



Universidade Estadual da Paraíba  
Centro de Ciências e Tecnologia  
Departamento de Estatística

Sidcleide Barbosa de Sousa

# **Análise de associação em dados relativos ao acompanhamento do crescimento infantil no município de Queimadas-PB**

Campina Grande  
Dezembro de 2012

Sidcleide Barbosa de Sousa

# **Análise de associação em dados relativos ao acompanhamento do crescimento infantil no município de Queimadas-PB**

Trabalho de Conclusão de Curso apresentado ao curso de Bacharelado em Estatística do Departamento de Estatística do Centro de Ciências e Tecnologia da Universidade Estadual da Paraíba em cumprimento às exigências legais para obtenção do título de bacharel em Estatística.

Orientador:

Gustavo Henrique Esteves

Campina Grande

Dezembro de 2012

FICHA CATALOGRÁFICA ELABORADA PELA BIBLIOTECA CENTRAL – UEPB

S725a Sousa, Sidcleide Barbosa de.  
Análise de associação em dados relativos ao acompanhamento do crescimento infantil no município de Queimadas-PB [manuscrito] / Sidcleide Barbosa de Sousa. – 2012 .  
32f.: il. color.

Trabalho de Conclusão de Curso (Graduação em Estatística)  
– Universidade Estadual da Paraíba, Centro de Ciências e Tecnologia, 2012.

“Orientação: Prof. Dr. Gustavo Henrique Esteves, Departamento de Estatística”.

1. Estatística - Área Médica. 2. Medidas de Associação. 3. Probabilidade. I. Título.

21. ed. CDD 519.52

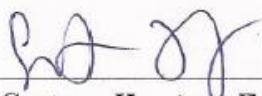
Sidcleide Barbosa de Sousa

# Análise de associação em dados relativos ao acompanhamento do crescimento infantil no município de Queimadas-PB

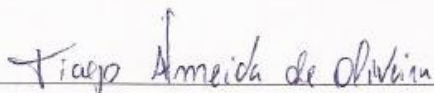
Trabalho de Conclusão de Curso apresentado ao curso de Bacharelado em Estatística do Departamento de Estatística do centro de Ciências e Tecnologia da Universidade Estadual da Paraíba em cumprimento às exigências legais para obtenção do título de bacharel em Estatística.

Aprovado em: 14 /12 /12

## Banca Examinadora:



Prof. Dr. Gustavo Henrique Esteves  
Orientador



Prof. Dr. Tiago Almeida de Oliveira  
Universidade Estadual da Paraíba



Profa. Msc. Ana Patrícia Bastos Peixoto  
Universidade Estadual da Paraíba

# Dedicatória

*Dedico este trabalho a DEUS o autor da minha VIDA*

*Aos meus pais Rosalia e Severino*

*Aos meus irmãos em especial a Sivoneide e Sivonaldo que colaboraram com  
compreensão, dedicação, amor e carinho nesta jornada*

*A meu sobrinho Deivisson Barbosa*

*Obrigado por vocês fazerem parte da minha vida*

*Por que família é tudo.*

# Agradecimentos

Em primeiro lugar agradeço a DEUS, por ter me dado força e pela vitória a mim concedida.

Ao professor Gustavo Henrique Esteves, pelo apoio, estímulo, confiança e dedicação durante todo o desenvolvimento da orientação.

A toda a minha família: Aos meus pais Rosalia e Severino, aos meus irmãos em especial a Sivoneide e Sivonaldo e a meu sobrinho Deivisson Barbosa que é o querido da Titia, pelo apoio, pelo amor incondicional em todos os momentos, fazendo-se sempre presente ao longo do curso, a todos vocês o meu muito obrigado, eu amo todos vocês.

A Universidade Estadual da Paraíba (UEPB), ao Centro de Ciências e Tecnologia (CCT), e em especial ao Departamento de Estatística.

A todos os professores da UEPB que contribuíram na minha vida acadêmica ao longo deste curso.

Aos amigos e colegas da graduação em especial a **Allana Lívia** pelos momentos de estudos em sua residência, pelas alegrias, conversas e estresse por ter me suportado, a toda sua família que são pessoas maravilhosas, a seus irmãos Alisson e Lúcio pelas caronas em altas horas, amei conhecer todos vocês. A Priscilla que veio estudar com a gente de enxada, mas é uma pessoa maravilhosa espero que estejamos sempre juntas e aos colegas de curso Rosendo, Diego e Fabio pelos poucos momentos que passamos juntos mais muito divertido e todos aqueles que de alguma forma tivemos algum tipo de contato, a todos vocês o meu muito obrigado.

# Resumo

As técnicas estatísticas são utilizadas em diversas áreas, principalmente na área da saúde, em que é normalmente conhecida como bioestatística, utilizando aplicação de técnicas estatísticas com o objetivo de solucionar problemas na área de ciência biológica. Neste trabalho verifica-se a existência de associação entre as variáveis envolvidas com as crianças, as mães e as famílias de acordo com a localização do domicílio e de acordo com o risco nutricional da criança, por meio do teste de qui-quadrado para amostra considerada suficientemente grande e o teste exato de Fisher para amostra considerada pequena para variáveis envolvidas com a estrutura do serviço e o processo de trabalho nas UBSF para o acompanhamento do crescimento das crianças de acordo com a localização do domicílio e de acordo com risco nutricional da criança, e em seguida utiliza-se a medida de associação da bioestatística por meio da razão de chances. A partir dos resultados obtidos pode-se observar que existe associação em algumas variáveis envolvidas como o risco nutricional versus peso ao nascer como se esperava, pois o nível descritivo do teste é 0,0003455 bem menor que o nível descritivo pré-fixado de 0,05 pelo teste qui-quadrado e já no teste exato de Fisher as variáveis envolvidas deram não significativas, ou seja, não existe associação entre as variáveis devido a amostra ser muito pequena, pois seu nível descritivo deram todos maiores que o nível descritivo pré-fixado que é de 0,05.

**Palavras-chaves:** Teste de Qui-quadrado, Teste exato de Fisher, Medidas de associação.

# Abstract

Statistical techniques are used in several scientific areas, mainly in Health, normally known as biostatistics, using and applying statistical models with the main focus to understand biological sciences area. In this work it was verified the existence of association between some variables with children, mothers and families according to the location of domicile and in accordance with the child's nutritional risk, using Chi Squared test for larger samples and Fisher's exact test for the smaller samples to variables involved with the structure of the service and the work process in UBSF for monitoring growth of children according to the location of domicile and nutritional risk according to the child's, following we used Odds Ratio as the association measure. From the observed results, it was observed evidence of association for some variables associated with nutritional risk versus birthweight as expected, with a p-value of 0.0003455, that is smaller than the 0.05 pre-defined, for chi squared tests. For Fisher's exact test any pair of variables presented evidence of association, mainly because the small sample size, showing p-values always greater than pre-defined 0.05.

