UNIVERSIDADE ESTADUAL DA PARAÍBA
CENTRO DE CIÊNCIAS E TECNOLOGIA
DEPARTAMENTO DE ESTATÍSTICA

MÁRIO SILVANO ALEXANDRE PEREIRA JÚNIOR

Testes de associação para dados de leucemia.

CAMPINA GRANDE
MAIO DE 2016

MÁRIO SILVANO ALEXANDRE PEREIRA JÚNIOR

Testes de associação para dados de leucemia.

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

Orientador: Prof. Drº Tiago Almeida de Oliveira

CAMPINA GRANDE

MAIO DE 2016

P436t Pereira Júnior, Mário Silvano Alexandre.
Testes de associação para dados de leucemia [manuscrito] / Mário Silvano Alexandre Pereira Júnior. - 2016.
30 p.

Digitado.
Trabalho de Conclusão de Curso (Graduação em Estatística) - Universidade Estadual da Paraíba, Centro de Ciências e Tecnologia, 2016.
"Orientação: Prof. Dr. Tiago Almeida de Oliveira, Departamento de Estatística".

1. Testes categorizados. 2. Dados categorizados. 3. Estatística. 4. Leucemia. I. Título.

21. ed. CDD 519.5

MÁRIO SILVANO ALEXANDRE PEREIRA JÚNIOR

Testes de associação para dados de leucemia.

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

Aprovado em: 17 / 05 / 2016

Banca Examinadora:

Prof. Drº Tiago Almeida de Oliveira
Universidade Estadual da Paraíba - DE/CCT
Orientador

Prof. Drº Ricardo Alves de Olinda - DE/CCT
Examinador

Prof. Drº Gustavo Henrique Esteves
Universidade Estadual da Paraíba - DE/CCT
Examinador

Dedicatória

Dedico este trabalho a todos que me incentivaram e colaboraram para que eu pudesse chegar até aqui.

MINHA HOMENAGEM

Agradecimentos

Agradeço primeiramente a Deus pela vida e por está presente em todos os momentos dela, pela família que ele me deu, por toda sabedoria e conhecimento que ele tem me dado.

Aos meus pais, por terem me dado uma ótima formação, por terem feito eu me tornar esse homem que eu sou, por sempre estarem comigo me apoiando e orientando em tudo que preciso.

A minha irmã por está comigo sempre me apoiando e me incentivando no que for preciso. A minha avó, por está presente em minha vida, sempre disposta a me ajudar em tudo.

E a todos os familiares e amigos que me apoiaram e me incentivaram.

A todos os professores do departamento de Estátistica da UEPB, principalmente ao professor Tiago que foi o meu orientador desse trabalho.

Resumo

Este trabalho de conclusão de curso tem como objetivo produzir um material bem prático sobre a análise de dados categorizados, que possa auxiliar a compreender o desenvolvimento de suas técnicas, especialmente quando aplicados os testes categorizados. Compreende-se por dados categóricos, dados estatísticos que podem ser representados por números, contagens de pessoas, objetos em diferentes áreas. Neste contexto, pretende-se com o trabalho fazer um estudo bem aprofundado sobre a análise de dados categorizados e seus respectivos testes. Os dados foram obtidos do site do INCA(Instituto Nacional do Câncer). Foi feito uma proporção para 800000 pessoas (400000LL,400000LM), os quais apresentam uma distinção para os tipos de cancêr: Leucemia Linfóide(LL) e Leucemia Mielóide(LM). Na realização deste trabalho serão aplicados os testes qui-quadrado, G^2, e o teste de Mantel-Haenszel para verificar a relação existente entre as variáveis e a doença, em seguida utilizaremos o teste da razão de chances para verificar a chance de aumento da doença com o passar do tempo. Além disso, utilizou-se o software R para realizar as análises estatísticas solucionando os problemas.

Palavras-chave: Dados categorizados, Testes categorizados, leucemia

Abstract

This course conclusion work aims to produce a very practical material on the analysis of categorical data, which can help to understand the development of their techniques, especially when categorized tests applied. It is understood by categorical data, statistical data that can be represented by numbers, headcounts, objects in different areas. In this context, we intend to work to make a very thorough study of the categorized data analysis and their respective tests. Data were obtained from INCA's website (National Cancer Institute). It was made to a proportion 800,000 people (400000LL, 400000LM), was made which have a distinction for Cancers: Lymphoblastic Leukemia (LL) and Myeloid Leukemia (LM). In this work will be applied the chi-square test, G^2, and the Mantel-Haenszel test to check the relationship between variables and disease, then we use the test odds ratio to check chance increase of the disease over time. In addition, we used the R software for statistical analysis solving the problems.

Keywords: Data categorized, categorized tests, leukemia

Sumário

Lista de Tabelas

1 Introdução

As leucemias são cânceres das células do sangue. As células cancerosas podem acometer toda a medula óssea, chegando ao ponto de impedir a produção de células normais do sangue (falência medular), o que levaria a quadros variáveis de sangramento, infecção e anemia. Elas são divididas em agudas e crônicas. O grupo das leucemias agudas é dividido em mieloblástica e linfocítica, sendo que essa diferenciação é feita na célula de origem de cada grupo. De forma geral as leucemias agudas apresentam uma evolução muito rápida, sendo necessário o diagnóstico precoce e o tratamento rápido.

A incidência das leucemias é semelhante por todo o mundo, sendo que, dentre as leucemias agudas, a mieloblástica tem ligeira predominância sobre a linfocítica. São mais predominantes nos homens, sendo maior o número de casos nos de raça branca. A idade de acometimento difere enormemente entre dois grupos, sendo a leucemia linfocítica aguda (LLA) muito comum até os 10 anos de idade e a leucemia mielóide aguda (LMA) muito comum na média de 65 anos de idade. O tratamento entre os dois grupos também é muito diferente. Além disso a leucemia mielóide aguda tem um pior prognóstico que a crônica.

Como a estatística tem sido uma ferramenta que tem ajudado nas pesquisas científicas, o uso de uma técnica apropriada para tal investigação torna as abordagem a respeito do tempo mais precisas. Dentre as técnicas utilizadas para tais questionamentos, tem-se a análise de dados categóricos, em que é frequente a necessidade de se incluir variáveis explanatórias (contínuas ou categorizadas) para levar em consideração certas características associadas aos indivíduos ou aos fatores sob estudo.

