemente, antes de selecionar uma das três fichas ao azar. Mostra uma das faces da ficha, mantendo a outra escondida, pedindo a seus alunos que adivinhem a cor do lado oculto. Uma vez feita as apostas, o professor mostrar o lado oculto. Cada aluno que tenha acertado a previsão efetuada consegue um ponto. (CARVALHO, 2015, p. 192)

Como resultados obtidos no estudo de Carvalho (2015), temos que “apontam lacunas no conhecimento comum e especializado do conteúdo deste grupo com o conceito de probabilidade condicional”. Teve caso de o participante acertar o valor da probabilidade, mas errou conceitualmente o que mostra que o mesmo tem dificuldades tanto no conhecimento comum, e especializado do conteúdo.

Os estudos envolvendo tabelas de contingência são importantes para a compreensão da associação de variáveis. Este conhecimento deve ser trabalhado na formação inicial e continuada do professor de matemática. Cañadas, Batanero, Contreras e Arteaga (2010), em sua pesquisa descrevem um estudo exploratório de distintas estratégias utilizadas para estabelecer uma possível associação entre duas variáveis, que estão apresentadas numa tabela de contingência.

Na busca de procurar saber e entender o que é uma tabela de contingência e para que elas são utilizadas, encontramos no website Wikipédia⁴ que as tabelas de contingência são utilizadas para registrar observações independentes de duas ou mais variáveis aleatórias, que normalmente são qualitativas. Cañadas, Batanero, Contreras e Arteaga (2010), afirmam em sua pesquisa que as tabelas de dupla entrada ou as tabelas de contingência se constituem em um método usual para apresentar informações estatísticas nas mídias, como por exemplo, a imprensa e na internet.

Acreditamos que as tabelas de contingência são tabelas que nos fornecem informações, onde as variáveis são geralmente qualitativas, ou seja, representam um significado de cunho qualitativo no contexto em que estão inseridas. Já as tabelas de dupla entrada são tabelas que fornece dados de variáveis normalmente quantitativas, ou seja, é uma variável que assume valores numérico, em outras palavras são variáveis que estão relacionadas a uma medição, a uma contagem, portanto são valores que podem ser medidos, por exemplo, temperatura, peso e altura.

Cañadas, Batanero, Contreras e Arteaga (2010), afirmam em seu estudo que as tabelas de contingência são um instrumento bastante comum no âmbito profissional de um psicólogo, especialmente nas análises diagnósticas e em uma avaliação psicológica, pois o psicólogo se depara com diferentes sintomas que podem estar associados ou não a uma patologia, ou algum hábito inadequado do paciente. A seguir apresentamos uma tabela de contingência 2x2.

Tabela 1 - Tabela de contingência 2x2

	<i>A</i>	<i>Não A</i>	<i>Total</i>
<i>B</i>	<i>a</i>	<i>b</i>	<i>a + b</i>
<i>Não B</i>	<i>c</i>	<i>d</i>	<i>c + d</i>
<i>Total</i>	<i>a + c</i>	<i>b + d</i>	<i>a + b + c + d</i>

Fonte: O autor (2018).

⁴ Disponível em: <https://pt.wikipedia.org/wiki/Tabela_de_conting%C3%Aancia>

A tabela 1, ilustra uma tabela de contingência 2x2, onde os dados podem ser interpretados da seguinte forma: a variável **A** representa uma possível patologia e **B** um sintoma possível, de acordo com ibidem (2010, p. 6) “la emisión de un diagnóstico sería equivalente a evaluar si existe o no asociación entre las correspondientes variables”. A célula **a** representa os indivíduos que apresentam o sintoma e a doença, a célula **b** representa o quantitativo de indivíduos que estão com o sintoma, mas não possuem a doença, já a célula **c** apresenta o número de indivíduos que não apresentam o sintoma, mas estão com a doença e a célula **d** são os indivíduos que não apresentam o sintoma e nem a doença.

Para facilitar na leitura e utilizando uma interpretação matemática desses dados apresentados na tabela acima, podemos ler esses dados da seguinte maneira: a célula **a** pode ser vista como “sim-sim”, a célula **b** como “sim-não”, a célula **c** como “não-sim” e a **d** como “não-não”. No decorrer deste tópico são abordadas as possíveis associações que se pode realizar com as variáveis em uma tabela de contingência.

As informações das células das tabelas de contingência normalmente são apresentadas como frequências absolutas, onde pode-se ser alterada para frequências relativas, fazendo com que fique mais fácil para o leitor interpretar os dados contidos nas células. Podemos também interpretar esses dados em forma de probabilidades conjuntas.

Retomando a tabela 1, podemos nos perguntar qual é a probabilidade de um indivíduo na tabela apresentar a patologia **A**? Da mesma forma podemos nos perguntar qual é a probabilidade de escolhermos um indivíduo aleatoriamente na tabela e ele apresentar o sintoma **B**? E qual é a probabilidade de uma pessoa da tabela não apresentar a doença **A**? E por fim, qual é a probabilidade de não apresentar o sintoma **B**?

Para respondermos as quatro perguntas feitas acima, iremos denominar por linha marginal a linha da tabela 1, onde, apresentar o total dos indivíduos com e sem a doença **A** respectivamente e de coluna marginal os totais de indivíduos que apresentam e não apresentam o sintoma **B**.

Sendo assim, para respondermos a primeira pergunta, sobre a probabilidade de um indivíduo na tabela ter a doença **A**, basta utilizar a fórmula laplaciana/clássica de probabilidade, onde, o número de indivíduos que apresentam a doença é $n(A) = (a + c)$ e o total de indivíduos que estão presentes na tabela é $n(S) = (a + b + c + d)$. Portanto, a probabilidade de escolher um indivíduo na tabela e ele apresentar a doença **A** é de

$$P(A) = \frac{n(A)}{n(S)} = \frac{a + c}{a + b + c + d}$$

Para a segunda pergunta, onde se quer saber a probabilidade de escolher uma pessoa aleatoriamente na tabela e ela apresentar o sintoma é de:

$$P(B) = \frac{n(B)}{n(S)} = \frac{a + b}{a + b + c + d}$$

desta maneira, utilizamos o total de indivíduos que apresentam o sintoma **B** em relação ao total de indivíduos presentes na tabela.

Como na questão anterior, podemos responder a terceira pergunta, onde se deseja saber qual é a probabilidade de a pessoa ser escolhida aleatoriamente e não apresentar a patologia **A**. Com isso, podemos representar algebricamente da seguinte forma:

$$P(\text{não } A) = \frac{n(\text{não } A)}{n(S)} = \frac{b + d}{a + b + c + d}$$

Para a quarta pergunta, a qual, queria saber qual é a probabilidade de um indivíduo ser escolhido aleatoriamente na tabela e não apresentar o sintoma **B**, podemos representar como:

$$P(\text{não } B) = \frac{n(\text{não } B)}{n(S)} = \frac{c + d}{a + b + c + d}$$

Se observarmos as quatro perguntas anteriores, podemos notar que as probabilidades calculadas são referentes a apenas uma variável, ou seja, são alusivas a doença **A** e ao sintoma **B**. Em outras palavras essas probabilidades são referentes as informações da linha e coluna marginais da tabela 1.

Para sabermos a probabilidade de cada célula a , b , c e d na tabela 1, precisamos nos perguntar qual é a probabilidade de escolhermos um indivíduo aleatoriamente na tabela e ele ter a doença **A** e apresentar o sintoma **B**? E qual é a probabilidade de escolhermos um indivíduo qualquer aleatoriamente e ele, está com a doença **A** e não apresentar o sintoma **B**? Também podemos nos perguntar sobre qual é a probabilidade de escolher o indivíduo na tabela aleatoriamente e ele não ter a doença **A** e apresentar o sintoma **B**? Por fim, podemos nos

perguntar qual é a probabilidade de escolher um indivíduo na tabela aleatoriamente e ele não apresentar doença **A** e nem o sintoma **B**?

Todas essas perguntas, referem-se à probabilidade de um indivíduo estar ou não com a doença **A** “e” apresentar ou não o sintoma **B**. Se nos detalharmos as quatro perguntas podemos perceber que todas elas têm em comum o artigo “e”. Como a teoria das probabilidades utiliza das propriedades da teoria dos conjuntos, temos que o artigo “e” têm um significado matemático, ou seja, está representando uma interseção entre esses dois eventos. Por exemplo, o indivíduo tem que está apresentando as duas variáveis.

Então, podemos responder todas essas perguntas utilizando a ideia de probabilidade da interseção entre dois eventos, onde a variável “patologia **A**” representa um evento e a variável “sintoma **B**” representa um outro evento. Portanto, podemos representar a fórmula da probabilidade da interseção de dois eventos como está de acordo com o livro de Giovanni, Giovanni Jr., Bonjorno e Sousa (2013, p. 157):

$$P(A \cap B) = \frac{n(A \cap B)}{n(S)}$$

onde, $n(A \cap B)$ é o número de indivíduos na interseção dos dois eventos e $n(S)$ é o número de indivíduos no espaço amostral.

Sendo assim, podemos responder as perguntas anteriores utilizando os dados da tabela 1, da seguinte forma respectivamente:

$$P(A \cap B) = \frac{n(A \cap B)}{n(S)} = \frac{a}{a + b + c + d}$$

$$P(A \cap \text{não}B) = \frac{n(A \cap \text{não}B)}{n(S)} = \frac{c}{a + b + c + d}$$

$$P(\text{não}A \cap B) = \frac{n(\text{não}A \cap B)}{n(S)} = \frac{b}{a + b + c + d}$$

$$P(\text{não}A \cap \text{não}B) = \frac{n(\text{não}A \cap \text{não}B)}{n(S)} = \frac{d}{a + b + c + d}$$

Conhecendo os valores das células podemos transformar a tabela de contingência em uma tabela de probabilidade conjunta, como iremos ilustrar a seguir:

Tabela 2 - Tabela de probabilidade conjunta

	<i>A</i>	<i>Não A</i>	<i>Total</i>
<i>B</i>	$P(A \cap B)$	$P(\text{não}A \cap B)$	$P(B)$
<i>Não B</i>	$P(A \cap \text{não}B)$	$P(\text{não}A \cap \text{não}B)$	$P(\text{não} B)$
<i>Total</i>	$P(A)$	$P(\text{não} A)$	$P(S) = 1$

Fonte: O autor (2018)

Esses resultados da tabela 2 nos mostra importante pois auxilia no cálculo das probabilidades condicionais que podem ser exploradas de acordo com os dados da tabela 1. Vamos nos deparar com situações que envolvem probabilidade condicional. Essa situação pode ser identificada quando quer saber a probabilidade de um indivíduo que tenha sido escolhido aleatoriamente na tabela 1 ter a patologia *A* dado que ele já tenha apresentado o sintoma *B*? Interpretando essa pergunta com atenção, podemos perceber que já sabemos que essa pessoa já possui o sintoma, sendo assim podemos expressar a solução dessa perguntar através da fórmula de probabilidade condicional, pois queremos saber a probabilidade de um evento que estar condicionado a outro evento que já ocorreu.

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{\frac{a}{a+b+c+d}}{\frac{a+b}{a+b+c+d}} = \frac{a}{a+b+c+d} \cdot \frac{a+b+c+d}{a+b} = \frac{a}{a+b}$$

Ao trabalhar com probabilidade condicional devemos tomar alguns cuidados matemáticos para não cometer alguns erros do tipo assumir que $P(A|B) = P(B|A)$, a seguir apresentaremos a explicação matemática do porquê dessa igualdade ser falsa.

$$P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{\frac{a}{a+b+c+d}}{\frac{a+c}{a+b+c+d}} = \frac{a}{a+b+c+d} \cdot \frac{a+b+c+d}{a+c} = \frac{a}{a+c}$$

O cálculo anterior nos remete a situação de que se quer calcular a probabilidade de um indivíduo ser escolhido aleatoriamente na tabela 1 ter o sintoma *B* dado que ele esteja com a

doença A . Podemos chegar ao resultado que igualdade de $P(A|B) = P(B|A)$ só é válida se $P(A) = P(B)$ ou se $b = c$.

Para uma tabela de contingência 2×2 podemos fazer até oito perguntas diferentes envolvendo a probabilidade condicional. Sempre seguindo o mesmo raciocínio que apresentamos até então. No exemplo posterior será feita uma análise mais profunda nesses possíveis cálculos.

A tabela 2 é importante também para auxiliar na tomada de decisões para saber se duas variáveis estão associadas entre si. Para verificar se existe associação basta verificarmos se elas são independentes e dizemos que “dois eventos A e B de um espaço amostral Ω (com $p(A) \neq 0$ e $p(B) \neq 0$) são independentes se, e somente se, $p(A/B) = p(A)$, ou, de modo equivalente: $p(A \cap B) = p(A) \cdot p(B)$ ”. (DANTE, 2016, p. 246)

Fica mais claro de ver essas probabilidades se utilizarmos exemplos com números. Consideremos o exemplo na figura 1, onde foi utilizado no livro de Lima, Carvalho, Wagner e Morgado (2006), para introduzir o conteúdo de probabilidade condicional.

Figura 1 - exemplo de uma tabela de contingência

Exemplo. A tabela abaixo dá a distribuição dos alunos de uma turma, por sexo e por carreira pretendida.

	Masculino	Feminino	Total
Científica	15	5	20
Humanística	3	17	20
Total	18	22	40

Escolhe-se ao acaso um aluno.

Fonte: Adaptado de Lima, Carvalho, Wagner e Morgado (2006, p. 132)

A partir dos dados da tabela na figura 1, podemos estabelecer algumas condições e conclusões para facilitar o cálculo matemático. Então a primeira conclusão é referente ao tipo de variável, onde, temos a variável sexo que é uma variável qualitativa nominal, que de acordo com Dante (2004) é uma variável que expõe uma qualidade, mas seus valores não seguem uma ordem, no caso do sexo podemos categorizar por masculino e feminino. A variável carreira

pretendida também é uma variável qualitativa nominal, no qual podemos preestabelecer uma qualidade para a carreira pretendida, que no caso acima é científica e humanística.

Quanto as condições, podemos considerar C, H, M e F como os eventos, pretende uma carreira científica, pretende uma carreira humanística, o aluno selecionado ser do sexo masculino e o aluno ser do sexo feminino, respectivamente. Observe que estamos trabalhando com dados presentes em uma tabela de contingência e a mesma ideia que foi apresentada anteriormente se aplicar a essa tabela, por exemplo, podemos orientar o sexo masculino como a variável A e o sexo feminino como a variável *não* A . O mesmo acontece para carreira pretendida. Assim temos que a probabilidade de cada evento desse acontecer é de:

$$P(C) = \frac{20}{40} = \frac{1}{2};$$

$$P(H) = \frac{20}{40} = \frac{1}{2};$$

$$P(M) = \frac{18}{40} = \frac{9}{20};$$

$$P(F) = \frac{22}{40} = \frac{11}{20}.$$

Se pensarmos na probabilidade de selecionar uma pessoa dessa turma ao acaso, em que essa pessoa pretenda uma carreira científica e seja do sexo feminino, teremos que pensar em um cálculo de probabilidade que contemple os dois eventos, fazendo uso do artigo “e”, podemos responder essa pergunta como a probabilidade da interseção de dois eventos. Para este caso temos o evento pretender uma carreira científica (C) e o evento ser do sexo feminino (F), em que apenas 5 das 40 pessoas apresentam a característica de interesse, portanto

$$P(C \cap F) = \frac{5}{40} = \frac{1}{8}.$$

Podemos também fazer esse mesmo procedimento analogamente para todas as situações, onde for perguntado qual é a probabilidade de escolhermos uma pessoa na turma aleatoriamente, onde ela pretenda uma carreira científica (C) e seja do sexo masculino (M)? Pretenda seguir uma carreira humanística (H) e for do sexo feminino (F)? Pretenda seguir uma carreira humanística (H) e seja do sexo masculino (M)? Para essas situações podemos fazer os cálculos da seguinte maneira respectivamente

$$P(C \cap M) = \frac{15}{40} = \frac{3}{8};$$

$$P(H \cap F) = \frac{17}{40};$$

$$P(H \cap M) = \frac{3}{40}.$$

Por outro lado, podemos fazer outras indagações referentes aos dados da tabela na figura 1, por exemplo, pode-se perguntar qual é a probabilidade de uma pessoa dessa turma ser escolhida aleatoriamente, onde ela pretenda seguir uma carreira humanística dado que essa pessoa é do sexo feminino?

Essa situação podemos resolver fazendo uso dos conceitos de probabilidade condicional, no exemplo dado se quer saber a probabilidade, dessa pessoa pretender seguir carreira humanística, sabendo que ela é do sexo feminino, em outras palavras, já sabemos que o evento (F), já aconteceu. Sendo assim podemos calcular essa probabilidade da seguinte forma:

$$P(H|F) = \frac{P(H \cap F)}{P(F)} = \frac{\frac{17}{40}}{\frac{11}{20}} = \frac{17}{40} \cdot \frac{20}{11} = \frac{17}{22}$$

Ao trabalharmos com o conceito de probabilidade condicional, devemos tomar alguns devidos cuidados, pois o mesmo pode nos induzir ao erro que é bastante conhecido, por exemplo, é comum que as pessoas associem que $P(A|B) = P(B|A)$, o que é uma conclusão falsa. Esse erro é chamado de *falácia da condicional transposta* (FALK, 1986). Essa dificuldade é trazida por Carvalho (2015, p. 189), onde, o autor afirma que:

Em pesquisas realizadas por diversos autores, envolvendo tanto professores como estudantes, uma das dificuldades com probabilidade condicional é discriminar adequadamente a direção da condicional $P(A|B)$ e $P(B|A)$ ou supor que $P(A|B)$ e $P(B|A)$ são iguais (Batanero, Contreras e Díaz, 2012; Oliveira, 2013). [...] pelo teorema de Bayes, estas probabilidades condicionais só são iguais se A e B tiverem a mesma probabilidade.

Dando continuidade ao exemplo da figura 1, ainda podemos calcular a probabilidade de uma pessoa escolhida aleatoriamente na turma ser do sexo feminino, sabendo-se que ela pretenda ter uma carreira humanística ($P(F|H)$). Outra analogia que podemos fazer é saber qual

é a probabilidade de um aluno escolhido ser do sexo masculino, sabendo-se que ele pretenda seguir uma carreira humanística. De acordo com o exemplo dado na tabela da figura 1, podemos calcular ao todo, oito probabilidades condicionais ($P(F|H)$, $P(H|F)$, $P(M|H)$, $P(H|M)$, $P(F|C)$, $P(C|F)$, $P(M|C)$ e $P(C|M)$). Essas probabilidades podem ser calculadas da seguinte maneira, respectivamente.

$$P(F|H) = \frac{P(F \cap H)}{P(H)} = \frac{\frac{17}{40}}{\frac{1}{2}} = \frac{17}{40} \cdot \frac{2}{1} = \frac{17}{20}$$

$$P(H|F) = \frac{P(H \cap F)}{P(F)} = \frac{\frac{17}{40}}{\frac{11}{20}} = \frac{17}{40} \cdot \frac{20}{11} = \frac{17}{22}$$

$$P(M|H) = \frac{P(M \cap H)}{P(H)} = \frac{\frac{3}{40}}{\frac{1}{2}} = \frac{3}{40} \cdot \frac{2}{1} = \frac{3}{20}$$

$$P(H|M) = \frac{P(H \cap M)}{P(M)} = \frac{\frac{3}{40}}{\frac{18}{40}} = \frac{3}{40} \cdot \frac{40}{18} = \frac{3}{18} = \frac{1}{6}$$

$$P(F|C) = \frac{P(F \cap C)}{P(C)} = \frac{\frac{1}{8}}{\frac{1}{2}} = \frac{1}{8} \cdot \frac{2}{1} = \frac{1}{4}$$

$$P(C|F) = \frac{P(C \cap F)}{P(F)} = \frac{\frac{1}{8}}{\frac{11}{20}} = \frac{1}{8} \cdot \frac{20}{11} = \frac{5}{22}$$

$$P(M|C) = \frac{P(M \cap C)}{P(C)} = \frac{\frac{3}{8}}{\frac{1}{2}} = \frac{3}{8} \cdot \frac{2}{1} = \frac{3}{4}$$

$$P(C|M) = \frac{P(C \cap M)}{P(M)} = \frac{\frac{3}{8}}{\frac{9}{20}} = \frac{3}{8} \cdot \frac{20}{9} = \frac{1}{2} \cdot \frac{5}{3} = \frac{5}{6}$$

Após realizar todos os cálculos de probabilidade condicional, é possível fazer algumas conclusões entre as duas variáveis presente na tabela da figura 1, uma seria afirmar que elas são dependentes.

Essa conclusão poderia ser vista através do conceito de eventos independentes, assim vamos usar como exemplo a seguinte igualdade $P(C \cap M) = P(C) \cdot P(M)$, sabemos que $P(C \cap M) = 15/40$, $P(C) = 1/2$ e $P(M) = 9/20$.

$$P(C) \cdot P(M) = \frac{1}{2} \cdot \frac{9}{20} = \frac{9}{40} \neq \frac{15}{40} = P(C \cap M)$$

Com esse resultado chegamos que o evento sexo e carreira pretendida não são independentes, a seguir apresentamos uma outra maneira correta de fazer a associação entre essas duas variáveis. Lembrando que a probabilidade condicional também é uma frequência relativa condicional.

Pode-se chegar a essa conclusão comparando as probabilidades, por exemplo, se comparamos $P(F/C) = 1/4$ com $P(M/C) = 3/4$, é possível afirmar que para cada 4 (quatro) alunos na turma que pretende uma carreira científica 3 (três) são do sexo masculino e apenas 1 (uma) é do sexo feminino.

Ainda com as comparações das probabilidades, podemos deduzir que para cada 20 alunos da turma que pretende uma carreira humanística apenas 3 (três) são do sexo masculino e 17 (dezesete) são do sexo feminino. Chegamos a essa conclusão comparando os valores de $P(M/H) = 3/20$ com $P(F/H) = 17/20$.

Para concretizar as possíveis associações necessárias para tomar uma decisão, é preciso fazer pelo menos mais duas comparações entre as probabilidades condicionais calculadas acima, nos dois parágrafos anteriores, foi feita comparações entre a variável carreira pretendida e a variável sexo. Em seguida será feito comparações entre a variável sexo e a variável carreira pretendida.

Primeiramente vamos comparar as probabilidades $P(C/M) = 5/6$ com $P(H/M) = 1/6$, essas probabilidades condicionais informam que de cada 6 (seis) alunos do sexo masculino, 5 (cinco) deles pretende uma carreira científica e apenas 1 (um) pretende uma carreira humanística. Agora se comparamos as probabilidades condicionais, $P(C/F) = 5/22$ com $P(H/F) = 17/22$, chegamos à conclusão que das 22 (vinte e duas) alunas do sexo feminino, 5 (cinco) almeja uma carreira científica e 17 (dezesete) pretende uma carreira humanística.

Após realizar todas essas comparações, ficar fácil ver que a maioria dos alunos do sexo masculino dessa turma pretende uma carreira científica e uma pequena parcela deles pretende seguir uma carreira humanística. Com as alunas do sexo feminino acontece da mesma forma,

só que a preferência de seguir uma carreira se inverte, grande parte das alunas almejam seguir a carreira humanística e uma minoria pretende seguir uma carreira científica.

Para mostrar que as variáveis são dependentes, basta usar a definição de eventos independentes, ao qual, mencionamos anteriormente, um evento é independente se, e somente se, $P(A|B) = P(A)$ essa formula é equivalente a

$$P(A|B) = P(A) \Rightarrow \frac{P(A \cap B)}{P(B)} = P(A) \Rightarrow P(A \cap B) = P(A) \cdot P(B).$$

caso os eventos não seja independente, deduzimos que são dependentes. Antes de apresentar todas possíveis situações usando o conceito de eventos independentes, a seguir é mostrado ao leitor que nos dados da tabela da figura 1, existe uma tendência entre as variáveis.

Olhando os dados da tabela com uma análise criteriosa é possível afirmar que as variáveis nesta amostra são dependentes, pois os alunos do sexo masculino dessa turma têm uma tendência de seguir uma carreira científica, enquanto as alunas do sexo feminino têm uma tendência maior para seguir uma carreira humanística. Não é possível afirmar o real motivo, que motivou os alunos e alunas dessa turma terem essa tendência por carreira pretendida, apenas podemos é levantar algumas hipóteses que levou a ocorrer essa situação, por exemplo, incentivo dos familiares, vivências realizadas nas aulas, enfim são “n fatores” que motivaram a chegar a este resultado. Só é possível afirmar que existe uma associação entre as duas variáveis.

