

Idalice Carvalho Figueiredo Rillo

Tabelas 2x2 Uma Revisão



1910000752



TABELAS 2×2.UMA REVISÃO

Idalice Carvalho Figueiredo Rillo

**Dissertação de Mestrado
Pós-Graduação em Matemática Aplicada e
Computacional
MAC-016**



TABELAS 2 X 2. UMA REVISÃO

Dissertação apresentada ao Instituto de Biociências, Letras e Ciências Exatas da Universidade Estadual Paulista "Júlio de Mesquita Filho", Câmpus de São José do Rio Preto, para a obtenção do título de Mestre em Ciências Matemáticas (Área de Concentração: Matemática Aplicada e Computacional).

Idalice Carvalho Figueiredo Rillo

sob orientação do Prof. Dr. José Antônio Cordeiro

São José do Rio Preto - S.P.



752/98

[1997]

Defesa em 20/01/1997

15



AGRADECIMENTOS:

A José Antônio Cordeiro pela orientação segura e pela confiança em mim depositada.

A Luiz Carlos Balda pelo valioso auxílio.

A amiga Aparecida Conceição Botoni Bueno da Silva pela revisão gramatical deste trabalho.

A Stela e Fernando pelo convívio e momentos de descontração.

A meus irmãos.

A minha mãe por possibilitar, em tempos difíceis, que eu fosse uma estudante.

***"Ao meu marido Paulo e às minhas filhas Luciana, Graziela e Paula,
que dão luz à minha existência."***

A Deus.

INTRODUÇÃO: AGRADecIMENTOS: 1

A José Antônio Cordeiro pela orientação segura e pela confiança em mim depositada; 2

A Luiz Carlos Baida pelo valioso auxílio; 3

À amiga Aparecida Conceição Botoni Bueno da Silva pela revisão gramatical deste trabalho; 4

À Stela e Fernando pelo convívio e momentos de descontração; 5

A meus irmãos; 6

À minha mãe por possibilitar, em tempos difíceis, que eu fosse uma estudante; 7

À minha família pela compreensão e ajuda em meus momentos de atropelo; 8

A Deus. 9

ÍNDICE

INTRODUÇÃO.....	I
1. Associação em tabelas 2 x 2	1
1.1 Métodos de Amostragens	3
1.2 Variáveis Independentes	5
1.3 Medidas de Associação	6
1.3.1 Diferença de proporções.....	7
1.3.2 Risco Relativo	8
1.3.3 “Odds”	10
1.3.4 “Odds ratio”	11
1.3.5 Coeficiente Q de Yule	15
1.4 Estimadores e Intervalos de Confiança para: “odds”, “odds ratio”, diferença de proporções e risco relativo	17
2. Inferências em tabelas 2 x 2	20
2.1 Distribuições amostrais	20
2.1.1 Modelo Poisson	20
2.1.2 Modelo Multinomial	21
2.1.3 Produto de Binomiais	21
2.1.4 Modelo Hipergeométrico	22
2.2 Testes de Independência para grandes amostras	24
2.2.1 Estatística Quiquadrado da Razão de Verossimilhança	24
2.2.2 Teste Quiquadrado de Pearson	26
2.2.3 Ambas Marginais fixadas	29
2.2.4 Uma Marginal fixada.....	29
2.3 Dificuldades relacionadas a pequenas amostras	31
Conclusão.....	84
Bibliografia.....	86

3. Teste Exato para pequenas amostras	33
3.1 Estatística Ancilar	33
3.2 Estatística Ancilar na presença de parâmetros incômodos.....	35
3.3 Estatística Aproximadamente Ancilar	35
3.4 Princípio de Condicionalidade	36
3.5 Teste Exato de Fisher	40
3.5.1 Formas para cálculo do <i>valor p</i>	44
3.5.2 Teste de Fisher com adaptação de Tocher	46
3.5.3 A função poder condicional do Teste de Fisher	47
3.6 Considerações contra o Teste de Fisher	49
3.7 Estatística Fracamente Ancilar	50
3.8 A Informação de Fisher em uma distribuição	52
3.9 Análise Condicional para a “odds ratio”	57
3.9.1 Estimador	57
3.9.2 Intervalo de Confiança	58
3.9.3 Aproximações	60
3.9.4 Aproximação para o estimador	61
3.9.5 Aproximação dos limites do intervalo de confiança	61
3.9.6 Teste de Hipótese	62
3.10 Enfoque Bayesiano	62
3.10.1 Teste de Bayes	66
4. Amostras Correlacionadas	69
4.1 Teste de McNemar	70
4.2 Teste de Gart para efeitos da ordem	72
4.3 Combinando tabelas 2 x 2	76
4.3.1 Método $\sqrt{\chi^2}$	77
4.3.2 Método de Cochran	79
4.3.3 Teste de Mantel-Hanszel.....	81
Conclusão.....	84
Bibliografia.....	86

INTRODUÇÃO

Muitos experimentos são caracterizados por tabelas de contingência 2×2 e, devido à aparente simplicidade delas, muitos pesquisadores consideram-se capacitados para analisá-las e o fazem, freqüentemente, sem qualquer consulta a um estatístico.

O fácil acesso aos programas de computadores favorece estes abusos, que têm se repetido sistematicamente. Geralmente, o teste *quiquadrado de Pearson* é o mais utilizado para verificar a igualdade das proporções entre as variáveis estudadas. Segundo Engeman e Swanson (1991), esta escolha se dá, principalmente, por ter sido este teste o primeiro a contar com software estatístico para a elaboração de seus cálculos.

Outro teste também bastante utilizado é o teste *exato de Fisher* que, desde sua criação, tem sido atacado e defendido por diversos estatísticos de renome tais como Yates, Barnard, Pearson e outros. Os bayesianos também têm dedicado severas críticas a este teste.

Engeman e Swanson (1991) apresentam dois exemplos de análises efetuadas na área biomédica que deveriam ter sido consideradas de formas diferentes. Uma, na qual o teste utilizado foi bicaudal quando o melhor procedimento, de acordo com os objetivos da situação em estudo, era, claramente, o de um teste unicaudal. O outro nos mostra a importância da presença de um profissional na escolha do teste a ser utilizado. Neste trabalho, os autores analisaram os dados de um artigo de Davis e Laurenson (1981), cujo objetivo era verificar se a proporção de mortalidade, em pacientes idosos que foram submetidos a cirurgia abaixo do quadril, era menor entre os que receberam anestesia na espinha do que entre os que receberam anestesia geral.

Davis e Laurenson utilizaram o quiquadrado de Pearson sem correção de continuidade e indicam o uso de anestesia na espinha como uma opção mais adequada, quando os pacientes forem idosos.

Entretanto, McLeskey (1985), ao analisar posteriormente os mesmos dados, considerou que a diferença de risco de mortalidade entre os dois métodos de anestesia não era estatisticamente significativa.

A resposta do resultado mais acertado só pode ser emitida se forem considerados os dados e a pesquisa no todo. A escolha do teste mais apropriado envolve critérios relacionados com os dados, os parâmetros de interesse, os parâmetros incômodos ("nuisance"), a qualidade do teste e, principalmente, uma grande integração entre o profissional da área em estudo e o estatístico que vai trabalhar os dados.



Yates (1984) apresentou diversas considerações favoráveis ao teste exato de Fisher e, pelas discussões que estão anexadas ao artigo, constatava-se a aceitação de seus argumentos por uma grande parte de conceituados estatísticos. Entretanto, Vos e Hudson (1995) apresentaram um estudo se colocando contrários ao uso desse teste quando o objetivo for testar a igualdade de proporções em situações de duas distribuições binomiais independentes e indicando como forma alternativa um teste incondicional.

Estas disputas nortearam nosso trabalho visando atingir dois objetivos: o primeiro, efetuar uma revisão de tabelas 2×2 de modo que possamos melhor entender sua estrutura e, o segundo, verificar o embasamento teórico que justifica as afirmativas de Vos e Hudson (1995). Este segundo objetivo se tornou a nossa meta principal.

No capítulo 1, apresentamos métodos de amostragens e medidas de associação com seus estimadores incondicionais. Destaque especial foi dado à “odds ratio” devido a seu uso intenso na área biomédica.

No capítulo 2, consideramos testes de hipóteses em situações de grandes amostras ressaltando o teste quiquadrado de Pearson e as dificuldades relacionadas às pequenas amostras são discutidas.

No capítulo 3, apresentamos o teste exato de Fisher como forma optativa para situações de pequenas amostras. Por ser este o teste mais controvertido por nós enfocado, demos uma ênfase especial ao princípio de condicionalidade e às estatísticas ancilares por ele utilizadas.

A estatística fracamente ancilar também é, neste capítulo, destacada, sendo apresentado um método para identificá-la e as vantagens de seu uso em inferências condicionais, quando o interesse for evitar perdas no número de informação de Fisher em relação ao número de informação contidos nos dados.

Devido ao fato de que o teste de Fisher utiliza a função de verossimilhança condicional, novamente apresentamos estimadores para a “odds ratio” obtidos pela análise condicional. Algumas opções de análise na linha bayesiana se encontram no final deste capítulo.

No capítulo 4, enfocamos situações de amostras correlacionadas e sugestões de algumas adaptações de estatísticas para testes de hipóteses. Os testes foram ilustrados com exemplos, mas não apresentamos suas deduções. Este capítulo é apenas um início de estudos que deve ser ainda completado.



A literatura estatística nos mostra a disputa ocorrida pela aceitação de um teste para associação em tabelas 2×2 . Descreveremos a seguir alguns fatos históricos desta disputa.

Karl Pearson, em 1900, ao apresentar o teste quiquadrado (χ^2), erroneamente afirmou que os graus de liberdade de uma tabela $r \times s$ era $(rs-1)$.

Udny Yule em 1911 também trabalhou em tabelas 2×2 e introduziu um teste para associação. Após algumas simulações, chegou a um teste equivalente ao quiquadrado de Pearson, só que com 1 (*um*) grau de liberdade. Percebendo a discrepância entre seu teste e o quiquadrado de Pearson e, após analisar 350 tabelas 2×2 e 100 tabelas 4×4 , comprovou o erro de Pearson, mas não publicou imediatamente seus resultados.

Em 1922, R.A. Fisher publicou um artigo chamando a atenção para este erro e afirmou: "*Em testes para associação em tabelas de contingências, a esperança dos valores das celas são estimados pelas marginais totais e o número de graus de liberdade para tabelas $r \times s$ é $(r-1)(s-1)$* ". Logo após, Yule também publicou seus resultados confirmando as conclusões de Fisher.

Como Pearson demorou para aceitar essas conclusões, iniciou-se uma disputa principalmente em situações envolvendo tabelas 2×2 .

O teste χ^2 é assintótico e não se ajusta bem em situações nas quais os valores esperados para as celas de uma tabela de contingência são pequenos. Analisando situações com essas peculiaridades, Fisher (1925) estabeleceu uma regra considerando que o valor esperado de qualquer cela não poderia ser menor que 5 para que o teste quiquadrado se ajustasse bem.

Yates (1934), após algumas investigações, mostrou que as aproximações obtidas pela estatística X^2 poderiam ser melhoradas caso sofressem uma correção de continuidade.

Este resultado já era conhecido por Fisher (1926) e, em 1934, ele fez uma revisão em seu livro *Statistical Methods for Research Workers* incluindo uma seção para correção de continuidade e outra para o seu teste exato. Em 1935 ele expôs suas razões para acreditar que o teste exato podia sempre ser usado para verificar a igualdade de proporções e a associação em tabelas de contingências 2×2 , com as marginais fixadas a priori ou não.

Segundo Yates (1984) as oposições ao teste exato de Fisher apoiam-se principalmente, por o considerarem muito conservativo e, pelo fato de que Neyman e Pearson, ao desenvolverem as teorias para teste de significância, estabeleceram que "*o nível de significância precisa ser igual para a frequência com as quais a hipótese é rejeitada em amostras repetidas de qualquer população fixada, permitida pela hipótese*".



A construção do teste exato de Fisher foi uma consequência de estudos que o levou a concluir que condicionar em uma estatística ancilar era uma preciosa contribuição para a teoria dos testes de significância e isto gerou um dos grandes motivos de controvérsia, pois a exigência da frequência para repetidas amostras não é satisfeita quando condicionamos aos totais marginais fixados.

Outros pontos são geradores de discórdia em torno desse teste. Um deles consiste na prática de duplicar o *valor p* para teste bilateral visto que o mesmo é unilateral. O outro é a questão se os valores marginais contêm ou não informações da existência de associação entre as variáveis em estudo, ocasionando conseqüentemente, perda de informações quando utilizamos o princípio de condicionalidade com os totais marginais.

Em 1945, Barnard criou um teste que ele afirmava ser mais poderoso que o teste exato de Fisher. Ele trabalhou com tabelas 2×2 , onde as duas linhas eram amostras de tamanhos n_1 e n_2 , retiradas de duas grandes populações, com distribuição binomial e, ao invés de impor a condição de que as proporções eram iguais ($p = p_1 = p_2$), considerou todas as combinações das possíveis amostras que podiam ser levantadas, associou suas probabilidades e ordenou pelos valores de $p_1 - p_2$, criando um teste alternativo para verificar se p_1 diferia significativamente de p_2 .

Este teste era incondicional e foi rejeitado pelo próprio Barnard (1949) em favor do teste exato condicional de Fisher.

Atualmente, as disputas por um teste de associação para tabelas 2×2 têm usado como base o poder de cada teste, sendo que as comparações, na maior parte dos artigos revisados, têm sido entre o teste exato de Fisher e alguns testes incondicionais. Berger (1994) apresenta uma boa comparação entre o poder dos testes acima citados, em situações de duas distribuições binomiais independentes. A análise deste artigo exige um estudo detalhado de seis testes incondicionais que não foram aqui abordados e, são motivo de interesse futuro para próximos trabalhos.



1. ASSOCIAÇÃO EM TABELAS 2X2

Consideremos X e Y duas variáveis categóricas, cada uma com dois níveis, que serão a presença ou a ausência de determinado atributo denotados por:

1: presença do atributo

0: ausência do atributo

Se classificarmos os elementos através das duas variáveis haverá 4 possíveis combinações na classificação.

As respostas (X,Y) de um elemento escolhido aleatoriamente em uma população têm uma distribuição de probabilidade. Se dispusermos as frequências absolutas obtidas em uma amostra de tamanho n e a distribuição de probabilidade em tabelas cruzadas, obteremos as tabelas 2×2 , representadas abaixo

Tabela 1: Frequências absolutas da classificação de n indivíduos segundo as categorias X e Y

X	Y		total
	1	0	
1	n_{11}	n_{12}	$n_{1.}$
0	n_{21}	n_{22}	$n_{2.}$
total	$n_{.1}$	$n_{.2}$	n

Tabela 2: Distribuições conjunta, condicionais e marginais de duas categorias X e Y em uma população

X	Y		total
	1	0	
1	p_{11}	p_{12}	$p_{1.}$
	$(p_{1/1})$	$(p_{2/1})$	1
0	p_{21}	p_{22}	$p_{2.}$
	$(p_{1/2})$	$(p_{2/2})$	1
total	$p_{.1}$	$p_{.2}$	1

As notações associadas às tabelas acima são:

i: número da linha

j: número da coluna

$i=1,2$ $j=1,2$

n_{ij} : frequência da cela na linha i e coluna j

$$n_{.j} = \sum_i n_{ij}$$

$$n_{i.} = \sum_j n_{ij}$$

$n = \sum_i \sum_j n_{ij}$ é o tamanho amostral total,

$\{p_{ij}\}$ é a distribuição conjunta da amostra.

Em muitas tabelas de contingência, uma das variáveis é uma variável resposta e a outra, uma variável preditora; e um dos principais objetivos é comparar a distribuição condicional da variável resposta nos dois níveis da variável preditora. Neste trabalho, usaremos, nestas situações, a variável Y como resposta e, na tabela, ela será a variável coluna.

Assim convencionado, as distribuições condicionadas de interesse são as apresentadas na Tabela 2.

Se considerarmos as frequências apresentadas na Tabela 1, os estimadores de máxima verossimilhança de p_{ij} e $p_{j|i}$ são:

$$\hat{p}_{ij} = \frac{n_{ij}}{n} \tag{1.1}$$

$$\hat{p}_{j|i} = \frac{\hat{p}_{ij}}{\hat{p}_{i.}} = \frac{n_{ij}}{n_{i.}} \tag{1.2}$$

Na prática, temos essencialmente três métodos de amostragem que geram um conjunto de frequências, distribuídos em uma tabela 2×2 .



1.1 Métodos de Amostragem

1º método : Amostragem Aleatória Simples ou "Cross- Over"

Quando selecionamos, ao acaso, n elementos de uma grande população e classificamos cada elemento pela presença ou ausência de uma característica X e pela presença ou ausência de uma característica Y , denominamos amostragem aleatória simples.

Neste caso, somente o tamanho amostral total n foi fixado de acordo com algum critério.

O objeto de interesse é verificar se a presença ou ausência da característica X está associada à presença ou à ausência da característica Y , ou seja, verificar a independência entre elas.

2º método : Estudos Comparativos Prospectivos e Estudos Comparativos Retrospectivos

Nos estudos comparativos, escolhemos duas amostras, sendo que, em uma delas, uma característica em estudo está presente e, na outra, está ausente.

Os elementos destas amostras são classificados de acordo com a presença ou a ausência de uma outra característica .

Estes estudos podem ser *prospectivos* ou *retrospectivos*. Do ponto de vista da amostragem, os dois processos são amostragem estratificada; no entanto o modo de estudar a associação entre as variáveis difere nos dois casos.

Nos *estudos comparativos prospectivos*, selecionamos duas amostras: uma na subpopulação onde um fator de risco suspeito, variável preditora, está presente; e outra, na subpopulação onde ele está ausente; e, depois, acompanhamos os elementos amostrados por um período de tempo e classificamos cada um deles pela presença ou ausência da variável resposta em análise.

Este estudo também recebe o nome de *coorte*.

Um estudo desta natureza foi efetuado por Doll e Hill (1964) para verificar a associação entre câncer no pulmão e fumo. Médicos da Inglaterra foram utilizados para



formarem os grupos comparativos. Após terem respondido a um questionário, selecionaram-se duas amostras, classificadas em fumantes e não fumantes.

Durante 10 anos, houve um acompanhamento e, nos dois grupos, foram anotados os casos de morte relacionada ao câncer de pulmão.

Neste exemplo, o fator de risco é o fumo e a variável resposta é a morte relacionada ao câncer pulmonar.

Nos *estudos comparativos retrospectivos*, selecionamos primeiro as duas amostras pela ausência ou presença da variável resposta e, tomando por base a história dos elementos, eles são classificados de acordo com a presença ou a ausência de um fator suspeito.

Este estudo também é conhecido como *caso-controle*.

Para diferenciarmos estes dois estudos, consideremos o mesmo interesse do exemplo anterior, onde o objetivo era verificar a associação entre fumo e morte relacionada ao câncer pulmonar.

Poderíamos considerar, em um grande hospital, casos de óbitos associados ao câncer pulmonar como uma amostra e, como grupo controle, óbitos por outro tipo de câncer. Através dos registros, estes elementos seriam classificados em fumantes ou não fumantes.

Este estudo tem sido preferido por ser mais fácil a obtenção dos dados e mais barato do que o prospectivo.

Após a 2^a guerra mundial, a convenção de Helsinque estabeleceu normas éticas para realização de experimentos envolvendo seres humanos. Os estudos comparativos retrospectivos e prospectivos não têm, em sua aplicação, restrições relativas à estas normas.

O objeto de interesse neste método é verificar se as proporções nas duas populações amostradas (p_{11} e p_{21}) são iguais, ou seja, se há homogeneidade de proporções de casos positivos da resposta nas duas populações.

3^o método : Estudos Comparativos Aleatorizados

Semelhante ao 2^o método, duas amostras de tamanhos determinados serão comparadas, contudo estas duas amostras serão compostas, aleatoriamente, de uma mesma população escolhida segundo algum critério. Uma destas amostras será utilizada para o tratamento e a outra, para controle do tratamento.

Este método é a base dos Ensaaios Clínicos Aleatorizados.



Quando se quer verificar, entre dois tratamentos, qual é o mais eficaz, escolhemos uma população e retiramos uma amostra. Estes elementos serão alocados, aleatoriamente, em dois grupos: um deles será o grupo tratamento e receberá a terapia que está sendo testada; o segundo grupo receberá a outra terapia e será denominado grupo controle.

Durante o experimento, os dois grupos deverão receber os mesmos cuidados; após o seu término, serão utilizadas técnicas estatísticas para avaliar a diferença entre a eficácia das duas terapias.

Neste caso, o objeto de interesse é verificar se as proporções do evento de interesse nos dois grupos experimentais são as mesmas.

Por ser uma metodologia experimental, está sujeita às normas éticas regidas pela convenção de Helsinque.

1.2 Variáveis Independentes

Ao analisarmos uma tabela 2×2 , é de grande interesse verificar a existência ou não de associação entre as variáveis em estudo e, caso ela exista, estabelecermos medidas de associação.

Por exemplo, nas pesquisas clínicas, um problema peculiar é estudar a associação entre o desenvolvimento de uma doença e as formas de tratamento; nas pesquisas eleitorais pode-se estar interessado na associação entre tipo de discurso e mudança na intenção do voto; na área pedagógica, verificar a associação entre determinada metodologia e a evasão escolar.

Vamos então apresentar o conceito de associação e algumas medidas de associação.

Consideremos uma população classificada de acordo com a presença ou a ausência de dois atributos X e Y descritas na Tabela 1.

Se ambas forem variáveis respostas, podemos descrever a associação através da distribuição conjunta ou das distribuições condicionais.

As variáveis X e Y são independentes se as probabilidades conjuntas são iguais ao produto de suas marginais, ou seja, se

$$p_{ij} = p_{i.} \cdot p_{.j}, \forall i, j \quad i, j = 1, 2 \quad (1.3)$$

Se X e Y forem independentes, temos

$$p_{j|i} = \frac{p_{ij}}{p_{i.}} = \frac{p_{i.} \cdot p_{.j}}{p_{i.}} = p_{.j} \quad (1.4)$$



para $i = 1, 2$. Logo, a distribuição condicional de Y é igual a sua marginal. Portanto, duas variáveis são independentes quando a probabilidade condicional da coluna resposta j é a mesma em cada linha, para $j = 1, 2$.

Segundo Agresti (1990), se j representa o nível da variável resposta Y , $j = 1, 2$ e X uma variável explanatória, a condição

$$P_{j|1} = P_{j|2} \quad (1.5)$$

produz uma definição mais natural de independência que (1.3). De fato, se a probabilidade de ocorrer um nível da variável resposta Y condicionada a cada nível da variável explanatória for a mesma, então é natural que aceitemos a idéia de que a ocorrência de determinado nível da variável X não interfere na ocorrência da variável resposta Y , ou seja, que elas são independentes.

1.3 Medidas de Associação

Uma medida de associação entre variáveis aleatórias é um parâmetro da população que nós estimamos através de uma amostra.

Segundo Bishop, Fienberg e Holland (1988), a utilidade das medidas de associação e dos testes de independência tem sido algumas vezes confundida. Os testes para independência devem ser utilizados para determinar se há relação entre as variáveis, enquanto que o uso de medidas de associação possibilita o dimensionamento e o entendimento de uma relação em particular.

De acordo com Bishop, Fienberg e Holland (1988), todas as medidas de associação devem satisfazer à seguinte propriedade:

Se a tabela apresentar independência total, isto é,

$$p_{ij} = p_{i.} p_{.j} \quad , \quad \forall i, j, \quad (1.6)$$

então a medida deverá indicar ausência de associação.

