Universidade Federal do Ceará
Departamento de Estatística e Matemática Aplicada
Coordenação do Curso de Estatística
Disciplina: CC0291- Estatística Não Paramétrica -2023.1
Professor: Maurício Mota

. Teste Exato de Fisher

1:Generalidades

O teste Exato de Fisher é apresentado no livro do professor Humberto de Campos na seção 4.4 páginas 149 a 156. Embora o teste de χ^2 seja o mais usual dos testes não-paramétricos aplicáveis ao caso de duas amostras independentes, ele não é apropriado para casos de pequenas amostras.

Se admitirmos um tabela de contingência 2×2, com os totais marginais fixos, e com o total geral(N) muito baixo, o teste de χ^2 não se aplica devido às suas restrições.

Em casos dessa natureza, recomenda-se a aplicação do Teste Exato de Fisher, que consiste em determinar a exata probabilidade de ocorrência de uma frequência observada, ou de valores ainda mais extremos.

São consideradas duas amostras (A e B), representativas de duas populações. Para ambas as amostras, são tomadas duas classes (I e II), constituindo, assim, a tabela de contingência.

	I	II	Total
A	a	b	$a+b=N_1$
B	c	d	$c+d=N_2$
Total	a+c	b+d	N

No livro do Conover esta tabela é apresentada como:

	Coluna 1	Coluna 2	Total
Linha 1	x	r-x	r
Linha2	c-x	N-r-c +x	$N-r$
Total	c	N-c	N

2:Pressuposições

As principais suposições do teste são:

a. As amostrais são casuais e independentes;

b. As duas classes são mutuamente exclusivas.

As principais suposições do Conover são:

a. Cada observação é classificada em exatamente uma cela;

b. Os totais de linhas e colunas são fixos

2:Hipóteses

Para estabelecermos H_0 , consideremos uma das classes e nela admitamos:

Sejam $P(A)$, a probabilidade de um elemento pertencer à população representada pela amostra A e $P(B)$, a probabilidade de um elemento pertencer à população representada pela amostra B.

Nestas condições podemos testar

$$H_0 : P(A) = P(B)$$

contra qualquer tipo de hipótese alternativa:

$$H_1 : P(A) < P(B), \quad H_1 : P(A) > P(B) \quad H_1 : P(A) \neq P(B).$$

As hipóteses do Conover são:

Sejam p_1 a probabilidade de uma observação da linha 1 ser classificada na coluna 1 e
p_2 a probabilidade de uma observação da linha 2 ser classificada na coluna 1.
Podemos ter um teste bilateral:

$$H_0 : p_1 = p_2 = p, \quad H_1 : p_1 \neq p_2.$$

Um teste unilateral à esquerda;

$$H_0 : p_1 = p_2, \quad H_1 : p_1 < p_2.$$

Ou um teste unilateral à direita:

$$H_0 : p_1 = p_2, \quad H_1 : p_1 > p_2.$$

3:Estatística

A estatística do teste T_2 é o número de observações na casela linha 1 e coluna 1.

Vamos provar que se H_0 é verdade:

$$T_2 \sim \text{Hipergeométrica}(N, r, c),$$

cuja função de probabilidade é dada por:

$$P(T_2 = x) = \frac{\binom{r}{x}\binom{N-r}{c-x}}{\binom{N}{c}} \; I_A\;(x),$$

com $A = \{max(0, c + r - N), max(0, c + r - N) + 1, \ldots, min(r, c)\}$.

A média de T_2 é dada por:

$$\mu = \frac{rc}{N},$$

e a variância de T_2 é dada por:

$$\sigma^2 = \frac{rc(N-r)(N-c)}{N^2(N-1)}.$$

Para grandes amostras temos

$$T_3 = \frac{T_2 - \mu}{\sigma} \sim N(0, 1),$$

aproximadamente.

Vamos mostrar como a distribuição hipergeométrica aparece na parada:
Sejam

$$X \sim Bin(m, p) \quad e \quad Y \sim Bin(n, p), \;\; \text{independentes.}$$

A distribuição de

$$S = X + Y \sim Bim(m + n, p).$$

Qual a distribuição de $X|S = s$?