Usando o conceito de eventos de independentes no exemplo dado na figura 1, temos que $P(C \cap M) = 3/8$, sabemos também que $P(C) = 1/2$ e que $P(M) = 9/20$, Assim $P(C) \cdot P(H) = 9/40$, logo verificamos que a igualdade $P(C \cap M) = P(C) \cdot P(H)$ não é satisfeita, sendo assim podemos concluir que as variáveis são dependentes. Note, que em todas as situações neste exemplo essa igualdade será falsa. Em seguida iremos fazer todas igualdades possíveis.

$$P(M|C) = \frac{3}{4} \neq \frac{1}{2} = P(C)$$

$$P(F|C) = \frac{1}{4} \neq \frac{1}{2} = P(C)$$

$$P(M|H) = \frac{3}{20} \neq \frac{1}{2} = P(H)$$

$$P(F|H) = \frac{17}{20} \neq \frac{1}{2} = P(H)$$

$$P(C|M) = \frac{5}{6} \neq \frac{9}{20} = P(M)$$

$$P(H|M) = \frac{1}{6} \neq \frac{9}{20} = P(M)$$

$$P(C|F) = \frac{5}{22} \neq \frac{11}{20} = P(F)$$

$$P(H|F) = \frac{17}{22} \neq \frac{11}{20} = P(F)$$

Portanto podemos concluir que a variável sexo e a variável carreira pretendida são dependentes. Essa é uma das estratégias corretas de fazer uma associação entre as duas variáveis. Uma outra estratégia correta de fazer associação seria de vez de usar as frações formadas pelas as probabilidades condicionais para tomar as decisões é usar os dados em forma de frequência relativas, ou seja, transformando as frações em porcentagem.

Sabendo que qualquer probabilidade é um número que tem valor entre 0 (zero) e 1 (um), também sabemos que quando o valor da probabilidade é 0 (zero) o evento é impossível de acontecer, em outras palavras, nunca acontecerá. Um exemplo de evento impossível é o lançamento de dois dados comuns, onde a soma de suas respectivas faces seja maior que 12 (doze).

E quando a probabilidade é igual a 1 (um), denominamos como evento certo, ou seja, tem 100% (cem portentos) de certeza que o evento vai acontecer, um exemplo prático de um evento certo é em um lançamento de um dado comum e obtemos faces com valores menores que 7 (sete). Como toda probabilidade é um número pertencente ao intervalo $[0, 1]$, podemos reescrever as probabilidades a partir das frequências relativas dos dados. Iremos abordar pesquisas que apresentam possíveis estratégias de associação entre duas variáveis.

Cañadas, Batanero, Contreras e Arteaga (2010), apresentam um estudo exploratório avaliando a compreensão intuitiva da associação entres duas variáveis presentes em uma tabela de contingência por parte dos estudantes que ingressaram no curso de Psicologia.

Os autores afirmam que se basearam em diversos estudos no campo da psicologia e da didática matemática entre esses é destacado Inhelder e Piaget (1955) no campo da psicologia, onde realiza estudos exploratórios sobre as tabelas de contingência em duas etapas, no qual, a primeira se destina a estudantes de 12 há 13 anos e a segunda etapa com estudantes de 13 há 15 anos. Inhelder e Piaget (1955) de acordo com Cañadas, Batanero, Contreras e Arteaga (2010) influenciaram outros autores, como Smedlund (1963) e Shaklee e Mins (1982).

Já no campo da didática matemática destacamos no estudo de Cañadas, Batanero, Contreras e Arteaga (2010) os estudos de Estepa (1993) e Estepa e Batanero (1995), que juntos realizaram um estudo qualitativo de estratégia com 213 estudantes do Ensino Médio.

Cañadas, Batanero, Contreras e Arteaga (2010), utilizam estratégias de estudos anteriores e criam algumas outras estratégias de acordo com sua análise dados. Essas estratégias estão divididas em três categorias (Estratégias corretas; Estratégias parcialmente corretas; e Estratégia incorretas). Todas as categorias de estratégias utilizadas para a análise dados do estudo de idem (2010), foram as seguintes:

- Estratégias corretas: (EC. 1) comparar todas as distribuições das frequências relativas condicionadas de uma variável para os distintos valores da outra variável; (EC. 2) comparar todas as frequências relativas condicionadas de uma variável para um único valor da outra variável com a frequência marginal da primeira variável; (EC. 3) comparação das possibilidades a favor e contra a B em cada valor da variável A e (EC. 4) o aluno calcula as frequências relativas condicionais por linhas na representação gráfica.
- Estratégias parcialmente corretas: (EP. 1) comparar a distribuição das duas frequências absolutas de uma variável para um único valor da outra variável com a frequência absoluta marginal da primeira variável; (EP. 2) comparar somente uma das duas de uma variável para um valor da segunda com a frequência marginal da primeira variável; (EP. 3) comparar as frequências absolutas duplas entre si; (EP. 4) comparar a somas das frequências nas diagonais da tabela; (EP. 5) equiparar a independência com equiprobabilidade nas células da tabela.
- Estratégias incorretas: (EI. 1) o único uso da célula de maior frequência; (EI. 2) o uso de apenas uma distribuição condicional para decidir a associação; (EI. 3) comparar frequências absolutas ou relativas entre si ou com o número total de observações; (EI. 4) comparar frequências marginais entre si; (EI. 5) utilizar apenas a célula em que os caracteres estão ausentes e, se a frequência for alta, supor que isto implica na inexistência de associação; (EI. 6) comparar a caixa com o resto das células, ou bem com as células b e c; (EI. 7) o uso células de maior e menor frequência; (EI. 8) o uso das células de menor frequências; (EI. 9) o aluno forma as tabelas de frequências marginais calculando frequências, probabilidade e porcentagens; (EI. 10) calculando a diferença das diagonais; (EI. 11) o aluno apenas considera sua teoria previa e não levar em conta os dados e (EI. 12) outros procedimentos incorretos.

De acordo, com Estepa (1993, apud CAÑADAS, BATANERO, CONTRERAS E ARTEAGA, 2010, p. 11) “sugiere que las estrategias incorrectas usadas por los estudiantes en los juicios de asociación dependen de algunas concepciones que poseen los alumnos sobre ésta”. Descrevemos essas concepções abaixo de acordo com o estudo de Cañadas, Batanero, Contreras e Arteaga (2010):

- **Concepção casual:** o aluno apenas considera a dependência entre variáveis, que podem ser explicados pela presença de uma relação casual entre elas.
- **Concepção determinista:** são os alunos que só percebem a dependência de tipo funcional e não são capazes de generalizar para os casos de dependência aleatória. Um exemplo é quando se exige uma expressão algébrica que associam as variáveis.
- **Concepção unidirecional:** nessa concepção, o aluno não admite a associação inversa, considera a intensidade da associação, porém não considera o seu sinal.
- **Concepção local:** está concepção se apresenta quando os sujeitos respondem se baseando em alguns casos isolados, por exemplo, só considera apenas uma distribuição condicional ou a célula de maior valor.

A pesquisa realizada por Cañadas, Batanero, Contreras e Arteaga (2010), se caracteriza como um estudo de campo, onde, os autores aplicam um questionário a 62 alunos do primeiro ano do curso de Bacharelado em Psicologia da Universidade de Huelva, na Espanha. Os autores salientam que os dados foram coletados antes dos alunos estudarem especificamente sobre tabelas de contingência. Segundo os autores essa decisão foi tomada “a fin de detectar posibles estrategias y concepciones intuitivas incorrectas y tenerlas en cuenta en el diseño de la enseñanza”. (CAÑADAS, BATANERO, CONTRERAS, ARTEAGA, 2010, p.12)

O questionário que foi utilizado no trabalho de Cañadas, Batanero, Contreras e Arteaga (2010), foi adaptado do estudo de Estepa e Batanero (1995), onde, se constitui em quatro problemas com resposta aberta com a mesma apresentação dos dados da pergunta apenas variando o contexto.

Onde o primeiro problema questiona se o sofrimento de estresse ou não, está associado a ter insônia ou não; o segundo problema perguntar se ser filho único ou não, está relacionado com ser ou não uma criança problemática; no terceiro item é questionado que se o habito de levar uma vida sedentária ou não, está associado a ter ou não crise de alergia; e na última questão

é perguntado se o número de horas estudadas para um exame tem uma relação direta com os números de aprovados ou não.

Cañadas, Batanero, Contreras e Arteaga (2010), apresenta em seu trabalho uma solução correta para responder e analisar cada item do seu questionário, essa solução se consiste em comparar as distribuições das frequências relativas condicionais, por exemplo, no primeiro item 1:

Figura 2 - item 1 do estudo de Cañadas, Batanero, Contreras e Arteaga (2010).

Se pretende estudiar si el sufrir insomnio tiene relación con los trastornos de estrés. En una muestra de 250 personas observadas, se obtuvieron los siguientes resultados:

	Padece estrés	No padece estrés
Tener insomnio	90	60
No tener insomnio	60	40

a) Utilizando estos datos, rãzonas si tener o no insomnio tiene relación con padecer estrés.

b) Indica cómo has usado los datos de la tabla para llegar a tu conclusión

Fonte: (CAÑADAS, BATANERO, CONTRERAS E ARTEAGA, 2010, p. 13)

Continuando com a resolução do item 1, os autores respondem que as frequências relativas de sofrer estresse quando se têm insônia é de $(90 / 150) = 0,6$ e a frequência relativa de ter estresse dado que não têm insônia é $(60 / 100) = 0,6$; o que são iguais. Da mesma forma, se comparamos a frequência relativa de não sofrer de estresse dado que tem insônia que é de $(60 / 150) = 0,4$ com a frequência relativa de não sofre de estresse dado que não tem insônia que também é de $(40 / 100) = 0,4$ verificamos que são iguais. Portanto, as variáveis no item 1 da pesquisa de Cañadas, Batanero, Contreras e Arteaga (2010), são independentes. Do ponto de vista matemático podemos usar a ideia de evento independente. Sabemos que um evento é independente a outro se essa igualdade seguinte for verdadeira $P(A \cap B) = P(A) \cdot P(B)$. Note que:

$$P(E \cap I) = \frac{90}{250} = \frac{9}{25}; P(E) = \frac{150}{250} = \frac{15}{25} = \frac{3}{5} \text{ e } P(I) = \frac{150}{250} = \frac{15}{25} = \frac{3}{5}.$$

$$P(E) \cdot P(I) = \frac{3}{5} \cdot \frac{3}{5} = \frac{9}{25} = P(E \cap I)$$

Como a igualdade foi verdadeira podemos afirmar que os eventos são independentes. Onde $P(E)$ é a probabilidade de uma pessoa ser escolhida aleatoriamente na tabela da figura 2 ter estresse; $P(I)$ é a probabilidade de uma pessoa ser escolhida aleatoriamente na tabela da figura 2 ter insônia; e $P(E \cap I)$ é a probabilidade dos dois eventos ocorrerem juntos. É bom salientar ao leitor que para ter a certeza que a pessoa que sofre de estresse ou não tem alguma relação com ter insônia ou não, é necessário analisar as demais igualdades.

$$P(E \cap \text{não}I) = P(E) \cdot P(\text{não}I);$$

$$P(\text{não}E \cap I) = P(\text{não}E) \cdot P(I);$$

$$P(\text{não}E \cap \text{não}I) = P(\text{não}E) \cdot P(\text{não}I).$$

Ao qual, $P(\text{não}I)$ é a probabilidade de uma pessoa ser escolhida aleatoriamente na tabela da figura 2 não ter insônia; $P(\text{não}E)$ é a probabilidade de uma pessoa ser escolhida aleatoriamente não apresentar estresse; $P(E \cap \text{não}I)$ é a probabilidade de uma pessoa ser escolhida aleatoriamente ter estresse e não apresente insônia; $P(\text{não}E \cap I)$ é a probabilidade de uma pessoa ser escolhida aleatoriamente não apresente estresse e tenha insônia; e $P(\text{não}E \cap \text{não}I)$ é a probabilidade de uma pessoa ser escolhida aleatoriamente não apresente estresse e nem apresente insônia.

Com essas quatro igualdades é possível afirmar que as duas variáveis estão relacionadas ou não. No caso das variáveis da tabela da figura 2, temos que ter estresse ou não ter, não está relacionado com ter insônia ou não.

Cañadas, Batanero, Contreras e Arteaga (2010), afirma que os resultados do seu estudo foram pobres em estratégias, porém destaca que o tamanho da amostra é moderada, onde os estudantes estão no primeiro ano da Universidade de Huelva, cujas características dos mesmo não são muitos diferentes dos demais estudantes espanhóis de outras universidades. Uns dos pontos positivos do estudo segundo os autores é que o trabalho proporcionou algumas estratégias novas que não foram descritas em trabalhos anteriores.

Assim como exploramos e detalhamos o estudo de Cañadas, Batanero, Contreras e Arteaga (2010), vamos destacar a seguir os estudos Nunes e Bryant (2011) e Carvalho (2017), ao qual, utilizamos esses três estudos como base para elaboração deste capítulo e também dessa pesquisa.

Em Nunes e Bryant (2011), encontramos um estudo que se aprofundar mais sobre a temática, ao qual, abordamos até então. Nesse artigo os autores, defendem que os alunos da Educação Básica, devem sair da escola com uma boa compreensão de matemática, especificamente sobre como lidar com incerteza. Nunes e Bryant (2011), advogam o porquê dos alunos devem sair da Educação Básica com uma boa compreensão matemática sobre incerteza, pois, como adultos terão que tomar decisões importantes em cima de uma probabilidade em certos eventos e sobre o risco que têm uma gama de perigos.

Os autores, afirmam que o artigo se resume a pesquisas sobre o entendimento de correlações, que avaliam a existência de associações mútuas entre variáveis, o tipo e o grau de associação entre elas (NUNES, BRYANT, 2011). Os autores nos submetem que a incerteza em eventos totalmente aleatórios é um mundo imperfeito, praticamente impossível de ter certeza se determinado evento irá acontecer, porém, podemos estabelecer associações importantes, o que nos permitirá prever se o evento é pouco provável de acontecer ou se é bem provável que aconteça.

Nunes e Bryant (2011), em seu estudo afirmam que as correlações entre duas variáveis, são medidas que medem a força de intensidade e a direção da associação entre essas duas variáveis. O coeficiente de correlação pode assumir valores entre (-1) e (1), quando o coeficiente de correlação é igual a (1), temos a informação que as duas variáveis estão perfeitamente e positivamente relacionadas e quando o coeficiente é igual a (-1), também podemos afirmar que as variáveis estão perfeitamente relacionadas, no entanto, estão negativamente relacionadas.

De acordo com Nunes e Bryant (2011), nenhuma das correlações apresentadas anteriormente deixa qualquer espaço de incerteza, porém, essas correlações perfeitas são muito raras no campo da estatística descritiva⁵. Como já vimos as correlações são maiores ou menores do que zero podendo variar de (-1) à (1), sendo assim, podem pertencer ao intervalo de [-1, 0] ou ao intervalo de [0, 1]. Com isso, elas nos mostram que existe uma associação ou não entre as duas variáveis, no entanto não podemos ter certeza de como é realizada a associação para cada caso (NUNES, BRYANT, 2011).

No trabalho realizado por Cañadas, Batanero, Contreras e Arteaga (2010), os autores afirmam que podemos tirar algumas conclusões de acordo com o valor do coeficiente de correlação/associação. A primeira é referente ao “sinal de associação entre as duas variáveis”.

⁵ A estatística descritiva é um ramo da estatística que aplica várias técnicas para descrever e sumarizar um conjunto de dados.

Sobre o sinal, consideramos três casos possíveis da dependência ser direta, quando o valor coeficiente é maior que zero, da dependência ser inversa, isto ocorre quando o valor do coeficiente de correlação for menor que zero e da associação ser independente quando o valor do coeficiente for igual a zero.

A segunda conclusão que se pode fazer é sobre o “grau de intensidade da dependência entre duas variáveis”, podemos estabelecer conclusões mediante a medida do coeficiente de Phi de Pearson⁶, para tabelas 2x2 e com o coeficiente de V de Cramér⁷ em tabela 2x3 (CAÑADAS, BATANERO, CONTRERAS, ARTEAGA, 2010). Podemos estabelecer alguns graus de correlação de acordo com a figura abaixo:

Figura 3 - Grau de intensidade do relacionamento de coeficiente de correlação positivos e negativos

Perfeito	+ 1		- 1
Forte	+ 0,9		- 0,9
	+ 0,8		- 0,8
	+ 0,7		- 0,7
Moderado	+ 0,6		- 0,6
	+ 0,5		- 0,5
	+ 0,4		- 0,4
Fraco	+ 0,3		- 0,3
	+ 0,2		- 0,2
	+ 0,1		- 0,1
Zero	0		

Fonte: Dancey e Reidy (2006, p. 186)

O diagrama da figura 3 ilustra que o coeficiente (-1) é tão forte quanto o coeficiente (+1). Dancey e Reidy (2006, p. 186), afirma que “só porque o relacionamento é negativo não significa que seja menos importante ou menos forte do que o relacionamento positivo”. Então, quando a correlação é inversa, significa dizer que quando uma variável cresce a outra variável

⁶ **Karl Pearson** (1857- 1936), foi um grande colaborador para o desenvolvimento da estatística como uma disciplina científica séria e independente. Foi o fundador do Departamento de Estatística Aplicada na University College London em 1911; foi o primeiro departamento universitário dedicado à estatística em todo o mundo.

⁷ **Harald Cramér** (1893 - 1985) foi um matemático, atuário e estatístico sueco, especializado em estatística matemática e teoria probabilística de números.

diminuir, e quando a correlação é direta, implicar dizer que quando uma variável aumentar a outra variável também aumentar.

No exemplo trazido por Nunes e Bryant (2011), é possível deduzir que existe uma relação entre o peso das pessoas e quantidade de comida ingerida por elas, ou seja, se o peso subiu ou baixou, porém não podemos afirmar que a associação entre essas duas variáveis é perfeita, já que esse resultado depende de pessoa a pessoa, por exemplo, algumas pessoas têm um metabolismo que colabora com o ganho de peso, ou com a perda de peso. Também existe a questão da ansiedade que atingem determinadas pessoas e outras não. Ou seja, são diversos fatores que podem estar associados ao ganho ou a perda de peso.

Nunes e Bryant (2011), afirma que a associação, mesmo não sendo perfeita, permite que os médicos e nutricionistas ajudem os pacientes dando conselhos fundamentais e importantes sobre os riscos que a obesidade trás para a saúde das pessoas. O raciocínio correlacional não só um parâmetro de medida, onde se mede a direção e o grau de relacionamento entre duas variáveis, mas como já vimos anteriormente envolve o raciocínio probabilístico. (NUNES, BRYANT, 2011).

Os autores, trazem em seu trabalho que o “risco” é uma incerteza. O risco pode ser expresso como uma probabilidade ou uma frequência com base em dados empíricos. Um exemplo clássico e trazido por idem (2011) sobre risco, é na medicina. Numa situação em que um médico pede um exame a um paciente, o qual servirá para diagnosticar um câncer, é comum que a sociedade e o próprio paciente só pensem no risco de estar com a doença (ou seja a probabilidade de ter câncer), por outro lado não pensam na possibilidade de não estar com a doença (ou seja a probabilidade de não está com a doença). Ou seja, é comum que a palavra risco esteja associada no dia a dia das pessoas a algo ruim/negativo.

Nunes e Bryant (2011), advogam a importância das correlações, afirmando que as correlações ajudam as pessoas a definir a probabilidade de um determinado evento acontecer. No exemplo dado pelos os autores sobre a importância das associações consideremos que um comitê está realizando um inquérito sobre a morte de crianças submetidas a um certo tipo de cirurgia em um determinado hospital particular. O comitê deve considerar diversos fatos que podem estarem ligados ou não a morte da criança neste hospital. Para isso é usado o raciocínio correlacional. A questão que deve ser colocada em pauta pelo o comitê é se as crianças operadas nesse hospital têm mais chances/propensas a óbito do que as outras crianças que receberam a mesma intervenção cirúrgica em outros hospitais. Ou seja, existe uma possível associação entre receber o tratamento no hospital e a morte?

Nesta situação trazida por Nunes e Bryant (2011), é improvável que a associação seja perfeita, pois nem todas as crianças que se submeteram ao procedimento cirúrgico neste hospital chegaram a óbito e nem todas a criança que fizeram a mesma cirurgia em outro hospital sobreviveram. Entretanto, a situação, nos submete que a questão é saber se as chances de as crianças morrerem são aumentadas quando elas são operadas neste hospital. (NUNES, BRYANT, 2011).

Diante a todos exemplos levantados e apresentados até então, podemos ver a importância e valor que as associações de variáveis tem no cotidiano de qualquer pessoa. Nunes e Bryant (2011), ressaltam que o valor e a importância das associações imperfeitas que são amplamente reconhecidas.

Nunes e Bryant (2011) defendem que existem várias exigências cognitivas para a compreensão do raciocínio correlacional, porém destaca três como fundamentais para uma associação correta em uma tabela de dupla entrada/contingência. São o entendimento de aleatoriedade, a compreensão do espaço amostral e por último a quantificação de probabilidades de formas proporcionais.

Sobre o entendimento de aleatoriedade Nunes e Bryant (2011), defendem que o objetivo de uma análise de correlação entre duas variáveis ou dois eventos é estabelecer se eles ocorrem com mais frequência do que se espera por acaso. Ou seja, se não existe uma relação entre dois eventos **A** e **B** por exemplo, ainda é possível que eles possam ocorrer juntos por acaso. Nunes e Bryant (2011), concluem que a compreensão da aleatoriedade é, deste modo, objeto da compreensão de correlações.

A respeito da compreensão do espaço amostral, Nunes e Bryant (2011, p.5, tradução nossa⁸) defendem que o raciocínio correlacional também “envolve a compreensão de espaço amostral. Para examinar se dois eventos estão associados, precisamos estabelecer não apenas se eles co-ocorrem, mas também o que todos os casos possíveis são: de **A** acontecer? Sim ou não. E de **B** acontecer? Sim ou não”.

Ou seja, o espaço amostral nessa situação mencionada por Nunes e Bryant (2011) em uma tabela de contingência, são (sim-sim), (sim-não), (não-sim) e (não-não). Esse resultado foi apresentado no início desta seção. Os autores, afirmam que é comum pensar que apenas o caso (sim-sim) é proeminente para a questão de uma correlação entre os dois eventos, porém a probabilidade dos eventos ocorrerem juntos deve ser entendida no contexto dos eventos que

⁸ involves understanding sample space. In order to examine whether two events are associated, we need to establish not only whether they co-occur but also what all possible the cases are: Did A happen? Yes or no. Did B happen? Yes or not.

não estão ocorrendo juntos também (NUNES, BRYANT, 2011). De acordo com Cañadas, Batanero, Contreras e Arteaga (2010), esse pensamento se caracteriza como uma estratégia incorreta, onde, se olhar apenas a uma célula da tabela neste caso a célula (sim-sim) e não se levar em conta todos os dados que a tabela traz consigo.

Referente a quantificação das probabilidades de forma proporcionais Nunes e Bryant (2011), advogam que

Uma vez que estabelecemos o espaço amostral, precisamos passar para a quantificação das probabilidades de forma proporcional. A frequência dos casos que apoiam a existência de uma associação (o caso sim-sim e o caso não-não) proporcionalmente muito maior do que a frequência dos casos que não suportam a associação, de modo que se pode assumir que essa frequência se afasta do que seria de espera por acaso? Se for esse o caso, concluímos que existe uma associação entre os dois eventos. (p. 6, tradução nossa⁹)

De acordo, com o que é colocado acima podemos afirmar que uma variável está associada com uma outra, se os dados na tabela obedecerem às condições trazidas por Nunes e Bryant (2011), ou seja, as frequências das células dos casos (sim-sim e não-não) que confirmam a associação devem ser maiores proporcionalmente que as células dos casos (sim-não e não-sim) que não confirmam a associação. Quando falamos “ser maiores proporcionalmente” queremos dizer que há uma disparidade muito grande entre os valores das respectivas frequências. Quando isso ocorre podemos concluir que existe associação entre as duas variáveis numa tabela de contingência.

Em resumo Nunes e Bryant (2011), advogam que para entender a associação entre duas variáveis a criança necessita saber de pelo menos de três exigências sobre este raciocínio: elas precisam entender aleatoriedade e espaço amostral, elas devem ser capazes de reconhecer quais são os casos que são a favor a uma associação e quais são os casos que são contra a associação entre as duas variáveis e com isso elas precisam ser capazes de quantificar os casos positivos em comparação com os casos negativos para avaliar se os casos positivos são suficiente para sugerir que a co-ocorrência observada não é devido ao acaso.