O recente desenvolvimento de métodos estatísticos para análise com dados categorizados tem estimulado a sofisticação metodológica em vários campos de interesse como: Agronomia, Biologia, Ciências da Educação, Economia, etc. Estes métodos estatísticos para dados categorizados tem apenas recentemente alcançado o mesmo nível de sofisticação, que a metodologia para dados contínuos (AGRESTI, 2002).

O objetivo desse trabalho é fazer um estudo com pessoas que apresentam dois tipos de leucemia: linfóide e mileóide, para isso utilizou-se os principais testes categorizados para identificar a relação existente entre a doença e as variáveis que estão envolvidas.

2 Fundamentação Teórica

Nesta sessão iremos abordar os principais métodos utilizados para compor este trabalho, desde a caracterização do problema até os métodos estatísticos que serão utilizados.

2.1 Tipos de leucemia

Leucemia é um termo geral dado a um grupo de cânceres que se desenvolve na medula óssea. As leucemias podem ser agrupadas com base em quão rapidamente a doença evolui e torna-se grave. Sob esse aspecto, a doença pode ser do tipo crônica (que geralmente agrava-se lentamente) ou aguda (que geralmente agrava-se rapidamente).

i) Crônica: No início da doença, as células leucêmicas ainda conseguem fazer algum trabalho dos glóbulos brancos normais. Médicos geralmente descobrem a doença durante exame de sangue de rotina. Lentamente, a leucemia crônica se agrava. À medida que o número de células leucêmicas aumenta, aparecem inchaço nos linfonodos (ínguas) ou infecções. Quando surgem, os sintomas são brandos, agravando-se gradualmente.

ii) Aguda: As células leucêmicas não podem fazer nenhum trabalho das células sanguíneas normais. O número de células leucêmicas cresce rapidamente e a doença agrava-se num curto intervalo de tempo.

2.1.1 Leucemia linfóide aguda

A leucemia linfoide aguda (LLA) é o tipo mais comum de câncer infantil onde constitui-se cerca de um terço dos tumores malignos da criança e a incidência da LLA nas crianças dos Estados Unidos é de aproximadamente 3,4 casos para 100.000 crianças menores de 15 anos de idade onde crianças com idades entre 3 e 4 anos apresentam maior incidência, a LLA têm sido mais comum em crianças brancas do que em negras e em meninos do que em meninas.

Com o desenvolvimento de combinações terapêuticas, utilizando diversas drogas citotóxicas com ou sem transplante de Stem-cell, tem aumentado o percentual de cura da

criança portadora de Leucemia linfóide aguda em mais de 80%. Essa acentuada melhora nos resultados tem produzido um aumento na população de sobreviventes. Anualmente cerca de 1.500 crianças com LLA, nos Estados Unidos, estão sendo curadas. É estimado que atualmente um em cada 1.000 adultos jovens, com menos de 20 anos de idade, seja um sobrevivente do câncer. As complicações tardias representam outra área de investigação e têm fornecido informações importantes para o planejamento do tratamento inicial no sentido de evitar tais complicações (PEDROSA E LINS, 2002)

2.1.2 Leucemia linfóide crônica

É um tipo de câncer no sangue que se dá por uma alteração no DNA de células-tronco dentro da medula óssea, nossa fábrica do sangue. Sem motivo conhecido, os linfócitos (glóbulos brancos que ajudam no combate a infecções) têm seu material genético alterado e começam a se multiplicar descontroladamente na medula óssea,acarretando o seu aumento no sangue.

O CLL é o tipo mais comum de leucemia em indivíduos caucasianos e incomum em asiáticos, em países de incidência Ocidental é de cerca de 3 novos casos por 100.000 habitantes por ano. Na Europa e nos Estados Unidos da América é responsável por 0,8% de todas as malignidades,e aproximadamente 30% de todas as leucemias. Sua incidência tem uma grande variação de acordo com a área geográfica, que varia de 2,5% de todas as leucemias no Japão a 38% na Dinamarca. O LLC tem várias manifestações clínicas. Com muita frequência, em cerca de metade dos casos, o diagnóstico é feito em um indivíduo completamente assintomáticos tem sido feito hemogramas de rotina ou que este teste não mostraram sinais de uma doença banais. Em outras ocasiões os primeiros sintomas são aparecimento de gânglios linfáticos, fadiga e mal estar. Ao contrário do que isso, acontece em linfomas, febre, transpiração e perda de peso são raros. Às vezes, elas são repetidas infecções virais, ou bacteriana, indicando que motivam um exame de sangue que revela a doença. É muito mais raro do que primeira manifestação de uma LLC ser um anemia hemolítica auto-imune (RAMIREZ, 1999).

2.1.3 Leucemia Mielóide Aguda

A Leucemia Mielóide Aguda (LMA) é uma doença maligna da medula óssea, onde mieloblastos expandem-se, acumulam-se e suprimem a atividade hematopoética normal. Tem incidência de 2,2 casos/100 mil pessoas/ano nos EUA, afetando 2,9 homens:1,9 mulheres.

É uma neoplasia hematológica heterogênea constituindo um enorme desafio diagnóstico e terapêutico. Apesar de todos os progressos no campo da onco-hematologia e da obtenção de índices de remissão pós-indução atingindo os 80%, a taxa de cura em LMA permanece em torno de 20%. Até a década de 70, somente 10% dos pacientes atingiam cinco anos livres de doença. Nos anos 80, a introdução de intensificações e, posteriormente, o transplante de medula óssea estenderam a taxa de sobrevida livre de doença em cinco anos para 40% dos pacientes. Entretanto, a LMA ainda é associada a alta taxa de mortalidade, sendo responsável pela morte de dois em cada 100 mil/indivíduos/ano nos EUA (BITTENCOURT et al.,2003).

2.1.4 Leucemia mileóide cronica

A leucemia mielóide crônica (LMC) foi descrita como forma independente de leucemia há 150 anos, em pacientes que morreram em consequência de intensa leucocitose e hepato-esplenomegalia. É uma doença clonal das células pluripotentes da medula óssea e constitui 14% de todas as leucemias,com uma incidência anual de 1,6 casos por 100 mil indivíduos. É mais frequente em adultos entre 40 e 60 anos de idade e afeta ambos os sexos,mas com predominância no sexo masculino. A progressão clínica da LMC pode ser dividida em três fases: crônica, acelerada e blástica início da fase e é "benigna". Alguns pacientes são assintomáticos, mas outros apresentam fadiga, fraqueza, dores de cabeça, irritabilidade, febre, suor noturno e perda de peso.