Prova:

Seja x um ponto do suporte de T_2 . Logo

$$P(X = x|S = s) = \frac{P(X = x, S = s)}{P(S = s)} = \frac{P(X = x, X + Y = s)}{P(S = s)} = \frac{P(X = x, Y = s - x)}{P(S = s)},$$

pela independência entre X e Y temos:

$$P(X = x|S = s) = \frac{P(X = x)\; P(Y = s - x)}{P(S = s)}.$$

$$P(X = x|S = s) = \frac{\binom{m}{x} p^x q^{m-x} \; \binom{n}{s-x} p^{s-x} q^{n-s+x}}{\binom{m+n}{s} \; p^x q^{m+n-s}},$$

$$P(X = x|S = s) = \frac{\binom{m}{x} \; \binom{n}{s-x} p^s q^{m+n-s}}{\binom{m+n}{s} p^s q^{m+n-s}},$$

$$P(X = x|S = s) = \frac{\binom{m}{x} \; \binom{n}{s-x}}{\binom{m+n}{s}} \; I_A(x), .$$

$A = \{max(0, s - n), max(0, s - n) + 1, \ldots, min(m, s)\}$.
note que

$$s - x \le n$$

e

$$x \geq s - n.$$

Para entender o oddsratio que aparecem na saída do R. Vamos trabalhar no caso geral:

Encontrei na página 502 do livro Rohatgi, e Saleh, An Introduction to Probability and Statistics.

Sejam $X \sim B(m, p_1)$, $Y \sim B(n, p_2)$, independentes , eles falam que o teste de Fisher-Irwin é baseado na distribuição condicional de $X|S = X + Y$. Vamos obtê-la agora:

Seja

$$\Psi = \frac{p_1 q_2}{p_2 q_1}.$$

$$P(S = s) = \sum \binom{m}{x} p_1^x q_1^{m-x} \binom{n}{s-x} p_2^{s-x} q_2^{n-s+x},$$

Mas

$$p_1^x q_1^{m-x} p_2^{s-x} q_2^{n-s+x} = p_1^x q_1^{-x} p_2^{-x} p_2^a q_2^{-s} q_2^x \; q_1^m \; q_2^{n-s}$$

$$= \left(\frac{p_1 q_2}{p_2 q_1}\right)^x q_1^m \; q_2^n \left(\frac{p_2}{q_2}\right)^s$$

$$= \Psi^x \; q_1^m \; q_2^n \left(\frac{p_2}{q_2}\right)^s$$

$$= \Psi^x \; q_1^m \; q_2^n a(m, n).$$

Assim

$$P(S = s) = \sum \binom{m}{x} \binom{n}{s-x} \Psi^x \; q_1^m \; q_2^n a(m, n)$$

Já vimos ante que:

$$P(X = x|S = s) = \frac{P(X = x) \; P(Y = s - x)}{P(S = s)}.$$

$$P(X = x|S = s) = \frac{\binom{m}{x} p_1^x q_1^{m-x} \; \binom{n}{s-x} p_2^{s-x} q_2^{n-s+x}}{P(S = s)},$$

$$P(X = x|S = s) = \frac{\binom{m}{x} \; \binom{n}{s-x} \Psi^x \; q_1^m \; q_2^n a(m, n)}{\sum \binom{m}{x} \binom{n}{s-x} \Psi^x \; q_1^m \; q_2^n a(m, n)},$$

Finalmente,

$$P(X = x|S = s) = \frac{\binom{m}{x} \; \binom{n}{s-x} \Psi^x}{\sum \binom{m}{x} \binom{n}{s-x} \Psi^x}.$$

Se $\Psi = 1$ caímos no caso anterior. É esta hipergeométrica generalizada que o R usa para calcular o oddsratio.

4:Exemplos:

Exemplo 1- Conover-pg 190:

Quatorze graduados em negócios recém-contratados, 10 homens e 4 mulheres, todos igualmente qualificados, estão sendo designados pelo presidente do banco para seus novos empregos. Dez dos novos empregos são como caixa e quatro como representantes de contas. A hipótese nula é que homens e mulheres têm chances iguais de conseguir os empregos de representante de conta mais desejáveis. A alternativa unilateral de interesse é que as mulheres têm mais probabilidade do que os homens de conseguir os empregos de representante de contas.