Carvalho (2017), apresenta alguns estudos envolvendo risco probabilístico. O autor ressalta a importância do tema e adverte sobre o nível de complexidade dele, que para abordar

⁹ Once we have established the sample space, we need to move on to the quantification of probabilities in a proportional manner. Is the frequency of cases that support the existence of an association (the yes-yes and the no-no cases) proportionally really larger than the frequency of the cases that do not support the association, so that one can assume that this frequency departs from what one would expect by chance? If this is the case, we conclude that there is an association between the two events.

o referido tema se faz necessário uma abordagem detalhada de probabilidade condicional em tabelas de dupla entrada.

Idem (2017), advoga que é importante compreender a noção de risco probabilístico para realizar investigações sobre os problemas probabilísticos. Carvalho (2017, p. 61), afirma que:

Compreender o risco – que é outro aspecto do pensamento probabilístico – tem estreita relação com o raciocínio correlacional. Esse raciocínio exige o reconhecimento que as relações entre variáveis não são absolutas, mas existem em graus (ROSS E COUSINS, 1993) e, assim, envolvem raciocínio probabilístico. Assim, o grau de relacionamento entre duas variáveis pode ser determinado pelas frequências relativas nas tabelas de dupla entrada.

Sabemos que as relações que envolvem variáveis relevantes estão presentes em nossa vida e se encontram presentes no nosso cotidiano, em diversos campos da sociedade por exemplo, no campo das finanças: se existe alguma relação entre o valor do salário ganho e o grau de instrução (1º grau, 2º grau, Ensino Superior, etc.), já no campo da saúde: se existe alguma relação entre a contração do vírus da HIV e a prevenção contra o vírus da HIV. Carvalho (2017, p. 61) advoga que “A análise dessas relações permeia o entendimento sobre risco probabilístico. Tal entendimento se torna significativo tanto para a matemática como para outras disciplinas do currículo escolar como ciências, geografia, biologia”.

Na tese de Carvalho (2017), o autor nos submete que existe uma grande quantidade de situações que envolvem as tabelas de dupla entrada e como as pessoas interpretam por meio das informações que são apresentadas por meio de frequências e relações que são estabelecidas pelo cálculo de probabilidade. O autor informa que as pessoas (tanto médicos e até mesmo crianças) tendem a interpretar e compreender melhor a noção de risco quando os dados nessas tabelas são apresentados em forma de relação (por exemplo 1:4) ao invés de porcentagens (25%) ou frações ($1/4$) e proporções (0,25).

O autor também aponta que há estudos que indicam que as pessoas de inteligência normal conseguem fazerem inferências bayesianas corretamente, quando os dados da tabela de dupla entrada são apresentados em forma de frequências. Carvalho (2017), considera em sua pesquisa que as pessoas de inteligência normal são as pessoas que não cursaram o curso de estatística. O autor também traz que as tabelas de dupla entrada revelam uma ligação mais abertamente entre o raciocínio correlacional e probabilístico.

Carvalho (2017), defende que para as pessoas tomarem uma decisão correta numa tabela de dupla entrada ou de contingência se faz necessário utilizar o raciocínio do tipo correlacional:

Para uma tomada de decisão correta é necessário utilizar um raciocínio do tipo correlacional envolvendo uma coordenação entre as noções de probabilidade e proporções. É comum as pessoas, mesmo quando as informações são apresentadas nas tabelas, olharem apenas para uma célula, o que eles podem julgar como relevante para confirmar a relação entre as duas variáveis, sem ter em conta a frequência desta célula em relação ao total. (p. 63)

O autor apresenta uma outra maneira mais sofisticada estatisticamente de analisar se uma variável A está associada ou não com outra variável B, que é a ideia de razão de chance, ao qual, está ilustrada da seguinte forma:

Estatisticamente, uma razão de chances 1 indica que a condição ou evento que se está analisando é igualmente provável de ocorrer nos dois grupos. Uma razão maior do que 1 indica que a condição ou evento tem maior probabilidade de ocorrer no primeiro grupo. Finalmente, uma razão de chances menor do que 1 indica que a probabilidade é menor no primeiro grupo do que no segundo. (CARVALHO, 2017, p. 64)

O autor ilustra uma tabela de dupla entrada, em que apresenta informações de duas variáveis que é padecer de estresses e ter insônia e em seguida ilustra a fórmula de razão de chance para a associação de variáveis em uma tabela de dupla entrada.

Tabela 3 - Exemplo de comparação de risco por meio da razão de chances

	Padecer de estresses	Não padecer de estresses	Total
Ter insônia	w	x	$w + x$
Não ter insônia	y	z	$y + z$
Total	$w + y$	$x + z$	n

Fonte: Carvalho (2017, p. 65)

O autor apresenta uma fórmula matemática que viabilizar fazer uma associação correta em uma tabela de dupla entrada ou de contingência semelhante a tabela 3.

$$\text{Razão de chance} = \frac{w \cdot z}{x \cdot y}$$

Note que, w e z , são exatamente as células que confirmam a associação, enquanto que as células x e y , são as células que não confirmam a associação. Nunes e Bryant (2011) apresentam essa ideia das células que apoiam a associação e as que não apoiam e é possível afirmar se duas variáveis estão associadas ou não através das disparidades das frequências dessas células. Porém, esse resultado é diferente ao que Carvalho (2017) apresenta com a fórmula de razão de chance, pois é um resultado mais bem estruturado e de simples entendimento, já que algebricamente é bem simples do que apenas olhar as disparidades das frequências, um exemplo é quando os dados das células tiverem uma pequena disparidade, não fica fácil ver a associação entre as duas variáveis. Entretanto, a razão de chance mostrar o quanto as variáveis estão associadas ou não.

4 O QUE A BASE NACIONAL COMUM CURRICULAR APONTA SOBRE ASSOCIAÇÃO DE VARIÁVEIS EM TABELAS DE CONTINGÊNCIA?

Ao realizar uma análise na Base Nacional Comum Curricular (BNCC) para o Ensino Fundamental é possível identificar que o documento evidencia a importância da estatística e probabilidade, onde, o documento defende que a Matemática não se restringe apenas a eventos de natureza determinística, mas também a eventos de natureza aleatória.

A BNCC também assegurar que a Matemática no Ensino Fundamental deve fazer com que os alunos desenvolvam algumas competências específicas da Matemática. Essas competências estão divididas em oito, onde, apresentamos todas a seguir:

1. Reconhecer que a Matemática é uma ciência humana, fruto das necessidades e preocupações de diferentes culturas, em diferentes momentos históricos, e é uma ciência viva, que contribui para solucionar problemas científicos e tecnológicos e para alicerçar descobertas e construções, inclusive com impactos no mundo do trabalho.
2. Desenvolver o raciocínio lógico, o espírito de investigação e a capacidade de produzir argumentos convincentes, recorrendo aos conhecimentos matemáticos para compreender e atuar no mundo.
3. Compreender as relações entre conceitos e procedimentos dos diferentes campos da Matemática (Aritmética, Álgebra, Geometria, Estatística e Probabilidade) e de outras áreas do conhecimento, sentido segurança quanto à própria capacidade de construir e aplicar conhecimentos matemáticos, desenvolvendo a autoestima e a perseverança na busca de soluções.
4. Fazer observações sistemáticas de aspectos quantitativos e qualitativos presentes nas práticas sociais e culturais, de modo a investigar, organizar, representar e comunicar informações relevantes, para interpretá-las e avaliá-las crítica e eticamente, produzindo argumentos convincentes.
5. Utilizar processos e ferramentas matemáticas, inclusive tecnologias digitais disponíveis, para modelar e resolver problemas cotidianos, sociais e de outras áreas de conhecimento, validando estratégias e resultados.
6. Enfrentar situações-problema em múltiplos contextos, incluindo-se situações imaginadas, não diretamente relacionadas com o aspecto prático-utilitário, expressar suas respostas e sintetizar conclusões, utilizando diferentes registros e linguagens (gráficos, tabelas, esquemas, além de texto escrito na língua materna e outras linguagens para descrever algoritmos, como fluxogramas, e dados).
7. Desenvolver e/ou discutir projetos que abordem, sobretudo, questões de urgência social, com base em princípios éticos, democráticos, sustentáveis e solidários, valorizando a diversidade de opiniões de indivíduos e de grupos sociais, sem preconceitos de qualquer natureza.

8. Interagir com seus pares de forma cooperativa, trabalhando coletivamente no planejamento e desenvolvimento de pesquisas para responder a questionamentos e na busca de soluções para problemas, de modo a identificar aspectos consensuais ou não na discussão de uma determinada questão, respeitando o modo de pensar dos colegas e aprendendo com eles. (BRASIL, 2018, p. 265)

Se olharmos as competências específicas que a BNCC traz, é perceptivo que o ensino associação de variáveis em uma tabela de contingência ou de dupla entrada é uma ferramenta importantíssima para o desenvolvimento de algumas dessas competências, como por exemplo, a competência 4, 5 e 6.

A BNCC sugere cinco unidades temáticas para o ensino de Matemática no Ensino Fundamental, fazendo com que todas unidades temáticas abordem e desenvolvam todas competências específicas que enunciamos anteriormente (BRASIL, 2018). Sendo assim, a Matemática no Ensino Fundamental está dividida em cinco temáticas que são Números, Álgebra, Geometria, Grandezas e medidas, Probabilidade e estatística. Como já explúissemos no título deste capítulo iremos apenas dar ênfase a unidade temática de Probabilidade e estatística, mais especificamente o assunto de associação de variáveis.

A respeito a unidade temática de Probabilidade e estatística a BNCC, afirmar que a incerteza e o tratamentos de dados devem ser estudados nesta unidade temática. Onde é proposto pelo o documento que deve ser realizada uma abordagem em conceitos, fatos e procedimentos presentes em problemas presentes em situações cotidianas, nas ciências e tecnologia (BRASIL, 2018). De acordo com o documento o aluno terá que desenvolver algumas habilidades para consolidar a abordagem feita do conteúdo. Essas habilidades são coletar, organizar, representar, interpretar e analisar dados em diversa situações do cotidiano e que consequentemente consiga tomar decisões corretas e consiga fazer julgamentos bem embasados mantendo coerência nas suas afirmações (BRASIL, 2018).

A BNCC afirmar que o estudo das noções de probabilidade tem a finalidade no Ensino Fundamental – Anos Iniciais de:

promover a compreensão de que nem todos os fenômenos são determinísticos. Para isso, o início da proposta de trabalho com probabilidade está centrado no desenvolvimento da noção de aleatoriedade, de modo que os alunos compreendam que há eventos certos, eventos impossíveis e eventos prováveis. (BRASIL, 2018, p. 272)

O documento já defende que o aluno entenda a noção de aleatoriedade, fazendo com que ele consiga distinguir eventos certos, eventos impossíveis e eventos prováveis. Julgamos importante essa noção, pois Nunes e Bryant (2011) afirmam que o entendimento de associação entre variáveis depende de três conhecimentos mobilizados pelo o aluno, a lembrar ao leitor são que os alunos precisam entender a ideia de aleatoriedade; entender a noção de espaço amostral para saber identificar os casos que apoiam e não apoiam a associação; e por fim precisam quantificar os casos positivos em comparação com os casos negativos para assim avaliar se os casos positivos ocorre com mais frequência nos casos negativos ou não.

Para os Anos Finais, a BNCC aponta que os conhecimentos dos Anos Iniciais devem ser aprofundados e ampliado por meio de atividades envolvendo experimentos aleatórios e simulações para poder confrontar por meio do cálculo de probabilidade clássico e frequentista.

Esse último se baseia em realizar um determinado experimento certo número de vezes, sobre as mesma condições e de forma independente e tomar a probabilidade do evento de interesse como a frequência relativa a ele.

5 MÉTODO

Parafrazeando Oliveira (2016, p. 43), a método de pesquisa é um processo que começa a partir da disposição inicial para escolha do tema da pesquisa até a análise dos dados. Assim, nesse capítulo apresentamos como ocorreu a escolha do tema, do objetivo geral e dos objetivos específicos. Apresentamos também todos os procedimentos metodológicos usados para atender os objetivos levantados e como ocorreu as análises de dados.

A escolha do tema dessa pesquisa se deu a partir de uma revisão de literatura (COSTA; NACARATO, 2011; LOPES, 2013 e CARVALHO, 2015) e a partir das leituras dessas obras, surgiu a vontade de encaminhar e direcionar o tema da pesquisa para a estocástica e através de algumas conversas durante as orientações que foi escolhido o objeto de pesquisa desta monografia.

O nosso objeto de pesquisa é o conhecimento dos participantes sobre a noção de risco probabilístico em tabelas de contingência. Para atender ao nosso objeto de pesquisa foi levantado três objetivos específicos, os objetivos levantados para se chegar ao objeto de pesquisa foram:

- Identificar as estratégias desenvolvidas pelos futuros professores de matemática nas situações envolvendo risco”.
- Investigar se os tipos de problemas com respeito a dependência das variáveis interferem nas respostas.
- Compreender os conhecimentos mobilizados por estudantes, em dois grupos distintos, uma do 1º período de licenciatura-matemática que ainda não cursou a disciplina de Estatística e a outra do 5º período que já cursou a disciplina.

Para alcançar a todos os objetivos específicos levantados, foi necessário elaborar um questionário contendo 4 (quatro) questões problemas, em que todas apresentam uma tabela de contingência. Esse questionário foi dividido em duas partes. A primeira parte questionava sobre o perfil dos participantes e denominamos por “caracterização do participante”, na mesma era indagado dos respondentes qual era a sua idade, seu sexo, curso ao qual está matriculado, em que período está no curso e por último se já cursou a disciplina de Estatística.

No questionário foi reservado um pequeno espaço para informar aos participantes envolvidos na pesquisa, onde foi pedido a colaboração e seriedade de todos, pois as respostas seriam utilizadas para as coletas de dados. Também informamos que os nomes ficariam

preservados na pesquisa, ou seja, optamos por utilizar códigos para cada protocolo, por exemplo, os participantes foram denominados por aluno 1, aluno 2, até o último participante.

Consideramos essa primeira parte do questionário importante, pois os participantes são alunos e vivem constante desenvolvimento do seu intelecto. Por exemplo o aluno talvez tenha um resultado melhor no teste por ter cursado a disciplina de Estatística, ou pode ser que com a experiência na disciplina o aluno pode querer modelar as situações do questionário ao que foi vivenciado em algum assunto estatístico em sua graduação, que por consequência o aluno acabar utilizando uma estratégia errada, por estar atrelado as situações vivenciadas durante a disciplina. Na verdade, não sabemos ao certo se existe uma relação deste tipo já que cada estudante tem uma forma de pensar e agir, nesse caso, são apenas especulações que podemos fazer para essa primeira parte do questionário.

A segunda parte do questionário constou de 4 (quatro) situações problemas adaptadas dos estudos de Cañadas, Batanero, Contreras e Arteaga (2010) e de Nunes e Bryant (2012). Cada problema apresentava aos participantes uma tabela de contingência, onde os problemas indagavam se existia alguma relação entre as duas variáveis e pedíamos que explicassem usando os dados da tabela se há ou não uma associação.

No capítulo seguinte, os quatro problemas do questionário serão detalhados sendo apresentado o que era esperado em cada questão e os resultados encontrados em cada problema com base nas respostas dos participantes da pesquisa. Na primeira questão era perguntado se existe alguma relação entre ter doença brônquica e o habito de fumar. A segunda questão constituía em perguntar se existia alguma relação entre não fazer exercícios físicos e ter alergias. Também era questionado se existe alguma relação entre o uso de inseticida numa determinada planta e a doença apresentada pela planta, ou seja, se o inseticida é eficaz, essa situação é da terceira questão. Na quarta e última questão perguntava aos participantes se existia alguma relação entre os números de aprovados e reprovados com o tempo dedicado para estudar para o exame.

A primeira, segunda e quarta questão foram do estudo de Cañadas, Batanero, Contreras e Arteaga (2010) e a terceira foi retirada do estudo de Nunes et al. (2012). Todas questões perguntavam se o participante concordava se uma variável estava relacionada com a outra, com exclusão da três. Nela era indagado se aplicação do inseticida naquela situação era eficaz. Após a resposta se existia relação ou não, pedíamos que explicassem, usando os dados da tabela, o porquê de eles pensarem assim.

Os sujeitos da pesquisa são os futuros professores de matemática, que estão cursando o curso de Matemática-Licenciatura, ofertado pela Universidade Federal de Pernambuco, no

Centro Acadêmico do Agreste (UFPE-CAA). Esta escolha dos participantes se deu pelo fato de acreditarmos que após a conclusão do curso, ou até mesmo antes, os participantes irão lecionar na Educação Básica e vão se deparar com situações de risco probabilístico em sala de aula ou até mesmo fora dela.

Carvalho (2017), em sua tese de doutorado evidencia que a noção de risco por meio do estudo da associação entre variáveis em tabelas de contingências como um conhecimento emergente para o ensino nos anos finais do Ensino Fundamental ou no Ensino Médio.

Ainda, os sujeitos da pesquisa poderão ser indagados pelos seus futuros alunos sobre qual é a melhor decisão a se tomar ou qual é a possível associação que se pode fazer em uma tabela de contingência. Nossa proposta com essa escolha é tentar amenizar os resultados obtidos por Costa e Nacarato (2011), ao qual denominamos no capítulo anterior por “ciclo vicioso”. Também usamos como critério a importância do assunto, visando as possíveis situações presentes no cotidiano.

Para alcançar o primeiro objetivo específico utilizamos como base três classificações para categorizar as estratégias da mesma maneira que foi categorizada no estudo de Cañadas, Batanero, Contreras, Arteaga (2010). Foi denominado, por “estratégia correta” a estratégia que em todas as situações envolvendo tabelas de contingência, não deixe sombra de dúvida no seu desenvolvimento. Ou seja, se aluno optar por responder por elas, terão grande chances, para se chegar ao êxito na associação entre as duas variáveis e no risco probabilístico. De acordo com idem (2010), são as estratégias que utiliza todos os dados da tabela mediante a comparação das probabilidades.

A “estratégia parcialmente correta”, são estratégias que podem levar a uma possível associação correta, porém em alguns casos ela pode induzir o aluno ao erro. Cañadas, Batanero, Contreras e Arteaga (2010), colocam que as estratégias parcialmente corretas são aquelas que utilizam todos os dados da tabela, porém ao invés de se fazer comparações com probabilidades utiliza apenas frequências absolutas.

Consideramos por “estratégias incorretas”, as estratégias que levam os estudantes ao erro. Ou seja, eles não utilizam todas informações da tabela e acabam tirando conclusões falsas. De acordo com Cañadas, Batanero, Contreras e Arteaga (2010), a estratégia incorreta ocorre quando o aluno usar somente uma parte dos dados necessários.

Nas análises, apenas foram mencionadas as estratégias que surgiram nas coletas de dados. E nosso foco maior é observar se os licenciandos utilizam a estratégia errônea de escolha da célula de maior valor. Essa estratégia se torna errônea pois o aluno não deve centrar suas conclusões através de uma única célula e desconsiderar as demais células da tabela.

No segundo objetivo específico, queríamos “investigar se os tipos de problemas com respeito a dependência das variáveis interferem nas respostas”. Neste objetivo esperamos analisar se a dependência/associação entre as variáveis influenciariam nas respostas dos participantes. No questionário as questões apresentavam três tipos de associação. No primeiro problema não existia nenhuma associação entre as duas variáveis. O segundo e quarto problema a associação é direta. E a terceira questão a associação era inversa entre as duas variáveis.

Assim como Cañadas, Batanero, Contreras e Arteaga (2010, p. 15), para a elaboração do questionário levamos em consideração algumas variáveis, onde, é apresentado a seguir:

- ✚ Sinal de associação entre as variáveis, levando em contas os três casos possíveis: dependência direta, dependência indireta e independência.
- ✚ O grau de intensidade de associação entre as duas variáveis, na qual foi medida pelo o coeficiente de correlação de Phi Pearson para as tabelas 2x2 e o coeficiente V de Cramér nas tabelas 2x3. Foi escolhido durante a elaboração do questionário que a *Questão 1* teria o grau de intensidade igual a zero, uma outra questão teria esse grau de intensidade fraca que no questionário é a *Questão 4*, e duas questões de grau intensidade forte que são as *Questões 2 e 3*.
- ✚ A concordância entre os dados e as teorias previas sugerido pelo o tema. Na elaboração do questionário foi levado em conta se a associação empírica coincide ou não nos dados apresentados na tabela com as teorias previas. No questionário as *Questões 3 e 4* coincidem, a *Questão 1* não coincide e a *Questão 2* não existe teoria previas.

A seguir são apresentadas em uma tabela as variáveis que foram levadas em consideração durante a elaboração do questionário.

Tabela 4 - Variáveis que foram levadas em conta na elaboração do questionário

	Tabela 2x2			Tabela 2x3
	Questão 1	Questão 2	Questão 3	Questão 4
Dependência	Independência	Direta	Inversa	Direta
Grau de intensidade de associação entre duas variáveis	0	0,67	-0,60	0,37
Concordância com a teoria previa	Não	Não existe teoria	Sim	Sim

Fonte: O autor (2018).

Para atender o terceiro objetivo específico, que para lembrar ao leitor é “compreender os conhecimentos mobilizados por estudantes, em dois grupos distintos, uma do 1º período de licenciatura-matemática que ainda não cursou a disciplina de Estatística e a outra do 5º período que já cursou a disciplina”. Este objetivo foi alcançado nas análises de dados, onde apresentamos os resultados gerais e por grupo de quem cursou e quem não cursou a disciplina de estatística. Após apresentar os resultados dos dois grupos fizemos comparações entre eles.

A aplicação do questionário para as coletas de dados ocorreu em dois momentos. Inicialmente foi pensado em aplicar o questionário apenas para os alunos do quinto período do curso de Matemática-Licenciatura, pois na teoria os estudantes já teriam cursado a disciplina de Estatística que é ofertada no segundo período do curso. A primeira aplicação do questionário, foi durante uma aula da disciplina de Metodologia do Ensino de Matemática I, onde 19 (dezenove) estudantes estavam presente na sala de aula no dia da aplicação.

Após a aplicação do questionário, constatamos que era interessante para a pesquisa fazer uma nova coleta de dados com estudantes do curso de Matemática-Licenciatura, ao qual esses estudantes não tivessem cursado a disciplina de Estatística. Com isso, poderíamos comparar o desempenho dos dois grupos que cursaram a disciplina de estatística e os que não cursaram a disciplina.

Sendo assim, para a segunda parte da aplicação do questionário, foram escolhidos estudantes que estavam cursando o primeiro período do curso. A aplicação ocorreu em uma aula da disciplina de Matemática Básica. No dia da aplicação estavam presentes em sala de aula 27 (vinte e sete) alunos.

É bom salientar que antes de qualquer uma das aplicações do questionário, consultamos cada professor das respectivas disciplinas mencionadas anteriormente, o professor Dr. José Ivanildo Felisberto de Carvalho e a professora Me. Debora Karyna dos Santos Araújo

Bernardino Da Silva. Ambos concordaram positivamente com a aplicação do questionário durante as suas aulas.

Ressaltamos que logo após a aplicação do questionário foi disponibilizado para os participantes da pesquisa uma carta de esclarecimento sobre a pesquisa e um termo de consentimento livre e esclarecido. Na carta, informava o tema desta pesquisa, o objetivo geral e por fim foi enfatizado que a pesquisa aconteceria em dois momentos, onde, o primeiro momento seria com a participação deles na coleta de dados e o segundo momento aconteceria com as análises dos dados coletados. No termo informava que os participantes iriam participar na pesquisa como voluntários, também asseguramos que suas identidades seriam preservadas. Salientamos ainda que uma cópia do termo ficou com cada um dos participantes e uma outra com o pesquisador.

Com isso, o estudo foi realizado com uma amostra de 46 estudantes do curso de Matemática-Licenciatura da Universidade Federal de Pernambuco, do Centro Acadêmico do Agreste (UFPE-CAA). Sendo que dos 46 estudantes, 27 estavam cursando a disciplina de Matemática Básica ofertada no primeiro período do curso e os demais estavam cursando a disciplina de Metodologia do Ensino de Matemática 1 ofertada no quinto período.

Para organizar as informações dos resultados nas análises de dados foi usado o software estatístico IBM SPSS Statistics em sua versão 2.0. Fizemos o uso do software para fazer tabelas e gráficos dos resultados obtidos nas análises.

Oliveira (2016, p. 37), conceitua que uma pesquisa tem uma abordagem qualitativa quando se é

um processo de reflexão e análise da realidade através da utilização de métodos e técnicas para compreensão detalhada do objeto de estudo em seu contexto histórico e/ou segundo a sua estruturação. Esse processo implica em estudos segundo a literatura pertinente ao tema, observações, aplicação de questionários, entrevistas e análise de dados, que deve ser apresentada de forma descritiva.