**Key-words:** Chi-Squared test, Fisher's exact test, association measures.



# Sumário

## Lista de Tabelas

<b>1</b>	<b>Introdução</b>	p. 11
<b>2</b>	<b>Fundamentação Teórica</b>	p. 13
2.1	Bioestatística . . . . .	p. 13
2.2	Tabelas de contingência . . . . .	p. 13
2.2.1	Tabela de contingência em ensaios clínicos . . . . .	p. 14
2.2.2	Tabela de contingência em estudos prospectivos . . . . .	p. 15
2.2.3	Tabela de contingência em estudos retrospectivos . . . . .	p. 15
2.2.4	Tabela de contingência em estudos transversais . . . . .	p. 16
2.3	Teste de Associação . . . . .	p. 17
2.3.1	Teste de qui-quadrado ( $\chi^2$ ) . . . . .	p. 17
2.3.1.1	Condições necessárias para aplicação do teste de qui-quadrado ( $\chi^2$ ) . . . . .	p. 18
2.3.1.2	Hipóteses a serem testadas . . . . .	p. 18
2.3.1.3	Como calcular . . . . .	p. 18
2.3.1.4	Conclusão do teste de $\chi^2$ . . . . .	p. 19
2.3.2	Correção de continuidade de Yates para o teste de $\chi^2$ . . . . .	p. 20
2.3.3	Teste exato de Fisher . . . . .	p. 21
2.3.3.1	As características do Teste Exato de Fisher . . . . .	p. 21
2.4	Medidas de associação nas Tabelas $2 \times 2$ . . . . .	p. 22
2.4.1	Coefficiente de contingência de Pearson . . . . .	p. 23

2.4.2	Razão de chances ou <i>Odds-Ratio</i> . . . . .	p. 23
<b>3</b>	<b>Aplicação</b>	p. 25
3.1	Descrição dos Dados . . . . .	p. 25
3.2	Análise dos Dados . . . . .	p. 25
<b>4</b>	<b>Conclusão</b>	p. 31
	<b>Referências</b>	p. 32

# Lista de Tabelas

1	Tabela $2 \times 2$ de frequências observadas. . . . .	p. 14
2	Tabela $r \times c$ . . . . .	p. 17
3	Tabela $2 \times 2$ de frequências esperadas. . . . .	p. 19
4	Teste de associação para as variáveis envolvidas com as crianças em estudo.	p. 27
5	Teste de associação para variáveis envolvidas com as mães em estudo. .	p. 27
6	Teste de associação para as variáveis envolvida com as famílias em estudo.	p. 28
7	Tabela de contingência de risco nutricional <i>versus</i> peso ao nascer . . . .	p. 28
8	Teste exato de Fisher para associação entre as variáveis, envolvidas com a UBSF. . . . .	p. 29

# 1 Introdução

Da mesma forma que a prática da medicina não é completamente objetiva, a da Estatística também não é. Não existem “receitas prontas” para tratar doentes, assim como não existem fluxogramas que indiquem as técnicas estatísticas que devem ser utilizadas em cada caso. O que existem são “práticas comuns” que podem ser aplicadas ou não, dependendo das condições do estudo. Por exemplo, um médico pode optar por um medicamento alternativo no lugar de um mais comum, devido às condições do paciente (pode ser um paciente idoso cujo medicamento não é recomendado). Portanto, o que pretendemos ressaltar é que cada caso deve envolver uma análise particular, assim como em um exame clínico, de forma que a escolha da técnica seja feita com critério e cuidado (PAES, 1998).

A Bioestatística também conhecida como estatística vital surgiu em 1894 quando Karl Pearson aplicou a probabilidade à biologia e criou a área de estudo denominada Biométricas, a estatística vital baseia-se antes de tudo no estudo da população do ponto de vista de sua composição e crescimento, a Bioestatística em si é a aplicação de técnicas estatísticas com o propósito de solucionar problemas nas áreas da saúde.

Os testes de associação utilizados neste trabalho para verificação de associação entre as variáveis envolvidas foram os teste de qui-quadrado para amostras maiores que 40, e o teste exato de Fisher que é o mais apropriado para amostra menores que 20. As medidas de associação são adequadas em determinados tipos de estudos, como a razão de chances que é mais adequada em estudo do tipo caso-controle, em que sempre divide os indivíduos em dois grupos, sendo em um grupo os indivíduos que possuem desfecho de interesse, ou seja, grupo de casos, e no outro grupo os indivíduos que não possuem o desfecho de interesse, ou seja, grupo controle. O principal objetivo desse estudo é comparar a probabilidade de um indivíduo estar exposto ao fator de risco dado que exista o desfecho, e a outra probabilidade é se o indivíduo não estar exposto ao fator de risco dado que não exista o desfecho.

Este trabalho tem por objetivo fundamental o de abordar e discutir a aplicação de métodos estatísticos como ferramenta nas análises preliminares dos dados coletados nas unidades de saúde no município de Queimadas no estado da Paraíba em 2011, em que foi feita uma colaboração para analisar os dados obtidos no projeto usando o *software* R. Comparando os resultados encontrados de acordo com os testes de associação, ou seja, o teste de qui-quadrado verifica-se que existe associação entre as variáveis, já no teste exato de Fisher foi verificado que não há associação entre as variáveis, utilizando as medidas de associação em razão de chances o que não foi possível obter o valor da estatística.

## 2 Fundamentação Teórica

### 2.1 Bioestatística

Estatística vital, também conhecida como bioestatística, foi definida por Trask como estudo estatístico da história vital das comunidade ou nações. A estatística vital baseia-se antes de tudo no estudo da população do ponto de vista de sua composição e crescimento. Seu estudo abrange, por conseguinte, os fatos vitais, como: casamentos, nascimentos, óbitos, doenças, etc. (SOUNIS, 1975).

Métodos estatísticos são essenciais no estudo de situações em que as variáveis de interesse estão sujeitas, inerentemente a flutuações aleatórias. Convencionou-se chamar de bioestatística o conjunto de métodos estatísticos usados no tratamento de dados nas ciências médicas e biológicas. A bioestatística fornece métodos para se tomar decisões adequadas na presença de incerteza, estabelecendo faixas de confiança para a eficácia dos tratamentos e verificando a influência de fatores de risco no aparecimento de doenças (SOARES; SIQUEIRA, 2002).

A estatística não se resume apenas na coleção de números, mas trata-se desde da coleta de dados, a apuração desses dados, de sua classificação e principalmente da análise dos dados. São várias as técnicas utilizadas na bioestatística por meio das aplicações dos métodos estatísticos para solucionar proplemas existentes na área da saúde.