No entanto, as medidas de associação diferem no modo como indicam associação perfeita e estágios intermediários de associação.

Para as tabelas 2×2 , como o grau de liberdade é 1, uma única medida é suficiente para descrever a associação.

De acordo com Renyi (1959), um coeficiente de associação razoável entre duas variáveis aleatórias X e Y deve satisfazer às seguintes condições:



Se $C(X,Y)$ é o coeficiente de associação:

- a) $C(X,Y)$ é definida para um par de variáveis aleatórias (X,Y) com probabilidade 1 de não serem constantes;
- b) $C(X,Y) = C(Y,X)$;
- c) $0 \leq C(X,Y) \leq 1$;
- d) $C(X,Y) = 0$ se, e somente se, X e Y são independentes;
- e) $C(X,Y) = 1$ se há uma forte dependência entre X e Y ;
- f) Se $(h(X),k(Y))$ são transformações 1 - 1 das variáveis envolvidas, então $C(h(X),k(Y)) = C(X,Y)$;
- g) $C(X,Y)$ é uma função estritamente monotônica de $|r|$, se a distribuição conjunta de X e Y for normal bivariada e r for o coeficiente de correlação entre X e Y .

Entretanto, existem coeficientes eficientes que não satisfazem estes critérios como por exemplo o R^2 que só satisfaria a condição f se fosse linear. Vejamos pois alguns outros critérios.

Segundo Kendall e Stuart (1979), ao estabelecermos uma medida de associação entre duas variáveis, devemos conhecer os limites de variação deste coeficiente e ele deve indicar a independência das variáveis no valor central ou no valor mínimo do intervalo de variação. Diferentemente de Renyi (1959), eles não estabelecem que os limites do intervalo de variação devam ser 0 e 1.

Segundo Bishop, Fienberg e Holland (1988), é usual normalizar as medidas de associação de maneira que elas fiquem linearmente entre 0 e 1 ou entre -1 e 1; e todas as medidas devem assumir o valor 0 quando a distribuição $\{p_{ij}\}$ satisfizer a hipótese de independência.

Ainda de acordo com os autores acima, estas medidas podem ser simétricas ou assimétricas, sendo que, se o interesse for a distribuição conjunta das duas variáveis, a medida deve ser simétrica; enquanto que, se o interesse for na distribuição condicional de uma das variáveis dada a outra, uma medida assimétrica pode refletir a situação mais adequadamente.

Vamos considerar algumas medidas de associação freqüentemente utilizadas.



1.3.1 Diferença de proporções

Podemos utilizar como medida de associação a diferença das proporções. Se somente a variável aleatória Y for uma variável resposta, nós podemos comparar as linhas utilizando a diferença $p_{11} - p_{12}$.

Quando ambas as variáveis são respostas e há uma distribuição conjunta $\{p_{ij}\}$ temos, ao compararmos as linhas,

$$P(\text{coluna1}|\text{linha1}) - P(\text{coluna1}|\text{linha2}) = p_{11} - p_{12} = \frac{p_{11}}{p_{1.}} - \frac{p_{21}}{p_{2.}} \quad (1.7)$$

onde as probabilidades condicionais são as convencionadas na Tabela 2.

A diferença de proporções é uma medida pertencente ao intervalo $[-1, 1]$ e, ela é igual a zero quando as duas linhas têm distribuições condicionais idênticas, ou seja, quando as variáveis são independentes.

Se quisermos utilizar a comparação entre colunas, teremos

$$P(\text{linha1}|\text{coluna1}) - P(\text{linha1}|\text{coluna2}) = \frac{p_{11}}{p_{.1}} - \frac{p_{12}}{p_{.2}} \quad (1.8)$$

e, neste caso as probabilidades condicionais serão obtidas por $p_{i|j} = \frac{p_{ij}}{p_{.j}}$ onde, i e j representam linhas e colunas, respectivamente.

1.3.2 Risco Relativo

Consideremos duas variáveis, sendo uma um fator de risco X e a outra, uma variável resposta Y .

Define-se como *risco relativo* da variável resposta Y a razão entre a proporção da incidência do nível 1 da variável resposta na *presença* do fator de risco e a proporção da incidência do nível 1 da variável resposta na *ausência* do fator de risco.

$$RR = \frac{p_{11}}{p_{12}} = \frac{P(Y=1|X=1)}{P(Y=1|X=0)} \quad (1.9)$$

O risco relativo é uma razão na qual comparamos as probabilidades de um nível da variável resposta dentro dos dois níveis da outra variável que, no caso, é a variável fator de risco que está sendo considerada na pesquisa.

Geralmente, o nível de interesse da variável resposta é o ruim : daí a denominação risco.

Ele nos dá a quantidade de vezes que a probabilidade p_{11} é maior ou menor que p_{12} e sua variação ocorre no intervalo real $[0, \infty[$.

As variáveis X e Y são independentes se, e somente se, o risco relativo for igual a 1.

Vamos demonstrar esta afirmativa.

a) Se as variáveis X e Y são independentes então:

$$RR = \frac{p_{11}}{p_{12}} = \frac{\frac{p_{11}}{p_{1.}}}{\frac{p_{21}}{p_{2.}}} = \frac{\frac{p_{1.} \cdot p_{.1}}{p_{1.}}}{\frac{p_{2.} \cdot p_{.1}}{p_{2.}}} = 1$$

b) Se o risco relativo for igual a 1 então :

$$RR = \frac{p_{11}}{p_{12}} = 1 \rightarrow \frac{\frac{p_{11}}{p_{1.}}}{\frac{p_{21}}{p_{2.}}} = 1 \rightarrow \frac{p_{11}}{p_{1.}} = \frac{p_{21}}{p_{2.}} \rightarrow \frac{p_{11}}{p_{1.}} = \frac{p_{.1} - p_{11}}{1 - p_{1.}} \rightarrow$$

$$p_{11} - p_{11} p_{1.} = p_{1.} p_{.1} - p_{1.} p_{11} \rightarrow p_{11} = p_{1.} p_{.1}$$

$$\text{Se } p_{11} = p_{1.} p_{.1}, \text{ então } p_{ij} = p_{i.} p_{.j}, \quad \forall i, j. \quad (1.10)$$

Portanto, se o valor do risco relativo for 1, haverá a independência entre as variáveis.

De acordo com Agresti (1990), o risco relativo não pode ser calculado nos *estudos comparativos retrospectivos* (caso e controle), pois as marginais fixadas não são apropriadas para este cálculo.

Exemplo : Consideremos a seguinte tabela de dados fictícios, que relaciona as variáveis transfusão de sangue e infecção.

O objetivo é verificar se a probabilidade de ocorrer infecção em um doente aumenta, se ele sofreu uma transfusão.

A variável preditora transfusão foi fixada e a variável resposta é a infecção.

transfusão	infecção		total
	1	0	
1	5	49	54
0	13	35	48
			102

1: presença
0: ausência

Calculando-se o risco relativo de infecção em relação à transfusão, tem-se:

$$RR = \frac{P(I = 1|T = 1)}{P(I = 1|T = 0)} = \frac{5/54}{13/48} = 0,3419$$

Portanto a probabilidade de um indivíduo que não sofreu transfusão ter uma infecção é 2,9 vezes a probabilidade de um indivíduo que sofreu transfusão ter uma infecção.

1.3.3 "Odds"

Sejam duas variáveis, cada uma com dois níveis. Considerando-se um determinado nível em uma das variáveis, define-se como "odds" da outra variável a razão entre a probabilidade de presença e a da ausência da mesma condicionadas ao nível determinado. "Odds" é um número real não negativo ilimitado.

Notação: $Od(T = t)$, onde t é o nível fixado de uma das variáveis que geralmente é da variável preditora.

$$\Omega_1 = Od(X=1) = \frac{p_{11}}{p_{21}} = \frac{p_{11}/p_{1.}}{p_{12}/p_{1.}} = \frac{p_{11}}{p_{12}} \quad \text{ou} \quad \Omega_1 = \frac{P(Y = 1|X = 1)}{P(Y = 0|X = 1)} \quad (1.11)$$

$$\Omega_2 = Od(X = 0) = \frac{p_{12}}{p_{22}} = \frac{p_{21}/p_{2.}}{p_{22}/p_{2.}} = \frac{p_{21}}{p_{22}} \quad \text{ou} \quad \Omega_2 = \frac{P(Y = 1|X = 0)}{P(Y = 0|X = 0)} \quad (1.12)$$

Se as variáveis forem independentes, então $\Omega_1 = \Omega_2$.

No exemplo de infecção e transfusão, temos:

a) A estimativa da "Odds" da infecção, considerando os que sofreram transfusão é :

$$\Omega_1 = \frac{P(I = 1|T = 1)}{P(I = 0|T = 1)} = \frac{5/102}{49/102} = \frac{5}{49} = 0,1020.$$

b) A estimativa da "Odds" da infecção dentro do nível dos que não sofreram transfusão é :

$$\Omega_2 = \frac{P(I = 1|T = 0)}{P(I = 0|T = 0)} = \frac{13}{35} = 0,3714.$$

Como $\Omega_1 \neq \Omega_2$ existem indícios de que as variáveis não são independentes.

Dos dois resultados, observa-se que a resposta presença de infecção é menos provável que a ausência de infecção, tanto entre os que sofreram transfusão como entre aqueles que não sofreram transfusão.

1.3.4 "Odds Ratio"

Denomina-se "odds ratio" a razão entre as "odds" de uma variável calculada nos dois diferentes níveis da outra variável. Em nosso trabalho ela será denotada por θ .

$$\theta = \frac{\Omega_1}{\Omega_2} = \frac{\frac{p_{11}}{p_{12}}}{\frac{p_{21}}{p_{22}}} = \frac{p_{11} \cdot p_{22}}{p_{12} \cdot p_{21}} \quad (1.13)$$

A "odds ratio" é também conhecida como *razão produto-cruzado*, devido ao fato de que o numerador e o denominador desta razão são os produtos das probabilidades das celas das diagonais da tabela 2×2 .

Se nenhuma cela tiver probabilidade zero, o valor de θ varia de 0 a ∞ .

As variáveis são independentes se, e somente se, "odds ratio" for igual a 1.

Demostremos esta afirmativa.

a) Se as variáveis X e Y são independentes temos:

$$\theta = \frac{p_{11} \cdot p_{22}}{p_{12} \cdot p_{21}} = \frac{p_{1.} \cdot p_{.1} \cdot p_{2.} \cdot p_{.2}}{p_{1.} \cdot p_{.2} \cdot p_{2.} \cdot p_{.1}} = 1$$

b) Se $\theta = 1$

$$\theta = \frac{p_{11} \cdot p_{22}}{p_{12} \cdot p_{21}} = 1 \rightarrow p_{12} p_{21} = p_{11} p_{22} \rightarrow$$

$$(p_{1.} - p_{11})(p_{.1} - p_{11}) = p_{11}(p_{2.} - p_{21}) \rightarrow$$

$$p_{1.} p_{.1} - p_{11} p_{.1} - p_{11} p_{.1} + p_{11}^2 = p_{11}[(1 - p_{1.}) - (p_{.1} - p_{11})] \rightarrow$$

$$p_{1.} p_{.1} - p_{11} p_{.1} - p_{11} p_{.1} + p_{11}^2 = p_{11} - p_{11} p_{1.} - p_{11} p_{.1} + p_{11}^2 \rightarrow$$

$$p_{11} = p_{1.} p_{.1}$$

$$\text{Se } p_{11} = p_{1.} p_{.1}, \text{ então } p_{ij} = p_{i.} p_{.j}, \quad \forall i, j. \quad (1.14)$$

Portanto, se a "odds ratio" for 1, as variáveis em estudo são independentes.

Se $0 < \theta < 1$, a primeira resposta é menos provável na linha 1 que na linha 2 ou seja $p_{1/1} < p_{1/2}$ e, se $\theta > 1$, a primeira resposta é mais provável na linha 1 do que na linha 2.

Por exemplo, se $\theta = 3$, então a “odds” da primeira resposta é duas vezes maior na linha 1 do que na linha 2. Observe que este valor não significa que probabilidade $p_{1/1}$ é três vezes a probabilidade $p_{1/2}$. Esta interpretação é do risco relativo.

Valores de θ mais distantes de 1 representam níveis de associação mais fortes.

Dois valores de θ que sejam números inversos representam o mesmo nível de associação, mas em sentidos contrários.

O valor estimado da “odds ratio” não muda quando multiplicamos todas as freqüências das celas de uma linha ou uma coluna por uma constante não nula. Conseqüentemente, a estimativa da “odds ratio” é a mesma, independentemente de termos selecionado marginais totais iguais ou não.

No exemplo de infecção e transfusão, se nós tivéssemos amostrado 50 pessoas que fizeram transfusão e 50 que não fizeram transfusão, ou 100 pessoas que fizeram transfusão e 50 que não fizeram e, em cada caso, classificá-los em indivíduos que tiveram ou não infecção, a “odds ratio” estimaria a mesma característica.

Do mesmo modo, nós poderíamos ter escolhido uma amostra de 100 indivíduos e classificar cada um relacionando aos pares (1,1) ; (1,0) ; (0,1); (0,0) que a “odds ratio” teria a mesma característica.

Calculando-se a “odds ratio” no exemplo de infecção e transfusão, temos

$$\text{“odds ratio”} = \frac{5 \times 35}{49 \times 13} = 0,2747$$

Como $0 < \theta < 1$, há indícios de que seja menos provável o aparecimento de infecção entre os que foram submetidos à transfusão.

Segundo Agresti (1990), a “odds ratio” é igualmente válida para estudos *retrospectivos, prospectivos ou cross-over*; isto foi demonstrado por Cornfield (1951), que destaca o fato de que, para efetuar-se uma análise fazendo uso da “odds ratio”, os dados de uma tabela tanto podem ter vindo de um estudo caso- controle ou estudo de coortes.

Cornfield (1951), em um estudo sobre métodos de análise de ensaios clínicos, utilizou os dados de Schrek (1950) os quais considera dois grupos de homens brancos com idade entre 40 e 49 anos, sendo que um dos grupos tinha câncer de pulmão e o outro era representativo de uma população sem câncer de pulmão.

Os dois grupos foram classificados em fumantes ou não fumantes, sendo classificado como fumante um indivíduo que fumava 10 ou mais cigarros por dia. Dos elementos que



tinham câncer de pulmão, 77% foram classificados como fumantes e 58% do grupo controle tiveram esta mesma classificação.

Considerando-se p_1 e p_2 estas proporções e K a proporção na população geral que teve câncer de pulmão durante um especificado período de tempo, pode-se obter, para a população geral, uma importante informação que ficará sintetizada na tabela abaixo.

X	Y	
	1	0
1	p_1K	$p_2(1-K)$
0	$(1-p_1)K$	$(1-p_2)(1-K)$
total	K	$1-K$

X: fumantes

Y: doentes com câncer de pulmão

Sendo p a proporção de fumantes na população, podemos obtê-la por

$$p = P(\text{fumantes com câncer}) + P(\text{fumantes sem câncer}).$$

Observando a tabela temos:

$$p = p_1 K + p_2 (1 - K) = p_2 + K(p_1 - p_2)$$

e a proporção de não fumantes na população será obtida por

$$1 - p = (1 - p_2) - K(p_1 - p_2).$$

As proporções de fumantes e não fumantes tendo câncer serão calculadas por

$$p_1K / [p_2 +K(p_1 -p_2)] \tag{1.15}$$

e

$$(1- p_1)K / [(1- p_2) -K(p_1- p_2)] \tag{1.16}$$

Estas fórmulas fornecerão as taxas verdadeiras.

Se a proporção da população geral, K , for relativamente pequena nas duas expressões, o termo $K(p_1 - p_2)$ pode ser desconsiderado e teremos a verdadeira taxa de câncer de pulmão

aproximada, entre fumantes dada pela seguinte expressão: $\frac{p_1K}{p_2}$; e , $\frac{(1-p_1)K}{1-p_2}$, a

correspondente taxa para os não fumantes.

Conseqüentemente, se K for conhecido, pode-se estimar a ocorrência da doença na população de fumantes e não fumantes.

Para a análise desses dados, Cornfield (1951) usou os resultados de Dorn (1944), que apontavam a ocorrência anual de câncer de pulmão, em homens brancos na faixa etária de 40 a 49 anos, de 15,5 por 100.000 e, por isto achou que um bom valor para K era $0,155 \times 10^{-3}$.

A seguinte Tabela foi, então, montada:

X	Y		
	1	0	
1	$0,119 \times 10^{-3}$	0,579910	0,580029
0	$0,036 \times 10^{-3}$	0,419935	0,419971
total	$0,155 \times 10^{-3}$	0,999845	1,000000

As proporções de fumantes e de não fumantes que têm câncer de pulmão calculadas pelas fórmulas (1.15) e (1.16) foram $0,205 \times 10^{-3}$ e $0,086 \times 10^{-3}$. Logo as correspondentes taxas anuais estimadas para estas categorias são de 20,5 fumantes e 8,6 de não fumantes por 100.000 doentes de câncer. Essas taxas produzem uma fonte básica para avaliar o efeito do fumo na ocorrência de câncer de pulmão.

Se não for conhecido o valor de K, pode-se também calcular a relação entre essas taxas somente através das freqüências relativas, pois a razão entre as taxas, quando K for pequeno, é $\frac{p_1(1-p_2)}{p_2(1-p_1)}$. Através desta expressão, somente utilizando os dados de Schrek (1950), obtém-se a informação de que a ocorrência de câncer de pulmão numa população de homens brancos entre 40 e 49 anos é 1,4 vezes maior entre os que fumam 10 ou mais cigarros por dia do que entre os que não fumam.

Outras análises semelhantes foram apresentadas por Cornfield (1951), o que o levou a concluir que

“...qualquer conjunto de dados que forneça estimativas de freqüências relativas pode ser usado para obter estimativas das taxas.”

Quando estivermos trabalhando nos estudos retrospectivos, geralmente não é possível calcular o risco relativo para a variável resposta. Nestes casos, podemos optar pela “odds ratio”, que pode ser calculada em qualquer direção.

Vamos analisar as situações onde a ocorrência de um nível de interesse seja rara. Se p_{11} e p_{21} são valores pequenos tem-se $p_{11} + p_{12} \cong p_{12}$ e $p_{21} + p_{22} \cong p_{22}$ e, então, calculando-se o risco relativo, temos:

$$RR = \frac{p_{11}}{p_{12}} = \frac{p_{11}}{p_{1.}} = \frac{p_{11}}{p_{11} + p_{12}} \cong \frac{p_{11}}{p_{12}} = \frac{p_{11}p_{22}}{p_{12}p_{21}} = \theta \quad (1.17)$$



Conseqüentemente, quando a probabilidade do resultado de interesse for muito pequena, os valores do risco relativo e “odds ratio” são muito próximos; logo, nestas situações, podemos obter conclusões para o risco relativo, analisando o valor obtido para a “odds ratio”.

Freqüentemente, o logaritmo natural da “odds ratio” é usado, ao invés da mesma.

Esta transformação associa medidas do intervalo de $[0, \infty[$ a valores simétricos em torno do zero no intervalo real.

Considerando que $-\ln \theta = \ln \frac{1}{\theta}$, portanto, $\ln \theta$ e $-\ln \theta$ são transformações obtidas de valores inversos; logo dois valores de $\ln \theta$ que sejam opostos representam o mesmo nível de associação.

Se há independência entre as variáveis, o valor de θ é igual a 1 e, conseqüentemente, $\ln \theta = 0$.

Outra vantagem do logaritmo da “odds ratio” é a normalidade assintótica demonstrada por Fleiss (1979); sob as condições de independência.

A utilização do $\ln \theta$ como medida de associação está também ligada à construção do modelo logito.

Devido a esses aspectos, prefere-se, muitas vezes, estimar o $\ln \theta$, mas a estimação de θ também deve ser considerada, visto que a “odds ratio” é, na prática, muito usada devido a sua fácil interpretação.

1.3.5 Coeficiente Q de Yule

Yule (1912) define o seguinte coeficiente de associação, que ele notou por Q, em homenagem à Quetelet, estatístico Belga que viveu no séc XIX;

$$Q = \frac{p_{11}p_{22} - p_{12}p_{21}}{p_{11}p_{22} + p_{12}p_{21}} \quad (1.18)$$

Se considerarmos a diferença entre o valor observado e o valor esperado sob a hipótese de independência para a cela c_{11} :

$$D = n_{11} - \frac{n_{1.}n_{.1}}{n} = \frac{n_{11}n - n_{1.}n_{.1}}{n}.$$

Como $n = n_{1.} + n_{2.}$, se a substituirmos na expressão acima, temos:



$$D = \frac{n_{11}n_{22} - n_{12}n_{21}}{n} \quad (1.19)$$

A estimativa do Q está relacionada ao D

$$\hat{Q} = \frac{n_{11}n_{22} - n_{12}n_{21}}{n_{11}n_{22} + n_{12}n_{21}} = \frac{nD}{n_{11}n_{22} + n_{12}n_{21}} \quad (1.20)$$

Se $\hat{Q} = 0$ as variáveis são independentes.

Se $n_{12} \cdot n_{21} = 0$ então $\hat{Q} = +1$. E, neste caso, há associação positiva completa entre as variáveis.

Se $n_{11} \cdot n_{22} = 0$ então $\hat{Q} = -1$. E, neste caso, há associação negativa completa entre as variáveis.

O coeficiente Q de Yule é uma transformação estritamente monotônica da "odds ratio" (θ) indo do intervalo $[0, \infty[$ para o intervalo $[-1, 1]$.

Dividindo-se o numerador e o denominador da expressão (1.20) por $n_{12} \cdot n_{21}$, temos que

$$\hat{Q} = \frac{\hat{\theta} - 1}{\hat{\theta} + 1}, \quad (1.21)$$

onde $\hat{\theta}$ é o estimador de máxima verossimilhança de θ , como será descrito abaixo.

Ao observarmos a Tabela 1, podemos classificar os elementos n_{11} e n_{22} como pares concordantes e, n_{12} e n_{21} como pares discordantes.

As probabilidades dos pares concordantes e dos pares discordantes são, respectivamente:

$$\frac{p_{11}p_{22}}{p_{11}p_{22} + p_{12}p_{21}} \quad e \quad \frac{p_{12}p_{21}}{p_{11}p_{22} + p_{12}p_{21}}$$

e, se subtrairmos estas expressões, temos:

$$\frac{p_{11}p_{22}}{p_{11}p_{22} + p_{12}p_{21}} - \frac{p_{12}p_{21}}{p_{11}p_{22} + p_{12}p_{21}} = \frac{p_{11}p_{22} - p_{12}p_{21}}{p_{11}p_{22} + p_{12}p_{21}} = Q.$$

Portanto, o coeficiente Q de Yule pode ser interpretado como a diferença entre as probabilidades condicionais dos pares concordantes e dos pares discordantes.

1.4 Estimadores e Intervalos de Confiança para :

“odds”, “odds ratio”, diferença de proporções e risco relativo.

Os estimadores dos parâmetros aqui apresentados têm, como referências, Haldane (1955), Fleiss (1981), Liebetrau (1983), Agresti (1990) e Walter e Cook (1991).

Para estimar a “odds ratio”, que foi por nós denotada por θ , vamos considerar o caso de duas amostras binomiais independentes de tamanhos n_1 e n_2 , com probabilidades de sucesso p_{11} e p_{21} , respectivamente.