Diante do que apresentamos na metodologia, julgamos que esta pesquisa atende a uma abordagem qualitativa. Além disso, acreditamos que esta pesquisa também atende a uma abordagem quantitativa. Idem (2016), afirmar que

A utilização de métodos e técnicas em pesquisa está associada aos objetivos, hipóteses e aos fundamentos teóricos do objeto de estudo. Esse procedimento exige uma escolha criteriosa e sistemática para se fazer a descrição, explicação e análise de fatos e fenômenos. Esse procedimento se faz através da *abordagem qualitativa* e da *abordagem quantitativa*, ou ainda com a aplicação

simultânea desses dois tipos de abordagem para interpretação e análise do objeto de estudo. (OLIVEIRA, 2016, p. 58)

De acordo com Oliveira (2016, p. 58) cada abordagem apresenta uma característica específica nos tratamentos dos dados coletados sendo a “abordagem quantitativa voltada para dados mensuráveis através da utilização de recursos e técnicas estatísticas. Já a abordagem qualitativa se preocupa com uma visão sistemática do problema ou objeto de estudo”.

Quanto ao tipo de pesquisa julgamos que esta pesquisa se caracteriza como sendo uma pesquisa diagnóstica, pois queremos apenas diagnosticar os conhecimentos dos futuros professores de matemática sobre a noção de risco probabilístico em tabelas de contingência, de acordo com as estratégias, ao qual, já apresentamos neste capítulo.

6 ANÁLISES E DISCUSSÃO DOS RESULTADOS

Neste capítulo são expostas ao leitor as análises e discussão dos resultados desta pesquisa, onde dividimos esse capítulo em dois momentos. O primeiro momento iremos destrinchar o perfil dos participantes dessa pesquisa e no segundo momento são apresentados os resultados e discussões das quatro situações problemas que foram trazidas no questionário. Durante as análises são apresentados os resultados gerais das duas turmas (quem cursou e quem não cursou a disciplina de Estatística) e em seguida os resultados individuais desses dois grupos.

Dos 46 (quarenta e seis) respondentes do questionário, 27 (vinte e sete) estudantes afirmaram não ter cursado a disciplina de Estatística, correspondendo a 58,7% do total de participantes desta pesquisa. E 19 (dezenove) estudantes asseguraram ter cursado a disciplina, totalizando 41,3% do total dos participantes.

A idade dos 46 (quarenta e seis) participantes variou entre 17 (dezessete) e 40 (quarenta) anos, obtendo uma média aritmética de 22,35 anos, também constatamos que a mediana foi de 20,50 anos. Esses dados são importantes para apontarmos alguns resultados sobre o perfil dos participantes da pesquisa, podemos concluir que, pela faixa etária a grande parte dos respondentes ao questionário concluiu o Ensino Médio recentemente. Esse resultado fica bem mais evidente quando apresentamos a moda das idades que foi de 18 anos.

Ainda sobre o perfil dos participantes, constatamos que 27 (vinte e sete) são do sexo masculino e 19 (dezenove) do sexo feminino. Em porcentagens temos que 58,7% são do sexo masculino e 41,3% são do sexo feminino. Essa informação não influenciará na análise de dados, apenas é uma informação de livre esclarecimento para o leitor.

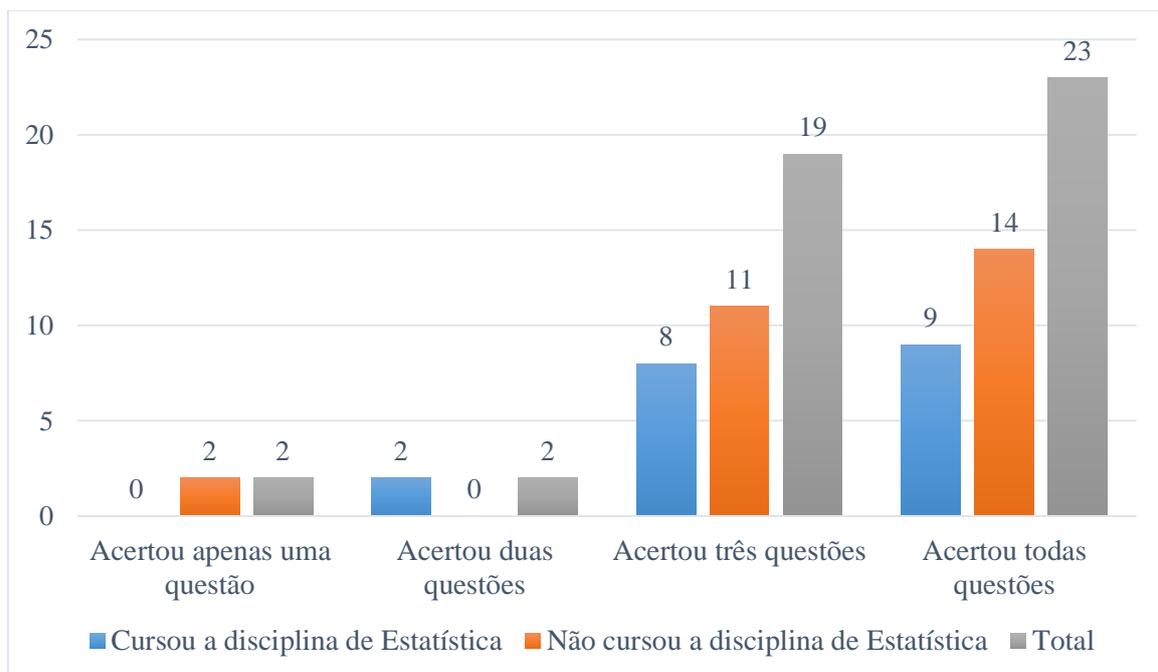
Ressaltamos que todos os respondentes do questionário são alunos do curso de Matemática-Licenciatura da Universidade Federal de Pernambuco – Centro Acadêmico do Agreste (UFPE-CAA). Dos 46 (quarenta e seis) participantes, 27 (vinte e sete) estavam no 1º período cursando a disciplina de Matemática Básica, sendo que um aluno dessa turma afirmou já ter cursado a disciplina de Estatística, 17 (dezessete) estavam matriculados no 5º período, tendo um aluno não tido cursado a disciplina de Estatística e apenas 2 (dois) estavam no 7º período, ambos cursando a disciplina de Metodologia do Ensino de Matemática 1.

É importante destacar que a amostra se distinguiu um pouco, pelo o fato de queremos confrontar os dados entre dois grupos o que cursou a disciplina e o que não cursou a disciplina de Estatística. Dando continuidade as análises e discussão dos resultados, iremos realizar uma análise geral das 4 (quatro) situações problemas, apresentando o número de acertos e erros de

cada questão e em seguida apresentamos os dados individuais por item. Também são apresentadas as principais estratégias utilizadas pelos participantes.

Apresentamos aqui algumas análises gerais do questionário. O total de acertos possível nos quatro itens é de 184, pois, é os quatro itens vezes o número de participantes da pesquisa que é 46. Essa pesquisa obteve um total de 155 acertos. Podemos afirmar que o resultado desse questionário foi bastante positivo uma vez que a média de acertos foi elevada, sendo igual a 3,37 acertos por questionário. A seguir apresentamos um gráfico contendo o número de questões acertadas por estudantes.

Gráfico 1 - Número de questões acertadas por estudantes.



Fonte: O autor (2018).

Esses resultados são importantes, pois podemos chegar a algumas conclusões através deles. Uma é que nenhum aluno errou todas questões. A segunda é que a metade dos participantes acertaram todas as questões. Se analisamos por grupo vemos que os alunos que não cursaram a disciplina obtiveram um total de 91 acertos, sendo que o máximo que eles poderiam alcançar é 108 acertos. Ou seja, resulta em um aproveitamento de 84,26% de acertos aproximadamente. Já no grupo dos estudantes que cursaram a disciplina de Estatística obteve um total de 64 acertos, sendo que o máximo de acertos que esse grupo poderia alcançar é 76 acertos, isso resulta em um percentual de 84,21% de acertos para este grupo. Com essa informação podemos chegar à conclusão que o desempenho nos dois grupos foi praticamente

igual e com isso podemos concluir com essa análise geral dos dois grupos, que pouco influenciou ter cursado a disciplina ou não, uma vez que os resultados foram semelhantes. Para melhor esclarecimento desses dados a seguir apresentaremos uma tabela, ao qual contempla as estratégias usadas pelos os licenciandos no geral e por grupo de quem cursou e quem não cursou.

Tabela 5 - Estratégias utilizadas pelos os licenciandos na resolução dos 4 itens.

Estratégias			Grupos dos estudantes		Total
			Que não cursaram a disciplina de Estatística	Que cursaram a disciplina de Estatística	
Estratégias corretas	EC. 1	Olhar apenas os dados de uma linha da tabela.	13	8	21
	EC. 2	Olhar apenas os dados de uma coluna da tabela.	13	9	22
	EC. 3	Fazer comparações entre as frequências (relativas e absolutas), proporções, razões e porcentagens.	35	10	45
	EC. 4	Acertar, porém não sabe justificar.	3	7	10
	EC. 5	Acertar, porém, errar nos cálculos de porcentagem.	2	2	4
	EC. 6	Faz uma comparação entre as células que apoiam e as que não apoiam a associação.	15	19	34
	EC. 7	Acertar utilizando a associação empírica.	2	2	4
	EC. 8	Acertar olhando apenas a célula de maior frequência na tabela.	7	5	12
	EC. 9	Acertar olhando apenas a célula de menor frequência na tabela.	1	-	1
	EC. 10	Acertar fazendo cálculos de razão entre cada célula e o seu total de indivíduos da tabela.	-	2	2
Estratégias incorretas	EI. 1	Acreditar apenas na sua crença.	4	2	6
	EI. 2	Olhar apenas a célula de maior valor.	4	4	8
	EI. 3	Não calcular as proporções corretamente.	4	-	4
	EI. 4	Olhar apenas os dados de uma linha da tabela.	2	-	2
	EI. 5	Errar e não tem certeza no que diz.	2	3	5
	EI. 6	Errar por não utilizar todas as informações da tabela.	1	3	4
Total			108	76	184

Fonte: O autor (2018).

Na tabela 5 apresentamos as estratégias que os licenciandos utilizaram na resolução dos quatro itens. Classificamos em duas categorias as estratégias identificadas durante as análises de dados. Classificamos como estratégia correta todas as estratégias que levaram os alunos ao êxito no item, ou seja, que fez com que fizesse a associação corretamente naquele item. Já, consideramos como estratégia incorreta, todas as estratégias que ocasionaram o aluno a responder erroneamente. Identificamos dez estratégias corretas e seis incorretas, a seguir será feita uma pequena explanação de cada estratégia.

EC. 1 - Olhar apenas os dados de uma linha da tabela: essa estratégia foi utilizada no total de vinte e uma vezes pelos os estudantes, salientamos que essa estratégia não garante ao aluno que em toda vez que ele for utilizar, a mesma para fazer uma associação não é garantida que acerte a associação entre as variáveis. Pois, essa estratégia consiste apenas em analisar uma parte dos dados da tabela, que no caso não é o mais aconselhado a se fazer, pois os restantes dos dados da tabela são importantes numa associação entre variáveis.

EC. 2 - Olhar apenas os dados de uma coluna na tabela: essa estratégia foi utilizada no total de vinte e duas vezes pelos os estudantes. Essa estratégia é semelhante a estratégia **EC. 1**, o que diferencia é que ao invés de utilizar a linha para se tomar uma decisão o aluno utiliza apenas uma coluna da tabela. Assim como a estratégia anterior, essa estratégia não garante ao aluno certeza numa associação.

EC. 3 - Fazer comparações entre as frequências (relativas e absolutas), proporções, razões e porcentagens: essa estratégia foi utilizada pelos os alunos quarenta e cinco vezes, esse resultado também expressar que essa estratégia foi a mais usadas pelos os participantes nesta pesquisa. Essa estratégia é considerada por Cañadas, Batanero, Contreras e Arteaga (2010), como uma estratégia correta. Ela levar o aluno a fazer uma associação entre variáveis de forma adequada e conseqüentemente o aluno acabar acertando o item. Essa estratégia se consiste em fazer comparações entre as frequências relativas ou absolutas de cada célula, inclusive das linhas e colunas marginais da tabela de contingência.

EC. 4 - Acertar, porém, não sabe justificar: essa estratégia foi utilizada 10 vezes no total pelos os participantes da pesquisa. Consideramos essa estratégia como correta pois fez com que os alunos que optaram por ela acertassem nos itens. Ou seja, os alunos que utilizaram essa estratégia acertam item, porém não dar nenhuma justificativa convincente o porquê havia pensado daquela forma.

EC. 5 - Acertar, porém, errar nos cálculos de porcentagem: essa estratégia foi utilizada apenas quatro vezes pelos os participantes. Consideramos por correta pois levou aos

alunos ao êxito, porém, essa estratégia se consiste em fazer associação através de porcentagem, no entanto os cálculos de porcentagem não fazem sentido matematicamente.

EC. 6 - Faz comparações entre as células que apoiam e as que não apoiam a associação: essa estratégia foi a segunda mais usada nesta pesquisa, obtendo um total de uso de trinta e quatro vezes. Essa estratégia é considerada por Nunes e Bryant (2011) como correta. Ela consiste em olhar as disparidades nos casos que apoiam e não apoiam e através dessa análise chegar a uma conclusão correta sobre a associação.

EC. 7 - Acertar utilizando a associação empírica: essa estratégia foi pouco utilizada, onde foi usada apenas quatro vezes. Essa estratégia segundo Carvalho (2017) e Cañadas, Batanero, Contreras e Arteaga (2010), é uma estratégia errônea, pois o aluno não analisar as informações da tabela e se deixar levar pela sua crença. Consideramos como correta pois em quatro situações levou o aluno ao êxito na associação.

EC. 8 – Acertar olhando apenas a célula de maior frequência tabela: essa estratégia foi usada no total de doze vezes, essa estratégia é apontada por Cañadas, Batanero, Contreras e Arteaga (2010) e por Carvalho (2017) como uma estratégia errônea. Porém consideramos como estratégia correta, por levar os estudantes ao êxito nas doze vezes, esse êxito pode estar associado pelo o tipo de questão onde em dois itens as variáveis estão positivamente relacionadas e em um item negativamente relacionadas. Essa estratégia se consiste em observar apenas a celular de maior valor e através dela tirar conclusões com respeito a dependência das variáveis se estão associadas ou não.

EC. 9 – Acertar olhando apenas a célula de menor frequência na tabela: essa estratégia se assemelhar bastante com a anterior, ao qual foi utilizada apenas uma vez para responder. O que diferencia é que ao invés do aluno tirar uma conclusão sobre a associação olhando para célula de maior valor, ele aqui observa apenas a célula de menor valor e acabar acertando o item.

EC. 10 - Acertar fazendo cálculos de razão entre cada célula e o total de indivíduos na tabela: essa estratégia foi utilizada apenas duas vezes, que na ocasião levou a êxito. Porém essa estratégia não é aconselhada a se usar para saber se as duas variáveis estão relacionadas, pois se consiste em fazer cálculos de razões entre cada célula com o total de indivíduos na tabela. Ao fazer isso, o aluno não sabe ao certo se estão associadas ou não, apenas está calculando a porcentagem de cada célula em relação ao total. Uma maneira mais aconselhada e correta para saber se as duas variáveis estão associadas ou não é fazer comparações das frequências relativas condicionais.

EI. 1 – Acreditar apenas na sua crença: essa estratégia foi utilizada seis vezes, ao qual induziu os alunos ao erro. Essa estratégia é praticamente igual a estratégia EC. 7 o que diferencia é que uma levou alguns estudantes ao êxito e outros ao erro. Cañadas, Batanero, Contreras e Arteaga (2010), apontar em seu estudo essa estratégia. Destacamos que ao analisar uma tabela qualquer não podemos desconsiderar os dados contidos na tabela e acreditar apenas na crença pessoal ou social em que as duas variáveis podem estarem associadas ou não.

EI. 2 – Olhar apenas a célula de maior valor: entre as estratégias errôneas essa foi a mais que foi utilizada, totalizando um total de oito vezes. Essa estratégia também é descrita no trabalho de idem (2010). Se consiste em olhar apenas a célula de maior valor, ou seja, o estudante tomar a decisão sem olhar o restante dos dados da tabela, ele apenas considera uma parte dos dados da tabela.

EI. 3 – Não calcular as proporções corretamente: essa estratégia se consiste em fazer cálculos de proporções erradamente e conseqüentemente acabar induzindo o aluno tirar conclusões falsas de uma possível associação ou não entre as duas variáveis. Essa estratégia foi utilizada no total de quatro vezes pelos os estudantes.

EI. 4 – Olhar apenas os dados de uma linha da tabela: apenas dois alunos utilizaram essa estratégia. Onde se consiste em verificar apenas uma linha da tabela e faz com que o aluno tome uma decisão se as duas variáveis estão associadas ou não, utilizando apenas uma parte dos dados da tabela e não todos os dados que é uma escolha mais sensata e correta.

EI. 5 – Errar e não tem certeza no que diz: essa estratégia se consiste em errar o item e não dar nenhuma justificativa convincente ou assinalar o item erroneamente e deixar em branco a justificativa. Essa estratégia foi utilizada no total de cinco vezes.

EI. 6 – Errar por não utilizar todas as informações da tabela: essa estratégia foi utilizada quatro vezes. Se consiste em tomar decisões com respeito a associação das variáveis utilizando apenas uma parte dos dados da tabela, o que faz com que o aluno não acerte a associação entre as variáveis no item.

Para o melhor esclarecimento dessas estratégias a seguir é apresentado uma análise por item, o qual apresentamos exemplos que melhor explicar as estratégias.

No primeiro item das situações problema do questionário é apresentada uma tabela de contingência com o número de pessoas que tinha o hábito de fumar ou não, e as pessoas que tinham ou não doença brônquica. Em seguida era feita uma indagação sobre se existe alguma relação entre ter a doença brônquica e o hábito de fumar. Após ser feita a indagação era pedido que os participantes explicassem o porquê havia pensado daquela maneira, usando os dados da tabela. Esse primeiro item está ilustrado abaixo na figura a seguir:

Figura 4 - item 1 das situações problemas do questionário.

	Pessoas com doença brônquica	Pessoas com nenhuma doença brônquica
Pessoas que têm hábito de fumar	90	60
Pessoas que não fumam	60	40

(Questão 1) Você acha que existe uma relação entre ter doença brônquica e o hábito de fumar?

() Sim () Não

Explique, **usando as informações da tabela**, por que você pensa assim

.....

.....

.....

Fonte: O autor (2018).

Neste item a resposta correta de acordo com os dados da tabela é a não relação entre as duas variáveis (hábito de fumar e ter doença brônquica). Uma solução correta para se chegar a essa conclusão é fichar a linha ou coluna marginal da tabela e em seguida fazer as comparações das frequências relativas. Para a melhor ilustração desse procedimento vamos construir a tabela de contingência com as linha e coluna marginais, lembrando que a linha e a coluna marginal representam o total dos indivíduos na coluna e na linha respectivamente. Na tabela 6 ilustramos a tabela de contingência do item 1 com as suas linha e coluna marginais.

Tabela 6 - Tabela do item 1 do questionário

	Pessoas com doença brônquica	Pessoas com nenhuma doença brônquica	Total
Pessoas que têm hábito de fumar	90	60	150
Pessoas que não fumam	60	40	100
Total	150	100	250

Fonte: O autor (2018).

Como já mencionamos é necessário fixar a linha ou a coluna marginal para fazer a análise correta. Se fixarmos a linha marginal temos que a frequência relativa entre a quantidade de pessoas que têm o hábito de fumar e têm a doença brônquica é de $90/150 = 0,6$ que equivale a 60% do total das pessoas que têm a doença. Por outro lado, temos que a frequência relativa das pessoas que não têm o hábito de fumar, mas têm a doença brônquica é de $60/150 = 0,4$ que corresponde a 40% do total das pessoas que têm doença.

Agora se olharmos apenas para o grupo de pessoas que não têm a doença brônquica, temos que a frequência relativa das pessoas que fumam é de $60/100 = 0,6$ o que é equivalente a 60%, ou seja, 60% das pessoas que não tem doença brônquica, têm o hábito de fumar e a frequência das pessoas que não fumam é de $40/100 = 0,4$ em outras palavras temos que 40% das pessoas que não possuem a doença, também não fumam.

Para concluir a análise da associação entre essas duas variáveis basta compararmos as informações da coluna dos totais das linhas com o total geral da tabela. Fazendo isso é a mesma coisa que de apenas considerar o grupo de pessoas que estão fazendo parte da coluna dos totais, assim temos que, a frequência relativa das pessoas que têm o hábito de fumar com o total geral de pessoas é de $150/250 = 0,6$ esse resultado apontar que 60% do total de pessoas da amostra apresentada na tabela têm o hábito de fumar e que 40% desse total de pessoas não fumam, sabemos desse resultado pois a frequência relativa das pessoas que não fumam em relação com o total de pessoas da tabela é de $100/250 = 0,4$.

Com essas informações podemos concluir que as duas variáveis não estão associadas, pois as frequências relativas dos diferentes grupos são iguais. Ou seja, a frequência relativa das pessoas que fumam dado que elas tem a doença brônquica é 60%, que é o mesmo valor da frequência relativa das pessoas que não estão com a doença dado que elas tem o hábito de fumar, que também é o mesmo valor da frequência relativa do total de pessoas que fumam em relação com o total de pessoas da pesquisa. Portanto, temos que nos três grupos (pessoas com doença brônquica, com nenhuma doença brônquica e o total) a frequência relativa é a mesma, o que implicar que as variáveis são independentes.

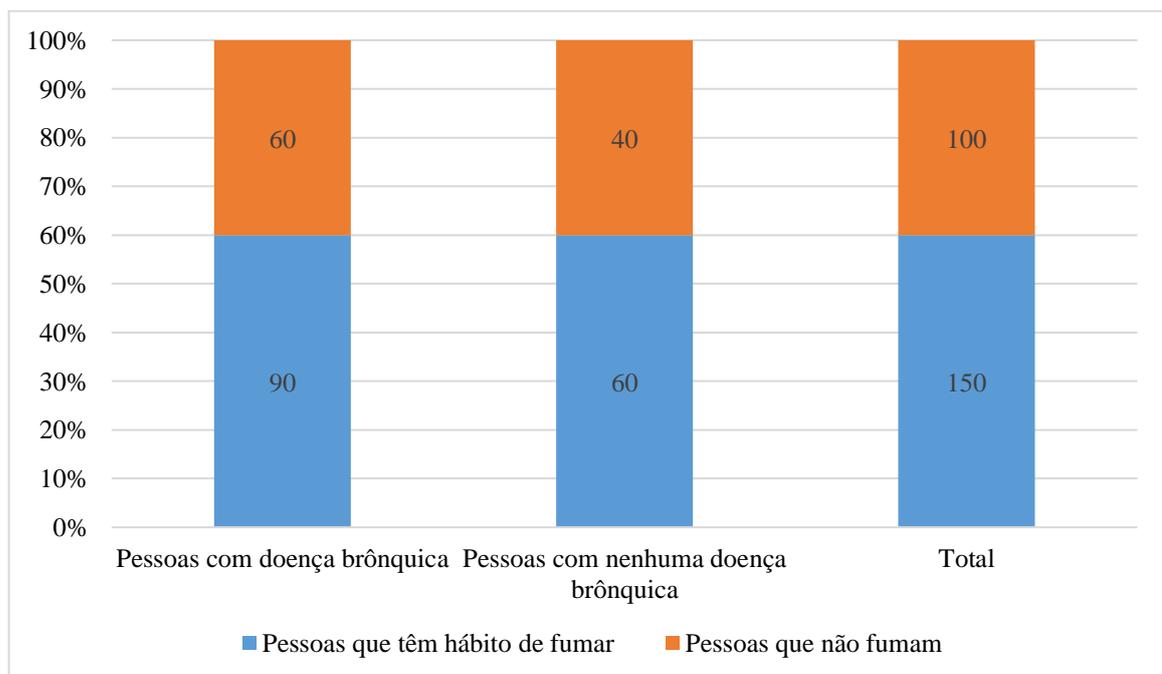
A interpretação matemática por trás desse fenômeno é que saber que a variável “ter doença brônquica” aconteceu, não afeta a probabilidade de ocorrência do evento “pessoa que tem o hábito de fumar”, ou vice-versa.