### 2.2 Tabelas de contingência

Segundo Pagano e Gauvreau (2004), ao trabalhar-se com dados nominais agrupados em categorias, frequentemente dispõe-se das contagens em um formato tabular normalmente conhecido como tabela de contingência. Em casos mais simples, duas variáveis aleatórias serão envolvidas, as linhas da tabela representam os valores de uma variável, e as colunas, os valores da outra. As entradas da tabela são as contagens que correspondem

a uma combinação particular de categorias.

Para se testar a existência de dependência entre duas variáveis categóricas utilizaremos métodos estatísticos para se obter melhores resultados, o caso mais simples para se testar independência entre variáveis aleatórias é quando temos duas variáveis binárias (ou dicotômicas), que assumem apenas dois valores possíveis. (ESTEVES, 2010).

Uma tabela  $2 \times 2$  é genericamente representada como na Tabela 1:

Variável Y	Variável X		Total
	$x_1$	$x_2$	
$y_1$	$n_{11}$	$n_{12}$	$n_{1.}$
$y_2$	$n_{21}$	$n_{22}$	$n_{2.}$
Total	$n_{.1}$	$n_{.2}$	$n$

Tabela 1: Tabela  $2 \times 2$  de frequências observadas.

Em que,  $n$  é o tamanho amostral total,  $n_i$  é a frequência absoluta de  $y_i$ ,  $i = 1, 2$ ,  $n_j$  é a frequência absoluta de  $x_j$ ,  $j = 1, 2$ ,  $n_{ij}$  é a frequência absoluta conjunta para cada  $Y = y_i$  e  $X = x_j$ ,  $i = 1, 2$  e  $j = 1, 2$ .

São vários os planejamentos amostrais para a construção da tabelas de contingência. entre eles tem-se:

- i) Modelo com os totais marginais das duas variáveis fixos;
- ii) Modelo com os totais marginais de apenas uma variável fixo;
- iii) Modelo com apenas o tamanho amostral fixo;
- iv) Modelo livre, isto é, sem nenhuma restrições no tamanho amostral nem nos totais marginais.

Segundo Vieira (2003), vários tipos de pesquisas científicas podem dar origem as frequências exibidas nas tabelas de contingência como: ensaios clínicos, estudos prospectivos, estudos retrospectivos, e estudos transversais. Como veremos nas seções a seguir.

### 2.2.1 Tabela de contingência em ensaios clínicos

Segundo Vieira (2003), em um ensaio clínico controlado com placebo, o pesquisador deve dividir uma amostra de  $n$  pacientes, em dois grupos de forma aleatória, sendo assim,

um grupo experimental recebe o tratamento em teste e o outro grupo controle recebe um placebo. A contribuição da aleatorização é em relação a homogeneidade dos grupos, em que o grupo experimental recebe o tratamento e ao final compara-se com a variável  $X$  observado no grupo de controle. Na variável  $X$  pode-se surgir algum tipo de doenças ou qualquer outro evento, que possa acontecer o processo de saúde-doença. Se o tratamento de cada paciente tem apenas duas possibilidades, logo os dados do ensaio deverão ser apresentados em uma tabela  $2 \times 2$ , como foi apresentado Tabela 1.

Com isso é possível comparar as probabilidades, sendo que a probabilidade de um paciente estar exposto aos tratamentos ser igual, quer seja no grupo experimental ou no grupo controle, denominada de hipótese nula ( $H_0$ ), ou se as probabilidades são diferentes, denominada hipótese alternativa ( $H_1$ ).

### 2.2.2 Tabela de contingência em estudos prospectivos

Em um estudo prospectivo, os dados são coletados ao decorrer do tempo da pesquisa, em que são acompanhados dois grupos diferentes de indivíduos, um grupo possua o fator de risco e o outro grupo sem o fator de risco. O fator de risco é uma variável  $Y$  que se supõe estar associada a variável  $X$ , estes dois grupos serão acompanhados durante um longo período de tempo, esperando-se que surjam novos casos da variável  $X$  nos dois grupos e assim pode-se comparar os grupos e calcular as taxas de incidência para o grupo exposto ao fator de risco e para o grupo não exposto. Os dados serão apresentado em uma tabela de contingência, se houver apenas a possibilidade de ter ou não algum tipo de doença, conforme visto na Tabela 1.

Esse tipo de estudo é caro e dura um longo período de tempo, sendo necessário acompanhar muitos indivíduos e quanto mais raro é a variável  $X$ , maior será o número de indivíduos que serão examinados. O objetivo desse tipo de pesquisa é comparar se a probabilidade do indivíduo exposto ao fator de risco na variável  $X$  é igual a probabilidade dos indivíduos que não estão expostos ao fator de risco, ou seja, a hipótese nula ( $H_0$ ) ou é diferente, ou seja a hipótese alternativa ( $H_1$ ).

### 2.2.3 Tabela de contingência em estudos retrospectivos

Os estudos retrospectivos são realizados após o efeito já ter ocorrido, ou seja, esse estudo parte do efeito para chegar as causas. Portanto é preciso que um grupo de indivíduos com determinada doença e outro grupo comparável de indivíduos sem essa determinada



doença, mas que seja comparável em todos os outros aspectos, tais como: sexo, idade e nível socioeconômico. Com isso o pesquisador calcula a proporção de indivíduos expostos ao fator de risco em cada grupo, logo após o pesquisador escolhe para cada caso um controle similar, dessa forma os efeitos das variáveis ficam automaticamente ajustados, Ao se ter apenas duas probabilidades a de ter a doença ou a de não ter a doença na variável  $X$ , teremos então uma tabela  $2 \times 2$ , ou seja, uma tabela de contingência similar à da Tabela 1.

Sendo assim, é possível comparar a probabilidade de um indivíduo ter a doença com a probabilidade do indivíduo não ter a doença, em que essa probabilidade poderá ser a hipótese nula ( $H_0$ ), ou é diferente, a hipótese alternativa ( $H_1$ ).

#### 2.2.4 Tabela de contingência em estudos transversais

Nos estudos transversais, o pesquisador terá uma grande população, no qual se obterá amostra, isto é, será classificado cada indivíduo de acordo com as variáveis de interesse seja com exposição de uma determinada doença ou sem essa determinada doença, se houver apenas a possibilidade de se ter a doença ou não, o pesquisador poderá apresentar seus dados em uma tabela  $2 \times 2$  similar à que foi apresentada na Tabela 1.

Os dados apresentados na tabela de contingência, ou seja, na Tabela 1 nos permite verificar se de fato existe algum tipo de associação em relação ao indivíduo ter certa doença ou não. Com isso pode-se testar as seguintes hipóteses: ( $H_0$ ) o aparecimento da doença não depende do fato do indivíduo estar exposto ao fator de risco, ou ( $H_1$ ) o aparecimento da doença está associado a exposição do indivíduo está exposto ao fator de risco.