Tabela 3

	1	0	total
X	n_{11}	n_{12}	$n_{1.}$
Y	n_{21}	n_{22}	$n_{2.}$

O estimador de máxima verossimilhança incondicional para θ é deduzido recorrendo-se aos tamanhos amostrais n_1 e n_2 fixados, e p_{11} e p_{21} são estimados pelos seus estimadores de máxima verossimilhança.

$$\hat{p}_{11} = \frac{n_{11}}{n_{1.}} \quad e \quad \hat{p}_{21} = \frac{n_{21}}{n_{2.}}.$$

Os estimadores para as “odds” Ω_1 e Ω_2 são feitos usando-se a propriedade de invariância dos estimadores de máxima verossimilhança, obtendo-se

$$\hat{\Omega}_1 = \frac{n_{11}}{n_{12}} \quad (1.22)$$

$$\hat{\Omega}_2 = \frac{n_{21}}{n_{22}} \quad (1.23)$$

Fazendo-se a razão entre os estimadores das “odds”, tem-se

$$\hat{\theta} = \frac{n_{11} \cdot n_{22}}{n_{12} \cdot n_{21}} \quad (1.24)$$

A “odds ratio” amostral $\hat{\theta}$ é igual a 0 ou ∞ , se algum $n_{ij} = 0$ e, ela não é definida se ambas as entradas de uma linha ou coluna são zero. Para que possamos fazer estimação sob estas condições, Haldane (1955) sugeriu que fosse adicionado um termo de correção igual a



0,5 em todas as quatro células. De acordo com Walter (1985), este procedimento acarreta melhora nas propriedades do Erro Médio Quadrático e no viés do $\ln\theta$, se o adicionarmos apenas nas células em que o zero ocorre.

Haldane (1955) mostrou que o estimador transformado

$$\tilde{\theta} = \frac{(n_{11} + 0,5)(n_{22} + 0,5)}{(n_{12} + 0,5)(n_{21} + 0,5)} e^{\ln \tilde{\theta}} \quad (1.25)$$

tem um bom comportamento em termos de viés e erro médio quadrático.

Walter e Cook (1991) comprovaram estas vantagens e, seus resultados foram apresentados em tabelas e gráficos onde estão comparados o viés e o erro médio quadrático dos estimadores de $\hat{\theta}$ com o de $\tilde{\theta}$ e o de $\ln \hat{\theta}$ com o de $\ln \tilde{\theta}$.

O acréscimo de 0,5 em cada célula desaparece quando $n \rightarrow \infty$ e, portanto, $\hat{\theta}$ e $\tilde{\theta}$ têm a mesma distribuição normal assintótica.

Segundo Agresti (1990), este estimador é o mesmo sob os modelos produto binomial, multinomial completa ou de Poisson dentro das linhas ou das colunas quando n tende ao infinito.

Apesar do estimador de θ ter uma distribuição normal assintótica, a variância assintótica deste estimador é geralmente apresentada pela variância de $\ln \hat{\theta}$.

$$Var[\ln \hat{\theta}] = \frac{1}{n_{11}} + \frac{1}{n_{12}} + \frac{1}{n_{21}} + \frac{1}{n_{22}} \quad (1.26)$$

e desvio padrão assintótico estimado

$$\hat{\sigma}(\ln \hat{\theta}) = \left(\frac{1}{n_{11}} + \frac{1}{n_{12}} + \frac{1}{n_{21}} + \frac{1}{n_{22}} \right)^{\frac{1}{2}} \quad (1.27)$$

Para construirmos um intervalo de confiança para $\ln\theta$, devemos considerar $z_{\alpha/2}$

onde $1 - \alpha/2$ é o quantil da distribuição normal padrão de ordem $\alpha/2$.

$$\ln \hat{\theta} \pm z_{\alpha/2} \hat{\sigma}(\ln \hat{\theta}) \quad (1.28)$$

são os limites de um intervalo de confiança 100 $(1 - \alpha)$ por cento para $\ln\theta$.

Exponencializando os extremos deste intervalo, obtemos um intervalo de confiança para θ .

Para calcularmos o estimador da diferença das proporções $p = p_{11} - p_{12}$, vamos considerar a Tabela 3.

O estimador de máxima verossimilhança de $p_{1|i}$ é $\hat{p}_{1|i} = \frac{n_{i1}}{n_i}$ e tem-se

$$E(\hat{p}_{1|i}) = p_{1|i} \quad i = 1, 2$$

$$Var(\hat{p}_{1|i}) = \frac{p_{1|i}(1-p_{1|i})}{n_i} \quad (1.29)$$

$$\text{logo } E(\hat{p}_{1|1} - \hat{p}_{1|2}) = p_{1|1} - p_{1|2} \quad (1.30)$$

$$\text{e o desvio padrão } \sigma(\hat{p}_{1|1} - \hat{p}_{1|2}) = \left[\frac{p_{1|1}(1-p_{1|1})}{n_{1.}} + \frac{p_{1|2}(1-p_{1|2})}{n_{2.}} \right]^{1/2} \quad (1.31)$$

Substituindo cada $p_{j|i}$ por seus estimadores, na expressão acima, obtemos, então,

$$(\hat{p}_{1|1} - \hat{p}_{1|2}) \pm z_{\alpha/2} \hat{\sigma}(\hat{p}_{1|1} - \hat{p}_{1|2}) \quad (1.32)$$

como limites de um intervalo de confiança para a *diferença de proporções*.

Consideremos agora um estimador para o risco relativo. Geralmente se utiliza o logaritmo do risco relativo para fazer sua estimação.

O desvio padrão assintótico para o logaritmo do risco relativo é

$$\sigma(\ln RR) = \left(\frac{1-p_{1|1}}{p_{1|1}n_{1.}} + \frac{1-p_{1|2}}{p_{1|2}n_{2.}} \right)^{1/2} \quad (1.33)$$

Mas $\ln RR$ e a versão amostral do desvio padrão $\sigma(\ln RR)$ não são definidos quando $\hat{p}_{1|1}$ ou $\hat{p}_{1|2}$ assumir o valor zero. Para contornar esta situação, um estimador de $\ln RR$ é

$$\ln(R\tilde{R}) = \ln\left(\frac{n_{11} + 0,5}{n_{1.} + 0,5}\right) - \ln\left(\frac{n_{21} + 0,5}{n_{2.} + 0,5}\right) \quad (1.34)$$

e um intervalo de confiança para o risco relativo tem limites:

$$\ln R\tilde{R} \pm z_{\alpha/2} \left[(n_{11} + 0,5)^{-1} - (n_{1.} + 0,5)^{-1} + (n_{21} + 0,5)^{-1} - (n_{2.} + 0,5)^{-1} \right]^{1/2} \quad (1.35)$$

Exponencializando os extremos, obtemos um intervalo de confiança para o risco relativo.

2 . INFERÊNCIAS EM TABELAS 2 X 2

Ao efetuarmos uma pesquisa, muitas interrogações são intrínsecas ao experimento. Para respondê-las, a estimação de parâmetros, de intervalos de confiança e testes de hipóteses serão parte integrante da pesquisa.

Quando utilizarmos um método inferencial, precisaremos conhecer o modelo amostral, sua distribuição e, algumas vezes, optar pela função de verossimilhança condicional ou incondicional .

Os modelos amostrais já foram descritos no capítulo 1, em 1-1. Vamos, agora, descrever as diferentes distribuições amostrais em tabelas 2x2.

Notaremos, por e_{ij} , a frequência esperada para cada cela.

$$E(n_{ij}) = e_{ij} \quad i = 1, 2 \quad j = 1, 2$$

2.1 Distribuições Amostrais

2.1.1 Modelo de Poisson (tempo ou espaço fixo)

O modelo probabilístico será de Poisson, quando somente o *tempo* ou o *espaço*, em uma pesquisa a ser realizada, for fixado.

Por exemplo, em uma cidade de porte médio, se quisermos verificar se há associação entre tipo de acidente de moto (fatal e não-fatal) com a idade do motorista acidentado (abaixo ou acima de 20 anos) e, supondo-se que foram anotados todos os n acidentes no decorrer de um ano e que eles foram classificados de acordo com estas variáveis descritas acima.

Estas variáveis são variáveis aleatórias com distribuição de Poisson, independentes, com parâmetros $\{ e_{11}, e_{12}, e_{21}, e_{22} \}$, se forem obedecidos os axiomas do modelo.

A distribuição das frequências amostradas para a Tabela 1 é :

$$L = \prod_{j=1}^2 \left(\prod_{i=1}^2 \frac{\exp(-e_{ij}) e_{ij}^{n_{ij}}}{n_{ij}!} \right) \quad (2.1)$$

O tamanho amostral n também tem uma distribuição Poisson com parâmetro $\sum e_{ij}$.



2.1.2 Modelo Multinomial (tamanho amostral total fixo)

Suponhamos que, na pesquisa de acidentes de moto, os pesquisadores tenham considerado os n acidentes de moto ocorridos no ano de estudos que passou e que cada motorista acidentado tenha sido classificado segundo a faixa etária e o tipo de acidente descrito anteriormente. Para este estudo do tipo “cross-over”, a ocorrência das frequências nas quatros celas tem uma distribuição multinomial.

Dado que $\sum n_{ij} = n$, a probabilidade de ocorrer o conjunto $\{n_{ij}\}$ é

$$L = \frac{n!}{\prod_j \left(\prod_i n_{ij}! \right)} \prod_j \left(\prod_i p_{ij}^{n_{ij}} \right), \quad (2.2)$$

onde $\{p_{ij}\}$ é o conjunto das probabilidades indicadas na Tabela 2.

2.1.3 Produto de Binomiais (uma marginal é fixada)

Se selecionarmos inicialmente duas amostras, uma com elementos que sofreram acidentes fatais e outra com elementos que sofreram acidentes não-fatais, e os classificarmos pela faixa etária, o modelo probabilístico é um produto de distribuições binomiais independentes.

Considerando, sem perda de generalidade, $n_{1.}$ e $n_{2.}$ fixas, e sendo n_{11} e n_{21} as variáveis independentes, conseqüentemente $n_{.1}$ será uma variável aleatória.

Tem-se, assim, a função de probabilidade

$$\begin{aligned} P(n_{11}, n_{21} | n_{1.}, n_{2.}) &= P(n_{11} | n_{1.}) P(n_{21} | n_{2.}) = \\ &= \binom{n_{1.}}{n_{11}} p_{11}^{n_{11}} (1 - p_{11})^{n_{1.} - n_{11}} \binom{n_{2.}}{n_{21}} p_{21}^{n_{21}} (1 - p_{21})^{n_{2.} - n_{21}} \end{aligned} \quad (2.3)$$



2.1.4 Modelo Hipergeométrico (ambas as marginais fixadas)

Em alguns projetos de experimentos, ambos os totais marginais são fixados.

Isto ocorre, se um corpo de jurados deve ser formado com 21 elementos, sendo que, pela natureza do crime, neste corpo deve haver 10 homens e 11 mulheres e, além disto, este corpo também deve ter 10 pessoas de raça negra e 11 pessoas de raça branca .

As tabelas abaixo são ilustrações de possíveis corpos de jurados.

sexo	raça		total
	B	N	
masculino	6	4	10
feminino	5	6	11
total	11	10	21

ou

sexo	raça		total
	B	N	
masculino	5	5	10
feminino	6	5	11
total	11	10	21

Fisher, em seu livro *The Design of Experiments* (1935), descreve um experimento comumente conhecido como o Teste do Chá, que ilustra uma situação onde todas as marginais são fixadas.

Uma senhora inglesa afirmava que, ao beber uma xícara de chá com leite, ela podia identificar se o leite ou o chá havia sido colocado primeiramente na xícara, ao ser preparada a bebida.

Para testar a veracidade desta afirmativa, foram oferecidas à senhora oito xícaras de chá com leite, sendo que o preparo, em quatro delas, fora feito na ordem leite e chá e, nas outras quatro, chá e leite. A senhora foi informada sobre estas quantidades.



As xícaras eram apresentadas em uma ordem aleatória e a senhora, após provar as bebidas, deveria classificá-las de acordo com a substância inicial que ela acreditava ter sido colocada na xícara. Dois possíveis resultados estão apresentados nas tabelas abaixo.

X	Y		total
	leite	chá	
leite	3	1	4
chá	1	3	4
total	4	4	8

X	Y		total
	leite	chá	
leite	4	0	4
chá	0	4	4
total	4	4	8

X: primeira substância colocada na xícara

Y: primeira substância indicada pela senhora

Este experimento, planejado por Fisher, produz um clássico e raro exemplo de modelo em tabelas 2x2, com ambas as marginais fixadas anteriormente.

Nestes modelos de experimentos, a distribuição conjunta utilizada para a análise deve ser obtida condicionando-a aos totais marginais.

Tanto nos modelos de Poisson, Multinomiais ou Binomiais Independentes quando condicionamos aos totais marginais, sob a hipótese nula de independência entre as variáveis X e Y, obtemos uma distribuição Hipergeométrica.

$$L = \frac{\binom{n_{1.}}{n_{11}} \binom{n_{.2}}{n_{.1} - n_{11}}}{\binom{n}{n_{.1}}} \tag{2.4}$$

Esta fórmula expressa a distribuição de ocorrência das quatro celas em termos de apenas um elemento: n_{11} .



2.2 Testes de Independência para Grandes Amostras em tabelas 2 x 2

Dois problemas fundamentais, em pesquisas, são a investigação de independência entre as duas variáveis e o julgamento da igualdade de um conjunto de probabilidades e de certos valores fixos.

Para testar a hipótese de independência em tabelas 2x2, em situações de grandes amostras, apresentaremos duas estatísticas. Uma obtida através da razão de verossimilhança e outra, baseada nas diferenças entre os valores observados e os valores esperados, sob H_0 , em cada cela da tabela.

2.2.1 Estatística Quiquadrado da Razão de Verossimilhança

Para construir esta estatística, consideraremos um modelo multinomial cujas observações foram retiradas de uma grande população ou retiradas com reposição.

Utilizando a função de verossimilhança, vamos construir a razão de verossimilhança generalizada.

A distribuição de probabilidades segundo as frequências amostradas (Tabela 1) é

$$L = \frac{n!}{n_{11}! n_{12}! n_{21}! n_{22}!} p_{11}^{n_{11}} p_{12}^{n_{12}} p_{21}^{n_{21}} p_{22}^{n_{22}} \quad (2.5)$$

e os estimadores de máxima verossimilhança de p_{ij} , são

$$\hat{p}_{ij} = \frac{n_{ij}}{n} \quad i = 1, 2 \quad j = 1, 2 \quad (2.6)$$

Se considerarmos a igualdade

$$p_{11}p_{22} = p_{12}p_{21} \rightarrow$$

$$p_{11}p_{22} = (p_{1.} - p_{11})(p_{.1} - p_{11}) = p_{1.}p_{.1} - p_{1.}p_{11} - p_{1.}p_{11} + p_{11}^2 \rightarrow$$

$$p_{11}p_{22} + p_{1.}p_{11} + p_{1.}p_{11} - p_{11}^2 = p_{1.}p_{.1} \rightarrow$$

$$p_{11}(p_{22} + p_{.1} + p_{1.} - p_{11}) = p_{1.}p_{.1} \rightarrow$$

$$p_{11}(p_{22} + p_{.1} + p_{21}) = p_{1.}p_{.1} \rightarrow$$

$$p_{11}(p_{1.} + p_{2.}) = p_{1.}p_{.1} \rightarrow$$



$$p_{11} = p_{1.} p_{.1}$$

Se $p_{11} = p_{1.} p_{.1}$, então $p_{ij} = p_{i.} p_{.j} \quad \forall i, j$.

Portanto, a hipótese de independência pode ser representada por

$$H_0 : p_{11} \cdot p_{22} = p_{12} \cdot p_{21}$$

Para expressar p_{21} e p_{22} , em função de p_{11} e p_{12} , utilizaremos as seguintes expressões

$$\begin{aligned} p_{21} &= \frac{p_{11}(1 - p_{11} - p_{12})}{p_{11} + p_{12}} \\ p_{22} &= \frac{p_{12}(1 - p_{11} - p_{12})}{p_{11} + p_{12}} \end{aligned} \quad (2.7)$$

Substituindo-as na função de verossimilhança (2.5) e aplicando o logaritmo, temos

$$\log L \propto n_{1.} \log p_{11} + n_{2.} \log p_{12} + n_{2.} (\log (1 - p_{1.}) - \log p_{1.})$$

Sob H_0 , os estimadores de máxima verossimilhança são:

$$\hat{p}_{11} = \frac{n_{1.}}{n} \frac{n_{.1}}{n}, \hat{p}_{12} = \frac{n_{1.}}{n} \frac{n_{.2}}{n}, \hat{p}_{21} = \frac{n_{2.}}{n} \frac{n_{.1}}{n}, \hat{p}_{22} = \frac{n_{2.}}{n} \frac{n_{.2}}{n} \quad (2.8)$$

Portanto, nós podemos estimar as probabilidades das celas pelo produto das frequências marginais proporcionais.

Substituindo (2.8) em (2.5), temos que, sob H_0 ,

$$L(n_{ij} | H_0, \hat{p}_{ij}) \propto (n_{1.} n_{.1})^{n_{11}} (n_{1.} n_{.2})^{n_{12}} (n_{2.} n_{.1})^{n_{21}} (n_{2.} n_{.2})^{n_{22}} / n^{2n} \quad (2.9)$$

e o máximo incondicional da função de verossimilhança se dá, substituindo (2.6) em (2.5),

$$L(n_{ij} | \hat{p}_{ij}) \propto n_{11}^{n_{11}} n_{12}^{n_{12}} n_{21}^{n_{21}} n_{22}^{n_{22}} / n^n, \quad (2.10)$$

o que nos dá a seguinte estatística teste:

$$\lambda = \left(\frac{\hat{e}_{11}}{n_{11}} \right)^{n_{11}} \left(\frac{\hat{e}_{12}}{n_{12}} \right)^{n_{12}} \left(\frac{\hat{e}_{21}}{n_{21}} \right)^{n_{21}} \left(\frac{\hat{e}_{22}}{n_{22}} \right)^{n_{22}}, \quad (2.11)$$

onde λ é a razão de verossimilhança generalizada e $\hat{e}_{ij} = \frac{n_{i.} n_{.j}}{n}$

Wilks (1935, 1938) mostrou que $-2 \log \lambda$ tem, assintoticamente, distribuição quiquadrado com 1 grau de liberdade.

Esta estatística, denotada por G^2 , tem a seguinte expressão:

$$G^2 = -2 \log \lambda = 2 \sum_j \sum_i n_{ij} \log \left(\frac{n_{ij}}{e_{ij}} \right) \quad (2.12)$$

e recebe o nome de *estatística quiquadrado da razão de verossimilhança*.

A expressão $-2 \log \lambda$, após aproximação por série de Taylor em torno de e_{ij} , pode ser expressa por:

$$-2 \log \lambda \doteq \sum_i \sum_j \frac{(n_{ij} - e_{ij})^2}{e_{ij}} \equiv X^2 \quad (2.13)$$

Portanto, existe uma equivalência assintótica da estatística quiquadrado da razão de verossimilhança G^2 com a estatística teste X^2 que será considerada a seguir.

Conseqüentemente, a *hipótese de independência*, nos casos de grandes amostras, pode ser testada através do ajustamento das observações com as frequências esperadas (X^2) ou através da estatística G^2 .

2.2.2 Teste quiquadrado de Pearson

Pearson (1904) propôs uma estatística já apresentada em (2.13) para testar a hipótese de independência entre linhas e colunas em tabelas de contingências de 2 entradas, ou seja, para testar a hipótese

$$H_0 : p_{ij} = p_i \cdot p_j \quad i = 1, 2, \dots, r \quad j = 1, 2, \dots, s$$

A estatística de Pearson é dada por

$$X^2 = \sum_j \sum_i \frac{(n_{ij} - e_{ij})^2}{e_{ij}} \quad (2.14)$$

onde e_{ij} é o valor esperado de n_{ij} sob a hipótese nula.

Para grandes amostras, em tabelas $r \times s$, X^2 é distribuída, assintoticamente, segundo uma distribuição quiquadrado (χ^2) com grau de liberdade dado por $gl = (r-1)(s-1)$.

Quando da proposta inicial deste teste, Pearson (1904) afirmara que o grau de liberdade era $gl = rs-1$; posteriormente Fisher (1925) demonstrou que o valor correto para o grau de liberdade é $gl = (r-1)(s-1)$.

Para o caso de uma tabela 2×2 , a estatística X^2 se resume à seguinte forma simplificada



$$X^2 = \frac{n(n_{11}n_{22} - n_{12}n_{21})^2}{n_{1.}n_{2.}n_{.1}n_{.2}} \quad (2.15)$$

Yates(1934) propôs a seguinte correção de continuidade para melhorar a aproximação com a distribuição χ^2 : o numerador da estatística X^2 deve sofrer um acréscimo de 0,5 em n_{12} e n_{21} e uma redução de 0,5 em n_{11} e n_{22} .

A estatística teste, com esta correção fica com a seguinte expressão

$$X_c^2 = \frac{n(|n_{11}n_{22} - n_{12}n_{21}| - 0,5n)^2}{n_{1.}n_{2.}n_{.1}n_{.2}} \quad (2.16)$$

Esta estatística é conhecida como *quiquadrado corrigido para continuidade*.

Conover (1968,1974) questionou a validade do seu uso rotineiro em todas as situações, mas Mantel e Greenhouse (1968), Fleiss (1981) e Mantel (1974) derrubaram os argumentos de Conover.

Não é simples descrever o tamanho amostral necessário para que a distribuição χ^2 seja suficientemente próxima da distribuição exata de X^2 e G^2 .

Para um número fixado de celas, a distribuição de X^2 geralmente converge mais rapidamente que a distribuição de G^2 para uma variável aleatória com distribuição χ^2 .

Quando $\frac{n}{rs} < 5$ a distribuição de G^2 não tem uma boa aproximação à uma distribuição χ^2 .

A distribuição da estatística X^2 também não tem uma boa aproximação com a distribuição χ^2 quando a esperança de alguma das celas for pequena. Isto foi exaustivamente estudado por R.A.Fisher e, em seu livro *Statistical Methods for Research Workers* (1925), ele estabeleceu que a frequência esperada em qualquer cela não poderia ser menor que 5 para que a aproximação da distribuição da estatística X^2 à uma distribuição χ^2 fosse razoável.

Snedecor e Cochran(1980) desaconselham o uso do teste quiquadrado se o tamanho amostral total for menor que 20 ou se o tamanho amostral está entre 20 e 40 e o menor valor esperado para as frequências de uma das celas for menor que 5.

O teste quiquadrado de Pearson é bilateral e, no caso de uma situação unilateral, a prática comum é dividirmos por dois o *valor p* obtido. Para os casos de teste de independência, não há necessidade de unilateralidade, pois é um teste de ajuste.

Exemplo 1: Os dados apresentados na tabela abaixo foram analisados por Clogg e Shockey (1988). Uma amostra de 1397 indivíduos foi classificada em favoráveis ou não ao porte de armas registradas por civis e favoráveis ou não à pena de morte.



registro de armas	pena de morte		total
	1	0	
1	784 (799,50)	236 (220,50)	1020
0	311 (294,69)	66 (81,50)	377
	1095	302	1397

1: favorável
0: contrário

Vamos testar a hipótese de independência.

$$H_0: p_{ij} = p_i \cdot p_j, \quad \forall i, j.$$

Neste caso, rejeitando H_0 temos $p_{ij} \neq p_i \cdot p_j$ para algum par (i, j) .