O diagnóstico é realizado pelos achados clínicos, citogenéticos e hematológicos do sangue periférico e medula óssea. A fase acelerada surge após um período variável do diagnóstico, de poucos meses a vários anos, e caracteriza-se pelo aumento de blastos na medula óssea e no sangue periférico, leucocitose e basofilia no sangue periférico, anemia e trombocitopenia. Clinicamente, o paciente torna-se refratário ao tratamento empregado na fase crônica e pode apresentar progressão da hepato-esplenomegalia. Posteriormente, a doença evolui para a fase blástica, definida hematologicamente pelo aumento de blastos leucêmicos (linfóides ou mielóides) no sangue periférico e/ ou medula óssea (mais de 20%). Nesse estágio da doença, muitos pacientes evoluem para o óbito entre três e seis meses. A progressão para a fase acelerada e blástica parece estar associada, principalmente, à instabilidade genômica, o que predispõe ao aparecimento de outras anormalidades moleculares (BERGANTINI et al., 2005).

2.2 Análise de dados categóricos

Segundo Pino (2001) o uso de tabelas de contingência para resumir dados de contagens se deu a partir de meados do século XIX, quando investigadores como o estatístico Belga Quetelet resumiam a informação de duas variáveis em tabelas de contigência 2×2. Pearson em 1900 formulou o teste de independência qui-quadrado para tabelas de duas vias, que constitui o primeiro resultado verdadeiramente importante neste campo e esse resultado juntamente com o coeficiente de correlação de Pearson foram a base para estudos posteriores.

Pearson supõem a existência de distribuições bivariantes contínuas subjacentes as tabelas bidimensinais, ele cría que a associação entre duas variáveis qualitativas devia ser descrita aproximando a correlação para esta distribuição contínua, a correlação tetracórica para uma tabela 2×2 é um exemplo de medida de associação baseada nessa ideia.

Ao mesmo tempo, Yulle publicou vários trabalhos nos que considerava as categorias de uma variável qualitativa como fixas e cría que se podiam definir coeficientes significativos sem a necessidade de supor distribuições contínuas subjacentes para uma tabela de contigência. Definiu assim medidas de associação que eram funções da razão de produto cruzado, por exemplo a Q de Yulle para tabelas 2×2. O debate entre Yulle e Pearson foi longo e árduo. Yulle argumentava que as variáveis qualitativas eram inerentemente discretas e que a introdução de hipóteses desnecessárias e não verificáveis, como a normalidade assumida por Pearson, não lhe parecia desejável em um trabalho científico. Já Pearson defendia que as teorias de Yulle eram inaceitáveis e causariam um dano irreparável no desenrolar da estatística moderna e na formação dos jovens estatísticos (PINO, 2001).

2.2.1 Tipos de variáveis

A Análise de Dados Categorizados visa evidenciar e interpretar a informação relevante que está contida em dados discretos provenientes de contagens de eventos ou de unidades de investigação (pessoas, lugares, objetos) possuindo certas características ou atributos definidos pela combinação das categorias de duas ou mais variáveis de interesse, ou ainda, apenas categorias de uma variável. A classificação do tipo da variável depende de como foi medida a característica de interesse, determinando assim a maneira que será analisada estatisticamente. As variáveis podem ser classificadas da seguinte forma:

i) Quantitativas - são aquelas em que os possíveis valores são numéricos por exemplo,

altura, peso, idade, temperatura, etc. Essas variáveis podem ser classificadas como contínuas e discretas. As variáveis contínuas são aquelas que podem ter um conjunto de valores infinitos não contáveis e as variáveis discretas são aquelas que podem tomar um conjunto finito ou infinito numeráveis.

ii) Qualitativas que são aquelas em que os valores são um conjunto de qualidade que não são numéricas e que geralmente podem ser chamadas de categorias, ou modalidades ou níveis, por exemplo, sexo (homem, mulher), cor de cabelo (loiro, moreno, castanho, ruivo), etc.

Quanto aos níveis de mensuração (ou escala) as variáveis podem ser classificadas como:

i) Nominal - São variáveis para as quais os níveis não tem uma ordenação natural e a ordem da listagem das categorias é irrelevante para a análise estatística.

ii) Ordinal - São variáveis que têm níveis ordenados, por exemplo, classe social (baixa, média, alta), atitude para legalização do aborto (desaprova fortemente, desaprova, aprova, aprova fortemente). Variáveis ordinais têm suas categorias ordenadas, porém a distância absoluta entre as categorias são desconhecidas;

iii) Intervalar - São variáveis que têm distância numérica entre qualquer dois níveis da escala, por exemplo, pressão sanguínea e Razão - São observações numéricas que têm as características de escala intervalar porém, possui um ponto zero verdadeiro e uma unidade de medida absoluta, por exemplo, rendimentos.

2.3 Testes Categorizados

Os testes para dados categorizados tem o objetivo de determinar, seguindo algum critério válido de decisão se o fator onde se quer estudar algo exerce alguma influência em um outro fator estudado. As hipóteses são construídas com H_0 sendo a hipótese nula onde se observa se as categorias de um certo grupo A exerce a mesma influência sobre as categorias de um certo grupo B; e a hipótese H_1 sendo a hipótese alternativa onde se observa se pelo menos uma categoria de um certo grupo A apresenta diferenças em relação as categorias de um certo grupo B. Podemos observar isso mais claramente em tabelas do tipo 2×2 (onde representa duas classificações para cada variável)(ARANGO, 2011).

2.3.1 Teste Qui-quadrado

É um teste de hipóteses que se destina a encontrar um valor da dispersão para duas variáveis nominais e avaliar a associação existente entre variáveis qualitativas. É um teste não paramétrico, ou seja, não depende de parâmetros populacionais, como média e variância.

O princípio básico deste método é comparar proporções, isto é, as possíveis divergências entre as frequências observadas e esperadas para um certo evento. Pode-se 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, próximas a zero. Com isso o teste é utilizado para verificar se a frequência com que um determinado acontecimento observado em uma amostra se desvia significativamente ou não da frequência com que ele é esperado e para fazer a comparação da distribuição de diversos acontecimentos em diferentes amostras afim de avaliar se as proporções observadas destes eventos apresentam ou não diferenças significativas ou se as amostras diferem significativamente quanto às proporções desses acontecimentos.