Da mesma maneira acontece com os outros casos, que são a frequência relativa dos três grupos mencionados acima com o número de pessoas que não fumam. Nos três grupos a

frequência relativa obteve o valor de 40%, o que evidencia a não associação entre as duas variáveis. A seguir apresentamos um gráfico de barras com a distribuição das frequências relativas dos três grupos.

É importante ressaltar que quando mencionamos frequência relativa das pessoas que têm o hábito de fumar e têm a doença é a mesma coisa de estarmos nos referindo a probabilidade condicional da pessoa ter o hábito de fumar (F) dado que essa pessoa tem a doença (D). Ou seja, a probabilidade condicional, ao qual nos referimos é $P(F|D)$.

Gráfico 2 - Distribuição das pessoas com/sem a doença brônquica com relação as pessoas que têm/não o hábito de fumar.



Fonte: O autor (2018).

Como é ilustrado no gráfico de barras na figura 4, temos que independentemente de qualquer grupo de pessoas as duas variáveis (ter doença brônquica e o hábito de fumar) não estão associada, ou seja, uma variável não influencia na outra. Sabemos que existe três tipos de associação de variáveis que são: dependência direta, dependência indireta e independência; no caso desse item a associação é de independência das variáveis. Com isso temos que o coeficiente de correlação entre as duas variáveis é igual a zero, pois elas são independentes.

Com esta análise podemos concluir que a associação empírica dos dados da tabela não coincide com os as teorias prévias, ou seja, denominamos por teorias prévias o senso comum do participante, por exemplo, de acordo com os dados da tabela, neste item, o cigarro não está

associado com ter doença brônquica, mas sabemos que o uso do cigarro está diretamente ligado a diversas doenças da atualidade. Ou seja, neste item os dados apresentados na tabela são contraditórios as teorias prévias.

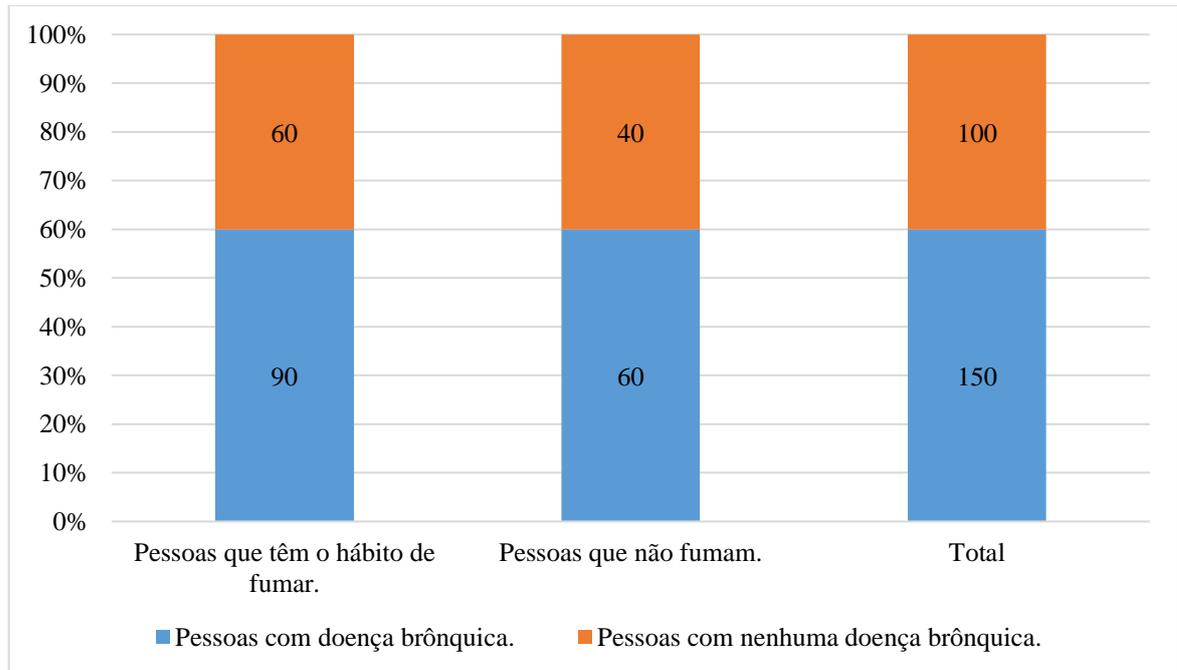
Como já mencionamos que é possível chegar a esse resultado fixando a coluna marginal da tabela 5, o procedimento é análogo ao que foi feito na linha marginal. Ilustraremos esse procedimento a seguir para o melhor entendimento do leitor e notar que tanto faz escolher a linha ou a coluna marginal para fazer a associação correta.

Sendo assim, agora fixaremos a coluna marginal, onde, apresentar os totais de cada linha da tabela. Após fixar a coluna marginal iremos calcular a frequência relativa dos três grupos de pessoas (que têm o hábito de fumar, não fumam e o total). Se calcularmos a frequência relativa das pessoas que têm o hábito de fumar e têm a doença brônquica é de $90/150 = 0,6$ e a frequência relativa das pessoas que têm o hábito de fumar mas não têm nenhuma doença é de $60/150 = 0,4$; ou seja, 60% das pessoas que têm o hábito de fumar, também têm doença brônquica e 40% não possuem a doença.

Temos o mesmo resultado se calcularmos a frequência relativa das pessoas que não fumam, porém possuem doença brônquica que é de $60/100 = 0,6$. Já a frequência relativa das pessoas que não fumam e também não possuem a doença é de $40/100 = 0,4$. O mesmo acontece com o total de pessoas da tabela se calcularmos a frequência relativas das pessoas com e sem doença brônquica que é de $150/250 = 0,6$ e $100/250 = 0,4$ respectivamente. Observe que, nos dois grupos podemos afirmar que, 60% das pessoas têm doença brônquica e 40% não têm a doença. Portanto, podemos concluir que as variáveis não possuem nenhuma relação. No gráfico 3 é ilustrado um gráfico de barras, todas as frequências relativas das três linhas em relação com a coluna marginal.

Novamente, saber que uma das duas variáveis aconteceu não afeta a probabilidade de que a outra aconteça. Isso faz com que possamos escolher a linha marginal ou a coluna marginal e o resultado de nossa análise deverá ser o mesmo.

Gráfico 3 - Distribuição das pessoas que têm o hábito de fumar em relação ao número de pessoas com doença brônquica.



Fonte: O autor (2018).

Esse procedimento, ao qual detalhamos anteriormente para analisar os dados da tabela e deduzir se uma variável está relacionada/associada a outra variável é uma das estratégias corretas trazida por Canãdas, Batanero, Contreras e Arteaga (2010).

Após apresentar uma das possíveis associações entre as duas variáveis do item 1, apresentamos agora os resultados desse item. Começamos apresentando o número de acertos e erros dos 46 (quarenta e seis) participantes da pesquisa. Julgamos como acerto quem assinalou a não relação entre as duas variáveis e como erro que achou que sim, sobre existir relação entre as duas variáveis.

Do total de participantes, temos que 19 (dezenove) assinalaram erroneamente que existe uma relação entre as duas variáveis, o que corresponde a 41,3% do total. Já 27 (vinte e sete) participantes afirmaram que de acordo com os dados da tabela ter a doença brônquica não está relacionada com o hábito de fumar, então temos que do total de respondentes do questionário, 58,7% acertaram a associação entre as duas variáveis do item 1.

Por outro lado, se compararmos os acertos e erros dos dois grupos (quem cursou e quem não cursou Estatística), temos que 10 (dez) alunos que cursaram a disciplina, assinalaram erroneamente que sim, as duas variáveis estão associadas e 9 (nove) marcaram corretamente

que as variáveis não estão associadas. Portanto, aproximadamente 52,6% dos alunos que cursaram a disciplina erraram e aproximadamente 47,4% dos alunos acertaram.

Já no grupo dos alunos que não cursaram a disciplina de Estatística, temos que 9 (nove) responderam erroneamente que sim existe uma relação entre as duas variáveis e 18 (dezoito) assinalaram corretamente que as variáveis não estão relacionadas. Ou seja, desse grupo de alunos que não cursaram a disciplina 66,7% aproximadamente acertaram a associação e 33,3% dos alunos desse grupo não acertaram a associação entre as duas variáveis.

Com essa análise de acerto e erro por grupo, podemos levantar algumas hipóteses para esse resultado, a primeira é que há uma disparidade considerável no número de acertos e erros nos dois grupos. Podemos perceber que os alunos que não cursaram a disciplina de Estatística tiveram um desempenho melhor do que o grupo dos alunos que não cursaram a disciplina. É bom salientar que dos 27 (vinte e sete) alunos do 1º período apenas um já tinha cursado a disciplina de Estatística, o que nos leva a acreditar que esse grupo de alunos ainda está com os conhecimentos do Ensino Médio fortemente atrelados e por esse motivo tenham se saído melhor do que o grupo de alunos que já cursaram a disciplina. Entretanto, para confirmar nossas suspeitas, um teste estatístico mais detalhado deve ser aplicado para confirmar essa hipótese, mas esse não é o foco do nosso trabalho. Esse resultado é no mínimo contraditório já que é de se esperar que quem tenha cursado a disciplina de Estatística tenha mais artifícios para interpretar e fazer a associação correta entre essas duas variáveis.

Para o melhor entendimento dos resultados iremos destrinchar as estratégias utilizadas pelos os participantes e em seguida apresentaremos por grupo. A priori identificamos 11 (onze) estratégias diferentes utilizadas pelos os participantes neste item 1, sendo que 5 (cinco) estratégias levaram os estudantes a assinalar corretamente que não existe associação entre ter o hábito de fumar e ter doença brônquica. E 6 (seis) estratégias que os alunos adotaram erroneamente marcando sim no questionário.

A seguir apresentamos as estratégias que levamos em consideração durante a análise do questionário, apresentando o quantitativo para o item 1, em uma tabela. Nesta tabela apresentamos também o quantitativo geral e por grupo que essas estratégias foram utilizadas.

Tabela 7 - Estratégias utilizadas pelos os estudantes para responder o item 1.

Estratégias			Grupos dos estudantes		Total
			Que não cursaram a disciplina de Estatística	Que cursaram a disciplina de Estatística	
Estratégias corretas	EC. 1	Olhar apenas os dados de uma linha da tabela.	1	1	2
	EC. 2	Olhar apenas os dados de uma coluna da tabela.	2	-	2
	EC. 3	Fazer comparações entre as frequências (relativas e absolutas), proporções, razões e porcentagens.	11	4	15
	EC. 4	Acertar, porém não sabe justificar.	2	4	6
	EC. 5	Acertar, porém, errar nos cálculos de porcentagem.	2	-	2
	EC. 6	Faz uma comparação entre as células que apoiam e as que não apoiam a associação.	-	-	-
	EC. 7	Acertar utilizando a associação empírica.	-	-	-
	EC. 8	Acertar olhando apenas a célula de maior frequência na tabela.	-	-	-
	EC. 9	Acertar olhando apenas a célula de menor frequência na tabela.	-	-	-
	EC. 10	Acertar fazendo cálculos de razão entre cada célula e o seu total de indivíduos da tabela.	-	-	-
Estratégias incorretas	EI. 1	Acreditar apenas na sua crença.	1	2	3
	EI. 2	Olhar apenas a célula de maior valor.	4	4	8
	EI. 3	Não calcular as proporções corretamente.	3	-	3
	EI. 4	Olhar apenas uma das linhas da tabela.	1	-	1
	EI. 5	Errar e não tem certeza no que diz.	-	2	2
	EI. 6	Errar por não utilizar todas as informações da tabela.	-	2	2
Total			27	19	46

Fonte: O autor (2018).

Na tabela 7, apresentamos todas as estratégias que identificamos durante a análise de dados desse item, apresentaremos uma explicação formal e detalhada de cada estratégia dessa, em alguns casos apresentaremos imagens de alguns protocolos que demonstre a estratégia.

EC. 1- Olhar apenas os dados de uma linha da tabela: essa estratégia foi utilizada por dois alunos. Ressaltamos que essa estratégia não garante ao aluno certeza numa associação, pois o aluno está apenas analisando uma parte dos dados da tabela, que no caso não é o mais aconselhado a se fazer. Um exemplo, dessa situação é analisar apenas a linha das pessoas que

não fumam, como resultado desse exemplo, temos que 3/5 dessas pessoas têm a doença brônquica e 2/5 não têm a doença, ou seja, isso pouco nos diz se existe uma associação ou não entre essas duas variáveis. Para confirmar, essa associação ou não era necessário que o aluno fizesse esse mesmo procedimento na linha dos fumantes. Ressaltamos que classificamos essa estratégia como correta, pois fez com que dois estudantes acertassem este item. A seguir apresentamos algumas respostas de alguns alunos referente a essa estratégia.

Aluno 01: Não entendo como uma relação, acredito que no caso acima mostra *que é mais fácil* ter bronquite sendo fumante, mas isso não exclui a porcentagem de que mesmo não fumando pode-se adquirir a doença.

Aluno 34: O hábito de fumar não implica no fato da pessoa ter doença brônquica pois, pessoas que não fumam e que não possui nenhuma doença é bem abaixo da porcentagens de pessoas que não fumam e possuem a doença. Pode sim, ser o causador para algumas pessoas, considerando os fumantes passivos, mas não é a única causa.

Note que, o aluno 34 tentar justificar o fato das pessoas que não fumam ter um número maior de pessoas com a doença, através da sua crença afirmando que este fato pode ter relação com a inalação da fumaça do cigarro, que no caso são os fumantes passivos referido pelo o aluno. Ressaltamos que a questão problema não se referia a fumantes passivos, acreditamos que o aluno justificar esse fato pela sua crença.

EC. 2 - Olhar apenas os dados de uma coluna da tabela: nessa estratégia o aluno utiliza o mesmo raciocínio da estratégia anterior, só que ao invés de usar a linha da tabela, utilizar-se a coluna da tabela. Ressaltamos que essa estratégia não garante ao aluno certeza numa associação, classificamos como correta pois, para dois alunos ela levou a êxito neste item. A seguir apresentamos uma resposta de um aluno que utilizou essa estratégia.

Aluno 02: A quantidade de pessoas que fumam e não tem doença brônquica é maior que a quantidade das pessoas que não fumam e não tem doenças brônquica.

EC. 3 - Fazer comparações entre as frequências (relativas e absolutas), proporções, razões e porcentagens: essa estratégia foi a mais que foi usada pelos os participantes neste item, ao todo foi utilizada pelos os participantes da pesquisa 15 (quinze) vezes. Essa estratégia é considerada por Cañadas, Batanero, Contreras e Arteaga (2010), como uma estratégia correta. Essa estratégia leva o aluno a fazer uma associação correta no item e em qualquer outra tabela

de contingência. A seguir apresentamos uma imagem de um protocolo desta pesquisa, que servirá para exemplificar melhor essa estratégia.

Figura 5 - Protocolo do item 1, aluno 18.

	Pessoas com doença brônquica	Pessoas com nenhuma doença brônquica
Pessoas que têm hábito de fumar	90	60
Pessoas que não fumam	60	40

(Questão 1) Você acha que existe uma relação entre ter doença brônquica e o hábito de fumar?

() Sim

(x) Não

Explique, usando as informações da tabela, por que você pensa assim

$150 - 100$ $150 - 100$ $100 - 100$ $100 - 100$
 $90 - x$ $60 - x$ $60 - x$ $40 - x$
 $150x = 9000$ $150x = 6000$ $x = 60\%$ $x = 40\%$
 $x = 90\%$ $x = 60\%$
 15 15 Com uso de porcentagem percebe-se
 $x = 60\%$ $x = 40\%$ que para se ter a doença supracitada
 NÃO É NECESSÁRIO SER FUMANTE.

Fonte: O autor (2018).

Note que, o estudante faz o cálculo de porcentagem para a primeira linha/coluna da tabela e em seguida faz para a segunda linha/coluna da tabela e em seguida compara essas porcentagens uma com a outra e tomar uma decisão correta sobre a associação entre as duas variáveis. Essa decisão é fácil ver quando o aluno 18 responde que “Com uso de porcentagem percebe-se que para se ter a doença supracitada não é necessário ser fumante”.

EC. 4- Acertar, porém não sabe justificar: essa estratégia foi utilizada por 6 (seis) estudantes, ou seja, assinalou corretamente que não existe nenhuma relação entre as duas variáveis, porém não deram uma justificativa convincente do porquê haviam pensado dessa forma. Na imagem a seguir exemplificaremos melhor o que acontece nessa estratégia.

Figura 6 - Protocolo do item 1, aluno 06.

(Questão 1) Você acha que existe uma relação entre ter doença brônquica e o hábito de fumar?

Sim

Não

Explique, usando as informações da tabela, por que você pensa assim

Se fossemos analisar apenas segundo dados da tabela, temos que levar em conta que o quantitativo de entrevistados que não fumam é consideravelmente menor do que as pessoas que fumam, em uma análise superficial, creio que se o quantitativo fosse bem igual.

Fonte: O autor (2018).

Vejamos que o aluno 06, até tentar justificar com uma certa coesão no começo de sua explicação, porém ele não conclui o porquê as duas variáveis não estão associadas. A seguir apresentamos algumas respostas de alguns alunos que responderam este item, mas não justificaram corretamente a sua resposta:

Aluno 06: Se fossemos analisar apenas segundo dados da tabela, temos que levar em conta que o quantitativo de entrevistados que não fumam é consideravelmente menor do que as pessoas que fumam, em uma análise superficial, creio que se o quantitativo fosse bem igual.

Aluno 31: 60 pessoas que não fumam têm doença brônquica já 40 pessoas que não fumam não têm nenhuma doença brônquica. Ou seja, nesse caso não dá para ter certeza que existe uma relação entre o hábito de fumar e ter doença brônquica.

Aluno 40: Podemos observar que a quantidade de pessoas não fumante e que têm a doença, é uma quantidade significativa assim como as pessoas fumantes. Porém não podemos afirmar que o hábito de fumar está, exclusivamente, relacionado a doença brônquica.

A justificativa do aluno 6 é incompleta, por isso não é possível fazer uma análise mais criteriosa a respeito dela. Na verdade, não importa se o quantitativo de fumantes é maior do que o de não fumantes, pois devemos levar em consideração a proporção nos dois grupos e a verificar essas proporções constataríamos que são iguais, ou seja, ter doença brônquica nesta amostra da tabela independe do hábito de fumar já que as chances são iguais nos dois grupos.

O aluno 31, levar em consideração na sua resposta apenas a linha de não ser fumante e conclui erroneamente que não dá para ter uma certeza se o hábito de fumar tem relação com a doença brônquica. Se o aluno tivesse analisado a linha das pessoas fumantes e tivesse calculado a frequência relativa nos dois grupos, poderia chegar a conclusão que as variáveis são independentes.

Já o aluno 40, tentar justificar parecido com o aluno 06, porém conclui que não pode afirmar que o hábito de fumar está diretamente, relacionado com ter doença brônquica. Ressaltamos que nesses três casos, se fazia necessário para os alunos fazer uma análise dos dados da tabela mais cautelosamente nas informações da tabela, e o cálculo das frequências relativas mostrariam que independe de quantitativo nos dois grupos as chances de ter a doença seria a mesma.

EC. 5- Acertar, porém, errar nos cálculos de porcentagem: dois alunos do grupo dos estudantes que não cursaram a disciplina de Estatística usaram essa estratégia, onde, acertaram na associação, porém, erraram nos cálculos de porcentagens. Para o melhor entendimento do leitor optamos por ilustrar um protocolo na imagem a seguir, onde o aluno utiliza essa estratégia.

Figura 7 - Protocolo do item 1, aluno 07

(Questão 1) Você acha que existe uma relação entre ter doença brônquica e o hábito de fumar?

Sim Não

Explique, usando as informações da tabela, por que você pensa assim

não porque olhando a porcentagem
o número de pessoas que não apresentam doença
brônquica é a mesma em ambos os casos, nas
pessoas que fumam e nas que não fumam
6,6, 6%

Handwritten calculations:

$$\begin{array}{l} 60 \rightarrow 100\% \\ 90 \rightarrow X \\ \frac{60}{90} = \frac{100}{X} \\ X = \frac{100 \cdot 90}{60} \\ X = \frac{9000}{60} \\ X = 150 \end{array}$$

$$\begin{array}{l} 60 \rightarrow 100\% \\ 40 \rightarrow X \\ \frac{60}{40} = \frac{100}{X} \\ X = \frac{100 \cdot 40}{60} \\ X = \frac{4000}{60} \\ X = \frac{400}{3} \end{array}$$

Fonte: O autor (2018).

Observe que a interpretação do aluno 07 está correta, porém o aluno comete um equívoco nos cálculos das porcentagens. Note que o aluno considera 90 como sendo 100% e calcular a porcentagem de 60 com base nessa relação. O erro que o aluno cometeu foi de que 90 é uma parte do grupo, assim como 60 é, ou seja, 90 não é o total daquele grupo e sim 150.

O mesmo acontece na linha dos não fumantes, ele considera 60 como sendo 100%, na verdade 100% seria o total de indivíduos naquela linha. Como as variáveis são independentes, os dados acabaram distribuídos proporcionalmente.

Como não foi identificado nenhum registro das estratégias EC. 6; EC. 7; EC. 8; EC. 9 e EC. 10 neste item não apresentaremos exemplos ou explicações das mesmas.

EI. 1- Acreditar apenas na sua crença: ao todo, três alunos responderam este item através de sua crença, esse resultado já era esperado, uma vez que nesse item especificamente os dados da tabela eram contraditórios com a associação empírica das pessoas. Sabemos que no cotidiano o cigarro é um dos principais causadores de várias doenças no mundo. Por esse motivo já era esperado que alguns alunos respondessem este item utilizando a sua crença. O que colocamos em jogo aqui é que como futuros professores de matemática não devemos tomar decisões em tabelas ou gráficos por acreditarmos em algo, devemos analisar e interpretar esses dados. Na figura a seguir apresentamos um exemplo de como a crença influenciou na resposta.

Figura 8 - Protocolo do item 1, aluno 32.

(Questão 1) Você acha que existe uma relação entre ter doença brônquica e o hábito de fumar?

Sim () Não

Explique, usando as informações da tabela, por que você pensa assim

..... Num mundo globalizado, existe (n)s fatores contribuindo para a formação de doenças, mas, o cigarro é o principal deles; pois o fumante adquire sérias doenças e graves e as bronquites e pulmões são os mais afetados dos órgãos.

$$\frac{60 - x}{90 - 100} \Rightarrow \frac{6000}{90} = 66,67\% \Rightarrow \text{ou seja } 66,67\% \text{ das pessoas com}$$

doenças brônquicas têm hábito de fumar; enquanto 66,67% dos não fumantes não têm esse tipo de doenças.

Fonte: O autor (2018).

Note que o aluno 32 ele contextualiza sua resposta acrescentando que no mundo globalizado existe (n)s fatores que contribuem para a formação de doença e que o cigarro é um dos principais responsáveis por ocasionar doenças sérias e graves como bronquites e os pulmões são os mais afetados. Em seguida o aluno usa os dados da tabela e faz uma razão errada, pois comete o mesmo erro da estratégia anterior e mesmo calculando as porcentagens erroneamente assume que a porcentagem é a mesma no caso de fumar e ter a doença e no caso de não fumar

e não ter a doença. A seguir apresentaremos duas respostas de alunos que se deixaram levar pela a sua crença.

Aluno 05: Considerando que ao inalar a fumaça do cigarro aumentam as chances de o indivíduo ter doenças brônquicas, isso justifica o maior número de pessoas que fumam e têm doença. E a questão de ter mais gente que não tem a doença e fuma do que os que não fumam pode ter relação com a quantidade de não fumantes que têm proximidade com os fumantes e inalam essa fumaça.

Aluno 38: Acredito que sim, por se tratar de uma doença do pulmão, mas, outros fatores devem ser levados em conta, pois essas doenças também têm a ver com a genética da pessoa. Contudo o hábito de fumar agrava a chances de ter a doença.

Note que a questão pedia aos participantes que explicassem usando os dados da tabela o por que você havia pensado assim, a crença está fortemente ligada para esses estudantes, pois eles justificaram através da sua associação empírica.

EI. 2- Olhar apenas a célula de maior valor: essa estratégia é bastante conhecida, Cañada, Batanero, Contreras e Arteaga (2010), apresenta no seu estudo como uma estratégia incorreta, ou seja, que levar o aluno ao erro. Ao todo 8 (oito) estudantes optaram por usar essa estratégia, que se constitui em observar apenas a célula na tabela de maior valor ou de maior frequência e fazer a associação através dela. Em seguida apresentamos algumas respostas de alguns alunos utilizando essa estratégia.

Aluno 09: Porque as pessoas que têm o hábito de fumar estão em maior número de doença brônquica com relação ao restante.