Observando o total amostral e as frequências esperadas (valores indicados entre parênteses), conclui-se que um teste, para grandes amostras, é adequado.

Os valores obtidos estão apresentados na tabela abaixo:

estatística	valor	valor p
X^2	5,15	0,0238
X_c^2	4,82	0,0292
G^2	7,027	0,0084

Consultando uma tabela padrão, constatamos que $\chi_1^2(0,95) = 3,84$ e, portanto, ao nível de significância $\alpha = 0,05$, rejeita-se a hipótese de independência. No entanto, se estabelecêssemos $\alpha = 0,01$, esta hipótese seria aceita se fossem usadas as estatísticas X^2 ou X_c^2 .

Os valores obtidos através das estatísticas testes nos levam a acreditar que existe uma pequena associação entre ser favorável à pena de morte e ser favorável a civis terem porte de arma registrada, na população pesquisada por Clogg e Shockey (1988).

O teste utilizando a estatística quiquadrado da razão de verossimilhança foi deduzido utilizando um modelo multinomial. No entanto o teste quiquadrado de Pearson também se mostra apropriado para os outros modelos de experimentos em casos de grandes amostras.

Vejamos para os outros modelos probabilísticos.



2.2.3 Ambas as marginais fixadas

Consideremos a Tabela 1 com a condição que os totais marginais estejam fixados, ou seja: n , $n_{1.}$ e $n_{.1}$ são fixas. Com estas suposições apenas n_{11} será uma variável livre.

A hipótese nula a ser testada é

$$H_0: \frac{p_{11}}{p_{1.}} = \frac{p_{21}}{p_{2.}} .$$

Considerando as hipóteses do modelo, a distribuição de n_{11} é da família hipergeométrica.

$$P\{n_{11}|n, n_{1.}, n_{.1}\} = \frac{\binom{n_{1.}}{n_{11}} \binom{n_{2.}}{n_{21}}}{\binom{n}{n_{.1}}} = \frac{n_{1.}! n_{.1}! n_{2.}! n_{2.}!}{n! n_{11}! n_{12}! n_{21}! n_{22}!} I(n_{11} \leq n_{.1} \wedge n_{1.}) \quad (2.17)$$

Portanto a esperança e variância são, respectivamente :

$$E(n_{11}) = \frac{n_{1.} n_{.1}}{n}$$
$$Var(n_{11}) = \frac{n_{1.} n_{.1} n_{2.} n_{2.}}{n^2 (n-1)}$$

O teorema central do limite garante que n_{11} é assintoticamente normal e, portanto, temos

$$t = \frac{n_{11} - E(n_{11})}{\sqrt{Var(n_{11})}} \sim N(0,1)$$

De acordo com Kendall e Stuart (1979), t^2 é, para n grande, equivalente à estatística X^2 e portanto, o teste quiquadrado será apropriado para testar a hipótese nula acima para as situações de ambas as marginais fixadas, em casos de grandes amostras.



2.2.4 Uma marginal é fixada

Na Tabela 1, consideremos, sem perda de generalidade, $n_{1.}$ e $n_{2.}$ fixas, n_{11} e n_{21} variáveis livres e, conseqüentemente, $n_{.1}$ será uma variável aleatória. Como descrito em 2.1.3, o modelo probabilístico é um produto de distribuições binomiais independentes.

A hipótese a ser testada será

$$H_0: \frac{p_{11}}{p_{1.}} = \frac{p_{21}}{p_{2.}}$$

ou seja, se as probabilidades do evento de interesse, nas duas populações, são iguais (homogeneidade).

Denotando por $p = \frac{p_{11}}{p_{1.}} = \frac{p_{21}}{p_{2.}}$ o valor hipotético comum, testaremos a hipótese nula,

considerando a diferença entre as proporções

$$d = \frac{n_{11}}{n_{1.}} - \frac{n_{21}}{n_{2.}},$$

que é distribuída, sob a hipótese e, n suficientemente grande, aproximadamente normal com esperança e variância, respectivamente

$$E(d) = 0$$

$$Var(d) = p(1-p) \left(\frac{1}{n_{1.}} + \frac{1}{n_{2.}} \right)$$

Substituindo $p(1-p)$ pelo seu estimador combinado

$$\frac{n}{n-1} \left(\frac{n_{1.}}{n} \right) \left(\frac{n_{2.}}{n} \right),$$

obtemos
$$\widehat{Var}(d) = \frac{n_{1.}n_{2.}}{(n-1)n_{1.}n_{2.}}.$$

Então

$$t = \frac{d - E(d)}{\sqrt{\widehat{Var}(d)}} = \frac{\frac{n_{11}}{n_{1.}} - \frac{n_{21}}{n_{2.}}}{\sqrt{\frac{n_{1.}n_{2.}}{(n-1)n_{1.}n_{2.}}}} \sim N(0,1)$$

Novamente, temos que t^2 é equivalente à estatística X^2 e o teste quiquadrado também será adequado para testar a homogeneidade entre as duas populações.



2.3 Dificuldades Relacionadas a Pequenas Amostras

Os testes até agora descritos foram assintóticos. De acordo com Fisher (1925), para situações de pequenas amostras eles não têm validade e, como forma alternativa, ele propôs um teste baseado na probabilidade exata de ocorrer um dado conjunto de frequências.

Esse teste se baseia na soma das probabilidades sob a “cauda” da distribuição de n_{11} . Essa soma vai nos dar a região crítica para o teste exato.

Mas, ao tentarmos calcular esta probabilidade exata, pode ser que nos deparemos com parâmetros incômodos na expressão.

Para o caso das *duas marginais fixadas* como visto em 2.2.3, a expressão

$$P\{n_{11}|n, n_{1.}, n_{.1}\} = \frac{\binom{n_{1.}}{n_{11}} \binom{n_{.2}}{n_{21}}}{\binom{n}{n_{.1}}} = \frac{n_{1.}! n_{.1}! n_{.2}! n_{.2}!}{n! n_{11}! n_{12}! n_{21}! n_{22}!} I(n_{11} \leq n_{.1} \wedge n_{.1})$$

possibilita-nos o cálculo da probabilidade exata. O teste exato de Fisher foi criado a partir desta hipótese.

Para a situação de uma *marginal fixada*, sob a hipótese nula, a probabilidade de observarmos a Tabela 1 é

$$\begin{aligned} P\{n_{11}|p, n_{1.}\} \cdot P\{n_{21}|p, n_{.2}\} &= \binom{n_{1.}}{n_{11}} p^{n_{11}} (1-p)^{n_{12}} \binom{n_{.2}}{n_{21}} p^{n_{21}} (1-p)^{n_{22}} = \\ &= \frac{n_{1.}! n_{.1}! n_{.2}! n_{.2}!}{n! n_{11}! n_{12}! n_{21}! n_{22}!} \binom{n}{n_{.1}} p^{n_{.1}} (1-p)^{n_{.2}} \end{aligned}$$

onde $p = \frac{p_{11}}{p_{1.}} = \frac{p_{21}}{p_{2.}}$.

A expressão, agora, depende de um parâmetro desconhecido p e a probabilidade exata não poderá ser avaliada.

Para o caso de apenas o *total amostral n fixo*, a probabilidade de n_{11} sob a hipótese nula, será dada por

$$\begin{aligned} P\{n_{11}|n_{1.}, n_{.1}, n\} P\{n_{.1}|p, n\} P\{n_{1.}|p', n\} &= \\ &= \binom{n_{1.}}{n_{11}} p^{n_{11}} (1-p)^{n_{12}} \binom{n_{.2}}{n_{21}} p^{n_{21}} (1-p)^{n_{22}} \binom{n}{n_{.1}} (p')^{n_{.1}} (1-p')^{n_{.2}} \end{aligned}$$



onde $p' = \frac{p_{11}}{p_{.1}} = \frac{p_{12}}{p_{.2}}$ e $p = \frac{p_{11}}{p_{1.}} = \frac{p_{21}}{p_{2.}}$.

Temos, então, uma expressão dependendo de dois parâmetros desconhecidos e não poderemos avaliar a probabilidade exata.

Estes parâmetros serão complicadores para a realização de inferências e portanto merecem uma atenção especial. No capítulo a seguir, apresentaremos, procedimentos indicados para a eliminação de parâmetros incômodos e testes de hipótese para pequenas amostras.



3. TESTE EXATO PARA PEQUENAS AMOSTRAS

Fisher (1925), no prefácio do seu livro *Statistical Methods for Research Workers*, já fazia alusões ao fato de que os métodos estatísticos tradicionais não eram adequados para análise de pesquisas práticas.

"... the traditional machinery of statistical processes is wholly unsuited to the needs of practical research. Not only does it take a cannon to shoot a sparrow, but it misses the sparrow! The elaborate mechanism built on the theory of infinitely large samples is not accurate enough for simple laboratory data. Only by systematically tackling small sample problems on their merits does it seem possible to apply accurate tests to practical data."

Se queremos verificar a independência ou a homogeneidade entre variáveis aleatórias na situação de tamanhos amostrais pequenos, em tabelas de duas entradas e, particularmente, em tabelas 2x2, as alternativas são os testes exatos.

Neles utilizaremos a distribuição exata ao invés de aproximações assintóticas. No entanto, como já foi visto, o cálculo das distribuições exatas, às vezes, é dificultado pela existência de parâmetros incômodos.

Em análise de tabelas 2x2, a escolha da função de verossimilhança não condicional ou da função de verossimilhança condicional é o primeiro passo para se definir qual a técnica apropriada a ser utilizada.

Após uma série de estudos, Fisher concluiu que condicionar em uma *estatística ancilar* era fundamental e valioso para a teoria dos testes de significância e, seguindo esta teoria, a função de verossimilhança condicional tem sido a preferida pelos especialistas.

Como o *princípio de Condicionalidade* requer utilização de *estatísticas ancilares* na sua aplicação, vamos apresentar os conceitos de estatística ancilar e o princípio de condicionalidade.

3.1 Estatística Ancilar

Inicialmente vamos definir estatística suficiente e estatística suficiente mínima.

Definição 1: Seja X_n um vetor amostral com densidade conjunta $p(x;\theta)$. Uma estatística $S = s(X_n)$ é uma *estatística suficiente* para θ se, e somente se, a distribuição condicional de T dada S não depende de θ para qualquer $T = t(X_n)$.



Definição 2 : Um conjunto de estatísticas conjuntamente suficientes é definido como *suficiente mínima* se, e somente se, ele é uma função de qualquer outro conjunto de estatística suficiente.

A definição de estatística ancilar aqui apresentada utiliza as hipóteses sugeridas por Cox e Hinkley (1974). Nela está imposta uma estatística suficiente mínima, enquanto que, na definição estabelecida inicialmente por Fisher, não existia esta hipótese. O por quê desta opção será exemplificado adiante.

Seja S uma estatística suficiente mínima para um parâmetro θ em uma família de distribuição uniparamétrica indexada por $\theta \in \Theta$.

Algumas vezes, podemos escrever $S = (T, C)$ onde C tem uma distribuição marginal não dependendo de θ e, então, C é denominada *estatística ancilar*.

A estatística T é denominada *condicionalmente suficiente*, pois ela é usada em inferência condicional com $C = c$.

Conseqüentemente, se C é uma estatística ancilar para um parâmetro θ , a distribuição de C não pode ser uma função de θ .

O seguinte exemplo é dado em Cox -Hinkley (1974 - pág 32)

Exemplo 1: Tamanho de uma amostra aleatória

Seja N uma variável aleatória, com uma distribuição conhecida

$p_n = P(N = n)$ ($n=1,2,3,\dots$) e seja Y_N um vetor amostral, com a densidade conjunta pertencente à família exponencial; ou seja:

$$f_Y(y, \theta) = e^{a(\theta) \sum b_j(y_j) + \sum c_j(\theta) + \sum d_j(y_j)}$$

Portanto, a função de verossimilhança dos dados $(n, y_1, y_2, y_3, \dots, y_n)$ é dada pela expressão:

$$f_{N,Y}(n, y) = p_n e^{a(\theta) \sum_{j=1}^n b(y_j) + nc(\theta) + \sum_{j=1}^n d(y_j)}$$

Deste modo observa-se que $\left\{ \sum_{j=1}^n b(y_j), N \right\}$ é suficiente para θ e N é uma *estatística ancilar*.

Qualquer tamanho amostral não fixado a priori, mas com distribuição conhecida independente de θ , é uma *estatística ancilar*.



Logo o valor observado do N , por si só, não fornece nenhuma informação sobre θ ; mas, se utilizarmos a estatística N no princípio de condicionalidade, poderemos obter informações sobre este parâmetro.

3.2 Estatística Ancilar na presença de parâmetros incômodos

Segundo Basu (1977), modelos estatísticos com a família de distribuição uniparamétricas são casos raros. Geralmente, nós temos que trabalhar com famílias multiparamétricas.

Se o problema requer inferência relacionada a um dos parâmetros, os outros são chamados de *parâmetros incômodos* ("nuisance parameter"). Vamos, pois, estender a definição de estatística ancilar na presença de parâmetros incômodos.

Suponhamos que o parâmetro desconhecido θ é particionado em $\theta = (\psi, \lambda)$ onde λ é um parâmetro incômodo.

O espaço paramétrico é $\Theta = \Psi \times \Lambda$.

Seja S uma estatística suficiente mínima para θ e suponhamos $S = (T, C)$, onde :

- i) a distribuição de C depende de λ , mas não depende de ψ ,
- ii) a distribuição condicional de T dado $C = c$ depende de ψ , mas não depende de λ (para todo valor de c).

Nestas condições, a estatística C será chamada de *ancilar* para ψ e T *condicionalmente suficiente* para ψ , na presença de um parâmetro incômodo λ .

3.3 Estatística Aproximadamente Ancilar

Uma estatística é considerada *aproximadamente ancilar* se a sua distribuição sofre apenas pequenas alterações quando o parâmetro de interesse percorre o seu espaço de variação.

Para ilustrar, consideremos o seguinte exemplo, dado em Cox e Hinkley (1974 -pág 34):



Exemplo 2: Suponhamos que $(Y_1, Z_1), (Y_2, Z_2), \dots, (Y_n, Z_n)$ sejam independentes e identicamente distribuídas, em uma distribuição normal bivariada com média zero, variância unitária e coeficiente de correlação ρ desconhecido.

A densidade conjunta é

$$f(y_j, z_j) = \frac{1}{(2\pi)^n (1 - \rho^2)^{n/2}} \exp \left\{ -\frac{\sum (y_j^2 + z_j^2)}{2(1 - \rho^2)} + \frac{\rho \sum y_j z_j}{1 - \rho^2} \right\}$$

A estatística $S = \left\{ \sum Y_j Z_j; \sum (Y_j^2 + Z_j^2) \right\}$ é suficiente mínima.

A estatística $C' = \sum (Y_j^2 + Z_j^2)$ é *aproximadamente ancilar*, pois :

$E(C') = 2n$ é independente de ρ e

$\text{Var}(C') = 4n(1 + \rho^2)$ também não depende fortemente de ρ , com exceção das situações em que existe forte correlação entre as variáveis, ou seja, nas vizinhanças de $+1$ ou -1 .

3-4 Princípio de Condicionalidade

O princípio de condicionalidade, aqui descrito, foi sugerido por Fisher (1956), para inferências estatísticas, principalmente para testes de hipóteses.

Seja $p(x; \theta)$ a densidade para a variável X com parâmetro θ onde $\theta \in \Theta = \Psi \times \Lambda$.

Seja $S = (T, C)$ uma estatística suficiente mínima para os parâmetros $\theta = (\psi, \lambda)$.

Suponhamos que C tenha uma distribuição livre de ψ , isto é, que ela seja ancilar para ψ .

Seja a fatoração da função de verossimilhança

$$p(x; \theta) = g(T, C; \psi, \lambda) h(x) = g_1(T|C, \psi, \lambda) g_2(C; \lambda) h(x). \quad (3.1)$$

De acordo com Fisher temos :

“ Se C é distribuída independente de ψ , então a distribuição condicional de $T|C$ é tudo o que se precisa considerar para fazer inferência sobre ψ ”

Particularmente, se C é suficiente para λ quando ψ é conhecido, então

$$p(x; \theta) = g_1(T|C; \psi) g_2(C; \lambda) h(x) \quad (3.2)$$

e cada uma das distribuições depende de parâmetros distintos.



Assim, para inferirmos sobre ψ , a função g_1 é tudo que precisamos analisar.

Portanto o princípio de condicionalidade estabelece que as conclusões sobre o parâmetro de interesse podem ser tiradas quando C for fixado em um valor observado c .

A estatística C , entretanto, tem a sua distribuição marginal livre de ψ . Mas, ela não é, necessariamente, suficiente para λ , e isto levou muitos estatísticos a pesquisarem como eliminar estes parâmetros incômodos.

Welch (1939) demonstrou que o teste condicional baseado em g_1 pode ser mais poderoso que um teste incondicional.

O exemplo abaixo, dado em Cox-Hinkley (1974-pág 38), ilustra a importância do princípio de condicionalidade.

Exemplo 3 : Dois instrumentos de medidas

Dois instrumentos, C_1 e C_2 , podem ser utilizados para efetuar uma medida com o objetivo de determinar se um parâmetro físico θ é igual a θ_1 ou θ_2 .

Y é a variável aleatória que representa a medida, $Y = 1, 2$.

Sabe-se também que:

$$P(Y = 1|C_1; \theta_2) = P(Y = 2|C_1; \theta_1) = 1$$

$$P(Y = 1|C_2; \theta_2) = P(Y = 2|C_2; \theta_1) = 0,01$$

$$P(\text{selecionar } C_1) = 0,9$$

$$P(\text{selecionar } C_2) = 0,1$$

No experimento, escolhe-se, aleatoriamente, um instrumento e efetua-se uma medida y .

Acima, nós sabíamos qual instrumento havia sido usado.

Supondo, agora, que $y=1$ e que C_2 tenha sido usada.

Considerando-se, coerentemente com a teoria bayesiana, que θ é uma variável aleatória que pode assumir os valores θ_1 ou θ_2 , então temos

$$P(Y = 1; \theta_1) = P(Y = 1 \wedge C_2; \theta_1) = P(C_2) \cdot P(Y = 1|C_2; \theta_1) = 0,1 \times 0,99 = 0,099$$

$$P(Y = 1; \theta_2) = 0,901$$

o qual sugere que $\theta = \theta_2$.

No entanto as probabilidades condicionais em C_2 sugeriam que $\theta = \theta_1$.

Teremos, pois, conclusões conflitantes sobre θ , se nós não pudermos fazer o condicionamento sobre a informação de que C_2 foi usada.

Observe que C_2 é uma *estatística ancilar* para o parâmetro θ .



Vamos expor a seguir um outro princípio de condicionalidade onde a estatística T utilizada não seja ancilar para o parâmetro de interesse mas, a distribuição condicional da variável X , dada T não dependa do parâmetro incômodo.

Consideremos, agora, uma variável aleatória X com densidade $p(x; \psi, \lambda)$ (ψ e λ parâmetros reais) e uma estatística $T = t(X)$. Então, podemos escrever:

$$p(x; \psi, \lambda) = g_1(x|t; \psi, \lambda)g_2(t; \psi, \lambda) \quad (3.3)$$

Se a distribuição condicional de X , dado $T = t$, não depender de ψ , teremos:

$$p(x; \psi, \lambda) = g_1(x|t; \lambda)g_2(t; \psi, \lambda) \quad (3.4)$$

e, assim, para fazermos inferências sobre λ , sendo ψ o parâmetro incômodo, devemos considerar apenas g_1 e ignorar g_2 ; desta forma, estaremos eliminando o parâmetro incômodo existente na situação em estudo.

O fator g_1 é chamado de *verossimilhança parcial*.

De acordo com Pereira e Lindley (1987) este método tem sido muito utilizado, mas uma questão relacionada a ele é como justificar o descarte do outro fator, principalmente quando o parâmetro incômodo é complicador na inferência e queremos eliminá-lo totalmente.

Basu (1977) sugere uma coletânea de métodos para eliminarmos estes parâmetros incômodos, destacando o método da Marginalização e o método de Condicionamento.

O método de Condicionamento, no qual estamos interessados, consiste em escolhermos a estatística T de modo que a distribuição condicional da amostra, dado $T = t$, dependa apenas do parâmetro de interesse. Mas não é suficiente optar por alguma estatística que faça o trabalho de eliminação; é necessário que ela seja de tal modo que, se possível, não se percam informações sobre o parâmetro de interesse.

De acordo com o Princípio de Condicionalidade proposto por Fisher, e aqui já descrito, esta estatística deve ser ancilar.

Deparamos, agora, com a questão das dificuldades relacionadas a uma estatística ancilar.

Primeiro, pelo fato de que não há um método geral para se construir esta estatística; segundo, porque ela pode não ser única; e, terceiro, porque pode ter ocorrido perda no número de informação de Fisher.

O seguinte exemplo é dado por Basu (1964), para ilustrar uma destas dificuldades.



Exemplo 4: Distribuição Multinomial Especial

Seja $Y_1, Y_2, Y_3, \dots, Y_n$ uma amostra aleatória com a seguinte distribuição discreta

$$p(Y_j = 1) = \frac{1}{6}(1 - \theta)$$

$$p(Y_j = 2) = \frac{1}{6}(1 + \theta)$$

$$p(Y_j = 3) = \frac{1}{6}(2 - \theta)$$

$$p(Y_j = 4) = \frac{1}{6}(2 + \theta)$$

com $j = 1, 2, \dots, n$.

Se n_l é o número de observações iguais a l ($l=1,2,3,4$), então a probabilidade conjunta de uma sequência particular é :

$$f_Y(y|\theta) = 6^{-n}(1 - \theta)^{n_1}(1 + \theta)^{n_2}(2 - \theta)^{n_3}(2 + \theta)^{n_4}.$$

A estatística $S = (N_1, N_2, N_3, N_4)$ é suficiente e mínima $\left(\sum_l N_l = n \right)$.

As estatísticas $T_1 = (N_1 + N_2, N_3 + N_4)$ e $T_2 = (N_1 + N_4, N_2 + N_3)$ são, ambas, ancilares para θ .

A questão, agora, é: qual das duas é melhor para se usar em inferências sobre θ ?

A definição de estatística ancilar aqui apresentada utiliza uma estatística suficiente mínima. Um conceito mais amplo, que foi apresentado por Fisher, define como ancilar qualquer estatística com a distribuição independente de θ .

Segundo Cox e Hinkley (1974), se esta definição for adotada, acentua-se o problema da não unicidade e, conseqüentemente aumentam as dificuldades relacionadas à opção sobre qual é a melhor para utilizarmos no princípio de condicionalidade. Se retornarmos ao *exemplo 2*, caso adotemos esta definição, podemos observar duas estatísticas ancilares: $\sum Y_j^2$ e $\sum Z_j^2$ e, não existe um critério razoável para justificar uma escolha entre as duas.



3.5 Teste Exato de Fisher

Como já citada na página III, a aproximação da distribuição da estatística X^2 com a distribuição χ^2 é lenta em n , se o valor esperado de alguma cela for pequeno. Como proposta alternativa, podemos usar, nestas situações, o teste exato de Fisher.

Vamos, então, descrever este teste.

Consideremos a tabela abaixo, com a suposição de que ela representa duas variáveis dicotômicas, tanto nas linhas como nas colunas.