#### Teste de independência de qui-quadrado em tabela $r \times c$

Segundo Pagano e Gauvreau (2004), numa tabela  $2 \times 2$  o teste qui-quadrado para proporções independentes é equivalente ao teste de hipótese que usa a aproximação normal para a distribuição binomial. Pode ser generalizado para acomodar a comparação de três ou mais proporções. Em situações similares, dispomos os dados a uma tabela de contingência  $r \times c$ , onde  $r$  é o número de linhas e  $c$  o número de colunas.

Como as variáveis  $X$  e  $Y$  podem assumir mais de dois valores, dadas as variáveis  $X = x_1, x_2, \dots, x_r$  e  $Y = y_1, y_2, \dots, y_c$ , pode-se construir a tabela de contingência  $r \times c$  como indicado na Tabela 2.

Variável $Y$	Variável $X$				Total
	$x_1$	$x_2$	$\dots$	$x_c$	
$y_1$	$n_{11}$	$n_{12}$	$\dots$	$n_{1c}$	$n_{1.}$
$y_2$	$n_{21}$	$n_{22}$	$\dots$	$n_{2c}$	$n_{2.}$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
$y_r$	$n_{r1}$	$n_{r2}$	$\dots$	$n_{rc}$	$n_{r.}$
Total	$n_{.1}$	$n_{.2}$	$\dots$	$n_{.c}$	$n$

Tabela 2: Tabela  $r \times c$ .

## 2.3 Teste de Associação

A seguir serão apresentados os principais testes de associação

### 2.3.1 Teste de qui-quadrado ( $\chi^2$ )

O qui-quadrado, simbolizado por ( $\chi^2$ ), conhecido em algumas literaturas como (Chi-quadrado), é um teste de hipóteses que se destina a encontrar um valor de dispersão para duas variáveis nominais, ou seja, um teste estatístico para avaliar a associação existente entre as variáveis qualitativas. O teste ( $\chi^2$ ) nos diz em que medida os valores observados se desviam dos valores esperados, onde as duas variáveis não são correlacionadas, é um teste não paramétrico, ou seja, não depende dos parâmetros populacionais como: média e variância (OLINDA, 2011).

Portanto, quanto maior for o valor da estatística de qui-quadrado, mas significativa é a relação entre as variáveis dependente e independentes, esse teste serve para verificar a hipótese de duas variáveis qualitativas, isto é categóricas, pode-se também testar se as variáveis quantitativas são independentes, porém elas devem ser agrupadas em classes, ou seja, elas devem ser categorizadas. O princípio básico desse teste é comparar proporções, isto é, as possíveis divergências entre as frequências observadas e as esperadas para certo tipo de evento, portanto podemos dizer que dois grupos se comportam de forma semelhante se as diferenças entre as frequências observadas e as esperadas em cada categoria forem muito pequenas, ou seja, próximas de zero.

A principal aplicabilidade do teste de qui-quadrado é, verificar se as frequências com que um determinado acontecimento que será observado em uma determinada amostra se desvia significativamente ou não da frequência com que é esperado, a fim de avaliar se as proporções observadas desses eventos mostram ou não as diferenças significativas ou se as

amostras diferem significativamente quanto as proporções desses acontecimentos.

### 2.3.1.1 Condições necessárias para aplicação do teste de qui-quadrado ( $\chi^2$ )

Para se aplicar corretamente o teste de qui-quadrado, algumas suposições precisam ser atendidas, entre as quais destacam-se as seguintes:

- i) As observações devem ser independentes;
- ii) As observações sejam selecionadas aleatoriamente;
- iii) As observações devem ser frequências ou contagens;
- iv) Cada observação pertence a uma única categoria;
- v) A amostra deve ser relativamente grande, ou seja, o tamanho da amostra deve ser superior a 40 (OLINDA, 2011).

### 2.3.1.2 Hipóteses a serem testadas

O teste de independência de  $\chi^2$  trabalha com duas hipóteses a serem testada em que  $n_{ij}$  é a frequência observada e  $e_{ij}$  é a frequência esperada, que são dadas por.

$$\begin{cases} H_0 : n_{ij} = e_{ij} \\ H_1 : n_{ij} \neq e_{ij} \end{cases}$$

( $H_0$ ): Entre as variáveis não há associação, em que as frequências observadas são iguais as frequências esperadas, ou seja as variáveis são independentes.

( $H_1$ ): Entre as variáveis há associação, em que as frequência observada não são iguais as frequências esperadas, ou seja, as variáveis são dependentes.

### 2.3.1.3 Como calcular

Na população em que a amostra é escolhida é denotado por  $p_{ij}$  a probabilidade real de que uma observação seja simultaneamente do tipo  $y_i$  e  $x_j$ ,  $i = 1,2$  e  $j = 1,2$ . Assim pode-se calcular o valor esperado para a casela  $(i, j)$  da tabela de contingência como:  $E_{ij} = np_{ij}$ , mas dada a hipótese nula de independência e pela definição de independência temos que sob  $H_0$ ,  $p_{ij} = p_i.p_j$  e logo,  $E_{ij} = np_i.p_j$ . O teste qui-quadrado compara as frequências observadas em cada categoria da tabela de contingência (representadas

por  $(n_{ij})$  com as frequências esperadas, sendo  $H_0$  verdadeira (denotadas por  $(e_{ij})$ ), é de grande importância montar a tabela com as frequências esperadas, em que as frequências esperadas são calculadas por meio da seguinte expressão:

$$e_{ij} = \frac{n_{i.}n_{.j}}{n_{..}}$$

ou seja, defini-se uma tabela de frequências esperadas para o caso de uma tabela  $2 \times 2$ , sob a hipótese de independência entre as variáveis  $X$  e  $Y$ . Pode-se observar que os totais marginais são mantidos fixos, conforme a Tabela 3.

Variável $Y$	Variável $X$		Total
	$x_1$	$x_2$	
$Y_1$	$e_{11}$	$e_{12}$	$n_{1.}$
$Y_2$	$e_{21}$	$e_{22}$	$n_{2.}$
Total	$n_{.1}$	$n_{.2}$	$n_{..}$

Tabela 3: Tabela  $2 \times 2$  de frequências esperadas.