Aluno 11: 90 pessoas que fumam têm doença brônquica, enquanto apenas 60 pessoas que não fumam têm doença brônquica.

Aluno 12: Há uma maior quantidade de pessoas que fumam, com doença brônquica.

Aluno 39: Sim, os dados mostram que pessoas com o hábito de fumar apresentam uma probabilidade maior de ter doença brônquica.

O aluno 39 até tenta associar a probabilidade do evento, ser fumante, com a questão da dependência ou independência, mas para isso é necessário avaliar as probabilidades condicionais, dado que o outro evento de interesse aconteceu ou não. Neste caso a probabilidade de ser fumante não é alterada.

EI. 3- Não calcular as proporções corretamente: 3 (três) alunos utilizaram essa estratégia, que se consisti em errar no cálculo das proporções e conseqüentemente acabar levando o aluno a fazer uma associação errada. Apresentamos na figura abaixo um protocolo que exemplificar essa estratégia.

Figura 9 - Protocolo do item 1, aluno 17.

(Questão 1) Você acha que existe uma relação entre ter doença brônquica e o habito de fumar?

Sim () Não

Explique, usando as informações da tabela, por que você pensa assim

Sim, porque analisando a tabela identifiquei que as pessoas que fumam tem 50% a mais de chances de ter uma doença brônquica.

Fonte: O autor (2018)

O aluno 17, afirma que as pessoas que fumam são mais propícias para ter a doença, ainda ressalta que as pessoas que fumam têm 50% a mais chances de ter bronquite do que uma pessoa não fumante. Vimos que isso não é verdade para esses dados, já que as chances são a mesma nos dois grupos.

EI. 4- Olhar apenas a linha das pessoas que não fumam ou a linha das pessoas que fumam: essa estratégia é praticamente igual a “EC. 1”, sendo que para esse estudante essa estratégia levou-o a cometer um erro na associação entre as duas variáveis. Ressaltamos que é necessário utilizar todos os dados da tabela, para chegar a uma conclusão. Nunes e Bryant (2011), advogam que o aluno tem que saber delimitar o seu espaço amostral e quantificar o seu espaço amostral para poder chegar a uma posição sobre a associação. O aluno 26 utiliza essa estratégia e acaba tomando uma decisão sobre precipitada sem ao menos analisar todas células da tabela.

Aluno 26: Se 60 não fumam e possuem a doença e 40 não tem nenhuma doença em um total de 100 pessoas. Ao aumentar o grupo pesquisado com os fumantes vemos uma quantidade significativa de pessoas com a doença.

EI. 5- Errar e não tem certeza no que diz: ao todo, dois alunos responderam de formar errônea que sim as duas variáveis estão relacionadas, porém não souberam explicar o porquê chegaram a essa conclusão. A seguir apresentamos essas respostas:

Aluno 41: A tabela me deixa um pouco confusa, pois ao mesmo tempo que pessoas que fumam tem uma maior quantidade de doenças também tem a maior quantidade de nenhuma doença.

Aluno 42: Porque de acordo com a tabela é possível verificar a inferência que está sendo mostradas e nos faz concluir uma ou mais informações através desta tabela.

EI. 6- Errar por não utilizar todas as informações da tabela: apenas dois alunos acabaram errando por não usar todas as informações da tabela, a seguir apresentamos a resposta dos alunos e destrinchamos melhor essa estratégia.

Aluno 43: Analisando a tabela percebo que o número de pessoas que fumam e com doença brônquica é superior ao número de pessoas que não fumam e tem a doença. Porém 40% das pessoas que fumam não possuem a doença.

Aluno 46: Pelos dados apresentados demonstram que a quantidade de pessoas com doenças brônquica é maior para aquelas pessoas que têm hábito de fumar.

Note que o aluno 43, até calcula uma porcentagem, restou fazer o mesmo para as demais células, o aluno apresenta nesta resposta apenas uma análise da coluna das pessoas que tem a doença brônquica. O aluno 46 apenas olha para a coluna das pessoas que têm bronquite, faltou ele analisar todos os dados da tabela.

O levantamento dessas estratégias se faz necessário na análise dessa pesquisa pois serve para interpretar melhor o que aconteceu nas respostas dos participantes e se realmente coincide com a análise quantitativa deste item. Vimos que alguns alunos acertaram usando uma estratégia não confiante para se fazer uma associação, pois os alunos se baseavam apenas em uma parte dos dados da tabela, essa mesma estratégia fez alguns alunos tirarem conclusões precipitadas, o que, ocasionou a fazer associações errôneas. Também constatamos que os erros se diferenciaram, por exemplo um grupo de estudantes se deixou levar pela a sua crença, outro grupo errou por olharem apenas a célula de maior valor.

No segundo item das situações problemas do questionário, é apresentada aos alunos uma tabela de contingência, ao qual, apresenta em suas colunas as pessoas que tem alergia de pele ou não e nas suas linhas as pessoas que não praticam muito exercício e as que fazem exercício regularmente. Em seguida pedíamos que o participante da pesquisa analisasse se existe alguma relação entre as duas variáveis (ter alergias de pele e não fazer muito exercício). Posteriormente era pedido que os alunos explicassem usando as informações da tabela o porquê havia pensado daquela maneira. O item 2 das situações problemas do questionário está ilustrado na figura 10:

Figura 10 - item 2 das situações problemas do questionário.

	Pessoas que tem alergias de pele	Pessoas que não tem alergias de pele
Pessoas que não fazem muito exercício	130	30
Pessoas que fazem exercício regularmente	20	120

(Questão 2) Você acha que existe uma relação entre não fazer exercícios e ter alergias?

() Sim () Não

Explique, **usando as informações da tabela**, por que você pensa assim

.....

.....

.....

Fonte: O autor (2018)

Observe que podemos afirmar que as duas variáveis estão associadas, apenas olhando para os dados da tabela. Note que, há uma disparidade considerável nos valores das células que confirmam a associação entre as duas variáveis, que no caso é a célula das pessoas que não fazem muito exercício (pessoas sedentárias) e as pessoas que tem alergias de pele, em termos matemáticos essa célula é sim-sim e que têm uma frequência absoluta correspondente ao valor de 130. A outra célula que confirmar a associação entre as duas variáveis é a não-não, que no caso da tabela do item 2 são as pessoas não sedentárias e as pessoas que não tem alergias de pele, que tem um valor correspondente a 120 pessoas. Enquanto, que as células que não confirma a associação entre as duas variáveis são as pessoas que tem uma vida sedentárias e não possuem alergias de pele (sim-não), que é correspondente a 30 pessoas, e a outra célula que não apoia a associação é as pessoas que praticam exercícios físicos regularmente e tem alergias de pele (não-sim), que equivale ao um total de 20 pessoas.

Então se comparamos os casos que apoiam a associação e os que não apoiam, podemos notar uma disparidade grande entre os valores, que no caso é 130 e 120 para os casos que apoiam a associação (sim-sim e não-não, respectivamente), e os que não apoiam a associação é 20 (não-sim) e 30 (sim-não). Nunes e Bryant (2011), em sua pesquisa defendem que quando há essa disparidade considerável entre os valores das frequências que apoiam e não apoiam a associação podemos concluir que as variáveis estão relacionadas. Na verdade, os autores

destacar que notar essa disparidade é saber definir o espaço amostral e fazer a quantificação das probabilidades, que no caso chamamos de frequências.

O que Nunes e Bryant (2011), querem dizer é que quando as probabilidades condicionais que confirmam a associação forem maiores ou menores proporcionalmente podemos afirmar que as variáveis estão associadas. Se olharmos para os dados da tabela do item 2, temos que, a probabilidade de escolhermos uma pessoa aleatoriamente que tem uma vida sedentária dado que já tenha alergias de pele é igual a $P(S|A)$, onde, S representa o grupo de pessoas que tem uma vida sedentária e A o grupo de pessoas que possuem alergia de pele. Assim, temos que:

$$P(S|A) = \frac{P(S \cap A)}{P(A)} = \frac{\frac{130}{300}}{\frac{150}{300}} = \frac{130}{300} \times \frac{300}{150} = \frac{130}{150} = \frac{13}{15} \cong 0,87$$

Pois, na tabela 2 vimos que, $P(S \cap A) = 130/300$ e $P(A) = 150/300$.

Por outro lado, se calcularmos as demais probabilidades condicionais podemos notar que as duas variáveis estão relacionadas e estão diretamente associadas, pois a chance de escolher uma pessoa que faça parte dos casos que confirmam a associação é bem mais provável do que nos casos que não apoiam a associação. A seguir apresentaremos os demais cálculos das probabilidades condicionais, que podemos fazer através da tabela do item 2.

$$P(\text{não}S|A) = \frac{P(\text{não}S \cap A)}{P(A)} = \frac{\frac{20}{300}}{\frac{150}{300}} = \frac{20}{300} \times \frac{300}{150} = \frac{20}{150} = \frac{2}{15} \cong 0,13$$

$$P(S|\text{não}A) = \frac{P(S \cap \text{não}A)}{P(\text{não}A)} = \frac{\frac{30}{300}}{\frac{150}{300}} = \frac{30}{300} \times \frac{300}{150} = \frac{30}{150} = \frac{1}{5} = 0,20$$

$$P(\text{não}S|\text{não}A) = \frac{P(\text{não}S \cap \text{não}A)}{P(\text{não}A)} = \frac{\frac{120}{300}}{\frac{150}{300}} = \frac{120}{300} \times \frac{300}{150} = \frac{120}{150} = \frac{4}{5} = 0,80$$

Observe que para esses dados fictícios da tabela do item 2, temos que, para cada 15 pessoas que tem alergia de pele 13 delas tem uma vida sedentárias e apenas 2 (duas) pessoas

fazem exercício físicos, ou seja, as pessoas que não tem uma vida sedentária. Também temos que, de cada 5 pessoas que fazem exercícios regularmente, apenas 1 (uma) pessoa possui alergia de pele e 4 (quatro) pessoas não tem alergia de pele. Essas informações nos dizem que as duas variáveis não estão perfeitamente relacionadas, pois não exclui a probabilidade de quem faz exercício físicos regularmente ter problema de alergia. Porém podemos concluir que existe sim uma correlação entre as duas variáveis.

Calcular a probabilidade condicional do evento “ter alergia”, dado que o evento “ser sedentário ou não” aconteceu nos dá resultados análogos. Uma outra forma de chegar a esses resultados é calcular as probabilidades condicionais, agora em relação ao grupo de pessoas sedentárias e não sedentárias. A seguir apresentaremos os referidos cálculos.

$$P(A|S) = \frac{P(A \cap S)}{P(S)} = \frac{\frac{130}{300}}{\frac{160}{300}} = \frac{130}{300} \times \frac{300}{160} = \frac{130}{160} = \frac{13}{16} \cong 0,81$$

$$P(\text{não}A|S) = \frac{P(\text{não}A \cap S)}{P(S)} = \frac{\frac{30}{300}}{\frac{160}{300}} = \frac{30}{300} \times \frac{300}{160} = \frac{30}{160} = \frac{3}{16} \cong 0,19$$

$$P(A|\text{não}S) = \frac{P(A \cap \text{não}S)}{P(\text{não}S)} = \frac{\frac{20}{300}}{\frac{140}{300}} = \frac{20}{300} \times \frac{300}{140} = \frac{20}{140} = \frac{1}{7} \cong 0,14$$

$$P(\text{não}A|\text{não}S) = \frac{P(\text{não}A \cap \text{não}S)}{P(\text{não}S)} = \frac{\frac{120}{300}}{\frac{140}{300}} = \frac{120}{300} \times \frac{300}{140} = \frac{120}{140} = \frac{6}{7} \cong 0,86$$

Com essa outra forma de identificar a associação podemos novamente chegar a mesma conclusão anterior que é a existência de uma relação entre as duas variáveis. Se pode chegar a essa conclusão apenas olhando para os resultados dos cálculos das probabilidades, temos que, de cada 16 pessoas sedentária, 13 tem alergias de pele e apenas 3 (três) pessoas não tem problema de alergia de pele. Também é possível notar que de cada 7 pessoas que não tem uma vida sedentária (ou seja, pessoas que fazem exercícios regularmente), 6 (seis) pessoas não tem alergias de pele e apenas 1 (uma) pessoa tem alergias de pele.

Portanto, é perceptível que existe uma relação entre quem tem alergias de pele e tem uma vida sedentária, pois é mais propício você apresentar alergias de pele se não realizar

exercício físicos com uma certa frequência. Da mesma forma, acontece com quem faz exercício regularmente e não tem a alergia de pele.

Quanto à existência das teorias prévias, entre essas duas variáveis podemos afirmar que no cotidiano não existe nenhuma associação entre não fazer exercícios e ter alergias de pele, porém os dados da tabela no item 2 afirmaram a associação entre as duas variáveis. Ou seja, levando em consideração o senso comum das pessoas há uma *inexistência* das teorias prévias entre essas duas variáveis.

Destacamos a seguir, os resultados das respostas do item 2 do questionário apenas olhando se o aluno acertou ou não a associação sem olhar as estratégias desenvolvidas por eles. Então, temos que dos 46 (quarenta e seis) respondentes desse item 42 (quarenta e dois) alunos acertaram afirmando que sim existe uma relação entre ter uma vida sedentária ou não e ter alergias de pele ou não, apenas 4 (quatro) alunos assinalaram incorretamente que não existe uma relação entre as duas variáveis. Transformando esses valores em porcentagens temos que 91,3% do total de alunos acertou assinalando que sim existe uma relação entre as duas variáveis e 8,7% do total erraram marcando que não existe uma relação entre as duas variáveis.

Se olharmos para o grupo de 19 (dezenove) alunos que cursaram a disciplina de Estatística, destacamos o resultado que todos os alunos acertaram o item 2, onde, assinalaram positivamente que sim para a existência de uma relação entre não fazer exercícios e ter alergias de pele. Do grupo de 27 (vinte e sete) alunos que não cursaram a disciplina, obtemos o resultado de que, 23 (vinte e três) desses estudantes acertaram o item 2 afirmando positivamente que sim para o relacionamento entre as duas variáveis e apenas 4 (quatro) alunos erraram este item afirmando o não relacionamento entre as duas variáveis.

Em termos de porcentagem para os dois grupos, temos que, para o grupo de alunos que cursou a disciplina 100% dos alunos acertaram a este item. Já no grupo dos alunos que não cursaram a disciplina, temos que aproximadamente 85,19% dos alunos acertaram o item 2 e 14,81% dos alunos erraram assinalando que não.

Neste item, os resultados nos dois grupos houve uma pequena diferença de desempenho, sendo que no grupo dos estudantes que cursaram a disciplina todos os estudantes acertaram este item, já no outro grupo dos que não cursaram a disciplina quatro erraram. Mesmo assim podemos afirmar que neste item os dois grupos se sobressaíram bem, pois obtiveram um índice de acertos elevadíssimos.

Para o melhor esclarecimento dos resultados desse item apresentamos a seguir as estratégias utilizadas pelos os participantes da pesquisa, onde usaremos uma tabela indicando

o quantitativo de uso pelo os participantes. Foi identificamos 11 (onze) estratégias diferentes para este item 2, sendo 4 (quatro) estratégias incorretas e 7 (sete) corretas.

Tabela 8 - Estratégias utilizadas pelos os estudantes para responder o item 2.

Estratégias			Grupos dos estudantes		Total
			Que não cursaram a disciplina de Estatística	Que cursaram a disciplina de Estatística	
Estratégias corretas	EC. 1	Olhar apenas os dados de uma linha da tabela.	4	1	5
	EC. 2	Olhar apenas os dados de uma coluna da tabela.	5	4	9
	EC. 3	Fazer comparações entre as frequências (relativas e absolutas), proporções, razões e porcentagens.	5	-	5
	EC. 4	Acertar, porém não sabe justificar.	-	1	1
	EC. 5	Acertar, porém, errar nos cálculos de porcentagem.	-	2	2
	EC. 6	Faz uma comparação entre as células que apoiam e as que não apoiam a associação.	6	6	12
	EC. 7	Acertar utilizando a associação empírica.	-	-	-
	EC. 8	Acertar olhando apenas a célula de maior frequência na tabela.	3	4	7
	EC. 9	Acertar olhando apenas a célula de menor frequência na tabela.	-	-	-
	EC. 10	Acertar fazendo cálculos de razão entre cada célula e o seu total de indivíduos da tabela.	-	1	1
Estratégias incorretas	EI. 1	Acreditar apenas na sua crença.	1	-	1
	EI. 2	Olhar apenas a célula de maior valor.	-	-	-
	EI. 3	Não calcular as proporções corretamente.	1	-	1
	EI. 4	Olhar apenas uma linha da tabela.	-	-	-
	EI. 5	Errar e não tem certeza no que diz.	1	-	1
	EI. 6	Errar por não utilizar todas as informações da tabela.	1	-	1
Total			27	19	46

Fonte: O autor (2018).

Na tabela 8 é apresentado o quantitativo de respostas utilizando algumas das estratégias que mencionamos anteriormente, apesar que houve um quantitativo pequeno de erro neste item, identificamos pela tabela que as estratégias foram bastantes diversificadas. A seguir

apresentaremos alguns exemplos, ao qual melhor explica as estratégias utilizadas por esses estudantes.

EC. 1 - Olhar apenas os dados de uma linha da tabela: essa estratégia foi utilizada no total de cinco vezes neste item. Salientamos que essa estratégia não é a mais indicada para fazer uma associação de forma correta, pois consiste em analisar uma parte dos dados da tabela a seguir apresentamos alguns exemplos da mesma.

Figura 11 - Protocolo do item 2, aluno 19.

	Pessoas que tem alergias de pele	Pessoas que não tem alergias de pele
Pessoas que não fazem muito exercício	130	30
Pessoas que fazem exercício regularmente	20	120

(Questão 2) Você acha que existe uma relação entre não fazer exercícios e ter alergias?

Sim () Não

Explique, usando as informações da tabela, por que você pensa assim

Porque quem não faz exercício tem maior chance de ter alergias de pele. Já quem faz exercício regularmente tem menos chances de ter alergias.

Fonte: O autor (2018).

O aluno 19 justifica, argumentando que as pessoas que não fazem exercícios têm maior número de alergia na pele, ainda concluir que há um menor número de pessoas sem alergia. Note que, o aluno apenas concluir que quem tem uma vida sedentária tem maior chances de ter alergias de pele, o aluno chegou a essa conclusão utilizando apenas uma parte dos dados da tabela, o que não é o mais aconselhado a se fazer. Para se ter a certeza que as duas variáveis estão associadas ou não era necessário analisar o restante dos dados da tabela. Neste caso o aluno 19, acertou o item pois a sua análise coincidi-o com a associação correta. Em seguida apresentamos outros exemplos de respostas que fez uso dessa estratégia neste item.

Aluno 9: Porque quem pratica mais exercícios tem menos chances de ter alergia.

Aluno 12: Há uma grande diferença nos números, quem não pratica exercício, tem risco de ter alergias de pele.

Aluno 35: Porque, na tabela mostra uma grande diferença para pessoas que não fazem muito exercício tem uma diferença grande de ter alergia na pele de que não tem que é de 100 pessoas.

Observe que os três alunos justificam que as variáveis estão relacionadas olhando apenas a primeira linha da tabela, onde há uma discrepância nos valores das células. Porém seria mais aconselhado e correto se olhassem todos os dados da tabela. Não se pode tomar uma decisão olhando uma parte dos dados da tabela, pois os outros dados também são importantes.

EC. 2 - Olhar apenas os dados de uma coluna da tabela: essa estratégia foi a segunda mais utilizada neste item, sendo identificado nove vezes durante as análises. Essa estratégia é semelhante a anterior. A diferença é que ao invés de tomar uma decisão baseado em apenas uma linha da tabela, o aluno passará a tomar uma decisão baseado numa coluna da tabela. Essa estratégia é considerada errônea por Cañadas, Batanero, Contreras e Arteaga (2010), pois não utilizar todos os dados da tabela. A seguir apresentamos alguns exemplos de respostas dessa estratégia neste item 2.

Figura 12 - Protocolo do item 2, aluno 22.

(Questão 2) Você acha que existe uma relação entre não fazer exercícios e ter alergias?

Sim () Não

Explique, usando as informações da tabela, por que você pensa assim

..... faz a chance de ter alergia de pele para pessoas.....
 que não fazem exercício e 13/15 modo para quem faz só.....
 é 20/150.

Fonte: O autor (2018).

Observe que o aluno 22, calcula a proporção na primeira coluna e afirma que a chance de a pessoa ter alergias de pele é bem maior para o grupo de pessoas que não fazem muito exercício. Porém, o aluno desconsidera o restante dos dados da tabela. Novamente salientamos que a estratégia mais adequada e correta para se fazer uma associação é utilizando todos os dados da tabela. Em seguida, apresentamos algumas outras respostas que enquadrados nessa estratégia.

Aluno 11: Há uma divergência absurda entre as pessoas que não se exercitam e tem alergia e as que se exercitam e tem alergia.

Aluno 15: Pela uma simples análise pois os indivíduos que não fazem exercício têm um alto índice de terem alergia ao contrário do que fazem exercício pois os indivíduos que se exercitam têm uma pequena probabilidade de terem a alergia.

Aluno 33: Segundo os dados da tabela pessoas que não fazem muito tem mais tendência.

Aluno 38: Acredito que tem relação, já que as pessoas que não fazem muito exercícios tem mais alergias de pele, do que as pessoas que praticam exercícios regularmente.

Aluno 44: Pessoas que não possuem o hábito de praticar exercícios físicos, de acordo como os dados apresentados, desenvolvem alergias na pele, notando-se uma grande diferença dos que fazem exercícios.

Se olharmos todas respostas evidencia que os alunos olharam apenas a coluna de ter alergia de pele e com isso chegaram à conclusão que as variáveis estão relacionadas, porém seria necessário analisar a outra coluna, o que confirmaria a hipótese da primeira.

EC. 3 - Fazer comparações entre as frequências (relativas e absolutas), proporções, razões e porcentagens: essa estratégia de acordo com Cañadas, Batanero, Contreras e Arteaga (2010) é considerada como sendo umas das possíveis estratégias corretas para resolver problemas envolvendo tabelas de contingências. No total essa estratégia foi utilizada cinco vezes neste item. A seguir apresentaremos um exemplo dessa estratégia.

Figura 13 - Protocolo do item 2, aluno 18.

(Questão 2) Você acha que existe uma relação entre não fazer exercícios e ter alergias?

Sim Não

Explique, usando as informações da tabela, por que você pensa assim

$160 - 100$	$160 - 100$	$140 - 100$	$140 - 100$	A PORCENTAGEM ULTRAPASSA 80%.
$130 - x$	$30 - x$	$20 - x$	$120 - x$	
$160x = 13000$	$160x = 3000$	$140x = 2000$	$140x = 12000$	
$x = 1300$	$x = 300$	$x = 200$	$x = 1200$	
16	16	14	14	
$x = 81,25\%$	$x = 18,75\%$	$x = 14,28\%$	$x = 85,72\%$	

Fonte: O autor (2018).

Observe que o aluno 18, consegue fazer todas relações necessárias antes de tomar a decisão de que as variáveis estão relacionadas, também é perceptivo que o aluno utilizar todas

as informações da tabela. A seguir apresentamos dois protocolos que também fazem uso dessa estratégia neste item.

Aluno 08: De acordo com a tabela, as pessoas que fazem exercício regularmente têm menor probabilidade de adquirir alergias de pele. Já as pessoas que não fazem muito exercício possuem uma maior tendência de adquirir alergias de pele.

Aluno 26: Usando o grupo de amostra das pessoas que praticam exercício regularmente vemos que só 15% tem alergia, já ao comparar o grupo dos não praticantes vemos a maior parte do grupo estudado contém alergia.

EC. 4 - Acertar, porém não sabe justificar: Apenas um aluno assinalou corretamente que sim as duas variáveis estão relacionadas, porém não soube justificar a sua resposta. Consideramos esse acontecimento como uma estratégia correta pois o aluno assinalou corretamente o item. A seguir é apresentado o protocolo com a resposta do aluno que não soube explicar o porquê as duas variáveis estão relacionadas.

Aluno 23: As pessoas que fazem exercícios estão com menor índice de alergias. E o foco da pesquisa entre quem apresente alergias, portanto quem não apresente alergias, indefere se faz ou não exercício.

Note que o aluno 23, observar que as pessoas que fazem exercício têm uma menor chance de ter alergia, porém em seguida o aluno se enrolar ao tentar explicar e concluir que ter alergia de pele indefere de fazer exercício ou não. Logo, ele conclui de forma equivocada e erroneamente.