	1	0	
X	n_{11}	n_{12}	$n_{1.}$
Y	n_{21}	n_{22}	$n_{2.}$
	$n_{.1}$	$n_{.2}$	n

1: sucesso

0: fracasso

Consideremos também X e Y variáveis aleatórias com distribuições binomiais

$$X \sim B(n_{1.}, p_{11}) \text{ e } Y \sim B(n_{2.}, p_{21}).$$

De acordo com as hipóteses estabelecidas para a construção do teste, a inferência deve ser feita utilizando a função de verossimilhança condicional, proveniente da distribuição que obtemos assumindo os *totais marginais fixos*.

O Princípio de Condicionalidade será utilizado com a estatística $X+Y=n_{.1}$ e, de acordo com Yates (1984), esta estatística é aproximadamente ancilar.

Vamos obter a verossimilhança condicional para os dados da tabela.

$$\begin{aligned}
 P[X = n_{11} | n_{1.}, n_{2.}, n_{.1}, n_{.2}, n] &= P[X = n_{11} | X + Y = n_{.1}] = \frac{P[X = n_{11}, X + Y = n_{.1}]}{P[X + Y = n_{.1}]} = \\
 &= \frac{P[X = n_{11}, Y = n_{.1} - n_{11}]}{P[X + Y = n_{.1}]} = \frac{\binom{n_{1.}}{n_{11}} p_{11}^{n_{11}} (1 - p_{11})^{n_{1.} - n_{11}} \binom{n_{2.}}{n_{21}} p_{21}^{n_{21}} (1 - p_{21})^{n_{2.} - n_{21}}}{\sum_j \binom{n_{1.}}{j} \binom{n_{2.}}{n_{.1} - j} p_{11}^j (1 - p_{11})^{n_{1.} - j} p_{21}^{n_{.1} - j} (1 - p_{21})^{n_{2.} - n_{.1} + j}}
 \end{aligned}$$

com j variando do Máx $\{0, n_{.1} - n_{2.}\}$ ao Mín $\{n_{.1}, n_{1.}\}$.

Após algumas simplificações, temos:

$$P[X = n_{11} | X + Y = n_{.1}] = \frac{\binom{n_{1.}}{n_{11}} \binom{n_{2.}}{n_{.1} - n_{11}} p^{n_{11}}}{\sum_j \binom{n_{1.}}{j} \binom{n_{2.}}{n_{.1} - j} p^j}, \quad (3.5)$$

onde $p = \frac{p_{11}(1 - p_{21})}{p_{21}(1 - p_{11})}$.

Se quisermos testar a homogeneidade, ou seja,

$$H_0: p_{11} = p_{21}$$

contra uma hipótese alternativa unilateral como por exemplo $H_1: p_{11} < p_{21}$ então, sob H_0 , teremos $p = 1$ e a distribuição da variável aleatória X será

$$P[X = n_{11} | X + Y = n_{.1}] = \frac{\binom{n_{1.}}{n_{11}} \binom{n_{2.}}{n_{.1} - n_{11}}}{\binom{n}{n_{.1}}} = \frac{n_{1.}! n_{.1}! n_{2.}! n_{.2}!}{n! n_{11}! n_{12}! n_{21}! n_{22}!}, \quad (3.6)$$

isto é, será uma distribuição Hipergeométrica (livre de parâmetros incômodos).

Assumindo que H_0 é verdadeira, podemos calcular a probabilidade de se observar um determinado conjunto de freqüências em uma tabela 2×2 .

O teste exato de Fisher consiste em adicionar o valor da probabilidade de ocorrência da freqüência observada aos valores das probabilidades das ocorrências ainda mais extremas que sejam coerentes com a hipótese alternativa. A este valor obtido chamaremos *valor p* e denotaremos por p .

A soma destas probabilidades é, então, comparada com o nível de significância α escolhido.

Se $p \leq \alpha$, rejeitamos a hipótese de homogeneidade; e, portanto, concluímos que há associação significativa entre as variáveis em estudo.

O teste exato de Fisher é unilateral, ao contrário do teste Quiquadrado que é bilateral, visto que a soma das probabilidades é feita, de forma consistente, com a hipótese alternativa.

Yates (1984), em seu artigo *Tests of Significance for 2 x 2 Contingency Tables* apresenta-nos as seguintes conclusões sobre o teste exato de Fisher :

1) Apesar de várias vezes constatararmos que o teste exato de Fisher, baseado em condicionamento dos valores marginais, é muito conservativo, ele aparenta ser o único teste



racional, com duas, uma ou nenhuma das marginais determinadas a priori. Os valores marginais determinam a sensibilidade do teste.

Tocher (1950) demonstrou que este teste é adequado para as situações de uma marginal fixa e para os casos de somente o total amostral ser fixado.

2) A correção de continuidade para o χ^2 dá uma boa aproximação com o teste exato, com exceção dos casos em que a distribuição hipergeométrica não seja simétrica.

3) Em geral, probabilidades unicaudais são usadas, mas se a probabilidade bicaudal for necessária, a melhor convenção é duplicar o valor p obtido na probabilidade unicaudal.

4) A responsabilidade do estatístico não termina com o cálculo da probabilidade. Ele precisa, também, comentar o real valor p e suas prováveis implicações.

Para melhor entender o teste, apresentaremos a situação dada por Everitt(1977):

Exemplo 5: De um modo geral, os pacientes psiquiátricos podem ser classificados em psicóticos ou neuróticos. Vinte (20) pacientes de cada um destes grupos serão analisados por um psiquiatra para constatar a presença ou a ausência do seguinte traço de comportamento : tendência suicida.

O objetivo é verificar se existe associação entre os dois grupos psiquiátricos e a ausência ou a presença de tendência suicida.

A tabela seguinte apresenta os dados obtidos:

tabela a

	tendência suicida		total
	sim	não	
psicóticos	2	18	20
neuróticos	6	14	20
total	8	32	40

A hipótese é de que a proporção de psicóticos com tendência suicida é igual à proporção de neuróticos com este traço de comportamento, ou seja, se p_p e p_n forem as verdadeiras proporções de tendência suicida entre os psicóticos e neuróticos, respectivamente, então as hipóteses serão:

$$H_0 : p_p = p_n \quad \times \quad H_1 : p_p < p_n .$$



Se estimarmos os valores esperados para as celas, constataremos que $e_{11} = \frac{20 \times 8}{40} = 4$; como este valor é menor que 5, indica-se o teste exato de Fisher como mais apropriado que o teste quiquadrado para se verificar a possível associação entre as variáveis estudadas.

Temos, então, duas tabelas de ocorrências mais extremas que a observada:

tabela b

	tendência 1	suicida 0	total
psicóticos	1	19	20
neuróticos	7	13	20
total	8	32	40

tabela c

	tendência 1	suicida 0	total
psicóticos	0	20	20
neuróticos	8	12	20
total	8	32	40

Calculando as probabilidades de cada tabela através da fórmula (3.6), teremos:

$$p_a = \frac{8!32!20!20!}{2!6!18!14!40!} = 0,095760$$

$$p_b = \frac{8!32!20!20!}{1!7!19!13!40!} = 0,020160$$

$$p_c = \frac{8!32!20!20!}{0!8!20!12!40!} = 0,001638$$

portanto o valor p é igual a:

$$p = p_a + p_b + p_c = 0,117558$$

Se considerarmos o nível de significância $\alpha = 0,05$, então $p > \alpha$ e, portanto, os dados não evidenciam que psicóticos e neuróticos diferem quanto ao traço de comportamento tendência suicida.



Se quisermos um teste bilateral, teremos $2p=0,235116$.

3.5.1 Formas para cálculo do valor p

i) Fórmula recursiva

Para explicar esta fórmula, utilizaremos a seguinte tabela:

tabela i

a_i	b_i	a_i+b_i
c_i	d_i	c_i+d_i
a_i+c_i	b_i+d_i	n

onde, para $i = 1$, temos a tabela dos dados observados.

A probabilidade da tabela observada será denotado por p_1 e obtido através da expressão (3.6).

Feldman e Klinger (1963) deduziram a seguinte fórmula recursiva para obter os valores das probabilidades das tabelas mais extremas, que são usadas no teste exato de Fisher:

$$p_{i+1} = \frac{a_i d_i}{b_{i+1} c_{i+1}} p_i \quad (3.7)$$

onde : p_i é a probabilidade da tabela i,

a_i, d_i são os valores na tabela i ,

b_{i+1}, c_{i+1} são os valores na tabela (i + 1),

i vai assumir valores compatíveis com a hipótese alternativa do teste.

Sachs (1984) ilustra a fórmula com as seguintes tabelas onde, a tabela 1 será considerada a observada e, as demais, as tabelas mais extremas.

tabela 1

2	8	10
10	4	14
12	12	24



tabela 2

1	9	10
11	3	14
12	12	24

tabela 3

0	10	10
12	2	14
12	12	24

Calculando inicialmente p_1 , que é a probabilidade da tabela observada, obtém-se:

$$p_1 = \frac{10!14!12!12!}{24!2!8!10!4!} = 0,016659$$

Utilizando a expressão (3.7) vamos calcular as probabilidades das tabelas mais extremas.

$$\text{Se } i = 1, \text{ então } p_{i+1} = p_2 = \frac{2 \times 4}{9 \times 11} p_1 = 0,001346.$$

$$\text{Se } i = 2, \text{ então } p_3 = \frac{1 \times 3}{12 \times 10} p_2 = 0,000034$$

portanto, o *valor p* para o teste exato de Fisher será igual a:

$$p = p_1 + p_2 + p_3 = 0,018039.$$

Algumas restrições devem ser consideradas sobre a fórmula dada acima.

A fórmula é dita como recursiva mas depende da construção prévia das tabelas que devem ser utilizadas para o cálculo do *valor p* no teste exato de Fisher.

Se construirmos a seguinte expressão:

$$p_{i+1} = \frac{a_i d_i}{(b_i + 1)(c_i + 1)} p_i$$

onde p_i é a probabilidade da tabela observada e obtida por (3.6).

Desta maneira teremos uma fórmula de recorrência onde só é necessário estabelecermos o valor final para i .

Este valor será estabelecido pelos totais marginais fixados.



ii) Tabelas e Processos Computacionais

Para calcular o *valor p* existem tabelas, tais como as dadas em *Biometrika Tables for Statisticians* (1967) e, também, em Finney (1963), embora, hoje, isto esteja ultrapassado, pois muitos programas computacionais estatísticos já trazem o teste exato de Fisher.

Alguns pacotes estatísticos que efetuam este teste são o SAS, através de PROC FREQ, o BMDP, através do 4F, o SPSS^x, o StatXact e o EPI Info.

Entretanto, qualquer pacote estatístico com possibilidades de programação pode ser utilizado para o cálculo do *valor p*.

Se necessitamos dos valores das probabilidades, temos também tabelas de fatoriais; mas isso, atualmente, é desnecessário, devido ao poder de cálculo das calculadoras científicas.

3.5.2 Teste de Fisher com adaptação de Tocher

Para pequenas amostras, a distribuição exata (3.6) é muito discretizada, visto que n_{11} assume poucos valores. Portanto é difícil construir um teste exato de tamanho α .

Tocher (1950) mostrou que o teste de Fisher, para o caso unilateral com uma extensão que o transforma em um teste aleatorizado, é uniformemente mais poderoso, não viesado.

A modificação proposta por Tocher (1950) consiste em determinar, primeiro, a probabilidade de todos os casos mais extremos do que o observado.

Para explicar esta modificação utilizaremos as notações p_e e p_o , para indicarem a soma das probabilidades das tabelas mais extremas e a probabilidade da tabela observada, respectivamente.

Se p_e for superior ao nível fixado α , não podemos rejeitar H_0 .

Se esta soma for menor que α , enquanto que o *valor p* obtido pelo teste de Fisher for maior do que α , Tocher (1950) recomenda o cálculo da razão δ definida abaixo:..

$$\delta = \frac{\alpha - p_e}{p_o}$$

Em seguida extrai-se um número aleatoriamente entre 0 e 1 (t_0).

Se $\delta > t_0$, rejeitamos H_0

Se $\delta < t_0$, aceitamos H_0



Com esta probabilidade incluída, o teste de Fisher é um pouco menos conservativo e de acordo com Cochran (1952), com esta modificação o teste exato de Fisher é o mais poderoso de todos os testes unilaterais de tamanho α , para tabelas 2×2 .

Agresti (1990) descreve um modo de se alcançar um nível de significância fixado empregando uma aleatorização nas bordas da região crítica.

Vamos considerar o exemplo dos pacientes psiquiátricos.

Se efetuarmos a soma das probabilidades das tabelas mais extremas, ou seja, $p_b + p_c$, o valor obtido é menor que o nível de significância $\alpha = 0,05$ estabelecido no exemplo.

$$p_e = p_b + p_c = 0,0217$$

No entanto o *valor p* obtido no teste exato de Fisher foi 0,117558, que ultrapassa o valor de α .

Calculando o valor de δ temos:

$$\delta = \frac{\alpha - p_e}{p_o} = \frac{0,05 - 0,0217}{0,0958} = 0,2954$$

O teste para rejeitar H_0 com os valores descritos acima será:

Rejeitamos H_0 com probabilidade 1,0 quando $n_{11} = 0$ ou 1;

Rejeitamos H_0 com probabilidade 0,2954 quando $n_{11} = 2$;

Com probabilidade 0, rejeitamos H_0 para os outros casos.

Temos, então:

$$P(\text{Rejeitar } H_0) = E[P(\text{Rejeitar } H_0 \mid n_{11})] =$$

$$= 1,0(0,0217) + 0,2954(0,0958) + 0,0 P(n_{11} \geq 3) = 0,05$$

De acordo com Agresti (1990), na prática, testes aleatorizados são raramente justificados e recomenda-se simplesmente recorrer ao *valor p*

3.5.3 A função poder condicional do teste exato de Fisher

Sob a hipótese alternativa H_1 considerada no teste exato de Fisher, a distribuição condicional de n_{11} dado $n_{.1}$ é dada por



$$P[X = n_{11} | X + Y = n_{.1}] = \frac{\binom{n_{1.}}{n_{11}} \binom{n_{2.}}{n_{.1} - n_{11}} p^{n_{11}}}{\sum_j \binom{n_{1.}}{j} \binom{n_{2.}}{n_{.1} - j} p^j},$$

onde $p = \frac{p_{11}(1 - p_{21})}{p_{21}(1 - p_{11})}$.

O teste exato de Fisher estabelece a região crítica para um teste de nível α através de um valor n_{11} tal que

$$P[X \leq n_{11} | X + Y = n_{.1}; H_0] \leq \alpha \quad e \quad P[X \leq n_{11} + 1 | X + Y = n_{.1}; H_0] > \alpha$$

A função poder condicional para este teste é dada por

$$\beta(p | n_{.1}) = P(X \leq n_{11} | X + Y = n_{.1}; H_1) \quad (3.9)$$

que é uma razão de dois polinômios em p .

Diversos pesquisadores têm investigado o poder do teste de Fisher e têm mostrado que tamanhos amostrais grandes são necessários para detectar diferenças muito pequenas entre duas proporções populacionais.

Bennett e Hsu (1960) apresentam gráficos da função poder (3.9) para diversos tamanhos amostrais.

Para o exemplo dos pacientes psiquiátricos, onde os tamanhos amostrais são 20 neuróticos e 20 psicóticos, testando ao nível 0,05 e os valores das proporções de tendência suicida na população, em cada grupo, forem 0,5 e 0,2, respectivamente, então o poder tem um valor de 0,53. Gart e Gail (1973) mostram que uma amostra de 42 pacientes em cada grupo seria necessária para atingir um poder de 0,9 de obtenção de uma diferença nas proporções populacionais.

O poder esperado do teste exato de Fisher para um tamanho amostral n é dado por

$$\beta(p) = \sum_{n_{.1}} \beta(p | n_{.1}) p(n_{.1}) \quad (3.10)$$

onde $p(n_{.1})$ representa a distribuição de $n_{.1}$.

Fu e Arnold (1992), baseados nesta expressão, apresentam tabelas para a obtenção do tamanho amostral exato necessário para alcançar o poder médio de 0,50 e 0,90 em um teste com nível de significância de 1% ou 5%.



Sillitto (1949) apresentou uma aproximação para $\beta(p)$ com uma distribuição normal padrão; mas de acordo com Bennett e Hsu (1960), esta aproximação não é muito precisa para situações de n_1 e n_2 pequenos.

Bennett e Hsu (1960) fizeram algumas comparações entre $\beta(p)$ e a aproximação de Sillitto e obtiveram, por exemplo, que, para

$$n_1 = 20 \quad n_2 = 15 \quad p_{11} = 0,8 \quad p_{21} = 0,2 \quad \alpha = 0,05,$$

têm-se $\beta(p) = 0,9681$ e na aproximação de Sillitto foi de 0,983.

3.6 Considerações contra o teste de Fisher

Vos e Hudson (1955) colocam-se contrários ao teste exato de Fisher quando o objetivo é testar a igualdade de proporções em distribuições binomiais independentes. Seus argumentos estão relacionados ao fato de este teste utilizar o princípio de condicionalidade com a estatística 'número total de sucessos', que é considerada como aproximadamente ancilar, e às perdas ocasionadas se o interesse estiver no número de informações de Fisher.

Vamos então apresentar os argumentos que embasam estas afirmativas.

De acordo com Vos e Hudson (1995) não é claro o que significa uma estatística aproximadamente ancilar. Se esta definição não é clara, como determinar quando uma estatística aproximadamente ancilar é melhor que outra?

Se uma medida de informação desejada for o número de informação de Fisher, uma boa estatística será aquela em que o número de informação de Fisher, contido na função de verossimilhança dos dados, não sofra perda ou que esta perda seja pequena.

Andersen (1970) propôs um conceito de ancilaridade que ele definiu como ancilaridade fraca para diferenciar do conceito proposto por Fisher. Esta será de grande importância, quando a situação envolver inferências sobre um parâmetro na presença de outros parâmetros incômodos e, estivermos interessados em analisar se houve perdas na informação de Fisher.

Liang (1983), apresenta uma série de resultados sobre esta estatística que vamos expor a seguir:



3.7 Estatística Fracamente Ancilar

Seja $p(x; \theta)$ a densidade ou a probabilidade no caso de uma distribuição discreta da variável X com parâmetro $\theta = (\psi, \lambda)$ com $\psi \in \Psi$ e $\lambda \in \Lambda$, Ψ e Λ intervalos da reta real.

Seja uma estatística T tal que a distribuição condicional de X dada T dependa de θ somente através de ψ , ou seja,

$$p(x; \theta) = f(x|t; \psi)g(t; \theta), \quad (3.11)$$

onde g é a densidade marginal de T .

Definição 3: Uma estatística T é chamada *fracamente ancilar* se para algum dado par $(\psi_0, \lambda_0) \in \Psi \times \Lambda$, existe uma função $\lambda = l(\psi)$, tal que

$$g\{t; (\psi_0, \lambda_0)\} \equiv g\{t; \{\psi, l(\psi)\}\} \quad (3.12)$$

para todo t e $\forall \psi \in \Psi$.

Exemplo 6: Seja $X = (X_1, X_2)$ um vetor de variáveis aleatórias independentes tais que

$$X_1 \sim N(\psi + \lambda, 1)$$

$$X_2 \sim N(\lambda, 1)$$

Então, $T = X_1 + X_2$ é também distribuída normalmente

$$T \sim N(\psi + 2\lambda, 2).$$

A distribuição conjunta de X_1 e X_2 é:

$$p(x; (\psi, \lambda)) = \frac{1}{2\pi} e^{-\frac{1}{2}[(x_1 - \psi - \lambda)^2 + (x_2 - \lambda)^2]}$$

A densidade marginal de T é:

$$g(t; (\psi, \lambda)) = \frac{1}{2\sqrt{\pi}} e^{-\frac{1}{2} \frac{(t - \psi - 2\lambda)^2}{2}}$$



e a distribuição condicional de X dado $T = t$ é:

$$f(x|t; \psi) = \frac{1}{\sqrt{\pi}} e^{-\frac{(x_1 - x_2 - \psi)^2}{2}} ;$$

logo a estatística T satisfaz (3.11) e para qualquer $(\psi_0, \lambda_0) \in R^2$ se fizermos

$$\begin{aligned} g\{t; (\psi_0, \lambda_0)\} &= \frac{1}{2\sqrt{\pi}} e^{-\frac{(t - \psi_0 - 2\lambda_0)^2}{2}} = \frac{1}{2\sqrt{\pi}} e^{-\frac{(t - \psi + \psi - \psi_0 - 2\lambda_0)^2}{2}} = \\ &= \frac{1}{2\sqrt{\pi}} e^{-\frac{\left(t - \psi - 2\left(-\frac{\psi}{2} + \frac{\psi_0}{2} + \lambda_0\right)\right)^2}{2}} \end{aligned}$$

Portanto para $\lambda = l(\psi) = -\frac{\psi}{2} + \frac{\psi_0}{2} + \lambda_0$, teremos

$$g\{t; (\psi_0, \lambda_0)\} \equiv g\{t; \{\psi, l(\psi)\}\}.$$

A estatística T é então *fracamente ancilar* para ψ .

De acordo com Andersen (1973), o estimador de máxima verossimilhança condicional para ψ , baseado em $f(x|t; \psi)$, coincide com o estimador de máxima verossimilhança incondicional se a estatística T for fracamente ancilar.

Basawa (1981) efetuou um estudo em famílias exponenciais e concluiu que o estimador de máxima verossimilhança condicional de um parâmetro em uma família exponencial bi-paramétrica é eficiente no sentido de satisfazer o limite de Cramér-Rao para a variância assintótica, se a estatística condicionante satisfizer a seguinte:

Condição I : Existe uma reparametrização 1 - 1 entre (ψ, λ) e (ψ, δ) , tal que $g(t; (\psi, \lambda))$ em (3.11) dependa de (ψ, λ) , somente através de δ .

Logo, de acordo com Basawa, devemos fazer inferência para ψ baseados na distribuição condicional $f(x|t; \psi)$, onde T deve ser fracamente ancilar (apesar de este termo não estar citado em seu artigo).



3.8 A Informação de Fisher em uma distribuição

A informação de Fisher para ψ na distribuição $p(x; \psi)$, sem a presença de parâmetros incômodos, é definida por:

$$I_p(\psi) = E \left[\frac{\partial \log p(x; \psi)}{\partial \psi} \right]^2 \quad (3.13)$$

Quando um parâmetro incômodo λ está presente, uma extensão da definição é dada por Efron (1977);

Definição : A informação de Fisher $I_p(\theta)$ sobre ψ , na distribuição $p(x; \theta)$, $\theta = (\psi, \lambda)$, é dada por:

$$I_p(\theta) = \inf_c E \left(\frac{\partial \log p}{\partial \psi} - c \frac{\partial \log p}{\partial \lambda} \right)^2 \quad (3.14)$$

onde $p = p(x; \theta)$.