Veja os dados:

	Rep. de Conta	Caixa	Total
Homens	1	9	10
Mulheres	3	1	4
Total	4	10	14

Solução Vamos testar as hipóteses:

$$H_0 : p_1 = p_2 = p,$$

$$H_1 : p_1 < p_2.$$

Aqui p_1 é a probabilidade de um homem ser representante de conta e p_2 de uma mulher.

O valor observado de $t_2 = 1$. O nível descritivo do teste é dado por:

$$nd = P(T_2 \leq 1).$$

A distribuição de T_2 é:

$$T_2 \sim HG(N = 14, r = 10, c = 4),$$

Analisando os dados temos que dos 4 representantes de contas temos 3 mulheres e um homem. Será que isto é suficiente para rejeitar H_0?

$$P(T_2 = x) = \frac{\binom{10}{x}\binom{4}{4-x}}{\binom{14}{4}}\ I_A\ (x),$$

com $A = \{0, 1, 2, 3, 4\}$ pois $min(r, c) = min(10, 4) = 4$ e $max(0, c + r_N) = max(0, 0) = 0$.

A probabilidade

$$P(T_2 = 0) = \frac{\binom{10}{0}\binom{4}{4}}{\binom{14}{4}} = 0,000999001$$

A probabilidade

$$P(T_2 = 1) = \frac{\binom{10}{1}\binom{4}{3}}{\binom{14}{4}} = 0,03996004$$

O nível descritivo é dado por:

$$\hat{\alpha} = nd = P(T_2 \leq 1) = 0,041.$$

```
> N=14;r=10;c=4
>
> choose(N,c)
[1] 1001
>
> p_0=( choose(r,0)*choose(N-r,4))/choose(N,c);p_0
[1] 0.000999001
> p_1=( choose(r,1)*choose(N-r,3))/choose(N,c);p_1
[1] 0.03996004
> nd=p_0+p_1;nd
[1] 0.04095904
> round(nd,3)
[1] 0.041
>
> alfa=0.05
> alfa >nd
[1] TRUE
>
>
> #####Use a distribuição Hipergeométrica do R.
>
> alfa_hat=phyper(1,r,N-r,c);alfa_hat
[1] 0.04095904
>
```

Devemos rejeitar a hipótese nula.

Vamos fazer direto no R:

```
 RC=c(1,3)
>
> Cai=c(9,1)
> Sexo=cbind(RC,Cai);Sexo
     RC Cai
[1,]  1   9
[2,]  3   1
> row.names(Sexo)=c("H","F")
> Sexo
  RC Cai
H  1   9
F  3   1
>
> fisher.test(Sexo, alternative="less")

Fisher's Exact Test for Count Data

data:  Sexo
p-value = 0.04096
alternative hypothesis: true odds ratio is less than 1
95 percent confidence interval:
0.000000 0.897734
sample estimates:
odds ratio
0.05545513
```

Exemplo 2: Humberto de Campos-pg 154

Numa classe de 24 alunos, foi feito um estudo a fim de verificar se os estudantes provenientes das escolas particulares tinham o mesmo aproveitamento que os das escolas públicas. Para isso foi tomada como referência a nota média da classe. O resultado obtido foi o que se segue:

	Acima da Média	Abaixo da Média	Total
Escolas Particulares(A)	5	7	12
Escolas Públicas (B)	10	2	12
Total	15	9	24

Solução Vamos testar as hipóteses:

$$H_0 : P(A) = P(B) \quad vs \quad H_0 : P(A) \neq P(B).$$

O valor observado de $t_2 = 5$. O nível descritivo do teste é dado por:

$$nd = 2 \times min\Big(P(T_2 \leq 5), P(T_2 \geq 5)\Big).$$

A distribuição de T_2 é:

$$T_2 \sim HG(N = 24, r = 12, c = 15),$$

Analisando os dados temos que dos 15 alunos com notas acima da média temos 5 se escolas particulares e 10 de escolas públicas. Será que isto é suficiente para rejeitar H_0?

$$P(T_2 = x) = \frac{\binom{12}{x}\binom{12}{15-x}}{\binom{24}{15}} \; I_A\;(x),$$

com $A = \{3, 4, \ldots, 12\}$ pois $min((r, c) = min(12, 15) = 15$. e $max(0, c + r - N) = max(0, 15 + 12 - 24) = max(0, 3) = 3$.