Para aplicar o teste as seguintes suposições precisam ser satisfeitas: Os grupos devem ser independentes; Os itens de cada grupo devem ser selecionados aleatoriamente; As observações devem ser frequências ou contagens; Cada observação pertence a uma e somente uma categoria e A amostra deve ser relativamente grande (pelo menos 5 observações em cada célula e, no caso de poucos grupos, pelo menos 10. Exemplo: em tabelas 2×2).

Devido as suas várias qualidades os testes qui-quadrado podem ser utilizados em várias aplicações onde envolvem os dados categóricos e podemos utilizar esses testes para: testar a qualidade do ajuste de uma distribuição multinomial e também para determinar se os dados seguem a mesma distribuição que seguiam anteriormente essa distribuição é chamada de distribuição multinomial com um conjunto de porcentagens históricas, onde definem a percentagem dos itens que se enquadram em cada categoria da resposta e o teste qui-quadrado vai testar se qualquer percentual vai diferir significadamente do respectivo percentual histórico.

Usa-se esse teste para testar a porcentagem de defeituosos para mais de 2 grupos para determinar se existe diferença entre os percentuais de defeituosos dos grupos testados, os grupos diferem por uma característica de interesse como por exemplo um produto produzido por fábricas diferentes ou em momentos diferentes, e o teste qui-quadrado irá testar se algum percentual de defeituosos difere significativamente de qualquer outro

percentual de defeituosos.

É possível usar este teste para Testar a associação entre duas variáveis categóricas: para determinar se uma variável resposta categórica (Y) está relacionada ou associada a outra variável preditora categórica (X). O qui - quadrado testa conjuntamente se existe uma associação entre a variável resposta e uma variável preditora. É possível realizar um Teste qui-quadrado de associação com uma variável preditora (X) que contém dois ou mais valores diferentes (duas ou mais amostras).

Para realizar o teste, o pesquisador trabalha com duas hipóteses:

$$\begin{cases} H_0: & \text{As frequências observadas não são diferentes das frequências esperadas.} \\ H_1: & \text{As frequências observadas são diferentes das frequências esperadas.} \end{cases}$$

Se a hipótese nula (H_0) for verdadeira, não existe diferença entre as frequências (contagens) dos grupos. Portanto, não há associação entre os grupos. Por outro lado se a hipótese alternativa (H_1) for a verdadeira, existe diferença entre as frequências. Portanto, há associação entre os grupos. É necessário obter duas estatísticas denominadas calculado e tabelado. As frequências observadas são obtidas diretamente dos dados das amostras, enquanto que as frequências esperadas são calculadas a partir destas.

Assim, o valor calculado é obtido a partir dos dados experimentais, levando-se em consideração os valores observados e os esperados, tendo em vista a hipótese. Já o valor tabelado depende do número de graus de liberdade e do nível de significância adotado. A tomada de decisão é feita comparando-se os dois valores de: Se o valor calculado for maior ou igual ao valor tabelado: Rejeita-se Ho. Se o valor calculado for menor que o valor tabelado: Aceita-se Ho. Quando se consulta a tabela de contigência observa-se que é determinada uma probabilidade de ocorrência daquele acontecimento. Portanto, rejeita-se uma hipótese quando a máxima probabilidade de erro ao rejeitar aquela hipótese for baixa (alfa baixo). Por outro lado, quando a probabilidade dos desvios terem ocorrido pelo simples acaso é baixa. O nível de significância (α) representa a máxima probabilidade de erro que se tem ao rejeitar uma hipótese. O número de graus de liberdade, nesse caso é assim calculado: G.L. = número de classes - 1. E, e videntemente, quanto maior for o valor mais significativa é a relação entre a variável dependente e a variável independente.

Para construir a matriz de valores esperados, E, de dimensões $r \times s$ ($linha \times coluna$).

Os valores da matriz E são calculados da seguinte forma:

$$E_{ij} = \frac{\sum_{j=1}^{s} O_{ij} \sum_{i=1}^{r} O_{ij}}{\sum_{i=1}^{r} \sum_{j=1}^{s} O_{ij}} = \frac{A_i.B_j}{T}$$

Posteriormente, com os valores da matriz O e da matriz E calcula-se a estatística:

$$x^2{}_c = \sum_{i=1}^{r} \sum_{j=1}^{s} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

Essa expressão corresponde ao teste qui-quadrado clássico, sua utilização, contudo, não pode ser indiscriminada. O teste qui-quadrado clássico pode ser usado quando o número total de dados é maior que 40. Se o número de dados estiver entre 20 e 40, e todos os valores esperados forem maiores que 5, o teste qui-quadrado deve ser aplicado, usando a expressão

$$x_c^2 = \sum_{i=1}^{r} \sum_{j=1}^{s} \frac{|O_{ij} - E_{ij}| - (0,5)^2}{E_{ij}}$$

denominada correção de Yates.

2.3.2 Teste exato de Fisher

O teste exato de Fisher é o mais adequado para se utilizar em casos de amostras pequenas e é representado em tabela de contigência $2x2$, e com o seu total geral(N) muito baixo. São nesses casos onde não se aplica o teste qui-quadrado apesar de ser um teste bastante utilizado em caso de duas amostras independentes, e então aplica-se o teste exato de Fisher que consiste em determinar a exata probabilidade de ocorrência de uma frequência observada e de acordo com suas pressuposições podemos destacar que: as suas amostras são casuais e independentes; e as duas classes são mutuamente exclusivas (CAMPOS,1983).

O teste exato de Fisher também pode ser usado para determinar se duas proporções da população são iguais. Para esta aplicação, a hipótese nula afirma que as duas proporções da população são iguais (H0: $p_1 = p_2$); a hipótese alternativa pode ter uma lateral esquerda ($p_1 < p_2$), direita ($p_1 > p_2$) ou ser bilateral ($p_1 \neq p_2$). O teste exato de Fisher é útil como teste de 2 proporções por ser preciso para todos os tamanhos amostrais, enquanto um teste de 2 proporções baseado em uma aproximação normal pode ser impreciso quando

o número de eventos é menor que 5 e quando o número de ensaios menos o número de eventos é menor que 5. O teste exato de Fisher é baseado na distribuição hipergeométrica. Portanto, o valor p depende dos totais marginais da tabela.