Calculemos o valor de c :

$$f(c) = E \left[\frac{\partial \log p}{\partial \psi} \right]^2 + c^2 E \left[\frac{\partial \log p}{\partial \lambda} \right]^2 - 2c E \left[\frac{\partial \log p}{\partial \psi} \frac{\partial \log p}{\partial \lambda} \right]$$

$$f'(c) = 2c E \left[\frac{\partial \log p}{\partial \lambda} \right]^2 - 2 E \left[\frac{\partial \log p}{\partial \psi} \frac{\partial \log p}{\partial \lambda} \right]$$

$$f'(c) = 0 \Rightarrow 2c E \left[\frac{\partial \log p}{\partial \lambda} \right]^2 = 2 E \left[\frac{\partial \log p}{\partial \psi} \frac{\partial \log p}{\partial \lambda} \right].$$

$$\text{Denotando } \begin{cases} I_{11} = E \left[\frac{\partial \log p}{\partial \psi} \right]^2 \\ I_{22} = E \left[\frac{\partial \log p}{\partial \lambda} \right]^2 \\ I_{12} = E \left[\frac{\partial \log p}{\partial \psi} \frac{\partial \log p}{\partial \lambda} \right] \end{cases}, \text{ teremos, então: } c = \frac{I_{12}}{I_{22}}$$

Logo ,

$$I_p(\theta) = I_{11} + \left(\frac{I_{12}}{I_{22}} \right)^2 \cdot I_{22} - 2 \frac{I_{12}}{I_{22}} \cdot I_{12}$$

$$I_p(\theta) = I_{11} + \frac{I_{12}^2}{I_{22}} - 2 \frac{I_{12}^2}{I_{22}}$$

$$I_p(\theta) = I_{11} - \frac{I_{12}^2}{I_{22}} \quad (3.15)$$

O inverso de $I_p(\theta)$ é o limite inferior de Cramér - Rao para a variância de um estimador não viesado de ψ na presença de λ .

Vamos denotar por $I_f(\theta)$ e $I_g(\theta)$ as informações de Fisher sobre ψ nas distribuições $f(x|t; \psi)$ e $g(t; \theta)$, respectivamente.

Como $f(x|t; \psi)$ depende de θ somente através de ψ , então :

$$I_f(\theta) = E \left(\frac{\partial \log f}{\partial \psi} \right)^2.$$

$$\text{Note que } E \left(\frac{\partial \log f}{\partial \psi} \right)^2 = E \left[E \left\{ \left(\frac{\partial \log f}{\partial \psi} \right)^2 \middle| T \right\} \right], \quad (3.16)$$

onde, para cada $T = t$ fixado, a esperança condicional $E \left\{ \left(\frac{\partial \log f}{\partial \psi} \right)^2 \middle| t \right\}$ é a informação de Fisher para ψ na distribuição condicional $f(x|t; \psi)$.

Portanto, $I_f(\theta)$ pode ser vista como a média da “ informação de Fisher condicional

”.

Além disto, $I_g(\theta) = 0$ se, e somente se, existir uma função $c(\psi, \lambda)$ tal que

$$\frac{\partial \log g(t; \theta)}{\partial \psi} = c(\psi, \lambda) \left\{ \frac{\partial \log g(t; \theta)}{\partial \lambda} \right\} \text{ para todo } t. \quad (3.17)$$

De acordo com Liang (1983), sob certas condições de regularidade que permitam a mudança dos limites e do sinal da integral, tem-se

$$I_p(\theta) = I_f(\theta) + I_g(\theta) \text{ para todo } t. \quad (3.18)$$

Logo, quando fazemos o condicionamento, resulta geralmente numa perda de informação relativa à já contida na verossimilhança total $p(x; \theta)$ e o resultado acima nos

mostra que a informação total $I_p(\theta)$ pode ser decomposta em duas partes, uma devido à verossimilhança condicional em t e outra devido à densidade da estatística condicionante.

Andersen (1970) obteve também alguns resultados sobre estatísticas fracamente ancilar em situações de famílias exponenciais. Esses lemas foram utilizados por Liang (1983) para identificar se uma estatística é fracamente ancilar, nesse caso. Estes resultados serão apresentados a seguir.

Consideremos $p(x; \theta)$ pertence à família exponencial, ou seja:

$$p(x; \theta) = k(\psi, \lambda) e^{s(x)\psi + t(x)\lambda + h(x)}, \quad (3.19)$$

onde s e t são funções reais.

De acordo com Lehmann (1959), a densidade marginal de $T = t(X)$ é

$$g(t; \theta) = k(\psi, \lambda) e^{t\lambda} r(t; \psi), \quad (3.20)$$

sendo $r(t; \psi)$ alguma função normalizadora.

Lema 1: A estatística condicionante T é fracamente ancilar para ψ se, e somente se,

$$\frac{\partial \log r(t; \psi)}{\partial \psi} = a(\psi)t + b(\psi), \quad (3.21)$$

onde a e b são funções unicamente de ψ .

Lema 2: A igualdade do lema 1 é válida se, e somente se, existe uma função $c(\psi, \lambda)$ de ψ e λ , tal que

$$\frac{\partial \log g(t; \theta)}{\partial \psi} = c(\psi, \lambda) \frac{\partial \log g(t; \theta)}{\partial \lambda} \text{ para todo } t \quad (3.22)$$

Liang (1983) relaciona as definições de Basawa (1981) e Andersen (1970) através do seguinte

Teorema : Supondo-se que $p(x; \theta)$ seja da família exponencial e que satisfaça (3.11) para alguma estatística $T = t(X)$, então T satisfaz a condição I se, e somente se, for fracamente ancilar.

Ainda segundo Liang (1983), temos o seguinte

Corolário : Sob as mesmas condições do teorema anterior, as seguintes afirmações são equivalentes :

- i) t é fracamente ancilar.
- ii) t satisfaz a condição I .
- iii) $\frac{\partial \log r(t; \psi)}{\partial \psi} = a(\psi)t + b(\psi)$, onde a e b são funções de ψ .
- iiii) $I_g(\theta) = 0$ para todo θ .

Então, em uma família exponencial, se queremos evitar uma possível perda de informação quando pretendemos fazer inferência sobre um parâmetro e estamos usando o princípio de condicionalidade, é necessário e suficiente que a estatística condicionante T satisfaça a condição I ou, equivalentemente, seja fracamente ancilar.

O corolário acima nos será de grande utilidade pois identificar se uma estatística é fracamente ancilar, pela definição, nem sempre é fácil. No entanto, utilizando o item iii), poderemos fazer esta classificação com mais facilidade.

Vejamos isto no exemplo dado por Liang (1983):

Exemplo 7: Seja $X = (X_1, X_2, \dots, X_n)$ um vetor de variáveis aleatórias independentes, com n fixo e com distribuições de Poisson, com parâmetros (ϕ, θ)

$$X_i \sim \wp(e^{\phi + \theta z_i}) \quad (i=1,2,3,\dots,n)$$

Seja a estatística $T = t(X) = \sum_i X_i$; portanto,

$$T \sim \wp\left(e^{\phi \sum_i e^{\theta z_i}}\right).$$

Calculada a distribuição condicional de X dado $T = t$, temos:

$$f(x|t; \theta) = t! \frac{e^{\theta \sum_i z_i x_i}}{\left(\sum_i e^{\theta z_i}\right)^t \prod x_i!} I\left(\sum_i X_i = 0,1,2,\dots,t\right)$$

A estatística condicionante T é tal que a distribuição condicional de X dado $T = t$ depende apenas de θ .

A distribuição marginal de T é :

$$g(t; \phi, \theta) = \frac{e^{-\phi \sum_i e^{\theta z_i}} \left(e^{\phi \sum_i e^{\theta z_i}} \right)^t}{t!} \quad I(t = 0, 1, 2, \dots)$$

Colocando g na forma (3.20), temos a seguinte função normalizadora

$$r(t; \theta) = \left(\sum_i e^{\theta z_i} \right)^t ;$$

e aplicando o item iii) do corolário, ficamos com

$$\log r(t; \theta) = t \left(\log \sum_i e^{\theta z_i} \right)$$

$$\frac{\partial \log r(t; \theta)}{\partial \theta} = t \cdot \frac{\partial \log \sum_i e^{\theta z_i}}{\partial \theta}$$

e, portanto, $a(\theta) = \frac{\partial \log \sum_i e^{\theta z_i}}{\partial \theta}$ e $b(\theta) = 0$.

Logo, T é fracamente ancilar para θ .

O nosso interesse específico está relacionado às tabelas de contingência 2×2 e particularmente à estatística $T = X_1 + X_2$, que é a utilizada como condicionante no teste exato de Fisher. De acordo com Vos e Hudson ao verificarmos a igualdade de proporções, em situações de duas distribuições binomiais independentes, haverá perdas no número de informação de Fisher se o teste escolhido for o teste exato de Fisher.

Analisemos, portanto, um exemplo nestas condições.

Exemplo 8: Seja a tabela de contingência 2×2 abaixo, com a suposição de que X_1 e X_2 tenham distribuições binomiais independentes:

	sucesso	fracasso	
X_1	x_1	$n_1 - x_1$	n_1
X_2	x_2	$n_2 - x_2$	n_2
	$x_1 + x_2$	$n - x_1 - x_2$	n

Seja p_i a probabilidade de sucesso de X_i , ($i=1,2$)

Sejam, também,

$$\theta = \log \frac{p_1(1-p_2)}{p_2(1-p_1)}, \text{ o logaritmo da "odds ratio" ;}$$

$$\phi = \log \frac{p_2}{1-p_2}, \text{ a transformação logito de } p_2 \text{ e}$$

$$T = X_1 + X_2, \text{ a estatística condicionante.}$$

A distribuição condicional de X dada T é :

$$f(x|t; \theta) = \frac{\binom{n_1}{x_1} \binom{n_2}{t-x_1} e^{\theta x_1}}{\sum_{u=0}^t \binom{n_1}{u} \binom{n_2}{t-u} e^{\theta u}}$$

e, neste exemplo, ao calcularmos a distribuição marginal de t obtemos a seguinte função

$$\text{normalizadora } r(t; \theta) = \left[\sum_{u=0}^t \binom{n_1}{u} \binom{n_2}{t-u} e^{\theta u} \right].$$

De acordo com Plackett (1977), a derivada do logaritmo desta expressão não é linear em termos de t . Conseqüentemente, T não é fracamente ancilar para θ .

Portanto alguma informação sobre θ está contida em $T = X_1 + X_2$ e, ao utilizarmos o teste exato de Fisher, neste caso específico, haverá perda de informações em relação às contidas na função de verossimilhança.

Liang, em sua dissertação de Ph.D (não publicada), afirma que a informação $I_g(\theta)$ contida em $T = X_1 + X_2$ é muito pequena comparada à contida na informação total $I_p(\theta)$, quando os tamanhos amostrais são grandes.

3.9 Análise Condicional para a "odds ratio"

3.9.1 Estimador condicional para "odds ratio"

No capítulo 2 apresentamos o estimador de verossimilhança incondicional da "odds ratio". No entanto, se estamos fazendo uma análise utilizando uma abordagem condicional,



nada mais lógico que efetuarmos esta estimativa através de um estimador obtido pela função de verossimilhança condicional.

As inferências exatas sobre a “odds ratio” serão baseadas na função de verossimilhança condicional apresentada em (3.5), onde o parâmetro p será aqui denotado por θ .

A expressão ficará então:

$$P[X = n_{11} | X + Y = n_{.1}; \theta] = \frac{\binom{n_{1.}}{n_{11}} \binom{n_{2.}}{n_{.1} - n_{11}} \theta^{n_{11}}}{\sum_j \binom{n_{1.}}{j} \binom{n_{2.}}{n_{.1} - j} \theta^j} \quad (3.23)$$

onde, $\theta = \frac{p_{11}(1-p_{21})}{p_{21}(1-p_{11})}$ e j varia do Máx $\{0, n_{.1} - n_{2.}\}$ ao Mín $\{n_{.1}, n_{1.}\}$.

O estimador de máxima verossimilhança condicional $\hat{\theta}_{cond}$ é o valor que maximiza (3.23) e será dado pela solução da equação

$$n_{11} = E[X | X + Y = n_{.1}; \theta] \quad (3.24)$$

3.9.2 Intervalo de confiança para “odds ratio”

O intervalo de confiança para θ é formado por todos os valores que são consistentes com os dados, no sentido de que $H_0: \theta = \theta_0$ não é rejeitada através do teste exato de Fisher ao nível α .

Os limites do intervalo (θ_l, θ_s) serão encontrados resolvendo-se as inequações

$$\frac{\alpha}{2} \leq \sum_{i \leq n_{11}} P[X = i | n_{1.}, n_{.1}, n_{2.}, n_{.2}; \theta_s] \quad (3.25)$$

$$\frac{\alpha}{2} \leq \sum_{i \geq n_{11}} P[X = i | n_{1.}, n_{.1}, n_{2.}, n_{.2}; \theta_l] \quad (3.26)$$

Para ilustrar, vamos considerar a seguinte tabela, com dados fictícios, apresentada por Breslow e Day (1980) :



			total
	2	1	3
	3	1	4
total	5	2	7

Considerando ambas as marginais fixadas, os possíveis valores que n_{11} pode assumir, consistente com as hipóteses do teste de Fisher, são $n_{11}=1, 2$ e 3 .

Para estes valores, as outras tabelas são :

			total
	1	2	3
	4	0	4
total	5	2	7

			total
	3	0	3
	2	2	4
total	5	2	7

A probabilidade para $n_{11} = 2$ será expressada por:

$$P[X = 2; \theta] = \frac{\binom{3}{2} \binom{4}{3} \theta^2}{\binom{3}{1} \binom{4}{4} \theta + \binom{3}{2} \binom{4}{3} \theta^2 + \binom{3}{3} \binom{4}{2} \theta^3} = \frac{4\theta}{1 + 4\theta + 2\theta^2}$$

Para $n_{11} = 1$ e $n_{11} = 3$ temos:

$$P[X = 1; \theta] = \frac{1}{1 + 4\theta + 2\theta^2}$$

$$P[X = 3; \theta] = \frac{2\theta^2}{1 + 4\theta + 2\theta^2}$$

A equação (3.24) será:

$$1 \cdot \frac{1}{1 + 4\theta + 2\theta^2} + 2 \cdot \frac{4\theta}{1 + 4\theta + 2\theta^2} + 3 \cdot \frac{2\theta^2}{1 + 4\theta + 2\theta^2} = n_{11}$$

$$\frac{1 + 8\theta + 6\theta^2}{1 + 4\theta + 2\theta^2} = 2 ,$$

cujas raízes são $\pm \sqrt{\frac{1}{2}}$. O valor para a “odds ratio” é um número real positivo; portanto, o

$$\text{valor de } \hat{\theta}_{cond} = \sqrt{\frac{1}{2}} \cong 0,707 .$$

Se fizermos esta estimativa através do estimador de máxima verossimilhança incondicional, teremos $\hat{\theta} = \frac{n_{11}n_{22}}{n_{21}n_{12}} = 0,667$, que difere ligeiramente do valor anterior.

As equações para obtermos os limites de um intervalo com 80% de confiança serão:

$$0,10 = \frac{1 + 4\theta_s}{1 + 4\theta_s + 2\theta_s^2}$$

$$0,10 = \frac{4\theta_l + 2\theta_l^2}{1 + 4\theta_l + 2\theta_l^2}$$

cujas soluções são $\theta_s = 18,25$ e $\theta_l = 0,0274$.

O intervalo obtido foi amplo devido ao fato de os tamanhos amostrais serem pequenos.

Estas equações, se os dados da tabela forem grandes, são polinomiais de alto grau e só podem ser resolvidas por métodos iterativos. Se todos os valores da tabela forem maiores do que 5, a procura da solução exata vai exigir um longo tempo computacional, visto que estas expressões envolvem somatórios e números combinatórios.

Devido a estas dificuldades, justifica-se a procura de aproximações para estes valores.

3.9.3 Aproximações

Cornfield (1956) e Hannan e Harkness (1963) demonstraram que, quando todas as frequências das quatro celas de uma tabela 2 x 2 são grandes, as probabilidades condicionais produzidas por (3.5) são aproximadamente iguais aos valores da densidade de uma distribuição normal com média $\mu(\theta)$ dada pela solução da equação:

$$\frac{\mu(\theta)(n_{2.} - n_{1.} + \mu(\theta))}{(n_{1.} - \mu(\theta))(n_{1.} - \mu(\theta))} = \theta \quad (3.27)$$

e variância dada por:



$$Var(\theta) = \left(\frac{1}{\mu(\theta)} + \frac{1}{n_{1.} - \mu(\theta)} + \frac{1}{n_{.1} - \mu(\theta)} + \frac{1}{n_{2.} - n_{.1} + \mu(\theta)} \right)^{-1} \quad (3.28)$$

3.9.4 Aproximação para o Estimador da "odds ratio"

Se substituirmos o lado direito da igualdade, na expressão (3.24), por $\mu(\theta)$, temos

$n_{11} = \mu(\theta)$ e obteremos a seguinte solução

$$\hat{\theta} = \frac{n_{11}n_{22}}{n_{21}n_{12}} \quad (3.29)$$

ou seja, a mesma expressão obtida na análise incondicional.

3.9.5 Aproximação dos limites do Intervalo de Confiança para "odds ratio"

Cornfield (1956), utilizando a aproximação normal da variável X , sugere que os limites do intervalo de confiança devem ser obtidos através das soluções das equações:

$$n_{11} - \mu(\theta_l) - \frac{1}{2} = Z_{\alpha/2} \sqrt{Var(\theta_l)} \quad e$$

(3.30)

$$n_{11} - \mu(\theta_s) + \frac{1}{2} = -Z_{\alpha/2} \sqrt{Var(\theta_s)} \quad (3.31)$$

De acordo com Cornfield (1956), estes resultados dão as melhores aproximações para os limites exatos dados em (3.25) e (3.26).

Segundo Soares e Colosimo (1995), não há mais vantagens em usar estas aproximações pois os seus cálculos também exigem programas computacionais e existem programas para computadores, de fácil localização, que fazem o cálculo da distribuição exata.



3.9.6 Teste de Hipótese

De acordo com Breslow e Day (1980), sob a hipótese de que $\theta = 1$, as equações (3.27) e (3.28) se simplificam e assumem as expressões:

$$\mu(1) = \frac{n_{.1}n_{1.}}{n}$$

$$Var(1) = \frac{n_{1.}n_{.1}n_{2.}n_{.2}}{n^3}$$

Portanto, sob esta hipótese nula, a distribuição de X pode ser aproximada pela distribuição $N(\mu(1), Var(1))$

Para testarmos a hipótese

$$H_0: p_{11} = p_{21} \quad \text{contra a alternativa} \quad H_1: p_{11} \neq p_{21},$$

vamos substituí-la pela hipótese equivalente

$$H_0: \theta = 1 \quad \text{contra a alternativa} \quad H_1: \theta \neq 1$$

Esta hipótese será rejeitada se o valor observado para a variável aleatória X for pouco provável, sob H_0 .

Como X é uma variável discreta, recomenda-se usar a correção de continuidade e, portanto, a hipótese nula será rejeitada se

$$\frac{(|n_{11} - \mu(1)| - \frac{1}{2})^2}{Var(1)} > \chi^2_{1-\alpha/2}, \text{ isto é, se}$$

$$\frac{n(|n_{11}n_{22} - n_{21}n_{12}| - \frac{n}{2})^2}{n_{1.}n_{.1}n_{2.}n_{.2}} > \chi^2_{1-\alpha/2}. \quad (3.32)$$

3.10 Enfoque Bayesiano

As análises condicionais e incondicionais que apresentamos nos tópicos anteriores são métodos da teoria Clássica que utilizam as frequências para obter conclusões sobre os parâmetros de interesse. Nesta teoria, os parâmetros são valores desconhecidos, mas fixos.

Na teoria Bayesiana, os parâmetros são variáveis aleatórias e, sendo assim, têm uma distribuição de probabilidade. À distribuição do parâmetro dá-se o nome de distribuição a priori.



Nesta linha teórica, as conclusões sobre o parâmetro de interesse são obtidas através de uma análise da distribuição a posteriori do parâmetro.

A distribuição a posteriori para o parâmetro será obtida combinando uma distribuição a priori com a função de verossimilhança dos dados.

Se considerarmos $\pi_{\Theta}(\theta)$ a distribuição a priori de um parâmetro de interesse Θ e $f_{X|\Theta}(x|\theta)$ a função de verossimilhança dos dados x , então a distribuição a posteriori para Θ é

$$\pi(\theta|x) \propto f_{X|\Theta}(x|\theta) \cdot \pi_{\Theta}(\theta) \tag{3.33}$$

O nosso objetivo não é comparar as análises procurando uma melhor, pois isto foge ao nosso tema; o intuito é apresentar opções de análise para tabelas 2x2.

Pericchi e Pérez (1994) citam o seguinte problema analisado por Kadane (1990):

Uma companhia reduziu sua força de trabalho em quatro etapas durante os anos 80. Um sindicato dos trabalhadores alegou que a companhia dispensou sem considerar a proporção para empregados acima de 40 anos, regulamentada por lei federal.

A situação pedia uma verificação da existência ou não de discriminação para o fator idade acima de 40 anos.

Os dados foram dispostos em quatro tabelas 2x2 e considerando-se todas as marginais fixadas.

tabela : Classificação dos dispensados e retidos nas quatro datas, segundo o fator idade.

data	idade	dispensados	retidos
6/10/82	+ 40	18	129
	- 39	0	102
11/10/82	+ 40	26	105
	- 39	10	82
5/31/83	+ 40	13	92
	- 39	14	66
6/28/84	+ 40	13	81
	- 39	2	52



A função de verossimilhança para cada tabela é:

$$f(n_{11}|L) = \frac{\binom{n}{n_{11}, n_{1.} - n_{11}, n_{.1} - n_{11}, n - n_{1.} - n_{.1} + n_{11}} e^{n_{11}L}}{\sum_j \binom{n}{j, n_{1.} - j, n_{.1} - j, n - n_{1.} - n_{.1} + j} e^{jL}}, j \in N \quad (3.34)$$

onde $L = \log \frac{p_{11}p_{22}}{p_{12}p_{21}}$ (logaritmo da "odds ratio"),

onde j assume valores segundo o intervalo $\max(0, n_{1.} + n_{.1} - n) \leq j \leq \min(n_{1.}, n_{.1})$ e

$$\binom{n}{a, b, c, d} = \frac{n!}{a!b!c!d!}.$$

Se $L = 0$, conclui-se que não houve ação da companhia favorável e nem desfavorável aos trabalhadores velhos.

Se a "odds ratio" for maior que 1 tem-se $L > 0$ e portanto, corresponderá a uma discriminação contra trabalhadores velhos e, se a "odds ratio" assumir um valor entre 0 e 1 tem-se $L < 0$ que corresponderá, a uma discriminação contra trabalhadores novos.

De acordo com Kadane (1990), é necessário neutralizar a distribuição a priori e, para isto, a escolha deve ser de uma distribuição para L que seja unimodal e simétrica em $L = 0$.

Sua opção foi por distribuições a priori normais $\pi(L) = N(0, \sigma^2)$, $\sigma^2 > 1$ de modo que os dados dominassem no cálculo da distribuição a posteriori.

De acordo com os resultados obtidos, ele concluiu que havia evidência de discriminação contra as pessoas acima de 40 anos na primeira, segunda e quarta data de dispensa; e discriminação contra as pessoas abaixo de 40 anos na terceira data de dispensa.

Para fazer uma análise combinada, foi construída a verossimilhança combinada, que é obtida através do produto das verossimilhanças das quatro tabelas. Novamente foram usadas distribuições a priori normais e constatou-se a evidência de discriminação contra funcionários velhos.

A questão levantada, em seguida, era avaliar a influência da forma normal das distribuições a priori no resultado inferencial, pois esta seleção pode não dar suporte à estrutura do problema.

Kadane, Pérez, Moreno e Pericchi (1993) propuseram que o mesmo problema fosse retomado, agora com análises conjugadas, análise não informativa e análise com classes de

distribuição a priori. O resumo destes resultados, aqui apresentados, estão em Pericchi e Pérez (1994).