A probabilidade

$$P(T_2 = 3) = \frac{\binom{12}{3}\binom{12}{12}}{\binom{24}{15}} = 0,0001682595$$

$$P(T_2 = 4) = \frac{\binom{12}{4}\binom{12}{11}}{\binom{24}{15}} = 0,004543007$$

$$P(T_2 = 5) = \frac{\binom{12}{5}\binom{12}{10}}{\binom{24}{15}} = 0,03997846$$

A probabilidade $p = P(T_2 \leq 5) = 0,04468973 < 0.5$ como o teste é bilateral temos:

$$nd = 2 \times p = 0,08937946.$$

assim, rejeitamos H_0 para um nível de significância $\alpha = 0,10$. Bons tempos!!!

```
>
> AcM=c(5,10) ###Acima da média
> AbM=c(7,2)   ###Abaixo da média
>
> Escola=cbind(AcM,AbM)
> row.names(Escola)=c("Particular","Pública")
> Escola
           AcM AbM
Particular   5   7
```

```
Pública     10   2
> fisher.test(Escola)

Fisher's Exact Test for Count Data

data:  Escola
p-value = 0.08938
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
0.01167257 1.23485892
sample estimates:
odds ratio
0.1563843

>
> ##T_2 ~HG(N,r,c)
>
> N=24;r=12;c=15
> p1=phyper(5,r,N-r,c);p1
[1] 0.04468973
> p1<0.5
[1] TRUE
> nd=2*p1;nd
[1] 0.08937946
> round(nd,4)
[1] 0.0894
>
> den=choose(N,c);den
[1] 1307504
> t=3
> num3=choose(r,t)*choose(N-r,c-t);num3
[1] 220
> p_3=num3/den;p_3
[1] 0.0001682595
> t=4
> num4=choose(r,t)*choose(N-r,c-t);num3
[1] 220
> p_4=num4/den;p_4
[1] 0.004543007
>
> t=5
> num5=choose(r,t)*choose(N-r,c-t);num5
[1] 52272
> p_5=num5/den;p_5
[1] 0.03997846
>
> p=p_3+p_4+p_5;p
[1] 0.04468973
>
> p<0.5
[1] TRUE
> nd=2*p;round(nd,4)
[1] 0.0894
>
> ###Direto usando a função phyper:
>
> pd=phyper(5,r,N-r,c);pd
[1] 0.04468973
```

```
>
```

Exemplo 3: Humberto de Campos-pg 156 Num estudo sobre fecundidade de duas raças bovinas foram feitos acasalamentos, obtendo-se os seguintes resultados:

	I=Fecundos	II=Não Fecundos	Total
Raça A	a=3	b=7	$a + b = N_1 = 10$
Raça B	c=4	d=1	$c + d = N_2 = 5$
Total	a+c=7	b+d=8	N=15

Verifique se as duas raças diferem quanto à fecundidade.

Solução

Sejam p_1 a probabilidade de um animal da raça A ser fecundo e p_2 a probabilidade de um animal da raça B ser fecundo. Vamos testar se:

$$H_0 : p_1 = p_2 = p, \qquad H_1 : p_1 \neq p_2.$$

O valor observado $t_2 = 3$. A distribuição de T_2 é

$$T_2 \sim HG(N = 15, r = 10, c = 7).$$

Logo

$$P(T_2 = x) = \frac{\binom{10}{x}\binom{5}{7-x}}{\binom{15}{7}} \; I_A \; (x),$$

com $A = \{2, 3, 4, 5, 6, 7\}$
pois $min((r, c) = min(10, 7) = 7$. e $max(0, c + r - N) = max(0, 7 + 10 - 15) = max(0, 2) = 2$.
Assim