EC. 5 - Acerta, porém, errar nos cálculos de porcentagem: identificamos neste item apenas duas respostas em que o aluno faz a associação correta, porém errar nos cálculos de porcentagem. Essa estratégia ficar mais nítida com o exemplo a seguir:

Figura 14 - Protocolo do item 2, aluno 30.

	Pessoas que tem alergias de pele	Pessoas que não tem alergias de pele
Pessoas que não fazem muito exercício	130 $\frac{3}{4}$	30
Pessoas que fazem exercício regularmente	20	120 $\frac{3}{4}$

(Questão 2) Você acha que existe uma relação entre não fazer exercícios e ter alergias?

Sim () Não

Explique, usando as informações da tabela, por que você pensa assim

Porque a mesma pessoa pode fazer ou não exercícios mesmo tendo a alergia.

61,5%
25%
81,25%

Fonte: O autor (2018).

Além do aluno 30 errar nos cálculos de porcentagem e de razão ele argumentar que uma pessoa específica pode fazer ou não exercícios físicos estando com alergias. Ou seja, o aluno não entendeu o que a questão pediu.

EC. 6 - Faz uma comparação entre as células que apoiam e as que não apoiam a associação: essa estratégia foi utilizada 12 vezes neste item, essa estratégia se consiste em identificar as células que apoiam a associação e as que não apoiam a associação e fazer uma comparação ao identificar uma discrepância entre os valores assumir que está associada ou não, para este item sabemos que as variáveis estão diretamente associadas. Neste item as células que apoiam a associação são as células “ter alergia de pele e não fazer muito exercício físicos” e a “não ter alergia de pele e fazer exercício regularmente” e as células que não apoiam a associação são “ter alergia de pele e fazer exercícios físicos regularmente” e “não ter alergia de pele e não fazer muito exercícios físicos regularmente”. A seguir, apresentamos um exemplo que melhor descreve essa estratégia.

Figura 15 - Protocolo do item 2, aluno 31.

FAZ exercícios e tem alergia = 20
 não " " " " = 130 } nesse caso os não fazem exercícios provocam a alergia

	Pessoas que tem alergias de pele	Pessoas que não tem alergias de pele
Pessoas que não fazem muito exercício	130	30
Pessoas que fazem exercício regularmente	20	120

(Questão 2) Você acha que existe uma relação entre não fazer exercícios e ter alergias?
 Sim Não não faz exercícios e não tem = 30

Explique, usando as informações da tabela, por que você pensa assim faz não " " = 120

através da tabela é possível perceber que existe uma relação entre fazer exercícios e ter alergias. A quantidade de pessoas que não fazem muito exercício e tem alergias é muito grande em relação as pessoas que fazem exercícios e tem alergia.

Fonte: O autor (2018).

Note que o aluno 31 percebe que há uma discrepância nos valores das células que apoiam e não apoiam a associação entre as variáveis. E assim concluir que as variáveis estão relacionadas, podemos ver essa afirmação quando o aluno 31 argumentar que “A quantidade de pessoas que não fazem muito exercícios e tem alergias e tem alergias é muito grande em relação as pessoas que fazem exercícios e tem alergia”.

EC. 8 - Acertar olhando a célula de menor frequência na tabela: essa estratégia foi utilizada sete vezes neste item. Ressaltamos que essa estratégia é considerada por Cañadas, Batanero, Contreras e Arteaga (2010) como uma estratégia errônea onde o aluno apenas considera a célula de maior valor numérico. Consideramos nestes sete casos como correta pois levaram ao êxito neste item. A seguir apresentamos alguns exemplos que exemplifica essa estratégia.

Aluno 14: Pois a quantidade é alta de pessoas que não fazem exercícios e consequentemente têm alergias de pele.

Aluno 16: O número de pessoas que não fazem muito exercício e tem alergia é muito superior.

Aluno 17: Sim, porque a quantidade de pessoas que não fazem exercícios e tem alergias é muito alta.

Aluno 36: A tabela mostra um número bem maior de pessoas que não fazem muito exercício com alergias de pele.

Aluno 39: Sim, os dados mostram que há uma probabilidade maior de pessoas que não fazem exercícios ter alergias de pele.

Aluno 46: Porque os dados mostram que pessoas que não fazem muito exercício tem maior possibilidade de alergias de pele.

Se observamos os exemplos podemos identificar que as respostas são semelhantes e onde também evidencia que os alunos responderam o item usando como justificativa a célula de maior valor.

EC. 10 - Acertar fazendo cálculos de razão entre a célula e o seu total de indivíduos da tabela: essa estratégia só foi utilizada apenas uma vez neste item. A seguir apresentamos um exemplo de como é essa estratégia.

Figura 16 - Protocolo do item 2, aluno 28.

	Pessoas que tem alergias de pele	Pessoas que não tem alergias de pele
Pessoas que não fazem muito exercício	130 $\frac{130}{300} = \frac{13}{30}$	$\frac{30}{300} = \frac{1}{10}$ 30
Pessoas que fazem exercício regularmente	20 $\frac{20}{300} = \frac{1}{15}$	$\frac{120}{300} = \frac{2}{5}$ 120

(Questão 2) Você acha que existe uma relação entre não fazer exercícios e ter alergias?

Sim () Não

Explique, **usando as informações da tabela**, por que você pensa assim

Pela quantidade de pessoas e pela razão de proporção.

Fonte: O autor (2018).

Se observamos os cálculos realizados pelo o aluno 28, notaremos que o aluno calcula algumas razões que é o valor da célula sobre o total de indivíduos da tabela. Ou seja, o aluno calcula a probabilidade da interseção de cada célula, o mais recomendado seria calcular a probabilidade condicional de cada célula da tabela.

EI. 1 - Acreditar apenas na sua crença: essa estratégia foi usada apenas uma vez neste item. Essa estratégia consiste em acreditar apenas na sua crença para responder o problema e desconsidera as informações contidas na tabela. A seguir apresentamos um exemplo que melhor explica essa estratégia.

Aluno 32: Alergias de um modo geral com relação a pele se deve na maioria das vezes por causa do ambiente onde as pessoas estão e o principal delas é a insolação; acho que o público da pesquisa foi selecionado em um ambiente específico; mas considerando os dados desse ambiente considero que há alguma relação especificamente para tal; mas no geral isso se contrapõe; tem outros fatores: não uso de protetor; vestimentas inadequadas; má alimentação; genética familiar...

É perceptível que o aluno 32 ao responder o item 2, só levar em consideração apenas a sua crença e demais problemas cotidianos. Note também que o aluno não levou em consideração para sua resposta os dados da tabela.

EI. 3 - Não calcular as proporções corretamente: essa estratégia foi utilizada apenas uma vez neste item. Ela consiste em errar nos cálculos de proporção, ou seja, o aluno não está habituado com esse tipo de problema acabar errando cálculo triviais numa associação de variáveis.

Figura 17 - Protocolo do item 2, aluno 7.

(Questão 2) Você acha que existe uma relação entre não fazer exercícios e ter alergias?

() Sim Não

Explique, usando as informações da tabela, por que você pensa assim

pois entre as pessoas que fazem exercício o número de pessoas que apresentam alergia é de 500% a mais enquanto que nas pessoas que não fazem o número é bem menor cerca de 23,76%

130 → 100% 130x = 30
30 → X X = 300

20 → 100%
120 → X
20 x 12000
x = 2000
2000 = 600%

300/13 = 23,76%
80
(9)

Fonte: O autor (2018).

Neste exemplo o aluno acabou se precipitando no cálculo das porcentagens e consequentemente acabou errando a questão.

EI. 5 - Errar e não tem certeza no que diz: essa estratégia foi usada apenas uma vez pelos os alunos neste item. Essa estratégia é errônea e consiste em tomar uma decisão tendo dúvidas e acabar justificando algo que não tenha convicção.

Aluno 13: Sim, provavelmente com a exposição diminui o risco de alergias, visto que pessoas que fazem exercício tem menos doenças alérgicas.

Consideramos que o aluno 13, errou e não teve certeza ao justificar o seu erro, pois o aluno assinala que não há nenhuma relação entre as duas variáveis e justifica que provavelmente possa ser que exista uma associação tendo em vista que, as pessoas que fazem exercícios regularmente têm menos chance de ter alergia de pele.

EI. 6 - Errar por não utilizar todas as informações da tabela: temos que essa estratégia só foi utilizada apenas uma vez neste item, lembramos que essa estratégia é descrita por Cañadas, Batanero, Contreras e Arteaga (2010), que também o classificam como incorreta. A seguir apresentamos um exemplo que melhor exemplifica essa estratégia.

Aluno 24: Pois se observarmos os números de pessoas que fazem exercícios e tem alergias e o número de pessoas que não fazem exercícios e tem alergias os números são praticamente iguais.

Veja que o aluno utilizar apenas duas células da tabela e acabar errando o item. Além, do aluno 24, usar apenas uma parte dos dados ele comete um erro de análise dos dados ao afirmar que 20 é praticamente igual a 130.

No item 3 das situações problemas do questionário, era apresentado aos participantes da pesquisa uma tabela de dupla entrada 2×2 , onde, na tabela apresentava alguns dados fictícios referentes a um estudo se o uso de inseticida é eficaz ou não quando é aplicado para combater uma doença em algumas plantas. Nesta tabela era apresentado o número de plantas com e sem a doença nas colunas. E nas suas linhas era apresentado o número de plantas que recebeu ou não o inseticida.

Posteriormente a visualização da tabela indagávamos aos participantes se o produto era eficaz, em seguida pedíamos que explicassem usando apenas as informações trazidas na tabela o por que você pensou se o produto é eficaz sim ou não. Na figura a seguir é apresentado o item 3 das situações problemas do questionário.

Figura 18 - item 3 das situações problemas do questionário

	Plantas com doença	Plantas sem doença
Recebeu inseticida	9	36
Não recebeu inseticida	40	10

(Questão 3) O produto é eficaz?

Sim Não

Explique, **usando as informações da tabela**, por que você pensa assim

.....

.....

.....

Fonte: O autor (2018).

Neste item 3, era esperado que os alunos notassem que, como no item 2 a desproporção entre os valores das células que apoiam a associação e as que não apoia a associação entre as variáveis. Também é possível identificar a associação entre as duas variáveis seguindo os procedimentos que realizamos no item 1. Aqui apresentaremos, uma forma mais sofisticada e bem mais elaborada de se saber se as duas variáveis estão associadas, utilizaremos o que Carvalho (2017) apresenta em sua tese de doutorado, que é a ideia de razão de chance:

$$\text{Razão de chance} = \frac{9 \cdot 10}{40 \cdot 36} = \frac{90}{1440} = \frac{1}{16} = 0,0625.$$

Como a razão de chance foi um número menor do que um, temos que as variáveis estão relacionadas e ainda podemos afirmar que a probabilidade no primeiro grupo é inferior a probabilidade do segundo grupo. Ou seja, o primeiro grupo ao que referimos é a diagonal principal da tabela, em outras palavras são o grupo das plantas que recebeu o inseticida e contém a doença, também faz parte do primeiro grupo as plantas que não recebeu o inseticida e não está com a doença. Consideramos como segundo grupo as plantas que estão na diagonal secundária, que no caso é as plantas que não recebeu o inseticida e estão com a doença, também fazem parte do segundo grupo as plantas que receberam o inseticida e estão sem a doença.

Com essas afirmações podemos chegar a alguns resultados importantes sobre o relacionamento dessas duas variáveis, uma é referente ao coeficiente de correlação dessas duas

variáveis que é aproximadamente igual a $(-0,60)$. Com isso podemos afirmar que as variáveis estão inversamente associadas e o grau de correlação entre as duas variáveis é considerado forte. Sobre a concordância com as teorias prévias temos que os dados da tabela estão de acordo com associação empírica.

Aqui apresentamos os resultados quantitativos de acertos e erros do item 3, sem olhar as estratégias. Então, temos que dos 46 (quarenta e seis) alunos que responderam o questionário neste item, 42 (quarenta e dois) assinalaram corretamente que sim o produto é eficaz e apenas 4 (quatro) marcaram erroneamente que o produto não é eficaz. Em termos de porcentagem temos que, 91,3% dos estudantes acertaram este item e 8,7% assinalaram erradamente este item.

Fazendo uma análise dos resultados por grupo de quem cursou e de quem não cursou a disciplina de Estatística, temos que dos 27 (vinte e sete) alunos que não cursaram a disciplina de Estatística, 24 (vinte e quatro) deles acertaram o item 3, assinalando que sim o produto é eficaz, o que totaliza um total 88,89% dos estudantes que não cursaram a disciplina. Apenas 3 (três) alunos que não cursaram a disciplina que marcaram erroneamente que o produto não é eficaz, o que totalizar uma porcentagem de 11,11%. No grupo dos estudantes que cursaram a disciplina, tínhamos um total de 19 (dezenove) estudantes, destes 18 (dezoito) acertaram o item 3 assinalando corretamente que sim o produto é eficaz e apenas 1 (um) aluno assinalou errado que o produto não é eficaz. Em porcentagem para esse grupo temos que 94,74% acertaram este item e 5,26% errou este item.

Olhando apenas para esses resultados podemos considerar que o desempenho dos estudantes nos dois, foi similar comparado com o total de estudantes que participaram da pesquisa. Consideramos esse resultado bastante positivo, já que o índice de acerto neste item foi próximo aos 90%.

A seguir apresentamos uma tabela que quantificar as estratégias utilizadas por grupo de quem cursou e os que não cursaram a disciplina de estatística. Ao todo identificamos para este item, sete estratégias onde consideramos corretas e três incorretas.

Tabela 9 - Estratégias utilizadas pelos os estudantes para responder o item 3.

Estratégias			Grupos dos estudantes		Total
			Que não cursaram a disciplina de Estatística	Que cursaram a disciplina de Estatística	
Estratégias corretas	EC. 1	Olhar apenas os dados de uma linha da tabela.	8	6	14
	EC. 2	Olhar apenas os dados de uma coluna da tabela.	2	1	3
	EC. 3	Fazer comparações entre as frequências (relativas e absolutas), proporções, razões e porcentagens.	8	4	12
	EC. 4	Acertar, porém não sabe justificar.	1	2	3
	EC. 5	Acertar, porém, errar nos cálculos de porcentagem.	-	-	-
	EC. 6	Faz uma comparação entre as células que apoiam e as que não apoiam a associação.	2	4	6
	EC. 7	Acertar utilizando a associação empírica.	-	-	-
	EC. 8	Acertar olhando apenas a célula de maior frequência na tabela.	2	1	3
	EC. 9	Acertar olhando apenas a célula de menor frequência na tabela.	1	-	1
	EC. 10	Acertar fazendo cálculos de razão entre cada célula e o seu total de indivíduos da tabela.	-	-	-
Estratégias incorretas	EI. 1	Acreditar apenas na sua crença.	1	-	1
	EI. 2	Olhar apenas a célula de maior valor.	-	-	-
	EI. 3	Não calcular as proporções corretamente.	-	-	-
	EI. 4	Olhar apenas uma linha da tabela.	1	-	1
	EI. 5	Errar e não tem certeza no que diz.	1	1	2
	EI. 6	Errar por não utilizar todas as informações da tabela.	-	-	-
Total			27	19	46

Fonte: O autor (2018).

A seguir apresentaremos apenas as estratégias EC. 9 e EI. 4, o motivo da escolha é que as resposta das demais estratégias são semelhante à dos itens 1 e 2, já essas duas não compareceram nos itens anteriores a este.

EC. 9 – Acertar olhando apenas a célula de menor frequência na tabela: essa estratégia foi utilizada apenas uma vez nesse item. Um fato curioso é que o aluno optou por tomar a decisão referente a associação olhando apenas a célula de menor frequência o que contradiz alguns estudos anteriores como o de Carvalho (2017); e Cañadas, Batanero, Contreras

e Arteaga (2010) que afirmam que é comum uma pessoa qualquer fazer uma associação baseada apenas na célula de maior frequência. Consideramos essa estratégia como correta pois levou o aluno ao êxito, mesmo usando uma parte dos dados da tabela. A seguir apresentamos o exemplo que melhor descreve essa estratégia.

Figura 19 - Protocolo do item 3, aluno 19.

	Plantas com doença	Plantas sem doença
Recebeu inseticida	9	36
Não recebeu inseticida	40	10

(Questão 3) O produto é eficaz?

Sim () Não

Explique, **usando as informações da tabela**, por que você pensa assim

..... as plantas sem inseticida morreram.....

..... as plantas com inseticida não morreram.....

Fonte: O autor (2018).

Note que o aluno olhou apenas a célula das plantas que receberam inseticida. Uma outra interpretação da resposta do aluno é que ele só observou o que aconteceu com a linha das plantas que receberam inseticida. O aluno acertou este item assinalando corretamente que sim o produto é eficaz, porém essa estratégia não é recomendada, já que utiliza apenas uma parte dos dados da tabela.

EI. 4 – Olhar apenas uma linha da tabela: essa estratégia foi utilizada apenas uma vez neste item. Essa estratégia é considerada errônea por ter induzido o aluno ao erro e por utilizar apenas uma parte dos dados da tabela. A seguir apresentamos um exemplo que melhor explica essa estratégia.

Figura 20 - Protocolo do item 3, aluno 8.

	Plantas com doença	Plantas sem doença
Recebeu inseticida	9	36
Não recebeu inseticida	40	10

(Questão 3) O produto é eficaz?
 Sim Não

Explique, usando as informações da tabela, por que você pensa assim

De acordo com as informações da tabela 40 com a doença não recebeu inseticida e 10 sem a doença não recebeu, logo não foi eficaz.

Fonte: O autor (2018).

Vejam que o aluno responde apenas transcrevendo as informações da linha da tabela, para ser mais específico a linha das plantas que não receberam inseticida e concluir que o produto não é eficaz. O aluno errou por não olhar todas as informações da tabela, se aluno olhasse e analisasse os dados da outra linha da tabela, talvez sua conclusão fosse diferente.

No item 4, das situações problemas do questionário, era apresentado ao leitor uma tabela de contingência 2x3. Onde, nesta tabela mostrava o número de alunos aprovados e reprovados em um certo exame, de acordo com o tempo que foi dedicado por cada aluno para estudar, para o exame. Indagávamos aos participantes se eles concordavam que o número de aprovados e reprovados no exame tinha relação com o tempo de estudo que foi dedicado para o exame. Logo, em seguida pedíamos ao estudante que explicasse usando os dados da tabela o por que eles haviam pensado daquela maneira. O item 4 está ilustrado na figura 21:

Figura 21- Item 4 das situações problema do questionário

(Questão 4) A tabela abaixo mostra o número de alunos aprovados e reprovados em um exame, levando em consideração o tempo que cada aluno se dedicou para estudar.

	Menos de 5 h	Entre 5 e 10 h	Mais de 10 h
Aprovados	5	15	51
Reprovados	20	7	2

Você concorda que o número de aprovados e reprovados no exame têm relação com o tempo dedicado para se estudar.

() Sim () Não

Explique, **usando as informações da tabela**, por que você pensa assim

.....

.....

.....

Fonte: O autor (2018).

Apresentaremos uma solução para este item 4 semelhante ao que foi feito no item 1. Antes de iniciar a solução deste item vamos destacar o tipo das duas variáveis, note que, a variável aprovação no exame ou não é uma variável qualitativa nominal, onde, os seus elementos não obedecem a uma ordenação. Já a variável tempo dedicado para estudar para o exame é uma variável qualitativa ordinal, perceba que o tempo em si é uma quantitativa contínua, ou seja, já existe uma medição para cada instante de tempo passado, porém no item 4 o tempo ao qual nos referimos está dividido em três categorias e por esse motivo é uma variável qualitativa ordinal. A seguir apresentaremos uma tabela idêntica ao do problema, porém acrescentamos a linha e coluna marginais dessa tabela.

Tabela 10 - Tabela do item 4 das situações problemas do questionário

	Menos de 5h	Entre 5 e 10 h	Mais de 10 h	Total
Aprovados	5	15	51	71
Reprovados	20	7	2	29
Total	25	22	53	100

Fonte: O autor (2018).

De maneira análoga ao que fizemos no item 1, fixaremos ou a coluna marginal ou a linha marginal. Neste caso optamos por fixar a coluna marginal e calcular as frequências relativas de cada célula por ela. Se olharmos a tabela acima, podemos observar que 5 (cinco) estudantes que estudaram menos de cinco horas, foi aprovado no exame, ou seja, esses 5 (cinco) estudantes corresponde a $5/71 \cong 0,07$, o que é equivalente a 7% do total de aprovados desse exame. Agora se olharmos para a célula de quem foi aprovado e estudou entre cinco a dez horas para o exame, temos um quantitativo de 15 (quinze) pessoas, essas pessoas representa um quantitativo do número de aprovados, que podemos encontrar essa porcentagem da seguinte maneira $15/71 \cong 0,21$, ou seja, as 15 (quinze) pessoas corresponde 21% do número de aprovados. Fazendo esse mesmo procedimento para os 51 (cinquenta e um) indivíduos da célula dos aprovados que estudaram mais de dez horas para o exame, temos que, esse quantitativo de alunos aprovados e que estudaram mais de dez horas é correspondente a $51/71 \cong 0,72$, em outras palavras é correspondente a 72% do número de aprovados.

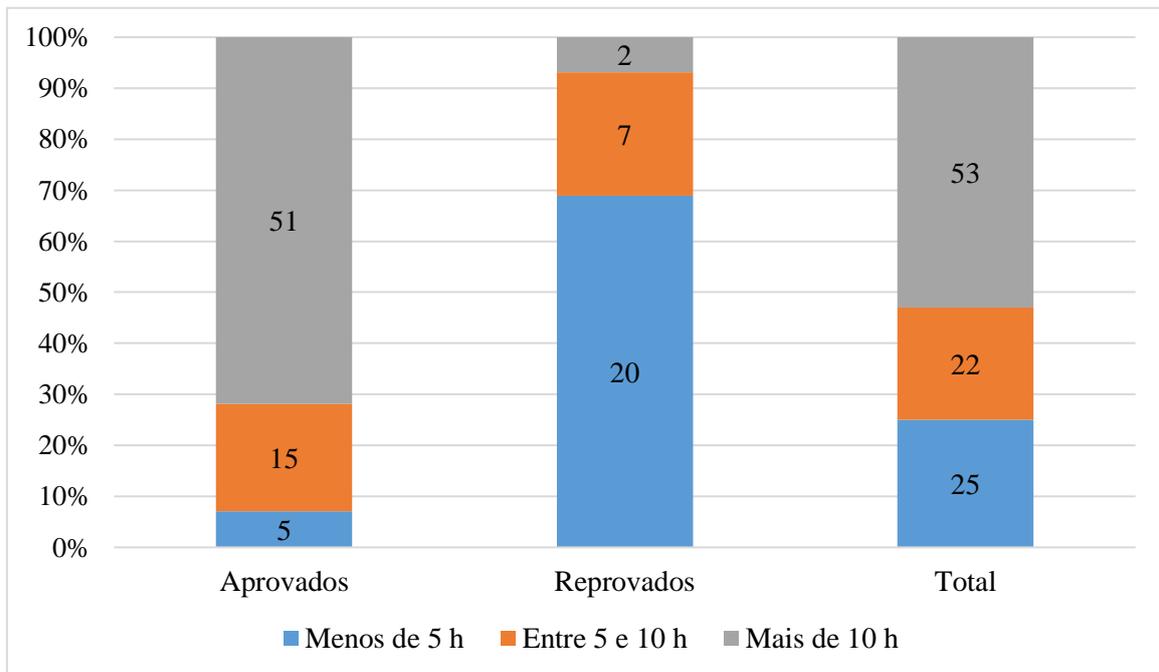
Olhando agora para linha com o número de reprovados no exame, temos que, 20 (vinte) alunos que foram reprovados no exame estudaram menos de cinco horas, ou seja, do total de reprovados que é 29 (vinte e nove) alunos, 20 (vinte) estudaram menos de cinco horas, podemos expressar essa informação como $20/29 \cong 0,69$, o que corresponde a um total de 69% do número de reprovados no exame. Já se olharmos a coluna dos estudantes que estudaram entre cinco e dez horas, o número de reprovados é de 7 (sete) alunos, o que corresponde a $7/29 \cong 0,24$, ou seja, é correspondente a 24% do total de reprovados. Já, na coluna dos alunos que se dedicaram mais de dez horas, apenas 2 (dois) alunos reprovaram o que é equivalente a $2/29 \cong 0,07$, o que corresponde 7% do total de reprovados no exame.