Uma distribuição a priori é conjugada se a distribuição a posteriori pertence à mesma família da distribuição a priori.

Cox e Hinkley (1974) apresentam uma tabela com indicações de distribuições a priori conjugadas para diferentes distribuições dos dados. Para a distribuição Binomial, a distribuição a priori conjugada deve ser Beta; para o caso de uma distribuição de Poisson, a distribuição a priori conjugada deve ser Gama; e, para o caso geral de uma distribuição pertencente à família exponencial, pode-se usar uma distribuição a priori da família exponencial.

Para a execução da análise conjugada feita por Pericchi e Pérez (1994), a verossimilhança (3.34) foi reescrita na forma:

$$f(n_{11}|L) = \exp\{n_{11}L - M(L)\}g(n) \quad , \quad (3.35)$$

onde

$$M(L) = \log \left[\sum_j \binom{n}{j, n_{11} - j, n_{11} - j, n - n_{11} - n_{11} + j} \right] e^{jL} \quad , j \in N \quad (3.36)$$

e

$$g(n) = \log \left[\binom{n}{n, n_{11} - n_{11}, n_{11} - n_{11}, n - n_{11} - n_{11} + n_{11}} \right] \quad (3.37)$$

Deste modo, verifica-se que a verossimilhança é da família exponencial e a priori conjugada para L utilizada foi a apresentada em um resultado de Diaconis e Ylvisaker (1979):

$$\pi(L|n_0, x_0) \propto \exp\{n_0 x_0 L - n_0 M(L)\} \quad , \quad (3.38)$$

onde n_0 e x_0 devem ser selecionados de modo a refletir neutralidade.

Para x_0 , sugere-se o valor $\hat{e}_{ij}$, e para n_0 , valores menores ou iguais a 1. Para $n_0 = 1$, ocorre a menor dispersão. Deste modo, tem-se uma distribuição a priori que não domine no cálculo da densidade a posteriori e portanto, o domínio será da função de verossimilhança dos dados.

Utilizando estas distribuições a priori com diferentes valores de n_0 , as distribuições a posteriori mostraram a evidência de discriminação contra pessoas acima de 40 anos, confirmando, com poucas diferenças, a análise feita com distribuições a priori normais.



Na opção por uma análise não informativa, muitas distribuições a priori não informativas têm sido indicadas; mas, na situação do problema de Kadane (1990), a escolha de Pericchi e Pérez foi pela distribuição a priori de Jeffreys que é dada por

$$\pi(\theta) \propto I^{\frac{1}{2}}(\theta) \quad , \quad (3.39)$$

onde $I(\theta)$ é a informação de Fisher do modelo probabilístico dos dados.

De acordo com Pérez (1994), a distribuição a priori de Jeffreys para tabela 2x2, com todas as marginais fixadas, é dada por

$$\pi_j(L) \propto [M''(L)]^{\frac{1}{2}} \quad (3.40)$$

onde $M(L)$ é a descrita em (3.36).

Os resultados obtidos novamente foram semelhantes aos obtidos nas duas análises anteriores. Nas análises bayesianas, a escolha da distribuição a priori é de fundamental ajuda para as inferências. Em situações de desconhecimento desta distribuição deve-se optar por uma distribuição a priori não informativa.

3-10.1 Teste de Bayes

Na linha clássica inferencial, para verificarmos, nas tabelas 2x2, a igualdade das proporções ou a independência das variáveis em situações onde os métodos assintóticos não são apropriados, a indicação é pelo teste exato de Fisher.

Na linha Bayesiana existe a indicação do teste de Bayes que, segundo Irony e Pereira (1986) é superior ao teste exato de Fisher.

Para chegar a estas conclusões, eles se basearam no conceito de qualidade de Dawid (1982), que estabelece as seguintes condições para classificar um teste como bom:

- i) O teste deve estabelecer corretamente a probabilidade dos erros tipo I e tipo II, α e β respectivamente.
- ii) Para um valor fixado $k > 0$, o valor de $\alpha + k \beta$ é minimizado pelo teste.

As referências para o teste de Bayes aqui descrito são Irony e Pereira (1986).



Para a descrição deste teste, usaremos a tabela abaixo com a suposição de que X e Y são duas variáveis aleatórias binomiais independentes e o objetivo é comparar as duas proporções p_{11} e p_{21} .

	1	0	total
X	n_{11}	n_{12}	$n_{1.}$
Y	n_{21}	n_{22}	$n_{2.}$
total	$n_{.1}$	$n_{.2}$	n

O espaço paramétrico é $\Omega = \{ (p_{11}, p_{21}) \mid 0 < p_{11} < 1, 0 < p_{21} < 1 \}$.

A hipótese nula H_0 define o conjunto $\Omega_0 = \{ (p_{11}, p_{21}) \mid 0 < p_{11} = p_{21} < 1 \}$

$$H_0 : p_{11} = p_{21} \quad \times \quad H_1 : p_{11} \neq p_{21}$$

As distribuições a priori utilizadas para os parâmetros são uniformes em Ω_0 e $\Omega_1 = \Omega - \Omega_0$.

Com estas distribuições a priori, as probabilidades preditivas para $(n_{11}, n_{.1})$, sob H_0 e H_1 , são, respectivamente:

$$f_{H_0}(n_{11}, n_{.1}) = \frac{\binom{n_{.1}}{n_{11}} \binom{n - n_{.1}}{n_{1.} - n_{11}}}{\binom{n}{n_{.1}}} \frac{1}{n + 1} \tag{3.41}$$

$$e \quad f_{H_1}(n_{11}, n_{.1}) = \frac{1}{(n_{.1} + 1)(n_{2.} + 1)} \tag{3.42}$$

Para que fosse dada a mesma importância para H_0 e H_1 , consideraram-se as probabilidades a priori $P_r(H_0) = P_r(H_1) = 0,5$.

Temos assim o seguinte fator de Bayes :

$$b(n_{11}, n_{.1}) = \frac{f_{H_0}}{f_{H_1}} = \frac{(n_{.1} + 1)(n_{2.} + 1)}{n + 1} \frac{\binom{n_{.1}}{n_{11}} \binom{n - n_{.1}}{n_{1.} - n_{11}}}{\binom{n}{n_{.1}}} \tag{3.43}$$

Para uma constante k fixada ($k > 0$), a função teste f_B para o teste de Bayes é

$$f_B(n_{11}, n_{.1}) = \begin{cases} 0 & \text{se } b(n_{11}, n_{.1}) \geq k \\ 1 & \text{se } b(n_{11}, n_{.1}) < k \end{cases} \quad (3.44)$$

A qualidade principal deste teste é minimizar $\alpha + k\beta$ onde

$$\begin{cases} \alpha = \sum f_{H_0}(n_{11}, n_{.1}) \\ \beta = 1 - \sum f_{H_1}(n_{11}, n_{.1}) \end{cases} \quad (3.45)$$

e o somatório é sobre a região crítica

$$R_{B_k} = \{(n_{11}, n_{.1}) | b(n_{11}, n_{.1}) < k\} \quad (3.46)$$

Para que não seja considerada nenhuma preferência entre as duas hipóteses, assume-se para k o valor 1.

Para exemplificar, consideremos $n_{1.} = n_{2.} = 5$

Substituindo os possíveis valores de n_{11} e $n_{.1}$ na expressão $b(n_{11}, n_{.1})$ dada em (3.43), foram obtidos os seguintes valores apresentados na tabela seguinte.

Valores assumidos por $f_B(n_{11}, n_{.1})$

$n_{11} \backslash n_{.1} - n_{11}$	0	1	2	3	4	5
0	0	0	1	1	1	1
1	0	0	0	1	1	1
2	1	0	0	0	1	1
3	1	1	0	0	0	1
4	1	1	1	0	0	0
5	1	1	1	1	0	0

Utilizando as expressões (3.45), foram obtidos os seguintes valores para os erros do tipo I e tipo II , $\alpha = 0,2251$ e $\beta = 0,4444$.

As comparações entre o teste de Fisher e o teste de Bayes estão em Irony e Pereira (1986) e, de acordo com seus resultados, o poder do teste de Bayes é maior que o do teste exato de Fisher.

4. AMOSTRAS CORRELACIONADAS

Ao realizarmos estudos comparativos, algumas vezes são observadas diferenças significativas que não são resultados da variável preditora.

Pode ocorrer, por exemplo, que um pesquisador, ao comparar dois métodos de ensino que tenham sido aplicados em dois grupos de alunos distintos, tenha obtido resultados que não reflitam fielmente a eficiência dos dois métodos, pois os fatores inteligência, aptidão e motivação dificilmente serão balanceados entre os dois grupos.

Para superar estas dificuldades, podemos formar pares de indivíduos ou utilizar cada indivíduo como seu próprio controle.

Se houver um emparelhamento, os dois elementos devem ser selecionados de modo que, em cada par, as amostras sejam tão semelhantes quanto possível em relação a quaisquer variáveis extrínsecas que possam influenciar no resultado da pesquisa. Cada elemento do par será submetido a um dos métodos.

No entanto, na formação de pares, existem muitas dificuldades relacionadas à homogeneidade dos elementos. No exemplo de comparação de métodos de ensino, como medir ou avaliar as variáveis inteligência, motivação e aptidão se os métodos de mensuração são discutíveis?

Nestes casos, o ideal seria que o mesmo indivíduo pudesse ser usado como seu próprio controle.

Se um indivíduo servir como seu próprio controle, ele será submetido aos dois métodos, em ocasiões diferentes, e teremos, assim, um experimento do tipo “antes e depois”.

Entretanto, isto nem sempre é possível. Por exemplo, se queremos analisar o comportamento de um indivíduo em uma situação que lhe seja nova. Em uma próxima vez não teria sentido pensarmos nesta situação como “situação nova” para este indivíduo.

Na maioria das situações do tipo “antes e depois”, a utilização do mesmo indivíduo é sempre a melhor opção. Se queremos comparar duas formas diferentes de tratamento ou eficiência sobre determinado tipo de discurso nas preferências eleitorais o mesmo indivíduo é o ideal para a nossa proposição.

Os testes de hipótese utilizados nestas situações geralmente são adaptações do teste quiquadrado ou teste exato de Fisher.



A seguir, descreveremos o teste sugerido por McNemar (1955), para comparar frequências em amostras pareadas, sendo que estes pares podem ser de elementos distintos ou o mesmo indivíduo na situação “antes e depois”.

4.1 Teste de McNemar

Vamos considerar a situação na qual duas amostras são pareadas e queremos analisar determinadas mudanças em um atributo nos pares.

Por exemplo, em um experimento onde queremos comparar duas formas diferentes de tratamento para uma determinada doença incurável, tal como alteração na taxa de hormônios pelo excesso de produção da glândula suprarrenal. Neste caso, cada indivíduo pode ser seu próprio controle e este experimento do tipo “cross-over” gera a seguinte tabela :

tratamento 2	tratamento 1	
	0	1
1	n_{11}	n_{12}
0	n_{21}	n_{22}

- 1: reação favorável no tratamento
0: reação desfavorável no tratamento

Um teste apropriado para comparar frequências pareadas é o de McNemar.

A hipótese nula será

$$H_0 : p_{11} = p_{22}$$

onde, p_{11} e p_{22} são as probabilidades de um indivíduo ser alocado nas celas c_{11} e c_{22} , respectivamente e o teste pode ser unilateral ou bilateral.

O nosso interesse está nas celas onde os elementos dos pares tiveram reações diferentes; no caso da nossa tabela, nas frequências n_{11} e n_{22} .

Como $n_{11} + n_{22}$ representa o número de pares que acusaram estas características, então o valor esperado para estas celas, sob a hipótese nula, é

$$e = \frac{p_{11} + p_{22}}{2} ,$$

e pode ser estimado por $\hat{e} = \frac{n_{11} + n_{22}}{2}$

A hipótese nula pode ser testada pela estatística quiquadrado de Pearson com as seguintes alterações

$$X^2 = \sum_{O=n_{11}, n_{22}} \frac{(O - \hat{e})^2}{\hat{e}}, \quad (4.1)$$

onde O representa o valor ocorrido nas celas onde houveram resposta (1,0) ou (0,1).

Desenvolvendo-se o somatório acima e efetuando-se as operações algébricas, temos

$$X^2 = \frac{(n_{11} - n_{22})^2}{n_{11} + n_{22}}. \quad (4.2)$$

Se uma correção de continuidade for aplicada, esta expressão ficará

$$X_c^2 = \frac{(|n_{11} - n_{22}| - 1)^2}{n_{11} + n_{22}}. \quad (4.3)$$

Esta é a estatística teste de McNemar utilizada para testar a associação ou a significância das mudanças em uma tabela 2 x 2 onde as amostras foram pareadas.

X^2 tem uma distribuição assintótica quiquadrado com 1 grau de liberdade.

Exemplo: Para ilustrar, vamos considerar a seguinte situação fictícia.

Um coordenador pedagógico está interessado em verificar se uma nova abordagem em uma determinada disciplina curricular, sugerida por uma equipe de pedagogos, aumenta a dedicação dos alunos em termos de tempo destinado aos estudos.

Em sua escola de atuação, ele escolhe, aleatoriamente, uma turma para ser a sua amostra. Os alunos serão questionados no 1º dia da pesquisa sobre o tempo médio diário que destinam ao estudo da disciplina que foi previamente escolhida para ter uma nova abordagem. Após 60 dias deste novo procedimento didático, os alunos serão novamente interpelados.

Os alunos serão classificados em duas categorias :

1: estudam acima de 1 hora diária (inclusive)

0: estudam menos de 1 hora diária

Os resultados obtidos estão na tabela abaixo



técnica anterior	nova técnica		total
	1	0	
0	6	10	16
1	5	3	8
total	11	13	24

Considerando p_{11} a probabilidade de um aluno que estudava menos de uma hora por dia ter passado a estudar acima de uma hora diária e, p_{22} a probabilidade de um aluno ter diminuído o tempo de estudo de acima de uma hora para menos de uma hora.

A hipótese nula será

$$H_0 : p_{11} = p_{22} ,$$

ou seja, a probabilidade de mudança de tempo de estudo é a mesma nas duas situações em que houve mudança de tempo de estudo.

O interesse em questão é verificar se os alunos dedicarão mais tempo aos estudos com esta mudança de abordagem; portanto a hipótese alternativa, no caso, é $H_1 : p_{11} > p_{22}$, ou seja, é um teste unilateral.

Aplicando a estatística teste dada em (4.3) temos,

$$X_c^2 = \frac{(|6 - 3| - 1)^2}{6 + 3} = \frac{4}{9} = 0,444...$$

Se considerarmos o nível de significância $\alpha = 0,05$, o valor obtido para a estatística cai na região de aceitação da hipótese nula.

Conseqüentemente, não existem evidências de que esta mudança de técnica pedagógica modifica o interesse pela disciplina.

4.2 Teste de Gart para Efeitos da Ordem

O teste de McNemar para mudanças nos dá algumas conclusões sobre a igualdade ou não das proporções para casos de respostas diferentes em duas etapas de um experimento. Porém este teste ignora a ordem em que ocorreram estas mudanças. No exemplo anterior, as mudanças consideradas favoráveis e as desfavoráveis têm o mesmo peso

Vamos, então, considerar a situação descrita por Everitt (1977) .(dados fictícios)

Um total de 100 indivíduos com depressão recebeu as drogas A e B em diferentes ocasiões e algumas respostas foram anotadas. O objetivo principal da pesquisa era investigar efeitos colaterais tal como náusea.

Se aplicarmos o teste de McNemar, poderemos obter algumas conclusões sobre a igualdade ou não da ocorrência de náusea na ingestão destas duas drogas.

No entanto este teste vai ignorar a ordem em que elas foram aplicadas.

As drogas foram aplicadas com um intervalo de tempo, de modo que a interferência de uma sobre a outra pudesse ser minimizada.

Verificou-se o efeito náusea e não náusea.

Os dados estão na tabela a seguir.

Droga B	Droga A		total
	0	1	
1	3	9	12
0	75	13	88
total	78	22	100

A hipótese a ser testada é se a ocorrência de náusea é igual na ingestão das duas drogas.

$$H_0 : p_{NA} = p_{NB} ,$$

onde p_{NA} é a probabilidade de ocorrer náusea só na droga A e p_{NB} é a probabilidade de ocorrer náusea só na droga B.

Aplicando-se a estatística teste (4.3), obteremos

$$X^2 = \frac{(|3 - 13| - 1)^2}{16} = 5,06.$$

Na distribuição do quiquadrado com um grau de liberdade, se estabelecermos um nível de significância $\alpha = 0,05$, o valor acima será significativo e conclui-se-á que a ocorrência de náusea é diferente para as duas drogas .

No entanto pode ser que a ordem em que as drogas foram ministradas possa ter interferido nas conclusões, apesar do intervalo de tempo em que foram aplicadas.

Vamos então reconsiderar este experimento .

A amostra será separada em duas sub-amostras de mesmo tamanho.



Em uma delas, aplica-se a droga A e, posteriormente, a droga B; e, na outra, as drogas serão aplicadas na ordem inversa.

Denotaremos pelo par (A,B) o elemento da amostra onde A foi a primeira droga aplicada e B, a segunda.

Suponhamos que tivéssemos obtido as seguintes frequências:

Náusea	Drogas		total
	(A,B)	(B,A)	
A	7	6	13
B	1	2	3
total	8	8	16

Gart (1969) apresentou um teste para analisar o *efeito ordem* e outro para o *efeito da droga*.

Para expô-los, vamos considerar as seguintes notações:

y_a : número de observações onde A produz náusea no par (A,B)

y_a' : número de observações onde A produz náusea no par (B,A)

y_b : número de observações onde B produz náusea no par (A,B)

y_b' : número de observações onde B produz náusea no par (B,A)

Estas frequências são, agora, arranjadas em duas tabelas 2 x 2.

Tabela a :

Dados organizados para testar o efeito do tratamento

Náusea	Drogas		total
	(A,B)	(B,A)	
1ª droga	y_a	y_b'	
2ª droga	y_b	y_a'	
total			

Tabela b:

Dados organizados para testar o efeito ordem

Náusea	Drogas		total
	(A,B)	(B,A)	
A	y_a	y_a'	
B	y_b	y_b'	
total			

No nosso exemplo, temos:

Tabela a

Náusea	Drogas		total
	(A,B)	(B,A)	
1ª droga	7	2	9
2ª droga	1	6	7
total	8	8	16

Tabela b

Náusea	Drogas		total
	(A,B)	(B,A)	
A	7	6	13
B	1	2	3
total	8	8	16

O teste de Gart sugere a aplicação do teste de Fisher nas duas tabelas

a) Teste para o *efeito droga*

Utilizando as tabelas apropriadas para o teste de Fisher, obteve-se o seguinte *valor p*

:

$$p = 0,02$$



Ao nível de significância $\alpha = 0,05$, rejeitamos a hipótese nula e aceitamos que a ocorrência de náusea é maior quando aplicamos a droga A do que quando aplicamos a droga B.

b) Teste para o *efeito ordem*

O resultado obtido através do teste de Fisher para o *valor p* foi:

$$p = 0,50$$

Logo aceita-se a hipótese nula e não temos indícios de que a ordem em que as drogas foram ministradas tenha efeito na ocorrência de náusea.

4.3 Combinando Tabelas 2x2

Em muitos estudos de associação entre 2 variáveis, podemos querer fazer levantamento de dados em diferentes locais.

Por exemplo, se queremos investigar a associação entre a ocorrência de câncer de mama entre mulheres que amamentaram os filhos no seio e a ocorrência de câncer de mama entre as que não amamentaram no seio. Podemos levantar estes dados em hospitais localizados em diferentes cidades e, em cada área, os dados serão arrumados em tabelas 2 x 2.

A questão é: como estas informações em diferentes tabelas podem ser fundidas e a associação ser investigada?

Um possível método é reunir todos os dados em uma única tabela e aplicar o teste quiquadrado de Pearson. Mas isto só será adequado se as proporções correspondentes em cada tabela forem iguais, o que quase inviabiliza esta opção.

Se elas forem diferentes, este processo não será correto, pois a tabela que reúne os dados pode nos dar uma informação diferente das tabelas originais.

Em nosso exemplo, pode ser que, em algumas cidades, a ocorrência de câncer de mama seja mais freqüente que em outras cidades.

Outro processo é calcular os valores dos quiquadrados separadamente, em cada tabela, e depois adicionar estes valores.

Para o caso de k tabelas

$$X^2 = \sum_{i=1}^k X_i^2 \quad (4.4)$$

O valor do X^2 obtido será, então, confrontado com os valor tabelado de uma distribuição quiquadrado central com n graus de liberdade.



Entretanto este processo não leva em consideração a direção das diferenças entre as proporções das diferentes tabelas.

De acordo com Everitt (1977), os métodos mais apropriados para reunir estas informações são o método $\sqrt{\chi^2}$ e o método de Cochran. Entretanto, segundo Agresti (1990) o teste de Mantel-Haenszel é o mais utilizado. Este teste consiste numa aproximação para as análises condicionais. A seguir vamos expôr estes métodos.

4.3.1 Método $\sqrt{\chi^2}$

Everitt (1977) indica este método para as situações nas quais os tamanhos amostrais nas tabelas individuais 2×2 obtidas em diferentes locais não diferem grandemente e os valores das proporções estão entre 0,2 e 0,8, aproximadamente.

O método consiste em:

- i) calcular as estatísticas X^2 das diferentes tabelas;
- ii) extrair as raízes quadradas de cada X^2 ;
- iii) atribuir um sinal, de acordo com a diferença das proporções, para cada valor obtido em ii);
- iv) adicionar estas raízes.

Sob a hipótese de que as proporções sejam iguais, as estatísticas $\sqrt{X^2}$ calculadas para cada tabela 2×2 são assintoticamente distribuídas como normal padrão $N(0,1)$ e, portanto, a soma dos valores X das n diferentes tabelas é assintoticamente distribuído como $N(0, n)$.

Logo uma estatística teste para a hipótese de que não haja diferença entre as proporções em todas as tabelas será dada por :

$$Z = \frac{\sum_{i=1}^n \chi_i}{\sqrt{n}}, \quad (4.5)$$

onde $\chi_i = \sqrt{X_i^2}$ é o valor obtido na i -ésima tabela, precedido do sinal apropriado (sentido da diferença das proporções).

Z é assintoticamente distribuído como uma variável aleatória normal padrão.

Para exemplificar, consideremos uma pesquisa sobre associação entre câncer de mama e amamentação no seio.



Suponhamos que, após consultar as fichas de 3 hospitais, tenhamos obtido os seguintes dados sobre tumor de mama maligno ou benigno em mães que amamentaram seus filhos no seio e que não amamentaram no seio:

(dados fictícios)

Incidência de tumor no seio em 3 hospitais

hospital	amamentação	tumor		proporção	χ^2	χ
	no seio	1	0	de tumor 0		
H ₁	1	16	4	0,2000		
	0	7	6	0,4615	2,5527	1,5977
		23	10			
H ₂	1	14	4	0,2222		
	0	6	4	0,4000	0,9958	0,9979
		20	8			
H ₃	1	12	4	0,2500		
	0	12	10	0,4545	4,7985	2,1905
		24	14			

amamentação no seio : 1 sim tumor : 1 benigno
 0 não 0 maligno

Os tamanhos amostrais nos três locais não diferem muito.

Para calcular a estatística Z, devemos, primeiro, observar as proporções de tumores malignos em cada tabela .