$$nd = 2 \times min\Big(P(T_2 \leq 6), P(T_2 \geq 6)\Big).$$

Seja

$$p_a = P(T_2 \geq 6) = P(T_2 = 6) + P(T_2 = 7) = \frac{\binom{10}{6}\binom{5}{7-6}}{\binom{15}{7}} + \frac{\binom{10}{7}\binom{5}{7-7}}{\binom{15}{7}}$$

$$p_a = 0,1002331 < 0,5.$$

O nível descritivo é dado por:

$$nd = 2p_a = 0,2004662.$$

```
>
> F=c(3,4)      ### Fecundo
> NF=c(7,1)   ### Não Fecundo
>
> Raça=cbind(F,NF)
> row.names(Raça)=c("A","B")
> Raça
F NF
```

```
A 3  7
B 4  1
> fisher.test(Raça)

Fisher's Exact Test for Count Data

data:  Raça
p-value = 0.1189
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
0.00191838 1.98894847
sample estimates:
odds ratio
0.1269625

>
>
> ##T_2 ~HG(N,r,c)
>
> N=15;r=10;c=7
> p1= phyper(3,r,N-r,c);p1
[1] 0.1002331
> p1<0.5
[1] TRUE
> nd=2*p1;nd
[1] 0.2004662
> round(nd,4)
[1] 0.2005
>
> alfa=0.05
> nd>alfa
[1] TRUE
> ###Não rejeitar H_0. As duas raças parece ter o mesmo potencial de fecundidade.
>
```

Exemplo 4: Introdução à Estatística Médica:Francisco Soares e Arminda Siqueira s-pg 156.

Para a verificação se o fato de ter sido amamentado pela mãe é um fator de proteção par o câncer de mama. Freudenhelm et al.(1994) realizaram estudo do tipo caso-controle nas condados de Erie e Niagára situados na parte oeste do estado de Nova York(EUA). As pacientes tomadas como controle foram escolhidas na população da região, não havendo emparelhamento. Os dados obtidos foram:

Grupo-Amamentação	Sim	Não	Total
Casos Sim	353	175	502
Controles	449	153	602
Total	802	328	1130

Solução O risco de desenvolver Câncer de mama entre mulheres amamentadas pela mãe, aproximado pela razão de chances, é estimado por:

$$\hat{\Psi} = \frac{353 \times 153}{175 \times 449} = 0,69,$$

ou seja nominalmente, o risco do grupo amamentado é apenas 69 % do risco do grupo não amamentado. Para obter um intervalo de confiança para Ψ temos que calcular.

$$teta_{est} = \log((\hat{\Psi})) = log(0,69) = -0,37,$$

$$Var_{est}(teta_{est}) = \frac{1}{353} = \frac{1}{353} + \frac{1}{175} + \frac{1}{449} + \frac{1}{153} = 0,02.$$

vamos obter um IC de 95%

$$-0,37 \pm 1,96\ \sqrt{0,02} = [-0,647; -0,093].$$

Exemplo 5: Introdução à Estatística Médica:Francisco Soares e Arminda Siqueira -pg 186. Pasternak et al,(1992) avaliaram a eficácia e segurança de dois antibióticos no tratamento de pneumonia bacteriana de origem comunitária em adultos. Foram avaliados 63 pacientes, sendo 32 tratados com cefadroxil e 31 com cefalexina. A avaliação da resposta terapêutica foi baseada na evolução do quadro clínico e do exame radiológico do toráx feito na admissão ao estudo e no décimo dia de tratamento.

Dos pacientes avaliados, a cura completa ocorreu em 31 dos 32 (96,9 %) dos pacientes do grupo cefadroxil e em 28 dos 31 (90,3%) pacientes do grupo cefalexina. Foram observados efeitos adversos em quatro casos: um(3,1 %) no grupo que recebeu cefadroxil, com ocorrências de náuseas que desapareceram com o tratamento sintomático instituído. todos os 3 casos do cefalexina tiveram diarréia, dois se recuperaram espontaneamente e um teve que interromper o tratamento. Faça uma análise dos dados.

Solução Vamos montar uma tabela de contingência 2× com as variáveis antibiótico com dois níveis (cefadroxil e cefalexina) e cura com dois níveis(Sim, Não).