Segundo Arango(2011) para a aplicação do teste exato de Fisher serão construídas duas matrizes a partir da tabela de contigência original que contém os valores observados $(O_{11}, O_{12}, ...)$, essas matrizes serão denominadas por X1 e X2 e serão elaboradas tomando o total de uma das linhas que pode ser escolhida por uma das duas linhas dos totais dos valores do fator onde se quer estudar algo considerando que os elementos de uma dessas linhas estão na condição do fator onde se estuda algo; e o segundo procedimento é que a partir da matriz original O e das matrizes extremas X1 e X2 é calculada uma estatística para cada matriz a partir da expressão:

$$F = \frac{A1!A2!B1!B2!}{|\prod_{i=1}^{r}\prod_{j=1}^{s} Oij!|T!}$$

Onde:

A são os valores das linhas e B são os valores das colunas.

Para obtermos as hipóteses vamos considerar uma das classes e admitir que:

i) P(A) é a probabilidade de um elemento pertencer à população representada pela amostra A;

ii) P(B) é a probabilidade de um elemento pertencer à população representada pela amostra B;

Com isso têm-se $H_0 : P(A) = P(B)$, $H_1 : P(A) > P(B)$, $P(A) < P(B)$, $P(A)$ diferente de $P(B)$

2.3.3 Teste de Mantel-Haenszel

O teste de Mantel-Haenszel se aplica a dados categorizados em situações parecidas com a do teste qui quadrado onde se estuda a associação entre duas variáveis com a presença de uma terceira variável associada a um fator onde se quer estudar algo exercendo influencia também sobre um fator estudado.(ARANGO, 2011).

Calcula-se de modo geral a estatística do teste de Mantel-Haensel da seguinte forma:

$$X^2MH = \frac{(|SO - SE| - 0,5)^2}{SV}$$

onde, SO a soma dos valores observados que relacionam positivamente o fator onde se quer estudar algo e o fator onde estuda algo; SE uma soma parecida para os valores esperados; SV a soma das variâncias para todas as possíveis condições da variável calculado pela fórmula:

$$V = \frac{A1 \times A2 \times B1 \times B2}{T^2 \times (T - 1)}$$

O teste de Mantel-Haenszel é adequado para situações em que queremos verificar a associação entre duas variáveis binárias controlando pelas demais, este teste é chamado de independência condicional e não é apropiado quando a associação varia muito entre as tabelas parciais, suas variáveis a serem controladas tem que ser categóricas ou categorizadas. O teste fica muito limitado na presença de muitas tabelas, com pequeno tamanho amostral.

2.3.4 Teste G^2

O Teste G^2 é um teste estatístico equivalente com o teste qui-quadrado onde a diferença dos dois converge em probabilidade para zero, em que seus grandes valores estatísticos proporcionam mais evidências contra H_0. Seus resultados obtidos para o limite de amostragem multinomial também são válidos para outros tipos de amostragens (PINO, 2001). Esse teste tem a desvantagem de usar logaritmos em seu cálculo mas pode ser decomposto para aumentar o poder do teste de contraste de independência condicional em múltiplas tabelas. Esta decomposição da estatística G^2 é especialmente útil na seleção de modelos log-lineares.

O valor da estatística G^2 depende dos totais marginais de linhas e colunas e não de ordem entre linhas e colunas. A invariância contra permutações de linhas e colunas leva a ignorar a informação adicional no caso de variáveis ordinais para o qual tem contrastes mais poderosos de independência com base em alternativas mais limitadas. Estes contrastes são válidos para tamanhos amostrais grandes. A eficiência da aproximação qui-quadrado depende do tamanho amostral e das frequências esperadas estimadas (PINO, 2001).

Para o caso de amostras pequenas serão consideradas duas soluções alternativas: corrigir o erro que se comete ao aproximar uma distribuição discreta por uma contínua,que será abordada a continuação, e construir contrastes de independência baseados em distri-

buições exatas em lugar de aproximadas para as frequências observadas. Para testar as hipóteses de independência usamos o teste da razão de verossimilhança:

$$Ho : p_{ij} = p_{i}.p_{j}.$$

Em uma tabela de contigência $I \times J$ gerada por uma amostra multinomial, se expressa em termos de frequências esperadas como:

$$Ho : m_{ij} = m_{i}.m_{.}j/n$$

Recordamos que se X_1,...,X_n,são uma amostra de tamanho n com uma v.a com família paramétrica de distribuições de probabilidade com parâmetro $\theta \ \epsilon \ \Theta \subseteq R^I$, a estatística de teste de razão de probabilidade para testar é:

$$Ho : \theta \epsilon \Theta_0$$

$$H1 : \theta \epsilon \Theta_0 = \Theta - \Theta_0$$

se define como:

$$\lambda(X_1, ..., X_n) = \frac{sup_{\theta \epsilon \Theta_0} L(X_1, ..., X_n; \theta)}{sup_{\theta \epsilon \Theta_0} L(X_1, ..., X_n; \theta)}$$

sendo L a função de verossimilhança dos dados observados. É claro que $0 \leq [\lambda(X_1, ..., X_n)] \leq 1$, e o teste consiste em rejeitar Ho se $\lambda(X_1, ..., X_n) \leq$ c determinando c a partir da restrição de tamanho

Certas condições de regularidade para a família paramétrica de distribuições, da v.a

$$-2log(\lambda(X_1, ..., X_n))$$

se distribui assintoticamente como uma v.a X^2 com graus de liberdade df=dimΘ - dimΘ_o A estatística de Wilks de razão de verossimilhanças para o teste de independência em uma tabela de contigência $I \times J$ gerada por uma amostragem multinomial completa é da forma

$$G^2 = -2log\Lambda = 2\Sigma_i \Sigma_j n_{ij} log \frac{n_{ij}}{m_{ij}}$$

sendo $mij = n_{i}.n_{.j}/n$ a estimação de máxima verossimilhança das frequências esperadas das hipóteses de independência. G^2 se chama a estatística qui-quadrado da razão de verossimilhanças e se distribui assintoticamente como uma v.a X^2 com (I-1)(J-1) graus de liberdade. Efetivamente, fazendo uso dos estimadores de máxima verossimilhança para

amostra multinomial se tem que a estatística de razão de verossimilhança é:

$$\Lambda = \frac{\prod_{i=1}^{I}\prod_{j=1}^{J}(n_i.n_{.j})}{n^n \prod_{i=1}^{I}\prod_{j=1}^{J} n_i j^{n_i j}}$$

Temos em conta que para obter os graus de liberdade onde o espaço paramétrico é p_{ij} sob a restrição $\Sigma_i\Sigma_j p_{ij} = 1$, de modo que sua dimensão é $(IJ-1)$. Sob Ho os parâmetros estão determinados pelas probabilidades marginais $p_i.$ e $p_{.j}$ sob as restrições $\Sigma_i p_{i.}$=$\Sigma_j p_{.j}$=1, portanto a dimensão do espaço paramétrico se reduz a $(I-1)$+$(J-1)$, então como demonstrado por Wilks o número de graus de liberdade de G^2 é a diferença em dimensões dada por $(I-1)(J-1)$.