Em todas as tabelas, a proporção de casos de tumores malignos em pessoas que não amamentaram é maior do que nas que amamentaram no seio; portanto aos valores χ de cada tabela será atribuído o mesmo sinal. Vamos atribuir a todos o sinal +.

$$Z= \frac{1,5977 + 0,9979 + 2,1905}{\sqrt{3}} = 2,7665$$

Ao nível de significância $\alpha = 0,05$, este valor é significativo. Logo, reunindo os dados das tabelas, concluímos que há associação entre tipo de câncer de mama nas mulheres que tiveram filhos e o fator amamentação no seio.



4.3.2 Método de Cochran

Ao combinarmos tabelas 2 x 2 para verificação de associação entre duas variáveis, se os tamanhos amostrais nas diferentes tabelas não forem próximos e as proporções forem extremas (não pertencerem ao intervalo de extremos 0,2 e 0,8) então, a adição dos valores dos χ_i não é uma boa opção, pois as tabelas com tamanhos totais maiores conseguem detectar melhor a diferença entre as proporções.

Cochran (1954) sugeriu um teste baseado na média ponderada das diferenças entre as proporções. A ponderação será efetuada utilizando pesos estabelecidos de acordo com os tamanhos amostrais.

A estatística teste usada será:

$$Y = \frac{\sum_{i=1}^n w_i d_i}{\left(\sum_{i=1}^n w_i P_i Q_i \right)^{1/2}} \quad (4.6)$$

onde para a i -ésima tabela temos as seguintes notações :

n : número de tabelas 2 x 2;

n_{i1}, n_{i2} são os tamanhos amostrais nos dois grupos;

p_{i1}, p_{i2} são as proporções observadas nas amostras;

$$P_i = \frac{n_{i1} p_{i1} + n_{i2} p_{i2}}{n_{i1} + n_{i2}} ;$$

$$Q_i = 1 - P_i ;$$

$$d_i = p_{i1} - p_{i2} ;$$

$$w_i = \frac{n_{i1} n_{i2}}{n_{i1} + n_{i2}} .$$

Y é uma ponderação dos valores d_i nos quais os pesos são maiores para as tabelas com tamanhos amostrais maiores.

Sob a hipótese de que os termos d_i são iguais a zero para $i = 1, 2, \dots, n$, temos

$$Y \sim N(0,1)$$

Para ilustrar, utilizaremos os dados apresentados por Everitt(1977) onde é feita uma avaliação da existência de associação entre a ocorrência de tiques em crianças mal-ajustadas e sexo feminino ou masculino.

Os dados foram coletados em 3 grupos de faixas etárias distintas.



tabela 1

Faixa Etária : 5 a 9 anos

sexo	tiques		total	proporção
	1	0		
masculino	13	57	70	0,1857
feminino	3	23	26	0,1154
total	16	80	96	0,1667

tabela 2

Faixa Etária : 10 a 12 anos

sexo	tiques		total	proporção
	1	0		
masculino	26	56	82	0,3171
feminino	11	29	40	0,2750
total	37	85	122	0,3033

tabela 3

Faixa Etária : 13 a 15 anos

sexo	tiques		total	proporção
	1	0		
masculino	15	56	71	0,2113
feminino	2	27	29	0,0690
total	17	83	100	0,1700

tique : 1 presença

0 ausência

Vamos apresentar os elementos para a obtenção da estatística Y

$n = 3$ tabela i $i=1,2,3$

para $i = 1$

$n_{11} = 70$

$n_{12} = 26$

$p_{11} = 0,1857$

$p_{12} = 0,1154$

$P_1 = 0,1667$

$Q_1 = 0,8333$

$d_1 = 0,0703$

$w_1 = 18,96$

para $i = 2$

$n_{21} = 82$

$n_{22} = 40$

$p_{21} = 0,3171$

$p_{22} = 0,2750$

$P_2 = 0,3033$

$Q_2 = 0,6967$

$d_2 = 0,0421$

$w_2 = 26,89$

para $i = 3$

$n_{31} = 71$

$n_{32} = 29$

$p_{31} = 0,2113$

$p_{32} = 0,0690$

$P_3 = 0,1700$

$Q_3 = 0,8300$

$d_3 = 0,1423$

$w_3 = 20,59$

Substituindo-se estes valores em (4.6), obtemos o seguinte valor para a estatística teste $Y = 1,61$.

Se considerarmos o nível de significância $\alpha = 0,10$ e compararmos com o quantil de ordem $(1 - \alpha/2)$ da distribuição normal padrão, este valor não será significativo. Portanto não há indícios de associação entre tipo de sexo e ocorrência de tiques em crianças mal-ajustadas.

4.3.3 Teste de Mantel- Haenszel

Quando os procedimentos de análise forem os condicionais o teste mais utilizado, para verificar a associação entre as duas variáveis em tabelas 2×2 em situações de K estratos, é o de Mantel e Haenszel (1959). Esse teste é uma aproximação para os procedimentos exatos condicionais.

Para descrevermos o teste utilizaremos a seguinte tabela:

tabela do estrato K

X	Y		total
	1	0	
1	n_{11k}	n_{12k}	$n_{1.k}$
0	n_{21k}	n_{22k}	$n_{2.k}$
total	$n_{.1k}$	$n_{.2k}$	$n_{..k}$

$K = 0, 1, 2, \dots, k.$

A análise dos dados em cada estrato é condicional e supondo-se os totais marginais fixados. Sob a hipótese de independência temos para cada estrato k , uma distribuição hipergeométrica. Logo, as esperanças e variâncias para cada n_{11k} são estimadas, respectivamente por:

$$\hat{e}_{11k} = \frac{n_{1.k} n_{.1k}}{n_{..k}}$$

$$\hat{Var}(n_{11k}) = \frac{n_{1.k} n_{2.k} n_{.1k} n_{.2k}}{n_{..k}^2 (n_{..k} - 1)}$$

Considerando o total de estratos, a ocorrência dos n_{11k} em cada um são independentes. Portanto, $\sum_k n_{11k}$ tem esperança e variância dadas por $\sum_k E(n_{11k})$ e $\sum_k Var(n_{11k})$, respectivamente e, a estatística proposta por Mantel e Haenszel (1959) para verificar a associação entre as variáveis, nestes casos é:

$$M^2 = \frac{\left(\left| \sum_k n_{11k} - \sum_k \hat{e}_{11k} \right| - \frac{1}{2} \right)^2}{\sum_k \hat{Var}(n_{11k})}. \quad (4.7)$$

Sob a hipótese de independência, a distribuição de M^2 se aproxima de uma distribuição χ^2 com 1 (um) grau de liberdade e, o teste proposto é rejeitar H_0 se

$$M^2 > \chi^2_{1-\frac{\alpha}{2}}.$$

De acordo com Agresti (1990) este teste é desaconselhado quando a associação muda drasticamente entre os estratos.

Para ilustrar, analisaremos novamente os dados apresentados por Everitt (1977) onde foi feita uma avaliação da existência de associação entre a ocorrência de tiques em crianças mal-ajustadas e sexo feminino ou masculino. Os estratos foram feitos utilizando diferentes faixas etárias.



Ocorrência de tiques em crianças mal-ajustadas classificadas de acordo com: sexo em diferentes faixas etárias.

faixas etárias	sexo	tiques		$\hat{e}_{11k}$	$\widehat{\text{Var}}(n_{11k})$
		1	0		
5 ____ 9	1	13	57	11,6666	2,6608
	0	3	23		
10 ____ 12	1	26	56	24,8688	5,7278
	0	11	29		
13 ____ 15	1	15	56	12,0700	2,9346
	0	2	27		
tiques: 1 presença		sexo: 1 masculino			
0 ausência		0 feminino			

A tabela apresenta as frequências e as estimativas do valor esperado e da variância para a presença de tiques no sexo masculino, sob a hipótese de independência, para cada faixa etária.

Calculemos a estatística de Mantel-Hanszel. Os valores das somatórias necessárias para o cálculo da mesma são $\sum_k n_{11k} = 54$, $\sum_k \hat{e}_{11k} = 48,6054$ e $\sum_k \widehat{Var}(n_{11k}) = 11,3232$ e, substituindo-os na expressão (4.7), obtemos o seguinte valor:

$$M^2 = \frac{(|54 - 48,6054| - 0,5)^2}{11,3232} = 2,1183.$$

Se considerarmos o nível de significância $\alpha = 0,10$ e compararmos com o quantil de ordem $(1 - \alpha/2)$ da distribuição quiquadrado com um grau de liberdade, este valor não será significativo. Portanto não há indícios de associação entre tipo de sexo e ocorrência de tiques em crianças mal-ajustadas.

Conclusão

As tabelas de contingência 2×2 aparecem muito amiúde em aplicações estatísticas, principalmente em estudos clínicos e estudos nas áreas sociais. Apesar da aparente simplicidade dessas tabelas, as técnicas de análise estatísticas são diversas e têm sido motivo de polêmica entre diversos estatísticos renomados.

Nos estudos comparativos, os retrospectivos são os mais comuns por necessitarem menos recursos financeiros que os demais e serem os de mais fácil obtenção dos dados. Contudo, esses estudos trazem o inconveniente de não possibilitarem o cálculo de um estimador para o risco relativo.

Entre as medidas de associação por nós estudadas, a “odds ratio” mereceu maior destaque pois:

- a) pode ser calculada em todos os métodos de amostragem;
- b) é facilmente estimada através da amostra;
- c) é muito usada na área biomédica e de simples interpretação;
- d) tem o valor muito próximo do risco relativo, quando a probabilidade de interesse for muito pequena.

Entre estes itens, destaca-se o último por permitir que possamos, através da “odds ratio”, obter conclusões que seriam peculiares do risco relativo que é de grande importância na área de epidemiologia.

O teste quiquadrado de Pearson tem se mostrado apropriado para verificar a existência de associação entre as variáveis em estudo para casos de grandes amostras, tanto nos modelos Multinomial, de Poisson, produto de Binomiais ou Hipergeométrico. Para verificar a igualdade de proporções ele também é aconselhado, em situações de grandes amostras.

No entanto, se alguma das celas tiver esperança menor que 5 ou se o total amostral for menor que 20, o uso desse teste é desaconselhado.

Em situações de pequenas amostras ou de ocorrência rara de um evento, aconselha-se o teste exato de Fisher, para verificar a associação. Esse teste tem diversas vantagens inerentes, tais como:

- a) condicionar em uma estatística aproximadamente ancilar, possibilitando assim a eliminação de parâmetros incômodos;
- b) utilizar o *valor p* para as análises estatísticas ao invés de nível de significância;



c) utilizar uma função de verossimilhança condicional que propicia a possibilidade de efetuarmos inferências sobre a “odds ratio”, que é uma medida de associação com grande uso na área biomédica;

d) aparecer em diversos pacotes estatísticos.

No entanto, o teste de exato de Fisher ocasiona perda no número de informação de Fisher se a situação envolver duas distribuições binomiais independentes. Como esta perda é decrescente a medida que o total amostral n cresce então, para situações de alguma cela ter frequência esperada menor que 5, consideramos adequado a opção pelo teste de Fisher para verificar a associação entre duas variáveis.

Após as conclusões de Yates (1984) sobre o teste exato de Fisher, a aceitação do mesmo tem sido generalizada pela grande maioria dos pesquisadores. Isso pode ser constatado através das discussões de diversos estatísticos anexadas ao artigo de Yates.

A correção de continuidade, proposta por Yates (1934), para a estatística X^2 , após Yates (1984), passou a ter aceitação da maioria dos estatísticos.

Apesar das dificuldades relacionadas à não unicidade na construção de uma estatística ancilar, são inegáveis as grandes vantagens que ela propicia em análises estatísticas.

Os resultados obtidos por Liang (1983) facilitam a classificação de uma estatística fracamente ancilar e devem ser elogiados pois, se uma medida de informação desejada for o número de informação de Fisher, condicionar em uma estatística fracamente ancilar se mostrou a melhor opção por nós estudada para que haja o mínimo de perda neste número de informação em relação ao número de informação contido nos dados.

Os testes para amostras correlacionadas na sua grande maioria são adaptações, para situações particulares, do teste quiquadrado de Pearson ou do teste exato de Fisher. Os estudos dos testes apresentados no capítulo 4 consistiu um início de pesquisa que necessitam uma complementação, principalmente para situações de combinar tabelas 2×2 .

Os testes incondicionais não foram abordados em nosso estudo, mas acreditamos que seria frutífero um estudo relacionando os estudos de Irony e Pereira (1986) com os de Berger (1994), visando comparar o poder entre teste de Bayes, teste exato de Fisher e alguns testes incondicionais.

As opções de análise bayesiana foram apresentadas como sugestões e não foram mais explanadas pois serão alvo de estudo detalhado em próximos trabalhos.

Muito se tem escrito e desenvolvido sobre tabelas 2×2 . Este trabalho é apenas um pequeno passo no sentido de obter um certo grau de proficiência em análise de tabelas 2×2 .



BIBLIOGRAFIA

- Agresti, A.(1990). *Categorical Data Analysis*. John Wiley, New York.
- Agresti, A. (1992). A survey of exact inference for contingency tables. *Statistical Science*. Vol 7, n^o 1, 131 -177.
- Andersen, E. L. (1970). Asymptotic properties of conditional maximum likelihood estimators. *J. Roy. Statist. Soc.B*. 32, 283-301.
- Andersen, E. L. (1973). *Conditional Inference and Models for Measuring*. Copenhagen Mentalhygiejnisk. Forlag.
- Barnard, G. A. (1949). Statistical inference. *J. Roy. Statist. Soc. B*. 11, 115-139.
- Barnard, G. A. (1984). Discussion of "Test of Significance for 2×2 contingency tables" by F. Yates. *J. Roy. Statist. Soc.A*. 147, 449 - 450.
- Barndorff - Nielsen, O. (1973). On M- ancillarity. *Biometrika* 60, 447 - 455.
- Barndorff - Nielsen, O. (1976). Nonformation. *Biometrika* 63, 567 - 571.
- Basawa, I. V. (1981). Efficiency of conditional maximum likelihood estimators and confidence limits for mixtures of exponencial families. *Biometrika* 68, 515 - 523.
- Basu, D. (1964). Recovery of ancillary information. *Sankhya A* 23.
- Basu, D. (1977). On the elimination of nuisance parameters. *J. Am. Statist. Assoc.* 72, 355 - 366.
- Bennett, B. M.and Hsu, P. (1960). On the power function of the exact test for the 2×2 contingency table. *Biometrika* 47, 393-398.
- Berger, R. L.(1994). Power comparison of exact unconditional tests for comparing two binomial proportions. *Technical Report, Institute of Statistics Mimeo Series n^o 2266* .North Carolina State University.
- Bishop, Y. V., Fienberg, S. E. and Holland, P. W. (1988). *Discrete Multivariate Analysis*. 9^a th edn. Cambridge. MIT Press.
- Breslow, N. and Day, N. E. (1980). *Statistical Methods in Cancer Research*.vol II. Lyon: IARC.
- Charles Griffin and Co. Ltd.(editor).(1967).*Biometrika Tables for Statisticians*. vol I , 3^a th end.
- Clogg, C. C. and Shockey, J. W. (1988). *Multivariate Analysis of Discrete Data*. in *Handbook of Multivariate Experimental Psychology*. J.R. Nesselroade and R.B. Cattell. New York.
- Cochran, W. G. (1952). The χ^2 test of goodness-of-fit. *Ann. Math. Statist.* 23, 315 - 345.



Cochran, W. G. (1954). Some methods for strengthening the common χ^2 tests. *Biometrics* 10, 417- 451.

Conover, W. J. (1968). Uses and abuses of the continuity correction. *Biometrics* 24, 1028.

Conover, W. J. (1974). Some reason for not using the Yates continuity correction on 2 x 2 contingency tables. *J. Am. Statist. Assoc.* 69, 347 - 376.

Cornfield, J. (1951). A method of estimating comparative rates from clinical data; applications to cancer of the lung, breast and cervix. *J. Nat. Canc. Inst.* 11, 1269 - 1275.

Cornfield, J. (1956). A statistical problem arising from retrospective studies. *J. Neyman* 4, 135 - 148.

Cox, D. R. and Hinkley, D.V. (1974). *Theoretical Statistics*. Chapman and Hall, New York.

Davis, F. M. and Laurenson, V. G. (1981). Spinal anaesthesia for emergency hip surgery in elderly patients. *Anaesthesia Intensive Care* 9.

Dawid, A. P. (1982). The well calibrated Bayesian. *J. Am. Statist. Assoc.* 77, 605-613.

Diaconis, P. and Ylvisaker, D. (1979). Conjugate priors for exponential families. *Ann. of Stat.* 7, 269- 281.

Doll, R. and Hill, A. B. (1964). Mortality of doctors in relation to their smoking habits; a preliminary report. *British Medical Journal* 1, 1451 -1455.

Dorn, H. F. (1944). Illness from cancer in the United States. *Publ. Heath. Rep.* 59 n^{os} 2,3,4.

Efron, B. (1977). The efficiency of Cox's likelihood function for censored data. *J. Am. Statist. Assoc.* 72

Engeman, R. and Swanson, G. (1991). On analytical methods and inferences for 2 x 2 contingency table data from medical studies. *Computers and Biomedical Research* 24, 509 - 513.

Everitt, B. S. (1977). *The Analysis of Contingency Tables*. Chapman and Hall.

Feldman, E. E. and Klinger, E. (1963). Short - cut calculation of the Fisher - Yates "exact test". *Psychometrika* 28, 289 - 291.

Finney, D. J.; Latscha, R.; Bennett, B. M. and Hsu, P. (1963). *Tables for testing significance in 2 x 2 table*. Cambridge University Press.

Fisher, R. A. (1925). *Statistical Methods for Research Workers*. Edinburgh: Oliver and Boyd.

Fisher, R. A. (1926). Bayes' theorem and the fourfold table. *Eugen. Rev.* 18, 32 -34.

Fisher, R. A. (1935). The logic of inductive inference.. *J. Roy. Statist. Soc.* 98, 39 -54.



Fisher, R. A. (1956). *Statistical Methods and Scientific Inference*. Edinburg: Oliver and Boyd.

Fleiss, J. L. (1979). Confidence intervals for the odds ratio in case - control studies: The state of the art. *Jour. Chronic Diseases* 32.

Fleiss, J. L. (1981). *Statistical methods for rates and proportions*. 2^ath edn. Wiley, New York.

Fu, Y. X. and Arnold, J. (1992). A table of exact sample sizes for use with Fisher's exact test for 2 x 2 tables. *Biometrics* 48, 1103 -1112

Gart, J. J. (1969). An exact test for comparing matched proportions in crossover designs. *Biometrika* 56, 75 -80.

Gail, M. H. and Gart, J. J. (1973). The determination of samples sizes for use with the exact conditional test in 2 x 2 comparative trials. *Biometrics* 29, 441 - 448.

Haldane, J. B. S. (1955). The estimation and significance of the logarithm of a ratio of frequencies. *Ann. Human Genet.* 20, 309 - 311.

Hannan, J. and Harkness, W. L. (1963). Normal approximation to the distribution of two independent binomials, conditional on fixed sum. *Ann. Math. Stat.* 34, 1593 - 1595.

Irony, T. Z. and Pereira, C. A. B. (1986). Exact tests for equality of two proportions: Fisher v Bayes. *Jour. Stat. Comput. Simul.* 25, 93 -114.

Kadane, J. (1990). A statistical analysis of adverse impact of employer decisions. *J. Am. Statist. Assoc.* 85, 925 - 933.

Kadane, J., Pérez, M. E., Moreno, E., Pericchi, L.R. (1993). Robust and neutral assessment of evidence in court. (não publicado).

Kendall, M. and Stuart, A. (1979). *The Advanced Theory of Statistics*. vol 2, 4th edn. Charles Griffin & Co.Ltd.

Lehmann, E. L. (1959). *Testing Statistical Hypotheses*. Wiley. New York.

Liang, K. L. (1983). On information and ancillarity in the presence of a nuisance parameter. *Biometrika* 70, 607-612.

Liebetrau, A. M. (1983). *Measures of Association*. Sage University Paper series on Quantitative Applications in the Social Sciences, 07-032. Beverly Hills and London: Sage Pubns.

Mantel, N. and Greenhouse, S. W. (1968). What is the continuity correction ? *Amer. Statistician* 22, 27 - 30.



Mantel, N. (1974). Comment and a suggestion on the Yates continuity correction. *J. Am. Statist. Assoc.* 69, 378- 380.

Mcleskey, C. H. (1985). *Anesthesia for the geriatric patient*. In " *Advances in Anesthesia*". Year BK. Med. Chicago.

McNemar, Q. (1955). *Psychological Statistics* .Wiley. New York.

Pearson, K. (1904). On the theory of contingency and its relation to association and normal correlation. *Drapers Co. Memoirs Biometric series n° 1*. London.

Pereira, C. A. and Lindley, D. V.(1987). Examples questioning the use of partial likelihood. *The Statistician* 36, 15 - 20.

Pérez, M. E. (1994). An automatic and proper analysis of 2 x 2 tables with one and two fixed margins. Test

Pericchi, L. J. and Pérez, M. E. (1994). A case study on the Bayesian analysis of 2 x 2 tables with all margins fixed. *Rev. Bras. Prob. Estat.* 1, 27 - 37.

Plackett, R. L. (1977). The marginal totals of a 2 x 2 table. *Biometrika* 64, 37 -42.

Renyi, A. (1959). On measures of dependence . *Acta. Math. Acad. Scien. Hangar* 10, 441 -451.

Sachs, L. (1984). *Applied Statistics Springer*. Verlag. New York.

Schrek, R. (1950). Tobacco smoking as an etiologic factor in disease. I. Cancer. *Cancer Research* 10, 49 -58.

Siegel, S. (1975). *Nonparametric Statistics for the Behavioral Sciences*. McGraw- Hill.

Sillito, P. (1949). Note on approximations to the power function of the "2x2 comparative trial". *Biometrika* 36, 347 - 352.

Snedecor, G. W. and Cochran, W. G.(1980). *Statistical methods*. 7 th. Iowa State Univ. Press. Ames. IA.

Soares, J. F. e Colosimo, E. A.(1995). *Métodos Estatísticos na Pesquisa Clínica*. 40^a RBRAS. Ribeirão Preto-SP.

Tocher, K. D. (1950). Extension of the Neyman-Pearson theory of tests to discontinuous variates. *Biometrika* 37, 130 - 144.

Vos, P. W. and Hudson, S. (1995). Signed G.L.R.test, not Fisher's exact test, in testing equality of independent binomial proportions. *Dept. of Biostat., East Carolina Univ. Greenville*.

Yates, F. (1934). Contingency tables involving small numbers and the χ^2 test.. *J. Roy. Statist. Soc. Suppl* 1, 217 - 235.



Yates, F. (1984). Tests of Significance for 2 x 2 Contingency Tables.. *J. Roy. Statist. Soc. A.* 147, 426 - 463.

Yule, G. U. (1912). On the methods of measuring association between two attributes. *J. Roy. Statist. Soc.* 75, 579.

Walter, S. D. (1985). Small- sample estimation of log odds ratio from logistic regression and fourfold tables. *Stat. in Medicine* 4, 437 -444.

Walter, S. D. and Cook, R. J. (1991). A comparison of several point estimators of the odds ratio in a single 2 x 2 contingency table. *Biometrics* 47, 795 - 811.

Welch, B. L. (1939). On confidence limits and sufficiency, with particular reference to parameters of location. *Ann. Math. Stat.* 10, 58.

Wilks, S. S. (1935 , 1938). The large - sample distribution of the likelihood ratio for testing composite hypotheses. *Ann. Math. Stat.* 9, 60.