	Não	Sim	Total
Cefadroxil	1	31	32
Cefalexina	3	28	31
Total	4	59	63

Vamos usar o teste exato de Fisher.
Sejam
p_1-proporção de não curados usando o antibiótico Cefadroxil e
p_2-proporção de não curados usando o antibiótico Cefalexina.
As hipóteses a serem testadas:

$$H_0 : p_1 = p_2 \quad vs \quad p_1 \neq_2 .$$

```
>
> Cefadroxil=c(1,31)
> Cefalexina=c(3,28)
> Antibi=rbind(Cefadroxil,Cefalexina);Antibi
           [,1] [,2]
Cefadroxil    1   31
Cefalexina    3   28
> colnames(Antibi)=c("Não","Sim")
> Antibi
           Não Sim
Cefadroxil   1  31
Cefalexina   3  28
>
> fisher.test(Antibi)

Fisher's Exact Test for Count Data

data:  Antibi
p-value = 0.3547
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
0.005564844 4.069911965
sample estimates:
odds ratio
0.3064652

> chisq.test(Antibi)
```

```
Pearson's Chi-squared test with Yates' continuity correction

data:  Antibi
X-squared = 0.302, df = 1, p-value = 0.5826

Warning message:
In chisq.test(Antibi) : Aproximação do qui-quadrado pode estar incorreta
>
> chisq.test(Antibi, correct=FALSE)

Pearson's Chi-squared test

data:  Antibi
X-squared = 1.137, df = 1, p-value = 0.2863

Warning message:
In chisq.test(Antibi, correct = FALSE) :
Aproximação do qui-quadrado pode estar incorreta
>
```

Analise cuidadosamente a saída do R. Foram aplicados 3 testes. O exato de Fisher e o qui-quadrado com ou sem correção de continuidade.

Exemplo 6: Introdução à Estatística Médica:Francisco Soares e Arminda Siqueira -pg 176.

Fischl et al(1987) publicaram o primeiro relato de um ensaio clínico que comprovou a eficácia de zidovudina(AZT) para prolongar a vida dos pacientes com AIDS. Os dados centrais do trabalho estão na tabela a seguir. Temos dois grupos o que tomou AZT e o Placebo e o desfecho(Morto,Vivo).

	Vivo	Morto	Total
AZT	144	1	145
Placebo	121	16	137
Total	265	17	282

Oa autores afirmam que este experimento foi cercado de muitos cuidados, embora a análise estatística do caso seja fundamental, a decisão final de liberação do AZT foi tomada levando-se em consideração muitos outros resultados pelos estudo, como aqueles referentes a efeitos colaterais. A análise de dados da tabela consiste basicamente na comparação de duas proporções:

p_1= proporção de curados no grupo que tomou AZT

e

p_2= proporção de curados no grupo Placebo.

A proporção estimada dos que estavam vivos depois de 24 semanas de tratamento foi de;

$$\hat{p}_1 = \frac{144}{145} = 0,993,$$

enquanto que para o grupo Placebo foi de:

$$\hat{p}_1 = \frac{121}{137} = 0,883.$$

Como alocação dos pacientes aos grupos foi feita de forma aleatória, a diferença entre essas duas proporções parece indicar que em pacientes com AIDS o AZT tem o efeito de prolongar a vida. Antes de aceitar esta conclusão, entretanto, é preciso afastar o acaso como explicação alternativa. Ou seja,deve-se responder à pergunta: será que este resultado ocorreu por mero acaso ou por ser o AZT de fato uma droga efetiva?

Sejam p_1 a proporção de vivos no grupo que tomou AZT e p_2 a proporção de vivos no grupo Placebo. Queremos testar:

$$H_0 : p_1 = p_2 = p$$

versus

$$H_1 : p_1 \neq p_2 = p,$$

Temos duas variáveis aleatórias independentes:

X_1= número de vivos no grupo que tomou AZT e

X_2= número de vivos no grupo Placebo.