2.3.5 Razão de Chances (odds ratio)

Segundo Agresti (2002) muitos estudos são projetados para comparar os grupos em uma variável resposta binária. Então Y tem apenas duas categorias, tais como o sucesso ou fracasso para resultado de um tratamento médico. Com dois grupos, uma tabela de contingência 2×2 exibe os resultados. As linhas são os grupos e as colunas são as categorias de Y. Esta seção apresenta parâmetros para comparação entre os grupos.

Para assuntos na linha i , $\Pi_{1/i}$ é a probabilidade de que a resposta tem resultado em categoria 1 ("sucesso ") . Com apenas dois resultados possíveis,$\Pi_{2/i} = 1 - \Pi_{1/i}$ e usamos a notação mais simples de Π_i para $\Pi_{1/i}$. A diferença de proporções de sucessos, é uma comparação de base das duas linhas . Comparação em falhas é equivalente a comparação sobre sucessos, desde:

$$(1 - \Pi_1) - (1 - \Pi_2) = \Pi_2 - \Pi_1$$

A diferença de proporções cai entre -1.0 e +1.0. Ele é igual a zero quando as linhas têm distribuições condicionais idênticas. A resposta é estatisticamente independente da classificação linha quando $\Pi_1 - \Pi_2 = 0$ Quando ambas as variáveis são respostas, as distribuições condicionais são aplicadas em qualquer direção. Pode-se também comparar as duas colunas, tais como pela diferença entre as proporções em linha 1. Isto geralmente não é igual a diferença $\Pi_1 - \Pi_2$ comparando as linhas. Um valor $\Pi_1 - \Pi_2$ de tamanho fixo pode ter uma maior importância quando ambos os Π estão perto de 0 ou 1. Para um estudo comparando dois tratamentos sobre a proporção de indivíduos que morrem, a diferença entre 0,010 e 0,001 pode ser mais digno de notar que a diferença entre 0,410 e 0,401, embora ambos são 0.009. Em tais casos, a razão entre as proporções Também é

informativo (AGRESTI, 2002). O risco relativo é definido como sendo a razão:

$$\Pi_1/\Pi_2$$

Pode ser qualquer número real não negativo. Um risco relativo de 1,0 para corresponde independência. Para as proporções dadas apenas, os riscos relativos são 0.010/0.001=10.0 e 0.410/0.401=1.02. Comparando-se as linhas na segunda categoria de resposta dá um risco relativo diferente, $(1 - \Pi_1)/(1 - \Pi_2)$
Para uma probabilidade Π de sucesso, as probabilidades estão a ser definidas como:

$$\Omega = \Pi/(1 - \Pi)$$

As chances são não negativos, com Ω=1,0 quando um sucesso é mais provável do que um falha, quando Π=0,75, por exemplo, em seguida, quando Ω=0,75/0,25=3,0; um sucesso é três vezes mais provável que um fracasso, e esperamos cerca de três sucessos para 1 cada um de falha. Quando Ω=1/3, uma falha é três vezes mais provável que um sucesso. Inversamente,

$$\Pi = \Omega/(\Omega + 1)$$

Por exemplo, quando $\Omega=\frac{1}{3}$, em seguida, Π é 0,25. Refira-se novamente a uma tabela 2x2. Dentro de linha i, as chances de sucesso, em vez de fracasso são $\Omega_i=\Pi_i/(1-\Pi_i)$. A razão das probabilidades Ω_1 e Ω_2 e nas duas linhas,

$$\theta = \frac{\Omega_1}{\Omega_2} = \frac{\Pi_1/(1 - \Pi_1)}{\Pi_2/(1 - \Pi_2)}$$

é chamada a razão de chances.

Para distribuições conjuntas com probabilidades de células Π_{ij}, a definição equivalente para as probabilidades em linha i é $\Omega_i=\Pi_{i1}/\Pi_{i2}$. Em seguida, a razão de chances é

$$\theta = \frac{\Pi_{11}/\Pi_{12}}{\Pi_{21}/\Pi_{22}} = \frac{\Pi_{11}\Pi_{22}}{\Pi_{12}\Pi_{21}}$$

Um nome alternativo para é a relação entre produtos, uma vez que é igual à razão dos produtos e das probabilidades de diagonalmente células oposta.

Segundo Agresti (2002) a razão de chances pode ser igual a qualquer número não negativo. A condição $\Omega_1=\Omega_2$ e quando todas as probabilidades celulares são positivas $\theta = 1$ corresponde para independência dos X e Y. Quando $1<0<\infty$, os indivíduos na linha 1 são mais propensos a ter um sucesso do que os que estão na linha 2; isso é, $\pi_1 > \pi_2$, por exemplo, quando θ=4, as chances de sucesso na linha 1 são quatro vezes mais chances do

que da linha 2. Os valores de mais de 1,0 em uma determinada direção representam mais forte Associação. Dois valores representam a mesma associação, mas em frente instruções, quando uma é o inverso do outro. Por exemplo, quando θ=0,25, as chances de sucesso na linha 1 são 0,25 vezes mais chances na linha 2, ou equivalentemente, as chances de sucesso na linha 2 são 1/0,25=4,0 vezes mais chances na linha 1.