Portanto, a estatística  $\chi^2$  é usada para determinar se os desvios entre as contagens observadas e as esperadas  $n_{ij} - e_{ij}$ , que é a diferença entre a frequência observada e a esperada em uma classe, são muito grandes para serem atribuídos ao acaso, uma vez que há mais de uma casela na tabela, esses desvios serão combinados de alguma forma para realizar o teste de  $\chi^2$  para as contagens de uma tabela de contingência com  $r$  que é o número de linhas e  $c$  o número de colunas. Assim, as hipóteses de homogeneidade podem ser generalizadas para  $r$  categorias da variável  $Y$  e  $c$  categorias da variável  $X$ , ou seja,  $i = 1, 2, \dots, r$  e  $j = 1, 2, \dots, c$  (OLINDA, 2011). Karl Pearson um importante estatístico, propôs a seguinte estatística para medir as diferenças entre os valores observados e os valores esperados que é dada por:

$$\chi^2 = \sum_{i=1}^r \sum_{j=1}^c \frac{(n_{ij} - e_{ij})^2}{e_{ij}}, \quad (2.1)$$

#### 2.3.1.4 Conclusão do teste de $\chi^2$

Se o  $\chi^2$  ficar próximos de zero tem-se a evidência a favor de  $(H_0)$ , pois a diferença entre os valores observados  $(n_{ij})$  e esperados  $(e_{ij})$  ficam próximo de zero, enquanto os

valores grandes de  $\chi^2$  mostram evidência contra ( $H_0$ ). Pode-se mostrar que a estatística  $\chi^2$  tem distribuição assintótica de qui-quadrado com  $(r-1)(c-1)$  que em tabelas de dupla entrada, ou seja, com variáveis binárias resulta em 1 grau de liberdade. Para um valor de  $\alpha$  pré-fixado, em que é definida a região crítica do teste, essa tomada de decisão é feita comparando-se o valor do  $\chi^2$  calculado com o valor do  $\chi^2$  tabelado.

- i Se  $\chi_{cal}^2 \leq \chi_{tab}^2$ , Se o valor calculado da estatística de  $\chi^2$  é menor que o valor tabelado do teste, há indícios para não rejeitar a hipótese da nulidade ( $H_0$ ) com  $\alpha$  de significância, ou seja, aceita-se a hipótese de independência.
- ii se  $\chi_{cal}^2 \geq \chi_{tab}^2$ , Se o valor calculado da estatística de  $\chi^2$  é maior que o valor tabelado do teste, há indícios para rejeitar a hipótese da nulidade ( $H_0$ ) a favor da alternativa ( $H_1$ ) com  $\alpha$  de significância, ou seja, não se aceita a hipótese de independência.

### 2.3.2 Correção de continuidade de Yates para o teste de $\chi^2$

Um problema muito importante associado aos testes de qui-quadrado está baseado no fato de a estatística de teste ser essencialmente discreta, mas aproximada por uma distribuição contínua chamada qui-quadrado, como as tabelas de contingência são geradas a partir de variáveis categóricas e com tamanhos amostrais finitos, os possíveis valores de  $\chi^2$  são enumeráveis.

Para tabelas  $2 \times 2$  com 1 grau de liberdade, essa aproximação de distribuição fornece bons resultados para amostras grandes, sendo assim para amostras pequenas o teste não fornece bons resultados, logo o uso do teste de qui-quadrado requer a aplicação de uma correção chamada correção de continuidade ou correção de Yates, na qual a diferença absoluta entre as frequências observadas ( $n_{ij}$ ) e as esperadas ( $e_{ij}$ ), tomadas em módulo, que consiste em subtrair 0,5 (OLINDA, 2011). Assim, a expressão estatística de  $\chi^2$  corrigida é dada por.

$$\chi_y^2 = \sum_{i=1}^s \sum_{j=1}^r \frac{(|n_{ij} - e_{ij}| - 0,5)^2}{e_{ij}}, \quad (2.2)$$

em que  $\chi_y^2$  é o qui-quadrado com correção de Yates, o que diminui um pouco o valor da estatística  $\chi^2$  e aumenta o nível de significância do teste. Alguns autores argumentam que a correção de Yates só pode ser aplicada em tabelas  $2 \times 2$ , ou seja, para variáveis dicotômicas, outros sugerem que ela pode ser usada para qualquer tipo de tabela de contingência, mas o fato principal é que esse tipo de método é muito conhecido na literatura

estatística.

Uma regra prática encontrada em muitos livros de estatística aplicada diz que se o tamanho amostral for maior que 40, pode-se utilizar o teste de  $\chi^2$  convencional, (sem correção de Yates), se  $n$  estiver entre 20 e 40, deve-se utilizar a correção de Yates e se  $n < 20$  deve-se usar o teste exato de Fisher (OLINDA, 2011).

### 2.3.3 Teste exato de Fisher

Como o teste de associação de qui-quadrado não é recomendável para amostras pequenas, o teste exato de Fisher é um teste de significância estatística utilizado na análise de tabelas  $2 \times 2$ , e é indicado quando a amostra é considerada pequena. O Teste exato de Fisher foi proposto por Ronald A. Fisher e por homenagem a ele o teste acabou sendo chamado por teste exato de Fisher, que pertence a uma classe de teste exatos, assim chamado por que o significado do desvio de uma hipótese nula pode ser calculado exatamente, ao invés de depender de uma aproximação que se torna exata, assim o teste é útil para dados categóricos que tem por objetivo testar a hipótese de duas variáveis para examinar se tem associação ou não, esse teste assume que os totais marginais das duas variáveis são fixos.

#### 2.3.3.1 As características do Teste Exato de Fisher

As principais características do teste são:

- i) Possuir duas populações;
- ii) Cada população ser subdividida em duas categorias exclusivas, estatisticamente conhecidas como Fator 1 e Fator 2, com as mesmas categorias para as duas populações;
- iii) esta subdivisão tem que ser mutuamente exclusiva, ou seja, cada elemento de uma população irá pertencer a exatamente uma destas categorias;
- iv) Observa-se, em cada população, a quantidade de elementos pertencentes a cada categoria;
- v) O Teste exato de Fisher visa comparar se as quantidades de ocorrências nestas categorias são ou não equivalentes nas duas populações.

Para realizar o teste exato de Fisher precisa-se fixar o nível de significância  $\alpha$ , manter fixas as frequências marginais da tabela de dados, ou seja, os totais de linhas e colunas,

variando os números dentro da tabela, escrevendo todas as tabelas possíveis mantendo os mesmo totais marginais. Consequentemente, calcula-se a probabilidade de cada tabela utilizando a seguinte expressão:

$$P = \frac{n_{1.}!n_{2.}!n_{.1}!n_{.2}!}{n_{..}!} \frac{1}{n_{11}!n_{12}!n_{21}!n_{22}!} \quad (2.3)$$

Em seguida calcula-se a probabilidade de ocorrerem frequências iguais ou menores do que as observadas, considerando todos os resultados possíveis, e somar todas as probabilidades das frequências observadas. O procedimento do teste rejeita a hipótese de independência, se a probabilidade de ocorrerem frequências iguais, ou menores do que a observada, for menor do que o nível de significância  $\alpha$  estabelecido, caso contrário não rejeita-se a hipótese de independência.