$$X_1 \sim Bin(n_1 = 145, p_1)$$

e

$$X_2 \sim Bin(n_2 = 137, p_2).$$

Se H_0 é verdade então $p_1 = p_2 = p$

$$S = X_1 + X_2 \sim Bin(n_1 + n_2 = 282, p).$$

O parâmetro p precisa ser estimado. Vamos achar o estimador pelo método dos momentos:

$$E(S) = (m + n)p$$

e

$$E\Big(\frac{S}{n_1+n_2}\Big) = p,$$

e um estimador não viciado para p é dado por:

$$\hat{p} = \frac{S}{n_1+n_2} = \frac{X_1+X_2}{n_1+n_2}.$$

Vamos estimar p:

$$p_e st = \frac{x_1+x_2}{n_1+n_2} = \frac{144+121}{145+137} = \frac{265}{282} = 0,94.$$

```
>
> n_1=145;n_2=137
> num=144+121;num
[1] 265
> den=145+ 137;den
[1] 282
>
> p_est=num/den;p_est
[1] 0.9397163
> round(p_est,3)
[1] 0.94
>
>
> ###Se não houver diferenças o número  esperado de vivos que tomaram AZT é:
>
> E_11=n_1*p_est;E_11;round(E_11,2)
[1] 136.2589
[1] 136.26
>
> ###Se não houver diferenças o número  esperado de mortos que tomaram AZT é:
>
> E_12=n_1*(1-p_est);E_12;round(E_12,2)
[1] 8.741135
[1] 8.74
>
> ##Note que:
>
> E_11+ E_12;n_1
[1] 145
[1] 145
>
>
> ###Se não houver diferenças o número  esperado de vivos no grupo Placebo é:
>
> E_21=n_2*p_est;E_21;round(E_21,2)
[1] 128.7411
[1] 128.74
>
> ###Se não houver diferenças o número  esperado de mortos que tomaram AZT é:
>
> E_22=n_2*(1-p_est);E_22;round(E_22,2)
[1] 8.258865
[1] 8.26
>
> ##Note que:
>
> E_21+ E_22;n_2
```

```
[1] 137
[1] 137
>
> ###Os cálculos necessarios para fazer o teste de qui-quadrado
> i=1:4;i
[1] 1 2 3 4
> O_i= c(144,121,1,16); E_i=c(136.26,128.74,8.74,8.26)
>
> D_i=O_i-E_i
>
> D2_i=D_i^2
> aux_i=
+
+
+ D2_i/E_i
>
> tab=cbind(i,O_i,E_i,D_i,D2_i,aux_i);round(tab,2)
     i O_i    E_i   D_i  D2_i aux_i
[1,] 1 144 136.26  7.74 59.91  0.44
[2,] 2 121 128.74 -7.74 59.91  0.47
[3,] 3   1   8.74 -7.74 59.91  6.85
[4,] 4  16   8.26  7.74 59.91  7.25
>
> S=apply(tab,2,sum);S
 i       O_i       E_i       D_i      D2_i     aux_i
 Total  282.00000 282.00000   0.00000 239.63040  15.01215
>
>
>
> tab1=rbind(tab,S);tab1
i O_i    E_i   D_i     D2_i      aux_i
1 144 136.26  7.74  59.9076  0.4396565
2 121 128.74 -7.74  59.9076  0.4653379
3   1   8.74 -7.74  59.9076  6.8544165
4  16   8.26  7.74  59.9076  7.2527361
Total  282.00  0.00 239.6304 15.0121470
>
> X2_cal=15.01
> alfa=0.05
>
> X2_tab=qchisq(1-alfa,1);X2_tab
[1] 3.841459
>
> X2_cal > X2_tab
[1] TRUE
>
>
> ##calcular o nível descritivo.
>
> nd=1-pchisq(X2_cal,1);nd
[1] 0.000106943
>
>
>
>
>
>
> ###Vamos fazer o teste exato de Fisher
```

```
>
> vivo=c(144,121)
> morto=c(1,16)
>
>
> Situação=cbind(vivo,morto);Situação
     vivo morto
[1,]  144     1
[2,]  121    16
> rownames(Situação)=c("AZT","Placebo")
> Situação
       vivo morto
AZT      144     1
Placebo  121    16
>
>
> W=fisher.test(Situação)
> W

Fisher's Exact Test for Count Data

data:  Situação
p-value = 6.564e-05
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
2.856659 800.904923
sample estimates:
odds ratio
18.91139

> names(W)
[1] "p.value"     "conf.int"    "estimate"    "null.value"  "alternative"
[6] "method"      "data.name"
> nd1=W$p.value;nd1
[1] 6.563822e-05
> round(nd1,4)
[1] 1e-04
>
> ###Vamos fazer o teste de qui-quadrado direto do R.
>
>
> ###Com Correção de Yates.
> QW=chisq.test(Situação);QW

Pearson's Chi-squared test with Yates' continuity correction

data:  Situação
X-squared = 13.139, df = 1, p-value = 0.0002891

>
>
> ###Sem Correção de Yates.
> QWS=chisq.test(Situação,correct=FALSE);QWS

Pearson's Chi-squared test

data:  Situação
X-squared = 15.017, df = 1, p-value = 0.0001066
```

```
>
>
```