A razão de chance não altera o valor quando as linhas tornam-se colunas e as colunas tornam-se linhas. É desnecessário identificar uma classificação como variável de resposta, a fim de utilizar θ, de fato, definimos em termos de probabilidades usando Com uma distribuição conjunta, existem distribuições condicionais em cada sentido, e

$$\theta = \frac{\pi_{11}\pi_{22}}{\pi_{12}\pi_{21}} = \frac{P(Y=1|X=1)/P(Y=2|X=1)}{P(Y=1|X=2)/P(Y=2|X=2)} = \frac{P(X=1|Y=1)/P(X=2|Y=1)}{P(X=1|Y=2)/P(X=2|Y=2)}$$

Na verdade, a razão de chances é igualmente válida para estudos prospectivos e retrospectivos, ou planos de amostragem transversais

2.4 Aplicação

2.5 Material e Métodos

Para o estudo utilizou-se informações do site do INCA[1], que traz as proporções de pessoas com diferentes tipos de câncer, para gerar as tabelas de pacientes que apresentaram os tipos de leucemia linfóide e mielóide utilizando os períodos de 1985 a 1991, 2005 à 2010, a partir das proporções disponíveis para estas duas doenças.

Para a primeira análise entre os anos (1985-1991), foi constituído um total de 200000 pacientes os quais 100000 são do sexo masculino e 810 apresentam a leucemia linfóide, e 100000 do sexo feminino onde 570 apresentam a leucemia linfóide. Para a segunda análise entre os anos (2005-2010) oi constituído por um total de 200000 pacientes os quais 100000 são do sexo masculino e 1030 apresentam a leucemia linfóide, e 100000 do sexo feminino onde 750 apresentam a leucemia linfóide. Para a terceira análise entre os anos(1985-1991) foi constituído um total de 200000 pacientes os quais 100000 são do sexo masculino e 1130 apresentaram a leucemia mileóide, e 100000 do sexo feminino onde 950 apresentaram a leucemia mileóide. E para a quarta análise tivemos 200000 os quais 100000 são do sexo masculino e 1270 apresentaram a leucemia mileóide, e 100000 do sexo feminino onde 1130

[1]https://mortalidade.inca.gov.br/MortalidadeWeb/pages/Modelo05/consultar.xhtml#panelResultado

apresentaram a leucemia mileóide.

Os métodos aplicados foram os testes qui-quadrado, o teste G^2, o teste Mantel-Haensel, e o teste da razão de chances. A aplicação dos métodos e das técnicas da análise de dados categorizados foi possível com o auxílio do programa estatístico R versão 3.0.3 (R CORE TEAM, 2016).

3 Resultados e Discussão

Na Tabela 1 tem-se os o número de casos de LL para 100000 mil pessoas por sexo, percebe-se inicialmente que a proporção de casos em homens é maior que em mulheres.

Tabela 1: Número de casos de Leucemia Linfóide (LL) no período de 1985 à 1991, por sexo.

Status	Masculino	Feminino
Presença (LL)	810,00	570,00
Ausência (LL)	99190,00	99430,00

Se observarmos a proporção entre observados e esperados, na Tabela 2, tudo nos leva a crer que há associação entre sexo e LL, pois a proporção de doentes se alterou com o sexo, ou seja, tanto masculino e/ou feminino, a proporção de doentes se altera, esta é uma conclusão subjetiva, na qual estamos afirmando algo sem fornecer uma probabilidade de certeza. Da Tabela 2 utilizaremos a estatística de teste qui-quadrado para avaliarmos a associação entre variáveis qualitativas provenientes de uma tabela de contingência. A estatística de teste de $\chi^2 = 42,02$ com um valor $P = 8,99 \times 10^{-11}$, nos da evidências que existe associação entre sexo e leucemia linfoblástica.

Tabela 2: Frequência observada e esperada do número de casos de Leucemia Linfóide (LL) no período de 1985 à 1991, por sexo.

	Status	Sexo	Freq. obs.	Freq. esp.	OmE	OmE^2	OmE^2dE
1	Presença (LL)	Masculino	810,00	690,00	120,00	14400,00	20,87
2	Ausência (LL)	Masculino	99190,00	99310,00	-120,00	14400,00	0,15
3	Presença (LL)	Feminino	570,00	690,00	-120,00	14400,00	20,87
4	Ausência (LL)	Feminino	99430,00	99310,00	120,00	14400,00	0,15

OmE- Observado menos Esperado

OmE^2- Observado menos Esperado ao quadrado

OmE^2dE- Observado menos Esperado ao quadrado dividido pelo Esperado

O teste da razão de verossimilhança G^2 de independência sem correção, foi igual a $G^2 = 42,24$, está estatística tem distribuição χ^2 com um grau de liberdade e o valor

P foi $8,06 \times 10^{-11}$. Segundo McDonald(2014 o teste qui-quadrado e o teste de G^2 são semelhantes em suas conclusões, a escolha entre um teste e outro depende um pouco do analista, o teste G^2 é mais elaborado em sua construção estatística, enquanto o teste de qui-quadrado é muito mais difundido na comunidade acadêmica, sendo sempre uma boa ideia na área aplicada utilizar um teste mais familiar quanto possível. McDonald (2014) não recomenda que se utilize ambos os testes, mas afirma que quando fizer ambos os testes que o pesquisador escolha o teste que tiver menor valor P, pois isso ajudaria a se evitar falsos positivos (presença de doença quando não existe).

Após se realizar o teste de qui-quadrado e o teste G^2 para os casos de LL no período de 1985 a 1991, procedeu-se os mesmos testes para o período de 2005 a 2010.

Tabela 3: Frequência observada e esperada do número de casos de Leucemia Linfóide (LL) no período de 2005 à 2010, por sexo.

	Status	Sexo	Freq. obs.	Freq. esp.	OmE	OmE2	OmE2dE
1	Presença (LL)	Masculino	1030,00	890,00	140,00	19600,00	22,02
2	Ausência (LL)	Masculino	98970,00	99110,00	-140,00	19600,00	0,20
3	Presença (LL)	Feminino	750,00	890,00	-140,00	19600,00	22,02
4	Ausência (LL)	Feminino	99250,00	99110,00	140,00	19600,00	0,20

Se observarmos novamente a proporção entre observados e esperados, na Tabela 3 percebemos uma alteração na proporção entre os sexos e a LL, havendo uma sinalização de associação entre sexo e LL, pois a proporção de doentes no sexo masculino é maior que no sexo feminino.

O valor da estatística de qui-quadrado foi de 44,44 com um valor P de $2,62 \times 10^{-11}$, que indica a existência de associação entre o sexo e a doença.