## 2.4 Medidas de associação nas Tabelas $2 \times 2$

De acordo com Vieira (2003), os pesquisadores consideram que a análise estatística está pronta logo que termina de aplicar o teste de  $\chi^2$ , mas isso não é verdade, pois, é importante estimar o grau de associação entre as duas variáveis, mesmo porque o teste de  $\chi^2$  serve para verificar a significância da associação, mas não para medir o grau de associação entre duas variáveis. Portanto a significância de todo teste estatístico depende do tamanho da amostra, ou seja, a significância não depende só das diferenças entre as proporções, mas do tamanho da amostra, o grau de associação, no entanto, independe do tamanho da amostra.

Ao rejeitar a hipótese nula, há indício de associação entre duas variáveis, a associação entre duas ou mais variáveis implica que o conhecimento de uma altera a probabilidade do resultado da outra. São várias as medidas de associação, uma das mais utilizadas é o coeficiente de contingência de Pearson, e a medida de associação mais utilizada na área da bioestatística é a razão de chances, em que é conhecido pelo termo em inglês, *odds ratio* (LOPES, 2011).

### 2.4.1 Coeficiente de contingência de Pearson

O coeficiente de contingência de Pearson é uma medida de associação baseada na estatística  $\chi^2$  do teste de qui-quadrado, e é dada pela expressão:

$$C = \sqrt{\frac{\chi^2}{\chi^2 + n}} \quad (2.4)$$

como esse coeficiente é falho em determinar o sentido da associação, porém ele é uma das medidas de associação mas utilizadas, em que  $n$  é o tamanho da amostra e os valores de  $C = 0$  indica independência e valores de  $C > 0$  sugere associação, ou seja, o coeficiente varia de um intervalo  $0 \leq C < 1$ , além disso não se pode determinar o sentido da associação, porém o coeficiente nunca apresenta o limite superior máximo, ou seja, ela nunca atinge o valor 1 esse é um tipo de falha. Portanto para contornar esse problema, sugere-se o uso do coeficiente corrigido que é expressado por.

$$C_{cor} = C \times \frac{\min(c, r)}{\min(c, r) - 1}, \quad (2.5)$$

o  $c$  é o número de categorias da variável  $Y$  e  $r$  é o número de categorias da variável  $X$ .

### 2.4.2 Razão de chances ou *Odds-Ratio*

Não há consenso em como traduzir o termo *odds-ratio*. Prefere-se o nome razão de chances que é criticado porque a palavra razão vem de produtos cruzados e chance está mais relacionada com probabilidade, conceito diferente daquele captado pela palavra *Odds*. Opta-se por criar um novo sentido para a palavra chances ao invés de criar uma nova expressão, como defendido por aqueles que usam termos como razão de *Odds* (SOARES; SIQUEIRA, 2002).

O evento ocorre com probabilidade  $p$ , quando a chance desse evento ocorrer é dada pela probabilidade do evento ocorrer dividida pela probabilidade desse evento não ocorrer, como podemos ver na expressão.

$$chance = \frac{p}{(1 - p)}$$

No contexto epidemiológico, pode-se definir a chance como uma probabilidade condicional, ou seja, a probabilidade do indivíduo possuir o desfecho dado que está exposto ao



fator de risco (JEWELL, 2004), que é expresso por.

$$P(D|E) = \frac{P(D \cap E)}{P(E)} = \frac{n_{11}}{n_{11} + n_{12}} = \frac{n_{11}}{n_1} = \hat{p}_1 \quad (2.6)$$

em que,  $n_1$  é o número de indivíduo expostos na amostra e  $\hat{p}_1$  é a proporção observada dos indivíduos que possuem o desfecho entre os expostos. Assim, quando o indivíduo é exposto ao fator de risco, a chance dele possuir o desfecho é expressa por.

$$chance(D|E) = \frac{n_{11}/n_1}{(n_1 - n_{11})/n_1} = \frac{n_{11}}{n_1 - n_{11}} = \frac{n_{11}}{n_{11} + n_{12} - n_{11}} = \frac{n_{11}}{n_{12}} \quad (2.7)$$

Da mesma forma calcula-se quando o indivíduo não é exposto ao fato de risco, a chance dele possuir o desfecho e a probabilidade de possuir o desfecho é expressa por.

$$P(D|\bar{E}) = \frac{P(D \cap \bar{E})}{P(\bar{E})} = \frac{n_{21}}{n_{21} + n_{22}} = \frac{n_{21}}{n_2} = \hat{p}_2 \quad (2.8)$$

em que,  $n_2$  é o número de indivíduo não é exposto na amostra e  $\hat{p}_2$  é a proporção observada dos indivíduos que possuem o desfecho entre os não exposto. Assim quando o indivíduo não é exposto ao fator de risco, a chance dele possuir o desfecho é expressa por.

$$chance(D|\bar{E}) = \frac{n_{21}/n_2}{(n_2 - n_{21})/n_2} = \frac{n_{21}}{n_2 - n_{21}} = \frac{n_{21}}{n_{21} + n_{22} - n_{21}} = \frac{n_{21}}{n_{22}} \quad (2.9)$$

Portanto, o *Odds-Ratio* é definido como a chance dos indivíduos que possuem o desfecho entre os expostos, dividido pela chance dos indivíduos que possuem o desfecho entre os não expostos, como é expresso abaixo.

$$\widehat{OR} = \frac{n_{11}/n_{12}}{n_{21}/n_{22}} = \frac{n_{11}n_{22}}{n_{12}n_{21}} \quad (2.10)$$

Portanto, se o  $\widehat{OR} > 1$  a exposição ao fator de risco contribui para o desenvolvimento do desfecho e se  $\widehat{OR} < 1$  a exposição ao fator não contribui para o desenvolvimento do desfecho, ou seja, a exposição é um fator de proteção.

## 3 Aplicação

### 3.1 Descrição dos Dados

Os dados analisados neste trabalho foram oriundo de uma pesquisa realizada por Rocha (2012), aluna do mestrado em Saúde Pública pela Universidade Estadual da Paraíba-UEPB, o conjunto de dados reais foi analisado por meio do *software* R versão 2.14.2 com a interface gráfica do RStudio, e que este trabalho gerou um manuscrito submetido para uma revista nacional da área de saúde do qual fiquei como co-autora. As variáveis de estudo foram situações alimentar, nutricional e atenção a saúde materna-infantil na cidade de Queimadas, Paraíba, em que esses dados foram coletados entre os meses de julho e agosto de 2011. Neste estudo foram incluídas todas as crianças nascidas durante o ano de 2009 residentes em Queimadas Paraíba e os enfermeiros de todas as unidades de saúde da família que compõe o sistema único de saúde do município.