Ainda podemos fazer algumas considerações entre a linha marginal e a coluna marginal, ou seja, temos que 25 (vinte e cinco) alunos optaram por estudar menos de cinco horas para o exame, isso correspondente a $25/100 = 0,25$, ou seja, corresponde a 25% do total de alunos. Na coluna dos estudantes que estudaram entre cinco e dez horas para o exame, temos um total de 22 (vinte e dois) estudantes, o que corresponde a $22/100 = 0,22$, o que equivale à 22% do total de estudantes que fizeram este exame. Por fim, temos que o quantitativos de estudantes que estudaram mais de dez horas é de 53 (cinquenta e três) alunos, o que pode ser representado por $53/100$, que é equivale a 53% do total de estudante que participaram desse exame.

Do ponto de vista matemático, ao observar as frequências mencionadas anteriormente, estamos analisando probabilidades condicionais, ou seja, saber a informação que uma variável aconteceu, afeta a probabilidade da outra.

Com esses dados podemos ver, que existe sim uma relação entre ser aprovado ou reprovado com o tempo de estudo, ao qual foi dedicado para estudar para o exame. Pois, é notável que o grupo de estudantes que estudaram mais de dez horas conseguiram ter um índice de aprovação alto no exame. Já quem estudou menos de cinco horas obteve um índice de reprovação alto. Ainda é mais notável de identificar essa associação, pois o quantitativo de estudantes que estudou menos de cinco horas é superior ao quantitativo de estudantes que estudou entre cinco e dez horas e mesmo assim o índice de aprovação foi maior no último grupo. Essa relação entre as variáveis fica fácil de identificar no gráfico a seguir:

Gráfico 4 - Distribuição dos estudantes aprovados e reprovados no exame em relação com o número de horas dedicadas por cada estudante para estudar para o exame.



Fonte: O autor (2018).

Observe que este gráfico ilustra bem que quem estuda mais horas para o exame maior é a sua chance de ser aprovado, porém não podemos afirmar que é certo sua aprovação no exame pelo o tempo que foi dedicado de estudo para o exame. A seguir iremos realizar esse mesmo procedimento que realizamos anteriormente, agora fixaremos a linha marginal. Para encurtar a leitura e para o melhor entendimento do leitor, optamos em apresentar em uma tabela com os valores de cada célula representada por sua frequência absoluta e a porcentagem dessa frequência de acordo com a linha marginal.

Tabela 11 - Frequência absolutas com suas respectivas porcentagens calculadas em relação a linha marginal do item 4.

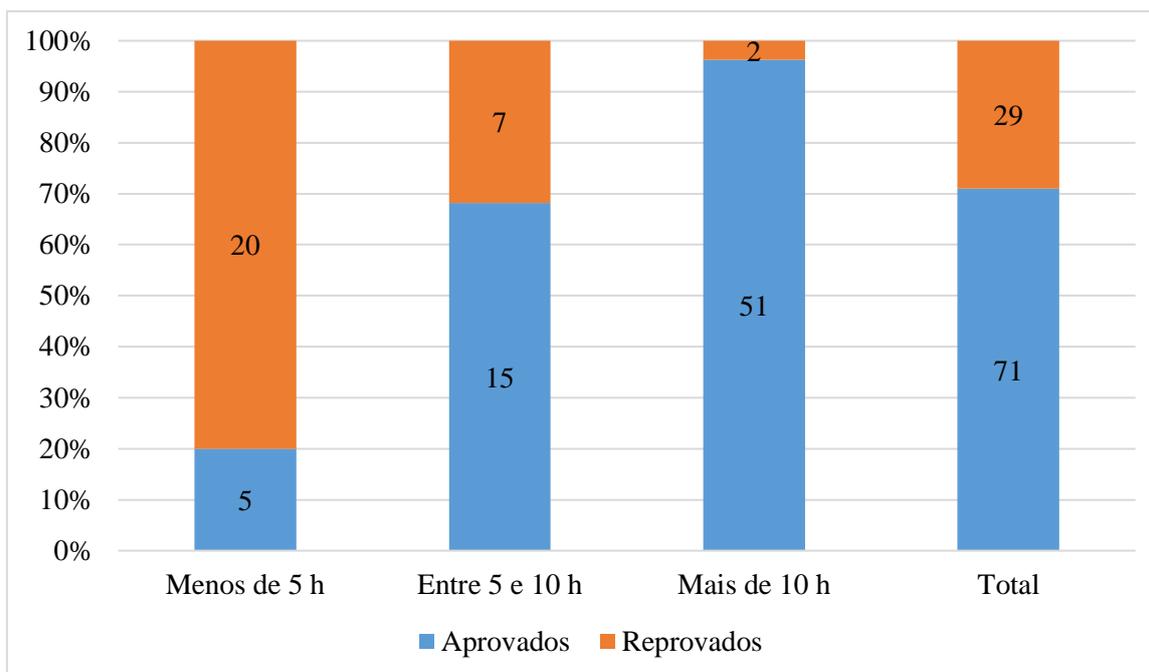
	Menos de 5 h	Entre 5 e 10 h	Mais de 10 h	Total
Aprovados	5 (20%)	15 (68,18%)	51 (96,23%)	71 (71%)
Reprovados	20 (80%)	7 (31,82%)	2 (3,77%)	29 (29%)
Total	25 (100%)	22 (100%)	53 (100%)	100 (100%)

Fonte: O autor (2018).

Com essas informações podemos notar a associação entre as duas variáveis, note que apenas 20% dos alunos que estudaram menos de cinco horas obtiveram uma aprovação, por outro lado o índice de reprovação neste grupo é alto, alcançando um índice 80%. Esses resultados se inverte na coluna dos alunos que estudaram entre cinco e dez horas, pois temos que, 68,18% desses estudantes obtiveram uma aprovação no exame e apenas 31,82% foram reprovados neste exame. Na terceira coluna, que é dos estudantes que se dedicaram mais de dez horas para o exame, temos que o índice de aprovações é bastante superior comparados aos dois últimos casos. O índice de aprovação neste grupo foi de 96,23%, já o índice de reprovação foi de apenas 3,77%. Já nos totais temos que o índice geral de aprovação no exame foi de 71% e o de reprovação foi de 29%.

Como vimos no procedimento que antecedeu a esse, vimos que sim existe uma relação entre as duas variáveis. Essa conclusão também pode ser identificada com esses dados acima, ou seja, quem mais estuda tem mais chances de ser aprovado no exame, enquanto que quem não estudar muito tem maior chance de ficar entre os reprovados. No gráfico seguinte apresentamos todos esses dados e com o auxílio do mesmo podemos entender melhor visualmente.

Gráfico 5 - Distribuição do número de horas dedicadas por cada estudante para estudar para o exame em relação com o número de estudantes aprovados e reprovados no exame.



Fonte: O autor (2018).

Esse gráfico ilustrar claramente a associação mostrando ao leitor que existe sim uma relação entre o tempo ao qual foi dedicado para estudar para o exame com o índice de aprovação e reprovação. Observando, esses dados podemos afirmar que as informações da tabela coincidem com a associação empírica dessas duas variáveis, ou seja, os dados da tabela estão de acordo com as nossas teorias prévias. O grau de associação entre essas duas variáveis é de 0,37, com este resultado podemos fazer algumas considerações uma é que a intensidade de associação entre as duas variáveis é fraca. A segunda é que pelo o sinal de associação, temos que a associação ocorre diretamente.

Apresentamos agora, o índice de acertos e erros neste item 4, ressaltamos ao leitor que apenas olhamos para esse momento se o aluno marcou certo ou errado este item, em seguida iremos fazer uma análise das estratégias desenvolvidas pelos os alunos neste item. Então, dos 46 (quarenta e seis) participantes, 44 (quarenta e quatro) assinalaram corretamente que sim, existe uma relação entre o tempo que foi dedicado para se estudar para o exame, com aprovação e reprovação no exame. Ou seja, neste item obtivemos um índice de acerto de 95,65%. Apenas 2 (dois) alunos erraram este item assinalando que não existe nenhuma relação entre essas duas variáveis, o que totalizar um índice de erro para este item de 4,35%.

Numa análise por grupo temos que no grupo dos estudantes que não cursaram a disciplina de Estatística, temos como resultado que apenas 1 (um) aluno errou este item e 26

(vinte e seis) acertaram este item. Em termos de porcentagem temos que, o índice de acertos foi de 96,3%, já o índice de erro neste grupo foi de aproximadamente 3,7%.

No grupo dos estudantes que cursaram a disciplina obtivemos que, dos 19 (dezenove) alunos, 18 (dezoito) acertaram este item assinalando que sim existe uma relação entre essas duas variáveis e apenas 1 (um) aluno marcou erroneamente que não existia nenhuma relação entre as duas variáveis. O índice de acerto deste grupo no item foi de 94,74% e o de erro foi de 5,26%.

Neste item 4, não podemos dizer que um grupo se sobressaiu melhor que o outro, pois os resultados são muitos parecidos. Quanto ao índice de acerto foi alto no geral e nos dois grupos, este índice de acerto elevado pode ter acontecido pois é do senso comum acreditarmos que quanto mais se estudar para um exame, melhor será o desempenho dessa pessoa no exame e os dados da tabelas coincidem com esse resultado, só poderemos ter essa convicção quando apresentarmos as análises da estratégias.

Na tabela 12 apresentamos o quantitativo de estudantes por grupo, os que cursaram e os que não cursaram a disciplina de estatística e também o total dos dois grupos juntos e as estratégias que eles utilizaram ao responder esse item. Para este item totalizamos oito estratégias distintas, sendo seis corretas e duas incorretas. Esse estudo é importante pois, descreverá um pouco de como os alunos pensaram ao resolver este item 4.

Tabela 12 - Estratégias utilizadas pelos os estudantes para responder o item 4.

Estratégias			Grupos dos estudantes		Total
			Que não cursaram a disciplina de Estatística	Que cursaram a disciplina de Estatística	
Estratégias corretas	EC. 1	Olhar apenas os dados de uma linha da tabela.	-	-	-
	EC. 2	Olhar apenas os dados de uma coluna da tabela.	4	4	8
	EC. 3	Fazer comparações entre as frequências (relativas e absolutas), proporções, razões e porcentagens.	11	2	13
	EC. 4	Acertar, porém não sabe justificar.	-	-	-
	EC. 5	Acertar, porém, errar nos cálculos de porcentagem.	-	-	-
	EC. 6	Faz uma comparação entre as células que apoiam e as que não apoiam a associação.	7	9	16
	EC. 7	Acertar utilizando a associação empírica.	2	2	4
	EC. 8	Acertar olhando apenas a célula de maior frequência na tabela.	2	-	2
	EC. 9	Acertar olhando apenas a célula de menor frequência na tabela.	-	-	-
	EC. 10	Acertar fazendo cálculos de razão entre cada célula e o seu total de indivíduos da tabela.	-	1	1
Estratégias incorretas	EI. 1	Acreditar apenas na sua crença.	1	-	1
	EI. 2	Olhar apenas a célula de maior valor.	-	-	-
	EI. 3	Não calcular as proporções corretamente.	-	-	-
	EI. 4	Olhar apenas uma linha da tabela.	-	-	-
	EI. 5	Errar e não tem certeza no que diz.	-	-	-
	EI. 6	Errar por não utilizar todas as informações da tabela.	-	1	1
Total			27	19	46

Fonte o autor (2018).

Apresentaremos a seguir apenas um exemplo da estratégia EC. 7, pois as demais já foram apresentadas nos itens anteriores.

EC. 7 - Acertar utilizando a associação empírica: essa estratégia foi utilizada no total de quatro vezes neste item. Apesar de alguns autores como Cañadas, Batanero, Contreras e Arteaga (2010); Carvalho (2017); e Nunes e Bryant (2011) afirmarem essa estratégia como errônea, consideramos como correta por terem levados os alunos neste item ao êxito. Essa estratégia é baseada apenas no senso comum do aluno, ou seja, o que ele pensar e traz consigo

do seu cotidiano para responder o item. Salientamos que obtivemos quatro acertos usando essa estratégia e essa ocorrência pode está diretamente ligada a concordância dos dados da tabela e as variáveis com a associação empírica. Em seguida apresentamos dois exemplos que evidência melhor essa estratégia.

Figura 22 - Protocolo do item 4, aluno 6.

(Questão 4) A tabela abaixo mostra o número de alunos aprovados e reprovados em um exame, levando em consideração o tempo que cada aluno se dedicou para estudar.

	Menos de 5 h	Entre 5 e 10 h	Mais de 10 h
Aprovados	5	15	51
Reprovados	20	7	2

Você concorda que o número de aprovados e reprovados no exame têm relação com o tempo dedicado para se estudar.

Sim () Não

Explique, usando as informações da tabela, por que você pensa assim

pois quanto mais tempo você dedica a algo mais satisfatório o resultado será, claro que se esse tempo for usado de maneira proveitosa.

Fonte: O autor (2018).

Observe que o aluno associa o tempo dedicado para o exame ao número de aprovação, essa falar ficar evidente quando o aluno diz quanto mais tempo você dedica a algo mais satisfatório o resultado, isso se o tempo for “usado de maneira proveitosa”. Ou seja, o aluno utilizou a sua associação empírica para resolver o problema.

Figura 23 - Protocolo do item 4, aluno 32.

(Questão 4) A tabela abaixo mostra o número de alunos aprovados e reprovados em um exame, levando em consideração o tempo que cada aluno se dedicou para estudar.

	Menos de 5 h	Entre 5 e 10 h	Mais de 10 h
Aprovados	5	15	51
Reprovados	20	7	2

Você concorda que o número de aprovados e reprovados no exame têm relação com o tempo dedicado para se estudar.

Sim

Não

Explique, usando as informações da tabela, por que você pensa assim

Para se fazer exame e obter êxito é necessário ter conhecimento acumulado; ter guardado muita ideia e informações na memória; para ser aprovado é ter que elaborar mesmo nos livros, jornais, revistas, tes referências; de fato quem estudar com dedicação obterá êxito; tem mais conhecimento e poder, quem conhece, quem busca, quem luta tem futuro promissor. Os números dá um indicativo e é real 51 estudantes que se dedicou e estudou mais de 10h foi aprovado, enquanto num conjunto de 25 estudantes que estudaram menos de 5h foram reprovados 80% destes. Quanto a palavra, ALUNO a exclua do seu vocabulário. Pois, A = não + LUNO = luz = Sem luz; substitua por ESTUDANTE, EDUCANDO, em aprendizagem, lutar, em busca de mais conhecimento.

Fonte: O autor (2018).

O aluno 32 deixa evidente sua opinião em toda sua resposta. O aluno explicar a importância de se estudar, de se manter atualizado e de ter foco e determinação na hora de estudar, porém não era o foco da questão e nem muito menos o significado da palavra aluno como o aluno explicou. Ou seja, é evidente que o aluno só levou em conta só a sua associação empírica.

De uma forma geral, considerando os quatro itens, os futuros professores de matemática, recorreram à estratégia errônea de tomar sua decisão baseado em uma única célula; normalmente a célula de maior valor. Nossos resultados estão em consonância com os estudos apontados em nossa revisão de literatura ao qual o uso desta estratégia já estava previsto em diversos estudos.

É importante destacar que optamos por apresentar diferentes ideias de soluções para cada item, pensando em subsidiar o trabalho com esses conceitos. No primeiro e quarto item optamos por apresentar a solução através das frequências marginais; o segundo item apresentamos uma solução através do cálculo de probabilidade condicional e utilizamos o conceito de razão utilizando as frações; no terceiro item apresentamos uma solução mais sofisticada estatisticamente que foi a ideia de razão de chance. O intuito de diferenciar na apresentação de cada solução é de mostrar ao leitor que não existe apenas uma forma correta de se fazer uma associação entre duas variáveis e sim uma gama de formas que podem ser usadas numa associação de variáveis.

7 CONSIDERAÇÕES FINAIS

Esse trabalho buscou analisar os conhecimentos de futuros professores de matemática sobre associação de variáveis em tabelas de contingência. Para isso, realizamos essa pesquisa em duas fases.

A primeira fase constituiu em fazer um levantamento em estudos antecedentes a essa pesquisa. Buscamos entender de que forma os futuros professores de matemática estão saindo do curso de licenciatura em matemática, referente a disciplina de estatística e se a mesma vem contribuindo para a formação desse professor. Utilizamos diversos estudos da literatura atual para discutir que pouco tem sido feito na formação do professor de matemática, pois a disciplina é vista através de fórmulas e cálculos mecanizados e acaba restando pouca contribuição dessa disciplina na formação do professor, pois os cursos de matemática ofertam uma ou duas disciplinas que abordam essa temática. Um outro fator que julgamos importante para essa defasagem na formação do professor de matemática, é no que se refere aos assuntos de estatística, pois são abordados separadamente dos assuntos de probabilidade e acreditamos que esses conteúdos devem ser expostos de forma interligada.

Apresentamos estudos que reportam o ensino de probabilidade e os seus diversos significados, também abordamos estudos sobre probabilidade condicional. Posteriormente, apresentamos vários estudos sobre associação de variáveis em tabelas de contingência, apresentamos também três formas de fazer essas associações, uma através do cálculo de probabilidade, outra através da comparação dos casos que apoiam a associação e os que não apoiam e por fim expomos ideia de razão de chance.

Em uma segunda fase, elaboramos e aplicamos um questionário para diagnóstico dos conhecimentos dos futuros professores de matemática. Este instrumento foi constituído por quatro situações problemas que contemplava a associação entre variáveis. Dessas questões duas delas apresentavam dados condizentes com a associação empírica; uma não existia associação empírica; e outra os dados da tabela iam contra a associação empírica.

Optamos em apresentar a análise e discussão dos resultados em dois momentos. No primeiro momento apresentamos resultados dos quatro itens juntos, no qual constatamos que o resultado dessa pesquisa foi bastante positivo, com uma média de acertos bastante alta. Também verificamos que o índice de acertos foi superior a 80% nos dois grupos de estudantes. Já no segundo momento que foi a análise por item, podemos ver que no primeiro item, ao qual nos referimos que os dados não coincidiam com a associação empírica, tivemos que os alunos que cursaram a disciplina de estatística obtiveram um desempenho abaixo de 50%. Já o grupo

de estudantes que não cursaram a disciplina de estatística obtiveram um índice de acertos superior há 65% neste mesmo item. Podemos afirmar que uns dos principais motivos para essa diferença de desempenho nesses dois grupos podem ter sido motivados pelo fato dos dados da tabela não coincidir com a associação empírica. Mas, julgamos e acreditamos que como futuros professores de matemática não podemos nos deixar levar pela crença ou até mesmo não podemos tirar conclusões olhando apenas para a célula de maior valor. Devemos levar em consideração todos os dados apresentados na tabela. Nos demais itens o índice de acertos foi bastante alto e não teve diferença de desempenho nos dois grupos. Com isso, podemos afirmar também que para esta pesquisa pouco influenciou ter cursado ou não a disciplina de estatística.

REFERÊNCIAS

- AMÂNCIO, J. R. **Planejamento e aplicação de uma sequência didática para o ensino de probabilidade no âmbito do PIBID**. 2012. 225 f. Dissertação de Mestrado em Ensino de Matemática. Programa de Pós-graduação em Ensino de Matemática. Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2012.
- BRASIL. Ministério da Educação. Secretaria Executiva. Secretaria da Educação Básica. **Base Nacional Comum Curricular**. Brasília: Secretaria da Educação Básica, 2018. Disponível em: < <http://basenacionalcomum.mec.gov.br/download-da-bncc> > Acesso em 25 de junho de 2018.
- CAÑADAS, G., BATANERO, C., CONTRERAS, J. M., ARTEAGA, P. Estrategias en el estudio de la asociación en tablas de contingencia por estudiantes de psicología. **Educación Matemática**, México, vol. 23, n.2, p. 5-31, 2011. Disponível em: < <http://www.scielo.org.mx/pdf/ed/v23n2/v23n2a2.pdf> > Acesso em 19 de março de 2018.
- CARVALHO, J. I. F. Conhecimentos de futuros professores de matemática sobre probabilidade condicional por meio do jogo das três fichas. **Actas: Segundas jornadas virtuales en didáctica de la estadística, probabilidad y combinatoria**, Granada, p. 189-196, 2015. Disponível em: < <http://www.estadis.net/3/actas/COM/11.%20Conhecimentos%20de%20futuros%20%20professores%20Probabilidade%20Condicional%20REVISADO.pdf> >. Acesso em 18 de abril de 2016.
- _____. **Um estudo sobre os conhecimentos didáticos-matemáticos de probabilidade com professores de matemática dos anos finais do Ensino Fundamental**. 2017. Tese (Doutorado em Educação Matemática) – Coordenadoria de Pós-graduação – Universidade Anhanguera, São Paulo, 2017.
- COSTA, A.; NACARATO, A. M. A estocástica na formação do professor de matemática: percepções de professores e de formadores. **Bolema- Boletim de Educação Matemática**, Rio Claro (SP), vol. 24, n.39, p. 367-386, 2011. Disponível em: < <http://www.periodicos.rc.biblioteca.unesp.br/index.php/bolema/article/view/5092/4011> >. Acesso em 15 de abril de 2016.
- COUTINHO, C. Q. S. Conceitos probabilísticos: quais contextos a história nos aponta? **REVEMAT – Revista Eletrônica de Educação Matemática**, Florianópolis, v. 2, n. 3, p. 50-67, 2007. Disponível em: < <https://periodicos.ufsc.br/index.php/revemat/article/view/12991> >. Acesso em 31 ago. 2017.
- DANCEY, Christine P. Estatística sem matemática para psicologia usando SPSS para Windows. In: DANCEY, Christine P., REIDY, John; tradução Lorí Viali. (Orgs.). **Análise de Correlação: o r de Pearson**. Porto Alegre: Artmed, ed. 3, 2006.
- DANTE, Luiz Roberto. **Matemática: livro do aluno**, volume 3, Ática, 1ª edição, São Paulo, 2004.

_____. **Matemática: contexto & aplicações:** Ensino Médio: volume 2, Ática, 3ª edição, São Paulo, 2016.

FALK, R. Conditional probabilities: insights and difficulties. En R. Davidson y J. Swift (Eds), **Proceedings of the Second International Conference on Teaching Statistics.** Victoria, Canada: International Statistical Institute, p. 292 - 297, 1986.

FERNANDES, J. A.; MUGABE, D. A.; CORREIA, P. F. Associação Estatística em Tabelas de Contingência de 2x2. **Acta Scientiae** – Revista de Ensino de Ciências e Matemática. Canoas (RS), vol. 14, n. 3, p. 374-390, 2012. Disponível em: < <http://www.periodicos.ulbra.br/index.php/acta/article/view/397/305> >. Acesso em 04 de maio de 2018.

GIOVANNI, J. R., GIOVANNI Jr., J. R., BONJORNO, J. R., SOUSA, P. R. C. **Matamática uma nova abordagem:** progressões: 2º ano, Ensino Médio, FTD, 3ª edição, São Paulo, 2013.

LIMA, Elon Lages. A Matemática do Ensino Médio - volume 2. In: LIMA, Elon Lages, CARVALHO, Paulo Cezar Pinto, WAGNER, Eduardo, MORGADO, Augusto César. (Orgs.). **Probabilidade.** Rio de Janeiro: SBM, ed. 6, 2006.

LOPES, C. E. Educação Estatística no Curso de Licenciatura em Matemática. **Bolema-** Boletim de Educação Matemática, Rio Claro (SP), vol. 27, n.47, p. 901-915, 2013. Disponível em: < <http://www.periodicos.rc.biblioteca.unesp.br/index.php/bolema/article/view/8285/5866> >. Acesso em 15 de abril de 2016.

NUNES, T., BRYANT, P. Understanding risk and uncertainty: The importance of correlations. **EM TEIA-** Revista de Educação Matemática e Tecnológica Iberoamericana, Recife (PE) vol. 2, n.2, 2011. Disponível em: < <https://periodicos.ufpe.br/revistas/emteia/article/view/2150/1719> >. Acesso em 28 agosto de 2017.

NUNES, T.; BRYANT, P.; EVANS, D.; GOTTARDIS, L.; TERLEKTSI, M. **Teaching primary school children about probability. Teacher handbook.** Departamento de Educação, Universidade de Oxford. [CD-ROM]. 2012.

OLIVEIRA, Maria Marly de. Como fazer pesquisa qualitativa. In: OLIVEIRA, Maria Marly de (Orgs.). **Metodologia, métodos e técnicas.** Petrópolis - RJ: Vozes, 2016.

SHULMAN, L. S. Knowledge and Teaching: foundations of the New Reform. **Harvard Educational Review**, vol.57, n.1. 1987.

SHULMAN, L.S. Those who understand: knowledge growth in teaching. **Education Researcher**, v.15, n.2, p.4-14, 1986.

SILVA, C. D. B. **Significados de probabilidade: uma análise em livros didáticos dos anos finais do ensino fundamental.** Monografia (Trabalho de Conclusão de Curso) – Universidade Federal de Pernambuco, CAA, Licenciatura em Matemática, 2015.