Baseado neste estudo podemos dizer com grande certeza que O AZT tem efeito de prolongar a vida de um paciente, com AIDS, primeira evidência necessária para liberação do medicamento.

5:Exercícios

1. Numa pesquisa sobre desquites, realizada entre as classes alta e média, foram obtidos os seguintes resultados para a variável: Número de divórcios.

	Amigáveis	não Amigáveis	Total
Classe Alta	6	4	10
Classe Média	2	8	10
Total	8	12	20

É admissível concluir que a proporção de divórcios amigáveis é maior na classe alta?

Comprove que se a tabela fosse reorganizada de todas as maneiras a conclusão não

1. Foi retirada uma amostra de 793 indivíduos envolvidos em acidentes ciclísticos durante um período de um ano na cidade de Mombaça. Destes 793 acidentes, 147 usavam capacetes de segurança no momento do acidente e 646 não. Entre os que usavam, 17 sofreram lesões na cabeça, exigindo a atenção de um médico. Entre os que não usavam 218 sofreram sérias lesões.

   Queremos testar a hipótese H_0 : A proporção de pessoas que sofre lesões na cabeça na população de indivíduos usuários desses capacetes no momento do acidente é igual à proporção de pessoas que sofre lesões na cabeça entre os não usuários.

   a. Monte uma tabela 2×2 usando as variáveis Lesão na Cabeça e Uso do Capacete com níveis do tipo Não, Sim.

   b. Teste a hipótese acima usando um teste aproximado bilateral com $\alpha = 5\%$ usando a distribuição normal. Caracterize bem as variáveis. Enuncie as suposições. Calcule o nível descritivo.

   c. Há associação entre as variáveis? Use um teste sem (com) correção usando a distribuição de qui-quadrado. Faça uma análise completa.

2. (Higgins-página 174) Um cirurgião está interessado em saber se a admissão de uma nova droga pode reduzir a incidência de embolia pulmonar(Sim ou Não) em pacientes de alto risco operatório. 19 pacientes foram selecionados para o estudo com 11 recebendo a droga e 8 o tratamento padrão. Os dados estão na tabela a seguir:

	Sim	Não Amigáveis	Total
Droga	3	8	11
Não Droga	4	4	8
Total	7	12	19

   Existe evidência suficiente para concluir que a droga é eficaz na redução da incidência de embolia pulmonar?

3. (Higgins-Exercício 7- página 190) Em um estudo de contaminação de poços agrícolas, a água contaminada foi classificada como baixa ou alta e a distância do poço de uma fonte potencial de contaminação orgânica foi classificada como próxima ou não próxima. Os resultados são mostrados na tabela. Teste se há associação significativa entre contaminação e distância.

	Próxima	Não Próxima	Total
Alta	4	3	7
Baixa	9	0	9
Total	13	3	16

4. Em um estudo sobre associação entre o uso corrente de contraceptivos e o infarto do miocárdio, Shapiro et al (1979) observaram os resultados entre pacientes com idade entre 30 e 34 anos mostrados na tabela a seguir: Temos dois grupos(Casos e Controles) e uso recente do contraceptivo (Sim, Não).

	Sim	Não	Total
Casos	9	12	21
Controles	33	390	423
Total	42	402	444

Existe associação entre o uso corrente de contraceptivos e o infarto do miocárdio para pacientes entre 30 e 34 anos?