### 3.2 Análise dos Dados

Como os dados já estavam disponíveis no *software Excel*, eles foram apenas importados para o *software* R usando o comando `read.xls` disponível através do pacote `gdata`. Para a realização do teste de qui-quadrado foi utilizada a função `chisq.test` disponível por padrão na instalação do R. Nos dados que tem apenas 16 observações, o teste de qui-quadrado não é o mais adequado. Desta forma uma alternativa nesta situação é o teste exato de Fisher, disponível no R através do comando `fisher.test`. Por fim, para os cálculos das razões de chances foi utilizada a função `epitab` disponível no pacote `epitools` (observe que este comando também calcula o nível descritivo do teste de qui-quadrado para as categorias individuais da variável explicativa). Para isso, é necessário que os pacotes sejam carregados no R.

```
library(gdata)
```

```
library(epitools)
```

A seguir, o comando utilizado para importar os dados do Excel no R.

```
dados1 = read.xls("../Banco_Final_Carol.xls", sheet=1)
dados3 = read.xls("../Banco_Final_Carol.xls", sheet=3)
```

Inicialmente, trabalhou-se com as variáveis envolvidas com as crianças, as mães e as famílias de acordo com a localização do domicílio (zona urbana ou zona rural) e de acordo com o risco nutricional (R.N.) da criança utilizaram-se as variáveis especificadas a seguir: relacionadas às crianças - sexo, peso ao nascer (P.N.), relação peso/idade com base na posição do último peso(U.P) registrado no cartão, direção da curva (D.C.) considerando os dois últimos pesos no gráfico de P/I no cartão da criança; relacionadas às mães - idade, anos de estudos (A.E.), situação conjugal (S.C.); relacionadas às condições sócio-econômicas das famílias - renda per capita (R.P.C); posse de bens (P.B.); número de cômodos no domicílio (N.C.D); número de pessoas no domicílio (N.P.D); abastecimento de água no domicílio(A.A.D.); tratamento da água (T.A.); destino do lixo (D.L.); escoamento sanitário (E.S.); benefício de programa social (B.P.S.).

Para se trabalhar com as variáveis envolvidas com a estrutura do serviço e o processo de trabalho nas Unidades Básicas de Saúde da Família (UBSF) para o acompanhamento do crescimento das crianças de acordo com a localização do domicílio (zona urbana ou zona rural) e de acordo com o risco nutricional (R.N.) da criança, utilizaram-se as variáveis especificadas a seguir: relacionadas à estrutura: completude da equipe mínima (E.M.) da estratégia de saúde da família (ESF); apoio às equipes por nutricionista e pediatra dos NASFs (E. NASF)(Núcleos de Apoio à Saúde da Família); existência de equipamento pra medir crianças (E.M.); existência de equipamento pra pesar crianças (E.P.); disponibilidade de Cartões/Cadernetas de saúde da criança (C.S.C.), existência de enfermeiro treinado para atendimento à criança (E.T.), utilização de protocolos para o acompanhamento do crescimento da criança (U.P.A).

Para estabelecer diferenças entre a zona urbana e a zona rural, assim como entre as crianças com risco nutricional e aquelas isentas desse risco foi utilizado o teste de qui-quadrado, considerando um valor de  $p < 0,05$  para denotar significância estatística.

Inicia-se a verificação da existência de associação entre as variáveis envolvidas por meio do teste de  $\chi^2$  e do respectivo nível descritivo para denotar a significância estatística. Foram analisadas 200 observações nesta pesquisa em situação alimentar, nutricional e

atenção à saúde materno-infantil, no município de Queimadas no estado da Paraíba, as crianças nascidas vivas em 2009 foram incluídas no estudo, assim serão apresentados as análises individuais para a existência de associação entre diversos pares de variáveis, na Tabela 4 iremos apresentar as variáveis envolvidas com as crianças do estudo.

Teste	Nível de Significância
Zona × Sexo	0,7218
R.N. × Sexo	0,8238
Zona × P.N.	0,5437
R.N. × P.N.	0,0003
Zona × U.P.	0,1673
R.N. × U.P.	< 0,0001
Zona × D.C.	0,3354
R.N. × D.C.	< 0,0001

Tabela 4: Teste de associação para as variáveis envolvidas com as crianças em estudo.

Na Tabela 4 estão descritas as características das crianças, na qual pode-se observar que algumas variáveis tem associação, ou seja, é estatisticamente significativo, pois o nível de significância é menor que  $\alpha$  pré-fixado de 0,05 de probabilidade, as variáveis envolvidas são risco nutricional com peso ao nascer, risco nutricional com último peso e risco nutricional com direção da curva que apresentam significância estatística para associação entre si.

As variáveis trabalhadas, que envolvem as mães em estudos para analisar o grau de associação entre as variáveis, pode-se observar na Tabela 5.

Teste	Nível de Significância
Zona × Idade	0,2967
R.N. × Idade	0,3812
Zona × A.E.	0,0376
R.N. × A.E.	0,2817
Zona × S.C.	0,4837
R.N. × S.C.	0,5801

Tabela 5: Teste de associação para variáveis envolvidas com as mães em estudo.

Na Tabela 5 estão descritas as características das mães, em que pode-se observar que apenas uma variável tem associação, ou seja, é estatisticamente significativo, pois o nível de significância é menor que  $\alpha$  pré-fixado de 0,05 de probabilidade, as variáveis envolvidas são zona vs anos de estudos da mãe, que apresentam significância estatística para associação entre si.

As variáveis envolvidas com às condições sócio-econômicas das famílias em estudos, pode ser observada na Tabela 6.

Teste	Nível de Significância
Zona × R.P.C	0,0342
R.N. × R.P.C	0,4432
Zona × P.B.	0,4817
R.N. × P.B.	0,0502
Zona × N.C.D	0,0595
R.N. × N.C.D	0,2955
Zona × N.P.C	0,7539
R.N. × N.P.C	0,3416
Zona × A.A.D	< 0,0001
R.N. × A.A.D	0,3876
Zona × T.A	0,8026
R.N. × T.A	0,5838
Zona × D.L.	< 0,0001
R.N. × D.L.	0,9193
Zona × E.S.	< 0,0001
R.N. × E.S.	0,0484
Zona × B.P.S	< 0,0001
R.N. × B.P.S	0,5548

Tabela 6: Teste de associação para as variáveis envolvida com as famílias em estudo.