Na Tabela 4 nos anos de 1985 a 1991 procedeu-se a construção da tabela de dupla entrada para sexo e Leucemia Mielóide.

Tabela 4: Número de casos de Leucemia Mielóide (LM) no período de 1985 à 1991, por sexo.

Status	Masculino	Feminino
Presença (LM)	1130,00	950,00
Ausência (LM)	98870,00	99050,00

Como aconteceu com a Tabela 1 as proporções entre os sexos foram diferentes, mas ao se comparar os valores das Tabelas 1 com 4, há indícios que ter sido mais pronunciada esta diferença na LM em relação a LL. Verificou-se na Tabela 5 os valores observados e esperados para construção do teste qui-quadrado, afim de verificar se existe associação entre sexo e LM.

Tabela 5: Frequência observada e esperada do número de casos de Leucemia Mielóide (LM) no período de 1985 à 1991, por sexo.

	Status	Sexo	Freq. obs.	Freq. esp.	OmE	OmE2	OmE2dE
1	Presença (LM)	Masculino	1130,00	1040,00	90,00	8100,00	7,79
2	Ausência (LM)	Masculino	98870,00	98960,00	-90,00	8100,00	0,08
3	Presença (LM)	Feminino	950,00	1040,00	-90,00	8100,00	7,79
4	Ausência (LM)	Feminino	99050,00	98960,00	90,00	8100,00	0,08

O valor da estatística de qui-quadrado foi de $15,74$ e o valor $P = 7,26 \times 10^{-5}$. O teste da razão de verossimilhança de independência sem correção G^2 foi igual a $G^2 = 15,76$, esta estatística tem distribuição χ^2 com um grau de liberdade e o valor P foi $7,19 \times 10^{-5}$, ambos os testes indicaram que existe associação entre sexo e LM. Procedeu-se também os mesmos testes para LM no período de 2005 a 2010, Tabela 6.

Tabela 6: Frequência observada e esperada do número de casos de Leucemia Mielóide (ML) no período de 2005 à 2010, por sexo.

	Status	Sexo	Freq. obs.	Freq.esp.	OmE	OmE2	OmE2dE
1	Presença (LM)	Masculino	1270,00	1193,97	76,03	5780,58	4,84
2	Ausência (LM)	Masculino	98730,00	97806,03	-76,03	5780,58	0,06
3	Presença (LM)	Feminino	1130,00	1206,03	-76,03	5780,58	4,79
4	Ausência (LM)	Feminino	98870,00	98793,97	76,03	5780,58	0,06

O valor da estatística de qui-quadrado foi de 9,75 com um valor P de 0,0017, que indica a existência de associação entre sexo e LM.

Foi aplicado também o teste de Mantel-Haenszeal para a leucemia linfoide com o objetivo de se comparar se havia diferença entre as janelas temporais estudadas e o valor P calculado foi de $2,2 \times 10^{-16}$ com uma razão de chances de 1,39, o que indica um acréscimo de quase 1,4 vezes na chance de se ficar doente ao longo dos anos (1985 - 1991 à 2005 à 2010). Para Mieloide nas mesmas condições o valor P foi de $4,67 \times 10^{-09}$ com uma razão de chance de 1,15, que indica que também há um acréscimo ao longo dos anos, mas este acréscimo é menor se comparado a LL em iguais períodos.

Deste modo, realizou-se o teste de Mantel-Haenszeal para comparar se havia diferença entre as leucemias mieloide e linfoide na janela temporal de 1985 a 1991. O valor P encontrado foi de $8,33 \times 10^{-13}$ para o teste de Mantel-Haenszeal e a razão de chances foi de 1,27, indicando que a chance de ficar doente é de quase 1,3 vezes maior na LM em realação a LL.

4 Considerações Finais

O objetivo desse estudo foi utilizar os principais testes estatísticos categorizados a dados de pacientes com leucemias linfóide e mielóide.

Inicialmente foi observado nas tabelas os pacientes que apresentaram os tipos de leucemia. Foram feitos os testes qui-quadrado, Mantel-Haenzel, G^2 e a razão de chances para os dois tipos de leucemia observados. E evidenciou-se que tanto para a leucemia linfóide quanto para a leucemia mileóide os testes foram significativos e concluímos que a variável sexo está relacionada com a variável doença e que a leucemia mielóide tem maior incidência que a linfóide .

A partir dos métodos e das técnicas usadas neste trabalho, evidenciou-se que a análise de dados categorizados é uma importante ferramenta na área da saúde desde que todos os critérios de cada técnica estatística possam ser seguidos corretamente, ajudando a entender o comportamento e as características dos pacientes sobre risco.

5 Referências

AGRESTI, A. **Categorial Data Analysis**. New York: Jonh Wiley & Sons, inc., 2002.

ARANGO, H.G **Bioeststística: teórica e computacional: com banco de dados reais em risco** 3.ed.-[Reimpr].- Rio de Janeiro: Guanabara Koogan, 2011.

BERGANTINI, ANA PAULA. F. et al. Leucemia mielóide crônica e o sistema Fas-FasL. **Rev. bras. hematol. hemoter**, v.27, n.2, p. 120-125 , 2005.

BITTENCOURT, ROSANE et al. Leucemia Mielóide Agura: perfil de duas décadas do Serviço de Hematologia do Hospital das Clínicas de Porto Alegre - RS. **Rev. Bras. Hematol. Hemoter**, v.25, n.1, São José do Rio Preto Jan./Mar, 2003.

CAMPOS, H. **Estatística Experimental Não-Paramétrica**. FEALQ.1983 4.ed. 1983. 349p.

PEDROSA, F.; LINS, M. Leucemia linfóide aguda: uma doença curável. **Rev. bras. saúde matern. infant**, v.2, n.1, p. 63-68, 2002.

PINO, A.N.A. **Tablas de Contigencia Bidimensionales**. Ed. La Muralla, S.A. 2001. 213p.

McDONALD, J.H. **Handbook of Biological Statistics** (3rd ed.). Sparky House Publishing, Baltimore, Maryland, p. 53-58, 2014.

RAMIREZ, DR.P.H Leucemia linfoide crónica. Aspectos clínicos y biológicos. **Rev Cubana Hematol Inmunol Hemoter**, v.15, n.1, Ciudad de la Habana ene.-abr, 1999.