Logo, na Tabela 6 estão descritas as características das famílias observando-se que as variáveis tem associação, ou seja, tem significância estatística, pois o nível de significância é menor que  $\alpha$  pré-fixado de 0,05 de probabilidade, as variáveis envolvidas são zona vs renda per capita, zona vs abastecimento de água no domicílio, zona vs destino do lixo, zona vs escoamento sanitário, risco nutricional vs escoamento sanitário, zona vs benefício do programa social que apresentam significancia estatistica para associação entre si.

O teste de qui-quadrado foi realizado de acordo com os resultados das Tabelas (4, 5, 6)

Risco Nutricional	Peso ao Nascer		Total
	Adequado	Baixo Peso	
Adequado	87	0	87
Risco	69	11	80
Total	156	11	167

Tabela 7: Tabela de contingência de risco nutricional *versus* peso ao nascer

Feito a análise no *software* R, pode-se observar na Tabela 7 de contingência, que foi calculado o teste de qui-quadrado e tem-se que o valor de  $\chi^2 = 12.806$  e seu nível descritivo é 0.0003455, afirmando-se que existe associação entre o risco nutricional e o peso ao nascer da criança. Na medida de associação verifica-se que não é possível executar a estatística do teste da razão de chance entre todas as variáveis que deram significativas devido não ter o aparecimento de informações, em que uma das caselas está zerada como pode-se observar na Tabela 7 e como a estatística da razão de chance é feita através do produto cruzado a estatística será sempre infinita.

Para se trabalhar com as variáveis envolvidas com a estrutura do serviço e o processo de trabalho nas (UBSF) para o acompanhamento do crescimento das crianças de acordo com a localização do domicílio (zona urbana ou zona rural) e de acordo com o risco nutricional (R.N.) da criança utilizaram-se as variáveis especificadas a seguir: relacionadas à estrutura: completude da equipe mínima da estratégia de saúde da família (ESF), (E.M.); apoio às equipes por nutricionista e pediatra dos NASFs (E. NASF)(Núcleos de Apoio à Saúde da Família); existência de equipamento pra medir crianças (E.M.); existência de equipamento pra pesar crianças (E.P.); disponibilidade de Cartões/Cadernetas de saúde da criança (C.S.C.), existência de enfermeiro treinado para atendimento à criança (E.T.), utilização de protocolos para o acompanhamento do crescimento da criança (U.P.A). Para situações com tamanhos amostrais muito pequenos (como é o caso aqui, com  $n = 16$ ), mesmo com o teste exato de Fisher, que é mais indicado nestas situações, é muito difícil conseguir encontrar resultados significativos em situações como estas. Na sequência são apresentadas as análises individuais para a existência de associação entre diversos pares de variáveis. Trabalhando com as variáveis envolvidas com a UBSF. Na Tabela 8 de caracterização mostra o resultado do teste exato de Fisher para associação e seus respectivos níveis descritivos entre as variáveis:

Teste	Nível de Significância
L. UBSF $\times$ E.M.	0,2143
L. UBSF $\times$ E. NASF	1,0000
L. UBSF $\times$ E.M.	1,0000
L. UBSF $\times$ E.P.	1,0000
L. UBSF $\times$ CSC	0,3125
L. UBSF $\times$ E.T	1,0000
L. UBSF $\times$ U.P.A	0,2445

Tabela 8: Teste exato de Fisher para associação entre as variáveis, envolvidas com a UBSF.

Pode-se observar que na Tabela 8, que mesmo utilizando o teste exato de Fisher que é o mais apropriado para amostra pequenas, mesmo assim o teste é não significativo para nenhuma das variáveis envolvidas a que está mais próxima de algum tipo de associação e a localização da UBSF vs equipe mínima, mas mesmo assim não podemos afirmar que possa existir algum tipo de associação devido a amostra ser muito pequena, pois o nível de significância foram superior a o  $\alpha$  pré-fixado de 0,05 de probabilidade.

## 4 Conclusão

A partir dos métodos utilizados neste trabalho verifica-se que as técnicas estatísticas são bastante úteis em diversas áreas principalmente na área da saúde, pode-se observar no decorrer deste trabalho a importância do teste de qui-quadrado em relação a verificação das evidências de existência de associação entre as variáveis envolvidas, e também o teste exato de Fisher que é mais apropriado para amostras pequenas como é o caso das variáveis envolvidas com as estruturas do serviço e o processo de trabalho na UBSF para o acompanhamento do crescimento das crianças de acordo com a localização do domicílio e de acordo com o risco nutricional da criança com  $n = 16$ , outra técnica é a medida de associação para medir o grau de associação utilizando a razão de chances que é a medida de associação mais importante utilizada na bioestatística.

Por meio da aplicação verificou-se a associação entre as variáveis envolvidas utilizando o teste de qui-quadrado para os casos em que se tinha um número grande de observações, e verificou-se que existe associação em algumas variáveis envolvidas, pois o nível descritivo foi menor que 0,05. Utilizou também o teste exato de Fisher, nos casos onde o número de observações era pequeno, as variáveis envolvidas não obteve associação mesmo utilizando o teste exato de Fisher que é o mais apropriado para amostras pequenas pois seu nível descritivo deram acima de 0,05.



# Referências

- ESTEVEVES, G. H. *Notas de aula*. Campina Grande, 2010.
- JEWELL, N. P. *Statistics for Epidemiology*. [S.l.]: Berkeley: Chapman & Hall, 2004.
- FIGUEIRÊDO BÁRBARA CAMBOIM LOPES de. *Construção do Intervalo de Confiança para a Razão de Chances e Risco Relativo*. Campina Grande, 2011.
- OLINDA, R. A. de. *Notas de aula*. Campina Grande, 2011.
- PAES, A. T. *Itens Essenciais em Bioestatística*. São Paulo: [s.n.], 1998. Disponível em: <[http://www.scielo.br/scielo.php?script=sci\\_arttext&pid=S0066-782X1998001000003](http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0066-782X1998001000003)>.
- PAGANO, M.; GAUVREAU, K. *Princípios de Bioestatística*. 2. ed. São Paulo: Thomson, 2004.
- ROCHA, A. C. D. *Assistência pré-natal e vigilância do crescimento infantil no contexto da estratégia Saúde da Família em Queimadas, Paraíba, Brasil*. Dissertação (Mestrado em Saúde Pública) — Universidade Estadual da Paraíba, Campina Grande, 2012.
- SOARES, J. F.; SIQUEIRA, A. L. *Introdução à Estatística Médica*. 2. ed. Minas Gerais: coopmed, 2002.
- SOUNIS, E. *Bioestatística: princípios fundamentais, metodologia, estatística, aplicação às ciências biológicas*. 2. ed. São Paulo: Mcgraw-hill do Brasil, LTDA, 1975.
- VIEIRA, S. *Bioestatística Tópicos Avançados*. 2. ed. Rio de Janeiro: Elsevier, 2003.